JMIR Human Factors

Journal Impact Factorâ ¢ (JIF) (2023): 5.8
Volume 11 (2024) ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Original Papers

Reviews

Short Papers

Viewpoints

Corrigenda and Addenda

Evaluating the Construct Validity of the Charité Alarm Fatigue Questionnaire using Confirmatory Factor Analysis

Maximilian Markus Wunderlich[1], MSc; Henning Krampe[2], PhD; Kristina Fuest[3], Dr med; Dominik Leicht[4], Dr med; Moriz Benedikt Probst[5], Dr med; Julian Runge[6], Dr med; Sebastian Schmid[7], Dr med; Claudia Spies[2], Prof Dr med; Björn Weiß[2], Prof Dr med; Felix Balzer[1*], Prof Dr med, Dr rer nat; Akira-Sebastian Poncette[1,2*], PD, Dr med; CAFQa Study Group Germany[8]

[1]Institute of Medical Informatics, Charité–Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Charitéplatz 1, Berlin, Germany

[2]Department of Anesthesiology and Intensive Care Medicine CVK/CCM, Charité–Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

[3]Department of Anaesthesiology & Intensive Care Medicine, Klinikum Rechts der Isar, School of Medicine, Technical University Munich, Munich, Germany

[4]Department for Anaesthesiology, Intensive Care Medicine and Pain Therapy, Justus Liebig University, Giessen, Germany

[5]Department for Anaesthesiology, Intensive Care Medicine and Pain Therapy, Vivantes Klinikum im Friedrichshain, Berlin, Germany

[6]Department for Anesthesiology, Surgical Intensive Care, Pain and Palliative Medicine, Marien Hospital Herne–Universitätsklinikum der Ruhr-Universität Bochum, Herne, Germany

[7]Department of Anesthesiology and Intensive Care Medicine, University Hospital Ulm, Ulm University, Ulm, Germany

[8]

[*]these authors contributed equally

Corresponding Author:
Akira-Sebastian Poncette, PD, Dr med
Institute of Medical Informatics, Charité–Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, , Charitéplatz 1, Berlin, , Germany

Abstract

Background: The Charité Alarm Fatigue Questionnaire (CAFQa) is a 9-item questionnaire that aims to standardize how alarm fatigue in nurses and physicians is measured. We previously hypothesized that it has 2 correlated scales, one on the psychosomatic effects of alarm fatigue and the other on staff's coping strategies in working with alarms.

Objective: We aimed to validate the hypothesized structure of the CAFQa and thus underpin the instrument's construct validity.

Methods: We conducted 2 independent studies with nurses and physicians from intensive care units in Germany (study 1: n=265; study 2: n=1212). Responses to the questionnaire were analyzed using confirmatory factor analysis with the unweighted least-squares algorithm based on polychoric covariances. Convergent validity was assessed by participants' estimation of their own alarm fatigue and exposure to false alarms as a percentage.

Results: In both studies, the χ^2 test reached statistical significance (study 1: χ^2_{26}=44.9; *P*=.01; study 2: χ^2_{26}=92.4; *P*<.001). Other fit indices suggested a good model fit (in both studies: root mean square error of approximation <0.05, standardized root mean squared residual <0.08, relative noncentrality index >0.95, Tucker-Lewis index >0.95, and comparative fit index >0.995). Participants' mean scores correlated moderately with self-reported alarm fatigue (study 1: *r*=0.45; study 2: *r*=0.53) and weakly with self-perceived exposure to false alarms (study 1: *r*=0.3; study 2: *r*=0.33).

Conclusions: The questionnaire measures the construct of alarm fatigue as proposed in our previous study. Researchers and clinicians can rely on the CAFQa to measure the alarm fatigue of nurses and physicians.

Trial Registration: ClinicalTrials.gov NCT04994600; https://www.clinicaltrials.gov/study/NCT04994600

(JMIR Hum Factors 2024;11:e57658) doi:10.2196/57658

KEYWORDS

patient monitoring; intensive care unit; alarm; alarms; validity; validation; safety; intensive; care; alarm fatigue; alarm management; patient safety; ICU; alarm system; alarm system quality; medical devices; clinical alarms; questionnaire; questionnaires; warning; factor analysis

Introduction

Background

Alarm fatigue is a phenomenon where health care workers in intensive care units (ICUs) become desensitized to alarms of medical devices [1]. It can make ICU staff feel stressed, and it is a substantial risk to patient safety, as it can lead to alarms being missed or acknowledged with delay [2]. When implementing interventions or IT solutions [3] that try to remedy alarm fatigue, clinicians and clinical alarm researchers need a reliable way to assess whether they were successful. However, they have not yet agreed on a standardized way of measuring alarm fatigue [4,5], even though it was recognized more than 2 decades ago [6].

Solely analyzing an ICU's alarm log data cannot serve as a measure of staff's alarm fatigue. While it is a valuable method for designing alarm management interventions [7], there is no clear association between the number of alarms on an ICU and staff's subjective alarm fatigue. For example, simply focusing on the number of alarms disregards their temporal distribution (eg, it might fatigue staff more if alarms came in random bursts than if they were evenly spaced out [8]). This could be one of the reasons why in their intervention study, Sowan et al [9] did not find that staff's alarm fatigue improved despite having managed to significantly reduce the number of alarms on their ICU.

Therefore, we recently developed the Charité Alarm Fatigue Questionnaire (CAFQa), which is a 9-item questionnaire that measures alarm fatigue in nurses and physicians [10]. Using exploratory factor analysis, we identified 2 correlated factors: one revolving around the psychophysiological effects of alarms (eg, headaches and feelings of distraction), and one revolving around ICU staff's alarm management strategies (eg, customization of alarm limits). We named the former the "alarm stress scale" and the latter the "alarm coping scale." The alarm coping scale consists of items that are reversely scored. Hence, a high score on either scale is indicative of alarm fatigue.

When developing a new questionnaire, it is essential to establish construct validity, that is, whether the questionnaire truly measures what it attempts to measure. One way to test an instrument's construct validity is to administer it to a different sample and test whether the originally proposed factor structure reemerges using confirmatory factor analysis [11,12] (for a recent example see Canivez et al [13]).

Aim

We aim to validate the exploratively derived factor structure of the CAFQa and thus underpin the instrument's construct validity.

Methods

Ethical Considerations

The ethical approval for this study was granted by the ethics committee of the Charité–Universitätsmedizin Berlin (EA4/218/20) and, if required, confirmed by the local ethics committee at the participating hospital. This study was conducted in compliance with the relevant guidelines and regulations. All participants voluntarily agreed to take part after being fully informed about the study. In study 1, as a reward for completing the questionnaire, we offered participants the chance to enter a draw where they could win a €50 (US $53) voucher for online shopping. Participants were asked to consent to have their data collected, analyzed, and stored anonymously.

Participants

In both studies, we included nurses, physicians, and nurses in training, while excluding other professions and non-ICU staff.

Study 1

We recruited participants from 9 ICUs of 5 large German hospitals. The questionnaire was administered on the web using REDCap (Research Electronic Data Capture; Vanderbilt University) between October 2021 and July 2022.

Study 2

Using a mailing list, we invited all members of the German Society of Anaesthesiology and Intensive Care Medicine [14] to fill out the web-based questionnaire (again using REDCap) between March 2023 and July 2023.

Questionnaire

The questionnaire used in both studies was identical and consisted of all 9 items from the CAFQa [10] and 5 general questions about the alarm situation in participants' ICUs. These general questions were not part of the analysis for this report. All 14 items were pseudorandomly arranged and required responses on a Likert scale ranging from −2 (indicating "I do not agree at all") to 2 (indicating "I very much agree"). Items with negative valences were reverse scored. Demographic items asked participants about their average number of workdays in an intensive care or monitoring area, their number of years and months of ICU experience, their workplace (campus and unit), and their profession. We made small adjustments to the original wording of 2 items (items 8 and 9) to improve readability: In item 8 we used "situation" instead of "urgency." In item 9 we used the phrase "clinical pictures" instead of "clinical symptoms."

Statistical Analysis

All analyses were conducted in R (version 4.2.1; R Foundation for Statistical Computing) using the following packages: *Tidyverse* [15], *reshape2* [16], *psych*, *semPlot* [17], and *lavaan* [18]. For study 1, we pooled the data from the participating hospitals.

Missing Data

In accordance with Heymans and Eekhout [19], we used predictive mean matching via the *mice* package [20] to impute missing data that were assumed to be missing at random (MAR). We did not impute questionnaires that were either completely empty or terminated prematurely (presumably due to survey fatigue), as the assumption of MAR was not met in these cases. We assumed that survey fatigue occurred if a participant failed to respond to at least the final 20% of the questionnaire (ie, the last 3 or more of the 9 items of the CAFQa plus the 5 general questions). In total, 0.3% of the data were MAR.

Testing Assumptions of Confirmatory Factor Analysis

In both studies, the results of the Mardia test indicated that the multivariate skew did not come from a normal distribution with *P*<.001. Outliers were identified using Mahalanobis distances, with none being detected in study 1 and 4 being detected in study 2 (for both studies: χ^2_9 cutoff=27.9; *P*<.001). Visual inspection of the data from all 4 cases revealed no unusual response patterns. Given the large sample size, we decided not to remove any outliers. The Kaiser-Meyer-Olkin statistic [21] in study 1 was 0.76, and in study 2 it was 0.8. In both studies, the Bartlett test of sphericity [22] rejected the null hypothesis that the correlation matrix was an identity matrix (study 1: χ^2_{36}=438.3; *P*<.001; study 2: $\chi^2_{36=}$2495.4). There was no evidence of multicollinearity in either study as the determinant of both *R* matrices was greater than 0.00001 [23] and no correlations were greater than |0.7|. Overall, these results suggest that the data of both studies were suitable for factor analysis.

Confirmatory Factor Analysis

For both studies, we specified the model in line with our previous findings [10], with 2 correlated latent factors, labeled "alarm stress" and "alarm coping." Items 1 - 5 were assigned to "alarm stress." Items 5 - 9 were assigned to "alarm coping." Since all CAFQa items are ordered categorical variables (due to being measured on a 5-point Likert scale) and because the Mardia tests indicated that the multivariate skew of both studies did not come from a normal distribution, we used the unweighted least-squares (ULS) algorithm based on polychoric covariances for estimating factor loadings [24-26]. We assessed the goodness-of-fit of the model using χ^2, and the following fit indices in line with the cutoff criteria defined by Hu and Bentler [27]: root mean square error of approximation (RMSEA), relative noncentrality index (RNI), Tucker-Lewis index (TLI), standardized root mean squared residual (SRMR), and comparative fit index (CFI).

Convergent Validity

At the end of the questionnaire in both studies, we provided participants with a brief description of alarm fatigue and asked them to estimate their personal alarm fatigue as a percentage (0% indicating no alarm fatigue and 100% indicating extreme alarm fatigue). We also asked participants to provide their perceived rate of false alarms in their ICU as a percentage (0% indicating no false alarms, 100% indicating no true alarms). To measure convergent validity, we correlated the participants' mean scores on the questionnaire with the self-provided alarm fatigue and false alarm rate estimations (in total and per factor).

Internal Consistency

As a measure of internal consistency, we report Cronbach coefficient α, the McDonald coefficient ω [28], and the mean interitem correlation for both factors.

Results

Participants

Study 1

We received 363 submissions. Among these, 23 came from participants who did not consent to have their data analyzed, 67 questionnaires were empty, and 8 showed signs of survey fatigue. Therefore, the sample size for this study was 265. The number of participants was roughly similar for each hospital (Giessen: n=43; Herne: n=50; Munich: n=64; Ulm: n=57; Vivantes: n=51). Most participants were nurses (n=150, 56.6%) and 35.8% (n=95) were physicians. A few participants (n=9, 3.4%) were supporting nurses, nurses in training, medical students, or interns, while 4.2% (n=11) did not state their profession.

Study 2

Of the 1564 submissions we received, 69 came from participants who refused to consent to have their data processed and 223 were empty questionnaires. We suspected survey fatigue in 60 cases. Hence, the sample size of study 2 was 1212. Contrary to study 1, more participants were physicians (n=1002, 82.7%) than nurses (n=186, 15.3%). Again, the group of supporting nurses, nurses in training, medical students, and interns was a minority (n=6, 0.5%). Among the participants, 1.5% (n=18) did not state their profession.

Confirmatory Factor Analysis

Descriptive statistics of both studies are presented in Table 1 for each item.

Table . Descriptive statistics for each item and the pattern coefficients found in the confirmatory factor analysis of the 2-factor model in both studies. All loadings were statistically significant at *P*<.001.

Item	Description	Study 1					Study 2				
		Factor 1 (95% CI)	Factor 2 (95% CI)	Mean (SD)	Kurtosis	Skew	Factor 1 (95% CI)	Factor 2 (95% CI)	Mean (SD)	Kurtosis	Skew
1	With too many alarms on my ward, my work performance, and motivation decrease.	0.730 (0.643 - 0.818)	—[a]	0.47 (1.01)	–0.62	–0.3	0.677 (0.636 - 0.717)	—	0.51 (1.08)	–0.63	–0.37
2	Too many alarms trigger physical symptoms for me, e.g., nervousness, headaches, and sleep disturbances.	0.706 (0.612 - 0.800)	—	0.23 (1.26)	–1.13	–0.16	0.694 (0.653 - 0.735)	—	0.22 (1.20)	–1.00	–0.14
3	Alarms reduce my concentration and attention.	0.725 (0.635 - 0.814)	—	0.43 (1.07)	–0.91	–0.21	0.813 (0.779 - 0.846)	—	0.63 (1.03)	–0.59	–0.39
4	My or neighboring patients' alarms or crisis alarms frequently interrupt my workflow.	0.432 (0.318 - 0.547)	—	0.87 (0.83)	–0.41	–0.39	0.519 (0.469 - 0.570)	—	0.70 (0.87)	–0.28	–0.37
5	There are situations when alarms confuse me.	0.488 (0.384 - 0.593)	—	0.08 (1.09)	–0.73	–0.09	0.634 (0.592 - 0.676)	—	0.19 (1.07)	–0.85	–0.04
6	In my ward, procedural instruction on how to deal with alarms is regularly updated and shared with all staff.[b]	—	0.434 (0.270 - 0.598)	0.77 (1.24)	–0.68	–0.7	—	0.449 (0.375 - 0.523	1.10 (1.02)	0.44	–1.06

Item	Description	Study 1					Study 2				
		Factor 1 (95% CI)	Factor 2 (95% CI)	Mean (SD)	Kurtosis	Skew	Factor 1 (95% CI)	Factor 2 (95% CI)	Mean (SD)	Kurtosis	Skew
7	Responsible personnel respond quickly and appropriately to alarms.[b]	—	0.587 (0.424 - 0.750)	-0.32 (0.82)	–0.12	–0.12	—	0.639 (0.567 - 0.711	–0.39 (0.90)	–0.05	0.19
8	The acoustic and visual monitor alarms used on my ward floor and in my nurses' station allow me to assign the patient, the device, and the situation clearly.[b]	—	0.349 (0.182 - 0.517)	-0.46 (1.06)	–0.52	0.36	—	0.428 (0.359 - 0.498	–0.36 (1.15)	–0.66	0.32
9	Alarm limits are regularly adjusted based on patients' clinical pictures (e.g., blood pressure limits for conditions after bypass surgery).[b]	—	0.581 (0.428 - 0.734)	-0.38 (0.93)	–0.24	0.26	—	0.575 (0.508 - 0.641	–0.35 (0.94)	–0.12	0.40

[a]Not applicable.

[b]Item with a negative valence that is reversely scored.

Study 1

While the χ^2 test was significant at α=.05 with χ^2_{26}=44.9 (*P*=.01), indicating that the model did not fit the data, all fit indices suggested a good model fit: RMSEA=0.03, SRMR=0.052, RNI=0.989, TLI=0.985, and CFI=0.989. All items loaded onto their hypothesized factors as expected, with factor loadings that were statistically significant at *P*<.001, ranging from 0.35 to 0.73. The factors were moderately correlated at 0.4 (95% CI 0.21 - 0.59; *P*<.001).

Study 2

As in study 1, the χ^2 test was significant (χ^2_{26}=92.4; *P*<.001), indicating that the model did not fit the data, while the fit indices showed a good model fit: RMSEA=0.046, SRMR=0.041, RNI=0.982, TLI=0.975, and CFI=0.982. Again, all items loaded onto their hypothesized factors as expected, with factor loadings that were statistically significant at *P*<.001, ranging from 0.43 to 0.81. The factors were moderately correlated at 0.44 (95% CI 0.36 - 0.51; *P*<.001) (Figure 1).

Figure 1. Diagram of the final model from each study. Factors are shown in circles (factor 1: alarm stress; factor 2: alarm coping), and items 1 - 9 in squares. In both studies, the variance of the factors was constrained to 1. The arrows connecting the factors with the items are the factor loadings and arrows pointing toward the items show the residuals. The arrows between the factors show their correlation.

Convergent Validity

In study 1, the participants' mean scores on the questionnaire correlated moderately with self-reported alarm fatigue (r_{242}=0.45, 95% CI 0.34-0.54; P<.001) and weakly with the perceived percentage of false alarms (r_{247}=0.3, 95% CI 0.18 to –0.41; P<.001). Similar patterns were observed in study 2 (Table 2 provides full details).

Table . Correlation coefficients, *P* values, and CIs used to investigate the convergent validity in each study.

Study and correlation measure		*r* (*df*; 95% CI)	*P* value
Study 1			
	MS-SRAF[a,b]	0.45 (242; 0.34 - 0.54)	<.001
	F1S-SRAF[c]	0.42 (242; 0.31 - 0.52)	<.001
	F2S-SRAF[d]	0.29 (242; 0.17 - 0.4)	<.001
	MS-PPFA[e]	0.30 (247; 0.18 - 0.41)	<.001
	F1S-PPFA	0.20 (247; 0.08 - 0.32)	.002
	F2S-PPFA	0.29 (247; 0.17 - 0.4)	<.001
Study 2			
	MS-SRAF	0.53 (1180; 0.49 - 0.57)	<.001
	F1S-SRAF	0.49 (1180; 0.45 - 0.53)	<.001
	F2S-SRAF	0.34 (1180; 0.29 - 0.39)	<.001
	MS-PPFA	0.33 (1182; 0.28 - 0.38)	<.001
	F1S-PPFA	0.26 (1182; 0.21 - 0.32)	<.001
	F2S-PPFA	0.28 (1182; 0.23 - 0.33)	<.001

[a]MS: mean score on the questionnaire.
[b]SRAF: self-reported alarm fatigue.
[c]F1S: scores on factor 1.
[d]F2S: scores on factor 2.
[e]PPFA: perceived percentage of false alarms.

Internal Consistency

In study 1, the Cronbach α of factor 1 was 0.72, and it was 0.49 for factor 2. Cronbach α across factors was 0.67. The mean interitem correlation on factor 1 was 0.38, and it was 0.23 on factor 2. The McDonald coefficient ω for factor 1 was 0.77, and for factor 2 it was 0.55. The overall coefficient ω for the assessment was 0.8.

Results were similar in study 2: the Cronbach α was 0.77 for factor 1 and 0.55 for factor 2. Cronbach α across factors was 0.72. The mean interitem correlation was 0.44 on factor 1 and 0.27 on factor 2. The McDonald coefficient ω was 0.8 for factor 1 and 0.59 for factor 2. The overall coefficient ω for the assessment was 0.85.

Discussion

We aimed to underpin the construct validity of the CAFQa by submitting the exploratively derived factor structure from our previous study to confirmatory factor analysis in 2 independent studies. While the χ^2 test rejected the model in both studies, all fit indices indicated a good model fit. The factor loadings ranged from 0.35 to 0.73 in study 1 and from 0.43 to 0.81 in study 2; all were statistically significant. Overall, these results support the hypothesized factor structure. The questionnaire seems to measure the construct of alarm fatigue as proposed in our previous work [10].

The χ^2 test is known to be sensitive to large sample sizes [29], which might explain its statistical significance. We did not modify the model because all fit indices indicated a good fit and because model modifications, no matter how small or plausible, can make a model less generalizable.

In both studies, the first factor, that is, the alarm stress scale, and the overall questionnaire were internally consistent. However, the second factor, that is, the alarm coping scale, seems to have issues with its internal consistency. Here, Cronbach α and McDonald ω were 0.49 and 0.55 in study 1, respectively, and 0.55 and 0.59 in study 2, respectively. A similar pattern can be found in our previous study, where the Cronbach α of factor 2 was 0.57 [10]. An internally consistent questionnaire is desirable. However, it can also mean that items are very similar. It was our ambition to create a questionnaire that is brief while measuring the many facets of alarm fatigue. Future studies using the CAFQa should routinely assess the internal consistency of both factors. If the second factor continues to show medium internal consistency, research should be done on how it can be improved (eg, by adding additional items, which would come at the cost of a longer questionnaire).

Because no other measures of alarm fatigue exist that would allow us to estimate the CAFQa's convergent validity, we asked participants to rate their own alarm fatigue as well as the rate of false alarms they perceived in their daily work. Participants who had a high mean score on the questionnaire also rated themselves as more alarm fatigued (r=0.45 in study 1 and r=53 in study 2). This positive correlation indicates the convergent validity of the questionnaire. Similarly, both studies demonstrated that participants with a high mean score on the questionnaire perceived more alarms to be false in their ICU. In study 1 this association was stronger for factor 2 than for factor 1. This makes sense since a high score on factor 2 (ie, the alarm coping scale) indicates that alarms are not properly managed (eg, by means of patient-specific threshold customizations), which typically leads to more false alarms [30]. However, study 2 could not replicate this pattern. Future research should find an answer to this question: Do ICU staff with a high perceived percentage of false alarms tend to develop stronger alarm fatigue, or do staff that are more alarm fatigued tend to perceive more alarms as being false?

Limitations

The fit indices RMSEA, CFI, and TLI have been shown to overestimate model fit when using the ULS estimator [31,32], potentially leading researchers to accept a bad-fitting model. Yet, in our case, other fit indices indicated a good model fit. As in our previous work [10], the assumption that participants can accurately reflect and express their own alarm fatigue as a percentage is likely flawed (otherwise, it would not be necessary to develop a questionnaire in the first place). However, most ICU nurses and physicians have heard of alarm fatigue, and we provided them with a brief recapitulation on the concept before having them answer the self-report item in each study. Likewise, it is also probably a flawed assumption that participants can accurately report the rate of false alarms, though Bliss et al [33] showed that participants were able to adapt their response frequencies to alarms based on to the perceived probability that an alarm was not false, thus suggesting that people might have an intuitive grasp of the rate of false alarms in their unit. All in all, we believe that asking these self-rating questions is a valuable method for assessing convergent validity when no other instrument is available.

Conclusion

Our results from 2 independent studies underpin the construct validity of the CAFQa. All items consistently loaded onto the factors, as we proposed in a previous publication [10]. When conducting research or quality improvement projects in ICUs, clinical alarm researchers and clinicians can rely on this instrument to measure, compare, and benchmark the alarm fatigue of nurses and physicians.

Acknowledgments

ASP is a participant in the Charité Digital Clinician Scientist Program funded by the Charité–Universitätsmedizin Berlin and the Berlin Institute of Health. We are grateful to Nicolas Frey for supervising the logistical aspects of the survey distribution, to Sandro Amende-Wolf for his valuable clinical insights, and to Daniel Schulze for his critical examination of our statistical analysis. We are grateful to the German Society of Anaesthesiology and Intensive Care Medicine for the distribution of the questionnaire to its members. This article was extracted from the doctoral thesis of MMW.

Data Availability

The data sets generated and analyzed during the current study are available in the Zenodo repository [34].

Authors' Contributions

MMW, CS, FB, and ASP conceptualized the study. MMW, KF, DL, MBP, JR, and SS conducted the investigation. MMW and HK developed the study design and ensured methodological rigor. MMW, BW, and ASP contributed to the formal data analysis. MMW and ASP drafted the initial manuscript and contributed to its revision. MMW created all figures and tables and managed the project together with ASP. HK, KF, DL, MBP, JR, SS, CS, BW, FB, and ASP reviewed the manuscript. FB and ASP provided oversight and leadership throughout the study.

Ulrich Frey (Department for Anesthesiology, Surgical Intensive Care, Pain and Palliative Medicine, Marien Hospital Herne–Universitätsklinikum der Ruhr-Universität Bochum, Herne, Germany), Christian von Heymann (Department for Anaesthesiology, Intensive Care Medicine and Pain Therapy, Vivantes Klinikum im Friedrichshain, Berlin, Germany), Jacqueline Holan (Department for Anaesthesiology, Intensive Care Medicine and Pain Therapy, Vivantes Klinikum im Friedrichshain, Berlin, Germany), Bettina Jungwirth (Department of Anesthesiology and Intensive Care Medicine, University Hospital Ulm, Ulm University, Ulm, Germany), Dragutin Popovic (Department for Anaesthesiology, Intensive Care Medicine and Pain Therapy, Vivantes Klinikum im Friedrichshain, Berlin, Germany), and Michael Sander (Department for Anaesthesiology, Intensive Care Medicine and Pain Therapy, Justus Liebig University, Giessen, Germany).

Conflicts of Interest

None declared.

References

1. Sendelbach S, Funk M. Alarm fatigue: a patient safety concern. AACN Adv Crit Care 2013;24(4):378-386. [doi: 10.1097/NCI.0b013e3182a903f9] [Medline: 24153215]
2. Clinical alarm management: compendium. AAMI Foundation. 2015. URL: https://www.aami.org/docs/default-source/foundation/alarms/alarm-compendium-2015.pdf?sfvrsn=2d2b53bd_2 [accessed 2024-04-08]
3. Chromik J, Klopfenstein SAI, Pfitzner B, et al. Computational approaches to alleviate alarm fatigue in intensive care medicine: a systematic literature review. Front Digit Health 2022;4:843747. [doi: 10.3389/fdgth.2022.843747] [Medline: 36052315]
4. Hüske-Kraus D, Wilken M, Röhrig R. Measuring alarm system quality in intensive care units. Presented at: Clusterkonferenz 2018; Apr 16 to 17, 2018; Freiburg, Germany URL: https://www.pflegeinnovationszentrum.de/wp-content/uploads/2018/12/17.-Measuring-Alarm-System-Quality-in-ICUs.pdf [accessed 2024-06-27]
5. Lewandowska K, Weisbrot M, Cieloszyk A, Mędrzycka-Dąbrowska W, Krupa S, Ozga D. Impact of alarm fatigue on the work of nurses in an intensive care environment-a systematic review. Int J Environ Res Public Health 2020 Nov 13;17(22):8409. [doi: 10.3390/ijerph17228409] [Medline: 33202907]
6. Wears RL, Perry SJ. Human factors and ergonomics in the emergency department. Ann Emerg Med 2002 Aug;40(2):206-212. [doi: 10.1067/mem.2002.124900] [Medline: 12140500]
7. Poncette AS, Wunderlich MM, Spies C, et al. Patient monitoring alarms in an intensive care unit: observational study with do-it-yourself instructions. J Med Internet Res 2021 May 28;23(5):e26494. [doi: 10.2196/26494] [Medline: 34047701]
8. Wilken M, Hüske-Kraus D, Röhrig R. Alarm fatigue: using alarm data from a patient data monitoring system on an intensive care unit to improve the alarm management. Stud Health Technol Inform 2019 Sep 3;267:273-281. [doi: 10.3233/SHTI190838] [Medline: 31483282]
9. Sowan AK, Gomez TM, Tarriela AF, Reed CC, Paper BM. Changes in default alarm settings and standard in-service are insufficient to improve alarm fatigue in an intensive care unit: a pilot project. JMIR Hum Factors 2016 Jan 11;3(1):e1. [doi: 10.2196/humanfactors.5098] [Medline: 27036170]
10. Wunderlich MM, Amende-Wolf S, Krampe H, et al. A brief questionnaire for measuring alarm fatigue in nurses and physicians in intensive care units. Sci Rep 2023 Aug 24;13(1):13860. [doi: 10.1038/s41598-023-40290-7] [Medline: 37620385]
11. DiStefano C, Hess B. Using confirmatory factor analysis for construct validation: an empirical review. J Psychoeduc Assess 2005 Sep;23(3):225-241. [doi: 10.1177/073428290502300303]
12. Bryant FB, Yarnold PR, Michelson EA. Statistical methodology: VIII. using confirmatory factor analysis (CFA) in emergency medicine research. Acad Emerg Med 1999 Jan;6(1):54-66. [doi: 10.1111/j.1553-2712.1999.tb00096.x] [Medline: 9928978]
13. Canivez GL, McGill RJ, Dombrowski SC, Watkins MW, Pritchard AE, Jacobson LA. Construct validity of the WISC-V in clinical cases: exploratory and confirmatory factor analyses of the 10 primary subtests. Assessment 2020 Mar;27(2):274-296. [doi: 10.1177/1073191118811609] [Medline: 30516059]
14. Deutsche Gesellschaft für Anästhesiologie & Intensivmedizin. URL: https://www.dgai.de/ [accessed 2023-08-28]
15. Wickham H, Averick M, Bryan J, et al. Welcome to the Tidyverse. J Open Source Softw 2019;4(43):1686. [doi: 10.21105/joss.01686]
16. Wickham H. Reshaping data with the reshape package. J Stat Soft 2007;21(12):1-20. [doi: 10.18637/jss.v021.i12]
17. Epskamp S. semPlot: unified visualizations of structural equation models. Struct Equ Model 2015 Jul 3;22(3):474-483. [doi: 10.1080/10705511.2014.937847]
18. Rosseel Y. Lavaan: an R package for structural equation modeling. J Stat Softw 2012;48:1-36. [doi: 10.18637/jss.v048.i02]
19. Heymans M, Eekhout I. Chapter 8: missing data in questionnaires. In: Applied Missing Data Analysis With SPSS and (R) Studio: Bookdown; 2019. URL: https://bookdown.org/mwheymans/bookmi/missing-data-in-questionnaires.html [accessed 2024-06-27]
20. Van Buuren S, Groothuis-Oudshoorn GOK. Mice: multivariate imputation by chained equations in R. J Stat Softw 2011;45(1):1-67. [doi: 10.18637/jss.v045.i03]
21. Kaiser HF. An index of factorial simplicity. Psychometrika 1974 Mar;39(1):31-36. [doi: 10.1007/BF02291575]
22. Watkins MW. Exploratory factor analysis: a guide to best practice. J Black Psychol 2018 Apr;44(3):219-246. [doi: 10.1177/0095798418771807]
23. Field A, Miles J, Field Z. Discovering Statistics Using R: SAGE Publications; 2012.
24. Flora DB, Labrish C, Chalmers RP. Old and new ideas for data screening and assumption testing for exploratory and confirmatory factor analysis. Front Psychol 2012;3:55. [doi: 10.3389/fpsyg.2012.00055] [Medline: 22403561]
25. Forero CG, Maydeu-Olivares A, Gallardo-Pujol D. Factor analysis with ordinal indicators: a Monte Carlo study comparing DWLS and ULS estimation. Struct Equ Model 2009 Oct 6;16(4):625-641. [doi: 10.1080/10705510903203573]
26. Koğar H, Yilmaz Koğar E. Comparison of different estimation methods for categorical and ordinal data in confirmatory factor analysis. Egit Psikol Olcme Deger Derg 2015 Dec;6(2). [doi: 10.21031/epod.94857]
27. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]

28. Kalkbrenner MT. Alpha, omega, and H internal consistency reliability estimates: reviewing these options and when to use them. Couns Outcome Res Eval 2023 Jan 2;14(1):77-88. [doi: 10.1080/21501378.2021.1940118]
29. Alavi M, Visentin DC, Thapa DK, Hunt GE, Watson R, Cleary M. Chi-square for model fit in confirmatory factor analysis. J Adv Nurs 2020 Sep;76(9):2209-2211. [doi: 10.1111/jan.14399] [Medline: 32323338]
30. Winters BD, Cvach MM, Bonafide CP, et al. Technological distractions (part 2): a summary of approaches to manage clinical alarms with intent to reduce alarm fatigue. Crit Care Med 2018 Jan;46(1):130-137. [doi: 10.1097/CCM.0000000000002803] [Medline: 29112077]
31. Savalei V. Improving fit indices in structural equation modeling with categorical data. Multivariate Behav Res 2021;56(3):390-407. [doi: 10.1080/00273171.2020.1717922] [Medline: 32054327]
32. Xia Y, Yang Y. RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: the story they tell depends on the estimation methods. Behav Res Methods 2019 Feb;51(1):409-428. [doi: 10.3758/s13428-018-1055-2] [Medline: 29869222]
33. Bliss JP, Gilson RD, Deaton JE. Human probability matching behaviour in response to alarms of varying reliability. Ergonomics 1995 Nov;38(11):2300-2312. [doi: 10.1080/00140139508925269] [Medline: 7498189]
34. Dataset for evaluating the construct validity of the Charité Alarm Fatigue Questionnaire. Zenodo. URL: https://zenodo.org/records/8296934 [accessed 2024-06-27]

Abbreviations

CAFQa: Charité Alarm Fatigue Questionnaire
CFA: confirmatory factor analysis
CFI: comparative fit index
EFA: exploratory factor analysis
ICU: intensive care unit
MAR: missing at random
REDCap: Research Electronic Data Capture
RMSEA: root mean square error of approximation
RNI: relative noncentrality index
SRAF: self-reported alarm fatigue
SRMR: standardized root mean squared residual
TLI: Tucker-Lewis index
ULS: unweighted least-squares

Edited by D Gooch; submitted 28.02.24; peer-reviewed by J Schlesinger, K Bin, K Takano; revised version received 02.05.24; accepted 02.05.24; published 08.08.24.

Please cite as:
Wunderlich MM, Krampe H, Fuest K, Leicht D, Probst MB, Runge J, Schmid S, Spies C, Weiß B, Balzer F, Poncette AS, CAFQa Study Group Germany
Evaluating the Construct Validity of the Charité Alarm Fatigue Questionnaire using Confirmatory Factor Analysis
JMIR Hum Factors 2024;11:e57658
URL: https://humanfactors.jmir.org/2024/1/e57658
doi:10.2196/57658

User-Centered Design for Designing and Evaluating a Prototype of a Data Collection Tool to Submit Information About Incidents of Violence Against Sex Workers: Multiple Methods Approach

Melissa H Ditmore, MS, PhD; Jose Fernando Florez-Arango, MS, MD, PhD

CUNY School of Public Health and Health Policy, 55 W 125th St, New York, NY, United States

Corresponding Author:
Melissa H Ditmore, MS, PhD
CUNY School of Public Health and Health Policy, , 55 W 125th St, New York, NY, , United States

Abstract

Background: Sex workers face an epidemic of violence in the United States. However, violence against sex workers in the United States is underreported. Sex workers hesitate to report it to the police because they are frequently punished themselves; therefore, an alternative for reporting is needed.

Objective: We aim to apply human-centered design methods to create and evaluate the usability of the prototype interface for ReportVASW (violence against sex worker, VASW) and identify opportunities for improvement.

Methods: This study explores ways to improve the prototype of ReportVASW, with particular attention to ways to improve the data collection tool. Evaluation methods included cognitive walkthrough, system usability scale, and heuristic evaluation.

Results: End users were enthusiastic about the idea of a website to document violence against sex workers. ReportVASW scored 90 on the system usability scale. The tool scored neutral on consistency, and all other responses were positive toward the app, with most being strong.

Conclusions: Many opportunities to improve the interface were identified. Multiple methods identified multiple issues to address. Most changes are not overly complex, and the majority were aesthetic or minor. Further development of the ReportVASW data collection tool is worth pursuing.

(JMIR Hum Factors 2024;11:e53557) doi:10.2196/53557

KEYWORDS

mobile health; sex worker; user-centered design methods; usability; heuristic analysis; cognitive walkthrough; aggression; abuse; occupational health; reporting; prototype; heuristics; human-centered design; implementation; barriers; enablers; data collection; digital health; underreporting

Introduction

Sex workers are frequently victimized by a variety of perpetrators [1] due to lack of empathy and because people who commit crimes against sex workers know that their crimes are unlikely to be reported. Sex workers are targets of violence, with 32% to 55% reporting violent experience in the previous year [2], but they often do not report violent incidents [2] in part because they encounter problems when they seek to report violence to police. In fact, police may arrest them for sex work instead of investigating violent crimes against sex workers; sex workers have also been victimized or harassed by police [3,4]. Those who seek care after victimization may not reveal their status as sex workers because health care providers may stigmatize and discriminate against them. These factors culminate in an unfortunate lack of knowledge in the domain. Documenting violence against sex workers is a necessary prerequisite to demonstrating the need to address interpersonal violence against sex workers, and will help develop more effective responses and raise funds for their implementation. Violence against sex workers is associated with sequelae including post traumatic stress disorder [5], HIV, and other sexually transmitted infections [6].

Opportunities to increase the likelihood of sex workers reporting violence include changing police practices such that they would accept reports and investigate violence, changing laws to decriminalize sex work or legalize sex work such that some sex work would be within the law. Legal support could help sex workers who want to report violence to police. There are at least two nonprofit organizations that offer legal services to sex workers in the United States. Sex workers share information about perpetrators of violence among themselves [7] separate from reporting to law enforcement, in order to share information about violent experiences and help others avoid their attackers; however, it is impossible to know how widely these mechanisms are used online, via text messages, or on paper. In contrast, an app was commissioned by the Asia Pacific Network of Sex

Workers and designed for sex workers in Myanmar to help them report violence committed against them to the police.

The app (iMonitor+, DureTechnology) was part of a larger program, and they used the app or a hotline that connected victims of violence to a service provider who accompanied them to the police to report violence and to health care providers [8]. Most sex workers in Myanmar have smartphones, but uptake of the app was not strong, in part because they preferred using an existing hotline to the app [8]. As the app was commissioned by sex workers who are part of a regional international network, it exhibited some aspects of user-centered design (UCD) because sex workers explained what they wanted to include; moreover, the context of use was well understood and the users specified the end requirements. The app was not used to share information about perpetrators of violence. Other apps have been developed for sex workers in Cambodia [9] and South Africa [10].

The lead author (MHD) was commissioned to evaluate the antiviolence program for sex workers in Myanmar. After conducting the evaluation and learning more about the Burmese app, MHD was inspired to try to develop something such as this for use in the United States. As a result, we developed a prototype of a data collection tool that we hope sex workers would feel confident using to report violent experiences. The prototype mobile health (mHealth) data collection tool *ReportVASW* is intended for sex workers in the United States to report violence committed against them. VASW stands for violence against sex worker, and the tool collects data reported by sex workers who have been victims of violence. The interface was designed to enable multiple options for reporting violence, including drop down menus, open text, and audio recording. ReportVASW can be pronounced Report Violence.

The victimization of sex workers has been a long-term focus, and we have published multiple reports and papers about violence against sex workers and documented human rights violations in multiple locations in Africa [11], Asia [12], and the United States [13,14].

Most systems are set up to be used repeatedly if not constantly; ReportVASW is different because as it addresses violent victimization of the end user, it is hoped that most people never need to use it, and that those who do will use it once or rarely. For this reason, we believe ReportVASW should be intuitive and easy to use, without a learning curve.

Methods

Design Process and Prototype

UCD was used to design this prototype because only active sex workers seeing clients face-to-face can describe their current methods for sharing information about violent people and what they do to try to avoid violence in their work. Considering that the user context is in the aftermath of a violent and possibly traumatic event, ease of use is paramount, and so developing an app that is easy to learn and quick to use is important. These are hallmarks of UCD.

The aim of UCD is to improve usability by maximizing effectiveness, efficiency, and satisfaction of end users in the specific aim of the product in question. Careful application of UCD methods at the earliest stages should reduce user error, and limit cost and time spent redesigning after developing software.

UCD follows specific principles, including focusing on users and tasks, measuring usability empirically, and iterative testing of design and usability. UCD has been used with success to develop mHealth apps and health record systems [15,16]. These principles were at the heart of the specific methods used for each of the 4 steps. The steps of the UCD approach align with more specific methods; for example, contextualization using functional analysis and consulting potential end users, and ideation through task analysis focused on the end user group and what steps would be required to successfully complete the task; prototyping using representational analysis of the tool; and finally, usability testing using scenario-based testing and heuristic analysis. This study's design process used the 4 methodological steps associated with UCD.

Step 1 - Contextualization

While violence against sex workers has been studied, few efforts to address this violence have been evaluated in the United States [17]. The authors seek to develop a new tool to respond to violence against sex workers. Contextualization was undertaken through desk research using functional analysis and consultation with active sex workers (user analysis) to ask whether such a tool would be useful; some were interested and offered opinions on the proposed tool. In total, 4 of 5 sex workers consulted were interested in the project at least in theory. UCD with multiple methods, such as considering both end user and design use context, has been used with success in mHealth [18,19] but it is not without difficulties [20,21]. Formative research using UCD can be time consuming, and users are not always easy to engage, but the literature shows the value of engaging end users in formative processes [20,21].

Step 2 - Ideation

The lead author thought through what this data collection tool should ask, how to collect information, and what information is most important. This process included task analysis, identifying the intended task and using a flowchart on paper to plan the way the tool would collect information.

Step 3 - Prototyping

The lead author engaged in representation analysis using the final flowchart to inform illustrations of what the screen would look like at each step. This was followed with the lead author attempting to apply Nielsen Heuristics to assess and refine the prototypes. Not all steps needed extensive revision. The initial paper prototype included 7 screens.

- The prototype was designed using the free version of Figma software online tool as of April 26, 2022, which allows the creation of 3 pages only. The heuristic analysis was based on Neilsen "10 usability heuristics." [22]
- Representational analysis consisted of heuristic analysis undertaken by the authors and 2 colleagues. The form used is a spreadsheet developed by the second author based on the work of Zhang and Wallji [23,24], in which a scale of

0 ‑ 4 is used to grade each issue, from minor (1) to catastrophic (4), and 0 used to indicate disagreement that the point is an issue. The spreadsheet also contains a column labeled “proposed solution.”

Step 4 - Usability Testing

Multiple methods were used to evaluate the ReportVASW interface. After the prototype was developed, usability testing using representation analysis was undertaken via heuristic analysis and task based and scenario testing. Heuristic analysis of the prototype was undertaken by the authors and 2 colleagues.

Additionally, the lead author and a developer recruited a convenience sample of 5 end users who self-identify as female with experience in a variety of sex work venues (escorts, brothel workers, sadomasochism professionals, and strippers) to test the usability by entering data from scenarios provided (scenario based testing) in a cognitive walkthrough [25], using the proposed app to enter data from a scenario taken from interviews with sex workers; all 5 agreed to do the walkthrough with the paper prototype and verbal consent was obtained. Paper prototypes have been used with good results in developing and testing prototypes [26]. In total, 4 of the 5 end users recruited did the cognitive walkthrough during the spring of 2022, three in private locations and one in an office (one was not available after contracting COVID-19). The scenarios used were taken from our previous research [13,14] and are outlined in Multimedia Appendix 1.

This protocol was submitted to the CUNY Graduate School of Public Health and Health Policy Human Research Protection Program, and was classified as exempt. No incentives were offered.

A 10 item system usability scale (SUS) was brought to the second cognitive walkthrough and asked the end user to rate the 10 statements, which were read aloud (by MHD), as a back-up evaluation method. SUS is a 10-item Likert scale, with each item’s score ranging from 0 to 4. Odd numbered items are scored at the scale position minus 1 and even numbered items are scored at 5 minus the scale position; the sum of the scores is then multiplied by 2.5 to obtain the overall score [27,28].

The following sections of this paper detail the evaluation methods undertaken in usability testing with a multidisciplinary team of informatics professionals and 4 potential end users based on the known theory that this number will generally expose the majority of problems with usability [25], in our attempt to evaluate the usability of the first prototype of ReportVASW.

Ethical Considerations

This project was deemed exempt by CUNY Graduate School of Public Health and Health Policy since it focused on the usability of a data collection tool and did not involve human participants or personal information.

Results

Step 1 - Contextualization

Sex workers consulted confirmed that the data collection tool should make it easy to report incidents of violence, including location, what violence occurred, and who committed this violence. They confirmed that the site should be easy for people who are in the aftermath of a traumatic event to use, demonstrating understanding of the end user population, sex workers.

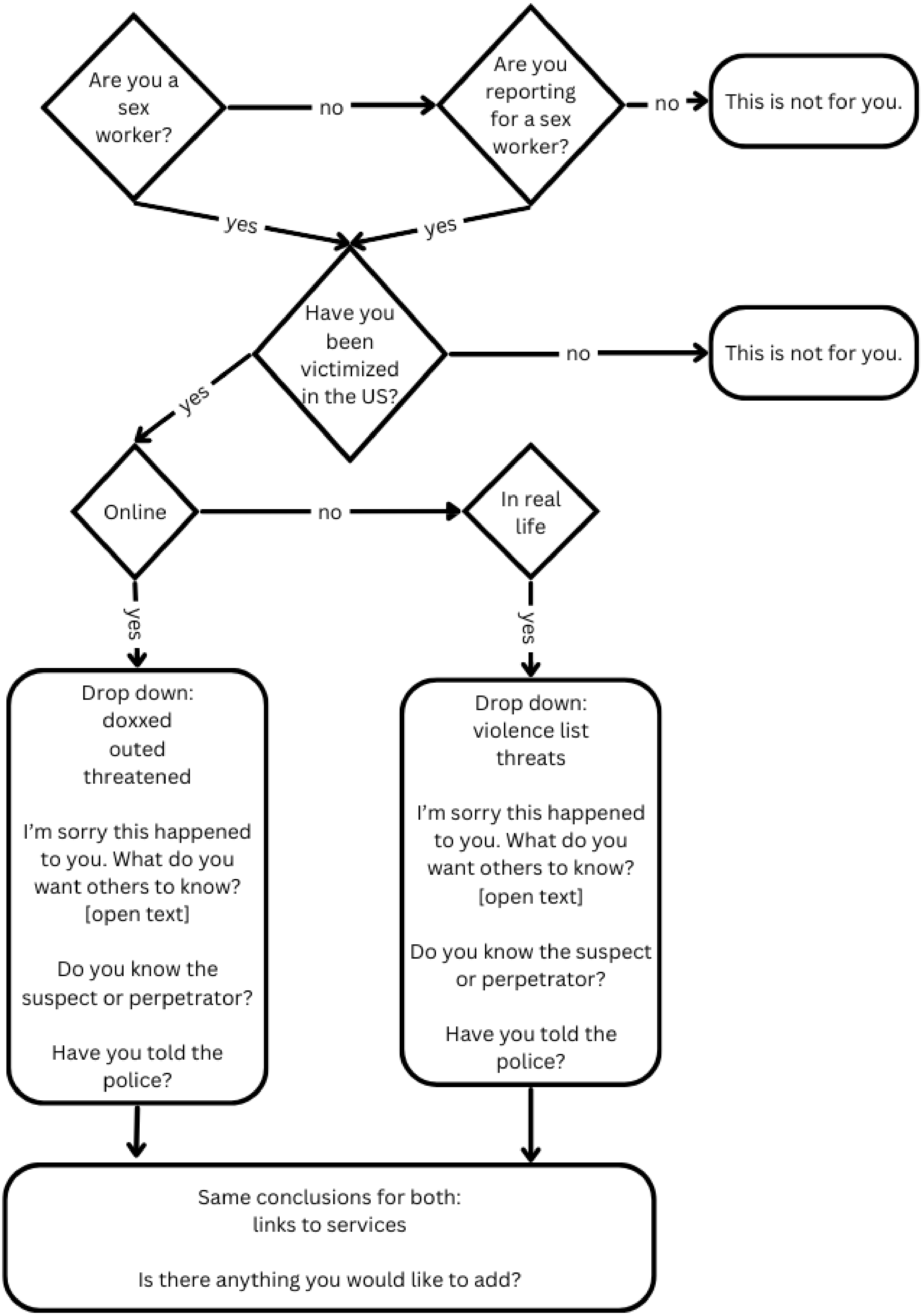
Step 1 is the functional analysis of ReportVASW. The information collected using ReportVASW could be used in multiple ways:

1. This site may facilitate information sharing in addition to documenting violence committed against sex workers and generate evidence to be used in reporting violent incidents in which sex workers are victimized.
2. Geolocation data about violent incidents can be used in the allocation of resources by organizations that work with sex workers, and in advocacy for additional resources.
3. The app could connect sex workers who have been victims of violence to an organization offering services to sex workers, perhaps including trauma-informed service providers, and possibly to attorneys in the area where the crime was committed. This would need to be determined by location; local sex worker groups would be consulted about friendly services to reach out to.

Step 2 - Ideation

The flowchart (Figure 1) went through 3 drafts in an iterative process. During this process, the ways the information would be collected and the order of questions were changed, including adding questions, each time making adjustments to the information presentation and order and ideas about how to collect it. The task analysis aspect of ideation was fruitful, because it forced the developers to clarify what should be identified and reported. The first component in the app is screening questions about sex work experience and victimization, the second component asks about the victimization, the final component would offer links to services.

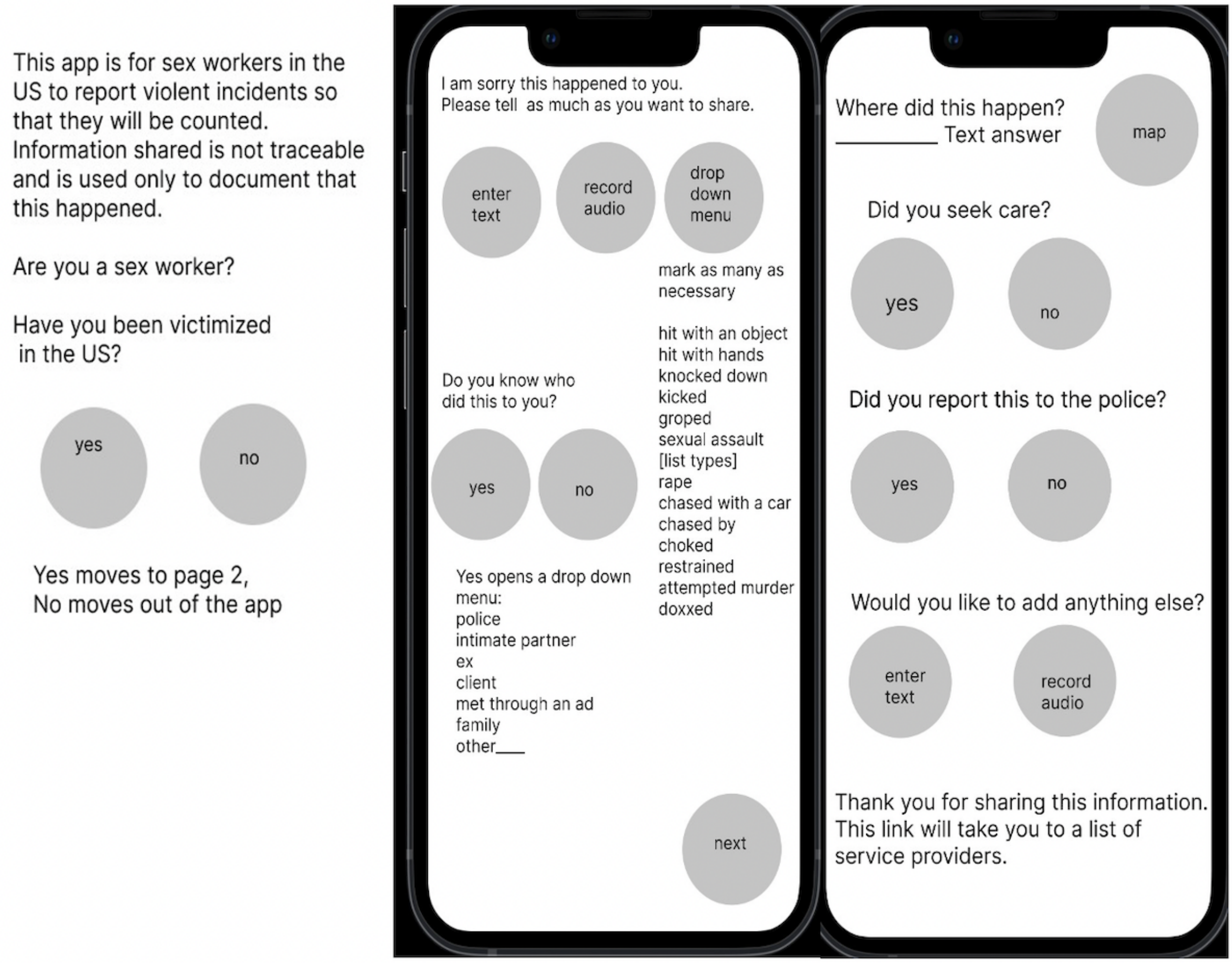
Figure 1. Ideation product: flowchart.



Step 3 - Prototyping: Representational Analysis

The heuristic analysis and cognitive walkthrough using scenario-based testing methods generated similar assessments about the ease of use of the prototype and that the aesthetics should be improved; this overall agreement would seem to indicate that the findings reflect the actual usability of the prototype ReportVASW (Figure 2).

Figure 2. Prototype screens (developed using Figma).



Evaluation Outcomes

Responses were compiled and average scores computed; heuristic analysis scores ranged from 0.5 to 3.25 (Table 1) and revealed that most problems are aesthetic or minor. The consistency inspection was the easiest to undertake, revealing that there were multiple problems with aesthetics:

- Buttons do not line up, looks messy.
- Labels are not uniform, labels are all over the place on the buttons.
- Color scheme is the default, and can be improved.

Table . Most severe problems identified by heuristic analysis, with scores over 2.5.

Problem	Heuristic violation characteristic	Average (SD)
Aesthetics	Aesthetics	3.25 (1.50)
No way for user to know if they can erase what they wrote or said or resubmit	Undo	3.25 (0.96)
Submit button missing	Minimalism	2.75 (1.26)
Wording is too direct, recommend using words that are sensitive to user so they can feel comfortable sharing	Language	2.75 (0.50)
Map: no indication if this is based on zip code, neighborhood, city, state	Visibility	2.67 (0.50)
Hard to read because there is too much on screen	Visibility	2.5 (1.29)

These are all aesthetic and minor changes, simple to address, and not an obstacle to development. The aesthetics, however, need more attention. Consolidating questions may have compromised usability by rendering the screen too crowded because they incorporate too many frames, thereby decreasing visibility and violating the heuristic of minimalism. The default color scheme should be changed.

The number of violations of each heuristic listed in the Heuristic Analysis chart were counted (Table 2). The most frequently violated heuristics were visibility (6), consistency (4), and minimalism (4). These are related: addressing violations of the heuristics of visibility and consistency will contribute to minimizing violations of the heuristic of minimalism.

Table . Frequency of heuristic violations.

Characteristic	Value (n)
Visibility	6
Consistency	4
Minimalism	4
Language	3
Control	2
Documentation	2
Flexibility	2
Feedback	1
Undo	1

Step 4 - Usability Testing: Scenario-Based Evaluation by Cognitive Walkthrough and Task Analysis

Overview

The first 2 cognitive walkthroughs and task analysis became in-depth discussions with input from the participating end user; end users confirmed the need for an application such as ReportVASW, and said that it could be a viable tool for collecting information about violence committed against sex workers, especially incidents that remain uncounted because they were not reported to police or victims did not present at hospitals, and is worth pursuing. The cognitive walkthrough process delivered positive feedback and users offered many ideas for improvements. Suggestions included new features, the collection of additional information, and aesthetic and functional comments, such as specific text for an introductory screen. Further, 4 end users completed the evaluation; the fifth was excluded because of illness.

Task Completion Time

In total, 2 people who completed the task did so in under 90 seconds. The other 2 end users offered more information than was asked during the cognitive walkthrough and so the timing of the actual task was not possible to measure; the discussion took over 20 minutes. All end users were very satisfied with the flexibility of multiple ways to submit information. Users felt that with a submit button, it would prove an effective way to collect information about the epidemic of violence against sex workers. End user testers of the prototype (Figure 2) made some recommendations, which are listed below.

Aesthetics

- Use universally recognized symbols where possible, for example, a microphone emoji for “record audio.”
- Improve attractiveness through color scheme and format; usability.
- Specify clearly that people can use any and all of these methods (text, record audio, or drop-down menu) to submit information.
- Consider 1 question per screen, which would advance without a “next” button. It may be possible to do this with multiple forms of data submission.
- Move open text field in drop-down menus to top, in order for people not to need to scroll to see it.
- Add a submit button.

Changes to Data Collection

- Change the screening question about victimization in order to capture data from people who do not identify as victims, but who have had violent experiences.
- Add a date field for the event, perhaps simply year, to distinguish recent events from long ago events.
- Add “drugged” to drop-down menu of types of victimization
- Change “did you seek care?” to specify “did you see a doctor or go to a hospital?” in order to be more clear, so that users will not include calling or visiting friends.
- Add branching questions after “did you report this incident to the police?” including “did they take your report?” and “were you treated respectfully?”
- Add demographic information about race and gender
- Ask whether the data should be shared with people collecting information about people sex workers should avoid (ie, a “bad date list”).

Additional Input

- Make more clearly anonymous, with an introductory page that emphasizes anonymity and the why of ReportVASW, emphasize that IP address and mobile numbers will not be stored, therefore “you can’t be traced or tracked” – this page must be clear, concise, and convincing.

- "Location" is ambiguous, and could be an address, as on a map, or a venue such as "car" or "brothel." Even place names can be unclear: Springfield is a town in every state.
- Ask whether the information was shared with other sex workers, as in a "bad date list."
- One end user tester suggested using speech to text with the audio function and enabling the speaker to edit in the moment.
- A critical point raised by the third tester was that the screening questions are good but will miss some people, highlighting that this person did not identify as someone who had been victimized, and so would be eliminated by the screening question "have you been victimized in the United States?" but had been drugged, threatened with a gun, and raped, at different times.
- A second critical point raised by the fourth tester was that the final question, "would you like to add anything else?" could lead to actionable information being shared without any way to act on this, including but not limited to suicidal ideation, violent impulses, and information about human trafficking situations. It was recommended that this open-ended question be eliminated to preclude the possibility of liability and to limit mental anguish for the person addressing these reports.
- All 4 agreed that entering the data from the scenarios was possible and easy, but nuance would lost without offering open-ended formats.
- One user pointed out that recounting violent experiences takes emotional energy, prompting her to ask, "Without a clear benefit to the victim, why do it? Can there be a way to connect to targeted services for the individual to report/record?"

About SUS

ReportVASW scored 90 on the SUS; SUS scores over 68 are considered good [27,28] (Table 3). The tool scored neutral on consistency, and all other responses were positive toward the app, with most being strong.

Table . System usability scale chart scores for ReportVASW.

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
I would like to use this	✓	—[a]	—	—	—
It is too complex	—	—	—	—	✓
Easy to use	—	✓	—	—	—
I need IT support	—	—	—	—	✓
Functions are well integrated	✓	—	—	—	—
Too much inconsistency	—	—	✓	—	—
Most would learn it fast	✓	—	—	—	—
Cumbersome	—	—	—	✓	—
I feel confident using it	✓	—	—	—	—
Requires much learning	—	—	—	—	✓

[a]Not applicable.

Discussion

Step 1 - Contextualization

UCD with multiple methods was successfully implemented in the development and evaluation of this first prototype, reflecting both success [15,16,18] and difficulties [20,21] reported by others, with the additional aspect of an atypical end user for informatics.

Step 2 - Ideation

Using multiple methods to evaluate the prototype enabled the collection of new information, including phrasing for screening questions, and positive reception of ReportVASW. New knowledge was gained from the evaluation, particularly through engagement with end users, even considering the lead author's significant expertise, particularly information about additional topics and language to incorporate in the next version of the prototype. The literature reflects the usefulness of multiple methods, despite challenges [20,21].

We believe we have sufficient input and information to proceed to significantly improve the next draft of the prototype, because the end user evaluation aligned with the heuristic analysis. End user comments offered solutions to issues identified in heuristic analysis particularly regarding aesthetic and functional issues; these solutions will be applied in the next steps. Additionally, UCD was useful in evaluating the ReportVASW prototype interface because adaptation is necessary to bring something designed for an Asian context to the American context; input from end users will improve this adaptation. Evaluations of the interface using heuristic analysis and end user scenario-based testing will inform the revisions to the prototype.

Step 3 - Prototyping: Representational Analysis (Heuristic Analysis)

Figma's 3 screen limit encouraged the consolidation of some of the screens, thereby making the app simpler. This involved consolidated screening questions on page 1, information about a violent event on page 2, and information about seeking care and reporting and asking if there is anything more they would like to share on the final screen. This reduced number of screens may be better than the initial drawings with 1 question per screen. Each step, from the flowchart (Figure 1) to the paper draft to the digital pictures (Figure 2) offered opportunities for improvement.

The heuristic analysis forms the baseline for the evaluation of the prototype through its next iteration. This analysis offered actionable recommendations and afforded interesting discussion related to the varied backgrounds of the analysts. For example, the most technically skilled of the analysts disagreed that some things were necessary, while an analyst with experience working with different communities offered important points about sensitive language, which the lead author is confident can be addressed through consultation with end users.

Step 4 - Usability Testing: Task Analysis by Users

As shown above, end users completed the task quickly, and offered substantive input. Their interested and substantive responses indicate a need for ReportVASW. Most information offered was concrete and included suggestions that can be easily incorporated, for example, each offered ways to formulate specific questions, about seeking care and about location. However, some of the input is not as easily addressed, such as how to phrase screening questions in order not to exclude people who do not see themselves as victimized. Other input pushes the developers to find ways to benefit participants, who are expending energy to share information about potentially traumatic experiences. Possibilities include offering a list of referrals to service providers around the United States, including clinical therapists and supportive trauma-informed health care professionals, in partnership with existing services used by sex workers. Further, 1 complication is that most services for victims of violence focus on women, usually cisgender women; however, sex workers of any gender may be victimized [1-3].

Sharing the information collected with "bad date lists" about people who commit violence against sex workers is more complicated that it sounds because of recent US legislation; law enforcement efforts have led to the closure of online venues for information sharing among sex workers [29,30]. While sex workers actively share their concerns online [31], US sex worker groups are decentralized, and sex worker groups alert their members about reports of bad dates; ultimately, ReportVASW should be managed by sex workers. Each could have copies of decentralized data, and in the future we will need to explore alternatives to manage this data, for example, using blockchain.

Adapting standardized methods to the end user population has been challenging to others, who recommend flexibility and accommodation of end users over rigidity about standardization [21]. The prompt for end users to begin the evaluation task must be chosen wisely. Cognitive walkthroughs with 2 end users featured interruptions with salient and helpful input. The third and fourth people who conducted the cognitive walkthrough each took approximately 90 seconds to complete the task. This time certainly does not account for the difficult nature of the material; none of the testers were recent victims of potentially traumatizing situations.

Next Steps

Follow-up is essential to the findings and implications of the project. We have received actionable recommendations through the cognitive walkthrough and the heuristic analysis that indicate clear urgent next steps. The agreement between the task analysis of the cognitive walkthrough and the input from the cognitive walkthroughs and the heuristic analysis included many recommendations addressing aesthetics, usability, data collection, and other input about ways to improve uptake and also increasing end users' confidence that ReportVASW is benign and not used for surveillance. Immediate next steps based on this input include:

- Adding all the input in changes to screening and data collection offered by end users, including adding a convincing introductory page about the use of the data and lack of tracking, as identified in heuristic analysis and with suggestions made during cognitive walkthrough,
- Making the urgent changes identified in the heuristic analysis including redesigning the interface for consistency and improving attractiveness.
- Exploring ways to link people providing input to services.

The most important next step will be to link users to services that could be helpful in the aftermath of violence, including the long-term aftermath, involving long-term effects of violence such as chronic disease [6] and post traumatic stress disorder [5]. It is not clear whether sharing links to legal and social services would meet this need. There are few low-barrier services for sex workers in the United States, presenting an obstacle to access. Considering this, it may prove beneficial to collaborate with an existing program offering legal and/or health services for sex workers. End users must be involved in the decisions about services included, in order to identify service providers that do not stigmatize or discriminate against sex workers. Additionally, geolocation data about violent events should be used to help identify where services are most urgently needed. Building more evidence will contribute to understanding reasons for sex workers to report violence against sex workers. However, police resistance to investigating violence against sex workers cannot be addressed by an app, and the data collected may be used in advocacy.

The next version of the prototype will also be evaluated using heuristic analysis. Comparing these sequential heuristic analyses will help determine priorities for changes to the following version. Using multiple methods for all 4 steps of the UCD process gave us richer information than we would have had using only 1 method at each step. While the time invested was significantly more than it might have taken using only 1 method, the benefits are great because the information gathered offers more certain next steps and reduces the chances of missing important elements that could require additional versions later.

Limitations

Figma constrained design possibilities that contributed to more creative ways to include information in less space. The number of individual end user evaluators was in the ideal range of 4 to 5 [25], while the cognitive walkthrough might benefit from another end user because of the interest in sharing additional information. The convenience sample of end users also presents a limitation, and data could have been different from people who were not familiar with the lead author, and the sample includes only self-identified women. SUS input may have been influenced by the lead author reading the questions and asking the answers, rather than the end user checking the boxes themselves. Additionally, the creator's bias impeded her ability to test the prototype, due to extreme familiarity. The small number of evaluators for the heuristic analysis presents a limitation.

Conclusions

The value of using multiple methods in UCD was clearly demonstrated in the process of designing and evaluating a prototype data collection to submit information about incidents of violence against sex workers. Using multiple methods in the initial steps of contextualization and ideation led to multiple revisions in these early stages. Using multiple methods in prototyping and evaluating the prototype afforded the opportunity to collect informative input from people in different roles, including end users and informatics professionals. The results from each method aligned such that the representational analysis, the consistency inspection, and heuristic analysis reinforced ways to improve the prototype, reinforcing the input from each source. End users confirmed the need for an application such as ReportVASW and that developing the data collection tool is worth pursuing, and informatics personnel reinforced the feasibility and offered insight to improve its design and utility. The use of multiple methods to evaluate the prototype contributed to a greater understanding than any single method alone.

Acknowledgments

We would like to thank the end users who agreed to undertake the usability testing. We would like to thank Sadaf Ahmad, MS, and Rafaela Villacres, MS, for undertaking the heuristic analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Scenarios used for the cognitive walkthrough.

[DOCX File, 14 KB - humanfactors_v11i1e53557_app1.docx]

References

1. Decker M, Rouhani S, Park JN, et al. Incidence and predictors of violence from clients, intimate partners and police in a prospective US-based cohort of women in sex work. Occup Environ Med 2020 Nov 3:oemed-2020-106487. [doi: 10.1136/oemed-2020-106487] [Medline: 33144359]
2. Deering KN, Amin A, Shoveller J, et al. A systematic review of the correlates of violence against sex workers. Am J Public Health 2014 May;104(5):e42-e54. [doi: 10.2105/AJPH.2014.301909] [Medline: 24625169]
3. Stenersen MR, Thomas K, McKee S. Police harassment and violence against transgender & gender diverse sex workers in the United States. J Homosex 2024 Feb 23;71(3):828-840. [doi: 10.1080/00918369.2022.2132578] [Medline: 36228168]
4. Fehrenbacher AE, Park JN, Footer KHA, Silberzahn BE, Allen ST, Sherman SG. Exposure to police and client violence among incarcerated female sex workers in Baltimore city, Maryland. Am J Public Health 2020 Jan;110(S1):S152-S159. [doi: 10.2105/AJPH.2019.305451] [Medline: 31967867]
5. Park JN, Decker MR, Bass JK, et al. Cumulative violence and PTSD symptom severity among urban street-based female sex workers. J Interpers Violence 2021 Nov;36(21-22):10383-10404. [doi: 10.1177/0886260519884694] [Medline: 31679445]
6. Decker MR, Park JN, Allen ST, et al. Inconsistent condom use among female sex workers: partner-specific influences of substance use, violence, and condom coercion. AIDS Behav 2020 Mar;24(3):762-774. [doi: 10.1007/s10461-019-02569-7] [Medline: 31254189]
7. CBC. News KL CBC-wide “bad date” reporting system aims to improve sex worker safety. : CBC News; 2021. URL: https://www.cbc.ca/news/canada/british-columbia/b-c-bad-datedatabase-sex-worker-safety-1.5897871 [accessed 2024-08-17]
8. Ditmore M. Aye myanmar association's community-based intervention to combat violence against sex workers, 2017-2020. United Nations Trust Fund to End Violence Against Women (UN Women). URL: https://untf.unwomen.org/en/digital-library/publications/2022/08/final-evaluation-community-based-intervention-to-combat-violence-against-sex-workers-myanmar#view [accessed 2024-08-27]

9. Brody C, Chhoun P, Tuot S, et al. A mobile intervention to link young female entertainment workers in Cambodia to health and gender-based violence services: randomized controlled trial. J Med Internet Res 2022 Jan 4;24(1):e27696. [doi: 10.2196/27696] [Medline: 34982716]
10. SANAC sex worker program. DureTechnology. URL: https://apps.apple.com/us/app/sanac-sex-worker-program/id1173413049 [accessed 2024-09-30]
11. Population size estimation of key populations. National AIDS Program of Sierra Leone. 2013. URL: https://www.nas.gov.sl/publication/140population-size-estimation-of-key-populations-august-2013 [accessed 2024-08-17]
12. Ditmore MH. Caught between the tiger and the crocodile: Cambodian sex workers' experiences of structural and physical violence. Stud Gend Sex 2014 Jan 2;15(1):22-31. [doi: 10.1080/15240657.2014.877726]
13. Thukral J, Ditmore MH. Revolving door. urban justice center sex workers project. 2003 URL: https://swp.urbanjustice.org/wp-content/uploads/sites/14/2019/09/RevolvingDoor.pdf [accessed 2024-08-27]
14. Ditmore MH, Maternick A, Zapert K. The Road North: The Role of Gender, Poverty and Violence in Trafficking from Mexico to the US.: Urban Justice Center Sex Workers Project; 2012. URL: https://swp.urbanjustice.org/wp-content/uploads/sites/14/2019/09/swp-2012-the-road-north-en.pdf [accessed 2024-08-17]
15. Schnall R, Rojas M, Bakken S, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251. [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
16. De Vito Dabbs A, Myers BA, Mc Curry KR, et al. User-centered design and interactive health technologies for patients. Comput Inform Nurs 2009;27(3):175-183. [doi: 10.1097/NCN.0b013e31819f7c7c] [Medline: 19411947]
17. Carr C, King LM, Maizel J, et al. Strategies and interventions used to prevent violence against sex workers in the United States: a scoping review using the social-ecological model. Trauma Violence Abuse 2024 Jul;25(3):2436-2451. [doi: 10.1177/15248380231214786] [Medline: 38054440]
18. Lyon AR, Munson SA, Renn BN, et al. Use of human-centered design to improve implementation of evidence-based psychotherapies in low-resource communities: protocol for studies applying a framework to assess usability. JMIR Res Protoc 2019 Oct 9;8(10):e14990. [doi: 10.2196/14990] [Medline: 31599736]
19. Graham AK, Munson SA, Reddy M, et al. Integrating user-centered design and behavioral science to design a mobile intervention for obesity and binge eating: mixed methods analysis. JMIR Form Res 2021 May 10;5(5):e23809. [doi: 10.2196/23809] [Medline: 33970114]
20. Noergaard B, Sandvei M, Rottmann N, et al. Development of a web-based health care intervention for patients with heart disease: lessons learned from a participatory design study. JMIR Res Protoc 2017 May 17;6(5):e75. [doi: 10.2196/resprot.7084] [Medline: 28526674]
21. Cornet VP, Toscos T, Bolchini D, et al. Untold stories in user-centered design of mobile health: practical challenges and strategies learned from the design and evaluation of an app for older adults with heart failure. JMIR Mhealth Uhealth 2020 Jul 21;8(7):e17703. [doi: 10.2196/17703] [Medline: 32706745]
22. Experience WL in RBU. 10 usability heuristics for user interface design. : Nielsen Norman Group URL: https://www.nngroup.com/articles/ten-usability-heuristics/ [accessed 2024-08-17]
23. Uribe-Ocampo S, Torres EA, Luna IF, Florez-Arango JF, Smith JW. SIM-CIG: a serious game to practice and improve clinical guidelines adoption based on computer-interpretable guidelines. Stud Health Technol Inform 2019 Aug 21;264:1997-1998. [doi: 10.3233/SHTI190753] [Medline: 31438447]
24. Zhang J, Wallji M. TURF: Toward a unified framework of EHR usability. J Biomed Inform. 2011. [doi: 10.1016/j.jbi.2011.08.005]
25. Nielsen J. Estimating the number of subjects needed for a thinking aloud test. Int J Hum-Comput Stud 1994 Sep;41(3):385-397. [doi: 10.1006/ijhc.1994.1065]
26. Or C, Tao D. Usability study of a computer-based self-management system for older adults with chronic diseases. JMIR Res Protoc 2012 Nov 8;1(2):e13. [doi: 10.2196/resprot.2184] [Medline: 23612015]
27. Brooke J. SUS - A quick and dirty usability scale. 2023. URL: https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html [accessed 2024-08-17]
28. Cunningham S, Sanders T, Scoular J, et al. Behind the screen: commercial sex, digital spaces and working online. Technol Soc 2018 May;53:47-54. [doi: 10.1016/j.techsoc.2017.11.004]
29. Campbell R, Sanders T, Scoular J, Pitcher J, Cunningham S. Risking safety and rights: online sex work, crimes and 'blended safety repertoires.'. Br J Sociol 2019 Sep;70(4):1539-1560. [doi: 10.1111/1468-4446.12493] [Medline: 30318604]
30. Moorman JD, Harrison K. Gender, race, and risk: intersectional risk management in the sale of sex online. J Sex Res 2016 Sep;53(7):816-824. [doi: 10.1080/00224499.2015.1065950] [Medline: 26488687]
31. Al-Rawi A, Zemenchik K. Sex workers' lived experiences with COVID-19 on social media: content analysis of Twitter posts. JMIR Form Res 2022 Jul 14;6(7):e36268. [doi: 10.2196/36268] [Medline: 35767693]

Abbreviations

mHealth: mobile health
SUS: system usability scale

UCD: user-centered design
VASW: violence against sex worker

Edited by A Parush; submitted 10.10.23; peer-reviewed by C Daley, T Lillie; revised version received 01.05.24; accepted 19.05.24; published 09.10.24.

Please cite as:
Ditmore MH, Florez-Arango JF
User-Centered Design for Designing and Evaluating a Prototype of a Data Collection Tool to Submit Information About Incidents of Violence Against Sex Workers: Multiple Methods Approach
JMIR Hum Factors 2024;11:e53557
URL: https://humanfactors.jmir.org/2024/1/e53557
doi:10.2196/53557

Development of a System for Predicting Hospitalization Time for Patients With Traumatic Brain Injury Based on Machine Learning Algorithms: User-Centered Design Case Study

Huan Zhou[1*], ME; Cheng Fang[2*], MM; Yifeng Pan[1], ME

[1]The School of Big Data and Artificial Intelligence, Anhui Xinhua University, Hefei, China

[2]Department of Neurosurgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Yifeng Pan, ME
The School of Big Data and Artificial Intelligence, Anhui Xinhua University, , Hefei, , China

Abstract

Background: Currently, the treatment and care of patients with traumatic brain injury (TBI) are intractable health problems worldwide and greatly increase the medical burden in society. However, machine learning–based algorithms and the use of a large amount of data accumulated in the clinic in the past can predict the hospitalization time of patients with brain injury in advance, so as to design a reasonable arrangement of resources and effectively reduce the medical burden of society. Especially in China, where medical resources are so tight, this method has important application value.

Objective: We aimed to develop a system based on a machine learning model for predicting the length of hospitalization of patients with TBI, which is available to patients, nurses, and physicians.

Methods: We collected information on 1128 patients who received treatment at the Neurosurgery Center of the Second Affiliated Hospital of Anhui Medical University from May 2017 to May 2022, and we trained and tested the machine learning model using 5 cross-validations to avoid overfitting; 28 types of independent variables were used as input variables in the machine learning model, and the length of hospitalization was used as the output variables. Once the models were trained, we obtained the error and goodness of fit (R^2) of each machine learning model from the 5 rounds of cross-validation and compared them to select the best predictive model to be encapsulated in the developed system. In addition, we externally tested the models using clinical data related to patients treated at the First Affiliated Hospital of Anhui Medical University from June 2021 to February 2022.

Results: Six machine learning models were built, including support vector regression machine, convolutional neural network, back propagation neural network, random forest, logistic regression, and multilayer perceptron. Among them, the support vector regression has the smallest error of 10.22% on the test set, the highest goodness of fit of 90.4%, and all performances are the best among the 6 models. In addition, we used external datasets to verify the experimental results of these 6 models in order to avoid experimental chance, and the support vector regression machine eventually performed the best in the external datasets. Therefore, we chose to encapsulate the support vector regression machine into our system for predicting the length of stay of patients with traumatic brain trauma. Finally, we made the developed system available to patients, nurses, and physicians, and the satisfaction questionnaire showed that patients, nurses, and physicians agreed that the system was effective in providing clinical decisions to help patients, nurses, and physicians.

Conclusions: This study shows that the support vector regression machine model developed using machine learning methods can accurately predict the length of hospitalization of patients with TBI, and the developed prediction system has strong clinical use.

(JMIR Hum Factors 2024;11:e62866) doi:10.2196/62866

KEYWORDS

machine learning; traumatic brain injury; support vector regression machine; predictive model; hospitalization

Introduction

Traumatic brain injury (TBI) surgery increases the risk of certain complications, posttraumatic head injury complications that can lead to severe acute and chronic morbidity and mortality, and these complications are known risk factors for prolonged hospital stays [1-4]. In addition, the average length of stay after surgery for patients with traumatic brain trauma is an important indicator of the quality of medical management in neurosurgery departments and the overall level of care in this disease area and to some extent reflects the severity and complexity of the patient's disease. In the current environment of insufficient supply of medical resources, it is important to predict the hospitalization time of patients with brain trauma in advance by technical means, and then combine the predicted hospitalization time with ward management methods to improve the bed turnover rate and medical service capacity, so as to improve the current situation of patients' difficulty in hospitalization, reduce unnecessary medical expenses, and alleviate the medical burden of society [5-8]. Therefore, the purpose of this study is to develop a system for predicting the length of stay of patients with brain trauma based on clinical data from hospital medical records and machine learning algorithms to provide clinical decision support for physicians and reference for nurses to coordinate ward management in advance.

Given the complexity and diversity of clinical data after brain injury, complex statistical features such as multiple nonlinearities between different factors are prevalent [9-11]; the use of traditional multifactor logistic regression (LR) analysis methods alone cannot establish a predictive model for the hospitalization time of patients with early brain trauma, at which time different machine learning models are required for statistical analysis and prediction [12,13]. Convolutional neural networks (CNNs) and support vector machines are currently the mainstream prediction models for processing and analyzing and building complex statistical data that enable feature extraction and building prediction models for clinical data information. In this study, multiple machine learning models were developed to predict the hospitalization time of patients with brain trauma, and the advantages and disadvantages of prediction models based on different machine learning algorithms were explored through a comparative analysis research method, and the model with the best performance was selected. Finally, based on the optimal machine learning models, we develop a system that can be applied to clinical decision-making.

This paper describes the development of a system for predicting the length of stay of patients with craniocerebral trauma based on a machine learning algorithm, using a machine learning method that was retrospectively applied to analyze clinical data of patients with craniocerebral trauma from the Second Affiliated Hospital of Anhui Medical University, and predicting the length of stay of patients with craniocerebral trauma from this dataset. Our objectives were to:

1. prospectively predict the length of stay of patients based on clinical data and
2. develop a system that can be applied to clinical decision-making by means of an optimal prediction model.

Methods

Data Sources and Exclusion Criteria

Our dataset was obtained from a total of 1128 case records from the Neurosurgery Center of the Second Affiliated Hospital of Anhui Medical University. The decision to discharge a patient requires discussion in the treatment group and assessment by an experienced neurosurgeon superior before a decision can be made. Therefore, in order to ensure that the output of the developed predictive length-of-stay model was determined by the assessment of experienced neurosurgeons and was not influenced by the subjective desire of the patient's family to abandon treatment or transfer to another department resulting in loss of follow-up, we set exclusion criteria for the medical records:

1. Automatic discharge of the patient at the request of the patient's family to forgo treatment.
2. Serious injuries in other areas requiring transfer to the relevant department for further treatment.
3. Patients with previous experience of craniocerebral injury.

The exclusion criteria were developed and the data that met the exclusion criteria were removed from the model we developed through natural language processing (NLP) techniques. A total of 1001 patients were successfully enrolled, and by random splitting, we used 70% of the data (700 items) for training the model and the remaining 30% (301 items) for testing the performance of the model. In addition, we also collected clinical data from 111 patients at the First Affiliated Hospital of Anhui Medical University as external test data for external validation of the model.

Ethical Considerations

This retrospective cohort study was approved by the Ethics Committee of the Second Affiliated Hospital of Anhui Medical University (S20210098). Participants or proxies signed the relevant informed consent forms within 24 hours of admission.

Feature Matrix

Table 1 summarizes the input data features used to predict length of stay. We selected a total of 28 features recorded in the medical record system that were available in our dataset, and these data can be used to prospectively determine the length of stay of patients with craniocerebral trauma in a practical application. These data were selected because of the experience provided by previous studies.

Table . Input dataset features.

Name	Type	Unit	Data availability (%)
Age	Integer	Years	100
Gender	Male/female	—[a]	100
Hypertension	Boolean	—	100
Diabetes mellitus	Boolean	—	100
Alzheimer disease	Boolean	—	100
Coronary disease	Boolean	—	100
Chronic bronchus	Boolean	—	100
Arthrolithiasis	Boolean	—	100
Hypothyroidism	Boolean	—	100
Hyperthyroidism	Boolean	—	100
Personal history of tumor	Boolean	—	100
Cirrhosis	Boolean	—	100
Pancreatitis	Boolean	—	100
Hyperlipidemia	Boolean	—	100
Cerebral infarction	Boolean	—	100
Chronic obstructive pulmonary disease	Boolean	—	100
Hepatitis	Boolean	—	100
Poliomyelitis	Boolean	—	100
Tuberculosis	Boolean	—	100
Nephrotic syndrome	Boolean	—	100
Atrial fibrillation	Boolean	—	100
Mechanism of brain injury	Motor vehicle accident/all on the same plane/falling from height/injuries caused by heavy objects/none	—	100
Is there any loss of consciousness after injury?	Boolean	—	100
Glasgow Coma Index score for admission	Integer	—	100
Head CT[b] examination on admission	Skull fracture/cerebral contusion/subdural hematoma/subarachnoid hemorrhage/intracranial pneumatosis	—	100
Brain surgery	Boolean	—	100
Intensive care treatment	Boolean	—	100
Complications during hospitalization	Bacterial infection/tracheotomy/anemia/gastrointestinal bleeding/liver function damage/electrolyte disorder/respiratory failure/abnormal coagulation function/thrombocytopenia/heart failure/peripheral facial paralysis/posttraumatic epilepsy/cerebrospinal fluid leak/acute coronary syndrome/none	—	100

[a]Not available.

[b]CT: computed tomography.

Estimation of Missing Data

For the missing data in this study, the model method is used to complete them. We will predict the missing fields as target variables based on other existing fields to obtain the most probable complementary values. If the column with missing values is a numerical variable, the regression model is used to complete it. If it is a categorical variable, the categorical model is used to complete it. The steps of [14] modeling method are as follows.

1. Determine the variables (characteristic columns) that fill in the missing values.
2. Splitting the original dataset: split the original dataset into 2 subsets according to the variables that need to be filled with missing values: (1) without missing values: dataset_train; and (2) with missing values only dataset_pred)

1. Identify and test the correlation of the variables of interest: empirical analysis determines which attribute columns are correlated with the variables filled with missing values, and statistical analysis tools are applied to view the correlations between the selected attribute columns on the dataset_train dataset for validation.

1. Modeling and prediction: use the dataset_train dataset to build a linear regression model and apply the built model to estimate predictions for the missing variables in the dataset_pred dataset,

1. Merge and reduce datasets: merge and reduce the 2 subsets into 1 dataset to prepare the data for subsequent modeling.

Model Establishment

Overview

In this study, we established 6 models: random forest (RF), CNN, support vector regression (SVR), multilayer perceptron (MLP), back propagation (BP) neural network, and LR, and compared the mean absolute percentage error (MAPE) and goodness of fit of the actual and predicted values to determine the optimal model for predicting the hospitalization time of patients with craniocerebral trauma in the system. The MAPE formula is shown in equation 1, where *n* is the sample size, yi` is the predicted value, and yi is the true value.

$$(1)\quad MAPE=\frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i'-y_i}{y_i}\right|$$

Random Forest

The RF model was chosen because we considered the following advantages of the RF model [15]:

1. It can handle very high-dimensional data and does not have to do feature selection because the subset of features is chosen randomly.
2. After training, it is able to derive feature importance.
3. When creating an RF, an unbiased estimate of the generalization error is used, and the model generalizes well.
4. The trees are independent of each other during training, which makes training fast and easy-to-make parallelization methods.

Convolutional Neural Network

CNN is a class of feedforward neural networks that contains convolutional computation and has a deep structure, which is one of the representative algorithms of deep learning and has been widely used in various fields. Compared with traditional neural network algorithms, CNN has stronger modeling ability to extract effective feature data from the input relevant data and learn the internal structure of the feature data for better prediction [16-18]. We consider that CNNs have the following characteristics, which are suitable for application in this study:

1. CNNs have a weight-sharing network structure, which reduces the complexity of the network model and reduces the number of weights.
2. The data (including image data) can be directly used as the input of the network, avoiding the complicated process of feature extraction and data reconstruction in traditional algorithms.

Support Vector Regression

SVR is suitable for solving various regression prediction problems thanks to kernel functions and a few support vectors that play a decisive role and has achieved excellent prediction results [19]. Drucker et al [20] proposed a new regression technique based on the Vapnik support vector concept in 1996. SVR was compared with regression techniques based on regression trees and ridge regression performed in the feature space. Based on these experiments, it is concluded that SVR will be advantageous in high-dimensional spaces because SVR optimization does not depend on the dimensionality of the input space. Considering that the input data of this study have 28 dimensions, the SVR machine was chosen to predict the length of stay of patients with craniocerebral trauma.

MLP, BP Neural Network, and LR

In addition, 3 classical regression models, namely, MLP, BP, and LR, were developed for application in predicting the length of stay in patients with craniocerebral trauma. These 3 models are also the most common prediction models in the field of applied clinical informatics and have achieved good results in a large number of related studies [21-29].

Training Set and Test Set Ratio

After establishing these 6 machine learning models, all valid original data samples are randomly disrupted and divided into a training set and a test set. In order to study the effect of different training set sample sizes on the modeling effect of different machine learning algorithms, the same rules are used to divide the training set and test set into 50%, 60%, 70%, 80%, and 90% of the total samples, and the MAPE of the test set samples is used to evaluate the model error.

External Dataset Validation Model

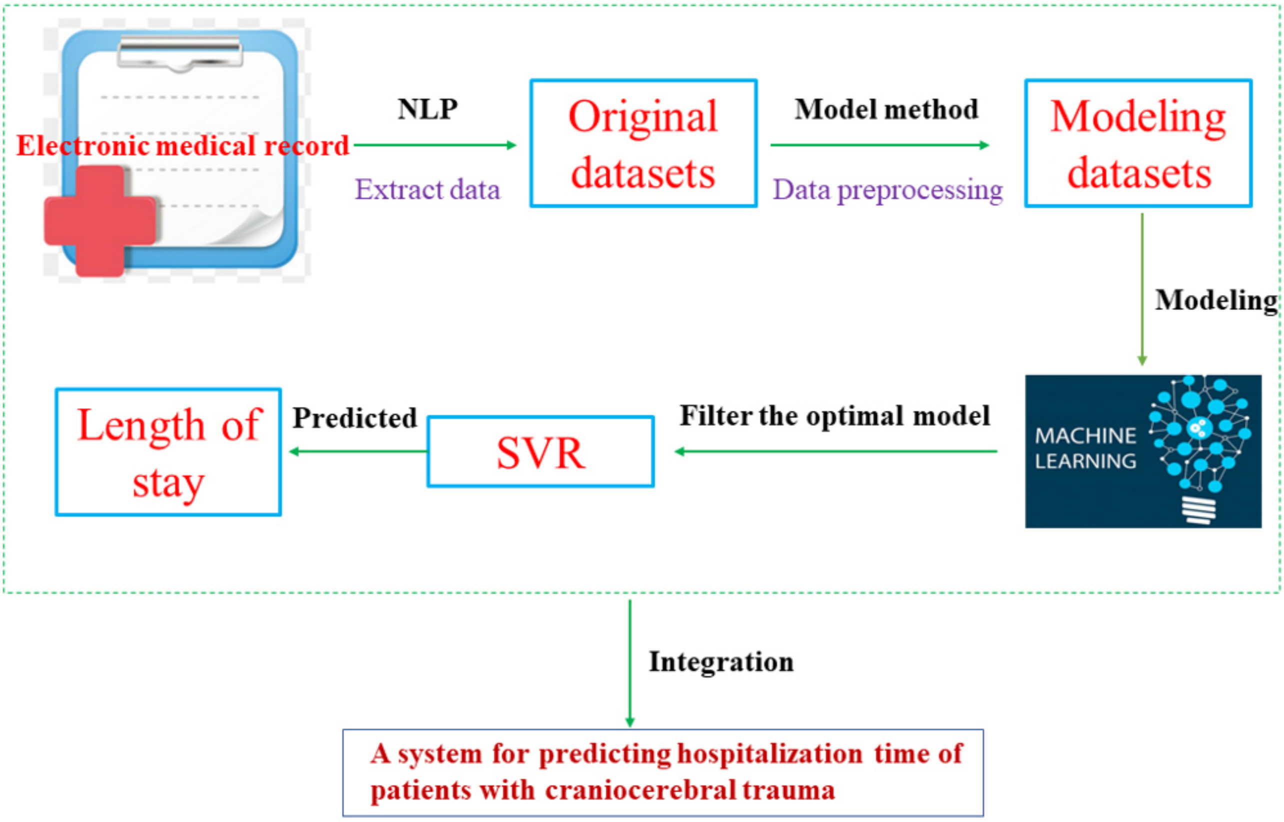
Considering that the ultimate goal of this study is to apply the system to clinical decision-making and ward bed management, external validation is needed to demonstrate that the optimal model selected in this study has a strong generalization capability, that is, the ability to predict datasets other than the modeled data. The reason for external validation is that overfitting may occur during the modeling process, in which case the model predicts the modeled dataset well, but does not work well for other datasets (test set). Such a model is obviously of no application value. Therefore, the model designed in this study uses 111 data from the First Affiliated Hospital of Anhui Medical University for external validation, and once the accuracy of the external validation of the model meets the requirements of the system application, the model can be encapsulated into the system developed in this study.

Predictive Modeling Pipeline

The development of our predictive modeling pipeline is based on Python 3.9 (Python Software Foundation), PyTorch 1.9.0 (Meta), and the django framework (Adrian Holovaty and Simon Willison), and consists of the following 4 main modules, and the process of developing the system is shown in Figure 1:

1. Data extraction module: Using NLP technology to extract the 28 features we need in the medical record system for patients with craniocerebral trauma, and based on the exclusion criteria we set, the data that meet the exclusion criteria are eliminated.
2. Data preprocessing module: The input feature data are automatically filled in with missing values using the model method to ensure the integrity of each patient record.
3. Model evaluation module: Visualize and evaluate the model-based predictions using the results of various standard key performance indicators.
4. Prediction of hospitalization time module: In this study, by comparing the prediction accuracy of the 6 models established, the model with the best prediction accuracy was established as the prediction model for system application, and finally the hospitalization time of patients with TBI was predicted by the input of the input features.

Figure 1. Development process of a system for predicting hospitalization time of patients with traumatic brain injury. NLP: natural language processing; SVR: support vector regression.



The predictive modeling pipeline ensures reproducibility and process stability and features many modules for processing data from medical record systems through machine learning algorithms to predicted length of stay data, which also supports the development of applied clinical systems for predicting the length of stay of patients with TBI.

Results

Proportion of Optimal Training and Testing Sets for Different Models

There is no fixed value for the partition ratio between the training set and the test set, and approximately 2 of 3 to 4 of 5 of the samples are usually used for training. The most common training set and test set ratios are 7:3 or 8:2. In order to achieve the best prediction accuracy of these 6 machine learning models designed in this study, the same rules are used to partition the training set and test set. The training set samples account for 50%, 60%, 70%, 80%, and 90% of the total samples. The MAPE of the test set samples was used to evaluate the prediction accuracy of these 6 models, and the experimental results are shown in Table 2.

The experimental results in Table 2 indicate that the optimal training and testing set ratios for different models may not necessarily be the same. Based on the experimental results, we selected the optimal training and testing set ratios for CNN, SVR, LR, RF, BP, and MLP as 0.7, 0.7, 0.8, 0.7, 0.8, and 0.7, respectively.

Table . Comparison of accuracy of 6 machine learning model test sets under different sample ratios of training sets.

Modeling set/test set	MAPE[a] (%)					
	CNN[b]	SVR[c]	LR[d]	RF[e]	BP[f]	MLP[g]
0.5	30.76	28.46	40.57	32.74	35.93	33.70
0.6	24.83	22.78	35.14	25.63	28.65	26.49
0.7	12.19	10.69	27.68	13.47	21.37	18.41
0.8	15.63	18.25	22.75	15.40	19.28	21.76
0.9	28.43	26.71	34.13	26.24	32.17	28.43

[a]MAPE: mean absolute percentage error.
[b]CNN: convolutional neural network.
[c]SVR: support vector regression.
[d]LR: logistic regression.
[e]RF: random forest.
[f]BP: back propagation.
[g]MLP: multilayer perceptron.

Model Accuracy

Divide all 1001 valid samples into training and testing sets. Divide the training and testing sets based on the results of the optimal ratio of the 6 models in Table 2. Then, start using these 6 algorithms to train the model for predicting hospitalization time, including cross-validation data. Repeat the training and testing 5 times and take the average value. For the prediction model of TBI patients' length of stay, the goodness of fit and the MAPE of the test set are used to check the model performance. The reason for using the test set results to evaluate the model is that the test set results can screen out models with strong generalization ability for us, which is universal in our system. Based on this, we can directly adjust the parameters of different models through the errors of the model on the test set, making the predictive ability of the model better and the applicability of the system for predicting hospitalization time stronger.

From Table 3, we can see that in the test set, SVR has the lowest MAPE and the best goodness of fit, and CNN and RF perform well in the test set. The higher the error of LR, BP, and MLP than that of the other 3 models, the poorer the applicability of these 3 models, which should be considered for exclusion.

Table . MAPE and R^2 of 6 models in the test set.

Model	MAPE[a] (%)	R^2
CNN[b]	11.98	0.862
SVR[c]	10.22	0.904
LR[d]	21.83	0.718
RF[e]	14.27	0.827
BP[f]	18.35	0.785
MLP[g]	19.14	0.772

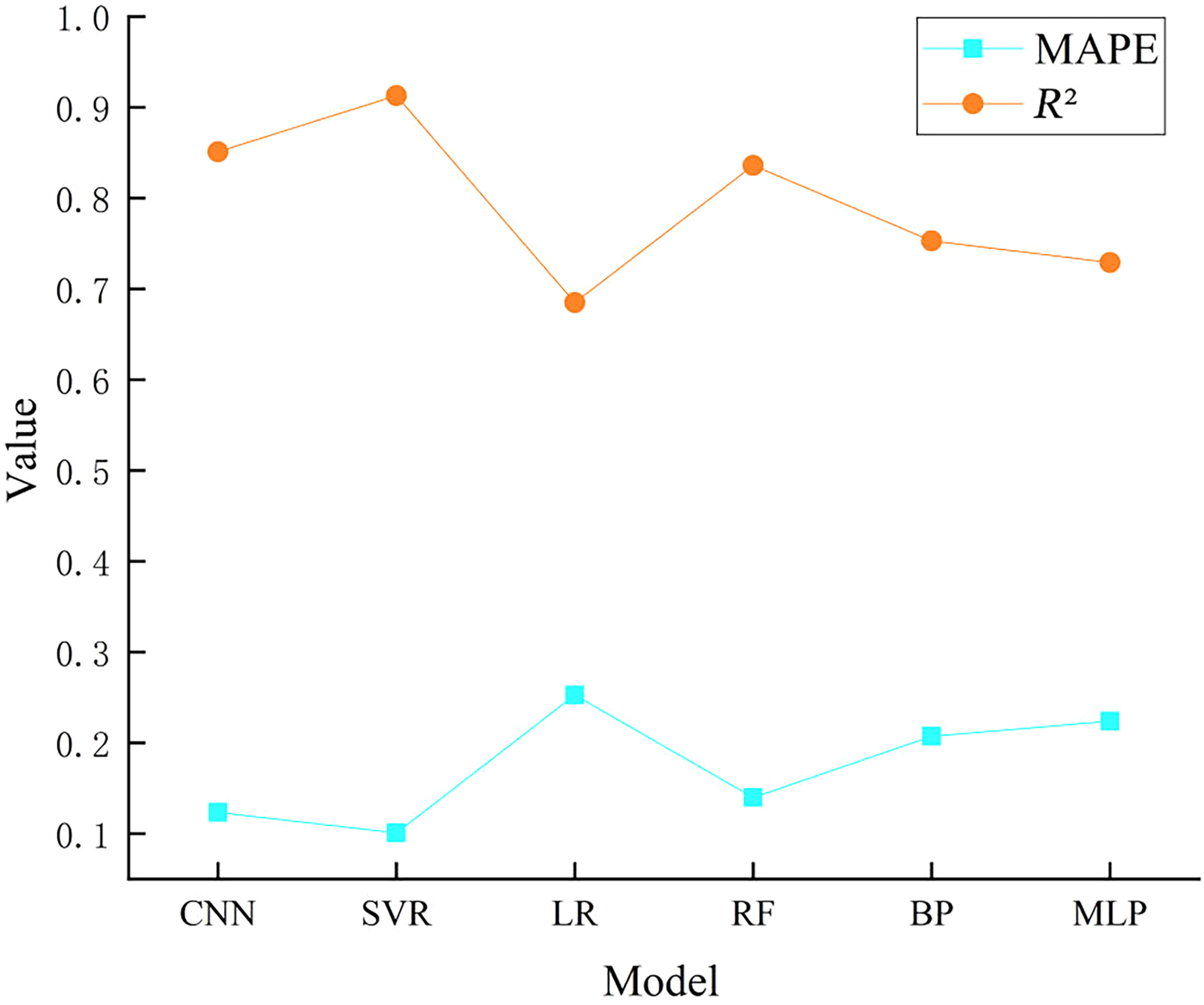
[a]MAPE: mean absolute percentage error.
[b]CNN: convolutional neural network.
[c]SVR: support vector regression.
[d]LR: logistic regression.
[e]RF: random forest.
[f]BP: back propagation.
[g]MLP: multilayer perceptron.

External Dataset Validation Results

Considering that the data of the test set and training set in this study are from the medical record system of the Second Affiliated Hospital of Anhui Medical University, in order to avoid contingency of the model prediction accuracy, consider using the data from the medical record system of other hospitals to assist in verification. If the predictive performance of data input from other hospitals into the system is poor, consider developing different models for data from different centers to ensure the high accuracy of the system. If the prediction accuracy is excellent, it proves the universality of the system and can be applied more widely.

Therefore, in order to further compare the reliability of the algorithm, 111 data records from the First Affiliated Hospital of Anhui Medical University were used for external validation. A total of 28 types of input variables required by our model were extracted through NLP from the collected medical records and input into 6 models to compare the advantages and disadvantages of different models in the external dataset. The experimental results are shown in Figure 2. The experimental results show that the SVR-based hospital stay prediction model has the lowest error and the highest R^2 in the external dataset, greatly maintaining the accuracy of the predicted data. It is the best machine learning model for predicting hospital stay. Therefore, the system we developed chose SVR as our prediction model.

Figure 2. MAPE and R^2 of 6 models in external test sets. BP: back propagation; CNN: convolutional neural network; LR: logistic regression; MAPE: mean absolute percentage error; MLP: multilayer perceptron; RF: random forest; SVR: support vector regression.



Satisfaction Evaluation of Patients, Nurses, and Doctors

After determining SVR as the final prediction model, this study ultimately designed a system to predict the hospitalization time of patients with TBI using data from an electronic medical record system. In order to verify the applicability of the clinical system, this study designed survey questionnaires for nurses and doctors to obtain satisfaction evaluations.

A total of 88 doctors and nurses from the First Affiliated Hospital of Anhui Medical University and the Second Affiliated Hospital of Anhui Medical University participated in the questionnaire. We set the following questions to assess health care workers' satisfaction with the system:

1. Have you learned about and used our newly developed system for predicting length of stay for patients with TBI?
2. How well do you think the system supports clinical decision-making?
3. Do you think the system can improve your work efficiency?
4. How accurate do you think the system is?
5. Do you think the system's predictions will help you better manage your patients' length of stay?
6. Do you think the user interface of the system is user-friendly?
7. How well do you think the system protects data security and privacy?
8. Would you recommend the system to other colleagues or health care organizations?
9. Whether the system's predicted outcomes will affect your treatment plan or patient management plan?
10. Do you feel the system integrates seamlessly with your existing workflow?

The above 10 questions will receive 1 point for agreement, 0 point for disagreement, and a maximum score of 10 points. The final average score is 9.18 points. The questionnaire results indicate that the system has strong practicality for patients. During the application process, most doctors and nurses stated that informing patients of the accurate length of stay during hospitalization can help alleviate their own stress, reduce fear of illness, and plan their life after discharge in advance.

Discussion

Principal Findings

The main finding of this study is that by using machine learning models, we can effectively predict the length of hospital stay for patients with TBI, which is of great significance for their rehabilitation and efficient use of medical resources. Specifically, the prediction system can help medical professionals more accurately evaluate the patient's condition and develop treatment plans, thereby arranging surgery, medication treatment, and rehabilitation training reasonably; reducing waste of medical resources; and avoiding frequent hospitalization due to unstable patient recovery, reducing medical costs. In addition, predicting hospitalization time can also help patients and their families better plan their lives, understand treatment progress and rehabilitation plans, and improve confidence in treatment and rehabilitation outcomes. This study used a large amount of clinical data on TBI accumulated over the years at the Second Affiliated Hospital of Anhui Medical University, combined with machine learning algorithms, to develop a complete system from data extraction, preprocessing, and hospital stay prediction to model evaluation. This system realizes fully automatic operations from electronic medical records to hospital stay prediction and visualization. Users only need to query the hospital stay number to obtain the prediction results, which is convenient and fast to use and has high practicality. This research result not only provides assistance to doctors in clinical decision support but also significantly improves the rehabilitation effect and quality of life of patients.

In the 4 modules of system development, this study focuses on the prediction of hospitalization time module. We use NLP technology to extract enough clinical data from the electronic medical record for processing and screening, as the input variable of the model, and the length of stay as the output variable. Using input-output datasets, 6 machine learning models (CNN, SVR, LR, RF, BP, and MLP) were compared and constructed. In order to avoid the impact of the same proportion of training and testing sets on the prediction accuracy of different machine learning algorithms, and to achieve the best prediction accuracy of these 6 machine learning models designed in this study, the same rules were used to divide the training and testing sets into 50%, 60%, 70%, 80%, and 90% of the total samples and use the MAPE of the test set samples to evaluate the prediction accuracy of these 6 models. For these 6 models (CNN, SVR, LR, RF, BP, and MLP), we ultimately chose the optimal training and testing set ratios of 0.7, 0.7, 0.8, 0.7, 0.8, and 0.7, respectively. Divide the optimal ratio of training and testing sets and then start using these 6 algorithms to train models for predicting hospitalization time. Evaluate the performance of the prediction model based on the results of the testing set. The experimental results show that the minimum MAPE of SVR is 10.22%, and the best goodness of fit is 0.904. The CNN and RF perform well on the test set. The errors of LR, BP, and MLP are very high, and the goodness of fit is low, which indicates that these 3 models have poor clinical applicability, and should be excluded. At the same time, in order to further validate the reliability of the algorithm, the universality of the system for predicting hospitalization time of patients with TBI was verified through external datasets. The experimental results showed that the SVR-based hospitalization time prediction model had the lowest error and the highest R^2 in the external dataset, maintaining the accuracy of the predicted data to a great extent. It is the best machine learning model for predicting hospitalization time. Therefore, the system we developed chose SVR as our prediction model.

Although in this study we compared and applied different machine learning algorithms to select the best algorithm SVR for application in the system and developed a system to predict the hospitalization time of patients with TBI through electronic medical records, the model adjustment and data-preprocessing steps were too specific, resulting in the current prediction mechanism being limited to local hospitals, while for hospitals in sparsely populated areas, it is not yet known whether the system has applicability. In addition, the differences in the electronic medical record systems of different hospitals can also lead to the inability to obtain the 28 input data required by this system through NLP, resulting in the inability to complete predictions or a decrease in prediction accuracy. Therefore, in the future, it is necessary to adopt a unified length of stay prediction framework to generate more reliable estimates and use them in different hospitals, where the system structure of electronic medical records for patient populations is similar. Finally, considering the inherent complexity and uncertainty of the system for predicting hospitalization time for TBI, as well as the relevant data currently being collected from different hospitals, we hope that the model needs to be widely applicable in subsequent system updates [30].

Nowadays, the application of machine learning models in the field of clinical decision support has become increasingly widespread [31-33]. However, this system can only predict the hospitalization time of patients with TBI. Therefore, more predictive indicators can be added in future research. In addition to hospitalization time, consideration should be given to increasing the treatment cost, readmission rate, rehabilitation time, and so on. Through these indicators, patients' rehabilitation status can be more comprehensively reflected, helping doctors make better treatment decisions, and at the same time, precise medical plans can be formulated to provide better support and assistance for patients' rehabilitation.

In addition to the fact that the system discussed above can only predict the length of hospital stay of patients with traumatic brain injury, the system itself still has certain limitations. First, the prediction accuracy of the system is too dependent on data quality, and the prediction accuracy of machine learning models depends on the quality of the data used. If the data quality is poor, it may affect the accuracy and stability of the model. Second, due to incomplete and insufficient data collection, there may be deviations in the data. This may affect the prediction results of the model and lead to misjudgment. Finally, machine learning models are often difficult to explain, which may lead to doctors not trusting the model's prediction results. Therefore, in order to avoid data quality issues and biases, it is necessary to use multisource, multicenter, and diverse data as much as possible to train the model. In addition, data augmentation techniques can be used to expand the scale and diversity of the dataset, in order to improve the generalization ability of the model. In addition, to improve the interpretability of the model, interpretable machine learning techniques such as decision trees, rule learning, linear regression, and other models can be used to construct predictive models. At the same time, methods such as model visualization and feature importance analysis can be used to explain the predicted results of the model, enabling doctors to better understand the model's decisions. In a word, to solve the limitations of the system, we need to combine domain knowledge and data analysis and mining technology to constantly optimize and adjust the model to meet the actual clinical needs.

Conclusions

This study successfully developed a prediction model based on SVR by applying machine learning methods, which can accurately predict the hospitalization time of patients with TBI. This achievement not only demonstrates the strong potential of machine learning in the field of medical prediction but also provides strong support for clinical practice. By analyzing the medical data of patients in depth, the model can capture key factors that affect hospitalization duration and make accurate predictions based on them. This prediction system not only helps doctors better plan patient treatment and rehabilitation plans but also helps optimize the allocation of medical resources and improve the efficiency and quality of medical services. Therefore, the prediction system developed in this study has strong clinical practicality and is expected to become an important auxiliary tool for medical decision-making in the future, bringing patients a more personalized and efficient treatment experience.

Acknowledgments

We would like to thank Bing Zhao for his assistance and guidance in this research. This study was funded by Natural Science Research Project of Anhui Province (KJ2021A1160), Quality Engineering Project of Anhui Province (2022sx060; 2022jyxm659), Natural Science Foundation of the Anhui Xinhua University (no. 2023zr013), and Anhui Provincial University Students Innovation and Entrepreneurship Program (S202212216037; S202212216022; S202312216119).

Authors' Contributions

YP and HZ researched literature and conceived the study. YP and CF were involved in protocol development, gaining ethical approval, patient recruitment, and data analysis. HZ wrote the first draft of the manuscript and is the guarantor. All authors reviewed and edited the manuscript and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Pilitsis JG, Rengachary SS. Complications of head injury. Neurol Res 2001 Mar;23(2-3):227-236. [doi: 10.1179/016164101101198389] [Medline: 11320604]
2. Iverson GL, Lange RT. Mild traumatic brain injury. In: Schoenberg M, Scott J, editors. The Little Black Book of Neuropsychology: A Syndrome-Based Approach: Springer; 2010:697-719. [doi: 10.1007/978-0-387-76978-3_22]
3. Guerra WK, Gaab MR, Dietz H, Mueller JU, Piek J, Fritsch MJ. Surgical decompression for traumatic brain swelling: indications and results. J Neurosurg 1999 Feb;90(2):187-196. [doi: 10.3171/jns.1999.90.2.0187] [Medline: 9950487]
4. Carroll LJ, Cassidy JD, Holm L, Kraus J, Coronado VG, WHO Collaborating Centre Task Force on Mild Traumatic Brain Injury. Methodological issues and research recommendations for mild traumatic brain injury: the WHO Collaborating Centre Task Force on Mild Traumatic Brain Injury. J Rehabil Med 2004 Feb;36(43):113-125. [doi: 10.1080/16501960410023877] [Medline: 15083875]

5. Salmond SW, Echevarria M. Healthcare transformation and changing roles for nursing. Orthop Nurs 2017 Jan;36(1):12-25. [doi: 10.1097/NOR.0000000000000308] [Medline: 28107295]
6. Chiles JA, Lambert MJ, Hatch AL. The impact of psychological interventions on medical cost offset: a meta-analytic review. Clin Psychol Sci 1999 Jun;6(2):204-220. [doi: 10.1093/clipsy.6.2.204]
7. Wu SW, Pan Q, Chen T. Research on diagnosis-related group grouping of inpatient medical expenditure in colorectal cancer patients based on a decision tree model. World J Clin Cases 2020 Jun 26;8(12):2484-2493. [doi: 10.12998/wjcc.v8.i12.2484] [Medline: 32607325]
8. Fang C, Pan Y, Zhao L, Niu Z, Guo Q, Zhao B. A machine learning-based approach to predict prognosis and length of hospital stay in adults and children with traumatic brain injury: retrospective cohort study. J Med Internet Res 2022 Dec 9;24(12):e41819. [doi: 10.2196/41819] [Medline: 36485032]
9. El Naqa I, Bradley J, Blanco AI, et al. Multivariable modeling of radiotherapy outcomes, including dose–volume and clinical factors. Int J Radiat Oncol Biol Phys 2006 Mar;64(4):1275-1286. [doi: 10.1016/j.ijrobp.2005.11.022]
10. Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eDoctor: machine learning and the future of medicine. J Intern Med 2018 Dec;284(6):603-619. [doi: 10.1111/joim.12822] [Medline: 30102808]
11. Horn SD, DeJong G, Deutscher D. Practice-based evidence research in rehabilitation: an alternative to randomized controlled trials and traditional observational studies. Arch Phys Med Rehabil 2012 Aug;93(8):S127-S137. [doi: 10.1016/j.apmr.2011.10.031] [Medline: 22840879]
12. Liu C, Zhao Z, Gu X, et al. Establishment and verification of a bagged-trees-based model for prediction of sentinel lymph node metastasis for early breast cancer patients. Front Oncol 2019 Apr;9:282. [doi: 10.3389/fonc.2019.00282] [Medline: 31041192]
13. Bai ZH, Guo XQ, Dong R, Lei N, Pei HH, Wang H. A nomogram to predict the 28-day mortality of critically ill patients with acute kidney injury and treated with continuous renal replacement therapy. Am J Med Sci 2021 May;361(5):607-615. [doi: 10.1016/j.amjms.2020.11.028] [Medline: 33288206]
14. Larose DT, Larose CD. An introduction to data mining. In: Discovering Knowledge in Data: John Wiley & Sons; 2014.
15. Ali J, Khan R, Ahmad N, et al. Random forests and decision trees. Int J Comput Sci Issues 2012 Sep;9(5):272.
16. Wang P, Fan E, Wang P. Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. Pattern Recognit Lett 2021 Jan;141:61-67. [doi: 10.1016/j.patrec.2020.07.042]
17. Yang J, Li J. Application of deep convolution neural network. Presented at: 2017 14th International Computer Conference on Wavelet Active Media Technology and Information Processing; Dec 15-17, 2017; Chengdu, China. [doi: 10.1109/ICCWAMTIP.2017.8301485]
18. Yao X, Wang X, Wang SH, Zhang YD. A comprehensive survey on convolutional neural network in medical image analysis. Multimedia Tools Applications 2022 Dec;81(29):41361-41405. [doi: 10.1007/s11042-020-09634-7]
19. De Leone R, Pietrini M, Giovannelli A. Photovoltaic energy production forecast using support vector regression. Neural Comput Applic 2015 Nov;26(8):1955-1962. [doi: 10.1007/s00521-015-1842-y]
20. Drucker H, Burges CJC, Kaufman L, Smola A, Vapnik V. Support vector regression machines. In: Mozer MC, Jordan M, Petsche T, editors. Advances in Neural Information Processing Systems 9 (NIPS 1996): The MIT Press; 1996.
21. Song X, Mitnitski A, Cox J, Rockwood K. Comparison of Machine Learning Techniques With Classical Statistical Models in Predicting Health Outcomes: IOS Press; 2004:736-740. [doi: 10.3233/978-1-60750-949-3-736]
22. Kurt I, Ture M, Kurum AT. Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease. Expert Syst Appl 2008 Jan;34(1):366-374. [doi: 10.1016/j.eswa.2006.09.004]
23. Miravitlles M, Guerrero T, Mayordomo C, Sánchez-Agudo L, Nicolau F, Segú JL. Factors associated with increased risk of exacerbation and hospital admission in a cohort of ambulatory COPD patients: a multiple logistic regression analysis. Respiration 2000 Oct;67(5):495-501. [doi: 10.1159/000067462] [Medline: 11070451]
24. Li JS, Tian Y, Liu YF, Shu T, Liang MH. Applying a BP neural network model to predict the length of hospital stay. In: Huang G, Liu X, He J, Klawonn F, Yao G, editors. Health Information Science: Second International Conference, HIS 2013, London, UK, March 25-27, 2013 Proceedings: Springer:18-29. [doi: 10.1007/978-3-642-37899-7_2]
25. Battula BP, Balaganesh D. Prediction of hospital re-admission using firefly based multi-layer perceptron. Ingénierie Systèmes Inf 2020 Sep 20;25(4):527-533. [doi: 10.18280/isi.250417]
26. Khan RU, Almakdi S, Alshehri M, et al. Probabilistic approach to COVID-19 data analysis and forecasting future outbreaks using a multi-layer perceptron neural network. Diagnostics (Basel) 2022 Oct 19;12(10):2539. [doi: 10.3390/diagnostics12102539] [Medline: 36292228]
27. Bagnasco A, Fresi F, Saviozzi M, Silvestro F, Vinci A. Electrical consumption forecasting in hospital facilities: an application case. Energy Build 2015 Sep;103:261-270. [doi: 10.1016/j.enbuild.2015.05.056]
28. Shi HY, Lee KT, Lee HH, et al. Comparison of artificial neural network and logistic regression models for predicting in-hospital mortality after primary liver cancer surgery. PLoS One 2012;7(4):e35781. [doi: 10.1371/journal.pone.0035781] [Medline: 22563399]
29. Wu C, Zha D, Gao H. Prediction of bronchopneumonia inpatients' total hospitalization expenses based on BP neural network and support vector machine models. Comput Math Methods Med 2022 May 18;2022:9275801. [doi: 10.1155/2022/9275801] [Medline: 35633928]

30. Stone K, Zwiggelaar R, Jones P, Mac Parthaláin N. A systematic review of the prediction of hospital length of stay: towards a unified framework. PLOS Dig Health 2022 Apr 14;1(4):e0000017. [doi: 10.1371/journal.pdig.0000017] [Medline: 36812502]
31. Hong S, Lee S, Lee J, Cha WC, Kim K. Prediction of cardiac arrest in the emergency department based on machine learning and sequential characteristics: model development and retrospective clinical validation study. JMIR Med Inform 2020 Aug 4;8(8):e15932. [doi: 10.2196/15932] [Medline: 32749227]
32. Lu YT, Chao HJ, Chiang YC, Chen HY. Explainable machine learning techniques to predict amiodarone-induced thyroid dysfunction risk: multicenter, retrospective study with external validation. J Med Internet Res 2023 Feb 7;25:e43734. [doi: 10.2196/43734] [Medline: 36749620]
33. Xiao J, Mo M, Wang Z, et al. The application and comparison of machine learning models for the prediction of breast cancer prognosis: retrospective cohort study. JMIR Med Inform 2022 Feb 18;10(2):e33440. [doi: 10.2196/33440] [Medline: 35179504]

Abbreviations

BP: back propagation
CNN: convolutional neural network
LR: logistic regression
MAPE: mean absolute percentage error
MLP: multilayer perceptron
NLP: natural language processing
RF: random forest
SVR: support vector regression
TBI: traumatic brain injury

Edited by A Kushniruk; submitted 03.06.24; peer-reviewed by T Cao, W Shen; revised version received 19.06.24; accepted 22.06.24; published 30.08.24.

Please cite as:
Zhou H, Fang C, Pan Y
Development of a System for Predicting Hospitalization Time for Patients With Traumatic Brain Injury Based on Machine Learning Algorithms: User-Centered Design Case Study
JMIR Hum Factors 2024;11:e62866
URL: https://humanfactors.jmir.org/2024/1/e62866
doi:10.2196/62866

A/B Testing of User Enrollment Forms to Enhance Diversity in the Biomedical Workforce via the National Research Mentoring Network: User-Centered Design Case Study

Toufeeq Ahmed Syed[1*], MS, PhD; Erika L Thompson[2*], PhD; Jason Johnson[1*], HS; Zainab Latif[3*], MCS; Nan Kennedy[4*], MLAS; Damaris Javier[5*], MA; Katie Stinson[1*], MLIS; Jamboor K Vishwanatha[5*], PhD

[1]University of Texas Health Science Center at Houston, Houston, TX, United States
[2]University of Texas Health Science Center at San Antonio, San Antonio, TX, United States
[3]Northwestern University, Evanston, IL, United States
[4]Vanderbilt University Medical Center, Nashville, TN, United States
[5]University of North Texas Health Science Center at Fort Worth, Fort Worth, TX, United States
[*]all authors contributed equally

Corresponding Author:
Toufeeq Ahmed Syed, MS, PhD
University of Texas Health Science Center at Houston, , Houston, TX, , United States

Abstract

Background: The National Research Mentoring Network (NRMN) is a National Institutes of Health–funded program for diversifying the science, technology, engineering, math, and medicine research workforce through the provision of mentoring, networking, and professional development resources. The NRMN provides mentoring resources to members through its online platform—MyNRMN.

Objective: MyNRMN helps members build a network of mentors. Our goal was to expand enrollment and mentoring connections, especially among those who have been historically underrepresented in biomedical training and the biomedical workforce.

Methods: To improve the ease of enrollment, we implemented the split testing of iterations of our user interface for platform registration. To increase mentoring connections, we developed multiple features that facilitate connecting via different pathways.

Results: Our improved user interface yielded significantly higher rates of completed registrations ($P<.001$). Our analysis showed improvement in completed enrollments that used the version 1 form when compared to those that used the legacy form (odds ratio 1.52, 95% CI 1.30-1.78). The version 2 form, with its simplified, 1-step process and fewer required fields, outperformed the legacy form (odds ratio 2.18, 95% CI 1.90-2.50). By improving the enrollment form, the rate of MyNRMN enrollment completion increased from 57.3% (784/1368) with the legacy form to 74.5% (2016/2706) with the version 2 form. Our newly developed features delivered an increase in connections between members.

Conclusions: Our technical efforts expanded MyNRMN's membership base and increased connections between members. Other platform development teams can learn from these efforts to increase enrollment among underrepresented groups and foster continuing, successful engagement.

(JMIR Hum Factors 2024;11:e54532) doi:10.2196/54532

KEYWORDS

diversity; mentoring; health workforce; underrepresented groups; online platform; user-computer interface; A/B testing; split testing; recommendation algorithm; network of mentors; groups; enrollment

Introduction

The need for mentoring networks is critical in the biomedical sciences, where the shortage of available mentors contributes to a scarcity of students pursuing biomedical careers. Funded by the National Institutes of Health (NIH) in 2014, the mission of the National Research Mentoring Network (NRMN) is to increase opportunities for mentorship and professional development in the biomedical sciences, especially among underrepresented populations. As part of the NIH-sponsored Diversity Program Consortium, the program works to diversify the biomedical research workforce through programming, training, mentorship, and advocacy efforts [1,2].

The MyNRMN platform [3] enables mentors and mentees nationwide to connect one-on-one or in groups with mentors and peer mentors. As of April 2024, MyNRMN has grown to include over 30,000 members from across the nation, engaging more than 15,700 mentees and 8400 mentors who represent more than 4000 institutions and organizations from all 50 US states and US territories. Within the platform, more than 7800 learners have taken online courses and training and have created more than 500 online collaboration and discussion groups with over 9900 participants. There have been more than 12,000 connections made between mentees and mentors in the network, with strong representation from underrepresented faculty, researchers, and students. To expand the reach of this mentorship platform, the MyNRMN development team continually revises the user interface and platform design and adds new features.

The MyNRMN platform faced the following two challenges: increasing enrollment and expanding network connections. Pivotal aspects of creating an online network are continually increasing the number of people in the network and providing opportunities for members to make connections with others and increase engagement. To increase enrollment, especially the enrollment of diverse users, our team sought to improve the ease and appeal of the enrollment process by using split testing (ie, A/B testing)—a method that is used widely to compare two variations of a form or content [4-9]. For web forms, this testing is accomplished by randomly serving one version of a form to half of the platform visitors and a different version to the other half. A post analysis reveals which variation is the most effective in motivating users to complete a task (eg, enrollment). Additionally, we designed and implemented an array of features to facilitate mentoring connections. This paper details the strategies and underlying technology that supported our efforts. By empowering members from diverse backgrounds to join MyNRMN and connect, our platform expands personal networks and increases career opportunities for those who have been traditionally underrepresented in the biomedical sciences.

In this paper, we describe how we conducted the iterative split testing of three enrollment forms to determine how to increase enrollment in a nationwide networking platform for the biomedical workforce. We also discuss the additional mentoring, networking, and professional development features that we created to further encourage enrollment and increase engagement in the platform.

Methods

Overview of MyNRMN

MyNRMN was developed by TAS in 2016 as part of the NRMN. Various forms of individual guided mentoring, free-form mentoring, and group mentoring are offered by the NRMN [1,2]. Prior to the development of MyNRMN, the NRMN online resources consisted of a website with information about the program. Since 2016, the MyNRMN platform has supported many online features and resources for mentors and mentees to connect, seek peer or group mentoring, provide guided online mentoring, and create a network of mentors.

Enrollment—Attracting Individuals to Join MyNRMN

The first step to finding a mentor was joining MyNRMN. To begin the enrollment process, the users registered by using a third-party service—Auth0 (Okta Inc)—to authenticate their sign-in information via Gmail, LinkedIn, Facebook, or email address and password. This service was especially beneficial when integrating the social sign-ins, particularly while modifying the required registration fields throughout this study.

To finish the process, users completed an online enrollment form that included demographic information. An enrollment form was already used at the time Vanderbilt University Medical Center inherited the MyNRMN registration process during the Phase II U24 award period. This “legacy” form, which was designed primarily for data collection, included 19 fields arranged in 2 columns, with all fields required to complete enrollment (Figure 1). The lengthy slate of questions contained potentially sensitive fields, such as fields for ethnicity, disability, and gender. We observed that many users dropped out of the enrollment process after creating an Auth0 account and before completing the form. We became concerned that having sensitive questions early in the process hindered our efforts to recruit diverse members. An additional fear was losing potential members by not offering alternative log-in methods, particularly logging in via social media, which younger users may prefer.

In hopes of increasing enrollment completion, we created the version 1 form and tested it against the original legacy form. The version 1 form asked the same questions in an improved user interface. In contrast to the legacy form, the version 1 form used a wizard format with 3 tabs. The first 2 tabs presented 11 required fields, while the third tab contained optional fields (Figure 2). When implementing the multipage format, we anticipated reduced perceptions of the form being burdensome.

Figure 1. Legacy enrollment form. NRMN: National Research Mentoring Network.

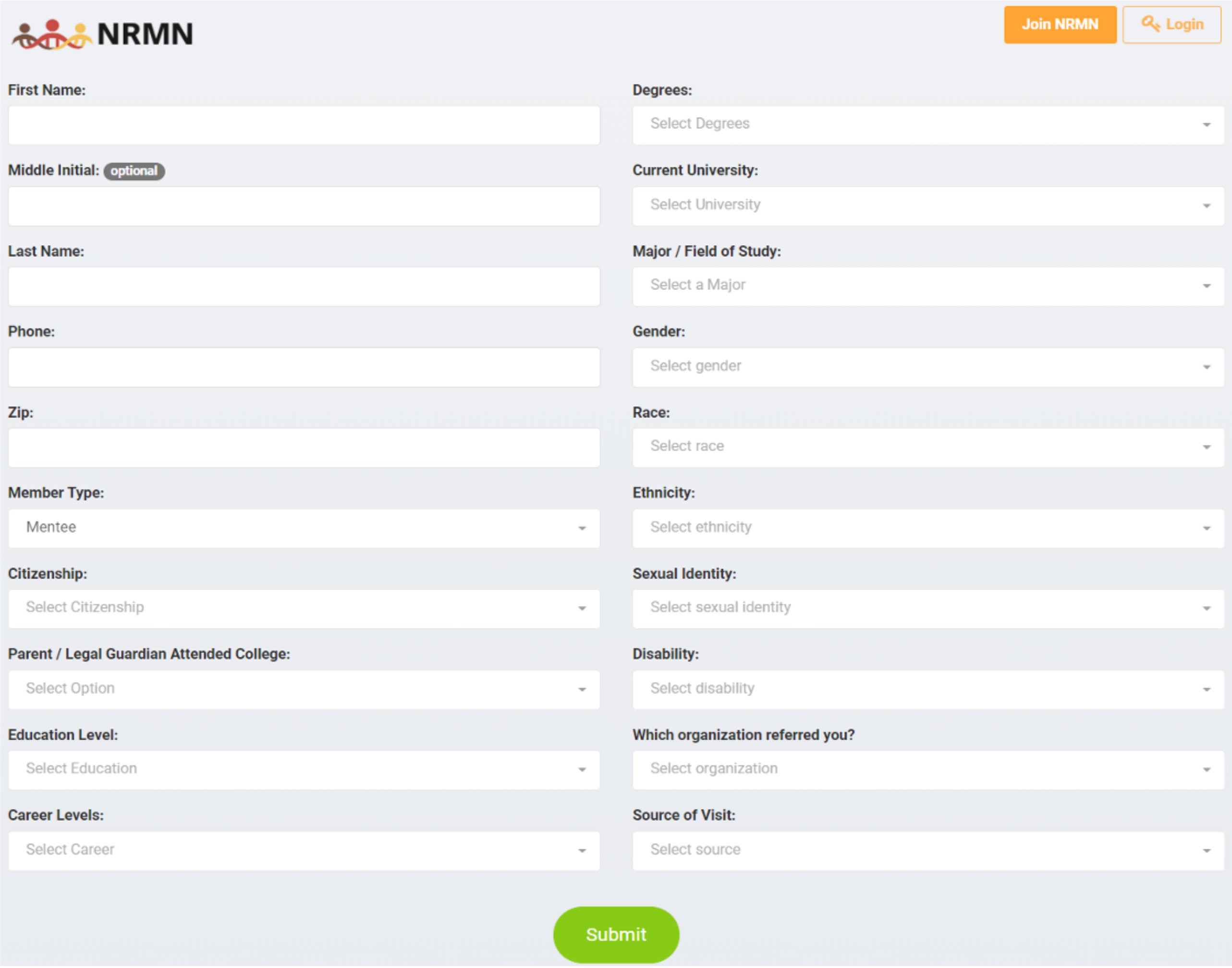
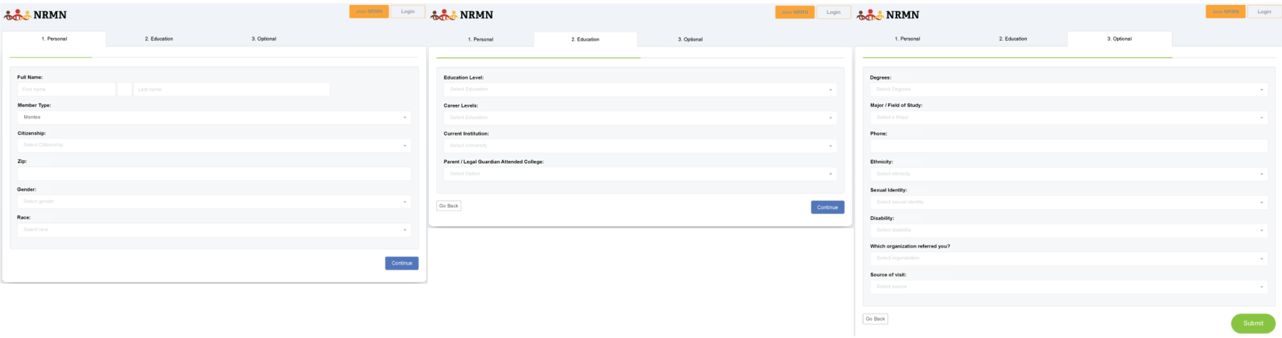
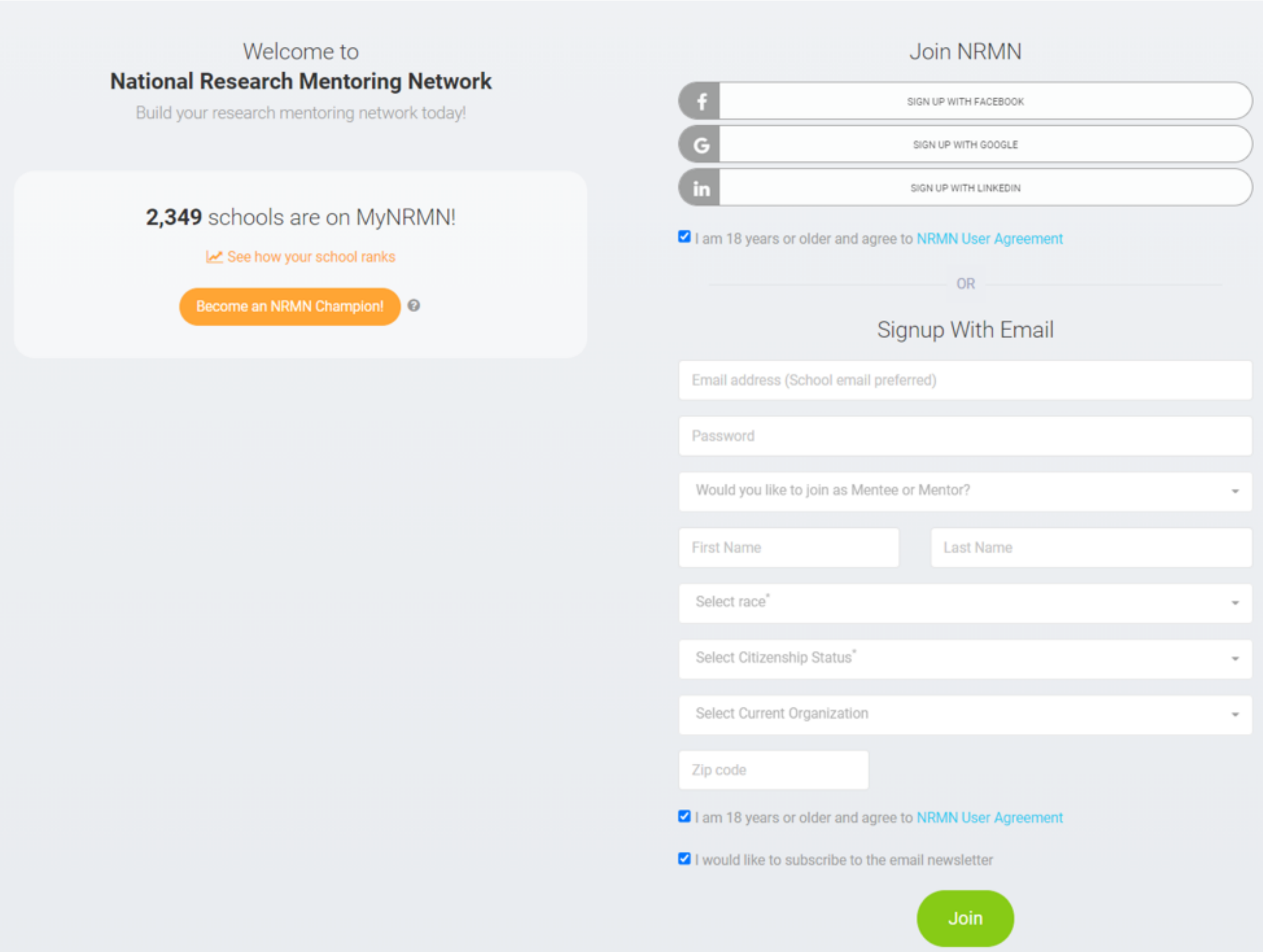


Figure 2. Version 1 enrollment form. NRMN: National Research Mentoring Network.



From July 1, 2019, to August 26, 2020, we used split testing to evaluate the different registration forms to optimize enrollment. To conduct split testing, we programmed the system to randomly redirect each user, after Auth0 sign-in, to 1 of 2 enrollment forms—the legacy form or the version 1 form. Despite seeing some improvement in enrollment among users who were provided with the version 1 form, the improvement was not as significant as anticipated. This motivated us to continue revising the form to develop the version 2 form (Figure 3). The version 2 form allowed users to create an Auth0 account and enroll synchronously. Sensitive questions were removed from the first steps in the version 2 form, and members were prompted, but not required, to complete these questions after enrollment. We observed that removing the requirement to complete the sensitive fields prior to enrollment significantly increased user enrollment. The NRMN's goal is to increase diversity in the biomedical workforce. To measure the impact of our mentoring platform, we collect demographic data during the user enrollment process. Beginning with the version 2 form, the only demographic information that has been collected during enrollment is race. Since August 26, 2020, all new users have been redirected to the version 2 form, which replaced the legacy and version 1 forms.

Figure 3. Version 2 enrollment form. NRMN: National Research Mentoring Network.



Connections—Finding a Mentor and Building a Personal Network of Mentors

The MyNRMN platform's goal is to help members increase connections to build and expand their networks. The larger the network, the greater the opportunity to engage with members who are diverse with respect to gender, racial and ethnic background, and other demographics. To facilitate finding and acquiring new connections on MyNRMN, we developed the following new features and modes:

1. Member search engine: Members use the powerful search engine within the *Find a Mentor* feature to search for mentors or peer mentors. Keywords are used to search by name, institution, location, or areas of interest. Additionally, the keyword search is applied to curricula vitae, résumés, and publications that are synced within the platform via natural language processing–based indexing and retrieval.
2. Recommendation engine: Through the use of Neo4j (Neo4j Inc)—a robust graph database technology—the platform's recommendation algorithms suggest new connections to members. Fresh recommendations appear when a member accesses their personal dashboard.
3. Profile: A member's profile page contains detailed information, including research interests, institutions, locations, and publications, and excludes any profile questions that may be deemed "sensitive." A mentee can view the profile of any mentor and request to connect.

Data Analysis

We estimated descriptive statistics for variables in SAS 9.4 (SAS Institute Inc). To examine the effect of form type on form completion, odds ratios (ORs) and 95% CIs were estimated for odds of completing enrollment by form type and within demographic subgroups. Subgroup analyses were conducted for demographic subgroups with sufficient data available (eg, White, African American, and Asian race subgroups) and on the basis of which fields were required for enrollment completion. After implementing the version 2 form, some demographic fields were converted to optional fields, which resulted in some users leaving these fields blank, hence the incomplete ("Missing from Total") data in our data set.

Ethical Considerations

This study was approved by the NRMN's institutional review board (reference number: 2015 - 0720). All user data were protected under and provided by the North Texas Regional Institutional Review Board and stored securely.

Results

Growth in Enrollment

Prior to August 26, 2020, a user joining MyNRMN was required to complete the following two steps: (1) create an account with Auth0 and (2) fill in all required fields in the enrollment form. Our analysis showed improvement in completed enrollments that used the version 1 form when compared to those that used the legacy form (OR 1.52, 95% CI 1.30-1.78). The version 2 form, with its simplified, 1-step process and fewer required fields, outperformed the legacy form (OR 2.18, 95% CI 1.90-2.50).

Table 1 describes the proportions of users who completed enrollment by form type and demographic characteristics. Figure 4 displays the users who completed enrollment by form type and self-identified race. In the version 1 form, the field for ethnicity was optional. In the version 2 form, the fields for ethnicity, gender, and education were optional.

Table . Proportions of users who completed enrollment by form type.

		Legacy form (N=1368)	Version 1 form (N=1396)	Version 2 form (N=2706)
Overall completion, n (%[a])		784 (57.3)	938 (67.2)	2016 (74.5)
Race[b]				
	White, n (%)	361 (80)	420 (96.1)	958 (87.4)
	Black or African American, n (%)	165 (78.6)	234 (95.9)	395 (90.8)
	Asian, n (%)	128 (84.2)	125 (94)	338 (91.6)
	American Indian or Alaska Native, n (%)	15 (88.2)	7 (70)	21 (87.5)
	Native Hawaiian or Pacific Islander, n (%)	2 (50)	3 (100)	14 (93.3)
	Two or more, n (%)	15 (83.3)	28 (96.6)	42 (95.5)
	Other, n (%)	30 (71.4)	54 (93.1)	114 (89.1)
	Prefer not to answer, n (%)	68 (73.1)	66 (98.5)	135 (86)
	"Missing from Total," n	381	415	438
Ethnicity[b]				
	Non-Hispanic, n (%)	590 (80.4)	632 (95.9)	358 (90)
	Hispanic, n (%)	92 (78)	102 (82.9)	53 (85.5)
	Other, n (%)	31 (77.5)	30 (88.2)	17 (81)
	Prefer not to report, n (%)	71 (80.7)	62 (95.4)	27 (90)
	"Missing from Total," n	388	515[c]	2195[c]
Role[b]				
	Mentee, n (%)	504 (76.4)	602 (86.4)	1313 (97.8)
	Mentor, n (%)	280 (75.7)	335 (85.2)	704 (98.6)
	"Missing from Total," n	338	306	649
Log-in method[b], n (%)				
	Username-password	556 (58.5)	650 (70.4)	1302 (83.9)
	Google	156 (54.7)	170 (58)	446 (56.7)
	Facebook	20 (48.8)	30 (57.7)	20 (45.5)
	LinkedIn	52 (57.1)	87 (68.5)	249 (77.1)
Gender[b]				
	Male, n (%)	236 (81.1)	260 (92.9)	168 (89.8)
	Female, n (%)	526 (77.7)	663 (92.3)	419 (89.2)
	Other, n (%)	2 (100)	3 (100)	7 (100)
	Prefer not to report, n (%)	20 (74.1)	11 (84.6)	2 (100)
	"Missing from Total," n	371	382	2040[c]
Education[b]				
	Undergraduate, n (%)	171 (84.2)	202 (87.5)	85 (84.2)
	Nondegree postbaccalaureate, n (%)	22 (88)	17 (89.5)	8 (88.9)
	Graduate, n (%)	153 (73.9)	175 (90.7)	95 (88)
	Postdoc, n (%)	114 (81.4)	127 (89.4)	89 (89.9)
	Other, n (%)	324 (76.1)	416 (86.1)	228 (91.2)

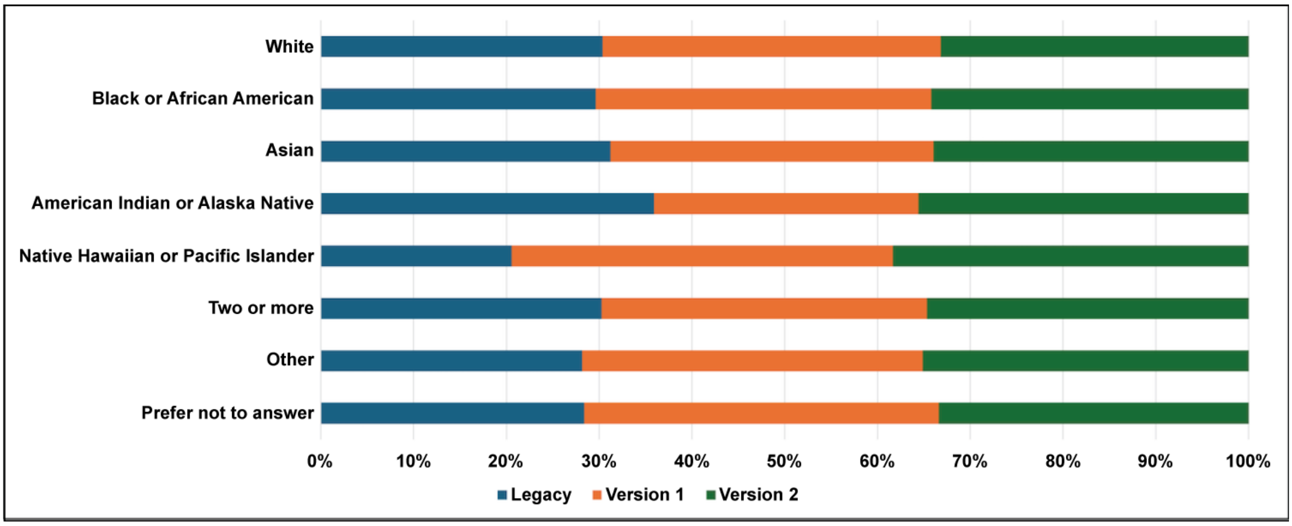
	Legacy form (N=1368)	Version 1 form (N=1396)	Version 2 form (N=2706)
"Missing from Total," n	367	328	2139[c]

[a]Percentages were calculated by using the N values in the corresponding row headings as the denominator.

[b]The percentages in these rows were calculated by using the total number of users who self-identified as each subgroup for each form type as the denominator.

[c]Became an optional field; this "Missing from Total" value includes users who completed enrollment and users who did not complete enrollment.

Figure 4. Proportions of users who completed the enrollment form by form type and race.



We performed subgroup analyses to determine if enrollment completion differed within demographic subgroups by form type (Table 2). We compared the legacy and version 1 forms for gender, race, education, and log-in method subgroups, since these were required fields. The version 1 form performed better than the legacy form for male users, female users, White users, Black or African American users, Asian users, graduate students, nonstudents, and persons who used the username-password method. We compared the version 2 and legacy forms for race and log-in method subgroups, since these were required fields. The version 2 form performed better than the legacy form when users used the username-password method (version 1: OR 1.69, 95% CI 1.39-2.04; version 2: OR 3.70, 95% CI 3.07-4.46) or LinkedIn (version 1: OR 1.63, 95% CI 0.93-2.85; version 2: OR 2.52, 95% CI 1.55-4.12). The version 2 form also performed better than the legacy form in terms of facilitating enrollment completion among White, Black or African American, and Asian users.

Table . Odds ratios (ORs) and 95% CIs for enrollment completion within demographic subgroups.

		Version 1 form vs legacy form (reference), OR (95% CI)	Version 2 form vs legacy form (reference)[a], OR (95% CI)
Gender			
	Male	3.03 (1.76-5.21)[b]	—[c]
	Female	3.46 (2.49-4.81)[b]	—
Race			
	White	6.16 (3.60-10.54)[b]	1.73 (1.29-2.32)[b]
	Black or African American	6.38 (3.13-13.03)[b]	2.69 (1.70-4.28)[b]
	Asian	2.93 (1.27-6.77)[b]	2.04 (1.16-3.62)[b]
Education			
	Undergraduate	1.30 (0.76-2.24)	—
	Graduate	3.43 (1.93-6.10)[b]	—
	Postdoc	1.93 (0.97-3.83)	—
	Other	1.96 (1.39-2.75)[b]	—
Log-in method			
	Username-password	1.69 (1.39-2.04)[b]	3.70 (3.07-4.46)[b]
	Google	1.14 (0.82-1.59)	1.08 (0.82-1.42)
	Facebook	1.43 (0.63-3.26)	0.88 (0.37-2.05)
	LinkedIn	1.63 (0.93-2.85)	2.52 (1.55-4.12)[b]

[a]Gender and education were not required fields in the version 2 form.
[b]Indicates statistical significance (*P*<.05).
[c]Not applicable.

Growth in Connections

After implementing the version 2 form, the amount of mentoring requests grew for all types of mentor-mentee connections, including peer-to-peer connections (Figure 5). However, the percentage of accepted connection requests has remained relatively consistent. Of the total requests to connect, roughly half have been accepted. Although very few requests have been actively declined, a large percentage remain pending.

Networking connections originated via several pathways (Table 3). Both mentors and mentees made extensive use of the *Recommendations* feature that suggests connections on the user dashboard. Via the *Search* and *Profile* connection pathways, mentees initiated more connections than mentors.

Figure 5. Cumulative mentoring connections.

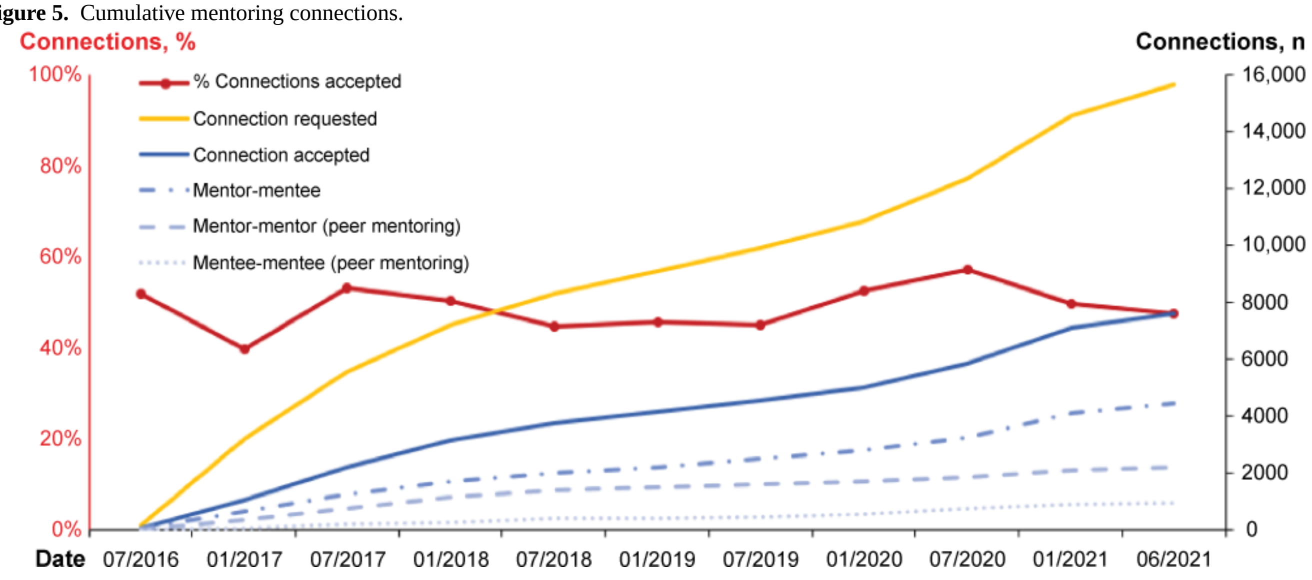


Table . Origins of connection requests.

Pathway	Mentees		Mentors	
	Number of connection requests	Requests accepted, n (%)	Number of connection requests	Requests accepted, n (%)
Recommendations feature	601	231 (38.4)	724	285 (39.4)
Search connection pathway	577	277 (48)	331	171 (51.7)
Profile connection pathway	633	320 (50.6)	148	74 (50)

Discussion

Principal Findings

Providing access to mentors and meeting the need for mentoring are essential to grow and enhance the biomedical workforce, and the value of mentorship in accessing the connections needed to advance professionally has been long established [10,11]. Research shows that the probability of individual success increases when multiple mentors with differing backgrounds and levels of expertise are tapped for support [12]. This *network of mentors* facilitates growth that is tailored to an individual's goals by providing the multiple perspectives needed for professional and personal growth [13-15].

Our results show that our enrollment forms work similarly well for female and male users, and the version 1 form performed better (enrollment rate of male users: 260/280, 92.9%; enrollment rate of female users: 663/718, 92.3%) for both groups than the legacy form (enrollment rate of male users: 236/291, 81.1%; enrollment rate of female users: 526/677, 77.7%). This is an important consideration in decreasing traditional gender disparities in the biomedical sciences—a situation that the NIH has recognized by designating female individuals as underrepresented. Although the version 1 form was established as superior to the legacy form in terms of enrollment completion, it did not perform as anticipated. We discussed the shortcomings of the existing forms with MyNRMN stakeholders, including the NRMN-Resource Center, the Product Council (which included representative NRMN members), and the NIH project scientist. After collaborative discussions, we simplified the enrollment process in the version 2 form. We discovered that the design of the user interface did not negatively affect enrollment, which was instead affected by the length and sensitive nature of the information required. After reducing the number of required questions and moving sensitive questions to an optional postenrollment form, our rate of registration completion increased significantly (P<.001; Table 1). Moreover, improved registration completion was observed for racial subgroups that used the version 2 form when compared to those that used the legacy form.

In fostering mentoring connections, we found that most mentors (724/1203, 60.2%) extensively used more passive features, such as the *Recommendations*. In contrast, mentees were more likely to proactively use the *Search* feature or the *Profile* feature to build connections when compared to mentors. Both of these features use the personalized recommendation algorithm and natural language processing when providing results. The recommendation algorithm pulls information from a user's profile, as well as from their network connections within the platform. To promote the further expansion of personal networks, we developed the following five new features: (1) the *Groups* feature (members can join discussion and collaboration groups to connect with others who shared common interests); (2) the *Ask Me A Question* feature (members can ask questions prior to requesting a connection and then initiate the connection request); (3) the *Invites* feature (members can invite peers, colleagues, or friends to join MyNRMN and connect); (4) the *My Cohort* feature (institutions and organizations can bring their members to the platform to connect within the cohort and throughout MyNRMN); and (5) the *Administrative Match* feature (mentees can request a system administrator to initiate a mentoring request to a mentor). Because these features are relatively new, we did not include them in our results. Favorable initial data, however, demonstrate the value of continually adding novel features. Providing the online platform to facilitate mentoring connections, in conjunction with providing additional professional development resources and multiple, easy-to-use enrollment options, has aided the growth of MyNRMN, which now has 30,000 users (as of April 2024). Developing new strategies to promote networking connections may further support MyNRMN platform engagement for additional resources.

We were successful in increasing enrollment on our platform and understanding users' preferences for the enrollment form. However, we identified a limitation in our A/B testing process, and we believe that a longer testing period and more iterations could provide us with more data for enhancing the user enrollment process. In our future work, we will aim to study how we can encourage participation from underrepresented groups, such as Native Hawaiian or Pacific Islander individuals or people who do not identify within the gender binary.

Future Directions

The MyNRMN platform's goal is to ensure that mentoring connection requests are seen and are acted upon. Requests to connect are too frequently languishing, especially those made through the *Recommendations* feature; these may appear less personally targeted than those made through other features. Therefore, we are examining how to strengthen the relevancy of our recommendation algorithm by using machine learning and our graph database for more tailored recommendations. To nudge those receiving a connection request, the platform will remind members immediately upon logging in and will also send email reminders. To ensure that the enrollment and connection emails are received, we will continually monitor spam filter rules and comply with changes and updates that are implemented by email services. In addition, pending requests for connections will be displayed prominently on the dashboard or as a pop-up.

Our goal is to quickly increase each new member's network from the initial node of 1 to an ever-widening web of connections. We plan to begin including embedded recommendations in a monthly email. In addition, when a connection is accepted, the confirmation email will include additional recommendations.

Conclusion

We built a platform for online mentoring to meet the needs of our members and to add value by increasing their connections. By conducting the A/B testing of enrollment forms, we were able to identify and overcome a barrier to enrollment and thus provide mentoring, networking, and professional development resources to a broader audience, which in turn promotes the diversification of the biomedical workforce. We continue to evaluate our paradigm and improve our engagement. Many of our ideas are generalizable to those building other membership networks. These entities can learn from our experience that creating multiple pathways, such as by providing social media options for account creation, achieves better results than a single track and that removing sensitive questions, such as those about gender and sexual identity, can attract a more diverse membership. For MyNRMN, our key value proposition is increasing network connections, which we achieve through the technology solutions described herein. By making the enrollment process less onerous and sensitive, ensuring that new members feel instantly welcome, and constantly developing and implementing novel engagement features, MyNRMN fosters an environment that engages an increasingly diverse population of mentors and mentees.

Acknowledgments

The research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award numbers U54GM119023 and U24GM132217. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

References

1. Sorkness CA, Pfund C, Ofili EO, et al. A new approach to mentoring for research careers: the National Research Mentoring Network. BMC Proc 2017 Dec 4;11(Suppl 12):22. [doi: 10.1186/s12919-017-0083-8] [Medline: 29375663]
2. Javier D, Stinson K, Zavala M, Ahmed T, Vishwanatha JK. NRMNet: building a national resource for mentorship, networking and professional development to enhance diversity. Ethn Dis 2021 Jul 15;31(3):469-480. [doi: 10.18865/ed.31.3.469] [Medline: 34295135]
3. Ahmed T, Johnson J, Latif Z, et al. MyNRMN: a national mentoring and networking platform to enhance connectivity and diversity in the biomedical sciences. FASEB Bioadv 2021 Mar 31;3(7):497-509. [doi: 10.1096/fba.2020-00102] [Medline: 34258519]
4. Cruz-Benito J, Vazquez-Ingelmo A, Sanchez-Prieto JC, Theron R, Garcia-Penalvo FJ, Martin-Gonzalez M. Enabling adaptability in web forms based on user characteristics detection through A/B testing and machine learning. IEEE Access 2018;6:2251-2265. [doi: 10.1109/ACCESS.2017.2782678]
5. King R, Churchill EF, Tan C. Designing With Data: Improving the User Experience With A/B Testing: O'Reilly Media, Inc; 2017.
6. Xu Y, Chen N, Fernandez A, Sinno O, Bhasin A. From infrastructure to culture: A/B testing challenges in large scale social networks. In: KDD '15: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining: Association for Computing Machinery; 2015:2227-2236. [doi: 10.1145/2783258.2788602]
7. Kohavi R, Longbotham R. Online controlled experiments and A/B testing. In: Sammut C, Webb GI, editors. Encyclopedia of Machine Learning and Data Mining: Springer; 2017:922-929. [doi: 10.1007/978-1-4899-7687-1_891]
8. Speicher M, Both A, Gaedke M. Ensuring web interface quality through usability-based split testing. In: Casteleyn S, Rossi G, Winckler M, editors. Web Engineering: 14th International Conference, ICWE 2014, Toulouse, France, July 1-4, 2014, Proceedings: Springer; 2014. [doi: 10.1007/978-3-319-08245-5_6]
9. Fisher RA, Yates F. Statistical Tables for Biological, Agricultural and Medical Research: Oliver & Boyd; 1938.
10. Packard BWL. Successful STEM Mentoring Initiatives for Underrepresented Students: A Research-Based Guide for Faculty and Administrators: Stylus Publishing, LLC; 2016. [doi: 10.4324/9781003447245]
11. Haggard DL, Dougherty TW, Turban DB, Wilbanks JE. Who is a mentor? a review of evolving definitions and implications for research. J Manage 2011 Jan;37(1):280-304. [doi: 10.1177/0149206310386227]
12. McGee R, Lee S, Pfund C, Branchaw J. Beyond "Finding Good Mentors" to "Building and Cultivating Your Mentoring Team". In: Huang BL, editor. Advancing Postdoc Women Guidebook: National Postdoctoral Association; 2015:23-33.
13. National Academies of Sciences, Engineering, and Medicine; Policy and Global Affairs; Board on Higher Education and Workforce; Committee on Effective Mentoring in STEMM. Findings and recommendations. In: Dahlberg ML, Byars-Winston

A, editors. The Science of Effective Mentorship in STEMM: National Academies Press; 2019. [doi: 10.17226/25568] [Medline: 31958221]
14. Higgins MC, Kram KE. Reconceptualizing mentoring at work: a developmental network perspective. Acad Manage Rev 2001 Apr;26(2):264-288. [doi: 10.2307/259122]
15. Sorcinelli MD, Yun J. From mentor to mentoring networks: mentoring in the new academy. Change 2007;39(6):58-61. [doi: 10.3200/CHNG.39.6.58-C4]

Abbreviations

NIH: National Institutes of Health
NRMN: National Research Mentoring Network
OR: odds ratio

Edited by A Kushniruk, E Borycki; submitted 13.11.23; peer-reviewed by A Anastasiou, M Gasmi, S chaudhry, T Adebambo; revised version received 17.04.24; accepted 17.04.24; published 02.07.24.

Please cite as:
Syed TA, Thompson EL, Johnson J, Latif Z, Kennedy N, Javier D, Stinson K, Vishwanatha JK
A/B Testing of User Enrollment Forms to Enhance Diversity in the Biomedical Workforce via the National Research Mentoring Network: User-Centered Design Case Study
JMIR Hum Factors 2024;11:e54532
URL: https://humanfactors.jmir.org/2024/1/e54532
doi:10.2196/54532

Using a Human-Centered Design Process to Evaluate and Optimize User Experience of a Website (InPACT at Home) to Promote Youth Physical Activity: Case Study

Rebecca E Hasson[1], PhD; Michelle Xie[2], BS; Dhiraj Tadikamalla[2], BS; Lexie R Beemer[1], PhD

[1]School of Kinesiology, University of Michigan, Ann Arbor, MI, United States
[2]School of Information, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Rebecca E Hasson, PhD
School of Kinesiology, University of Michigan, , Ann Arbor, MI, , United States

Abstract

Background: Web-based physical activity interventions often fail to reach the anticipated public health impact due to insufficient use by the intended audiences.

Objective: The purpose of this study was to use a human-centered design process to optimize the user experience of the Interrupting Prolonged sitting with ACTivity (InPACT) at Home website to promote youth physical activity participation.

Methods: Qualitative interviews were conducted to assess engagement and pain points with the InPACT at Home website. Interview data were used to create affinity maps to identify themes of user responses, conduct a heuristic evaluation according to Nielsen's usability heuristics framework, and complete a competitive analysis to identify the strengths and weaknesses of competitors who offered similar products.

Results: Key themes from end user interviews included liking the website design, finding the website difficult to navigate, and wanting additional features (eg, library of watched videos). The website usability issues identified were lack of labeling and categorization of exercise videos, hidden necessary actions and options hindering users from decision-making, error-prone conditions, and high cognitive load of the website. Competitive analysis results revealed that YouTube received the highest usability ratings followed by the Just Dance and Presidential Youth Fitness Program websites.

Conclusions: Human-centered design approaches are useful for bringing end users and developers together to optimize user experience and impact public health. Future research is needed to examine the effectiveness of the InPACT at Home website redesign to attract new users and retain current users, with the end goal of increasing youth physical activity engagement.

(JMIR Hum Factors 2024;11:e52496) doi:10.2196/52496

KEYWORDS

web-based interventions; children; adolescents; child; adolescent; youth; user experience; website; websites; implementation science; human-centered design; human-centred design; HCD; web based; home based; interview; heuristics; interviews; heuristic; competitive analysis; video; videos; YouTube; physical activity; exercise; fitness

Introduction

Physical activity is one of the most efficacious pathways to promote child health, well-being, and academic achievement [1,2], yet most children and adolescents in the United States are classified as inactive. Less than half (42%) of children ages 6-11 years participate in the recommended 60 minutes of daily physical activity, and this percentage declines as children transition into adolescence [3-5]. Children living in low-resource communities report even lower rates of physical activity [6,7], and the recent COVID-19 pandemic exacerbated these disparities [8,9] by contributing to a 17-minute decline in youth physical activity [8-10]. We have an urgent and unmet need to increase youth physical activity engagement to improve child and adolescent health.

The COVID-19 pandemic also brought renewed attention to prioritizing virtual methods of physical activity promotion as families were sheltering in place and children were attending schools online [11]. Web-based interventions have the potential to improve youth physical activity participation because of their extensive reach, high convenience, immediate feedback, diverse delivery formats, anonymity, and use across different contexts [12-14]. Web-based interventions can reach children and adolescents nearly anywhere at any time through desktops, laptops, and mobile devices [15,16]. Because this generation of children and adolescents spend large amounts of time watching or using screens (4‐6 hours per day for children ages 8‐12 years and up to 9 hours for teens) [17], web-based interventions represent a feasible and accessible strategy to promote youth physical activity engagement.

Evidence is lacking, however, for achieving sustainable physical activity behavior change through the internet. A recent review of web-based physical activity interventions highlighted that despite large developments in internet technology and knowledge of how to design and implement web-based physical activity interventions, website quality remains low [18]. These websites also provided limited social support and educational content [18].

Families play a crucial role in shaping a child's activity levels by providing various forms of social support [19,20]. This support includes encouragement, participating in activities together, and observing a family member's involvement in physical activities or sports. For instance, Tandon et al [21] found that in predominantly White households, parental support was linked to an extra 12 minutes of moderate-to-vigorous physical activity per day. Similarly, a study by Graham et al [22] in a diverse population showed that parental support, including role modeling, influenced adolescent activity levels.

Our Interrupting Prolonged sitting with ACTivity (InPACT) at Home program, for example, confirmed the importance of parental support. InPACT at Home was a television and web-based intervention designed to help elementary and secondary school students in kindergarten through grade 12 (K-12) stay physically active and maintain healthy lifestyles during the COVID-19 pandemic [23]. Our preliminary research assessing the feasibility of children using InPACT at Home exercise videos demonstrated that parents encouraging their children, reminding them, and establishing schedules and routines significantly facilitated participation in the program (LR Beemer et al, unpublished data, 2024). These findings underscore the pivotal role of parents in promoting virtual home-based physical activity for youth. Consequently, the design of the InPACT at Home website targets parents, where parental engagement leads to parent support and subsequent increased participation by their children in the exercise videos.

Problems associated with attracting, engaging, and retaining participants in web-based interventions have also been observed [24,25]. While the reach of the InPACT at Home program through public television broadcasting averaged 15,000-20,000 daily viewers, there were only 23 registered users on the program website 1 year after the program launch (website data unpublished). These findings illustrated the potential need for enhanced website design quality and the incorporation of end user input to reach the intended audiences and achieve the planned behavior change.

Proper design has become a critical element needed to engage website users. Poorly designed websites may frustrate users and result in a high "bounce rate," or people visiting the home page without exploring other pages within the site [26]. On the other hand, a well-designed website with high usability has been found to positively influence visitor retention (revisit rates) and engagement behavior [27,28]. A comprehensive analysis of the usability heuristics of the InPACT at Home program website was not conducted before its launch. Human-centered design represents a unique approach for tailoring web platforms to fit end users, narrowing the gap between efficacious interventions and public health impact. This approach places end users (ie, real people) at the center of the development process, enabling website developers to create programs and platforms that are tailored to the intended audiences' needs. The end users' wants, pain points, and preferences are prioritized during every phase of the process to enhance engagement and accessibility of the web-based program [29]. Given the problems associated with attracting, engaging, and retaining users to the InPACT at Home program website, the purpose of this study was to use a human-centered design process to evaluate and optimize user experience to promote website engagement and subsequent youth physical activity participation.

Methods

InPACT at Home Program

The InPACT at Home program is an evidence-informed family physical activity program that uses high-quality instructor-led physical activity videos to promote exercise in the home [23,30]. The InPACT at Home website is run on a WordPress platform, hosted by the university, and was developed by a professional web design company. The website is published publicly with log-in features to allow both the program developers and end users to track their completed activities and rewards. End users are awarded badges upon completion of the exercise videos. This feature was added based on previous research conducted in classroom settings that demonstrated significant improvements in youth moderate-to-vigorous intensity physical activity engagement when game design elements were added to the program [31]. Rewards have also been identified as facilitators of youth participation in virtual reality exergaming interventions [32].

A "Challenges" page was added to the website to highlight one health theme each month. Finally, to encourage mindfulness after each workout a postworkout survey was added to the website. End users are encouraged to answer the following questions: "In one or two words, please describe how participating in physical activity makes you feel?" and "If you could tell your friends one or two words about why physical activity is important, what would you say?" Previous research has demonstrated that engaging in self-reflection activities after a positive exercise experience can aid in the continuance of the behavior [33].

Physical education teachers, fitness professionals, pediatric exercise physiologists, athletes, and high school students from across the state of Michigan were recruited and hired to develop exercise videos that were developmentally appropriate and could be completed at home with no or minimal equipment. The types of exercises included were aerobic, isometric strength training, motor skills, sports skills, yoga, and mindfulness training. To supplement the exercise videos, physical activity play cards and family engagement tool kits were developed to provide another opportunity for children and families to move and play together. The movement-based play cards included cardio, strength, mindfulness, flexibility, and "with a buddy" activities. School psychologists, regional school health coordinators, and teachers from across the state of Michigan were hired to develop family engagement tool kits that focused on the following topics: resilience, well-being, focus, nutrition, sleep, family team building, family discussion, personal best, health choices, lifelong skills, substance misuse, schedules, and routines. Each module also included a 20-day challenge that incorporated movement activities. All program materials are hosted on the InPACT at Home website [34].

Recruitment

Parents of children and adolescents in grades K-12 (ages 5 - 17 years) were recruited to participate in this study using a variety of methods: registration opt-in for user research on the parent permission form on the program website, advertisements on the university clinical trials website, and sending out email advertisements to current users. Participants were eligible for inclusion in this study if they had a desktop computer and internet access in their homes, were able to answer questions and complete tasks on a computer, and were able to understand English. Participant eligibility was determined by parents answering a screening questionnaire, after which a member of the research team contacted them by email to confirm their eligibility and schedule the interview. Informed consents were obtained before the start of the study via a web-based survey using Qualtrics software (Qualtrics International Inc). If participants did not agree to participate in the study, the survey ended.

Qualitative Interviews

One-on-one semistructured interviews with parents of child users were conducted to assess engagement and pain points with the InPACT at Home website. Pain points were defined as specific problems faced by current or prospective website users and included any problems the user experienced when engaging with the website [35]. Interviews were conducted by trained research staff using videoconferencing. Four trained research staff members, all fourth-year undergraduate students at the University of Michigan School of Information, conducted the data analysis. Their training involved four years of coursework and real-world experiential learning through internships within the university's school. The same research staff conducted all aspects of the research study.

Purposive sampling was used to select participants from the pool of participants who responded to the advertisement. The criteria used for selection included accounting for diversity in race, gender, and age of their child. Using a standardized interview schedule, all participants were asked the same interview questions. Interviewers also asked additional unplanned questions to further assess new information introduced by participants. During the interview, interviewers were able to see participants' computer screens. Interviews ranged from 45 to 60 minutes in duration and were audio recorded and transcribed verbatim using a transcription company. Participants were compensated US $25 for their time.

Affinity Mapping

An affinity map involves gathering qualitative information about a target population and organizing it into categories. Initially, it is a useful method for compiling extensive information and data about users from various stages of development, such as user testing, surveys, observations, and feedback collection. The goal is to create an affinity diagram, a tool that visualizes the brainstorming process.

Professional user experience teams typically follow a flexible set of instructions, starting with selecting a topic, forming a cross-functional team, gathering facts and ideas, categorizing items, and devising an action plan. Throughout the session, team members collaborate to generate ideas pertinent to the chosen topic, with each brainstorming session yielding potentially different outcomes. An essential principle of affinity mapping is the absence of absolute right or wrong ways to categorize data; different teams may interpret the data differently and create distinct groups of data points based on collective decisions.

Approaching the data with a fresh perspective is advisable, avoiding premature labeling based on past experiences, as each data set is unique. Moreover, there are no rigid rules on how observations should be articulated; the focus is on gathering data in a manner that aligns with the team's dynamics.

Using a phenomenological approach, a thematic content analysis was used to examine the data and identify themes that elucidate each participant's experiences with physical activity programming and the InPACT at Home website. Qualitative data from semistructured interviews were organized into an affinity map using Miro software version 3.11.8 (RealTimeBoard, Inc).

The research staff reviewed the qualitative information, jotting down each observation on a movable card (ie, sticky note). The visual aspect of using sticky notes aids the team in physically visualizing connections between key data points, facilitating a literal connection between ideas. Sticky notes also allow for easy rearrangement and modification of groupings throughout the brainstorming process. The raters collaborated in a single room to jot down observations and identify themes, benefiting from collective brainstorming and free exchange of thoughts.

Using a large whiteboard, patterns in the observations were identified and categorized into groups. Each group was named, and a summary statement was provided regarding what was learned about each group. The analysis team also looked for outlier observations to understand instances where individual participant perspectives differed from the main findings, thus allowing for multiple perspectives and mitigating bias. Regular research meetings were held throughout the data analysis process with research team members possessing qualitative expertise to discuss the progress.

Heuristic Evaluation

A heuristic evaluation serves to systematically review the current state of a product, identifying usability and experience issues [36,37]. This evaluation is conducted based on Jakob Nielsen's 10 usability heuristics, which are high-level guidelines grounded in an understanding of human behavior, psychology, and information processing. These principles cover various aspects such as system status visibility, matching with the real world, user control, consistency, error prevention, recognition over recall, flexibility, minimalist design, error recovery assistance, and the provision of help and documentation [38]. These heuristics can be grouped into four main quality components: learnability, efficiency, memorability, and error management.

The term "heuristic" refers to a rule of thumb, and this process is particularly valuable in the early stages of a project due to its cost-effectiveness in analyzing the product being worked on. While it does not replace user research, it aids in identifying and defining the problems within a product. For instance, during evaluations of the InPACT at Home website using Figma software version 3.30 (Figma, Inc), usability issues were identified through the Nielsen process. These issues, such as dead links leading to a blank screen, were detected during internal product evaluations.

All issues identified during evaluations are based on team member observations, while the affinity map consolidates key data points from various sources collected before the evaluation. These data points, derived from user surveys and interviews, represent insights from the target users. Initially, all identified issues are assessed for severity to prioritize them effectively.

Each evaluator assigned a severity rating to usability issues on a scale of 0 (ie, no issue) to 4 (ie, usability catastrophe), accompanied by documentation of the specific violation and recommendations for fixing the problem. These ratings reflect a consensus reached by the group of evaluators, and they help guide decision-making regarding issue resolution.

Competitive Analysis

The purpose of conducting a competitive analysis is to gain strategic insights into how your product compares to the design solutions offered by competitors. This analysis covers various aspects such as functions, features, user flows, and the emotional response elicited by competitors' products. The goal is to strategically design your product to outperform the competition. Typically, this analysis is conducted initially to understand how you want your new product to differentiate itself. However, it is beneficial to approach this process iteratively, as competitors are constantly evolving. The key is to draw inspiration from competitors' solutions and determine what aligns best with your product and its intended users.

We specifically selected the Presidential Youth Fitness Program, YouTube, and Just Dance because we believe their features closely align with those of InPACT at Home. Presidential Youth Fitness Program offers a youth fitness training program aimed at promoting health-related fitness and providing quality resources for physical education, which aligns well with InPACT at Home's goals of engaging families and promoting physical activity. Similarly, YouTube offers a vast array of functionalities, including promoting healthy lifestyles, and its user-friendly video experience and large user base make it a strong competitor for analysis. Just Dance targets a younger audience and encourages active engagement through video platforms, aligning with InPACT at Home's objectives. Therefore, we identified these three competitors for comparison based on their alignment with InPACT at Home's goals and features.

The research team conducted a competitive analysis using Figma software to identify the strengths and weaknesses of InPACT at Home's competitors offering similar online products promoting physical activities. Research staff analyzed both direct and indirect competitors to identify gaps or opportunities that could give InPACT at Home an edge over its competitors [39]. Five aspects of each website were compared: target audience, first impressions, interactions, visual design, and content, chosen based on their relevance to InPACT at Home's goals.

Each aspect was rated as "Outstanding," "Good," "Okay," or "Needs work" based on the observed pros and cons. An example of a con for first impressions would be "too many features and complicated user flow," while a pro would be a "clean, minimalist design." Our approach involved individual reporting followed by consolidation to generate comprehensive insights based on key takeaways. Information from the competitive analysis was not compared against the InPACT at Home website for benchmarking but instead used as inspiration to determine what aligns best with our website and intended users.

Ethical Considerations

This study was approved by the institutional review board of the University of Michigan (HUM00192745).

Results

Overview

Of the 98 eligible participants who responded to study advertisements, seven parents of children in grades K-12 were contacted and agreed to be interviewed. There were three non-Hispanic White male participants, 1 Asian female participant, two non-Hispanic Black female, and one non-Hispanic White female (average age 41.3, SD 10.2 years). On average, parents had 1.6 (SD 0.8) school-age children (average age 8.4, SD 4.5 years) residing in their household. Of the seven parents, three reported being regularly physically active, and five parents reported their children participated in regular physical activity.

Affinity Mapping

Thematic saturation was achieved, and Textbox 1 displays the themes and supporting quotes from participants derived from the qualitative interviews conducted. Participant interview responses were categorized into specific website components and included the following: landing/home page, video, current progress and badges, play cards, and overall experience with the website. Responses were further categorized into "likes," "dislikes/struggled with," and "wants" as they related to each component. The following themes emerged from the interview responses. The first theme related to "website likes" included the *website design*. Parents noted that they liked that the website was gamified, colorful, and included pictures. Parents also commented on the variety of exercises and resources available to parents. The second theme related to "website dislikes" included *difficult navigation*. Parents noted that there was too much scrolling on the home page. The reflection/record progress survey was difficult to find, and some parents were unable to find the play cards. Finally, the third theme related to "website wants" included *added features*. Parents suggested adding a progress button, a library of watched videos, and more information about the "Challenge" page.

Textbox 1. Themes emerging from the end user qualitative interviews.

Theme related to website likes: liked website design (eg, gamified, colorful, pictures, exercise variation, resources for parents)

- “The design is simple and a very colorful website.” (P02 Sarah)
- “More motivated to earn badges.” “Like how it provides resources for parents, not only kids.” (P07 Molly)

Theme related to website dislikes: difficult navigation (eg, too much scrolling on home page, reflection/record progress survey difficult to find, unable to locate play cards)

- “Too much information and scrolling on the home and landing page.” (P02 Sarah)
- “Could not find the record your progress survey button.” “Was expecting the survey to pop up immediately after finishing the video.” (P03 Matt)
- “Thinks the reflection survey process was challenging because had to click on the back button to go back to the original page.” (P04 Janice)
- “Did not know what a play card is so it was hard to find, and the search bar did not work on the website.” (P06 Nina)

Theme related to website wants: added features (eg, progress button, library of watched videos, challenge information)

- “It would be better if there was a feature that stated my progress to put the individual’s current physical activity progress.” “It would be nice to have a library to show already watched videos.” (P03 Matt)
- “Give more information about what kind of badges and what you can do with the challenges.” (P07 Molly)

Heuristic Evaluation

Table 1 presents the results of the heuristic evaluation conducted on the InPACT at Home website. For the heuristics of consistency and standards, and helping users recognize, diagnose, and recover from errors, the website received a score of 0, indicating that the evaluators did not perceive these as usability issues. Two other heuristic categories, recognition rather than recall and help and documentation were assigned a score of 2, indicating minor usability problems. For three categories, namely match between the system and the real world, flexibility and efficiency of use, and aesthetic and minimalist design, the website received a score of 3, indicating major usability problems. Finally, the heuristics of visibility of system status, user control and freedom, and error prevention were rated with a severity score of 4, representing usability catastrophes.

Table . Heuristic evaluation for the Interrupting Prolonged sitting with ACTivity (InPACT) at Home website.

Heuristics	Violation	Recommendation	Severity
Visibility of system status	• Does not show the user how much time they must wait before a new page is loaded. • When the user clicks on the tab, there's nothing that indicates that the user has clicked on it or is currently clicking on it.	• Having a loading icon that pops up when the user clicks on a tab to go to another page to show the user that the new page is loading. • When the user clicks on a tab in the navigation bar, have the tab color change to a different color to show the user that the system knows they're clicking on the right tab.	4
Match between system and the real world	• Require the user to think hard about what the category means and what the language implies. For instance, "Topics, Challenges" are not familiar categories to the user. They'll be thinking about what topics and challenges the website is referring to.	• Replace category names with more familiar categories to the user such as replacing "Topics" with "Explore."	3
User control and freedom	• Users do not have the control to exit the reflection survey after finishing watching a video.	• Add a button that allows the user to exit the survey if they do not want to fill it out.	4
Consistency and standards	• All pages are consistent.	• Nothing to change.	0
Error prevention	• The reflection survey leads to a dead screen, so users must click the back button to return to the InPACT at Home website. • Users can accidentally click log-in on the register page, leading to users having no actual registration and needing to enter all personal information again.	• Include another button that allows the user to go back to the home page or have the original button add a new tab that can be closed. • Remove the log-in option on the register page.	4
Recognition rather than recall	• Users must remember to scroll down to complete the reflection survey since it is not on a screen once you complete an activity. • Users must remember which videos are their favorites for future uses.	• Move the reflection survey to within the screen when users finish a video/challenge. • Incorporate a favorite section where users can easily see which videos they have enjoyed.	2
Flexibility and efficiency of use	• While there is an option for progress recording, the information is not displayed on the screen, hindering users from decision-making. • Video titles are not descriptive enough to communicate the video content (eg, "Scott Przystas-Short Video 2").	• Notify the users when they complete the video and guide them for progress recording. • Subcategorize videos and make each title distinct to one another.	3
Aesthetic and minimalist design	• The profile section is not designed with proper grouping and colors—badge colors do not communicate their meaning.	• Include section that explains what each color represents or substitute the colored badges into word tags.	3

Heuristics	Violation	Recommendation	Severity
Help users recognize, diagnose, and recover from errors	• InPACT at Home uses this heuristic by providing error messages when the user tries to log in, and if the log-in is incorrect, there will be a message that says why the log-in is not working.	• Nothing to change.	0
Help and documentation	• Users must understand that their most recent badges and ranks are on their profile page. Users not only have to navigate back from the reflection survey but also must remember which video was most recently watched.	• Incorporate a notification that users have earned a badge or reward right after it was achieved. Either a banner or pop-up notification, so users do not need to remember or navigate anywhere else.	2

Competitive Audit

Table 2 displays the competitive audit comparing the online physical activity experience of each website. YouTube received the highest ratings of the three competitors with "outstanding" ratings in six of the seven categories (ie, desktop web/game experience, accessibility, user flow, navigation, brand identity, and descriptiveness). Just Dance received the second highest ratings with "outstanding" ratings in five of the seven categories (ie, desktop web/game experience, user flow, navigation, brand identity, and descriptiveness). The Presidential Youth Fitness Program website was the lowest-rated website with "good" ratings in three of the seven categories (ie, navigation, brand identity, and descriptiveness).

Table . Competitive audit comparing the online physical activity experience of the Presidential Youth Fitness Program (PYFP), YouTube, and Just Dance websites.

	First impressions	Interactions				Visual design	Content	
	Desktop web/game experience	Features	Accessibility	User flow	Navigation	Brand identity	Tone	Descriptiveness
PYFP	Okay • (+) Clean design • (–) Too many features and complicated user flow	Okay • (+) Resource guide for parents and educators • (+) Awards store for recognition • (–) No progress recorder • (–) Not able to log in unless users are part of the organization	Needs work • (+) Video speed options • (+) Only offers 1 language: English • (–) No subtitles or closed captions • (–) No colorblind mode	Needs work • (–) Overwhelming number of user interface elements and content	Good • (+) Clear indication of clickable elements • (–) Some unfamiliar navigation patterns	Good • (+) Visual design communicates organization ethos • (–) Visual design does not always support content intuitively	Professional and informative	Good • (+) All key information is present • (–) Too descriptive
YouTube	Outstanding • (+) Well designed and easy to use • (+) Modern minimalistic design	Good • (+) Any user can create own videos, comment, like, save, and share. • (+) Filtering and recommendations features • (+) YouTube Kids, providing content that is age appropriate • (–) Not categorized	Outstanding • (+) Subtitles and closed captions • (+) Screen reader, interaction controls, display settings, audio and on-screen text options • (+) Offers 75 different languages for site navigation	Outstanding: • (+) Straightforward user flow • (+) One click sign-up • (+) Easy video selection process due to recommendations	Outstanding: • (+) Easy basic navigation • (+) Clear indication of clickable elements • (+) Understandable link labels	Outstanding • (+) Strong brand identity including colors, fonts, style, and imagery • (+) Visual design communicates company ethos	Sophisticated and informative	Outstanding • (+) All key information is present

	First impressions	Interactions				Visual design	Content	
	Desktop web/game experience	Features	Accessibility	User flow	Navigation	Brand identity	Tone	Descriptiveness
Just Dance	Outstanding • (+) Well designed and easy to navigate • (+) Cheerful theme with vibrant colors	Outstanding • (+) Support both single and multiplayer mode • (+) Kids mode: dancers of any age can enjoy • (+) Multiple levels: easy to hard • (+) Leaderboards and forums	Good • (+) Game can be paused at all times • (+) Subtitles for lyrics of the songs • (–) Offers 15 different language options • (–) No color-blind mode • (–) No assist features or ability to set the game speed	Outstanding • (+) Fun and easy to use • (+) Display song options in digestible categories	Outstanding • (+) Easy basic navigation • (–) No link labels—only present as icons	Good • (+) Modern and trendy design • (+) Visual design communicates company ethos	Engaging, concise, and friendly	Outstanding • (+) All key information is present

Discussion

Principal Results

Given the problems associated with attracting, engaging, and retaining users to the InPACT at Home program website (15,000-20,000 daily viewers through public television broadcasting vs 23 registered website users) [23], the purpose of this study was to use a human-centered design process to evaluate and optimize the user experience to promote website engagement and subsequent youth physical activity participation. Using qualitative methodologies and evidence-based heuristic evaluation approaches, we conducted a series of assessments to examine end user engagement and pain points as well as completed a competitive analysis to identify the strengths and weaknesses of competitors who offered similar products. Both the affinity maps developed from end user interviews and the heuristic evaluation of the InPACT at Home program website revealed several major problems and usability catastrophes in three of the four Nielsen quality components: *learnability*, *efficiency*, and *errors*. All these issues resulted in low usability (difficult to navigate) of the InPACT at Home program website and likely contributed to the low user retention (registration rates) and engagement behavior previously observed. The competitive analysis identified YouTube as the highest-rated competitor with "outstanding" ratings and revealed key features that the InPACT at Home program website could benchmark (ie, desktop web/game experience, accessibility, user flow, navigation, brand identity, and descriptiveness). Taken together, these findings suggest the InPACT at Home website needed numerous modifications to enhance usability. Appropriately, the YouTube website interface provided a road map by which we could improve our design interface to fit end user goals and preferences.

Comparison With Prior Work

Previous research has demonstrated that parent support is an important determinant of child and adolescent physical activity participation. Data from vEngage, a virtual reality exergaming intervention, suggest that while parents would rather their child perform "real-world" physical activity, they believed the key to engagement was through technology and were willing to support their child's participation in exergaming [40]. Our recent findings from the InPACT at Home program demonstrated that parent support in the form of encouraging their children, reminding them, and establishing schedules and routines significantly facilitated participation in the program (LR Beemer et al, unpublished data, 2024). These findings provide the rationale for why parents were selected as the target audience. Understanding the pain points of parents in using the InPACT at Home website was vitally important to achieving a "trickle-down effect" for child engagement, and accordingly, issues with learnability, efficiency, and error needed to be addressed.

Nielsen's usability heuristics framework conceptualizes *learnability* as the ease with which users can accomplish basic tasks the first time they encounter the website design [41]. The goal is to design a clear interface that users can quickly learn and understand. Previous research has demonstrated that users can receive more value from a website with high learnability compared to websites with lower learnability [42]. This is due in part to users being able to adopt the learnable interface much quicker and subsequently accomplish their goals in a shorter amount of time using the website. By having an easier time navigating the website, users will also have an overall better experience with the website, which can contribute to a better retention rate and lower bounce rates [43,44]. Best practices for creating a learnable interface include consistency (eg, giving all the web pages a similar look by positioning elements in the same location), feedback (eg, link color changes that tell the user that an element is clickable), using well-known user interface elements (eg, sticking to industry design best practices), familiarity (eg, user can learn the new interface based on previous knowledge), and testimonials (eg, visual storytelling enabling users to learn and remember information).

In this study, the *learnability* of the InPACT at Home website was deemed low as the program website did not provide timely feedback and used unfamiliar concepts, thereby increasing the time needed to learn how to use the website. The website also contained extraneous information that competed with relevant information needed to complete a task, making it difficult for the end user to understand how to use the website. Themes from end user interviews also confirmed that the website was difficult to navigate. Accordingly, substantial attention to creating a learnable user interface on the InPACT at Home program website was needed to optimize user experience and engagement with program resources.

The *efficiency* of the InPACT at Home website was also deemed low, and *errors* were deemed high. Efficiency measures the speed (or quickness) with which a user can accomplish a task once they have become familiarized with the website design [45]. In other words, *efficiency* is the number of keystrokes or clicks it takes a user to complete a task. Like *learnability*, the more efficient an interface design is, the greater value a user can gain from a website as they can complete a task in a shorter amount of time [27,42]. *Errors* on the other hand are software problems that come from a misconfigured website design, making it difficult to complete a task resulting in user frustration [45]. The InPACT at Home program website did not enable users to have control to exit the reflection survey after finishing watching a video; there were no options for progress reporting on the website, and video titles were nondescriptive; all these factors led to website inefficiencies. In addition, some pages on the website led to dead screens, and there were several error-prone conditions on the registration page; these factors contributed to errors on the website.

Optimization of the InPACT at Home Website

Based on the recommendations provided by the end users and website evaluators, we have made several updates to the InPACT at Home program website. To overcome the barriers related to *learnability*, we have created custom module content that can easily be searched and filtered by topic and automatically archived. Users can select the type of exercise videos they want to engage in as well as select the family engagement tool kit topics they are most interested in. A recent review identified personalization as a key mechanism of web-based interventions that elicited positive changes in physical activity behaviors [14]. To overcome the barriers associated with *efficiency*, we have created a modified log-in process to direct the visitor to their content. Rather than having to scroll through all 132 exercise videos, their personalized profile page now hosts their preferred content, making resources quicker to access. To overcome the barriers with *errors*, we have identified and removed dead screens throughout the website and redesigned the registration page. The removal of these errors and error-prone conditions should reduce user frustration with the website.

To also be responsive to user preferences, we added exercise intensity levels to each video and a brief description of the video content to each video so that children know which activities they will be doing and what equipment is needed. We have also created QR codes for customized workouts to further enhance the personalization of the site. Many of these improvements were modeled after YouTube features identified in the competitive analysis for website design. Our next step in the website optimization process is to conduct additional user testing to confirm these updates are meeting end user needs. We will begin to monitor engagement with the website and program resources.

This study has several important strengths that are worth mentioning. First, we used a common evidence-based heuristic evaluation and competitive analysis to determine the usability of the InPACT at Home website. Second, analyses were conducted by experienced user interface evaluators and researchers with expertise in qualitative interviewing. Finally, our human-centered design approach enabled end users, evaluators, and website developers to come together to evaluate and optimize the user experience to increase website engagement and eventual youth physical activity engagement.

Limitations

Limitations of this study also warrant attention. First, we acknowledge that our parent sample exhibits some diversity; however, it is essential to consider other characteristics to ensure a truly diverse sample in this context. These include the age range of children from K-12, geographical distribution, levels of digital literacy, income levels, and patterns of technology use. Further testing may be necessary to ensure that the InPACT at Home website adequately caters to the diverse needs of parents and families across the state of Michigan and beyond. Second, there were few existing registered users at the initiation of this analysis; hence, most of the interviews were conducted with parents who were unfamiliar with or not currently using the InPACT at Home website. This could have led to biased responses and recommendations that are only appropriate for first-time users. Third, while 98 eligible participants responded to study advertisements, time and cost constraints limited our ability to conduct more interviews. Nevertheless, we did achieve thematic saturation, and many of the themes identified in the qualitative interviews were confirmed in the heuristic evaluation. Fourth, we used subjective assessments (evaluator ratings) to determine the accessibility and usability of the InPACT at Home program website, thereby increasing the potential for inconsistency in scoring. To overcome this limitation, all four evaluators completed the affinity mapping and heuristic evaluation together; scores reflected a group consensus. The competitive analysis was completed independently (ie, each team member researched one competitor) and then discussed as a group. Finally, the observational nature of the study precluded our ability to directly conduct comparative user testing of the InPACT at Home website along with its competitors. Despite these limitations, our analysis provided valuable information to our website developers from experienced evaluators and end users that enabled us to make substantive changes to the website to improve usability.

Conclusions

Most children in the United States are classified as inactive because they do not participate in the recommended 60 minutes of physical activity per day [1,2]. Online and web-based interventions have the potential to improve physical activity engagement because of their extensive reach and accessibility across different contexts [14]. Because this generation of children and adolescents is the first to have their entire childhood influenced by the internet and mobile devices [46], web-based interventions may be uniquely positioned to promote sustainable physical activity participation in this age group. Like most other web-based physical activity interventions, the InPACT at Home program website failed to reach its anticipated impact due to insufficient use by the intended audiences. Problems associated with attracting, engaging, and retaining participants in web-based interventions were likely the result of using a website design with low *learnability*, low *efficiency*, and high *errors*. Human-centered design was an evidence-based approach for optimizing the InPACT at Home program website to fit end user goals and preferences. Behavioral interventionists should consider conducting a comprehensive usability heuristic evaluation *before* the website launch to narrow the gap between efficacious interventions and public health impact.

Acknowledgments

We would like to thank Insun Kim, Alaa Shahin, Dr Florian Schaub, and the University of Michigan School of Information client engagement team for their helpful assistance throughout the research process. We are also grateful to our study participants and their families for their involvement. Finally, we are appreciative of the team of website developers who created the website and made subsequent improvements. The results of this study are presented clearly, honestly, and without fabrication or inappropriate data manipulation.

Conflicts of Interest

None declared.

References

1. Piercy KL, Troiano RP, Ballard RM, et al. The physical activity guidelines for Americans. JAMA 2018 Nov 20;320(19):2020-2028. [doi: 10.1001/jama.2018.14854] [Medline: 30418471]
2. Physical activity guidelines for Americans. Health.gov. URL: https://health.gov/our-work/nutrition-physical-activity/physical-activity-guidelines [accessed 2024-02-07]
3. Merlo CL, Jones SE, Michael SL, et al. Dietary and physical activity behaviors among high school students - youth risk behavior survey, United States, 2019. MMWR Suppl 2020 Aug 21;69(1):64-76. [doi: 10.15585/mmwr.su6901a8] [Medline: 32817612]
4. Nader PR, Bradley RH, Houts RM, McRitchie SL, O'Brien M. Moderate-to-vigorous physical activity from ages 9 to 15 years. JAMA 2008 Jul 16;300(3):295-305. [doi: 10.1001/jama.300.3.295] [Medline: 18632544]
5. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, Mcdowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008 Jan;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3]
6. Davison KK, Lawson CT. Do attributes of the physical environment influence children's level of physical activity? Int J Behav Nutr Phys Act 2006 Jul 27;3:1-17. [doi: 10.1186/1479-5868-3-19] [Medline: 16872543]
7. Richmond TK, Hayward RA, Gahagan S, Field AE, Heisler M. Can school income and racial/ethnic composition explain the racial/ethnic disparity in adolescent physical activity participation? Pediatrics 2006 Jun 1;117(6):2158-2166. [doi: 10.1542/peds.2005-1920]
8. Hasson R, Sallis JF, Coleman N, Kaushal N, Nocera VG, Keith N. COVID-19: implications for physical activity, health disparities, and health equity. Am J Lifestyle Med 2021 Jul 27;16(4):420-433. [doi: 10.1177/15598276211029222] [Medline: 35855783]
9. Neville RD, Lakes KD, Hopkins WG, et al. Global changes in child and adolescent physical activity during the COVID-19 pandemic: a systematic review and meta-analysis. JAMA Pediatr 2022 Sep 1;176(9):886-894. [doi: 10.1001/jamapediatrics.2022.2313] [Medline: 35816330]
10. Woolford SJ, Sidell M, Li X, et al. Changes in body mass index among children and adolescents during the COVID-19 pandemic. JAMA 2021 Oct 12;326(14):1434-1436. [doi: 10.1001/jama.2021.15036] [Medline: 34448817]
11. Peng X, Menhas R, Dai J, Younas M. The COVID-19 pandemic and overall wellbeing: mediating role of virtual reality fitness for physical-psychological health and physical activity. Psychol Res Behav Manag 2022 Jul 13;15:1741-1756. [doi: 10.2147/PRBM.S369020] [Medline: 35860203]

12. Atkinson NL, Gold RS. The promise and challenge of eHealth interventions. Am J Health Behav 2002;26(6):494-503. [doi: 10.5993/ajhb.26.6.10] [Medline: 12437024]
13. Evers KE, Prochaska JM, Prochaska JO, Driskell MM, Cummins CO, Velicer WF. Strengths and weaknesses of health behavior change programs on the internet. J Health Psychol 2003 Jan;8(1):63-70. [doi: 10.1177/1359105303008001435] [Medline: 22113901]
14. Goodyear VA, Skinner B, McKeever J, Griffiths M. The influence of online physical activity interventions on children and young people’s engagement with physical activity: a systematic review. Phys Education Sport Pedagogy 2021 Aug 26;28(1):94-108. [doi: 10.1080/17408989.2021.1953459]
15. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27. [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]
16. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The effectiveness of web-based vs. non-web-based interventions: a meta-analysis of behavioral change outcomes. J Med Internet Res 2004 Nov 10;6(4):e40. [doi: 10.2196/jmir.6.4.e40] [Medline: 15631964]
17. Screen time and children. American Academy of Child and Adolescent Psychiatry. 2024 May. URL: https://www.aacap.org/AACAP/Families_and_Youth/Facts_for_Families/FFF-Guide/Children-And-Watching-TV-054.aspx [accessed 2024-06-28]
18. Vandelanotte C, Kirwan M, Rebar A, et al. Examining the use of evidence-based and social media supported tools in freely accessible physical activity intervention websites. Int J Behav Nutr Phys Act 2014 Aug 17;11:105. [doi: 10.1186/s12966-014-0105-0] [Medline: 25128330]
19. Hutchens A, Lee RE. Parenting practices and children’s physical activity: an integrative review. J Sch Nurs 2018 Feb;34(1):68-85. [doi: 10.1177/1059840517714852] [Medline: 28631518]
20. Yao CA, Rhodes RE. Parental correlates in child and adolescent physical activity: a meta-analysis. Int J Behav Nutr Phys Act 2015 Feb 11;12:10. [doi: 10.1186/s12966-015-0163-y] [Medline: 25890040]
21. Tandon P, Grow HM, Couch S, et al. Physical and social home environment in relation to children's overall and home-based physical activity and sedentary time. Prev Med 2014 Sep;66:39-44. [doi: 10.1016/j.ypmed.2014.05.019] [Medline: 24887496]
22. Graham DJ, Wall MM, Larson N, Neumark-Sztainer D. Multicontextual correlates of adolescent leisure-time physical activity. Am J Prev Med 2014 Jun;46(6):605-616. [doi: 10.1016/j.amepre.2014.01.009] [Medline: 24842737]
23. Hasson RE, Eisman AB, Wassmann A, et al. Rapid cycle adaptation of a classroom-based intervention to promote equity in access to youth physical activity. Transl Behav Med 2022 Oct 7;12(9):945-955. [doi: 10.1093/tbm/ibac049] [Medline: 36205474]
24. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [doi: 10.1016/j.amepre.2007.02.041] [Medline: 17572313]
25. Kohl LFM, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013 Jul 16;15(7):e146. [doi: 10.2196/jmir.2665] [Medline: 23859884]
26. [UA] Bounce rate. Google Help. 2015. URL: https://support.google.com/analytics/answer/1009409?hl=en [accessed 2024-06-28]
27. Garett R, Chiu J, Zhang L, Young SD. A literature review: website design and user engagement. Online J Commun Media Technol 2016 Jul;6(3):1-14. [Medline: 27499833]
28. AIContentfy Team. The impact of website design on customer acquisition and retention. AIContentfy. 2023 Nov 6. URL: https://aicontentfy.com/en/blog/impact-of-website-design-on-customer-acquisition-and-retention#:~:text=A%20well%2Ddesigned%20website%20should,drive%20customer%20acquisition%20and%20retention [accessed 2024-06-28]
29. Landry L. What is human-centered design? Harvard Business School Online. 2020 Dec 15. URL: https://online.hbs.edu/blog/post/what-is-human-centered-design [accessed 2024-06-28]
30. Beemer LR, Tackett W, Schwartz A, et al. Use of a novel theory-based pragmatic tool to evaluate the quality of instructor-led exercise videos to promote youth physical activity at home: preliminary findings. Int J Environ Res Public Health 2023 Aug 11;20(16):6561. [doi: 10.3390/ijerph20166561] [Medline: 37623147]
31. Beemer LR, Ajibewa TA, DellaVecchia G, Hasson RE. A pilot intervention using gamification to enhance student participation in classroom activity breaks. Int J Environ Res Public Health 2019 Oct 24;16(21):4082. [doi: 10.3390/ijerph16214082] [Medline: 31652885]
32. Farič N, Yorke E, Varnes L, et al. Younger adolescents' perceptions of physical activity, exergaming, and virtual reality: qualitative intervention development study. JMIR Serious Games 2019 Jun 17;7(2):e11960. [doi: 10.2196/11960] [Medline: 31210135]
33. Brand R, Ekkekakis P. Affective-reflective theory of physical inactivity and exercise: foundations and preliminary evidence. German J Exerc Sport Res 2017 Nov 14;48:48-58. [doi: 10.1007/s12662-017-0477-9]
34. InPACT at Home. URL: https://inpactathome.umich.edu/ [accessed 2024-06-27]
35. Pain points. Gartner. 2023. URL: https://www.gartner.com/en/sales/glossary/pain-points [accessed 2024-06-28]
36. Molich R, Nielsen J. Improving a human-computer dialogue. Commun ACM 1990 Mar 1;33(3):338-348. [doi: 10.1145/77481.77486]
37. Nielsen J. Heuristic evaluation. In: Mack RL, editor. Usability Inspection Studies: John Wiley & Sons; 1994:25-62.

38. Neil T. 6 tips for a great flex UX: part 5. Designing Web Interfaces. 2009 May 17. URL: http://designingwebinterfaces.com/6-tips-for-a-great-flex-ux-part-5 [accessed 2024-06-28]
39. Competitive analysis for UX - top 6 research methods. UXPin. 2022 Jun 1. URL: https://www.uxpin.com/studio/blog/competitive-analysis-for-ux/#:~:text=UX%20researchers%20can%20use%20competitive,making%20their%20product%20more%20desirable [accessed 2024-06-28]
40. McMichael L, Farič N, Newby K, et al. Parents of adolescents perspectives of physical activity, gaming and virtual reality: qualitative study. JMIR Serious Games 2020 Aug 25;8(3):e14920. [doi: 10.2196/14920] [Medline: 32840487]
41. Mulders M. Learnability in web design: 5 best practices. SitePoint. 2020 Oct 22. URL: https://www.sitepoint.com/learnability-in-web-design [accessed 2024-06-28]
42. Evison B. The importance of learnability in web design. StrategiQ. 2016 Jul 4. URL: https://strategiq.co/importance-learnability-web-design [accessed 2024-06-28]
43. Zantal-Wiener A. What is bounce rate? (And how can I fix mine?). HubSpot Blog. 2021 Dec 22. URL: https://blog.hubspot.com/marketing/what-is-bounce-rate-fix#:~:text=Let's%20say%20you%20were%20comparing,re%20losing%20out%20on%20conversions [accessed 2024-06-28]
44. Willson A. 12 reasons your website can have a high bounce rate. Search Engine Journal. 2022 Nov 24. URL: https://www.searchenginejournal.com/website-bounce-rate/332439/#close [accessed 2024-06-28]
45. Shukairy A. 5 usability design tips for a better user experience design. Invesp. 2023 Dec 8. URL: https://www.invespcro.com/blog/usability-design-for-a-better-user-experience [accessed 2024-06-28]
46. Buckingham D. Making sense of the 'Digital Generation': growing up with digital media. Self Soc 2015 Jan 21;40(3):7-15. [doi: 10.1080/03060497.2013.11084274]

Abbreviations

InPACT: Interrupting Prolonged sitting with ACTivity
K-12: kindergarten through grade 12

Edited by A Kushniruk; submitted 05.09.23; peer-reviewed by H Chambliss, N Farič; revised version received 01.04.24; accepted 05.05.24; published 15.07.24.

Please cite as:
Hasson RE, Xie M, Tadikamalla D, Beemer LR
Using a Human-Centered Design Process to Evaluate and Optimize User Experience of a Website (InPACT at Home) to Promote Youth Physical Activity: Case Study
JMIR Hum Factors 2024;11:e52496
URL: https://humanfactors.jmir.org/2024/1/e52496
doi:10.2196/52496

Original Paper

Acceptance and User Experiences of a Wearable Device for the Management of Hospitalized Patients in COVID-19–Designated Wards in Ho Chi Minh City, Vietnam: Action Learning Project

An Phuoc Luu[1], MPH; Truong Thanh Nguyen[2], MD; Van Thi Cam Cao[2], RN; Trinh Hoang Diem Ha[2], RN; Lien Thi Thu Chung[2], RN; Trung Ngoc Truong[2], MD; Tung Nguyen Le Nhu[2], MD; Khoa Bach Dao[2], MD; Hao Van Nguyen[2], MD; Phan Nguyen Quoc Khanh[1], MD; Khanh Thuy Thuy Le[1], MD; Luu Hoai Bao Tran[1], MD; Phung Tran Huy Nhat[1], BSc; Duc Minh Tran[1], PMD; Yen Minh Lam[1], MD; Catherine Louise Thwaites[1,3], BSc, MBBS, MD; Jacob Mcknight[3], BEng, MSc, DPhil; Nguyen Van Vinh Chau[1,2], MD; Jennifer Ilo Van Nuil[1,3], BA, MA, PhD; Vietnam ICU Translational Applications Laboratory (VITAL)[4]

[1]Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam

[2]Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam

[3]Centre for Tropical Medicine and Global Health, Nuffleld Department of Medicine, University of Oxford, Oxford, United Kingdom

[4]See Acknowledgments

Corresponding Author:
Jennifer Ilo Van Nuil, BA, MA, PhD
Oxford University Clinical Research Unit
764 Vo Van Kiet
Ho Chi Minh City, 70000
Vietnam
Phone: 84 362620124
Email: jvannuil@oucru.org

Abstract

Background: Wearable devices have been used extensively both inside and outside of the hospital setting. During the COVID-19 pandemic, in some contexts, there was an increased need to remotely monitor pulse and saturated oxygen for patients due to the lack of staff and bedside monitors.

Objective: A prototype of a remote monitoring system using wearable pulse oximeter devices was implemented at the Hospital for Tropical Diseases in Ho Chi Minh City, Vietnam, from August to December 2021. The aim of this work was to support the ongoing implementation of the remote monitoring system.

Methods: We used an action learning approach with rapid pragmatic methods, including informal discussions and observations as well as a feedback survey form designed based on the technology acceptance model to assess the use and acceptability of the system. Based on these results, we facilitated a meeting using user-centered design principles to explore user needs and ideas about its development in more detail.

Results: In total, 21 users filled in the feedback form. The mean technology acceptance model scores ranged from 3.5 (for perceived ease of use) to 4.4 (for attitude) with behavioral intention (3.8) and perceived usefulness (4.2) scoring in between. Those working as nurses scored higher on perceived usefulness, attitude, and behavioral intention than did physicians. Based on informal discussions, we realized there was a mismatch between how we (ie, the research team) and the ward teams perceived the use and wider purpose of the technology.

Conclusions: Designing and implementing the devices to be more nurse-centric from their introduction could have helped to increase their efficiency and use during the complex pandemic period.

(JMIR Hum Factors 2024;11:e44619) doi:10.2196/44619

KEYWORDS

vital signs; wearable devices; action learning; technology acceptance model; TAM; COVID-19; user-centered design; wearables; remote monitoring; technology acceptance; oximeter

Introduction

The popularity of portable wearable technologies that monitor health has increased substantially over the past decade due to their perceived utility, relatively simple implementation, and immediate feedback [1]. Wearable technology is used in both personal and clinical settings, and more recently in the context of the COVID-19 pandemic for diagnosis, remote monitoring, and other applications in both inpatient and outpatient settings [1-3]. Using wearable devices for COVID-19 care can result in infection control by reducing the amount of time that health care workers (HCWs) are physically with patients and providing continuous monitoring of vital signs for the early identification and potential treatment of deteriorating patients [2]. Specifically, remote monitoring of oxygen saturation using wearable devices became increasingly common during COVID-19 in hospital settings [4,5].

Despite the potential benefits, there have been many challenges noted in implementing and using wearable devices during COVID-19, including technical, social, and political spheres [1]. Technical challenges often include battery life, Wi-Fi or Bluetooth connections, and device communication. A few examples of social challenges are patients lacking technological confidence (eg, in older patients) and repeated device alerts or continuous monitoring making patients nervous, while political challenges could include regulatory issues for expanding the approval of devices for COVID-19–related medical situations [1]. Many of these challenges may be enhanced in low- and middle-income countries (LMICs), while the need for such integration is crucial, especially during pandemic situations [6,7].

There are several studies exploring the technical challenges of integrating wearable devices in trial settings during COVID-19 (eg, see [8]), but there is a lack of research surrounding the acceptability of such devices within these contexts and how attitudes may impact actual use [9]. Portable wearable devices could be a potential solution to allow for continuous monitoring of vital signs remotely and affordably for COVID-19 wards in LMIC settings; however, while advantageous, these devices cannot meet their full potential if the users do not agree to use them or realize their potential value [10]. Understanding user perceptions and needs as well as the context in which the technology is implemented is crucial for successful implementation [1]. User-centered approaches stress the importance of integrating both human factors and technical factors [11] while also paying attention to avoid excluding certain populations in the design [7]. User-centered approaches have been cited as a “critical success factor” in a variety of health-related technology projects [12].

From August to December 2021, when COVID-19 cases were increasing more rapidly than at any time previously in Ho Chi Minh City, Vietnam, there was an opportunity to integrate a prototype wearable device and monitoring system into the COVID-19–designated wards at the Hospital for Tropical Diseases (HTD). At this time, the HTD was overwhelmed with patients with COVID-19 and we needed to deploy something urgently that could help. Using pragmatic methods during the rollout of the device, we describe stakeholders’ use of the wearable device, aspects of acceptability, and under which circumstances its use would be most beneficial for improving the care of patients with COVID-19. The primary objective of this work was to support the implementation process of the wearable device in the hospital to improve patient care during a catastrophic period of the COVID-19 pandemic in Ho Chi Minh City, Vietnam.

Methods

Study Setting

This work took place within a larger project called the Vietnam ICU Translation Application Laboratory (VITAL) at the Oxford University Clinical Research Unit (OUCRU) and HTD. The goal of VITAL is to design and implement innovative technologies to improve patient care within the intensive care unit (ICU) at the HTD, with a longer-term goal of expanding these technologies regionally. In addition to the clinical and technological studies, there is an ethnographic study to explore the sociotechnical contexts of the ICU at the HTD and within ICUs in Vietnam more broadly. The VITAL multidisciplinary team was in place at the start of the COVID-19 pandemic.

In the first 100 days of the COVID-19 pandemic, Vietnam rapidly implemented a variety of public health measures resulting in relatively few cases and zero deaths [13]. Since that time, there were a few concentrated outbreaks (for example, in Da Nang in July 2020 and December 2020 in northern Vietnam). In May 2021, the cases started to increase on a countrywide basis, and by August 2021, the hospitals began to fill with patients with COVID-19. It was within this pandemic context that the wearable device was implemented at the HTD, and the VITAL study teams worked together throughout to improve its implementation.

The wearable device was selected by the company and was already integrated into a locally developed platform based on an available application programming interface, licensing, and availability. The device was medical grade and measured heart rate and blood oxygen levels, similar to a pulse oximeter. The wearable device was battery powered and each one connected to a tablet that was kept at the patients’ bedside. The tablets had a 3G or 4G sim card and sent the data to a cloud where multiple patients’ data were viewable by HCWs outside the patients’ rooms and isolation area.

Study Design

The aim of this work was to support the ongoing implementation of the wearable device rather than to follow a predefined, replicable study protocol, as would be used in trial settings, for example. Therefore, the work here describes the pragmatic rollout of the device. We used an action learning approach, including integration of multiple methods to assess the use and acceptability of the wearable device [14]. Action learning approaches rely on an iterative process of assessing local contexts, learning from relevant stakeholders, and using the information to improve an implementation or further develop a technology specific to the context [15,16]. As the wearable device started to be implemented in the HTD wards, our team of HCWs, social scientists, and technology developers took the opportunity to work together to inform the implementation. Therefore, we adapted the methods as the situation changed and more insights were gained [14].

Participants

Potential participants included the HCWs from the HTD who were using the device in the wards during the implementation and corresponding ward heads. We estimated that a total of 30 doctors and 60 nurses would have worked in the wards where the wearable device was implemented and potentially used it in some form; therefore, we planned to recruit participants from this larger sample.

Data Collection Methods

Informal Discussions and Observations

We used an iterative process of engaging in informal discussions coupled with sense-checking discussions and observations during the implementation period. The informal and sense-checking discussions and observations were conducted with the team who was working directly in the wards, as well as with head nurses from the wards where the wearable device was being implemented. The informal discussions and observations were conducted during the implementation of the device.

Feedback Survey Form

We created the feedback form based on the components of the technology acceptance model (TAM) to assess the use and acceptance of the device. The TAM is used in a variety of disciplines to determine how individuals accept (or not) and use (or not) a given technology. Davis [17] developed this model based on components from the theory of reasoned action [18] and it consists of the following variables: use motivation (with perceived ease of use and perceived usefulness) and behavioral intention [17,19]. The model suggests that an individual will accept the use of a technology (ie, their behavioral intention) based on their perception of the technology's usefulness and ease of use. Perceived usefulness refers to the perception that using the technology will enhance one's work; for example, the wearable device will provide physicians and nurses some advantages (eg, remote monitoring). Perceived ease of use refers to the perception that the use does not add more work or effort to the work that could be enhanced; for example, using the wearable device will not increase nurses' workload, despite its utility and simplicity [17]. The TAM framework was expanded twice to include attitudes as well as several other external factors [20]. The use of the TAM in health research has shown how perceived usefulness and perceived ease of use relates positively to attitude and behavioral intention [21]. The TAM has been criticized for being insensitive to the context or social factors, being simplistic, and following an assumption that users are rational decision makers, when indeed other factors play into decision making [22-24]. We used the TAM framework for its simplicity and because the categories of perceived usefulness and perceived ease of use were of relevance, but we also integrated other data collection methods alongside it to counter these limitations to some extent.

Based on the components of the TAM, we included 23 questions related to usefulness (n=5), ease of use (n=5), attitude (n=5), and behavioral intent (n=8) [25]. We asked these questions using a 5-point Likert scale (with scores of 5 being more favorable). We also added 2 open-ended questions and collected a variety of relevant demographic information (Multimedia Appendix 1). We piloted the tool in both English and Vietnamese and adjusted the form as needed. We used Google forms for electronic self-completion of the form and offered paper forms for hand-written self-completion. We explained the feedback form to the ward staff during team meetings and provided the link. The feedback form was distributed and completed in Vietnamese. We kept the feedback form link open for 7 weeks in total and started data collection after the implementation had been integrated into the wards so that users would have had experience using the device.

User-Centered Workshop

We held a user-centered workshop with a selection of HTD ward staff to explore user needs and ideas for development in more detail. Because we already had the technology and knew the spaces where implementation would be held, we followed an adapted version of the process described by Cooper et al [26]. With this approach, the workshop participants and facilitators set the scene as a busy COVID-19 ward during the peak of the pandemic. Then, the facilitators described the shells of users (personas), including a nurse and a doctor persona shell, and we had the workshop attendees describe who they imagined the nurse and doctor to be, as well as their behaviors and needs and the values each user group would find most essential. We based the conversation on the wearable technology that the participants had already used. Then, the group discussed solutions to the issues identified [26].

Data Analysis

Using the principles of action learning, we integrated the responses from informal discussions and observations into subsequent data collection, as well as summarized the content and grouped it into themes. For the analysis of the feedback survey form, we calculated mean scores for each variable and compared scores by profession. For the open-ended survey questions, we used content coding to summarize the responses topically. We presented the demographic data descriptively. We documented the responses from the user-centered design workshop as notes and summarized the results into main themes.

Ethical Considerations

In this paper, we are describing the processes that occurred as part of the development and implementation of a monitoring system; therefore, the work did not require ethics approval. Prior to the initiation of any activities, we held a meeting with ward heads to describe the work in more detail and obtain their agreement.

Results

Device Implementation Within the HTD Context

The wearable device was implemented in 3 wards starting in August 2021, including the adult ICU, Ward A, and Ward E.

We describe the implementation over a 5-month period from August to December 2021. During this period, these wards changed from COVID-19–designated and then back again to routine patient care settings, depending on the number of patients. Although the HTD was one of the COVID-19–designated hospitals, throughout the pandemic they offered routine patient care for specific diseases (eg, tetanus).

In addition to the rapidly changing physical spaces, the hospital management quickly deployed remote monitoring capacity using existing closed-circuit television cameras as a temporary solution to monitor very sick patients from outside the patients' rooms. The remote monitoring was useful as it allowed for multitasking and prevented nurses and doctors from checking on patients more routinely in person. The hospital wards were at capacity during the study period. Prior to the pandemic, however, it was not unusual for the wards at the HTD to often be at maximum patient capacity. For example, in the adult ICU or during the rainy season, the number of dengue patients increases dramatically and the wards tend to be full.

Also, the workflow was organized differently during the pandemic period. Instead of nurses taking care of a few specific patients for the whole shift, 2 nurses and 1 doctor would instead go into the ward (in full personal protective equipment) as a team for 3 hours at a time while the other 2 nurses on shift completed admin work in the office. This meant that more coordination was needed, and often the team with the patients "need[ed] someone else to be [their] memory" as it was not easy to remember everything about all patients. The health care team's workload, especially that of the nurses, ended up being more extensive for many reasons. One important reason is that, because of COVID-19 restrictions, there were also no families allowed in the wards who would help to look after patients in non–COVID-19 times; therefore, the majority of the care was left to the nurses. The patients were also more severely ill than previously in these wards and required more care by fewer staff.

Device Use and Acceptability

When we first distributed the feedback form, out of 90 potential participants, only 22 completed the survey (19 electronic and 3 paper forms), and 1 person stated that they did not use the technology and therefore no responses were recorded for that participant. Of the 21 respondents who completed the feedback form, 48% (n=10) were doctors and 48% (n=10) were nurses, with 52% (n=11) of the participants coming from Ward E (Table 1).

Overall, when assessing the TAM variables, the mean (SD) scores ranged from 3.6 (0.8) for perceived ease of use to 4.4 (0.6) for attitude, with behavioral intention (mean 3.9, SD 0.6) and perceived usefulness (mean 4.2, SD 0.7) scoring in between. Those working as nurses scored higher on perceived usefulness, perceived ease of use, attitude, and behavioral intention than did physicians (Table 2).

When asked, as an open-ended question, why participants would or would not use the wearable device in the future, of the 19 responses inputted, 15 participants wrote that they would use the system because of its convenience and usefulness in monitoring patients. However, in 2 of those responses, they also added comments that the device had limited perceived accuracy and transmission problems. Of the remaining 4 participants, 1 participant simply stated that the monitor was still in use, 2 participants wrote that they did not use the system anymore due to job location changes, and 1 participant wrote a few sentences about why the wearable device is not the "best choice," highlighting its limited battery life, how the system had become additional work for the already overworked staff, and how it is not yet completely implemented.

Table 1. Demographic characteristics of the survey respondents (n=21).

Characteristic	Value
Gender, n (%)	
Women	13 (62)
Men	8 (38)
Age (years), median (IQR)	35 (30-38)
Occupation, n (%)	
Doctor	10 (48)
Nurse	10 (48)
Other: nurses' aid	1 (5)
Primary ward during the implementation phase, n (%)	
Adult intensive care unit	6 (29)
Ward A	1 (5)
Ward D	3 (14)
Ward E	11 (52)

Table 2. Mean technology acceptance model (TAM) scores by variable. The maximum score was 5.

TAM variable	All participants, mean (SD)	Nurses, mean (SD)	Doctors, mean (SD)
Perceived usefulness	4.2 (0.7)	4.3 (0.8)	4.1 (0.6)
Perceived ease of use	3.6 (0.8)	3.8 (0.7)	3.4 (0.8)
Attitude	4.4 (0.6)	4.6 (0.6)	4.2 (0.6)
Behavioral intention	3.9 (0.6)	4.0 (0.6)	3.7 (0.7)

Integrating User Perceptions for Improved Implementation

As part of the action learning process, we supplemented the feedback form results with data from the observations and informal discussions during the 5-month period. There were 3 main observations. First, there was a mismatch between how we (ie, the research team) and the ward teams perceived the use of the technology. We quickly realized, from our observations and from informal discussions with the implementation team, that many of the nurses either did not use the wearable device or did not think that they used it even if they used it in some aspect (eg, connecting the device for the patients or changing batteries). Even after we clarified what we meant by "use," there were still not additional participants who filled in the feedback form because they felt like they did not use the technology.

Second, the ward teams had varying perceptions of the technologies that are routinely implemented by the OUCRU team in the HTD wards as part of research projects. We heard from informal discussions with colleagues that the nurses assumed the wearable devices were from a research project, as is often the case with OUCRU projects, and therefore the nurses, in particular, ignored the device even if they had some role in its use. They did not see its potential benefit.

Finally, in order to make the device more useful for the ward staff, we realized during the meetings and informal discussions with the team that we needed to make the implementation and use of the device more "pro-nurse," meaning we would need to emphasize how the device and its data were also useful and relevant to them. When discussing with the head nurse, the data were only displayed on the main screen in the staff room for one department. One suggestion was to move the tablet to the wall so that the nurses and others in the room (including the patients) could potentially see their vital signs. Because the devices and corresponding data were not in sight, it was easy to think that it was not relevant for the nurses and made it easier for them to ignore the device while with the patients.

User-Centered Design Workshop

With the information we had learned from the informal discussions, observations, and feedback form, we held a follow-up workshop on January 17, 2022, to discuss how we could make better use of the technology in the wards in COVID-19 situations in the future. The attendees included 2 doctors (1 man and 1 woman) and 3 nurses (2 women and 1 man). The participants discussed the behaviors and needs of the nurse and doctor persona. For both roles, the needs centered on having equipment and improved coordination. The nurses also mentioned more training needs, while the doctors' needs were about the accuracy of monitoring (Textboxes 1 and 2).

There were 3 main value prop themes, including medical, technical, and patient themes. For medical aspects, the attendees discussed how the device should be able to provide highly accurate data, with appropriate alarms and cut-offs. For the technical theme, the device and software should be simple to connect and use, with a long battery life and stable connections during charging or switching devices. The display should be large and clear, and the data should be stored for a long period of time (ie, 7-10 days). Finally, for the patient theme, the device should be comfortable for the patients to wear to avoid them removing it.

There were several solutions discussed in the group to improve the use and efficacy of the wearable device (Table 3). Solutions included improving the credibility of the data, ideas to improve the ease of use, ways to make the alarms more consistent, and ideas for more ideal placement. One very specific issue that the group mentioned was that the alarms went off too much on the large display and the alarms were always red or black and blinking, and it was difficult to know if the device was turned off (due to patient discharge) or actually disconnected, which would require an intervention. The solution was to refresh the devices; however, if the alarms were excessive and not always indicating a real issue, trust in the device would remain low, so this was an important priority. They also suggested that the alarms and display on the tablet should be the same as the big screen, as they preferred screen consistency.

Another in-depth discussion was about moving the tablets to the walls and having the device plugged in all the time, which would solve the battery issues. They felt that the tablet could be set up on the wall but that brought up other issues about how to keep the device and watch safe after use. For some of the topics, the group used features of another wearable device that they had used in the wards in the past to inform their solutions (eg, device graphs and a line on the device for finger placement).

Textbox 1. Behaviors and needs of the nurses.

Participant: Nurse Van is a 36-year-old woman. She is an administrative nurse and has a management job. She likes to have fun and has a family and 2 children. She is also responsible for bringing the kids to school and back.

Behaviors:

- Visit and provide direct patient care and monitor vital signs
- Carry out medical orders (ie, medications, blood tests, and nutrition)
- Assess, monitor, and hand over patients
- Work night duty
- Night shifts inform doctors on vital signs as prescribed

Needs:

- Equipment (eg, to measure blood pressure, temperature, oxygen levels, and heart rate)
- Training on diseases
- Teamwork and coordination

Textbox 2. Behaviors and needs of the doctors.

Participant: Doctor Huong is a 30-year-old woman. She is flexible and very active. She is not married and has no children and currently lives in a hotel. She is on night shift every 4 nights, and at times she visits her home in another town in Ho Chi Minh City, which is far from the Hospital for Tropical Diseases.

Behaviors:

- Prescribe medications
- Update medical records
- Perform examinations and change treatments
- Data entry
- Check vital signs in patient rooms (with a portable monitor that they move around) for examination and to detect abnormalities

Needs:

- Equipment (eg, monitors)
- Coordination with nurses (progress: medical records)
- Re-evaluation and working with other doctors
- Accuracy of vital sign monitoring

Table 3. Solutions for improvement.

Topic	Specific solution
Data credibility	Adding a graph for signal strength
Ease of use	• Adding a finger placement mark on the device • Increasing the font size on the watch and tablet • Tablets should be fixed on the wall
Alarms	• Reduce the alarm colors and blinking on the screen • Use the same display on the screen and the tablets for consistency • Refresh the tablets for more accurate alarms
Battery issues	Keep the tablet plugged in
Device placement	Placement on the wall (but only with an increase in font size)

Discussion

The HTD and OUCRU teams, along with the technology company, rolled out the wearable device in an extremely complex pandemic situation with a prototype system. In the end, the team used the device on over 100 patients. We assessed the usability and acceptance of the device over the implementation period when COVID-19 cases were peaking in the hospital and into the period when the COVID-19 cases were reducing. Similar to the literature on the topics, we found that the importance of understanding the users and their experiences using the device was crucial to get the most use out of these technologies.

There was a mismatch between our perception of who was using the device and those who thought that they were using or benefiting from the device on the ground. From the start, the device was designed and set up with doctors in mind, but in practice, the nurses' roles and use were overlooked, even though they could also routinely use and benefit from the device. In our study, we found that the nurses who filled in the feedback form, on average, had slightly higher scores on 3 of the 4 TAM domains (ie, perceived usefulness, attitude, and behavioral intention), while the doctors, on average, scored the perceived ease of use slightly higher than the nurses. We know from the challenges with acquiring feedback that many nurses did not feel that they used the device even though they had some role in the device set-up and monitoring. Designing the device to be more nurse-centric from the early phases could have helped to increase the efficiency and definition of who is meant to use it. In the future, it is important to consider that the way the device is used might be dependent on the form of its use (eg, for triage, use in a pandemic emergency, or routine hospital use). We recommend the involvement of staff who could benefit from the technology, especially nurses in the hospital context, in the full implementation process. This could help to avoid mismatches in the perceptions of who the users are and who could and should benefit from the new technology. Research on integrating wearable devices during COVID-19 in Singapore also highlighted that device simplicity would encourage its use and the importance of making the technology fit into the current environment while not increasing or disrupting workflows [27].

The trust in the device and its data was an issue brought up several times during the implementation and feedback sessions. There are a variety of potential explanations for inconsistent data (eg, incorrect device placement or averaging of data); however, it reduced the credibility of the device for both doctors and, importantly, nurses. Data concerns about technology in clinic settings has been noted in other studies. For example, Faria et al [28] found that study clinicians reported that 36% of the data from a remote monitoring project were "invalid" for a variety of reasons, including low literacy of the patients and complexity of the device. Involvement from users from the beginning of the design and implementation process is crucial for design purposes but also to build trust and confidence in the devices [11]. While this project took place during COVID-19, which is a very specific circumstance, the broader findings resonate with research conducted prior to COVID-19 that focused on the implementation and scaling up of digital health technologies in LMICs. The recommendations also included integration of end-user feedback and engagement with all stakeholders throughout the design and implementation process [12]

There are limitations to this work. First, we did not collect data on the clinical worth or the accuracy of the data transmitted from the devices. Second, we focused on feedback from only heath care staff (ie, doctors and nurses), and from only a subset of those who perceived that they used the device, which may have excluded some users and limited the overall sample size. We did not include patients who could also inform device acceptance, especially if used in noncritical cases where patients are moving around and conscious. Finally, the implementation setting for this work is not typical of other hospital settings in Vietnam or possibly other LMICs, as the HTD is a large referral hospital with an international research institute attached to it.

In anticipation of future (novel) pandemic situations or integration of wearable technologies into a range of clinical settings more broadly, it is important to fully understand if and how the wearable devices could be used more effectively by doctors, and importantly, nurses in the wards, for monitoring of deteriorating patients, especially in LMICs where resources are already stretched. Using an action learning approach during the implementation process highlights the importance of integrating user perspectives, ideas, and solutions into development and design.

Acknowledgments

We would like to acknowledge Nguy n Đ i Phú and Nguy n Th Tin from the Hospital for Tropical Diseases and Nguy n Th Di m Trinh and Đ ng Ph ng Th o from the Oxford University Clinical Research Unit for their support and commitment in operating the remote monitoring system.

The members of the Vietnam ICU Translational Applications Laboratory (VITAL) group are as follows: Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam: An Phuoc Luu, Chanh Quang Ho, Duc Hong Du, Duc Minh Tran, Dung Thi Phuong Nguyen, Giang Thi Nguyen, Hai Bich Ho, Hien Van Ho, Hung Manh Trinh, Huy Quang Nguyen, Khanh Nguyen Quoc Phan, Khoa Dinh Van Le, Kien Trung Dang, Lam Khanh Phung, Lieu Thi Pham, Ngoc Thanh Nguyen, Nhat Tran Huy Phung, Phuong Thanh Le, Quyen Than Ha Nguyen, Thanh Thi Le Nguyen, Thy Bui Xuan Doan, Trieu Trung Huynh, Trinh Huu Khanh Dong, Van Minh Tu Hoang, Van Thi Thanh Ninh, Vuong Lam Nguyen, Yen Minh Lam, Sayem Ahmed, Joseph Donovan, Ronald Geskus, Evelyne Kestelyn, Angela Mcbride, Guy Thwaites, Louise Thwaites, Hugo Turner, Jennifer Ilo Van Nuil, and Sophie Yacoub; Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam: Tam Thi Cao, Thuy Bich Duong, Duong Thi Hai

Ha, Nghia Dang Trung Ha, Chau Buu Le, Thu Ngoc Minh Le, Thao Thi Mai Le, Tai Thi Hue Luong, Phu Hoan Nguyen, Viet Quoc Nguyen, Nguyen Thanh Nguyen, Phong Thanh Nguyen, Anh Thi Kim Nguyen, Hao Van Nguyen, Duoc Van Thanh Nguyen, Chau Van Vinh Nguyen, Oanh Kieu Nguyet Pham, Van Thi Hong Phan, Qui Tu Phan, Tho Vinh Phan, and Thao Thi Phuong Truong; University of Oxford, Oxford, United Kingdom: David Clifton, Mike English, Shadi Ghiasi, Heloise Greeff, Jannis Hagenah, Ping Lu, Jacob McKnight, Chris Paton, and Tingting Zhu; Imperial College London, London, United Kingdom: Pantellis Georgiou, Bernard Hernandez Perez, Kerri Hill-Cawthorne, Alison Holmes, Stefan Karolcik, Damien Ming, Nicolas Moser, and Jesus Rodriguez Manzano; King's College London, London, United Kingdom: Alberto Gomez, Hamideh Kerdegari, Marc Modat, and Reza Razavi; ETH Zurich, Zurich, Switzerland: Abhilash Guru Dutt, Walter Karlen, Michaela Verling, and Elias Wicki; The University of Melbourne, Melbourne, Australia: Linda Denehy and Thomas Rollinson.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study tools in English and Vietnamese.
[PDF File (Adobe PDF File), 240 KB - humanfactors_v11i1e44619_app1.pdf]

References

1. Channa A, Popescu N, Skibinska J, Burget R. The rise of wearable devices during the COVID-19 pandemic: a systematic review. Sensors (Basel) 2021 Aug 28;21(17):5787 [FREE Full text] [doi: 10.3390/s21175787] [Medline: 34502679]
2. Atta RM. Cost-effective vital signs monitoring system for COVID-19 patients in smart hospital. Health Technol (Berl) 2022;12(1):239-253 [FREE Full text] [doi: 10.1007/s12553-021-00621-y] [Medline: 34786323]
3. Iqbal FM, Joshi M, Davies G, Khan S, Ashrafian H, Darzi A. Design of the pilot, proof of concept REMOTE-COVID trial: remote monitoring use in suspected cases of COVID-19 (SARS-CoV-2). Pilot Feasibility Stud 2021 Mar 05;7(1):62 [FREE Full text] [doi: 10.1186/s40814-021-00804-4] [Medline: 33673868]
4. O'Carroll O, MacCann R, O'Reilly A, Dunican EM, Feeney ER, Ryan S, et al. Remote monitoring of oxygen saturation in individuals with COVID-19 pneumonia. Eur Respir J 2020 Aug;56(2):2001492 [FREE Full text] [doi: 10.1183/13993003.01492-2020] [Medline: 32616588]
5. Seshadri DR, Davies EV, Harlow ER, Hsu JJ, Knighton SC, Walker TA, et al. Wearable sensors for COVID-19: a call to action to harness our digital infrastructure for remote patient monitoring and virtual assessments. Front Digit Health 2020;2:8 [FREE Full text] [doi: 10.3389/fdgth.2020.00008] [Medline: 34713021]
6. Local production and technology transfer to increase access to medical devices addressing the barriers and challenges in low-and middle-income countries. World Health Organization. URL: https://www.who.int/publications/i/item/9789241504546 [accessed 2023-12-07]
7. Mitgang EA, Blaya JA, Chopra M. Digital health in response to COVID-19 in low- and middle-income countries: opportunities and challenges. Glob Policy 2021 Jul;12(Suppl 6):107-109 [FREE Full text] [doi: 10.1111/1758-5899.12880] [Medline: 34230840]
8. Un K, Wong C, Lau Y, Lee JC, Tam FC, Lai W, et al. Observational study on wearable biosensors and machine learning-based remote monitoring of COVID-19 patients. Sci Rep 2021 Feb 23;11(1):4388 [FREE Full text] [doi: 10.1038/s41598-021-82771-7] [Medline: 33623096]
9. Rosic T, Petrina N, Baysari M, Ritchie A, Poon SK. Patient and clinician use characteristics and perceptions of pulse oximeter use: a scoping review. Int J Med Inform 2022 Mar 18;162:104735 [FREE Full text] [doi: 10.1016/j.ijmedinf.2022.104735] [Medline: 35325661]
10. Binyamin SS, Hoque MR. Understanding the drivers of wearable health monitoring technology: an extension of the unified theory of acceptance and use of technology. Sustainability 2020 Nov 18;12(22):9605. [doi: 10.3390/su12229605]
11. Morcillo LF, Morer-Camo P, Rodriguez Ferradas MI, Cazón Martín A. The wearable co-design domino: a user-centered methodology to co-design and co-evaluate wearables. Sensors (Basel) 2020 May 21;20(10):2934 [FREE Full text] [doi: 10.3390/s20102934] [Medline: 32455756]
12. Labrique AB, Wadhwani C, Williams KA, Lamptey P, Hesp C, Luk R, et al. Best practices in scaling digital health in low and middle income countries. Global Health 2018 Nov 03;14(1):103 [FREE Full text] [doi: 10.1186/s12992-018-0424-z] [Medline: 30390686]
13. Thai PQ, Rabaa M, Luong DH, Tan DQ, Quang TD, Quach HL, et al. The first 100 days of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) control in Vietnam. Clin Infect Dis 2021 May 04;72(9):e334-e342 [FREE Full text] [doi: 10.1093/cid/ciaa1130] [Medline: 32738143]
14. Revans R. Action learning: tts originsnature. In: Pedler M, editor. Action Learning in Practice 4th edition. London: Routledge; 2011.
15. Brook C. What does action learning look like today? Action Learn Res Pract 2021 Jul 31;18(2):89-90. [doi: 10.1080/14767333.2021.1935058]

16. Boak G. Action learning and healthcare 2011–2022. Action Learn Res Pract 2022 Oct 17;19(3):251-268. [doi: 10.1080/14767333.2022.2133376]
17. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
18. Fishbein M, Ajzen I. Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research. Reading: Addison-Wesley; 1975.
19. Hong X, Zhang M, Liu Q. Preschool teachers' technology acceptance during the COVID-19: an adapted technology acceptance model. Front Psychol 2021;12:691492 [FREE Full text] [doi: 10.3389/fpsyg.2021.691492] [Medline: 34163416]
20. Lai P. The literature review of technology adoption models and theories for the novelty technology. JISTEM 2017 Jun 08;14(1):21-38. [doi: 10.4301/s1807-17752017000100002]
21. Tsai T, Lin W, Chang Y, Chang P, Lee M. Technology anxiety and resistance to change behavioral study of a wearable cardiac warming system using an extended TAM for older adults. PLoS One 2020;15(1):e0227270 [FREE Full text] [doi: 10.1371/journal.pone.0227270] [Medline: 31929560]
22. Legris P, Ingham J, Collerette P. Why do people use information technology? A critical review of the technology acceptance model. Inf Manag 2003 Jan;40(3):191-204. [doi: 10.1016/s0378-7206(01)00143-4]
23. Benbasat I, Barki H. Quo vadis, TAM? J Assoc Inf Syst 2007;8(4):211-218. [doi: 10.17705/1jais.00126]
24. Bagozzi R. The legacy of the technology acceptance model and a proposal for a paradigm shift. J Assoc Inf Syst 2007 Apr;8(4):244-254. [doi: 10.17705/1jais.00122]
25. Venkatesh V, Bala H. Technology acceptance model 3 and a research agenda on interventions. Decis Sci 2008 May 09;39(2):273-315. [doi: 10.1111/j.1540-5915.2008.00192.x]
26. Cooper A, Reimann R, Cronin D, Noessel C. About Face: The Essentials of Interaction Design. 4th edition. Indianapolis: Wiley; 2014.
27. Fan EMPJ, Ang SY, Phua GC, Chen Ee L, Wong KC, Tan FCP, et al. Factors to consider in the use of vital signs wearables to minimize contact with stable COVID-19 patients: experience of its implementation during the pandemic. Front Digit Health 2021;3:639827 [FREE Full text] [doi: 10.3389/fdgth.2021.639827] [Medline: 34713111]
28. Faria I, Gaspar C, Zamith M, Matias I, das Neves RC, Rodrigues F, et al. TELEMOLD project: oximetry and exercise telemonitoring to improve long-term oxygen therapy. Telemed J E Health 2014 Jul;20(7):626-632. [doi: 10.1089/tmj.2013.0248] [Medline: 24796364]

Abbreviations

HCW: health care worker
HTD: Hospital for Tropical Diseases
ICU: intensive care unit
LMIC: low- and middle-income country
OUCRU: Oxford University Clinical Research Unit
TAM: technology acceptance model
VITAL: Vietnam ICU Translation Application Laboratory

Edited by A Kushniruk; submitted 28.11.22; peer-reviewed by K Liu, M Harris, J Claggett; comments to author 15.04.23; revised version received 15.06.23; accepted 20.11.23; published 05.01.24.

Please cite as:
Luu AP, Nguyen TT, Cao VTC, Ha THD, Chung LTT, Truong TN, Nguyen Le Nhu T, Dao KB, Nguyen HV, Khanh PNQ, Le KTT, Tran LHB, Nhat PTH, Tran DM, Lam YM, Thwaites CL, Mcknight J, Vinh Chau NV, Van Nuil JI, Vietnam ICU Translational Applications Laboratory (VITAL)
Acceptance and User Experiences of a Wearable Device for the Management of Hospitalized Patients in COVID-19–Designated Wards in Ho Chi Minh City, Vietnam: Action Learning Project
JMIR Hum Factors 2024;11:e44619
URL: https://humanfactors.jmir.org/2024/1/e44619
doi:10.2196/44619
PMID:38180799

Original Paper

Exploring a Gaming-Based Intervention for Unemployed Young Adults: Thematic Analysis

Ingjerd Jevnaker Straand[1], MSc; Asbjørn Følstad[2], PhD; Jone Ravndal Bjørnestad[1], PhD

[1]Department of Social Work, University of Stavanger, Stavanger, Norway

[2]Department of Sustainable Communication Technologies, SINTEF Digital, Oslo, Norway

Corresponding Author:
Ingjerd Jevnaker Straand, MSc
Department of Social Work
University of Stavanger
Kjell Arholms hus
Stavanger, 4021
Norway
Phone: 47 93 22 22 89
Email: ingjerd.j.straand@uis.no

Abstract

Background: Promoting positive psychologies that promote resilience such as a growth mindset could be beneficial for young, unemployed adults, as many lack the self-esteem and self-efficacy to cope with job search adversity. These young people may be reached at scale through the web-based delivery of self-administered positive psychology interventions. However, past studies report unsatisfying user experiences and a lack of user engagement. A gaming-based experience could be an approach to overcoming these challenges.

Objective: Our research objective was to explore how young, unemployed adults experience a positive psychology intervention designed as a game to extract learning and principles for future intervention research and development.

Methods: To respond to the research question, a team of researchers at the University of Stavanger worked with designers and developers to conceptualize and build a gaming-based intervention. Feedback from the users was collected through formative usability testing with 18 young adults in the target group. Retrospectively, recordings and notes were transcribed and subjected to thematic analysis to extract learnings for the purposes of this paper.

Results: A total of 3 themes were identified that pinpoint what we consider to be key priorities for future gaming interventions for unemployed young adults: adaptation to user preferences (eg, need for responding to user preferences), empathic player interaction (eg, need for responsiveness to user inputs and a diverse set of interaction modes), and sensemaking of experience and context (eg, need for explicit presentation of game objectives and need for management of user expectations related to genre).

Conclusions: Feedback from end users in usability-testing sessions was vital to understanding user preferences and needs, as well as to inform ongoing intervention design and development. Our study also shows that game design could make interventions more entertaining and engaging but may distort the intervention if the game narrative is not properly aligned with the intervention intent and objectives. By contrast, a lack of adaptation to user needs may cause a less motivating user experience. Thus, we propose a structured approach to promote alignment between user preferences and needs, intervention objectives, and gameplay.

(JMIR Hum Factors 2024;11:e44423) doi:10.2196/44423

KEYWORDS

positive psychology intervention; digital mental health; serious gaming; intervention design; research through design; gaming-based intervention

Introduction

Positive Psychology for Unemployed Young Adults

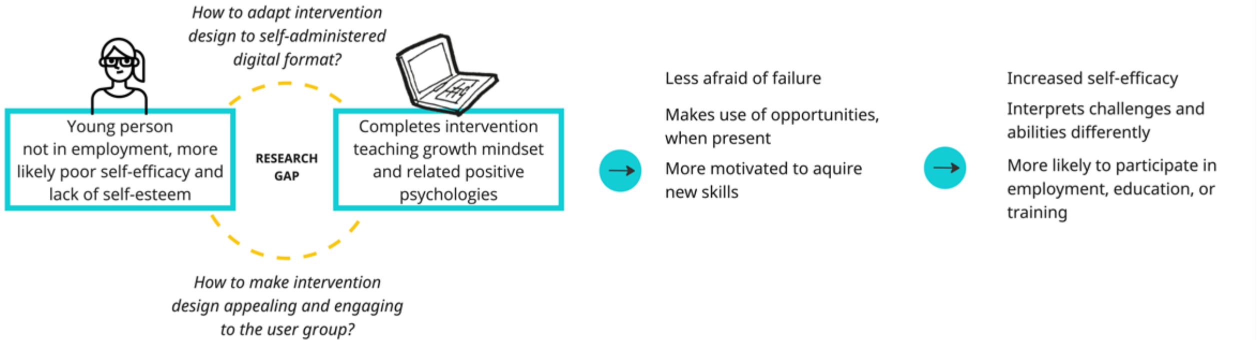
Young people who are not in education, employment, or training (NEET) comprise, on average, 12.8% aged between 15 and 29 years in the Organisation for Economic Cooperation and Development countries [1]. Studies show that negative self-perceptions and a lack of perseverance are barriers to successful labor market inclusion [2-4], as the new labor market requires highly skilled workers who are not afraid of change, challenges, and acquiring new skills [5,6]. For young people with weak beliefs in their capacity to learn, this could be a major risk factor for labor market exclusion, and this may in turn impact their overall well-being. Several researchers have studied the relationship between unemployment and mental health. McGee and Thompson [7] found a relationship between unemployment and depression in young adults and suggested the use of psychological interventions for the young and unemployed. The Norwegian NEET group is more likely to be recipients of health-related benefits, have poorer mental health, and lower levels of education compared with the average of the Organisation for Economic Cooperation and Development [8,9]. A qualitative inquiry into young people's own experience of unemployment in Norway points to poor self-efficacy and lack of self-esteem that are reinforced through challenges and setbacks, even when these initially occur beyond the individual's control [10,11], such as when there are insufficient training placements on offer for the vocational school pupils, a problem leading to a relatively large number of unqualified school dropouts in Norway [9]. Thus, there is a substantial rationale for exploring further how the public can offer training, not only in job-seeking skills, such as curriculum vitae (CV) writing and gaining work skills, but also in building psychological well-being and resilience to cope with such setbacks and challenges [12,13].

In the context of a broader research project, the Career Learning App, our study investigates the design and development of a web-based intervention using positive psychology to achieve beneficial changes [14]. Our broader research idea is that young people in the NEET group, henceforth referred to as *young, unemployed adults* (for the sake of simplicity and to reduce stigma), could benefit from building confidence in the possibility of learning and improving. The research idea stems from a body of work that has demonstrated positive results from offering high school students self-administered positive psychology interventions (PPIs) centered on growth mindset and challenge-seeking behaviors [15,16]. A growth mindset is the belief that human capacities are not fixed but can be developed and increased in response to one's own efforts, good strategies, and help from others [17]. If a simple web-based PPI can influence high schoolers' mindsets in ways that lead to positive academic outcomes [15,16,18-20], then it could also likely be beneficial to the young unemployed, leading to changes in how they engage with their contexts. Despite this strong rationale for the applicability of PPIs to facilitate well-being and personal growth in vulnerable populations, they have only, to a limited extent, been tested and used in the context of unemployment [13,21]. However, we cannot simply apply the PPIs designed for educational contexts; they need substantial adaptation to be relevant or usable for this new target group of young, unemployed adults [17]. For instance, the school-related examples used within the PPI to make them relatable are not relevant to this new target population. Furthermore, there is a lack of shared context to piggyback on to deliver the intervention and ensure that users will adhere to it. Thus, there is a need to design and develop a web-based PPI designed specifically for young, unemployed adults and their context. If successful in user studies, a resulting intervention may be used in forthcoming large-scale randomized controlled trials in Norway.

Problems With Self-Administered Interventions

Self-administered web-based interventions have the potential to support well-being and positive health changes in a large number of people at a moderate cost [22]. However, this introduces new challenges, illustrated in Figure 1. First, there is the challenge of adapting current PPIs to self-administered digital formats that are fit for the purpose of the intended user population. Second, there is the challenge of user motivation to obtain the users to complete and adhere to the intervention [23,24]. Past research suggests that users are not interested in or do not enjoy using digital mental health interventions [25], suggesting a need to work on the actual interventions themselves to increase engagement and user motivation.

Figure 1. How mindset change may positively impact young, unemployed adults and the challenges of adherence, which we see as related to a research gap with a lack of knowledge of intervention designing.



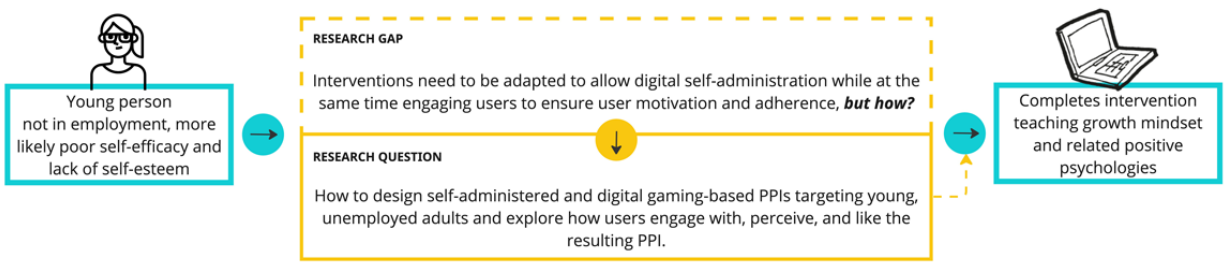
Exploring PPI as Gameplay to Be Relevant for Young People

Past research suggests a need to adapt to the media preferences of young people and make the apps more visual and interactive to increase engagement and motivation among young people [26]. One possible approach to increasing engagement is to explore games and game elements. Starting from "where the young people are at" makes pedagogical sense [27], thus the application of game design is founded on young people's own interests as a way to foster engagement and learning of positive psychologies. Although play and games are not unique to young humans [28,29], the average age of video game players is now 33 [30]. However, playing video games continues to be popular among young people [30,31]. Interactive digital games are increasingly used for purposes beyond entertainment, as exemplified by the rise of health gaming apps for video gaming consoles. Game design elements are also increasingly applied to nongame contexts, for instance, by adding points and badges to nongame experiences, such as social media networks [32,33] or learning contexts [34]. When game design and game concepts are being applied for purposes beyond fun, they may be termed "serious games," "learning games," or "gamification" [35-40]. Game design applied to learning may be seen as a form of experiential learning (eg, learning-by-doing) [41,42]. Game design has been successfully applied to mental health interventions [43,44] and educational contexts [34,38] in the past. Game design offers an approach to creating engaging experiences. Engagement is a complex and ambiguous term [45]. Our use of the term is in the sense of "emotional involvement" as in offering a pleasurable experience [32] and to describe how motivational, usable, and acceptable [46] the game would be in the eyes of the target audience, because this could be an important predictor of adherence. In general, there is insufficient research on the application of gaming and gamification to mental health, particularly in the well-being domain [47], and we have not found empirical studies that pursue to gamify positive psychology targeted specifically toward unemployed young people. There were no available gaming-based PPIs that could be used for the purposes of this study.

Research Objective

Our limited knowledge of how to adapt the intervention from an educational setting to a game-based format suggested a highly explorative approach, where we identified the need to design a game to explore this topic and to overcome the gaps in knowledge summarized in Figure 2. Thus, the research objective and question of this study were how to design a self-administered and digital PPI in a gaming format targeting young, unemployed adults and to explore how they engage with the game and whether they like using it.

Figure 2. How the research question is linked with the research gap identified. PPI: positive psychology intervention.

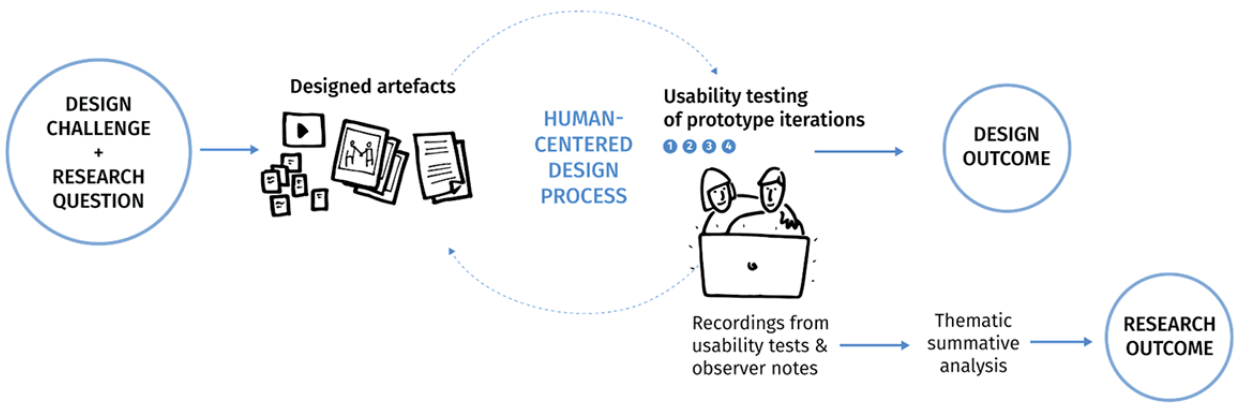


Methods

Setting: Learning in "Action"

To answer our research question, this study used a human-centered design process [48,49], an approach that allows input from target users during design and development. The human-centered design process is a form of research with design as the primary outcome [50]. The collection of feedback from users during design and development impacts not only the design of the game but also the production of knowledge. As such, it is a form of participatory action-research [51,52] where "doing research" and "doing action" happen simultaneously. In the design research literature, this may also be referred to as "Research through Design" [53,54], where knowledge is produced through the design of the artifact and through the experience of the artifact. At the end of the project, we analyzed user feedback data thematically [55] to extract the learnings and design principles for the potential application toward creating user-friendly and user-relevant gaming-based PPIs for vulnerable populations. Figure 3 summarizes the "Research through Design" approach of this study.

Figure 3. A Research through Design process with design outcomes and research outcomes.



Data Collection Through User Testing

We used testing with users to capture the user experience. Testing with users is important as a tool for obtaining the “design right” in game design [56]. Our project-specific usability study [57] was formative, and we were qualitatively evaluating early prototypes to answer how and why questions as a means to improve the design [58,59]. We aimed to capture participants’ *thinking*, for example, opinions, reasoning, and attitudes toward the prototype experience. The various prototypes thus functioned as exploratory “hypothesis testing” [60,61] and as boundary objects [62-64] that framed conversations with end users. These sessions with the end users lasted from 45 to 90 minutes, where users were asked to briefly describe their background and interests, followed by open “think-aloud” questions [65] related to a prototype experience, such as “What do you think about what you see here?” “What do you expect will happen now?” and “What do you think this is?” Thus, the participants were encouraged to verbalize their thoughts and experiences. Afterward, the participants were asked follow-up questions that were equally open, such as “What are you thinking now that you have seen this?” and ‘How would you describe this to a friend?” The objectives of usability testing were to collect feedback related to broad aspects of the intervention experience, namely its (1) engagement, (2) relatability, (3) understandability, and (4) potential for improvement.

Sample and Recruitment

The NEET group includes anyone not in economic activity from the age of 15 to 29 years [1]. The European Foundation for the Improvement of Living and Working Conditions has defined 7 subgroups of NEETs [66], ranging from the “classically unemployed” to young people who are caretakers, unable to work, or simply listed as “inactive.” How long an individual remains a NEET also varies significantly. The diversity of the population is not necessarily problematic for our study; when designing with users, one would usually strive for variance [49,67] rather than representativeness. Our study is exploratory and does not require large samples [68]. Our estimate required 20 users; however, this sample size was highly approximate, in line with qualitative studies in general [69,70]. Pragmatic needs in the design process guided the number of participants to a large degree and not, for instance, theoretical saturation.

We recruited from the Norwegian Labour and Welfare Administration (NAV), the national welfare institution of Norway that pays out unemployment wages and social support, and from a regional Individual Placement and Support (IPS) program, which offers placement support to young people with first-episode psychosis. We also recruited through the user testing platform, Teston (UserTesting). Our inclusion criteria were an age range of 18 to 29 years, with a “NEET background,” and because of the language in the game prototype, living in Norway and speaking Norwegian. We say “NEET background” and not “NEET status” because our participants from the IPS program were no longer in the NEET group by definition. We made a deliberate choice not to exclude based on the length of NEET status and unemployment. Although those who are entering the NEET group in the short term, the “in-betweeners” [66], often find new employment without assistance [71] relatively quickly, even a short time out-of-work may increase the risk of exclusion [72]. We did not recruit participants on permanent disability allowance. Using multiple channels enabled quicker recruitment during the COVID-19 pandemic and increased the variance in our sample (as desired), as young people without the rights to receive unemployment benefits have fewer incentives to register with NAV [73]. We did not compare the experience of the intervention based on the recruitment channel because the groups were overlapping and experienced varying prototypes depending on the stages in the design process.

Participants

In total, 18 participants (12/18, 67% females, 6/18, 33% males) took part in the study during the 21 testing sessions; thus, some participants were involved more than once. Recruitment was particularly challenging because of the COVID-19 pandemic, and we found that it was difficult to recruit young, unemployed men. Remote participation through web-based technologies, such as the Zoom (Zoom Video Communications) platform, enabled the study to continue during the lockdown. We also experienced that this user group continued to prefer remote participation, even when restrictions were lifted. Table 1 summarizes the participant statistics and format of the usability test.

Table 1. Age distribution of participants and format of user testing.

Recruitment channel	Format user testing	Age (years)[a]
IPS[b]	In person+Remote	19
IPS	In person+Remote	28
IPS	In person+Remote	18
IPS	In person+Remote	27
IPS	In person	22
IPS	Remote	19
IPS	Remote	18
IPS	Remote	23
Teston	Remote	18-29
Teston	Remote	18-29
Teston	Remote	18-23
NAV[c]	Remote	21
NAV	Remote	18
NAV	In person	21
NAV	Remote	21
NAV	Remote	18
NAV	Remote	18
NAV	In person	22

[a]For 3 (17%) of the 18 participants, we only had an age interval provided to us.

[b]IPS: Individual Placement and Support.

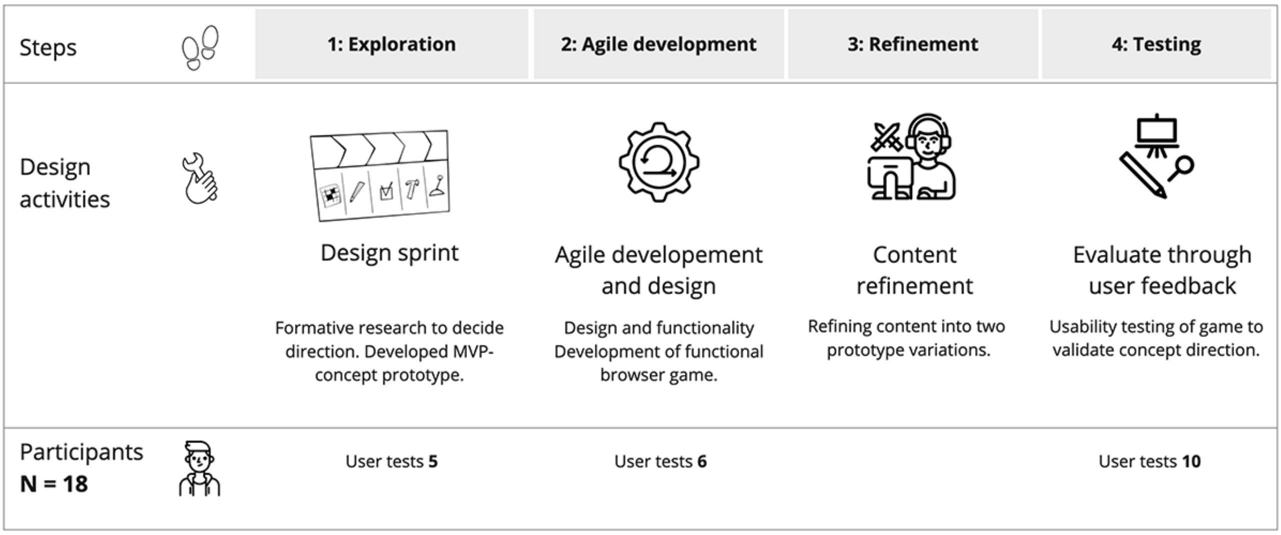
[c]NAV: Norwegian Labour and Welfare Administration.

The Design and Development Process

Researchers at the University of Stavanger worked with designers and developers from a consulting company and potential future end users in an agile [60] and human-centered design [48,49] process. The objective of the process was to create an enjoyable gaming-based PPI targeting unemployed young adults. This process took place in 4 steps over approximately 9 months, from 2019 to 2020. In the first step, (1) *design exploration* in the form of a design sprint, a 5-day design and prototyping process [61], produced a minimum viable product of a gaming concept that was tested with 5 participants. Following a brief period for planning and procurement, we moved on to (2) *agile development*, consisting of 3 sprints, each lasting about a month. During this step, usability tests were conducted on 6 participants. This was followed by (3) *refinement* of content and prototypes, with involvement from behavioral intervention researchers to further develop and "add-in" the necessary intervention content. Finally, the prototypes were evaluated using (4) *testing*. In this step, 10 participants participated in usability testing and provided feedback on the final set of prototypes. Figure 4 summarizes the stepwise design and research process.

The intervention content needed to be adapted to be meaningful and relatable to the user group [74]. The basis for our gaming-based PPIs were growth mindset interventions from the "National Study Learning Mindset" [16] and its translated Norwegian version, "U-SAY" [5,15]. These are interventions that target high school students [15,16]. We added selected parts of cognitive behavioral therapy [75], specifically management of negative emotions, panic, and anxiety, to offer a more productive interpretation of stressors [76] that may occur during job search adversity [77,78]. During the first few days of the initial design sprint, a gaming concept, VitaNova, was developed where players can build a "new life" in a fictional narrative.

Figure 4. The stepwise design and development process.



The Gaming-Based Intervention

The VitaNova gaming concept is a one-player fantasy game where the user plays a "no name," an android character that can take on any skin to morph into another character with their skills and abilities. In particular, the player can choose between being Noomi and Twizzlesprock. However, as we learn through the game, your character has a backstory as the male character Abel, a former engineer and an outcast who sees himself as a failure. The game is designed in 3D and split into episodes (missions). The game starts with very little information and instructions, landing the user right into action. This was a design tactic to spark curiosity and make the users intrigued by the game so that they would want to explore it further. As a player, your first task is to escape from captivity, and then, gradually, more and more information is being revealed to you as the game progresses. Refer to Textbox 1 for an overview of the game narrative.

The visual design of games can influence how motivational and acceptable they are to the target population [46]. Therefore, to make the game look polished, cool, and professional and to keep users immersed and engaged [79], emphasis was placed on 3D design and detailing. The game was divided into episodes, or missions. The first 3 episodes were developed into a nearly fully functional game in Unity WebGL, a platform for building 3D games that can be used in a web browser. During user testing, we also showed the prototypes and the wireframes that were made in Figma. Figure 5 shows the prototype iterations of the game design.

Psychological content and tasks are entered into the gameplay to foster psychological well-being, teach a growth mindset, and offer psychoeducation and mental health tips. Some of this is interwoven into action in the form of interactive quizzes, dialogues, or other forms of interaction, such as a CV builder applied within the game. This was intended to be transferable to the end user situation to increase relevance, although the acquisition of such practical skills was not a target of the intervention. Furthermore, there was also psychological content that was external to the gameplay, such as embedded videos. When using externally sourced content, this was implemented in the game as "ruins from the past," which the player could "find" in the game. The player would need to watch this content and use the information provided to complete the challenges and the in-game quizzes. Upon completing an episode of the game, the player was requested to write an answer to a reflective question where the user should answer as himself or herself, to encourage internalization of the messages that had been taught in the intervention through self-persuasion [80], and to transfer learning to the user's own situation. We have included a further description of the prototypes in Multimedia Appendix 1.

Textbox 1. Description of the game narrative. (The 2 final missions were not included in the user testing.)

1. Introduction

Wake up in the trunk of a moving vehicle. Use hacking skills to hack the lock. Find a tavern and interact with a bodyguard who refuses your entry as "no name." Find the Noomi skin and power unit and turn them on to enter the tavern.

2. Tavern

Interact with Griff and Mia at the tavern to learn about this world. Try to receive help. Tell them that you must go to an old public office to pick up energy bars if you are going to receive any help.

3. Learn

Find an abandoned public office where, among other things, you will discover many pieces of ancient psychological knowledge as ruins from the past.

4. The sidekick

Use newly acquired knowledge to help out the depressed and anxiety-ridden droid Griff, who now becomes your sidekick, and help Mia the bartender.

5. The bully

A bounty hunter is out looking for you to receive a reward from the boss at BetterJu Janus. Threatens your new friends at the tavern. Why are they after you?

6. The chance

You discover that the only way to obtain all the answers is to try to join into BetterJu. Your friends tell you of a job posting that is open. Interactive curriculum vitae and job application process.

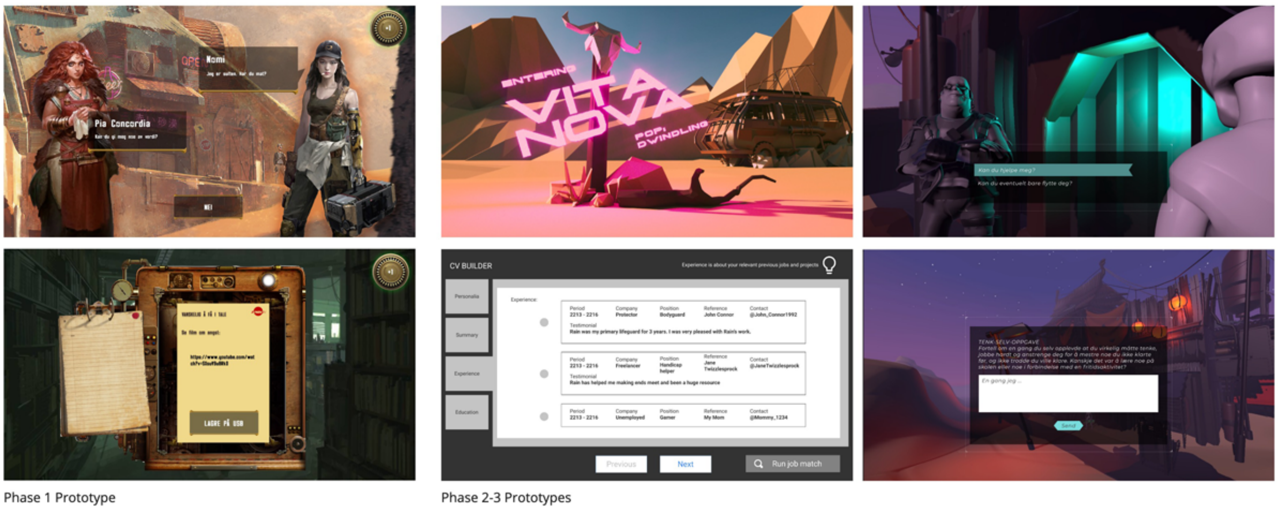
7. The job interview

The job requires a different set of skills, which you acquire through entering the skin of Twizzlesprock. Job interview at the company BetterYou as Twizzlesprock.

8. The escape

Discover who you really are from the overhearing conversation between your new boss, Janus, and a droid. Find your ex-girlfriend, who has been trapped but confirms your true identity. You both escape from the evil boss. The end.

Figure 5. Prototype iterations of VitaNova.



Analysis

The purpose of the usability testing was to synthesize findings that led to improvements and changes in the designed outcome. Retrospectively, we also conducted thematic analysis with the steps from Braun and Clark [55,81] as a practical guide: (1) data familiarization, (2) initial code generation, (3) search for themes, (4) review themes, (5) define and name themes, and (6) produce reports [81]. All 21 usability-testing sessions were recorded. The 7 most comprehensive usability tests were transcribed verbatim. The remaining data were analyzed based on recordings, researcher notes, and memos. Specifically, we used the transcribed data as our starting point and went back to recordings and memos to review codes and themes. The analytic process was iterative and creative, where we often moved back and forth between the data and codes [82]. All authors independently familiarized themselves with the data. Author 1 started with coding using the qualitative analysis software ATLAS.ti (ATLAS.ti Scientific Software). As a group, we discussed the findings and the initial codes in a workshop before moving over to paper-based coding and printing quotes from participants organized on large paper sheets. We used diagramming, both digitally in Miro and in pen-and-paper sketches, to iterate themes and review codes in consensus meetings. Findings and concepts were discussed with other researchers, some of whom had acted as observers for the user testing or had watched recorded sessions. The quality of the analysis was ensured by researcher reflexivity, end user involvement, and method triangulation. *Researcher reflexivity* concerns activities that consider how researchers might have informed the research or biased outcomes [69]. Reflexivity was enabled through critical discussion of assumptions, themes, and codes in the team of researchers. The team also involved researchers not involved in the user testing or in the design process as an approach to validate the analysis based on these methods. *End user involvement* was supported through the iterative design process, where participant perspectives were sought at different levels of concept and design maturity. Specifically, we found it valuable to involve users from different recruitment sources (IPS, Teston, and NAV) to strengthen the credibility and transferability of the findings. *Method triangulation* was conducted by applying different approaches to design and user involvement at different phases of the process, allowing the assessment of themes or constructs from different perspectives. In particular, including data from the different phases of exploration, design, and evaluation was found to strengthen the credibility of the findings.

Ethical Considerations

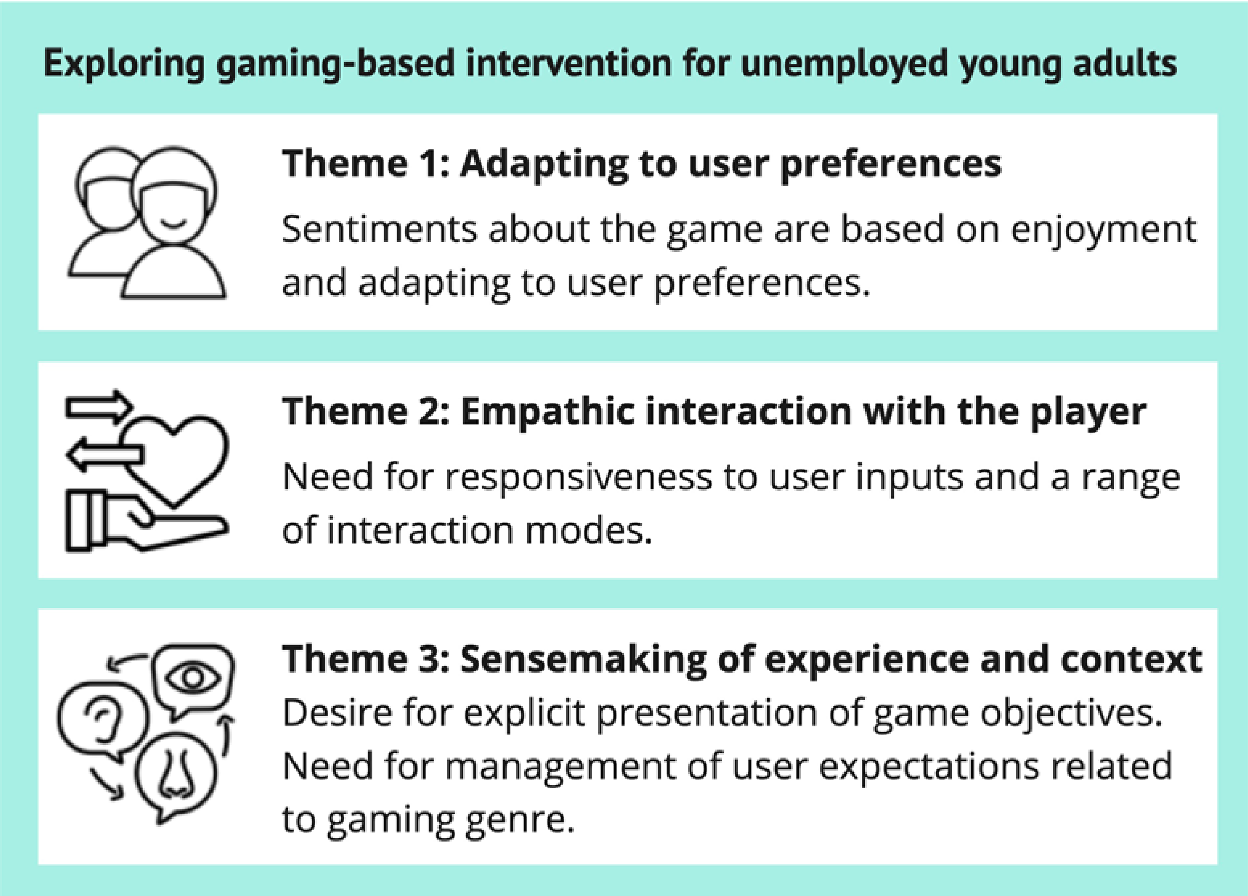
All participants provided explicit and written consent to participate in the study and were rewarded with gift cards (approximately US $30/session) for their participation. The study was evaluated and approved by the Norwegian Centre for Research Data (approval number 131074) and the regional committees for medical and health research ethics in Norway (approval number 42128).

Results

Overview

As a tool to motivate and help young adults engage in work or education, the idea of an interactive and digital game was regarded by all participants as a "good idea"; it was described as "cool," "unexpected," and "motivating" upon first impression. Upon closer experience with the different gaming prototypes, we received different and more specific feedback. A total of 3 themes were constructed by the researchers through active analytic engagement with the data [55]. The themes pinpoint what we consider key priorities for future gaming interventions for unemployed young adults: (1) adapting to user preferences, (2) empathic player interaction, and (3) sensemaking of experience and context. Refer to Figure 6 for an overview of the themes and their key characteristics.

Figure 6. Themes and theme characteristics.



Theme 1: Adapting to User Preferences

Sentiments about the game are based on enjoyment of the game experience, the genre, and whether the game meets user preferences, either by being targeted toward them or by allowing for experience customization. The participants expressed positive feelings and excitement related to the game. It was described as much more of an *actual game* than was expected:

> *I liked it; it was unexpected.* [P3]

> *Seems like a cool concept. Never heard of it before.* [P17]

The participants pointed to the use of humor and compared the game to pure entertainment media, such as commercially available games, or other parts of popular culture, such as films or television series:

> *It was actually pretty exciting [laughs]. Kind of funny, considering that you have included the welfare administration here.* [P6]

> *The story was funny, it seemed a bit like a video game.* [P10]

The game concept was further described as “something to do” or “something to allow the time to fly,” as pure entertainment or for relaxation purposes. The relaxing features were described as something that could make one more receptive:

> *If it’s like that you get new assignments once a week, then it can seem exciting. At least that is something to do.* [P7]

> *So you are pretty relaxed when you get these questions, so then it is probably a little easier to answer... a little easier to reflect over this.* [P4]

Others felt that the gameplay story could motivate them to do something by creating a sense of urgency, where the story would drive them to make more effort. A participant wanted proof within the game that it would be worth the effort to perform mundane tasks, such as updating their CV:

> *That’s how it is in life too: You have to make an effort yourself to get ahead; when you have to do something here [...], then you have to hurry, because someone is after you. Then you get the adrenaline to do it.* [P16]

> *If there’s something in the game that can prove to me that it’s worth it, like writing a resume is worth it....* [P18]

However, several participants expressed uncertainty about whether the game would be “something for them,” suggesting either other preferences or not quite the right conceptual fit. In particular, the concept likely needed “aging up,” as it was perceived as something for a younger population. Furthermore, we learned from the participants, who spent more time gaming, that they had started out playing “adventure games” and “role-playing games” when they were younger, but that they had since moved on to playing “first-person shooter” games or other kinds of games:

> *It’s not a game I would have bought in the store.* [P15]

> *I think if I had been younger then, yes. Because I have a lot of different types of games that I like to play, and now I like to play games where you shoot people, but earlier I liked playing games like that, where you follow a story, for example, it’s very different in... it’s very different for people what kind of games they like to play.* [P2]

Designing for mobile phone use was considered important and was brought up as an improvement suggestion by nearly all users. Participants expressed that mobile phone use would make it easier to meet their own user preferences or the user preferences that they expected other young people to have:

> *I would have chosen an app or a mobile game. Or a course, if it was on [a] mobile [phone].* [P7]

> *Not everyone has a PC with them everywhere, so I wondered if this was on mobile.* [P17]

> *If the game could be turned into a mobile game, that would be better.* [P1]

Adapting to user preferences could also mean designing a customizable or more personalized experience. In particular, the choice of characters in the game is usually an arena for customization and personalization. One user commented that she would prefer to customize a character by selecting hair length, and body shape, etc rather than choosing between predefined female and male characters:

> *if you can choose male or female or... you don’t have to have these two, but you can choose what that person looks like. Since now a lot of young people... there are some young people who don’t want to be a man or a woman, so I think you... it’s very smart to make something like ‘do you want long hair’ ‘do you want short hair’ ‘do you want...’ [laughs]* [P2]

Others mentioned different strategies for choosing characters by either choosing a character that resembled themselves or identifying with them. Others would deliberately choose the extreme opposite of themselves. As one participant mentioned, “if your choice was the big male character, then you were likely ‘more vulnerable on the inside’ and ‘in need of protection.’”

Theme 2: Empathic Interaction With the Player

This theme describes the need for responsiveness to user inputs and the desire for a range of interaction modes. Although initially intrigued and enthused by the game design, users quickly became disappointed by the lack of functionality. Thus, this theme is based on the need for empathic interaction with the player in the game, where the game needs to take the user seriously by being responsive to user inputs, thereby allowing for actual contribution to the experience:

> *It’s very... you can see very easily that your answers don’t make much of a difference. It doesn't matter what you choose. And I think if you’re going to have a game like this, you have to have a little change in what you say, how will it affect the game.* [P2]

> *No, it’s just that I want to see that the people you’re talking to have something else or something more to*

> *say. That they have other reactions than just "no [laughs]."* [P2]

> *It was, it's not a challenging game, it's not a difficult game... the game itself [laughs]. It's designed to be clicked through. You don't need to spend a week on it.* [P3]

Participants find the graphics visually appealing, although they do not feel that this is the most important aspect, saying that how the game works and how exciting and entertaining the game is are the most important parts:

> *I think it looks pretty nice. There are many different types of games that can be in like... many different types of 'art' [styles] and so, yes, there are many games that can look one way, but can still be really good, but some games that look really nice, can be boring. So, don't worry about how the game looks, but how the game works, that's very important.* [P2]

> *Visually it was very nice... The story I am more unsure of....* [P12]

Some of the feedback indicates that our game was perhaps not fully developed as a game with the necessary combination of rules, goals, feedback, fantasy, and fun [32]. Participants wish for a wider range of interaction modes, such as moving more freely around in the game, having more challenges and tasks in the game, and having different ways of interacting with the characters. Users expect game-like interactions, not just choosing answers, reading, and writing:

> *It is unusual for me that you cannot move around [in the game].* [P4]

> *But I think that it would be a bit boring if it was just like that you had to read, and then click to choose answer options. [...] if I was sitting at home, and this was something I had to go through every week[...] then I'd just click quickly through it. And then I hadn't properly read what it said.* [P7]

To support the learning objectives of the game, players had to answer reflective questions at the end of each gaming session, where they answered as themselves and not as the gaming character. This felt a bit "off" to the participants. Furthermore, several participants expressed a problem with articulating answers to those kinds of questions, expressing that they would not know what to write when asked:

> *I didn't expect that an assignment came up where you have to write about an experience from reality, sort of, which seems a little unusual to me since you are sitting inside this alternate world. But I had probably only written something about skateboarding. But it was very unusual for it to be like that.* [P4]

> *I don't know... at least I struggle a lot with tasks like that[...] I probably wouldn't have written anything here.* [P5]

Although not as exciting as hoped for, the challenges and tasks in the game can still provide the user with a sense of achievement and act as an awakening for new thoughts; if not for them, then perhaps for someone else:

> *You kind of get a little more confidence in yourself then. That you have actually managed something.* [P6]

> *...after all, it raises thoughts and yes... new ways of looking at things, I think. that it can start something in someone.* [P7]

Theme 3: Sensemaking of Experience and Context

This theme comprises a desire for an explicit presentation of the game objectives and a need to manage user expectations related to the gaming genre. The game was described as "cool but confusing:"

> *Uh, well, it seems kind of cool, but it was a little hard to understand, I felt.* [P5]

Many participants mentioned *sensemaking* or lack of understanding in some form or another; they struggled to understand the point, the objective, or the mission to be completed in the game. Some participants pointed to a lack of logic or strangeness in the storyline and over-the-top reactions to what they perceived as minor happenings:

> *So people want to buy parts of dead people so they can look how they want? Hum. That's a very strange concept! [laughs].* [P2]

> *She is stabbed now! ... And the taverna is burning. That was over-the-top. She just came for some food. This is over-the-top.* [P12]

The intervention messages in the game were not perceived by any of the participants. They were uncertain about what they were learning from the game experience. They were focusing on the details of the game narrative, trying to make sense of that, and, thus, the intervention part seemed "part of the fiction" and not clear what this was meant for:

> *Didn't learn anything. Well... I learned that there can be different ways to solve things, but I didn't really learn anything[...] It was a bit difficult to understand the whole story, that is the whole thing.* [P1]

> *Lots of talk about the brain, that the brain is a muscle, but don't know what it can help with, it doesn't make sense.* [P2]

The gameplay added complexity and was confusing to the participants. For most participants, there was a desire for an explicit presentation of game objectives, both in terms of what it should ultimately achieve for the end user (as an intervention) and what the objectives in the game narrative are. However, other participants felt that this uncertainty was part of the excitement:

> *...I should have known a little more what the goal was and what the meaning behind the game was. Because it seemed a bit like that, yes... a bit out of the blue. And you didn't quite know what an anonymous person was and whether this was the future or whether this was a completely different world.* [P7]

> *It was very interesting. It was very unusual for me with that kind of game. But I liked how it was. And you didn't have very much information about what*

> *you were doing so you kind of had to find out a bit about the skins and such. And I liked that.* [P4]

One user expressed explicit concern about how relatable the contents could be if you use a context that is far removed from everyday life:

> *If it becomes too sci-fi, I think it might be difficult to transfer to reality.* [P18]

Furthermore, we also identified a need to manage gaming genre expectations. The participants expressed a preference for certain genres over others; it may be difficult to cater to different preferences in terms of what games they like best to play. There are also certain expectations connected with different gaming genres that we were not so aware of in the research team, where users were trying to make sense of the game prototypes *in relation to* established genres, with expectations of gaming interaction to be similar to games in that genre. The participants asked us about the game in relation to genre concepts such as “open world,” “adventure,” and “role-playing” games:

> *So, I have a question, is this an open-world type of game where you go out to different places and pick up things or is this a text where you just follow what happens in the story?* [P2]
>
> *I have a question: Is this open world—or just to follow a track, like?* [P17]

It seems that the game genre was not clear to the participants, who pointed to different features of the gameplay prototypes that would take the game in different genre directions.

Discussion

Principal Findings

This study has used an iterative design process with active participation from potential users to develop an interactive game that aims to be user-friendly and engaging to be able to provide a vulnerable population with positive psychologies. As pointed out by past research, there is a strong rationale for promoting psychological well-being, for instance, to improve resilience [83] in the face of setbacks and challenges that occur as part of job search and being “out-of-work” [12,13,78], and may thus alleviate suffering [84]. A total of 3 themes were constructed from the user-based research that occurred through the design process of the gaming-based intervention: (1) adapting to user preferences, (2) empathic player interaction, and (3) sensemaking of experience and context. In the following section, we discuss the themes and how they could potentially be applied as designing principles for future self-administered gamified PPIs. Thus, the study sheds light on the application of game design for PPIs that aim to promote well-being and increase challenge-seeking in young, unemployed adults.

Comparisons With Previous Work

Adapting to User Preferences

This study expands the knowledge found in other studies on PPIs for young people, where the need to offer interactive, visual, and more engaging experiences has been identified [26,85,86]. Most participants expressed uncertainty about whether the game would be “something for them,” suggesting other preferences or not quite the right conceptual fit with their preferences. We interpret this to mean that there is a need to consider gaming genres and user preferences specifically, where a more refined user segmentation may be necessary [87,88]. For instance, in our group, some participants said that they only had an interest in certain kinds of games. Future studies could consider a more fine-grained targeting strategy based on preferences and interests and not simply age and employment status. A possibility would be to segment the population based on player types [89,90] or motivation [89,91-93], and to think more carefully about user preferences for different gaming genres before choosing a concept. In our study, we found that the selected game genre was perceived as engaging “not to them,” but “someone younger.” This was particularly true for the active gamers, who found the genre to be immature. Furthermore, most participants stated that they would have preferred a game designed for mobile use, indicating another kind of context of use than our initially planned use on PCs at home.

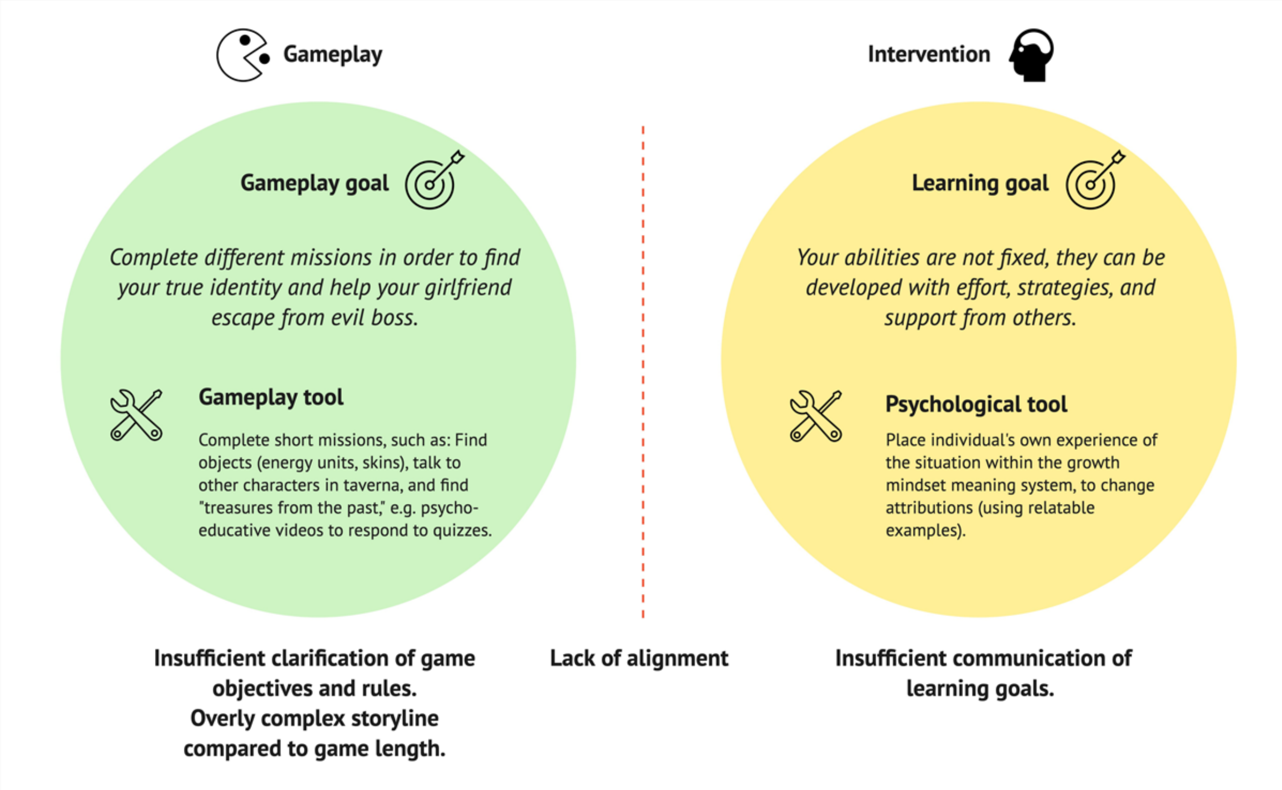
Empathic Player Interaction

Inside the game experience itself, the participants in our study were disappointed by the lack of features, functionalities, and opportunities to influence what was going to happen in the game, for example, player autonomy. A lack of autonomy may cause a more negative interpretation of an experience [94]. We interpret this as underdelivery, partly because of the overpromise of the first impression and the esthetics of the graphics [95,96]. The participants expressed bleakly that in-game actions “do not matter” because, as players, they experienced an insufficient influence on the string of events in the game. Autonomy is an important motivator in self-determination theory (SDT) [97] and a lack thereof may contribute to reduced user motivation [13,35,97]. In part, the lack of autonomy and interactivity was caused by the requirements for a structured intervention set by the broader research project; each player needed to experience the same sequence of events. However, even within this frame, the game should be built to cater for somewhat more variation and focus on the interaction between the game and the player to meet user expectations. There is also a more specific need to be aware that within this target group of young, unemployed adults, many may feel *in general* that what they do does not matter [98]. Gameplay with insufficient df may unintentionally reinforce that message.

Sensemaking of Experience and Context

Participants had trouble making sense of the experience: our prototypes did not meet the participants’ expectations for genre, not fitting with role-playing games based on narratives and dialogues or open-world type games, where you move around freely in a 3D world to pick up items and battle with other characters. Furthermore, the game objective and rules were either not clearly presented to the participants or did not cater to sufficient player-game dialogue and manipulation of the experience. We found that there were tensions between the gameplay and the messages of the intervention, which could undermine the intervention and potentially threaten its effectiveness. This finding is depicted in Figure 7. The PPI gameplay had a complex storyline, which confused the participants and made them miss out on their learning objectives. Past studies have pointed to psychological affordances and the importance of a "fertile soil" to make positive psychological interventions more likely to work [99-101]. In instructional or serious games used in education, Young et al [38] concluded the need to ensure that game objectives and learning objectives correspond and, further, that an overly complex gameplay can lead to misunderstandings and interfere with understanding. This seems transferable to gaming-based interventions. Other authors have referred to this as "relevant narrative," which states that the narrative of the game should be relevant to the subject matter [102]. The choice of gameplay as a strategy for creating engagement for an intervention introduces a new context, which becomes the background for interpreting the messages of the intervention. The game design concept should be selected carefully and tested early with inexpensive methods, such as roleplay or paper sketches [32,56], to explore whether the gameplay is supportive of the intervention. In VitaNova, the gameplay goals implicitly reflected the learning goals, as the development of abilities was presented through the completion of in-game missions. However, because these learning goals were not explicitly communicated, the effectiveness of the intervention depended on the users themselves seeing the connection and transferring this knowledge to their own situation. Combined with the lack of clarity of game objectives and rules as well as an overly complex storyline, this led to confusion.

Figure 7. Mismatch of goals between intervention and gameplay, in combination with insufficient clarity overall.



Practical Implications for Future Gaming-Based Interventions

When revisiting the 3 themes and comparing them to related work, there appears to be a similarity between the identified findings of this study and the 3 basic psychological needs in SDT [103], which are: needs for autonomy, relatedness, and competence [92]; refer to Figure 8. As such, this study provides a form of bottom-up support for the usefulness of these constructs in designing and evaluating future gaming-based PPIs to understand how they might be more motivating to the user [92]. Further research should also investigate how and if a gaming-based PPI experience that does satisfy the relevant needs of autonomy, relatedness, and competence may contribute *in itself* to positive psychological outcomes for this population, as has been suggested [13].

Striking the right balance between learning and fun is a significant challenge, along with producing a relevant narrative [102] that supports intended learning. Ferrara [32] suggests a strategy for identifying the "gameness" that already exists in a context or situation rather than trying to *tack it on*. This may make it easier to transfer learning from the gaming space to everyday life [56,104]. However, moving from an idea to a game that works conceptually is challenging [105], and good intentions may be undermined by a seemingly fun yet unfit idea or concept, for example in the case of Disney and their first version of the game and exhibit "Habit Heroes," intended to support healthy eating but rather reinforced stereotypes and made children feel bad about themselves [106]. Choosing an approach that "gamifies life" should thus be done with empathy, care, and frequent testing with users to avoid banalizing the situation and experiences of a vulnerable population, such as the young and unemployed. As such, a human-centered design approach is ideal because it starts with empathy [48]. However, frequent playtesting [56] and usability evaluation [57] are also needed to reduce the risk of developing a concept that is not engaging with the intended audience [88] or that undermines or does not foster learning. Established game genres and concepts could be used as inspiration in early explorative ideation. The characteristics of existing games may be viewed as opposing values on a spectrum [32], and by imagining what the game-based PPI would look like in the form of existing game genres, a large volume of different ideas can be formed that may be tested early for fit with the PPI objectives and user preferences.

In Figure 9, we propose a broad but structured approach for how game-based PPI exploration may be executed, based on the lessons learned from our study and the discussion points in the preceding section. In this approach, insight into the user, context, gaming preferences, and gaming interests frames the design problem. It is also necessary to establish a clear and precise definition of PPI, including its underpinning mechanics, theories, and strategies that can help make the intervention effective. An alignment between the 2, forms the necessary "fertile soil" for the intervention game, where we ask how gameplay may support both user preferences and goals and PPI goals. Next, we propose working with existing genres and games to quickly generate many different ideas of what our gaming-based PPI may look and feel like. Promising concepts should be evaluated against relevant criteria, such as gameplay and user experience objectives and PPI objectives, and then made into prototypes for validation with user research.

The approach outlined here should be further detailed, refined, and validated in future research.

Figure 8. How the 3 themes correspond to basic psychological needs for relatedness, autonomy, and competence, in line with self-determination theory (SDT).

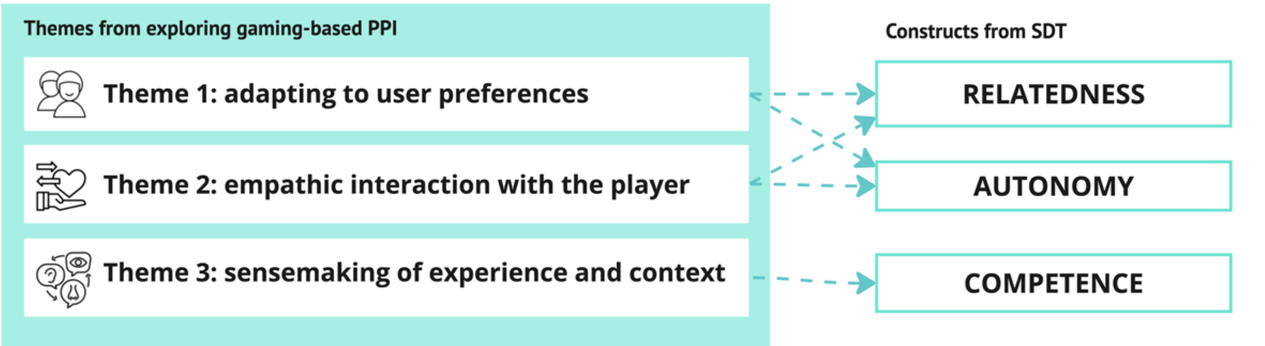
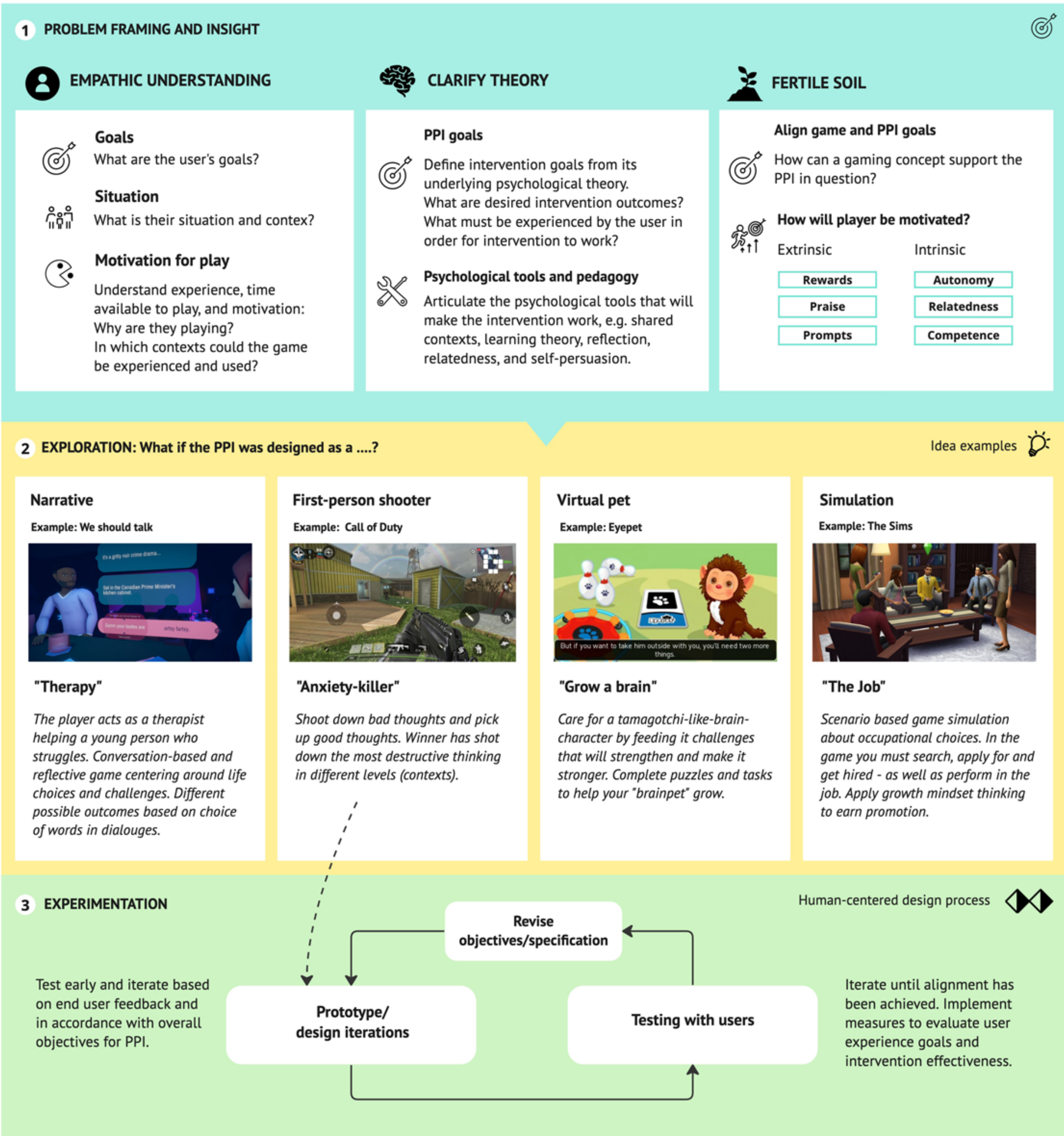


Figure 9. A proposed approach for designing future gaming-based positive psychology interventions (PPIs).



Pointers for Future Research

There is a range of possible strategies to choose from to work to improve an intervention and increase user motivation. In this specific study and for the purposes of this paper, we explored one possible strategy: to attempt to make the PPI more engaging, user-friendly, and relevant for young, unemployed adults by creating a gaming-based intervention. There are other alternative strategies that could increase relevance, appeal, and adaptation to the needs of the target audience. Kelders et al [24] suggest the use of design and persuasive design techniques, including reward, praise and reminders [107] as a tool to increase motivation and retention. Others [13] suggest the use of SDT [97], as we also found some support for this study. These strategies should be explored further in future studies.

Furthermore, the alignment between gameplay and intervention does not rely solely on the crafting of the game. Although our study grounded ideas on learning from past empirical research where PPIs had been applied to different contexts, there was a lack of clarity and theoretical grounding for the user experience in itself, including a clear definition of the learning [35] that should happen within the game design space. Incorporating learning theory, such as experiential learning [41], along with motivation theory, such as SDT [97] and persuasive design [108], as a more complete theoretical framework for the game designing process may provide a stronger direction to the conceptual work for the practitioners involved. Designing for behavioral and mindset change is increasingly relevant for design research and professional design practice [109], and there seem to be several gaps in understanding for design researchers and design teams who find themselves grappling with psychological and behavioral theories to produce interventions to support problem-solving of societal problems, such as youth unemployment.

Limitations

In this study, relevant participants were involved in a design process to capture their experience with designs in-the-making and take feedback into consideration in the design of revisions. We consider such early involvement a strength of the study. However, it also holds limitations; as the results are drawn from user experiences with *prototype* PPIs, the study does not provide user experiences resulting from a completed and verified PPI. Although the knowledge gained through the different stages of the design process is of substantial value to this area of research, future work is needed on experiences with fully functional gameplay PPIs to validate the findings of this study and to measure engagement, effectiveness, and adherence to the intervention. In addition, considering the fact that the positive psychologies implemented in the game mechanics were, to some extent, unclear to the participants after exposure, we cannot draw any conclusions on the experience of these in themselves at this stage. However, this was also not the purpose of this study.

Another limitation is the choice and availability of participants in the study. The population of unemployed young adults is highly heterogeneous. With our recruitment strategy, we are aware that we do not cover the entire range of end users, especially because we in part relied on voluntary registration and on contact with specific public welfare systems. Nevertheless, we find the involved participants to be within the scope of the studied PPI, and their feedback, hence, is of substantial benefit in understanding how the PPI may be experienced by representatives of this target group.

The third limitation concerns the context of the usability testing. Being observed by another person influences behavior (eg, Hawthorne effect), and participants likely spent much more time considering the prototypes than they would normally have. However, this approach was chosen because our interventions were prototypes and had unfinished functionality, which required a moderator to "fill the gaps" [110,111]. A fully self-administered and unmoderated use of a gameplay PPI would be a natural next step in future research.

Finally, it is important to note that, although our exploratory approach to insight into user perceptions of a game-based intervention for this target group is an important starting point for this area of investigation, future research is needed to establish the knowledge base needed to reliably provide such interventions. As part of this, we envision future studies with larger sample sizes and established scales as part of randomized controlled trials to gain further knowledge of the effectiveness of game-based interventions for this group and a basis for improvements in intervention design.

Conclusions

The study contributes insights into key user perceptions of game-based interventions for unemployed young adults. The contribution has implications for future game-like intervention design for this purpose. Our principal contribution is to explore engagement through a PPI, designed as an interactive game. We have described the iterative process of the development of a 3D-game concept, VitaNova, and have explored participants' thoughts and feedback on their experiences. Although the participants were positive about the general idea of a game targeted toward unemployed young people, we found tensions between a PPI and an exciting game play and 3 themes that pinpoint priorities for future gaming implementations. Our study shows that interactive game design could make interventions more entertaining and engaging but can easily come into conflict with or undermine the intervention. We recommend aligning the gameplay narrative, objectives, and mechanics with intervention content and objectives to create engaging, relevant, and effective gaming-based PPIs that promote a more productive view of the challenges experienced by the young and unemployed.

Acknowledgments

The authors would like to thank the researchers at the University of Stavanger who contributed to this study: Venke F Haaland, Espen Sagen, Mari Rege, Hilde Ness Sandvold, and Wenche Ten Velden Hegelstad. The authors would also like to thank the Norwegian Welfare Administration in Sandnes and Skole- og Jobbresept at the University Hospital of Stavanger, as well as all the participants who took part in the study and gave their feedback on the different prototypes. The project is funded by the Norwegian Research Council (grant number 296390).

Conflicts of Interest

None declared.

Multimedia Appendix 1

A presentation of the different game prototypes.

[PDF File (Adobe PDF File), 2796 KB - humanfactors_v11i1e44423_app1.pdf]

References

1. Youth and the labour market. Organisation for Economic Cooperation and Development. URL: https://www.oecd-ilibrary.org/education/youth-not-in-employment-education-or-training-neet/indicator/english_72d1033a-en [accessed 2024-01-05]
2. Borghans L, Duckworth AL, Heckman JJ, ter Weel B. The economics and psychology of personality traits. J Hum Res 2008;43(4):972-1059. [doi: 10.1353/jhr.2008.0017]
3. Heckman JJ, Stixrud J, Urzua S. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. J Labor Econ 2006 Jul;24(3):411-482. [doi: 10.1086/504455]
4. Roberts BW, Kuncel NR, Shiner R, Caspi A, Goldberg LR. The power of personality: the comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. Perspect Psychol Sci 2007 Dec;2(4):313-345 [FREE Full text] [doi: 10.1111/j.1745-6916.2007.00047.x] [Medline: 26151971]
5. Bettinger E, Ludvigsen S, Rege M, Solli IF, Yeager D. Increasing perseverance in math: evidence from a field experiment in Norway. J Econ Behav Org 2018 Feb;146:1-15. [doi: 10.1016/j.jebo.2017.11.032]
6. Deming DJ. The growing importance of social skills in the labor market*. Q J Econ 2017;132(4):1593-1640. [doi: 10.1093/qje/qjx022]
7. McGee RE, Thompson NJ. Unemployment and depression among emerging adults in 12 states, behavioral risk factor surveillance system, 2010. Prev Chronic Dis 2015 Mar 19;12:E38 [FREE Full text] [doi: 10.5888/pcd12.140451] [Medline: 25789499]
8. Fyhn T, Radlick RL, Sveinsdottir V. Unge som står utenfor arbeid, opplæring og utdanning (NEET): en analyse av unge i NEET-kategorien. Norce. 2021. URL: https://www.ks.no/contentassets/25072fbf9d4d4577899c335e431117d7/Unge-utenfor-arbeid-opplering-og-utdanning.pdf [accessed 2023-08-02]
9. Investing in youth. Organisation for Economic Cooperation and Development. URL: https://www.oecd-ilibrary.org/social-issues-migration-health/investing-in-youth-norway_9789264283671-en [accessed 2024-01-05]
10. Bunting M, Moshuus G. Young peoples' own stories about dropping out in Norway: an indirect qualitative approach. Acta Didactica Norge 2017 Aug 25;11(2):3. [doi: 10.5617/adno.3182]
11. Moshuus GH, Ask TS, Bordevich K, Bunting M, Gulløy E, Halvorsen TA, et al. Inaktive ungdommer - en av vår tids største utfordringer. Sluttrapport (2015 – 2018). FARVE-Fondet, NAV, Kunnskapsavdelingen. 2019. URL: https://openarchive.usn.no/usn-xmlui/handle/11250/2687785 [accessed 2023-08-02]
12. Isherwood K. Examining the effect of positive psychology interventions to improve psychological wellbeing, workplace motivation and re-employment prospects in North Wales. Bangor University. 2019. URL: http://tinyurl.com/3zvm6vsx [accessed 2023-08-02]
13. van der Vaart L, Van den Broeck A. Self-determination and positive psychology interventions: an extension of the positive activity model in the context of unemployment. In: Theoretical Approaches to Multi-Cultural Positive Psychological Interventions. Cham, Switzerland: Springer; 2019.
14. Seligman ME, Csikszentmihalyi M. Positive psychology: an introduction. Am Psychol 2000;55(1):5-14. [doi: 10.1037/0003-066X.55.1.5]
15. Rege M, Hanselman P, Solli IF, Dweck CS, Ludvigsen S, Bettinger E, et al. How can we inspire nations of learners? An investigation of growth mindset and challenge-seeking in two countries. Am Psychol 2021 Jul;76(5):755-767 [FREE Full text] [doi: 10.1037/amp0000647] [Medline: 33180534]
16. Yeager DS, Hanselman P, Walton GM, Murray JS, Crosnoe R, Muller C, et al. A national experiment reveals where a growth mindset improves achievement. Nature 2019 Sep 7;573(7774):364-369 [FREE Full text] [doi: 10.1038/s41586-019-1466-y] [Medline: 31391586]
17. Dweck CS, Yeager DS. Mindsets: a view from two eras. Perspect Psychol Sci 2019 May 01;14(3):481-496 [FREE Full text] [doi: 10.1177/1745691618804166] [Medline: 30707853]
18. Paunesku D, Walton GM, Romero C, Smith EN, Yeager DS, Dweck CS. Mind-set interventions are a scalable treatment for academic underachievement. Psychol Sci 2015 Jun;26(6):784-793. [doi: 10.1177/0956797615571017] [Medline: 25862544]
19. Walton GM, Crum AJ. Handbook of Wise Interventions: How Social Psychology Can Help People Change. New York, NY: Guilford Publications; 2020.
20. Yeager DS, Romero C, Paunesku D, Hulleman CS, Schneider B, Hinojosa C, et al. Using design thinking to improve psychological interventions: the case of the growth mindset during the transition to high school. J Educ Psychol 2016 Apr;108(3):374-391 [FREE Full text] [doi: 10.1037/edu0000098] [Medline: 27524832]
21. Dambrun M, Dubuy AL. A positive psychology intervention among long‐term unemployed people and its effects on psychological distress and well‐being. J Employ Couns 2014 Jun 02;51(2):75-88. [doi: 10.1002/j.2161-1920.2014.00043.x]
22. Patel V, Saxena S, Lund C, Thornicroft G, Baingana F, Bolton P, et al. The Lancet Commission on global mental health and sustainable development. Lancet 2018 Oct 27;392(10157):1553-1598. [doi: 10.1016/S0140-6736(18)31612-X] [Medline: 30314863]
23. Christensen H, Reynolds J, Griffiths KM. The use of e-health applications for anxiety and depression in young people: challenges and solutions. Early Interv Psychiatry 2011 Feb;5 Suppl 1:58-62. [doi: 10.1111/j.1751-7893.2010.00242.x] [Medline: 21208393]

24. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
25. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. J Med Internet Res 2019 Jul 22;21(7):e12869 [FREE Full text] [doi: 10.2196/12869] [Medline: 31333198]
26. Michel T, Tachtler F, Slovak P, Fitzpatrick G. A review of youth mental health promotion apps towards their fit with youth media preferences. Pervasive Health Technol 2019 Feb 12;19(17):e2. [doi: 10.4108/eai.13-7-2018.161419]
27. Smyth J. Critically Engaged Learning: Connecting to Young Lives. Lausanne, Switzerland: Peter Lang Publishing Group; 2008.
28. Vygotsky LS, Vygotskiĭ LS, Cole M. Mind in Society: The Development of Higher Psychological Processes. Cambridge, MA: Harvard University Press; 1978.
29. Gee JP. What video games have to teach us about learning and literacy. Comput Entertain 2003 Oct 01;1(1):20. [doi: 10.1145/950566.950595]
30. 2022 Essential facts about the video game industry. Entertainment Software Association. URL: https://www.theesa.com/resource/2022-essential-facts-about-the-video-game-industry/ [accessed 2023-08-07]
31. Adachi PJ, Willoughby T. The link between playing video games and positive youth outcomes. Child Dev Perspect 2017 Mar 17;11(3):202-206. [doi: 10.1111/cdep.12232]
32. Ferrara J. Playful Design: Creating Game Experiences in Everyday Interfaces. Brooklyn, NY: Rosenfeld Media; 2012.
33. Anderson SP. Seductive Interaction Design: Creating Playful, Fun, and Effective User Experiences. Indianapolis, IN: New Riders; 2011.
34. Tan M, Hew KF. Incorporating meaningful gamification in a blended learning research methods class: examining student learning, engagement, and affective outcomes. Australas J Educ Technol 2016 Dec 01;32(5). [doi: 10.14742/ajet.2232]
35. Grund CK. How games and game elements facilitate learning and motivation: a literature review. In: Proceedings of the Informatik 2015. 2015 Presented at: Informatik 2015; September 28-October 2, 2015; Cottbus, Germany URL: https://dl.gi.de/handle/20.500.12116/2112
36. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013 Dec 12;1(1):e3 [FREE Full text] [doi: 10.2196/games.3139] [Medline: 25658754]
37. Abt CC. Serious Games. Lanham, Maryland: University Press of America; 1987.
38. Young MF, Slota ST, Cutter A, Jalette GC, Mullin G, Lai B, et al. Our princess is in another castle: a review of trends in serious gaming for education. Rev Educ Res 2012 Mar;82(1):61-89. [doi: 10.3102/0034654312436980]
39. Djaouti D, Alvarez J, Jessel JP, Rampnoux O. Origins of serious games. In: Ma M, Oikonomou A, Jain L, editors. Serious Games and Edutainment Applications. London, UK: Springer; Oct 13, 2011.
40. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining "gamification". In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. 2011 Presented at: MindTrek '11: Academic MindTrek 2011; September 28-30, 2011; Tampere, Finland. [doi: 10.1145/2181037.2181040]
41. Kolb DA. Experiential Learning: Experience as the Source of Learning and Development, Second Edition. Upper Saddle River, NJ: FT Press; 2014.
42. Sharlanova V. Experiential learning. Trakia J Sci 2004;2(4):36-39 [FREE Full text] [doi: 10.4135/9781412950596.n56]
43. Fleming T, Dixon R, Frampton C, Merry S. A pragmatic randomized controlled trial of computerized CBT (SPARX) for symptoms of depression among adolescents excluded from mainstream education. Behav Cogn Psychother 2012 Oct;40(5):529-541. [doi: 10.1017/S1352465811000695] [Medline: 22137185]
44. Burns JM, Webb M, Durkin LA, Hickie IB. Reach out central: a serious game designed to engage young men to improve mental health and wellbeing. Med J Aust 2010 Jun 07;192(S11):S27-S30. [doi: 10.5694/j.1326-5377.2010.tb03689.x] [Medline: 20528704]
45. Zepke N. Student engagement in neo-liberal times: what is missing? High Educ Res Dev 2017 Sep 04;37(2):433-446. [doi: 10.1080/07294360.2017.1370440]
46. Whitton N. Digital Games and Learning: Research and Theory. Milton Park, UK: Taylor & Francis; 2014.
47. Cheng VW, Davenport T, Johnson D, Vella K, Hickie IB. Gamification in apps and technologies for improving mental health and well-being: systematic review. JMIR Ment Health 2019 Jun 26;6(6):e13717 [FREE Full text] [doi: 10.2196/13717] [Medline: 31244479]
48. van der Bijl-Brouwer M, Dorst K. Advancing the strategic impact of human-centred design. Design Stud 2017 Nov;53:1-23. [doi: 10.1016/j.destud.2017.06.003]
49. The Field Guide to Human-centered Design: Design Kit. San Francisco, CA: IDEO; 2015.
50. Swann C. Action research and the practice of design. Des Issues 2002;18(1):49-61 [FREE Full text] [doi: 10.1162/07479360252756287]
51. Lewin K. A research approach to leadership problems. J Educ Sociol 1944 Mar;17(7):392-398. [doi: 10.2307/2262546]

52. Kock N. Action research. Interaction Design Foundation - IxDF. 2014. URL: http://tinyurl.com/yvxwcv8a [accessed 2023-08-08]
53. Stappers P, Giaccardi E. Research through design. Interaction Design Foundation - IxDF. 2014. URL: https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/research-through-design [accessed 2023-08-08]
54. Frayling C. Research in art and design. Royal College of Art. 1994. URL: https://researchonline.rca.ac.uk/id/eprint/384 [accessed 2023-08-27]
55. Braun V, Clarke V. Thematic Analysis: A Practical Guide. Thousand Oaks, CA: SAGE Publications, Inc; 2021.
56. Tekinbas K, Zimmerman E. Rules of Play: Game Design Fundamentals. Cambridge, MA: MIT Press; 2003.
57. Cockton G. Usability evaluation. Interaction Design Foundation - IxDF. 2014. URL: https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/usability-evaluation [accessed 2023-08-08]
58. Lewis JR. Usability: lessons learned … and yet to be learned. Int J Hum Comput Interact 2014 Jun 18;30(9):663-684. [doi: 10.1080/10447318.2014.930311]
59. Lewis JR. Sample sizes for usability tests: mostly math, not magic. Interactions 2006 Nov 01;13(6):29-33. [doi: 10.1145/1167948.1167973]
60. Gothelf J, Seiden J. Lean UX: Applying Lean Principles to Improve User Experience. Sebastopol, CA: O'Reilly Media; 2013.
61. Knapp J, Zeratsky J, Kowitz B. Sprint: How to Solve Big Problems and Test New Ideas in Just Five Days. New York, NY: Simon & Schuster; 2016.
62. Star SL. Chapter 2 - the structure of ill-structured solutions: boundary objects and heterogeneous distributed problem solving. In: Gasser L, Huhns MN, editors. Distributed Artificial Intelligence. Burlington, MA: Morgan Kaufmann; 1989:37-54.
63. Huvila I, Anderson TD, Jansen EH, McKenzie P, Worrall A. Boundary objects in information science. J Assoc Inf Sci Technol 2017 Jun 08;68(8):1807-1822. [doi: 10.1002/asi.23817]
64. Akkerman SF, Bakker A. Boundary crossing and boundary objects. Rev Educ Res 2011 Jun;81(2):132-169 [FREE Full text] [doi: 10.3102/0034654311404435]
65. Ericson KA, Simon HA. Protocol analysis. In: Protocol Analysis: Verbal Reports as Data Revised Edition. Cambridge, MA: MIT Press; 1993.
66. Mascherini M, Ledermaier S, Eurofound. Exploring the diversity of NEETs. Eurofound. 2016. URL: https://policycommons.net/artifacts/1840981/exploring-the-diversity-of-neets/2584049/ [accessed 2023-08-08]
67. Chen E, Leos C, Kowitt SD, Moracco KE. Enhancing community-based participatory research through human-centered design strategies. Health Promot Pract 2020 Jan 25;21(1):37-48. [doi: 10.1177/1524839919850557] [Medline: 31131633]
68. Nielsen J. How many test users in a usability study? Nielsen Norman Group. 2012 Jun 3. URL: https://www.nngroup.com/articles/how-many-test-users/ [accessed 2023-08-10]
69. Smith JA, Flowers P, Larkin M. Interpretative Phenomenological Analysis: Theory, Method and Research. Thousand Oaks, CA: SAGE Publications, Inc; 2009.
70. Patton MQ. Qualitative Evaluation and Research Methods, 2nd Edition. Thousand Oaks, CA: SAGE Publications, Inc; 1990.
71. Unge som registrerer seg hos NAV: 4 av 10 kommer raskt i jobb uten hjelp. NAV. 2021 Nov 30. URL: http://tinyurl.com/54dzea2z [accessed 2023-09-01]
72. Raaum O, Rogstad J, Røed K, Westlie L. Young and out: an application of a prospects-based concept of social exclusion. J Soc Econ 2009 Jan;38(1):173-187. [doi: 10.1016/j.socec.2008.08.003]
73. Horgen EH. Hvorfor ulike arbeidsledighetstall? Statistisk sentralbyrå. 2020 Jan 23. URL: https://www.ssb.no/arbeid-og-lonn/artikler-og-publikasjoner/hvorfor-ulike-arbeidsledighetstall--410344 [accessed 2023-08-10]
74. Dweck CS, Yeager DS. A growth mindset about intelligence. In: Walton GM, Crum AJ, editors. Handbook of Wise Interventions: How Social Psychology Can Help People Change. New York, NY: The Guilford Press; 2021.
75. Beck JS. Cognitive Behavior Therapy: Basics and Beyond Second Edition. New York, NY: Guilford Publications; 2011.
76. Yeager DS, Bryan CJ, Gross JJ, Murray JS, Krettek Cobb D, H F Santos P, et al. A synergistic mindsets intervention protects adolescents from stress. Nature 2022 Jul 06;607(7919):512-520 [FREE Full text] [doi: 10.1038/s41586-022-04907-7] [Medline: 35794485]
77. Chen DJ, Lim VK. Strength in adversity: the influence of psychological capital on job search. J Organ Behav 2012 Jun 25;33(6):811-839. [doi: 10.1002/job.1814]
78. Moorhouse A, Caltabiano ML. Resilience and unemployment: exploring risk and protective influences for the outcome variables of depression and assertive job searching. J Employ Couns 2011 Dec 23;44(3):115-125. [doi: 10.1002/j.2161-1920.2007.tb00030.x]
79. Bellotti F, Berta R, De Gloria A, Primavera L. Enhancing the educational value of video games. Comput Entertain 2009 Jun 24;7(2):1-18. [doi: 10.1145/1541895.1541903]
80. Aronson E. The power of self-persuasion. Am Psychol 1999 Nov;54(11):875-884 [FREE Full text] [doi: 10.1037/h0088188]
81. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

82. Braun V, Clarke V. Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. Couns Psychother Res 2020 Oct 18;21(1):37-47. [doi: 10.1002/capr.12360]
83. Herrman H, Stewart DE, Diaz-Granados N, Berger EL, Jackson B, Yuen T. What is resilience? Can J Psychiatry 2011 May 01;56(5):258-265. [doi: 10.1177/070674371105600504] [Medline: 21586191]
84. Duckworth AL, Steen TA, Seligman ME. Positive psychology in clinical practice. Annu Rev Clin Psychol 2005 Apr;1(1):629-651. [doi: 10.1146/annurev.clinpsy.1.102803.144154] [Medline: 17716102]
85. Michel T, Tachtler F, Slovak P, Fitzpatrick G. Young people's attitude toward positive psychology interventions: thematic analysis. JMIR Hum Factors 2020 Nov 09;7(4):e21145 [FREE Full text] [doi: 10.2196/21145] [Medline: 33164908]
86. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug 05;21(3):116-119 [FREE Full text] [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
87. Hainey T. Moving digital games for learning forward. Horiz 2016 Feb 08;24(1):132-136. [doi: 10.1108/oth-08-2015-0037]
88. Fleming T, Merry S, Stasiak K, Hopkins S, Patolo T, Ruru S, et al. The importance of user segmentation for designing digital therapy for adolescent mental health: findings from scoping processes. JMIR Ment Health 2019 May 08;6(5):e12656 [FREE Full text] [doi: 10.2196/12656] [Medline: 31066705]
89. Bartle R. Hearts, clubs, diamonds, spades: players who suit muds. Players Who Suit MUDs. URL: https://mud.co.uk/richard/hcds.htm [accessed 2022-11-10]
90. Kort-Butler LA. The well-being of gamers, video game players, and non-players. Soc Sci J 2021 Apr 15:1-17. [doi: 10.1080/03623319.2021.1910774]
91. Williams D, Yee N, Caplan SE. Who plays, how much, and why? Debunking the stereotypical gamer profile. J Comput Mediat Commun 2008 Jul 1;13(4):993-1018. [doi: 10.1111/j.1083-6101.2008.00428.x]
92. Ryan RM, Rigby CS, Przybylski A. The motivational pull of video games: a self-determination theory approach. Motiv Emot 2006 Nov 29;30(4):344-360. [doi: 10.1007/s11031-006-9051-8]
93. Yee N. Motivations of play in MMORPGs. In: Proceedings of DiGRA 2005 Conference: Changing Views – Worlds in Play. 2005 Presented at: DiGRA 2005 Conference: Changing Views – Worlds in Play; June 16-20, 2005; Vancouver, BC URL: http://www.digra.org/dl/db/06276.26370.pdf
94. Deci EL, Koestner R, Ryan RM. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. Psychol Bull 1999 Nov;125(6):627-68; discussion 692. [doi: 10.1037/0033-2909.125.6.627] [Medline: 10589297]
95. Hekkert PP. Design aesthetics: principles of pleasure in design. Psychol Sci 2006;48(2):157-172 [FREE Full text]
96. Honea H, Horsky S. The power of plain: intensifying product experience with neutral aesthetic context. Mark Lett 2011 Oct 7;23(1):223-235. [doi: 10.1007/s11002-011-9149-y]
97. Deci EL, Ryan RM. Facilitating optimal motivation and psychological well-being across life's domains. Can Psychol 2008 Feb;49(1):14-23. [doi: 10.1037/0708-5591.49.1.14]
98. Baum A, Fleming R, Reddy DM. Unemployment stress: loss of control, reactance and learned helplessness. Soc Sci Med 1986 Jan;22(5):509-516. [doi: 10.1016/0277-9536(86)90016-x] [Medline: 3704688]
99. Bronfenbrenner U. The Ecology of Human Development: Experiments by Nature and Design. Cambridge, MA: Harvard University Press; 1979.
100. Lewin K. Field theory and experiment in social psychology: concepts and methods. Am J Sociol 1939 May;44(6):868-896. [doi: 10.1086/218177]
101. Walton GM, Yeager DS. Seed and soil: psychological affordances in contexts help to explain where wise interventions succeed or fail. Curr Dir Psychol Sci 2020 Jun 14;29(3):219-226 [FREE Full text] [doi: 10.1177/0963721420904453] [Medline: 32719574]
102. Shandley K, Austin D, Klein B, Kyrios M. An evaluation of 'Reach Out Central': an online gaming program for supporting the mental health of young people. Health Educ Res 2010 Aug;25(4):563-574. [doi: 10.1093/her/cyq002] [Medline: 20150530]
103. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000;55(1):68-78. [doi: 10.1037//0003-066X.55.1.68]
104. Stenros J. In defence of a magic circle: the social and mental boundaries of play. In: Proceedings of 2012 International DiGRA Nordic Conference. 2012 Presented at: DiGRA Nordic '12; June 6-8, 2012; Tampere, Finland URL: http://www.digra.org/digital-library/publications/in-defence-of-a-magic-circle-the-social-and-mental-boundaries-of-play/
105. Baker C. Trying to design a truly entertaining game can defeat even a certified genius. Wired. 2008 Mar 24. URL: https://www.wired.com/2008/03/pl-games-25/ [accessed 2023-08-10]
106. Habit heroes. Wikipedia. URL: https://en.wikipedia.org/w/index.php?title=Habit_Heroes&oldid=1120538825 [accessed 2022-11-07]
107. Ludden GD, van Rompay TJ, Kelders SM, van Gemert-Pijnen JE. How to increase reach and adherence of web-based interventions: a design research viewpoint. J Med Internet Res 2015 Jul 10;17(7):e172 [FREE Full text] [doi: 10.2196/jmir.4201] [Medline: 26163456]

108. Fogg BJ. A behavior model for persuasive design. In: Proceedings of the 4th International Conference on Persuasive Technology. 2009 Presented at: Persuasive '09; April 26-29, 2009; Claremont, CA. [doi: 10.1145/1541948.1541999]
109. Cash PJ, Hartlev CG, Durazo CB. Behavioural design: a process for integrating behaviour change and design. Design Stud 2017 Jan;48:96-128. [doi: 10.1016/j.destud.2016.10.001]
110. Schade A. Remote usability tests: moderated and unmoderated. Nielsen Norman Group. 2013 Oct 12. URL: https://www.nngroup.com/articles/remote-usability-tests/ [accessed 2023-08-22]
111. Vasalou A, Ng B, Wiemer-Hastings P, Oshlyansky L. Human-moderated remote user testing: protocols and applications. In: Proceedings of the 8th ERCIM Workshop, User Interfaces for All. 2004 Presented at: 8th ERCIM Workshop, User Interfaces for All; June 28-29, 2004; Vienna, Austria URL: https://reed.cs.depaul.edu/peterh/papers/b2hd-VasalouERCIM2004.html

Abbreviations

CV: curriculum vitae
IPS: Individual Placement and Support
NAV: Norwegian Labour and Welfare Administration
NEET: not in education, employment, or training
PPI: positive psychology intervention
SDT: self-determination theory

Edited by A Kushniruk; submitted 18.11.22; peer-reviewed by S Azer, B Jeong, C Rodriguez-Paras; comments to author 07.06.23; revised version received 02.09.23; accepted 20.11.23; published 18.01.24.

Please cite as:
Straand IJ, Følstad A, Bjørnestad JR
Exploring a Gaming-Based Intervention for Unemployed Young Adults: Thematic Analysis
JMIR Hum Factors 2024;11:e44423
URL: https://humanfactors.jmir.org/2024/1/e44423
doi:10.2196/44423
PMID:38236624

Original Paper

Evaluating the User Experience of a Smartphone-Delivered Sexual Health Promotion Program for Older Adults in the Netherlands: Single-Arm Pilot Study

Ana Correia de Barros[1], PhD; Mariëtte Bergmans[2], BSc; Kreshnik Hasanaj[3], HS; Drianë Krasniqi[3], MD; Catarina Nóbrega[4], MSc; Bruna Carvalho Carneiro[4], MSc; Priscila A Vasconcelos[4], MSc; Ana Luísa Quinta-Gomes[4], PhD; Pedro J Nobre[4], PhD; Joana Couto da Silva[1], BSc; Cristina Mendes-Santos[1], PhD

[1]Fraunhofer Portugal AICOS, Porto, Portugal

[2]PCOB, Utrecht, Netherlands

[3]SPRIGS RND, Pristina, Kosovo

[4]Center for Psychology, Faculty of Psychology and Education Sciences, University of Porto, Porto, Portugal

Corresponding Author:
Ana Correia de Barros, PhD
Fraunhofer Portugal AICOS
Rua Alfredo Allen, 455
Porto, 4200-135
Portugal
Phone: 351 220430306
Email: ana.barros@fraunhofer.pt

Abstract

Background: Sexual health is an important component of quality of life in older adults. However, older adults often face barriers to attaining a fulfilling sexual life because of issues such as stigma, lack of information, or difficult access to adequate support.

Objective: We aimed to evaluate the user experience of a self-guided, smartphone-delivered program to promote sexual health among older adults.

Methods: The mobile app was made available to community-dwelling older adults in the Netherlands, who freely used the app for 8 weeks. User experience and its respective components were assessed using self-developed questionnaires, the System Usability Scale, and semistructured interviews. Quantitative and qualitative data were descriptively and thematically analyzed, respectively.

Results: In total, 15 participants (mean age 71.7, SD 9.5 years) completed the trial. Participants showed a neutral to positive stance regarding the mobile app's usefulness and ease of use. Usability was assessed as "Ok/Fair." The participants felt confident about using the mobile app. To increase user experience, participants offered suggestions to improve content and interaction, including access to specialized sexual health services.

Conclusions: The sexual health promotion program delivered through a smartphone in a self-guided mode was usable. Participants' perception is that improvements to user experience, namely in content and interaction, as well as connection to external services, will likely improve usefulness and acceptance.

(JMIR Hum Factors 2024;11:e56206) doi:10.2196/56206

KEYWORDS

internet interventions; mobile health; mHealth; older adults; sexual health; smartphone; user experience; pilot study; mobile phone

Introduction

Background

Sexual health is a component of general health [1] and quality of life in older age [2]. However, older age is also associated with barriers to a fulfilling sexual life [3-5]. Many older adults are sexually active [6] but are at a higher risk than the general population to present sexual difficulties and dysfunctions. Older women often report decreased libido or lack of vaginal lubrication, whereas erection issues, reduced sexual desire, or being unable to reach orgasm are difficulties regularly reported by men [7]. In health care services, sexual difficulties are often untreated [8] and aggravated by poor communication related to lack of appropriate and case-specific information, lack of training among clinicians, or negative social beliefs and societal stigma, which makes it difficult for both patients and clinicians to bring about the topic [9]. Therefore, identifying the means of circumventing societal stigma and providing timely and adequate support are 2 important courses of action to promote sexual health among older adults.

As the prevalence of smartphone ownership and access to the internet increase [10], there is an opportunity to use these technologies to deliver ubiquitous sexual health support in an inconspicuous manner, that is, one that does not overly expose support seekers to fear of social judgment. Smartphones, as they are intimate technologies that ubiquitously accompany their owners, seem to be an adequate means for the delivery of sexual health promotion programs. Although there is evidence of the efficacy of internet-based sexual health interventions for sexual dysfunction [11] or sexual health education [12], the literature is nonexistent on smartphone-based sexual health interventions targeting older adults [13].

Critical to the acceptance and adoption of such technologies is the user experience they provide [14,15]. Coined by Don Norman [16], the term “user experience” was used by the author to characterize all the sets of experiences a user has with a product throughout a user journey, from intention to use until postuse reflections [17]. Therefore, the concept goes beyond usability, defined by International Organization for Standardization as “the extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” [18]. Designing positive user experiences with mobile digital technologies for older adult users has been a focus of many studies because the levels of engagement have been low, thus hampering the potential health benefits of such technologies [19]. Research has found that older adults’ user experience with mobile digital health could be improved if the technology considered potential user sensorimotor and cognitive issues, users’ motivation, and social support [19] as well as if it promoted more personalized experiences and trust [20]. Although there are general guidelines on designing for accessibility and inclusive design [21,22], best practices for designing digital technologies for sensitive topics such as sexuality and intimacy are lacking [23]. Understanding older adults’ experiences with such technologies in the topic of sexual health is critical to improving their acceptability, usability, and adoption, so that they can deliver positive outcomes. However, no study has yet reported on older adults’ user experiences with smartphone-delivered sexual health promotion programs.

To address these gaps, we have designed a smartphone-based sexual health promotion program [24] under a European project called Anathema (reference AAL-2020-7-133-CP). This program was made available to older adults in a longitudinal study during which we assessed the participants’ user experience with the software. The findings contribute to the body of knowledge on older adults’ preferences, use, and appropriation of digital technologies for sexual health and the design of smartphone-based sexual health promotion programs targeting this population.

Aim

The aim of this study was to evaluate the user experience of Anathema, a self-guided, smartphone-delivered program to promote sexual health among older adults.

Anathema Mobile App Overview

The mobile app used in this study was developed using a participatory design approach [25], which involved users from 3 European countries using the following methods: questionnaires, interviews, focus groups, usability tests, and co-design workshops [23,26].

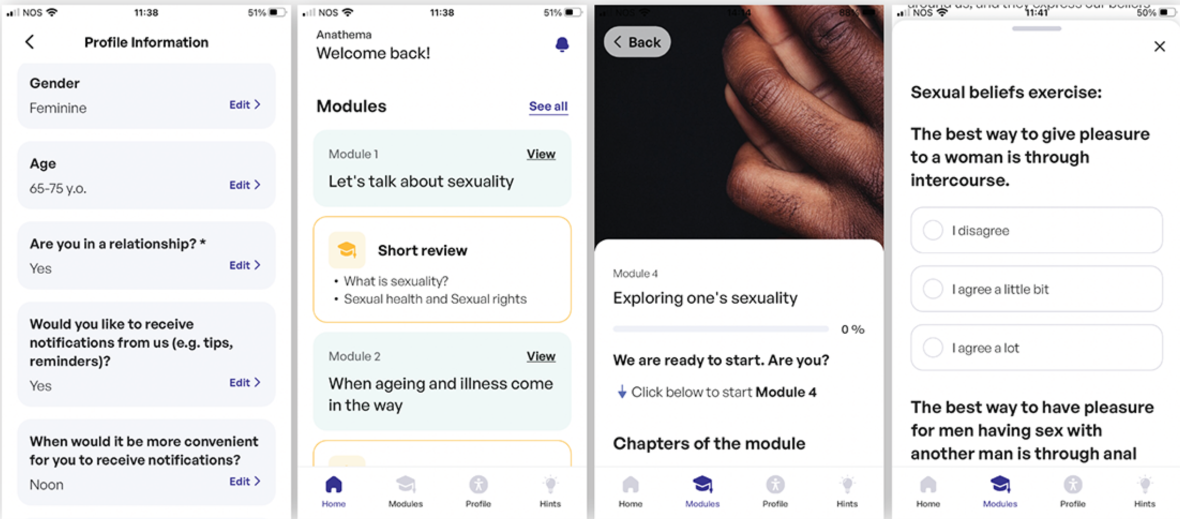
The app is available for Android and iOS operating systems and contains a sexual health promotion program tailored to older adults. The program, which has an 8-week duration, is organized into 5 modules (which include chapters and subchapters):

- *Module 1—Let’s talk about sexuality (week 1)*: features information on male and female anatomies, sexual response, the importance of sexual pleasure, and sexual rights.
- *Module 2—When age and illness come in the way (week 2)*: addresses successful aging; the physiological, cognitive, and emotional changes in older age; and the main sexual problems and sexual dysfunction in older age.
- *Module 3—Emotional and physical intimacy (weeks 3-6)*: covers psychoeducation on the cognitive behavioral therapy model and the impact of sexual beliefs, thoughts, and emotions on sexuality. It includes exercises for cognitive restructuring, mindfulness, and communication skills training.
- *Module 4—Exploring one’s sexuality (week 7)*: delivers information on sex aids and strategies to enhance sexual pleasure and satisfaction and includes sexual skills training and mindfulness exercises.
- *Module 5—Planning for a long-term fulfilling sex life (week 8)*: targets on relapse prevention with a focus on strategies to maintain progress and prevent setbacks. It also shares strategies to promote a healthy lifestyle and sexual health.

Each module is unlocked upon the completion of the previous module to ensure knowledge and skills acquisition. The chapters and subchapters are made of content in the form of text, images, and videos. The program also includes exercises such as written reflections or answers to multiple-choice questions using radio buttons (Figure 1). The app is available in English, European Portuguese, German, and Dutch languages.

The mobile app performs passive data collection through timestamp logs of interactions (eg, module completion date) as well as active data collection through logs of users' inputs on exercises. Another tool, Trial Monitor [27], fetches data from the database and shows visualizations thereof to the research or therapist teams.

Figure 1. Sample screenshots from the Anathema app (left to right): personal information, overview of modules, introduction to module, exercise.



Methods

Study Design

The study design was a single-arm pilot study with older adults (aged ≥55 years) testing the self-guided format of the sexual health promotion program in its Dutch version. The pilot study was conducted to assess user experience of the program. The content, structure, and format were also preliminarily evaluated toward the identification of improvements to the program and technological means of its delivery.

Inclusion and Exclusion Criteria

The inclusion criteria for participation in this trial were as follows: (1) being able to provide informed consent, (2) being aged ≥55 years, and (3) having digital skills and internet access. The exclusion criteria were as follows: (1) having a severe psychiatric disorder or alcohol or substance abuse; (2) taking medication that could interfere with sexual response; (3) having an uncontrolled medical condition that could interfere with sexual health; and (4) currently being on psychotherapy for sexual or intimate problems or for other psychological problems or current participation in another intervention study or clinical trial (or both).

Study Procedures

In a previous phase of this research, 1119 older adults, recruited through the contact list of the Dutch senior organization Katholieke Bond van Ouderen - Protestant Christelijke Ouderen Bond (KBO-PCOB), answered a questionnaire on unmet sexual needs [26]. In this questionnaire, the respondents were asked to indicate whether they would be available for future research within the same research project. Respondents who gave a positive reply were regularly invited to participate in user research activities throughout the research project [23], including the pilot study described in this paper. The majority of this subsample (N=346) were men (69.4%), had a high education level (53.2%), and were retired (89.9%). For the pilot study, further potential participants were contacted via other KBO-PCOB channels, including KBO-PCOB's employees.

Upon signing the informed consent form, participants were asked to complete a web-based screening questionnaire. If deemed eligible to participate in the study, the participants were asked to answer a web-based sociodemographic questionnaire. After completing the questionnaire, participants were provided access to the Anathema app and were prompted to complete the program in 8 weeks.

Once they had completed the 5 modules in the app, participants were asked to fill in a web-based, self-developed user experience questionnaire (Multimedia Appendix 1), which also included the System Usability Scale (SUS). Participants were then invited to participate in a semistructured debriefing interview about (Multimedia Appendix 2) their experiences with the program.

Metrics and Data Analyses

The main outcome of the study was user experience, which included dimensions of usefulness and usability. User experience was assessed after the intervention with a self-developed multiple-choice list of characteristics (answer options: *accessible*, *arousing curiosity*, *attractive*, *boring*, *elegant*, *fascinating*, *helpful*, *instructive*, *meets expectations*, and *strenuous*), a question on free grading of the app from 1 to 10, with 10 being the highest grade, a Net Promoter Score question (answer options: *Yes*/*No*/*Don't know*), and a semistructured debrief interview with questions addressing usefulness, usability, feasibility, clinical aspects, and implementation (Multimedia Appendix 2). Perceived usefulness was assessed using a self-developed 5-point Likert scale assessing the program in general, each module, and exercises. Usability was assessed using a self-developed 5-point Likert scale on perceived ease of use and perceived readability, as well as with the Dutch version of the SUS [28]. Assessment of the self-perceived contribution of the program to changes in satisfaction and pleasure in sex life was also performed postintervention with a single-item question (4-item descriptive rating scale).

To characterize the study sample, sociodemographic variables were collected using a self-developed questionnaire assessing age, education, professional status, gender, sexual orientation, marital status, current sexual partnership status, satisfaction with current sex life (5-point Likert scale), self-rated quality of life, and degree of satisfaction with their own health (based on items 1 and 2 from World Health Organization Quality of Life Brief Version [29]).

The interviews were audio-recorded and partially transcribed for relevant content. The transcriptions, written in Dutch, were then translated into English by a native Dutch speaker (MB) for analysis by a non-Dutch speaker (ACB). The questionnaire and the interview data were analyzed descriptively and thematically, respectively.

Ethical Considerations

The study was approved by the Ethics Committee of the Faculty of Psychology and Educational Sciences, University of Porto (reference 2022/01-05b). All potential participants were informed about the study objectives and procedures. The participants who agreed to participate signed the informed consent form. There was no compensation or payment offered to the participants.

Results

Participants

A total of 400 participants were approached to participate in this study. Most participants did not provide a reason for declining or not answering the invitation. Among those who did (n=47), the reasons given were that participants were no longer interested (n=15), considered the pilot required too much commitment or effort (n=12), felt uncomfortable with the topic (n=9), considered they did not meet the criteria (n=5), or had a malfunctioning email (n=5). We also received information that one person had died.

In total, 23 participants agreed to participate and completed a web-based screening questionnaire to confirm the eligibility criteria. All participants were deemed eligible and were given access to Anathema after answering a sociodemographic questionnaire.

A total of 8 participants dropped out of the study. Of them, 4 participants did not provide any reasons for abandoning the study. Those who did shared the following reasons: discontinued access to the internet (n=1), dissatisfaction with the fact that future content modules were locked (n=1), inability to install and open the app (n=1), and lost motivation to use the app (n=1). A total of 15 participants used the Anathema app, having completed all the modules and completed the final questionnaire on user experience and usability. In total, 8 participants agreed to participate in a debriefing interview.

The 15 participants who used the app and answered the final questionnaire were 7 cisgender women and 8 cisgender men aged between 56 and 85 years (mean 68.3, SD 9.5 years). Most (n=12) were retired, and most (n=10) had completed higher professional education. Overall, 6 participants were married, 4 were single, 3 were cohabiting, and 2 were widowed (Table 1).

Table 1. Sociodemographic characteristics of the sample (N=15).

Characteristics	Values
Gender, n (%)	
Female	7 (47)
Male	8 (53)
Marital status, n (%)	
Single	4 (27)
Cohabiting	3 (20)
Married	6 (40)
Widowed	2 (13)
Professional status, n (%)	
Employed	3 (20)
Retired	12 (80)
Education, n (%)	
Secondary professional education	2 (13)
Higher professional education	10 (67)
University or scientific training	3 (20)
Age (years), mean (SD; range)	71.7 (9.5; 56-85)

Most of the 15 participants were exclusively heterosexual (n=12), most had sex with a partner in the context of an exclusive relationship with that person (n=11), and the level of sexual satisfaction was heterogeneously distributed, as shown, together with complete sexual characteristics (Table 2). The sample comprised participants who tended to positively rate their quality of life and health (Table 3).

Table 2. Sexual characteristics of the sample (N=15).

Characteristics	Baseline, n (%)
Sexual orientation or preference	
Exclusively heterosexual	12 (80)
Mainly heterosexual	2 (13)
Exclusively homosexual	1 (7)
Current sexual partners	
Sex with a partner, in the context of my exclusive relationship with him or her	11 (73)
Casual sex with a partner	1 (7)
No sexual partner	3 (20)
Satisfaction with current sex life	
Very satisfied	3 (20)
Satisfied	5 (33)
Neither satisfied nor dissatisfied	4 (27)
Dissatisfied	3 (20)

Table 3. Perceived quality of life and health satisfaction (N=15).

	Baseline, n (%)	After the test, n (%)
Rating of quality of life[a]		
Very good	8 (53)	9 (60)
Fairly good	7 (47)	5 (33)
Neither good nor bad	—[b]	1 (7)
Satisfaction with health[c]		
Very satisfied	8 (53)	7 (47)
Satisfied	7 (47)	7 (47)
Neither satisfied nor dissatisfied	—	1 (7)

[a]Original wording: How would you rate your quality of life? Responses were rated on a 5-point Likert scale: 1=very bad to 5=very good.

[b]Not available.

[c]Original wording: How satisfied are you with your health? Responses were rated on a 5-point Likert scale: 1=very dissatisfied to 5=very satisfied.

User Experience

In this section, we present the quantitative and qualitative results of the participants' user experience (Table 4). As we do so, we provide interpretations of the results mostly because of the interpretation required by the analysis of the interview data. Therefore, we discuss some of the results as we present them.

Table 4. Results of the user experience questionnaire (N=15).

	Values
Would recommend Anathema to friends or family (net promoter score)[a], n (%)	
Yes	6 (40)
No	6 (40)
Doesn't know	3 (20)
Perceived usefulness of app[b], n (%)	
Very useful	2 (13)
Useful	6 (40)
Neither useful nor useless	6 (40)
Useless	0 (0)
Extremely useless	1 (7)
Perceived usefulness of exercises[c], n (%)	
Very useful	0 (0)
Useful	7 (47)
Neither useful nor useless	3 (20)
Useless	3 (20)
Extremely useless	2 (13)
Perceived ease of use[d], n (%)	
Very easy	1 (7)
Easy	7 (47)
Neither easy nor difficult	5 (33)
Difficult	2 (13)
Readability[e], n (%)	
Very easy	1 (7)
Easy	7 (47)
Neither easy nor difficult	5 (33)
Difficult	2 (13)
System Usability Scale score, mean (SD; range)	56.3 (19.1; 20-85)
Score (1-10) given to Anathema app, mean (SD)	6.5 (1.8; 2-9)
Perceived impact of Anathema app in satisfaction and pleasure[f], n (%)	
Positive impact	4 (27)
No change	7 (47)
Negative impact	1 (7)
Doesn't know	3 (20)

[a]Original wording: Would you recommend the Anathema app to friends and/or family members?

[b]Original wording: How useful do you think the Anathema app is for older adults? Responses were rated on a 5-point Likert scale: 1=extremely useless to 5=very useful.

[c]Original wording: How useful did you find the (writing) exercises you were offered? Responses were rated on a 5-point Likert scale: 1=extremely useless to 5=very useful.

[d]Original wording: How easy was it for you to use the Anathema app without any help from others? Responses were rated on a 5-point Likert scale: 1=extremely difficult to 5=very easy.

[e]Original wording: How readable did you find the content of the Anathema app? Responses were rated on a 5-point Likert scale: 1=extremely difficult to 5=very easy.

[f]Original wording: Do you have the impression that the Anathema app can help you change satisfaction and pleasure in your sex life? Rated using a

descriptive scale: Don’t know; No, no change; Yes, namely less satisfying and fun; Yes, namely more satisfying and fun.

Most participants showed a neutral to positive stance toward the app regarding its *usefulness*. There are some nuances when analyzing the perceived usefulness per module, as illustrated in Figure 2. Modules 2 and 3 had slightly more polarized responses. Modules 1 and 2 were found to be “very useful” for more participants, likely because of the reasons given in the interviews: participants learned new concepts, learned to understand what is normal in aging (“I end up thinking about the part about body ageing. That’s reliable information that I can’t easily get anywhere else today” [P03]), were made to rethink the way in which they faced sexuality, and also learned about the genitalia of other sexes:

> *Nice to read some details about genitals[...] also from the opposite sex, how something works.* [P11]

> *Enlightening. I did benefit from seeing what a prostate looked like.* [P04]

Figure 2. Visualization of perceived usefulness by module.

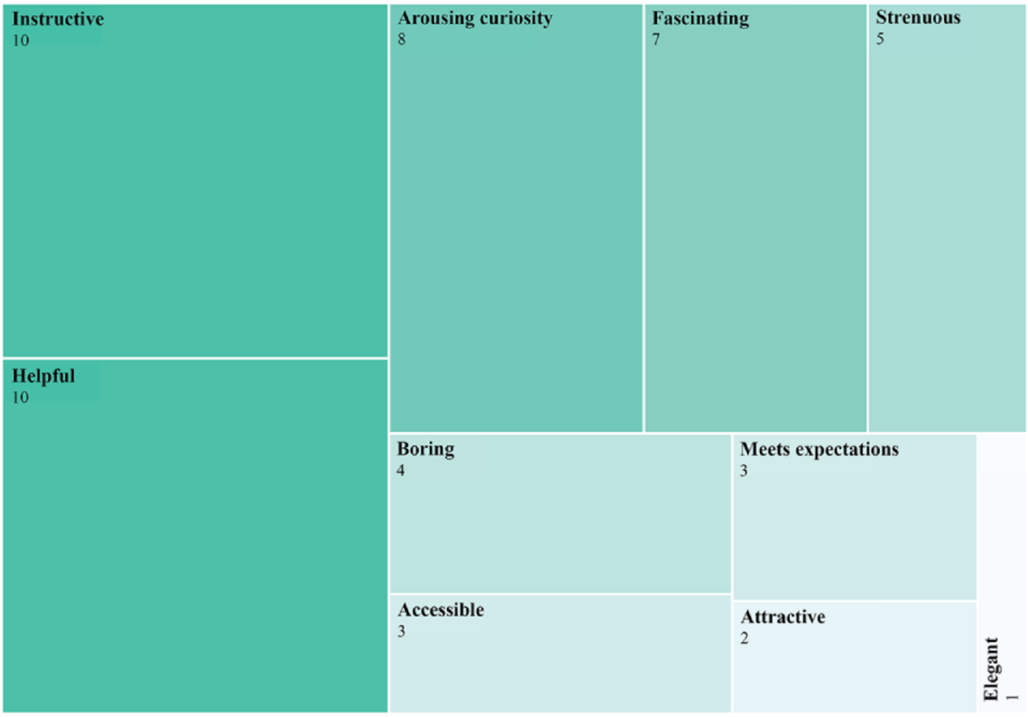
Module 1	Module 2	Module 3	Module 4	Module 5
Extremely useless	Extremely useless	Extremely useless	Extremely useless	Extremely useless
Useless	Useless	Useless	Neither useful nor useless	Neither useful nor useless
Neither useful nor useless	Useless	Useless	Neither useful nor useless	Neither useful nor useless
Neither useful nor useless	Neither useful nor useless	Neither useful nor useless	Neither useful nor useless	Neither useful nor useless
Neither useful nor useless	Neither useful nor useless	Neither useful nor useless	Neither useful nor useless	Neither useful nor useless
Neither useful nor useless	Neither useful nor useless	Neither useful nor useless	Neither useful nor useless	Neither useful nor useless
Neither useful nor useless	Neither useful nor useless	Neither useful nor useless	Useful	Useful
Neither useful nor useless	Useful	Neither useful nor useless	Useful	Useful
Useful	Useful	Useful	Useful	Useful
Useful	Useful	Useful	Useful	Useful
Useful	Useful	Useful	Useful	Useful
Very useful	Very useful	Useful	Useful	Useful
Very useful	Very useful	Very useful	Very useful	Useful
Very useful	Very useful	Very useful	Very useful	Useful
Very useful	Very useful	Very useful	Very useful	Very useful

Other highlighted learning points from the program are the importance of communication and the fact that sexuality does not need to be equated with penetration. Something some participants missed was the possibility to ask the app questions about their specific problems, ask questions anonymously, or to be able to search for certain themes that could be of more interest to them.

Seven participants evaluated the *exercises* as useful, whereas the other 8 found them neutral (n=3), useless (n=3), or extremely useless (n=2). Crossing these results with the information provided in the interviews, one can infer that there were 2 aspects that hindered the experience with the exercises: on the one hand, participants struggled with long text input on their smartphone keyboards; on the other hand, for this group, the feeling of being “schooled” by the app was not equated with positive emotions, thus negatively impacting the experience. Finally, in the interviews, participants revealed that some exercises helped them think of sexuality in a different way, which they experienced as being positive.

When asked to attribute *characteristics* to the *Anathema app*, most participants selected a set of descriptors displayed in Figure 3, but the number of choices varied from a single adjective to 6 adjectives. Most of the qualifiers have positive valence, with the exception of “boring” and “strenuous,” with 4 and 5 mentions, respectively. In line with the data collected through the interviews, the participants perceived that they had learned from the app. However, only 5 participants assessed the app as having the potential to help change their sexual satisfaction and pleasure.

Figure 3. Visualization of qualifiers attributed to the Anathema app and how often they were attributed.



The interviews also revealed that participants appreciated the app aesthetically, which connects to the descriptors that were chosen, as well as the tone of voice that was adopted for the content, which, in some cases, helped them deal with a sensitive topic:

> *I admire that this can be done in an app. Good looking and doesn't scare someone. I managed to deal with such a sensitive topic. [It's] friendly and nicely constructed.* [P22]

For 2 participants, the communication style options were not the most appropriate, for example, when showing an animated video of an anthropomorphized clitoris. Although the photographs were selected based on a survey conducted by the research team about the characteristics of photos that were appreciated by Dutch older adults, 2 interviewees did not find them totally appropriate, for example, some having a comical or childish tone, representing too young people, or not representing enough diversity.

Taking the *net promotor score* as an indicator of satisfaction, we can see that opinions were divided. Three participants did not know whether they would recommend the app to friends or family, whereas the remaining 12 participants were equally divided between wanting to recommend and not wanting to do so. In the debriefing section of the interviews, participants who were not certain whether to recommend Anathema expanded on this. They explained that they think the app has potential but that it needs certain improvements, as described earlier, for them to confidently recommend it to others.

The average SUS score, which measures *usability*, stood at 56.3, which, according to the scoring standards, corresponds to an assessment of "OK/Fair" [30]. Based on the averages per item, we can see that participants tend to think that they do not need help in using the system, although usability is not perceived to be at the excellent level. The level of confidence felt by participants while operating the app was high. Participants generally showed a neutral to positive stance toward the app regarding its *ease of use* and its *readability*. Although most people did not experience trouble reading because of font size or contrast, this was an issue for one of the participants who dropped out:

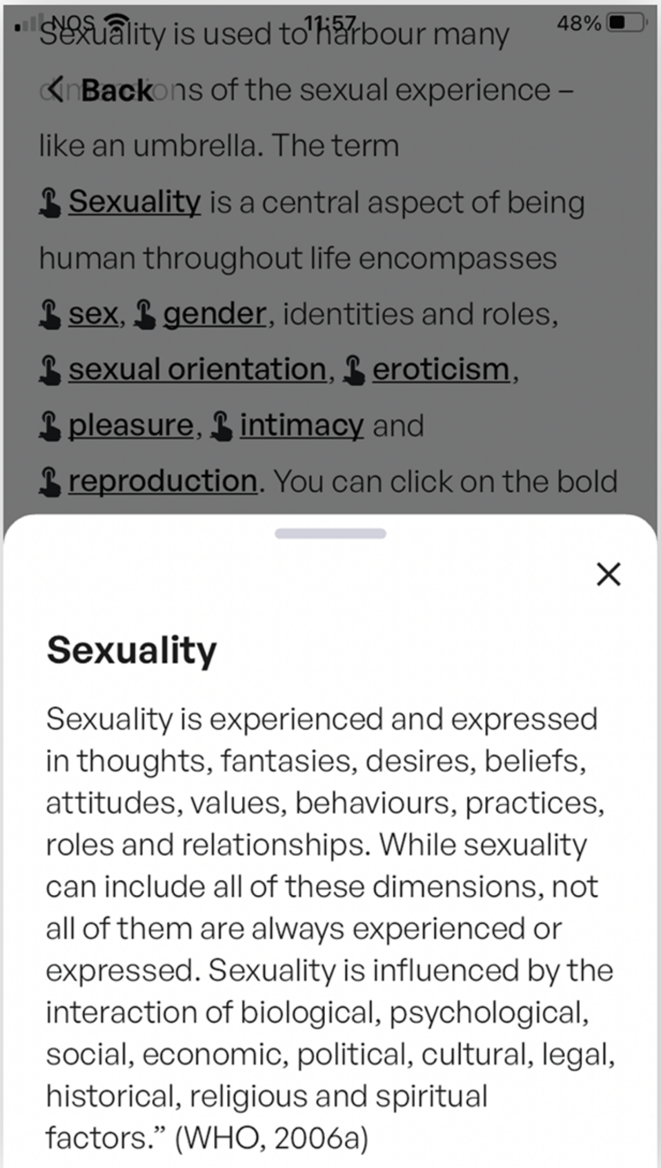
> *I also found the fine print difficult. They are clear but with deteriorating eyes good reading requires more effort.* [P04]

The interviews revealed that the app worked well on participants' phones and that they found it very convenient. However, participants often wished that the app would also be easy to use on a tablet device or desktop:

> *Excellent [the experience of using the app on the phone]. Preferably on an iPad, because of the larger screen. On the phone it worked. The smaller keyboard asked more caution, but [it] went fine.* [P11]

With the exception of 2 participants, who suggested direct speech, easier wording, and shorter sentences, interviewees found the wording easy to understand. In total, 2 participants reflected on whether the scientific explanation should be highlighted as is (Figure 4), for instance, on starting the first module with the definition of sexuality or whether it should be made more digestible to engage readers.

Figure 4. Screenshot from the definition of sexuality, which appears in a slide-up pop-up when the user taps the word “Sexuality” shown in the gray-faded part of the image. WHO: World Health Organization.



Another aspect of readability that was touched on was finding one’s place in the content structure. For 2 participants, it was hard to understand at which stage they were in navigating the app, originating the feeling of being lost: “In a book you can browse through that and then you see where you are. In the app this overview is not so clear” (P06). For those who felt lost, as well as for one participant who would like to revisit specific parts of the content, a possible solution was provided by one participant who said that they missed a way to bookmark “Favorites.”

The participants *appropriated* the app in different ways. There were reports of people using the app only randomly when they found the time, defining a fixed schedule (eg, evening, late at night), or defining a place to use the app (eg, kitchen, home). Common to the participants was the need to use the app alone and undisturbed.

One participant asked the partner to also go through the app, but they did not want it because it was a taboo topic. Three mentioned how they talked to friends or their partners later about the app and what they had learned, for example:

> *Through the app I can easily talk with my partner about sexuality. The participation, together with my partner, in the previous workshops for Anathema, also contributed to this. The openness of other participants was a good example for me.* [P04]

While 5 participants stated that there would be no place or time when they felt uncomfortable using the app, the remaining 3 gave some indications thereof. For these participants, it would be important to use the app alone and in a private place. Participants also reflected on how they would like to discuss what they were learning in the app with others but found it stigmatizing:

> *When I talk to friends about food, for example, all the experiences can be discussed. Apparently, that is not possible when talking about sex.* [P20]

> *When I try to discuss with seniors of an association with a Catholic background that I am participating in this project, the reaction is that does not suit our people.* [P18]

Half of the participants found an *8-week period* to be too long, whereas the other half felt it was an acceptable or good duration. However, participants struggled with the idea of this being time bound in some way because they could not understand why this specific duration was chosen. In 1 case, the participant felt that this indication of duration could send the wrong message: "I have to be ready in 8 weeks" (P03).

Being presented with *content that had a specific reading order* was cumbersome to some participants. This was because, on the one hand, they could feel schooled, and, on the other hand, they did not want to feel that they were losing time in content that they were not interested in. One participant shared their technique for when something like this happened: they just scrolled the content very quickly to get to the bottom and move to the next chapter. Despite negative comments about the locked content (eg, "I wanted to look at a topic in Module 4. But didn't do that out of irritation at the locks in the extended Module 3" [P04]), participants generally agreed that the content is well structured, being easy to follow. Mindfulness is something that some interviewees found unnecessary. On the other hand, some interviewees would expect to read more about love and affection. There were also other suggestions of curated lists of contacts for further support (eg, participants stated they would like to be able to ask questions to therapists over email) and fitness exercises (eg, pelvic floor muscle exercises).

Although 9 of 15 participants in the questionnaire assessed module 5 as useful or very useful, the interviews revealed a slightly different picture. The interviewees had mixed opinions regarding the usefulness of the last module. With the exception of 1 person, those who found it useful as a recap also reflected on the possibilities of *coupling the app* with curated contacts to therapists to continue exploring the topic or to find tailored help to a specific issue. One interviewee thought about accompanying the app with television or radio shows, stating that this was the reason why they bought a book on sexuality. Another possible extension would be a course, workshops, or group activities that would let people discuss and further explore what they had learned and experienced:

> *In addition to using the app, it could be interesting to be in a discussion group with other couples as a couple. That could help improve communication about sex. The app provides plenty of conversation material for that.* [P11]

Interviewees had mixed opinions regarding *whether the app should be paid.* On the one hand, participants shared that they are not used to paying for apps, but on the other hand, they recognized that they might pay for extra services (eg, consultations) and that free apps do not have as much credibility. Credibility is something that participants cling to when reflecting with the interviewer about how to make the app available to more people. Participants concluded that the app could be credibly made available through medical doctors, therapists, or reliable associations. Although this was not asked, participants also shared ideas on how to raise awareness about the Anathema app, for example, through advertisement, television or radio shows, or leaflets.

Discussion

Principal Findings

The pilot study conducted in the Netherlands with a group of 15 community-dwelling older adults was a novel study in the field of mobile health apps in sexual health. Although the dropout rate was high (65%), no participants were lost to follow-up or nonuse cases, that is, participants answering the questionnaires without having used the mobile app until the end. We found that the app was usable, that participants showed high levels of self-confidence in using it, that the smartphone can be a useful and private way to have access to reliable sexual health information, that participants foresee how extra services could help tailor the program to their specific needs, and that certain improvements in content and in interaction are likely to increase user experience for this smartphone-delivered sexual health promotion program.

As with other studies in the literature [20], the user experience was negatively affected by a lack of social support for users' specific issues. In the interviews, participants gave examples of further content on love and affection, a curated list of resources and fitness exercises that they would like to see, and options to search through the content to get the information they were looking for. The lack of social support, ranging from relatives to professionals, also seems to have negatively affected participants' user experiences. In their systematic review, van Acker et al [19] noted how social support (ranging from relatives to professionals) was an important factor in user experience. In our study, with the exception of 1 participant who could not convince their sexual partner to also use the app, there were no reports of available or lacking support from relatives, but participants specifically mentioned that professional support would be useful in addition to the existing offer. A nuance with relation to the literature [19] is that participants in our study did not require much professional support to interact with the program, but rather as an extension to it, often to attain the personalization requirement we have just described earlier. Furthermore, the participants struggled with the locked content. Although the tunneling technique has been used to increase engagement with intervention or technology, in our study, it did not seem to have this effect. This is similar to recent findings with an intervention for a younger generation [31].

As noted in the literature [20], trust is also an important dimension in user experience. Although not directly asked about it, our interviewees alluded to the element of credibility regarding willingness to pay, which was considered by Hurmuz et al [20] as a metric of user experience. For the participants in our sample, the channel via which they access the app is an important factor at the time of choosing whether to use and ultimately pay for the app.

As measured by the SUS instrument, self-confidence among the participants in our study was high. This might also have been influenced by the level of education and digital literacy of the sample. The level of education might also explain why the participants often alluded to the experience of "being schooled" as a negative valence. Although the tone of voice for the program regarding visual and written content was co-designed [23], it might not have been implemented properly to eliminate this negative experience. This aspect is further discussed in the "Limitations" section below. On the other hand, some users also reacted negatively to content that seemed "too scientific," and some commented that some terms might not be easy to understand for the wider population. This is at odds with the higher educational level of this sample, but the explanation for the dislike might be related not to the understandability of the content but rather to a kind of experience that users expect when they are using an app that is related to sexuality.

The participants stated that the topic of sexuality was not embarrassing. However, there were some accounts of users requiring privacy when going through the content, one user whose partner did not want to use the app because of the topic, or users commenting on how they did not feel free or at ease discussing the topic with their peers. Therefore, the topic of taboo still requires further research in terms of how much of a barrier it is to accept and use technology around this topic. Participants' statements in the interviews suggested that a smartphone-based intervention can bring the advantages of ubiquity, intimacy, and anonymity to an intervention that is likely to elicit stigma in some contexts. The program itself was regarded as a trustworthy source of information that participants think is difficult to find on this topic. On the other hand, it could be coupled with more targeted personal services for users who would like to interact with therapists or even join groups willing to openly discuss topics of sexual health. Future research should study the provision of such discussion groups either in person or through moderated and anonymous forums inside the app.

Our study included participants interested in sexual health. In any case, even within our small sample, we witnessed a wide spectrum regarding taboo. For instance, some participants felt blurring genitalia photographs by default with overlaid text: "Sensitive content. Click to view" was condescending, whereas others felt that suggesting exercises for sexual pleasure was going too far. As with other types of apps targeting older adult users, our study saw a large heterogeneity in user preferences. Even if resources are allowed for the software development team to implement ultrapersonalization, we could place a large burden on users upon onboarding to set up preferences, which, in itself, would have a negative effect on technology acceptance. One way of addressing this could be to create certain user profiles and adapt scaffolding techniques that have been used for usability [32] for the purpose of conspicuousness degrees. Future research should work on this balance between a certain level of tailoring to one's needs and preferences, with time invested in customizing the app.

Strengths

This was the first study to evaluate the user experience of a self-guided, smartphone-delivered program to promote sexual health among older adults. The mixed methods approach was a strength of this study in the sense that it provided a rich description of participants' experiences with the app and the program. Without the interviews, we would hardly have had such detailed information that would indicate how to improve the app and the program, as well as a first understanding of how participants appropriated the app.

Our study did not aim at generalizability but rather at an in-depth understanding of user experience, which justified the emphasis on the qualitative data. Through this approach, we derived actionable insights to improve the content, structure, and format of the program.

Although our study was conducted with a small and specific sample of older adults in the Netherlands, the methodology we used allowed us to unveil nuances that can be useful for researchers to consider when implementing smartphone-based programs for sexual health in different populations: the relevance of social support, the credibility of the program, the opportunities that smartphone-based interventions may bring to sexual health interventions in terms of privacy or convenience, and the variability among program users about what might be considered a taboo and how this might impact users' preferences, practices, and attitudes toward the programs.

Limitations

As we conducted a user experience pilot study to obtain in-depth feedback, the results might not reflect the characteristics of the older adult population in the Netherlands. Although further research is needed to reach generalizability, this study constitutes a stepping stone in this journey.

The sample characteristics in our study are its greatest limitation. Only one-fifth of our participants were dissatisfied with the current state of their sex lives, and most considered themselves to be in fairly good or good health, which may not be representative of the older adult population. These characteristics may have biased how participants responded to a sexual health promotion program tailored to help users identify and cope with issues related to their sexual health. Our sample also comprised participants with a high level of education. This might explain why some participants felt schooled, as they were already in possession of information that was provided by the program. As participants have suggested, for a future pilot study, it would be advisable to increase the number and type of channels used for dissemination and recruitment, such as the mainstream media. This would help increase the visibility of Anathema and reduce, if not altogether, prevent, selection bias.

The features implemented on the app responded as much as possible to the user research requirements, but this was not always possible or perhaps implemented at its best. In some cases, there were technical limitations that did not allow their implementation. For instance, the app began to be implemented as web based so that it would also run on desktop browsers if participants preferred, but the identification of a problem in a technical component ahead of the implementation process forced the software development team to develop natively for Android and iOS.

We expected this lack of flexibility in the device type to be a negative aspect for some participants. On the other hand, at least once, the preferences collected from participants in user research studies preceding the pilot study were not aligned with the preferences of the pilot study sample. We describe 2 instances of this problem.

The first example relates to the choice of imagery. To select the photos for the app, we conducted a survey with 111 older adults in the Netherlands, in which we showed 10 different pictures and asked participants to rate the pictures, select their favorites, and justify their choices. The survey revealed that participants preferred uplifting, cheerful, and romantic images of participants who were not young but also not too old. The interviews in the pilot study revealed that, for some participants, these images were not appropriate.

Another example was the language used: a series of tests on the preferred tone of voice were used to create the original content in English [23]. The content was translated into Dutch, which went through content reviews from native speakers with experience in older adult care. Nevertheless, for some participants in the sample, the language was described as "too scientific." It is also possible that the research team was not able to correctly implement the insights from the user research phase, thus causing a mismatch between the users' expectations and the implemented app. Further research should revise the feedback from the user research phases and cross it with the results from the pilot study to understand where the app can be improved to meet users' expectations.

Further research should also focus on interaction and content issues to improve current mobile apps toward improving user experience. In particular, there is a need to understand how to balance the quantity and type of content with an engaging user experience. Once an improvement in user experience has been noted through further formative testing, the pilot should be repeated. As there was a mix of negative and positive comments provided by the participants and because the results from the SUS score are aligned with the comments from the interviews, we do not think that social desirability influenced participants' answers. However, as social desirability plays an important role in sex research surveys, a future pilot study could include a questionnaire (eg, [33]) to control for this effect. Further research should focus on a larger and more diverse sample regarding sexual satisfaction, health status, and literacy level.

Future pilots should include study designs that enable the collection of fine-grained data about the user experience combined with an assessment of the program's efficacy in improving sexual health so that the aspects of appropriation and how the app fits into participants' practices could be better understood and, in turn, inform strategies to improve sexual health outcomes, engagement, and user experience with such an intervention.

Conclusions and Implications for Design

The mobile app of Anathema with a sexual health promotion program delivered in a self-guided mode to a sample of older adults in the Netherlands was assessed as usable. Most participants tended to assess the app and program as useful, but both the app and the program would benefit from certain improvements, which we group under "content" and "interaction" as possible guidelines.

Content wise, readability and engagement can be improved by using plainer language in general, revising sections that sound "too scientific" (eg, definition of sexual health) or too medical (eg, content regarding erectile problems was very focused on the urological aspects). Although for some participants, the content was too long, and participants would also prefer not to have locked content, in which case the length would not be a barrier to engagement. It is clear that participants would appreciate more curated content that would refer them to support the community or to further services.

Regarding *interaction*, there are suggestions to enable searching and asking questions so that the user could be directly guided to the content that is of most interest to them or so that they could center their learning in their own experiences. This implies that content is unlocked by default. Participants would also like to bookmark certain sections and have the means to know where they are in the app. Finally, although participants shared that some exercises made them reflect—they saw this as positive—they struggled with the exercises that involved text input. Therefore, the interaction modes in the exercises can be improved. Although participants state they do not want to be schooled, they highlight "learning" as one of the advantages of using the app. In the future, the Anathema app should meet the goal of teaching without resembling a schoolbook. This was highlighted by participants who were expecting more interactivity from the app rather than an app that reads like a book.

Acknowledgments

Beyond the participants in this study, ≥1400 older adults from 3 different countries were involved in the design and testing of the program. The authors are very thankful to all the participants and to the secondary and tertiary users who provided valuable feedback.

The authors would like to acknowledge the European Commission through the Active Assisted Living (AAL) Programme (reference AAL-2020-7-133-CP) and the Member States: Fundação para a Ciência e a Tecnologia, in Portugal; the Austrian Research Promotion Agency (FFG), in Austria; and ZonMw, the Netherlands Organisation for Health Research and Development, in the Netherlands for funding this project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Posttest questionnaire.
[DOCX File , 45 KB - humanfactors_v11i1e56206_app1.docx]

Multimedia Appendix 2
Semistructured interview script.
[DOCX File , 39 KB - humanfactors_v11i1e56206_app2.docx]

References

1. World Health Organization. Defining sexual health: report of a technical consultation on sexual health. World Health Organization. 2006. URL: https://tinyurl.com/ye2aep8c [accessed 2024-03-11]
2. Vasconcelos P, Paúl C, Serruya SJ, Ponce de León RG, Nobre P. A systematic review of sexual health and subjective well-being in older age groups. Rev Panam Salud Publica 2022;46:e179 [FREE Full text] [doi: 10.26633/RPSP.2022.179] [Medline: 36320206]
3. Træen B, Hald GM, Graham CA, Enzlin P, Janssen E, Kvalem IL, et al. Sexuality in older adults (65+)—an overview of the literature, part 1: sexual function and its difficulties. Int J Sex Health 2016 Sep 14;29(1):1-10. [doi: 10.1080/19317611.2016.1224286]
4. Towler LB, Graham CA, Bishop F, Hinchliff S. Older adults' embodied experiences of aging and their perceptions of societal stigmas toward sexuality in later life. Soc Sci Med 2021 Oct;287:114355. [doi: 10.1016/j.socscimed.2021.114355] [Medline: 34474307]
5. Sinković M, Towler L. Sexual aging: a systematic review of qualitative research on the sexuality and sexual health of older adults. Qual Health Res 2019 Jul 25;29(9):1239-1254. [doi: 10.1177/1049732318819834] [Medline: 30584788]
6. Wang V, Depp CA, Ceglowski J, Thompson WK, Rock D, Jeste DV. Sexual health and function in later life: a population-based study of 606 older adults with a partner. Am J Geriatr Psychiatry 2015 Mar;23(3):227-233. [doi: 10.1016/j.jagp.2014.03.006] [Medline: 24840610]
7. Bertschi I, Meier F, Bodenmann G. Disability as an interpersonal experience: a systematic review on dyadic challenges and dyadic coping when one partner has a chronic physical or sensory impairment. Front Psychol 2021;12:624609 [FREE Full text] [doi: 10.3389/fpsyg.2021.624609] [Medline: 33732189]
8. Smith L, Yang L, Veronese N, Soysal P, Stubbs B, Jackson SE. Sexual activity is associated with greater enjoyment of life in older adults. Sex Med 2019 Mar;7(1):11-18 [FREE Full text] [doi: 10.1016/j.esxm.2018.11.001] [Medline: 30554952]
9. Landers S, Kapadia F. The public health of pleasure: going beyond disease prevention. Am J Public Health 2020 Feb;110(2):140-141. [doi: 10.2105/ajph.2019.305495]
10. Hunsaker A, Hargittai E. A review of internet use among older adults. New Media Soc 2018 Jul 16;20(10):3937-3954. [doi: 10.1177/1461444818787348]
11. van Lankveld J. Internet-based interventions for women's sexual dysfunction. Curr Sex Health Rep 2016 Jun 30;8(3):136-143 [FREE Full text] [doi: 10.1007/s11930-016-0087-9] [Medline: 27512346]
12. Martin P, Cousin L, Gottot S, Bourmaud A, de La Rochebrochard E, Alberti C. Participatory interventions for sexual health promotion for adolescents and young adults on the internet: systematic review. J Med Internet Res 2020 Jul 31;22(7):e15378 [FREE Full text] [doi: 10.2196/15378] [Medline: 32735217]
13. Karim H, Choobineh H, Kheradbin N, Ravandi MH, Naserpor A, Safdari R. Mobile health applications for improving the sexual health outcomes among adults with chronic diseases: a systematic review. Digit Health 2020 Feb 20;6:2055207620906956 [FREE Full text] [doi: 10.1177/2055207620906956] [Medline: 32128234]
14. Mlekus L, Bentler D, Paruzel A, Kato-Beiderwieden AL, Maier GW. How to raise technology acceptance: user experience characteristics as technology-inherent determinants. Gr Interakt Org 2020 Aug 06;51(3):273-283. [doi: 10.1007/s11612-020-00529-7]
15. Hornbæk K, Hertzum M. Technology acceptance and user experience: a review of the experiential component in HCI. ACM Trans Comput Hum Interact 2017 Oct 06;24(5):1-30. [doi: 10.1145/3127358]
16. Norman D. The Design of Everyday Things. Cambridge, MA: MIT Press; 1988.
17. Interaction Design Foundation. User experience (UX) design. The Interaction Design Foundation. 2023. URL: https://www.interaction-design.org/literature/topics/ux-design?page=7 [accessed 2023-11-27]
18. International Organization for Standardization. ISO 9241-11:2018(en), ergonomics of human-system interaction: part 11: usability: definitions and concepts. International Organization for Standardization. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso:9241:-11:ed-2:v1:en:sec:A [accessed 2023-11-27]
19. van Acker J, Maenhout L, Compernolle S. Older adults' user engagement with mobile health: a systematic review of qualitative and mixed-methods studies. Innov Aging 2023;7(2):igad007 [FREE Full text] [doi: 10.1093/geroni/igad007] [Medline: 37007638]

20. Hurmuz MZ, Jansen-Kosterink SM, Beinema T, Fischer K, Op den Akker H, Hermens HJ. Evaluation of a virtual coaching system eHealth intervention: a mixed methods observational cohort study in the Netherlands. Internet Interv 2022 Mar;27:100501 [FREE Full text] [doi: 10.1016/j.invent.2022.100501] [Medline: 35198411]
21. Abascal J, Nicolle C. Moving towards inclusive design guidelines for socially and ethically aware HCI. Interact Comput 2005 Sep;17(5):484-505. [doi: 10.1016/j.intcom.2005.03.002]
22. Kascak LR, Rébola CB, Sanford JA. Integrating universal design (UD) principles and mobile design guidelines to improve design of mobile health applications for older adults. In: Proceedings of the 2014 IEEE International Conference on Healthcare Informatics. 2014 Presented at: ICHI '14; September 15-17, 2014; Verona, Italy p. 343-348 URL: https://ieeexplore.ieee.org/document/7052509 [doi: 10.1109/ichi.2014.54]
23. Correia de Barros D, Couto da Silva J, Ramadani N, Mendes Santos C. Towards the inclusive design of a digital sexual health promotion programme. In: Proceedings of the 2014 IEEE International Conference on Healthcare Informatics. 2014 Presented at: ICHI '14; September 15-17, 2014; Verona, Italy p. 18 URL: https://tinyurl.com/52tj29jj
24. Mendes-Santos C, Quinta-Gomes AL, Pereira R, Vasconcelos P, Nobre P, Couto J, et al. A smartphone-delivered program (Anathema) to promote the sexual health of older adults, colorectal cancer survivors, and stroke survivors: protocol for a feasibility pilot randomized controlled trial. JMIR Res Protoc 2023 Jun 27;12:e46734 [FREE Full text] [doi: 10.2196/46734] [Medline: 37368469]
25. Bødker S, Kyng M. Participatory design that matters—facing the big issues. ACM Trans Comput Hum Interact 2018 Feb 13;25(1):1-31. [doi: 10.1145/3152421]
26. Mendes SC, Ramadani N, Pereira R, Vasconcelos P, Quinta-Gomes A, Nobre P. Older adults' acceptability of smartphone-based sexual health promotion programs. In: Proceedings of the 11th Scientific Meeting of the International Society for Research on Internet Interventions. 2022 Presented at: ISRII '22; September 18-21, 2022; Pittsburgh, PA URL: https://anathemaproject.eu/wp-content/uploads/2022/10/Anathema_poster_ISRII-scaled.jpg
27. Ribeiro J, Lima P, Nunes F. Trial monitor: scaffolding personalised web dashboards for human–computer interaction field trials. SoftwareX 2021 Dec;16:100883. [doi: 10.1016/j.softx.2021.100883]
28. Ensink CJ, Keijsers NL, Groen BE. Translation and validation of the system usability scale to a Dutch version: D-SUS. Disabil Rehabil 2024 Jan 27;46(2):395-400. [doi: 10.1080/09638288.2022.2160837] [Medline: 36573399]
29. The WHOQOL group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. Psychol Med 1998 May 01;28(3):551-558. [doi: 10.1017/s0033291798006667] [Medline: 9626712]
30. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud Exp 2009;4(3):23.
31. Kornfield R, Stamatis CA, Bhattacharjee A, Pang B, Nguyen T, Williams JJ, et al. A text messaging intervention to support the mental health of young adults: user engagement and feedback from a field trial of an intervention prototype. Internet Interv 2023 Dec;34:100667 [FREE Full text] [doi: 10.1016/j.invent.2023.100667] [Medline: 37746639]
32. Ribeiro J, Correia de Barros A. Efficiency of a video and a tutorial in teaching older adults to interact with smartphones. In: Proceedings of the 8th International Conference, UAHCI 2014, Held as Part of HCI International 2014 on Universal Access in Human-Computer Interaction: Aging and Assistive Environments. 2014 Presented at: UAHCI HCI '14; June 22-27, 2014; Crete, Greece p. 34-45. [doi: 10.1007/978-3-319-07446-7_4]
33. Hays RD, Hayashi T, Stewart AL. A five-item measure of socially desirable response set. Educ Psychol Meas 2016 Sep 07;49(3):629-636. [doi: 10.1177/001316448904900315]

Abbreviations

KBO-PCOB: Katholieke Bond van Ouderen - Protestant Christelijke Ouderen Bond
SUS: System Usability Scale

Edited by D Gooch; submitted 11.01.24; peer-reviewed by C Klein, C Abbatantuono, A Batrakoulis, S Markham; comments to author 09.02.24; revised version received 22.02.24; accepted 22.02.24; published 03.04.24.

Please cite as:
Correia de Barros A, Bergmans M, Hasanaj K, Krasniqi D, Nóbrega C, Carvalho Carneiro B, Vasconcelos PA, Quinta-Gomes AL, Nobre PJ, Couto da Silva J, Mendes-Santos C
Evaluating the User Experience of a Smartphone-Delivered Sexual Health Promotion Program for Older Adults in the Netherlands: Single-Arm Pilot Study
JMIR Hum Factors 2024;11:e56206
URL: https://humanfactors.jmir.org/2024/1/e56206
doi:10.2196/56206
PMID:38568726

Original Paper

Designing for Improved Patient Experiences in Home Dialysis: Usability and User Experience Findings From User-Based Evaluation Study With Patients With Chronic Conditions

Anna Aspelund[1], MSc; Paula Valkonen[1], MA; Johanna Viitanen[1], PhD; Virpi Rauta[2,3], PhD

[1]Department of Computer Science, Aalto University, Espoo, Finland

[2]Department of Nephrology, Helsinki University Central Hospital, University of Helsinki, Helsinki, Finland

[3]Abdominal Center, Nephrology, Helsinki University Hospital, Helsinki, Finland

Corresponding Author:
Anna Aspelund, MSc
Department of Computer Science
Aalto University
PO Box 15400
Espoo, FI-00076
Finland
Phone: 358 504763381
Email: anna.aspelund@aalto.fi

Abstract

Background: Chronic kidney disease affects 10% of the population worldwide, and the number of patients receiving treatment for end-stage kidney disease is forecasted to increase. Therefore, there is a pressing need for innovative digital solutions that increase the efficiency of care and improve patients' quality of life. The aim of the eHealth in Home Dialysis project is to create a novel eHealth solution, called eC4Me, to facilitate predialysis and home dialysis care for patients with chronic kidney disease.

Objective: Our study aimed to evaluate the usability, user experience (UX), and patient experience (PX) of the first version of the eC4Me solution.

Methods: We used a user-based evaluation approach involving usability testing, questionnaire, and interview methods. The test sessions were conducted remotely with 10 patients with chronic kidney disease, 5 of whom had used the solution in their home environment before the tests, while the rest were using it for the first time. Thematic analysis was used to analyze user test and questionnaire data, and descriptive statistics were calculated for the UMUX (Usability Metric for User Experience) scores.

Results: Most usability problems were related to navigation, the use of terminology, and the presentation of health-related data. Despite usability challenges, UMUX ratings of the solution were positive overall. The results showed noteworthy variation in the expected benefits and perceived effort of using the solution. From a PX perspective, it is important that the solution supports patients' own health-related goals and fits with the needs of their everyday lives with the disease.

Conclusions: A user-based evaluation is a useful and necessary part of the eHealth solution development process. Our study findings can be used to improve the usability and UX of the evaluated eC4Me solution. Patients should be actively involved in the solution development process when specifying what information is relevant for them. Traditional usability tests complemented with questionnaire and interview methods can serve as a meaningful methodological approach for gaining insight not only into usability but also into UX- and PX-related aspects of digital health solutions.

(JMIR Hum Factors 2024;11:e53691) doi:10.2196/53691

KEYWORDS

usability; UX; user experience; PX; patient experience; user-based evaluation; patients; eHealth; digital health solution; kidney disease; home dialysis

Introduction

Chronic kidney disease is a global health problem that leads to kidney failure, cardiovascular disease, and premature death. Chronic kidney disease affects 10% of the population and is one of the leading causes of mortality worldwide [1]. Dialysis, along with kidney transplant, is a lifesaving treatment for people with end-stage kidney disease. Dialysis can be delivered in hospital or home settings, and home dialysis is associated with a higher or equal quality of life for patients [2] and lower costs for the health care system [3]. As the number of patients receiving treatment for end-stage kidney disease is forecasted to rise [4], innovative digital solutions that maximize efficiency, improve patients' quality of life, and facilitate care delivery and monitoring are needed.

eHealth solutions, such as digital patient engagement platforms (DPEPs), are increasingly developed to support self-care, enhance patient-clinician collaboration, and increase the efficiency of care delivery [5-7]. In dialysis care, new DPEP solutions have the potential to improve disease management, health outcomes, and patient experience (PX) among patients with chronic conditions [8,9]. To achieve these goals, a human-centered design approach to development is a necessity. Human-centered design is an approach that aims to make digital systems usable and useful by applying human factors and usability techniques, such as user-based testing, guidelines for interaction design, prototypes, user observations, and user requirements specifications [10]. Usability refers to the interaction between the end user and the system, whereas the user experience (UX) includes aspects like emotions, beliefs, and perceptions [10,11]. Originating from the UX, the PX has also become an important and acknowledged concept as the health care sector has shifted to a more customer-oriented approach. PX has been used to describe patients' interactions and care experiences across the care continuum [12,13], but it lacks a consensus definition [14]. Regarding eHealth solutions, numerous factors influence PX, such as the solution type and quality, risks and concerns, communication, remote interaction, and patients' attitudes toward digital solutions [14].

Several studies have evaluated the usability of eHealth solutions aimed at patients with chronic and serious conditions. These have included solutions targeted to patients with cancer for monitoring and managing their illness or treatment-related symptoms [15-17], digital self-management programs for patients with juvenile idiopathic arthritis [18] and chronic obstructive pulmonary disease [19], and an electronic patient-reported outcome tool for patients with complex chronic disease and disability to set and monitor their health-related goals [20]. Common usability problems identified across these studies have included terminology issues [15,18], navigation problems [15,17], and challenges with the way information is presented to the patients [16,18]. Regarding UX, studies have found that patients' illness-related problems and limitations should be taken into account when designing eHealth solutions for patients with chronic and serious conditions [16,19,20]. Further, customization of the solutions, for example, based on the stage or severity of the illness or type of treatment should be possible to provide a pleasant UX [16,17,19]. Some prior studies have also reported PX-related findings, such as patients fearing that the eHealth solutions will replace in-person consultations with clinicians [20], and patients generally welcoming the additional digital communication channel [16,17]. However, these results have not been analyzed or described in relation to PX, and it seems that PX-related aspects were not systematically explored in the evaluation studies.

In this paper, we report a user-based evaluation study of the novel eHealth solution: a DPEP targeted to patients with chronic kidney disease in CKD stages 4-5, for example, to patients undergoing predialysis and patients undergoing home dialysis (both peritoneal dialysis and home hemodialysis). Patients with functioning renal transplants were excluded. This study is part of the larger eHealth in Home Dialysis project [21], which is coordinated by HUS Helsinki University Hospital, Finland. The solution is designed to facilitate advanced home care: enable patients with chronic kidney disease to document their treatment data, monitor their clinical and health data, order dialysis supplies, and report their symptoms as well as enhance patient-provider communication. The objective of our study was to evaluate the first version of the DPEP solution, called eC4Me, and support deployment of the solution and promote end user participation in later phases of the development. The research questions are as follows: (1) What kind of usability problems does the evaluated DPEP solution have? (2) What kind of UXs, expectations, and improvement ideas do patients with chronic kidney disease have about the new DPEP solution? and (3) How can the new DPEP solution support positive PX for patients with chronic conditions?

Our user-based evaluation study aims to widen the scope of usability evaluations of eHealth solutions targeted at patients with chronic and serious conditions to include PX-related aspects alongside usability and UX. Additionally, to our best knowledge, this is the first study to evaluate a DPEP solution specifically targeted to patients with chronic kidney disease.

Methods

Study Design

Our study design was based on a formative user-based evaluation approach [22]. The formative evaluation aims to support the improvements of the system, particularly the user interface, as part of an iterative design process [22]. A typical method of formative evaluation is a think-aloud usability test, which includes 4 stages: preparation, introduction, the test itself, and debriefing [22]. Usability testing is stated to be the most fundamental usability method since it provides direct information about how people use the systems and what their exact problems are [22]. In practice, the think-aloud method involves the test participants continuously thinking aloud while performing the predefined test tasks. The researcher's role is to make observations and continuously prompt the participant to think aloud by asking general questions [22].

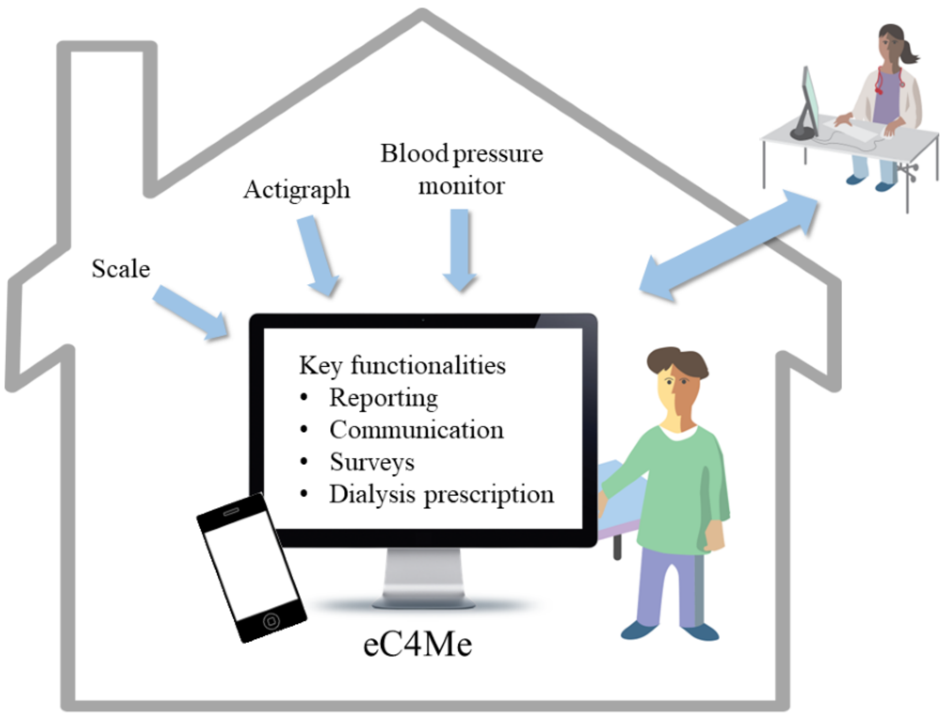
The usability assessment methods recommended for gathering supplementary data are observations, questionnaires, and interviews [22]. For the questionnaire, we used the UMUX (Usability Metric for User Experience) [23], which closely conforms to the 3 widely acknowledged attributes of usability—effectiveness, efficiency, and satisfaction [10]—and strongly correlates with other commonly used usability metrics such as the System Usability Scale [24,25]. UMUX questionnaire is considered compact since it includes four question items: (1) the system's capabilities meet my requirements, (2) using the system is a frustrating experience, (3) the system is easy to use, and (4) I have to spend too much time correcting things with this system [23].

Due to the COVID-19 pandemic, we conducted the evaluation sessions remotely. Experiences from previous studies have shown that high-quality research data can be collected remotely [26]. However, compared to traditional face-to-face usability testing, remote evaluations require more pedantic preparation. Researchers must pay attention to building trust and confidentiality [26], choose tools that are familiar and easily accessible for the participants [27], focus on body language and facial expressions to establish rapport [26,28], and provide the participants with technical support as needed [29]. Other recommendations include having multiple researchers participate in remote test sessions [28], and using a synchronous approach, which makes it possible to observe participants' screens in real-time [30].

Evaluated DPEP Solution

The first version of the eC4Me solution was introduced in autumn 2021. A core part of the solution was an app, which had both computer and mobile interfaces and included the following key functionalities: monitoring of treatment-related data (reporting functionality), messaging between patients and nurses (communication functionality), answering quality-of-life surveys (survey functionality), and access to patient's dialysis prescriptions (dialysis prescription functionality; see Figure 1). In addition, the solution delivered to the patients included external monitoring devices, such as a blood pressure monitor, a scale, and an actigraph, which, together with clinical data collected from electronic health records, enabled patients to monitor their conditions.

Figure 1. First version of the eC4Me solution.



Participants

The participants were recruited from a large university-affiliated nephrology clinic with the help of research nurses. All participants were familiar with the eHealth in Home Dialysis project since they had participated in an interview study that was conducted earlier as part of the project.

Eleven patients with chronic kidney disease were originally invited to participate in this study. As 1 test participant was particularly interested in technology development and came to his test session with predrafted design ideas, we decided to modify his test session to focus on discussing these ideas. Consequently, he did not perform the test tasks in our test procedure, and the data from his session was omitted from this study. Therefore, data from 10 user-based evaluation sessions were included in this study.

The background information collected from the participants included age, gender, type of treatment, occupational status, and technical skills. Technical skills were evaluated by asking participants to give their own estimation of their technical skills, with response options being "good," "basic," and "weak." Half of the participants (n=5) had the solution delivered to their homes 1-3 weeks before the tests, whereas the other half (n=5) were using the solution for the first time in their test session. Participant characteristics are shown in Table 1.

Table 1. Participant characteristics (N=10).

Characteristics	Participants, n (%)
Type of treatment	
Predialysis (not yet in dialysis treatment)	4 (40)
Peritoneal home dialysis	3 (30)
Home hemodialysis	3 (30)
Experience with the solution before the test session	
No experience	5 (50)
1-3 weeks of experience	5 (50)
Sex	
Male	6 (60)
Female	4 (40)
Age (y)	
30-60	4 (40)
>60	6 (60)
Occupational status	
Fully working	2 (20)
Partially working	2 (20)
Not working (fully retired or sick leave)	6 (60)
Technical skills (own estimation)	
Good	2 (20)
Basic	7 (70)
Weak	1 (10)

Test Procedure

In our study, we used synchronous remote usability testing with Microsoft Teams as a tool, and 2 researchers were present in each session. During the test sessions, patients used the eC4Me solution with a computer, enabling the researchers to monitor their task performance via screen share on Microsoft Teams.

Each test session followed the traditional structure and stages of usability testing [22], lasted about 2 hours, and included the following phases: (1) introduction—participants were introduced to the evaluation study and given an opportunity to become familiar with Microsoft Teams; (2) test tasks—run-through of predefined usability test tasks, which included logging in, searching for information and functionalities, viewing and interpreting health-related data, reporting treatment-related data, and filling in the surveys; (3) questionnaire—participants answered the UMUX questionnaire; and (4) interview—participants answered semistructured interview questions to elaborate their UMUX scores and give overall feedback on the solution based on the usability test tasks. The interview consisted of 4 open-ended "Why did you give this score?" questions, which were asked for each of the UMUX items separately, and a question on how participants would improve the solution.

Before the actual tests, the test procedure was piloted with 2 research nurses. To ensure privacy, all patients used the solution with test login IDs and dummy health data during the test sessions. The exact test tasks varied slightly between the participants, depending on their prior experience with the solution, illness stage, and type of treatment, as not all functionalities of the solution were relevant for all patients. The participants who had used the solution before the tests were also encouraged to provide feedback on the entire solution including the research devices and a mobile interface. This study was performed in the Finnish language.

Data Analysis

The qualitative data included video recordings from remote usability tests, observation notes, and transcripts from semistructured interviews. The qualitative data were analyzed following a thematic analysis method [31], which involved collaboration between 3 researchers (AA, PV, and JV). The data were first coded by 1 researcher (PV), and the findings were discussed by the 3 researchers. Further, 2 researchers (AA and JV) then continued the analysis by categorizing the codes into thematic groups, following the principles of the affinity diagram method [32]. The following main three thematic groups were used: (1) *usability*, which includes findings about users' interactions with the DPEP solution; (2) *UX*, which includes findings about users' experiences and feelings toward using the DPEP solution; and (3) *PX*, which includes findings about how the DPEP solution can support patients' interactions and care experiences across the care continuum.

The researchers (AA and JV) then continued the analysis with several rounds of iterations. Along with other data, improvement ideas expressed by the participants were thematically grouped. At the end of the analysis, the thematic grouping of observations was discussed, approved, and finalized collaboratively by the 3 researchers.

The quantitative data consisted of UMUX item scores, which were analyzed following the UMUX scoring scheme [23]: to obtain the overall UMUX score, items 1 and 3 were scored as [score−1] and items 2 and 4 as [7−score], and the sum of the item scores was then divided by 24 and multiplied by 100. In addition to the overall score, the means and SDs for each of 4 question items were calculated separately for 2 participant groups (patients who had or had not used the solution before the test). The differences between the groups were analyzed using *t* tests for independent samples. The tools used for data analysis were ATLAS.ti (ATLAS.ti Scientific Software Development GmbH) and Microsoft Excel for qualitative data analysis and Microsoft Excel for statistical analysis.

Ethical Considerations

This study has a research permit from the ethical committee of the Hospital District of Helsinki and Uusimaa (HUS/1649/2020).

Results

Overview

The results are divided into 5 topics: usability, UX and PX findings each, UMUX results, and improvement ideas.

Usability

The *usability* findings of the evaluated eC4Me solution consisted of 8 subthemes (Table 2).

Navigation includes findings about whether patients could locate the functionalities, content, and commands that they were looking for. Nine out of 10 users had at least some problems navigating the app, and the most common navigation challenges were related to users not understanding the content structure or the terminology used in the menus.

Table 2. *Usability*, *UX*, and *PX* findings of the user-based evaluation. “All findings” includes positive, negative, and neutral findings. For the *usability* theme, negative findings, that is, the identified usability problems, are also reported separately under “problems.”

Subtheme	All findings		Problems	
	Codes, n	Users, n	Codes, n	Users, n
Usability				
Navigation	41	10	19	9
Terminology	35	10	22	8
Front page	21	10	3	2
Presentation of data	35	9	17	5
Login	10	9	4	4
Survey functionality[a]	12	7	2	2
Reporting functionality[a]	50	6	29	6
Dialysis prescription functionality[a]	6	5	0	0
UX[b]				
Technical functionality	26	9	—[c]	—
Use of access devices (computer, tablet, or mobile)	26	9	—	—
Workload and effort	16	7	—	—
Perceived benefits	12	7	—	—
Security	8	4	—	—
PX[d]				
Content-related needs	67	10	—	—
Situation of use	57	10	—	—
Communication with clinicians	52	10	—	—

[a]The survey, reporting, and dialysis prescription functionalities were tested with some of the participants only (n=7, n=6, and n=5, respectively).

[b]UX: user experience

[c]Not applicable.

[d]PX: patient experience

Terminology includes findings about the comprehensibility and clarity of the terminology used. Eight users (80%) had problems understanding the terminology, and approximately half (10/22) of the terminology challenges were related to problems with understanding medical- or treatment-related terminology. Other terminology issues included problems with the terms used in the menus as well as the use of a foreign language.

Front page includes findings about the comprehensibility and clarity of the front-page contents. The front page of the tested version contained relatively little information and functionalities, and most users found it simple and clear.

Presentation of data includes findings about the comprehensibility and clarity of the presentation of health data, such as health measurements. Five users (50%) had issues understanding or viewing the data. The most common challenges were not comprehending the data or graphs or not knowing how to adjust the scales and timelines to view the data in a meaningful way.

Login includes findings about the ease of logging in. Four users (40%) had problems logging into the system. Typical challenges included not understanding where to input the login information or making errors while typing the login details.

Surveys, reporting, and dialysis prescription functionalities include findings about the ease of use of these functionalities. All users who tested the reporting functionality (6/6) had problems using it. Users struggled with not understanding what they should type in the input fields, feeling that options in the fields did not match the way treatment was provided in the real world, or not comprehending the medical- or treatment-related terminology. In this study, there were few usability issues in the survey functionality and none in the dialysis prescription functionality.

About UX

The *UX* findings of the evaluated eC4Me solution consisted of 5 subthemes (Table 2).

Technical functionality includes patients' experiences and feedback regarding the technical aspects of the eC4Me solution. Four users (40%) expressed frustration because some information they thought should be transferred automatically between the app, the research devices, the home dialysis machine, and patient information systems had to be typed manually. For the same reason, 2 users (20%) felt that they needed to use several systems for essentially the same purpose, such as monitoring their health data.

Use of access devices includes patients' expressed preferences regarding using the solution with different access devices: desktop computer, tablet, or mobile phone. Two users (20%) said they would prefer to use the computer interface, as they have found tablet and mobile keyboards difficult to use or feel that the mobile interface would give them less information. In contrast, 3 (30%) users indicated that they preferred a mobile phone or tablet as they are readily at hand and easier to use during the treatment, while another 3 (30%) said that their choice of access device would depend on the task they were performing.

Workload and effort includes findings about the perceived time and effort required to use the solution. Six users (60%) felt that the solution was not burdensome to use as such and that filling in the surveys or documenting treatment details did not take too much time. However, 4 users (40%) expressed concern that the solution might nevertheless increase their burden if it does not replace any other service, thus becoming one more thing to use and keep track of on top of all the other health-related solutions.

Perceived benefits includes patients' thoughts about the benefits and added value of the eHealth solution. One (10%) user saw value in using the solution primarily for the benefit of the health care personnel, while 2 (20%) others said that they needed to see clear benefits for themselves to be motivated to use the solution. Yet another user mentioned that the data generated by the solution could benefit all patients, as it could be used for research and treatment development.

> *I'm uncertain what this is meant for, is it for my benefit or someone else's? The remote measuring devices that I have had, I have found the data very useful for myself.... But I don't understand the thinking behind this (the solution), do I benefit or is it someone else?* [P7]

Security includes findings about potential security issues and patients' concerns regarding the use of the eHealth solution. Only 2 (20%) users gave direct comments on security aspects, while most findings related to security were observations of behaviors that could introduce potential security risks, such as the user closing the browser instead of logging out when asked to do so.

About PX

The *PX* findings of the evaluated eC4Me solution consisted of 3 subthemes (Table 2).

Content-related needs includes patients' comments regarding health-related data that they want to see so they can monitor and manage their treatment and health. The expressed needs and what was considered most important varied between the users, but overall, patients were interested in seeing all the types of data that the tested version of the solution provided. Only 1 user (10%) gave a general comment that the solution "should not contain anything unnecessary or useless," but other than that, none of the users reported that they would not need or want to see some of the information or data that was available to them.

Situation of use includes patients' comments and feedback about how well the solution fits their situations and supports their everyday lives with the disease. Users had numerous, often variable comments regarding how often and in what situations they would likely use the solution. They also commented on how well the functionalities fit their care and treatment schedules, as in the following quote:

> *I fill these during my home dialysis treatment, so I may write notes about yesterday's treatment. I don't necessarily have time to use [the solution] after the treatment.* [P8]

Communication with clinicians includes findings about how the new solution supports patient-clinician communication. Users expressed interest in using the messaging function and saw benefits in using the documented data to facilitate their communication with clinicians during face-to-face appointments. It was not clear to the users how actively and by whom their data were being monitored and if messages were noticed and replied to. Three users (30%) were hoping for immediate feedback, while 4 others (40%) considered the messaging function appropriate for nonurgent communication. In addition, 5 users (50%) expected their own nurse to read and respond to their messages, while 3 users (30%) thought that the work was handled by a care team.

UMUX Results

The UMUX score of the first version of eC4Me was 70.6 (SD 18.6), which indicates an average level of usability [25].

Means for individual UMUX items are presented in Table 3. Users with 1-3 weeks of prior experience with the solution rated it more favorable overall compared to users without prior experience. However, the differences between the groups were not statistically significant.

Table 3. UMUX[a] item scores per user groups on a scale of 1 "strongly disagree" to 7 "strongly agree."

UMUX questionnaire item	All users (n=10), mean (SD) score	Users with no prior experience using the solution (n=5), mean (SD) score	Users with 1-3 weeks of experience using the solution (n=5), mean (SD) score
The solution's capabilities meet my requirements	4.5 (1.4)	4.0 (1.6)	5.0 (1.0)
Using the solution is a frustrating experience	2.5 (1.9)	3.0 (2.3)	1.9 (3.0)
The solution is easy to use	5.7 (1.5)	5.0 (1.9)	6.4 (0.5)
I have to spend too much time correcting things with the solution	2.8 (1.9)	3.0 (1.4)	2.6 (2.5)

[a]UMUX: Usability Metric for User Experience.

Improvement Ideas

In total, 66 improvement ideas (Table 4) for the eC4Me solution were identified from the data, with all 10 users expressing at least one improvement idea. Two-thirds of the ideas (40/66) came from users who had used the solution before the test.

The most common theme for improvement ideas was *content-related needs*. Seven patients (70%) expressed interest in monitoring some health-related measurements that were not available in the tested version, and 3 (30%) patients wanted to see benchmark values or descriptions that would enable them to better understand their health data.

Table 4. Improvement ideas and their most common subthemes.

Subtheme	All improvement ideas	
	Codes, n	Users, n
Content-related needs	28	9
Situation of use	15	7
Communication with clinicians	15	6
Presentation of data	8	6
Ease of using reporting	13	5
Technical functionality	8	5
Other[a]	26	8

[a]Subthemes with fewer than 5 ideas (combined).

In addition, the participants brought up improvement ideas related to the following: (1) *situation of use*—ideas on how the solution could be improved to better fit the patient's situation, everyday life, and treatment schedule; (2) *communication with clinicians*—ideas on how the solution could better support communication and data exchange between patients and clinicians; (3) *presentation of data*—ideas on how health data could be presented to make them more meaningful for the patients; (4) *ease of using reporting*—ideas on how to improve the reporting functionality to make it easier to use; and (5) *technical functionality*—ideas regarding automatic data exchange between the solution and other devices or services.

Discussion

Main Contribution

Our user-based evaluation study of the novel DPEP solution targeted to patients with chronic kidney disease with 10 participants resulted in a wide variety of usability-, UX-, and PX-related findings.

Most usability problems of the first version of the solution were related to navigation, the use of terminology, and the presentation of health data. Many participants struggled with the reporting functionality, which was one of the key functionalities of the solution. A considerable number of patient participants also expressed improvement ideas related to these themes. We decided not to classify usability problems by severity, as a proper severity rating should consider not only usability aspects but also potential medical- and health-related consequences of users' mistakes and misunderstandings. However, the usability challenges identified in our study were remarkably similar to those found in evaluations of other eHealth solutions aimed at patients with chronic and serious conditions [15-18]. Our findings thus emphasize the importance of using terminology and presenting health data in a way that is understandable and meaningful to patients. Our results also highlight the need to consider patients as end users when designing user interfaces for eHealth solutions.

Our study identified several challenges related to the UX of the evaluated DPEP solution. Largely due to deficiencies in integration and data exchange, the participants feared that the solution might create additional tasks and thus increase their burden. Our results also showed considerable variation in the expected benefits of the solution. Some patients wanted to see direct value for themselves, whereas others mentioned benefits for the health care professionals as their primary motivation of use.

Despite the usability and UX challenges, the patients' overall ratings of the evaluated solution were surprisingly positive. This may be at least partially explained by findings from previous studies, which have shown that patients with chronic and serious conditions often express high interest in disease-specific eHealth solutions [17,18], even when experiencing severe usability challenges [20]. In our study, patients who had used the solution for a few weeks in a home setting evaluated it more positively than patients who were using the solution for the first time. Although the differences were not statistically significant due to the small number of participants, these initial findings could simply be explained by the fact that learning to use the solution makes it easier and thus more pleasant to use. However, they could also indicate that after having used the solution in their home setting with their own health data, patients have a better understanding of the benefits and potential value of the solution.

Regarding PX, our study generated insights on how the DPEP solution can support patients with chronic conditions in monitoring and managing their conditions and how the DPEP solution could better fit their everyday lives with a disease. From the patients' perspective, it is not enough that an eHealth solution is easy or pleasant to use if it does not support their health-related goals, feel meaningful, and fit their real-life situations and daily care activities. Special attention needs to be paid to ensure that these PX-related considerations are included in user-based evaluations of eHealth solutions, as generic usability questionnaires, classifications, and frameworks do not adequately capture these aspects [33,34]. As our study shows, traditional usability tests, complemented with questionnaire and related interview methods, can serve as a meaningful methodological approach for collecting information about PX-related aspects of eHealth solutions.

In our study, the participants generated a considerable amount of improvement ideas. In particular, nearly all patients had ideas on what health-related data they would like to see to better manage their condition. This implies 2 things. First, many patients with chronic conditions are interested in taking responsibility for their own care. Although the participants selected for our study are likely to represent the most motivated and active patients with chronic kidney disease, it would seem meaningful to support and empower these motivated patients to take more responsibility by providing them with the information they view as important and meaningful, not only the information that makes the most sense from a health care professionals' point of view. Second, patients should be actively involved in the co-design process in the early phases of solution development and when specifying what kind of information is relevant for them.

The user-based evaluation was a crucial step in the eHealth solution development process and generated findings that helped to make substantial changes in the solution to make it more suitable for the end users (patients), thus helping the solution reach its goals. The evaluation of usability-, UX-, and PX-related aspects of the solution will continue in a future research project. We aim to conduct a similar study in a further phase of the development project to examine how the usability and UX of the solution have been improved.

Limitations

Due to the COVID-19 pandemic, our study used remote testing as an evaluation method. In comparison to in-context evaluations, this limited the scope of our evaluation, as we could not fully observe participants in their home environments. We also decided not to include research devices that were part of the DPEP solution in the test procedure, as this would have been difficult to realize in the remote setup. However, when compared to face-to-face usability testing in laboratory settings, our arrangement also had some advantages. As contextual factors are well known to influence emotional experiences and expressions [35], allowing patients to remain in their natural home settings during test sessions likely produced more reliable data, especially regarding experience-related topics like UX and PX.

Further Research

Our findings indicate that including participants who have used the evaluated solution before the test can have a nonnegligible effect on the quality and amount of information that the evaluation study generates. Half of the participants in our study had used the evaluated solution in their home environment with their own health data, while the other half were using the solution for the first time in their test session. As the number of participants was small (n=10) and the groups were heterogeneous in terms of other background variables, it was not meaningful to make more comprehensive comparisons between the 2 groups. However, participants with prior experience evaluated the solution more favorably and generated more improvement ideas. Many UX- and PX-related aspects, such as perceived benefits, workload, and compatibility with everyday life, can be difficult to assess using a solution only in a test setting, especially considering that privacy issues often prevent researchers from using patients' own health data in user tests. This could have implications on how user-based evaluation studies of eHealth solutions should ideally be arranged, and it is therefore an important topic for further research.

In addition, further research is needed to explore the relationship and connections between the concepts of UX and PX, as suggested by recent review studies [14,36]. This includes planning and practicalities of user-based evaluation studies, considering the PX perspective, for the assessment and improvement of eHealth services.

Conclusions

User-based evaluation can produce valuable findings about usability aspects but also about the UX and PX of the evaluated DPEP solution. The findings of our study can be used in the development process to improve the evaluated solution from the perspective of patients with chronic conditions. Evaluation is a useful and necessary part of the solution development process, especially considering the high number of novel eHealth solutions that are currently being developed.

Our study also highlights the importance of understanding how digital health solutions for patients with chronic and serious conditions support patients' own health-related goals and fit their lives with disease. To fully understand the motivation for using such solutions, it is necessary to understand how patients perceive the benefits versus the effort required to use the solution in their everyday lives.

Acknowledgments

The eHealth in Home Dialysis project was part of the CleverHealth Network ecosystem. The project was funded by participating companies, the Hospital District of Helsinki and Uusimaa (HUS), and Business Finland. We thank our participants, project partners, and colleagues: S Hölsä, S Leinonen, and N Karisalmi from Aalto University and T Puurtinen, P Silvast-Äikäs, and A Varila from HUS.

Conflicts of Interest

None declared.

References

1. Kovesdy CP. Epidemiology of chronic kidney disease: an update 2022. Kidney Int Suppl 2022;12(1):7-11 [FREE Full text] [doi: 10.1016/j.kisu.2021.11.003] [Medline: 35529086]
2. Bonenkamp AA, van Eck van der Sluijs A, Hoekstra T, Verhaar MC, van Ittersum FJ, Abrahams AC, et al. Health-related quality of life in home dialysis patients compared to in-center hemodialysis patients: a systematic review and meta-analysis. Kidney Med 2020;2(2):139-154 [FREE Full text] [doi: 10.1016/j.xkme.2019.11.005] [Medline: 32734235]
3. Howell M, Walker RC, Howard K. Cost effectiveness of dialysis modalities: a systematic review of economic evaluations. Appl Health Econ Health Policy 2019;17(3):315-330. [doi: 10.1007/s40258-018-00455-2] [Medline: 30714086]
4. Liyanage T, Ninomiya T, Jha V, Neal B, Patrice HM, Okpechi I, et al. Worldwide access to treatment for end-stage kidney disease: a systematic review. Lancet 2015;385(9981):1975-1982. [doi: 10.1016/S0140-6736(14)61601-9] [Medline: 25777665]
5. Shaw T, McGregor D, Brunner M, Keep M, Janssen A, Barnet S. What is eHealth (6)? Development of a conceptual model for eHealth: qualitative study with key informants. J Med Internet Res 2017;19(10):e324 [FREE Full text] [doi: 10.2196/jmir.8106] [Medline: 29066429]
6. Campbell K, Louie P, Levine B, Gililland J. Using patient engagement platforms in the postoperative management of patients. Curr Rev Musculoskelet Med 2020;13(4):479-484 [FREE Full text] [doi: 10.1007/s12178-020-09638-8] [Medline: 32388724]
7. Hölsä S, Viitanen J, Valkonen P, Lääveri T, Rauta V. Developing eHealth for Home Dialysis: Clinicians' Needs for a Digital Patient Engagement Platform. Studies in health technology and informatics 2024;310:1111-1115. [doi: 10.3233/shti231137]
8. Shen H, van der Kleij RMJJ, van der Boog PJM, Chang X, Chavannes NH. Electronic health self-management interventions for patients with chronic kidney disease: systematic review of quantitative and qualitative evidence. J Med Internet Res 2019;21(11):e12384 [FREE Full text] [doi: 10.2196/12384] [Medline: 31687937]
9. Cartwright EJ, Zs Goh Z, Foo M, Chan CM, Htay H, Griva K. eHealth interventions to support patients in delivering and managing peritoneal dialysis at home: a systematic review. Perit Dial Int 2021;41(1):32-41 [FREE Full text] [doi: 10.1177/0896860820918135] [Medline: 32301380]
10. ISO 9241-210:2019 ergonomics of human-system interaction—part 210: human-centred design for interactive systems. International Organization for Standardization. URL: https://www.iso.org/standard/77520.html [accessed 2023-07-19]
11. Hassenzahl M. User experience and experience design. In: Soegaard M, Dam RF, editors. The Encyclopedia of Human-Computer Interaction, 2nd Edition. Aarhus, Denmark: The Interaction Design Foundation Publisher; 2013.
12. Wolf JA, Niederhauser V, Marshburn D, LaVela SL. Defining patient experience. Patient Exp J 2014;1(1):7-19. [doi: 10.1097/naq.0000000000000293]
13. Meyer MA. COVID-19 pandemic accelerates need to improve online patient engagement practices to enhance patient experience. J Patient Exp 2020;7(5):657-664 [FREE Full text] [doi: 10.1177/2374373520959486] [Medline: 33294595]

14. Viitanen J, Valkonen P, Savolainen K, Karisalmi N, Hölsä S, Kujala S. Patient experience from an eHealth perspective: a scoping review of approaches and recent trends. Yearb Med Inform 2022;31(1):136-145 [FREE Full text] [doi: 10.1055/s-0042-1742515] [Medline: 36463871]
15. Hanghøj S, Boisen KA, Hjerming M, Elsbernd A, Pappot H. Usability of a mobile phone app aimed at adolescents and young adults during and after cancer treatment: qualitative study. JMIR Cancer 2020;6(1):e15008 [FREE Full text] [doi: 10.2196/15008] [Medline: 31895046]
16. Mirkovic J, Kaufman DR, Ruland CM. Supporting cancer patients in illness management: usability evaluation of a mobile app. JMIR mHealth uHealth 2014;2(3):e33 [FREE Full text] [doi: 10.2196/mhealth.3359] [Medline: 25119490]
17. Moradian S, Krzyzanowska MK, Maguire R, Morita PP, Kukreti V, Avery J, et al. Usability evaluation of a mobile phone-based system for remote monitoring and management of chemotherapy-related side effects in cancer patients: mixed-methods study. JMIR Cancer 2018;4(2):e10932 [FREE Full text] [doi: 10.2196/10932] [Medline: 30578238]
18. Stinson J, McGrath P, Hodnett E, Feldman B, Duffy C, Huber A, et al. Usability testing of an online self-management program for adolescents with juvenile idiopathic arthritis. J Med Internet Res 2010;12(3):e30 [FREE Full text] [doi: 10.2196/jmir.1349] [Medline: 20675293]
19. Voncken-Brewster V, Moser A, van der Weijden T, Nagykaldi Z, de Vries H, Tange H. Usability evaluation of an online, tailored self-management intervention for chronic obstructive pulmonary disease patients incorporating behavior change techniques. JMIR Res Protoc 2013;2(1):e3 [FREE Full text] [doi: 10.2196/resprot.2246] [Medline: 23612363]
20. Gray CS, Gill A, Khan AI, Hans PK, Kuluski K, Cott C. The electronic patient reported outcome tool: testing usability and feasibility of a mobile app and portal to support care for patients with complex chronic disease and disability in primary care settings. JMIR mHealth uHealth 2016;4(2):e58 [FREE Full text] [doi: 10.2196/mhealth.5331] [Medline: 27256035]
21. Home dialysis. CleverHealth Network. URL: https://www.cleverhealth.fi/en/development-of-home-dialysis/ [accessed 2023-07-19]
22. Nielsen J. Usability Engineering. San Francisco, CA: Morgan Kaufmann; 1994.
23. Finstad K. The usability metric for user experience. Interact Comput 2010;22(5):323-327. [doi: 10.1016/j.intcom.2010.04.004]
24. Borsci S, Federici S, Bacci S, Gnaldi M, Bartolucci F. Assessing user satisfaction in the era of user experience: comparison of the SUS, UMUX, and UMUX-LITE as a function of product experience. Int J Hum-Comput Interact 2015;31(8):484-495. [doi: 10.1080/10447318.2015.1064648]
25. Lewis JR. Measuring perceived usability: the CSUQ, SUS, and UMUX. Int J Human-Computer Interact 2018;34(12):1148-1156. [doi: 10.1080/10447318.2017.1418805]
26. Valkonen P, Karisalmi N, Kaipio J, Kujala S. Remote interviews and visual timelines with patients: lessons learned. Stud Health Technol Inf 2021;281:845-849. [doi: 10.3233/shti281]
27. Amberkar S, Bennett K, Delchamps A, Edwards R. Expanding the playtest method for UX research beyond gaming. Proc Hum Factors Ergon Soc Annu Meet 2019;63(1):2239-2243. [doi: 10.1177/1071181319631119]
28. Judge TK, Neustaedter C. Studying and Designing Technology for Domestic Life: Lessons from Home, 1st Edition. Burlington, MA: Morgan Kaufmann; 2014.
29. Petrie H, Hamilton F, King N, Pavan P. Remote usability evaluations with disabled people. 2006 Presented at: CHI '06: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; April 22-27, 2006; Montreal, QC p. 1133-1141 URL: https://dl.acm.org/doi/proceedings/10.1145/1124772 [doi: 10.1145/1124772.1124942]
30. Vasalou A, Ng B, Wiemer-Hastings P, Oshlyansky L. Human-moderated remote user testing: protocols and applications. In: Proceedings of the 8th ERCIM Workshop, User Interfaces for All. 2004 Presented at: 8th ERCIM Workshop, User Interfaces for All; June 28-29, 2004; Vienna, Austria URL: https://reed.cs.depaul.edu/peterh/papers/b2hd-VasalouERCIM2004.html
31. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
32. Holtzblatt K, Beyer H. Contextual Design: Design for Life. Second Edition. Cambridge, MA: Elsevier; 2017.
33. Broekhuis M, van Velsen L, Hermens H. Assessing usability of eHealth technology: a comparison of usability benchmarking instruments. Int J Med Inform 2019;128:24-31. [doi: 10.1016/j.ijmedinf.2019.05.001] [Medline: 31160008]
34. Broekhuis M, van Velsen L. Improving usability benchmarking for the eHealth domain: the development of the eHealth Usability Benchmarking Instrument (HUBBI). PLoS One 2022;17(2):e0262036 [FREE Full text] [doi: 10.1371/journal.pone.0262036] [Medline: 35176033]
35. Greenaway KH, Kalokerinos EK, Williams LA. Context is everything (in emotion research). Soc Personal Psychol Compass 2018;12(6):e12393. [doi: 10.1111/spc3.12393]
36. Wang T, Giunti G, Melles M, Goossens R. Digital patient experience: umbrella systematic review. J Med Internet Res 2022;24(8):e37952 [FREE Full text] [doi: 10.2196/37952] [Medline: 35925651]

Abbreviations

DPEP: digital patient engagement platform
PX: patient experience

UMUX: Usability Metric for User Experience
UX: user experience

Edited by P Santana-Mancilla; submitted 16.10.23; peer-reviewed by F Wang, H Shen; comments to author 29.12.23; revised version received 21.02.24; accepted 28.03.24; published 14.05.24.

Please cite as:
Aspelund A, Valkonen P, Viitanen J, Rauta V
Designing for Improved Patient Experiences in Home Dialysis: Usability and User Experience Findings From User-Based Evaluation Study With Patients With Chronic Conditions
JMIR Hum Factors 2024;11:e53691
URL: https://humanfactors.jmir.org/2024/1/e53691
doi:10.2196/53691
PMID:38743476

Original Paper

Enabling Health Information Recommendation Using Crowdsourced Refinement in Web-Based Health Information Applications: User-Centered Design Approach and EndoZone Informatics Case Study

Wenhao Li[1,2], PhD; Rebecca O'Hara[3], PhD; M Louise Hull[3], PhD, CREI; Helen Slater[4], PhD; Diksha Sirohi[3], MPH, MBA; Melissa A Parker[5], MPhil, RN, RM; Niranjan Bidargaddi[1], PhD

[1]College of Medicine and Public Health, Flinders University, Clovelly Park, Australia
[2]School of Information Science and Engineering, Shandong Normal University, Jinan, China
[3]Robinson Research Institute, Adelaide Medical School, University of Adelaide, Adelaide, Australia
[4]Curtin School of Allied Health, Curtin University, Perth, Australia
[5]Canberra Endometriosis Centre, Centenary Hospital for Women and Children, Canberra, Australia

Corresponding Author:
Niranjan Bidargaddi, PhD
College of Medicine and Public Health
Flinders University
1284 South Road
Clovelly Park, 5042
Australia
Phone: 61 423416543
Email: niranjan.bidargaddi@flinders.edu.au

Abstract

Background: In the digital age, search engines and social media platforms are primary sources for health information, yet their commercial interests–focused algorithms often prioritize irrelevant content. Web-based health applications by reputable sources offer a solution to circumvent these biased algorithms. Despite this advantage, there remains a significant gap in research on the effective integration of content-ranking algorithms within these specialized health applications to ensure the delivery of personalized and relevant health information.

Objective: This study introduces a generic methodology designed to facilitate the development and implementation of health information recommendation features within web-based health applications.

Methods: We detail our proposed methodology, covering conceptual foundation and practical considerations through the stages of design, development, operation, review, and optimization in the software development life cycle. Using a case study, we demonstrate the practical application of the proposed methodology through the implementation of recommendation functionalities in the EndoZone platform, a platform dedicated to providing targeted health information on endometriosis.

Results: Application of the proposed methodology in the EndoZone platform led to the creation of a tailored health information recommendation system known as EndoZone Informatics. Feedback from EndoZone stakeholders as well as insights from the implementation process validate the methodology's utility in enabling advanced recommendation features in health information applications. Preliminary assessments indicate that the system successfully delivers personalized content, adeptly incorporates user feedback, and exhibits considerable flexibility in adjusting its recommendation logic. While certain project-specific design flaws were not caught in the initial stages, these issues were subsequently identified and rectified in the review and optimization stages.

Conclusions: We propose a generic methodology to guide the design and implementation of health information recommendation functionality within web-based health information applications. By harnessing user characteristics and feedback for content ranking, this methodology enables the creation of personalized recommendations that align with individual user needs within trusted health applications. The successful application of our methodology in the development of EndoZone Informatics marks a significant progress toward personalized health information delivery at scale, tailored to the specific needs of users.

(JMIR Hum Factors 2024;11:e52027) doi:10.2196/52027

KEYWORDS

information recommendation; crowdsourcing; health informatics; digital health; endometriosis

Introduction

Background

Members of the general public predominantly resort to search engines such as Google or social media platforms such as Facebook, YouTube, and TikTok as their initial source of health information [1-7]. These platforms use intricate recommendation algorithms to curate the information made accessible to users [8]. The algorithms are designed to rank information based on certain criteria, presenting it in the order of the ranking score. However, the underlying architecture of these ranking systems is by default crafted with commercial intent as opposed to health-centered intent. As a result, information that entices interactions that lead to increased revenue, such as more time spent on the platform or increased traffic to advertisements, gets ranked more prominently. Meanwhile, the information that accurately reflects people's medical needs is buried under large amounts of unrelated articles and posts and becomes difficult to find [9,10].

As the preference for web-based information seeking continues to grow, the development of web-based health information applications by trusted sources has become increasingly popular [11,12]. Through these applications (eg, websites or mobile apps) [13], people can readily access a wealth of health information generated by trusted sources. These interactions present an opportunity to shape an alternative ranking architecture for recommending web-based health content, one that is grounded in health outcomes. The information curated by these trusted platforms is considered superior in quality. Using user behavior after content access to rank health information could pave the way for more effective algorithms. This improved method could be integrated into search engine and social media algorithms through regulatory measures, challenging the current prioritization of web-based health content.

The existing body of research lacks comprehensive guidance on integrating content-ranking algorithms into applications centered around health information delivery. In this paper, we outline a generic methodology to guide the design and implementation of health information recommendation functionality within web-based health information applications. In this methodology, the health information recommendation interface and logic are co-designed with medical experts and application users such as patients and their supporters. This ensures the credibility of the health information provided, as well as accurate reflection of users' preference when interacting with the application. The health information recommended to users is ranked and presented using crowdsourcing technology based on feedback from users who have similar demographic and medical profiles. This ensures that health information can be delivered to people according to their situations and needs. The methodology can be easily integrated into new or existing health information applications. By implementing this ranked health information recommendation feature, we foresee improvements in user experience (UX) and the relevance of health information provided.

This methodology for enabling health information recommendation was first formulated based on our experience and expertise in informatics system development and implementation. It was then further refined and validated through the process of designing and implementing the informatics features of a medical information platform named EndoZone [14]. The platform is funded by the Australian government and Jean Hailes for Women's Health and provides evidence-based information to address symptoms and strategies for managing endometriosis. We illustrate the applicability of the methodology through its application in the EndoZone platform to enable its tailored health information recommendation system known as EndoZone Informatics. The implementation process shows that the methodology is practical for enabling information recommendation functionalities for web-based health information applications that have targeted health content–sharing requirements. Early data show that the solution built using this methodology is effective in reflecting users' feedback and providing highly personalized information recommendations and is also highly flexible in adjusting information recommendation logic. It has also been observed that the design of the user engagement process and user interface (UI) is highly relevant to the rate of users providing feedback and hence can affect the outcome of an information recommendation solution significantly.

The aim of this paper was to outline a generic methodology to guide the design and implementation of health information recommendation functionality within web-based health information applications and demonstrate its application in designing and implementing the informatics features of the EndoZone health information platform.

Related Research

A substantial amount of the articles and videos recommended by search engines and social media platforms have quality issues. They may contain biased content, are not comprehensive enough to cover the topic, are not evidence based, and provide limited coverage or content irrelevant to the topic [5,6,9,10,15]. A review by Osman et al [6] highlighted that >40% of the videos on YouTube on lumbar discectomy, cardiopulmonary resuscitation, and stroke are not useful, while more than half of the videos about vaccination as well as phototherapy and excimer laser treatment for psoriasis reflect bias due to commercial interests. A study assessing the quality of diabetes-related content on TikTok found that the quality of the content varies significantly depending on the types of creators and does not fully meet the health information needs of patients [5]. From billions of web pages and videos on the internet, commercial recommendation algorithms of search engines and social media platforms show those with the highest rank first, where the ranking criteria often have nothing to do with whether the content could meet people's medical needs [16]. To obtain a higher rank, which leads to a higher visibility rate and eventually a better commercial outcome, billions of dollars have been invested by companies for search engine optimization [8]. This compounds the situation because trusted health information sources such as research organizations and noncommercial health organizations often do not have the financial capacity to compete with commercial companies. As a result, the recommendations made by search engines and social media platforms lead people to unrelated articles, commercial advertisements, or even misinformation. As people generally lack the skills and experience to evaluate the accuracy of the information they are recommended [17], incorrect and harmful medical decisions could be made.

In comparison, web-based health information applications developed by trusted sources such as governments, credited health organizations, and universities provide health information with criteria that people value, such as trustworthiness, expertise, and objectivity [18]. In recent years, many of these applications have been developed globally to bypass the information recommendation algorithms of search engines and social media platforms [11,12,19,20]. Several applications contain mechanisms that provide personalized recommendations of nutritional information, medications, treatment plans, diagnoses or disease predictions, physical activities, or other health care services, based on users' profiles and inputs [21]. However, these recommendation features have not yet been applied extensively in health informatics and medical scenarios [22] and are typically created on an app-to-app basis, targeting a specific disease or recommendation context [12].

The lack of effective information recommendation functionality can be eliminated by enabling health information recommendation capability at scale. Many web-based health information applications could apply similar methodologies in design, development, and evaluation in terms of health information recommendation functionality due to their similarities in context, purpose, and category of recommended items. Tran et al [12] summarized 4 basic recommendation techniques: collaborative filtering, content-based filtering, knowledge-based recommendation, and a hybrid recommendation that combines these 3 techniques. In terms of evaluating the recommendation quality and the effectiveness of the recommendation mechanism, users' feedback is considered to be a major quality criterion [23]. Crowdsourcing technology has been applied in health care and has proven to be an effective approach to collecting retrospective data, such as user feedback, from a large number of dispersed participants [24]. With the development of health informatics technology and current trends of population preferences toward seeking information on the web, the use of crowdsourcing technologies for validating the effectiveness of health information recommendations is promising.

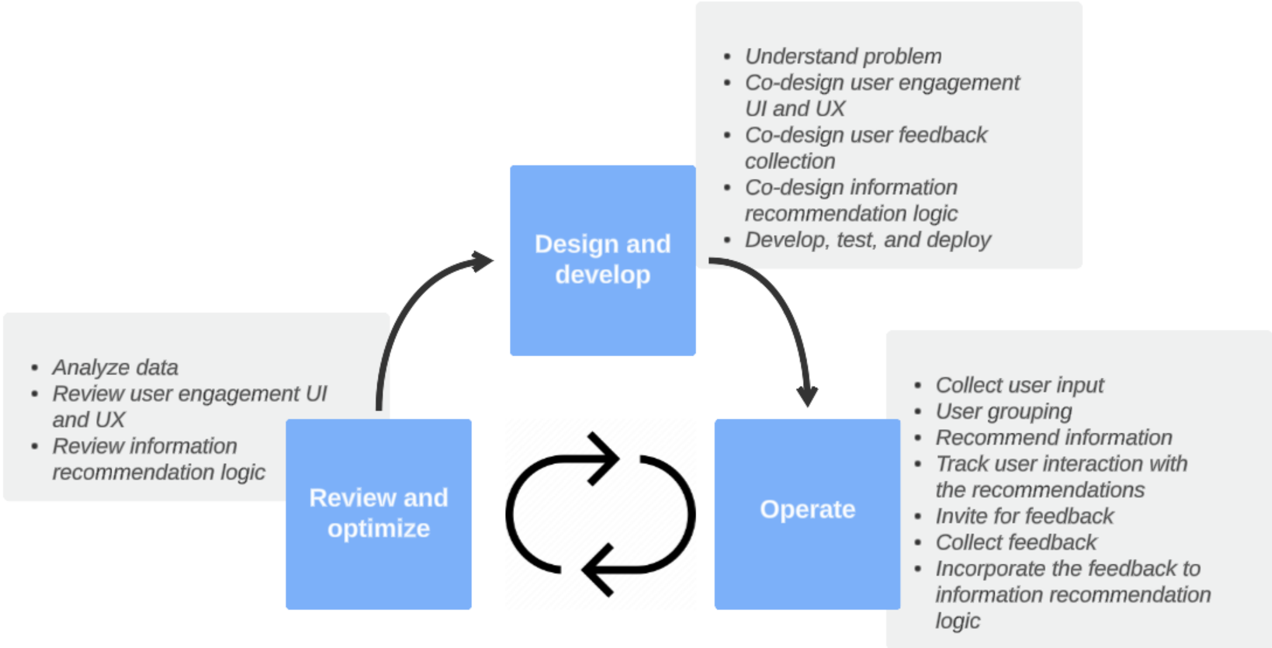
Methods

Overview

In this section, we present the concept of the methodology as well as implementation-related design, including software components, software development, and maintenance considerations, during 2 different implementation phases.

The methodology for enabling health information recommendation functionalities involves medical experts, researchers or data analysts, software developers, designers, and users of the web-based health information application. As shown in Figure 1, the methodology consists of 3 stages: *design and develop*, *operate*, and *review andoptimize*. At a high level, the methodology can be summarized thus: first, co-design and codevelop the information recommendation solution; second, recommend information to, and collect feedback from, users to improve the recommendation logic; and third and last, periodically review the statistical data to identify issues and continually adjust the solution.

Figure 1. The information recommendation solution life cycle. UI: user interface; UX: user experience.



Design and Develop

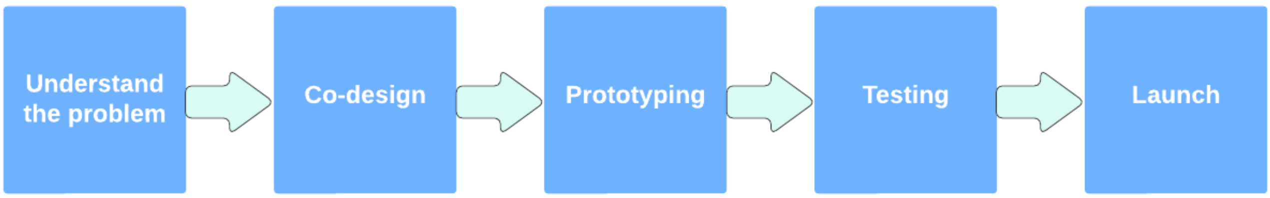
Overview

The implementation of the health information recommendation functionalities in the web-based health information application starts with the design and development of a solution that meets the specific requirements of the application. The design and development process adopts a human-centered design thinking model [25,26]. It considers the need of users to be the main factor that drives the design decision-making process. Figure 2 shows the design and development process for the methodology.

The solution designer first needs to understand the problem and develop detailed requirements for the information recommendation functionalities. To engage in the cocreation process, users can be invited to participate in observation sessions or interview sessions to explore the key challenges and their needs. During these sessions, the questions to be answered could include *What problem is the application trying to solve? What is the status of the application (launched/unbuilt)? What information will be recommended? How is the information expected to be recommended? What data are available for the recommendation logic to be based on?* and *If the application already exists, what does it look like and how is the information recommendation component expected to be integrated?*

Once all requirements have been clearly defined, a series of co-design or ideate sessions are carried out by the solution designer and the medical experts who have comprehensive domain knowledge about the condition or disease the application covers. The co-design process aims to deliver several outcomes, as outlined in the following subsections.

Figure 2. The design and development process.



The Engagement Process

This is also called UI or UX, that is, the UI and the process that the user interoperates with the application's information recommendation functionalities. It will need to be co-designed with medical experts to fully consider users' medical needs.

The Feedback Collection Method

This includes the feedback questions to be asked and the form of questions to be delivered, and it needs to be co-designed with the medical experts as well as the researchers or data analysts, making sure that good UXs and the data collected can properly serve the purpose of the feedback collection.

The Information Recommendation Logic

This specifies how health information that is recommended to users can be realized via different data structures and algorithms. On the basis of the EndoZone Informatics example that we will present later, the recommendation logic could include things such as a list of expert-verified information, a set of rules for information recommendation, an algorithm for user grouping, an algorithm for feedback analysis, an algorithm for feedback incorporation, and an algorithm for information recommendation. In the co-design process for this deliverable, medical experts should be closely involved in the design of all included components, providing insights that are as detailed as possible and making sure that the recommendations are appropriate (ie, evidence based) and meet users' medical needs. Specifically, the list of expert-verified information and the set of rules for information recommendation should be based on medical experts' input and available research data. Taking a rule in EndoZone as an example, a recommendation of yoga as a self-management strategy is made for a user who has severe pelvic pain and does not experience heavy bleeding during menstruation. In addition to medical experts, researchers or data analysts should also participate in the co-design process, making sure that the algorithms are correctly designed.

The Develop, Test, and Deploy Processes

To ensure that the design fully reflects the users' needs while fully considering the complexity and professionalism of the design activities, a smaller group of user representatives can be invited for consultation, where staged co-design outcomes, as mentioned previously, are sent for review and feedback.

After the co-design process is completed, the solution designer translates the outcomes into system design and architecture specifications, which are then used by software developers to develop, test, and deploy the system. How the develop, test, and deploy processes are carried out depends on the preference of the software development team, where no restrictions are imposed by the methodology. However, it is necessary for stakeholders, including researchers, medical experts, and users, to participate in testing early versions of the solution and provide feedback, where design issues and recommendation logic issues can be identified and resolved in time. The tests can be carried out differently by different stakeholder groups; for example, medical experts and researchers can be asked to test specific features that are closely related to their expertise, while for users of the application, a series of tasks that match their needs and expectations (provided in the initial requirement collection or understanding sessions) can be preset, making sure that their feedback is relevant and targeted.

This *design and develop* stage may be conducted multiple times throughout the lifespan of an information recommendation solution in which the solution is updated to fix issues that are identified and rectified or to incorporate new features.

Operate

Overview

The methodology shifts to the *operate* stage once the information recommendation functionalities are launched. In this stage, the solution performs activities such as recommending health information to users, collecting user feedback on recommendations, and incorporating the feedback into the recommendation logic. The first entry point of the users to the solution should be an event that is related to the content of the application; for example, it can be a click on a button on a web page, an action when using a digital tool, or a click on a link included in an invitation email. Such an event triggers a series of activities to generate a list of recommendations to the user, as outlined in the following subsections.

Collect User Data on Entry

Data stored in the application, such as user account profiles, user input in digital tools, and user browsing history, contain information about the unique circumstance of a user that is needed for personalizing recommendations. When the entry-point event happens, such data are collected for subsequent algorithms.

Group Users

This is an essential step for recommending personalized information to users with different conditions. In this step, the users are grouped by the algorithm for user grouping, based on a set of predefined attributes. Members of a group could have similar demographic and medical profiles, such as condition, age, educational background, symptoms, treatments, and so on.

Recommend Information

Using the data collected on entry as input, an algorithm generates a list of recommended information according to the information recommendation logic. In the algorithm, first, user data collected on entry are checked against the rules for information recommendation; for example, if a user *U* has symptom *S*, and the rule *R* indicates that all users with symptom *S* will be recommended information *I*, then *I* will be recommended to user *U*. Second, recommendations that match the rules will be ranked according to previous feedback from all users in the same user group. Third and last, the information is shown to the user in the order of the rank, where information with the best feedback (eg, the highest positive feedback rate) is presented first and has a better chance to be viewed.

Track User Interaction

After the information has been recommended, users are likely to read not all but a subset of the recommendations. It is necessary to track which recommendations are read by a user so that in the later step of collecting user feedback to recommendations, questions can be asked effectively. It is assumed that a recommendation has been read by the user if the content has been exposed to the user (eg, the user clicks on a link to an article). Therefore, any interaction that indicates exposure of the information to the user is recorded. Depending on the UI or UX design, recorded interactions could include clicks on a recommendation link or button, the opening of the web page of the recommendation, and so on.

The evaluation of whether the recommendations meet users' medical needs relies on feedback from the users themselves, supported by the power of crowdsourcing. After a certain period of making the recommendations, attempts are made to collect summative feedback from users who may have read the recommendations and potentially carried out practical activities based on the recommendations.

Invite for Feedback

The collection of feedback starts with sending an invitation to the user for participation. If the user has accessed any of the recommendations, an invitation for feedback is sent. Invitations can be sent in the form of oral invitations (eg, telephone invitations or opportunistic face-to-face invitations) or written letter invitations (by post or via email), which will vary from case to case [27]. The method of sending invitations is determined according to medical experts' suggestions to approach users with specific medical conditions appropriately and maximize the response rate.

Collect Feedback

There are a few ways in which user feedback data can be collected on the web (eg, conducting web-based surveys and allowing user ratings) [28]. Conducting a web-based survey is one of the most popular ways to collect user feedback, is easy to implement, and can meet the requirements of a web-based health information application in many cases. Questions in the survey can be asked from a UX perspective in terms of the helpfulness of the recommendations; for example, questions against a therapy recommendation could include *Did you try this therapy? Did you find the therapy easy to do? How difficult did you find fitting this therapy into your life with your other activities?* and *Did you find this therapy helped in managing your symptoms?* One of the known issues of web-based surveys is the low completion rate [29]. Some strategies to incentivize completion rates can be found in existing studies [30,31].

Incorporate

After a user's feedback is collected and digitized, first, an algorithm for analyzing feedback executes to convert the feedback data into measurable attributes. Second, an algorithm for feedback incorporation deploys these attributes into the information recommendation logic. Depending on the design of the algorithm for feedback incorporation, the outcomes of the incorporation could include an updated set of rules for information recommendation, updated ranks of recommendations, updated descriptions for each recommendation, and so on. After the incorporation process finishes, the user's journey with the information recommendation solution is completed. The updated information recommendation logic will then be applied when other users engage with the information recommendation solution.

Review and Optimize

Overview

As the information recommendation solution operates, system operation data and user engagement data accumulate. Besides using the user interaction data for improving the information recommendation logic in the *operate* stage, an in-depth review and optimization of the solution can be conducted. The purpose is to identify issues based on the analytical outcome of the accumulated operation data set and the experience gained from the continuous operation and maintenance of the solution. Whether the *review andoptimize* stage needs to be carried out depends on several factors, such as the amount of analyzable data accumulated, the urgency of major optimization of the solution to address emerging requirements, and the operation status of the current information recommendation solution. Researchers and software developers need to decide when a formal *review andoptimize* stage is needed. The outcome of the *review andoptimize* stage should include an optimization plan, where detailed redesign and development can be carried out in the following *design and develop* stage.

Analyze Data

User engagement data such as user profile data, data of user interaction with the information recommendation solution, and user feedback as well as system operation data such as operation logs and web-based traffic data are accumulated and of statistical value to the optimization of the information recommendation solution. Depending on the sufficiency of the accumulated data, research questions such as *Do the users engage well with the solution? Is the recommendation solution effective in helping users to find the information that meet their medical needs?* and *Are the recommendations appropriate and suitable for the user to practice?* can be answered, and potential issues in the engagement process and recommendation content can be identified.

Review User Engagement UI and UX

At the *review andoptimize* stage, a retrospective review can be conducted toward the user engagement UI and UX. The review can be based on 2 sources of input: first, it can be based on the researchers' experience gained while continuously operating the user engagement UI and UX. Second, it can be based on user feedback, such as volunteer user group feedback when asked to test and promote the solution. This review could identify design issues in the UI and user engagement process that cause difficulty for users in accessing the features of the solution and the health information they are recommended.

Review Information Recommendation Logic

It is difficult to provide the best configuration to elements of the information recommendation logic and achieve the optimal recommendation outcome during the *design and develop* stage. The reasons include users' composition, uncertainty in user interaction patterns with the application, and a lack of analyzable data. Thus, continuous adjustments to the configuration of algorithms and data structures are needed; for example, grouping attributes and the logic of the algorithm for user grouping, rules for information recommendation, the list of expert-verified information, and the logic for user feedback evaluation can all be fine-tuned to reflect issues identified from the data analysis. In the *review andoptimize* stage, the best configuration for the information recommendation logic should be determined based on testing different configurations. It is most practical to conduct tests on different parts of the information recommendation in parallel with tests in the operate stage of the software to minimize impact to the existing system. To achieve this, a staging infrastructure can be set up, where a mirror copy of the solution can be created for test-related activities.

Implementation Considerations

Overview

This methodology can be adopted for the implementation of health information recommendation functionalities, either with already launched applications or when the application is still under development. In the next 2 subsections, we present implementation-related considerations of the methodology: first, components need to be developed and how these interoperate with other application components is described; and second, a 2-phased implementation strategy that aims to provide the optimal UX is described.

Software Components

The architecture design for the information recommendation solutions could vary vastly due to factors such as user requirements, the software technology stack being applied, the skill sets of developers, and governance restrictions (eg, the General Data Protection Regulation applicable in the European Union). However, when adopting the methodology, logical components for the health information recommendation functionality should be consistent. Figure 3 shows a high-level software component diagram that implements health information recommendation functionalities in a web-based health information application. The diagram consists of 3 sections: components of a typical web-based health information application (Figure 3A), backend components of the information recommendation solution (Figure 3B), and front-end components of the information recommendation solution (Figure 3C).

Figure 3. Software component design. (A) Components of a typical web-based health information application. (B) Backend components of the information recommendation solution. (C) Front-end components of the information recommendation solution.

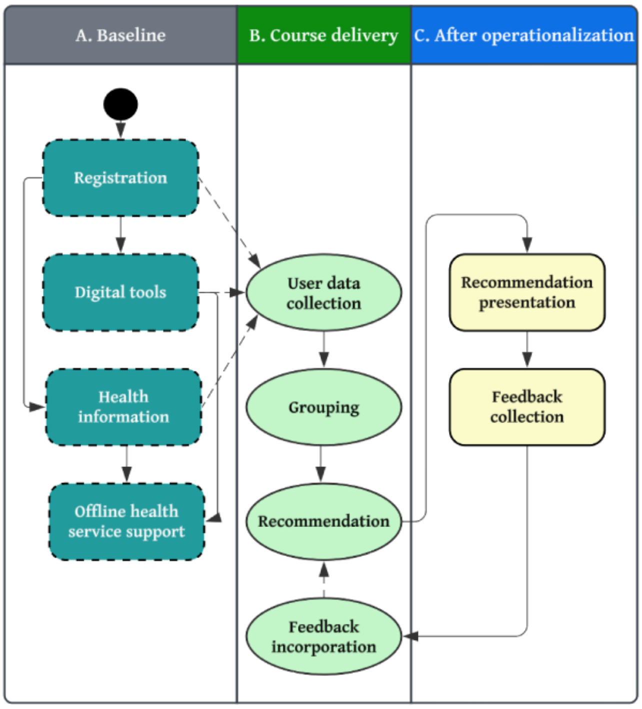


Figure 3A shows that a typical web-based health information application in the form of a website could include web pages of health information; instructions on local offline support; and, optionally, digital tools for certain informatic purposes and a registration component that is often needed by the digital tools. The 4 backend components shown in Figure 3B are needed for an information recommendation, which collects and processes all user input data, makes recommendations, and incorporates user feedback. Depending on the design of the information recommendation logic, the user input data collected could include user account data, health information browsing history, digital tool input, and user feedback on recommendations. The output of the 4 backend components includes updated information recommendation logic and a list of recommendations ranked based on user feedback. The front-end of the information recommendation solution shown in Figure 3C comprises 2 components: one for presenting recommendations and the other for collecting recommendation feedback. Depending on the design of the user engagement process, these 2 components can be either allocated on dedicated web pages or integrated into the web pages of any web-based health information application component.

The software components are designed in a loosely coupled fashion, where all functions and algorithms are independently maintained. Such a design pattern makes the adjustment of the information recommendation logic possible, from fine-tuning to a total replacement of the recommendation model. This feature is critical to a phased implementation of the solution, as will be discussed in the next subsection. One additional advantage of such a design is that it enables the potential of the web-based health information application to become a test bed of information recommendation algorithms, where algorithms can be easily alternated to test performance.

Implementation Phases

When the information recommendation solution is first launched, the number of users is small, and feedback on recommendations is not yet provided. Here, an information recommendation model that relies heavily on crowdsourced data for recommendation evaluation could produce suboptimal recommendations, impacting the UXs with the web-based health information application. To ensure the quality of recommendations before crowdsourced data are sufficiently accumulated, an implementation strategy is applied with the following two phases: (1) an initialization phase, in which crowdsourced data are not yet sufficiently collected, and an initial version of the algorithm for recommending information is used, where the information recommendation logic does not rank the recommendations based on user feedback data; and (2) an execution phase, where crowdsourced data are sufficient, and an execution version of the algorithm for recommending information is used, where user feedback is incorporated into the information recommendation logic for ranking recommendations based on user feedback data.

The main difference between these 2 versions of the algorithms is their logic in dealing with user feedback. Specifically, in the initial version of the algorithm, the list of recommendations is generated purely based on medical experts' input (ie, a set of predefined rules for information recommendation), whereas in the execution version, the list of recommendations is generated based on medical experts' input and further ranked based on user feedback data. Due to the loosely coupled software design, the algorithm for recommending information can be easily replaced. Researchers and data analysts can decide it is time to replace the algorithm when the amount of user feedback data is sufficient for the execution version of the algorithm to execute effectively.

Ethical Considerations

The development of the platform and analysis of EndoZone data was approved by the University of Adelaide Human Research Ethics Committee (H-2020-013 & H-2023-054). Informed consent was obtained from community members participating in the design and development phase of the EndoZone informatics platform, and all users accessing the tool after it was launched online. The extraction and analysis of de-identified EndoZone platform data for this study was in accordance with the guidelines approved by the ethics committee.

Results

Case Study: EndoZone Informatics

The methodology for enabling health information recommendation functionalities has been successfully applied in the development of the information recommendation functionalities of a co-designed endometriosis information platform called EndoZone [14]. Endometriosis is a chronic condition, where tissue similar to the lining of the uterus develops in places outside the uterus. Symptoms of endometriosis may include pain with menstruation, chronic pelvic pain, fatigue, and subfertility. Globally, it is estimated that endometriosis affects approximately 190 million women and people presumed female at birth [32]. To address the wide-ranging impact of endometriosis, the Australian government and Jean Hailes for Women's Health funded the development of EndoZone to improve knowledge, address symptoms, and provide strategies for managing endometriosis. This platform was designed for people affected by the condition as well as their supporters, such as parents, partners, teachers, and coworkers. The platform was cocreated and developed using the design thinking framework. During the cocreation process of the EndoZone platform, endometriosis community focus groups (n=36) were held to explore the key challenges and needs of the endometriosis community; in addition, a community priorities survey was conducted with 347 community member responses. On the basis of the key priorities identified, it was decided that functionalities would be developed to facilitate interaction and to support people experiencing endometriosis symptoms through the recommendation of strategies based on their symptoms, that is, EndoZone Informatics. The design, development, and implementation of EndoZone Informatics strictly follows the health information recommendation methodology. The solution was co-designed with other components of the EndoZone platform and integrated into the platform in April 2023. The solution is currently fully implemented and operating in the execution phase. In the following subsections, we present the design, development, and implementation of EndoZone Informatics to showcase the practicality of adopting the methodology for the design and implementation of information recommendation functionalities in a web-based health information application.

Design and Develop

The design of EndoZone Informatics was part of the broader platform development process, which follows the broader co-design process of EndoZone. It was designed in consultation with 5 community representatives from endometriosis associations (patients, advocates, and supporters), clinicians (endometriosis or fertility specialist, physiotherapist or pain researcher, and endometriosis nurse), researchers, and 2 health informatics specialists. This involved a series of workshops and meetings to discuss details of the user engagement process as well as a smaller working group with clinicians to develop the initial information recommendation logic. The design was mocked up in consultation with the UI or UX designer and then integrated into the EndoZone platform. The outcome of this co-design process includes the user engagement process and the corresponding UI or UX prototype, a feedback collection method using email invitations and web-based surveys, and the information recommendation logic. Specifically, the information recommendation logic includes a list of 16 expert-verified articles for different endometriosis self-management therapies; a set of 27 rules that match symptoms to the recommendation of therapies (eg, one rule is that if the user experiences severe menstrual cramps, an article on transcutaneous electrical nerve stimulation therapy will be recommended); and algorithms for user grouping, analyzing feedback, feedback incorporation, and recommending information. After the design was ready to be reviewed, a review meeting was carried out for all stakeholders, where feedback on the design outcome was collected for adjustment. After the design outcome was adjusted and agreed upon, the solution designer and the UI or UX designer translated the outcomes to formal UI or UX design and architecture specifications for the development work to be carried out.

The develop, test, and deploy process was carried out using an agile approach, more specifically, the Scrum development process [33], which is preferred by the development team due to the existing software technology stack and developer skill sets. The development progress was regularly reported to, and closely monitored by, the digital health solution transformation experts. One issue that was encountered during this stage was some previously unforeseeable dependencies of the informatics components on several other components of the platform, which caused a 3-month delay in the release date of EndoZone Informatics. However, the development process is in general smooth.

To test EndoZone Informatics, medical and health experts participated in 2 demonstrations of the platform and tested the ready-to-launch version. In all, 9 community users participated in testing the platform's early versions through a beta version with restricted access. Specifically, the user test was conducted after a series of tasks in which user testers were audio and video recorded while they completed the tasks and provided verbal feedback as they were using the platform. They also completed a series of questions related to their feedback on the platform (eg, what they liked, what they did not like, and suggestions for improvement), the usability of the platform, and whether they would recommend the platform to a friend or colleague. The feedback obtained from the user test was then incorporated into the further development of EndoZone Informatics features.

Operate

A user starts to engage with the EndoZone information recommendation solution from the submission of a health questionnaire named *My Endo Report*. The questionnaire contains a series of questions related to self-reported endometriosis symptoms and treatments that have been tried to manage symptoms, as well as a brief medical history. After the user has submitted the questionnaire, the backend algorithms are triggered to produce a list of recommended self-management therapies, where the recommendations are presented as part of *My Endo Report* (Figure 4). In EndoZone, the crowdsourced input (ie, user-provided feedback on recommended therapies) is used to determine the order of the recommendations being presented: among users with similar symptoms, therapies that are rated as "more useful" are given a higher rank and shown first in the list of recommendations. The description text of each recommendation contains a ranking to highlight this order. If the user clicks on a recommended therapy, the solution assumes that the user has viewed the content and records the click event. Next, 30 days after the recommendations are made, an email is sent to the user, inviting the user to complete a follow-up survey regarding the recommendations (Figure 5). When the user accepts the invitation, a follow-up survey is generated, containing questions related only to the recommendations that the user has clicked on. For each recommended self-management strategy, the survey contains 10 questions. It asks the user about the usefulness of the strategy, including their feelings after practicing the strategy, the practicality of the strategy, the effectiveness of the strategy in improving their symptoms, and so on. Figure 6 shows an example follow-up survey for the recommendation of pelvic health physiotherapy. Once the user has submitted the follow-up survey, their engagement with the EndoZone information recommendation solution is complete.

Figure 4. Recommended self-management therapies.

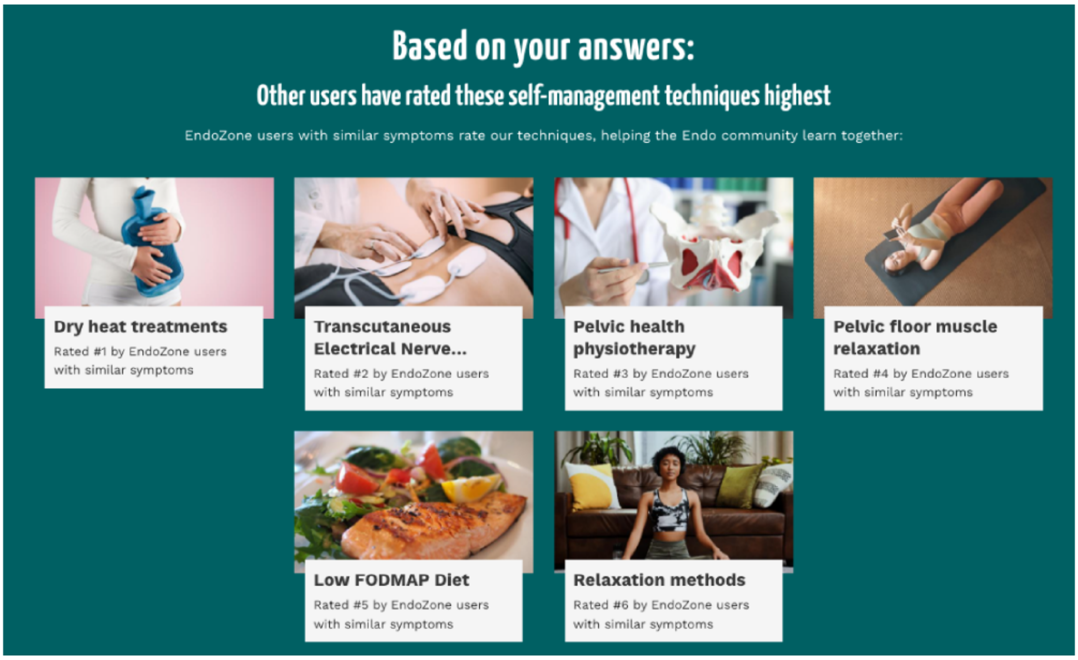
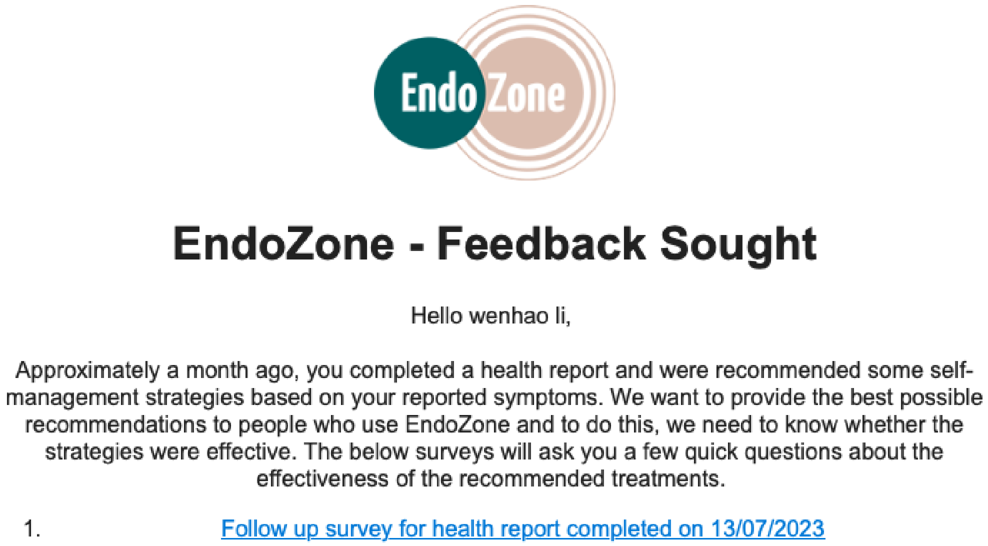


Figure 5. Invitation email to a follow-up survey.



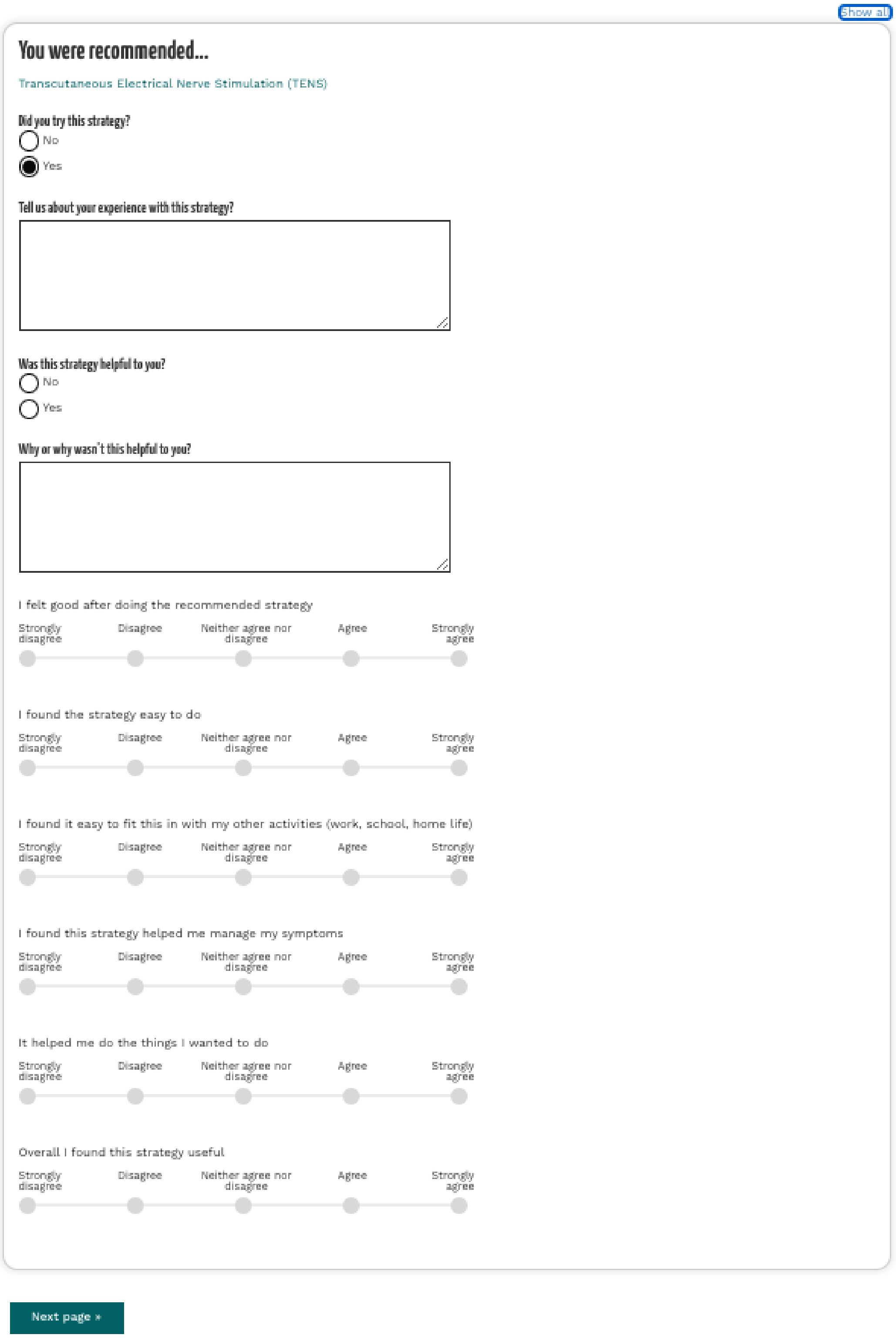
EndoZone - Feedback Sought

Hello wenhao li,

Approximately a month ago, you completed a health report and were recommended some self-management strategies based on your reported symptoms. We want to provide the best possible recommendations to people who use EndoZone and to do this, we need to know whether the strategies were effective. The below surveys will ask you a few quick questions about the effectiveness of the recommended treatments.

1. Follow up survey for health report completed on 13/07/2023

Figure 6. A follow-up survey example.



Software Architecture

The architecture design in the EndoZone information recommendation solution strictly followed the component design shown in Figure 3 but was customized to fit the specific requirements of the application. First, based on the needs of the EndoZone information recommendation logic, the user data collection component only collects user registration data (ie, demographic profile data) and user input to the digital tool (ie, *My Endo Report* submission data). Second, the recommendation presentation component is integrated into the *My Endo Report* summary page of the application as part of the *My Endo Report* outcome.

Deployment wise, based on best practice, the EndoZone information recommendation solution is designed to be cloud based. It operates on cloud-based infrastructure using Amazon Web Services. All recommendation-related components are deployed in the form of microservices using Amazon Web

Services Lambda, where each microservice contains components that are needed for a single application programming interface call. Specifically, the user data collection, grouping, and recommendation components are deployed in 1 microservice. Once the *My Endo Report* questionnaire is submitted, this microservice is called and responds with a list of recommended self-management therapies. The feedback incorporation component is deployed in another microservice. Once the follow-up survey is submitted, this microservice is called to update the information recommendation logic.

Implementation

The implementation process of the EndoZone information recommendation solution followed the 2-phased process. Compared with what is described in the *Overview* subsection in the *Methods* section, an alternative data accumulation approach was conducted in the initialization phase to accelerate the transition to the execution phase. After the platform was launched, a targeted social media campaign on Instagram and Facebook was conducted to promote initial use of the platform. During the campaign, the initial version of the algorithm for recommending information was executed based on the expert-derived set of information recommendation rules that were matched to self-management therapies and symptoms that were indicated in *My Endo Report*. In completing the report, users are contacted via various channels to self-rate how helpful each self-management therapy or strategy was to manage their symptoms using a 3-point scale ("Didn't work," "Helped a bit," and "Helped a lot").

At the time of reviewing the data, the EndoZone platform had had 57,000 visitors (Google Analytics; February 20, 2024), predominantly from Australia (n=32,000, 56.14%), the United States (n=6000, 10.53%), the United Kingdom (n=5200, 9.12%), and New Zealand (n=5000, 8.77%), of whom 5756 (10.1%) completed *My Endo Report* and submitted it through the platform. User feedback data were aggregated to count the number of reports that indicated that a particular strategy either "Helped a bit" or "Helped a lot." This feedback was then considered to be the initial rating of therapies on which the execution version of the algorithm could rely; for example, *yoga* was rated by 682 people, of whom 404 (59.24%) rated it as either "Helped a bit" or "Helped a lot." These feedback data were then manually incorporated, where a rating of "404/682 (59.24%)" was set as the initial rating of the therapy *yoga* for all user groups. A further analysis of the data collected through the platform is being conducted to feed into the next iteration of EndoZone Informatics.

Review and Optimize

The EndoZone information recommendation solution was integrated into the EndoZone platform in April 2023. Tests and feedback from the volunteer group have shown that the overall user engagement process can be carried out well, with a good UX. Meanwhile, based on early data accumulated, several design issues have been identified; for example, the participation rate for providing feedback is lower than expected. We suspect that the UI or UX design could be the major cause for this outcome: first, in the current design, only registered users are invited to complete the follow-up survey (unregistered users cannot be invited because they are not asked for their email address). Currently, most users use the site anonymously, which means that most users of the platform who decided not to create an account in EndoZone are not able to experience the full recommendation functionalities and provide recommendation feedback. Second, the recommendation section is in a relatively inconspicuous position on the *My Endo Report* summary page. This may lead to reduced visibility and hence less user participation. The finding indicates that the design of web pages (UI or UX) is highly relevant to the effectiveness of the solution. It also indicates that the methodology is limited in identifying specific design defects during the initial *design and develop* stage. However, such defects can be addressed in the *review andoptimize* stage, where issues that crop up during the execution of the solution are reviewed. In the context of the EndoZone platform, remedial development work has been planned in the second phase of the project from 2024 to 2026.

Furthermore, the logic for tracking user engagement with recommended therapies (ie, once the article is opened, the recommended therapy is considered to have been read) is not consistent with the industrial standard that large IT companies have applied; for example, in Google Analytics, a user is considered to have engaged with a web page if they stay on the page for >10 seconds [34]. How user engagement is tracked is not defined by the methodology and could vary from case to case. However, in the context of the EndoZone platform, the solution logic does not cause a loss of user feedback data. The impact on the UX (ie, several more survey questions are asked regarding a therapy that the user has not practiced) is limited and can be eliminated by adjusting the questions in the follow-up survey.

Discussion

Outcome

In the previous sections, we have presented a methodology that enables health information recommendation functionalities in web-based health information applications. The concept of the methodology as well as the implementation considerations, including the software component design and the 2-phased implementation process, are described in detail, based on which information recommendation solutions can be created and operationalized. The methodology has been refined and validated through its application to create EndoZone Informatics, that is, the information recommendation solution of an endometriosis information platform named EndoZone. Early data from the execution of the EndoZone Informatics solution shows that using this methodology was effective in recommending medical expert–verified information while incorporating crowdsourced input from users with similar conditions. This methodology helped users to find the information that could be of most use to them. The loosely coupled software component design enabled high flexibility in adjusting the information recommendation model, which makes the 2-phased implementation process easy to carry out.

During the application of the methodology for EndoZone Informatics, we encountered several issues. To recap, first, the dependencies of the information recommendation components on other components of the web-based health information application caused a 3-month delay in the development progress of EndoZone Informatics. Second, the UI or UX design flaws, such as unregistered users not being able to experience the full recommendation functionalities and *underexposure* of the recommendation section in the *My Endo Report* summary page, have resulted in a lower-than-expected participation rate for providing feedback. These issues reveal a limitation of the methodology, that is, it is not able to address some specific software engineering problems. These issues also show the significance of the *review and optimize* stage, where design and development issues can be identified, and repair plans can be created.

In general, the application of the methodology for designing and implementing EndoZone Informatics is successful. It is a solid step toward enabling personalized information recommendation at scale. The solution indicates a promising approach where personalized health information recommendation can be enabled in all web-based health information applications. Compared with accessing health information via recommendations derived from commercial algorithms of search engines and social media platforms, a health information–access approach provides people with an alternative health information–ranking and –recommendation path, which ranks information based on people's medical needs; provides them with trustworthy, credible, and evidence-based recommendations; and aims for the best health outcomes.

Potential of the Methodology

The methodology is proposed to be applied in web-based health information applications targeting personalized health information recommendations for educational and knowledge-sharing purposes. As showcased by the EndoZone platform, this methodology is applicable and works well for web-based health information applications that share health information such as chronic disease self-management strategies. However, the practicality of applying the methodology in creating solutions for applications that target acute diseases is yet to be proven. Another area for further research is the practicality of applying this methodology for recommending clinical treatments. This requires systematic study of what the impact could be if the methodology was applied for recommending clinical treatments (eg, medication use). What kind of care decisions (safety or risk of harm or relative benefits) need to be considered? What are the ethical issues involved? Answers to these questions are not yet clear.

A promising area for applying the methodology concerns creating solutions for recommending other medical and health services, such as links to local medical experts, health services, advocacy organizations, and related web-based applications [35,36]. Exploring how solutions created by applying this methodology could help in connecting web-based services to local offline services to improve the quality and scope of user support would also be of value.

Another potential application of this methodology is to generate test beds for information recommendation algorithms and their suitability for different medical scenarios. As described in the *Overview* subsection of the *Methods* section, the software component design allows all key logic components to be independently maintained and easily replaced. This feature can be leveraged for new information recommendation algorithms or models to be tested; for example, by applying different information recommendation models and monitoring user interactions under each model, the performance of different information recommendation models can be analyzed.

Conclusions

This study introduces a novel methodology that enriches web-based health applications with personalized information recommendation capabilities. Tested through the development of the EndoZone platform, our approach successfully merges expert knowledge with user insights to provide targeted health information. While we encountered developmental and design challenges, these experiences highlighted the importance of adaptability and continuous refinement. The methodology's potential extends beyond the specific case of EndoZone, offering a scalable solution for tailoring health information across various authoritative health websites, with implications for improving patient education and engagement in a digital era.

Acknowledgments

The authors are grateful for the support and expertise provided by Karl Fradley and Laura Morgan from SRA Information Technology and the team at Brainium Labs, who contributed to the development of EndoZone, which has greatly facilitated the completion of this research. EndoZone was funded by the Australian Government and Jean Hailes for Women's Health.

Conflicts of Interest

None declared.

References

1. Bernard E, Arnould M, Saint-Lary O, Duhot D, Hebbrecht G. Internet use for information seeking in clinical practice: a cross-sectional survey among French general practitioners. Int J Med Inform 2012 Jul;81(7):493-499. [doi: 10.1016/j.ijmedinf.2012.02.001] [Medline: 22425281]
2. Armstrong-Heimsoth A, Johnson ML, McCulley A, Basinger M, Maki K, Davison D. Good googling: a consumer health literacy program empowering parents to find quality health information online. J Consum Health Internet 2017 Jun 12;21(2):111-124. [doi: 10.1080/15398285.2017.1308191]

3. Wong DK, Cheung MK. Online health information seeking and eHealth literacy among patients attending a primary care clinic in Hong Kong: a cross-sectional survey. J Med Internet Res 2019 Mar 27;21(3):e10831. [doi: 10.2196/10831] [Medline: 30916666]
4. Hassan S, Masoud O. Online health information seeking and health literacy among non-medical college students: gender differences. J Public Health (Berl.) 2020 Mar 09;29(6):1267-1273. [doi: 10.1007/s10389-020-01243-w]
5. Kong W, Song S, Zhao YC, Zhu Q, Sha L. TikTok as a health information source: assessment of the quality of information in diabetes-related videos. J Med Internet Res 2021 Sep 01;23(9):e30409. [doi: 10.2196/30409] [Medline: 34468327]
6. Osman W, Mohamed F, Elhassan M, Shoufan A. Is YouTube a reliable source of health-related information? a systematic review. BMC Med Educ 2022 May 19;22(1):382. [doi: 10.1186/s12909-022-03446-z] [Medline: 35590410]
7. Zhao Y, Zhang J. Consumer health information seeking in social media: a literature review. Health Info Libr J 2017 Dec;34(4):268-283. [doi: 10.1111/hir.12192] [Medline: 29045011]
8. Singh Rajawat A, Barhanpurkar K, Goyal SB, Bedi P, Shaw RN, Ghosh A. Efficient deep learning for reforming authentic content searching on big data. In: Proceedings of the Advanced Computing and Intelligent Technologies. 2021 Presented at: ICACIT 2021; March 20-21, 2021; New Delhi, India. [doi: 10.1007/978-981-16-2164-2_26]
9. Hirsch M, Aggarwal S, Barker C, Davis CJ, Duffy JM. Googling endometriosis: a systematic review of information available on the internet. Am J Obstet Gynecol 2017 May;216(5):451-8.e1. [doi: 10.1016/j.ajog.2016.11.1007] [Medline: 27840143]
10. Daraz L, Morrow AS, Ponce OJ, Beuschel B, Farah MH, Katabi A, et al. Can patients trust online health information? a meta-narrative systematic review addressing the quality of health information on the internet. J Gen Intern Med 2019 Sep;34(9):1884-1891. [doi: 10.1007/s11606-019-05109-0] [Medline: 31228051]
11. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis. J Med Internet Res 2013 Jun 14;15(6):e120. [doi: 10.2196/jmir.2600] [Medline: 23770578]
12. Tran TN, Felfernig A, Trattner C, Holzinger A. Recommender systems in the healthcare domain: state-of-the-art and research issues. J Intell Inf Syst 2020 Dec 17;57(1):171-201. [doi: 10.1007/s10844-020-00633-6]
13. Baumel A, Kane JM. Examining predictors of real-world user engagement with self-guided ehealth interventions: analysis of mobile apps and websites using a novel dataset. J Med Internet Res 2018 Dec 14;20(12):e11491. [doi: 10.2196/11491] [Medline: 30552077]
14. Welcome to EndoZone. EndoZone. URL: https://www.endozone.com.au [accessed 2023-06-07]
15. Cuan-Baltazar JY, Muñoz-Perez MJ, Robledo-Vega C, Pérez-Zepeda MF, Soto-Vega E. Misinformation of COVID-19 on the internet: infodemiology study. JMIR Public Health Surveill 2020 Apr 09;6(2):e18444. [doi: 10.2196/18444] [Medline: 32250960]
16. Schultheiß S, Lewandowski D. “Outside the industry, nobody knows what we do” SEO as seen by search engine optimizers and content providers. JD 2020 Dec 08;77(2):542-557. [doi: 10.1108/JD-07-2020-0127]
17. Paakkari L, Okan O. COVID-19: health literacy is an underestimated problem. Lancet Public Health 2020 May;5(5):e249-e250. [doi: 10.1016/S2468-2667(20)30086-4] [Medline: 32302535]
18. Sun Y, Zhang Y, Gwizdka J, Trace CB. Consumer evaluation of the quality of online health information: systematic literature review of relevant criteria and indicators. J Med Internet Res 2019 May 02;21(5):e12522. [doi: 10.2196/12522] [Medline: 31045507]
19. Zhao L, Bidargaddi N, Vakulin A, Li W, Luscombe-Marsh N, Benton F, et al. A micro-randomized pilot study to examine the impact of just-in-time nudging on after-dinner snacking in adults with type 2 diabetes: a study protocol. Diabetes Obes Metab 2023 Sep;25(9):2439-2446. [doi: 10.1111/dom.15159] [Medline: 37385960]
20. Zarocostas J. How to fight an infodemic. Lancet 2020 Feb 29;395(10225):676. [doi: 10.1016/S0140-6736(20)30461-X] [Medline: 32113495]
21. Wiesner M, Pfeifer D. Health recommender systems: concepts, requirements, technical basics and challenges. Int J Environ Res Public Health 2014 Mar 03;11(3):2580-2607. [doi: 10.3390/ijerph110302580] [Medline: 24595212]
22. Calero Valdez A, Ziefle M, Verbert K, Felfernig A, Holzinger A. Recommender systems for health informatics: state-of-the-art and future perspectives. In: Machine Learning for Health Informatics. Cham, Switzerland: Springer; 2016.
23. Pu P, Chen L, Hu R. A user-centric evaluation framework for recommender systems. In: Proceedings of the fifth ACM conference on Recommender systems. 2011 Presented at: RecSys '11; October 23-27, 2011; Chicago, IL. [doi: 10.1145/2043932.2043962]
24. McCartney P. Crowdsourcing in healthcare. MCN Am J Matern Child Nurs 2013;38(6):392. [doi: 10.1097/NMC.0b013e3182a41571] [Medline: 24145496]
25. Marko-Holguin M, Cordel SL, van Voorhees BW, Fogel J, Sykes E, Fitzgibbon M, et al. A two-way interactive text messaging application for low-income patients with chronic medical conditions: design-thinking development approach. JMIR Mhealth Uhealth 2019 May 01;7(5):e11833. [doi: 10.2196/11833] [Medline: 31042152]
26. Woods L, Cummings E, Duff J, Walker K. Design thinking for mHealth application co-design to support heart failure self-management. Stud Health Technol Inform 2017;241:97-102. [Medline: 28809190]

27. Bunten A, Porter L, Gold N, Bogle V. A systematic review of factors influencing NHS health check uptake: invitation methods, patient characteristics, and the impact of interventions. BMC Public Health 2020 Jan 21;20(1):93. [doi: 10.1186/s12889-019-7889-4] [Medline: 31964366]
28. Boylan AM, Williams V, Powell J. Online patient feedback: a scoping review and stakeholder consultation to guide health policy. J Health Serv Res Policy 2020 Apr;25(2):122-129. [doi: 10.1177/1355819619870837] [Medline: 31495226]
29. Liu M, Wronski L. Examining completion rates in web surveys via over 25,000 real-world surveys. Soc Sci Comput Rev 2017 Feb 23;36(1):116-124. [doi: 10.1177/0894439317695581]
30. Sammut R, Griscti O, Norman IJ. Strategies to improve response rates to web surveys: a literature review. Int J Nurs Stud 2021 Nov;123:104058. [doi: 10.1016/j.ijnurstu.2021.104058] [Medline: 34454334]
31. Booker QS, Austin JD, Balasubramanian BA. Survey strategies to increase participant response rates in primary care research studies. Fam Pract 2021 Sep 25;38(5):699-702. [doi: 10.1093/fampra/cmab070] [Medline: 34213547]
32. Horne AW, Missmer SA. Pathophysiology, diagnosis, and management of endometriosis. BMJ 2022 Nov 14;379:e070750. [doi: 10.1136/bmj-2022-070750] [Medline: 36375827]
33. Schwaber K. SCRUM development process. In: Sutherland J, Casanave C, Miller J, Patel P, Hollowell G, editors. Business Object Design and Implementation. London, UK: Springer; 1997.
34. What's a good engagement rate in Google Analytics 4? First Page Sage. 2022 Nov 4. URL: https://firstpagesage.com/reports/whats-a-good-engagement-rate-fc/ [accessed 2023-02-03]
35. Hors-Fraile S, Rivera-Romero O, Schneider F, Fernandez-Luque L, Luna-Perejon F, Civit-Balcells A, et al. Analyzing recommender systems for health promotion using a multidisciplinary taxonomy: a scoping review. Int J Med Inform 2018 Jun;114:143-155. [doi: 10.1016/j.ijmedinf.2017.12.018] [Medline: 29331276]
36. Ismail N, Kasim S, Yah Jusoh Y, Hassan R, Alyani A. Medical appointment application. Acta Electron Malays 2017 Nov 01;1(2):05-09. [doi: 10.26480/aem.02.2017.05.09]

Abbreviations

UI: user interface
UX: user experience

Edited by B Price; submitted 21.08.23; peer-reviewed by Y Yang, B Crocker, M Herron; comments to author 12.01.24; revised version received 03.03.24; accepted 31.03.24; published 29.05.24.

Please cite as:
Li W, O'Hara R, Hull ML, Slater H, Sirohi D, Parker MA, Bidargaddi N
Enabling Health Information Recommendation Using Crowdsourced Refinement in Web-Based Health Information Applications: User-Centered Design Approach and EndoZone Informatics Case Study
JMIR Hum Factors 2024;11:e52027
URL: https://humanfactors.jmir.org/2024/1/e52027
doi:10.2196/52027
PMID:38809588

Original Paper

Participatory Development of an Integrated, eHealth-Supported, Educational Care Pathway (Diabetes Box) for People With Type 2 Diabetes: Development and Usability Study

Daan Leonhard de Frel[1], MD; Mariëlle A Schroijen[2], MD, PhD; Jiska J Aardoom[3,4], PhD; Wesley van Gils[2], RD, MSc; Sasja D Huisman[2], MSc, PhD; Veronica R Janssen[1], MSc, PhD; Anke Versluis[3,4], MSc, PhD; Maaike S Kleinsmann[5], Prof Dr; Douwe E Atsma[1,4,5], MD, Prof Dr; Hanno Pijl[2,3], MD, Prof Dr

[1]Department of Cardiology, Leiden University Medical Center, Leiden, Netherlands

[2]Division of Endocrinology, Department of Internal Medicine, Leiden University Medical Center, Leiden, Netherlands

[3]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, Netherlands

[4]National eHealth Living Lab, Leiden, Netherlands

[5]Department of Design, Organization and Strategy, Faculty of Industrial Design Engineering, Delft University of Technology, Delft, Netherlands

Corresponding Author:
Daan Leonhard de Frel, MD
Department of Cardiology
Leiden University Medical Center
B4-39g
Albinusdreef 2
Leiden, 2333 ZA
Netherlands
Phone: 31 715262020
Email: d.l.de_frel@lumc.nl

Abstract

Background: Type 2 diabetes (T2D) tremendously affects patient health and health care globally. Changing lifestyle behaviors can help curb the burden of T2D. However, health behavior change is a complex interplay of medical, behavioral, and psychological factors. Personalized lifestyle advice and promotion of self-management can help patients change their health behavior and improve glucose regulation. Digital tools are effective in areas of self-management and have great potential to support patient self-management due to low costs, 24/7 availability, and the option of dynamic automated feedback. To develop successful eHealth solutions, it is important to include stakeholders throughout the development and use a structured approach to guide the development team in planning, coordinating, and executing the development process.

Objective: The aim of this study is to develop an integrated, eHealth-supported, educational care pathway for patients with T2D.

Methods: The educational care pathway was developed using the first 3 phases of the Center for eHealth and Wellbeing Research roadmap: the contextual inquiry, the value specification, and the design phase. Following this roadmap, we used a scoping review about diabetes self-management education and eHealth, past experiences of eHealth practices in our hospital, focus groups with health care professionals (HCPs), and a patient panel to develop a prototype of an educational care pathway. This care pathway is called the Diabetes Box (Leiden University Medical Center) and consists of personalized education, digital educational material, self-measurements of glucose, blood pressure, activity, and sleep, and a smartphone app to bring it all together.

Results: The scoping review highlights the importance of self-management education and the potential of telemonitoring and mobile apps for blood glucose regulation in patients with T2D. Focus groups with HCPs revealed the importance of including all relevant lifestyle factors, using a tailored approach, and using digital consultations. The contextual inquiry led to a set of values that stakeholders found important to include in the educational care pathway. All values were specified in biweekly meetings with key stakeholders, and a prototype was designed. This prototype was evaluated in a patient panel that revealed an overall positive impression of the care pathway but stressed that the number of apps should be restricted to one, that there should be no delay in glucose value visualization, and that insulin use should be incorporated into the app. Both patients and HCPs stressed the importance of direct automated feedback in the Diabetes Box.

Conclusions: After developing the Diabetes Box prototype using the Center for eHealth and Wellbeing Research roadmap, all stakeholders believe that the concept of the Diabetes Box is useful and feasible and that direct automated feedback and education on stress and sleep are essential. A pilot study is planned to assess feasibility, acceptability, and usefulness in more detail.

(JMIR Hum Factors 2024;11:e45055) doi:10.2196/45055

KEYWORDS

diabetes mellitus; type 2; telemedicine; self-management; patient education as topic; activation; glucose regulation; Center for eHealth and Wellbeing Research; CeHRes; type 2 diabetes; tele; patient education; CeHRes roadmap; diabetes; glucose; insulin; education; development; usability; medical; behavioral; psychological; digital consultation; feasibility; endocrinology; endocrine; focus group; dietitian; psychologist; nurse; lifestyle factor; diet; exercise; stress; sleep; cardiovascular disease; heath care professional; mobile phone

Introduction

Background

Around 1 in 11 adults in Europe have diabetes mellitus, and the number of people with diabetes is increasing [1]. Currently, in more than 95% (n>1 million) of diabetes cases, it concerns type 2 diabetes (T2D). In people who are genetically predisposed to diabetes, adverse eating habits, excess body weight, and physical inactivity induce disruption of glucose control [2]. Hence, healthy lifestyle behaviors play a critical role in preventing and managing T2D. Indeed, quitting smoking, being more physically active, eating healthier, and losing weight when overweight can significantly reduce the risk of developing T2D [3,4]. Furthermore, in patients with recently diagnosed T2D, it was demonstrated that dietary lifestyle interventions can lead to persistent diabetes remission after 24 months in 36% (n=149) of patients [5].

Despite the obvious benefits of healthier lifestyles, adherence to healthy lifestyle behaviors in patients with T2D is poor [6-8]. This is alarming, as worse adherence obviously hampers therapeutic efficacy [9]. In Europe, glucose control is inadequate in at least half of the people with T2D. Inadequate glycemic regulation increases the risk of diabetes-related complications and mortality, and it increases medication use and health care costs [10,11]. Immediate action is needed to halt the rising incidence of T2D as well as to decrease the burden of T2D and curb health care costs [1].

A Cochrane review showed that diabetes self-management education (DSME) in people with T2D can improve glucose regulation [12]. In addition, it potentially improves blood pressure and reduces body weight and the requirement for diabetes medication. Yet, another systematic review reported that encouraging patients to play an active role in self-management, so-called patient activation or empowerment, can also improve glucose regulation [13]. Notably, mounting evidence clearly shows that the physiological response to lifestyle change is highly personal [14,15]. Moreover, it seems obvious to suppose that home monitoring of medical and behavioral parameters stimulates and improves self-management. Indeed, integrative monitoring of lifestyle behaviors and physiology using direct action-feedback loops potentially allows for the provision of informative personalized lifestyle advice [16].

Traditionally, DSME is done face-to-face, but digital tools can facilitate health behavior change and significantly improve glucose regulation in patients with T2D [17]. Effective digital tools are self-monitoring (eg, continuous glucose monitoring [CGM]) and telemonitoring by health care professionals (HCPs) [18,19]. Mobile phone apps providing automated feedback can also be effective in improving lifestyle modification and glucose regulation for people with T2D [20]. Indeed, due to low costs as compared to health care consultations and the 24/7 availability of HCPs, mobile phone apps have a lot of potential in diabetes management [21]. Currently available eHealth tools usually focus on one particular lifestyle component or relevant clinical parameter, such as CGM devices or apps that facilitate counting carbohydrates. Examples of digital tools that combine different lifestyle and biometrical parameters to improve self-management and glycemic control exist [22-24]. However, only a few digital tools exist that combine behavioral as well as biological data to provide informed, personalized lifestyle advice to people with T2D [24]. Most of the existing tools are one-size-fits-all lifestyle solutions. Personalized interventions are preferred as clinicians and patients together can choose the treatment plan that contributes most to favorable patient outcomes [25]. Here, we aimed to develop an eHealth-supported educational pathway using integrated behavioral and biological data collected by home monitoring to provide personalized lifestyle advice and promote self-management of people with T2D. Early involvement of stakeholders in the development process of eHealth tools is paramount for successful implementation in health care [26-30]. To assist in the construction of successful eHealth technologies, the Center for eHealth and Wellbeing Research (CeHReS) designed a roadmap to guide eHealth device development, implementation, and evaluation. The CeHReS roadmap consists of 5 phases and emphasizes stakeholder involvement throughout all of these phases [31]. The CeHReS roadmap was used to construct our educational program.

Objectives

Our aim is to empower patients with T2D to manage their disease by developing an integrated, eHealth-supported, blended educational pathway called the Diabetes Box (Leiden University Medical Center). In this paper, we delineate the different phases of the participatory development of the Diabetes Box using the CeHReS roadmap, and the lessons learned are shared.

Methods

Ethical Considerations

The accredited medical research ethics committee Leiden den Haag Delft (MREC registration P21.045) has reviewed the research protocol and gave its approval. The patients participating in the panel provided informed consent for their feedback and input to be used in scientific publication. Input data were deidentified. No compensation was provided for participating in the panel.

CeHReS Roadmap

The CeHReS roadmap was used to guide the development process of the Diabetes Box [31]. The CeHReS roadmap was designed to assist in planning, coordinating, and executing the development process of eHealth tools. The roadmap has a participatory dynamic and consists of 5 intertwined phases and continuous formative evaluation (Figure 1). The first 4 phases (ie, contextual inquiry, value specification, design, and operationalization) of the development process of the Diabetes Box are presented in this paper. The summative evaluation will be performed when the Diabetes Box is launched.

The contextual inquiry is meant to understand the challenges faced by the main stakeholders and how they could be solved. To this end, a literature review was performed. We followed the stages of a scoping review according to the revised Arksey and O'Malley framework [32]. We specified the research question "What diabetes self-management education strategies are being used in regular medical care?" The search strategies combined the terms "diabetes self management education," "technology/telemonitoring/glucose monitoring," and "healthcare/medical care." We searched PubMed and used Google for a broader search. One researcher (DLF) selected the studies and discussed these with a team consisting of 2 endocrinologists, a psychologist, a dietitian, and 2 diabetes nurses, all experienced in the field of DSME. Relevant studies were selected and summarized after which the team discussed the report. To elaborate further on the review of literature, previous experiences with eHealth in our center were evaluated, and important stakeholders were identified and interviewed in focus groups. Previous experiences mainly included technological and practical considerations from implementations of eHealth for patients with myocardial infarction, cardiac surgery, and COVID-19 [33-35]. The main stakeholders were patients, medical specialists, dietitians, psychologists, and diabetes nurses. The latter 4 would later form the development team and partake in the first 2 focus groups.

During the second phase, the value specification, the values gathered in the first phase were translated into (technological) requirements. What problems should the tool solve and how should it work? Weekly meetings with the relevant stakeholders (identified during the contextual inquiry) were used to refine the values and specify the technological requirements of the Diabetes Box.

Using these requirements, prototypes of the Diabetes Box were created during a highly dynamic, iterative, and collaborative design phase. Through biweekly meetings, the development team and stakeholders collaborated closely to ideate, create, and discuss ideas. A panel of patients with T2D gave feedback on the prototype. The entire development team was present on the web during the patient panel. Two members of the team wrote a summary of the recording, after which the recording was deleted. The entire team came together to discuss the outcomes of the patient panel, extract the most important aspects, and set out to change the prototype accordingly. Throughout the development process, the development team looked back on values and knowledge from previous phases to check the integrity of the design. Furthermore, at any point, incoming information could lead to adaptions in the process. This formative evaluation was enabled by constantly involving stakeholders in evaluations and decision-making.

When the design satisfied all stakeholders, the operationalization phase began. During this phase, the Diabetes Box was put into practice. First, a plan was made to implement the newly developed technology into the context defined by the contextual inquiry. The plan was made in close cooperation with the stakeholders to ensure a good fit. Second, the technology is launched.

In the fifth and last phase, the summative evaluation, the tool will be tested in the real world. Currently, the development team is setting up a pilot study to evaluate the feasibility, acceptability, and usability in clinical practice and get an impression of the clinical effects. It is important to note that the technology is quite versatile and adaptable to suit the practical demands of stakeholders as revealed during phase 5. A summary overview of all phases in this study is presented in Figure 2.

Figure 1. Overview of the CeHRes roadmap showing the different phases and formative evaluation (adapted from van Gemert-Pijnen et al [31]). CeHReS: Center for eHealth and Wellbeing Research.

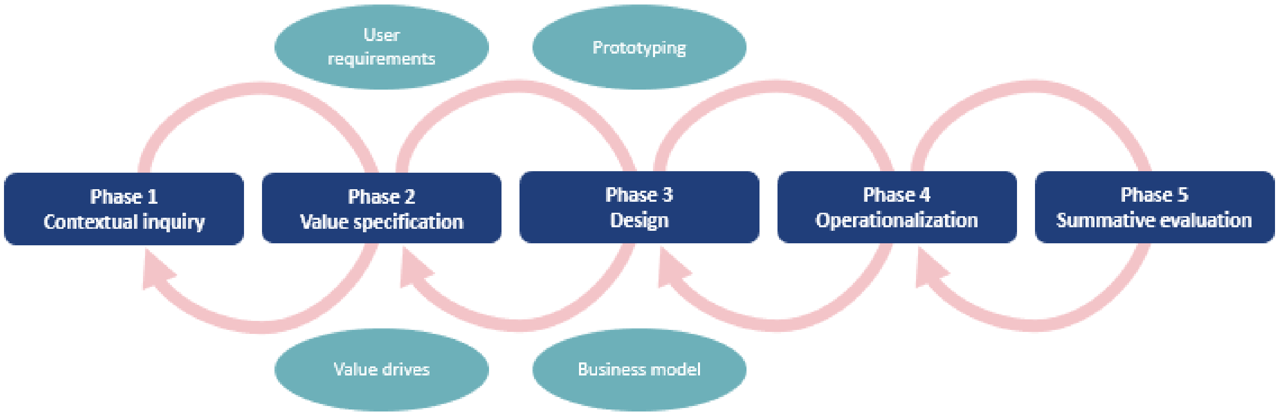
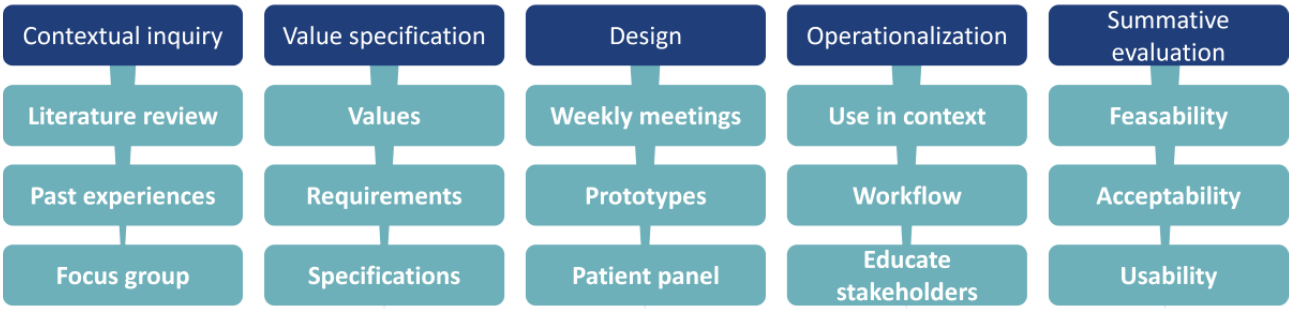


Figure 2. Overview of aspects in every phase. Note that this paper focuses on the contextual inquiry, the value specification, and the design phase [31].



Results

Phase 1: Contextual Inquiry

The contextual inquiry was meant to understand the challenges faced by the main stakeholders and how these challenges could be solved. We used a review of the literature and evaluated past experiences and focus groups with important stakeholders.

Literature Review

The information gathered from the literature review is summarized in Table 1. Studies have outlined several ways to support self-management, which can be categorized as education, monitoring, and modalities. Evidence shows that group-based education about disease pathophysiology, the influence of lifestyle (diet, exercise, stress, and sleep), self-management, and patient activation can improve glucose regulation, reduce body weight, and reduce the need for diabetes medication [4,12,17]. However, HCPs generally feel that they are insufficiently equipped to provide patients with T2D with the insights required to facilitate their health-related behavior change [36-38]. Furthermore, studies suggest that dietician-led lifestyle intervention as compared to interventions led by other HCPs achieves greater weight reductions [39].

Studies on monitoring indicate that self-monitoring (patients monitoring their own health parameters) and telemonitoring (using information technology to monitor patients at a distance) can significantly increase glucose regulation and reduce T2D-related complications [19]. For example, CGM significantly improves glucose regulation and reinforces patient satisfaction [18,48]. In addition, even though activity tracking has ambiguous effects on glucose regulation, it appears to reduce mortality and CVD risk in patients with T2D as well as the incidence of T2D in a general population [51,52]. Furthermore, blood pressure monitoring can decrease systolic blood pressure in patients with T2D when supported by an HCP [53]. To our knowledge, weight monitoring has not been assessed as a stand-alone intervention, but focus on weight can lead to stigma in patients with T2D, potentially leading to increased emotional distress [55].

Table 1. Outcomes of literature review.

Effects	Comments
DSME[a]	
• Self-management ↑[b] • Glucose regulation ↑ [4,12,13,39] • Knowledge or insight ↑ [12] • Body weight ↓[c] [12,39] • Need for medication ↓ [12] • Blood pressure ↓ [12]	• The inclusion of disease pathophysiology contributes to the effect of DSME [12] • Discussing the influence of lifestyle factors (diet, exercise, sleep, and stress) in DSME is decisive for its improvements [4,12,40-46] • Empowerment and patient activation beyond mere education are important in DSME [13] • Digital components of education can be effective [17,39] • The involvement of a dietitian increases the effect on body weight [39]
Telemonitoring	
• Glucose regulation ↑ [19] • Diabetes-related complications ↓ [19]	• Manual input may lead to erroneous input and can lower compliance [47]
CGM[d]	
• Glucose regulation ↑ [18,48] • Patient satisfaction ↑ [49]	• Failing to integrate well-structured education in glucose monitoring can diminish the effects on glucose regulation [50]
Activity tracker	
• Glucose regulation ↑/↓[e] [51] • Incidence T2D[f] ↓ [52] • Mortality ↓ [51] • CVD[g] risk ↓ [51]	• N/A[h]
Blood pressure monitor	
• Systolic blood pressure ↓ [53]	• HCP[i] support increases the effect [53,54]
Weight monitoring	
• Not assessed as a stand-alone	• Focus on weight monitoring and loss can be stigmatizing and lead to increased diabetes-related distress [55]
Mobile apps	
• Glucose regulation ↑ [56] • Monitoring or education ↑ • Lifestyle modification ↑ [20]	• HCP support increases the effect [56-58] • User-friendliness is an important aspect for success [58] • Apps can provide insight into self-management [57]
Dietary journal	
• Inform patients ↑ • Evaluate interventions ↑	• An easier, less time-consuming method would be beneficial to adherence • Photos have equal results as food weighing [59]

[a]DSME: diabetes self-management education.

[b]↑: improves.

[c]↓: deteriorates.

[d]CGM: continuous glucose monitoring.

[e]↑/↓: ambiguous results.

[f]T2D: type 2 diabetes.

[g]CVD: cardiovascular disease.

[h]N/A: not applicable.

[i]HCP: health care professional.

As far as modalities are concerned, mobile phone apps have a lot of potential in T2D management due to low costs, 24/7 availability, and dynamic automated feedback [21]. Evidence points out that mobile phone apps providing lifestyle advice can improve glucose regulation and facilitate lifestyle modification, particularly when they are supported by high-frequency HCP feedback [20,56,57]. Furthermore, keeping electronic dietary records effectively informs patients about the impact of food on glucose levels, but easy-to-use technology is needed [60,61]. For example, taking pictures of meals may be an adequate alternative of time-consuming, labor-intensive recording of dietary components [59].

What Experiences do we Have With Digital Tools Supporting Self-Management?

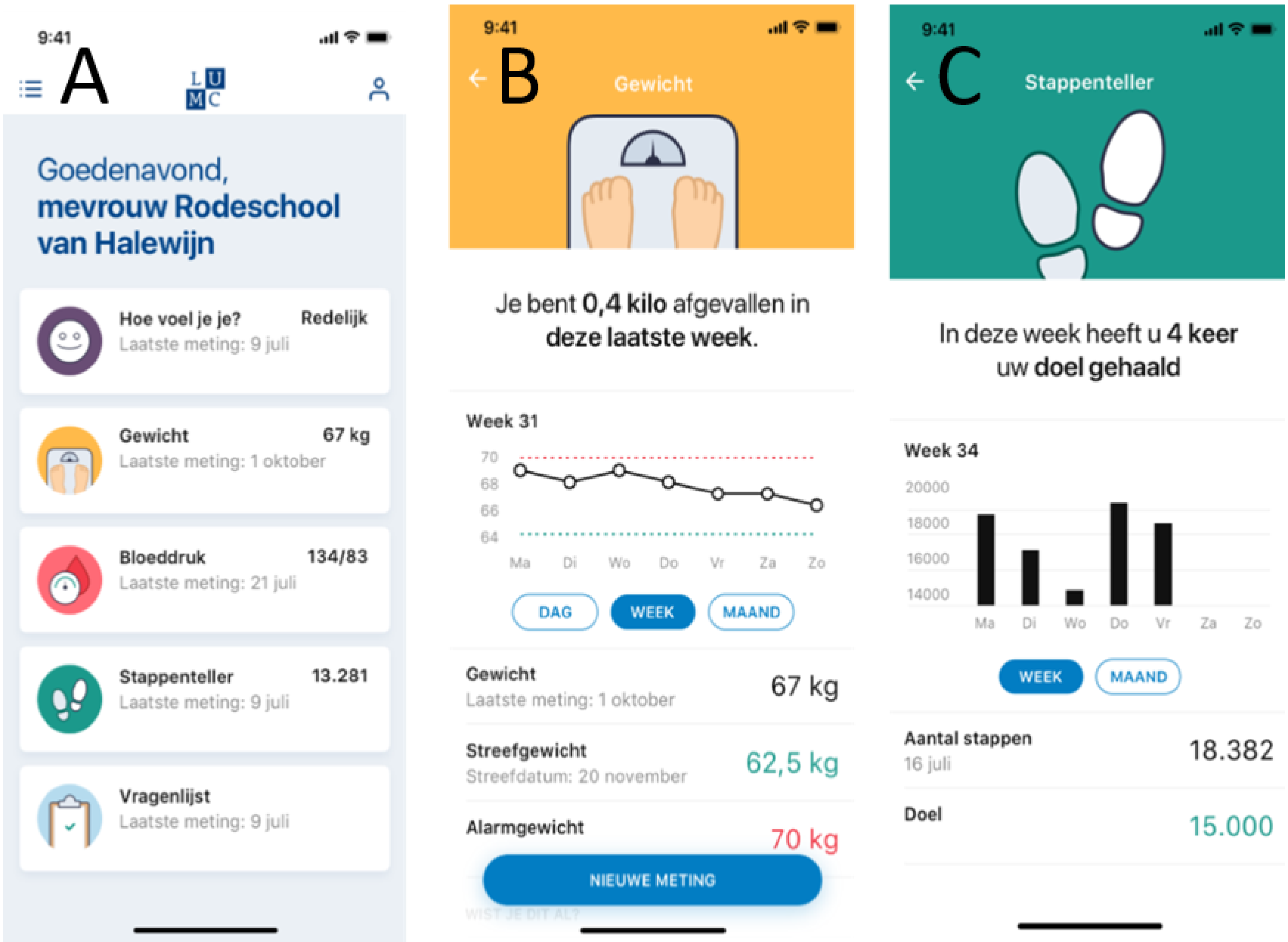
The Box

The Box (Leiden University Medical Center) comprises a set of eHealth tools that aim to improve self-management skills for a specific chronic condition. It includes devices for home monitoring of biological and behavioral parameters relevant to health (eg, glucose concentrations, physical activity, or blood pressure). The data are presented to the patient in a smartphone app called the LUMCCare app (discussed in LUMCCare App section). The data are also sent to the patients' electronic medical records in the hospital to allow evaluation by HCPs. The efficacy and safety of the Box have been examined in the follow-up care of patients with myocardial infarction. We recently reported that patient satisfaction with the Box was equal to regular medical care and that 96% (n=100) of participants appreciated that they could view their health data [62]. Furthermore, the Box has been shown to reduce hospital admissions by effectively surveying clinical symptoms and vital signs at home in patients with COVID-19 [35]. In conclusion, the Box appears to be an effective and appreciated prototype instrument for home monitoring that can be tailored to the health care needs of different conditions.

LUMCCare App

The LUMCCare app (Leiden University Medical Center) is a smartphone app available for Android and iOS. All data collected by the devices in the Box are automatically sent to the LUMCCare app via Bluetooth. The LUMCCare app was codeveloped with people with low health literacy, ensuring a good understanding and usability also in those individuals. Currently, the app was developed in the Dutch language and can display measurements of weight, blood pressure, heart rate, electrocardiograms, steps, temperature, and oxygen saturation (Figure 3). Moreover, users can indicate their level of well-being and provide a brief explanation. HCPs can also send questionnaires to patients through the app. The vast majority of patients with myocardial infarction report intensive and consistent use and high satisfaction with the app [62].

Figure 3. Screenshots of the LUMCCare app from left to right: (A) the home screen showing general well-being, weight, blood pressure, activity, and questionnaires, (B) the weight screen, and (C) the activity screen.



Stakeholders, Current Situation, and Experiences

The main stakeholders included patients, medical specialists, dietitians, psychologists, and diabetes nurses. To evaluate and confirm the findings of our scoping review and past experiences, we organized focus groups with a professor of diabetology, a clinical endocrinologist, a dietician, a psychologist, 2 diabetes nurses, an IT specialist, and a researcher.

According to international guidelines, people with T2D at least annually visit a physician (endocrinologist or general practitioner) and a nurse specialized in diabetes care. These HCPs should educate patients on (the role of lifestyle in) the pathophysiology of diabetes and on the types and dosing of available medication. Based on patient needs and health parameters, they decide if the patient requires a consult with the dietician or psychologist. All international guidelines advocate lifestyle intervention as a first step in the treatment of T2D. However, health care systems generally lack the means to adequately support patients trying to change deeply engrained habits. This is made exceedingly difficulty by an environment that relentlessly entices them to make unhealthy choices. All HCPs confirmed that continuous home monitoring of subcutaneous glucose concentrations has been a significant advance in supporting and motivating patients with diabetes to enhance their own grip on disease management. The notion that home monitoring of various relevant behavioral and biological parameters and integrating the data to yield personalized feedback would enhance patient empowerment and potentially improve self-management was broadly shared. To these ends, the contents of the Box and LUMCCare app were envisioned to require specific features as further defined in the next stage of development. A summary of the current situation is provided in Textbox 1.

Textbox 1. Current situation of health care for patients with type 2 diabetes (T2D) according to health care professionals (HCPs).

What is going well?

- Knowledge on diet, exercise, stress, and sleep is intermittently conveyed to patients by HCPs.
- All patients see an endocrinologist and diabetes nurse. If deemed necessary, a dietician and a psychologist are available.
- Adequate optimization of medication use.
- Close interdisciplinary collaboration between doctors, dieticians, psychologists, and diabetes nurses in the care for patients with diabetes.
- Health care can be delivered through digital means.

What can be improved?

- Patients' knowledge regarding the influence of lifestyle behaviors (diet, exercise, sleep, and stress) on glucose regulation.
- Activation of patients with T2D to improve lifestyle behaviors.
- Personalized lifestyle advice.
- Focus on personal goal setting.
- Shared decision-making regarding the timing and intensity of consultations with HCPs.
- Home monitoring of relevant parameters.
- Digital group consultations.

Phase 2: Value Specification

After a thorough exploration of the context and potential improvements, the next step was to translate the requirements of eHealth tools that were identified by the HCPs into specific technological properties. First, HCPs emphasized the need to more extensively convey the importance of lifestyle behaviors, including diet, exercise, stress, and sleep, in the control of glucose metabolism and the treatment of T2D. HCPs also stressed that the tool should tailor information and advice to the needs and wishes of the patient and that it should be easy to use for both patients and HCPs. Moreover, it should have features that activate patients to appropriately adapt their lifestyle. Activation was listed as a separate capacity of the eHealth tool. Finally, and importantly, the capability to monitor relevant parameters at home and easy accessibility to collected data for patients and HCPs were defined as prerequisites of an effective tool. This leads to a complete list of values, tool requirements, and tool specifications (Table 2).

Table 2. User perspective, user values, tool requirements, and tool specifications of the Diabetes Box.

Values	Tool requirements	Tool specifications
Provide insight, holistic view	Provide tailored education on the relationship between specific lifestyle factors and glucose regulation	• Include a graph of glucose combined with relevant lifestyle factors (diet, activity, sleep, and stress). • Include education on these topics.
Activate and stimulate	Help stimulate patients to adopt healthy behaviors	• Include goal setting in all aspects of self-management education. • Provide direct behavior–related feedback and education.
Personalized	Tailored to the patient	• Provide a place where patients can monitor their own personal and combined parameters.
24/7 Availability	Rely mostly on apps and e-learnings that are available 24/7	• Provide a digital resource that patients can use in their own time to measure glucose, diet, activity, stress, and sleep. • Provide links for access. • Resources or videos.
Integrated in health care	Integrated in health care	• Add Diabetes Box dashboard to the electronic medical record. • Plan education by HCPs[a] in work hours.
User-friendly	Easy to use, logical, and understandable	• Use B1-level language throughout the tool. • Simplify user interface. • Incorporate dashboard into electronic medical record. • Align education contents with the expertise of HCPs.
Monitoring patients	Monitoring of patient parameters and making them available for patient and HCP	• Provide a place where patients can gain insight into personal parameters of lifestyle factors and glucose. • Provide a dashboard where HCPs can monitor combined patient parameters to provide tailored support.
Low costs	No extra costs for patients, lower costs for health care	• Build on existing app content. • Use group meetings. • Free for patients (insurance covered).
Service desk	Enable a patient support desk	• Two separate phone numbers were provided for difficulties. First, the outpatient clinic number for diabetes-related questions, and second, the Box support desk for technology-related questions.

[a]HCP: health care professional.

Phase 3: Design

Overview

The Diabetes Box was developed using the participatory development method guided by the CeHReS roadmap. The design revolves around making prototypes of the Diabetes Box based on the tool specifications identified in the previous phases and gathering feedback from stakeholders. Our initial prototype was presented to HCPs and shared in a patient panel described below. After feedback from the HCPs, the Diabetes Box comprised digital self-measurement tools, an app, and DSME in the form of consultations and instructive videos (Figure 4). The tools included a continuous glucose monitor (Abbott Freestyle Libre), a sleep or activity tracker (Withings HR Steel), and a blood pressure monitor (Withings BPM Connect). The data collected were presented in the LUMCCare app. Subjective stress could also be registered, and food intake could be monitored by pictures taken of all that was consumed (Figure 5). All data were easily visualized in daily, weekly, and monthly overviews. The data of diet, activity, sleep, and subjective stress could also be plotted on the continuous glucose graph to provide insight into the relation between lifestyle factors and glucose regulation. An expert-led educational program was developed to further promote knowledge of the relationships between various components of lifestyle and glucose control. The overarching goal of the Diabetes Box is to empower patients and facilitate self-management of their disease. The program combines knowledge from routine diabetes care provided by dietitians, psychologists, endocrinologists, and specialized nurses. All educational material was developed, aiming to promote patient self-management. Therefore, multiple behavior change techniques were included in the development of the Diabetes Box. These included information provision, goal setting, action planning, self-monitoring, feedback provision, social comparison, and motivational interviewing. In Multimedia Appendix 1, we provide a list of behavior change techniques as described by Michie et al [63], including a description of the app in the Diabetes Box. There were nine 3- to 5-minute educational videos combining live feed and animations. The topics of these videos were an introduction to the Diabetes Box, the pathophysiology of diabetes, CGM, diet, exercise, sleep, stress, goal setting, and self-management. The educational program also entailed five consultations: (1) a group consultation introducing the digital tools, (2) an individual consultation focusing on diet, (3) a group consultation regarding diet and exercise, (4) a group consultation regarding sleep and stress, and (5) an individual consultation to evaluate, conclude, and set up future goals. Goal setting and patient activation were present in all videos and all educational consultations. The consultations lasted 45-90 minutes. Prior to each consultation, participants were asked to watch 1 or 2 videos.

Figure 4. Overview of the educational pathway for patients with T2D using the Diabetes Box.

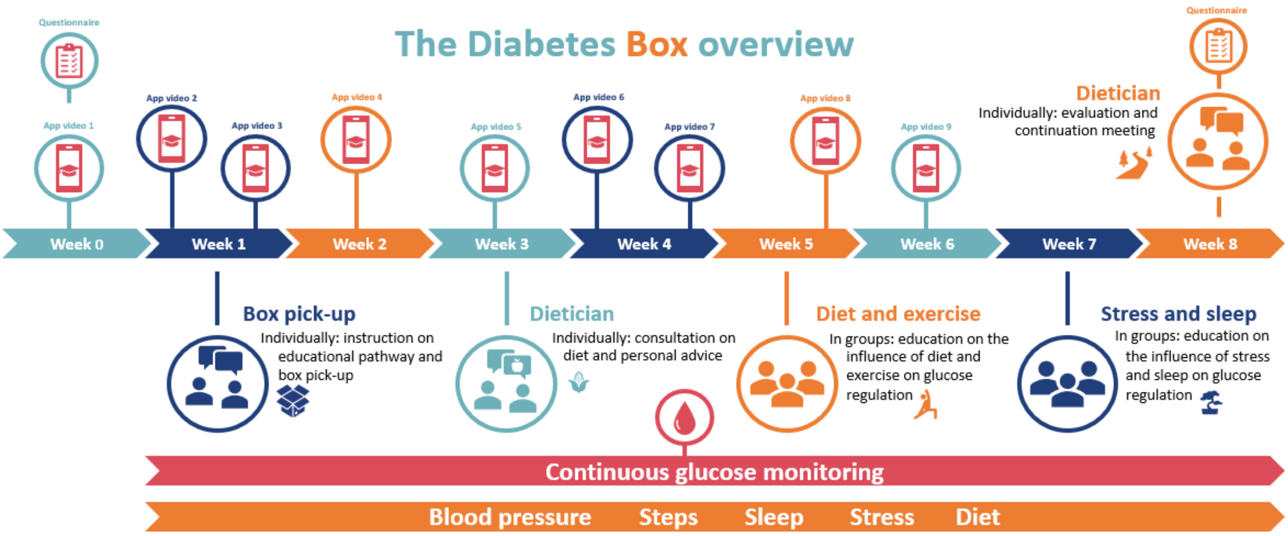
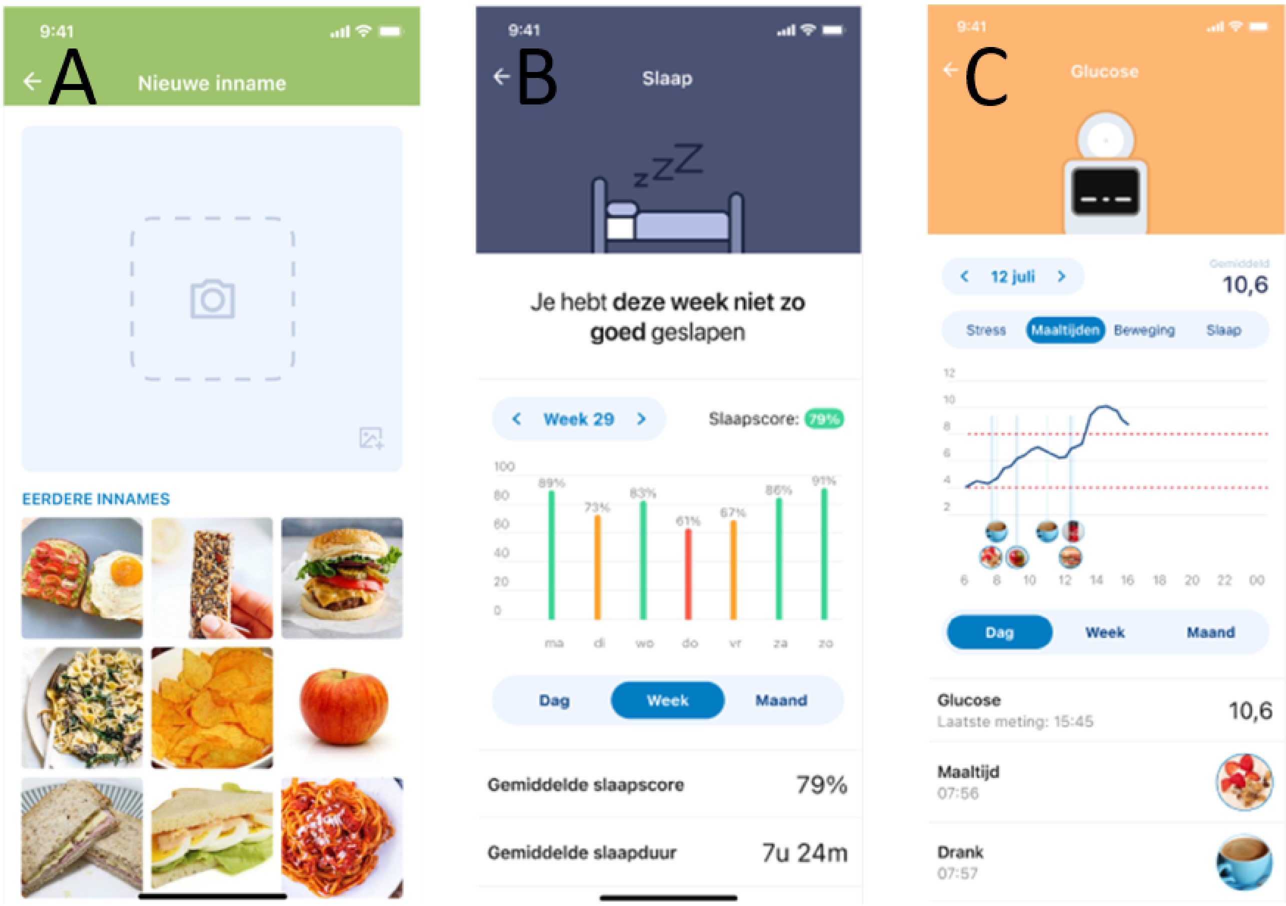


Figure 5. Screenshots of the new modalities of the LUMCCare app from left to right: (A) the screen to register a new intake, either food or drink, with a photo; (B) the sleep screen showing an average score based on estimated duration, interruptions, and regularity; and (C) the glucose screen combined with the diet screen showing the photos of intakes in the glucose graph.



Patient Panel

The prototype of the Box was shared with people with T2D to gather feedback on the preliminary design. Due to COVID-19 measures, only 4 people with diabetes were present during the session. The session took 90 minutes. The concept of the Box program was explained, 2 of the 9 educational videos were shown, the app and its functionalities were demonstrated, and people could try out the eHealth tools. The panel was asked questions covering 4 domains: general opinion, contents and clarity of the educational videos, functionality of eHealth instruments, and usability of the app. Overall, the evaluation was positive, while several potential improvements were suggested (Table 3).

Based on the feedback provided by the patient panel, multiple adaptations were made. First, a leaflet was added to the Box, explaining the flow of the program and anticipated time investment from the side of patients in more detail. Second, a web page was made, displaying the videos accompanied by instructions on when to watch which video. Third, a handout was made with detailing information about the different types of diabetes medications, their uses, and their common side effects. Fourth, as patients preferred a mix of consultation types, 2 consultations were planned online and 3 face-to-face. Fifth, the video of sleep was cut into 2 halves of 3 minutes to prevent viewers from quitting halfway through. Last, the LUMCCare was further developed to also accommodate insulin registration and other activities than steps (eg, cycling or swimming).

Table 3. Outcomes of patient panel.

Comments	Actions
General opinion	
• Positive first impression • Useful and feasible • Questions about total time investment, duration of the education, and loan of the devices • Advise to restrict the number of needed apps to one	• Informative material was made for participants of the Diabetes Box addressing the expected time investment, duration of the pathway, and the fact that participants can keep the devices. • Technologically we still need 2 apps. However, 1 only needs to run on the background and does not have to be opened.
Education	
• Positive • Duration and frequency seem feasible • Half preferred online (travel distance and comfort of home) and half preferred live (connection with HCP[a]) • Early evening is the best time • Videos are appreciated (up to 3 minutes) • A clear overview of videos and when to watch them is needed • Advice in education has to be consistent • Extra attention for diabetes medication, not all glucose levels can be related to behavior, correct use of the CGM[b]	• Hybrid pathway, part of consultations live, part online. • The video of sleep was cut into 2 parts. • A web page was made with an overview of all videos explaining when to watch which. • An extra handout was made about diabetes medication. • Attention to explaining glucose levels was added to the information.
Devices or measurements	
• Doable, clear, and easy to use • Frequency of measurement was regarded positively • Activities other than steps would be great • A (3 hours) delay in showing glucose values was deemed very impractical	• Added activity tracking other than steps in the app. • Focused on displaying real-time glucose data in the app.
LUMCCare app	
• Positive about layout and readability • Stress measurement and diet photos were deemed useful • Diet photos were deemed confronting in a helpful way • Incorporate insulin use in the app	• A functionality to register insulin use was added to the LUMCCare app.

[a]HCP: health care professional.

[b]CGM: continuous glucose monitoring.

Phases 4 and 5: Operationalization and Summative Evaluation

Operationalization involves the introduction of eHealth technology into practice. To test our design, we are currently planning a pilot study in 32 people with T2D to assess the feasibility, acceptability, and usability of the Box in clinical practice. Secondary objectives are evaluation of time in range and perceived learning. The study duration will be 2 months (as the concept Box program lasts 2 months). Participants will fill out a questionnaire before and after the study, and they will be interviewed about their experience as well. Patient satisfaction, user-friendliness of Box components, added value of the program in terms of disease management, and eventual use of the help desk will be evaluated. Consultation attendance, the use of eHealth tools and apps, and eventual replacement of glucose monitors will be registered. HCPs will be asked for their opinion regarding clinical practicalities in a structured interview, and average health care costs will be calculated.

Discussion

Principal Findings

Personalized lifestyle advice and promotion of self-management can help patients change their health behavior and improve glucose regulation. Digital tools have great potential in supporting patient self-management due to the effectiveness, low costs, 24/7 availability, and the option of dynamic automated feedback. However, reports documenting the impact of interventions incorporating multiple lifestyle modalities on glucose control are, to our knowledge, not available. Here, we developed an integrated, eHealth-supported, educational care pathway for people with T2D following the CeHReS roadmap and using a scoping review about DSME and eHealth, past experiences of eHealth practices in our hospital, focus groups with HCPs, and a patient panel. The care pathway aims to empower patients with T2D to self-manage their disease by providing them with direct feedback on their personal health behavior in relation to contemporaneous glucose levels.

HCPs and patients thought the concept of the Diabetes Box to be feasible, acceptable, and useful. The main strengths of the Diabetes Box were considered to be the integration of direct biofeedback on personal behavior, the focus on goal setting, and patient activation.

Comparison to Prior Work

The direct biofeedback regarding the impact of behavior on glucose concentrations was believed to be crucial to provide patients with insight into the relationships between their health behavior and glycemic control. A similar conclusion was drawn in an earlier study where patients with T2D were motivated to exercise while using CGM and accelerometer technology [64]. In many studies, data on lifestyle parameters were entered manually or via voice recording [22,24]. A Korean study showed input rates of diet and exercise of 24.9% and 5.3%, respectively [22]. Our study uses automatic input of steps, sleep, blood pressure, and glucose levels facilitating data gathering by participants. Diet was tracked using photographs. These photos were not analyzed for caloric content or carbohydrates but were used to provide insight into glucose-level fluctuations caused by certain food types. Beyond automated recording of behavioral and biological parameters, our app enables combining all lifestyle parameters with continuous glucose levels to create easily interpretable relations between lifestyle and glucose levels.

Regarding these lifestyle components, other interventions for people with T2D focus primarily on diet and exercise [23,24,65]. In one German study, stress management was included in the educational material, but stress or mood was not measured during the study [66]. The lifestyle components on which feedback should be provided include diet, physical activity, stress management, and sleep. Chronic stress may be a less obvious yet important disruptor of glucose control, as indicated by previous research [67,68]. Indeed, a recent meta-analysis revealed that stress reduction therapy improves glycemic control in people with diabetes [46]. The Diabetes Box gives direct biofeedback on diet, exercise, sleep, and stress. The effect of direct biofeedback on personal behavior is further enhanced by structured and tailored education. This is important, as physiological responses to lifestyle changes are often determined by personal characteristics [14,15]. Other studies often use one-size-fits-all education or even automated SMS text messages [24,65,69]. During the educational consultations in the Diabetes Box, HCPs can inform patients regarding the effects of their personal health behavior on metabolic control. The education in the Diabetes Box was designed to be simple, patient-centered, and multimodal, which is in line with the literature on successful patient education [70,71].

In addition, most of the existing tools are used in a research setting, and the challenge is to integrate these tools into regular medical care. A recent Dutch study showed that following a 2-year multicomponent lifestyle program outside of regular medical care could reduce medication use. In this setting, 71% of insulin users could stop insulin, and 28% of participants could stop glucose-lowering medication altogether. It must be stated though that only 234 of 438 starting participants were used in the final analyses [23]. When using these tools as regular medical care, all patients with T2D will follow the program instead of a selection of the more motivated patients. Our setup is to use this tool as regular medical care for all patients with T2D. The participatory development with an entire endocrinology team working in diabetes care can improve the chance of successful implementation. We are curious if similar results will be achieved when a multicomponent lifestyle program is integrated into regular medical care.

Studies have shown that people with higher levels of lifestyle-related knowledge (eg, influence of diet on glucose levels) tend to make healthier choices to improve their glycemic control [72]. However, better education and insight do not necessarily translate into behavior change [73]. To stimulate patients to change their behavior, goal setting and activation are integrated into all components of the Diabetes Box. All videos end with an assignment to self-monitor specific behavioral and biological parameters in preparation of the next consultation, and a separate video about goal setting is included. Furthermore, the individual consultations with the dietitian revolve around diet but also cover reflection on goals set by a patient. It is difficult to empower people with insufficient diabetes-related knowledge to manage their disease [74]. Therefore, we believe it is the combination of direct feedback, structured and tailored education, and targeted patient activation that grants the Diabetes Box its great potential.

The participatory development process played a critical role in the realization of the Diabetes Box. Involving all stakeholders from the start proved very fruitful, as it clearly facilitated the creation of a program that fits all stakeholder demands. The CeHReS roadmap was very helpful as well. It provided handholds and courses of action, which make it easier to make and measure progress. In addition, the value specification generated a concrete set of wishes from the key stakeholders that could be used to fall back and make decisions.

Limitations and Strengths

Obviously, there are at least 2 issues that limit the broad-based application of the Diabetes Box for the time being. First of all, the feasibility of the program as well as its impact on metabolic control and quality of life of patients with T2D needs to be evaluated in clinical practice. As the program was primarily developed by stakeholders employed by a third-line, academic medical institution, it also needs to be tested if it works for patients under regular surveillance by primary care (n>1 million, 90% of patients with T2D are treated by their general practitioners in the Netherlands). Second, although the LUMCCare app was created as a "white label" app, which means that it is relatively easy to adapt external characteristics, it was designed as part of the local (Leiden University Medical Center) infrastructure. Use by other institutions would therefore probably require modifications. We are willing to help and assist hospitals and other health care institutions that want to implement the Diabetes Box into their regular medical care for people with T2D. The challenges we foresee are training personnel and integrating the Diabetes Box into their daily workflow. The type of specialist who provides the consultations can be changed depending on what professionals are motivated and at hand. In addition, the content of the educational material can be altered to better fit the personal approach of the professional providing the education. On a technological basis, challenges also exist. The app is white label and can be easily adapted to accommodate the look and feel of other institutions. However, the data generated in the app have to be made available for the HCPs involved. This will most commonly involve integration into the electronic medical records, which is a process that costs both time and money. In the near future, the Diabetes Box will be tested in a single-center, mixed methods, sequential explanatory pilot study including approximately 32 patients with T2D, with the primary aim to assess its feasibility, acceptability, and usability. Secondary objectives will be to evaluate its impact on the "time in range" of glucose levels and perceived learning. Subsequently, in case of promising results, the Diabetes Box will be tested for efficacy in a larger, multicenter (including primary care) intervention study.

Conclusions

We have developed a unique care pathway in close collaboration with relevant stakeholders in order to ensure a good fit. The combined effects of direct biofeedback on personal behavior, structured and tailored education, and goal setting should empower people with T2D to improve their self-management and glycemic control. A pilot study is planned to assess feasibility, acceptability, and usability in more detail.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Description of behavior change techniques used in the Diabetes Box.

[DOCX File , 18 KB - humanfactors_v11i1e45055_app1.docx]

References

1. IDF diabetes atlas. International Diabetes Federation. 2019. URL: https://diabetesatlas.org/idfawp/resource-files/2019/07/IDF_diabetes_atlas_ninth_edition_en.pdf [accessed 2023-04-12]
2. Diabetes fact sheet. World Health Organization. 2020. URL: https://www.who.int/news-room/fact-sheets/detail/diabetes [accessed 2023-03-11]
3. Ford ES, Bergmann MM, Kröger J, Schienkiewitz A, Weikert C, Boeing H. Healthy living is the best revenge: findings from the European Prospective Investigation Into Cancer and Nutrition-Potsdam study. Arch Intern Med 2009;169(15):1355-1362. [doi: 10.1001/archinternmed.2009.237] [Medline: 19667296]
4. Uusitupa M, Khan TA, Viguiliouk E, Kahleova H, Rivellese AA, Hermansen K, et al. Prevention of type 2 diabetes by lifestyle changes: a systematic review and meta-analysis. Nutrients 2019;11(11):2611 [FREE Full text] [doi: 10.3390/nu11112611] [Medline: 31683759]
5. Churuangsuk C, Hall J, Reynolds A, Griffin SJ, Combet E, Lean MEJ. Diets for weight management in adults with type 2 diabetes: an umbrella review of published meta-analyses and systematic review of trials of diets for diabetes remission. Diabetologia 2022;65(1):14-36 [FREE Full text] [doi: 10.1007/s00125-021-05577-2] [Medline: 34796367]
6. da Rocha RB, Silva CS, Cardoso VS. Self-care in adults with type 2 diabetes mellitus: a systematic review. Curr Diabetes Rev 2020;16(6):598-607. [doi: 10.2174/1573399815666190702161849] [Medline: 31267873]
7. Jarvie JL, Pandey A, Ayers CR, McGavock JM, Sénéchal M, Berry JD, et al. Aerobic fitness and adherence to guideline-recommended minimum physical activity among ambulatory patients with type 2 diabetes mellitus. Diabetes Care 2019;42(7):1333-1339 [FREE Full text] [doi: 10.2337/dc18-2634] [Medline: 31221698]
8. Rivellese AA, Boemi M, Cavalot F, Costagliola L, De Feo P, Miccoli R, et al. Dietary habits in type II diabetes mellitus: how is adherence to dietary recommendations? Eur J Clin Nutr 2008;62(5):660-664. [doi: 10.1038/sj.ejcn.1602755] [Medline: 17426738]

9. Houle J, Beaulieu MD, Chiasson J, Lespérance F, Côté J, Strychar I, et al. Glycaemic control and self-management behaviours in type 2 diabetes: results from a 1-year longitudinal cohort study. Diabet Med 2015;32(9):1247-1254. [doi: 10.1111/dme.12686] [Medline: 25581545]
10. Alonso-Fernández M, Mancera-Romero J, Mediavilla-Bravo JJ, Comas-Samper JM, López-Simarro F, Pérez-Unanua MP, et al. Glycemic control and use of A1c in primary care patients with type 2 diabetes mellitus. Prim Care Diabetes 2015;9(5):385-391. [doi: 10.1016/j.pcd.2015.01.006] [Medline: 25686480]
11. American Diabetes Association. 6. Glycemic targets: standards of medical care in diabetes-2021. Diabetes Care 2021;44(Suppl 1):S73-S84 [FREE Full text] [doi: 10.2337/dc21-S006] [Medline: 33298417]
12. Deakin T, McShane CE, Cade JE, Williams RDRR. Group based training for self-management strategies in people with type 2 diabetes mellitus. Cochrane Database Syst Rev 2005(2):CD003417. [doi: 10.1002/14651858.CD003417.pub2] [Medline: 15846663]
13. Almutairi N, Hosseinzadeh H, Gopaldasani V. The effectiveness of patient activation intervention on type 2 diabetes mellitus glycemic control and self-management behaviors: a systematic review of RCTs. Prim Care Diabetes 2020;14(1):12-20. [doi: 10.1016/j.pcd.2019.08.009] [Medline: 31543458]
14. Berry SE, Valdes AM, Drew DA, Asnicar F, Mazidi M, Wolf J, et al. Human postprandial responses to food and potential for precision nutrition. Nat Med 2020;26(6):964-973 [FREE Full text] [doi: 10.1038/s41591-020-0934-0] [Medline: 32528151]
15. Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A, et al. Personalized nutrition by prediction of glycemic responses. Cell 2015;163(5):1079-1094 [FREE Full text] [doi: 10.1016/j.cell.2015.11.001] [Medline: 26590418]
16. Klein HA, Lippa KD. Type 2 diabetes self-management: controlling a dynamic system. J Cogn Eng Decis Mak 2008;2(1):48-62. [doi: 10.1518/155534308x284363]
17. Gershkowitz BD, Hillert CJ, Crotty BH. Digital coaching strategies to facilitate behavioral change in type 2 diabetes: a systematic review. J Clin Endocrinol Metab 2021;106(4):e1513-e1520 [FREE Full text] [doi: 10.1210/clinem/dgaa850] [Medline: 33206975]
18. Park C, Le QA. The effectiveness of continuous glucose monitoring in patients with type 2 diabetes: a systematic review of literature and meta-analysis. Diabetes Technol Ther 2018;20(9):613-621. [doi: 10.1089/dia.2018.0177] [Medline: 30095980]
19. Jaana M, Paré G. Home telemonitoring of patients with diabetes: a systematic assessment of observed effects. J Eval Clin Pract 2007;13(2):242-253. [doi: 10.1111/j.1365-2753.2006.00686.x] [Medline: 17378871]
20. Wu X, Guo X, Zhang Z. The efficacy of mobile phone apps for lifestyle modification in diabetes: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019;7(1):e12297 [FREE Full text] [doi: 10.2196/12297] [Medline: 30664494]
21. Hou C, Carter B, Hewitt J, Francisa T, Mayor S. Do mobile phone applications improve glycemic control (HbA1c) in the self-management of diabetes? A systematic review, meta-analysis, and GRADE of 14 randomized trials. Diabetes Care 2016;39(11):2089-2095 [FREE Full text] [doi: 10.2337/dc16-0346] [Medline: 27926892]
22. Oh SW, Kim KK, Kim SS, Park SK, Park S. Effect of an integrative mobile health intervention in patients with hypertension and diabetes: crossover study. JMIR Mhealth Uhealth 2022;10(1):e27192 [FREE Full text] [doi: 10.2196/27192] [Medline: 35014961]
23. Pot GK, Battjes-Fries MC, Patijn ON, van der Zijl N, Pijl H, Voshol P. Lifestyle medicine for type 2 diabetes: practice-based evidence for long-term efficacy of a multicomponent lifestyle intervention (Reverse Diabetes2 Now). BMJ Nutr Prev Health 2020;3(2):188-195 [FREE Full text] [doi: 10.1136/bmjnph-2020-000081] [Medline: 33521528]
24. Lee EY, Cha SA, Yun JS, Lim SY, Lee JH, Ahn YB, et al. Efficacy of personalized diabetes self-care using an electronic medical record-integrated mobile app in patients with type 2 diabetes: 6-month randomized controlled trial. J Med Internet Res 2022;24(7):e37430 [FREE Full text] [doi: 10.2196/37430] [Medline: 35900817]
25. Ruissen MM, Rodriguez-Gutierrez R, Montori VM, Kunneman M. Making diabetes care fit—are we making progress? Front Clin Diabetes Healthc 2021;2:658817 [FREE Full text] [doi: 10.3389/fcdhc.2021.658817] [Medline: 36994329]
26. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]
27. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of eHealth interventions: systematic review of the literature. J Med Internet Res 2018;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
28. Wachter RM. Making IT work: harnessing the power of health information technology to improve care in England. National Advisory Group on Health Information Technology in England. 2016. URL: https://assets.publishing.service.gov.uk/media/5a8091afe5274a2e87dba8f2/Wachter_Review_Accessible.pdf [accessed 2023-01-26]
29. van Velsen L, Evers M, Bara C, Op den Akker H, Boerema S, Hermens H. Understanding the acceptance of an eHealth technology in the early stages of development: an end-user walkthrough approach and two case studies. JMIR Form Res 2018;2(1):e10474 [FREE Full text] [doi: 10.2196/10474] [Medline: 30684434]
30. Hennemann S, Beutel ME, Zwerenz R. Ready for eHealth? Health professionals' acceptance and adoption of eHealth interventions in inpatient routine care. J Health Commun 2017;22(3):274-284. [doi: 10.1080/10810730.2017.1284286] [Medline: 28248626]

31. van Gemert-Pijnen JEWC, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 05;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
32. Westphaln KK, Regoeczi W, Masotya M, Vazquez-Westphaln B, Lounsbury K, McDavid L, et al. From Arksey and O'Malley and beyond: customizations to enhance a team-based, mixed approach to scoping review methodology. MethodsX 2021;8:101375 [FREE Full text] [doi: 10.1016/j.mex.2021.101375] [Medline: 34430271]
33. Treskes RW, van den Akker-van Marle ME, van Winden L, van Keulen N, van der Velde ET, Beeres S, et al. The Box—eHealth in the outpatient clinic follow-up of patients with acute myocardial infarction: cost-utility analysis. J Med Internet Res 2022;24(4):e30236 [FREE Full text] [doi: 10.2196/30236] [Medline: 35468091]
34. Biersteker TE, Boogers MJ, Schalij MJ, Penning de Vries B, Groenwold RHH, van Alem AP, et al. Mobile health vs. standard care after cardiac surgery: results of The Box 2.0 study. Europace 2023;25(1):49-58 [FREE Full text] [doi: 10.1093/europace/euac115] [Medline: 35951658]
35. Dirikgil E, Roos R, Groeneveld GH, Heringhaus C, Silven AV, Petrus AHJ, et al. Home monitoring reduced short stay admissions in suspected COVID-19 patients: COVID-box project. Eur Respir J 2021;58(2):2100636 [FREE Full text] [doi: 10.1183/13993003.00636-2021] [Medline: 33795321]
36. Stuckey HL, Vallis M, Burns KK, Mullan-Jensen CB, Reading JM, Kalra S, et al. "I do my best to listen to patients": qualitative insights into DAWN2 (diabetes psychosocial care from the perspective of health care professionals in the second diabetes attitudes, wishes and needs study). Clin Ther 2015;37(9):1986-1998.e12 [FREE Full text] [doi: 10.1016/j.clinthera.2015.06.010] [Medline: 26169765]
37. Raaijmakers LG, Hamers FJ, Martens MK, Bagchus C, de Vries NK, Kremers SP. Perceived facilitators and barriers in diabetes care: a qualitative study among health care professionals in the Netherlands. BMC Fam Pract 2013;14:114 [FREE Full text] [doi: 10.1186/1471-2296-14-114] [Medline: 23937325]
38. Yoon S, Ng JH, Kwan YH, Low LL. Healthcare professionals' views of factors influencing diabetes self-management and the utility of a mHealth application and its features to support self-care. Front Endocrinol (Lausanne) 2022;13:793473 [FREE Full text] [doi: 10.3389/fendo.2022.793473] [Medline: 35282452]
39. Sun Y, You W, Almeida F, Estabrooks P, Davy B. The effectiveness and cost of lifestyle interventions including nutrition education for diabetes prevention: a systematic review and meta-analysis. J Acad Nutr Diet 2017;117(3):404-421.e36 [FREE Full text] [doi: 10.1016/j.jand.2016.11.016] [Medline: 28236962]
40. García-Molina L, Lewis-Mikhael AM, Riquelme-Gallego B, Cano-Ibáñez N, Oliveras-López MJ, Bueno-Cavanillas A. Improving type 2 diabetes mellitus glycaemic control through lifestyle modification implementing diet intervention: a systematic review and meta-analysis. Eur J Nutr 2020;59(4):1313-1328. [doi: 10.1007/s00394-019-02147-6] [Medline: 31781857]
41. Hemmingsen B, Gimenez-Perez G, Mauricio D, Figuls MRI, Metzendorf MI, Richter B. Diet, physical activity or both for prevention or delay of type 2 diabetes mellitus and its associated complications in people at increased risk of developing type 2 diabetes mellitus. Cochrane Database Syst Rev 2017;12(12):CD003054 [FREE Full text] [doi: 10.1002/14651858.CD003054.pub4] [Medline: 29205264]
42. Kumar AS, Maiya AG, Shastry BA, Vaishali K, Ravishankar N, Hazari A, et al. Exercise and insulin resistance in type 2 diabetes mellitus: a systematic review and meta-analysis. Ann Phys Rehabil Med 2019;62(2):98-103 [FREE Full text] [doi: 10.1016/j.rehab.2018.11.001] [Medline: 30553010]
43. Esposito K, Maiorino MI, Bellastella G, Chiodini P, Panagiotakos D, Giugliano D. A journey into a Mediterranean diet and type 2 diabetes: a systematic review with meta-analyses. BMJ Open 2015;5(8):e008222 [FREE Full text] [doi: 10.1136/bmjopen-2015-008222] [Medline: 26260349]
44. Cappuccio FP, D'Elia L, Strazzullo P, Miller MA. Quantity and quality of sleep and incidence of type 2 diabetes: a systematic review and meta-analysis. Diabetes Care 2010;33(2):414-420 [FREE Full text] [doi: 10.2337/dc09-1124] [Medline: 19910503]
45. Uchendu C, Blake H. Effectiveness of cognitive-behavioural therapy on glycaemic control and psychological outcomes in adults with diabetes mellitus: a systematic review and meta-analysis of randomized controlled trials. Diabet Med 2017;34(3):328-339. [doi: 10.1111/dme.13195] [Medline: 27472405]
46. Ni Y, Ma L, Li J. Effects of mindfulness-based stress reduction and mindfulness-based cognitive therapy in people with diabetes: a systematic review and meta-analysis. J Nurs Scholarsh 2020;52(4):379-388. [doi: 10.1111/jnu.12560] [Medline: 32406186]
47. Seto E, Istepanian RSH, Cafazzo JA, Logan A, Sungoor A. UK and Canadian perspectives of the effectiveness of mobile diabetes management systems. Annu Int Conf IEEE Eng Med Biol Soc 2009;2009:6584-6587. [doi: 10.1109/IEMBS.2009.5333998] [Medline: 19964700]
48. Janapala RN, Jayaraj JS, Fathima N, Kashif T, Usman N, Dasari A, et al. Continuous glucose monitoring versus self-monitoring of blood glucose in type 2 diabetes mellitus: a systematic review with meta-analysis. Cureus 2019;11(9):e5634 [FREE Full text] [doi: 10.7759/cureus.5634] [Medline: 31700737]

49. Beck RW, Riddlesworth TD, Ruedy K, Ahmann A, Haller S, Kruger D, et al. Continuous glucose monitoring versus usual care in patients with type 2 diabetes receiving multiple daily insulin injections: a randomized trial. Ann Intern Med 2017;167(6):365-374. [doi: 10.7326/M16-2855] [Medline: 28828487]
50. Chomutare T, Fernandez-Luque L, Arsand E, Hartvigsen G. Features of mobile diabetes applications: review of the literature and analysis of current applications compared against evidence-based guidelines. J Med Internet Res 2011;13(3):e65 [FREE Full text] [doi: 10.2196/jmir.1874] [Medline: 21979293]
51. Hall KS, Hyde ET, Bassett DR, Carlson SA, Carnethon MR, Ekelund U, et al. Systematic review of the prospective association of daily step counts with risk of mortality, cardiovascular disease, and dysglycemia. Int J Behav Nutr Phys Act 2020;17(1):78 [FREE Full text] [doi: 10.1186/s12966-020-00978-9] [Medline: 32563261]
52. Kraus WE, Janz KF, Powell KE, Campbell WW, Jakicic JM, Troiano RP, et al. Daily step counts for measuring physical activity exposure and its relation to health. Med Sci Sports Exerc 2019;51(6):1206-1212 [FREE Full text] [doi: 10.1249/MSS.0000000000001932] [Medline: 31095077]
53. Tucker KL, Sheppard JP, Stevens R, Bosworth HB, Bove A, Bray EP, et al. Self-monitoring of blood pressure in hypertension: a systematic review and individual patient data meta-analysis. PLoS Med 2017;14(9):e1002389 [FREE Full text] [doi: 10.1371/journal.pmed.1002389] [Medline: 28926573]
54. Logan AG, Irvine MJ, McIsaac WJ, Tisler A, Rossos PG, Easty A, et al. Effect of home blood pressure telemonitoring with self-care support on uncontrolled systolic hypertension in diabetics. Hypertension 2012;60(1):51-517 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.111.188409] [Medline: 22615116]
55. Puhl RM, Himmelstein MS, Hateley-Browne JL, Speight J. Weight stigma and diabetes stigma in U.S. adults with type 2 diabetes: associations with diabetes self-care behaviors and perceptions of health care. Diabetes Res Clin Pract 2020;168:108387. [doi: 10.1016/j.diabres.2020.108387] [Medline: 32858100]
56. Hou C, Xu Q, Diao S, Hewitt J, Li J, Carter B. Mobile phone applications and self-management of diabetes: a systematic review with meta-analysis, meta-regression of 21 randomized trials and GRADE. Diabetes Obes Metab 2018;20(8):2009-2013. [doi: 10.1111/dom.13307] [Medline: 29582538]
57. Torbjørnsen A, Ribu L, Rønnevig M, Grøttland A, Helseth S. Users' acceptability of a mobile application for persons with type 2 diabetes: a qualitative study. BMC Health Serv Res 2019;19(1):641 [FREE Full text] [doi: 10.1186/s12913-019-4486-2] [Medline: 31492176]
58. Jeffrey B, Bagala M, Creighton A, Leavey T, Nicholls S, Wood C, et al. Mobile phone applications and their use in the self-management of Type 2 Diabetes Mellitus: a qualitative study among app users and non-app users. Diabetol Metab Syndr 2019;11:84 [FREE Full text] [doi: 10.1186/s13098-019-0480-4] [Medline: 31636719]
59. Rollo ME, Ash S, Lyons-Wall P, Russell AW. Evaluation of a mobile phone image-based dietary assessment method in adults with type 2 diabetes. Nutrients 2015;7(6):4897-4910 [FREE Full text] [doi: 10.3390/nu7064897] [Medline: 26091234]
60. Thompson FE, Subar AF. Chapter 1—Dietary assessment methodology. In: Coulston AM, Boushey CJ, Ferruzzi MG, Delahanty LM, editors. Nutrition in the Prevention and Treatment of Disease. 4th Edition. Cambridge, MA: Academic Press; 2017:5-48.
61. Schap TE, Six BL, Delp EJ, Ebert DS, Kerr DA, Boushey CJ. Adolescents in the United States can identify familiar foods at the time of consumption and when prompted with an image 14 h postprandial, but poorly estimate portions. Public Health Nutr 2011;14(7):1184-1191 [FREE Full text] [doi: 10.1017/S1368980010003794] [Medline: 21324224]
62. Treskes RW, van Winden LAM, van Keulen N, van der Velde ET, Beeres SLMA, Atsma DE, et al. Effect of smartphone-enabled health monitoring devices vs regular follow-up on blood pressure control among patients after myocardial infarction: a randomized clinical trial. JAMA Netw Open 2020;3(4):e202165 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.2165] [Medline: 32297946]
63. Michie S, Ashford SA, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
64. Allen NA, Jacelon CS, Chipkin SR. Feasibility and acceptability of continuous glucose monitoring and accelerometer technology in exercising individuals with type 2 diabetes. J Clin Nurs 2009;18(3):373-383 [FREE Full text] [doi: 10.1111/j.1365-2702.2008.02533.x] [Medline: 19191984]
65. Chen L, Chuang LM, Chang CH, Wang CS, Wang IC, Chung Y, et al. Evaluating self-management behaviors of diabetic patients in a telehealthcare program: longitudinal study over 18 months. J Med Internet Res 2013;15(12):e266 [FREE Full text] [doi: 10.2196/jmir.2699] [Medline: 24323283]
66. von Storch K, Graaf E, Wunderlich M, Rietz C, Polidori MC, Woopen C. Telemedicine-assisted self-management program for type 2 diabetes patients. Diabetes Technol Ther 2019;21(9):514-521. [doi: 10.1089/dia.2019.0056] [Medline: 31287736]
67. Derosa G, Salvadeo SAT, Mereu R, D'Angelo A, Ciccarelli L, Piccinni MN, et al. Continuous glucose monitoring system in free-living healthy subjects: results from a pilot study. Diabetes Technol Ther 2009;11(3):159-169. [doi: 10.1089/dia.2008.0101] [Medline: 19267581]
68. Sevil M, Rashid M, Hajizadeh I, Park M, Quinn L, Cinar A. Physical activity and psychological stress detection and assessment of their effects on glucose concentration predictions in diabetes management. IEEE Trans Biomed Eng 2021;68(7):2251-2260 [FREE Full text] [doi: 10.1109/TBME.2020.3049109] [Medline: 33400644]

69. Bretschneider MP, Klásek J, Karbanová M, Timpel P, Herrmann S, Schwarz PEH. Impact of a digital lifestyle intervention on diabetes self-management: a pilot study. Nutrients 2022;14(9):1810 [FREE Full text] [doi: 10.3390/nu14091810] [Medline: 35565782]
70. Cutilli CC. Excellence in patient education: evidence-based education that "sticks" and improves patient outcomes. Nurs Clin North Am 2020;55(2):267-282. [doi: 10.1016/j.cnur.2020.02.007] [Medline: 32389259]
71. Friedman AJ, Cosby R, Boyko S, Hatton-Bauer J, Turnbull G. Effective teaching strategies and methods of delivery for patient education: a systematic review and practice guideline recommendations. J Cancer Educ 2011;26(1):12-21. [doi: 10.1007/s13187-010-0183-x] [Medline: 21161465]
72. Cheng LJ, Wang W, Lim ST, Wu VX. Factors associated with Glycaemic control in patients with diabetes mellitus: a systematic literature review. J Clin Nurs 2019;28(9-10):1433-1450. [doi: 10.1111/jocn.14795] [Medline: 30667583]
73. Kelly MP, Barker M. Why is changing health-related behaviour so difficult? Public Health 2016;136:109-116 [FREE Full text] [doi: 10.1016/j.puhe.2016.03.030] [Medline: 27184821]
74. Lambrinou E, Hansen TB, Beulens JW. Lifestyle factors, self-management and patient empowerment in diabetes care. Eur J Prev Cardiol 2019;26(2_suppl):55-63 [FREE Full text] [doi: 10.1177/2047487319885455] [Medline: 31766913]

Abbreviations

CeHReS: Center for eHealth and Wellbeing Research
CGM: continuous glucose monitoring
DSME: diabetes self-management education
HCP: health care professional
T2D: type 2 diabetes

Edited by A Kushniruk; submitted 15.12.22; peer-reviewed by V Buss, SM Ching , SLS Drop; comments to author 14.05.23; revised version received 17.09.23; accepted 20.11.23; published 31.05.24.

Please cite as:
de Frel DL, Schroijen MA, Aardoom JJ, van Gils W, Huisman SD, Janssen VR, Versluis A, Kleinsmann MS, Atsma DE, Pijl H
Participatory Development of an Integrated, eHealth-Supported, Educational Care Pathway (Diabetes Box) for People With Type 2 Diabetes: Development and Usability Study
JMIR Hum Factors 2024;11:e45055
URL: https://humanfactors.jmir.org/2024/1/e45055
doi:10.2196/45055
PMID:38819880

Original Paper

Assessing the Utility, Impact, and Adoption Challenges of an Artificial Intelligence–Enabled Prescription Advisory Tool for Type 2 Diabetes Management: Qualitative Study

Sungwon Yoon[1,2], MPH, PhD; Hendra Goh[1], MPH; Phong Ching Lee[3], MBChB, MRCP; Hong Chang Tan[3], MBBS, MRCP; Ming Ming Teh[3], MB BCh BAO, MRCP, MD; Dawn Shao Ting Lim[3], MBBS, MRCP, MMed; Ann Kwee[3], MB BCh BAO, MRCP, MMed; Chandran Suresh[3], MBBS, MRCP; David Carmody[3], MB BCh BAO, LRCP & SI, MRCP, MD; Du Soon Swee[3], MBBS, MRCP; Sarah Ying Tse Tan[3], MBBS, MRCP, MMed; Andy Jun-Wei Wong[3], MBBS; Charlotte Hui-Min Choo[3], MBBS; Zongwen Wee[3], MBBS, MRCP, MMed; Yong Mong Bee[3], MBBS, MRCP

[1]Health Services and Systems Research, Duke-NUS Medical School, Singapore, Singapore
[2]Centre for Population Health Research and Implementation, SingHealth Regional Health System, SingHealth, Singapore, Singapore
[3]Department of Endocrinology, Singapore General Hospital, Singapore, Singapore

Corresponding Author:
Sungwon Yoon, MPH, PhD
Health Services and Systems Research
Duke-NUS Medical School
8 College Road
Singapore, 169857
Singapore
Phone: 65 66013198
Email: sungwon.yoon@duke-nus.edu.sg

Abstract

Background: The clinical management of type 2 diabetes mellitus (T2DM) presents a significant challenge due to the constantly evolving clinical practice guidelines and growing array of drug classes available. Evidence suggests that artificial intelligence (AI)–enabled clinical decision support systems (CDSSs) have proven to be effective in assisting clinicians with informed decision-making. Despite the merits of AI-driven CDSSs, a significant research gap exists concerning the early-stage implementation and adoption of AI-enabled CDSSs in T2DM management.

Objective: This study aimed to explore the perspectives of clinicians on the use and impact of the AI-enabled Prescription Advisory (APA) tool, developed using a multi-institution diabetes registry and implemented in specialist endocrinology clinics, and the challenges to its adoption and application.

Methods: We conducted focus group discussions using a semistructured interview guide with purposively selected endocrinologists from a tertiary hospital. The focus group discussions were audio-recorded and transcribed verbatim. Data were thematically analyzed.

Results: A total of 13 clinicians participated in 4 focus group discussions. Our findings suggest that the APA tool offered several useful features to assist clinicians in effectively managing T2DM. Specifically, clinicians viewed the AI-generated medication alterations as a good knowledge resource in supporting the clinician's decision-making on drug modifications at the point of care, particularly for patients with comorbidities. The complication risk prediction was seen as positively impacting patient care by facilitating early doctor-patient communication and initiating prompt clinical responses. However, the interpretability of the risk scores, concerns about overreliance and automation bias, and issues surrounding accountability and liability hindered the adoption of the APA tool in clinical practice.

Conclusions: Although the APA tool holds great potential as a valuable resource for improving patient care, further efforts are required to address clinicians' concerns and improve the tool's acceptance and applicability in relevant contexts.

(JMIR Hum Factors 2024;11:e50939) doi:10.2196/50939

KEYWORDS
clinical decision support system; artificial intelligence; endocrinology; diabetes management; human factors

Introduction

Diabetes mellitus is a chronic condition that affects millions of people worldwide. In Singapore, the prevalence of diabetes is estimated to surpass 400,000, with 1 out of 3 individuals at risk of developing the condition [1]. Uncontrolled diabetes can lead to various complications, such as neuropathy, retinopathy, and nephropathy. Diabetes is primarily associated with cardiovascular diseases, particularly ischemic heart disease and myocardial infarction, which account for most of the mortality cases in patients with diabetes [2,3].

Managing diabetes clinically poses a considerable challenge due to its complex nature. The treatment of diabetes involves achieving specific targets, such as optimal control of glycemia, blood pressure, and lipid levels, primarily relying on laboratory tests [4]. Regular review of test results and subsequent treatment adjustments are important in minimizing the risk of long-term complications and aligning with recommended targets [5]. However, the need to monitor multiple laboratory markers during clinical consultations can impose a cognitive burden. Furthermore, incomplete integration of critical patient data into the electronic medical record (EMR) can lead to errors in disease monitoring, compromising the quality of patient care [6].

Evidence suggests that clinical decision support systems (CDSSs) can assist clinicians in effectively monitoring patient data and making accurate and informed treatment decisions [7-9]. Traditionally, CDSSs have relied on medical expertise and clinical practice guidelines. However, keeping CDSS content and knowledge up-to-date is increasingly challenging due to the evolving nature of clinical practices [10]. The advent of big data and machine learning has enabled the development of artificial intelligence (AI)-powered CDSSs, capable of diagnosing conditions, suggesting evidence-based treatment options and aiding in care planning [11,12]. Research shows that AI-powered CDSSs have improved the quality of diabetes care and patient outcomes [12,13].

Despite the positive impacts of AI-driven CDSSs on health care, fewer studies have examined human factors. In addition, several critical issues surrounding AI-powered clinical tools have been brought to attention, including concerns regarding the transparency of underlying algorithms, accountability, data privacy, and limited trust and applicability [10,14]. Although these studies provide essential insights into implementing AI-based CDSSs, a significant gap exists in research concerning the early-stage implementation of an AI-enabled CDSS specifically for type 2 diabetes mellitus (T2DM). Evaluating a CDSS in the early stages of implementation is of utmost importance to optimize its benefits and mitigate potential drawbacks, as it offers vital information on use, acceptability, and the challenges pertaining to human factors in real-world clinical settings.

To support clinicians in making better treatment decisions in T2DM management, the AI-enabled Prescription Advisory (APA) tool was developed and integrated within the endocrinology specialist clinics at the Diabetes & Metabolism Centre in Singapore General Hospital. To ensure that the tool is capable of scaling up and meeting the needs of its users, it is crucial to assess its appropriateness within the clinical context. Therefore, this study aims to explore the perspectives of clinicians regarding the use and impact of the APA tool while also identifying potential challenges associated with its adoption and application.

Methods

Overview

This study adopted qualitative research methodology to assess the usability of the tool. For rigor and transparency, we anchored our study according to the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist [15].

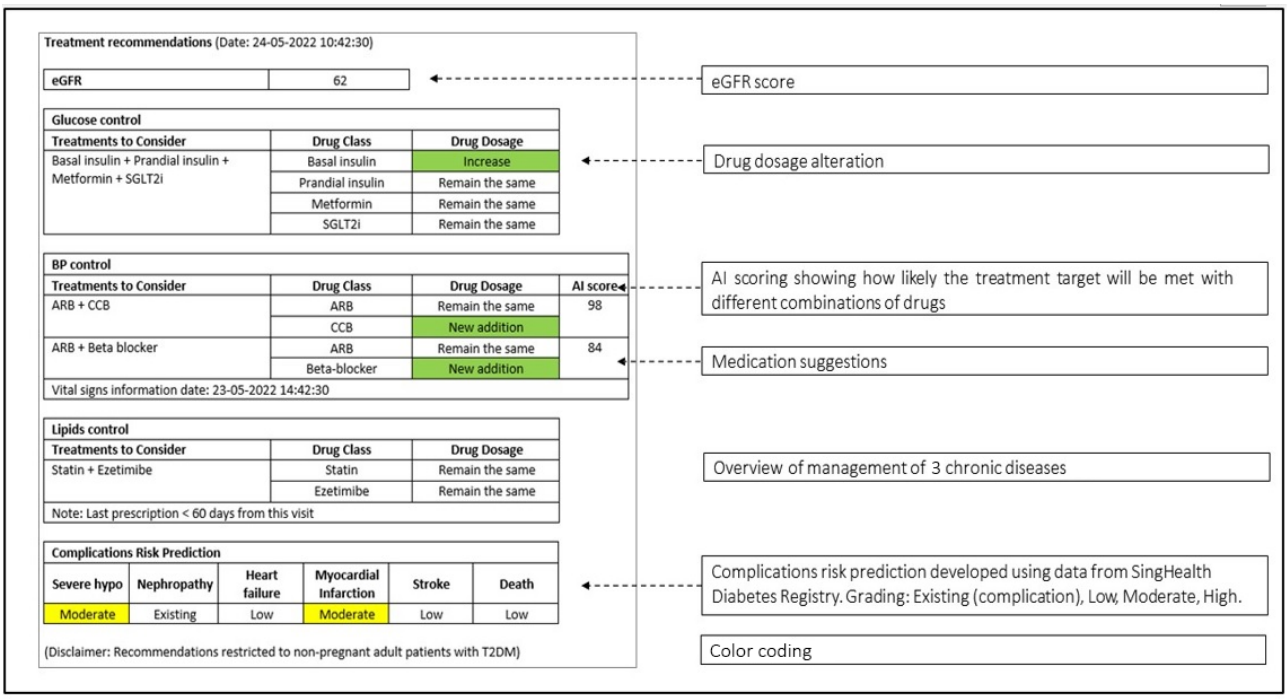
Development of an AI-Enabled Diabetes CDSS

The diabetes CDSS, also known as the APA tool, was developed using data gathered from the Singapore Health Services Diabetes Registry that comprised a total of 189,520 patients with diabetes. This data set included 6,407,958 outpatient visits spanning over 5 years from 2013 to 2018 [16]. For model development, 80% of the data set was used to build therapeutic recommendations, while the remaining 20% was used to test and validate the trained models. Three distinct therapeutic recommendation models were formulated for antiglycemic, antihypertensive, and lipid-lowering treatments. These models were created by integrating both a knowledge-driven approach and a data-driven approach. The knowledge-driven approach, initially drawing inputs from clinical guidelines and expert opinions, was used to identify potential therapeutic options. Subsequently, the data-driven approach that used deep learning techniques was used to select the identified therapeutic options based on anticipated clinical outcomes. To assess the performance of model's prediction, short-term outcomes compared therapeutic options between treatments that aligned with the model's recommendations and those that did not. Confounding factors were also accounted along the way and adjusted by stratification and multivariate regression. For evaluation of long-term outcomes, the rates of model-concordant treatments were computed by multivariate logistic regression to determine whether the combined treatments exhibited a positive impact on reducing the occurrence of long-term complications and mortality.

Features of the AI-Enabled Diabetes CDSS

Presently, the APA tool has been integrated into the EMR system at the Singapore General Hospital to provide tailored treatment recommendations for achieving target glucose, blood pressure, and cholesterol levels. It features 3 distinct AI components designed to improve patient outcomes and support clinical decision-making. The first component recommends drug classes based on laboratory markers such as glycated hemoglobin (HbA_{1c}), low-density lipoprotein cholesterol, and blood pressure measurements, aiding clinicians in selecting the most appropriate drug classes to achieve glycemic, low-density lipoprotein cholesterol and blood pressure treatment targets. The second component generates an AI score that indicates the likelihood of reaching treatment targets when adopting a suggested new drug therapy. The third component generates AI-based diabetic complications risk predictions, providing specific complication risks associated with suboptimal diabetes treatment. All outputs generated by the AI model are color-coded for enhanced visibility. Figure 1 provides a snapshot of the APA tool's outputs, illustrating how clinical results extracted from a patient's EMR are presented.

Figure 1. Outputs of APA tool. The APA tool features include laboratory markers related to diabetes care, medication prescribing recommendations, color-coding to highlight changes, and diabetic complications risk prediction. AI: artificial intelligence; APA: AI-enabled Prescription Advisory; ARB: angiotensin receptor blockers; BP: blood pressure; CCB: calcium channel blockers; eGFR: estimated glomerular filtration rate; T2DM: type 2 diabetes mellitus.



***Note**: The APA tool features include laboratory markers related to diabetes care, medication prescribing recommendations, color-coding to highlight changes and diabetic complications risk prediction.*

Participants

Eligible participants were (1) clinicians trained in endocrinology, (2) currently employed full-time by the institution, (3) completed training in APA, and (4) used the APA for a minimum duration of 4 weeks. We purposively selected participants according to age, gender, and seniority level in the workplace to gain a range of perspectives. Participants were approached via email, and informed consent was obtained prior to enrollment into the study.

Study Procedure

Prospective participants were purposefully selected and approached by the research team to ensure their engagement in the study. Consented participants took part in a comprehensive group training session, encompassing the following key components: (1) a comprehensive overview of the tool's development and validation process, (2) exploration of the specific features and underlying knowledge rules that inform the therapeutic recommendations, (3) familiarization with the tool's output (ie, therapeutic recommendations), and (4) an interactive question and answer segment addressing general inquiries. The entirety of the training session was conducted over the course of 1 hour. Following the training, the participants were granted immediate access to the APA tool during their clinical consultations. It was emphasized that the participants had the freedom to decide on the final treatments for their patients with T2DM, regardless of the clinical recommendations provided by the APA tool. Each participant was given a minimum of 4 weeks of exposure to the tool in the clinic setting before his or her involvement in this study. Following the 4-week period, the participants were invited to partake in a focus group discussion (FGD).

Data Collection

A semistructured FGD guide was developed by drawing upon relevant literature and leveraging the expert knowledge of the study team [17,18]. FGD was chosen to foster a dynamic and interactive environment that encouraged the exploration of shared experiences and perspectives within the professional context, thereby providing a comprehensive understanding of the collective viewpoints users had on the APA tool. To attain a variety of opinions, the participants were recruited according to their seniority (registrars, associate consultants, consultants, and senior consultants). FGDs were conducted exclusively among participants holding equivalent hierarchical positions within the workplace, with the specific intention of minimizing the potential for power differentials and fostering open and candid dialogue. Key topics of interest included (1) participants' firsthand experiences while using the APA tool, (2) perceptions and evaluations of the various features, (3) impacts of the APA tool on clinical practice, and (4) challenges related to the adoption and application. The interview guide underwent pilot-testing multiple iterations. Consented individuals were then invited to participate in virtual FGD (2-6 participants per session according to seniority) over Zoom (Zoom Video Communications, Inc) by a facilitator (HG) trained in qualitative research methodology. Reflections were recorded after each FGD to capture and document valuable insights shared during the discussions. The duration of the FGD ranged from 50 minutes to 75 minutes.

Data Analysis

Interviews were audio-recorded following verbal consent and transcribed verbatim. Transcripts were checked for validity, and any identified errors were corrected. Two coders (SY and HG) reviewed the transcripts independently and thematically analyzed the data using NVivo (version 12; Lumivero). We used reflexive thematic analysis following each completed interview [19], contributing to the ongoing refinement and direction of the interview guide for subsequent interviews. The coding categories evolved from initial open coding to more analytical coding of the text, ultimately revealing a series of interconnected themes and patterns. The analysis and interviews continued until no new emerging themes were identified. In case of discrepancies, iterative discussions involving study team members were conducted to resolve any differences and ensure consistency in the analysis process.

Recognizing the inherent influence of the coders' subjective perspectives in the research process, the study team prioritized strategies aimed at effectively managing these preconceptions and upholding the integrity and credibility of the analysis. Specifically, we implemented the following measures: (1) the establishment of an elaborate coding protocol, meticulously designed to promote consistency and minimize potential subjective interpretations and (2) regular engagement in peer-debriefing sessions and member checking to validate the interpretations and enhance the credibility and confirmability of our findings.

Ethical Considerations

Ethics approval was obtained from SingHealth Centralized institutional review board (2022/2329). Prior to the FGD, a proficient research coordinator (HG) engaged with each participant individually to meticulously review the participant information sheet and consent form. Particular emphasis was placed on elucidating the study's objectives, along with an extensive exploration of potential foreseeable risks and benefits. Upon satisfactory comprehension, the participants were invited to provide their informed consent by endorsing the documentation. Furthermore, they were duly informed that all collected data would undergo a stringent deidentification process to preserve anonymity. To uphold transparency and equity in the compensation process, the participants were explicitly notified beforehand that no form of compensation would be provided for their involvement in this study. It was reiterated that participation in the study was entirely voluntary.

Results

Characteristics of Participants

In total, 18 clinicians were contacted, and 13 responded positively to the invitation and participated in the FGDs. The remaining 5 clinicians declined to participate, citing time constraints and lack of interest as the reasons. The FGDs were conducted based on participants' seniority at work. Most participants were male (n=8, 61%) and aged between 31 and 50 years (n=12, 92%). The participants held various designations in their respective roles, including resident (n=3, 23%), associate consultant (n=1, 8%), consultant (n=5, 38%), and senior consultant (n=4, 31%). Notably, more than three-quarters (n=10, 77%) of the participants had no prior experience with AI-enabled CDSSs. Detailed characteristics of the study participants are shown in Table 1.

Table 1. Characteristics of participants (N=13).

Characteristics	Participants, n (%)
Age (years)	
31-40	6 (46)
41-50	6 (46)
Older than 50 years	1 (8)
Sex	
Female	5 (38)
Male	8 (61)
Seniority at work	
Resident	3 (23)
Associate consultant	1 (8)
Consultant	5 (38)
Senior consultant	4 (31)
Prior experience with AI[a]-enabled CDSSs[b]	
Yes	3 (23)
No	10 (77)

[a]AI: artificial intelligence.

[b]CDSS: clinical decision support system.

Our analysis yielded 2 themes and 9 subthemes that represented the participants' perspectives concerning the use and impact of the APA tool on patient care and clinical practice and the challenges to adoption and application of the APA tool. Descriptions of themes and subthemes are presented in Textbox 1.

Textbox 1. Main themes and subthemes.

Use and impact of the AI–enabled Prescription Advisory (APA) tool on patient care and clinical practice

- Supporting decision-making for patients with comorbidities (artificial intelligence [AI]–powered drug class recommendations)
- Facilitating doctor-patient communication (diabetic complications risk predictions)
- Enhancing clinical confidence through cross-checking (color-coded AI-generated recommendations)
- Serving as a gatekeeper against medical negligence

Challenges concerning adoption and application of the APA tool

- Interpretability issues due to the lack of standardized guidelines on AI risk predictions
- Mistrust in the system driven by perceived lack of transparency around system development and information sourcing
- Limited applicability in a specialist setting given extensive expertise and patient care accountability of endocrinologists
- Concerns about potential harm in light of occasional contradictions between the APA tool recommendation and a clinician's professional judgment
- Frustration with technical issues associated with the tool implementation

Use and Impact of APA Tool on Patient Care and Clinical Practice

When asked about their experience with the APA tool, most participants expressed a positive impact, highlighting its potential to guide clinical decision-making as a key benefit. Specifically, the participants appreciated the AI algorithm's ability to provide drug class recommendations based on patients' laboratory markers. Notably, the tool not only simplified chronic disease management but also assisted in identifying instances of suboptimal disease control that might have otherwise gone unnoticed during consultations. This streamlined approach proved invaluable in guiding clinicians toward effectively managing comorbidities and reducing the risk of long-term complications in patients with diabetes.

> *I think the most valuable part for me is the lipid control feature. Sometimes when engrossed in discussing patients' diabetes treatment plans, which is anyway their primary reasons for seeking consultation, I may overlook the assessment of their LDL-c levels. With the APA tool, a quick glance provides a clear indication of whether they are on*

> *target or not. There isn't much extra clinical information that is required by the tool, so I am able to rely on the medication recommendations to appropriately adjust the medication for hyperlipidemia.* [FGD 2, senior consultant]

Furthermore, some participants saw the AI-generated complication risk predictions as a helpful resource in “convincing patients to adhere to certain treatments or treatment plans.” By presenting visible evidence regarding the potential risks linked to noncompliance or inadequate treatment, the tool showed considerable potential in facilitating doctor-patient communication based on risk prediction.

> *The complications risk prediction feature stands out as particularly beneficial to me. For example, it provides an alert regarding the risk of hypoglycemia. When the risk level is classified as moderate or high, this information helped me better persuade patients to consider specific treatments or to improve their compliance with the recommended approach.* [FGD 2, senior consultant]

Some participants pointed out the lack of quantitative representation for the AI-generated complication risk prediction scores. They proposed an interactive time series graph that would visually illustrate the fluctuations in risk scores over time following the adoption of the tool’s recommendations. Participants believed that integrating visual aids would enhance patients’ understanding of their current risk levels associated with complications and promote the benefits of adhering to treatment plans.

> *The ability to visually present individual risk in a quantitative way through graphical or pictorial means and illustrate the potential changes that may occur after adopting the systems' recommendations would improve information delivery. I personally believe that patients are more inclined to accept the recommendations when they see their risk in a pictorial or a graphical format.* [FGD 1, associate consultant]

The color-coded AI-generated recommendations served as an additional point of reference during consultations, particularly when discrepancies emerged between the tool’s recommendations and the clinician’s own knowledge. This feature not only fostered critical thinking but also prompted clinicians to consider additional clinical histories that might have been overlooked initially. Overall, clinicians reported an enhanced level of confidence in their clinical decisions, thereby “improving the quality of patient care.”

> *So, the tool helps to reinforce my decision-making. The color-coded recommendations provide a clear visual indication, prompting me to address any discrepancies that may arise between the tool's suggestions and my own clinical plan. In this case, I delve into additional clinical histories that the tool does not have access to and elucidate the rationale behind my decisions. This process enhances my confidence and guides better decision-making during the clinical visit, which can improve the quality of patient care.* [FGD 1, consultant]

By and large, the participants perceived the APA tool as a mechanism to prevent the risk of negligence, especially in fast-paced clinical environments. Acting as a “gatekeeper for patient safety,” the APA tool effectively identified and flagged abnormal results, mitigating the risk of overlooking important tasks. The tool was regarded as a valuable partner in pursuit of delivering high-quality and safe care to patients.

> *I like the idea of the tool as a gatekeeper for patient safety. Making sure doctors don't forget things, reminding us to check and act on abnormal results. I think that is useful for busy clinics.* [FGD 1, associate consultant]

Challenges to the Adoption and Application of APA Tool

Although participants generally acknowledged the beneficial effects of the APA tool on quality patient care and clinical practice, they equally expressed reservations about incorporating and using the tool in their own clinical settings. One major concern centered around the interpretability of the automatically generated AI score when new drugs were recommended. While participants appreciated the availability of the scoring system to inform the likelihood of achieving treatment targets based on the recommendations, they remained unsure about the interpretability of the AI score.

> *I think it is quite interesting that the system is able to provide different percentages of achieving optimum blood pressure when different combinations of new drugs are used. However, my question is if plan A gives a score of 48 while plan B gives a score of 45, are these recommendations still clinically relevant? I mean, of course, the situation is more direct in cases with scores such as 98 and 88, then it will make more sense to pick the plan with 98% of likelihood.* [FGD 1, consultant]

A sense of mistrust in the APA system emerged, which appeared to stem from the unfamiliarity surrounding AI-based recommendations and concerns regarding transparency of the information sourcing and system development. Some participants openly expressed their hesitancy in adopting the APA tool due to the absence of essential clinical data. Without access to this information, they were not confident enough to use the tool.

> *When it [APA] was launched, a lot of us were not very sure how it was developed. I think part of the reason why we did not use it very much is also because we are not so familiar with how this system came about, what kind of information was used, and where the information came from. Is it also possible that critical information was not captured in the system? I can't trust totally, and [I am] not confident with what I'm seeing at the moment.* [FGD 3, senior consultant]

However, participants expressed openness to embrace the tool if they were presented with additional information. They emphasized the importance of transparent communication regarding the evidence supporting the system and the sources of information used. By gaining a clearer understanding of the logic and rules behind the recommendations, they would be more inclined to use the tool in their clinical practice.

> *[T]hat being said, if more information or transparent communication is given to us, I might be more inclined to use it in clinics. As I know the logic and rules behind these recommendations and where they are sourced from.* [FGD 3, senior consultant]

While a minority, some clinicians exhibited strong confidence in their own clinical judgment and thus did not see the necessity to rely on the APA tool. They felt that their experience and specialist training surpassed the assistance provided by the AI-driven system. In addition, they highlighted the potential ramifications of relying on CDSS recommendations, emphasizing that the responsibility for patient outcomes ultimately rested with the clinician. Consequently, this attitude led to a reluctance to use the tool, particularly among those who believed that they possessed the requisite expertise to make well-informed decisions in patient care.

> *I would say that I'm as good or even better than the system. I don't feel the need to rely on it; I'll just do what I do. We are all trained endocrinologists, so we trust our judgment because that has been our bread and butter for many years. At the end of the day, we bear the responsibility for our patients, so you know, if the algorithm makes a sound decision, but something unfortunate ever happens to the patient, then it's still our own accountability on the line.* [FGD 1, consultant]

These clinicians suggested the potential for the APA tool to bring benefits to the wider primary care community, particularly those who may be "less familiar with endocrinology clinical practices." They believed that the tool could assist general practitioners in effectively managing patients with complex cases and improve patient engagement.

> *These recommendations would be more valuable in a primary healthcare setting, where doctors may not have extensive knowledge of clinical practices related to novel glucose-lowering medications and insulin titration, especially in complex cases. I think implementing the AP tool in such settings would greatly help doctors in improving patient engagement and care.* [FGD 1, consultant]

Another important theme was related to the potential harm of the APA tool's drug recommendations on patients. Participants noted that the recommendation occasionally contradicted their own professional judgment. They cautioned against solely relying on algorithmic recommendations for clinical decision-making.

> *Some of the recommendations go against your clinical judgement. For example, I have two patients and the AI recommendation was to add a beta blocker to someone who doesn't have ischemic heart disease as a second line agent. That's just not something that we would normally do. So have to exercise caution too!* [FGD 1, consultant]

Finally, the participants expressed their frustration with the technical issues associated with the integration of the APA tool into the EMR system. The slow loading of clinical notes resulted in delayed clinical consultations, which added unnecessary mental burdens for some participants. Moreover, there were instances in which the clinical notes failed to load entirely, thereby affecting the quality of patient care.

> *One significant issue we encounter after implementing the CDSS is the considerable lag in loading clinical notes. It takes a few minutes to retrieve the clinical notes. So, by that time, I'm typically already engaged in a conversation with the patient, and we may even come up with a plan without the notes being available. In some instances, I can't even see the clinical notes at all.* [FDG 4, registrar]

Discussion

Principal Findings

This qualitative study explored clinicians' perspectives on the use and impact of the APA tool, as well as challenges to its adoption and application in clinical practice. In terms of use, the APA tool offers several useful features to assist clinicians in effectively managing diabetes. As shown in the literature, patients with T2DM frequently experience multiple comorbidities, which may add complexity to pharmacotherapy management and increase the mental burden of prescribing practices [20]. Our findings suggest that the AI algorithms for drug alteration embedded in the APA tool were generally viewed as a good knowledge resource in supporting the clinician's decision-making on drug modifications at the point of care, particularly for patients with T2DM with comorbidities.

Complications arising from diabetes pose a significant burden on the public health care system [21,22]. In light of this, an important feature developed in the APA tool was the diabetic complications risk prediction that provides information on the likelihood of developing the 6 most common diabetic complications in patients with T2DM [23,24]. We found that participants viewed the risk prediction as having a positive impact on patient care by facilitating early doctor-patient communication and initiating prompt clinical responses to delay the progression of complications associated with diabetes. This finding is similar to that of other research that AI-enabled CDSSs had a positive impact on patient-provider encounters and shared decision-making [25,26]. Therefore, appropriate use of risk prediction could enable clinicians to take early proactive measures to reduce the risk of developing diabetic complications, ultimately reducing the health care costs associated with diabetes [27,28].

Despite the perceived merits of AI-generated risk scores, the absence of clear frameworks (or the scientific basis from which recommendations were derived) limited the interpretability and usability of the risk scores and subsequent follow-up actions. This has been similarly identified in the literature as a key hindrance to clinical adoption [29,30]. As knowledge is deciphered differently based on personal experience and beliefs, the interpretation of scores could be dependent on the subjective attitudes of clinicians in decision-making [31]. Indeed, recent research indicates that the varying levels of knowledge and self-reported behavior among clinicians affect their approach in clinical practice, leading to potential noncompliance with the system recommendations [13]. Furthermore, as shown in our study, some clinicians chose to abstain from using the APA tool entirely because of their lack of trust in the quality of model inputs and parameters, as well as their concerns regarding the logic behind the AI outputs, often referred to as the “black box” situation [32,33]. To ensure a successful expansion of the APA tool within the clinical ecosystem, more effort should be directed to obtain a better comprehension of clinicians regarding the AI technology’s capabilities and the use of explainable frameworks to enhance transparency and clinician engagement [34-36].

As with the literature, clinicians in our study cautioned against being overly reliant on the APA tool, as occasional erroneous recommendations generated by the systems might prompt users to override a correct decision they have already made [26,37]. When users are subjected to automation bias, a tendency to overaccept system recommendations as a heuristic replacement of vigilant information processing [38] and medical errors ensue from following incorrect recommendations. Not only does it predispose patients to even greater harm, but it also diminishes the intention of using AI-enabled CDSSs [39]. Our results underscore the importance of collaborative intelligence, where users and AI work synergistically to enhance patient care. The human-in-the-loop concept suggests that while human oversight is active, overdependence on AI-enabled CDSSs is equally harmful. The optimal approach involves granting clinicians full control over the decision-making process while using AI to offer recommendations and inputs [40]. Clinical decisions, therefore, cannot be made without active involvement from clinicians to serve as gatekeepers, prevent negligence, and ensure patient safety. The seemingly conflicting recommendations identified in this study should be viewed as a catalyst that prompts critical thinking, and more effort should be made to confront meaningful disagreements. Also, encouraging clinicians to check on discrepancies may enhance their confidence in decision-making [41].

Finally, a significant obstacle that hindered the adoption of APA tool pertains to concerns surrounding accountability and liability, which is in line with the literature [36,42]. While ethical considerations regarding the use of AI persist, establishing well-defined clinical standards and codes of conduct for adopting APA tools can foster a culture of shared responsibility, moving away from a single form of attribution of responsibility. In addition, instead of focusing on assigning blame, it would be more constructive to acknowledge and commend clinicians’ efforts in integrating the outputs of AI-enabled CDSSs into their decision-making process, as long as the adoption of recommendations adheres to clinical standards, legal obligations, and ethical principles. This approach would motivate clinicians to embrace the most advanced medical technologies available in clinical practice, even if it means having to make a judgment call.

Collectively, the findings underscore the promising impact of adopting the APA tool within clinical settings and its potential to usher in notable enhancements in health policy. While the rapid integration of AI-based CDSSs in health care has presented promising potential for improved patient outcomes and streamlined clinical workflows, the persistent liability concerns among clinicians have created a barrier to the widespread adoption of these advanced technologies. With clinicians ultimately bearing the responsibility for any medical negligence, even after consulting with AI-enabled CDSS recommendations, there arises an urgent need for a comprehensive medicolegal framework. Such a framework must emphasize the allocation of liability among users, while also ensuring transparency in the decision-making processes of these AI tools [43]. For instance, the policy should delineate clear protocols for the documentation of AI-based recommendations, ensuring that the decision-making process is well-documented and easily accessible for medicolegal reviews. In addition, it is crucial to establish standardized protocols for the continuous evaluation and improvement of AI algorithms to minimize the risk of errors and improve the accuracy of recommendations. Creating an environment that fosters trust in AI technologies through a robust medicolegal framework will ultimately encourage clinicians to embrace these tools, leading to enhanced health care delivery and patient outcomes.

Limitations

This study provides valuable insights into the benefits and adoption challenges of an AI-based CDSS in its early stage of implementation. This study has some limitations. The study participants were limited to endocrinologists in a tertiary hospital; therefore, the generalizability of the findings to other health care settings may be limited. The sample size of the study is small, which may hinder the generation of comprehensive insights that better represent the broader context. As adoption of AI technology in clinical settings is still in its early stage, assembling a large cohort of clinicians for an in-depth analysis of AI-enabled CDSS implementation can be challenging due to the limited number of early adopters. Nevertheless, our findings shed light on the initial experiences and perceptions of a key group of clinicians, offering a foundation for future research and more extensive investigations. Our study’s sample size also aligns with the systematic review, which found that empirical studies, especially those with homogenous populations and narrowly defined objectives, typically achieve data saturation with 9-17 interviews [44]. Further research is needed to explore the use and impact of the APA tool in different clinical settings, such as primary care. As suggested by our participants, the use of the APA tool can be particularly beneficial to general practitioners who are responsible for managing a wide range of conditions and require access to a breadth of knowledge base across various specialty areas. Despite early findings on the APA tool’s use and adoption challenges, its long-term impacts on clinical and economic outcomes remain unknown. A subsequent larger evaluation is warranted to compare the APA tool with a standard of care. Finally, we did not explore the perspectives of patients with diabetes as the important end users of the APA tool.

Incorporating their perspectives may have contributed to a richer understanding.

Conclusions

AI-enabled CDSSs, such as the APA tool, has the potential to enhance clinical practice and patient care. Clinicians found certain features such as AI algorithms on medication adjustment and complication risk predictions useful in managing patients with T2DM with comorbidities and facilitating doctor-patient communication. However, interpretability of the risk scores, concerns about overreliance and automation bias, and issues surrounding accountability and liability were commonly cited as challenges inhibiting the adoption and application of the APA tool in endocrinology clinical settings. Further work is required to address these concerns effectively to enhance the tool's acceptance and applicability in relevant contexts.

Acknowledgments

We thank all participants for their participation in this study.

Data Availability

The data that support the findings of this study are available upon reasonable request from the corresponding author. The data are not publicly available due to information that could compromise the privacy of research participants.

Authors' Contributions

SY and YMB were responsible for the conception, design, and the whole protocol of the study. SY oversaw the study. HG, PCL, HCT, MMT, DSTL, AK, CS, DC, DSS, SYTT, AJWW, CHMC, and ZW were responsible for the acquisition of study data. SY and HG were responsible for data analysis and interpretation of study data. SY and HG drafted the manuscript. All authors critiqued the output and read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Er B, James L, Ma S, Ministry of Health. The Burden of Disease in Singapore, 1990-2017. Singapore: Ministry of Health; 2019.
2. Forbes JM, Cooper ME. Mechanisms of diabetic complications. Physiol Rev 2013;93(1):137-188 [FREE Full text] [doi: 10.1152/physrev.00045.2011] [Medline: 23303908]
3. Papatheodorou K, Banach M, Bekiari E, Rizzo M, Edmonds M. Complications of diabetes 2017. J Diabetes Res 2018;2018:3086167 [FREE Full text] [doi: 10.1155/2018/3086167] [Medline: 29713648]
4. Crowley K. Patient Education: The ABCs of Diabetes (The Basics). UpToDate. URL: https://www.uptodate.com/contents/the-abcs-of-diabetes-the-basics?search=abc%20diabetes&source=search_result&selectedTitle=1~150&usage_type=default&display_rank=1 [accessed 2024-05-23]
5. American Diabetes Association. Standards of medical care in diabetes—2016 abridged for primary care providers. Clin Diabetes 2016;34(1):3-21 [FREE Full text] [doi: 10.2337/diaclin.34.1.3] [Medline: 26807004]
6. Sim LLW, Ban KHK, Tan TW, Sethi SK, Loh TP. Development of a clinical decision support system for diabetes care: a pilot study. PLoS One 2017;12(2):e0173021 [FREE Full text] [doi: 10.1371/journal.pone.0173021] [Medline: 28235017]
7. Chen R, Jiang HQ, Huang C, Bau C. Clinical decision support system for diabetes based on ontology reasoning and TOPSIS analysis. J Healthc Eng 2017;2017:4307508 [FREE Full text] [doi: 10.1155/2017/4307508] [Medline: 29312655]
8. Jia P, Zhao P, Chen J, Zhang M. Evaluation of clinical decision support systems for diabetes care: an overview of current evidence. J Eval Clin Pract 2019 Feb;25(1):66-77. [doi: 10.1111/jep.12968] [Medline: 29947136]
9. Chien T, Ting H, Chen C, Yang C, Chen C. A clinical decision support system for diabetes patients with deep learning: experience of a Taiwan medical center. Int J Med Sci 2022;19(6):1049-1055 [FREE Full text] [doi: 10.7150/ijms.71341] [Medline: 35813300]
10. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020;3:17 [FREE Full text] [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
11. Contreras I, Vehi J. Artificial intelligence for diabetes management and decision support: literature review. J Med Internet Res 2018;20(5):e10775 [FREE Full text] [doi: 10.2196/10775] [Medline: 29848472]
12. Spänig S, Emberger-Klein A, Sowa J, Canbay A, Menrad K, Heider D. The virtual doctor: an interactive clinical-decision-support system based on deep learning for non-invasive prediction of diabetes. Artif Intell Med 2019;100:101706. [doi: 10.1016/j.artmed.2019.101706] [Medline: 31607340]

13. Nimri R, Tirosh A, Muller I, Shtrit Y, Kraljevic I, Alonso MM, et al. Comparison of insulin dose adjustments made by artificial intelligence-based decision support systems and by physicians in people with type 1 diabetes using multiple daily injections therapy. Diabetes Technol Ther 2022;24(8):564-572. [doi: 10.1089/dia.2021.0566] [Medline: 35325567]
14. Bajgain B, Lorenzetti D, Lee J, Sauro K. Determinants of implementing artificial intelligence-based clinical decision support tools in healthcare: a scoping review protocol. BMJ Open 2023;13(2):e068373. [doi: 10.1136/bmjopen-2022-068373] [Medline: 36822813]
15. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
16. Sun X, Bee YM, Lam SW, Liu Z, Zhao W, Chia SY, et al. Effective treatment recommendations for type 2 diabetes management using reinforcement learning: treatment recommendation model development and validation. J Med Internet Res 2021;23(7):e27858 [FREE Full text] [doi: 10.2196/27858] [Medline: 34292166]
17. Wan Q, Makeham M, Zwar NA, Petche S. Qualitative evaluation of a diabetes electronic decision support tool: views of users. BMC Med Inform Decis Mak 2012;12:61 [FREE Full text] [doi: 10.1186/1472-6947-12-61] [Medline: 22759239]
18. Pratt R, Saman DM, Allen C, Crabtree B, Ohnsorg K, Sperl-Hillen JM, et al. Assessing the implementation of a clinical decision support tool in primary care for diabetes prevention: a qualitative interview study using the Consolidated Framework for Implementation Science. BMC Med Inform Decis Mak 2022;22(1):15 [FREE Full text] [doi: 10.1186/s12911-021-01745-x] [Medline: 35033029]
19. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
20. Foo KM, Sundram M, Legido-Quigley H. Facilitators and barriers of managing patients with multiple chronic conditions in the community: a qualitative study. BMC Public Health 2020;20(1):273 [FREE Full text] [doi: 10.1186/s12889-020-8375-8] [Medline: 32106838]
21. McEwen LN, Herman WH. Health care utilization and costs of diabetes. In: Cowie CC, Casagrande SS, Menke A, Cissell MA, Eberhardt MS, Meigs JB, editors. Diabetes in America. Bethesda, MD: National Institute of Diabetes and Digestive and Kidney Diseases; 2018.
22. Disdier-Flores OM, Rodríguez-Lugo LA, Pérez-Perdomo R, Pérez-Cardona CM. The public health burden of diabetes: a comprehensive review. P R Health Sci J 2001;20(2):123-130. [Medline: 11561471]
23. Deshpande AD, Harris-Hayes M, Schootman M. Epidemiology of diabetes and diabetes-related complications. Phys Ther 2008;88(11):1254-1264 [FREE Full text] [doi: 10.2522/ptj.20080020] [Medline: 18801858]
24. Sagoo MK, Gnudi L. Diabetic nephropathy: an overview. Methods Mol Biol 2020;2067:3-7. [doi: 10.1007/978-1-4939-9841-8_1] [Medline: 31701441]
25. Labinsky H, Ukalovic D, Hartmann F, Runft V, Wichmann A, Jakubcik J, et al. An AI-powered clinical decision support system to predict flares in rheumatoid arthritis: a pilot study. Diagnostics (Basel) 2023;13(1):148 [FREE Full text] [doi: 10.3390/diagnostics13010148] [Medline: 36611439]
26. Romero-Brufau S, Wyatt K, Boyum P, Mickelson M, Moore M, Cognetta-Rieke C. A lesson in implementation: a pre-post study of providers' experience with artificial intelligence-based clinical decision support. Int J Med Inform 2020;137:104072. [doi: 10.1016/j.ijmedinf.2019.104072] [Medline: 32200295]
27. Lloyd A, Sawyer W, Hopkinson P. Impact of long-term complications on quality of life in patients with type 2 diabetes not using insulin. Value Health 2001;4(5):392-400 [FREE Full text] [doi: 10.1046/j.1524-4733.2001.45029.x] [Medline: 11705130]
28. Trikkalinou A, Papazafiropoulou AK, Melidonis A. Type 2 diabetes and quality of life. World J Diabetes 2017;8(4):120-129 [FREE Full text] [doi: 10.4239/wjd.v8.i4.120] [Medline: 28465788]
29. Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, et al. Opportunities and obstacles for deep learning in biology and medicine. J R Soc Interface 2018;15(141):20170387 [FREE Full text] [doi: 10.1098/rsif.2017.0387] [Medline: 29618526]
30. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 2019;1(5):206-215 [FREE Full text] [doi: 10.1038/s42256-019-0048-x] [Medline: 35603010]
31. Laws RA, Kemp LA, Harris MF, Davies GP, Williams AM, Eames-Brown R. An exploration of how clinician attitudes and beliefs influence the implementation of lifestyle risk factor management in primary healthcare: a grounded theory study. Implement Sci 2009;4:66 [FREE Full text] [doi: 10.1186/1748-5908-4-66] [Medline: 19825189]
32. Amann J, Blasimme A, Vayena E, Frey D, Madai VI, Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak 2020;20(1):310 [FREE Full text] [doi: 10.1186/s12911-020-01332-6] [Medline: 33256715]
33. Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. JAMA 2018;320(21):2199-2200. [doi: 10.1001/jama.2018.17163] [Medline: 30398550]
34. Cutillo CM, Sharma KR, Foschini L, Kundu S, Mackintosh M, Mandl KD, MI in Healthcare Workshop Working Group. Machine intelligence in healthcare-perspectives on trustworthiness, explainability, usability, and transparency. NPJ Digit Med 2020;3:47 [FREE Full text] [doi: 10.1038/s41746-020-0254-2] [Medline: 32258429]

35. Reddy S. Explainability and artificial intelligence in medicine. Lancet Digit Health 2022;4(4):e214-e215 [FREE Full text] [doi: 10.1016/S2589-7500(22)00029-2] [Medline: 35337639]
36. Jones C, Thornton J, Wyatt J. Artificial intelligence and clinical decision support: clinicians' perspectives on trust, trustworthiness, and liability. Med Law Rev 2023;31(4):501-520 [FREE Full text] [doi: 10.1093/medlaw/fwad013] [Medline: 37218368]
37. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. J Am Med Inform Assoc 2012;19(1):121-127 [FREE Full text] [doi: 10.1136/amiajnl-2011-000089] [Medline: 21685142]
38. Skitka L, Mosier K, Burdick M. Does automation bias decision-making? Int J Hum Comput Stud 1999;51(5):991-1006 [FREE Full text] [doi: 10.1006/ijhc.1999.0252]
39. Mosier K, Skitka L. Human Decision Makers and Automated Decision Aids: Made for Each Other?. Mahwah, NJ: Erlbaum; 1996:201-220.
40. Haselager P, Schraffenberger H, Thill S, Fischer S, Lanillos P, van de Groes S, et al. Reflection machines: supporting effective human oversight over medical decision support systems. Camb Q Healthc Ethics 2023:1-10. [doi: 10.1017/S0963180122000718] [Medline: 36624620]
41. Muhiyaddin R, Abd-Alrazaq AA, Househ M, Alam T, Shah Z. The impact of clinical decision support systems (CDSS) on physicians: a scoping review. Stud Health Technol Inform 2020;272:470-473. [doi: 10.3233/SHTI200597] [Medline: 32604704]
42. Funer F, Liedtke W, Tinnemeyer S, Klausen AD, Schneider D, Zacharias HU, et al. Responsibility and decision-making authority in using clinical decision support systems: an empirical-ethical exploration of German prospective professionals' preferences and concerns. J Med Ethics 2023;50(1):6-11 [FREE Full text] [doi: 10.1136/jme-2022-108814] [Medline: 37217277]
43. Garell C, Svedberg P, Nygren JM. A legal framework to support development and assessment of digital health services. JMIR Med Inform 2016;4(2):e17 [FREE Full text] [doi: 10.2196/medinform.5401] [Medline: 27226391]
44. Hennink M, Kaiser B. Sample sizes for saturation in qualitative research: a systematic review of empirical tests. Soc Sci Med 2022;292:114523 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114523] [Medline: 34785096]

Abbreviations

AI: artificial intelligence
APA: AI-enabled Prescription Advisory
CDSS: clinical decision support system
COREQ: Consolidated Criteria for Reporting Qualitative Research
EMR: electronic medical record
FGD: focus group discussion
T2DM: type 2 diabetes mellitus

Edited by A Kushniruk; submitted 17.07.23; peer-reviewed by D Heider, N Minian; comments to author 03.11.23; revised version received 07.11.23; accepted 05.05.24; published 13.06.24.

Please cite as:
Yoon S, Goh H, Lee PC, Tan HC, Teh MM, Lim DST, Kwee A, Suresh C, Carmody D, Swee DS, Tan SYT, Wong AJW, Choo CHM, Wee Z, Bee YM
Assessing the Utility, Impact, and Adoption Challenges of an Artificial Intelligence–Enabled Prescription Advisory Tool for Type 2 Diabetes Management: Qualitative Study
JMIR Hum Factors 2024;11:e50939
URL: https://humanfactors.jmir.org/2024/1/e50939
doi:10.2196/50939
PMID:38869934

Original Paper

Preferences, Needs, and Values of Patients With Chronic Obstructive Pulmonary Disease Attending a Telehealth Service: Qualitative Interview Study

Camilla Wong Schmidt[1,2], MSc; Karen Borgnakke[2], MA, PhD, DHA; Anne Frølich[3,4], MD, PhD; Lars Kayser[2], MD, PhD

[1]Medical Department, Holbæk Sygehus, Region Zealand, Holbæk, Denmark

[2]Section of Health Services Research, University of Copenhagen, Copenhagen, Denmark

[3]Innovation and Research Centre for Multimorbidity, Slagelse Hospital, Region Zealand, Slagelse, Denmark

[4]Department of Public Health, University of Copenhagen, Copenhagen, Denmark

Corresponding Author:
Camilla Wong Schmidt, MSc
Medical Department
Holbæk Sygehus
Region Zealand
Smedelundsgade 60
Holbæk, 4300
Denmark
Phone: 45 59484000
Email: cws@sund.ku.dk

Abstract

Background: Digitally assisted health care services and technologies are gaining popularity. They assist patients in managing their conditions, thereby reducing the burden on health care staff. Digital health care enables individuals to receive care that is more tailored to their needs and preferences. When implemented properly, it can promote equity by considering each person's opportunities and limitations in the context of health care needs, preferences, values, and capabilities.

Objective: This study aims to understand the needs, values, and preferences of individuals with chronic obstructive pulmonary disease (COPD) who are provided with a 24/7 digital health care service. Furthermore, we aim to understand the dynamics of the communities to which they belong and how these communities intersect. This will provide us with the essential knowledge to establish new methods of providing education, including the development of educational activities for health professionals to engage, train, and empower people living with COPD.

Methods: The study included 7 informants diagnosed with COPD who received 24/7 digital health care service support from a regional project in Region Zealand, Denmark. The informants were visited 4 times during 2 months, including a "Hello" visit, a day with a semistructured interview, and 2 days with field observations. The informants participated in a semistructured interview, following participant observation and an ethnographic approach. The interview content was analyzed using an inductive methodology to categorize the empirical data.

Results: Using the inductive approach, we identified 3 main categories related to the informants' needs, values, and preferences: (1) Health, (2) Value Creation, and (3) Resources. These 3 main categories were based on 9 subcategories: (1) health and barriers, (2) self-monitoring, (3) medication, (4) behavior, (5) motivation, (6) hobbies, (7) social networks, (8) health professionals, and (9) technology. These findings revealed that the informants placed value on maintaining their daily activities and preserving their sense of identity before the onset of COPD. Furthermore, they expressed a desire not to be defined by their COPD, as conversations about COPD often shifted away from the topic.

Conclusions: Digital health solutions and the health care professionals who offer them should prioritize the individuals they serve, considering their needs, values, and preferences rather than solely focusing on the medical condition. This approach ensures the highest level of daily living and empowerment for those living with long-term health conditions. The communities surrounding individuals must engage in constant interaction and collaboration. They should work together to incorporate people's needs, values, and preferences into future digital health services, thereby promoting empowerment and self-management. New educational programs aimed at developing the digital health service competencies of registered nurses should facilitate collaboration between the 2 communities. This collaboration is essential for supporting patients with long-term health conditions in their daily activities.

(JMIR Hum Factors 2024;11:e53131) doi:10.2196/53131

KEYWORDS

people with long-term health condition; patient education; COPD; digital health; ethnography; inductive; ethnographic; chronic; lung; lungs; pulmonary; respiratory; self-management; interview; interviews; qualitative; experience; experiences; attitude; attitudes; opinion; perception; perceptions; perspective; perspectives; acceptance

Introduction

Background

In recent years, a transformation has been occurring with the increased use of digitally assisted health care services and technologies. These advancements aim to reduce the burden on the health care work force by enabling patients to better manage their own conditions [1]. Digitally based health care offers an opportunity for personalized and tailored health care services that better meet the needs of individual patients. Digitalization can also help reduce inequities if introduced thoughtfully, with an awareness of both the opportunities and barriers for individuals, considering their health care needs, preferences, values, and capabilities [1].

When health care professionals have the appropriate knowledge about these factors and are trained to address them, it can facilitate meaningful conversations and better connections with those they serve. This, in turn, can increase patient motivation and ease their access to digital services [2,3]. This necessitates educational programs for health professionals that focus on understanding their patient's needs and values, capability, person-centered services, self-management, and communities surrounding the patient. These programs should be based on evidence from empirical data obtained through interviews and observations of individuals with firsthand experience using digitally enabled health care services [4].

Digital Health Care and Chronic Obstructive Pulmonary Disease

Chronic obstructive pulmonary disease (COPD) is a leading cause of death, responsible for over 3 million deaths worldwide in 2019 [5]. In Denmark, 3355 individuals died from COPD in 2020, making it the second most common cause of death in the nation [6]. As a result of the progressive loss of pulmonary function, people with COPD experience impairments in their daily activities. These impairments can inhibit mobility, leading to a sedentary lifestyle [7]. The Global Initiative for Chronic Obstructive Lung Disease (GOLD), established in 1998, has developed a set of recommendations for managing COPD. Evidence shows that self-management improves outcomes for patients with COPD and reduces the likelihood of hospitalization [8,9]. The recommendations by GOLD also address how increased self-management can help motivate and engage people, leading to positive adaptations in their health behaviors.

Digital health monitoring is one of the most widely used tools for self-management of COPD and other long-term health conditions (LTHCs) and appears to reduce the risk of both hospitalization and acute visits [10]. In 2015, a randomized controlled study was conducted with 100 people with COPD, with 48 randomly selected for home monitoring and 52 for usual care. The study found that people with COPD who used a home monitoring kit for 6 months had improved health-related quality of life and reduced anxiety scores compared with people who received usual care. The study also showed that people with home monitoring kits had fewer and shorter hospitalizations than those receiving usual care [11]. Furthermore, people with COPD who participate in telemedicine-based interventions feel safer, more empowered, and more in control of their own disease [11].

People with LTHCs who possess an enhanced ability to manage their condition themselves and experience a higher level of empowerment tend to have a reduced risk for hospitalization [12]. A likely explanation for this can be found in a qualitative study from 2017, which identified 3 main themes for individuals participating in a telemedicine intervention: (1) a sense of improved security and control; (2) a better understanding of their disease; and (3) the benefits of virtual conversations [13]. However, these studies do not take place within person-centered digital health communities and do not address the significance of the participants' sociodemographic characteristics. Evidence of sociodemographic characteristics is important, as research indicates that resourceful individuals benefit more from available health care, and those with an interest in technology may derive greater benefits from digital health solutions [14]. In addition, the prevalence of COPD and other chronic conditions, such as ischemic heart disease and type 2 diabetes mellitus, is higher in areas with populations characterized by lower sociodemographic status than in the average population [15,16]. This may contribute to the risk of inequity, as lower sociodemographic status often correlates with both lower levels of education and lower digital health literacy [15,17].

Educational Programs for the Digital Transformation

Another problem often overlooked concerning disadvantaged individuals living with 1 or more LTHCs is their reluctance to attend traditional educational services structured around scholastic planning. To reduce inequity, there is a need to develop new approaches to include this segment of the population, utilizing educational methods that are not scholastic or built on classical teaching methods such as classroom-based education. In response to this need, we have initiated a project where, based on an ethnographic approach involving interviews, observations, and the cocreation of educational materials, we will develop a new methodology inspired by social learning theory [18,19]. To provide guidance for the development of new educational programs and curricula tailored for digital transformation, we have examined, on an individual level, the preferences, values, and needs of people with COPD in the context of unrestricted access to a person-centered, digitally enabled health service available 24/7. Informed by social learning theories, we have also explored the existence of communities these individuals are part of in relation to their everyday lives and their ability to identify potential sources of support.

The purpose of this study is to gather essential information about individuals living with COPD within the context of accessing support from a regular 24/7 digital health care service and the dynamics of the communities to which they belong. This information will empower us to design new methods for providing education, including the development of educational activities for health professionals. These activities will enable professionals to effectively engage, train, and empower people with COPD.

This has led us to the following research questions:

- Research question 1: What matters for people with COPD with respect to their needs, values, and preferences in the context of using a 24/7 digital health service?
- Research question 2: What is the role of the health care and social networks, respectively, and how are these a potential part of communities formed around the patient's health condition?

Methods

Design

This report is a part of a larger PhD project and constitutes the first of 4 articles. The overarching aim of this work is to obtain insights into the lives of people with COPD, supported by a 24/7 digital health service, and to use this information to develop a patient-case–based curriculum to educate health professionals on effectively engaging, enabling, and empowering individuals living with COPD.

The first study reported here was conducted from August 1, 2020, to January 31, 2021. This period coincided with the COVID-19 pandemic in Denmark before vaccinations were introduced to protect against severe cases. The study is qualitative in nature, inspired by ethnographic research methods, and includes semistructured interviews and field observations [20]. The field study involved visiting the homes of people with COPD 4 times over 2 months. The visits included a preliminary "Hello" visit, a day dedicated to a semistructured interview, and 2 days focused on field observations. This article focuses on the data and results gathered during the second visit, which involved conducting the semistructured interview.

Context

The digital health service utilized in this study is provided by an innovation project in Region Zealand, Denmark, called PreCare. This project is built on the Epital Care Model (ECM) [21,22]. The ECM, developed in 2016, offers a 24/7 digital health care service where individuals with LTHCs monitor their own health with the assistance of nurses from a response and coordination center (RCC). The ECM consists of 6 stages: citizens with unknown LTHCs, active and independent living, virtual assisted living, virtual assisted living with support from health care professionals, outpatient care at home, and admission to a local health clinic or hospital. It serves as a template for digital health services based on patients' medical needs [22].

In total, approximately 400 participants with COPD or ischemic heart disease were enrolled in the PreCare project over a 4-year period. At any given time, there are approximately 150 participants. Each participant was provided with a tele-home monitoring kit, which included a tablet. Additionally, for participants with COPD, the kit included a spirometer, a thermometer, a pulse oximeter, and a box containing acute medicine. The participants were supported by an RCC, which was staffed with registered nurses (RNs) and an eDoctor. According to the project protocol, participants monitored their condition daily. They could always call the RCC to discuss their condition, and the RCC regularly initiated contact to ensure participants felt safe and confident. During these conversations, self-management was also supported (C Schmidt, MSc, personal communication, 2020). In the event of deterioration, the tablet would indicate a yellow or red code and send a message to the RCC. The nurses would then respond to the code and call the participant to follow-up on the reported condition. If needed, participants can take medicine from the box to treat exacerbations. In cases of further need, the nurses would contact the eDoctor [23].

Informants

The PreCare project initially provided a list of 15 participants diagnosed with COPD, each with varying degrees of severity, all of whom expressed interest in participating in the research related to the PreCare project. Subsequently, over a 4-month period, 10 of these participants were contacted by phone, selected from the top of the list. After receiving oral information about the project, 8 of these agreed to participate. However, 2 participants were not interested and the remaining 5 were not contacted as the recruitment period had exceeded. Furthermore, 1 potential informant expressed disinterest in participating after the initial meeting. The selected number of informants was determined by the limitations of the study design. The informants were invited in 3 separate periods: 3 informants were invited from August to September for the first period, 2 informants were invited from October to November for the second period, and 2 more were invited from December to January for the third period. In total, 7 informants were recruited. After obtaining oral consent via phone, further information was sent by email to 3 informants, while the other 4 did not require this. After 1 week, all 7 informants were contacted again by phone to schedule the first in-person meeting, which took place within 1-2 weeks. This study's inclusion and exclusion criteria followed the PreCare protocol [23].

Data Collection

To establish an emotional and trustworthy relationship with the informants, we scheduled a visit to their homes (the initial visit). This approach aimed to strengthen the connection between the researcher and the informants, fostering an informal and friendly atmosphere during the interview. To conduct the semistructured interview, we used an interview guide inspired by Spradley's [20] ethnographic interview techniques. The interview was conducted in a friendly and casual approach, allowing the informants to share their experiences and discuss their everyday lives with a chronic condition as they deemed appropriate [20,24]. All interviews were conducted in the informants' homes and were audio-recorded with their consent. As a result of the informants' background, the interviews were conducted in Danish, and only quotes were later translated into English by the first author (CWS). All informants participated in the interviews; 1 participant had his spouse present during the interview.

Interview Guide

The interview guide was developed based on sociotechnical ecosystem thinking, our concept of technology readiness, and an attempt to identify how individuals belong to 1 or more communities, inspired by social learning theory (Multimedia Appendix 1) [25-28]. We conducted the interview with an open-minded approach, including "how" questions, to enable the informants to respond as they found suitable. The interview guide was structured around 6 thematic areas: daily activities, health, measurements, communities, RCC and PreCare, and literacies and digital literacies. For each of the thematic areas, we included 1 main question and underlying questions to sustain the conversation throughout the interview. For example, the theme "daily activities" included the main question: "Can you tell me how a typical day is for you?" In the theme "health," the main question was "Can you tell me how COPD has affected your life?." The 6 themes were defined by the authors and were written in Danish.

Data Analysis

The interviews were conducted, transcribed, and analyzed by the author CWS. The transcripts were analyzed using content analysis, a method for systematically and objectively describing and quantifying phenomena. An inductive approach was used, beginning with open coding to create categories, followed by abstraction to generate main categories [29]. A 3-step content analysis was used to identify the main categories.

Analysis of Interviews

Each interview was transcribed and carefully reviewed to understand the context of the data. Subsequently, the transcripts were uploaded and coded using NVivo 12 (QSR International) [30] by CWS. Over 700 codes were identified and categorized into 66 subcategories. These subcategories were then merged to create an affinity diagram initially using paper and later repeated using NVivo. This process resulted in 9 categories, each containing 4-12 subcategories, respectively. The category "Self-Monitoring" had the fewest subcategories, while "Health Professionals and Social Network" had the most subcategories. The 9 categories were analyzed by CWS and the last author (LK) to synthesize the data into 3 main categories. CWS, who holds an MSc degree in health informatics and has been educated in qualitative methods, collaborated with LK, a professor in health service research with experience in both qualitative and quantitative analyses, for this process.

The 3 main categories identified were health, value creation, and resources (Textbox 1 and Multimedia Appendix 2). The category of health consisted of 3 subcategories: health and barriers, medication, and measurements. The category of value creation was formed from hobbies, behavior, and motivation. The category of resources was merged from 3 subcategories: social networks, health professionals, and technology. In our analysis, we paid particular attention to what matters to people with COPD, supported by the theories upon which the interview guide was built.

Textbox 1. Overview of the 3 main categories and subcategories.

1. Health

- Health and barriers
- Self-monitoring
- Medication

2. Value Creation

- Behavior
- Motivations
- Hobbies

3. Resources

- Social network
- Health professionals
- Technology

Ethical Consideration

Information regarding the study, partnerships, and data handling complies with the Helsinki Declaration and was communicated to the informants in both written and oral forms. They were informed that their participation was voluntary and anonymous and that they could revoke their consent at any time. Furthermore, they were assured that their involvement would not prevent them from participating in the PreCare project. All consent was obtained before the interview, through the signing of a consent form. The Danish National Center for Ethics was not required to approve the study as no biological material was used. Any data obtained from the informants were treated as personal health information and handled in accordance with Danish legislation (General Data Protection Regulation [GDPR]) and securely stored on drives. Health science questionnaire surveys and interview studies that do not involve human biological material [section 14(2) of the Danish Act on Committees] do not require reporting or approval from the Danish National Centre for Ethics [31].

Results

Characteristics of the Informants

A total of 4 men and 3 women participated in the interviews (age range 52-81 years). Two informants lived with their spouses. Despite having had COPD for an average of more than 2 years, the severity of each participant's COPD varied. Some participants continued to smoke daily despite being aware of the health risks. One male participant was unable to provide information to categorize his level of education, 2 had only completed elementary school, while 4 had completed higher education. There was no evidence of their usage of technology, such as websites and participation in online communities, in relation to their medical concerns. All informants had been included in the PreCare Project for more than 6 months.

The Three Main Categories

The main categories and subcategories identified in the content analysis provide insight into and offer a comprehensive understanding of the daily life situations and experiences of the informants living with COPD. Upon reviewing these categories, attention is drawn to both the specific consequences of a COPD diagnosis and how practical hurdles and activity levels are affected in the daily lives of the informants. These impacts are described in the interviews as limitations on activities the informants were accustomed to participating in, as well as a determination to carry out specific household duties despite a decreased energy level. The duality between "restrictions" and "experimental salvage" is evident in the category of activity but is also observed when interviews approach questions such as self-monitoring. Here, they take on different meanings, tasked with reclaiming self-discipline and control on one hand, while also being concerned that daily measurements can serve as a reminder of one's limitations, akin to "being reminded of having a chronic disease." Thus, through the interviews, it becomes apparent how the informants encounter difficulties and impediments in carrying out daily tasks due to their condition. In everyday life, this translates to tasks that were once feasible but now being difficult or impossible to complete. The distinction between "then" and "now" is frequently referenced, highlighting the contrast between the condition "before I got COPD" and "the situation as a chronic."

Health

Health and Barriers

This category describes the experience of living with COPD, detailing how it has impacted daily life and outlining the physical and mental barriers experienced throughout the day or in general.

The informants did not express interest in delving deeper into their everyday lives with COPD. Instead, they prioritized discussing other aspects of their daily life or past experiences. They responded quickly to questions about COPD and then redirected the conversation toward other topics. This deliberate redirection indicated their reluctance to discuss their chronic illness.

> *Interviewer: ...Can you tell us how your diagnosis has affected your life?*
>
> *M3: So, I'm crushed. One positive thing is that I had to sell my motorcycle and all my stuffs, I used to gather a lot. We had a 400 kvm house with basement and ceiling, which was filled with enamel sign, books, magazines, tech cars and bicycles...*

The informant swiftly and effectively shifts his focus away from negative thoughts about COPD's interference and begins discussing his previous interest in used objects. He demonstrates a clear desire not to dwell on the negative aspects of his existence, opting instead to redirect the conversation toward something positive and reminiscent of happier times.

The limitations imposed by COPD forced the informants to forego certain daily activities, some of which could have contributed to an improved quality of life. The frequent shortness of breath and coughing prevented them from engaging in activities such as walking outside, performing household duties, or general personal care needs.

> *Interviewer: ...but is there other things COPD had done, that you cannot do anymore, completely?*
>
> *M1: Well, I cannot go to the city and get me a cup of coffee at the street restaurant.*
>
> *Interviewer: No, that's true...*
>
> *M1: I'm not even sure I'd be able to go to the garbage cans anymore (coughing), but when my friend comes and the weather is good (...) he drives me in that wheelchair over there, and then we sit together and drink a cup of coffee and talk...*

The informant's worry about his capacity to take out the trash underscores the profound impact that COPD has on his life, to the extent that he feels unable to leave the house without assistance. Conversations with the informant were replete with stories where social interactions played a significant role.

For some individuals, participating in a community became challenging due to shortness of breath caused by COPD. Additionally, for others, COPD had led to the complete exclusion of previous acquaintances.

> *M3 wife: In return, you have thought about how many of them you have helped (...), you don't really hear from them anymore, because now you can't help them anymore.*
>
> *M3: Yes, there are many of those whom I have been calling, "Great that we are talking to you, we were just thinking of you, by the way we have a locker that*

doesn't work". You never heard from these people again, and I have been discussing this with others, and it is true...

The informant noted that his inability to visit friends anymore, coupled with their failure to reciprocate, has made it increasingly difficult for him to maintain relationships with them. This situation has surprised him, particularly because he is no longer able to provide assistance, as reported by his wife.

Self-Monitoring

This category highlights how the informants manage and self-monitor their condition. Furthermore, it explores how the outcomes of their monitoring efforts may impact their day and their motivation to engage in activities.

In the informants' descriptions of their daily lives, the topic of self-monitoring for the PreCare project was not initially mentioned. It was only during the conversation around this subject that the activity itself was explained and, in some cases, mentioned.

M1: Yes but, it's not interesting, no(...) and then I hope in the end it can help other people too. So, I take it with pleasure, but I could still think of something more exciting things to do...

W1: Yes, but they have changes it (pause), I'm just going to write something today, I'm not in for it. It's not correct anyway (temperature)

The self-monitoring is described here as uninteresting, with 1 informant considering it a waste of time because he could find more engaging activities to do during his challenging day. The second informant emphasizes the importance of accurate measurements for individuals to actively participate in self-monitoring.

The outcomes of self-monitoring had a significant impact on the informants. The results were displayed as 1 of 3 colors—red, yellow, or green—on the tablet's display. The meaning of the color had a tangible effect on the informant's day.

Interviewer: Yes, exactly when, but then how? Because now you said that you had a red measurement yesterday was it then a difficult day when you have a red measurement?

W2: Yes, that is a stupid day, at first the mood is going down, and I am going, well yes I usually get restless, because I can't, because a day like that, I am thinking about...is it now it's going in the wrong direction...

The informant faces difficulties getting through the day when the red color appears, disrupting their daily routine. Additionally, the informant begins to question whether their COPD-related health is deteriorating or if the red color indicates a negative trend.

Conversely, the green color holds significance for the informant, particularly in contrast to the red color. Seeing a green measurement might enhance the informant's enjoyment of daily life, especially if it has been a while since they last saw a green result.

W1: It's green! (happy/excited)

W1: It haven been that for a long time....

Although there was considerable excitement surrounding the green measurement, its implications for the maintenance of the day remained unclear. However, the informant did clarify early in the conversation that she felt more motivated to venture outside into the garden on good days.

Medication

Being chronically ill entails the necessity of medication and its management which, for most of the informants, has become integrated into daily life. This category elucidates how the informants handle their medication and who supports them in managing it. The informants varied in their approaches to and understanding of medication, and the availability of help and support was crucial.

W2: Thus, those prescriptions, I also have one lying here, and this is the new medicine I got, and I don't understand it, because I should have asked about it.

W3: Yes, I have these blue folders, you probably don't know them, but those blue boxes for morning, midday, evening, and there is for (cough)...think there is for eight days, probably, that can be right? Eight days, I believe that, and I sort them every second week. I sit by the dining table, and line the whole thing up, and then I sort them.

The aforementioned examples depict 2 different scenarios of handling and understanding medicine. In the first scenario, one of the informants blames herself for not seeking information about the new medication when she first started taking it because she is unfamiliar with it. While she accepts responsibility for her medication, she still requires assistance in understanding it. By contrast, the second informant has established routines for managing her medication, and therefore, understands what she deposits into the pillbox.

As a result of errors and inconsistent care from municipal employees, the informants began to question their trust in the municipality's care team. Although the informants could receive assistance from the municipality with their prescriptions and medication management, they found that the assistance and knowledge provided by municipal employees lacked the necessary qualifications.

W2: ...but she wasn't, and then she made the mistake of repeating after the other, and I quickly notice it, and it's not, it is not calcium tablets we sit and play with.

This is a serious concern, as indicated by the informant's statement that the pills are not calcium supplements, and taking medication in incorrect amounts could have negative effects on her health. Therefore, it is vital that her medication is prescribed correctly for her condition.

When it comes to medication, there was a strong tendency among the informants to rely on the RCC nurses, especially during exacerbations. The RCC nurses use telephone communication to reach out to the informants and inquire about their health. If necessary, the nurses may advise the informants to take additional medication. The informants comply with the nursing advice and adjust their medication accordingly because they have a high level of trust in the digital nurses.

> *M2: Yes, "Nærklinikken" is the ones who change it now, yes, they just say you have to take two breaths in the morning, and they do it regularly if I have felt worse for a little while. I'll just get more, double up.*
>
> *M2: Yes, I feel very safe.*

When it comes to medication, the informants trust the guidance of the digital nurses because they feel it is their responsibility to adjust their medication. They feel secure knowing that others are assisting them and providing direction with their medication management.

When the RNs oversee the health condition of the informants, their independence in managing their chronic disease and their understanding of their medication do not seem to improve. It appears that the RNs are still somewhat paternalistic. However, the informants do experience a sense of safety, particularly when it comes to their health and medication.

Value Creation

Behavior

This category identifies the former daily routines that had to change or be excluded from the informants' lives because of COPD. Furthermore, it highlights the new routines that should be adapted because of the weakened ability caused by their condition.

The informants must adapt their daily routines to accommodate their diminished capacity because of COPD, necessitating the establishment of new habits. Consequently, they may take fewer walks, experience reduced appetite, or sleep longer than usual. This limitation often confines the informants to their residences. When queried about their daily lives, the informants provided a range of responses. Some spoke very briefly and exhibited a negative attitude, while others believed that obtaining a comprehensive understanding of how COPD impacts daily activities was crucial.

> *Interviewer: ...oh if you should tell me how a typical everyday looks for you K1, what do you do on a general day?*
>
> *W1: Sitting here*
>
> *Interview: You sit there*
>
> *W1: Yes (cough), but sometimes when I'm well, I go out in the garden.*

In this case, the informant primarily spends time sitting on the same couch and does not elaborate much on her everyday activities. She finds joy in moving outside and into the garden whenever possible. Previously, she engaged in various artistic activities and housework as part of her daily routine, but these tasks are no longer feasible due to her health.

Some informants expressed that it was still important to maintain cleanliness in their own homes. While the municipality provides cleaning assistance to the majority of the informants, some individuals still prefer to handle specific tasks on their own.

> *W2: ...I said to her, now don't think I'm crazy but I've been standing and ironed my bedsheets for several days, then she was about to faint (...). Only the elderly irons their bedsheets. I have always done it, and I will not stop doing it as long as I can stand on my feet.*

She continues to prioritize tasks such as making her bed and changing the sheets, as she has always done. Despite the challenges posed by her health, she decides to persist because these tasks hold significant importance for her. However, she acknowledges that it may take several days to complete them. By contrast, most of the informants expressed overall dissatisfaction with the cleaning assistance provided by the municipality, stating that they often had to make numerous corrections.

Another aspect they felt had changed because of their condition was the rhythm of the day. They found that getting out of bed in the morning was becoming more challenging, or they noticed that they were waking up earlier. This change could be attributed to their increased frequency of sleeping and reduced engagement in everyday activities.

> *M1: Yes. Well, but I wake up before Satan gets his shoes on, because I am used to doing something, and I cannot really more, so I never get really really tired (coughing), so I do not get so terrible many hours of sleep (coughing).*

Here it is highlighted that the informant's daily rhythm has shifted from its previous pattern, and the indication is that their lack of sleep stems from both reduced activity and diminished tiredness. None of the informants mentioned experiencing anxiety or shortness of breath during the night, which could also contribute to a different daytime rhythm. However, the increased need for sleep during the day was frustrating, as it could result in missing out on certain activities.

Hobbies

The informants engaged in different activities in their lives that held personal value for them, and some had to alter their activities because of COPD. This category focuses on the activities that the informants currently undertake and have previously engaged in.

The activities and interests of the informants varied depending on their weekly or daily routines. However, their condition often took precedence over their interests, and the activities that were feasible differed among the informants. Additionally, there was a gender disparity, with men favoring fishing and other outdoor activities, while women tended to engage in activities such as handicrafts.

> *M2: So now that I have been sick, yes, I go out and fish a little, then I go and help a little with some horses.*
>
> *W3: ...and then I'm knitting or doing the crosswords or trying to sew on the sewing machine (laughing).*
>
> *M1: yeah, I'm trying to do the things I care about and like, unfortunately I can't paint anymore, as I cannot stand the smell of turpentine anymore, it's sad*

> *because I've spent a lot of time painting, I don't have the energy to start writing more books. So I, I read a little, it's a bit difficult now with these glasses, but I've read a lot, and I get a lot of pleasure from it.*

The different accounts provided by the informants offer insights into how interests are possible for individuals with chronic illness as well as how the condition can prevent them from pursuing activities they like/enjoy. Despite certain interests being curtailed by the condition, informants still strive to engage in activities that bring them joy and hold value for them. However, in some cases, informants found it challenging to pursue their interests because of the awareness of potential shortness of breath.

The informants expressed similar daily desires and willingness to go outside, but their walks had become shorter over time. Occasionally, they cited the weather as an excuse to stay indoors.

> *W2: Then the little dog and I go in and rest for an hour or half an hour, and lull a little and sleep a little. Perhaps it is something completely different. Then I get up and get ready, and then I go with my little dog and pick up the newspaper in the mailbox. We used to go for longer walks, but I don't unfortunately, I can't do it anymore.*

> *W3: For just such a trip, so there is not much nature to go and look at from here and down to the municipal office, but just to get out and get some fresh air*

The informant acknowledges here that the challenge of going on longer walks is something that annoys her, but she simply cannot manage it anymore due to shortness of breath. Despite this limitation, going for a walk can bring relief to the informants. Even though nature may not always be visually stunning, the informant finds solace in being able to get outside, especially on slightly gloomier days.

Motivation

This category underscores how the informants experienced a lack of motivation to engage in daily activities. Throughout the conversations, there was a tendency to discuss things they would love to do but lacked the competence or strength to accomplish, or they invented excuses because the tasks seemed overwhelming. These could range from simple tasks such as planting a rose to more complex endeavors such as writing a book, attending gym classes, or cooking.

> *W1: Not at all, and then you lose the motivation.*

> *Interviewer: Yes, I can understand, if you have been somewhere where you think it was good, and then you come to something else that you don't think is at the same standard...*

> *W1: but then, okay, I am not...there, but oh its hunting me, when I have to go. I will probably just get it over with, right?*

The informant's desire and motivation to participate in a COPD exercise team depend on how the teacher conducts the sessions. A negative experience with teaching methods in the past has diminished the informant's motivation, making it difficult for them to participate. It has transformed from an activity that brings joy to feeling like an obligation, something that the informant feels they must do rather than something they want to do.

Resources

Social Networks

This category explores the social networks that the informants are a part of and how they use them in their day-to-day activities.

The size of the social network varied among the informants, but the significance of social interactions was equally important to all of them. Family relationships showed considerable diversity among the informants, with some maintaining close contact with their family through daily conversations, weekly scheduled visits, or having their spouse present. By contrast, there were some informants who had limited contact with their families and spent much of their daily lives alone.

> *W2: I miss him very much. We lived in Fyn, and we talked over the phone several times a week. I miss my family very much, and I also miss my friends. And they all passed away...*

> *W2: ...I talk to my daughter. So so I don't talk about..., she can say to me, I think you sound a little stalled mother, because then I'm just for the moment and then, and then we're not talking about it anymore. She knows what it is, but we don't need to.*

The above description indicates that the informant is alone due to deaths in the family and social circles, and there are a significant number of people missing from her family. Interestingly, the informant does not mention at the present time that she still has 2 daughters, which she only brings up later in the interview. The informant's description of the varying family relations indicates that she is left more isolated and alone, which telephone conversations with her daughters cannot fully mitigate. The conversations surrounding the family and friends of the "lonely" informants were marked by a sense of sadness and depression over the lack of contact.

Informants with close family relations expressed how their family and close relationships maintain continuous contact with them. The conversations were even interrupted by phone calls from their families, highlighting the frequent and ongoing nature of their communication.

> *M2: Yeah, she is calling, or she has stopped a little, but otherwise she calls every morning, around 9 o'clock or something like, "How do we breathe today?" She says then, (laughing)*

> *W1: Then I also have my granddaughter, I talk a lot with her, but I also take care of what I said to her, because she is a little unstable.*

There are 2 different scenarios for contact described here, both indicating that the informants have contact with their families, signifying close relationships. In one instance, it is the informant's mother showing interest in their self-monitoring and health status. In the second scenario, the informant not only maintains close contact, but also plays a protective role for her grandchild, who also suffers from a diagnosis. Despite varying family dynamics, it is evident that the informants can be divided into 2 groups: those with close family relations and those lacking such connections.

Their interest in engaging in social activities was also significant, but the informants often found themselves coming up with excuses for not participating or found it challenging to leave their homes.

> *W3: Yes, I haven't reached it yet, but I'm probably getting enough. There has been something on Friday, because otherwise I had set myself up for, I have otherwise gone to gym down in Vig, but.. that, which is quite far from the station off and down to Balsagård (...), of course I can go down there from time to time, but as I have it, oh for the last season there, I was not there quite many times, but it costs no matter what, they do not pull anything from because I have not been there, and that.. it annoys me a little. Then there is Red Cross that has something like this in high town.. exercise, sport is known enough, and it is every Friday morning, but there has been something here the last couple of Fridays, and I also have to just get into the rhythm that I have to go there until half 10 p.m.*

As emphasized, the informant highlights that traveling a long distance to attend gymnastics is a major obstacle. Despite continuing to pay for it, this does not motivate her to attend regularly. She also mentions the challenge of incorporating it into her Friday routine and making it a regular part of her schedule.

They did not envision themselves participating in social events related to their COPD. Some of them were members of Lungeforeningen, the Danish Association for Lung Diseases. Although the Lung Association organizes various gatherings for those with COPD, none of them appeared interested in attending.

Health Professionals

The category focuses on the informants' interaction with various health professionals and how those relationships hold significance for them.

The relationship with health professionals was highly significant, as it was essential for the informants to feel secure while also being with mutual respect and seriousness. The informants interacted with various health professionals in their daily lives, and this analysis distinguished the difference between "ordinary" health professionals and PreCare nurses.

The informant's relationship with the assistance offered is crucial.

> *W2: After a hospitalization for yeah I don't know, let's just say a year ago. There seems to be, I can't remember who thinks that there should be a home care and dosing the medicine. And now you must not misunderstand me because I am not a racist, but there comes a little colored girl who could not really speak Danish and she was not very sweet if she had been sweet and smiling, pleasant, then it would have been something else, but it was she not, and then she made mistakes twice after the other, and I discovered it quickly, and it is not, it isn't the tablets we sit and play with.*

According to the informant, the connection with home care has been challenging because of mistakes and uncertainty. As a result of this, the informant has lost faith in the home care, which should be there to lend a helping hand and not cause her problems on a regular basis.

Unlike other health care professionals, the nurses in PreCare have succeeded in establishing a sense of security and mutual respect with the informants. The collaboration with those involved in their COPD care instills confidence in the PreCare project and the nurses among the informants.

> *M1: There is most of the contact through the nurses, just to start, just when I started up there was a doctor who was here, and so I have nothing bad to say about him, oh and it is also those who prescribe some medicine if I lack it, and such some things not too (slang). I think I have a good relationship with them, and are really pleased to have them, oh...and feel there is a great confidence to be with them. So, as I said earlier, I was sure I would have become a burden for the hospital if I had not known them. The society saves money, and that's not bad.*
>
> *W2: But Nærklinikken has helped me, exceptionally. I'm glad I got in touch with you, you can believe. I don't know what I would have done without medical care. They do nothing.*

The informant expresses happiness for the nurses and thinks that their connection is good. The following description includes several elements though. The informant mentions that their participation in PreCare makes them feel like less of a burden for the hospital and the municipality, which holds significant meaning for them. He also expresses faith in the nurses, which he had expressed several times in the conversation. Being a part of a project that highlights the superior care provided by nurses compared with general medical care has been particularly significant for the female informant.

Nurses are not only available to informants, but also offer support if an informant's condition deteriorates. When an exacerbation happens, trust means the informant has no reason to question the nurses and takes the prescribed prescription without a second thought.

> *M2: Yes, Nærklinikken says it, it is the ones who change it now, yes, they say you just have to have two breaths in the morning, and they do it regularly if I have had it a little bad for a while. No, then I'll just be put up, double up.*

Without hesitation, the informant chooses to follow the nurses' recommendations. He has completely surrendered to the project, giving them full responsibility for his condition.

Technology

This category covers the informants' daily technology and their search for health-related information.

All informants admitted to having technology at home, although the way they used it and its purpose varied. For some, technology provides entertainment during moments of boredom in their daily lives.

> *W1: Yes, I'm mostly on the computer when I get bored.*
>
> *Interviewer: What are you doing on the computer?*
>
> *W1: I am playing games*
>
> *W1: I have two different games I have discovered.*

Technology was not utilized by the informants as a means to gather information about COPD. The informants felt they already possessed all the information they needed about COPD, and they were concerned that obtaining more information might increase their anxiety.

> *M1: It is very very rare; it is very rare. If I happen to hear that there is something new about it, then I can well find out to look it up, that it is not so exciting to read about, so*
>
> *W1: I think the more you read, the more nervous you become.*

The informants do not use technology to seek information about their condition. The informants did not mention being part of online groups where information could be shared during the interviews. None of the informants mentioned using social media platforms such as Facebook as a community for sharing information about their condition. They also emphasized that they generally did not share information about their condition through digital solutions.

> *M3: It irritates me sometimes when we sit, sometimes I cut through and say, now we don't want to talk about illness, because, oh, then such a short evening can go*

The medical equipment provided by PreCare did not pose any problem for the informants to use on a daily basis. They all expressed how easy it is to use it and how it takes only a few seconds to use the technology. They appeared confident and stated that they performed the measurements every day.

> *M4: it's so easy, that's in order...the only thing is now just, the crazy computer goes out, or (...), I can't restart, even though I have PIN code...no matter what I do, it won't, so I wait when it comes a past...so it can restart again, the only problem...*

Despite the ease, they experienced some issues with the devices. The informants encountered issues with logging in, forgot to charge the tablet, or even misused the thermometer.

Discussion

Principal Findings

This study offers valuable insights into understanding the needs, values, and preferences of individuals living with COPD as well as which communities they identify with in a digital context. Indeed, the findings highlight that while fluctuations in their health condition significantly affect the daily lives of the informants, factors such as having hobbies, old habits, and social connections play a crucial role in their overall well-being. This underscores the importance of recognizing individuals with COPD as complete human beings beyond their medical condition.

It is interesting to note the distinction between the 2 communities the informants belong to. The community centered around the RCC represents a vital support network for them, where they feel included and have developed trust with the staff. This highlights the importance of such digital health services in providing continuous support and guidance for individuals managing chronic conditions such as COPD. Involving close relatives in the community centered around the RCC can further enhance the support system for individuals with COPD. The other community involves participation in social activities outside the context of their condition which provides informants with opportunities for social interaction, enjoyment, and connection with others beyond their health concerns. These activities offer a sense of normalcy and contribute to their overall quality of life, allowing them to engage in meaningful relationships and experiences beyond the realm of COPD management. Even though the informants may face constraints due to their condition, they still find value in participating in social activities, even if their involvement is limited.

The Needs, Values, and Preferences in the Digital Context

The informant's emphasis on maintaining their daily activities underscores the significance of preserving their sense of normalcy and independence despite their COPD symptoms. It reflects their desire to continue living fulfilling lives and not be defined solely by their health condition. Symptoms often contribute to a lower quality of life and well-being. The findings from an earlier study [32] resonate with the experiences reported by individuals with COPD in this study. Breathlessness, a common symptom of COPD, can significantly impact an individual's quality of life by limiting their ability to engage in daily activities and causing distress. This aligns with the participants' reports of reduced quality of life related to breathlessness, highlighting the importance of addressing this symptom to improve overall well-being for individuals living with COPD. While outdoor activities may be affected by COPD, the focus of the informants seems to be more on how the disease impacts their ability to engage in everyday tasks and maintain their hobbies or household chores. This also resonates well with another study including interviews of patients with COPD [7]. The study revealed that for women, being active in housekeeping was important and valued, while for men, maintaining the garden held similar significance. They also reported that people with COPD prefer activities within their home or immediate vicinity, showing little interest in engaging in social activities located far away due to their reduced physical abilities. The need to be close to a safe environment was similar for some of the informants in this study. However, some reported that they did leave their house. For example, one informant was helping with some horses, while others needed to take their dog to the dog groomer. Access to virtual support may play an important role here, as the informant could call the RN in the RCC at any time if anything were to happen.

The COPD-Related Conditions' Impact on the Identity

For the informants, it was important to avoid discussing or being associated with their COPD condition by shifting the topic during the interviews. This aligns with a previous study where people with COPD were not interested in being identified solely by their illness [33]. They found that their identity related to their condition, termed "illness identity," was affecting their roles and could potentially separate them from the social network or community. An additional contributing factor to social segregation was a sense of having a self-inflicted disease, leading to feelings of shame and guilt [33]. However, this was not evident in our data.

Most, but not all, of the informants appear to be able to cope with their diagnosis and condition. They try to continue the same kind of interactions and activities while considering their condition's restrictions. As a result, individuals experience a sense of maintaining their own identity within their communities, yet occasionally feel the sense of lacking something. This can involve engaging in distant activities, such as joining a choir.

The Two Communities of Practice

The Community of COPD Practice

In relation to the community with its formal caregivers, the informants experienced a genuine interest from the RNs in their well-being and they provided them with support in an empathic way. This experience may be attributed to the PreCare environment, with free access to RNs 24/7, where they always kept an eye on the informants. This may explain the absence of "anxiety" in the interviews with the informants. These could be attributed to the RN's *ability* to provide immediate support in response to changes in their health condition, with medication for deterioration accessible at home [32]. Anxiety, which often dominates the daily lives of individuals with COPD, is thus better managed [32]. Therefore, it is necessary for the RNs to be trained to instill confidence in individuals with COPD or other LTHCs, enabling them to feel more independent and socially active with the RCC, their equipment, and a medicine box readily available, thereby reducing anxiety and maximizing the benefits of their resources.

The informants' immediate access to the RNs in the RCC appears crucial, as it enhances their self-efficacy and confidence in how their equipment aids them in managing deterioration.

However, despite feeling secure in their use of the equipment, the informants did not appear to be influenced by their ability for self-management and did not feel more empowered, likely due to experiencing the RNs as being paternalistic. The informants were unable to fully benefit from the virtual support environment due to the influence of the RNs and instead remained in a passive role.

RNs and other clinicians will need to be aware of how they communicate to facilitate a dialog that is not experienced as paternalistic, but rather as a coaching conversation.

The Community of Social Practice

In relation to their social communities, the extent of social relationships varied among the informants. The importance of a family community aligns with the findings of Nicolson and Anderson [32], who showed that family and relatives significantly influence the quality of life for people with COPD. Nicolson and Anderson [32] identified that COPD impacted how individuals connected with relatives and perceived their ability to fulfill their roles within the family. In contrast to the study by Nicolson and Anderson [32], our findings indicate that the informants' family roles were not influenced by their COPD. They maintained their roles and continued normal interaction with their relatives and families.

Not all informants participated in social activities or were part of a local community. Those without support from friends and families experienced difficulties engaging in and finding motivation for social activities. They felt lonely when left alone in their homes, whereas those with a social network experienced loneliness to a lesser degree. This contrasts with another study [7], which found that loneliness was also a major issue for those with family support, such as spouses and friends [7]. Those who felt lonely because of their lack of participation in social activities found some comfort in the availability and contact with the RN, which to a certain extent reduced their experience of loneliness.

This underscores the necessity for RNs to possess skills in mental and social support, which should be included in the education programs for nurses.

Online resources such as patient portals and social media (eg, Facebook) can constitute a community for people with COPD. However, despite the availability of these platforms to our informants and their daily use of tablets, none of them considered these online opportunities in relation to their COPD condition. This may be due to various reasons. Some informants felt they had sufficient knowledge or were unsure how to interpret the overwhelming information on the internet. Additionally, they may have wanted to avoid exposing their diagnosis or involving others outside their close network [34]. The role of the PreCare environment and the RNs may substitute the need for a social media platform or COPD-related conversations on platforms such as Facebook.

When the informants do not participate in online communities, they may miss the opportunity to access new knowledge or learn from others with similar conditions. This lack of engagement can reduce their ability to manage their condition effectively and hinder their empowerment. Joining an online community and being actively involved can help transform newcomers into "super users" and "experts" [35]. These "online experts" can then help other members of the community, forming a virtual community of practice [25]. Thus, participants in the PreCare project may miss the opportunity to develop into experts through online activities but may instead develop this competence through participation in other communities or collaboration with the RNs.

Limitations

The study is based on 7 informants. This may be considered a limitation, as the relatively small number may result in some perspectives of people living with COPD not being expressed in the data. However, as all informants are exposed to the same PreCare environment, have the same diagnosis of COPD, and live within the same area, we find that the necessary number of participants to have enough power of information is met [36]. This is supported by the presence of common patterns among the informants and the alignment of the overall findings with the data obtained in the PreCare project. Further studies are needed to confirm our findings before they can be considered valid for scaling up and evaluating the impact of working with landscapes or communities of practice. This support aims to foster a sense of more active and independent living based on existing values.

Perspective

The findings suggest that education for RNs and other health professionals should focus on their roles as professionals while also acting as facilitators. They should avoid being paternalistic to create a space for the development of self-efficacy and self-management. A motivating factor will help develop self-efficacy and confidence, enabling people with COPD to be more socially active and encouraging them to pursue their desires. Health professionals play a key role here, as they can provide the means to help individuals become more active, thereby increasing their well-being.

Conclusions

When using digital health solutions, people's needs, values, and preferences should be considered, focusing primarily on addressing the whole person rather than just the "illness." This approach creates the best opportunity for individuals to maintain their daily activities and feel empowered. The 2 communities the informants take part must work together and will intersect in their daily lives. They should support each other, involving the needs, values, and preferences of the individuals, and ensuring that upcoming digital health services include and embrace situated learning to enhance people's empowerment and self-management. Furthermore, new educational programs should be developed or considered to enhance the competencies of RNs who are involved in digital health services. This will provide the best opportunity for the 2 communities to collaborate and support the daily activities of people with chronic conditions.

Acknowledgments

The authors thank the PreCare Project for assisting with the recruitment of informants and the informants for participating in the study. The first author has received funding from Region Zealand. CWS, KB, and LK have been partly financed by the EU H2020 Research & Innovation Programme (grant agreement number 101016848) for the period 2021-2024.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.
[DOCX File , 16 KB - humanfactors_v11i1e53131_app1.docx]

Multimedia Appendix 2
Categorization of the data.
[XLSX File (Microsoft Excel File), 12 KB - humanfactors_v11i1e53131_app2.xlsx]

References

1. Kayser L, Phanareth K. Leave no one behind: the role of digital health literacy. European Respiratory Society 2023:79-94. [doi: 10.1183/2312508x.10001023]
2. Cavanagh R, Gerson SM, Gleason A, Mackey R, Ciulla R. Competencies Needed for Behavioral Health Professionals to Integrate Digital Health Technologies into Clinical Care: a Rapid Review. J Technol Behav Sci 2022 Feb 17;8(4):1-14 [FREE Full text] [doi: 10.1007/s41347-022-00242-w] [Medline: 35194561]
3. Jarva E, Oikarinen A, Andersson J, Tuomikoski A, Kääriäinen M, Meriläinen M, et al. Healthcare professionals' perceptions of digital health competence: A qualitative descriptive study. Nursing Open 2022 Jan 30;9(2):1379-1393. [doi: 10.1002/nop2.1184]
4. Longhini J, Rossettini G, Palese A. Digital Health Competencies Among Health Care Professionals: Systematic Review. J Med Internet Res 2022 Aug 18;24(8):e36414. [doi: 10.2196/36414]
5. Chronic obstructive pulmonary disease (COPD). Chronic obstructive pulmonary disease (COPD). URL: https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd) [accessed 2023-01-30]
6. Nøgletal OL. Nøgletal om lungesygdomme. Lungeforeningen. 2016. URL: https://tinyurl.com/5dsabf2s [accessed 2023-01-30]
7. Williams V, Bruton A, Ellis-Hill C, McPherson K. What really matters to patients living with chronic obstructive pulmonary disease? An exploratory study. Chron Respir Dis 2016 Jun 29;4(2):77-85. [doi: 10.1177/1479972307078482]

8. Global Initiative for Chronic Obstructive Lung Disease - GOLD. GOLD Report 2023. URL: https://goldcopd.org/2023-gold-report-2/ [accessed 2023-05-25]
9. GOLD. Global strategy for the diagnosis, management, and preventien of chronic obstructive pulmonary disease (2018 report). GOLD. 2018. URL: https://tinyurl.com/3r6mkuuk [accessed 2023-05-25]
10. Pare G, Jaana M, Sicotte C. Systematic Review of Home Telemonitoring for Chronic Diseases: The Evidence Base. Journal of the American Medical Informatics Association 2007 May 01;14(3):269-277. [doi: 10.1197/jamia.m2270]
11. McDowell JE, McClean S, FitzGibbon F, Tate S. A randomised clinical trial of the effectiveness of home-based health care with telemonitoring in patients with COPD. J Telemed Telecare 2015 Mar;21(2):80-87. [doi: 10.1177/1357633X14566575] [Medline: 25586812]
12. Worth H, Dhein Y. Does patient education modify behaviour in the management of COPD? Patient Education and Counseling 2004 Mar;52(3):267-270. [doi: 10.1016/s0738-3991(03)00101-0]
13. Nissen L, Lindhardt T. A qualitative study of COPD-patients' experience of a telemedicine intervention. International Journal of Medical Informatics 2017 Nov;107:11-17. [doi: 10.1016/j.ijmedinf.2017.08.004]
14. McCabe C, McCann M, Brady AM. Computer and mobile technology interventions for self-management in chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2017 Dec 23;5:CD011425. [doi: 10.1002/14651858.CD011425.pub2] [Medline: 28535331]
15. Hardman R, Begg S, Spelten E. What impact do chronic disease self-management support interventions have on health inequity gaps related to socioeconomic status: a systematic review. BMC Health Serv Res 2020 Feb 27;20(1):150 [FREE Full text] [doi: 10.1186/s12913-020-5010-4] [Medline: 32106889]
16. WHO. Social determinants of health. URL: https://www.who.int/health-topics/social-determinants-of-health [accessed 2023-09-14]
17. Madsen J, Kanstrup AM, Josephsson S. The assumed relation between occupation and inequality in health. Scand J Occup Ther 2016 Aug 28;23(1):1-12. [doi: 10.3109/11038128.2015.1075065] [Medline: 26314738]
18. Lave J, Wenger E. Situated Learning. Cambridge, United Kingdom: Cambridge University Press; 1991.
19. Bandura A. Social Learning Theory. Canadian Journal of Sociology / Cahiers canadiens de sociologie 1977;2(3):321. [doi: 10.2307/3340496]
20. Spradley JP. Participant Observation. Anthropological Quarterly 1980 Oct;53(4):260. [doi: 10.2307/3318111]
21. WHO. WHO Framework on integrated people-centred health services. 2016 Aug. URL: http://www.who.int/servicedeliverysafety/areas/people-centred-care/en/ [accessed 2023-09-14]
22. Phanareth K, Vingtoft S, Christensen AS, Nielsen JS, Svenstrup J, Berntsen GKR, et al. The Epital Care Model: A New Person-Centered Model of Technology-Enabled Integrated Care for People With Long Term Conditions. JMIR Res Protoc 2017 Jan 16;6(1):e6 [FREE Full text] [doi: 10.2196/resprot.6506] [Medline: 28093379]
23. Data Development, Region Zealand. PRECARE: ViosKlinikken - Sector Neutral, Data Driven and Person-Centered Health and Care Services. PreCare Clinic 2024:140.
24. Kvale S. Interview, en introduktion til det kvalitative forskningsinterview. In: Interview, en introduktion til det kvalitative forskningsinterview. Denmark: Hans Reitzel Forlag; 1990.
25. Wenger E. Communities of Practice. Cambridge: Cambridge University Press; 1999.
26. Kayser L, Rossen S, Karnoe A, Elsworth G, Vibe-Petersen J, Christensen JF, et al. Development of the Multidimensional Readiness and Enablement Index for Health Technology (READHY) Tool to Measure Individuals' Health Technology Readiness: Initial Testing in a Cancer Rehabilitation Setting. J Med Internet Res 2019 Feb 12;21(2):e10377 [FREE Full text] [doi: 10.2196/10377] [Medline: 30747717]
27. Terp R, Kayser L, Lindhardt T. Older Patients' Competence, Preferences, and Attitudes Toward Digital Technology Use: Explorative Study. JMIR Hum Factors 2021 May 14;8(2):e27005 [FREE Full text] [doi: 10.2196/27005] [Medline: 33988512]
28. Terp R, Lindhardt T, Kayser L. Theory-driven development of an educative nutritional intervention (ENI) supporting older hospital patients to eat sufficiently, assisted by an eHealth solution: an intervention mapping approach. BMC Health Serv Res 2022 Nov 28;22(1):1435 [FREE Full text] [doi: 10.1186/s12913-022-08679-8] [Medline: 36443804]
29. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
30. Nvivo. Nvivo. Alfasoft. URL: https://help-nv.qsrinternational.com/12/win/v12.1.115-d3ea61/Content/welcome.htm [accessed 2021-01-30]
31. The Danish National Center for Ethics. Overview of Mandatory Reporting. 2024 Feb 01. URL: https://researchethics.dk/information-for-researchers/overview-of-mandatory-reporting [accessed 2024-03-25]
32. Nicolson P, Anderson P. Quality of life, distress and self-esteem: a focus group study of people with chronic bronchitis. Br J Health Psychol 2003 Sep 16;8(Pt 3):251-270. [doi: 10.1348/135910703322370842] [Medline: 14606972]
33. Lippiett K, Richardson A, May CR. How do illness identity, patient workload and agentic capacity interact to shape patient and caregiver experience? Comparative analysis of lung cancer and chronic obstructive pulmonary disease. Health Soc Care Community 2022 Nov 28;30(6):e4545-e4555 [FREE Full text] [doi: 10.1111/hsc.13858] [Medline: 35633149]
34. Andreou A, Dhand A, Vassilev I, Griffiths C, Panzarasa P, De Simoni A. Understanding Online and Offline Social Networks in Illness Management of Older Patients With Asthma and Chronic Obstructive Pulmonary Disease: Mixed Methods Study

Using Quantitative Social Network Assessment and Qualitative Analysis. JMIR Form Res 2022 May 17;6(5):e35244 [FREE Full text] [doi: 10.2196/35244] [Medline: 35579933]
35. Joglekar S, Sastry N, Coulson NS, Taylor SJ, Patel A, Duschinsky R, et al. How Online Communities of People With Long-Term Conditions Function and Evolve: Network Analysis of the Structure and Dynamics of the Asthma UK and British Lung Foundation Online Communities. J Med Internet Res 2018 Jul 11;20(7):e238 [FREE Full text] [doi: 10.2196/jmir.9952] [Medline: 29997105]
36. Malterud K, Siersma VD, Guassora AD. Sample Size in Qualitative Interview Studies: Guided by Information Power. Qual Health Res 2016 Nov 10;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]

Abbreviations

COPD: chronic obstructive pulmonary disease
ECM: Epital Care Model
GOLD: The Global Initiative for Chronic Obstructive Lung Disease
LTHC: long-term health condition
RCC: response and coordination center
RN: registered nurse

Edited by A Choudhury; submitted 27.09.23; peer-reviewed by H Salim, J Turnbull; comments to author 31.12.23; revised version received 29.04.24; accepted 24.05.24; published 21.06.24.

Please cite as:
Schmidt CW, Borgnakke K, Frølich A, Kayser L
Preferences, Needs, and Values of Patients With Chronic Obstructive Pulmonary Disease Attending a Telehealth Service: Qualitative Interview Study
JMIR Hum Factors 2024;11:e53131
URL: https://humanfactors.jmir.org/2024/1/e53131
doi:10.2196/53131
PMID:

Original Paper

Novel Approach to Personalized Physician Recommendations Using Semantic Features and Response Metrics: Model Evaluation Study

Yingbin Zheng[1], MS; Yunping Cai[2], ME; Yiwei Yan[1], MA; Sai Chen[2], ME; Kai Gong[1], MD

[1]Biomedical Big Data Center, The First Affiliated Hospital of Xiamen University, School of Medicine, Xiamen University, Xiamen City, China
[2]Meteorological Disaster Prevention Technology Center, Xiamen Meteorological Bureau, Xiamen City, China

Corresponding Author:
Kai Gong, MD
Biomedical Big Data Center, The First Affiliated Hospital of Xiamen University, School of Medicine, Xiamen University
10 Shanggu Road, Siming District
Xiamen City, 361003
China
Phone: 86 15160003918
Email: freatink@xmu.edu.cn

Abstract

Background: The rapid growth of web-based medical services has highlighted the significance of smart triage systems in helping patients find the most appropriate physicians. However, traditional triage methods often rely on department recommendations and are insufficient to accurately match patients' textual questions with physicians' specialties. Therefore, there is an urgent need to develop algorithms for recommending physicians.

Objective: This study aims to develop and validate a patient-physician hybrid recommendation (PPHR) model with response metrics for better triage performance.

Methods: A total of 646,383 web-based medical consultation records from the Internet Hospital of the First Affiliated Hospital of Xiamen University were collected. Semantic features representing patients and physicians were developed to identify the set of most similar questions and semantically expand the pool of recommended physician candidates, respectively. The physicians' response rate feature was designed to improve candidate rankings. These 3 characteristics combine to create the PPHR model. Overall, 5 physicians participated in the evaluation of the efficiency of the PPHR model through multiple metrics and questionnaires as well as the performance of Sentence Bidirectional Encoder Representations from Transformers and Doc2Vec in text embedding.

Results: The PPHR model reaches the best recommendation performance when the number of recommended physicians is 14. At this point, the model has an F_1-score of 76.25%, a proportion of high-quality services of 41.05%, and a rating of 3.90. After removing physicians' characteristics and response rates from the PPHR model, the F_1-score decreased by 12.05%, the proportion of high-quality services fell by 10.87%, the average hit ratio dropped by 1.06%, and the rating declined by 11.43%. According to whether those 5 physicians were recommended by the PPHR model, Sentence Bidirectional Encoder Representations from Transformers achieved an average hit ratio of 88.6%, while Doc2Vec achieved an average hit ratio of 53.4%.

Conclusions: The PPHR model uses semantic features and response metrics to enable patients to accurately find the physician who best suits their needs.

(JMIR Hum Factors 2024;11:e57670) doi:10.2196/57670

KEYWORDS

web-based medical service; text analysis; Sentence Bidirectional Encoder Representations From Transformers; SBERT; smart triage systems; patient-physician hybrid recommendation; PPHR; PPHR model

Introduction

Background

Web-based medical consultation is increasingly popular as an alternative to traditional health care services because it is convenient, accessible, and affordable [1]. This type of patient-physician interaction takes place electronically, connecting both parties through text, images, and videos. Its advantages include eliminating time and space constraints and accurately documenting the medication process [2], making it more attractive to many patients than in-person medical visits. As of the end of 2022, the number of users in China's internet medical and health market reached 363 million [3]. The rapid growth of web-based medical services and the vast amount of information available have created considerable difficulties for patients in finding the physicians best suited to their needs, leading to potentially mismatched consultations [4].

At present, most existing triage procedures rely on manual recommendation from schedulers to select departments for patients. As the number of consultations increases, manual provision of advice does not guarantee the professionalism and quality of medical services [5]. In addition, schedulers are unable to provide 24-hour service, resulting in gaps in health care access and the continuity of services. At this point, a common approach might be to develop an intelligent department recommendation model. Advancements in technology, particularly in the field of machine learning, present opportunities to improve the accuracy and efficiency of patient department assignment in health care systems. For example, Mullenbach et al [6] integrated the attention mechanism and used long short-term memory to predict the patient's disease type for further triage. Li and Yu [7] used multifilter residual convolutional neural networks to investigate the issue of department recommendation. Wang et al [8] used the Bidirectional Encoder Representations from Transformers (BERT) model to study disease diagnosis and department recommendations. These approaches can potentially automate the process of assigning patients to appropriate departments, reducing the burden on schedulers and improving patient outcomes through more accurate and timely care.

However, due to the ongoing subdivision of departments, these department recommendation models still cannot accurately match medical needs with physicians' specialties. For example, obstetricians and gynecologists further specialize in subfields such as gynecology, obstetrics, reproductive endocrinology, infertility, prenatal diagnosis, and genetic counseling. This refined division not only improves the effectiveness of diagnosis and treatment but also ensures that patients receive the most cutting-edge and professional care plans. In addition, even if the diseases treated are similar or the same, different medical institutions may have different department names. These problems have placed higher demands on hospital management, requiring more precise resource allocation to adapt to increasingly specialized services. Therefore, there is an urgent need to design personalized physician recommendation models.

Personalized recommendation methods can help users manage massive amounts of information and knowledge [9] and are crucial for providing personalized medical services that meet the patient's needs [10]. For instance, Ju and Zhang [11] integrated geographical location and patients' questions to generate personalized recommendations. Liu et al [12] proposed a physician recommendation model that considers the characteristics of patients and physicians. Lu et al [5] proposed a self-adaptive physician recommendation system that considers physician activity and patient feedback. These methods can be advantageous for both patients and web-based health care providers, as they minimize the time and effort required to find a suitable match, thus ensuring efficient delivery of health care services [13].

However, there are still some shortcomings in previous studies. Most existing studies use satisfaction as a measure of physician performance. However, the authenticity of satisfaction ratings across different platforms is not always reliable, as many users tend to habitually provide positive feedback. In terms of the evaluating indicators for recommended physicians, most studies used accuracy as a single indicator and did not consider the service quality of recommended physicians. These limitations may result in consultation mismatches, longer patient waiting times, and potentially reduced patient satisfaction. To the best of our knowledge, previous studies have not developed a triage system for recommending physicians that uses the transformer-based models, which are the cutting-edge models for natural language processing. BERT [14] is a popular transformer-based model that has been pretrained on common texts, such as Wikipedia and the Brown Corpus. BERT is a state-of-the-art model that uses an attention-based mechanism [15,16] to accurately understand the context of words, enabling unsupervised learning by linking text input and output through a decoder-encoder framework [17,18]. However, the BERT model is not suitable for semantic similarity searches or clustering, which has led to the creation of a different sentence-embedding model called the Sentence BERT (SBERT) model [19]. This modified version of the BERT model was designed to be semantically meaningful and suitable for sentence similarity tasks. It works by integrating a Siamese network and a pretrained BERT model, along with a pooling layer that generates a fixed-sized representation. The SBERT model can accurately identify whether there is a significant match between 2 sentences, making it a useful tool for data mining, information retrieval, and text matching [20].

This Study

The objective of this study was to develop a more precise algorithm that can better recommend professional and highly engaged physicians and thus improve the effective use of medical resources and the medical experience of patients by reducing the mismatches between medical needs and services. The practical benefits expected from our findings include the enhanced ability of web-based health care platforms to provide timely, relevant, and professional medical consultations that are closely tailored to each patient's unique needs. By implementing our advanced recommendation algorithm, we expect to not only identify the most appropriate specialists based on patient input but also incorporate a comprehensive evaluation of physician performance metrics. This will ensure that patients are recommended physicians who are not only experts in their field but also highly engaged and responsive, resulting in higher-quality care.

We seek to answer these three questions: (1) how can we effectively construct features for patients and physicians to facilitate efficient physician recommendations? (2) how can we incorporate the physicians' performance metrics into recommendation strategies to increase the chance of recommending highly active physicians? and (3) how can the effectiveness of the recommendation strategy be verified considering both accuracy and service quality?

Methods

Data Collection

This research collected a total of 646,383 web-based medical consultation records from the Internet Hospital of the First Affiliated Hospital of Xiamen University between 2016 and 2023. Each record contains the textual question, deidentified codes for the physician and patient, the physician's department, and the response status and time. Response status refers to whether a consultation request has received a reply from the corresponding physicians. Response time is the duration between submitting a request and getting a response. A total of 5 examples of the questions generated during web-based medical consultations are displayed in Table 1.

These records were divided into 2 test data sets and 1 training data set. For the first test data set, the physician with the highest number of consultations was selected from each of the following departments with the most inquiries: gastroenterology, obstetrics, respiratory medicine, pediatrics, and dermatology. Their codes were 98, 141, 202, 512, and 601, respectively. A total of 400 consultation records were randomly selected from each of the aforementioned physicians. These physicians then reviewed these textual questions to determine whether they were within their expertise. Any questions that a physician is proficient in was tagged, and eventually we randomly selected 200 records for each physician from these tagged questions to compile a test data set consisting of 1000 records. For the second test data set, a sample of 10,000 consultations was randomly chosen from the total data set, excluding the consultation samples from the first test data set. The training data set consisted of the consultations remaining after the removal of the first and second test data sets. The random seed for this study was set to 2023.

Table 1. Examples of patients' consultation questions.

Sample number	Patient code	Questions	Physician code	Department	Response status[a]	Response time
1	200321	You initially diagnosed me with left varicocele and ordered a color Doppler ultrasound examination. The results showed a moderate left varicocele with reflux. I would like to inquire whether, aside from surgery, this condition can be treated through medication, injections, or other noninvasive methods?	208	Urology	True	8 h 12 min
2	306878	My child has been experiencing discomfort in the throat and recurring fevers for four days before visiting a physician, who considered it was pneumonia and started administering azithromycin. Today is the fifth day of treatment. After the first day of intravenous azithromycin, the fever subsided, but there is still occasional coughing with phlegm. I am worried about the potential for significant side effects. Should the child continue taking azithromycin?	372	Pediatrics	False	null
3	447138	I am currently on my period and have scheduled an ultrasound and mammography for this afternoon. Could you please tell me if there is a recommended waiting period before trying to conceive after a mammography?	133	Breast surgery	True	20 min
4	591872	What does the glucose tolerance test report indicate? Could you please explain it to me?	423	Obstetrics	True	9 min
5	603826	My chin is red without feeling painful or itchy, and it has been like this for over a month. I've tried Clotrimazole but no obvious effects were achieved. Could you please tell me what condition this might be and what medication I should use?	418	Dermatology	True	2 h 20 min

[a]True: the physician has responded to the consultation request; false: the physician has not responded to the consultation request.

Data Preprocessing

Data related to patients' consultation questions were collected and presented in the form of natural language. Preprocessing of these unstructured data is crucial in machine learning framework [21] to remove unnecessary, duplicated, irrelevant, and noisy data [22]. This study involved several steps to process these consultation questions, including normalization, tokenization, part-of-speech tagging, and stop-word removal, thereby forming a reliable corpus.

The study calculated response rates and times for all physicians as shown in equations 1 and 2, where N_R denotes the number of consultation requests that physician P_i has responded to, with "responded" indicating that the response status is confirmed as true. Meanwhile, N indicates the total number of consultation requests that physician P_i has received. Furthermore, S_T refers to the total response time for all the consultation requests that physician P_i has responded to:

(1)

(2)

Upper and lower bounds on response times were established to minimize the impact of extremely high and low values on the experiment. Response times >95% were capped at 8 hours and 6 minutes, while those below the fifth percentile were raised to 9 minutes.

Feature Extraction

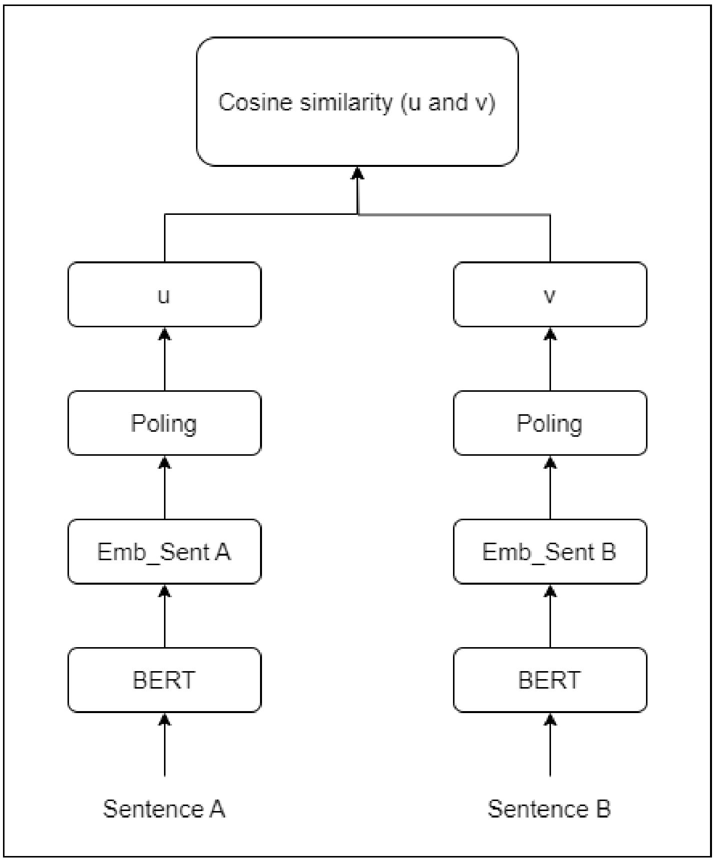
Feature extraction is the process of converting raw input data into a meaningful set of features [23] that can be understood by machine learning classifiers. In the feature extraction stage, 2 unique features for both patients and physicians were introduced.

Patients' Feature Modeling

This study used the pretrained SBERT model known as "distiluse-base-multilingual-cased" to convert all consultation questions into semantic representations and then calculate sentence embeddings for further analysis. As shown in Figure 1 [19], the SBERT model processed sentences A and B through BERT pooling to generate their respective embeddings, *u* and *v*. The similarity between these embeddings is then calculated using the cosine similarity method, which effectively measures how similar sentences are. The cosine similarity is expressed by equation 3, where *u* and *v* represent 2 vectors:

(3)

Figure 1. Sentence Bidirectional Encoder Representations from Transformers (BERT) architecture.



Physicians' Feature Modeling

Term frequency–inverse document frequency (TF-IDF) model [24,25] is a commonly used method in text mining and information retrieval because it can capture the importance of words and has the potential to extract features from multiple texts. The formula of this algorithm is shown in equation 4:

(4)

where *TF* (*t*,*d*) represents the frequency of a specific keyword *t* in document *d*, while *IDF* (*t*) signifies the inverse document frequency. According to this formula, the higher the *TF-IDF* (*t*,*d*) value, the more significant the feature is in the document.

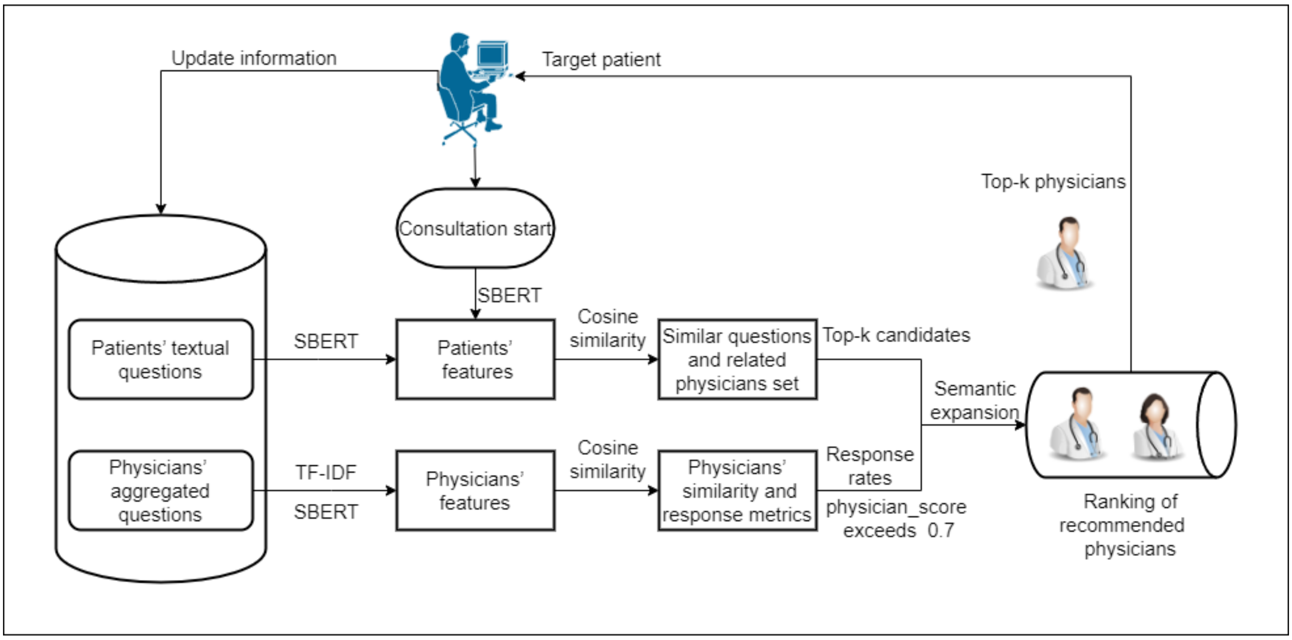
This study used the TF-IDF model to extract crucial information from a collection of patients' consultation questions aggregated by physician codes, selecting the top 20 with the highest TF-IDF weights. This extracted information was then fed into an SBERT model to compute cosine similarity among physicians.

Recommendation

Overview

A patient-physician hybrid recommendation (PPHR) model with response metrics was developed by combining features of both patients and physicians. This model also considers the physician's response rate in the recommendation strategy. The PPHR model, which is a type of top-*k* recommendation system, is designed to provide patients with a list of the top-*k* physicians who are most likely to meet their medical needs, as illustrated in Figure 2.

Figure 2. The architecture of the patient-physician hybrid recommendation model. SBERT: Sentence Bidirectional Encoder Representations from Transformers; TF-IDF: term frequency–inverse document frequency.



Step 1: Generate a Candidate Set Using the Patient Feature–Based Model

The patient feature–based (PFB) model was developed to identify sets of the most similar questions. When a new consultation starts, the consultation questions will be processed using SBERT to generate the corresponding embeddings. The patients' features were used to construct a similarity matrix among questions using cosine similarity. Then, consultation questions that are similar to the new consultation will be identified by comparing the patients' features. The physicians who are associated with these similar questions are considered potential candidates. The similarity score, known as the init_score, serves as the baseline for making recommendations. The top-*k* physicians are selected as candidates from the set of similar questions, where *k* is an adjustable hyperparameter in this model.

Step 2: Expand the Candidate Set Based on the Patient-Physician Hybrid Model

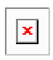
As patients' textual questions are unprofessional, setting a similarity threshold based solely on the patient characteristics may limit the recommendation results. The patient-physician hybrid (PPH) model ensures that all potential physician recommendations are considered. This model is formed by combining the physicians' features with the PFB model to semantically expand the scope of candidates. It does this by creating an index called the expand_score, which is derived from the physicians' features. This index reflects the degree of similarity among physicians and helps determine which physicians have the necessary expertise and qualifications to provide the right care for a given patient. This approach can adjust biases in the system that may arise from recommending physicians based solely on similarities to patients' questions. The PPH model is shown in equation 5:

$$\text{physician\_score}_i = \text{init\_score}_i \times \text{expand\_score}_i \quad \textbf{(5)}$$

If the physician_score exceeds 0.7, it will be used to semantically expand the range of candidates. When the physician is not derived from the PPH model, the expand_score is assumed to be 1.

Step 3: Optimize the Ranking of the Candidate Set by Incorporating the Response Rate

The response rate can serve as an indicator to measure the efficacy of physicians' performance. An increase in the response rate suggests that physicians are more willing to treat patients. This can be viewed as a positive feedback loop, as higher response rates lead to more motivated physicians. Therefore, it is crucial to consider the physicians' activity level along with the similarity index, as this can skew the recommendation results toward more active physicians, increasing the chance that inquiries will be answered and thus improving patient satisfaction. The final PPHR model is displayed in equation 6. The top-*k* physicians are selected for recommendation based on the scores, where *n* represents the number of times physician D_i is recommended:



(6)

Evaluation

The proposed PPHR model's effectiveness was evaluated using the following metrics: hit ratio, precision, recall, F_1-score, and high-quality service proportion. In the first test data set, a recommendation was considered correct when the selected physician was among the top-*k* recommended physicians. The hit ratio refers to the proportion of correct recommendations to the total number of recommendations. In the second test data set, a recommendation was regarded as accurate if the recommended physician's maximum physician_score is >0.7. Precision refers to the proportion of correctly recommended physicians to the total number of recommended physicians. By contrast, recall is the ratio of correctly recommended physicians to the number of physicians who should have been retrieved in the sample. The F_1-score is a valuable metric for assessing the recommendation algorithm's effectiveness, as it merges precision and recall to yield the best results. A higher F_1-score signifies a more efficient algorithm. Precision, recall, and F_1-score were calculated using the formulas described in equations 7 to 9, where *TP* is a true positive, *FP* is a false positive, and *FN* is a false negative:

(7)

(8)

(9)

A quick response time is a critical element for high user satisfaction, allowing the system to promptly provide services that meet patient expectations. If a physician responds quickly, the patient will perceive the quality of the physician's service to be better than that of a physician who takes a longer time to respond. Therefore, the proportion of physicians who respond quickly among all recommended physicians, known as the high-quality service proportion, is a significant measure of evaluation. The calculation for high-quality service proportion is determined by the formula shown in equation 10:

(10)

The term N_f represents the number of physicians whose response time is faster than the average response time, while N denotes the number of recommended physicians.

Baseline Experiments

For the PFB model, the purpose of baseline experiments was to determine whether excluding physicians' features from the PPHR model would degrade performance. A total of 3 steps were taken to assess its performance compared to the PPHR model. First, the hit ratio and ranking of the selected physician in the recommendation set were calculated in the first test data set. Second, the precision, recall, and F_1-score of the recommendation results were computed in the second test data set, and the consultation questions that were recommended to the selected physician (codes 98, 141, 202, 512, and 601) were collected. Finally, a questionnaire for assessing the rationality of the recommendations was administered by randomly selecting 200 consultation questions (100 for each model) for each physician from those consultation questions. The questionnaire included an evaluation of the relevance of each selected physician with the consultation questions. The survey question was as follows: "Based on your area of expertise, how would you rate the match between you and consultation question?" The questionnaire used a Likert 5-point scale [26] for measurement, with scores ranging from 1 (very inappropriate) to 5 (very appropriate). The Mann-Whitney *U* test [27] was used to determine whether there was a statistical difference in the physicians' perceptions of how well the consultation questions from these 2 models matched their area of expertise.

For the PPH model, the proportion of high-quality services in the recommendation results was calculated to assess whether eliminating the response rate in the PPHR model will reduce service quality.

Doc2Vec [28] was used to create text embeddings for all patients' consultation questions to reconstruct the PPHR model. The performances of Doc2Vec against SBERT were evaluated in the first test data set to determine the effectiveness of transfer learning without contextual modeling. The model's performance is measured by the hit ratio and the ranking of the selected physician within the set of recommendations.

Ethical Considerations

This study complied with all relevant ethical regulations. All the available data sets have been deidentified and anonymized. The First Affiliated Hospital of Xiamen University Ethics Committee approved this study (approval number SL-2021KY044-01), and no informed consent was necessary.

Results

Data Set Summary

Among the 646,383 consultation records, there were 193,675 patients and 858 physicians across 44 departments. According to Table 2, which provides a summary at the record level, 32.95% (n=212,983) of the records were created by male patients, while female patients accounted for 67.05% (n=433,400) of the records. The predominant age group among patients was 20 to 39 years, representing 54.6% (n=352,907) of the total number of consultations. Patients most frequently consult senior physicians, who account for 62.65% (n=404,958) of all consultations. Most consultations were initiated between 12 and 17 hours, accounting for 37.04% (n=239,401) of the total, while the bulk of responses were received between 18 and 23 hours, accounting for 40.94% (n=208,417). The average response time of the physicians was 3 hours and 40 minutes, with an average response rate of 65.2% (n=421,441).

Table 2. Summary of the characteristics of the collected data records (N=646,383).

Characteristic	Value, n (%)
Gender	
Male	212,983 (32.95)
Female	433,400 (67.05)
Intersex	0 (0)
Age group (y)	
<20	118,484 (18.33)
20-39	352,907 (54.6)
40-59	125,957 (19.49)
>60	49,035 (7.58)
P hy sician s' professional title	
Junior	10,766 (1.67)
Intermediate	45,892 (7.1)
Subsenior	184,767 (28.58)
Senior	404,958 (62.65)
Consulted created moment (h)	
0-5	15,686 (2.43)
6-11	195,297 (30.21)
12-17	239,401 (37.04)
18-23	195,999 (30.32)
Consultation responded moment (h)	
0-5	13,394 (2.61)
6-11	115,684 (22.73)
12-17	171,634 (33.72)
18-23	208,417 (40.94)
Response status	
True	421,441 (65.2)
False	224,942 (34.8)

Case Analysis

This study recommends the following question, as shown in sample 4 in Table 1: "What does the glucose tolerance test report indicate? Could you please explain it to me?" This was done to verify the feasibility of the PPHR model. Questions similar to the target patient's question and related candidate physicians are displayed in Table 3.

The physicians who were similar to the candidate physicians were identified and included in the recommendation strategy along with response indicators. The codes and scores of the recommended physicians of the PPHR model are displayed in Table 4.

To evaluate the precision of the recommended results, we compared the diagnoses for the consultation question within the recommended results. For example, physician 141's diagnosis includes "gestational diabetes" and "glucose tolerance," while physician 164's diagnosis includes "glucose tolerance" and "diabetes." This information matches the consultation question of sample 4, suggesting that the recommended results are likely accurate.

Table 3. Top-10 most similar questions to the target patient's question and related physicians.

Question code	Cosine similarity	Physician code
394946	0.9766	178
559249	0.9765	423
317643	0.9618	141
409326	0.9419	164
238700	0.9214	456
2416	0.9004	707
173519	0.8990	304
551580	0.8976	330
466072	0.8906	632
93556	0.8861	391

Table 4. Top-10 recommended physicians.

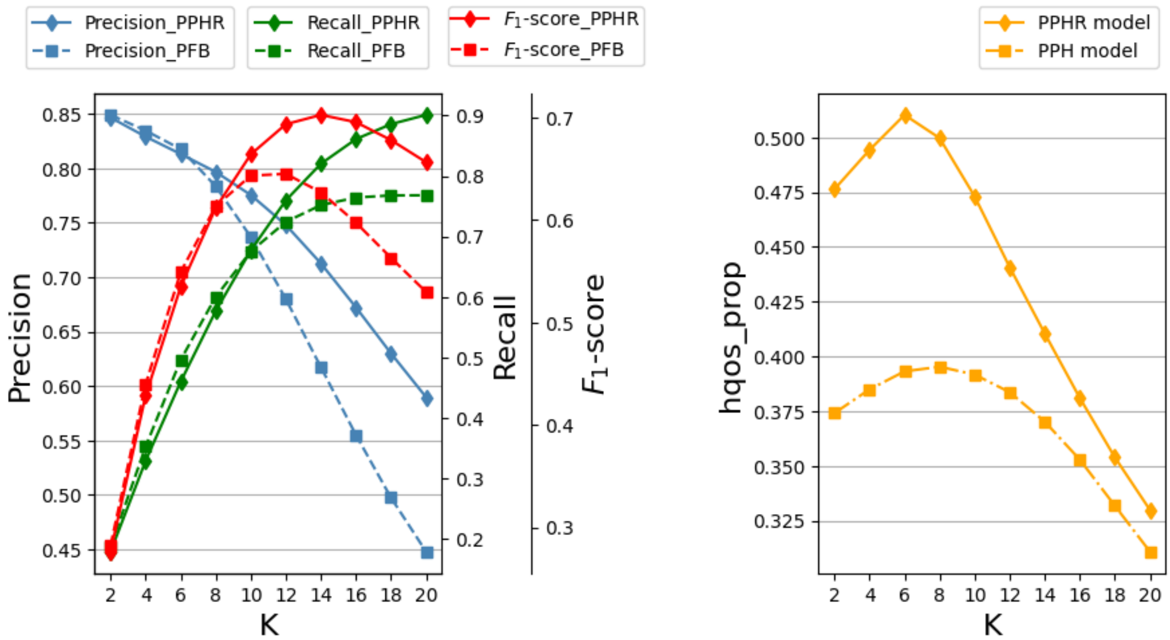
Physician code	score	physician_score
141	7.1877	0.9618
164	5.9507	0.9419
423	4.1309	0.9765
456	4.0309	0.9214
166	3.9408	0.8978
178	3.2916	0.9766
169	3.2082	0.7851
181	2.7208	0.7310
335	2.7096	0.7672
189	2.4849	0.7358

Evaluating the Effectiveness and Service Quality

Figure 3 shows a comparison of the PPHR model with the PFB and PPH models in terms of various indexes. The x-axis represents the number of recommended physicians (K), with values ranging from 2 to 20 and increasing in increments of 2. The PPHR and PFB models exhibit comparable performance when K is <10. However, as K increases to ≥10, the PPHR model demonstrates a marked improvement over the PFB model. The PPHR model achieves its highest F_1-score when K is 14, indicating optimal performance at this level. At this stage, the model has a precision of 71.26%, a recall of 82.02%, and an F_1-score of 76.25%. Compared to the PFB model, the precision has increased by 15.43%, the recall has increased by 9.10%, and the F_1-score has increased by 12.05%. The results presented in Table 4 indicate that although there are minor fluctuations in the rankings of the selected physicians of the PPHR model, its hit ratio has increased by 2.19% compared to the PFB model. These indicate that incorporating physicians' features into the recommendation strategy can improve the effectiveness of the recommendation system.

Regardless of the value of K, the PPHR model provides better hqos_prop than the PPH model. When K is set to 14, the high-quality service ratio of the PPHR model is 41.05%. This is an improvement of 10.87% over the PPH model. This suggests that incorporating the response rate into the recommendation strategy can enhance service quality.

Figure 3. Comparison of our proposed patient-physician hybrid recommendation (PPHR) model with the patient feature–based (PFB) model and patient-physician hybrid (PPH) model in terms of various indexes, including precision (blue), recall (green), F1-score (red), and hqos_prop (orange).



Evaluating the Performance of Text Embedding

The results displayed in Table 5 compare the SBERT model with the Doc2Vec model in terms of text embedding, with K set to 14. The ranking indicates the position of the selected physician within the recommendation set, while the hit ratio represents the percentage of successful recommendations. The SBERT model surpasses the Doc2Vec model by 65.92% in hit ratio, and the rankings of the selected physicians improved by 1.57. These findings suggest that using the SBERT model to create text embeddings could improve the performance of the recommendation system.

Table 5. Performance comparisons were conducted using different models for the selected physicians.

Physician code	Hit ratio			Ranking		
	Doc2Vec-PPHR[a]	SBERT[b]-PPHR	SBERT-PFB[c]	Doc2Vec-PPHR	SBERT-PPHR	SBERT-PFB
98	0.485	0.850	0.820	4.55	3.19	3.04
141	0.555	0.915	0.890	3.99	2.36	2.47
202	0.430	0.800	0.780	4.93	3.47	3.33
512	0.505	0.940	0.925	5.95	3.37	2.73
601	0.695	0.925	0.920	2.78	1.97	1.90
Combined, mean (SD)	0.534 (0.101)	0.886 (0.059)	0.867 (0.064)	4.44 (1.172)	2.87 (0.668)	2.69 (0.549)

[a]PPHR: patient-physician hybrid recommendation.

[b]SBERT: Sentence Bidirectional Encoder Representations from Transformers.

[c]PFB: patient feature–based.

Rationality Evaluation

As can be seen in Table 6, for each physician, the PPHR model consistently received higher ratings than the PFB model. The average rating of the PPHR model is 3.90, which is 11.43% higher than that of the PFB model. The *P* values indicate that the differences in ratings between these 2 models are statistically significant, indicating that the PPHR model is capable of recommending better-performing physicians compared to the PFB model.

Table 6. Rationality evaluations of patient-physician hybrid recommendation (PPHR) model and patient feature–based (PFB) model for selected physicians.

Physician code	Rating		*P* value
	PPHR	PFB	
98	3.89	3.48	.01
141	4.00	3.57	.02
202	3.79	3.33	.02
512	3.83	3.54	.049
601	3.98	3.56	.03
Combined, mean (SD)	3.90 (0.09)	3.50 (0.10)	—[a]

[a]Not applicable.

Discussion

Principal Findings

In this study, we developed an innovative physician triage algorithm named the PPHR model. This model improves the accuracy of matching patients' textual questions with physicians' specialties and optimizes the ranking of candidates according to the physicians' service performance. Consequently, the PPHR model may help increase both the efficiency and the quality of web-based medical services by recommending active physicians with the most appropriate specialties.

Challenges and Solutions for Web-Based Triage Systems

Triage service is a preliminary service in medical diagnosis [29], serving as the first point of contact for patients in health care. It is crucial for improving the efficiency and precision of medical services. In offline outpatient clinics, patients' choices are limited due to the physicians' fixed schedules, especially if the appointment times cannot be changed. Therefore, triage is usually performed at the department level [30]. In contrast, web-based consultation services typically do not adhere to a fixed schedule, and all physicians can provide services on the web, so patients have a wider range of choices. However, the current triage systems have inherited the offline departmental recommendation form, which provides limited assistance to patients. In addition, due to the ongoing division of departments, the naming conventions of these departments have become confusing, resulting in possible overlap in disease areas that physicians specialize in across various departments. Furthermore, with the development of regional medical platforms [31], physicians from different regions and multiple hospitals may share the same web-based consultation platform, which complicates the supply of medical services. Therefore, it is imperative to develop a new type of physician recommendation system.

The construction of a triage system must first consider the matching of the physicians' specialties with the patients' medical needs, which is a prerequisite for the effective operation of web-based medical services [32]. The triage system needs to accurately create user profiles for physicians and patients, analyzing their characteristics, and achieve precise matching. Traditional triage systems typically use the profiles of professional-level physicians. These profiles mainly include the names of diseases and medical fields that physicians specialize in, which can be difficult to match with patients' textual questions. Patients often ask their physicians questions using nonprofessional, colloquial descriptions of symptoms rather than precise disease names. Therefore, using professional descriptions to create physician profiles does not semantically match well with patients' questions. Some studies have attempted to extract physicians' features using textual questions from patients [5,11,12]. Our study draws on this approach, using natural language processing technology to build physician characteristics based on a large corpus of patient inquiries, thus constructing the profiles from the patient's perspective and aligning more closely with patient needs.

In addition to expertise, the quality of service provided by physicians is equally important in ensuring effective web-based diagnosis and treatment. In web-based services, the quality of service is particularly reflected in the response time and rate, as well as the thoroughness of the content provided. Formally, response time and rate are obvious and accessible indicators. Previous research did not consider these indicators when developing triage systems. Therefore, we have included the consideration of the response rate in the scoring calculation of our model's ranking. As there is a correlation between response rate and time, our results showed that this approach also significantly improved the recommended physicians' response time.

Feasibility and Potential Extensions of the Proposed Model

The most significant difference between web-based and offline medical consultations is the limited availability of data. When physicians cannot physically examine patients, the dialogue generated during the consultation becomes the primary source of usable information. Due to potential incompatibilities and lack of data sharing between web-based and offline systems [26], patients' medical histories are often missing on most web-based consultation platforms, making it more challenging to extract patient characteristics. In terms of the quality of medical services, patients' satisfaction with physicians is an indicator that can be referenced. However, the authenticity of satisfaction ratings on different platforms is not always reliable, as many users tend to habitually give positive feedback. Therefore, whether satisfaction ratings should be included in the model remains to be studied and verified. The PPHR model was designed to use minimal information to match physicians with patients. Despite the limited number of variables included, the advantage is that the algorithm is portable across different platforms, offering greater versatility and suitability for widespread adoption. Subsequently, different platforms can modify the model to suit their specific circumstances, including by incorporating past medical histories and satisfaction ratings. They can also adjust various hyperparameters within this framework, such as adjusting the weight of the response rates or setting a different number of recommendations to meet the needs of different platforms. In addition, the user interface can display information that the model used or disregarded, providing additional support for patient decision-making. Considering that some physicians may not be familiar with web-based platforms, it is also feasible to show indicators of their offline services. For instance, it is important to consider whether a physician has a sufficient number of in-person appointments and the level of satisfaction expressed by patients regarding those services. If a physician's expertise is a good match for a specific type of consultation and they have outstanding offline reviews, despite not being highly active on the web, patients could consider switching to offline consultations during clinic hours.

Real-World Application Challenges

Implementing the PPHR model in a real-world scenario presents a number of challenges. First and foremost, ethical issues related to privacy, consent, and confidentiality are major concerns. It is of the utmost importance that patients' medical histories and personal information are handled with the utmost care, in accordance with the patients' consent and in compliance with local and international laws. The PPHR model training process includes the pseudonymization or anonymization of data by removing or replacing personally identifiable information, as well as the use of SBERT for text vectorization, which transforms textual data into numerical vectors that represent the semantic meaning of the text but do not contain explicitly identifiable information, preventing the extraction of personal information directly from the vectors to protect individual privacy. It is also important to consider data confidentiality. Encrypting data at rest and in transit protects against unauthorized access. Using strong encryption standards, such as Advanced Encryption Standard for data at rest and Transport Layer Security for data in transit, can ensure that even if data are intercepted or accessed, they remain unreadable and secure. It is imperative to implement strong authentication mechanisms to verify the identity of users accessing the system; use multifactor authentication to add an extra layer of security; and implement role-based access control to ensure that users can access only the data relevant to their role, maintaining the principle of least privilege.

Integration with existing health care IT systems is another challenge. Many health care providers use legacy systems that may not be immediately compatible with newer models, such as PPHR. This requires the development of interfaces or middleware that can seamlessly connect the model to various health care IT infrastructures without disrupting existing workflows.

In addition, patient and physician acceptance is an integral part of the implementation process. Many users may be skeptical or resistant to changing traditional consultation methods. Educating both physicians and patients about the benefits of the PPHR model, such as increased efficiency; better physician-patient matching; and, ultimately, improved health care outcomes, is critical to facilitating adoption.

Evolving Text Feature Extraction

Text embedding is a fundamental method for text feature extraction, where Doc2Vec is an effective means of implementing text embedding [33]. However, with the advent of transformer-based models, previous text embedding methods are gradually being replaced by SBERT in the industrial service sector. First, Doc2Vec provides a static embedding for each word, best used for tasks that can benefit from representations without the need for understanding word-context relationships [34]. SBERT provides dynamic contextual embeddings that allow for a deeper understanding of the meaning of words in context. It also has the ability to transfer knowledge and analyze subwords [35], which are essential for more complex language comprehension tasks.

Second, the computational efficiency of SBERT compared to traditional methods such as Doc2Vec is primarily influenced by its transformer architecture. Transformers take advantage of parallel processing, which significantly speeds up the training phase. However, they also tend to be resource-intensive, primarily due to the need for larger memory footprints to handle the contextual embeddings and underlying mechanisms. For the PPHR model, this means that there could be increased computational requirements, especially when processing a large corpus of patient queries or generating physician profiles.

Third, when considering the scalability of SBERT within the PPHR model for widespread use, several strategies can help mitigate potential challenges. Horizontal scaling, or adding more machines to spread the load, is a straightforward approach but can increase costs. More efficient strategies include the use of cloud-based services that offer dynamic scaling options to accommodate fluctuating demand without the need for constant, high-capacity infrastructure. Another key consideration is optimizing input sequences. By limiting the length of textual input without losing critical information, the PPHR model can reduce the processing required for each query, making the system more responsive. In addition, caching frequently accessed embeddings and using batch processing for embedding generation can significantly reduce the overall computational load.

It also indicates that as technology progresses, the underlying technical components of models must be regularly updated and refined to enhance the system's overall efficiency. This is a real-world challenge that any web-based medical triage system in operation will encounter.

Limitations and Future Directions

There are some limitations and further solutions. First, it is important to collect data from multiple sources. This study was limited to 1 hospital, which casts doubt on whether the findings are relevant in different contexts. To overcome this limitation, subsequent research should aim to collect information from various sources to evaluate the efficiency of the proposed algorithm. Second, there were some irrelevant contents in our data sets, for example, questions such as “Doctor, will you be available tomorrow? Where can I find you?” These business process–related questions are often mixed with medical questions describing symptoms and represent noises in the data set [36]. Even though preprocessing methods may reduce these noises, manual involvement might still be necessary to enhance the data quality. Third, a common limitation of deep neural networks [37] is the lack of a natural method to explain their predictive results, which makes it difficult to understand why specific samples are predicted to be similar. Models based on transformers make it very challenging to identify when unfair biases or spurious correlations might drive predictions. Therefore, we have introduced the involvement of physicians. If physicians could provide more information based on order details, such as scores based on the perspectives of professional suitability and willingness to accept orders, it might effectively improve the final performance of the model. Fourth, it is critical to regularly update the physicians’ professional information because this information changes over time. Relying on outdated data can result in less-than-ideal recommendations. To ensure that physicians’ profiles are up to date, a time range feature can be implemented. This feature automatically deletes data beyond the specified time range and periodically updates the model with only the latest data. This approach can improve the chances of making accurate recommendations for active physicians and reduce the chances of those still in training or changing areas of expertise. Fifth, obtaining valid feedback from patients and physicians is essential to validating the model’s benefits to patients in real-world settings on a larger scale. For instance, surveys can be conducted on patients’ use of the system, whether patients have adopted the system’s recommendations, and patients’ feedback on whether the system has been helpful. Observing changes in metrics such as the number of consultations for the same condition before and after using the system, comparing their outcomes, and examining the health economic effects of the system are also important. Finally, we were unfortunately only able to obtain textual data to develop the PPHR model. However, the triage system framework proposed in this paper has the potential to incorporate various types of data beyond text. It is possible to integrate multimodal information, such as text, images, audio, and video, using vector embedding techniques to create new vector features. On the basis of this, calculating similarities could potentially achieve more precise matching.

Conclusions

This paper presents a PPHR model with response metrics that uses natural language processing techniques to tackle web-based medical triage tasks. The system filters out relevant physicians, aiding patients in finding the physicians who best suit their actual medical requirements. This approach has significant practical value and can be incorporated into various health website systems to enhance the quality of physician recommendations.

Acknowledgments

This study was funded by 2021 Xiamen Medical and Health Guidance Project “Research on Theoretical and Empirical Evaluation of Internet Hospital Services” (3502Z20214ZD1003); Xiamen Municipal Health Commission, Xiamen Municipal Bureau of Science and Technology (3502Z20209005); Fujian Provincial Clinical Research Center for Brain Diseases (2021FJSLCYX01); and Xiamen Clinical Research Center for Neurological Diseases (2021XMSLCYX0). The authors are grateful to all the participators who contributed to this study.

Data Availability

The data sets generated during and analyzed during this study are not publicly available due to privacy protection and data security regulations but are available from the corresponding author on reasonable request.

Authors' Contributions

KG contributed to the study concept and provided overall guidance. YY contributed to the acquisition of data. YZ contributed to model development and manuscript drafting. YC and SC contributed to data annotation. KG and YZ critically revised the manuscript. All authors contributed to the interpretation of the results and gave final approval for the manuscript.

Conflicts of Interest

None declared.

References

1. Haleem A, Javaid M, Singh RP, Suman R. Telemedicine for healthcare: capabilities, features, barriers, and applications. Sens Int 2021;2:100117 [FREE Full text] [doi: 10.1016/j.sintl.2021.100117] [Medline: 34806053]
2. Hajli MN. Developing online health communities through digital media. Int J Inf Manage 2014 Apr;34(2):311-314 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2014.01.006]
3. The 52nd statistical report on China’s internet development. China Internet Network Information Center (CNNIC). 2023. URL: https://www.cnnic.com.cn/IDR/ReportDownloads/202311/P020231121355042476714.pdf [accessed 2024-04-29]

4. Mosadeghrad AM. Factors affecting medical service quality. Iran J Public Health 2014 Feb;43(2):210-220 [FREE Full text] [Medline: 26060745]
5. Lu W, Zhai Y. Self-adaptive telemedicine specialist recommendation considering specialist activity and patient feedback. Int J Environ Res Public Health 2022 May 05;19(9):5594 [FREE Full text] [doi: 10.3390/ijerph19095594] [Medline: 35564988]
6. Mullenbach J, Wiegreffe S, Duke J, Sun J, Eisenstein J. Explainable prediction of medical codes from clinical text. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,. 2018 Jun Presented at: NAACL '18; June 1-6, 2018; New Orleans, LA p. 1101-1111 URL: https://aclanthology.org/N18-1100.pdf [doi: 10.18653/v1/n18-1100]
7. Li F, Yu H. ICD coding from clinical text using multi-filter residual convolutional neural network. Proc AAAI Conf Artif Intell 2020 Feb 03;34(5):8180-8187 [FREE Full text] [doi: 10.1609/aaai.v34i05.6331] [Medline: 34322282]
8. Wang J, Zhang G, Wang W, Zhang K, Sheng Y. Cloud-based intelligent self-diagnosis and department recommendation service using Chinese medical BERT. J Cloud Comp 2021 Jan 15;10(1):4. [doi: 10.1186/S13677-020-00218-2]
9. Liu H, He J, Wang T, Song W, Du X. Combining user preferences and user opinions for accurate recommendation. Electron Commer Res Appl 2013 Jan;12(1):14-23 [FREE Full text] [doi: 10.1016/j.elerap.2012.05.002]
10. Sharma D, Singh Aujla G, Bajaj R. Evolution from ancient medication to human - centered Healthcare 4.0: a review on health care recommender systems. Int J Communication 2019 Sep 09;36(12):e4058 [FREE Full text] [doi: 10.1002/dac.4058]
11. Ju C, Zhang S. Doctor recommendation model based on ontology characteristics and disease text mining perspective. Biomed Res Int 2021;2021:7431199 [FREE Full text] [doi: 10.1155/2021/7431199] [Medline: 34426788]
12. Liu J, Li C, Huang Y, Han J. An intelligent medical guidance and recommendation model driven by patient-physician communication data. Front Public Health 2023 Jan 26;11:1098206 [FREE Full text] [doi: 10.3389/fpubh.2023.1098206] [Medline: 36778565]
13. Bayeh Y, Tsehay CT, Negash WD. Health system responsiveness and associated factors for delivery care in public health facilities, Dessie City Administration, South Wollo zone, Ethiopia: cross-sectional study design. BMJ Open 2023 Jul 21;13(7):e069655 [FREE Full text] [doi: 10.1136/bmjopen-2022-069655] [Medline: 37479512]
14. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics. 2019 Presented at: NAACL-HLT '19; June 2-7, 2019; Minneapolis, MN p. 4171-4186 URL: https://aclanthology.org/N19-1423.pdf
15. Liu X, Esser D, Wagstaff B, Zavodni A, Matsuura N, Kelly J, et al. Capsule robot pose and mechanism state detection in ultrasound using attention-based hierarchical deep learning. Sci Rep 2022 Dec 07;12(1):21130 [FREE Full text] [doi: 10.1038/s41598-022-25572-w] [Medline: 36476715]
16. Mitchell JR, Szepietowski P, Howard R, Reisman P, Jones JD, Lewis P, et al. A question-and-answer system to extract data from free-text oncological pathology reports (CancerBERT network): development study. J Med Internet Res 2022 Mar 23;24(3):e27210 [FREE Full text] [doi: 10.2196/27210] [Medline: 35319481]
17. Chen W, Zhong C, Peng J, Wei Z. DxFormer: a decoupled automatic diagnostic system based on decoder-encoder transformer with dense symptom representations. Bioinformatics 2023 Jan 01;39(1):btac744 [FREE Full text] [doi: 10.1093/bioinformatics/btac744] [Medline: 36409016]
18. Arnaud É, Elbattah M, Gignon M, Dequen G. Learning embeddings from free-text triage notes using pretrained transformer models. In: Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies. 2022 Presented at: BIOSTEC '22; February 9-11, 2022; Virtual Event p. 835-841 URL: https://uwe-repository.worktribe.com/index.php/output/9187291/learning-embeddings-from-free-text-triage-notes-using-pretrained-transformer-models [doi: 10.5220/0011012800003123]
19. Reimers N, Gurevych I. Sentence-BERT: sentence embeddings using Siamese BERT-networks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. 2019 Presented at: EMNLP-IJCNLP '19; November 3-7, 2019; Hong Kong, China p. 3982-3992 URL: https://aclanthology.org/D19-1410.pdf [doi: 10.18653/v1/d19-1410]
20. Alksher MA, Azman A, Yaakob R, Kadir RA, Mohamed A, Alshari EM. A review of methods for mining idea from text. In: Proceedings of the 2016 3rd International Conference on Information Retrieval and Knowledge Management. 2016 Presented at: CAMP '16; August 23-24, 2016; Malacca, Malaysia p. 88-93 URL: https://ieeexplore.ieee.org/document/7806341 [doi: 10.1109/infrkm.2016.7806341]
21. Abduljabbar DA, Omar N. Exam questions classification based on bloom’s taxonomy cognitive level using classifiers combination. J Theor Appl Inf Technol 2015;78:447-455 [FREE Full text]
22. Neugebauer J, Tóthová V, Doležalová J. Use of standardized and non-standardized tools for measuring the risk of falls and independence in clinical practice. Int J Environ Res Public Health 2021 Mar 20;18(6):3226 [FREE Full text] [doi: 10.3390/ijerph18063226] [Medline: 33804715]
23. Scherer R, Rao R. Non-manual control devices: direct brain-computer interaction. In: Pereira J, editor. Handbook of Research on Personal Autonomy Technologies and Disability Informatics. New York, NY: IGI Global; 2011:233-250.

24. Xu S, Leng Y, Feng G, Zhang C, Chen M. A gene pathway enrichment method based on improved TF-IDF algorithm. Biochem Biophys Rep 2023 Jul;34:101421 [FREE Full text] [doi: 10.1016/j.bbrep.2023.101421] [Medline: 36923007]
25. Jones KS. A statistical interpretation of term specificity and its application in retrieval. In: Willett P, editor. Document Retrieval System. New York, NY: Taylor Graham Publishing; 1988:132-142.
26. Li C, Liu J, Qian G, Wang Z, Han J. Double chain system for online and offline medical data sharing private and consortium blockchain: a system design study. Front Public Health 2022;10:1012202 [FREE Full text] [doi: 10.3389/fpubh.2022.1012202] [Medline: 36304235]
27. Fay MP, Malinovsky Y. Confidence intervals of the Mann Whitney parameter that are compatible with the Wilcoxon-Mann-Whitney test. Stat Med 2018 Nov 30;37(27):3991-4006 [FREE Full text] [doi: 10.1002/sim.7890] [Medline: 29984411]
28. Le Q, Mikolov T. Distributed representations of sentences and documents. In: Proceedings of the 31st International Conference on International Conference on Machine Learning. 2014 Presented at: ICML '14; June 21-26, 2014; Beijing, China p. 1188-1196 URL: https://dl.acm.org/doi/10.5555/3044805.3045025
29. Sexton V, Dale J, Bryce C, Barry J, Sellers E, Atherton H. Service use, clinical outcomes and user experience associated with urgent care services that use telephone-based digital triage: a systematic review. BMJ Open 2022 Jan 03;12(1):e051569 [FREE Full text] [doi: 10.1136/bmjopen-2021-051569] [Medline: 34980613]
30. Liu TT, Cheng CT, Hsu CP, Chaou CH, Ng CJ, Jeng MJ, et al. Validation of a five-level triage system in pediatric trauma and the effectiveness of triage nurse modification: a multi-center cohort analysis. Front Med (Lausanne) 2022;9:947501 [FREE Full text] [doi: 10.3389/fmed.2022.947501] [Medline: 36388924]
31. Zhang L, Wang X, Xiao H, Ma C, Li X, Dai G, et al. Governance mechanisms for chronic disease diagnosis and treatment systems in the post-pandemic era. Front Public Health 2022;10:1023022 [FREE Full text] [doi: 10.3389/fpubh.2022.1023022] [Medline: 36582374]
32. Fekonja U, Strnad M, Fekonja Z. Association between triage nurses' job satisfaction and professional capability: results of a mixed-method study. J Nurs Manag 2022 Nov;30(8):4364-4377 [FREE Full text] [doi: 10.1111/jonm.13860] [Medline: 36206481]
33. Asudani DS, Nagwani NK, Singh P. Impact of word embedding models on text analytics in deep learning environment: a review. Artif Intell Rev 2023 Feb 22:1-81 [FREE Full text] [doi: 10.1007/s10462-023-10419-1] [Medline: 36844886]
34. Kinariwala S, Deshmukh S. Short text topic modelling using local and global word-context semantic correlation. Multimed Tools Appl 2023 Feb 02:1-23 [FREE Full text] [doi: 10.1007/s11042-023-14352-x] [Medline: 36747894]
35. Argote L. Knowledge transfer within organizations: mechanisms, motivation, and consideration. Annu Rev Psychol 2024 Jan 18;75:405-431 [FREE Full text] [doi: 10.1146/annurev-psych-022123-105424] [Medline: 37788572]
36. Wang YM, Sun Y, Wang B, Wu Z, He XY, Zhao Y. Transfer learning for clustering single-cell RNA-seq data crossing-species and batch, case on uterine fibroids. Brief Bioinform 2023 Nov 22;25(1):bbad426 [FREE Full text] [doi: 10.1093/bib/bbad426] [Medline: 37991248]
37. Kuperwajs I, Schütt HH, Ma WJ. Using deep neural networks as a guide for modeling human planning. Sci Rep 2023 Nov 20;13(1):20269 [FREE Full text] [doi: 10.1038/s41598-023-46850-1] [Medline: 37985896]

Abbreviations

BERT: Bidirectional Encoder Representations from Transformers
PFB: patient feature–based
PPH: patient-physician hybrid
PPHR: patient-physician hybrid recommendation
SBERT: Sentence Bidirectional Encoder Representations from Transformers
TF-IDF: term frequency–inverse document frequency

Edited by C Jacob; submitted 23.02.24; peer-reviewed by M Elbattah, G Kang; comments to author 01.05.24; revised version received 02.06.24; accepted 22.06.24; published 15.08.24.

Please cite as:
Zheng Y, Cai Y, Yan Y, Chen S, Gong K
Novel Approach to Personalized Physician Recommendations Using Semantic Features and Response Metrics: Model Evaluation Study
JMIR Hum Factors 2024;11:e57670
URL: https://humanfactors.jmir.org/2024/1/e57670
doi:10.2196/57670
PMID:39146009

Original Paper

Shaping Adoption and Sustained Use Across the Maternal Journey: Qualitative Study on Perceived Usability and Credibility in Digital Health Tools

Wei Ying Ng[1,2*], BSocSci; Ni Yin Lau[1,2*], BSocSci; V Vien Lee[1,2], PhD; Smrithi Vijayakumar[1,2], PhD; Qiao Ying Leong[1,2], BSc; Shu Qin Delicia Ooi[3], PhD; Lin Lin Su[4,5], MBBS; Yung Seng Lee[3,6], MBBS, PhD; Shiao-Yng Chan[4,7], MBBCHIR, PhD; Agata Blasiak[1,2,8,9], PhD; Dean Ho[1,2,8,9,10], PhD

[1]The N.1 Institute for Health, National University of Singapore, Singapore, Singapore

[2]The Institute for Digital Medicine (WisDM), Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[3]Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[4]Department of Obstetrics & Gynaecology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[5]Department of Obstetrics & Gynaecology, National University Hospital, Singapore, Singapore

[6]Division of Paediatric Endocrinology, Khoo Teck Puat-National University Children's Medical Institute, National University Hospital, Singapore, Singapore

[7]Singapore Institute for Clinical Sciences, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore

[8]Department of Biomedical Engineering, National University of Singapore, Singapore, Singapore

[9]Department of Pharmacology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[10]The Bia-Echo Asia Centre for Reproductive Longevity and Equality, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[*]these authors contributed equally

Corresponding Author:
Dean Ho, PhD
The N.1 Institute for Health
National University of Singapore
28 Medical Drive
#05-COR
Singapore, 117456
Singapore
Phone: 65 66017766
Email: biedh@nus.edu.sg

Abstract

Background: Maternal and child health outcomes are positively influenced by early intervention, and digital health (DH) tools provide the potential for a low-cost and scalable solution such as informational platforms or digital tracking tools. Despite the wide availability of DH tools out there for women from before to after pregnancy, user engagement remains low.

Objective: This study aims to explore the factors that shape women's DH adoption and sustained use across the maternal journey from preconception to postbirth, to improve user engagement with DH tools.

Methods: One-hour semistructured qualitative interviews were conducted with 44 women from before to after pregnancy (age range 21-40 years) about their experiences with DH. This study is part of a larger study on women's maternal experiences with health care and DH and focuses on the factors that affected women's DH adoption and sustained use. Interviews were audio recorded, transcribed verbatim, and analyzed using inductive thematic analysis.

Results: Five main themes and 10 subthemes were identified that affected women's adoption and sustained use of DH tools. These included themes on their preexisting attitudes to DH, perceived ease of use, perceived usefulness, perceived credibility, and perceived value of the tool.

Conclusions: The themes that emerged were fully or partially mapped according to the Unified Theory of Acceptance and Use of Technology 2 model. The applicability of the model and the need to consider specific cultural nuances in the Asian context (such as the importance of trust and social influence) are discussed. The interaction of the 5 themes with DH adoption and sustained use are explored with different themes being relevant at various points of the DH adoption journey. The insights gained serve to inform future DH design and implementation of tools for women to optimize their DH engagement and the benefits they derive from it.

Trial Registration: ClinicalTrials.gov NCT05099900; https://clinicaltrials.gov/study/NCT05099900

(JMIR Hum Factors 2024;11:e59269) doi:10.2196/59269

KEYWORDS

maternal and child health; conception; pregnancy; perinatal care; postpartum; maternal care; obstetric care; user engagement; Unified Theory of Acceptance and Use of Technology; femtech

Introduction

The lifelong health of an individual can be linked to periods even before conception and birth, as maternal health and behaviors have been shown to influence child health [1]. Many countries including Singapore, have worked on improving the quality of maternal health care [2]. In recent years, the challenges in maternal health have shifted from lowering maternal and infant mortality rates to managing complicated health outcomes such as obesity in women before, during, and after pregnancy [2]. Unhealthy weight in women can have implications for both mothers (eg, attempts at conception, pregnancy-related complications, and mental health disorders) and their children (eg, increased risks of childhood obesity and developmental problems later in life) [1,2]. These are known as the “twin metabolic and mental health challenges” as characterized in the S-PRESTO and GUSTO mother-offspring cohort studies in Singapore [3,4]. Evidently, managing maternal health outcomes has both direct and indirect consequences on maternal and child health [1-4]. As such, early intervention in maternal health decisions, as early as preconception, has the potential to positively influence maternal and child health outcomes [2].

A multitude of digital health (DH) tools have been developed to improve multiple maternal and child outcomes [5-14]. DH tools can take the form of digital platforms, such as smartphones, websites, forums, social media, and wearables, which also enable tracking of health data [6,15]. Due to their accessibility and the potential for low-cost implementation at scale, DH tools have the capacity to easily disseminate information and interventions for behavioral change to a large audience [15]. For maternal and birth outcomes, DH tools have been applied in numerous ways to improve and manage several conditions beyond obesity. These included fertility education apps [10], tracking of symptoms and complications through a continuum from before to after pregnancy [5,7], improving the delivery of pregnancy and antenatal care [6,11,14], early labor support apps [13], and promoting postpartum health screening via DH platforms [12]. These have been widely implemented in many geographies, from European to African countries [7,10,16-18]. Accelerated by the recent COVID lockdowns [19], DH tools present an opportunity to satisfy the health care, informational, and support needs of women throughout their journey from preconception to postbirth.

To realize the potential of DH tools, attention needs to be given to their adoption and retention by the users. User engagement with the DH tool is vital in maximizing the potential benefits derived from it [18,20]. The degree of user engagement determines the depth, breadth, and length of tool use [9,21]. Studies investigating the engagement of DH tools among pregnant women report a lack of user engagement. One study examining the uptake of digital antenatal care services by pregnant mothers found poor adoption of DH due to a lack of awareness and cost barriers [16]. Other studies found that despite incentives to motivate engagement, women struggled to sustain long-term use [18]. These findings have been observed across diverse settings, ranging from high-income countries like Germany to low-income countries like Uganda. Evidently, user engagement is affected by differences in women’s accessibility to and knowledge of the DH tools [17], emphasizing the need to actively identify reasons driving unequal adoption of DH tools [22]. The needs of women evolve as they progress through their reproductive stages and thus require varying interventions at different time points [23,24].This calls for more research to understand the factors that affect the adoption and sustained use of DH tools by women to supplement current research on the perspectives of clinicians [25]. Insights gained from women through their experiences can serve to better inform DH innovators, from the early design and development process to the implementation phase.

As such, this study seeks to explore and understand the factors that influence women in their adoption and sustained use of DH tools before, during, and after pregnancy. This study provides a continuation of research from a previous study by Lee et al [26], using the same pool of data obtained from performing 44 interviews with women across preconception to postbirth phases but focusing on responses specific to technology adoption. While Lee et al [26] primarily focused on key elements in DH interventions that promote healthy behavior change in women from before to after pregnancy, this study identified 5 main themes that support DH uptake and sustained use. The translational insights derived from this study aim to inform the future design and implementation of DH tools for women across preconception to postbirth, to create a positive impact on behavioral and health outcomes in both mother and child.

Methods

Ethical Considerations

The study procedures were approved by the National Healthcare Group Domain Specific Review Board (reference 2021/00034). Informed consent was obtained from participants by providing the consent form during initial contact and prior to the interview session. Additionally, the researcher explained the consent form fully before participants consented. Participation was voluntary and participants could withdraw at any point of the study. There were no direct risks involved for participants and any data collected were deidentified, encrypted, and stored in accordance with the institution's data management policies. Participants were reimbursed SG $30 (US $23.06) for their time spent on the study.

Recruitment

This study recruited women who were trying to conceive, were pregnant at the time of contact, or had a child aged 0 to 2 years through purposive sampling from the National University Hospital and the community in Singapore (ClinicalTrials.gov Identifier: NCT05099900). Study advertisements were distributed around National University Hospital, public places (eg, bus stops, housing estates, and learning institutions), and social media platforms (eg, Telegram and Facebook). The research team provided an overview of the study to potential participants who responded to the advertisement via email, phone call, or messaging. The team screened interested participants based on the following inclusion criteria: (1) English fluency; (2) aged 21 to 45 years; and (3) actively trying to conceive or currently in first to third trimester of pregnancy or have a child aged 0 to 2 years. Participants were not eligible for the study if they met the following exclusion criteria: (1) evidence or diagnosis of cognitive impairment; (2) current diagnosis of psychiatric disorder; (3) significant hearing impairment; (4) women requiring or who had any form of assisted conception; and (5) inability to complete the study at the judgment of the clinician investigators. The research team established communication with eligible participants to schedule the interview sessions. Recruitment took place over a period of 9 months (from November 2021 to July 2022) and ended when data saturation was achieved for each group (ie, preconception, during pregnancy, and postbirth). No participants declined to participate or dropped out after consenting to the study.

Data Collection

This study adopted a qualitative approach to explore women's experiences from preconception to postbirth, including their DH experiences and expectations. Prior to the interview session, participants completed a web-based questionnaire (Multimedia Appendix 1) regarding their demographics and technology use patterns. Before commencing the interview, participants were informed about study goals and researchers' interest in the research topic. All participants provided informed consent to the audio recording of the interview for the research team's transcription purposes. A 60- to 90-minute semistructured interview was conducted either in-person or virtually via Zoom, depending on participants' preference. Open-ended questions were used to facilitate discussion surrounding participants' experiences with their preconception, pregnancy, or postbirth journey, and their use or expectations of DH to support their journey. Both the questionnaires and interview guide have been detailed in Lee et al [26]. Guiding topics for the interview are presented in Multimedia Appendix 1 [26]. This manuscript is focused on the DH segment of the interview, specifically exploring women's experiences with DH and current expectations of how DH could support their maternal journey. Given that this manuscript presented a different focus on technology adoption from Lee et al [26], responses included in this work are different and were not reused between the manuscripts. All interviews were conducted in English with at least 2 researchers present. The interviewing team comprised of VVL, SV, WYN, NYL, and QYL—all female and trained in qualitative research.

Following the interview session, participants completed another web-based questionnaire (Multimedia Appendix 1) regarding their pregnancy concerns and DH expectations. All participants were reimbursed for their time. No repeat interviews were carried out and transcripts were not returned to participants. Reflective notes were taken after the interview to consider additional guiding prompts for future interviews.

All data collected, including signed consent forms, interview recordings, and questionnaire data, were deidentified, encrypted, and stored in a secure database.

Data Analysis

For the questionnaire, descriptive analyses were conducted using SPSS (IBM Corp). Pairwise deletion was employed to handle missing data in the questionnaire responses. For the interviews, audio recordings were transcribed verbatim. Inductive thematic analysis was conducted to identify emerging and recurring themes. First, transcripts were randomly assigned to the 5 interviewing researchers to conduct primary coding to descriptively label the data. All generated primary codes were then compared and discussed among all researchers to resolve any discrepancies. Thereafter, secondary coding, where labeled data were grouped into categories, was conducted independently for each group (preconception, pregnancy, and postbirth) using Microsoft Excel. As the categories that emerged from secondary coding were similar across all 3 groups, the categories were analyzed into broader, overarching themes. The final set of codes and broader themes was concluded for the study data after discussions and iterations by all researchers. No feedback was provided by participants on the findings.

Results

Participant Characteristics

The cohort used in this study was previously described in Lee et al [26], where detailed demographic characteristics were provided (Multimedia Appendix 1). A total of 44 participants (age range 21-40 years; mean age 31.6, SD 4.0 years) completed the study. Participants across preconception (13/44, 29.5%), pregnancy (16/44, 36.4%), and postbirth (15/44, 34.1%) phases were recruited. Participants were Chinese (33/44, 75%), Malay (4/44, 9.1%), Indian (4/44, 9.1%), and other (3/44, 6.8%) ethnicities. In terms of education, there was an equal distribution of participants who had 15 years of education or less (22/44, 50%) and those who had more than 15 years of education (22/44, 50%). Socioeconomic status (SES) varied, with participants categorized as low SES (9/44, 20.5%), middle SES (13/44, 29.5%), and high SES (22/44, 50%). Family size ranged widely, with participants having had no children (17/44, 38.6%), 1 child (16/44, 36.4%), and 2 or more children (11/44, 25%) at the point of the study. All participants spoke fluent English.

The technology use patterns and expectations obtained from the questionnaire responses are shown in Table 1. One participant from the postbirth group was entirely excluded from the table due to nonresponse to the questionnaire. Two participants (one from the pregnancy and one from the postbirth group) had missing data due to incomplete questionnaire responses. Women indicated preferred online sources of information and information topics. For instance, Google and child development–related information were commonly perceived as a useful information source and topic respectively. Most women could accept using DH tools that required weekly logging of information, and receiving feedback based on data logged was also rated by women as a feature that they were likely to use DH tools for. Participants across the 3 phases also used a variety of DH apps, including apps for physical health, mental health, fertility, pregnancy, and child-related care, which has been detailed in Lee et al [26]. While the tables presented in Lee et al [26] focused on categories of apps being used, Table 1 shows specific technology use patterns and aspects of DH tools participants would like to see. Multimedia Appendix 1 provides a breakdown of the categories of mobile phone application use across the 3 phases.

Table 1. Technology use patterns and expectations of study participants from questionnaire responses.

Technology use and category	Preconception (n=13)	Pregnancy (n=16)	Postbirth (n=15)
Useful online sources of information (n=42), n (%)			
Chat group	6 (46.2)	8 (53.3)	6 (42.9)
Online forum	11 (84.6)	7 (46.7)	8 (57.1)
Google	9 (69.2)	12 (80)	12 (85.7)
Mobile phone apps	5 (38.5)	10 (66.7)	9 (64.3)
Social media	8 (61.5)	9 (60)	7 (50)
Acceptable frequency of logging information (n=42), n (%)			
Daily	5 (38.5)	5 (31.3)	4 (30.8)
Weekly	8 (61.5)	10 (62.5)	7 (53.8)
Monthly	0 (0)	1 (6.3)	1 (7.7)
Biweekly	0 (0)	0 (0)	1 (7.7)
Likelihood of using DH[a] platform for respective features (n=43; Scoring: 0=Extremely unlikely, 2=Somewhat unlikely, 3=Neither unlikely or likely, 4=Somewhat likely, 5=Extremely likely), mean (SD)			
Complete questionnaires	2.85 (1.07)	2.38 (0.96)	2.53 (1.03)
Log physical health data	3.46 (0.52)	3.00 (0.52)	2.93 (1.14)
Log mental health data	2.69 (0.95)	2.75 (0.78)	2.36 (1.15)
Feedback based on data logged	3.00 (0.58)	3.13 (0.50)	3.07 (0.83)
Lifestyle guidelines and advice	3.38 (0.65)	3.00 (0.37)	3.14 (0.86)
Peer support	2.85 (1.07)	2.88 (0.89)	3.21 (0.80)
Breastfeeding and weaning information	3.08 (1.12)	3.00 (0,73)	3.5 (0.65)
Connect to wearables	2.54 (1.27)	2.75 (1.29)	3.14 (1.10)
Pair with digital tools	2.54 (1.33)	2.44 (1.21)	2.71 (1.33)
Preferred information topics on DH platform (n=42), n (%)			
Developmental information of child	13 (100)	16 (100)	11 (84.6)
Mental health resources	6 (46.2)	11 (68.8)	11 (84.6)
Physical activity	10 (76.9)	13 (81.3)	8 (61.5)
Helpline and health provider contact details	10 (76.9)	9 (56.3)	10 (76.9)
Others	Types and cost of local fertility or maternity services	Postpartum recovery resources, gestational diabetes mellitus meal plan suggestions, activities with newborn, coping with changes after delivery (ie, physical, mental, work, and social aspects)	—[b]

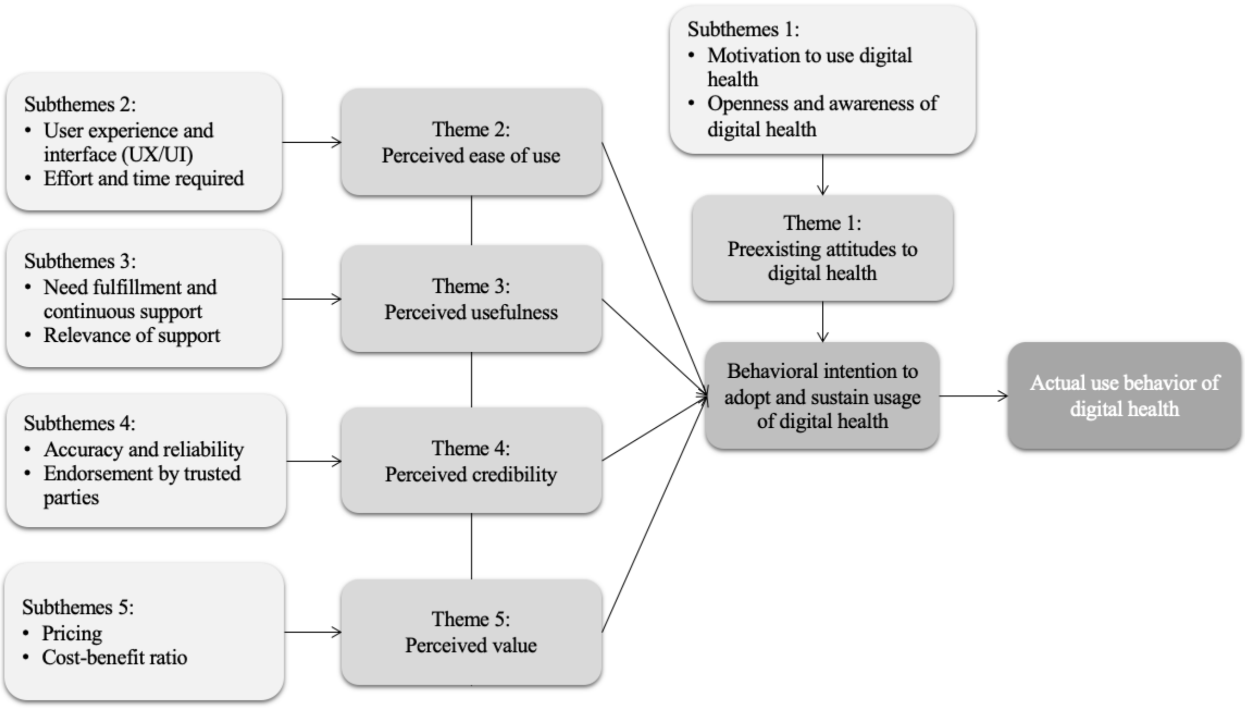
[a]DH: digital health.

[b]Not applicable.

Interview Data

A total of 5 themes and 10 subthemes were identified as factors that motivate or hinder women's behavioral intention toward adoption and sustained use of DH tools (Figure 1).

Figure 1. Factors influencing women's behavioral intention to adopt and sustain use of digital health tools.



Theme 1: Preexisting Attitudes Toward DH

Existing attitudes are defined as women's motivation to use DH tools, as well as their openness and awareness of DH tools.

Motivation to Use DH

Prior experiences and understanding of DH shaped women's current motivations to adopt DH tools. Women expressed having low motivation due to previous unpleasant experiences with DH tracking tools whereby tracking lifestyle metrics such as diet, physical activity, and water intake felt like a chore. Some women disengaged because they lacked the motivation to enter data manually, while others found it stressful to adhere to the requirement of regular data entry.

> *Last time I used to track my water intake, but got a bit lazy...because [you've] got to key in manually so I stopped that.* [Participant 5, preconception]
>
> *Previously I was tracking my diet but it got too troublesome.* [Participant 18, postbirth]
>
> *After like one, two months, I start to get very stressed in order to meet this routine [and] having to comply to the app, I think that would be more stressful to me...Do a food diary, nope that's [going to] be more stressful...* [Participant 41, preconception]

Such poor motivation was also attributed to poor understanding of the impact of lifestyle factors on maternal and child health, and how tracking lifestyle metrics might be useful.

> *I don't really know how sleep can impact, [so] I wouldn't really track sleep.* [Participant 33, preconception]
>
> *So like of course [diet tracking is] good to have but I don't have [the] confidence I will use it. I also don't know what [features of diet tracking are] good.* [Participant 18, postbirth]

Openness and Awareness of DH

Women were generally open to using DH tools to manage their pregnancy journey if they were aware of such tools, provided that the tool met their need for it to be easy to use, require little time, and provide health monitoring benefits.

> *But of course, if there's an app that can help me monitor my health better, I might be willing to try it.* [Participant 34, preconception]
>
> *I think if it was too complex to use, or if the app made me spend too much time on it, then I might not want to be so hooked on my digital device, especially if I have young kids.* [Participant 23, postbirth]

Women also suggested that outreach efforts to increase knowledge and awareness of a DH tool and its benefits might also encourage the adoption of the tool.

> *I think that if they come and engage me...like do you know that this app allows you to do this and do that. I'm like "okay!"* [Participant 12, preconception]

Theme 2: Perceived Ease of Use

Based on the interviews, DH tools' perceived ease of use was contingent upon factors such as user experience and interface of the digital platform, as well as how much effort and time were required from the user in their use of the DH tool.

User Experience and Interface (UX/UI)

Women's intention to continue the use of a tool after initial adoption was highly dependent on the user experience and interface of the DH tool. Unappealing interfaces that were challenging to navigate, slow loading times, or discomfort in user experiences were major sources of frustration that led to disengagement with a digital tool.

> *Just [that] the outlook was not very appealing...maybe the colour, the outlook, the way it is placed, not very user friendly. It's a bit confusing on how to use it.* [Participant 31, during pregnancy]

> *For me like the UX is important also, the ease of use and the features and functionalities of the app.* [Participant 4, during pregnancy]

> *When I needed to track those information [for my baby], I downloaded a few apps to try.*

> *If its' hard to use, or the interface is not so...erm I don't feel comfortable, I'll just straightaway delete. Until I find one that's comfortable, then I just stick to it.* [Participant 1, postbirth]

Hardware and software issues of a DH tool such as the extent of battery consumption, data use, storage space, and data syncing problems lowered the likelihood of adoption and sustained use. Women were also highly likely to delete an app if they experienced frequent technical glitches.

> *[I used to use] HealthHub, but I'm not using [it] now because the syncing has some issues.* [Participant 37, preconception]

> *Too much data, memory space, [and also] it will be good if [it] doesn't drain too much battery. I think Fitbit drains quite a bit [of battery life].* [Participant 5, preconception]

> *What would stop me [is] if it's always buggy and always hanging, and then like half the time you can't get the app to work, then obviously I would just delete it.* [Participant 38, postbirth]

Level of Effort and Time Required

The perceived ease of use was influenced by the level of effort and time required on the part of the user. Women identified several barriers that prevented continued use of a DH tool, including data entry that required too much time and effort, especially postbirth when women were busy caring for their children. DH tools with a steep learning curve and long onboarding time in the initial stage of use were also the reasons that led women to disengage from digital tools.

> *But after a while I stopped [tracking my diet] because I feel that it was a bit troublesome to have to key in what I eat, then find the exact food and portion size that I ate. I mean, although that did help me to manage my diet, but it was troublesome for me to use in the long run.* [Participant 18, postbirth]

> *So in the beginning I use [an app for breathing exercises] more, but I think when [my child] came, well I had no time to do that anymore* [Participant 8, postbirth]

> *When you first [begin to use] an app, [you] probably need to load quite a bit of information in there. So if that first stage [takes] me a little longer to actually set it all up, I might just give up in the end.* [Participant 34, preconception]

Theme 3: Perceived Usefulness

Another crucial determinant in women's adoption and ongoing use of DH tools was perceived usefulness, which was shaped by the extent of need fulfillment and continuous support provided, as well as the relevance of support.

Need Fulfillment and Continuous Support

It was important to women that DH tools fulfilled their unmet needs and delivered comprehensive support. The idea of providing an all-in-one solution was especially appealing, due to issues like app fatigue as women found it tedious to manage several apps for different health-related functions, or having to filter through overwhelming amounts of information on the internet. Having an all-in-one support tool was pointed to as an enabler of better access to quality information and increased ease in tracking health and lifestyle metrics.

> *I would say the only thing that would make [the DH tool] superior would be the fact that it has everything I need. So I can just delete my Flo app, I don't even have to do my Google search anymore. I mean [I can] reduce my Google search at least.* [Participant 35, preconception]

> *Yea, as in I think that would be helpful because sometimes along the way in the journey then you have questions, and you don't pre-empt the questions way before, so if the, the app covers the whole span then you can obviously uh, look through it at the different stages that you're at in your journey* [Participant 26, postbirth]

Furthermore, it was critical that DH tools delivered complementary support to the current continuum of care provided by the health care system. This was especially prominent for women in the pregnancy and postbirth phases, as many of their needs included touch points with the health care system, such as understanding their pregnancy trimesters or their child's development through doctor consultations or health booklets. Offering complementary support emerged as a feature to increase the likelihood of DH adoption. For example, scheduling appointments and integrating DH records of the child were mentioned as useful tools.

> *If I can use this app as like an all-in-one [tool], [including] scheduling appointments [or] even uploading my ultrasound photos...And you can download the digital copy of [the ultrasound scan], instead of having just the physical [copy], which you might lose or it might degrade over time. These ultrasound scans are quite precious to like the family during the pregnancy [because] that is [the] only time where you can like see the baby.* [Participant 4, during pregnancy]

Relevance of Support

Women highlighted the importance of the relevance of support provided by DH tools. They valued being able to receive tailored information based on their individual needs and preferences. This included information that was relevant to women's current phase in the pregnancy journey, topics of interest, or areas of child development in which they had concerns at that time.

> *What would stop me from using the app is if I'm not interested to know about that particular topic or maybe I don't see that my child has any issues at all in that aspect, then I wouldn't bother to check.* [Participant 26, postbirth]

The desired feedback from tracking maternal indicators was envisioned as being aware of women's conditions and sensitive to their emotional well-being, ensuring that it would not induce additional stress.

> *It really depends on the circumstances. Let's say I'm producing a lot of breast milk [and] I'm keying it into the app, then I wouldn't mind seeing it. But let's say I'm a mother who doesn't produce enough [breastmilk] and I'm keying it into the app. Then, I look at it and I start to feel depressed. Then, I probably would not enter it again.* [Participant 33, preconception]

The relevance of support was particularly important as women were often overwhelmed by the sheer amount of information available through technology. Often, such "knowledge burden" (participant 30) became unhelpful and a source of anxiety, as women needed to have sufficient mental capacity to filter through information and determine what was useful and appropriate for their circumstances.

> *Maybe you read too much and then you give yourself anxiety. Let's say [the information states] you cannot eat [too much of this], then you'll [wonder if you're] eating too much. So sometimes, in a way it's more like [a] knowledge burden. Yeah but you need to know how to filter.* [Participant 30, preconception]

Theme 4: Perceived Credibility

The extent to which DH tools were deemed credible was described to have a notable impact on their adoption and long-term use. Accuracy and reliability of information and support, along with endorsement by trusted parties through social networks, played a role in shaping the perceived credibility of a DH tool.

Accuracy and Reliability

Women indicated the need to perceive the information and support provided as accurate and reliable. Women brought up challenges in navigating conflicting information from online sources, their social networks, and health care professionals. They suggested that it was essential for DH tools to provide accurate information that aligned with clinical recommendations from health care professionals. Perceived reliability of information was also influenced by the source of information, as women expressed greater trust in the advice provided by medical professionals.

> *Sometimes [information] can be contradicting so I'm not sure which source to trust.* [Participant 9, during pregnancy]

> *Let's just say they have a moderator, a medical professional [in the online forum], that would answer questions here and there, that could be a game changer. Because then, I would look out for this badge – okay this [medical professional] says this. So, that would be an edge over a sea of apps.* [Participant 14, during pregnancy]

> *But I was hoping [for information about] medication [that you take] before pregnancy or after pregnancy, or medication [you take] while you are pregnant, breastfeeding and stuff. But sometimes, [information] is inaccurate. Let's say there's some medication, [like] Panadol [that] you can't take while breastfeeding, but the doctor [says] you can take while you are breastfeeding...Trust the doctor or trust the app?* [Participant 42, postbirth]

Endorsement by Trusted Parties

Women were more inclined to use DH tools recommended by trusted networks. Most women agreed that gaining awareness about DH tools from their clinicians or health care providers was important in influencing their decision to adopt them.

> *Definitely if a healthcare provider recommended it. So a lot of the things I have on my phone are to do with [physiotherapy] and all that, they've all come from the [physiotherapist], the lactation consultant, birth class teacher.* [Participant 7, during pregnancy]

Social networks that women were already situated in also had a strong influence on their decision-making. Recommendations via word-of-mouth from friends, female communities on online support groups or forums, and key influencers on social media platforms were persuasive in encouraging the adoption of DH tools. Additionally, it was important that these trusted parties were situated in the local context. Another commonly brought-up factor that influenced women's choice of DH tools was the sharing of experiences, such as how other women or mothers overcame their struggles with the support of DH tools. Such positive reviews were perceived as credible information that affirmed DH tools' ability to provide adequate and effective support to help women manage similar challenges.

> *Asian Parent [online platform with parenting-related content] was recommended because I know of friends who used it. So, I think word of mouth is always more powerful.* [Participant 14, during pregnancy]

> *Recommendations from friends or mommies. If I was on the internet, and it popped up, [saying that]this is the number 1[app]. Not trending app but in your area, maybe, within Asia or within Singapore, this is the app most mommies are using.* [Participant 7, during pregnancy]

Women also considered satisfaction ratings and reviews by other users on the internet or application download platforms when deciding which DH tools to adopt.

> *If it's on Play store, they have reviews and stars, so I kind of read those first. I mean some of them have issues like after an update, it just [does not] work anymore, then I just don't download [those apps]. But I generally go for those that are generally highly ranked. If they are highly ranked, but their latest reviews are actually quite bad, then I don't really [download those too]. Then if I do have one in mind*

> *already, I kind of go and Google for that specifically to see what are the reviews.* [Participant 34, preconception]

Theme 5: Perceived Value

The pricing of a DH tool, together with the balance of costs and benefits, played a pivotal role in shaping the perceived value of the tool.

Pricing

Women raised concerns regarding DH tools that required payment as there was a plethora of free applications and information on the internet. Specifically, DH tools that required an upfront fee were a strong deterrence to adoption as women preferred to experience the tool to understand its utility before deciding on the financial commitment.

> *[I think] what would stop [me from using an app would be] the pricing. So, I'm not too sure how much I would be willing to pay for such an app, because like [I said], everything can be found on the Internet.* [Participant 5, preconception]

> *I think if it's paid, like straight-up a paid app, and I am not sure what the features are, that might prohibit me from paying for it or downloading it.* [Participant 4, during pregnancy]

Instead, women were more open to using DH tools that were free but had paid features embedded within the tool, as they were able to make more informed purchase decisions based on its value.

> *But if it's like Flo, [which is] free to download and then there is paid features that I feel like might be helpful, then I might consider paying for it. So, price might be a challenge.* [Participant 4, during pregnancy]

Cost-Benefit Ratio

Women emphasized that their willingness to bear the expenses of DH tools is dependent on whether the cost is supported by its perceived benefits. One feature that women were willing to pay a premium for included teleconsultation features with health care professionals and receiving medical advice.

> *If it's just to calculate my ovulation, I don't think [I will pay for an app]. But if it's an app that has [in-built] chatbots, where I am able to consult doctors or fertility experts on any subject, then yes, I will [pay for the app].* [Participant 39, preconception]

Other factors, such as perceived usefulness and perceived credibility would increase the perceived value of DH tools. Benefits such as continuous support from preconception to postbirth, including long-term support for their child's health and development, and personalization were brought up to outweigh price considerations. The ability of the DH tool to provide an all-in-one comprehensive platform—integrating health records and incorporating functions of other health apps to reduce app fatigue—was pointed out to further enhance the tool's value.

> *[I would pay for the app] if it's going to override [other apps] like Health Hub. Otherwise, it would kind of defeat the purpose if I'm on this app to track my baby's progression, [but] at the same time I have to track my baby's health records on Health Hub.* [Participant 41, preconception]

In addition, the cost of DH tools needed to be justified by its credibility, through evidence-based information and support.

> *It depends on the [expanse] of the app. So if it follows me until the baby is two years old, then the amount that I would pay would be [different from] if it allows me to follow through to 8 years instead.* [Participant, 41, preconception]

> *To garner a premium subscription rate, the app must really be personalized, or must really have sufficient evidence, [so] that it's worth the subscription fee.* [Participant 5, preconception]

Discussion

Principal Findings

Our study identified the major themes that influenced women's adoption of DH tools from before, during, to after pregnancy. Depending on the stages they were in, women applied different considerations of those themes. For instance, the level of time and effort required to engage in DH tools was more salient to women as they were unlikely to devote time out of their busy schedules preconception, or out of their priorities to caring for their newborn postbirth. By accounting for the varying challenges faced by women in different stages, DH tool developers can improve user engagement with the tool and sustain use over time.

Unified Theory of Acceptance and Use of Technology (UTAUT) 2 Model

The 5 main themes and 10 subthemes discussed were mapped based on existing literature on IT use, specifically the Unified Theory of Acceptance and Use of Technology 2 (UTAUT 2) model. UTAUT 2 consists of 7 factors identified to contribute to the acceptance of technological tools: performance expectancy, effort expectancy, social influence, facilitating conditions, hedonic motivation, price value, and habit [27]. UTAUT 2 has its roots in the UTAUT model, which was formed by empirically comparing 8 other models and deriving the primary factors affecting technology use [27]. UTAUT 2 expands UTAUT with factors focused on behavioral intention to use the technology [27].

The themes of perceived usefulness and perceived ease of use identified in our study can be mapped to performance expectancy and effort expectancy respectively in UTAUT 2. The quality of information women received and the extent to which they believed the information fulfilled their needs determined their perceived usefulness of the DH tool and was therefore likely to influence their intention to adopt it. As women progress through different stages, their informational needs would change accordingly (eg, different information relevant to specific trimesters across pregnancy) [23,24]. Having a DH tool that caters to information and features specific to a particular stage in time would naturally improve engagement [26]. The predictability of informational needs from before to after pregnancy can offer a guide to DH developers in providing relevant content to women, thus improving their perceived usefulness of the DH tool. Additionally, enhanced antenatal education for women on improving long-term health outcomes would serve as motivation for higher quality information [16], thus highlighting the importance of perceived usefulness in the use of DH tools. Likewise, for effort expectancy, their experiences and the resources required of them in terms of time and effort would determine their perceived ease of use of the tool.

Under the theme of perceived credibility, there were two subthemes that emerged as follows: (1) the endorsement by trusted parties and (2) the accuracy and reliability of information. The subtheme of endorsement by trusted parties can be likened to the factor of social influence under the UTAUT 2 model. In addition to social influence being a key factor for the adoption of DH tools (as in UTAUT 2), the trust it invokes and the way it is marketed also influenced women's intentions to adopt it. Interestingly, while the UTAUT 2 model predicts the significance of the social influence factor when tool use is mandated [27], in our study, endorsement by trusted parties was highlighted where the use of DH tools by women was not mandatory. The subtheme on the perceived quality of information is not directly linked to any of the factors in UTAUT 2. It is plausible that women in our study who looked for information online were exposed to information that was not verified or reliably accurate [28]. This contrasted with other available trusted sources and may have led to the emergence of this specific subtheme. Overall, while perceived credibility is not a factor in UTAUT 2 in its entirety, current literature shows the rising importance of perceived credibility as a predictor of intention, over factors such as DH literacy in DH tool adoption [22,28]. The theme of perceived credibility and its relationship with social influence prompts further research into the significance of social influence and its nuances for technology adoption specific to a particular deployment context. In one example, public health care is a widely provided good in Singapore, and Singapore residents express trust in public health care and government services. As such, based on our study, promotion, and endorsement by government or public health care professionals may support the perceived credibility of the DH tool and in turn, promote adoption of the tool [29].

Another theme that emerged was perceived value, similar to the factor of price value on the UTAUT 2. Perceived value focuses on the subjective value that women assign to the DH tool, rather than the absolute price value of it as defined in UTAUT 2. Women revealed that they often performed a subjective cost-benefit analysis before deriving the absolute price value they were willing to pay for the DH tool. In addition to considering the pricing of the tool, women would weigh costs against the subjective benefits they derive from the tool. When pricing the tool, it is important to understand what women take into consideration to estimate their willingness to pay in any given tool deployment context. In Singapore, as public health care programs are subsidized or provided for free, women also tend to expect the prices of tools to be subsidized, or free-of-charge [30,31].

Another theme that emerged in our study, women's preexisting attitudes toward DH, can be reflected by the UTAUT 2 factor of facilitating conditions, stipulating that women's attitudes and their prior experiences would influence their perception of the new tool. Women in the preconception, pregnancy, and postbirth groups collectively used over 34 different DH apps for a variety of functions such as physical health monitoring; government and public health–related; mental health; and fertility-, pregnancy-, and child-related health (Multimedia Appendix 1). This indicates that women are experienced with DH use, which is aligned with the growing prevalence of DH use by women in Singapore [32]. Women who have had negative experiences with app use may hesitate when faced with another similar app again [33]. Analogically, positive experiences may reinforce positive attitudes toward a new DH tool.

Out of the 5 themes from our study, 2 themes can be fully (ie, perceived ease of use and perceived usefulness) and 3 themes can be partially (ie, perceived credibility, perceived value, and preexisting attitudes to DH) mapped to the factors listed in UTAUT 2 [27]. In addition, the model has to be contextualized to the deployment environment, and the differences between the themes that emerged in our study and the factors of UTAUT 2 warrant further research. Additionally, there were 2 other factors in the UTAUT 2 model that did not emerge as themes in our study: hedonic motivation and habit. It is possible that while habitual use may determine DH use, women could end up switching between DH tools catering to different needs too often to form a habit. Women may also perceive the use of DH tools to be limited to their reproductive journey and hence may not form a habit as opposed to DH tools for long-term health [34]. It is also possible that while those factors were at play, the participants' lack of awareness or expressiveness did not allow us to capture these factors with the deployed methodology. This area warrants further research. Nevertheless, UTAUT 2 shows high potential as a model used to understand women's DH engagement during their reproductive journey.

Adoption Intention and Sustained Use

The behavioral intention to use a DH tool has been previously shown to vary across the user journey, from the initial adoption to continued use [35]. It was the case for women in our study as well. The 5 identified themes exert different influences on their decision-making at different timepoints in their journey. Themes such as preexisting attitudes toward DH, perceived credibility, and perceived value are more pertinent in influencing adoption intention. For instance, in the early adoption stages, women's motivation to adopt a DH tool was more significantly influenced by their preexisting attitudes toward DH, shaped by previous unpleasant experiences or lack of understanding of DH tools. Women also sought recommendations from their surroundings to form the perceived credibility of the DH tool in determining if they would want to adopt it. Similarly, the UTAUT 2 model on technology adoption identified social influence to be pertinent to women, especially in the early stages of adoption [27]. This was confirmed by our study where women mentioned that recommendations about DH tools from their trusted social networks would have enhanced their early adoption of the DH tool. Additionally, the theme of perceived value has been shown to factor into user intention [36], where women preferred free-to-use apps with only a small minority willing to pay for the tool.

Themes such as perceived usefulness and perceived ease of use are relevant to both early adoption and sustained use of the DH tools. In terms of early adoption, the perceived usefulness and perceived ease of use have been shown to influence the intention to use [35]. The ability of the tool to meet women's needs in specific phases and transition across phases (eg, from pregnancy to postbirth), and a smooth experience using the tool emerged as desired factors that motivated sustained use in the long run. Additionally, women in our study highlighted the expected effort required for the DH tool as a determinant of its long-term use. This could be due to the changing needs and bandwidths of women over different phases of motherhood, which affect their capacity to manage the expected effort in sustained use of a DH tool, which calls for individualized requirements and informational needs [23,24]. As women navigate each phase, this would necessitate adapting to new features of the tools and resources required of them. The relatively rapid and predictable timing of the transition from one phase to the next is inherently useful and can be harnessed to sustain engagement from one phase to the next. Interestingly, prior studies suggest that effort expectancy is not a predictor of continuous use [35], and such discrepancy would require further research.

There are many studies examining the difference between the early adoption and sustained use of information systems [35,37]. There is a gap, however, in the understanding of these 2 elements for DH tools adoption by women from before to after pregnancy. In this study, we propose how the 5 discovered themes interact with these 2 elements and note the role of the deployment context in the realization of these elements.

Future Directions

This study works as a foundational inquiry into understanding women's needs and expectations, in line with the principles of co-design [38]. The potential next step could include the generation of the content-strategy-design guide—a deep dive into the specific components of the tools, such as content, appearance, workflow, and integration with the system and the existing user journey [39]. By eliciting themes that are the most salient to women in their reproductive journey, these insights can serve to inform DH tool developers in designing ways to improve engagement with a DH tool. Such a strategy has been used before, where qualitative methodology was employed to co-design a DH tool for women in pregnancy [30]. Table 2 summarizes the insights and recommendations that emerged from our study. Along with technology expectations in Table 1, this could provide specific guidance in designing frequency of use, sources of information, and the preferred information topics at each phase from preconception to postbirth.

Table 2. List of recommendations based on women's identified challenges and motivators in the adoption and sustained use of digital health tools.

Theme and subtheme	Recommendation
Perceived ease of use	
User experience and interface (UX/UI)	Consistent updating and maintenance of apps
Level of effort and time required	Adjusting the use of the app to user's availability
Perceived usefulness	
Need fulfillment and continuous support	All in one tool allowing women to use it consistently from preconception to postbirth, covering multiple functions
Perceived credibility	
Endorsement by trusted parties	Encouraging adoption by getting trusted endorsements from fellow mothers and health care providers (gynae, etc)
Perceived value	
Pricing	Accounting for localized context where users expect free-to-use apps

Limitations

As 79.5% (35/44) of our participants were of middle to high SES, the results of our study may have limited generalizability, especially stemming from their prior familiarity with technology. Additionally, the recruitment criteria included English fluency. Further research can investigate the perceptions of lower SES groups and non-English speakers to examine the validity of the identified factors affecting sustained adoption. Last, women who had used or were using assisted conception were excluded, similarly to women who had current diagnoses of psychiatric disorders. These women might face vastly different struggles and considerations to the population we examined which could potentially be supported by various DH tools as well. This calls for future research to examine the applicability of our findings to these populations of women from before to after pregnancy.

Conclusion

The study examined the factors affecting sustained adoption of DH tools for women from before to after pregnancy and mapped them according to the UTAUT 2 model. Our study shows that the UTAUT 2 model can broadly reflect the adoption of DH tools in women from before to after pregnancy. Specific considerations of the cultural implications are needed in its application to the context of Singapore, a highly tech-pervasive Asian society, which is different from the European and African contexts where prior DH adoption studies have been conducted. Additionally, understanding the interactions of the themes with early adoption and sustained use could support DH design in ensuring both uptake and long-term use. There are various considerations to be noted in the design of DH tools for women's preconception to postbirth to promote engagement, and tool design should be a regular iterative process to continuously improve the ability of the tool to meet women's needs.

Acknowledgments

DH gratefully acknowledges funding from the Institute for Digital Medicine (WisDM) Translational Research Programme (grant number R-719-000-037-733) at the Yong Loo Lin School of Medicine. The funder played no role in the study design, data collection, analysis, and interpretation of data, or the writing of this manuscript.

Data Availability

The data sets used or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

VVL, SV, DSQO, LLS, YSL, SYC, AB, and DH conceived the study. VVL, SV, and DSQO were involved in protocol development and gaining ethical approval. VVL, SV, WYN, NYL, and QYL were involved in participant recruitment, data collection, and data analysis. WYN and NYL wrote the first draft of the manuscript. All authors reviewed and edited the manuscript and approved the final version of the manuscript.

Conflicts of Interest

AB and DH are coinventors or previously filed pending patents on artificial intelligence–based therapy development. DH is a shareholder of KYAN Therapeutics, which has licensed intellectual property pertaining to AI-based drug development. SC reports grants from Société Des Produits Nestlé S.A. outside the submitted work and is a coinventor on patent filings by Nestlé S.A. relating to maternal supplements. SC has received reimbursement from the Expert Group on Inositol in Basic and Clinical Research (EGOI; a not-for-profit academic organization) and Nestlé Nutrition Institute for speaking at conferences. All other authors declare no financial or nonfinancial competing interests.

Multimedia Appendix 1
Annexes of pre- and postinterview questionnaires, interview framework, demographic data, and mobile application usage data of participants.
[DOCX File , 25 KB - humanfactors_v11i1e59269_app1.docx]

References

1. Poston L, Harthoorn LF, van der Beek EM, Contributors to the ILSI Europe Workshop. Obesity in pregnancy: implications for the mother and lifelong health of the child. a consensus statement. Pediatr Res 2011;69(2):175-180. [doi: 10.1203/PDR.0b013e3182055ede] [Medline: 21076366]
2. Yap F, Loy SL, Ku CW, Chua MC, Godfrey KM, Chan JKY. A golden thread approach to transforming maternal and child health in Singapore. BMC Pregnancy Childbirth 2022;22(1):561 [FREE Full text] [doi: 10.1186/s12884-022-04893-8] [Medline: 35836151]
3. Loo EXL, Soh S, Loy SL, Ng S, Tint MT, Chan S, S-PRESTO Study Group, et al. Cohort profile: Singapore preconception study of long-term maternal and child outcomes (S-PRESTO). Eur J Epidemiol 2021;36(1):129-142 [FREE Full text] [doi: 10.1007/s10654-020-00697-2] [Medline: 33222050]
4. Soh SE, Tint MT, Gluckman PD, Godfrey KM, Rifkin-Graboi A, Chan YH, et al. Cohort profile: growing up in Singapore towards healthy outcomes (GUSTO) birth cohort study. Int J Epidemiol 2014;43(5):1401-1409 [FREE Full text] [doi: 10.1093/ije/dyt125] [Medline: 23912809]
5. Shahil Feroz A, Afzal N, Seto E. Exploring digital health interventions for pregnant women at high risk for pre-eclampsia and eclampsia in low-income and-middle-income countries: a scoping review. BMJ Open 2022;12(2):e056130 [FREE Full text] [doi: 10.1136/bmjopen-2021-056130] [Medline: 35135777]
6. Moise IK, Ivanova N, Wilson C, Wilson S, Halwindi H, Spika VM. Lessons from digital technology-enabled health interventions implemented during the coronavirus pandemic to improve maternal and birth outcomes: a global scoping review. BMC Pregnancy Childbirth 2023;23(1):195 [FREE Full text] [doi: 10.1186/s12884-023-05454-3] [Medline: 36941565]
7. Goodday SM, Karlin E, Brooks A, Chapman C, Karlin DR, Foschini L, et al. Better understanding of the metamorphosis of pregnancy (BUMP): protocol for a digital feasibility study in women from preconception to postpartum. NPJ Digit Med 2022;5(1):40 [FREE Full text] [doi: 10.1038/s41746-022-00579-9] [Medline: 35354895]

8. Frid G, Bogaert K, Chen KT. Mobile health apps for pregnant women: systematic search, evaluation, and analysis of features. J Med Internet Res 2021;23(10):e25667 [FREE Full text] [doi: 10.2196/25667] [Medline: 34524100]
9. Henriksson P, Migueles JH, Söderström E, Sandborg J, Maddison R, Löf M. User engagement in relation to effectiveness of a digital lifestyle intervention (the HealthyMoms app) in pregnancy. Sci Rep 2022;12(1):13793 [FREE Full text] [doi: 10.1038/s41598-022-17554-9] [Medline: 35963935]
10. Grenfell P, Tilouche N, Shawe J, French RS. Fertility and digital technology: narratives of using smartphone app 'Natural Cycles' while trying to conceive. Sociol Health Illn 2021;43(1):116-132 [FREE Full text] [doi: 10.1111/1467-9566.13199] [Medline: 33147647]
11. Halila A, Ismail A, Sutan R, Rahman R, Juval K. A scoping review of digital technologies in antenatal care: recent progress and applications of digital technologies. Research Square. Preprint posted online on November 2, 2023 . [doi: 10.21203/rs.3.rs-3492596/v1]
12. Brown SD, Garcia BL, Ritchie JL, Tsai A, Millman A, Greenberg M, et al. Digital health outreach to promote postpartum screening after gestational diabetes: a randomized factorial pilot study. PEC Innov 2024;4:100256 [FREE Full text] [doi: 10.1016/j.pecinn.2024.100256] [Medline: 38282680]
13. Iwaya LH, Nordin A, Fritsch L, Børøsund E, Johansson M, Varsi C, et al. Early labour app: developing a practice-based mobile health application for digital early labour support. Int J Med Inform 2023;177:105139 [FREE Full text] [doi: 10.1016/j.ijmedinf.2023.105139] [Medline: 37406571]
14. Galle A, Semaan A, Huysmans E, Audet C, Asefa A, Delvaux T, et al. A double-edged sword-telemedicine for maternal care during COVID-19: findings from a global mixed-methods study of healthcare providers. BMJ Glob Health 2021;6(2):e004575 [FREE Full text] [doi: 10.1136/bmjgh-2020-004575] [Medline: 33632772]
15. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
16. Namatovu HK, Oyana TJ, Sol HG. Barriers to eHealth adoption in routine antenatal care practices: perspectives of expectant mothers in Uganda - a qualitative study using the unified theory of acceptance and use of technology model. Digit Health 2021;7:20552076211064406 [FREE Full text] [doi: 10.1177/20552076211064406] [Medline: 34900326]
17. Topçu S, Brown P. The impact of technology on pregnancy and childbirth: creating and managing obstetrical risk in different cultural and socio-economic contexts. Health Risk Soc 2019;21(3-4):89-99. [doi: 10.1080/13698575.2019.1649922]
18. Brusniak K, Arndt HM, Feisst M, Haßdenteufel K, Matthies LM, Deutsch TM, et al. Challenges in acceptance and compliance in digital health assessments during pregnancy: prospective cohort study. JMIR Mhealth Uhealth 2020;8(10):e17377 [FREE Full text] [doi: 10.2196/17377] [Medline: 33052134]
19. Rhodes A, Kheireddine S, Smith AD. Experiences, attitudes, and needs of users of a pregnancy and parenting app (Baby Buddy) during the COVID-19 pandemic: mixed methods study. JMIR Mhealth Uhealth 2020;8(12):e23157 [FREE Full text] [doi: 10.2196/23157] [Medline: 33264100]
20. Tolonen H, Dobson A, Kulathinal S, WHO MONICA Project. Effect on trend estimates of the difference between survey respondents and non-respondents: results from 27 populations in the WHO MONICA Project. Eur J Epidemiol 2005;20(11):887-898. [doi: 10.1007/s10654-005-2672-5] [Medline: 16284866]
21. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
22. Goetz M, Müller M, Matthies LM, Hansen J, Doster A, Szabo A, et al. Perceptions of patient engagement applications during pregnancy: a qualitative assessment of the patient's perspective. JMIR Mhealth Uhealth 2017;5(5):e73 [FREE Full text] [doi: 10.2196/mhealth.7040] [Medline: 28550005]
23. Madden SK, Skouteris H, Bailey C, Hills AP, Ahuja KDK, Hill B. Women in the workplace: promoting healthy lifestyles and mitigating weight gain during the preconception, pregnancy, and postpartum periods. Int J Environ Res Public Health 2020;17(3):821 [FREE Full text] [doi: 10.3390/ijerph17030821] [Medline: 32013002]
24. Madden SK, Blewitt CA, Ahuja KDK, Skouteris H, Bailey CM, Hills AP, et al. Workplace healthy lifestyle determinants and wellbeing needs across the preconception and pregnancy periods: a qualitative study informed by the COM-B model. Int J Environ Res Public Health 2021;18(8):4154 [FREE Full text] [doi: 10.3390/ijerph18084154] [Medline: 33919967]
25. Abekah-Nkrumah G, Guerriero M, Purohit P. ICTs and maternal healthcare utilization. evidence from Ghana. Int J Soc Econ 2014;41(7):518-541. [doi: 10.1108/ijse-11-2012-0218]
26. Lee VV, Vijayakumar S, Ng WY, Lau NY, Leong QY, Ooi DSQ, et al. Personalization and localization as key expectations of digital health intervention in women pre- to post-pregnancy. NPJ Digit Med 2023;6(1):183 [FREE Full text] [doi: 10.1038/s41746-023-00924-6] [Medline: 37775533]
27. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
28. Adjekum A, Blasimme A, Vayena E. Elements of trust in digital health systems: scoping review. J Med Internet Res 2018;20(12):e11254 [FREE Full text] [doi: 10.2196/11254] [Medline: 30545807]

29. Lysaght T, Ballantyne A, Xafis V, Ong S, Schaefer GO, Ling JMT, et al. "Who is watching the watchdog?": ethical perspectives of sharing health-related data for precision medicine in Singapore. BMC Med Ethics 2020;21(1):118 [FREE Full text] [doi: 10.1186/s12910-020-00561-8] [Medline: 33213433]
30. Shorey S, Tan TC, Thilagamangai, Mathews J, Yu CY, Lim SH, et al. Development of a supportive parenting app to improve parent and infant outcomes in the perinatal period: development study. J Med Internet Res 2021;23(12):e27033 [FREE Full text] [doi: 10.2196/27033] [Medline: 36260376]
31. Shorey S, Law E, Thilagamangai, Mathews J, Lim SH, Shi L, et al. Evaluating the effectiveness of the supportive parenting app on parental outcomes: randomized controlled trial. J Med Internet Res 2023;25:e41859 [FREE Full text] [doi: 10.2196/41859] [Medline: 36645699]
32. Hossain I, Ang YN, Chng HT, Wong PS. Patients' attitudes towards mobile health in Singapore: a cross-sectional study. Mhealth 2019;5:34 [FREE Full text] [doi: 10.21037/mhealth.2019.08.07] [Medline: 31620461]
33. Sampat B, Sharma A, Prabhakar B. Understanding factors influencing the usage intention of mobile pregnancy applications. In: Sharma SK, Dwivedi YK, Metri B, Rana NP, editors. Re-imagining Diffusion and Adoption of Information Technology and Systems: A Continuing Conversation. Cham: Springer International Publishing; 2020:641-654.
34. Wutz M, Hermes M, Winter V, Köberlein-Neu J. Factors influencing the acceptability, acceptance, and adoption of conversational agents in health care: integrative review. J Med Internet Res 2023;25:e46548 [FREE Full text] [doi: 10.2196/46548] [Medline: 37751279]
35. Karahanna E, Straub DW, Chervany NL. Information technology adoption across time: a cross-sectional comparison of pre-adoption and post-adoption beliefs. MIS Quarterly 1999;23(2):183-213. [doi: 10.2307/249751]
36. Hossain I, Lim ZZ, Ng JJL, Koh WJ, Wong PS. Public attitudes towards mobile health in Singapore: a cross-sectional study. Mhealth 2018;4:41 [FREE Full text] [doi: 10.21037/mhealth.2018.09.02] [Medline: 30363776]
37. Lim S, Xue L, Yen CC, Chang L, Chan HC, Tai BC, et al. A study on Singaporean women's acceptance of using mobile phones to seek health information. Int J Med Inform 2011;80(12):e189-e202. [doi: 10.1016/j.ijmedinf.2011.08.007] [Medline: 21956003]
38. Tremblay M, Hamel C, Viau-Guay A, Giroux D. User experience of the co-design research approach in eHealth: activity analysis with the course-of-action framework. JMIR Hum Factors 2022;9(3):e35577 [FREE Full text] [doi: 10.2196/35577] [Medline: 35943783]
39. Blewitt C, Savaglio M, Madden SK, Meechan D, O'Connor A, Skouteris H, et al. Using intervention mapping to develop a workplace digital health intervention for preconception, pregnant, and postpartum women: the health in planning, pregnancy and postpartum (HiPPP) portal. Int J Environ Res Public Health 2022;19(22):15078 [FREE Full text] [doi: 10.3390/ijerph192215078] [Medline: 36429795]

Abbreviations

DH: digital health
SES: socioeconomic status
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by A Kushniruk; submitted 08.04.24; peer-reviewed by S Goodday; comments to author 06.05.24; revised version received 08.07.24; accepted 16.07.24; published 01.10.24.

Please cite as:
Ng WY, Lau NY, Lee VV, Vijayakumar S, Leong QY, Ooi SQD, Su LL, Lee YS, Chan SY, Blasiak A, Ho D
Shaping Adoption and Sustained Use Across the Maternal Journey: Qualitative Study on Perceived Usability and Credibility in Digital Health Tools
JMIR Hum Factors 2024;11:e59269
URL: https://humanfactors.jmir.org/2024/1/e59269
doi:10.2196/59269
PMID:39352732

Original Paper

Co-Designing a Conversational Agent With Older Adults With Chronic Obstructive Pulmonary Disease Who Age in Place: Qualitative Study

Emilie Kauffeldt Wegener[1], MA, MSc; Jenny M Bergschöld[2], PhD; Tina Kramer[1], PhD; Camilla Wong Schmidt[3], MSc; Karen Borgnakke[1], PhD

[1]Department of Public Health, University of Copenhagen, København, Denmark
[2]Department of Health, SINTEF Digital, Oslo, Norway
[3]Region Zealand, Sorø, Denmark

Corresponding Author:
Emilie Kauffeldt Wegener, MA, MSc
Department of Public Health
University of Copenhagen
Øster Farimagsgade 5
København, 1353
Denmark
Phone: 45 26202168
Email: emiliewegener@gmail.com

Abstract

Background: As a reaction to the global demographic increase in older adults (aged 60+ years), policy makers call for initiatives to enable healthy aging. This includes a focus on person-centered care and access to long-term care for older adults, such as developing different services and digital health technologies. This can enable patients to engage in their health and reduce the burden on the health care systems and health care professionals. The European Union project Smart Inclusive Living Environments (SMILE) focuses on well-being and aging in place using new digital health technologies. The novelty of the SMILE project is the use of a cocreational approach focused on the needs and preferences of older adults with chronic obstructive pulmonary disease (COPD) in technology development, to enhance access, adaptation, and usability and to reduce stigma.

Objective: The study aimed to describe the perspective, needs, and preferences of older adults living with COPD in the context of the design and development of a conversational agent.

Methods: This study carried out a data-driven thematic analysis of interview data from 11 cocreation workshops with 33 older adults living with COPD.

Results: The three particular features that the workshop participants wanted to implement in a new technology were (1) a "my health" function, to use technology to manage and learn more about their condition; (2) a "daily activities" function, including an overview and information about social and physical activities in their local area; and (3) a "sleep" function, to manage circadian rhythm and enhance sleep quality, for example, through online video guides. In total, 2 overarching themes were identified for the 3 functions: measurements, which were actively discussed and received mixed interest among the participants, and health literacy, due to an overall interest in learning more about their condition in relation to everyday life.

Conclusions: The future design of digital health technology must embrace the complexities of the everyday life of an older adult living with COPD and cater to their needs and preferences. Measurements should be optional and personalized, and digital solutions should be used as a supplement to health care professionals, not as substitute.

(JMIR Hum Factors 2024;11:e63222) doi:10.2196/63222

KEYWORDS

eHealth; aging in place; digital health technology; health literacy; everyday life; co-design; co-designing; conversational agent; older adults; elderly; COPD; thematic analysis; design; development; interview data; cocreation; chronic obstructive pulmonary disease; mobile phone

Introduction

Background

Global policy makers have demanded maintained action and initiatives to enable healthy aging in all parts of the world. The United Nations and the World Health Organization (WHO) describe healthy aging as "developing and maintaining the functional ability that enables well-being in older age" [1]. The proportion of older adults will double from 1 billion to 2.1 billion between 2019 and 2050 globally [1]. As a reaction to this demographic change, areas of action suggested in the *UN Decade of Healthy Ageing: Plan of Action 2021-2030* include person-centered care and access to long-term care for older adults, such as developing different services including assistive technologies [1].

Cocreating Technology With Older Adults

The European Union Horizon 2020 project Smart Inclusive Living Environments (SMILE), a collaborative project among Denmark, Norway, Canada, the Netherlands, and Greece, draws on practical situations that older adults face in daily life to create smart living environments (SLEs) with novel eHealth solutions. The project intends to assist older adults with chronic obstructive pulmonary disease (COPD) to maintain their independence in the comfort of their own homes. The project emphasizes the involvement of older adults by identifying their needs and preferences and by involving them in co-designing and testing a new technological solution [2]. The technological solution that is cocreated with end users in the SMILE project is a conversational agent (CA), designed to facilitate daily communication online through a weblink that can be accessed on a smartphone, tablet, or computer.

In SMILE, qualitative cocreation workshops are used to enhance the practical involvement of older adults in the development of technology. This highlights one of the novelties in the SMILE project, as earlier cocreation or participatory design processes have not succeeded in facilitating active participation in all phases of a design process, from the initial needs identification to the final evaluation phase [2]. The aim of the cocreation workshops is to provide a secure environment where participants can share their daily experiences and thoughts about their health conditions, preferences, needs, and daily activities. These workshops also serve as exploratory sessions for both the participants and the researchers in facilitating the generation of ideas for future technology design and development [3].

The Living Labs

The SMILE project uses the concept of "living labs" to study the use and impact of new digital technology on older adults' daily lives. A "living lab" is a user-centered, open ecosystem that enables the integration of research and innovation processes into real-life communities and settings [4]. Using living labs has the advantage of being independent of existing economic models and policies, making it possible to challenge existing societal structures by providing suitable arenas for testing new ideas [5].

The Danish SMILE living lab "PreCare," was launched in 2018 and operates in a collaboration between the Odsherred municipality and Region Zealand. PreCare offers digital and outward-bound homecare treatments to older adults living with COPD and has approximately 150 older adults enrolled. PreCare works in dynamics between innovation, development, implementation, operation, and analysis based on an action research–based "plan-do-study-act cycle" [6] and is therefore in constant development. The Norwegian SMILE living lab is located in Gudbrandsdalen, a rural valley consisting of 12 municipalities. In Norway, the number of people over the age of 80 years will almost double in rural areas by 2040 [7,8], posing significant challenges in the context of demographic aging. In rural Norway, the long distances between people, workplaces, public services, care services, and institutions are challenging for older adults who age in place, especially those who have chronic conditions [9]. In SMILE, these challenges are used to collaborate closely with older adults with COPD, engaging them as co-designers and using qualitative methods to explore their lived experiences and needs [10].

Interprofessional Collaboration

With the SMILE project's intent, interprofessional collaboration is promoted, and existing structures and roles among patients, health care professionals, and informal caregivers are challenged [11]. Older adults, their social network, and health care professionals must be capable and willing to use these new technologies for the "smart inclusive living environment" to fully work.

Traditionally, health care professionals provide care to patients, with minimal involvement of the patients. This has gradually changed to an increased collaboration among health care professionals, patients, and informal caregivers. In the WHO's global strategy on integrated and people-centered health services (2016-2026), people, patients, and communities are placed at the center of health services. The strategy emphasizes that the care should be organized around the health needs and expectations of people and communities rather than a specific disease. People have individual preferences, needs, and abilities, and therefore should not be understood as patients defined by their disease [12]. Therefore, innovative initiatives toward more patient-oriented activities in health care are needed, such as the co-design of new digital solutions as carried out in the SMILE project, as well as the development of new digital solutions to support both patients, citizens, and formal and informal caregivers.

Methods

Overview

The empirical data presented and analyzed in this study were collected through 11 cocreational workshop sessions in Denmark and Norway with people living with COPD. The analysis was thematic and privileged the participants' descriptions and perspectives that were related to their reasons and relevance criteria for the design and use of the CA.

Aim

The aim is to describe the perspective, needs, and preferences of older adults living with COPD in the context of the design and development of a CA.

Participants, Recruitment, and Eligibility

A total of 33 older adults living with COPD participated in the SMILE cocreation workshops. In Denmark, 21 participated, and in Norway, 12 participated. Participants were recruited through health care partners as a part of their enrolment in the SMILE living labs. Access to the participants was established through the registered nurses at each living lab, and participants were then contacted through email or phone by the researchers. Participants were eligible to participate if they were aged 65 years or older, lived with COPD, and were able to fill in the Readiness and Enablement Index for Health Technology (READHY) instrument and take an active part in the cocreational design workshops and evaluation. Oral and written consent was obtained from all participants before the workshops (Multimedia appendix 1).

The saturation or information power of the sample size in this study was addressed based on the 5 items impacting information power in qualitative research presented by Malterud et al [13]. These included the following: (1) the study aim is specific and therefore can hold fewer participants, with its focus on identifying needs and preferences and with the purpose of cocreating a technology, through an iterative process; (2) the sample specificity is relevant to the study aim and is limited to older adults with COPD, but has variations regarding technology readiness; (3) theoretical frameworks included READHY and concepts presented by Yock et al [14] to cocreate technology with end users; (4) the dialogue with the participants was clear and followed a semistructured interview guide, which often requires fewer participants; and (5) the strategy for the analysis is an in-depth analysis of themes and narratives, for example, not a cross-case analysis where more participants often would be needed [13]. Based on these, saturation for this study was achieved.

Ethical Considerations

Informed consent was obtained from all humans participating in this study. Data obtained from participants was handled according to General Data Protection Regulation (GDPR). Our research has followed the declaration of Helsinki's ethical principles for medical research involving human participants. In Denmark, health science questionnaire surveys and interview studies that do not involve human biological material (section 14(2) of the Danish Act on Committees) do not require reporting or approval from the Danish National Centre for Ethics [15]. In Norway, the field studies were approved by the data protection officer at Innlandet Hospital Trust (journal 14832226).

Data Collection

Cocreation Workshops

The empirical data were gathered through 11 cocreation workshops: 5 in Denmark and 6 in Norway. Each workshop had a duration of 2 hours, and between 2 and 5 participants were present at each workshop. The CA was developed through 3 design iterations and the workshops were conducted before each iteration, meaning that there were 3 phases of the workshops. Each workshop phase had different purposes, which corresponded to the development of the CA.

The purpose of the first phase of the workshops was to elaborate on early ideas, to present and discuss 9 possible functions, and to narrow these down to 3 or 4 prioritized functions based on input from the participants. In total, 4 prioritized functions were identified in the first phase of the workshops, which were (1) physical activity, (2) clinical condition, (3) social activity, and (4) sleep. These 4 functions were then implemented in a CA mock-up. After each workshop, a summary report was written and provided to the technology developers, to inform the development of the CA based on input from the participants.

The purpose of the second phase of the workshops was to present the CA mock-up to the participants and gather input that could be used to develop the CA prototype version 1. The mock-up presented to the participants was built in Botpress [16] and consisted of a basic chatbot with predetermined dialogues that concentrated on the 4 functions identified as relevant by the participants during the first phase of the workshops. The functions and mock-up were discussed with the workshop participants. This phase resulted in changing the features, that is, social activity and physical activity were merged into one feature called "daily activity." There were changes to the interface and the conversation strings, and an onboarding feature that teaches users how to interact with the CA was included.

The third phase of the workshops was used to elaborate on the participants' perceptions of the content and functions of the CA prototype version 1. The prototype contained an onboarding feature and three main functions: (1) my health, (2) daily activity, and (3) sleep, which were developed based on input from the participants from phase 2. Findings from the third phase were used to develop the CA prototype version 2.

All workshops were audiotaped. Before data from the workshops were analyzed, sound recordings from the cocreation workshops were transcribed verbatim in the native languages at each living lab site [17]. In analyzing the transcriptions, relevant parts presented in this study were translated into English. EKW carried out the transcription and translation of the Danish material, and JMB carried out the transcription and translation of the Norwegian material. All translations privilege clarity of meaning over the verbatim transcripts.

Thematic Analysis

This study conducted a thematic analysis of the data from 11 cocreation workshops. The thematic analysis was inspired by Braun and Clarke [18]; concepts of data-driven analysis; and a realist approach to give voice to the data and the realities, meanings, and experiences presented in the data. As a part of the workshops, the themes "my health," "daily activity," and "sleep" were established, which corresponded to the most desired functions among the participants to implement in the CA. The analysis focused on identifying and reporting themes and patterns within these 3 functions. This was not based on coding or specific prevalence of the themes identified but through addressing overarching themes of relevance for the 3 most desired functions addressed by the participants [18]. Finally, the thematic analytical approach was inspired by Borgnakke [19], to strengthen the innovative and thematical analytical approach, inspired by practice-oriented research methods.

All researchers took an active part in processing the data, enhancing the process of addressing possible blind spots, and understanding several ways to see the data [20]. In this process, the researchers also acknowledge their active role in the identification, selection, and presentation of themes and patterns from the data in this study [18].

The thematic analytic approach corresponds to the interest in strengthening the relationship between technology development and the participant's needs and preferences, as well as in cocreating digital health technology that can support older adults' healthy aging and everyday life. By starting technology development from the participant's preferences and needs, focusing on meaning-making rather than on adopting a consensus-driven approach, the thematic analysis can help sharpen the sense of themes, problems, and dilemmas seen from the older adult's perspective, reasons, and relevance criteria. With this analysis, the aim was to capture the relevant themes and layers based on the participants' thematizations and problematizations and thus increase insight into everyday use situations and challenges.

The transcribed and translated data have not been presented to the participants, and member checking of the analysis has not been conducted [21]. However, due to the iterative approach in the cocreation workshop process, the findings were addressed and validated through the 3 ongoing workshop phases.

Results

Overview

In this section, results from the thematic analysis of data from 11 cocreation workshops are presented. The purpose of the analysis was to acquire knowledge and insight into older adults' own experiences and formulations on important topics and dilemmas connected to their everyday lives in the development of the CA. The results are summarized according to the 3 embedded functions in the CA: "my health," which addresses the user's health condition; "daily activity," which informs about both physical and social activities; and "sleep," which addresses circadian rhythm.

My Health

My health was one of the top-rated functions from the beginning of the cocreational process. The participants argued that their health is the most important thing they have. One of the first things addressed by some of the participants was that they want to learn more about their health condition, that is, they want to increase their health literacy. Examples included general information about COPD, for example, "when were you diagnosed with COPD?" "What are the different levels of COPD?" and other commonly asked questions that could be provided to them through the CA. Doing inhalations correctly was also highlighted as an important aspect of their daily life with COPD. Therefore, participants suggested that the CA could contain guides, written or video, on how to do inhale measurements.

> *An important thing [in relation to my health] is inhalation-technique...it is critical that a patient inhales correctly and is measured on whether you get it down where it needs to be...you can look up some videos on YouTube, they can show it...if there was a video link to a nurse and the citizen had a whistle at home. Then I am sure that it could happen on a video that the person sat and breathed, and the nurse instructed.* [M1]

An important aspect of the CA development and the function "my health" was that the participants raised concerns about the technology replacing their contact with health care professionals. A lot of the participants were positive toward new technologies in general, and over half of them already used different types of technologies in their everyday life including smartwatches, smartphones, and in-home measuring tools for their condition, for example, technologies provided to them by PreCare. But if a technology was implemented to replace their contact with health care professionals, they were not interested in it.

> *Well, I absolutely don't agree that technology is a good thing [if the idea is to replace human visits], because my daughter works in healthcare, and she said they're developing things now so they can call you and have this setup where they can check if you have a fever and see everything. What about the poor old folks at home who are isolated? Maybe the home care personnel are the only person they see?* [M2]

> *I was also offered when I came into PreCare that I could have a conversation on that tablet, right, then I said I wouldn't agree to that. If I'm going to have a conversation, it has to be face-to-face...but I may be old-fashioned.* [F2]

The participants mostly trust information from health care professionals. However, they were open to reading health-related information through the CA but required that this information refers to a trusted site, such as *medicin.dk* or *sundhed.dk* in Denmark and *Folkehelseinstituttet* or *Helsenorge* in Norway.

The participants had different needs and interests about the extent and frequency of monitoring their condition. Some participants were interested in measuring a few health data, for example, blood pressure and oxygen saturation, and some were interested in measuring more. Moreover, some participants were interested in measuring their condition daily, and some preferred measuring whenever they felt like it.

> *I do measurements as much as I can remember to do it...and I feel good about that.* [M2]

> *I do measurements every day.* [F2]

> *I do my saturation measurements at about 10 or 12...then I can see if the measures are not bad...or think why I am feeling the way I am, oh yes, I can see that on the measurements. That is actually a good indicator for me...* [M2]

Arguments for wanting to measure their condition daily included getting a feeling of insight into their condition and a general feeling of being aware of their current health state. Arguments against daily measurements included being reminded about their condition and having to identify with it daily. This shows the conflicting feelings and needs when being a patient who receives treatment by the health care system and a person living everyday life with a chronic condition.

> *I feel a bit ambivalent about it because it annoys me so much that I have to measure my condition every day because then I become aware of it and sometimes, I can't quite get it to hit the numbers I want to. I try again and then it gets even worse and then I get irritated and stop.* [F2]

The different preferences in terms of usage led to a discussion on how to personalize the CA. The participants suggested a "my profile" function that can help personalize the CA and add or remove measurement units as wanted. This function could also hold data such as height, weight, and information on current medicine intake, etc, that can be entered if relevant.

The participants also suggested the option of sharing their data stored in the CA with health care professionals and informal caregivers. This statement was however not shared by all participants; some found it very relevant, to be able to share data with family or their general practitioner, and some found more value in the data being private, as they did not want to worry or burden their loved ones.

In sum, the "my health" function is highly relevant to several workshop participants, as they state that "our health is the most important thing we have." The relevance of the function is highlighted in the participant's wish for more information and knowledge about their condition, guides to do proper in-home measurements daily and comprehensive measurements of different parameters, and the possibility of sharing this data with health care professionals and informal caregivers. However, not all participants shared this view fully, and nuances were identified. Some participants did not wish to measure daily, some wished to measure selected parameters, and others did not want to share their data with others. Therefore, the need for personalization of the app in a "my profile" function was identified.

Daily Activity

The daily activity function was created as a merge of the 2 top-rated functions—physical activity and social activity—based on ongoing input from the workshop participants. In a brainstorming session, participants suggested that the "daily activity" function should be able to suggest social activities in the local area; motivate them to be active, for example, through text or voice reminders; be able to monitor their activity, for example, through step count; and have an incorporated walking community feature, as some participants found it motivating to walk with others.

> *I have a "walking-friend"...and we walk together once a week, that could be a thing? Maybe that could be a part of this [the technology].* [M1]
>
> *It [the technology] could for example suggest you contact your walking friend.* [M2]

Some participants also suggested incorporating knowledge or management of lung capacity in this function, both to learn about how much physical activity they can do in a day and for how long based on their individual condition to feel safe when going outside, as well as inspire easy physical exercises and activities at home.

For the social activities, participants proposed receiving suggestions of social activities in the local area; receiving reminders about activities, for example, pop-up reminders or through a calendar function; or inspiring them to attend social communities for example, "COPD friends," knitting clubs, and lung choirs.

> *It's about knowing what is happening in the local area. I started in a knitting club, and I can't knit, and I am probably 20 years younger than the other people there, but I have fun, then I just draw or do pearls. It's the social aspect, and just getting out of the house.* [F1]

Participants suggested merging physical and social activities, among others, because the participants had different needs and preferences related to physical activity. Some participants had a high activity level and associated physical activity with exercising in a gym. They wanted reminders on planned workout sessions, motivational messages, and inspiration for different exercises; wanted to use sensors, wearables, or smartphones to track their progress; and suggested features such as "my goals" and "my progress."

> *It's a bit funny when you now have the option [to count steps] I catch myself getting annoyed when I've forgotten my phone when I've gone places, then you don't know how much you've walked.* [F1]
>
> *You somehow get happy when it says "you did it"...that means something.* [F2]

The other group of participants had a lower activity level and associated physical activity with taking out the trash or cleaning the house. This group was also open for activity tracking and goal setting, but also highlighted adding breathing exercises for people who use oxygen supply.

> *The function could be called daily activity, because doing the dishes or something else, everything counts, anything that gets you out of the couch....it is an exercise just going to the bathroom, at least for me it is, everything counts, so you don't feel like you are not doing anything at all.* [F2]

The participants also emphasized that physical and social activities for them are often connected, since just getting out of the house or going for a walk with someone covered both functions. Also, some participants struggled to see how the CA could add value when it measured the same things that they already did on their smartphones, for example, step count. Therefore, the social activity function was appealing to a lot of participants, as they did not have a lot of experience in using technology to find social activities. The daily activity function should ideally according to the participants be able to suggest social activities, both online and physical, for example, cinema, theatre, concerts, lung choir, etc. This would motivate and inspire them to do new things and preserve or boost their social life. A calendar function was suggested to help them keep track of upcoming activities, including social activities, daily household activities, physical activities, doctors' appointments, medicine, etc.

In sum, both social and physical activities are important to several of the workshop participants. The participants had different needs in relation to physical activities, as some had a high activity level and some had a lower activity level according to their condition and personal wishes. Participants were open to receiving motivational SMS text messages, but this should be optional; thus, the need for personalization of the CA is still relevant for this function. The participants highlight that physical and social activities, for them, are widely connected and therefore wanted input on social activities and gatherings in the neighborhood. The social activity element adds value to the function, as some participants already measured their steps daily through other technologies but were not familiar with using technology to initiate social activities.

Sleep

The sleep function was well discussed from the beginning of the workshop phases. Some participants did not think that the function was relevant, and others claimed sleep quality to be very important. For the thematic analytical framework, it is important not to stop at this finding. The technology development depends both on the findings of the different opinions and on the specific reasons given by the participants. Those who found it important that "sleep" was included in the CA development justified it with a meaningful aspect, which was "if you don't sleep well, your whole next day is ruined."

> *It [sleep] is important for how you feel during the day, that you sleep well. Personally, I often sleep like hell, and then the day after is ruined. Then you can't get anything done the next day...* [M1]

But what this simple justification refers to is, in the perspective of everyday life, a large and complex problem. If the series of weekdays is ruined by poor sleep, then the seriousness and magnitude of the reason are understood. Participants asked for ideas and help to fall asleep, as this was often a struggle. When it comes to measuring sleep, participants had mixed opinions; some stressed that they did not need technology to tell them that they had a poor night's sleep, and others already used a wearable such as Fitbit (Google) to track their circadian rhythm. Something that was further stressed by this discussion was that the measurements should not be forced; they should be based on need and interest, and not mandatory day-to-day measurements.

> *I know how much I sleep and whether I've slept, I don't need that.* [F1]

> *I think it is probably very different...because there are many who have a little bad sleep, so if they could find out why you have it then you can do something about it, because it means a lot for the whole day, if you have had a bad night, you have a bad day too.* [F2]

Meanwhile, some participants said that they have gotten the right medication and therefore sleep was not something they worry about. Another participant highlighted that if he wakes up during the night, this is not an issue for him as he is retired and can just sleep in if he wants to. Another participant says that this feature could be useful—only if the CA could tell him when he needs to consult a professional for help solving problems related to sleep, instead of him trying to solve it himself.

Together, these descriptions are illustrative of the heterogeneity of needs in relation to sleep. This may indicate that the sleep feature may need a longer co-design process to be able to catch all nuances that evolve as the participants learn about what the technology may do for them. Specifically, the following features were suggested for the sleep function: advice on how to fall asleep, advice on when to consult a medical professional, reminders to wake up and get oxygen, information on how to get better sleep quality, information on the impact of different medicines on sleep, and ideas for calming breathing exercises.

By starting with the older adults' reasons and preferences, the thematic analysis sharpens the sense of proportion. Next, the practical suggestions linked to the technical aids or informative COPD-related details allow themselves to be placed with greater precision in relation to the older adults' perception of technological measures as desirable or not, or useful or useless in everyday use.

Importantly, measurements of sleep were raised several times in conjunction with discussions about this feature. These discussions highlighted that the participants had mixed feelings toward measurements. All participants imagined that such measurements could be useful, but that it must also be optional.

Discussion

Principal Findings

In this section, the principal findings based on the 3 themes—my health, daily activity, and sleep—are summarized and discussed in relation to the overarching identified themes, measurements, and health literacy.

The workshop participants expressed strong interest in the "my health" function; some asked for comprehensive information and knowledge about their health guides to proper in-home measurements, highlighting the interest in increasing health literacy, which was identified as an overarching theme, and the ability to share data with health care professionals and informal caregivers. Measurements were also identified as an overarching theme, and it received mixed input; the participants had different preferences regarding the frequency and capacity of measurements. Therefore, personalization of the CA is necessary, which can be achieved through a "my profile" function enabling individualized setup.

Participants emphasized the importance of social and physical activities within the "daily activity" function. Different needs were identified, as some had a high activity level and others had a low activity level due to their condition. Personalized motivational messages were welcomed optionally. The participants highlighted that physical and social activities for them are widely connected; therefore, social activities and gatherings in the local area to get them out of the house were requested. Social activities added increased value to the function, as some participants already tracked their steps daily through other technologies.

Participants had a mixed interest in the “sleep” function. They argued that if you do not sleep well, your whole day is ruined. Some participants also wanted information on how to fall asleep easier, sleep patterns, and tips for better sleep. Measuring sleep was debated, with some finding it useful and others invasive if not optional.

Measurements

SMILE focuses on older adults’ needs and preferences in undertaking a cocreational approach to technology development. The older adults in the project are participants in developing the CA, patients with chronic conditions, citizens in a community, and users of the CA. As identified in the workshops, not all participants want to be reminded that they are patients every day, for example, through daily measurements, and a lot of emphasis was placed on the importance of social activities being a part of the CA. However, with the implementation of new health technologies, the borders between private homes and the health care system are torn down, including technologies entering citizens’ homes, enabling daily measurements.

Measurements are a cross-cutting theme and a common focal point for the SMILE project’s technology development and for the workshop participants’ decision on technological aids. Daily measurements seem to be an important feature that can optimize and aid participants through, for example, novel health technologies, such as smartwatches. However, daily measurements did receive mixed interest. The workshop participants highlighted that being measured every day, as a person living with COPD, is for some valuable and others harmful. Regardless of this, measurements must be seen in conjunction with other themes, such as physical activity. Some participants already measured their physical activity through Fitbit or on their smartphone (eg, step count).

Moreover, the participants who already monitored their steps daily doubted how the CA could bring them further value. The participants also emphasized that measurement must be easy to do and ideally automatically, as with smartwatches. Participants were open in using one solution, such as the CA, to do daily measurements; if it could collect all the measures, they found valuable as they did not have to measure everything twice.

Based on this, it can be emphasized that measurement must be needs based. Not all participants are interested in measuring several parameters every day. For some participants, the motivation to measure is that they feel ill. During those periods, they feel the need to monitor themselves. Others are motivated to do measurements to follow their condition daily. Regardless, it is important to emphasize that the main attitude of the workshop participants toward the CA was formulated almost as a statement: whatever may be offered to them by technological aids, they do not want it to become a substitute for their face-to-face contact with health care professionals.

Health Literacy

The long-term aim of the development and implementation of the CA is that it can be used by citizens in their own homes, regardless of health conditions and independently of social and professional networks. Studies have addressed how digital solutions and a focus on more hybrid care models have the possibility of relieving the global burden on health care systems and health care staff [22]. However, these digital solutions need to be perceived as valuable by several actors in the specific health care setting and need to be used “correctly” to prevent adding an extra burden on the staff and to enable widespread adoption among patients [23].

As illustrated in the analysis, the participants are interested in learning more about their health, tracking their physical activities, and getting advice and techniques to improve their sleep. In addition, they are interested in reflecting on their life situation as older adults living with COPD, from an everyday perspective.

At the same time, the thematization possibilities can be seen in relation to the development of professional education, which includes both health skills and digital literacy and deals with cocreational technology development. Theoretical approaches and concepts, such as Negt’s [24] conceptualizations of identity and technological competence, are relevant and can be put into the perspective of the themes and issues raised by the workshop participants. Negt’s statements that “identity competence relates to a lifelong learning process” and that “identity is not determined once and for all” resonate in relation to the results from the first phase of the workshops. In relation to the focus of SMILE and the workshop participants, this means that the orientation and perspective of the participants are challenged, partly by what Negt [24] describes as a reality test between “the well-ordered society” and “the marginalized environments, groups” and partly by the development of health literacy as a, didactically speaking, lifelong interaction among case, knowledge, and experience. Concerning technological competence, the challenges refer both to the development of an ability to distinguish between technologies and to the understanding of technology as a social project. In SMILE auspices, it also refers to the dilemma between the technology development system and patient orientation, with the latter guided by human needs, closer to the participant’s everyday life and experiences.

To clarify the technological user value, there is an implicit demand that older adults with COPD have some degree of both health literacy and digital literacy. Correspondingly, it requires a similar and explicit demand that health professionals have digital literacy, not only in terms of individual competence but in terms of collaborative and patient-oriented competence across professions. Ethnographic studies close to health care practice show examples of patient-oriented interprofessional cooperation as well as how interprofessional competence development can support the bridging between the hospital context and the home context [20,21].

Against this background, the results show that participants in the SMILE workshops prefer not to replace their interaction with health care professionals with technology but are open to using it as an additional tool in care. Thus, if the CA is able to inform the users about health-related topics, it needs to be from trusted sources. The participants emphasize that health care professionals must be active contributors to address health challenges alongside technology.

Conclusions

Future development of digital health services should encompass the complexities of everyday life among older adults to enhance usability, adoption, and successful use. Participants in the SMILE workshops are interested in reflecting on their life situation as older adults with COPD from an everyday perspective. Participants are open to using new technology, the CA, to learn about their health, monitor physical activity, and discover social activities in their local area. The CA should not only focus on the "patient" and the condition but also acknowledge the complexities of everyday life as an older adult living with COPD. This means grasping the different needs and preferences of individuals, regardless of their condition, in designing technology is important. Daily measurements should be optional, and digital solutions should be used as a supplement, not a replacement for in-person interaction with health care professionals.

Acknowledgments

The authors thank European Union Horizon 2020 and Smart Inclusive Living Environment (SMILE) for making this work possible. They also thank PreCare, Sykehuset Innlandet Trust, and the municipalities in the Norwegian living lab in Gudbrandsdalen for helping with the recruitment of participants, and above all, they thank the participants for their contribution to this study. This research was funded by the European Union's Horizon 2020 Research and Innovation Program under grant 101016848.

Data Availability

The datasets used and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

EKW wrote the main manuscript text. JBM, KB, CWS, and TK edited and contributed to the manuscript during the writing process, and all authors reviewed the final manuscript. EKW, JBM, and CWS collected the empirical data and analyzed this study. EKW, JBM, KB, CWS, and TK took an active part in processing and analyzing the data in the article.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Consent form.
[DOCX File , 25 KB - humanfactors_v11i1e63222_app1.docx]

References

1. United Nations. UN decade of healthy ageing: plan of action 2021-2030. World Health Organization. 2021. URL: https://cdn.who.int/media/docs/default-source/decade-of-healthy-ageing/decade-proposal-final-apr2020-en.pdf?sfvrsn=b4b75ebc_28 [accessed 2023-06-19]
2. Wegener EK, Bergschöld JM, Whitmore C, Winters M, Kayser L. Involving older people with frailty or impairment in the design process of digital health technologies to enable aging in place: scoping review. JMIR Hum Factors 2023;10:e37785 [FREE Full text] [doi: 10.2196/37785] [Medline: 36705959]
3. Pradhan A, Jelen B, Siek KA, Chan J, Lazar A. Understanding older adults' participation in design workshops. 2020 Presented at: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems; April 25-30, 2020; Honolulu, HI p. 1-15 URL: https://dl.acm.org/doi/10.1145/3313831.3376299 [doi: 10.1145/3313831.3376299]
4. Bilgram V, Brem A, Voigt KI. User-centric innovations in new product development - systematic identification of lead users harnessing interactive and collaborative online-tools. Int J Innov Manag 2008;12(03):419-458. [doi: 10.1142/s1363919608002096]
5. Hesseldal L, Kayser L. Healthcare innovation—the epital: a living lab in the intersection between the informal and formal structures. Qualitative Sociology Review 2016;12(2):60-80. [doi: 10.18778/1733-8077.12.2.04]
6. Kayser L, Furstrand D, Nyman RE, Monberg A, Karnoe A. GoTO: a process-navigation tool for telehealth and - care solutions, designed to ensure an efficient trajectory from goal setting to outcome evaluation. Informatics 2022;9(3):69. [doi: 10.3390/informatics9030069]
7. Rogne AF, Syse A. Framtidens eldre i by og bygd. Befolkningsframskrivinger, sosiodemografiske mønstre og helse. 2017. URL: https://www.ssb.no/befolkning/artikler-og-publikasjoner/_attachment/328026?_ts=167c714ea80 [accessed 2023-06-20]
8. Strand BH, Berg CL, Syse A, Nielsen CS, Skirbekk VF, Hansen T, et al. Helse hos elder I Norge. Health among the elderly in Norway. Norwegian Institute of Public Health. 2023 Mar 13. URL: https://www.fhi.no/nettpub/hin/grupper/eldre/ [accessed 2023-06-20]

9. NOU 2020: 15 (2020) det handler om norge - utredning om konsekvenser av demografiutfordringer i distriktene. It's about Norway - investigation of consequences of demographic challenges in the districts. Norwegian Ministry of Municipal Affairs and Modernization. 2020. URL: https://www.regjeringen.no/no/dokumenter/nou-2020-15/id2788079/?ch=3 [accessed 2024-09-17]
10. Bergschöld JM, Granås SB, Høgseth LD. Challenging the assumption that medical conditions are the prime drivers of older people's needs: a qualitative baseline study about the experienced needs of older people who age in place with COPD. Forthcoming 2023.
11. Kayser L, Karnoe A, Duminski E, Somekh D, Vera-Muñoz C. A new understanding of health related empowerment in the context of an active and healthy ageing. BMC Health Serv Res 2019;19(1):242 [FREE Full text] [doi: 10.1186/s12913-019-4082-5] [Medline: 31014350]
12. WHO global strategy on integrated people-centred health services 2016-2026. World Health Organization. 2015. URL: https://interprofessional.global/wp-content/uploads/2019/11/WHO-2015-Global-strategy-on-integrated-people-centred-health-services-2016-2026.pdf [accessed 2023-11-20]
13. Malterud K, Siersma VK, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
14. Yock P, Zenios S, Makower J, Brinton TJ, Kumar UN, Watkins FTJ. Biodesign - The Process of Innovating Medical Technologies. Cambridge, United Kingdom: Cambridge University Press; 2015.
15. Danish Research Ethics Committees. Overview of mandatory reporting. The Danish National Center for Ethics. 2024. URL: https://researchethics.dk/information-for-researchers/overview-of-mandatory-reporting [accessed 2024-09-17]
16. OpenAI. Botpress. 2023. URL: https://botpress.com/ [accessed 2024-09-17]
17. Kvale S. Doing Interviews. London, United Kingdom: SAGE Publications; 2023.
18. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
19. Borgnakke K. Ethnographic Methods for Researching Innovative Education. Oxford, United Kingdom: Oxford Research Encyclopedia of Education Internet. Oxford University Press; 2023.
20. Patton MQ. Enhancing the quality and credibility of qualitative analysis. Health Serv Res 1999;34(5 Pt 2):1189-1208 [FREE Full text] [Medline: 10591279]
21. Sandelowski M. Rigor or rigor mortis: the problem of rigor in qualitative research revisited. ANS Adv Nurs Sci 1993;16(2):1-8. [doi: 10.1097/00012272-199312000-00002] [Medline: 8311428]
22. Willems SH, Rao J, Bhambere S, Patel D, Biggins Y, Guite JW. Digital solutions to alleviate the burden on health systems during a public health care crisis: COVID-19 as an opportunity. JMIR Mhealth Uhealth 2021;9(6):e25021 [FREE Full text] [doi: 10.2196/25021] [Medline: 34033575]
23. Safi S, Thiessen T, Schmailzl KJ. Acceptance and resistance of new digital technologies in medicine: qualitative study. JMIR Res Protoc 2018;7(12):e11072 [FREE Full text] [doi: 10.2196/11072] [Medline: 30514693]
24. Negt O. Dannelse og demokrati. In: Dannelse og Demokrati. 1. Udgave. Frederiksberg, Denmark: Frydenlund; 2019.

Abbreviations

CA: conversational agent
COPD: chronic obstructive pulmonary disease
READHY: Readiness and Enablement Index for Health Technology
SLE: smart living environment
SMILE: Smart Inclusive Living Environment
WHO: World Health Organization

Edited by A Kushniruk, E Borycki; submitted 13.06.24; peer-reviewed by K Esquivel; comments to author 11.08.24; revised version received 26.08.24; accepted 13.09.24; published 08.10.24.

Please cite as:
Wegener EK, M Bergschöld J, Kramer T, Schmidt CW, Borgnakke K
Co-Designing a Conversational Agent With Older Adults With Chronic Obstructive Pulmonary Disease Who Age in Place: Qualitative Study
JMIR Hum Factors 2024;11:e63222
URL: https://humanfactors.jmir.org/2024/1/e63222
doi:10.2196/63222
PMID:39378067

Original Paper

Views and Uses of Sepsis Digital Alerts in National Health Service Trusts in England: Qualitative Study With Health Care Professionals

Runa Lazzarino[1], PhD; Aleksandra J Borek[1,2], PhD; Kate Honeyford[3], PhD; John Welch[4], MSc; Andrew J Brent[5], PhD; Anne Kinderlerer[6], MBChB; Graham Cooke[6], Prof Dr, FRCP; Shashank Patil[7], FRCEM; Anthony Gordon[6], MD; Ben Glampson[6], MSc; Philippa Goodman[8], MPhil; Peter Ghazal[9†], Prof Dr; Ron Daniels[10], MBChB; Céire E Costelloe[3], Prof Dr; Sarah Tonkin-Crine[1,2], PhD

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

[2]National Institute for Health and Care Research (NIHR) Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance, University of Oxford, Oxford, United Kingdom

[3]Team Health Informatics, Institute of Cancer Research, London, United Kingdom

[4]University College Hospital, London, United Kingdom

[5]Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom

[6]Imperial College Healthcare NHS Trust, London, United Kingdom

[7]Chelsea and Westminster Hospital, London, United Kingdom

[8]PPI representative, Oxford, United Kingdom

[9]School of Medicine, Cardiff University, Cardiff, United Kingdom

[10]UK Sepsis Trust and Global Sepsis Alliance, Birmingham, United Kingdom

[†]deceased

Corresponding Author:
Runa Lazzarino, PhD
Nuffield Department of Primary Care Health Sciences
Medical Division
University of Oxford
Radcliffe Observatory Quarter
Woodstock Road
Oxford, OX2 6GG
United Kingdom
Phone: 44 01865 289300
Email: r.lazzarino@ymail.com

Abstract

Background: Sepsis is a common cause of serious illness and death. Sepsis management remains challenging and suboptimal. To support rapid sepsis diagnosis and treatment, screening tools have been embedded into hospital digital systems to appear as digital alerts. The implementation of digital alerts to improve the management of sepsis and deterioration is a complex intervention that has to fit with team workflow and the views and practices of hospital staff. Despite the importance of human decision-making and behavior in optimal implementation, there are limited qualitative studies that explore the views and experiences of health care professionals regarding digital alerts as sepsis or deterioration computerized clinician decision support systems (CCDSSs).

Objective: This study aims to explore the views and experiences of health care professionals on the use of sepsis or deterioration CCDSSs and to identify barriers and facilitators to their implementation and use in National Health Service (NHS) hospitals.

Methods: We conducted a qualitative, multisite study with unstructured observations and semistructured interviews with health care professionals from emergency departments, outreach teams, and intensive or acute units in 3 NHS hospital trusts in England. Data from both interviews and observations were analyzed together inductively using thematic analysis.

Results: A total of 22 health care professionals were interviewed, and 12 observation sessions were undertaken. A total of four themes regarding digital alerts were identified: (1) support decision-making as nested in electronic health records, but never substitute professionals' knowledge and experience; (2) remind to take action according to the context, such as the hospital unit and the job role; (3) improve the alerts and their introduction, by making them more accessible, easy to use, not intrusive, more accurate, as well as integrated across the whole health care system; and (4) contextual factors affecting views and use of alerts in the NHS trusts. Digital alerts are more optimally used in general hospital units with a lower senior decision maker:patient ratio and by health care professionals with experience of a similar technology. Better use of the alerts was associated with quality improvement initiatives and continuous sepsis training. The trusts' features, such as the presence of a 24/7 emergency outreach team, good technological resources, and staffing and teamwork, favored a more optimal use.

Conclusions: Trust implementation of sepsis or deterioration CCDSSs requires support on multiple levels and at all phases of the intervention, starting from a prego-live analysis addressing organizational needs and readiness. Advancements toward minimally disruptive and smart digital alerts as sepsis or deterioration CCDSSs, which are more accurate and specific but at the same time scalable and accessible, require policy changes and investments in multidisciplinary research.

(JMIR Hum Factors 2024;11:e56949) doi:10.2196/56949

KEYWORDS

digital alerts; electronic health records; computerized clinical decision support systems; sepsis; patient deterioration; decision-making; secondary care; emergency care; intensive care; England; qualitative study

Introduction

Background

Sepsis is an uncontrollable response of the body to an infection, whereby the immune system starts attacking its own tissues and organs leading to organ dysfunction [1]. Sepsis is a common cause of serious illness and death. There are an estimated 918,000 hospital admissions with suspected sepsis, up to 250,000 cases of sepsis, and 48,500 deaths associated with sepsis in the United Kingdom each year [2]. Similarly, high levels of sepsis are reported internationally [3,4], with 11 million sepsis-related deaths calculated for 2017, representing approximately 20% of all deaths globally [5,6]. Sepsis is recognized by the World Health Organization as a global health priority [7].

Sepsis symptomatology can look like that of other conditions, potentially delaying and misleading diagnosis and treatment. As such, sepsis demands prompt intervention to avert poor outcomes [8]. Evidence has shown that timely, appropriately targeted, intravenous antibiotics are effective in improving outcomes for patients with sepsis [9-11]. On the basis of this evidence, hospitals in the United Kingdom have been set targets to rapidly diagnose and administer intravenous antibiotics [12,13]. Other countries have introduced national guidance for sepsis [14]; and the International Surviving Sepsis Campaign updated guidelines for the management of sepsis in 2021. These latter guidelines recommend the adoption of performance and quality improvement initiatives and of protocols and screening tools to improve and accelerate the identification and treatment of sepsis in hospitals [8,15], although these processes remain challenging and suboptimal [16,17].

Prior Work

Interlocking with the complexity of sepsis and the clinical intervention required [1], factors linked to the professional profile and the work environment affect the timely management of sepsis in secondary care. Previous qualitative research with health care professionals (HCPs) has highlighted limits in professionals' capacity to identify sepsis, difficulties in hand over of patients, and errors in communication [18-22]. Time pressures, the intense workload, and the complex clinical environment are all barriers which need to be overcome; simultaneously, well-coordinated multidisciplinary work and effective leadership are necessary to combine smoothly for effective sepsis management [2,18,20,21].

To support rapid sepsis diagnosis and treatment, screening tools have been proposed. These include the quick sepsis-related organ failure assessment [qSOFA; 1]; systemic inflammatory response syndrome (SIRS) criteria [23,24]; and in the United Kingdom, the national early warning score (NEWS) in 2012 [25,26], updated in 2017 as NEWS2 [12,27]. More recently, machine learning algorithms for early recognition of sepsis have been developed, performing better when compared with the aforementioned sepsis scoring tools [28,29].

With the expansion of electronic health records (EHRs) across the UK National Health Service (NHS), screening tools for sepsis have been embedded into hospital digital systems (DSs) as digital alerts. These digital tools are also known as computerized clinical decision support systems (CCDSSs) and are meant to enhance clinical decision with both targeted and general information [30], but there are also considerable concerns about alert fatigue [31]. There are different ways in which digital alerts, including sepsis or deterioration CCDSSs, interface with users. A total of 3 categories have been identified [32]. Hard-stop intrusive alerts pop up and either prevent users from taking any action or allow them to proceed only with the override of a third party, such as a senior decision maker. Soft-stop intrusive alerts allow the user to proceed after entering a response, often from a multiple-choice box. Nonintrusive, passive alerts do not pop up or interrupt the user workflow or require any interaction on the part of the user. Sometimes these are in-line, for example, an alert appearing in a row of an electronic patient list.

Previous evidence on the effect of digital alerts as sepsis or deterioration CCDSSs (henceforth, sepsis digital alert) on care quality and patient outcomes offered mixed results [33]. Work in NHS trusts in England demonstrated that the introduction of digital sepsis screening tools with their accompanying alerts was associated with an increase in timely treatment and reduction in risk of mortality in 2 trusts [34,35]. Some studies, mainly conducted in the United States, have demonstrated an increase in patients receiving timely intravenous antibiotics [36,37] and a decrease in in-hospital mortality [38-40] and length of stay [36,41,42], but others have not found any impact either on mortality [37,41,43,44] or hospital length of stay [39].

Sepsis digital alerts, which are embedded into the hospital EHRs DS, fit into the highly complex workflow of the management of sepsis, as any other CCDSSs. Qualitative literature exploring HCPs' views and experiences with sepsis digital alerts is scarce. A recent study [45] found that HCPs' perceptions of a hypothetical sepsis digital alert are linked to the level of trust professionals had in the alert; HCPs' trust was enhanced both by their previous experience with similar alerts and by being engaged in implementation and training initiatives about the sepsis digital alert. In another US-based study, physicians in a pediatric emergency department (ED) were interviewed about reasons for accepting or rejecting a sepsis electronic best practice alert: two-thirds of participants considered nonpatient factors relevant specific to the ED environment, individualized practice patterns, the digital tool design, and education [46].

Goal of This Study

The emerging evidence of the contribution of sepsis digital alerts is promising but still limited; there are no validated digital tools available to NHS trusts, with scarce evidence as to which tool to use and their effect on patient outcomes. In addition, the introduction of sepsis digital alerts is often accompanied by treatment plans, under the aegis of antimicrobial stewardship, and by implementation and quality improvement initiatives and education and training for HCPs; while these have been shown to yield positive results [34,47,48], little is known on how these actions affect the implementation and impact of the alerts.

This study aimed to explore the views and experiences of HCPs on the use of sepsis digital alerts and to identify barriers and facilitators to the implementation and use of sepsis digital alerts in NHS hospitals.

Methods

Study Design

This is a qualitative, multisite study with semistructured interviews with HCPs and unstructured observations of HCPs working in hospitals. It is nested within a broader program of research seeking to evaluate the impact of sepsis digital alerts on patient outcomes and staff activity in NHS hospitals, the Digital Alerting for Sepsis (DiAlS) study.

Sample and Settings

A combination of purposive and convenience sampling was used. A total of 3 NHS trusts were selected as sites for this qualitative study from the 6 hospitals involved in the DiAlS study. The 3 sites were chosen with consideration of diversity in the EHRs DS and in the sepsis digital alerts, previous evidence on the evaluation of sepsis digital alerts [34], and implementation and quality improvement initiatives. Trusts' coinvestigators were asked to identify and invite potential participants for interviews from their trusts. Participants were sampled to include a variety of job roles and a range of hospital units (EDs, outreach emergency response teams, and acute and intensive care units [ICUs]). In 2 sites (1 and 2), observations were conducted, and coinvestigators supported the research team to this end providing initial liaising with relevant colleagues in ED and the outreach team.

Data Collection

All data were collected between November 2022 and July 2023 by RL and AJ Borek. Most semistructured interviews with HCPs were conducted remotely using videoconferencing software (Microsoft Teams); the remaining were conducted in person at the participant's place of work. The topic guide was developed from the study objectives with input from the wider research team and the study patient and public involvement representative (Multimedia Appendix 1). Questions asked HCPs about their experiences of identifying and managing patients with sepsis and about their views and experiences of using sepsis digital alerts. Where relevant, HCPs were asked about their experience of developing and implementing sepsis digital alerts. Interviews lasted between 30 and 75 (mean 52) minutes, were audio recorded, and were professionally transcribed verbatim and pseudonymized.

Unstructured observations [49] sought to observe clinical practice to see how sepsis digital alerts fitted into the workflows of HCPs in different roles. We sought to assess what impact they had on clinical decision-making and to identify whether sepsis digital alerts were used differently by HCPs in different roles. A total of two types of observation were done: (1) observations of practice in EDs, with occasional informal conversations, and (2) one-to-one shadowing of an HCP in outreach teams, with more frequent dialogue. Either the HCP who expressed interest in being shadowed or the ED head or manager consented in writing before the observations. The observations lasted between 2 and 4 hours. Paper-and-pencil notes taken during the observations were anonymized and typed up.

Data Analysis

Data analysis began concurrently during data collection and was supported by NVivo (version 12; QSR International). Data from both interviews and observations were analyzed together inductively using thematic analysis [50,51]. Similarities and differences between transcripts were assessed using a constant comparison approach [52]. A codebook was developed to code the whole data set, across the 3 sites. Codes were compared with one another to create categories, grouping similar codes together. All categories were clearly named to ensure that only related data were included in that category. However, some codes and categories remained site specific to allow for comparisons between sites. Regular core team (RL, AJ Borek, and STC) meetings accompanied data collection and analysis phases to deliberate on the codebook and the data analysis and to follow an iterative approach, which is documented in a number of Microsoft Word documents with annotations and comments for auditing purposes. We referred to the 8 quality criteria for qualitative research to ensure rigor and trustworthiness of our process [53].

Ethical Considerations

The wider program of research, the DiAlS study, of which this study is a part, was reviewed and approved by the UK NHS Health Research Authority (Project ID—288,328). This qualitative work stream was further separately reviewed and approved by the Research Governance, Ethics and Assurance Team of the University of Oxford and the UK NHS Health Research Authority of England and Wales (Project ID 313699-22/PR/1020). The research teams of each trust reviewed and approved the study at the site level. Full verbal or written consent was obtained from all participants included in this study. Data were pseudonymized. Participants were offered a voucher to thank them for their time. The monetary value of the voucher was £20; US $25.

Results

Overview

We interviewed 22 HCPs: 8 (36%) from site 1, and 7 (32%) from each of the other 2 sites (Table 1). In site 1 and 2, a total of 12 observation sessions were also undertaken: 5 (42%) in the EDs and 7 (58%) sessions involved shadowing professionals, 3 (25%) of whom completed an interview. We identified 3 themes about HCPs' views and use of their sepsis digital alerts and a fourth theme capturing the complexity of how the hospital environment affected sepsis digital alerts' use across the 3 sites. The first 3 themes feed into the fourth theme, which allows a comparison between the 3 NHS trusts involved in this study. The presentation of the results was structured in this way as each section is seen as building on the previous one and supporting the following one.

Table 1. Characteristics of interview participants (N=22).

Site	Participants, n (%)	Declared gender (women), n (%)	Age (y), mean (range)	Job title and hospital unit (n)	Years of experience, mean (range)	Years of use of digital alerts, mean (range)
Site 1	8 (36)	3 (14)	43.5 (34-55)	• Nurses and OT[a]: n=3 • Nurse and ED[b]: n=1 • Consultants and ICU[c]: n=3 • Consultant and ED: n=1	10 (7-18)	6 (4-10)
Site 2	7 (32)	5 (23)	36 (29-42)	• Nurse and OT: n=2 • Nurse and ED: n=1 • Nurse and AMU[d]: n=1 • Consultant and OT: n=1 • Consultant and ED: n=1 • Consultant and ICU n=1	6 (3-8)	5 (1-7)
Site 3	7 (32)	7 (32)	43 (30-54)	• Nurses and OT: n=3 • Nurse and ED: n=1 • Nurse and AMU: n=1 • Consultants and ED: n=2	9 (1-20)	7 (5-10)
Total	22 (100)	15 (68)	41 (29-55)	• Nurses: n=13 • Consultants: n=9 • OT: n=9 • ED: n=7 • AMU: n=2 • ICU: n=4	11 (1-20)	6 (1-10)

[a]OT: outreach team.

[b]ED: emergency department.

[c]ICU: intensive care unit.

[d]AMU: acute medical unit.

Theme 1: Alerts Nested Within EHRs Support Decision-Making

All participants liked that sepsis digital alerts were nested within the DS of the EHRs shared across a trust. Having patient data all in 1 place, including clinical history, patients' trends during their hospital stay, preconditions, comorbidities, test results, and various digital alerts or scores, was described as useful to more quickly and safely build a picture of the patients who are flagged by the sepsis digital alert. This was regarded as enabling better decision-making and quality care:

> *We've been electronic in ICU for a long time but, of course, the interaction has been difficult because of the wards being paper-based and now having everything available everywhere is incredibly helpful. So, the cutting out of searching for information, it makes things much more efficient and I think as a result, much safer.* [ICU consultant, site 1]

All participants underlined that the sepsis digital alert in the patients' EHRs were supporting, and by no means substituting, clinical decision-makers, whose knowledge and experience were of paramount importance in sepsis identification and management. Sepsis digital alerts were seen as "a piece of the puzzle" (outreach team nurse, site 3), neither intended for, nor leading to, the formulation of a diagnosis or clinical decision in isolation:

> *I think it's another objective piece of the puzzle that will support junior members to ask for more help because they cannot ignore it, which is helpful I think, when they are wondering how to identify that someone is unwell.* [Outreach team nurse, site 3]

> *I will go and look in the obs chart for the patient to figure out what obs would have triggered the alerts and if there's congruence. So, does the presentation and the obs match the sepsis digital alert?...The more senior you get the more nuance you're looking for.* [ED consultant, site 3]

Some participants placed greater emphasis on the teamwork support function of the sepsis digital alert. They valued that they could access information related to clinical observations and the actions of their colleagues, reach out to them if needed, and factor these data into their own decisions. Other participants emphasized the monitoring support function of the sepsis digital alert; having a quick synopsis of the condition of all the patients in a hospital unit allowed better patient prioritization and management of workload:

> *Nowadays, I am not looking at six patients, which I can go you, you, you, you, you. I'm looking at 60 and how do I look at those patients? I can't. How do I look at them? I look at them electronically and you know, the digital system and the alert enabled me, despite not flawlessly, to build a really good impression about what the acuity is and where the danger is within my department.* [ED consultant, site 1]

At the operational level, participants described how sepsis digital alerts are embedded within a complex, multimodal way of working, which includes bleepers, mobile phones, and landline phones for communications between hospital units' team and with the laboratories; paper notes in absence of available computers to quickly annotate patients' observations before uploading them onto the DS; and face-to-face interactions, including for very urgent escalations:

> *The triage nurse is quite far from us...So, if they're very worried and they can't get through on our phone, then they come and find us in person and walk to us and ask for advice. Usually, because they're also quite busy, they usually try to ring us and just make us aware of the patient on the screen and they just say, please can you prescribe paracetamol, antibiotics, and request patient review. Then I can do it over the phone. In A&E, we don't really hold bleeps. Bleeps are more for the ward team.* [ED consultant, site 3]

Theme 2: Alerts Are Reminders That Lead to Context-Dependent Actions

Participants viewed sepsis digital alerts as useful to remind and prompt HCPs of a number of actions, from reviewing a patient's information in the EHRs to visiting them in person. In general, participants reported that sepsis digital alerts' utility decreased with increasing staff training and experience with sepsis cases, and with the higher senior HCP:patient ratio of certain units, such as acute or intensive care. More junior HCPs said that sepsis digital alerts afforded them with greater confidence to further investigate and interpret why sepsis digital alerts triggered, which could lead them to follow the sepsis protocol and to escalate to senior decision-makers:

> *Thinking back [to] being a junior, one of the big lessons is learning to recognise a critically sick patient, and that takes experience, so having some hard parameters to hang your hat on is really helpful, because I think we all remember you know running to get help when actually things were fine.* [ICU consultant, site 1]

For senior participants who "have a greater cognitive load" (acute medical unit nurse, site 3), the sepsis digital alerts were described as reminders to avoid missing actions for patients in their department:

> *I think the alert itself is really helpful, I think particularly for when you're looking through a patient list, the whiteboard of patients, to have the visual prompt there of somebody that may be way down the list in view of time to be seen, waiting to be seen, but if they've got that alert on the system then generally a senior registrar will pick those patients out and review their case and potentially start the right treatment before the patient is fully assessed, so I think the prompt certainly helps with that identification.* [AMU nurse, site 3]

HCPs' specialty and experience, and hospital units hosting different patient cohorts—with varying conditions, treatments, and lengths of stay—were seen to play a role in the use and views of sepsis digital alerts. Thus, some participants commented that they found sepsis digital alerts unspecific and oversensitive, which potentially could lead to overtriaging, overreferring, and, more rarely, overtreating patients:

> *In a major trauma ward the alert could be triggering because, I don't know, they've got local pressure...It doesn't necessarily mean they've got sepsis, it's very injury-related. Those nurses might find it a little bit frustrating because they will be like, I don't need to go and give antibiotics straight away, with the sepsis six treatment, because it is irrelevant for this quota of patients.* [AMU nurse, site 3]

Nevertheless, participants highlighted that this also had the positive implication of increasing inter- and intrateam communication, and reducing the risk of missing patients:

> *The threshold for which EHR alerts are generated is very, very low so there's different criteria that could match together to generate that alert and, therefore,*

> *we do get a degree of inappropriate patient alerts and referrals. But equally it's much better to because if we screen those patients, it's better that we have more rather than missing some...with less sensitive criteria.* [Outreach team nurse, site 2]

Theme 3: Improving Alerts and Their Introduction

Most participants expressed the importance of sepsis digital alerts being easy to use and accessible; some participants underlined that a user-friendly sepsis digital alert was important so that new staff could be trained more easily and quickly in its correct use. Other participants felt that ergonomic sepsis digital alerts were more likely to facilitate HCPs' work and teamwork and to be acknowledged. Conversely, complicated interfaces and pop-up sepsis digital alerts were seen to interrupt HCPs' work, requiring several steps and with the risk of confusing, desensitizing, and irritating users. In result, sepsis digital alerts were reported to be overridden and ignored. By contrast, some participants warned against the alerts being used in excess and deskilling hands-on practice because HCPs, especially those more junior, may "become so fixated on the number that they forget the core part of some of their nursing skills" (outreach team nurse, site 1).

Thinking about potential improvements of sepsis digital alerts, participants suggested adding a checklist of what a colleague has or has not done, or key pieces of information regarding the patient that could be sent in an SMS text message when referring a patient. Participants also suggested quick training that could be accessed via a smartphone:

> *Currently to learn about sepsis, it requires you to log off the computer, to watch a video, to sit in front of a computer that you can only use with a plug, then it's one of the barriers, but if you have a QR code, that anyone can access—because everyone got a smartphone at the moment—with a quick reference guide to what you need to do, but also some information about sepsis and also some training videos, what you've got to do, that will really help.* [AMU nurse, site 3]

In this respect, several participants, especially from the outreach teams, envisioned sepsis digital alerts embedded into portable devices so that HCPs could be reached when they are on the move. Other features of more advanced sepsis digital alerts that participants suggested included alerts being adjusted to the patient, factoring in their baseline parameters, comorbidities, and previous conditions or addressing relevant HCPs team; and alerts processing more information regarding the patient, transparently showing why they triggered and allowing for greater interaction with users:

> *What needs to happen is that when results come back and they're horribly abnormal, the doctor or the nurse looking after the patient is alerted to that fact without them having to log into a computer...without having to remember the patient's name and hospital number and then clicking through a load of tabs to find what the blood test is. Wouldn't it be nice if there was some way that a person could be alerted directly that this particular patient has a particular problem?* [ICU consultant, site 1]

Many participants wished to have a sepsis digital alert embedded in a DS shared beyond the trust and across the community, ambulance service, and primary care. Participants felt that a widely integrated sepsis digital alert would reduce ED waiting, triaging, and handover times:

> *If there was a system whereby we could link up our different services and have those observations pulled through so that those patients that don't get the alert when they have a set of normal obs in hospital, still have some other way of being flagged on the system that actually this concerning presentation was the case an hour ago in the community.* [AMU senior nurse, site 2]

Finally, participants found the implementation optimization and quality improvement initiatives around the sepsis digital alert useful and emphasized the need for ongoing training and education on sepsis identification and management. In all 3 trusts, at piloting and rollout, the sepsis digital alerts were iteratively adjusted based on staff feedback and regular training around sepsis. However, as some participants highlighted, several other "human factors" should be targeted by broader training aiming to change the organizational culture:

> *The predominant obstacle's definitely human factors so where you get cultural norms within a ward and certain clinical areas...there's just that kind of like, "Oh, we take care of our patients really well and they would never get sepsis" kind of attitude. There are clinical areas that we go to and there seems to be like a resistance [to] intervention from a specialist team because it's like, "Oh, well, if it was that, then we would have noticed"...it's that kind of assumption that they would know if the patient was going to deteriorate, which actually when you look at the evidence is not the case.* [Outreach team nurse, site 2]

Theme 4: Contextual Factors Affecting Views and Uses of Alerts in the Trusts

This fourth theme brings together and present the first 3 themes against 4 sets of factors that affect the views and uses of sepsis digital alerts. Some of these factors differently combine in the context of each of the 3 trusts, allowing comparative reflections. The following quote encapsulates several of these factors constituting the complexity of the hospital environment:

> *It's all human factor stuff. So, yes, we've got the digital alert, yes people know it might be sepsis, but what gets in the way is people. The context in which they're working, their environment, the business of the ward, the acute conditions of the other patients, the demands on them. It swings both ways, if people are really not busy, not that that ever happens, but then people tend to do less. Then when they're really, really busy things also get missed and they get a bit swamped.* [Outreach team nurse, site 1]

The first set is that of the factors pertaining to individual HCPs. As reported in theme 2, HCPs' seniority and clinical specialty affect the use of sepsis digital alerts. On occasion, personal circumstances (eg, childcare duties and other nonprofessional commitments) can play a role in their decision-making.

The second set of factors relate to the hospital unit or department; these will have HCPs with specific training and specialty caring for certain types of patients, and this influences the use of sepsis digital alerts in these patient populations. For example, in intensive units, such as ED resuscitation and ICU, sepsis cases are seen more frequently, and more senior staff look after more severely ill patients so observations are taken more frequently. Participants reported that these aspects make the sepsis digital alert less relevant.

The third set is that of the factors pertaining to the trust. The workload:staffing ratio, along with the presence of senior decision-makers per patient, are specific to the hospital unit but also dependent on the overall management and resources of the trust. Several participants, from sites 2 and 3, raised the issue of delayed actioning of the sepsis digital alert due to heavy workload, and to the retrospective uploading of patients' observations or clinical actions performed, resulting in lower performance toward meeting targets (eg, antibiotics within 1 h) at time of auditing:

> *The nursing notes said "IV access obtained. Antibiotics given" and then an hour or so later you see at 13:30pm antibiotic prescribed. 13:31pm antibiotic given. It all just seems a little bit like that's all been done after the fact. So sometimes you just have to infer that it sounds like they were really on top of this, and they just left the documentation, rightly, till the end, but we can usually tell, and you can see that gap between prescription and administration, that's often the bit that tips it over the 60 minutes.* [Outreach team consultant, site 2]

As highlighted in theme 3, the trust plays a role by investing in staff education or training about sepsis per se and sepsis digital alerts, in the implementation of quality improvement and in technological equipment, spanning from the number of computers available to introducing useful and accessible software programs.

The fourth set of factors is that of the digital tool itself, its features, and functionalities, which correlate with optimal use of sepsis digital alerts, as theme 3 encapsulated. Some aspects inherent to the sepsis digital alert are closely related to other factors, such as how ergonomic and accessible an alert is, which are linked to both trust and individual factors.

The 4 sets of factors differently combine in each of the trusts included in this study. Reading these in conjunction with the characteristics of each trust (Table 2) allows some comparative results, to which the unstructured observations have proved particularly enlightening.

In site 1, the DS was more recently introduced, with ED as the leading unit in the implementation; it is a straightforward system with hardly any intrusive alerts, but with a number of linkable phone apps and functions, such as the DS chat. The sepsis digital alert is a patient deterioration, nonsepsis specific, and passive alert. A nurse-led outreach team operates 24/7. The combination of these elements results in a perceived easiness-to-use of the DS and usefulness of the deterioration digital alert. Results indicate that the deterioration digital alert is acknowledged by HCPs and ignites the intended behaviors (eg, review, escalate, investigate, and visit). This positive pattern appears to occur even in the apparent absence of a major focus from the Trust on sepsis per se. Of note is that these same elements are related to, for some participants in site 1, an excessive reliance and use of the deterioration digital alert as a patient referral trigger; this means that the deterioration digital alert has been perceived to raise the number of patients' referrals, sometimes unnecessarily.

In site 2, the sepsis digital alert is nested within a DS considered "clunky" (ED consultant, site 2) by some participants; some professionals expressed the opinion that digital alerts were too numerous and could cause them an "alert fatigue" (outreach team nurse, ED nurse, ICU consultant, and acute medical unit nurse). Sepsis has been a priority on the trust's agenda, including via the development of an ad hoc algorithm for the sepsis digital alert, the establishment of an outreach team dedicated to sepsis, and several collateral initiatives related to sepsis management and antibiotic prescribing.

Nevertheless, some ED professionals reported that, within the framework of a very busy workflow and workload, there were team communication issues in the department, which meant that the sepsis digital alert could not always be acknowledged and acted upon timely

The pandemic and staff changes in the outreach team for sepsis meant that they felt that their role was not always known in ED; this might also be because at the time of this study, the 2 nurses in this team had been in their role for 8 or 9 months only. The team for sepsis was infrequently in person in the ED or onward, reporting occasional feelings of being negatively perceived as wanting to interfere with the work of colleagues. Some site 2 ED participants wished they had a 24/7 outreach team for more support with emergencies to alleviate workload which was perceived as untenable. The planned expansion of the outreach team for sepsis, with more staff and aiming to be operative 24/7, is a promising response on behalf of this trust to the heavy workload of professionals in ED and trust-wise.

Site 3 presents yet another different configuration of elements. Over the decade of DS use, lessons have been learned about the counterproductive effects of having too many soft-stop and pop-up alerts, and these have been actively reduced. The rollout of the sepsis digital alert was accompanied by, and optimized via, weekly flat-hierarchy, multidisciplinary meetings on sepsis awareness and optimized based on how the sepsis digital alert should work and look like. Site-3 participants involved in those meetings found them informative and useful. This initiative, together with the trust's ongoing investment in sepsis education and training in the DS and the sepsis digital alert, appeared to support better use of the digital tool. This was the case even though the sepsis digital alert being nested within the same DS as site 2 and also described as not user-friendly. In site 3, however, the sepsis digital alert dialogue box is different, with tailored interface to the HCP role (whether a prescriber or a nonprescriber is logged-in). An aspect that was reported as needing improvement was the Situation-Background-Assessment-Recommendation form, which was often done ex post as paperwork, and therefore, did not fulfill its full potential in speeding up referrals. Finally, the support of a nurse-led, 24/7 outreach team was an extra resource for staff and the NEWS2 score was fruitfully used in combination with the sepsis digital alert.

Table 2. Main characteristics of National Health Service trusts' research sites.

Characteristics	Site 1	Site 2	Site 3
The EHRs[a] DS[b]			
When DS introduced	DS was introduced in 2019	DS was introduced in 2015	DS (same as site 2) was progressively introduced from 2013
DS interface and access	DS interface user-friendly and the same across the site and job roles	DS interface not straightforward, with numerous tabs and colours; DS is different in intensive care	DS interface not straightforward, with numerous tabs and colours; different DS login for nurses and doctors
Digital alerts in the DS	Other alerts in use, mainly passive ones (eg, for ED[c] triaging)	Other soft-stop digital alerts, as well as passive ones (eg, for ED triaging)	Other soft-stop digital alerts, as well as passive ones (eg, for ED triaging)
Additional functions of the DS	DS has several functions, used as a phone app and linked to tablets for patients' observations	Use of DS's additional functions was not found	Other software and applications are in use, eg, on sepsis management and antibiotics administration
The sepsis or deterioration digital alert			
When sepsis digital alerts introduced	Alerts was introduced in 2019	Alerts was introduced in 2016	Alerts' was introduced in phases from 2016
Features of the sepsis digital alert	Sepsis digital alert is NEW2[d] (sepsis is suspected with a score of 5+ and a confirmed or suspected infection)	Sepsis digital alert is based on in-house adaptation of sepsis red flag screening tool	Sepsis digital alert is built in the DS, based on St John Sepsis Algorithm; alert informed by a binary alarm: (1) for potential sepsis and (2) for potential severe sepsis
Sepsis digital alert's functioning	Passive icon that changes color according to the score	Alert is both soft-stop and passive; prescribers and nonprescribers can respond differently	Alert is both soft-stop and passive; prescribers and nonprescribers can respond differently
Sepsis digital alert's interface	On the wards, alert embedded in EHR and does not pop up; in ED, alert appears on the digital list of patients	On the wards, the alert box pops up when an EHR is opened and closed; in ED, the alert is also a passive, colorful icon on patients' digital dashboard	On the wards, the alert box pops up when an EHR is opened and closed; in ED, the alert is also a passive, colorful icon on patients' digital dashboard
Trust actions on sepsis, the DS and the sepsis digital alerts			
Trust's position	Deterioration and acuity are the priority	Sepsis was among the trust's priorities up until COVID-19; there were several initiatives for sepsis awareness raising and education (eg, sites' performance competitions, videos and information on the intranet, sepsis trolleys, and champions)	Sepsis was among the trust's priorities up until COVID-19; there were several initiatives for sepsis awareness raising and education (eg, sites' performance competitions, videos and information on the intranet, sepsis trolleys, and champions)
Sepsis digital alerts' roll out	ED led on alerts' implementation	ED led on alerts' implementation	Weekly flat-hierarchy, multidisciplinary meetings on sepsis awareness and alert's optimization
Approach to digital alerts	Keep a minimal number of soft-stop digital alerts	Digital alerts were reported as too numerous	Keep a minimal number of soft-stop digital alerts
Trust's deterioration or sepsis team	NEWS2-based, nurse-led, 24/7 outreach team	Sepsis team for awareness and training, and to support hospital units following sepsis digital alerts; active during office hours.	NEWS2-based, nurse-led, 24/7 outreach team

[a]EHR: electronic health record.

[b]DS: digital system.

[c]ED: emergency department.

[d]NEWS2: National early warning score 2.

Discussion

Principal Findings

Participants generally viewed sepsis digital alerts positively but emphasized that they cannot substitute the HCP’s knowledge and experience in the identification and management of sepsis. Sepsis digital alerts are only a piece of the puzzle in a patient’s presentation, as well as in the complex, multimodal, and multidisciplinary clinical practice. Participants considered them as useful, context-specific reminders prompting HCPs to take a range of actions, from reviewing to escalating a patient.

Participants identified features of better sepsis digital alerts, such as accessibility and user-friendliness. More sophisticated sepsis digital alerts should be more specific; be patient based; target HCP teams; be portable and remotely accessible; and integrate community, ambulance, and primary care with secondary care to accelerate ED triaging.

Factors pertaining to the individual HCP, the hospital unit, the trust, and the digital tool itself differently combine and were seen to affect the use of sepsis digital alerts in the 3 trusts in this study. The combination of these 4 sets of factors leading to the more optimal use of a sepsis digital alert include a general, non-ICU with a lower senior decision maker:patient ratio; HCP’s previous experience with the sepsis digital alert; presence of a 24/7 emergency outreach team; sepsis digital alert’s quality improvement initiatives and continuous sepsis training; strong technological resources in the trust; good staffing and teamwork; digital tool’s ease-of-use; and digital tools that are not numerous and intrusive.

Strengths and Limitations

We included 3 NHS hospital trusts in this study to explore differences in relation to the EHRs DS and the sepsis digital alert, the implementation optimization strategies, and the approach and training in relation to sepsis. The study set out to involve ED, outreach or sepsis teams, and ICU professionals in different roles and career stage; as a result, we obtained a varied sample. The multisite design and the varied professional profiles afforded meaningful comparisons across job roles, units, and trusts toward a more nuanced identification of factors affecting the use of the sepsis digital alerts. Observations provided insights on how the sepsis digital alert fitted with workflows that were richer than the self-reported descriptions of sepsis digital alert in the interviews. Finally, 2 researchers (RL and AJ Borek) with different disciplinary backgrounds conducted interviews and observations in 1 of the sites. This added rigor to the process of data collection and analysis.

All 3 sites in this study are large, high-resourced, and urban university hospitals; research in trusts with different characteristics and contexts (eg, smaller district hospitals and hospitals in lower-resources settings) could convey different results. HCPs from other hospital wards and units may have different experiences and provide a contrasting example to further understand how sepsis digital alerts are viewed and used but was beyond the resources available for this study. Due to research team capacity and time constraints dictated by the project timeline, we could not include all the 3 trusts for the observations. Finally, the study recruitment strategy resulted in sampling and nonresponse biases which contributed to the fact that certain professional categories, such as junior doctors, were absent. Although we made efforts to recruit junior doctors, none volunteered to participate. This was influenced by the high workload and limited time and capacity as the study was conducted at the time of junior doctor strikes. The absence of junior doctors is a remarkable limitation as sepsis digital alerts, as this study has also found, may be more useful for less senior health professionals.

Comparison With Prior Work

The introduction of sepsis digital alerts rests on the evidence of the challenges to optimally identify and manage sepsis in hospitals, in particular in EDs, where HCPs have been found to need more confidence and time to assess and escalate septic patients [18,21]. Our study found that sepsis digital alerts support HCPs in identifying and making quicker decisions about deteriorating patients, which might be particularly important and helpful to new and less experienced HCPs in ED and in the general wards. The sepsis digital alert is an additional element that HCPs factor in their practice; it is not a substitute for their judgment. Similarly, a US study found that participants were more inclined to accept the machine learning–based system for sepsis if they perceived the tool as a partner, supporting their autonomy and workflow, and not a surrogate of their clinical judgment [54]. For this same reason, another work involving hospital leaders found that participant tended to distrust more machine learning than rule-based sepsis CCDSSs [55]. The literature on CCDSS implementation supports this finding: a study across 4 Italian hospitals concluded that the perception that an advanced CCDSS could reduce HCPs autonomy was the most significant barrier to implementation [56].

Users’ attitudes and perceptions about the ease of use and usefulness have been at the center of established technology acceptance theories [57,58]. Trust in the sepsis digital alert and its uptake were found to be affected by individual factors, such as previous experience with the DS, as 2 recent systematic reviews on the implementation of a CCDSS [59] and of an EHRs DS have corroborated [60]. Although our participants did not directly discuss trust in the sepsis digital alert, its importance can be inferred from other aspects they raised, especially when thinking about better sepsis digital alerts. Participants would welcome more sophisticated and reliable sepsis digital alerts, which would be more accurate and transparent, as previous work observed [61,62]. At the same time, our study revealed that perceived usefulness in the sepsis digital alert depended on the HCPs’ professional experience and specialty training. More senior and emergency HCPs as well as intensivists tended to take less advantage from the sepsis digital alert. Similarly, previous work has demonstrated that HCPs can disregard the evidence underpinning CCDSSs for fear that their critical reasoning and, again, their professional autonomy are challenged [56]; but also, because the evidence embedded in the digital tool may be seen as jeopardizing hierarchical, power relations based on medical specialty and seniority [63].

Concomitantly, features and functionalities of the digital tool appear to influence the use of the sepsis digital alert [45,46]. A study highlighted that easy-to-use tablet applications as part of the DS for sepsis were important mediators facilitating implementation [62]. We also found that accessible, mobile phone apps and functions, such as the chat of the EHRs DS in site 1, appear to support better use of the sepsis digital alert. The usability of the digital tool has been a focal aspect in theories of ergonomics and human-technology interaction [64,65]. Work on the uptake of CCDSSs found that scarcity of available computers, unfriendly user interface, and excessive number of intrusive alerts lead to disengagement and fatigue [31,56,64,66]. Our study corroborates the importance of factors inherent to the design of the sepsis digital alert and the EHRs DS; both sites 1 and 3 made the deliberate choice of keeping minimal or reducing soft-stop digital alerts.

Previous work has shown the importance of functional teamworking; this should be based on high standards of coordination and communication to ensure the smooth journey of the septic patient and improve clinical outcomes [20,61]. Our results indicate the sepsis digital alert contributes to prompter patients' referral, escalation, and treatment. However, participants felt that lack in communication among staff could hamper the proper use of the sepsis digital alert, as some participants in site 2 raised. This resonates with the findings of the aforementioned study in the pediatric ED where professionals' acknowledgment of the sepsis digital alert was based on factors specific to the ED environment [46]. Significantly, another study found that a discontinuous flow of communication and teamwork among clinicians was a barrier to the integration of a machine learning sepsis early warning system in ED [62].

Organizational factors were identified as a significant obstacle to recognizing and responding to patients with sepsis in ED [21]. Significantly, our study corroborated how trust-level factors, such as good level of staffing, staff training and involvement in the sepsis digital alert' optimization, and appropriate technological resources, linked with more optimal use of these CCDSSs. In line with our results, other work concluded that organizational factors affected HCPs' trust in the sepsis digital alert; this connection was facilitated by engagement and education activities fostering sepsis digital alerts' understanding and acceptance [45,46,55,62]. The importance of the context in affecting individual HCPs' decisions and practices related to CCDSSs has been demonstrated in studies using the normalization process theory [67-69]. Accordingly, organizational and practice theories applied to the introduction of technology in the complex health care environment maintain that implementation processes are connected with the interaction between the technology, on the one hand, and HCPs' practices, teamwork relationships, organization's policies and priorities, on the other hand [70-72]. The uniqueness of the hospital context makes it difficult to compare the effectiveness of a sepsis digital alerts in isolation.

This study highlights that the introduction of sepsis digital alerts necessitates a multilevel approach [56,72] that includes understanding and actions at 4 sets of factors: the HCP, the hospital unit or department, the hospital, and the digital tool. Multilevel approaches to innovations, including digital ones as in this study, have been captured by several process frameworks developed in implementation science. Comprehensive innovation process frameworks have provided research logic models factoring in several determinants which we found in our work. These determinants spanned from the intervention characteristics, the inner and outer settings—such as leadership, networks, and communication—and learning climate, to the characteristics of individuals and the process [73-75]. Other tools also embedded a plethora of factors to be considered when assessing intervention scalability [76]; factors worth analyzing to assess scaling readiness include the strategic or political context, the intervention costs and benefits, delivery setting, and workforce [76]. Other work has concentrated on specific factors of the scale and spread journey, such as the types of innovations [77], specific context and processes, such as the NHS innovation pathways and its accelerators [78]; the importance of the context [79]; of patient and public involvement [80]; and of innovation intermediaries [81]. The necessity of adopting a multipronged, evolving strategy which goes beyond mechanistic logics of change have been promoted [82,83]. This same approach has emerged as mandatory from our retrospective, descriptive qualitative study based on the views and uses of sepsis digital alerts already in place in 3 hospital trusts in England.

Implications

An a priori analysis of the organizational environment to assess the hospital readiness and unique feasibility for the introduction of the sepsis digital alert is recommended. Mapping areas demanding change in the trust and planning resources and actions for their improvement mitigate the risk that sepsis or deterioration CCDSSs fail to offer the intended benefits [72]. In addition to more structural factors, such as resources, staffing, sepsis training, and a successful leadership-teamwork dynamic, it is also advisable that organizational cultural factors are factored in. Cultural factors that trusts should consider assessing include the trust's prioritization of sepsis or of deterioration and acuity, staff retention trends and sociodemographic profile, attitudes and readiness toward technological innovations, and the more impalpable norms regulating hierarchies and power among staff.

Hospital trusts should aim to plan ongoing implementation optimization initiatives in the pilot and rollout phases [55]. These initiatives should be flat-hierarchical and multidisciplinary so that HCPs with different job roles and training, based in different hospital units, can voice their unique perspective and support needs they expect to be met by the sepsis digital alert [84,85]. The trust should ensure the continuous involvement of information technology developers [86], and that staff feedback is appropriately collected and analyzed so that it materializes into context-based modifications of the tool. The engagement of staff should go hand-in-hand with education and training to aim at the maximization of behavior changes toward improved patient care. These activities should be reinforced by the establishment of champions and other strategic communication and educational campaigns, whereby staff can easily and remotely access information about the sepsis digital alert. It would be useful if trusts established dedicated advisory groups monitoring and managing all the actions necessary for a more successful post go-live [86].

Design, content, and technical aspects can act as barriers or facilitators to changing the clinical behavior toward better patient care in technology-based interventions [87]. Further multidisciplinary research should inform the development of sepsis digital alerts which are easy to use but at the same time are more sophisticated, able to target specific HCPs and hospital units, and simultaneously become more patient-specific, transparent, and interactive [88]. Sepsis digital alerts should also be more effectively linked to guidance on sepsis protocols, escalation practice, and antibiotics prescribing. Researchers and developers should work in conjunction with HCPs and policy makers to refine technology-based behavior change techniques that *effectively* support HCPs' decision-making, care practice, and improve patients' outcomes as a result.

Conclusions

Current sepsis digital alerts nested within the EHRs are introduced to support the identification and management of sepsis or deterioration and improve patient outcomes. These sepsis or deterioration CCDSSs fulfill their purpose but not entirely and not equally in all hospitals; an organic, multilevel framework to enhance tailored implementation of sepsis digital alerts is needed, along with the simultaneous validation of their effect on patient outcomes. No technological innovation in the health care setting can be a solo driver of change, and sepsis digital alerts are not magic wands able to dissipate issues that instead become important barriers to their optimal use in the rollout phase. Sepsis digital alerts implemented in trusts with good levels of staffing, resources, and functional teamwork are likely to be taken up more optimally and become good partners for HCPs. Equally, where the roll out of sepsis digital alerts is accompanied by multidisciplinary quality optimization initiatives, and by training, education, and other sepsis awareness actions that continually engage staff, HCPs are more likely to accept and embed the sepsis digital alert in their practice. Trust implementation of sepsis digital alert requires changes on multiple levels and at all phases of the intervention, starting from a prego-live analysis assessing and addressing organizational needs and readiness. Advancements toward minimally disruptive and smart sepsis digital alerts, which are more accurate and specific, but at the same time scalable and accessible, have to see policy changes and investments in multidisciplinary research agendas.

Acknowledgments

This study has received funding from the National Institute for Health and Care Research (NIHR), Programme Health and Social Care Delivery Research, Research Call: HSDR Digital Technologies to Improve Health and Care (award number NIHR129082). AJ Borek and STC have received support from the NIHR Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance (NIHR200915) at the University of Oxford in partnership with the UK Health Security Agency (UKHSA). The views expressed here are those of the authors and not necessarily those of the NIHR or the institutions of affiliations of the authors. The authors would like to acknowledge the work and contribution of the wider study team, the Digital Alerting for Sepsis team, as well as of the staff of the three NHS trust sites where the study was conducted. The important contribution of the late P Ghazal to this work is also acknowledged. P Ghazal was a champion of thinking creatively, and speaking to improve care for patients.

Authors' Contributions

CEC, KH, P Goodman, and STC contributed to conception and design. RL, AJ Borek, JW, AJ Brent, AK, RD, and STC contributed to data acquisition. RL, AJ Borek, KH, JW, AJ Brent, AK, GC, SP, AG, BG, P Ghazal, and STC contributed to the analysis and interpretation of data. RL drafted the article and all authors revised it critically for important intellectual content. All authors granted approval for the final version of the manuscript to be published.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview topic guide.
[DOCX File , 39 KB - humanfactors_v11i1e56949_app1.docx]

References

1. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). JAMA 2016 Feb 23;315(8):801-810 [FREE Full text] [doi: 10.1001/jama.2016.0287] [Medline: 26903338]
2. The sepsis manual 6th edition. United Kingdom Sepsis Trust. URL: https://sepsistrust.org/wp-content/uploads/2022/06/Sepsis-Manual-Sixth-Edition.pdf [accessed 2023-10-16]
3. Paoli CJ, Reynolds MA, Sinha M, Gitlin M, Crouser E. Epidemiology and costs of sepsis in the United States-an analysis based on timing of diagnosis and severity level. Crit Care Med 2018 Dec;46(12):1889-1897 [FREE Full text] [doi: 10.1097/CCM.0000000000003342] [Medline: 30048332]

4. Prescott HC. The epidemiology of sepsis. In: Wiersinga WJ, Seymour CW, editors. Handbook of Sepsis. Cham, Switzerland: Springer; 2018:15-28.
5. Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, et al. Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the Global Burden of Disease Study. Lancet 2020 Jan 18;395(10219):200-211 [FREE Full text] [doi: 10.1016/S0140-6736(19)32989-7] [Medline: 31954465]
6. Chiu C, Legrand M. Epidemiology of sepsis and septic shock. Curr Opin Anaesthesiol 2021 Apr 01;34(2):71-76. [doi: 10.1097/ACO.0000000000000958] [Medline: 33492864]
7. Reinhart K, Daniels R, Kissoon N, Machado FR, Schachter RD, Finfer S. Recognizing sepsis as a global health priority - a WHO resolution. N Engl J Med 2017 Aug 03;377(5):414-417. [doi: 10.1056/NEJMp1707170] [Medline: 28658587]
8. Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, et al. Surviving sepsis campaign: International Guidelines for Management of Sepsis and Septic Shock 2021. Crit Care Med 2021 Nov 01;49(11):e1063-e1143. [doi: 10.1097/CCM.0000000000005337] [Medline: 34605781]
9. Kumar A, Roberts D, Wood KE, Light B, Parrillo JE, Sharma S, et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock*. Critical Care Medicine 2006;34(6):1589-1596. [doi: 10.1097/01.ccm.0000217961.75225.e9]
10. Ferrer R, Martin-Loeches I, Phillips G, Osborn TM, Townsend S, Dellinger RP, et al. Empiric antibiotic treatment reduces mortality in severe sepsis and septic shock from the first hour: results from a guideline-based performance improvement program. Crit Care Med 2014 Aug;42(8):1749-1755. [doi: 10.1097/CCM.0000000000000330] [Medline: 24717459]
11. Seymour CW, Gesten F, Prescott HC, Friedrich ME, Iwashyna TJ, Phillips GS, et al. Time to treatment and mortality during mandated emergency care for sepsis. N Engl J Med 2017 Jun 08;376(23):2235-2244 [FREE Full text] [doi: 10.1056/NEJMoa1703058] [Medline: 28528569]
12. Sepsis: recognition, diagnosis and early management. National Institute for Health and Care Excellence. 2017. URL: https://www.nice.org.uk/guidance/NG51/chapter/Recommendations#managing-and-treating-suspected-sepsis-in-acute-hospital-settings [accessed 2023-10-16]
13. Suspected sepsis in people aged 16 or over who are not and have not recently been pregnant. National Institute for Health and Care Excellence. 2023. URL: https://www.nice.org.uk/guidance/ng51/chapter/Over-16s-not-pregnant-or-recently-pregnant-evaluating-risk-and-managing-suspected-sepsis [accessed 2024-03-16]
14. Schlapbach LJ, Zimmermann EA, Meylan S, Stocker M, Suter PM, Jakob SM, et al. Swiss Sepsis National Action Plan: a coordinated national action plan to stop sepsis-related preventable deaths and to improve the support of people affected by sepsis in Switzerland. Front Med (Lausanne) 2023;10:1114546 [FREE Full text] [doi: 10.3389/fmed.2023.1114546] [Medline: 36891186]
15. Schinkel M, Nanayakkara PW, Wiersinga WJ. Sepsis performance improvement programs: from evidence toward clinical implementation. Crit Care 2022 Mar 22;26(1):77 [FREE Full text] [doi: 10.1186/s13054-022-03917-1] [Medline: 35337358]
16. Guarino M, Perna B, Cesaro AE, Maritati M, Spampinato MD, Contini C, et al. 2023 update on sepsis and septic shock in adult patients: management in the emergency department. J Clin Med 2023 Apr 28;12(9):3188 [FREE Full text] [doi: 10.3390/jcm12093188] [Medline: 37176628]
17. Vincent JL, Singer M, Einav S, Moreno R, Wendon J, Teboul JL, et al. Equilibrating SSC guidelines with individualized care. Crit Care 2021 Nov 17;25(1):397 [FREE Full text] [doi: 10.1186/s13054-021-03813-0] [Medline: 34789298]
18. Rogan A, Lockett J, Peckler B, Robinson B, Raymond N. Exploring nursing and medical perceptions of sepsis management in a New Zealand emergency department: a qualitative study. Emerg Med Australas 2022 Jun;34(3):417-427. [doi: 10.1111/1742-6723.13911] [Medline: 34889063]
19. Olander A, Bremer A, Sundler AJ, Hagiwara MA, Andersson H. Assessment of patients with suspected sepsis in ambulance services: a qualitative interview study. BMC Emerg Med 2021 Apr 09;21(1):45 [FREE Full text] [doi: 10.1186/s12873-021-00440-4] [Medline: 33836665]
20. Matthaeus-Kraemer CT, Thomas-Rueddel DO, Schwarzkopf D, Rueddel H, Poidinger B, Reinhart K, et al. Crossing the handover chasm: clinicians' perceptions of barriers to the early detection and timely management of severe sepsis and septic shock. J Crit Care 2016 Dec;36:85-91. [doi: 10.1016/j.jcrc.2016.06.034] [Medline: 27546753]
21. Harley A, Johnston AN, Denny KJ, Keijzers G, Crilly J, Massey D. Emergency nurses' knowledge and understanding of their role in recognising and responding to patients with sepsis: a qualitative study. Int Emerg Nurs 2019 Mar;43:106-112. [doi: 10.1016/j.ienj.2019.01.005] [Medline: 30733005]
22. Roberts N, Hooper G, Lorencatto F, Storr W, Spivey M. Barriers and facilitators towards implementing the Sepsis Six care bundle (BLISS-1): a mixed methods investigation using the theoretical domains framework. Scand J Trauma Resusc Emerg Med 2017 Sep 19;25(1):96 [FREE Full text] [doi: 10.1186/s13049-017-0437-2] [Medline: 28927439]
23. Beesley SJ, Lanspa MJ. Why we need a new definition of sepsis. Ann Transl Med 2015 Nov;3(19):296 [FREE Full text] [doi: 10.3978/j.issn.2305-5839.2015.11.02] [Medline: 26697456]
24. Yoon J, Kym D, Hur J, Cho YS, Chun W, Yoon D. Validation of Sepsis-3 using survival analysis and clinical evaluation of quick SOFA, SIRS, and burn-specific SIRS for sepsis in burn patients with suspected infection. PLoS One 2023;18(1):e0276597 [FREE Full text] [doi: 10.1371/journal.pone.0276597] [Medline: 36595535]

25. National early warning score (NEWS): standardising the assessment of acute-illness severity in the NHS. Royal College of Physicians. 2012. URL: https://www.ombudsman.org.uk/sites/default/files/National%20Early%20Warning%20Score%20%28NEWS%29%20-%20Standardising%20the%20assessment%20of%20acute-illness%20severity%20in%20the%20NHS_0.pdf [accessed 2024-04-29]
26. Inada-Kim M. NEWS2 and improving outcomes from sepsis. Clin Med (Lond) 2022 Nov;22(6):514-517 [FREE Full text] [doi: 10.7861/clinmed.2022-0450] [Medline: 36427883]
27. National Early Warning Score (NEWS) 2. Royal College of Physicians. 2017. URL: https://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news-2 [accessed 2023-10-03]
28. Islam MM, Nasrin T, Walther BA, Wu CC, Yang HC, Li YC. Prediction of sepsis patients using machine learning approach: a meta-analysis. Comput Methods Programs Biomed 2019 Mar;170:1-9. [doi: 10.1016/j.cmpb.2018.12.027] [Medline: 30712598]
29. Adams R, Henry KE, Sridharan A, Soleimani H, Zhan A, Rawat N, et al. Prospective, multi-site study of patient outcomes after implementation of the TREWS machine learning-based early warning system for sepsis. Nat Med 2022 Jul 21;28(7):1455-1460. [doi: 10.1038/s41591-022-01894-0] [Medline: 35864252]
30. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020;3:17 [FREE Full text] [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
31. Backman R, Bayliss S, Moore D, Litchfield I. Clinical reminder alert fatigue in healthcare: a systematic literature review protocol using qualitative evidence. Syst Rev 2017 Dec 13;6(1):255-220 [FREE Full text] [doi: 10.1186/s13643-017-0627-z] [Medline: 29237488]
32. Powers EM, Shiffman RN, Melnick ER, Hickner A, Sharifi M. Efficacy and unintended consequences of hard-stop alerts in electronic health record systems: a systematic review. J Am Med Inform Assoc 2018 Nov 01;25(11):1556-1566 [FREE Full text] [doi: 10.1093/jamia/ocy112] [Medline: 30239810]
33. Ruppel H, Liu V. To catch a killer: electronic sepsis alert tools reaching a fever pitch? BMJ Qual Saf 2019 Sep;28(9):693-696 [FREE Full text] [doi: 10.1136/bmjqs-2019-009463] [Medline: 31015377]
34. Honeyford K, Cooke GS, Kinderlerer A, Williamson E, Gilchrist M, Holmes A, The Sepsis Big Room, et al. Evaluating a digital sepsis alert in a London multisite hospital network: a natural experiment using electronic health record data. J Am Med Inform Assoc 2020 Feb 01;27(2):274-283 [FREE Full text] [doi: 10.1093/jamia/ocz186] [Medline: 31743934]
35. Honeyford K, Timney A, Lazzarino R, Welch J, Brent AJ, Kinderlerer A, et al. Digital innovation in healthcare: quantifying the impact of digital sepsis screening tools on patient outcomes, a multisite natural experiment. SSRN Journal Preprint posted online February 16, 2024 [FREE Full text]
36. Narayanan N, Gross AK, Pintens M, Fee C, MacDougall C. Effect of an electronic medical record alert for severe sepsis among ED patients. Am J Emerg Med 2016 Feb;34(2):185-188. [doi: 10.1016/j.ajem.2015.10.005] [Medline: 26573784]
37. Shah T, Sterk E, Rech MA. Emergency department sepsis screening tool decreases time to antibiotics in patients with sepsis. Am J Emerg Med 2018 Oct;36(10):1745-1748. [doi: 10.1016/j.ajem.2018.01.060] [Medline: 29395762]
38. Manaktala S, Claypool SR. Evaluating the impact of a computerized surveillance algorithm and decision support system on sepsis mortality. J Am Med Inform Assoc 2017 Jan;24(1):88-95 [FREE Full text] [doi: 10.1093/jamia/ocw056] [Medline: 27225197]
39. McRee L, Thanavaro JL, Moore K, Goldsmith M, Pasvogel A. The impact of an electronic medical record surveillance program on outcomes for patients with sepsis. Heart Lung 2014 Nov;43(6):546-549. [doi: 10.1016/j.hrtlng.2014.05.009] [Medline: 24961633]
40. Cull J, Brevetta R, Gerac J, Kothari S, Blackhurst D. Epic sepsis model inpatient predictive analytic tool: a validation study. Crit Care Explor 2023 Jul;5(7):e0941 [FREE Full text] [doi: 10.1097/CCE.0000000000000941] [Medline: 37405252]
41. Austrian JS, Jamin CT, Doty GR, Blecker S. Impact of an emergency department electronic sepsis surveillance system on patient mortality and length of stay. J Am Med Inform Assoc 2018 May 01;25(5):523-529 [FREE Full text] [doi: 10.1093/jamia/ocx072] [Medline: 29025165]
42. Joshi M, Ashrafian H, Arora S, Khan S, Cooke G, Darzi A. Digital alerting and outcomes in patients with sepsis: systematic review and meta-analysis. J Med Internet Res 2019 Dec 20;21(12):e15166 [FREE Full text] [doi: 10.2196/15166] [Medline: 31859672]
43. Downing NL, Rolnick J, Poole SF, Hall E, Wessels AJ, Heidenreich P, et al. Electronic health record-based clinical decision support alert for severe sepsis: a randomised evaluation. BMJ Qual Saf 2019 Sep;28(9):762-768 [FREE Full text] [doi: 10.1136/bmjqs-2018-008765] [Medline: 30872387]
44. Hwang MI, Bond WF, Powell ES. Sepsis alerts in emergency departments: a systematic review of accuracy and quality measure impact. West J Emerg Med 2020 Aug 24;21(5):1201-1210 [FREE Full text] [doi: 10.5811/westjem.2020.5.46010] [Medline: 32970576]
45. Silvestri JA, Kmiec TE, Bishop NS, Regli SH, Weissman GE. Desired characteristics of a clinical decision support system for early sepsis recognition: interview study among hospital-based clinicians. JMIR Hum Factors 2022 Oct 21;9(4):e36976 [FREE Full text] [doi: 10.2196/36976] [Medline: 36269653]

46. Souganidis ES, Patel B, Sampayo EM. Physician-specific utilization of an electronic best practice alert for pediatric sepsis in the emergency department. Pediatr Emer Care 2022 Jun 13;38(8):e1417-e1422. [doi: 10.1097/pec.0000000000002778]
47. Guirgis FW, Jones L, Esma R, Weiss A, McCurdy K, Ferreira J, et al. Managing sepsis: electronic recognition, rapid response teams, and standardized care save lives. J Crit Care 2017 Aug;40:296-302 [FREE Full text] [doi: 10.1016/j.jcrc.2017.04.005] [Medline: 28412015]
48. Westra BL, Landman S, Yadav P, Steinbach M. Secondary analysis of an electronic surveillance system combined with multi-focal interventions for early detection of sepsis. Appl Clin Inform 2017 Jan 18;8(1):47-66 [FREE Full text] [doi: 10.4338/ACI-2016-07-RA-0112] [Medline: 28097288]
49. Fetters MD, Rubinstein EB. The 3 Cs of content, context, and concepts: a practical approach to recording unstructured field observations. Ann Fam Med 2019 Nov;17(6):554-560 [FREE Full text] [doi: 10.1370/afm.2453] [Medline: 31712294]
50. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
51. Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. Int J Qual Methods 2016 Nov 29;5(1):80-92. [doi: 10.1177/160940690600500107]
52. Glaser BG, Strauss AL. Discovery of Grounded Theory: Strategies for Qualitative Research. New Brunswick, QC: Aldine Transaction; 2000.
53. Tracy SJ. Qualitative quality: eight “big-tent” criteria for excellent qualitative research. Qual Inq 2010 Oct 01;16(10):837-851. [doi: 10.1177/1077800410383121]
54. Henry KE, Kornfield R, Sridharan A, Linton R, Groh C, Wang T, et al. Human-machine teaming is key to AI adoption: clinicians' experiences with a deployed machine learning system. NPJ Digit Med 2022 Jul 21;5(1):97 [FREE Full text] [doi: 10.1038/s41746-022-00597-7] [Medline: 35864312]
55. Joshi M, Mecklai K, Rozenblum R, Samal L. Implementation approaches and barriers for rule-based and machine learning-based sepsis risk prediction tools: a qualitative study. JAMIA Open 2022 Jul;5(2):ooac022 [FREE Full text] [doi: 10.1093/jamiaopen/ooac022] [Medline: 35474719]
56. Liberati EG, Ruggiero F, Galuppo L, Gorli M, González-Lorenzo M, Maraldi M, et al. What hinders the uptake of computerized decision support systems in hospitals? A qualitative study and framework for implementation. Implement Sci 2017 Sep 15;12(1):113 [FREE Full text] [doi: 10.1186/s13012-017-0644-2] [Medline: 28915822]
57. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
58. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manage Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
59. Tokgöz P, Hafner J, Dockweiler C. Factors influencing the implementation of decision support systems for antibiotic prescription in hospitals: a systematic review. BMC Med Inform Decis Mak 2023 Feb 06;23(1):27 [FREE Full text] [doi: 10.1186/s12911-023-02124-4] [Medline: 36747193]
60. Sung M, He J, Zhou Q, Chen Y, Ji JS, Chen H, et al. Using an integrated framework to investigate the facilitators and barriers of health information technology implementation in noncommunicable disease management: systematic review. J Med Internet Res 2022 Jul 20;24(7):e37338 [FREE Full text] [doi: 10.2196/37338] [Medline: 35857364]
61. Downey CL, Tahir W, Randell R, Brown JM, Jayne DG. Strengths and limitations of early warning scores: a systematic review and narrative synthesis. Int J Nurs Stud 2017 Nov;76:106-119 [FREE Full text] [doi: 10.1016/j.ijnurstu.2017.09.003] [Medline: 28950188]
62. Sandhu S, Lin AL, Brajer N, Sperling J, Ratliff W, Bedoya AD, et al. Integrating a machine learning system into clinical workflows: qualitative study. J Med Internet Res 2020 Nov 19;22(11):e22421 [FREE Full text] [doi: 10.2196/22421] [Medline: 33211015]
63. Lipman T. Power and influence in clinical effectiveness and evidence-based medicine. Fam Pract 2000 Dec;17(6):557-563. [doi: 10.1093/fampra/17.6.557] [Medline: 11120731]
64. Paton C, Kushniruk AW, Borycki EM, English M, Warren J. Improving the usability and safety of digital health systems: the role of predictive human-computer interaction modeling. J Med Internet Res 2021 May 27;23(5):e25281 [FREE Full text] [doi: 10.2196/25281] [Medline: 34042590]
65. Pertiwi AA, Fraczkowski D, Stogis SL, Lopez KD. Using heuristic evaluation to improve sepsis alert usability. Crit Care Nurs Clin North Am 2018 Jun;30(2):297-309. [doi: 10.1016/j.cnc.2018.02.011] [Medline: 29724447]
66. Kesselheim AS, Cresswell K, Phansalkar S, Bates DW, Sheikh A. Clinical decision support systems could be modified to reduce 'alert fatigue' while still minimizing the risk of litigation. Health Aff (Millwood) 2011 Dec;30(12):2310-2317. [doi: 10.1377/hlthaff.2010.1111] [Medline: 22147858]
67. Murray E, Burns J, May C, Finch T, O'Donnell C, Wallace P, et al. Why is it difficult to implement e-health initiatives? A qualitative study. Implement Sci 2011 Jan 19;6(1):6 [FREE Full text] [doi: 10.1186/1748-5908-6-6] [Medline: 21244714]
68. Pope C, Halford S, Turnbull J, Prichard J, Calestani M, May C. Using computer decision support systems in NHS emergency and urgent care: ethnographic study using normalisation process theory. BMC Health Serv Res 2013 Mar 23;13:111 [FREE Full text] [doi: 10.1186/1472-6963-13-111] [Medline: 23522021]

69. Jeffries M, Salema NE, Laing L, Shamsuddin A, Sheikh A, Avery A, et al. The implementation, use and sustainability of a clinical decision support system for medication optimisation in primary care: a qualitative evaluation. PLoS One 2021;16(5):e0250946 [FREE Full text] [doi: 10.1371/journal.pone.0250946] [Medline: 33939750]
70. Nicolini D. The work to make telemedicine work: a social and articulative view. Soc Sci Med 2006 Jun;62(11):2754-2767. [doi: 10.1016/j.socscimed.2005.11.001] [Medline: 16343724]
71. Timmermans S, Berg M. The practice of medical technology. Sociol Health Illn 2003 Mar 14;25(3):97-114 [FREE Full text] [doi: 10.1111/1467-9566.00342] [Medline: 14498932]
72. Cresswell KM, Bates DW, Sheikh A. Ten key considerations for the successful implementation and adoption of large-scale health information technology. J Am Med Inform Assoc 2013 Jun;20(e1):e9-13 [FREE Full text] [doi: 10.1136/amiajnl-2013-001684] [Medline: 23599226]
73. Smith JD, Li DH, Rafferty MR. The Implementation Research Logic Model: a method for planning, executing, reporting, and synthesizing implementation projects. Implement Sci 2020 Sep 25;15(1):84 [FREE Full text] [doi: 10.1186/s13012-020-01041-8] [Medline: 32988389]
74. Damschroder LJ, Lowery JC. Evaluation of a large-scale weight management program using the consolidated framework for implementation research (CFIR). Implement Sci 2013 May 10;8:51 [FREE Full text] [doi: 10.1186/1748-5908-8-51] [Medline: 23663819]
75. Oliveira T, Barrenho E, Vernet A, Autio E, Barlow J. Developing a global healthcare innovation index. Policy Innovation Research Unit (PIRU). 2017. URL: https://piru.ac.uk/assets/files/Healthcare%20innovation%20index%20report.pdf [accessed 2024-04-29]
76. Milat A, Lee K, Conte K, Grunseit A, Wolfenden L, van Nassau F, et al. Intervention Scalability Assessment Tool: a decision support tool for health policy makers and implementers. Health Res Policy Syst 2020 Jan 03;18(1):1 [FREE Full text] [doi: 10.1186/s12961-019-0494-2] [Medline: 31900230]
77. Flessa S, Huebner C. Innovations in health care-a conceptual framework. Int J Environ Res Public Health 2021 Sep 24;18(19):10026 [FREE Full text] [doi: 10.3390/ijerph181910026] [Medline: 34639328]
78. Arora A, Wright A, Cheng TK, Khwaja Z, Seah M. Innovation pathways in the NHS: an introductory review. Ther Innov Regul Sci 2021 Sep 19;55(5):1045-1058 [FREE Full text] [doi: 10.1007/s43441-021-00304-w] [Medline: 34009551]
79. Fulop N, Robert G. Context for successful quality improvement. The Health Foundation. 2015. URL: https://www.health.org.uk/publications/context-for-successful-quality-improvement [accessed 2024-04-29]
80. Cluley V, Ziemann A, Feeley C, Olander EK, Shamah S, Stavropoulou C. Mapping the role of patient and public involvement during the different stages of healthcare innovation: a scoping review. Health Expect 2022 Jun 17;25(3):840-855 [FREE Full text] [doi: 10.1111/hex.13437] [Medline: 35174585]
81. Scarbrough H, Sanfilippo KR, Ziemann A, Stavropoulou C. Mobilizing pilot-based evidence for the spread and sustainability of innovations in healthcare: the role of innovation intermediaries. Soc Sci Med 2024 Jan;340:116394 [FREE Full text] [doi: 10.1016/j.socscimed.2023.116394] [Medline: 38000177]
82. Hemmings N, Hutchings R, Castle-Clarke S, Palmer DW. Achieving Scale and Spread: Learning for Innovators and Policy-Makers. London, UK: Nuffield Trust; 2020.
83. Greenhalgh T, Papoutsi C. Spreading and scaling up innovation and improvement. BMJ 2019 May 10;365:l2068 [FREE Full text] [doi: 10.1136/bmj.l2068] [Medline: 31076440]
84. Introducing a sepsis power plan - case study: the Imperial College Trust Sepsis Big Room. Imperial College Trust, Flow Coaching Academy, and The Health Foundation. 2019. URL: https://flowcoaching.academy/UserFiles/File/Case_Studies/CS_Sepsis.pdf [accessed 2024-01-09]
85. Improving the awareness and response to sepsis. a case study from Sheffield’s Sepsis Big Room. Flow Coaching Academy. 2024 Mar 12. URL: https://flowcoaching.academy/resources/improving-the-awareness-and-response-to-sepsis-a-case-study-from-sheffields-sepsis-big-room/ [accessed 2024-08-27]
86. Moon MC, Hills R, Demiris G. Understanding optimization processes of electronic health records (EHR) in select leading hospitals: a qualitative study. J Innov Health Inform 2018 Jun 30;25(2):109-125 [FREE Full text] [doi: 10.14236/jhi.v25i2.1011] [Medline: 30398450]
87. Keyworth C, Hart J, Armitage CJ, Tully MP. What maximizes the effectiveness and implementation of technology-based interventions to support healthcare professional practice? A systematic literature review. BMC Med Inform Decis Mak 2018 Nov 07;18(1):93 [FREE Full text] [doi: 10.1186/s12911-018-0661-3] [Medline: 30404638]
88. Joshi M, Ashrafian H, Arora S, Sharabiani M, McAndrew K, Khan SN, et al. A pilot study to investigate real-time digital alerting from wearable sensors in surgical patients. Pilot Feasibility Stud 2022 Jul 06;8(1):140 [FREE Full text] [doi: 10.1186/s40814-022-01084-2] [Medline: 35794669]

Abbreviations

CCDSS: computerized clinical decision support system
DiAlS: Digital Alerting for Sepsis
DS: digital system

ED: emergency department
EHR: electronic health record
HCP: health care professional
ICU: intensive care unit
NEWS: national early warning score
NHS: National Health System

Edited by A Kushniruk; submitted 31.01.24; peer-reviewed by J Silvestri, J Zhao; comments to author 12.03.24; revised version received 26.03.24; accepted 11.07.24; published 15.10.24.

Please cite as:
Lazzarino R, Borek AJ, Honeyford K, Welch J, Brent AJ, Kinderlerer A, Cooke G, Patil S, Gordon A, Glampson B, Goodman P, Ghazal P, Daniels R, Costelloe CE, Tonkin-Crine S
Views and Uses of Sepsis Digital Alerts in National Health Service Trusts in England: Qualitative Study With Health Care Professionals
JMIR Hum Factors 2024;11:e56949
URL: https://humanfactors.jmir.org/2024/1/e56949
doi:10.2196/56949
PMID:

Original Paper

Exploring the Needs of People With Chronic Low Back Pain and Health Care Professionals for mHealth Devices to Support Self-Managed Physical Activity and Pain: User-Centered Design Approach

Mathilde Berger[1,2], OT, MSc; Anne Deblock-Bellamy[1], OT, PhD; Laurence Chèze[2], PhD; Thomas Robert[2], PhD; Julie J Desrosiers[1], OT, PhD; Guillaume Christe[3], PT, PhD; Anne Martine Bertrand[1], OT, PhD

[1]Department of Occupational Therapy, University of Applied Sciences and Arts Western Switzerland (HETSL | HES-SO), Lausanne, Switzerland

[2]LBMC, Claude Bernard Lyon 1 University, Gustave Eiffel University, Bron, France

[3]Department of Physiotherapy, School of Health Sciences (HESAV), University of Applied Sciences and Arts of Western Switzerland (HES-SO), Lausanne, Switzerland

Corresponding Author:
Mathilde Berger, OT, MSc
Department of Occupational Therapy
University of Applied Sciences and Arts Western Switzerland (HETSL | HES-SO)
Chemin des Abeilles 14
Lausanne, 1010
Switzerland
Phone: 41 (0)216510330
Email: mathilde.berger@hetsl.ch

Abstract

Background: Chronic low back pain (CLBP) is a major economic and social problem worldwide. Despite the variety of recommended treatments, long-term self-management of this condition is complex and requires the development of innovative interventions. Mobile health (mHealth) technologies hold great promise for the management of chronic pain, particularly to support physical activity. However, their implementation is challenged by a lack of user compliance and limited engagement, which may be due to insufficient consideration of the needs of potential users during development.

Objective: This study aims to explore the needs of people with CLBP and health care professionals regarding mHealth technologies to support self-managed physical activity, and to delineate design recommendations based on identified needs.

Methods: A participatory study was conducted using a 3-phase, user-centered design approach: needs investigation with a group of experts in a workshop (phase 1), needs exploration with end users in focus groups (phase 2), and validation of needs using Delphi questionnaires followed by the development of a set of recommendations (phase 3).

Results: A total of 121 people with CLBP, expert patients, health care professionals, rehabilitation researchers, and biomechanical engineers participated in this study. The results indicated how technology could help people with CLBP overcome their difficulties with managing physical activity. Specific needs were formulated concerning device objectives, expected strategies, functionalities, technical features, conditions of use, and potential facilitators and barriers to use. These needs were validated by consensus from the potential end users and translated into design recommendations.

Conclusions: This study provides design recommendations for the development of an mHealth device specifically adapted for people with CLBP.

(JMIR Hum Factors 2024;11:e59897) doi:10.2196/59897

KEYWORDS

chronic low back pain; needs; self-management; physical activity; mobile health; mHealth; user-centered design

Introduction

Background

Chronic low back pain (CLBP) is a major public health concern worldwide because of its economic and social consequences [1,2]. It is the leading cause of disability and work absenteeism, and its prevalence is still increasing [3,4]. Management of daily symptoms and activity are major challenges for people with this condition [5-7].

Self-management strategies and maintaining an active lifestyle have been consistently recommended for people with CLBP [8-12]. Self-management can be defined as the ongoing and dynamic ability to handle symptoms, such as pain, physical and psychological consequences, and lifestyle adjustments [13,14]. Self-management programs include educational and psychosocial interventions; maintaining an active lifestyle is a core component of these programs. People with CLBP are encouraged to engage in regular physical activity and to adopt healthy behaviors to manage their symptoms [14]. Although self-management and physical activity–based interventions improve pain [15] and reduce disability [16,17] in the long term, as well as promote the development of self-management skills [18,19], the effect of such interventions are small to moderate. Moreover, it is difficult to support self-management in clinical practice because of constraints in time, service organization, and follow-up [20]. Therefore, there is a need to identify innovative interventions to promote sustainable self-managed physical activity in people with CLBP.

Recent advances in technology like mobile health (mHealth) may offer new opportunities to support self-managed physical activity and pain in people with CLBP. mHealth is defined as a "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices" [21]. It can help people change their health behaviors [22] and help people with CLBP pursue specific physical activity goals while receiving continuous feedback on their physical performance [23,24]. It could be used to help them manage their condition and maintain an active lifestyle [25,26]. From the point of view of health care professionals, mHealth technology can provide individualized interventions with real-time feedback, support their coaching role [27], and support the development of behavior change [28,29].

Although mHealth technologies are promising, their content and the context in which they are offered need to be explored further. Many mobile apps are available to support people with CLBP in self-management, mainly providing exercise recommendations or information about the mechanisms of CLBP. These apps score poorly on the Mobile Application Rating Scale (a scale to assess app quality), and do not mention the theoretical approaches on which they are based [30-32]. Sometimes these mobile apps are combined with a physical activity tracker; however, they have shown limited effectiveness in disability and pain management in people with CLBP [33,34]. For example, several studies using the Fitbit device found nonsignificant results for pain [35,36] and disability [37]. This lack of an effect could relate to the fact that the target population for the tool is healthy individuals. Devices, such as the Fitbit were developed primarily to help young people improve their physical condition and they do not specifically consider people with CLBP and their context. They do not appear to be suitable for supporting health care professionals or people with CLBP who may benefit from increasing their participation in unstructured physical activity, like walking, and reducing sedentary lifestyle habits [38-41]. Moreover, studies of rehabilitation programs that integrated mobile technology found mixed results because of participants' lack of adherence to the wearable technology [35,36].

It is now widely recognized that user-centered design (UCD) approaches are needed to facilitate the development, acceptability, and implementation of mHealth technologies [42-45]. In the field of CLBP, such designs are seldom used and rarely rigorously applied [46], particularly in the initial stages of device design [47].

This Study

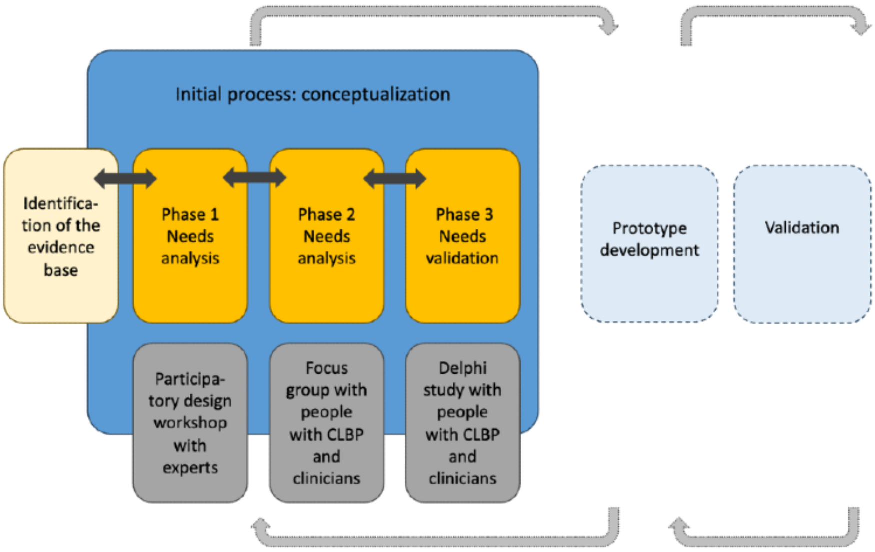
The main aim of this study was to explore the needs of people with CLBP and health care professionals regarding the use of mHealth technologies to support self-managed physical activity and pain. The specific aims were to (1) investigate how technology could help people with CLBP overcome the difficulties of managing their own physical activity with an expert group, (2) explore the needs, experiences, and preferences of people with CLBP and health care professionals regarding technology to support self-managed physical activity, and (3) validate the identified needs and develop a set of mHealth technology design recommendations.

Methods

Design

A participatory study was conducted using a UCD approach. UCD is an iterative process that focuses on users and their contexts in all stages of development design [44,48,49]. This study focused only on the initial process of the approach to investigate the needs of people with CLBP and to guide future prototype designs (Figure 1) [50-52]. The study was conducted in France and Switzerland and involved three successive data collection phases: (1) needs investigation with a group of experts, (2) needs exploration with people with CLBP and health care professionals, and (3) validation of the needs identified and development of recommendations. It was conducted after a preliminary literature review that aimed to identify existing evidence on mHealth devices used to support self-managed physical activity in people with CLBP.

Figure 1. Illustration of the 3 phases of data collection based on the initial process of a user-centered approach. CLBP: chronic low back pain.



Ethical Considerations

The study was approved by the research ethics committee of the University of Lyon (CER-UDL 2022-10-13-003). The Cantonal Ethics Committee of Swissethics stipulated that this research was outside their jurisdiction and did not require ethics approval (Req-2022-00733).

Data Collection

Phase 1: Needs Investigation With an Expert Group

Overview

The first phase investigated how technologies could support people with CLBP in overcoming difficulties with physical activity and pain self-management. A web-based participatory design workshop involving an expert group was set up for this purpose [50].

Sample

The expert group consisted of patients who were experts, clinicians with expertise in low back pain management, rehabilitation researchers, and biomechanical engineers in the rehabilitation field. Inclusion criteria for patients who were experts in self-management of their disease acquired over several years, and experience with several treatment modalities, including participation in at least 1 multidisciplinary rehabilitation program. Clinicians, rehabilitation researchers, and biomechanical engineers were required to have at least 5 years of experience in their respective fields.

Recruitment

The expert group was recruited through the investigators' clinical networks. Potential experts were contacted via email and telephone by the research team who explained the workshop to them, invited them to participate, and obtained their consent.

Data Collection

Data were collected during a 3-hour web-based participatory design workshop via teleconference in October 2022, moderated by 2 members of the research team (AD-B and MB). The workshop was organized into a preliminary activity followed by 3 phases. Before the workshop began, each participant received a summary of the existing evidence on mHealth devices used to support self-managed physical activity and pain in people with CLBP. In the first part of the workshop, the demographic characteristics of the participants were collected using a short questionnaire. Each of the 4 expert subgroups (patients who were experts, clinicians, rehabilitation researchers, and biomechanical engineers) was asked to reflect on the difficulties and factors limiting self-managed physical activity and pain in people with CLBP. In the second part, 2 subgroups were formed with different expert representatives (each subgroup consisted of at least 1 patient who was an expert, 1 clinician, 1 rehabilitation researcher, and 1 biomechanical engineer) to encourage discussion and the generation of several ideas on how technology could help people with CLBP overcome the difficulties identified. Each subgroup explored potential solutions to address the difficulties identified for a described persona. All exchanges with the group were recorded and transcribed.

Phase 2: Needs Exploration With End Users

Overview

The second phase explored the needs, experiences, and preferences of people with CLBP and health care professionals for using technology to support self-managed physical activity and pain. Two series of focus groups were run for this purpose [51]. The first explored end-user experiences and needs relating to physical activity and pain self-management. The second focused on potential technological solutions to meet the identified needs.

Sample

A sample of people with CLBP and a sample of health care professionals were recruited. Inclusion criteria for people with CLBP were a clinical diagnosis of CLBP and participation in ongoing rehabilitation within a multidisciplinary program.

Inclusion criteria for health care professionals were more than 2 years of experience caring for people with CLBP and participation in a multidisciplinary rehabilitation program for CLBP.

Recruitment

Recruitment was done in collaboration with 2 rehabilitation centers, one in Switzerland and the other in France. Health care professionals involved in the care of people with CLBP at both rehabilitation centers were informed of the study via an email from the research team with information about the study attached. People with CLBP were informed about the study by their health care professionals and received written information about the study. All people with CLBP and health care professionals willing to participate were contacted by the research team who provided further information and collected signed informed consent for participation. All participants included in phase 2 were different from those in phase 1 to explore diverse perspectives.

Data Collection

The 2 sets of focus groups were conducted separately for people with CLBP and health care professionals at the rehabilitation centers in Switzerland and France, with 4 to 8 participants as recommended by Kitzinger et al [51]. Each participant attended 2 focus groups held 2 weeks apart between January and February 2023. In the first series of focus groups, participants shared their experiences and needs regarding self-managed physical activity and pain in relation to the issues outlined by the expert group. In the second series of focus groups, participants discussed the potential technological solutions to meet the identified needs, in detail. A total of 8 focus groups were conducted by the research team. Participants were asked about their past experiences with health care technologies and their potential future needs for using technology. The potential facilitators and barriers to using technology to support self-managed physical activity and pain were also discussed. Two members of the research team moderated all the focus groups using a standardized interview guide developed by the research team (MB, JJD, or AD-B). An interview guide was developed for people with CLBP and health care professionals, asking the same questions but rephrased specifically for each group. Each focus group lasted approximately 1 hour 30 minutes and was recorded and transcribed. Participant's demographic characteristics were collected using a short questionnaire before the focus groups, and all data were coded to ensure anonymity.

Phase 3: Needs Validation

Overview

The third phase involved validating the identified needs and delineating a set of design recommendations for mHealth technologies. For this, a two-round Delphi questionnaire approach was used [52]. The first round assessed the needs expressed by participants during the previous consultation phases, and the second round consisted of evaluating the responses that did not reach a consensus based on the suggestions from the first round.

Sample

People with CLBP and health care professionals were selected. The inclusion criterion for people with CLBP was having already completed or in the process of completing a multidisciplinary rehabilitation program. The inclusion criterion for health care professionals was more than 2 years' experience caring for people with CLBP.

Recruitment

Participants were recruited through poster advertising, social media, and word of mouth, in Switzerland and France. Participants from phase 2 were invited to join phase 3, allowing them to evaluate all the needs identified in the earlier phases, not just those discussed in their focus groups. All recruitment materials directed interested individuals to a web-based survey (LimeSurvey) to determine their eligibility, register their characteristics, and provide informed consent.

Data Collection

The Delphi questionnaire was developed from all the themes identified from the results of the 2 previous phases of the study. Two researchers (MB and AD-B) selected each statement for the questionnaire. The research team then validated the statements according to the elements discussed during the previous 2 study phases. Before completing this questionnaire, the research team provided participants with a standardized summary of the methodology and the preliminary results of the first 2 phases. All participants completed a short sociodemographic questionnaire. Each participant was asked to rate their level of agreement with each statement on a 4-point bidirectional scale: strongly disagree (1), disagree (2), agree (3), or strongly agree (4). If participants "disagreed" or "strongly disagreed," they were invited to propose a new suggestion corresponding to their own opinion in open text boxes. The first round of questionnaires remained open for 4 weeks. Consensus was considered achieved if 90% of the participants agreed or strongly agreed with a statement. The second-round questionnaire was based on suggestions for statements for which there was no consensus in the first round. The second round was open for 12 weeks and only people who had already responded to the first round were invited to respond to the second round. This third phase lasted 16 weeks, from June to September 2023.

Data Analysis

Descriptive statistics were used to characterize the participants in the 3 phases, using frequencies and percentages. In phase 1, participant discussions were recorded and transcribed verbatim for analysis using an inductive thematic analysis [53,54]. First, 2 researchers (MB and AD-B) independently read the transcripts to familiarize themselves with the content. Second, they separately coded the data using NVivo (version 20, Lumivero) to create an initial codebook. After the initial coding, the codebook was discussed between the 2 researchers, and segments of the content with similar meaning were assigned to the same code. Third, the coded transcripts were used to refine the concepts of the initial codebook and combine the codes into key themes. When new codes or themes emerged, the codebook was revised, and the previous transcripts were recoded. Any discrepancies between the researchers were resolved by discussion, and any necessary adjustments were made. This thematic analysis was then conducted again to reach a consensus. Finally, the content of the themes and subthemes of the coding scheme was discussed with the entire research group.

In phase 2, the data analysis procedure from phase 1 was repeated using recorded material from all the focus groups [53,54] and the field notes written by the moderator in charge of describing the group interactions. The final version of the codebook from phase 1 was reused and revised by 2 researchers for use in phase 2 (MB and AD-B). All themes were validated by the entire research group.

In phase 3, the level of expert consensus was calculated for each item after each round [55]. Consensus was considered achieved when 90% of the participants agreed or strongly agreed with a statement. All suggestions were reviewed by 2 researchers (MB and AD-B), and the data were coded using the same codebook used for phases 1 and 2.

Results

Overview

A total of 121 participants were involved in the study. Of the 121 participants, 9 experts took part in the first phase, including 4 subgroups of 2 to 3 experts in the first part of the workshop and 2 subgroups of 4 to 5 experts in the second part of the workshop (one of the subgroups was composed of 1 patient who was an expert, 1 clinician, 1 rehabilitation researcher, and 1 biomechanical engineer while the other was composed of 1 patient who was an expert, 2 clinicians, 1 rehabilitation researcher, and 1 biomechanical engineer). Of 26 participants who took part in the second phase, 11 people with CLBP and 15 health care professionals (including 4 and 7 participants, respectively, for the 2 sets of focus groups with people with CLBP, and 7 and 8 participants, respectively, for the 2 sets of focus groups with health care professionals). Of 86 participants who took part in the third phase, 45 people with CLBP and 41 health care professionals. About 63% (7/11) of the participants with CLBP and 46% (7/15) of health care professionals from phase 2, participated in phase 3. The demographic characteristics of the participants are presented in Table 1 and more details about the participants are available in Multimedia Appendix 1.

More than 12 hours of participatory design workshop data, focus group data, and suggestions from 2 questionnaires were coded into 5 themes that are described in more detail below: (1) difficulties experienced by people with CLBP in relation to self-managed physical activity and pain, (2) device concept (aim and strategies), (3) device content, (4) condition of use, and (5) facilitators and barriers to use.

Table 1. Characteristics of the participants in each phase of the study.

Participant characteristics	Phase 1 (n=9), n (%)	Phase 2 (n=26), n (%)	Phase 3 round 1 (n=86), n (%)	Phase 3 round 2 (n=61), n (%)
Gender				
Women	6 (67)	16 (62)	46 (54)	36 (59)
Men	3 (33)	10 (38)	39 (45)	24 (39)
Other	0 (0)	0 (0)	1 (1)	1 (2)
Age (y)				
20-29	0 (0)	1 (4)	8 (9)	3 (5)
30-39	4 (44)	6 (21)	20 (23)	14 (23)
40-49	0 (0)	9 (35)	17 (20)	12 (20)
50-59	5 (56)	8 (31)	30 (35)	22 (36)
≥60	0 (0)	2 (8)	11 (13)	10 (16)
Role				
People with CLBP[a]	2 (22)	11 (42)	45 (52)	33 (54)
Health care professional	3 (33)	15 (58)	41 (48)	28 (46)
Rehabilitation researcher	2 (22)	—[b]	—	—
Research engineer	2 (22)	—	—	—
Location				
Switzerland	5 (56)	11 (42)	63 (73)	42 (69)
France	4 (44)	15 (58)	23 (27)	19 (31)

[a]CLBP: chronic low back pain.

[b]Not applicable.

Difficulties Encountered by People With CLBP in Relation to Self-Managed Physical Activity

The difficulties faced by people with CLBP in relation to self-managed physical activity and pain were specifically explored in phase 1 by the expert group composed of patients who were experts, clinicians, rehabilitation researchers, and biomechanical engineers. All participants were asked to discuss these difficulties to reach a consensus on the problems faced by people with CLBP. These difficulties were then iteratively addressed by people with CLBP and health care professionals in phases 2 and 3. These difficulties are presented in Table 2. The more prevalent ones included pain, personal beliefs, difficulty pacing physical activity, and a lack of sustained, individualized, long-term support from health care professionals. Only these prevalent difficulties were used as the basis for the assessment of the needs of people with CLBP and health care professionals, as they were shared by both groups.

Table 2. Difficulties and factors limiting self-managed physical activity and pain in people with chronic low back pain (CLBP) identified by the different groups of participants in the different phases.

Difficulties reported	Phase 1				Phase 2		Phase 3	
	eP[a]	eHp[b]	Rr[c]	Re[d]	People with CLBP	Hp[e]	People with CLBP	Hp
Chronic pain	✓[f]	✓	✓	✓	✓	✓		
Personal beliefs: fear of pain and underlying injury, kinesiophobia, difficulty understanding that the activity can be beneficial, loss of purpose, etc	✓	✓	✓	✓	✓	✓	✓	
Difficulty pacing physical activity	✓	✓	✓		✓	✓		
Sedentary lifestyle and lack of time	✓	✓	✓		✓	✓	✓	
Lack of long-term support	✓	✓			✓	✓	✓	
Lack of individualized support	✓		✓		✓		✓	
Poorly coordinated health care pathway		✓	✓			✓		
Lack of motivation				✓	✓	✓		
Negative experiences with physical activity	✓		✓		✓			
Difficulty making long-term plans	✓				✓	✓		
Beliefs of those around you and of society in general		✓	✓					
Depression, loss of confidence in one's own abilities		✓		✓				
Neurophysiological damage		✓						
Socioprofessional and safety problems		✓						
Medication side effects		✓						
Body image disorders		✓						
Feeling overwhelmed		✓				✓		
Fatigue, sleep disorders		✓						
Eating disorders		✓						
Laziness		✓						
Frustration with one's own abilities					✓			
Forgetting the principles of activity management						✓		

[a]eP: patient who was an expert.

[b]eHp: expert health care professional.

[c]Rr: rehabilitation researcher.

[d]Re: research engineer.

[e]Hp: health care professional.

[f]Difficulty mentioned by the group.

Device Concept

The needs of people with CLBP and health care professionals regarding the device concept are described in Tables 3 and 4 and illustrated with verbatim. All participants highlighted the importance of having a device that would not only help increase physical activity level of people with CLBP but also better pace it. They insisted on the need for this device to also help people with CLBP with long-term pain management. To achieve these objectives, they suggested the device should support the following behavior change strategies: goal setting, self-monitoring, feedback, positive reinforcement, reward delivery, and social support.

Table 3. Needs relating to device concept, especially objectives of the device, identified by experts, people with chronic low back pain (CLBP), and health care professionals.

Needs identified	Quotes (translated from French into English)
Help to pace physical activity	• "[The device should] allow self-pacing." [Person with CLBP; phase 2] • "I want it to be regular...even if performance increases progressively, the fact that it's regular is precisely to maintain the benefits for the back in the long term." [Person with CLBP; phase 2]
Support long-term self-management of pain and not only physical activity	• "That it's not just a device to improve physical activity, but also symptom management, pain management." [Ieg[a]; phase 1] • "The device must really improve the patient's well-being." [Ieg; phase 1]

[a]Ieg: interdisciplinary expert group.

Table 4. Needs relating to device concept, especially expected behavior change strategies supported by the device, identified by experts, people with chronic low back pain (CLBP), and health care professionals.

Needs identified	Quotes (translated from French into English)
Goal setting	• "What seemed important was that this device already allowed a certain degree of individualization, for example, based on patient goals." [Ieg[a]; phase 1] • "There must be clear goals. Without clear objectives, we'll just keep going. We'll get the story we want." [Person with CLBP; phase 2]
Self-monitoring	• "The system should allow the patient to monitor their progress." [Ieg; phase 1] • "To move forward, you need to know what's happened recently. How can you know if you don't have any data? Yes, it's been the same for 6 months, but the same, better, less? Well, I don't know. I can't remember." [Person with CLBP; phase 2]
Feedback and coaching advice sent by the device	• "We could also imagine motivational alerts if you've walked a little less or have been a little less active in the last few days. Not in a negative way, but in a positive way." [Ieg; phase 1] • "To have that on the smartphone to say, 'Ah today you basically did what you needed to do.' Or at the end of the day 'you haven't done all your steps' or 'you're missing some targeted exercises' and that would be magic." [Person with CLBP; phase 2]
Positive reinforcement	• "It's important that the words are kind and positive. You don't want to hear, 'Are you in pain?'" [Person with CLBP; phase 2] • "I think it's better to take into account positive emotions and parameters than to point out negative parameters such as stress!" [Hp[b]; phase 3]
Reward delivery	• "We wondered if something a little bit like a game could also stimulate activity...if the device could collect points or have something to help motivate to do exercises." [Ieg; phase 1] • "You took longer, you did more' suits me well, I've won stars, I've changed levels, super, I'm super happy. I mean, it's working really well for me." [Person with CLBP; phase 2]
Education	• "We know that when it comes to pain; simply put, the more you know, the less pain there is. So I don't know if a questionnaire that picks up things that have been covered in a program...to see if people still remember them, are able to remember them." [Hp; phase 2] • "To maintain our level of knowledge. All knowledge, in the end, we tend to go back to the representations we had in the past." [Hp; phase 2]
Social support through testimonials	• "Maybe if you can have patient testimonials, examples of people who have found solutions or things that are difficult and then they've been able to solve them. Maybe that can be motivating." [Ieg; phase 1] • "People with low back pain sometimes suffer from negative images, and to see that others have been successful, in quotes, I think can motivate them, even if it's anonymous." [Hp; phase 2]

[a]Ieg: interdisciplinary expert group.

[b]Hp: health care professional.

Device Content

The needs of people with CLBP and health care professionals regarding device content are described in Tables 5 and 6 and illustrated with verbatim. Participants expressed the need for a device that would allow people with CLBP to automatically track the following data: number of steps, heart rate, activity versus rest time, intensity of the activity, stress, and sleep. They emphasized the need for the collection of additional data recorded by the user, such as an activity diary, well-being, pleasure, satisfaction, comfort, medications taken, pain, sleep, and stress. They insisted that the user should specifically select automatically-collected and self-reported data. Participants also wanted the device to be able to send notifications to help people with CLBP manage their physical activity and pain. They highlighted the usefulness of receiving regular activity reports and of being able to access specific resources, such as physical exercises, relaxation exercises, and questionnaires to refresh knowledge about pain. They suggested the use of a wristband combined with a digital app that is easy to use, discreet, comfortable, waterproof, and robust.

Table 5. Needs relating to device content, especially functionalities, identified by experts, people with chronic low back pain (CLBP), and health care professionals.

Needs identified	Quotes (translated from French into English)
Automatically and effortlessly collect the user's activity data: steps, heart rate, active-resting time, intensity of activity, stress, and sleep	• "I think, on the contrary, to be able to objectively go back to your daily activities and then read, 'Oh well, yes, actually I'm doing a lot more than I feel I'm doing' or 'I'm doing a lot less than I feel I'm doing,' that's important." [Hp[a]; phase 2]
Recording of data entered by the user: activity diary, well-being, pleasure, satisfaction, comfort, medication taken, pain, stress, and sleep	• "To be able to quantify pain." [Person with CLBP; phase 2] • "All measures, obviously pleasure activities." [Hp; phase 2] • "I am not in favor of focusing on pain, but on the positive management of pain, what is the point of quantifying it for the sake of quantifying it, the nocebo vocabulary is still used too much, and everything is still based too much on negative criteria." [Hp; phase 3, round 2]
Notifications, alerts, and messages tailored to the user	• "If we think about artificial intelligence, we can imagine that the application will gradually be able to personalize advice based on the data collected." [Ieg[b]; phase 1] • "It could say: 'You're too stressed, do what's necessary to reduce it.'" [Person with CLBP; phase 2]
Reports: history accessible over a variable period (day, week, and month)	• "Have a PDF report of what we do at the end." [Person with CLBP; phase 2] • "The device captures all this data and then we make a summary to get an overview over several days, weeks, or even months." [Hp; phase 2]
Information transfer: provision of personalized physical exercises, provision of relaxation exercises, and provision of pain reminder questionnaires	• "I think with targeted exercises, but clearly individually, because we all do, we all have different activities." [Person with CLBP; phase 2] • "I've used a lot of digital tools, meditation tools, instant meditation support tools. It's been a great help." [Person with CLBP; phase 2] • "We tend to go back to the representations we had in the past. Questionnaire reminders." [Hp; phase 2]
Personalization: initial settings for automatically-collected data and data entered by the user, setting an activity goal	• "It's important to be able to personalize all the elements that the patient sees in a visual format or personalize them according to what they want to see." [Ieg; phase 1] • "Adapt the measured data." [Ieg; phase 1] • "Any personalization, tracking, and visualization options are welcome. The important thing is to give patients the freedom to choose what they want to track and share." [Person with CLBP; phase 3]

[a]Hp: health care professional.

[b]Ieg: interdisciplinary expert group.

Table 6. Needs relating to device content, especially characteristics, identified by experts, people with chronic low back pain (CLBP), and health care professionals.

Needs identified	Quotes (translated from French into English)
Support: wristband combined with a digital app on a smartphone	• "Just a wristband can be more discreet, with just one sensor. And then on your phone you'll see what you want to see." [Person with CLBP; phase 2]
Ease of use	• "For me, the keyword in all of this is simplicity. In all that it can mean in terms of tools, in terms of use, in terms of presentation. I really mean simplicity in the broadest sense." [Ieg[a]; phase 1] • "Something that's hyper user friendly, because otherwise it's not worth it." [Person with CLBP; phase 2]
Esthetics: discreet, comfortable, attractive, different colors, waterproof, and robust	• "In terms of comfort, something that's light, that's flexible." [Person with CLBP; phase 2] • "It must also be an attractive object." [Hp[b]; phase 2] • "The device would have to be waterproof because swimming is good for my pain and it would be a shame to lose my physical activity data. Also, if you're going to wear it 24 hours a day, it needs to recharge quickly." [Person with CLBP; phase 3]

[a]Ieg: interdisciplinary expert group.

[b]Hp: health care professional.

Conditions of Use

The needs of people with CLBP and health care professionals regarding the conditions of use are described in Table 7 and illustrated with verbatim. Participants indicated that this device may be used either 24 hours per day or only during the day, depending on the user's preference. They wanted it to be provided by a health care professional as part of a rehabilitation program, and they wanted health care professional supervision for its use. They wanted to be able to consult the data collected in real time, but some health care professionals expressed concern about misinterpretation and suggested that the data should initially be only available in the presence of a therapist. All participants insisted on the importance of long-term follow-up by a health care professional to check the use of the device and to help adjust the goals (either face-to-face or by videoconference).

Table 7. Needs relating to conditions of use identified by experts, people with chronic low back pain (CLBP), and health care professionals.

Conditions of use	Needs identified	Quotes (translated from French into English)
Frequency of use of the device	Can be used continuously all day, or night and day (according to the user's needs)	• "It has to be worn 24/7." [Person with CLBP; phase 2] • "I'd say all day, 24 hours a day." [Hp[a]; phase 2] • "Be careful about wearing 24/7—to be determined according to the objectives and the patient—I think wearing it all the time is difficult over time too." [Hp; phase 3]
Conditions for providing the device	By a health professional, as part of a rehabilitation program, after educational therapy	• "The devices are not that easy to use, so it's also part of the therapist's role to teach how to use them, or even provide tutorials." [Ieg[b]; phase 1] • "Something that will be implemented during therapeutic education workshops when we're doing nondrug tools." [Hp; phase 2]
Conditions for consulting the data for the user	In real time or when the user is a novice: possibly accompanied by a health care professional the first few times	• "You go to the app, you look when you feel like it." [Person with CLBP; phase 2] • "That they (patients) have access to information a posteriori, that they don't have access to information in real time, that is, that the watch or eventually the device. The application will record the data, but it won't be visible. It will only be visible after x amount of time." [Hp; phase 2] • "Real time also has its advantages. It can be motivating, depending on the parameters. There's the whole biofeedback side, which has proven to be very useful in certain situations." [Hp; phase 2]
Follow-up conditions related to device use	Follow-up session with a health care professional after a period of device use, possibly by videoconference	• "And maybe the digital tool and, at some point, a human relay that comes back on time that we can readjust." [Person with CLBP; phase 2] • "That's when I tell myself that teleconsultation makes sense, even in a group. At the end of a program or afterward, when you've really got an appointment." [Hp; phase 2] • "To have a regular appointment with a professional to take stock." [Person with CLBP; phase 3] • "For it to be beneficial, you'd have to have a regular review with a health professional, readjust the objectives." [Person with CLBP; phase 3]

[a]Hp: health care professional.

[b]Ieg: interdisciplinary expert group.

Facilitators and Barriers to Device Use

The facilitators and barriers to the use of a device mentioned by the participants in all the study phases are described in Table 8. The facilitators mentioned corresponded to the functionalities and features considered necessary for the content of the device. The potential barriers were related to concerns about the validity of the data collected, the storage and confidentiality of that data, the need to recharge the device, the additional availability of health care professionals to provide and monitor the device, and the risk of adverse effects associated with the use of the device.

Table 8. Facilitators and barriers to the use of a device identified by people with chronic low back pain (CLBP) and health care professionals in phases 2 and 3.

Facilitators and barriers	Quotes (translated from French into English)
Facilitators	
Personalization	• Provided in Table 5
Ease of use	• Provided in Table 6
Esthetic	• Provided in Table 6
Conditions of use	• Provided in Table 7
Barriers	
Validity of automatic data collection	• "(The device) is not ready for sleep yet." [Person with CLBP; phase 3] • "In addition, the measurement accuracy of these objects is still very random." [Hp[a]; phase 2] • "Heart rate is a very interesting variable, but the measuring devices are often unreliable." [Hp; phase 3]
Data storage and security	• "I'm not so much a fan of data in the cloud. I'm not a big fan of data that just magically disappears." [Person with CLBP; phase 2] • "To have this tool that measures all our activities for weeks on end? Well, there are things that are a little bit in the realm of privacy that will also be measured." [Person with CLBP; phase 2]
Instrument: setting up and reloading the device	• "The debate is about the battery and recharging." [Person with CLBP; phase 2] • "The only drawback, like any electronic device, is that it needs to be recharged, which creates a time lapse where the data is not, cannot be collected automatically and subjectively by the device." [Hp; phase 2]
Conditions for making the device available and monitoring its use	• "In the feasibility of the current programs. We'd need more time to integrate such a tool, because if we have to add things. If we have to add another tool to integrate. On top of what we're already asked to do. Of course, that means extra time." [Hp; phase 2]
Adverse reactions linked to the use of the device: misinterpretation, guilt	• "They make cause-and-effect relationships that are very random. And these objects also have a very random precision of measurement, so sometimes they come to conclusions that are completely off the mark, and sometimes, in my opinion, in some cases, I'm not saying for all, but it can be counterproductive." [Hp; phase 2] • "Wearing a connected bracelet 24 hours a day can be stimulating for some profiles, but it can also be guilt-inducing for others." [Hp; phase 3]

[a]Hp: health care professional.

Needs Validation

The needs of people with CLBP and health care professionals regarding device concept and content and the conditions of use were confirmed in 2 rounds of Delphi questionnaires with 86 and 61 participants, respectively (Tables 9 and 10). Twenty-eight needs were selected from those identified in the previous 2 study phases. A lot of attention was given to the variables that could be collected or recorded by the device, as much of the discussion in phases 1 and 2 focused on these aspects. In the first round, consensus was reached on 13 (46%) of the 28 needs (the 13 validated needs are italicized in Table 9). In the second round, the 15 needs for which there was no consensus were modified based on the suggestions made in the first round of consultation: 11 (73%) of the 15 needs were reformulated, and 3 (20%) of the 15 needs were merged into a single need (changes in needs are italicized in Table 10). In the second round, consensus was reached on 11 (91%) of the 12 needs. Out of 25 needs in total, a consensus was reached on 24 (96%) needs.

Table 9. Degree of consensus on the items proposed in the Delphi questionnaire for each participant group (round 1).

Items: "The device should make it possible to"	All, (N=86; %)	People with CLBP[a] (n=45; %)	Hp[b] (n=41; %)
Achieve a personalized physical activity goal [c] *(Cc* [d]*)*	99	98	100
Produce a report on changes in the user's physical activity[c] *(Ct* [e]*)*	99	98	100
Be set up with the support of a Hp as part of a rehabilitation program[c] *(Cu* [f]*)*	98	98	98
Provide various pain management tools and techniques (eg, relaxation, cardiac coherence, etc)[c] *(Ct)*	97	98	95
Generate a report of the variables collected that can be given to a Hp[c] *(Cu)*	97	96	98
Assess and monitor the physical performance of the user[c] *(Cc)*	95	96	95
Quantify heart rate automatically and in real time[c] *(Ct)*	93	100	85
Quantify the time spent on activities of different intensities automatically and in real time[c] *(Ct)*	93	98	88
Qualify the quality of sleep at a set frequency (eg, NS [g]*: 0-10)*[c] *(Ct)*	93	98	88
Qualify the activities carried out at a set frequency[c] *(Ct)*	93	98	88
Quantify activity and rest periods automatically and in real time[c] *(Ct)*	93	96	90
Quantify the sleep duration automatically and in real time[c] *(Ct)*	92	96	88
Monitor physical performance according to variables chosen by the user[c] *(Ct)*	91	98	83
Display changes in the user's activity in real time (Cu)	90	96	83
Qualify satisfaction at a set frequency (eg, NS: 0-10) (Ct)	88	93	83
Be offered on an optional basis, as a complement to the rehabilitation (Cu)	88	93	81
Qualify level of stress at a set frequency (eg, NS: 0-10) (Ct)	87	96	78
Consist of a connected wristband combined with a digital application (Ct)	87	90	85
Qualify the level of pleasure at a set frequency (eg, NS: 0-10) (Ct)	86	91	81
Qualify pain intensity at a set frequency (eg, NS: 0-10) (Ct)	85	96	73
Quantify the number of steps taken automatically and in real time (Ct)	84	96	71
Qualify feelings at a set frequency (eg, NS: 0-10) (Ct)	84	91	73
Qualify "weather emotions" at a set frequency (eg, by ticking the weather image for actual feeling) (Ct)	84	84	83
Quantify stress levels automatically and in real time (eg, automatic score based on heart rate variability) (Ct)	83	96	76
Qualify medication taken at a set frequency (eg, type and dose) (Ct)	81	89	73
Obtain rewards (eg, badges or points) (Cc)	81	76	88
Be available, on an optional basis, after the end of the rehabilitation (Cu)	78	96	61
Be wearable 24 hours a day (Cu)	66	78	54

[a]CLBP: chronic low back pain.

[b]Hp: health care professionals.

[c]Validated needs are italicized.

[d]Cc: concept device.

[e]Ct: content device.

[f]Cu: condition of use.

[g]NS: numeric scale.

Table 10. Degree of consensus on the items proposed in the Delphi questionnaire for each participant group (round 2).

Items modified according to the results of the first round (changes are italicized). The device should be able to...	All (N=61; %)	People with CLBP[a] (n=33; %)	Hp[b] (n=28; %)
Consist of an object, such as a *connected watch* or bracelet combined with a digital application (Ct[c])	100	100	100
Qualify "weather emotions" at a set frequency *if the user wishes* (eg, by ticking the weather image representing actual feelings). *The choice to take this data into account may be discussed with a Hp* (Ct)	98	100	96
Be offered on an optional basis, as a complement to the rehabilitation program, *by a Hp* (Cu[d])	98	100	96
Be wearable every day *during the day*, or 24 hours a day *(depending on the data the user wishes to collect)* (Cu)	98	97	100
Qualify *the pain treatments used, if the user wishes,* at a set defined frequency (eg, the *type and dose of medicinal treatments, the type of stretching and duration, breathing techniques, and the time taken to perform them). The choice of taking this information into account may be discussed with a Hp* (Ct)	97	100	93
See activity progression in real time, *if the user wishes. It could also enable users to see how their activity is progressing at certain stages in their support, in partnership with a Hp involved in the rehabilitation program* (Cu)	97	100	93
Quantify the number of steps taken automatically and in real time. *The choice of how this data is taken into account may be discussed with a Hp* (Ct)	97	100	93
Qualify, at a set frequency, a *parameter with a positive connotation,* such as satisfaction, pleasure, or feeling, *if the user wishes. The choice of how this data is taken into account could be discussed with a Hp* (Ct)	97	97	96
Qualify according to a set frequency the level of stress, *if the user wishes* (eg, NS[e]: 0-10). *The choice of how this data is taken into account may be discussed with a Hp* (Ct)	95	97	93
Qualify the intensity of the pain at a set frequency, *if the user wishes* (eg, NS: 0-10). *The choice of how this data is taken into account may be discussed with a Hp* (Ct)	93	100	86
Have an attractive "challenge" section (eg, with a badge or points system), *that the user would be free to consult or not, to motivate him or her* (Cc[f])	92	91	93
Be available, on an optional basis, after the end of the rehabilitation program, *depending on the person's degree of motivation to use it and their desire to make a financial investment in the device* (Cu)	89	100	75

[a]CLBP: chronic low back pain.

[b]Hp: health care professionals.

[c]Ct: content device.

[d]Cu: condition of use.

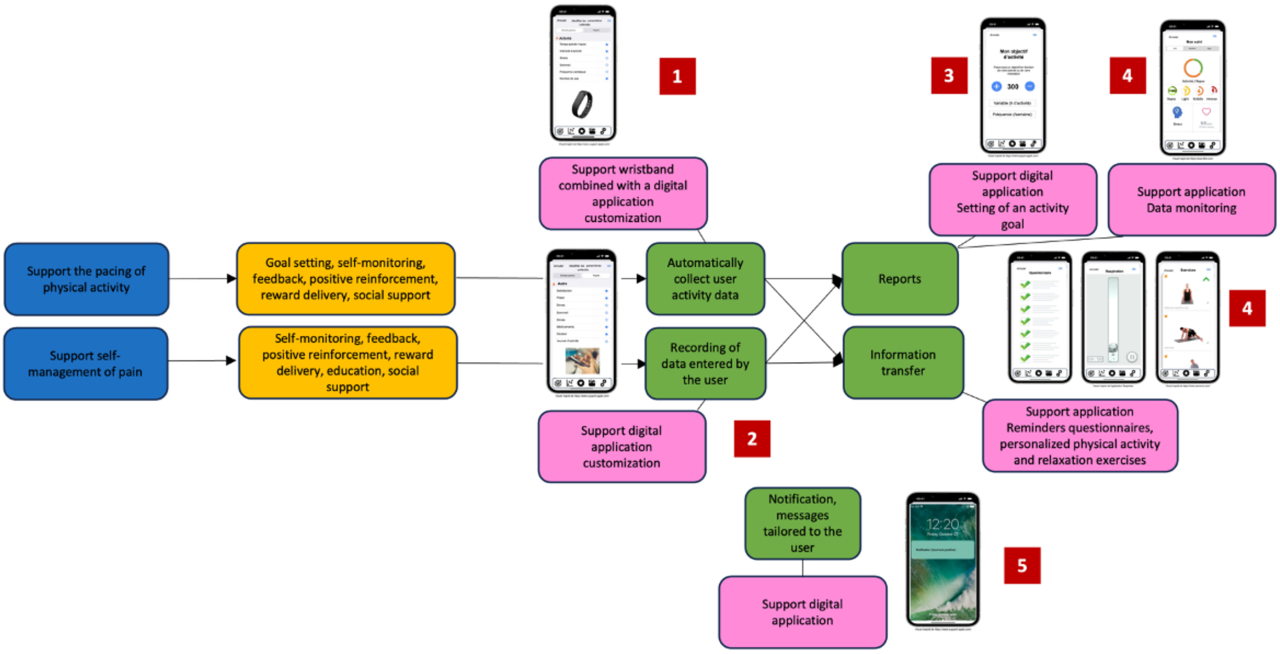
[e]NS: numeric scale.

[f]Cc: concept device.

Synthesis of Design Recommendations

The needs validated by the participants were synthesized as design recommendations using the Behavioral Intervention Technology model and are illustrated in Figure 2. This is based on the illustration of the MyFitnessPal mHealth app by Mohr et al [56] and describes the expected functionalities and technical features of the device. The numbers shown illustrate the conditions of use and the expected workflow: (1) first, the automatic data collected by the device can be selected, then (2) the data collected by the user can also be selected. From this, (3) an activity objective can be defined, (4) and the user can follow changes in the values of variables collected or recorded, referring, if he or she wishes, to the physical or relaxation exercises proposed. (5) Reception of regular notifications and alarms can be set.

Figure 2. Illustration of design recommendations for a mobile health device that specifically meets the identified needs of people with chronic low back pain and health care professionals to support self-managed physical activity.



Discussion

Principal Findings

This study explored the needs of people with CLBP and health care professionals regarding the use of mHealth technology to support self-managed physical activity and pain. It elucidated how people with CLBP believed the adoption of an mHealth device could mitigate their difficulties in managing physical activity in relation to their pain, beliefs, difficulties in pacing physical activity, and the need for long-term individualized support.

The results of this study show that participants (people with CLBP and health care professionals) expressed interest in using an mHealth system to help people with CLBP manage their physical activity and pain over the long term. They wanted a device that would allow them to set personalized activity goals, monitor their activity, receive feedback on their performance, receive positive reinforcement, possibly receive rewards, learn information, and receive social support. These results provide a more precise definition of the needs identified by Merolli et al [46], who highlighted the usefulness of developing a technological solution to support the self-management of CLBP with tracking, notifications, feedback, provision of educational resources, and exercises. These results, collected using a rigorous process, accurately describe the expectations of end users in terms of the characteristics of a potential device.

Special attention was paid to the functionalities and variables recorded by the device. This is a fundamental step in the development of effective technology [57]. People with CLBP and health care professionals described the need to set realistic activity goals and to track them using an mHealth device, while also having complementary tools to help them manage their pain daily (eg, a physical activity exercise bank and relaxation exercises). They strongly emphasized that physical activity should be monitored using self-selected variables, stating that the need was not always to increase physical activity level, but rather to better pace it daily through achievable goals. Several participants with CLBP highlighted their tendency to overdo activity, debunking the assumption that physical activity levels in people with CLBP are often lower than in healthy people [58-60].

Overall, people with CLBP and health care professionals expressed similar needs in terms of an mHealth device. They strongly emphasized that the settings should be personalized, and the device should be easy to use. However, the opinions of these groups differed regarding the variables that should be collected and the components that should be provided by the device. Health care professionals believed the focus should not be on the assessment of negative variables, such as pain, whereas people with CLBP wanted to collect this type of variable. Health care professionals suggested that a “challenge” module should be included, with a system to collect points or badges, to encourage users to interact with the device. The opinions of the people with CLBP were more divided about such a module, with some very interested in this type of reward and others completely opposed to it because they were afraid of being involved in competitive challenges. The gamification of mHealth is attracting increasing interest as a means of encouraging changes in physical activity behavior, although current evidence on the effectiveness of this type of device is still limited [61,62].

The potential disadvantages and adverse effects associated with the use of a device to self-manage physical activity were discussed in detail during the 3 phases of the study. On the one hand, the effects of such a device on the development of new behaviors and lifestyles may be directly linked to its maintenance. For example, behavior change is less likely to occur if the device monitor is lost, broken, not charged, or forgotten. In contrast, if the device is used consistently and integrated into an intervention aimed at maintaining new behaviors, new habits could emerge in the medium term [63,64]. In addition, some health care professionals were concerned that people with CLBP could misinterpret the data collected by the device, and that this could induce feelings of guilt. Indeed, some health care professionals were against users being able to consult their data in real time, whereas people with CLBP saw this as an obvious possibility. The ability of the mHealth device to provide feedback to the user is considered important for encouraging behavioral change [61,65]. Health care professionals also mentioned the additional time required to teach their patients to use the device. In addition, both health care professionals and people with CLBP expressed concern about the validity of automatically-collected data, its security, and the individual's privacy.

The discussions went well beyond the functionality of the device. The results highlighted the importance of developing an intervention around the mHealth device to help people with CLBP manage their activity and pain, rather than simply relying on the mHealth device. Human support was stated as an essential component of long-term care for people with CLBP. These results confirmed the findings of Svendsen et al [66] that technology alone was not enough to encourage people to engage in self-management. All participants were very clear about the conditions in which the device should be made available, that is, its implementation should be supported by a health care professional, and education should be provided on the importance of physical activity, its role in pain management, and the operation of the device. Education is important to increase the effectiveness of device use because people's beliefs determine their behavior [67]. These findings go beyond those of the pilot study by Ellingson et al [68], which showed the promising effects of a device-based intervention combined with minimal human support for people with CLBP; because they specified the expected characteristics of human support, in particular, the context in which the device should be offered, how it should be supported by health care professionals, and the frequency at which health care professional support should be provided.

After validating the needs of people with CLBP and health care professionals regarding mHealth devices, we developed design recommendations using the Behavioral Intervention Technology model (Figure 2) [56]. This model helped to structure the needs identified during the 3 phases of the study and to link them to the behavior change strategies that could be integrated within a new intervention encompassing the use of an mHealth device. These design recommendations should now be further developed in the light of behavior change theories and taxonomies currently published in the literature (eg, Michie's taxonomy) [14,69,70]. Future research should also seek to discuss and refine the prototype development phase with end users. Indeed, after this initial conceptualization phase, this synthesis could be revisited at a later stage to continue the prototype development phase, still using a user-centered approach.

Limitations

This study was based on the experiences of experts and potential end users. It first identified the needs of 9 experts in phase 1, followed by 26 people with CLBP and health care professionals in phase 2, and then validated these needs with a large number of participants (86 in the first round and 62 in the second round). Despite the sample size, the first limitation concerns the population studied, and its impact on the representativeness and transferability of the results. The participants with CLBP and the health care professionals were all involved in multidisciplinary rehabilitation; therefore, the results may not be generalizable to other settings. This choice was made to have a comparable, homogeneous sample, but the needs and experiences of people with CLBP who do not have access to multidisciplinary rehabilitation (ie, only physiotherapy sessions) may be different.

The Delphi method used in phase 3 may have been subject to some bias. Although the needs for device design were developed during phases 1 and 2, not all the elements that emerged in the first 2 phases were specifically examined during phase 3. The elements selected by the research team for the Delphi questionnaire are those that have been the subject of significant debate among the participants in the previous phases. Thus, in phase 3, particular attention was paid to the variables collected by the device, the content, and the condition of use. This process may have left out elements that would have been also important to present in the Delphi questionnaire. In addition, participants involved in phase 2 were invited to take part in phase 3, which might have introduced bias because of their previous involvement in defining the needs. This involvement may have resulted in an overestimation of the consensus level achieved in the 2 rounds of Delphi questionnaires.

Conclusions

This study used a UCD approach to explore the needs of people with CLBP and health care professionals regarding the use of a device to support self-managed physical activity. It identified potential users' expectations regarding the device's objectives, behavior-change strategies supported by the device, the device's functionalities, its technical features, and facilitators and barriers that may influence its implementation in a clinical context. The results were used to delineate design recommendations for the development of an mHealth device specifically for people with CLBP and health care professionals. These recommendations will be used in the future to prototype innovative devices that could be offered by health care professionals to people with CLBP in a follow-up rehabilitation context.

Acknowledgments

This study was supported by the University of Applied Sciences and Arts Western Switzerland (Haute école spécialisée de Suisse occidentale, HES-SO). It was also supported by the Claude Bernard Lyon 1 University and Ezus association. The authors thank Emmanuelle Chaléat-Valayer, Stéphane Genevay, and Damiano Salvi for their help in recruiting participants.

Authors' Contributions

MB contributed to conceptualization, funding acquisition, data collection, analysis of results, and writing and editing the manuscript. AD-B contributed to conceptualization, data collection, analysis of results, and reviewing the manuscript. JJD contributed to conceptualization, data collection, analysis of results, and reviewing the manuscript. GC contributed to conceptualization, analysis of results, and reviewing the manuscript. LC contributed to conceptualization, funding acquisition, analysis of results, and reviewing the manuscript. TR contributed to conceptualization, funding acquisition, analysis of results, and reviewing the manuscript. AMB contributed to conceptualization, funding acquisition, project administration, supervision, analysis of results, and reviewing and editing the manuscript. All authors have approved the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of people with chronic low back pain and healthcare professionals who participated in each phase of the study.
[DOCX File , 43 KB - humanfactors_v11i1e59897_app1.docx]

References

1. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1789-1858 [FREE Full text] [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
2. GBD 2021 Low Back Pain Collaborators. Global, regional, and national burden of low back pain, 1990-2020, its attributable risk factors, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021. Lancet Rheumatol 2023 Jun;5(6):e316-e329 [FREE Full text] [doi: 10.1016/S2665-9913(23)00098-X] [Medline: 37273833]
3. Hartvigsen J, Hancock MJ, Kongsted A, Louw Q, Ferreira ML, Genevay S, et al. What low back pain is and why we need to pay attention. Lancet 2018 Jun 09;391(10137):2356-2367 [FREE Full text] [doi: 10.1016/S0140-6736(18)30480-X] [Medline: 29573870]
4. GBD 2021 Other Musculoskeletal Disorders Collaborators. Global, regional, and national burden of other musculoskeletal disorders, 1990-2020, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021. Lancet Rheumatol 2023 Nov;5(11):e670-e682 [FREE Full text] [doi: 10.1016/S2665-9913(23)00232-1] [Medline: 37927903]
5. Vlaeyen JW, Linton SJ. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. Pain 2000 Apr;85(3):317-332 [FREE Full text] [doi: 10.1016/S0304-3959(99)00242-0] [Medline: 10781906]
6. de Moraes Vieira EB, de Góes Salvetti M, Damiani LP, de Mattos Pimenta CA. Self-efficacy and fear avoidance beliefs in chronic low back pain patients: coexistence and associated factors. Pain Manag Nurs 2014 Sep;15(3):593-602. [doi: 10.1016/j.pmn.2013.04.004] [Medline: 23891180]
7. Burbridge C, Randall JA, Abraham L, Bush EN. Measuring the impact of chronic low back pain on everyday functioning: content validity of the Roland Morris disability questionnaire. J Patient Rep Outcomes 2020 Aug 28;4(1):70 [FREE Full text] [doi: 10.1186/s41687-020-00234-5] [Medline: 32857224]
8. Foster NE, Anema JR, Cherkin D, Chou R, Cohen SP, Gross DP, et al. Prevention and treatment of low back pain: evidence, challenges, and promising directions. The Lancet 2018 Jun;391(10137):2368-2383. [doi: 10.1016/s0140-6736(18)30489-6] [Medline: 29573872]
9. Hayden JA, Ellis J, Ogilvie R, Malmivaara A, van Tulder MW. Exercise therapy for chronic low back pain. Cochrane Database Syst Rev 2021 Sep 28;9(9):CD009790 [FREE Full text] [doi: 10.1002/14651858.CD009790.pub2] [Medline: 34580864]
10. Low back pain and sciatica in over 16s: assessment and management. National Institute for Health and Care Excellence. 2016 Nov 30. URL: https://www.nice.org.uk/guidance/ng59 [accessed 2024-02-06]
11. Management of the patient with common low back pain. Haute Autorité de Santé. 2019 Apr 4. URL: https://tinyurl.com/4a6mfpt3 [accessed 2024-03-23]
12. Qaseem A, Wilt TJ, McLean RM, Forciea MA, Clinical Guidelines Committee of the American College of Physicians, Denberg TD, et al. Noninvasive treatments for acute, subacute, and chronic low back pain: a clinical practice guideline from the American College of Physicians. Ann Intern Med 2017 Apr 04;166(7):514-530 [FREE Full text] [doi: 10.7326/M16-2367] [Medline: 28192789]
13. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Counsel 2002 Oct;48(2):177-187. [doi: 10.1016/S0738-3991(02)00032-0] [Medline: 12401421]
14. Mansell G, Hall A, Toomey E. Behaviour change and self-management interventions in persistent low back pain. Best Pract Res Clin Rheumatol 2016 Dec;30(6):994-1002. [doi: 10.1016/j.berh.2017.07.004] [Medline: 29103556]

15. Bagg MK, Wand BM, Cashin AG, Lee H, Hübscher M, Stanton TR, et al. Effect of graded sensorimotor retraining on pain intensity in patients with chronic low back pain: a randomized clinical trial. JAMA 2022 Aug 02;328(5):430-439 [FREE Full text] [doi: 10.1001/jama.2022.9930] [Medline: 35916848]
16. Chou R, Deyo R, Friedly J, Skelly A, Hashimoto R, Weimer M, et al. Nonpharmacologic therapies for low back pain: a systematic review for an American College of Physicians clinical practice guideline. Ann Intern Med 2017 Apr 04;166(7):493-505 [FREE Full text] [doi: 10.7326/M16-2459] [Medline: 28192793]
17. Rizzo RR, Wand BM, Leake HB, O'Hagan ET, Bagg MK, Bunzli S, et al. "My back is fit for movement": a qualitative study alongside a randomized controlled trial for chronic low back pain. J Pain 2023 May;24(5):824-839 [FREE Full text] [doi: 10.1016/j.jpain.2022.12.009] [Medline: 36577460]
18. Keogh A, Tully MA, Matthews J, Hurley DA. A review of behaviour change theories and techniques used in group based self-management programmes for chronic low back pain and arthritis. Man Ther 2015 Dec;20(6):727-735. [doi: 10.1016/j.math.2015.03.014] [Medline: 25865062]
19. Briggs AM, Jordan JE, O'Sullivan PB, Buchbinder R, Burnett AF, Osborne RH, et al. Individuals with chronic low back pain have greater difficulty in engaging in positive lifestyle behaviours than those without back pain: an assessment of health literacy. BMC Musculoskelet Disord 2011 Jul 15;12(1):161 [FREE Full text] [doi: 10.1186/1471-2474-12-161] [Medline: 21756363]
20. Protheroe J, Rogers A, Kennedy AP, Macdonald W, Lee V. Promoting patient engagement with self-management support information: a qualitative meta-synthesis of processes influencing uptake. Implement Sci 2008 Oct 13;3(1):44 [FREE Full text] [doi: 10.1186/1748-5908-3-44] [Medline: 18851743]
21. mHealth programmes are sponsored by government. World Health Organization. URL: https://www.who.int/data/gho/indicator-metadata-registry/imr-details/4774 [accessed 2024-01-30]
22. Oosterveen E, Tzelepis F, Ashton L, Hutchesson MJ. A systematic review of eHealth behavioral interventions targeting smoking, nutrition, alcohol, physical activity and/or obesity for young adults. Prev Med 2017 Jun;99:197-206. [doi: 10.1016/j.ypmed.2017.01.009] [Medline: 28130046]
23. Routhier F, Duclos NC, Lacroix É, Lettre J, Turcotte E, Hamel N, et al. Clinicians' perspectives on inertial measurement units in clinical practice. PLoS One 2020 Nov 13;15(11):e0241922 [FREE Full text] [doi: 10.1371/journal.pone.0241922] [Medline: 33186363]
24. Smuck M, Tomkins-Lane C, Ith MA, Jarosz R, Kao MC. Physical performance analysis: a new approach to assessing free-living physical activity in musculoskeletal pain and mobility-limited populations. PLoS One 2017 Feb 24;12(2):e0172804 [FREE Full text] [doi: 10.1371/journal.pone.0172804] [Medline: 28235039]
25. Du S, Liu W, Cai S, Hu Y, Dong J. The efficacy of e-health in the self-management of chronic low back pain: a meta analysis. Int J Nurs Stud 2020 Jun;106:103507. [doi: 10.1016/j.ijnurstu.2019.103507] [Medline: 32320936]
26. Papi E, Murtagh GM, McGregor AH. Wearable technologies in osteoarthritis: a qualitative study of clinicians' preferences. BMJ Open 2016 Jan 25;6(1):e009544 [FREE Full text] [doi: 10.1136/bmjopen-2015-009544] [Medline: 26810998]
27. van der Weegen S, Verwey R, Spreeuwenberg M, Tange H, van der Weijden T, de Witte L. The development of a mobile monitoring and feedback tool to stimulate physical activity of people with a chronic disease in primary care: a user-centered design. JMIR Mhealth Uhealth 2013 Jul 02;1(2):e8 [FREE Full text] [doi: 10.2196/mhealth.2526] [Medline: 25099556]
28. Appelboom G, Camacho E, Abraham ME, Bruce SS, Dumont EL, Zacharia BE, et al. Smart wearable body sensors for patient self-assessment and monitoring. Arch Public Health 2014;72(1):28 [FREE Full text] [doi: 10.1186/2049-3258-72-28] [Medline: 25232478]
29. Korpershoek YJ, Hermsen S, Schoonhoven L, Schuurmans MJ, Trappenburg JC. User-centered design of a mobile health intervention to enhance exacerbation-related self-management in patients with chronic obstructive pulmonary disease (Copilot): mixed methods study. J Med Internet Res 2020 Jun 15;22(6):e15449 [FREE Full text] [doi: 10.2196/15449] [Medline: 32538793]
30. Rintala A, Rantalainen R, Kaksonen A, Luomajoki H, Kauranen K. mHealth apps for low back pain self-management: scoping review. JMIR Mhealth Uhealth 2022 Aug 26;10(8):e39682 [FREE Full text] [doi: 10.2196/39682] [Medline: 36018713]
31. de Melo Santana B, Raffin Moura J, Martins de Toledo A, Burke TN, Fernandes Probst L, Pasinato F, et al. Efficacy of mHealth interventions for improving the pain and disability of individuals with chronic low back pain: systematic review and meta-analysis. JMIR Mhealth Uhealth 2023 Nov 02;11:e48204 [FREE Full text] [doi: 10.2196/48204] [Medline: 37962085]
32. Zhou T, Salman D, McGregor A. mHealth apps for the self-management of low back pain: systematic search in app stores and content analysis. JMIR Mhealth Uhealth 2024 Feb 01;12:e53262 [FREE Full text] [doi: 10.2196/53262] [Medline: 38300700]
33. Berger M, Bertrand AM, Robert T, Chèze L. Measuring objective physical activity in people with chronic low back pain using accelerometers: a scoping review. Front Sports Act Living 2023 Nov 1;5:1236143 [FREE Full text] [doi: 10.3389/fspor.2023.1236143] [Medline: 38022769]
34. Marcuzzi A, Nordstoga AL, Bach K, Aasdahl L, Nilsen TIL, Bardal EM, et al. Effect of an artificial intelligence-based self-management app on musculoskeletal health in patients with neck and/or low back pain referred to specialist care: a

randomized clinical trial. JAMA Netw Open 2023 Jun 01;6(6):e2320400 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.20400] [Medline: 37368401]
35. Amorim AB, Pappas E, Simic M, Ferreira ML, Jennings M, Tiedemann A, et al. Integrating mobile-health, health coaching, and physical activity to reduce the burden of chronic low back pain trial (IMPACT): a pilot randomised controlled trial. BMC Musculoskelet Disord 2019 Feb 11;20(1):71 [FREE Full text] [doi: 10.1186/s12891-019-2454-y] [Medline: 30744606]
36. Oliveira CB, Christofaro DG, Maher CG, Franco MR, Tiedemann A, Silva FG, et al. Adding physical activity coaching and an activity monitor was no more effective than adding an attention control intervention to group exercise for patients with chronic nonspecific low back pain (PAyBACK Trial): a randomized trial. J Orthop Sports Phys Ther 2022 May;52(5):287-299. [doi: 10.2519/jospt.2022.10874] [Medline: 35536245]
37. Gordon R, Bloxham S. Influence of the Fitbit Charge HR on physical activity, aerobic fitness and disability in non-specific back pain participants. J Sports Med Phys Fitness 2017 Dec;57(12):1669-1675. [doi: 10.23736/s0022-4707.17.06688-9]
38. Zhuo LX, Macedo LG. Feasibility and convergent validity of an activity tracker for low back pain within a clinical study: cross-sectional study. JMIR Rehabil Assist Technol 2021 Mar 26;8(1):e18942 [FREE Full text] [doi: 10.2196/18942] [Medline: 33769301]
39. Hurley DA, Tully MA, Lonsdale C, Boreham CA, van Mechelen W, Daly L, et al. Supervised walking in comparison with fitness training for chronic back pain in physiotherapy: results of the SWIFT single-blinded randomized controlled trial (ISRCTN17592092). Pain 2015 Jan;156(1):131-147. [doi: 10.1016/j.pain.0000000000000013] [Medline: 25599309]
40. Lansing JE, Ellingson LD, DeShaw KJ, Cruz-Maldonado G, Hurt TR, Meyer JD. A qualitative analysis of barriers and facilitators to reducing sedentary time in adults with chronic low back pain. BMC Public Health 2021 Jan 26;21(1):215 [FREE Full text] [doi: 10.1186/s12889-021-10238-5] [Medline: 33499812]
41. Sitthipornvorakul E, Klinsophon T, Sihawong R, Janwantanakul P. The effects of walking intervention in patients with chronic low back pain: a meta-analysis of randomized controlled trials. Musculoskelet Sci Pract 2018 Apr;34:38-46. [doi: 10.1016/j.msksp.2017.12.003] [Medline: 29257996]
42. Duffy A, Christie GJ, Moreno S. The challenges toward real-world implementation of digital health design approaches: narrative review. JMIR Hum Factors 2022 Sep 09;9(3):e35693 [FREE Full text] [doi: 10.2196/35693] [Medline: 36083628]
43. Moore G, Wilding H, Gray K, Castle D. Participatory methods to engage health service users in the development of electronic health resources: systematic review. J Particip Med 2019 Feb 22;11(1):e11474 [FREE Full text] [doi: 10.2196/11474] [Medline: 33055069]
44. Oliveira M, Zancul E, Fleury AL. Design thinking as an approach for innovation in healthcare: systematic review and research avenues. BMJ Innov 2020 Dec 14;7(2):491-498. [doi: 10.1136/bmjinnov-2020-000428]
45. Cucciniello M, Petracca F, Ciani O, Tarricone R. Development features and study characteristics of mobile health apps in the management of chronic conditions: a systematic review of randomised trials. NPJ Digit Med 2021 Oct 05;4(1):144 [FREE Full text] [doi: 10.1038/s41746-021-00517-1] [Medline: 34611287]
46. Merolli M, Marshall CJ, Pranata A, Paay J, Sterling L. User-centered value specifications for technologies supporting chronic low-back pain management. Stud Health Technol Inform 2019 Aug 21;264:1288-1292. [doi: 10.3233/SHTI190434] [Medline: 31438133]
47. Grolier M, Arefyev A, Pereira B, Tavares Figueiredo I, Gerbaud L, Coudeyre E. Refining the design of a smartphone application for people with chronic low back pain using mixed quantitative and qualitative approaches. Disabil Rehabil Assist Technol 2023 Feb 05;18(2):145-150. [doi: 10.1080/17483107.2020.1839575] [Medline: 33151760]
48. Norman DA, Draper SW. User Centered System Design: New Perspectives on Human-Computer Interaction. Mahwah, NJ: Lawrence Erlbaum Associates; 1986.
49. Norman D. The Design of Everyday Objects. Paris, France: Eyrolles Publisher; 2020.
50. Ten Klooster I, Wentzel J, Sieverink F, Linssen G, Wesselink R, van Gemert-Pijnen L. Personas for better targeted eHealth technologies: user-centered design approach. JMIR Hum Factors 2022 Mar 15;9(1):e24172 [FREE Full text] [doi: 10.2196/24172] [Medline: 35289759]
51. Kitzinger J, Marková I, Kalampalikis N. Qu'est-ce que les focus groups? Bull Psychol 2004;57(3):237-243. [doi: 10.3406/bupsy.2004.15339]
52. Nasa P, Jain R, Juneja D. Delphi methodology in healthcare research: how to decide its appropriateness. World J Methodol 2021 Jul 20;11(4):116-129 [FREE Full text] [doi: 10.5662/wjm.v11.i4.116] [Medline: 34322364]
53. Guest G, MacQueen KM, Namey EE. Introduction to applied thematic analysis. In: Applied Thematic Analysis. Thousand Oaks, CA: SAGE Publications; 2014.
54. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis: striving to meet the trustworthiness. Int J Qual Methods 2017 Oct 02;16(1). [doi: 10.1177/1609406917733847]
55. Diamond IR, Grant RC, Feldman BM, Pencharz PB, Ling SC, Moore AM, et al. Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. J Clin Epidemiol 2014 Apr;67(4):401-409. [doi: 10.1016/j.jclinepi.2013.12.002] [Medline: 24581294]
56. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 05;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]

57. Molina-Recio G, Molina-Luque R, Jiménez-García AM, Ventura-Puertos PE, Hernández-Reyes A, Romero-Saldaña M. Proposal for the user-centered design approach for health apps based on successful experiences: integrative review. JMIR Mhealth Uhealth 2020 Apr 22;8(4):e14376 [FREE Full text] [doi: 10.2196/14376] [Medline: 32319965]
58. van Weering MG, Vollenbroek-Hutten MM, Hermens HJ. The relationship between objectively and subjectively measured activity levels in people with chronic low back pain. Clin Rehabil 2011 Mar 04;25(3):256-263. [doi: 10.1177/0269215510380828] [Medline: 20921030]
59. Lin CW, McAuley JH, Macedo L, Barnett DC, Smeets RJ, Verbunt JA. Relationship between physical activity and disability in low back pain: a systematic review and meta-analysis. Pain 2011 Mar;152(3):607-613. [doi: 10.1016/j.pain.2010.11.034] [Medline: 21251757]
60. van Rooij WM, Senden R, Heyligers IC, Cuppen PM, van Hemert WL, Grimm B. Physical functioning of low back pain patients: perceived physical functioning and functional capacity, but not physical activity is affected. Disabil Rehabil 2015 Mar 04;37(24):2257-2263. [doi: 10.3109/09638288.2015.1019010] [Medline: 25738908]
61. Gosetto L, Ehrler F, Falquet G. Personalization dimensions for mHealth to improve behavior change: a scoping review. Stud Health Technol Inform 2020 Nov 23;275:77-81. [doi: 10.3233/SHTI200698] [Medline: 33227744]
62. Xu L, Shi H, Shen M, Ni Y, Zhang X, Pang Y, et al. The effects of mHealth-based gamification interventions on participation in physical activity: systematic review. JMIR Mhealth Uhealth 2022 Feb 03;10(2):e27794 [FREE Full text] [doi: 10.2196/27794] [Medline: 35113034]
63. Davis TL, DiClemente R, Prietula M. Taking mHealth forward: examining the core characteristics. JMIR Mhealth Uhealth 2016 Aug 10;4(3):e97 [FREE Full text] [doi: 10.2196/mhealth.5659] [Medline: 27511612]
64. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of eHealth interventions: systematic review of the literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
65. Vitali D, Olugbade T, Eccleston C, Keogh E, Bianchi-Berthouze N, de C Williams AC. Sensing behavior change in chronic pain: a scoping review of sensor technology for use in daily life. Pain 2024 Jun 01;165(6):1348-1360. [doi: 10.1097/j.pain.0000000000003134] [Medline: 38258888]
66. Svendsen MJ, Nicholl BI, Mair FS, Wood K, Rasmussen CD, Stochkendahl MJ. One size does not fit all: participants' experiences of the selfBACK app to support self-management of low back pain-a qualitative interview study. Chiropr Man Therap 2022 Oct 03;30(1):41 [FREE Full text] [doi: 10.1186/s12998-022-00452-2] [Medline: 36192724]
67. Ellingson LD, Lansing JE, DeShaw KJ, Peyer KL, Bai Y, Perez M, et al. Evaluating motivational interviewing and habit formation to enhance the effect of activity trackers on healthy adults' activity levels: randomized intervention. JMIR Mhealth Uhealth 2019 Feb 14;7(2):e10988 [FREE Full text] [doi: 10.2196/10988] [Medline: 30762582]
68. Ellingson LD, Lansing JE, Perez ML, DeShaw KJ, Meyer JD, Welk GJ. Facilitated health coaching improves activity level and chronic low back pain symptoms. Transl J ACSM 2022 Apr 24;7(2):e000192. [doi: 10.1249/tjx.0000000000000192]
69. Michie S, Johnston M. Theories and techniques of behaviour change: developing a cumulative science of behaviour change. Health Psychol Rev 2012 Mar;6(1):1-6. [doi: 10.1080/17437199.2012.654964]
70. Eisele A, Schagg D, Krämer LV, Bengel J, Göhner W. Behaviour change techniques applied in interventions to enhance physical activity adherence in patients with chronic musculoskeletal conditions: a systematic review and meta-analysis. Patient Educ Couns 2019 Jan;102(1):25-36. [doi: 10.1016/j.pec.2018.09.018] [Medline: 30279029]

Abbreviations

CLBP: chronic low back pain
mHealth: mobile health
UCD: user-centered design

Edited by A Kushniruk, E Borycki; submitted 01.05.24; peer-reviewed by JB Lechauve, C Sitges; comments to author 04.08.24; revised version received 30.08.24; accepted 13.09.24; published 07.11.24.

Please cite as:
Berger M, Deblock-Bellamy A, Chèze L, Robert T, Desrosiers JJ, Christe G, Bertrand AM
Exploring the Needs of People With Chronic Low Back Pain and Health Care Professionals for mHealth Devices to Support Self-Managed Physical Activity and Pain: User-Centered Design Approach
JMIR Hum Factors 2024;11:e59897
URL: https://humanfactors.jmir.org/2024/1/e59897
doi:10.2196/59897
PMID:

Original Paper

An Exercise-Based Precision Medicine Tool and Smartphone App for Managing Achilles Tendinopathy (the 'PhysViz' System): User-Centered Development Study

Kohle Merry[1], MASc, PEng; Megan M MacPherson[2], OT, PhD; Jackie L Whittaker[1,3], PT, PhD; Christopher Napier[4], PT, PhD; Liisa Holsti[5], OT, PhD; Alex Scott[1], PT, PhD

[1]Department of Physical Therapy, University of British Columbia, Vancouver, BC, Canada

[2]Virtual Health Department, Fraser Health Authority, Surrey, BC, Canada

[3]Arthritis Research Canada, Vancouver, BC, Canada

[4]Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, BC, Canada

[5]Department of Occupational Science and Occupational Therapy, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Kohle Merry, MASc, PEng
Department of Physical Therapy
University of British Columbia
212 Friedman Building
2177 Wesbrook Mall
Vancouver, BC, V6T 1Z3
Canada
Phone: 1 (604) 822 8225
Email: kohle.merry@hiphealth.ca

Abstract

Background: People with Achilles tendinopathy (AT) experience persistent pain that can limit engagement with daily occupations and negatively impact mental health. Current therapeutic exercise approaches vary in success, with many people experiencing reinjury, leading to a cycle of chronic tendinopathy often lasting years. High-magnitude precision loading may help people exit this feedback cycle, but applying these principles clinically is challenging.

Objective: This user-centered design case study aims to provide an overview on how the PhysViz (a prototype for a novel remote rehabilitation intervention for AT management) was developed and evaluated following the development phase of the Framework for Accelerated and Systematic Technology-Based Intervention Development and Evaluation Research (FASTER).

Methods: The development process engaged a multidisciplinary team comprising people with AT experiences, clinicians, and engineers. It followed the 5 stages within the FASTER development phase: empathize, define, ideate, prototype, and test. The PhysViz development and evaluation were informed by needs assessments, surveys, literature reviews, validation studies, case studies, roundtable discussions, and usability testing (some of which have been published previously). The FASTER systematically guided the integration of evidence-based features and behavior change theory.

Results: By using the FASTER and ensuring that the PhysViz system was underpinned by diverse stakeholder needs, this work resulted in the development of a working prototype for both the PhysViz physical exercise tool and the accompanying PhysViz software package (mobile app and web application). A variety of study designs informed user-desired features that were integrated into the PhysViz prototype, including real-time biofeedback in the form of precision load monitoring, customizable exercise programs, and pain tracking. In addition, clinicians can visualize client data longitudinally and make changes to client exercise prescriptions remotely based on objective data. The identified areas for improvement, such as upgrading the user interface and user experience and expanding clinical applications, provide valuable insights for future PhysViz iterations. Further research is warranted to assess the long-term efficacy and feasibility of the PhysViz in diverse clinical settings and its potential to improve AT symptoms.

Conclusions: Being one of the first technology development initiatives guided by the FASTER, this study exemplifies a systematic and multidisciplinary approach to creating a remote rehabilitation intervention. By incorporating stakeholder feedback and evidence-based features, the PhysViz addresses key challenges in AT rehabilitation, offering a novel solution for precision loading and therapeutic exercise engagement. Positive feedback from users and clinicians underscores the potential impact of the PhysViz in improving AT management outcomes. The PhysViz serves as a model for technology-based intervention development, with potential implications for other tendinopathies and remote rehabilitation strategies.

(JMIR Hum Factors 2024;11:e57873) doi:10.2196/57873

KEYWORDS

exercise therapy; physical therapy modalities; rehabilitation; tendons; tendinopathy; mobile health; mHealth; mobile phone

Introduction

Background

Achilles tendinopathy (AT)—a chronic condition characterized by ongoing pain in the Achilles tendon and loss of function related to mechanical loading [1]—may limit participation in daily activities while negatively impacting mental health and reducing quality of life [2]. Although therapeutic exercise is the cornerstone of nonsurgical AT management [3-5], success varies [6-9], and rates of reinjury are high (eg, up to 27% in elite athletes) [10]. Tendinopathy may be recalcitrant to treatment [11,12] due, in part, to inappropriate treatment selection, inadequate dosing of therapeutic exercises, or psychosocial or contextual factors that are barriers to therapeutic exercise adherence [13,14]. Taken together, these facts highlight the need for research to optimize exercise-based AT treatment protocols and methods to personalize and deliver these protocols to promote adherence to therapeutic exercise and improve clinical outcomes.

Therapeutic exercise using high-magnitude loading (ie, >70% of maximal voluntary contraction [MVC]) can elicit positive adaptation in healthy tendons [15,16]. The extent to which these findings transfer to tendinopathic tendons remains largely unknown [17]. An individualized dose of 70% MVC exercise is higher than what is typically used in AT rehabilitation [6,18-20], and clinicians such as physical therapists (PTs) may struggle to achieve both this load magnitude and precision in the clinic. Although a 70% MVC prescription may be achieved clinically by basing an exercise load on a patient’s 1-repetition maximum, remote monitoring systems may improve access to adjustable high-magnitude precision loading and promote exercise adherence.

Remote monitoring systems can improve access to care, reduce the need for in-person follow-up and associated costs, and increase patient engagement with self-monitoring and built-in opportunities for clinician feedback [21,22]. As pain behavior is 1 method clinicians use to moderate AT exercise programs [6,18-20] (eg, instructions to not exceed a certain rating on a 10-point numerical pain rating scale during therapeutic exercises), home-based precision exercise dosing incorporating biofeedback and remote clinician monitoring could provide patients reassurance that their pain is acceptable and that they are operating in the correct loading range.

Existing dosing feedback systems are not yet practical or scalable due to prohibitive costs, space requirements (ie, isokinetic dynamometers), functional limitations (eg, positioning using handheld dynamometry as well as appropriate biofeedback and user interface [UI] and user experience [UX]) [23,24], and a lack of tailoring for home use by nonprofessionals. Given these limitations, and within the scope of AT rehabilitation, the development of a new system to facilitate precision dose–based therapeutic exercise is warranted. Knowing that the magnitude and precision of loading are important factors for inducing positive tendon adaptation [15,16], a home-based exercise-dosing system incorporating biofeedback and remote monitoring could potentially improve outcomes for those with AT. This paper provides a high-level overview of the development of 1 such system using the Framework for Accelerated and Systematic Technology-Based Intervention Development and Evaluation Research (FASTER) [25].

Prior Work

The concept for a home-based training system to enable high-magnitude loading of the triceps surae (ie, the gastrocnemius, soleus, and plantaris muscles, which act through the Achilles tendon to plantarflex the foot) comes from researchers at the Institute of Sport Sciences at the Humboldt University of Berlin [26]. In 2020, the Tendon Injury Prevention and Rehabilitation Group at the University of British Columbia consulted with the Humboldt group and obtained permission to adapt the Humboldt researchers’ idea for the development of an AT remote rehabilitation system, called the 'PhysViz'.

The PhysViz comprises a physical exercise tool incorporating a Bluetooth-enabled load sensor and the PhysViz software package consisting of a patient-facing mobile phone app and clinician-facing web application. The purpose of the PhysViz is to facilitate high-magnitude loading of the Achilles tendon, collect loading data, provide real-time biofeedback during loading to promote exercise engagement, centralize all patient data for review, and enable the remote modification of a patient’s exercise prescription by an overseeing clinician. The goal of the system is to enhance AT rehabilitation through high-load precision exercise–dosing approaches and empower active participation of patients in their rehabilitation through the dissemination of actionable information to both patients and clinicians.

The Goal of This Study

Designing a new system to facilitate precision dose–based therapeutic exercise in an evidence-based and user-informed manner is challenging. Traditional design often prioritizes time to market, potentially at the expense of rigor [27]. Rigorous, evidence-based, user-centric development takes time; however, such a time delay may lead to technologies becoming obsolete, implementation conditions changing, or new interventions emerging [25]. Unfortunately, a significant portion of medical research funding (up to 80%) ends up as “research waste,” failing to make a notable impact on public health [28]. In addition, 50% of clinical innovations never reach widespread clinical adoption [29], and the translation of research-based solutions into clinical practice can take upwards of 17 years [30]. As such, technology-based clinical interventions must strategically develop solutions and generate supporting evidence in a timely manner [25]. This work aims to describe the approach taken to develop the PhysViz, which features user-centric development and theoretical underpinning. This manuscript not only presents the systematic approach taken in developing the PhysViz system but also highlights both previously published findings (primarily in the empathize stage within the FASTER development phase) and new findings (primarily in the prototype and test stages).

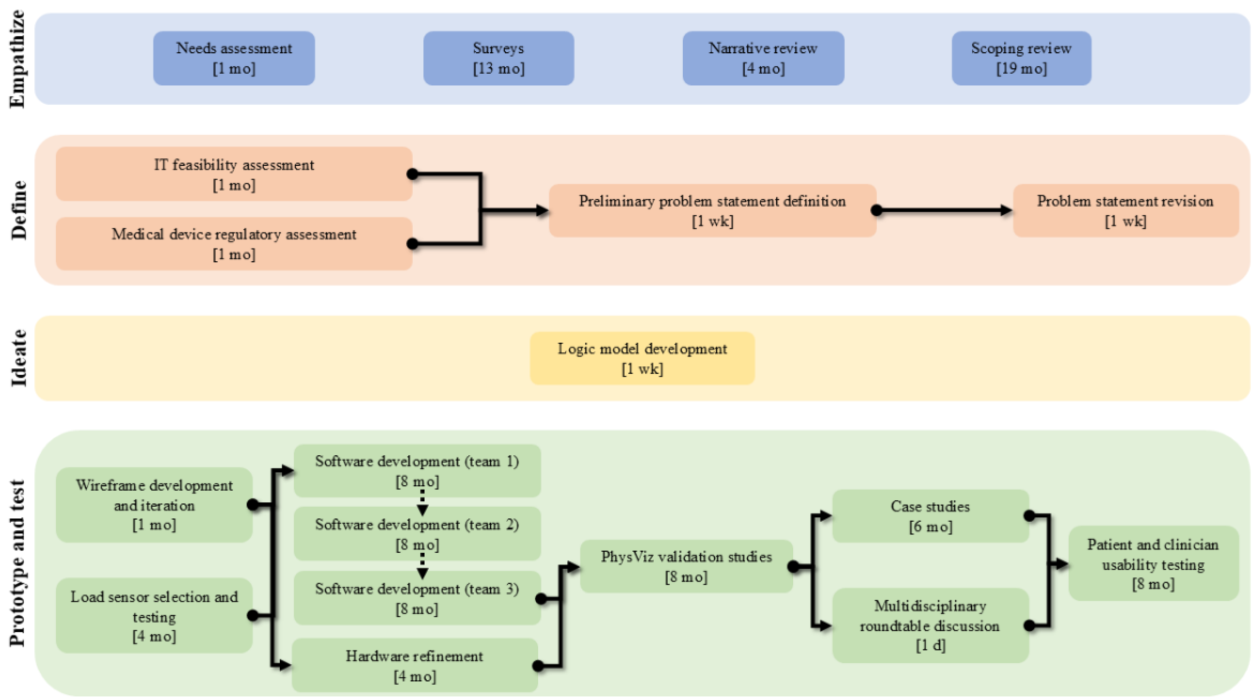
Methods

Design Approach

To promote the timely and strategic generation of supporting evidence, the PhysViz development is guided by the FASTER [25]. The FASTER includes 3 phases: development, progressive usability and feasibility evaluation, and scaled evaluation and implementation. This work is situated within the development phase, which aims to engage with end users and use transdisciplinary thinking (eg, engineering design and clinical experience) to generate actionable evidence appropriate for the efficient design and implementation of technology-based interventions. Specifically, within this phase, the FASTER underscores the importance of empathizing with users to understand their needs, preferences, and experiences while also using established design theories to generate ideas, develop prototypes, test them, and iterate. The FASTER development phase can be further broken down into 5 stages derived from Stanford Design Thinking: empathize, define, ideate, prototype, and test [31].

This work describes 1 of the first case studies applying the FASTER for the development of a remote rehabilitation system. An overview of the design approach and timeline is presented in Figure 1. By integrating previously published studies with new research, this manuscript offers a comprehensive perspective on the PhysViz development process guided by the FASTER. It is important to note that this manuscript offers a broad overview of how various research studies contributed to the PhysViz design through the FASTER development phase. It does not provide a detailed account of the methods and results of each individual study involved in the PhysViz development. To ensure transparency and facilitate replication of the development process, studies that have not been previously published in peer-reviewed journals have been deposited on the Open Science Framework website [32-34].

Figure 1. PhysViz development timeline according to the development phase of the Framework for Accelerated and Systematic Technology-Based Intervention Development and Evaluation Research. The arrows denote the progression of tasks. When arrows are not present, this denotes tasks completed concurrently. Development broadly progressed from the empathize stage to the prototype and test stages.



Stages of the FASTER Development Phase

Empathize

Within the FASTER empathize stage, researchers should determine end-user needs and preferences while also examining the broader context to define the research problem in detail [25]. As such, this stage consists of a needs assessment and surveys to garner information on end-user needs, while also including both narrative and scoping reviews to gain a broader understanding of exercise-based AT rehabilitation.

Needs Assessment

To assess end-user needs for the patient-facing PhysViz mobile app, our research team completed a needs assessment featuring clinicians, engineers, patient partners, and researchers. Due to the transdisciplinary nature of developing a software tool for home-based use by patients, it was vital to get multiple perspectives to establish the initial direction of the app development. Specifically, the research team was asked the following in reference to the PhysViz app: (1) what features they must see; (2) what features they would like to see; (3) what information they would like tracked; (4) what they would like the UI to look like; (5) whether they expected to see an overview or summary dashboard, and if so, what information they would like displayed there; (6) what type of security protection they expected to see; and (7) whether they had any other comments or relevant information pertaining to the development. All suggestions were collated and presented back to the research team for communal ranking before starting the development process.

Surveys

As described in Merry et al [35], to further identify end-user needs and preferences within the specific implementation context, we conducted 2 surveys to identify perceived barriers and facilitators to participating in and prescribing exercise-based therapies for AT among people with AT and PTs, respectively. While the PT survey assessed both clinical practice patterns and barriers and facilitators to developing, prescribing, and monitoring exercise-based treatments for AT, the survey of people with AT expanded upon the barriers and facilitators pertaining to engagement and adherence to exercise-based rehabilitation for AT through a series of questions using the capability, opportunity, motivation, and behavior (COM-B) model [36,37]. According to the COM-B model, for a behavior to occur, an individual must have sufficient *capability* (physical and psychological), *opportunity* (social and physical), and *motivation* (automatic and reflective). In addition, we asked participants what potential app features are desirable in the context of AT rehabilitation.

Narrative and Scoping Reviews

As described in Merry et al [17], we completed a narrative review to provide an overview of the broader AT context by (1) synthesizing the principles of tendon remodeling under resistance exercise–induced loading among both healthy and pathological tendons and (2) commenting on the biomechanical principles of Achilles tendon loading that may impact a therapeutic exercise prescription for AT. This work not only served to identify research gaps and potential future use cases for the PhysViz, but it was also undertaken to inform aspects of the PhysViz development based on prior evidence.

As described in Merry et al [32,38] and in parallel to the narrative review, we conducted a scoping review to synthesize how current resistance exercise–based AT interventions are being designed and implemented clinically. Given that therapeutic exercise is considered a primary clinical management strategy for AT [3,4], this review sought to understand what aspects of therapeutic exercise may be common across the clinical research evidence (in contrast to the largely mechanistic evidence encompassed in the narrative review) and how these aspects could potentially be built into the PhysViz design [32,38].

Define

In the define stage, information from the empathize stage is used to specify the problem in detail, including its location, the potential users and stakeholders involved, and any other pertinent details. Using both the contextual and research evidence identified in the empathize stage, we defined a preliminary problem statement. The problem statement was then refined after a medical device regulatory assessment and an IT feasibility assessment.

Ideate

The ideate stage then focuses on the generation of ideas and the exploration of possible solutions to the problem. Specifically, we developed a logic model for the PhysViz based on the therapeutic exercise treatment principles (both mechanistic and clinical) identified in the empathize stage coupled with the barriers and facilitators to participating in and prescribing therapeutic exercise for AT identified using the COM-B model. The purpose of developing the logic model was to link potential PhysViz features to desirable behavioral and treatment outcomes a priori; in this way, features could be strategically developed based on evidence and proposed treatment pathways rather than on personal opinions or “it sounded like a good idea at the time” logic.

Prototype and Test

Overview

The prototype stage describes the creation of low-fidelity prototypes using select ideas from the ideate stage to help the team members visualize and communicate their ideas. Select prototypes are then advanced to the test stage where feedback from end users and stakeholders is garnered to evaluate the feasibility of the design solution. The prototype and test stages are critical for ensuring that the team’s design solution is feasible and meets the needs of the end users. The iterative nature of these stages allows for continual improvements and adjustments, ultimately leading to the timely development of a more appropriate and user-friendly final product.

Using the logic models as a feature guide, we developed a series of wireframe designs for the PhysViz patient-facing mobile app and clinician-facing web application. The wireframes were then presented to the research team for review and iteration before beginning the software development. We completed the development of the PhysViz software consisting of both a mobile app and a web application in collaboration with 3 teams of senior computer engineering students from the University of British Columbia. Each team, consisting of 4 to 5 students, completed 8 months of development before passing the development to the next team. Building upon the initial design developed by the Humboldt group [26] and given the incorporation of a Bluetooth-enabled load sensor in the PhysViz design, the first PhysViz physical exercise tool was tested by the research team to assess viability and potential modifications based on learnings from the empathize stage.

Load Sensor Selection and Testing

As described in Merry et al [39], a fundamental aspect of the PhysViz design was the inclusion of a Bluetooth-enabled load sensor enabling mobile phone connectivity with the physical exercise tool. To start, we searched for an appropriate load sensor to meet the following criteria: (1) the product must be Bluetooth ready, (2) both tensile and compressive sensors were considered, (3) there were no restrictions on price, and (4) high-load tensile load sensors (eg, crane scales) were not considered. We reviewed 11 load sensors and selected 2 for more rigorous testing based on relevant technical specifications. Advanced testing consisted of comparing each load cell's output against a gold standard mechanical testing machine to assess validity and reliability [39]. After the completion of this study, a single load sensor was selected for inclusion in the physical exercise tool.

PhysViz Validation

In conjunction with work undertaken by Pratt et al [40], with an initial prototype of the PhysViz system complete and before using it for AT rehabilitation purposes, it was important to determine whether the system could consistently and accurately measure in vivo muscle strength compared to computerized dynamometry, which is the gold standard for assessing muscle function; for example, if a particular exercise prescription was meant to be based on an individual's MVC, was the MVC recorded using the PhysViz valid and reliable, thus promoting safe use among individuals who were symptomatic? We completed a validation study with healthy individuals to establish the efficacy of the PhysViz as a potential home-based exercise tool [40].

In conjunction with work undertaken by Schreiber et al [33], given the fundamental roles of precision load management and objective feedback (via the app) in the PhysViz system, it was also important to quantify whether real-time biofeedback has the potential to enhance the execution of AT rehabilitation exercises; for example, if an exercise program prescribed isometric plantar flexion exercises at 70% of an individual's MVC, would the biofeedback provided by the PhysViz to the individual improve their ability to execute the exercises at the prescribed load? We completed a cross-sectional study with healthy individuals to study the effect of biofeedback within this context [33].

Case Studies

Having completed basic validation of the PhysViz system, we conducted three 12-week case studies among people with current AT symptoms under the oversight of a clinician-scientist. Case studies began with an introductory session where we gave each participant basic instructions on how to use the PhysViz. We then tested the participant's plantarflexion MVC using the PhysViz, which informed the first therapeutic exercise dose intensity (70% of MVC). We asked each participant to exercise 2 to 3 times a week using the "exercise mode" of the PhysViz, which guides users through a therapeutic exercise session using audiovisual biofeedback. The research clinician checked in with participants approximately every 3 weeks and adjusted the therapeutic exercise dose based on participant tolerability. At the end of the 12 weeks, we held a semistructured debrief interview with each participant to discuss intervention acceptability, ergonomics, UI design, and comfort.

Multidisciplinary Roundtable Discussion

In parallel to the case studies, we assembled a stakeholder group for an open discussion to (1) refine the design through communal knowledge with a focus on proposals for improvement (customization and flexibility) and utility, (2) consider potential feature additions identified during the survey, and (3) comment on potential use cases beyond AT. The stakeholder group consisted of the research team in addition to external clinician-scientists, practicing clinicians (PTs, occupational therapists, and physicians), people with lived AT experiences, engineers, and insurers or regulators.

Patient and Clinician Usability Testing

As described in Merry et al [34], to obtain further user feedback on the PhysViz prototype, we completed a cross-sectional usability study including both people with AT and PTs. We used semistructured 1-on-1 interviews with embedded usability testing featuring a concurrent think-aloud method [41] to inform iteration on the PhysViz system in conjunction with the case studies and roundtable discussion. Combined results from the usability testing and concurrent think-aloud findings were used to create a list of potential design alterations based on common usability issues and end-user suggestions for improvement. The proposed modifications were then evaluated independently by 2 researchers (KM and MMM) using the affordability, practicability, effectiveness, acceptability, side effects and safety, and equity (APEASE) criteria [42]. Interrater agreement was assessed using Cohen's κ and prevalence-adjusted bias-adjusted κ. Values between 0.61 and 0.80 indicate substantial agreement, and values exceeding 0.80 indicate almost perfect agreement [43,44]. A critical appraisal of each proposed modification using the APEASE criteria facilitated the inclusion of only those modifications deemed most feasible for potential implementation in the second version of the PhysViz system.

Ethical Considerations

Ethics approval for this study was obtained from the University of British Columbia Clinical Research Ethics Board (approval number: H21-02879). Participants provided informed consent to participate in the study before taking part.

Results

Empathize

Needs Assessment

Five members of the research team completed the needs assessment, which led to establishing an initial direction for the PhysViz software development. Common "must-see" features included load monitoring and biofeedback, a customizable exercise program, app availability for both iOS and Android devices, and the ability to sync data between the mobile app and the web application. Some examples of "would be nice" features included instructions for participants (eg, video demonstrations or animations), reminders or notifications, a program calendar, and 2-way interaction capabilities between the patient and clinician. A desirable UI was defined by participants as simple and minimalistic, with an emphasis on ease of use for users of different ages and digital literacy levels.

User safety (people with AT) was noted as particularly important with a need to avoid placing any patient in a potentially harmful position by trying to complete therapeutic exercises exceeding recommended intensities.

Surveys

The survey results, which are published elsewhere [35], led to the following features being prioritized within the initial conceptualization of the PhysViz: exercise demonstrations, education on AT and treatment strategies, guided exercise sessions (ie, automatically counts sets, repetitions, and rest periods), and pop-up reminders for completing the exercise program.

Narrative and Scoping Reviews

The narrative review results, which are published elsewhere [17], highlighted the importance of high-load tolerability with the PhysViz design. The long sitting position was identified as advantageous to allow for ankle dorsiflexion and knee extension within the PhysViz design.

The findings from the scoping review, which are published elsewhere [38], emphasized the importance of modifiable exercise programs (both in terms of adjustable resistance and exercise volume) and the incorporation of techniques that allow for improved tracking of treatment fidelity and adherence within the PhysViz design.

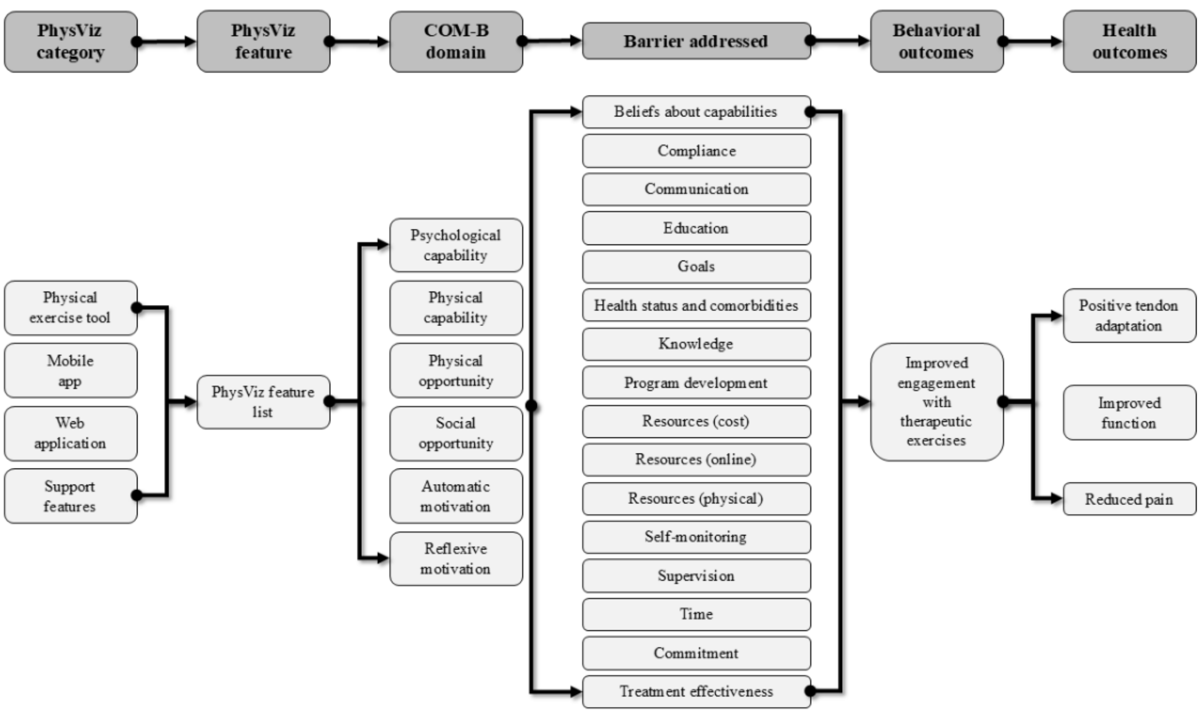
Define

Primary users of the PhysViz system are expected to be people with AT and PTs or other clinicians (eg, occupational therapists and physicians) who oversee and guide AT management. Secondary users will include clinical practice regulators, health care insurers, software engineers, and researchers. The initial problem statement was as follows: “People with chronic AT need a better way to deliver heavy loads to the Achilles tendon because not everyone has access to exercise equipment, and sufficient loading intensity appears to be important for tendon rehabilitation.” After regulatory and feasibility assessments, the problem statement was revised as follows: “People with chronic AT want to get back to their sports and daily occupations quickly but often struggle to adhere to therapeutic exercise programs and load the tendon heavily enough. Emerging exercise technologies have the potential to empower users, providing autonomy and objective feedback, thereby facilitating a quicker return to their activities.”

Ideate

We used findings from the empathize stage to inform the development of a logic model to address the revised problem statement and guide the initial development of the PhysViz system by linking potential evidence-based features with behavioral and treatment outcomes. Figure 2 presents a general logic model, while detailed logic models are presented in Multimedia Appendix 1.

Figure 2. PhysViz development general logic model. COM-B: capability, opportunity, motivation, and behavior.



Prototype and Test

Overview

On the basis of findings from the preceding stages and the operationalization of these findings in the logic model, we developed wireframe designs for both the PhysViz mobile app and web application (Figure 3 and Figure 4, respectively). The wireframe designs evolved based on circulation and iteration among the research team and were ultimately developed into working software packages for testing. In addition to the custom mobile app and web application, the physical exercise tool rounds out the PhysViz system architecture (Figure 5).

Figure 3. (A-C) PhysViz mobile app home page. (D-F) "Exercise mode" page with real-time biofeedback. (A and D) Initial wireframe. (B and E) First prototype. (C and F) Current prototype.

Figure 4. PhysViz web application sample exercise page. (A) Initial wireframe. (B) First prototype. (C) Current prototype.

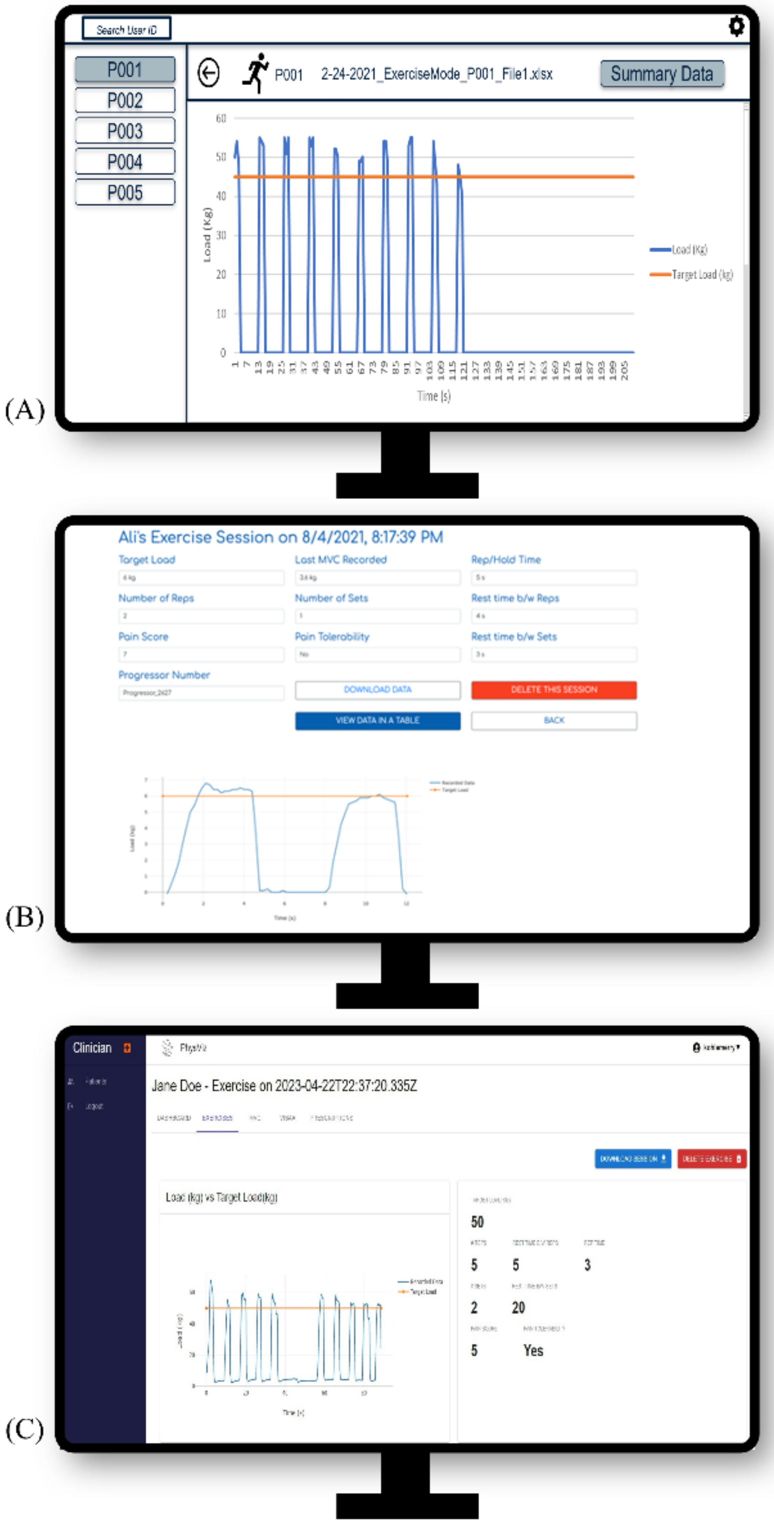
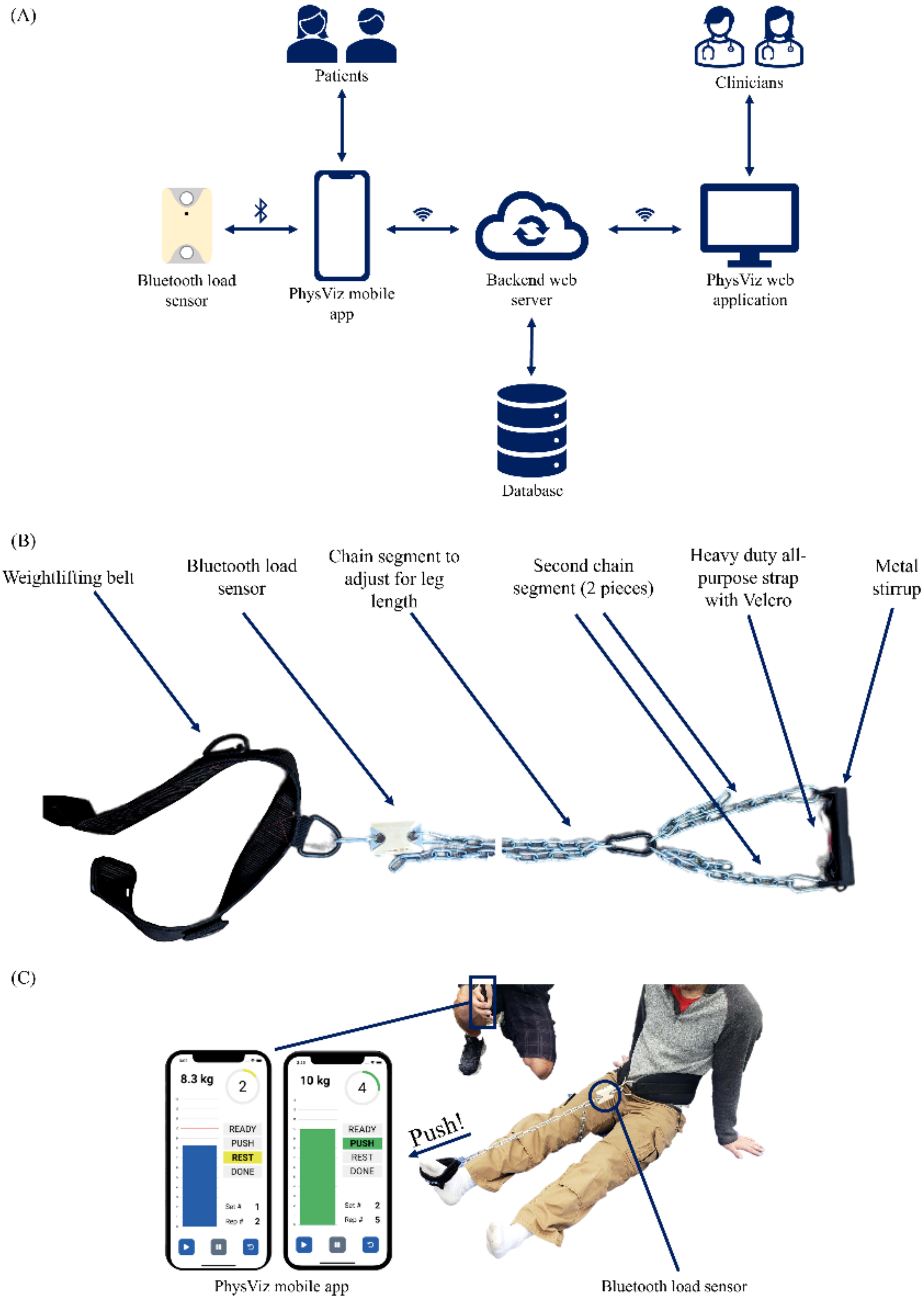


Figure 5. (A) System architecture diagram. (B) Current iteration of the PhysViz physical exercise tool. (C) Demonstration of intended system use with participant plantarflexing into the stirrup and imparting a load through the Achilles tendon. Participant simultaneously receives real-time biofeedback from the PhysViz mobile app.



Load Sensor Selection and Testing

We selected the Progressor load sensor (Tindeq) for inclusion in the PhysViz system due to its superior performance across its measurement range compared with the Activ5 load sensor (Activbody). Complete findings are provided in the study by Merry et al [39].

PhysViz Validation

The findings indicate that the PhysViz is a valid and reliable tool for isometric plantar flexor MVC measurement; complete findings are provided in the study by Pratt et al [40]. In addition, according to the findings presented in the study by Schreiber et al [33], PhysViz biofeedback was found to enhance the execution of AT therapeutic exercises by lessening the percentage difference between recorded and prescribed load across a brief exercise protocol (ie, 2 sets of 10 five-second repetitions) compared to a nonbiofeedback condition.

Case Studies

Over the 12 weeks, participant 1, a male individual aged 35 years experiencing unilateral AT for 5 months, completed 27 therapeutic exercise sessions using the PhysViz. In this time, participant 1's plantarflexion MVC on his affected side increased by 29% despite his pain score remaining unchanged.

Due to an additional sports-related injury impacting the affected side, participant 2, a male individual aged 40 years experiencing unilateral AT for 6 years, elected to discontinue the case study after 6 weeks. Over these 6 weeks, participant 2 completed 7 exercise sessions with the PhysViz. Only a single MVC test was recorded, precluding any inquiry into strength increases.

Participant 3, a female individual aged 57 years experiencing bilateral AT for 1 year, completed 8 exercise sessions with the PhysViz over the 12 weeks and increased her plantarflexion MVC by 20%. Participant 3's pain score and self-reported pain tolerability was variable, with no clear trend.

Overall, information from the interviews showed that the PhysViz system was well received by all 3 participants. Specifically, the participants identified that compared to standard treatment options (ie, heel drops or calf raises), the real-time biofeedback was particularly useful because it promoted engagement in the rehabilitation process through the visualization of load intensity and the pacing of repetitions. Nevertheless, all 3 participants noted lower extremity pain while using the system, particularly in the knee, calf, and Achilles tendon. Despite reporting that the pain was tolerable, they expressed a desire to refine the positioning to mitigate the pain and improve overall comfort during use. Finally, the participants generally felt that the UI of the mobile app was adequate and accessible, although simplistic in comparison to other apps. If the plan is to disseminate the system beyond a research-based audience, the participants suggested improving the UI and adding additional features to bolster the somewhat rudimentary UX (eg, brief animations, reminders to complete exercises, and an in-app calendar showing completed and upcoming exercise sessions).

Multidisciplinary Roundtable Discussion

Of the 18 people invited to the discussion, 10 (56%) attended, including the research team members, external clinician-scientists, practicing clinicians (PTs and occupational therapists), people with lived AT experiences, and engineers. Overall, the group identified the main value proposition of the PhysViz system as being able to deliver high-magnitude loading precisely, either in the clinic or at home. Given the importance of loading intensity when managing AT, the invitees suggested that the PhysViz offered a useful alternative for patients who may not have time to go to the gym or lack access to the gym equipment often required to deliver the high loads needed to rehabilitate the tendon. Furthermore, the group identified that both PTs and individuals with AT stood to benefit from using the PhysViz: PTs can individualize and monitor care better when they have objective measurements such as those provided by the PhysViz; and individuals with AT are more likely to adhere to, and subsequently benefit from, treatment when they are able to see objective progress.

The group did identify that the design of the PhysViz physical exercise tool could be improved; in particular, the group sought to reduce the system complexity to make it accessible for more potential users. Specifically, the invitees reiterated the importance of modifying the design while maintaining flexibility in the system such that it can still be individualized to suit an array of patient needs (eg, body size, flexibility, and preexisting injuries). Depending on the revised design of the physical exercise tool, the invitees also suggested that the PhysViz could be used to help manage other lower limb tendinopathies (eg, patellar tendinopathy) and could be used as a general dynamometer to help augment care.

Patient and Clinician Usability Testing

According to the findings presented in the study by Merry et al [34], 15 people with AT and 4 PTs participated in the usability testing study. During the usability testing across the 8 tasks completed with the physical exercise tool and mobile app, people with AT encountered an average of 12 (SD 3) total usability issues. Most of the issues (102/183, 55.7%) were "severity 2—minor problems" according to the scale developed by Nielsen [45] and were most commonly issues associated with the "meaning of labels" (47/183, 25.7%), the "understanding of system instructions/error messages" (37/183, 20.2%), and the "visibility of system status" (31/183, 16.9%) according to the coding scheme postulated by Kushniruk and Patel [46]. During the clinician usability testing of the 3 tasks completed using the web application, PTs encountered an average of 3 (SD 3) total usability issues. Most of the issues (7/13, 54%) were "severity 2—minor problems" and were most commonly issues associated with the "meaning of labels" (8/13, 62%) and "navigation" (3/13, 23%).

During the debrief, 73% (11/15) of the people with AT stated that the PhysViz system was worth improving, with an additional 20% (3/15) stating that it is worth improving, provided research demonstrates its efficacy. Most of the participants with AT (13/15, 87%) stated that they would be willing to use a system such as the PhysViz during their AT rehabilitation. When asked to rate how helpful the PhysViz would be in supporting them to achieve their AT rehabilitation goals on a scale ranging from 0=*not helpful at all* to 10=*extremely helpful*, the system received a mean score of 7.2 (SD 1.2). By contrast, 50% (2/4) of the clinician participants stated that the PhysViz system was worth improving, with the other 50% (2/4) suggesting that it might be worth improving, provided changes were made to accommodate other musculoskeletal conditions besides AT.

Using the APEASE criteria, 25 proposed modifications were considered, and of these, 16 (64%) were deemed feasible for implementation in the next version of the PhysViz system (ie, having met all APEASE criteria). Cohen's κ and prevalence-adjusted bias-adjusted κ values describing interrater agreement of the proposed PhysViz modifications were 0.39 (fair agreement) and 0.83 (almost perfect agreement), respectively.

Discussion

Principal Findings

Overview

This paper reports the formative development of the PhysViz, a novel remote rehabilitation system for people with AT using the FASTER. In addition to addressing the rationale and stakeholder needs underpinning the PhysViz, this project led to working prototypes of both the PhysViz physical exercise tool and the PhysViz software package that was tested by the potential end users (ie, both people with AT and PTs). Taken together, we think the development process outlined in this work represents a meaningful addition to the digital health landscape by providing a case study in systematic technology-based intervention development.

PhysViz development was framed within the development stage of the FASTER [25]. By comprehensively reviewing the literature and by engaging stakeholders through cross-sectional surveys and usability testing studies, this work leverages prior research knowledge, clinical insights and practice strategies, and patient-identified needs to improve the chances of intervention efficacy and uptake.

Empathize

The needs assessment, survey results [35], narrative review [17], and scoping review [38] collectively revealed critical insights into user expectations and preferences. Significantly, information was woven together from (1) research team members, (2) external stakeholders to whom we appealed directly, and (3) previous research, to inform system development in a robust way; for example, users expressed a demand for a versatile app compatible with both iOS and Android devices, with data-syncing capabilities for remote clinician monitoring. The emphasis on a simple and minimalistic UI aligns with the goal of inclusivity across different age groups and digital literacy levels. The incorporation of high-load tolerability and the use of the long sitting position in the design addressed specific biomechanical considerations highlighted in the narrative and scoping reviews.

Define and Ideate

The problem statement identified in the define stage highlighted the challenge of enabling individuals with AT to return to their activities quickly, emphasizing the importance of autonomy and objective feedback. The importance of autonomy and objective feedback identified aligns with past research [47], which found that personalized feedback can improve patient outcomes. This led to the formulation of a logic model, linking evidence-based features with behavioral and treatment outcomes. Taken together, both stages allowed the research team to critically reflect on the value proposition of the PhysViz system from the perspective of multiple end users, where it may fit along the treatment pathway, and how it may integrate into existing clinical workflows. Furthermore, mapping how different PhysViz system categories (eg, physical exercise tool and mobile app) could potentially address current treatment barriers helped mitigate potential design pitfalls such as unnecessary complexity and redundancy. Mapping also helped to delineate what features were needed in the PhysViz technologies (ie, physical exercise tool, mobile app, and web application) and what could be considered intervention support features (eg, the exercise protocol completed using the PhysViz, user training, and support for use).

Prototype and Test

Overview

Complexity was gradually built into the PhysViz system over months of development with the computer engineering student teams; varying levels of prototype fidelity were used to mock-up potential features and prove concepts in a timely manner. The resulting PhysViz system prototype and more specifically the UI and UX were then holistically evaluated toward the end of the prototype and test stages. Several features built into the PhysViz system prototype emerged directly from gaps in the literature and current standard care practices.

PhysViz UI and UX

The UI and UX of the PhysViz, tailored to address gaps and preferences identified in the emphasize stage, mirror recommendations from past literature that advocate for simplicity and intuitive design [48,49]. The needs assessment results underscored the importance of a simple and minimalistic UI, with a focus on ease of use for users across various age groups and digital literacy levels. Informed by the needs assessment and by the literature [45,50-53], we strategically designed the app and web portal with the goal of having simple screens and buttons as well as a predictable structure.

Moreover, the scoping review highlighted a significant gap in reporting related to adherence, compliance, and fidelity to prescribed exercises in AT interventions [38]. This poor reporting of adherence, compliance, and fidelity is also reflected in the existing literature [54,55], highlighting a need for improved monitoring systems. The UI and UX of the PhysViz addresses this gap by incorporating features for automated tracking of exercise sessions (ie, there is no “save/upload data” button present; the system does this automatically), accompanied by intuitive visual cues and real-time biofeedback in the form of repetition timing and load intensity monitoring. The various forms of exercise cueing and data visualization aim to empower users to increase their adherence, fostering a more engaging and motivating rehabilitation experience. The case studies underscored participants’ positive experiences with the PhysViz UI and UX, particularly appreciating aspects such as the color change effect when “target load” intensity was reached, as well as the pacing of repetitions. These results are further supported by previous literature, which highlights the utility of biofeedback to enhance user engagement and adherence [56,57].

During usability testing, most of the usability issues encountered by people with AT related to the “meaning of labels” (47/183, 25.7%), “understanding of system instructions” (37/183, 20.2%), and “visibility of system status” (31/183, 16.9%) [34]. When assessing the web application, most of the usability issues PTs faced were associated with the “meaning of labels” (8/13, 62%) and “navigation” (3/13, 23%). Usability testing results such as these will help inform further UI and UX iteration to improve the PhysViz by specifically identifying problematic features or areas according to end users.

In the case studies, participants generally perceived the UI of the mobile app as adequate and accessible, although they acknowledged its simplicity compared to other apps. Notably, suggestions emerged for enhancing the UI, particularly if the goal is to extend the reach of the PhysViz system beyond a research-based audience. Recommendations included incorporating brief animations, reminders for completing exercises, and an in-app calendar displaying completed and upcoming exercise sessions, aligning with behavior change literature [58,59]. These suggestions, if implemented, could address the reported simplicity of the UI and enhance the overall UX, making the PhysViz more accessible to a wider audience.

Biofeedback

Biofeedback emerged as a promising feature with the potential to address several gaps in existing treatment strategies. The need for precision loading, a key consideration highlighted in the narrative review [17], is directly addressed through real-time biofeedback in the form of load intensity monitoring, allowing users to achieve and maintain high-magnitude loading with greater accuracy. The results from the validation studies substantiate this claim, demonstrating that the PhysViz system is a valid and reliable tool for isometric plantar flexor MVC measurement and that biofeedback is a useful feature for promoting the accurate execution of exercise parameters such as load intensity [33,40]. Given the lack of reporting on exercise fidelity identified in the scoping review, biofeedback and the automated tracking of exercise sessions incorporated in the PhysViz may enhance program fidelity for individuals with AT while also facilitating transparent and automated reporting for clinicians.

The importance of transparent reporting was also identified in the surveys, which revealed a perceived lack of patient compliance among PTs [35]. Biofeedback has the potential to bridge this gap by supplying both people with AT and clinicians with objective, measurable data, fostering a collaborative approach to rehabilitation and improving overall treatment compliance. In the case studies, participants commended the real-time biofeedback provided by the PhysViz, citing its utility in promoting engagement through the visualization of load intensity and the pacing of repetitions. Moreover, insights from the multidisciplinary roundtable discussion, including perspectives from clinicians and individuals with AT, underscored the value proposition of the PhysViz in delivering high-magnitude loading precisely, thus enhancing its potential to address compliance challenges. Usability testing further supported the positive impact of biofeedback, with individuals expressing willingness to use the PhysViz during their rehabilitation, highlighting its potential to improve overall treatment compliance [34].

This multifaceted evidence stemming from both sets of end users surveyed [35] and existing literature [17,38] highlights the potential for precision loading and the need for improved therapeutic exercise engagement among people with AT. The findings were substantiated during the prototype and test stages of this work, including the validation studies [33,40], case studies, roundtable discussions, and usability testing [34]. Together, this body of evidence supports the notion that the biofeedback offered by the PhysViz serves as a valuable instrument for strengthening collaborative rehabilitation initiatives and promoting increased fidelity and, potentially, engagement with therapeutic exercises for AT rehabilitation.

Comparison to Previous Work

This work aligns with and builds upon existing literature regarding telehealth development, particularly within the context of human-centered design for health care interventions. Emerging research suggests that integrating human-centered design into telehealth research results in more usable, acceptable, and effective health care interventions [60,61]; however, there is significant underreporting and a lack of guidance in the development of sustainable health care apps [62], and much of the telehealth app research focuses more on laboratory and field evaluations than on the design and development phases [63].

Our work addresses these gaps by emphasizing the early stages of the development process, specifically the empathize and define stages within the FASTER, where thorough needs assessments and stakeholder engagement initiatives were conducted. This approach allowed us to prioritize end-user needs and design features that are both practical and meaningful for the target population. By focusing on the development phase, we ensured that the PhysViz system was not only functional but also aligned with the preferences and constraints of potential end users. In addition, our iterative development process, which included usability testing and multidisciplinary discussions, contributed to creating a more sustainable telehealth solution that can adapt to the evolving needs of its end users. This proactive approach contrasts with much of the existing telehealth research, which often underrepresents the importance of early-stage design and development in favor of later-stage evaluations [63].

A recent review of telehealth app development highlighted that lengthy development times and the use of limited qualitative research methods are often cited as constraints associated with the long-term efficacy of the apps [63]. Telehealth app development frequently uses interviews and focus groups [63], and while these methods are effective for gathering direct feedback from end users, they have been criticized for their limited ability to elicit tacit knowledge [64]. While our work did use these methods, we also relied on multiple smaller-scale and time-efficient studies, using diverse methods to capture both explicit insights (through the surveys and multidisciplinary roundtable discussion) and tacit knowledge (through case studies and usability testing). This mixed methods approach not only allowed us to gather comprehensive data on end-user needs and preferences but also facilitated the iterative refinement of the PhysViz system, ensuring that it was both user centered and responsive to the complexities of real-world use. By integrating feedback from a variety of stakeholders—patients, clinicians, and engineers—we were able to design a telerehabilitation solution that is more likely to be effective, sustainable, and widely accepted in clinical practice. This approach highlights the importance of combining traditional qualitative methods with more agile techniques to enhance the development and adoption of telehealth technologies.

While telehealth apps are often developed with the inclusion of end users, this engagement tends to be limited to a single user group such as patients or clinicians [65]. By contrast, our work involved multiple end-user groups throughout the development process, including people with AT as well as clinicians and stakeholders responsible for the implementation and dissemination of the PhysViz. This comprehensive approach allowed for a more holistic understanding of the diverse needs, preferences, and challenges associated with both the use and the integration of the system. By engaging these varied perspectives, we were able to design a more adaptable and user-friendly telehealth solution that not only meets the therapeutic needs of patients but also aligns with the practical and logistical requirements of clinicians and health care systems. This multistakeholder engagement is critical in ensuring that the resulting technology is effective, sustainable, and capable of being seamlessly integrated into routine clinical practice.

Strengths and Implications

This work represents one of the first uses of the FASTER development phase in developing a remote rehabilitation intervention (including a physical exercise tool, a mobile app, and a web application). The collaborative and multidisciplinary nature of this work is a strength, particularly in integrating perspectives from people with the condition of interest and clinician partners throughout the timely and rigorous development of the PhysViz system prototype. Involving a diverse team, including those with lived experiences, has not only enriched the development process but also ensured that the resulting intervention is attuned to the real-world needs and preferences of end users. This inclusive approach enhances the credibility and applicability of the system and serves as a model for future technology-based rehabilitation development efforts.

Another notable strength lies in the diversity of study types used. Tailoring study designs to suit development needs is a core feature of the FASTER, which suggests a variety of methods for informing and evaluating the proposed intervention. By incorporating narrative and scoping reviews, case studies, and both qualitative and quantitative data, this research amalgamation effectively addresses the limitations inherent in a single-study approach. This comprehensive methodological approach allows for a more nuanced understanding of the intervention's usability according to end users and potential effectiveness in managing AT through the mobilization of high-magnitude precision loading. It also contributes to a more holistic assessment of the PhysViz system, capturing a range of perspectives and insights that might be overlooked in a single-study design.

Furthermore, the systematic identification of theory-based techniques, such as the integration of the COM-B model, adds a layer of robustness to the intervention development process. By aligning the PhysViz system with behavior change theory, we enhance our understanding of the underlying mechanisms and pave the way for more targeted and potentially effective interventions. This methodological rigor ensures that the developed intervention is technologically sound and theoretically grounded, increasing its potential for successful implementation in real-world rehabilitation settings.

Future Directions

The positive reception from users and clinicians underscores the potential impact of the PhysViz system; however, the identified areas for improvement, such as improving consistency and intuitive navigation, should guide future iterations. More specifically, mobile app clarity can be improved by adding a designated "Home" button within the app UI to anchor users, adding an in-app back button, and avoiding potentially confusing words such as "tare." Further efforts seem necessary to enhance the UI and UX of the biofeedback interface to present users with crucial information while preventing information overload. Furthermore, the potential extension of the PhysViz to manage other tendinopathies and its use as a general purpose dynamometer offer exciting possibilities for broader clinical applications. Future research should explore these potential applications and assess the feasibility and effectiveness of the PhysViz in diverse clinical and home-based settings.

Limitations

This study has several limitations. First, the PhysViz development detailed in this work constitutes only 1 cycle through the FASTER development phase; further iterations of the system are recommended before proceeding to FASTER phase 2 (progressive usability and feasibility evaluation). In addition, because the FASTER provides little guidance regarding when it is appropriate to move between phases, some research and development discretion is needed. Second, long-term exposure to the PhysViz system by end users (people with AT or clinicians) was limited. Three people with AT completed case studies, although the use of the system during these studies was relatively low, limiting the comprehensiveness of feedback. Moreover, the lack of any long-duration (ie, >1 session) testing by the clinicians makes it challenging to comment on their opinions of the PhysViz within the context of their working environment. Although the clinicians were asked how they might integrate the PhysViz into their practice, the study did not specifically explore potential logistical or legal challenges they might face, which presents an area for future research. Third, working with the computer engineering student teams helped control costs when completing development and provided students with experiential learning opportunities; however, because of the course structure, coupled with other commitments, the students were limited in their development capacity. Furthermore, we did not have a software developer actively working on the project. As a result, development iterations took a significant amount of time to implement, limiting the rapid prototyping and feedback cycle typical of engineering design. We tried to circumvent this by using wireframe models and other low-resolution strategies where possible.

Conclusions

This paper describes the development process of the novel PhysViz system, an exercise-based remote rehabilitation system for AT management. This is one of the first studies to describe using the FASTER for technology-based intervention development; by following the FASTER development phase guidelines, the PhysViz incorporates past evidence, end-user needs and opinions, and behavior change science. Our goal is that the PhysViz will be used to improve the current knowledge of appropriate therapeutic exercise dosing for tendinopathic tissue rehabilitation as well as improve the pain, functional status, and activity of people with chronic AT. Future iterations of the PhysViz may be adapted for other home-based treatment strategies that provide user autonomy with remote clinical supervision, thereby decreasing in-clinic time and associated costs.

Acknowledgments

The authors would like to first acknowledge the researchers at the Institute of Sport Sciences at the Humboldt University of Berlin—specifically Univ-Prof Dr Adamantios Arampatzis, Dr Sebastian Bohm, Prof Kirsten Legerlotz, and colleagues in the Department of Training and Movement Sciences—for inspiring this technology and graciously letting the authors adapt their idea. The authors would also like to thank members of the research team who have assisted at various stages of the PhysViz development process: Vivian Chung, Rob Morgan, and Kip Squier. This work was supported by the WorkSafeBC research program (RS2020-TG05), the Natural Sciences and Engineering Research Council of Canada (PGSD3-559905-2021), and by a Natural Sciences and Engineering Research Council of Canada Discovery Grant (AWD-007759). CN is supported by a Michael Smith Foundation for Health Research Health Professional-Investigator Award (HPI-2020- 0719). JLW is supported by a Michael Smith Foundation for Health Research Scholar Award (SCH-2020- 0403) and an Arthritis Society STAR Career Development Award (STAR-19-0493).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Detailed logic models.

[DOCX File , 425 KB - humanfactors_v11i1e57873_app1.docx]

References

1. Scott A, Squier K, Alfredson H, Bahr R, Cook JL, Coombes B, et al. ICON 2019: International Scientific Tendinopathy Symposium consensus: clinical terminology. Br J Sports Med 2020 Mar;54(5):260-262. [doi: 10.1136/bjsports-2019-100885] [Medline: 31399426]
2. Lohrer H, David S, Nauck T. Surgical treatment for achilles tendinopathy - a systematic review. BMC Musculoskelet Disord 2016 May 10;17:207 [FREE Full text] [doi: 10.1186/s12891-016-1061-4] [Medline: 27165287]
3. de Vos RJ, van der Vlist AC, Zwerver J, Meuffels DE, Smithuis F, van Ingen R, et al. Dutch multidisciplinary guideline on Achilles tendinopathy. Br J Sports Med 2021 Oct;55(20):1125-1134 [FREE Full text] [doi: 10.1136/bjsports-2020-103867] [Medline: 34187784]
4. Martin RL, Chimenti R, Cuddeford T, Houck J, Matheson J, McDonough CM, et al. Achilles pain, stiffness, and muscle power deficits: midportion Achilles tendinopathy revision 2018. J Orthop Sports Phys Ther 2018 May;48(5):A1-38. [doi: 10.2519/jospt.2018.0302] [Medline: 29712543]
5. van der Vlist AC, Winters M, Weir A, Ardern CL, Welton NJ, Caldwell DM, et al. Which treatment is most effective for patients with achilles tendinopathy? A living systematic review with network meta-analysis of 29 randomised controlled trials. Br J Sports Med 2021 Mar;55(5):249-256 [FREE Full text] [doi: 10.1136/bjsports-2019-101872] [Medline: 32522732]
6. Silbernagel KG, Thomeé R, Thomeé P, Karlsson J. Eccentric overload training for patients with chronic Achilles tendon pain--a randomised controlled study with reliability testing of the evaluation methods. Scand J Med Sci Sports 2001 Aug;11(4):197-206. [doi: 10.1034/j.1600-0838.2001.110402.x] [Medline: 11476424]
7. Nørregaard J, Larsen CC, Bieler T, Langberg H. Eccentric exercise in treatment of Achilles tendinopathy. Scand J Med Sci Sports 2007 Apr;17(2):133-138. [doi: 10.1111/j.1600-0838.2006.00545.x] [Medline: 17394474]
8. Munteanu SE, Scott LA, Bonanno DR, Landorf KB, Pizzari T, Cook JL, et al. Effectiveness of customised foot orthoses for Achilles tendinopathy: a randomised controlled trial. Br J Sports Med 2015 Aug;49(15):989-994. [doi: 10.1136/bjsports-2014-093845] [Medline: 25246441]
9. Habets B, van Cingel RE, Backx FJ, van Elten HJ, Zuithoff P, Huisstede BM. No difference in clinical effects when comparing Alfredson eccentric and Silbernagel combined concentric-eccentric loading in Achilles tendinopathy: a randomized controlled trial. Orthop J Sports Med 2021 Oct 27;9(10):23259671211031254 [FREE Full text] [doi: 10.1177/23259671211031254] [Medline: 34722783]
10. Gajhede-Knudsen M, Ekstrand J, Magnusson H, Maffulli N. Recurrence of Achilles tendon injuries in elite male football players is more common after early return to play: an 11-year follow-up of the UEFA Champions League injury study. Br J Sports Med 2013 Aug;47(12):763-768. [doi: 10.1136/bjsports-2013-092271] [Medline: 23770660]

11. Silbernagel KG, Hanlon S, Sprague A. Current clinical concepts: conservative management of achilles tendinopathy. J Athl Train 2020 May;55(5):438-447 [FREE Full text] [doi: 10.4085/1062-6050-356-19] [Medline: 32267723]
12. Alfredson H, Cook J. A treatment algorithm for managing Achilles tendinopathy: new treatment options. Br J Sports Med 2007 Apr;41(4):211-216 [FREE Full text] [doi: 10.1136/bjsm.2007.035543] [Medline: 17311806]
13. Jack K, McLean SM, Moffett JK, Gardiner E. Barriers to treatment adherence in physiotherapy outpatient clinics: a systematic review. Man Ther 2010 Jun;15(3):220-228 [FREE Full text] [doi: 10.1016/j.math.2009.12.004] [Medline: 20163979]
14. Room J, Boulton M, Dawes H, Archer K, Barker K. Physiotherapists' perceptions of how patient adherence and non-adherence to recommended exercise for musculoskeletal conditions affects their practice: a qualitative study. Physiotherapy 2021 Dec;113:107-115. [doi: 10.1016/j.physio.2021.06.001] [Medline: 34571284]
15. Bohm S, Mersmann F, Arampatzis A. Human tendon adaptation in response to mechanical loading: a systematic review and meta-analysis of exercise intervention studies on healthy adults. Sports Med Open 2015 Dec;1(1):7 [FREE Full text] [doi: 10.1186/s40798-015-0009-9] [Medline: 27747846]
16. Lazarczuk SL, Maniar N, Opar DA, Duhig SJ, Shield A, Barrett RS, et al. Mechanical, material and morphological adaptations of healthy lower limb tendons to mechanical loading: a systematic review and meta-analysis. Sports Med 2022 Oct;52(10):2405-2429 [FREE Full text] [doi: 10.1007/s40279-022-01695-y] [Medline: 35657492]
17. Merry K, Napier C, Waugh CM, Scott A. Foundational principles and adaptation of the healthy and pathological Achilles tendon in response to resistance exercise: a narrative review and clinical implications. J Clin Med 2022 Aug 12;11(16):4722 [FREE Full text] [doi: 10.3390/jcm11164722] [Medline: 36012960]
18. Alfredson H, Pietilä T, Jonsson P, Lorentzon R. Heavy-load eccentric calf muscle training for the treatment of chronic Achilles tendinosis. Am J Sports Med 1998;26(3):360-366. [doi: 10.1177/03635465980260030301] [Medline: 9617396]
19. Kongsgaard M, Kovanen V, Aagaard P, Doessing S, Hansen P, Laursen AH, et al. Corticosteroid injections, eccentric decline squat training and heavy slow resistance training in patellar tendinopathy. Scand J Med Sci Sports 2009 Dec;19(6):790-802. [doi: 10.1111/j.1600-0838.2009.00949.x] [Medline: 19793213]
20. Stanish WD, Rubinovich RM, Curwin S. Eccentric exercise in chronic tendinitis. Clin Orthop Relat Res 1986 Jul(208):65-68. [Medline: 3720143]
21. Haleem A, Javaid M, Singh RP, Suman R. Telemedicine for healthcare: capabilities, features, barriers, and applications. Sens Int 2021;2:100117 [FREE Full text] [doi: 10.1016/j.sintl.2021.100117] [Medline: 34806053]
22. Boikanyo K, Zungeru AM, Sigweni B, Yahya A, Lebekwe C. Remote patient monitoring systems: applications, architecture, and challenges. Sci Afr 2023 Jul;20:e01638. [doi: 10.1016/j.sciaf.2023.e01638]
23. Reese NB. Muscle and Sensory Testing. Amsterdam, The Netherlands: Elsevier; 2020.
24. Stark T, Walker B, Phillips JK, Fejer R, Beck R. Hand-held dynamometry correlation with the gold standard isokinetic dynamometry: a systematic review. PM R 2011 May;3(5):472-479. [doi: 10.1016/j.pmrj.2010.10.025] [Medline: 21570036]
25. Wang RH, Kenyon LK, McGilton KS, Miller WC, Hovanec N, Boger J, et al. The time is now: a FASTER approach to generate research evidence for technology-based interventions in the field of disability and rehabilitation. Arch Phys Med Rehabil 2021 Sep;102(9):1848-1859. [doi: 10.1016/j.apmr.2021.04.009] [Medline: 33992634]
26. Radovanović G, Kunz J, Bohm S, Arampatzis A, Legerlotz K. Reliable and effective novel home-based training set-up for application of an evidence-based high-loading stimulus to improve triceps surae function. J Sports Sci 2021 Dec;39(24):2786-2795. [doi: 10.1080/02640414.2021.1959981] [Medline: 34378493]
27. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Digital health: a path to validation. NPJ Digit Med 2019 May 13;2:38 [FREE Full text] [doi: 10.1038/s41746-019-0111-3] [Medline: 31304384]
28. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. Lancet 2009 Jul 04;374(9683):86-89. [doi: 10.1016/S0140-6736(09)60329-9] [Medline: 19525005]
29. Balas EA, Boren SA. Managing clinical knowledge for health care improvement. Yearb Med Inform 2000(1):65-70. [doi: 10.1055/S-0038-1637943] [Medline: 27699347]
30. Morris ZS, Wooding S, Grant J. The answer is 17 years, what is the question: understanding time lags in translational research. J R Soc Med 2011 Dec;104(12):510-520 [FREE Full text] [doi: 10.1258/jrsm.2011.110180] [Medline: 22179294]
31. Doorley S, Holcomb S, Klebahn P, Segovia K, Utley J. Design thinking bootleg. d.school at Stanford University. 2018. URL: https://static1.squarespace.com/static/57c6b79629687fde090a0fdd/t/5b19b2f2aa4a99e99b26b6bb/1528410876119/dschool_bootleg_deck_2018_final_sm+%282%29.pdf [accessed 2021-11-02]
32. Merry KJ, MacPherson M, Vis-Dunbar M, Whittaker JL, Grävare Silbernagel K, Scott A. Identifying Intervention Characteristics of Exercise-Based Achilles Tendinopathy (AT) Treatments using TIDieR: A Scoping Review Protocol. OSF Home. 2021 May 14. URL: https://osf.io/t2vh9/ [accessed 2024-10-28]
33. Schreiber C, Ferguson A, Ip D, Moberg Q, Heydari M, Scott A. The effect of biofeedback on the execution of achilles tendinopathy training protocol parameters. OSF Home. 2024 Feb 8. URL: https://osf.io/z96mj/ [accessed 2024-10-28]
34. Merry KJ, MacPherson M, Holsti L, Napier C, Whittaker J, Scott A. Usability testing of the 'PhysViz': a novel tendinopathy telerehabilitation intervention. OSF Home. 2024 Feb 8. URL: https://osf.io/5kaqz/ [accessed 2024-10-28]
35. Merry K, MacPherson MM, Blazey P, Fearon A, Hunt M, Morrissey D, et al. Current practice, guideline adherence, and barriers to implementation for Achilles tendinopathy rehabilitation: a survey of physical therapists and people with Achilles

tendinopathy. BMJ Open Sport Exerc Med 2024 Feb 05;10(1):e001678 [FREE Full text] [doi: 10.1136/bmjsem-2023-001678] [Medline: 38347858]
36. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
37. Michie S, West R, Campbell R, Brown J, Gainforth H. ABC of Behaviour Change Theories. Surrey, UK: Silverback Publishing; 2014.
38. Merry K, MacPherson M, Vis-Dunbar M, Whittaker JL, Grävare Silbernagel K, Scott A. Identifying characteristics of resistance-based therapeutic exercise interventions for Achilles tendinopathy: a scoping review. Phys Ther Sport 2023 Sep;63:73-94 [FREE Full text] [doi: 10.1016/j.ptsp.2023.06.002] [Medline: 37536026]
39. Merry K, Napier C, Chung V, Hannigan BC, MacPherson M, Menon C, et al. The validity and reliability of two commercially available load sensors for clinical strength assessment. Sensors (Basel) 2021 Dec 16;21(24):8399 [FREE Full text] [doi: 10.3390/s21248399] [Medline: 34960492]
40. Pratt S, Seto J, Scott A, Merry K, Mitchell CJ, Waugh CM. Validation of a bluetooth-enabled load cell for objective and accessible assessment of lower extremity strength. J Sports Sci 2023 Sep;41(13):1263-1270. [doi: 10.1080/02640414.2023.2266266] [Medline: 37880983]
41. van den Haak MJ, de jong Menno DT, Schellens PJ. Evaluation of an informational web site: three variants of the think-aloud method compared. Tech Commun 2007 Feb;54(1):58-71 [FREE Full text]
42. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide to Designing Interventions. Surrey, UK: Silverback Publishing; 2014.
43. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 1968 Oct;70(4):213-220. [doi: 10.1037/h0026256] [Medline: 19673146]
44. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. J Clin Epidemiol 1993 May;46(5):423-429. [doi: 10.1016/0895-4356(93)90018-v] [Medline: 8501467]
45. Nielsen J. Usability Engineering. Amsterdam, The Netherlands: Elsevier Science; 1994.
46. Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. J Biomed Inform 2004 Feb;37(1):56-76 [FREE Full text] [doi: 10.1016/j.jbi.2004.01.003] [Medline: 15016386]
47. Giggins OM, Persson UM, Caulfield B. Biofeedback in rehabilitation. J Neuroeng Rehabil 2013 Jun 18;10:60 [FREE Full text] [doi: 10.1186/1743-0003-10-60] [Medline: 23777436]
48. Engelsma T, Jaspers MW, Peute LW. Design principles and guidelines for inclusive mHealth design for people living with dementia. In: Design for Dementia, Mental Health and Wellbeing. Milton Park, UK: Routledge; 2024.
49. Ruzic L, Sanford JA. Universal design mobile interface guidelines (UDMIG) for an aging population. In: Marston H, Freeman S, Musselwhite C, editors. Mobile e-Health. Cham, Switzerland: Springer; Dec 02, 2017.
50. Choi J, Bakken S. Web-based education for low-literate parents in Neonatal Intensive Care Unit: development of a website and heuristic evaluation and usability testing. Int J Med Inform 2010 Aug;79(8):565-575 [FREE Full text] [doi: 10.1016/j.ijmedinf.2010.05.001] [Medline: 20617546]
51. Norman D. The Design of Everyday Things: Revised and Expanded Edition. New York City, NY: Basic Books; 2013.
52. Dalal NP, Quible Z, Wyatt K. Cognitive design of home pages: an experimental study of comprehension on the World Wide Web. Inf Process Manag 2000 Jul 01;36(4):607-621. [doi: 10.1016/S0306-4573(99)00071-0]
53. Stephanidis C, Akoumianakis D, Antona M, Bannon L. User Interfaces for All: Concepts, Methods, and Tools. Mahwah, NJ: Lawrence Erlbaum Associates; 2000.
54. MacPherson MM, Johnston C, Cranston KD, Der S, Sim JA, Jung ME. Identification of intervention characteristics within diabetes prevention programs using the template for intervention description and replication: a scoping review. Can J Diabetes 2024 Jun;48(4):273-280. [doi: 10.1016/j.jcjd.2024.02.004] [Medline: 38417737]
55. Ginsburg LR, Hoben M, Easterbrook A, Anderson RA, Estabrooks CA, Norton PG. Fidelity is not easy! Challenges and guidelines for assessing fidelity in complex interventions. Trials 2021 May 29;22(1):372 [FREE Full text] [doi: 10.1186/s13063-021-05322-5] [Medline: 34051830]
56. Alneyadi M, Drissi N, Almeqbaali M, Ouhbi S. Biofeedback-based connected mental health interventions for anxiety: systematic literature review. JMIR Mhealth Uhealth 2021 Apr 22;9(4):e26038 [FREE Full text] [doi: 10.2196/26038] [Medline: 33792548]
57. Rodrigues RN, Tucci AM, de Barros Viana M. The use of biofeedback on students: a systematic review. Trends Psychol 2021 Oct 26;30:345-366. [doi: 10.1007/s43076-021-00109-8]
58. Garnett C, Crane D, West R, Brown J, Michie S. Identification of behavior change techniques and engagement strategies to design a smartphone app to reduce alcohol consumption using a formal consensus method. JMIR Mhealth Uhealth 2015 Jun 29;3(2):e73 [FREE Full text] [doi: 10.2196/mhealth.3895] [Medline: 26123578]
59. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
60. Altman M, Huang TT, Breland JY. Design thinking in health care. Prev Chronic Dis 2018 Sep 27;15:E117 [FREE Full text] [doi: 10.5888/pcd15.180128] [Medline: 30264690]

61. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
62. Choukou MA. Sustainability of mHealth solutions for healthcare system strengthening. In: Digital Health: Mobile and Wearable Devices for Participatory Health Applications. Amsterdam, The Netherlands: Elsevier; 2021.
63. An Q, Kelley MM, Hanners A, Yen PY. Sustainable development for mobile health apps using the human-centered design process. JMIR Form Res 2023 Aug 25;7:e45694 [FREE Full text] [doi: 10.2196/45694] [Medline: 37624639]
64. van Velsen L, Ludden G, Grünloh C. The limitations of user-and human-centered design in an eHealth context and how to move beyond them. J Med Internet Res 2022 Oct 05;24(10):e37341 [FREE Full text] [doi: 10.2196/37341] [Medline: 36197718]
65. Thornton L, Gardner LA, Osman B, Green O, Champion KE, Bryant Z, et al. A multiple health behavior change, self-monitoring mobile app for adolescents: development and usability study of the Health4Life app. JMIR Form Res 2021 Apr 12;5(4):e25513 [FREE Full text] [doi: 10.2196/25513] [Medline: 33843590]

Abbreviations

APEASE: affordability, practicability, effectiveness, acceptability, side effects and safety, and equity
AT: Achilles tendinopathy
COM-B: capability, opportunity, motivation, and behavior
FASTER: Framework for Accelerated and Systematic Technology-Based Intervention Development and Evaluation Research
MVC: maximal voluntary contraction
PT: physical therapist
UI: user interface
UX: user experience

Edited by C Jacob; submitted 28.02.24; peer-reviewed by N Hovanec, S Markham, MJ Simmich; comments to author 26.08.24; revised version received 04.09.24; accepted 02.10.24; published 13.11.24.

Please cite as:
Merry K, MacPherson MM, Whittaker JL, Napier C, Holsti L, Scott A
An Exercise-Based Precision Medicine Tool and Smartphone App for Managing Achilles Tendinopathy (the 'PhysViz' System): User-Centered Development Study
JMIR Hum Factors 2024;11:e57873
URL: https://humanfactors.jmir.org/2024/1/e57873
doi:10.2196/57873
PMID:

Original Paper

Predictive Factors of Physicians' Satisfaction and Quality of Work Under Teleconsultation Conditions: Structural Equation Analysis

Liliana Hawrysz[1*], PhD, DSC; Magdalena Kludacz-Alessandri[2*], PhD, DSC; Renata Walczak[3*], PhD, DSC

[1]Faculty of Management, Wrocław University of Science and Technology, Wroclaw, Poland

[2]College of Economics and Social Sciences, Warsaw University of Technology, Plock, Poland

[3]Faculty of Civil Engineering, Mechanics and Petrochemistry, Warsaw University of Technology, Plock, Poland

[*]all authors contributed equally

Corresponding Author:
Renata Walczak, PhD, DSC
Faculty of Civil Engineering, Mechanics and Petrochemistry
Warsaw University of Technology
17 Lukasiewicza St
Plock, 09-400
Poland
Phone: 48 887780100
Email: renata.walczak@pw.edu.pl

Abstract

Background: The COVID-19 pandemic contributed to an increase in teleconsultation adoption in the Polish primary health care system. It is expected that in the long run, teleconsultations will successfully replace a significant part of face-to-face visits. Therefore, a significant challenge facing primary health care facilities (PHCs) is the acceptance of teleconsultations by their users, especially physicians.

Objective: This study aimed to explore physicians' acceptance of teleconsultations during the COVID-19 pandemic in Poland.

Methods: A representative survey was conducted among 361 physicians of PHCs across Poland in 2021. For the purposes of the study, we developed a modified Technology Acceptance Model (TAM) model. Based on the modified TAM, we analyzed the impact of perceived usefulness (PU), perceived ease of use (PEU), and intention to use teleconsultation (INT) on physicians' satisfaction (SAT) and quality of work (Q). The psychometric properties of the research instrument were examined using exploratory factor analysis. Finally, structural equation modeling was used for data analysis.

Results: The results indicated a generally high level of PU (mean 3.85-4.36, SD 0.87-1.18), PEU (mean 3.81-4.60, SD 0.60-1.42), INT (mean 3.87-4.22, SD 0.89-1.12), and SAT (mean 3.55-4.13, SD 0.88-1.16); the lowest rated dimension in TAM was Q (mean 3.28-3.73, SD 1.06-1.26). The most important independent variable was PU. The influence of PU on INT (estimate=0.63, critical ratio [CR]=15.84, *P*<.001) and of PU on SAT (estimate=0.44, CR= 9.53, *P*<.001) was strong. INT was also a key factor influencing SAT (estimate=0.4, CR=8.57, *P*<.001). A weaker relationship was noted in the effect of PEU on INT (estimate=0.17, CR=4.31, *P*<.001). In turn, Q was positively influenced by INT (estimate=0.179, CR=3.64, *P*<.001), PU (estimate=0.246, CR=4.79, *P*<.001), PEU (estimate=0.18, CR=4.93, *P*<.001), and SAT (estimate=0.357, CR=6.97, *P*<.001). All paths between the constructs (PU, PEU, INT, SAT, and Q) were statistically significant, which highlights the multifaceted nature of the adoption of teleconsultations among physicians.

Conclusions: Our findings provide strong empirical support for the hypothesized relationships in TAM. The findings suggest that the PU and PEU of teleconsultation have a significant impact on the intention of physicians to adopt teleconsultation. This results in an improvement in the satisfaction of Polish physicians with the use of teleconsultation and an increase in Q. The study contributes to both theory and practice by identifying important prognostic factors affecting physicians' acceptance of teleconsultation systems.

(JMIR Hum Factors 2024;11:e47810) doi:10.2196/47810

KEYWORDS

perceived ease of use; perceived usefulness; physicians' satisfaction; behavioral intention to use telemedicine; health care quality; technology acceptance model; TAM; COVID-19: telemedicine

Introduction

Background

Telemedicine is an IT-based method that has the potential to support and enhance physicians' patient care [1]. Telemedicine is defined as a tool using information and communication technology (ICT) that is used to support and promote remote care, health-related professional education, and public health administration [2]. One of the basic forms of telemedicine is teleconsultations. Medical consultations have been provided remotely (teleconsultations), by telephone, or by instant messaging for many years, but the COVID-19 pandemic has forced widespread use of this form of communication with patients. Teleconsultations have reduced the high costs of medical services and the queues of patients in clinics [3]. In Poland, the need for teleconsultations began at the beginning of the pandemic. Most physicians had to adapt to the requirements of COVID-19 rules and regulations. Even though the need for teleconsultations has already been officially abolished, telemedicine is still the future of medicine. Not all advice has to be given to patients on-site in a clinic. In the face of a decreasing number of physicians, an aging society, and an increased demand for health care, telemedicine seems to be a solution that will solve the problems of personnel shortage. However, the use of telemedicine tools requires the medical staff to be proficient in using them and to accept and support such solutions.

The aim of our study was to examine the physicians' satisfaction (SAT) and quality of work (Q) in conditions of teleconsultations during the COVID-19 pandemic in Poland. For the purposes of the study, we developed a modified Technology Acceptance Model (TAM) and tested it in primary health care facilities (PHCs) in Poland. This model allowed us to analyze the impact of perceived usefulness (PU), perceived ease of use (PEU), and intention to use teleconsultation (INT) on SAT and Q. In our study, we focused primarily on teleconsultations, consisting of telephone and video conversations between physicians and patients [2]. Teleconsultations are the basic form of telemedicine in Poland, and during the COVID-19 pandemic, they were the only form of PHC in the country [4].

The first part of this paper analyzes the literature focusing on TAM in relation to the research hypotheses. The second part contains a description of the research methodology used in this study, with emphasis on the validation of the research tool developed for measuring the modified elements of TAM. The following section presents the results regarding the analyzed constructs of the model and the results of structural equation modeling (SEM). The last part of the paper contains the discussion, conclusions and practical implications.

Literature Review

Until March 2020, the use of telemedicine was negligible and mainly concerned patients in rural areas. In 2019, teleconsultations accounted for as much as 8% of all medical visits in the United States [5]. The situation has changed significantly after the outbreak of the COVID-19 pandemic, especially in the case of emergency visits. In the United States, there was a 683% increase in teleconsultations between March 2 and April 14, 2020 [5]. Researchers agree that the COVID-19 pandemic has changed the way we think about telemedicine. First and foremost, it popularized this way of providing medical services. However, its primary advantage is to protect both patients and physicians from the risk of virus infection [6,7]. In the long term, researchers believe that telemedicine can successfully replace a significant portion of face-to-face (F2F) visits; however, it will not eliminate them completely [4,8,9].

Telemedicine is not a new concept. Decades ago, pioneering projects emerged to test the concept of telemedicine or evaluate its applicability. However, most of these telemedicine projects failed to meet expectations. The failure was blamed on an underdeveloped and mostly primitive IT infrastructure, immature technology, and ineffective use [1]. The failure of first-generation telemedicine projects prompted an in-depth analysis and rigorous assessment of the technological, social, cultural, and organizational dimensions surrounding their introduction.

Resistance from users of new technologies in the medical community is as natural as possible, even if users are aware of the benefits this technology brings with it [10]. Therefore, an essential organizational challenge facing health care organizations considering or planning to provide telemedicine-enabled health care services is technology acceptance by users [1]. The problem regarding acceptance of technology has been discussed for a long time [11], and many models have been developed to assess users' attitudes toward new solutions [12-15]. Acceptance of ICT by physicians providing health care has also been assessed [16-18].

TAM is the most popular model in the literature for testing the acceptance of technology [19]. Van Schaik et al [20] used TAM to assess the attitudes of physiotherapists toward new medical technologies. Chau and Hu [21] studied the acceptance of telemedicine technologies by physicians. Holden and Karsh [22] extensively reviewed the literature on TAM applications and related models for ICT acceptance by health care professionals and, more specifically, health care information systems. They noted that in health care, there is a need for a complete approach to technology acceptance testing than for other professionals from companies or ICT organizations [22]. We chose to use TAM in our study because it is general, parsimonious, and ICT specific. It is designed to provide an explanation and prediction of the acceptance of a wide range of ICTs by a diverse population of users in different organizational contexts. In addition, the model has a well-researched and validated list of psychometric measurements, and this makes its use operationally attractive. Finally, TAM is the dominant model for studying user acceptance of technology, and over the years, it has accumulated satisfactory empirical support for its overall explanatory power and assumed individual causal relationships [1,2]. Over the past few decades, many researchers have proven that TAM enriched with certain other constructs is better suited to research and explain the acceptance of new information technologies by users [23].

TAM analyzes the influence of various factors on the intention to use new technology, among which the main role is played by PU and PEU [24]. PU is defined as the extent to which the user's work performance is expected to improve through the use of new technology [2,25]. Similarly, Davis [11] defined PU as an individual's perceptions regarding the outcome of the experience with technology. In the area of health care, Kissi et al [26] defined PU as "physicians' belief regarding the benefits of telemedicine services that they improve access to medical care, the flow of medical records and patients' health."

PEU, in contrast, is the degree to which using new technology is expected to be effortless [2,25,27]. In the area of health care, PEU describes how physicians perceive telemedicine services in terms of their ease of use and learning [26].

TAM suggests that actual technology usage is determined by individuals' INT [2]. INT is understood as a motivation encouraging the system's user to use the system continuously, and in the case of physicians, it concerns their motivation to use telemedicine services, including, above all, teleconsultations [26]. INT is affected by PU, PEU, and users' attitudes toward technology [2,19]. Lin et al [28] used an integrated approach with the key elements from TAM and assessed the technology acceptance by health professionals of what they called "personal digital assistance (PDA)." The main variables from TAM (PU and self-efficacy) determined INT [28]. Similar conclusions were reached by Zayyad and Toycan [29], Chau and Hu [30], Tubaishat [31], and Vitari and Ologeanu-Taddei [32]. Thus, our first research hypotheses were proposed as follows:

- Hypothesis (H)1: PU has a direct effect on INT.
- H2: PEU has a direct effect on INT.

The results of Lin et al [28] showed that the traditional variables of TAM can be effectively integrated with variables from other theoretical approaches, which may help better understand the acceptance of new technologies by health care professionals [33]. There are relatively fewer TAM tests and modifications in the health care sector. Therefore, it is worth making such attempts to broaden the knowledge about new factors affecting physicians' acceptance of technologies. In our research, we decided to include 2 new constructs: SAT and Q. SAT explains how satisfied users are with using a particular service [19]; Q explains how physicians assess the value and worth of their work with the use of teleconsultations. The addition of these constructs was a consequence of both literature reviews and interviews with physicians (pilot survey). Bhattacherjee and Premkumar [10,34-36] noted that in the use phase of technology (postacceptance), PU is positively associated with user SAT. Alsohime et al [37] also confirmed the effect of PU on SAT, noting the significant impact of training courses before the implementation of new technology. Petter and Fruhling [38] confirmed that the SAT that users have with the information system positively affects their INT. As a consequence, we presented the subsequent hypotheses:

- H3: PU has a direct effect on SAT.
- H4: INT has a direct effect on SAT.

In our previous studies examining technology acceptance in Polish PHC facilities, we developed a conceptual framework defining the impact of PU and PEU on the need for teleconsultation adoption and examined the influence of selected behavioral factors on these constructs [2]. In this paper, we enriched these previous studies by analyzing the impact of PU and PEU as independent variables on SAT and Q. In our analysis, we additionally considered INT as a mediating variable in the model. This is the first study that extends TAM to include an analysis of SAT and Q in the teleconsultation condition.

Padilha et al [39] surveyed students and nurses' ease, usefulness, and intention to use a Massive Open Online Course. Findings confirmed the significant impact of PU, PEU, and INT on the current and future Q the groups studied [39]. Similar conclusions were reached by Saputra et al [40] and Chirchir et al [41]. Souza et al [42] proposed a process model for the evaluation of the Q of clinical decision support systems following the ISO/IEC 25022 and ISO/IEC 25010 standards, part of which was to identify the effect of SAT on Q. Given these considerations, we proposed the following hypotheses:

- H5: PU has a direct effect on Q.
- H6: PEU has a direct effect on Q.
- H7: INT has a direct effect on Q.
- H8: SAT has a direct effect on Q.

Although according to the research conducted so far, TAM is a reliable model for examining technology acceptance in PHC facilities, we can always try to supplement it with new research constructs [2]. TAM, in our study, was supplemented with 2 constructs, SAT and Q, which will contribute to the health care literature.

The Polish health care system is based on an insurance model. PHC physicians in Poland must be health insurance physicians who have a contract with the National Health Fund (NHF) to provide health care services. The functioning of PHCs in Poland is based on the right of patients to personally choose a preferred physician. The selected physicians receive an annual capitation fee for each registered patient [43]. PHC facilities in Poland function as both state-owned and private facilities. Both sign contracts to provide services that are free to the patient and paid for by the NHF. Each facility is managed according to its own rules. Private facilities have more flexibility in making decisions and in hiring and paying employees. Public facilities are subject to top-down regulations governing their operations. As private facilities provide fee-based services, in addition to free services, they have more resources to pay salaries and run their operations.

Methods

Data Collection

In this cross-sectional study, the survey followed a multimodal approach, integrating computer-assisted web interviewing (CAWI), computer-assisted telephone interviewing (CATI), and paper-and-pencil interviewing (PAPI) techniques across a statistically representative sample of 371 PHC facilities. This number was derived from a total of 5503 outpatient PHC facilities in Poland, calculated to be representative at a 95% CI level, with a 50% response distribution and a 5% margin of error, for the aforementioned assumptions, and the minimum survey sample size was 359 [44-47]. The survey sample was randomly selected from the BISNODE database, which includes comprehensive information on all Polish PHC facilities. Of 5503 outpatient facilities in Poland, 371 (6.7%) were successfully surveyed, with each representing 1 physician providing remote medical advice. The survey process entailed replacing nonparticipating facilities with other randomly chosen facilities, ensuring the integrity and representativeness of the sample. Quality control measures were rigorously followed, with a certified polling company overseeing the survey execution. Instances of schematic responses and unusually short survey durations led to the exclusion and replacement of certain responses, resulting in a final analytical sample of 361 (97.3%) records. Before filling the questionnaire, the physicians were informed that the questionnaire is aimed at PHC physicians and concerns the evaluation of their satisfaction with the use of the teleconsultation system for the provision of patient care.

The sample was limited to 1 PHC physician from each randomly selected facility in Poland. This approach was adopted for several reasons. Conducting a survey that included multiple physicians from each facility would have significantly increased the scale and complexity of the study. Given resource constraints, such as funding, time, and personnel, it was more feasible to limit the number of participants, while still achieving a representative sample. The aim was to obtain a broad overview of the acceptance and satisfaction with teleconsultation across a wide range of PHC facilities in Poland. By selecting 1 physician from each facility, we ensured a diverse and statistically representative sample of the entire population of PHC facilities, which may not have been possible with a more concentrated sample from fewer facilities. The study was primarily designed to assess the impact of system-level factors (eg, PU and PUE) on the acceptance of telemedicine. Although individual characteristics, such as age and gender, are important, the primary focus was on broader systemic issues that could be generalized across the population. Conducting an extensive survey during the COVID-19 pandemic posed unique challenges, including limited access to PHC facilities and the need to minimize contact. The study design did not include a detailed examination of individual physician factors and their impact on the acceptance of telemedicine, and the approach was strategically chosen to balance comprehensiveness, feasibility, and the overarching research objectives.

Survey Instrument and Measures

The survey instrument contained 2 groups of statements and questions: statements about analyzed latent factors and general questions about age and gender of the respondent, legal status of the PHC facility, voluntariness of providing remote advice, and ways in which the respondent provided remote advice. The questionnaire included 48 statements and questions, but only the statements used in the modified TAM are presented in Multimedia Appendix 1. A 5-point Likert scale was used in this study: 1 (I do not agree), 2 (I do not agree somewhat), 3 (I neither agree nor disagree), 4 (I agree somewhat), and 5 (I agree).

SEM in this study was adapted from the original TAM, which identifies PU and PEU as the principal determinants of technology use and acceptance. The PU variable measures the degree to which physicians believe that teleconsultations improve their work efficiency and patient care. It encompasses aspects such as enhanced health care delivery, better documentation, and cost-effective monitoring. PU in this context is gauged through 6 survey statements (PU1-PU6), which assess various dimensions of utility that teleconsultations provide in the health care setting [11,48-51]. PEU variables are defined as the extent to which physicians believe that teleconsultations are effortless to learn and implement. The variables relate to the ease of use of the teleconsultation system and the use of medical data. This construct is evaluated via 7 survey statements (PEU1-PEU7), focusing on the usability and accessibility of the teleconsultation system [2,25,27]. INT, similar to that in TAM, used as a mediating variable, represents the likelihood of physicians continuing to use teleconsultations in the future. It is measured through 5 survey statements (INT1-INT5), focusing on the perceived long-term utility and effectiveness of teleconsultations in patient diagnosis and care [11,12,41]. Q, used as an independent variable, refers to the perceived enhancement in work value and worth due to teleconsultations and is gauged through 3 survey statements (Q1-Q3). It assesses whether teleconsultations uphold the standard of traditional visits and enable comprehensive patient care [22]. The SAT variable measures the overall contentment of physicians with teleconsultations. It includes aspects such as convenience compared to traditional visits and comfort in providing remote advice [11,52] and is assessed through 4 survey statements (SAT1-SAT4) [12,13,53,54].

Exploratory Factor Analysis

The study required validating the structure and dynamics within the adapted TAM, because the original model was extended to encompass SAT and Q. The evaluation of the survey statements for inclusion in the factors measuring the assessed dimensions was based on exploratory factor analysis (EFA). EFA was used to select the final variables for the structural model. For each dimension, EFA was separately carried out to assess the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy. The value of this index should be >0.7 [55]. For all dimensions, this condition was fulfilled, but the KMO value of the PEU dimension was <0.7. The EFA results for each dimension are presented in Table 1.

In addition, the ability of each dimension to be represented by individual survey statements was assessed using Bartlett's test of sphericity. For each dimension, the chi-square value was significant, and in each case, $P<.001$. Based on EFA, the PEU dimension was finally divided into 2 separate factors: PEU_1 and PEU_2. Questions PEU1, PEU2, PEU3, and PEU4 were about the technical ease of use of the system (PEU_1), and questions PEU5, PEU6, and PEU7 were about the ease of use of the system from the point of view of handling medical data (PEU_2). Reliability analysis was conducted for all dimensions of the validity of using the adopted statements to measure each factor. Cronbach coefficients were determined for each factor, with acceptable values falling within the range of 0.7-0.95 [56]. The constructs were confirmed to possess suitable psychometric properties, enabling their effective use in SEM analysis. For the PEU dimension, reliability analysis did not give a clear answer as to which statement should be removed to improve the Cronbach and KMO coefficient values. The use of survey statements to measure the PEU dimension requires confirmation in the structural model. The EFA results for the PEU dimension are presented in Table 2, and component factor loadings are presented in Table 3.

Based on imputed factors, a structural model was prepared, where the effects of PU, PEU_1, and PEU_2 on INT, SAT, and Q were studied. Figure 1 shows the final tested model with only significant dependencies.

Table 1. Component factor loadings.[a]

Variable	Factor					
	PU[b]	PEU[c]_1	PEU_2	INT[d]	SAT[e]	Q[f]
KMO[g]	0.90	0.69	0.69	0.80	0.77	0.73
Cronbach α	0.89	0.71	0.71	0.81	0..89	0.86
PU1	0.69	—[h]	—	—	—	—
PU2	0.84	—	—	—	—	—
PU3	0.89	—	—	—	—	—
PU4	0.87	—	—	—	—	—
PU5	0.79	—	—	—	—	—
PU6	0.79	—	—	—	—	—
PEU1	—	0.79	—	—	—	—
PEU2	—	0.69	—	—	—	—
PEU3	—	0.59	—	—	—	—
PEU4	—	0.81	—	—	—	—
PEU5	—	—	0.67	—	—	—
PEU6	—	—	0.86	—	—	—
PEU7	—	—	0.78	—	—	—
INT1	—	—	—	0.80	—	—
INT3	—	—	—	0.82	—	—
INT4	—	—	—	0.76	—	—
INT5	—	—	—	0.83	—	—
SAT1	—	—	—	—	0.78	—
SAT2	—	—	—	—	0.88	—
SAT3	—	—	—	—	0.89	—
SAT4	—	—	—	—	0.82	—
Q1	—	—	—	—	—	0.86
Q2	—	—	—	—	—	0.90
Q3	—	—	—	—	—	0.90

[a]Extraction method: principal component analysis; rotation method: varimax with Kaiser normalization.
[b]PU: perceived usefulness.
[c]PEU: perceived ease of use.
[d]INT: intention to use teleconsultation.
[e]SAT: satisfaction.
[f]Q: quality of work.
[g]KMO: Kaiser-Meyer-Olkin.
[h]Not applicable.

Table 2. Total variance of the PEU[a] dimension explained by EFA[b].[c]

Component	Initial eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.71	38.68	38.68	2.71	38.68	38.68	2.20	31.40	31.40
2	1.40	19.98	58.65	1.40	19.98	58.65	1.91	27.25	58.65
3	0.90	12.86	71.52	—[d]	—	—	—	—	—
4	0.69	9.83	81.35	—	—	—	—	—	—
5	0.58	8.34	89.69	—	—	—	—	—	—
6	0.37	5.28	94.96	—	—	—	—	—	—
7	0.35	5.04	100.00	—	—	—	—	—	—

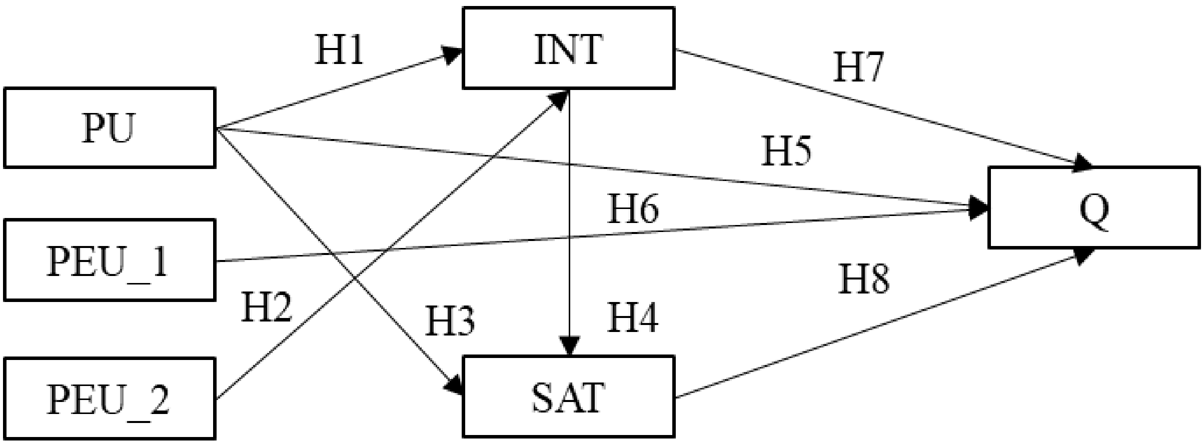
[a]PEU: perceived ease of use.
[b]EFA: exploratory factor analysis.
[c]Extraction method: principal component analysis.
[d]Not applicable.

Table 3. PEU[a] factors loadings.[b]

Variable	Factor	
	PEU_1	PEU_2
PEU1	0.787	—[c]
PEU2	0.685	—
PEU3	0.589	—
PEU4	0.808	—
PEU5	—	0.670
PEU6	—	0.863
PEU7	—	0.784

[a]PEU: perceived ease of use.
[b]Extraction method: principal component analysis; rotation method: varimax with Kaiser normalization.
[c]Not applicable.

Figure 1. Structural model depicting the factors affecting SAT with teleconsultation and Q. Research hypotheses H1-H8 for direct paths. H: hypothesis; INT: intention to use teleconsultation; PEU: perceived ease of use; PU: perceived usefulness; Q: quality of work; SAT: satisfaction.



In the original TAM, the actual use of technology is the independent model variable. In this study, physicians were required to conduct teleconsultations, so INT and SAT variables were treated as mediators affecting the dependent variable, Q. Because of this, the effect of PU, PEU1, and PEU_2 variables on Q was also studied indirectly. The research hypotheses H9-H14 regarding indirect effects were proposed as follows:

- H9: PU has an indirect effect on Q through INT.
- H10: PU has an indirect effect on SAT through INT.

- H11: PU has an indirect effect on Q through INT and SAT.
- H12: PEU_2 has an indirect effect on Q through INT.
- H13: PEU_2 has an indirect effect on SAT through INT.
- H14: PEU_2 has an indirect effect on Q through INT and SAT.

Ethical Considerations

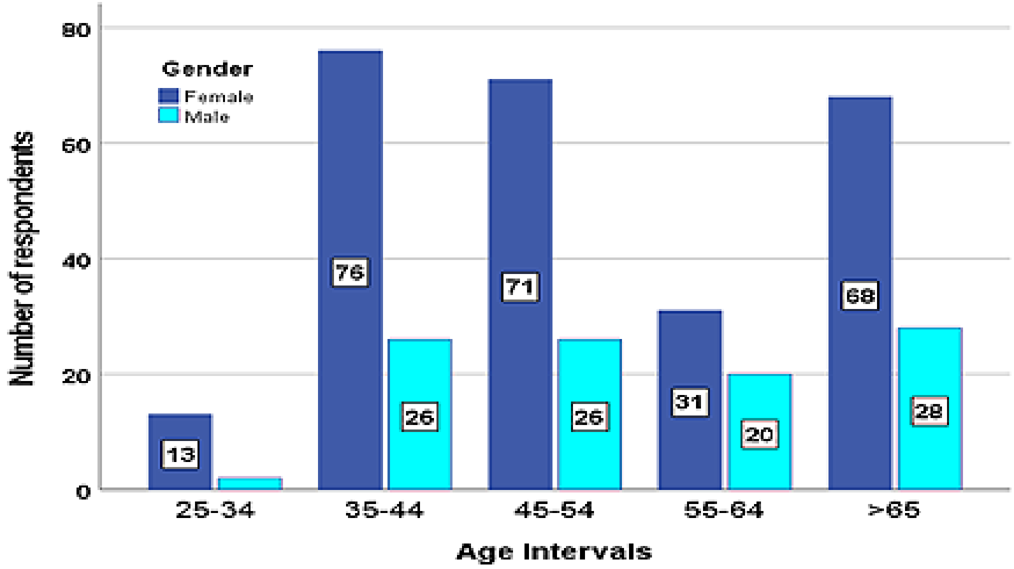
The survey instrument was approved by the Ethics Committee of the Warsaw University of Technology that issued the Certificate of Ethics Approval (certificate dated January 15, 2021). As a result of the contact, 587 physicians provided consent to participate in the study, of which, despite consent, in 216 (36.8%) PHC facilities, the complete set of surveys could not be completed. Respondents were informed about the purpose of the research before starting the survey. They could withdraw from completing the survey at any time. Study data were anonymous and deidentified. No compensation from respondents was taken for the research.

Results

Participant Characteristics

Of the 361 physicians, 199 (55.1%) were in the 35-54–year age group, 94 (26%) were of retirement age and over 65 years old, 260 (72%) were women, and only 101 (28%) were men. The age distribution of the surveyed physicians is presented in Figure 2.

Figure 2. Physicians' age distribution by gender.



Evaluation of the Level of TAM Dimensions

The research considered the dimensions originally defined in TAM (PU, PEU, and INT) [11]. Two dimensions were added to the model: SAT and Q.

The PU variable assesses physicians' perceptions of the utility and benefits of teleconsultation services. The mean scores for PU1-PU6 ranged from 3.85 to 4.36 (SD 0.87-1.18), indicating a generally high level of agreement among physicians that teleconsultations are beneficial to their work. The highest mean score was for PU1 (4.36, SD 0.94), suggesting that the physicians particularly valued teleconsultations during challenging times, such as pandemics. The SDs, ranging from 0.87 (mean 4.24) for PU4 to 1.18 (mean 3.85) for PU3, implied some variability in how the physicians perceived the usefulness of different aspects of teleconsultations. The higher deviation in PU3 indicated more varied opinions about the efficiency enhancement brought by teleconsultations. The skewness for all PU statements was negative, ranging from –0.81 to –1.73, suggesting a tendency among the physicians to agree that teleconsultations are useful in their work. The most pronounced skewness was observed in PU1, indicating strong agreement about the difficulty of work during a pandemic without teleconsultations. The kurtosis values ranged from –0.38 for PU5 to 2.77 for PU1, indicating varied distribution patterns. The higher kurtosis in PU1 reflected a more peaked distribution, suggesting more consistent agreement among physicians regarding its statement. The data suggested that physicians perceive teleconsultations as a valuable tool in their professional practice. The high mean scores across all PU items reflected positive perceptions of the teleconsultation system's usefulness, particularly in aiding work during a pandemic (PU1). The variation in SDs pointed to some differences in individual opinions about the specific benefits of teleconsultations, such as work efficiency and time-saving aspects. The negative skewness across all items highlighted a general agreement on the usefulness of teleconsultations, with a stronger consensus in areas such as coping with pandemic challenges. The kurtosis values, particularly for PU1, indicated that most responses were concentrated around higher ratings, showing strong agreement in specific areas of usefulness. The PU variable demonstrated that physicians generally regard teleconsultations as a beneficial tool in their practice, particularly under challenging conditions, such as a pandemic. Although there was overall agreement on their utility, the variation in perceptions across different aspects suggests areas where experiences and expectations of teleconsultations may differ among individual physicians.

The PEU variable examines physicians' perceptions of the ease and effortlessness associated with using teleconsultation systems. The mean scores for PEU1-PEU7 ranged from 3.81

to 4.60 (SD 0.60-1.42), respectively, and indicated a generally high level of agreement among the physicians that teleconsultation systems are user friendly and easy to use. The highest mean score was for PEU7 (4.60, SD 0.60), suggesting that using external systems during teleconsultations was perceived as particularly straightforward by most of the physicians. The SDs, ranging from 0.60 (mean 4.60) for PEU7 to 1.42 (mean 3.81) for PEU2, suggested a variation in the perceptions of ease of use, with the greatest variation in responses relating to the intellectual effort required (PEU2). The skewness for all PEU statements was negative, ranging from –0.99 to –1.99. This indicated a tendency among the physicians to rate the ease of use of teleconsultation systems highly, with a pronounced leaning toward agreement for most statements, particularly PEU7. The kurtosis values for the PEU variables were varied, with some (eg, PEU7) indicating a highly peaked distribution (7.13), which suggests that responses were more consistently clustered around the higher end of the scale. Overall, the data indicated that physicians find teleconsultation systems relatively easy to use. The high mean scores across all PEU items reflected a positive perception of the teleconsultation system's usability. The variation in SDs, especially the higher deviation for PEU2, suggested that although using the teleconsultation system is generally perceived as easy to use, there are aspects, such as the intellectual effort required, where opinions vary more widely. The pronounced negative skewness, especially for items such as PEU7, underlined a strong agreement in the ease of integrating and using external systems, which might be due to prior familiarity and necessity in clinical practice. The high kurtosis value for PEU7 pointed to a strong consensus among the respondents about the ease of this particular aspect of teleconsultation systems. The physicians generally found teleconsultation systems to be user friendly, with some areas, such as integration with external systems, being particularly well received. Around 253-325 (70.1%-90%) respondents rated the questions in the PEU group positively. However, there was notable variability in perceptions regarding the intellectual effort required, suggesting areas for potential improvement or further training.

The INT variable reflects physicians' intentions and willingness to continue using teleconsultation services in the future. The mean scores for INT1, INT3, INT4, and INT5 were 4.22 (SD 0.89), 3.99 (SD 1.01), 3.87 (SD 1.12), and 4.06 (SD 0.94), respectively. These scores, being close to or above 4 on a 5-point scale, suggested a generally positive inclination among the physicians toward the continued use of teleconsultations. The SDs for INT1, INT3, INT4, and INT5 indicated a moderate level of variation in responses. This variation signified that although there is a general trend of positive intention, individual physicians' perspectives on the future use of teleconsultations vary. The skewness values for these variables ranged from –0.97 to –1.20, which are negative. This negative skewness indicated a tendency among respondents to agree with the statements related to INT, suggesting that a larger proportion of physicians are inclined to continue using these services. The kurtosis values for INT1 (1.26), INT3 (0.46), INT4 (0.22), and INT5 (0.79) showed a relatively normal to slightly peaked distribution. This indicated a consistent pattern in physicians' responses, with a tendency toward agreement on the future use of teleconsultations. The data indicated a positive attitude among physicians toward continuing the use of teleconsultations. The inclination to add video consultations to telephone conversations and to use teleconsultations for patient diagnosis and collaboration with other physicians was evident. The moderate spread in responses, however, pointed to some differences in enthusiasm or confidence about teleconsultations among individual physicians. These differences could be attributed to factors such as personal experience, familiarity with technology, or specific demands of their medical practice. Around 253-289 (70.1%-80%) of respondents intend to use teleconsulting in the future.

The Q variable was rated by physicians as the lowest level of all dimensions. The mean scores ranged from 3.28 (SD 1.26) to 3.73 (SD 1.06). The spread of responses was not even. The mean scores for Q1, Q2, and Q3 were 3.73 (SD 1.06), 3.43 (SD 1.19), and 3.28 (SD 1.26), respectively. These scores indicated a moderate level of agreement among physicians regarding Q, with some variability in the perception of different aspects of Q. The SDs for Q1-Q3 suggested a significant spread in responses. This indicated a varied perception of Q among the physicians, reflecting diverse experiences and expectations regarding teleconsultations. The skewness values for Q1 (–0.82), Q2 (–0.55), and Q3 (–0.27) were negative, implying a tendency for responses to lean toward agreeing with the statements about Q, although this tendency was less pronounced compared to other variables, such as SAT or PU. The kurtosis values for Q1 (–0.01), Q2 (–0.82), and Q3 (–1.18) suggested a relatively flat distribution, particularly for Q2 and Q3. This flatness indicated that the responses were more evenly spread across the scale, reflecting a wide range of opinions on the quality of work. The data suggested that although there is a general trend of moderate satisfaction with Q, there is considerable variation in how physicians perceive this Q. This variability could be influenced by different factors, such as the type of teleconsultation services used, the technological infrastructure in place, and the specific needs of the patient population being served. The more even distribution of responses, especially for Q2 and Q3, indicated that opinions on Q are diverse. This diversity might reflect the complexity of evaluating health care quality in a remote setting, where factors such as patient interaction, diagnostic accuracy, and treatment effectiveness play a crucial role. Only 199 (55.1%) respondents felt that teleconsultations are suitable for holistic patient care, 130 (36%) felt that remote consultations do not allow for comprehensive care, 108 (29.9%) felt that the Q by both methods is not the same, and 224 (62%) believed that the Q using remote visits is similar to that of visits in the clinic. Most physicians (n=260, 72%) agreed with the statement that teleconsultations improve Q.

The SAT variable offers valuable insights into physicians' contentment and approval levels regarding the use of teleconsultations. Physicians positively rated statements regarding SAT. The mean scores for the 4 statements of the SAT variable (SAT1, SAT2, SAT3, and SAT4) were 3.55 (SD 1.16), 4.13 (SD 0.88), 4.00 (SD 0.99), and 3.91 (SD 1.01), respectively. These scores, hovering around or above 4 on a 5-point scale, indicated a general trend of SAT among physicians with teleconsultation services. The SDs for

SAT1-SAT4 suggested a moderate level of variation in responses. This indicated that although there is an overall sense of SAT, there are differences in individual experiences and perceptions regarding teleconsultation services. The skewness values for these variables ranged from –0.62 to –1.14, which are negative. This negative skewness implied a leaning toward higher SAT ratings among the respondents, indicating that more physicians agree with the positive aspects of teleconsultations. The kurtosis values for SAT1 (–0.52), SAT2 (1.51), SAT3 (0.82), and SAT4 (0.78) suggested a mixed distribution pattern. Although SAT1 indicated a relatively normal distribution, SAT2 showed a more peaked distribution, suggesting more consistent high ratings among physicians. The data suggested that physicians are generally satisfied with teleconsultation services, as indicated by the mean scores leaning toward the higher end of the scale. The moderate variation in responses indicated differing levels of SAT, which may be influenced by individual experiences, technological proficiency, or specific needs in their practice. The skewness toward higher SAT ratings suggested that a larger proportion of physicians find teleconsultations to be a convenient and effective medium for providing health care services.

Descriptive statistics of all model variables are presented in Table 4, and distributions of the responses are presented in Table 5.

Table 4. Descriptive statistics of model variables.

	Mean (SD)	Variance	Skewness	Kurtosis
PU1[a]	4.36 (0.94)	0.87	–1.73	2.77
PU2	4.04 (1.03)	1.06	–1.25	1.07
PU3	3.85 (1.18)	1.38	–0.91	–0.14
PU4	4.24 (0.87)	0.75	–1.34	1.93
PU5	3.89 (1.14)	1.30	–0.81	–0.38
PU6	4.03 (1.01)	1.01	–1.16	0.99
PEU1[b]	4.28 (0.83)	0.70	–1.34	1.88
PEU2	3.81 (1.42)	2.02	–0.99	–0.47
PEU3	4.51 (0.67)	0.45	–1.82	5.27
PEU4	4.06 (1.04)	1.08	–1.27	1.14
PEU5	4.19 (0.88)	0.77	–1.63	3.44
PEU6	4.42 (0.69)	0.48	–1.48	3.82
PEU7	4.60 (0.60)	0.36	–1.99	7.13
INT1[c]	4.22 (0.89)	0.80	–1.20	1.26
INT3	3.99 (1.01)	1.01	–0.97	0.46
INT4	3.87 (1.12)	1.25	–0.97	0.22
INT5	4.06 (0.94)	0.88	–1.04	0.79
Q1[d]	3.73 (1.06)	1.13	–0.82	–0.01
Q2	3.43 (1.19)	1.42	–0.55	–0.82
Q3	3.28 (1.26)	1.58	–0.27	–1.18
SAT1[e]	3.55 (1.16)	1.35	–0.62	–0.52
SAT2	4.13 (0.88)	0.78	–1.14	1.51
SAT3	4.00 (0.99)	0.98	–1.02	0.82
SAT4	3.91 (1.01)	1.02	–1.07	0.78

[a]PU: perceived usefulness.

[b]PEU: perceived ease of use.

[c]INT: intention to use teleconsultation.

[d]Q: quality of work.

[e]SAT: satisfaction.

Table 5. Participant (N=361) response distributions.

Likert scale response	I do not agree, n (%)	I do not agree somewhat, n (%)	I neither agree nor disagree, n (%)	I agree somewhat, n (%)	I agree, n (%)
PU1[a]	7 (1.9)	19 (5.3)	17 (4.7)	113 (31.3)	205 (56.8)
PU2	11 (3.0)	32 (8.9)	22 (6.1)	162 (44.9)	134 (37.1)
PU3	18 (5.0)	45 (12.5)	37 (10.2)	134 (37.1)	127 (35.2)
PU4	4 (1.1)	17 (4.7)	27 (7.5)	152 (42.1)	161 (44.6)
PU5	10 (2.8)	51 (14.1)	43 (11.9)	123 (34.1)	134 (37.1)
PU6	10 (2.8)	27 (7.5)	36 (10.0)	157 (43.5)	131 (36.3)
PEU1[b]	2 (0.6)	19 (5.3)	20 (5.5)	156 (43.2)	164 (45.4)
PEU2	48 (13.3)	35 (9.7)	11 (3.0)	111 (30.7)	156 (43.2)
PEU3	2 (0.6)	6 (1.7)	6 (1.7)	138 (38.2)	209 (57.9)
PEU4	12 (3.3)	29 (8.0)	25 (6.9)	154 (42.7)	141 (39.1)
PEU5	9 (2.5)	14 (3.9)	14 (3.9)	186 (51.5)	138 (38.2)
PEU6	2 (0.6)	6 (1.7)	12 (3.3)	160 (44.3)	181 (50.1)
PEU7	2 (0.6)	2 (0.6)	4 (1.1)	123 (34.1)	230 (63.7)
INT1[c]	4 (1.1)	16 (4.4)	41 (11.4)	136 (37.7)	164 (45.4)
INT3	8 (2.2)	28 (7.8)	52 (14.4)	145 (40.2)	128 (35.5)
INT4	17 (4.7)	35 (9.7)	45 (12.5)	146 (40.4)	118 (32.7)
INT5	5 (1.4)	25 (6.9)	43 (11.9)	157 (43.5)	131 (36.3)
Q1[d]	13 (3.6)	47 (13.0)	47 (13.0)	172 (47.6)	82 (22.7)
Q2	25 (6.9)	78 (21.6)	33 (9.1)	167 (46.3)	58 (16.1)
Q3	29 (8.0)	99 (27.4)	36 (10.0)	136 (37.7)	61 (16.9)
SAT1[e]	23 (6.4)	55 (15.2)	60 (16.6)	148 (41.0)	75 (20.8)
SAT2	6 (1.7)	12 (3.3)	48 (13.3)	157 (43.5)	138 (38.2)
SAT3	10 (2.8)	19 (5.3)	59 (16.3)	146 (40.4)	127 (35.2)
SAT4	11 (3.0)	33 (9.1)	36 (10.0)	178 (49.3)	103 (28.5)

[a]PU: perceived usefulness.

[b]PEU: perceived ease of use.

[c]INT: intention to use teleconsultation.

[d]Q: quality of work.

[e]SAT: satisfaction.

Structural Equation Modeling

Based on the factors extracted in EFA, a structural model was developed (Figure 3). The model demonstrated excellent fit with the data, as indicated by indices such as the chi-square-to-degrees-of-freedom index: PCMIN/DF=0.91 (<5 is acceptable), comparative fit index (CFI)=1 (>0.9 is acceptable), goodness-of-fit index (GFI)=0.99 (>0.9 is acceptable), and root mean square error of approximation (RMSEA)=0 (<0.08 is acceptable). These values suggested that our model is robust and accurately represents the observed data.

In assessing the relationships between key constructs in our teleconsultation acceptance model, we used maximum likelihood estimates to derive regression weights (Table 6). Our analysis revealed significant relationships between the constructs, as evidenced by the *P* values and critical ratios (CRs) in the regression weights. The influence of PU on INT was strong (estimate=0.63, CR=15.84, *P*<.001), suggesting that physicians' PU of teleconsultations significantly predicts their INT. PEU_2 also positively influenced INT (estimate=0.17, CR=4.31, *P*<.001), albeit to a lesser extent than PU. Both PU (estimate=0.44, CR=9.53, *P*<.001) and INT (estimate=0.4, CR=8.57, *P*<.001) significantly predicted SAT, indicating that PU and INT are crucial determinants of SAT with teleconsultations. Q was positively influenced by INT (estimate=0.179, CR=3.64, *P*<.001), PU (estimate=0.246, CR=4.79, *P*<.001), PEU_1 (estimate=0.18, CR=4.93, *P*<.001), and SAT (estimate=0.357, CR=6.97, *P*<.001), highlighting a multifaceted impact on Q. Since on each path of the model, the values of the regression parameters had *P*<.05, it can be said that at a significance level of =.05, all the direct dependencies of the model are significant, and hence, H1-H8 are supported.

Standardized regression weights underscore the relative strength of these relationships in standardized format, which is particularly useful for comparing the effects across different predictors within our model. A significant positive covariance was observed between PU and PEU_1 (estimate=0.37, CR=6.52, *P*<.001) and between PU and PEU_2 (estimate=0.28, CR=5.05, *P*<.001), suggesting that PU and PEU are interrelated constructs. However, no significant correlation was found between PEU_1 and PEU_2 (estimate=0, *P*=.99), indicating these aspects of PEU may independently influence the model. The squared multiple correlations for INT (0.48), SAT (0.58), and Q (0.6) indicate a substantial proportion of variance in these endogenous variables, explained by their respective predictors.

Our findings provide strong empirical support for the hypothesized relationships within TAM. The significant paths between constructs such as PU, PEU, INT, SAT, and Q highlight the multifaceted nature of teleconsultation acceptance among physicians.

These results underscore the importance of both PU and PEU in influencing INT and SAT, which, in turn, impact Q.

Figure 3. Structural model presenting standardized estimates. BI: measurement error; INT: intention to use teleconsultation; PEU: perceived ease of use; PU: perceived usefulness; Q: quality of work; SAT: satisfaction.

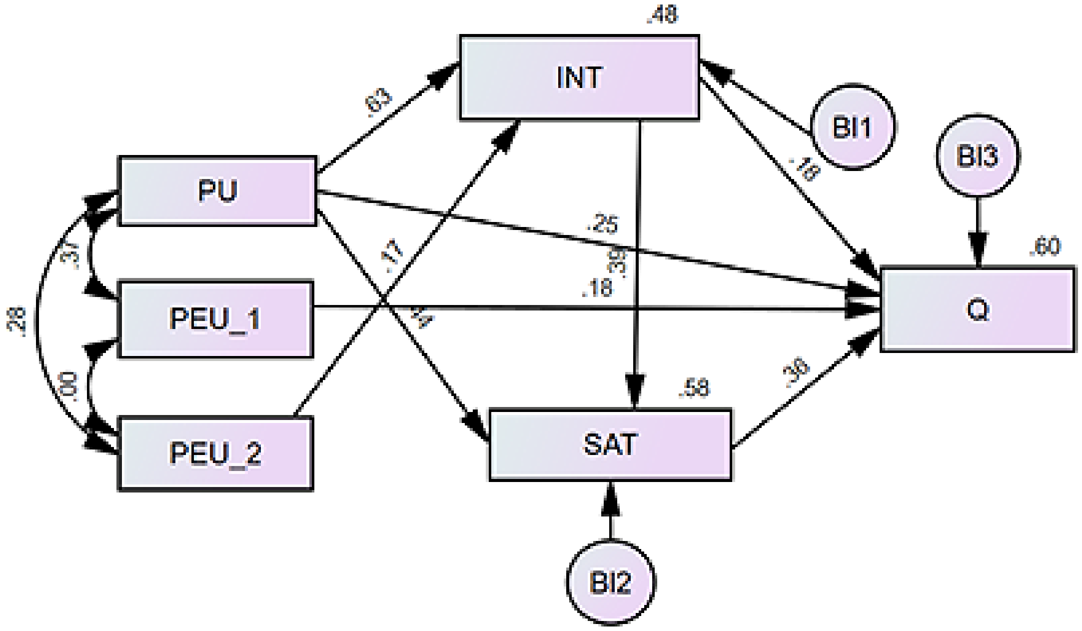


Table 6. Standardized regression weights of direct model paths.

Model paths	Estimate	*P* value
PU[a]→INT[b]	0.626	<.001
PEU[c]_2→INT	0.17	<.001
PU→SAT[d]	0.439	<.001
INT→SAT	0.395	<.001
INT→Q[e]	0.179	<.001
PU→Q	0.246	<.001
PEU_1→Q	0.176	<.001
SAT→Q	0.357	<.001

[a]PU: perceived usefulness.

[b]INT: intention to use teleconsultation.

[c]PEU: perceived ease of use.

[d]SAT: satisfaction.

[e]Q: quality of work.

Mediation Analysis

In addition to the analysis of the model's direct paths, a mediation analysis was conducted. The values of regression factors for indirect model paths were determined (Table 7). The mediating effects within the structural model for the teleconsultation research provide insights into the indirect pathways through which independent variables influence dependent variables. PU had a substantial indirect effect on SAT through INT, with an estimate of 0.25 (*P*<.001). This suggests that the PU of teleconsultation systems significantly influences SAT, mediated by INT. Additionally, PU indirectly affected Q through both INT (estimate=0.112, *P*<.002) and SAT (estimate=0.16, *P*<.001). These paths indicate that PU leads to higher Q, as it influences INT and SAT with it. PEU_2 indirectly influenced SAT via INT, with an estimate of 0.07 (*P*<.001). This effect signifies that the PEU contributes to SAT through INT. For Q, PEU_2 had an indirect effect through the mediation of INT (estimate=0.03, *P*<.001) and SAT (estimate=0.02, *P*<.001). These findings imply that PEU of teleconsultations not only impacts INT but also enhances the Q delivered through SAT. The path from INT to Q mediated by SAT was significant, with an estimate of 0.14 (*P*<.001), indicating that INT contributes to Q, as mediated by SAT levels. The bootstrap 95% CIs for these indirect effects reinforced their significance, as they did not include 0, and the *P* values were well below the .001 threshold, indicating robustness in these mediating relationships. The mediating effects elucidate the important role of INT and SAT with teleconsultation services in enhancing the PU and PEU.

Our findings from the SEM analysis corroborate several key hypotheses concerning direct, indirect, and mediating effects within the teleconsultation context. In accordance with H1, PU exhibited a significant direct effect on INT. Supporting H2, PU was also found to directly affect Q. In line with H3, a direct effect of PU on SAT was substantiated. PEU_1's direct effect on Q affirmed H4, demonstrating the technical influence of PEU on Q. PEU_2 was confirmed to directly impact INT, lending credence to H5. H6, positing a direct effect of INT on SAT, was also validated. Similarly, H7 was supported, with INT having a direct influence on Q. SAT was found to directly affect Q, confirming H8. The indirect influence of PU on Q through INT, posited in H9, was also substantiated. PU was found to indirectly affect SAT through INT, supporting H10. Moreover, the hypothesized indirect effect of PU on Q via the mediating roles of INT and SAT, as stated in H11, was validated. PEU_2's indirect impact on Q through INT, as hypothesized in H12, was confirmed. The indirect effect of PEU_2 on SAT via INT, detailed in H13, was likewise corroborated (Table 8).

Since on each path of the model, the values of the regression parameters had *P*<.05, it can be said that at a significance level of =.05, all the indirect dependencies of the model are significant, and hence, H9-H14 are supported.

Table 7. Standardized regression weights of indirect model paths.

Indirect model paths	Estimate	*P* value	Significance
PU[a]→INT[b]→SAT[c]	0.247	<.001	Significant
PU→INT→SAT→Q[d]	0.088	<.001	Significant
PU→INT→Q	0.112	<.002	Significant
PU→SAT→Q	0.157	<.001	Significant
PEU[e]_2→INT→SAT	0.067	<.001	Significant
PEU_2→INT→SAT→Q	0.024	<.001	Significant
PEU_2→INT→Q	0.03	<.001	Significant
INT→SAT→Q	0.141	<.001	Significant

[a]PU: perceived usefulness.

[b]INT: intention to use teleconsultation.

[c]SAT: satisfaction.

[d]Q: quality of work.

[e]PEU: perceived ease of use.

Table 8. Corroboration of hypotheses concerning direct, indirect, and mediating effects.

Hypothesis (H)	Description	Supported (yes/no)
H1	PU[a] has a direct effect on INT[b].	Yes
H2	PEU[c] has a direct effect on INT.	Yes
H3	PU has a direct effect on SAT[d].	Yes
H4	INT has a direct effect on SAT.	Yes
H5	PU has a direct effect on Q[e].	Yes
H6	PEU has a direct effect on Q.	Yes
H7	INT has a direct effect on Q.	Yes
H8	SAT has a direct effect on Q.	Yes
H9	PU has an indirect effect on Q through INT.	Yes
H10	PU has an indirect effect on SAT through INT.	Yes
H11	PU has an indirect effect on Q through INT and SAT.	Yes
H12	PEU_2 has an indirect effect on Q through INT.	Yes
H13	PEU_2 has an indirect effect on SAT through INT.	Yes
H14	PEU_2 has an indirect effect on Q through INT and SAT.	Yes

[a]PU: perceived usefulness.

[b]INT: intention to use teleconsultation.

[c]PEU: perceived ease of use.

[d]SAT: satisfaction.

[e]Q: quality of work.

Discussion

Principal Findings

This study evaluated Polish physicians' acceptance of teleconsultations during the COVID-19 pandemic in Poland. Most of the physicians positively assessed the PU of teleconsultations. The majority of physicians believed that their work during the COVID-19 pandemic would have been difficult without teleconsultations (88%) and that teleconsultations turned out to be a useful system enabling medical care (87%). The least number of physicians said that teleconsultations save time (61%) and improve performance (72%). Physicians are willing to use new technologies if they do not require additional time and effort, which is in line with other studies [25]. Similar results regarding the usefulness of teleconsultations during a pandemic and the ease of using them were obtained in a cross-sectional study conducted in 2020 in one of the Romanian counties using a questionnaire that assessed, among other things, the perception of teleconsultations by physicians. The study showed a positive perception of telemedicine by Romanian physicians. However, the researchers also highlighted the cons of teleconsultations, such as the time-consuming process, fear of making medical errors remotely, and communication difficulties on the part of patients [57]. The time-consuming nature of teleconsultations has also been confirmed in Great Britain; British physicians reported on time-consuming daily phone calls, emails, and complex electronic medical record protocols [58].

The PEU was also highly rated (average above 4). Most of the surveyed physicians (97%) declared that they know how to connect to external systems during teleconsultations. Using teleconsultations was understandable for most of the respondents (96%), and most of them (94%) could easily prepare all necessary documents (prescriptions, sick leave, referrals for tests, etc) during the teleconsultations. Our results confirm those of other studies according to which the teleconsultations are simple and support physicians' responsibility in their work and medical decisions [30,59].

Polish physicians also positively assessed the future of teleconsultations and declared their intent to this form of work with patients (83%) and other physicians to agree on the diagnosis (73%). According to the majority of respondents (79%), remote monitoring of patients' health would improve the performance of teleconsultations. In the future, they (75%) would also willingly use video visits to facilitate contact and diagnosis of patients. Similar findings were obtained in a Romanian study, in which physicians concluded that telemedicine should be used continuously, not just during the COVID-19 pandemic. Most physicians (91.1%) considered it necessary to provide care using telemedicine after the pandemic [57]. In addition, in Brazil, most physicians want to continue remote care and demand regulations on the use of telemedicine that would allow the extension of remote services [60].

Teleconsultation became popular during the COVID-19 pandemic, and now, there are expectations that it will become a permanent part of the health care system [61]. The development and integration of ICT in health care delivery have great potential for patients, providers, and payers in future health care systems [62].

Polish physicians positively assessed SAT and felt comfortable giving the system a high SAT score. The mean value of responses for all statements regarding SAT was approximately 4. Only the statement regarding the identity of remote and traditional visits was rated lower. Only 62% of physicians believed that both forms of medical consultations are equivalent, and the average for this dimension was 3.55. This is probably due to the influence of teleconsultations on Q, which Polish physicians assessed as the lowest of all the dimensions of TAM. The average response ranged from 3.28 to 3.73. Only 55.1% of respondents stated that teleconsultation is suitable for holistic patient care, and 36% stated that teleconsultation does not allow for comprehensive care. The literature also emphasizes that teleconsultations will never replace F2F meetings. The large-scale and urgent introduction of teleconsultation into our practice is likely to be redefined in the post–COVID-19 era [7]. Another opinion is that teleconsultation is not inferior to personal visits to the office in terms of the preferences and satisfaction of patients and physicians. It should, therefore, be an effective complement to F2F office visits as a mechanism for segregation and long-term continuity of care [63].

Comparison With Previous Studies

This is the first such study conducted on SAT and Q in PHCs. SAT and Q have already been studied in other medical specialties (eg, urology, dermatology, psychiatry, and oncology). In a study conducted among dermatologists, almost all categories regarding SAT with remote dermatological teleconsultations were rated at about 9 on a 0-10 scale [64]. Schubert et al [65] assessed SAT with teleconsultations in psychiatry, which turned out to be at a high level. Providers were satisfied with telepsychiatry, and both believed that telepsychiatry provides patients with better access to care. Urological teleconsultation introduced quickly during the COVID-19 closure has achieved a high level of satisfaction among both patients and physicians [7]. Physicians are interested in using telemedicine tools that increase improved access to health and differentiate their clinical practice [66]. Telemedicine benefits all physicians' patients by increasing access to health care services and remotely managing elderly people with chronic conditions [67]. Therefore, the findings of our research are in accordance with other studies documenting the openness of physicians to the use of teleconsultations in providing health services to patients [68].

However, teleconsultations have limitations regarding the uncertainty caused by the inability to physically check the patient's health condition, and this is something physicians should be aware of [69]. Studies so far show that almost two-thirds of physicians report uncertainty about the correctness of a diagnosis made with telemedicine, and only one-fourth have confidence in making remote decisions [57]. Teleconsultations will never fully replace a personal visit, due to the inability to check the physical symptoms of the disease and the lack of nonverbal signals expressing trust and empathy during remote contact.

SEM results substantiate the significant influence of PU on INT, SAT with this technology, and Q. A possible reason for this may be the availability and effectiveness of teleconsultation, the time saving in this system, and Q. Thus, a positive effect from PU will result in better SAT with teleconsultations and INT. This result is in line with other studies conducted in the field of telemedicine [30,70-73].

The medical PEU from the point of view of handling medical data (PEU_2) has a minor but significant impact on INT. Notably, the technical PEU_1 has a significant impact on Q. The easier it is to use teleconsultation, the better physicians are at assessing Q. The less effort users put in to handle medical data, the more positive their INT to use the system. This is in line with other studies, showing that the acceptance of telemedicine is greater when it provides faster health care, cost savings, better documentation, and time savings [52]. However, a study conducted in the United States found that the role of the influence of PEU on INT is insignificant. The study focused on pediatricians' INT to use of online health apps. The reason for this could be the longer contact of physicians with telemedicine technology [64]. Another explanation is that for highly competent physicians, the effect of PEU on INT is of little importance [74].

PU and PEU are considered the main determinants that directly explain the intent to use ("accept") a new technology [75]. In this study, we, therefore, confirmed the hypotheses that the constructs described in the traditional TAM are appropriate for measuring the intent of physicians to use teleconsultations.

INT emerged as a factor influencing SAT and Q and a pivotal mediator linking PU with both SAT and Q, thus underscoring the importance of intentions in the acceptance and effective use of telemedicine. This finding is in harmony with the existing literature that emphasizes the mediating role of INT in the context of technology acceptance [26]. When physicians believe that using teleconsultation will be effortless and useful, their attitude and INT will improve. This system, with less effort, encourages physicians to use teleconsultations and improves SAT and Q [31,70,76,77]. Therefore, the condition for the implementation of telemedicine technologies should be ensuring its understanding by health care providers in order to gain their acceptance and ensure the use of these technologies in the future [78]. SAT also turned out to be a significant mediator between PU, PEU, and Q. Our research, therefore, showed that the main elements of TAM viewed as PU and PEU have a significant impact on INT, which has been confirmed in other studies [30,70-73,79-82].

The model reaffirms the significant direct and indirect roles of PU and PEU in shaping INT, SAT, and Q. These findings contribute to the extant literature on teleconsultation acceptance and underscore the nuanced factors that influence the acceptance and satisfaction of teleconsultation services among physicians.

Implications

When planning a new teleconsultation system, PHC facilities should be able to predict whether the new system will be acceptable and satisfactory for medical staff, investigate the reasons why the planned system may not be fully acceptable, and then take action to increase the system's acceptance. The results of this study show that the PU of a system is a key determinant of medical professionals' INT. Therefore, before introducing a new system to PHC facilities, managers of these facilities can increase the acceptance of the system by involving medical personnel in the implementation process, assessing the medical personnel's perception of the system (PU and PEU) and taking appropriate actions based on this assessment. Training should also be provided to medical staff to highlight the effectiveness and usefulness of teleconsultation in PHCs. Information and training sessions should primarily focus on how teleconsultation can help improve the quality of PHCs.

Intention as an intermediary variable has a significant and positive impact on users' SAT with teleconsultation and its Q. To increase the expected results in the Q of teleconsultations, the teleconsultation system should be useful for health checking, improving the quality of life, and increasing the capacity for self-care. To increase PEU, the teleconsultation system should be clear, understandable, easy to learn, easy to implement, and easy to perform health checks with. To increase the perceived utility, the teleconsultation system should positively influence the treatment plan, provide more comprehensive care services, and efficiently diagnose and efficiently plan and precisely monitor the patient's condition. The system should make physicians willing to use it to increase their INT. All this will contribute to greater SAT of physicians with their work and better quality of care [83].

This study is a contribution to the field of teleconsultation acceptance research. The modified TAM and its psychometric properties verified in this study can be used as a research framework to understand the acceptance of teleconsultation, especially in the population of PHC workers. The model can also be used for future TAM research in a variety of contexts in identifying, explaining, and predicting the intention of PHC professionals to use teleconsultation. Therefore, it is highly recommended to replicate this study in different environments to generalize the results across domains.

Limitations

In the discussion of research findings, it is crucial to acknowledge the limitations of this study to provide a comprehensive understanding of its context and implications. Although the study offers valuable insights into the factors influencing the acceptance and SAT of medical professionals with teleconsultation systems, several limitations must be considered.

The study was conducted across a specific number of PHC facilities in Poland. Although efforts were made to ensure a representative sample of PHC facilities, the findings might not fully encapsulate the diverse range of experiences and perceptions of all medical professionals nationwide. Regional variations, different health care settings, and varying levels of teleconsultation acceptance could influence SAT and acceptance levels.

The cross-sectional nature of the survey limits our ability to infer causality or changes over time. Longitudinal studies would be required to understand how perceptions and SAT with teleconsultation evolve as users gain more experience and as the technology itself advances.

The reliance on self-reported data can introduce biases, such as social desirability or recall bias. Participants' responses might not accurately reflect their true experiences or feelings toward teleconsultation.

Although the study focused on PU and PEU, other factors could influence SAT and INT. These might include individual technological proficiency, prior experiences with teleconsultation, or organizational support, which were not extensively explored in this study.

The study primarily addressed teleconsultation in PHC settings. The findings might not be generalizable to other forms of teleconsultation or to specialists' use of teleconsultation, where different factors could be more influential. The study also did not deeply explore the technological and operational constraints that might impact the effectiveness and user SAT with teleconsultation systems, such as system reliability, user interface design, and integration with existing health information systems.

The study was conducted during a period potentially influenced by the COVID-19 pandemic, which might have affected attitudes toward teleconsultation. The urgency and necessity of teleconsultation during the pandemic might not reflect standard operational conditions.

By addressing these limitations, future research can build upon the findings to develop a more nuanced understanding of the factors influencing the successful implementation and adoption of teleconsultation systems in various health care settings. In the forthcoming models, we also intend to include constructs such as compatibility, self-efficacy, social norms, perceived behavioral control [25,84], social interaction, invasiveness, and relevance [85].

Conclusion

After the outbreak of the COVID-19 pandemic, there was a dynamic development of teleconsultations in PHCs in Poland. Therefore, we conducted satisfaction surveys of Polish physicians based on a modified TAM, which we extended with new constructs, including physicians' SAT and Q, considering INT as a mediating variable. The tool developed for this model was verified in terms of psychometric properties. Therefore, it has the potential to be used in both research and practice, especially to assess the SAT and Q of PHC physicians who use teleconsultations in Poland.

The findings highlight significant relationships between PEU, PU, INT, and physicians' SAT with teleconsultation and their Q. The study showed that the PEU and PU of teleconsultations are predictive determinants of the acceptance of teleconsultation, which in turn influences physicians' SAT and Q.

Identification of the most important factors influencing physicians' SAT and Q can provide important information to managers of PHC facilities and help them make the right decisions. This study provides information for the strategies of PHCs and policy makers to accept and encourage the use of teleconsultations in Poland.

Acknowledgments

This research was supported by the Warsaw University of Technology under the Excellence Initiative: Research University (IDUB) program.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

RW, MK-A, and LH contributed to study conception and design and data collection and wrote the draft manuscript. RW contributed to the formal analysis and created the figures. All authors have reviewed the results and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire statements used in the modified Technology Acceptance Model.
[DOCX File , 17 KB - humanfactors_v11i1e47810_app1.docx]

References

1. Hu PJ, Chau PY, Sheng ORL, Tam KY. Examining the technology acceptance model using physician acceptance of telemedicine technology. J Manag Inf Syst 2015 Dec 02;16(2):91-112. [doi: 10.1080/07421222.1999.11518247]
2. Walczak R, Kludacz-Alessandri M, Hawrysz L. Use of telemedicine technology among general practitioners during COVID-19: a modified technology acceptance model study in Poland. Int J Environ Res Public Health 2022 Sep 01;19(17):10937 [FREE Full text] [doi: 10.3390/ijerph191710937] [Medline: 36078650]
3. Kalaivani K, Anjalipriya V, Sivakumar R, Srimeena R. An efficient bio-key management scheme for telemedicine applications. 2015 Jul Presented at: 2015 IEEE Technological Innovation in ICT for Agriculture and Rural Development (TIAR); July 10-12, 2015; Chennai, India. [doi: 10.1109/TIAR.2015.7358543]
4. Kludacz-Alessandri M, Hawrysz L, Korneta P, Gierszewska G, Pomaranik W, Walczak R. The impact of medical teleconsultations on general practitioner-patient communication during COVID- 19: a case study from Poland. PLoS One 2021 Jul 16;16(7):e0254960 [FREE Full text] [doi: 10.1371/journal.pone.0254960] [Medline: 34270587]
5. Mann DM, Chen J, Chunara R, Testa PA, Nov O. COVID-19 transforms health care through telemedicine: evidence from the field. J Am Med Inform Assoc 2020 Jul 01;27(7):1132-1135 [FREE Full text] [doi: 10.1093/jamia/ocaa072] [Medline: 32324855]
6. Liu L, Gu J, Shao F, Liang X, Yue L, Cheng Q, et al. Application and preliminary outcomes of remote diagnosis and treatment during the COVID-19 outbreak: retrospective cohort study. JMIR Mhealth Uhealth 2020 Jul 03;8(7):e19417 [FREE Full text] [doi: 10.2196/19417] [Medline: 32568722]
7. Pinar U, Anract J, Perrot O, Tabourin T, Chartier-Kastler E, Parra J, et al. Preliminary assessment of patient and physician satisfaction with the use of teleconsultation in urology during the COVID-19 pandemic. World J Urol 2021 Jun 09;39(6):1991-1996 [FREE Full text] [doi: 10.1007/s00345-020-03432-4] [Medline: 32909174]
8. Layfield E, Triantafillou V, Prasad A, Deng J, Shanti RM, Newman JG, et al. Telemedicine for head and neck ambulatory visits during COVID-19: evaluating usability and patient satisfaction. Head Neck 2020 Jul;42(7):1681-1689 [FREE Full text] [doi: 10.1002/hed.26285] [Medline: 32476228]
9. Tenforde AS, Borgstrom H, Polich G, Steere H, Davis IS, Cotton K, et al. Outpatient physical, occupational, and speech therapy synchronous telemedicine: . a survey study of patient satisfaction with virtual visits during the COVID-19 pandemic. Am J Phys Med Rehabil 2020 Nov;99(11):977-981 [FREE Full text] [doi: 10.1097/PHM.0000000000001571] [Medline: 32804713]
10. Bhattacherjee A, Hikmet N. Physicians' resistance toward healthcare information technology: a theoretical model and empirical test. Eur J Inf Syst 2017 Dec 19;16(6):725-737. [doi: 10.1057/palgrave.ejis.3000717]
11. Davis F. Usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989;13(3):319-340 [FREE Full text] [doi: 10.2307/249008]
12. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003 Sep;27(3):425-478. [doi: 10.2307/30036540]
13. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manag Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
14. Beldad AD, Hegner SM. Expanding the technology acceptance model with the inclusion of trust, social influence, and health valuation to determine the predictors of German users' willingness to continue using a fitness app: a structural equation modeling approach. Int J Hum–Comput Interact 2017 Nov 30;34(9):882-893. [doi: 10.1080/10447318.2017.1403220]

15. Legris P, Ingham J, Collerette P. Why do people use information technology? A critical review of the technology acceptance model. Inf Manag 2003 Jan;40(3):191-204. [doi: 10.1016/S0378-7206(01)00143-4]
16. Yarbrough AK, Smith TB. Technology acceptance among physicians: a new take on TAM. Med Care Res Rev 2007 Dec;64(6):650-672. [doi: 10.1177/1077558707305942] [Medline: 17717378]
17. Huang J. Innovative health care delivery system--a questionnaire survey to evaluate the influence of behavioral factors on individuals' acceptance of telecare. Comput Biol Med 2013 May;43(4):281-286. [doi: 10.1016/j.compbiomed.2012.12.011] [Medline: 23375377]
18. Rho MJ, Choi IY, Lee J. Predictive factors of telemedicine service acceptance and behavioral intention of physicians. Int J Med Inform 2014 Aug;83(8):559-571. [doi: 10.1016/j.ijmedinf.2014.05.005] [Medline: 24961820]
19. Ladan MA, Wharrad H, Windle R. Towards understanding healthcare professionals' adoption and use of technologies in clinical practice: using Qmethodology and models of technology acceptance. J Innov Health Inform 2018 Mar 09;25(1):965-937 [FREE Full text] [doi: 10.14236/jhi.v25i1.965] [Medline: 29717952]
20. Van Schaik P, Bettany-Saltikov JA, Warren JG. Clinical acceptance of a low-cost portable system for postural assessment. Behav Inf Technol 2002 Jan;21(1):47-57. [doi: 10.1080/01449290110107236]
21. Chau PYK, Hu PJ. Information technology acceptance by individual professionals: a model comparison approach. Decis Sci 2007 Jun 07;32(4):699-719. [doi: 10.1111/j.1540-5915.2001.tb00978.x]
22. Holden RJ, Karsh B. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
23. Croteau A, Vieru D. Telemedicine adoption by different groups of physicians. 2002 Presented at: 35th Annual Hawaii International Conference on System Sciences; January 7-10, 2002; Big Island, HI p. 1985-1993. [doi: 10.1109/hicss.2002.994121]
24. Hsiao J, Chen R. Critical factors influencing physicians' intention to use computerized clinical practice guidelines: an integrative model of activity theory and the technology acceptance model. BMC Med Inform Decis Mak 2016 Jan 16;16(1):3 [FREE Full text] [doi: 10.1186/s12911-016-0241-3] [Medline: 26772169]
25. Deng Q, Zheng Y, Lu J, Zeng Z, Liu W. What factors predict physicians' utilization behavior of contrast-enhanced ultrasound? Evidence from the integration of the Theory of Planned Behavior and Technology Acceptance Model using a structural equation modeling approach. BMC Med Inform Decis Mak 2021 May 31;21(1):173 [FREE Full text] [doi: 10.1186/s12911-021-01540-8] [Medline: 34059025]
26. Kissi J, Dai B, Dogbe CS, Banahene J, Ernest O. Predictive factors of physicians' satisfaction with telemedicine services acceptance. Health Informat J 2020 Sep 19;26(3):1866-1880 [FREE Full text] [doi: 10.1177/1460458219892162] [Medline: 31854222]
27. Metallo C, Agrifoglio R, Lepore L, Landriani L. Explaing users' technology acceptance through national cultural values in the hospital context. BMC Health Serv Res 2022 Jan 17;22(1):84 [FREE Full text] [doi: 10.1186/s12913-022-07488-3] [Medline: 35039014]
28. Lin Y, Yu C, Yi C. The effects of positive affect, person-job fit, and well-being on job performance. Soc Behav Pers 2014 Oct 26;42(9):1537-1547. [doi: 10.2224/sbp.2014.42.9.1537]
29. Zayyad M, Toycan M. Factors affecting sustainable adoption of e-health technology in developing countries: an exploratory survey of Nigerian hospitals from the perspective of healthcare professionals. PeerJ 2018;6:e4436 [FREE Full text] [doi: 10.7717/peerj.4436] [Medline: 29507830]
30. Chau PY, Hu PJ. Investigating healthcare professionals' decisions to accept telemedicine technology: an empirical test of competing theories. Inf Manag 2002 Jan;39(4):297-311. [doi: 10.1016/S0378-7206(01)00098-2]
31. Tubaishat A. Perceived usefulness and perceived ease of use of electronic health records among nurses: application of technology acceptance model. Inform Health Soc Care 2018 Dec;43(4):379-389. [doi: 10.1080/17538157.2017.1363761] [Medline: 28920708]
32. Vitari C, Ologeanu-Taddei R. The intention to use an electronic health record and its antecedents among three different categories of clinical staff. BMC Health Serv Res 2018 Mar 21;18(1):194 [FREE Full text] [doi: 10.1186/s12913-018-3022-0] [Medline: 29562942]
33. Yi MY, Jackson JD, Park JS, Probst JC. Understanding information technology acceptance by individual professionals: toward an integrative view. Inf Manag 2006 Apr;43(3):350-363. [doi: 10.1016/j.im.2005.08.006]
34. Premkumar G, Bhattacherjee A. Explaining information technology usage: a test of competing models. Omega 2008 Feb;36(1):64-75. [doi: 10.1016/j.omega.2005.12.002]
35. Bhattacherjee A, Premkumar G. Understanding changes in belief and attitude toward information technology usage: a theoretical model and longitudinal test. MIS Quarterly 2004;28(2):229. [doi: 10.2307/25148634]
36. Bhattacherjee A. Understanding information systems continuance: an expectation-confirmation model. MIS Quarterly 2001 Sep;25(3):351. [doi: 10.2307/3250921]
37. Alsohime F, Temsah MH, Al-Eyadhy A, Bashiri FA, Househ M, Jamal A, et al. Satisfaction and perceived usefulness with newly-implemented electronic health records system among pediatricians at a university hospital. Comput Methods Programs Biomed 2019 Feb;169:51-57. [doi: 10.1016/j.cmpb.2018.12.026] [Medline: 30638591]

38. Petter S, Fruhling A. Evaluating the success of an emergency response medical information system. Int J Med Inform 2011 Jul;80(7):480-489. [doi: 10.1016/j.ijmedinf.2011.03.010] [Medline: 21501969]
39. Padilha JM, Machado PP, Ribeiro AL, Ribeiro R, Vieira F, Costa P. Easiness, usefulness and intention to use a MOOC in nursing. Nurse Educ Today 2021 Feb;97:104705. [doi: 10.1016/j.nedt.2020.104705] [Medline: 33341525]
40. Saputra DEE, Maulida Rahma V, Eliyana A, Pratama AS, Anggraini RD, Kamil NLM, et al. Do system quality and information quality affect job performance? The mediation role of users' perceptions. PLoS One 2023;18(6):e0285293 [FREE Full text] [doi: 10.1371/journal.pone.0285293] [Medline: 37352227]
41. Chirchir LK, Aruasa WK, Chebon SK. Perceived usefulness and ease of use as mediators of the effect of health information systems on user performance. Eur J Comput Sci Inf Technol 2019 Feb;7(1):22-37. [doi: 10.1017/CBO9781107415324.004]
42. Souza-Pereira L, Ouhbi S, Pombo N. A process model for quality in use evaluation of clinical decision support systems. J Biomed Inform 2021 Nov;123:103917 [FREE Full text] [doi: 10.1016/j.jbi.2021.103917] [Medline: 34563692]
43. Korneta P, Kludacz-Alessandri M, Walczak R. The impact of COVID-19 on the performance of primary health care service providers in a capitation payment system: a case study from Poland. Int J Environ Res Public Health 2021 Feb 03;18(4):1-19 [FREE Full text] [doi: 10.3390/ijerph18041407] [Medline: 33546467]
44. Davis JP, Crombie IK, Davies HT. Understanding sampling: representativeness matters. Hosp Med 2001 Apr;62(4):237-240. [doi: 10.12968/hosp.2001.62.4.2052] [Medline: 11338957]
45. Wetterberg H, Rydén L, Ahlner F, Falk Erhag H, Gudmundsson P, Guo X, et al. Representativeness in population-based studies of older adults: five waves of cross-sectional examinations in the Gothenburg H70 Birth Cohort Study. BMJ Open 2022 Dec 16;12(12):e068165 [FREE Full text] [doi: 10.1136/bmjopen-2022-068165] [Medline: 36526314]
46. Lynn P. Maintaining cross-sectional representativeness in a longitudinal general population survey. Understanding Society Working Paper Series No. 2011 – 04. 2011 Jun. URL: https://www.understandingsociety.ac.uk/wp-content/uploads/working-papers/2011-04.pdf [accessed 2023-12-15]
47. Hoffmann R, Lange M, Butschalowsky H, Houben R, Schmich P, Allen J, et al. KiGGS wave 2 cross-sectional study - participant acquisition, response rates and representativeness. J Health Monit 2018 Mar;3(1):78-91 [FREE Full text] [doi: 10.17886/RKI-GBE-2018-032] [Medline: 35586176]
48. Martínez A, Everss E, Rojo-Alvarez JL, Figal DP, García-Alberola A. A systematic review of the literature on home monitoring for patients with heart failure. J Telemed Telecare 2006;12(5):234-241. [doi: 10.1258/135763306777889109] [Medline: 16848935]
49. Rimmer MP, Al Wattar BH, UKARCOG Members. Provision of obstetrics and gynaecology services during the COVID-19 pandemic: a survey of junior doctors in the UK National Health Service. BJOG 2020 Aug;127(9):1123-1128 [FREE Full text] [doi: 10.1111/1471-0528.16313] [Medline: 32460422]
50. Bakken S, Grullon-Figueroa L, Izquierdo R, Lee N, Morin P, Palmas W, IDEATel Consortium. Development, validation, and use of English and Spanish versions of the telemedicine satisfaction and usefulness questionnaire. J Am Med Inform Assoc 2006;13(6):660-667 [FREE Full text] [doi: 10.1197/jamia.M2146] [Medline: 16929036]
51. Rho MJ, Kim SR, Kim H, Cho J, Yoon K, Mun SK, et al. Exploring the relationship among user satisfaction, compliance, and clinical outcomes of telemedicine services for glucose control. Telemed J E Health 2014 Aug;20(8):712-720. [doi: 10.1089/tmj.2013.0309] [Medline: 24901576]
52. Kamal SA, Shafiq M, Kakria P. Investigating acceptance of telemedicine services through an extended technology acceptance model (TAM). Technol Soc 2020 Feb;60:101212. [doi: 10.1016/j.techsoc.2019.101212]
53. Schepers J, Wetzels M. A meta-analysis of the technology acceptance model: investigating subjective norm and moderation effects. Inf Manag 2007 Jan;44(1):90-103. [doi: 10.1016/j.im.2006.10.007]
54. Kohnke A, Cole ML, Bush R. Incorporating UTAUT predictors for understanding home care patients' and clinician's acceptance of healthcare telemedicine equipment. J Technol Manag Innovat 2014 Jul;9(2):29-41. [doi: 10.4067/S0718-27242014000200003]
55. Hair J, Babin B, Anderson R, Black W. Multivariate Data Analysis. 8th Edition. Noida, Uttar Pradesh, India: Cengage; 2018.
56. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ 2011 Jun 27;2:53-55 [FREE Full text] [doi: 10.5116/ijme.4dfb.8dfd] [Medline: 28029643]
57. Florea M, Lazea C, Gaga R, Sur G, Lotrean L, Puia A, et al. Lights and shadows of the perception of the use of telemedicine by Romanian family doctors during the COVID-19 pandemic. Int J Gen Med 2021;14:1575-1587 [FREE Full text] [doi: 10.2147/IJGM.S309519] [Medline: 33953605]
58. Thornton J. Covid-19: how coronavirus will change the face of general practice forever. BMJ 2020 Mar 30;368:m1279. [doi: 10.1136/bmj.m1279] [Medline: 32229477]
59. Craig J, Patterson V. Introduction to the practice of telemedicine. J Telemed Telecare 2005 Jun;11(1):3-9. [doi: 10.1177/1357633X0501100102] [Medline: 15829036]
60. Caetano R, Silva AB, Guedes ACCM, Paiva CCND, Ribeiro GDR, Santos DL, et al. Challenges and opportunities for telehealth during the COVID-19 pandemic: ideas on spaces and initiatives in the Brazilian context. Cad Saude Publica 2020;36(5):e00088920 [FREE Full text] [doi: 10.1590/0102-311x00088920] [Medline: 32490913]

61. Hoff T, Lee DR. Physician satisfaction with telehealth: a systematic review and agenda for future research. Qual Manag Health Care 2022;31(3):160-169 [FREE Full text] [doi: 10.1097/QMH.0000000000000359] [Medline: 35132008]
62. Abdel-Wahab M, Rosenblatt E, Prajogi B, Zubizarretta E, Mikhail M. Opportunities in telemedicine, lessons learned after COVID-19 and the way into the future. Int J Radiat Oncol Biol Phys 2020 Oct 01;108(2):438-443 [FREE Full text] [doi: 10.1016/j.ijrobp.2020.07.006] [Medline: 32890528]
63. Melian C, Kieser D, Frampton C, C Wyatt M. Teleconsultation in orthopaedic surgery: a systematic review and meta-analysis of patient and physician experiences. J Telemed Telecare 2022 Aug 01;28(7):471-480. [doi: 10.1177/1357633X20950995] [Medline: 32873138]
64. Fluhr JW, Gueguen A, Legoupil D, Brenaut E, Abasq C, Araújo H, et al. Teledermatology in times of COVID-19 confinement: comparing patients' and physicians' satisfaction by the standardized brest teledermatology questionnaire. Dermatology 2021 Feb 10;237(2):1-6 [FREE Full text] [doi: 10.1159/000514029] [Medline: 33567427]
65. Schubert NJ, Backman PJ, Bhatla R, Corace KM. Telepsychiatry and patient-provider concordance. Can J Rural Med 2019;24(3):75-82 [FREE Full text] [doi: 10.4103/CJRM.CJRM_9_18] [Medline: 31249155]
66. Hodgkins M, Barron M, Jevaji S, Lloyd S. Physician requirements for adoption of telehealth following the SARS-CoV-2 pandemic. NPJ Digit Med 2021 Feb 09;4(1):19 [FREE Full text] [doi: 10.1038/s41746-021-00390-y] [Medline: 33564102]
67. Florea M, Puia A, Pop R. The family as recipient and provider of home care: a primary care perspective. In: Cascella M, Stones MJ, editors. Suggestions for Addressing Clinical and Non-Clinical Issues in Palliative Care. London, UK: IntechOpen; 2021.
68. Helou S, El Helou E, Abou-Khalil V, Wakim J, El Helou J, Daher A, et al. The effect of the COVID-19 pandemic on physicians' use and perception of telehealth: the case of Lebanon. Int J Environ Res Public Health 2020 Jul 06;17(13):1-17 [FREE Full text] [doi: 10.3390/ijerph17134866] [Medline: 32640652]
69. Iyengar K, Jain VK, Vaishya R. Pitfalls in telemedicine consultations in the era of COVID 19 and how to avoid them. Diabetes Metab Syndr 2020;14(5):797-799 [FREE Full text] [doi: 10.1016/j.dsx.2020.06.007] [Medline: 32534432]
70. Abdullah F, Ward R, Ahmed E. Investigating the influence of the most commonly used external variables of TAM on students' perceived ease of use (PEOU) and perceived usefulness (PU) of e-portfolios. Comput Hum Behav 2016 Oct;63:75-90. [doi: 10.1016/j.chb.2016.05.014]
71. Marangunić N, Granić A. Technology acceptance model: a literature review from 1986 to 2013. Univ Access Inf Soc 2014 Feb 16;14(1):81-95. [doi: 10.1007/s10209-014-0348-1]
72. Khatun F, Heywood AE, Ray PK, Hanifi SMA, Bhuiya A, Liaw S. Determinants of readiness to adopt mHealth in a rural community of Bangladesh. Int J Med Inform 2015 Oct;84(10):847-856. [doi: 10.1016/j.ijmedinf.2015.06.008] [Medline: 26194141]
73. Venkatesh V, Sykes T, Zhang X. 'Just what the doctor ordered': a revised UTAUT for EMR system adoption and use by doctors. 2021 Presented at: 2011 44th Hawaii International Conference on System Sciences; January 4-7, 2011; Kauai, HI. [doi: 10.1109/hicss.2011.1]
74. Chismar WG, Wiley-Patton S. Test of the technology acceptance model for the internet in pediatrics. Proc AMIA Symp 2002:155-159 [FREE Full text] [Medline: 12463806]
75. Scherer R, Siddiq F, Tondeur J. The technology acceptance model (TAM): a meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. Comput Educ 2019 Jan;128:13-35. [doi: 10.1016/j.compedu.2018.09.009]
76. Hoque MR, Bao Y, Sorwar G. Investigating factors influencing the adoption of e-Health in developing countries: a patient's perspective. Inform Health Soc Care 2017 Jan;42(1):1-17. [doi: 10.3109/17538157.2015.1075541] [Medline: 26865037]
77. Wai ISH, Ng SSY, Chiu DKW, Ho KKW, Lo P. Exploring undergraduate students' usage pattern of mobile apps for education. J Librariansh Inf Sci 2016 Aug 30;50(1):34-47 [FREE Full text] [doi: 10.1177/0961000616662699]
78. Kalayou MH, Endehabtu BF, Tilahun B. The applicability of the modified technology acceptance model (TAM) on the sustainable adoption of eHealth systems in resource-limited settings. J Multidiscip Healthc 2020;13:1827-1837 [FREE Full text] [doi: 10.2147/JMDH.S284973] [Medline: 33299320]
79. Faber S, van Geenhuizen M, de Reuver M. eHealth adoption factors in medical hospitals: a focus on the Netherlands. Int J Med Inform 2017 Apr;100:77-89. [doi: 10.1016/j.ijmedinf.2017.01.009] [Medline: 28241940]
80. Melas CD, Zampetakis LA, Dimopoulou A, Moustakis V. Modeling the acceptance of clinical information systems among hospital medical staff: an extended TAM model. J Biomed Inform 2011 Aug;44(4):553-564 [FREE Full text] [doi: 10.1016/j.jbi.2011.01.009] [Medline: 21292029]
81. Khan I, Xitong G, Ahmad Z, Shahzad F. Investigating factors impelling the adoption of e-Health: a perspective of African expats in China. SAGE Open 2019 Jul 23;9(3):215824401986580 [FREE Full text] [doi: 10.1177/2158244019865803]
82. Briz-Ponce L, García-Peñalvo FJ. An empirical assessment of a technology acceptance model for apps in medical education. J Med Syst 2015 Nov;39(11):176. [doi: 10.1007/s10916-015-0352-x] [Medline: 26411928]
83. Andriani A, Pink BM. Acceptance of Halodoc's online teleconsultation during covid-19. Manag Sci Field 2022;12(2):1566-1574 [FREE Full text] [doi: 10.1016/b978-0-323-99878-9.09997-0]
84. Garavand A, Aslani N, Nadri H, Abedini S, Dehghan S. Acceptance of telemedicine technology among physicians: a systematic review. Informat Med Unlock 2022;30:100943. [doi: 10.1016/j.imu.2022.100943]

85. Zhao J, Wang J. Health advertising on short-video social media: a study on user attitudes based on the extended technology acceptance model. Int J Environ Res Public Health 2020 Feb 26;17(5):1501 [FREE Full text] [doi: 10.3390/ijerph17051501] [Medline: 32110950]

Abbreviations

CR: critical ratio
EFA: exploratory factor analysis
F2F: face-to-face
ICT: information and communication technology
INT: intention to use teleconsultation
KMO: Kaiser-Meyer-Olkin
NHF: National Health Fund
PEU: perceived ease of use
PHC: primary health care facility
PU: perceived usefulness
Q: quality of work
SAT: satisfaction
SEM: structural equation modeling
TAM: Technology Acceptance Model

Edited by A Kushniruk; submitted 02.04.23; peer-reviewed by E Kim, J Mira; comments to author 01.08.23; revised version received 18.12.23; accepted 20.03.24; published 10.06.24.

Please cite as:
Hawrysz L, Kludacz-Alessandri M, Walczak R
Predictive Factors of Physicians' Satisfaction and Quality of Work Under Teleconsultation Conditions: Structural Equation Analysis
JMIR Hum Factors 2024;11:e47810
URL: https://humanfactors.jmir.org/2024/1/e47810
doi:10.2196/47810
PMID:38857081

Medication Management Initiatives Using Wearable Devices: Scoping Review

Haru Iino, BS; Hayato Kizaki, PhD; Shungo Imai, PhD; Satoko Hori, PhD

Division of Drug Informatics, Faculty of Pharmacy and Graduate School of Pharmaceutical Sciences, Keio University, Tokyo, Japan

Corresponding Author:
Satoko Hori, PhD
Division of Drug Informatics, Faculty of Pharmacy and Graduate School of Pharmaceutical Sciences, Keio University, , Tokyo, , Japan

Abstract

Background: Wearable devices (WDs) have evolved beyond simple fitness trackers to sophisticated health monitors capable of measuring vital signs, such as heart rate and blood oxygen levels. Their application in health care, particularly medication management, is an emerging field poised to significantly enhance patient adherence to treatment regimens. Despite their widespread use and increasing incorporation into clinical trials, a comprehensive review of WDs in terms of medication adherence has not been conducted.

Objective: This study aimed to conduct a comprehensive scoping review to evaluate the impact of WDs on medication adherence across a variety of diseases, summarizing key research findings, outcomes, and challenges encountered.

Methods: Adhering to PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines, a structured search was conducted across MEDLINE, Web of Science, and Embase databases, covering the literature from January 1, 2010, to September 30, 2022. The search strategy was based on terms related to WDs and medication adherence, specifically focusing on empirical studies to ensure the inclusion of original research findings. Studies were selected based on their relevance to medication adherence, usage of WDs in detecting medication-taking actions, and their role in integrated medication management systems.

Results: We screened 657 articles and identified 18 articles. The identified studies demonstrated the diverse applications of WDs in enhancing medication adherence across diseases such as Parkinson disease, diabetes, and cardiovascular conditions. The geographical distribution and publication years of these studies indicate a growing interest in this research area. The studies were divided into three types: (1) studies reporting a correlation between data from WDs or their usage and medication adherence or drug usage as outcomes, (2) studies using WDs to detect the act of medication-taking itself, and (3) studies proposing an integrated medication management system that uses WDs in managing medication.

Conclusions: WDs are increasingly being recognized for their potential to enhance medication management and adherence. This review underscores the need for further empirical research to validate the effectiveness of WDs in real-life settings and explore their use in predicting adherence based on activity rhythms and activities. Despite technological advancements, challenges remain regarding the integration of WDs into routine clinical practice. Future research should focus on leveraging the comprehensive data provided by WDs to develop personalized medication management strategies that can improve patient outcomes.

(JMIR Hum Factors 2024;11:e57652) doi:10.2196/57652

KEYWORDS

medication adherence; scoping review; database search; integrated medication management; drug; pharmacy; pharmacology; pharmacotherapy; pharmaceutics; medication; adherence; wearable; synthesis; review methods; digital health

Introduction

The term “medication adherence” refers to the extent to which patients correctly follow their medication regimens [1-4]. Medication adherence is influenced by lifestyle habits and the potential for using wearable devices (WDs) that record daily activities as activity data has been broadly considered [5-8].

WDs are currently equipped with various types of sensors [9-12]. Beyond standard heart rate or gyro sensors for activity and sleep tracking, some devices can measure or estimate blood pressure, blood glucose levels, blood oxygen levels, and various biomarkers [13-20]. Along with their widespread use, there has been an increasing trend in the use of these devices in health care. According to ClinicalTrials.gov, the number of clinical trials using WDs as outcome measures in drug-related interventions increased from 5, between January 1, 2016, and December 31, 2016, to 21 by 2021, marking a more than fourfold increase [21]. Due to their ability to provide long-term continuous monitoring with minimal burden on patients, there are reports of WDs being used not only in clinical trials but also in managing chronic diseases such as obesity, diabetes, and Parkinson disease, as well as in detecting conditions like COVID-19, cardiovascular diseases, and managing the side effects of chemotherapy [22-28].

In recent years, efforts have been made to use WDs to manage patient medication adherence. Medication management with WDs encompasses a wide range of approaches, including detecting medication-taking actions using motion sensors, using the device as an interface for information notifications, and analyzing various data recorded on the WDs to assess how accurately patients take their medications [29-31].

On the other hand, in the field of medication adherence, there has been a wide range of research methodologies and a lack of focus on specific disease areas, resulting in the absence of a review that provides research guidelines for the field. Therefore, this study seeks to conduct a comprehensive scoping review of WD applications across a spectrum of diseases, aiming to summarize key research findings, outcomes, and challenges encountered in medication management and adherence.

Methods

Databases and Search Strategy

The investigation was conducted in strict adherence to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines (Checklist 1) [32]. Overall, 3 major databases were searched: MEDLINE (through PubMed), Web of Science Core Collection (through Web of Science), and Embase (through ProQuest Dialog). Terms related to WDs and adherence were established separately for the search. These terms were appropriately divided into Medical Subject Headings (MeSH) terms and free words. For instance, in the MEDLINE search, the MeSH terms “wearable electronic devices” and “patient compliance” were used. Notably, under “patient compliance,” the inclusion of “medication adherence” was confirmed (Table 1). The search period was limited from January 1, 2010, to September 30, 2022. When filters were available, literature types such as reviews and editorial materials were excluded from the search.

Table . Search detail (MEDLINE).

Query number	Word	Search details
#1	Wearable electronic devices	"wearable electronic devices"[MeSH[a] Terms] OR ("wearable"[All Fields] AND "electronic"[All Fields] AND "devices"[All Fields]) OR "wearable electronic devices"[All Fields]
#2	wearable devices	"wearable electronic devices"[MeSH Terms] OR ("wearable"[All Fields] AND "electronic"[All Fields] AND "devices"[All Fields]) OR "wearable electronic devices"[All Fields] OR ("wearable"[All Fields] AND "devices"[All Fields]) OR "wearable devices"[All Fields]
#3	"wearable device*"	"wearable device*"[All Fields]
#4	"smart wearable*"	"smart wearable*"[All Fields]
#5	"smart watch"	"smart watch"[All Fields]
#6	fitbit	"fitbit"[All Fields] OR "fitbits"[All Fields]
#7	"apple watch"	"apple watch"[All Fields]
#8	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7
#9	Patient compliance	"patient compliance"[MeSH Terms] OR ("patient"[All Fields] AND "compliance"[All Fields]) OR "patient compliance"[All Fields]
#10	Medication compliance	"medication adherence"[MeSH Terms] OR ("medication"[All Fields] AND "adherence"[All Fields]) OR "medication adherence"[All Fields] OR ("medication"[All Fields] AND "compliance"[All Fields]) OR "medication compliance"[All Fields]
#11	#9 OR #10	#9 OR #10
#12	2010/01/01:2022/09/30 [Date - Publication]	2010/01/01:2022/09/30[Date - Publication]
#13	#8 AND #11 AND #12	#8 AND #11 AND #12
#14	review [Publication type]	"review"[Publication Type]
#15	systematic review [Publication type]	"systematic review"[Publication Type]
#16	#13 NOT (#14 OR #15)	#13 NOT (#14 OR #15)

[a]MeSH: Medical Subject Headings.

Database searches were conducted from November 2022 until December 2023. The search, removal of duplicates, and initial screening were performed by a single author (HI). Screening of potentially compatible references was independently performed by 2 authors (HI and HK). In case of any disagreements, the last author (SH) provided advice. The search process is shown in detail in Multimedia Appendix 1. Management of the literature and removal of duplicates were performed using Zotero (version 6.0.30; open-source software of the Corporation for Digital Scholarship).

Inclusion and Exclusion Criteria for Papers

In this survey, no strict criteria were set for the research design or outcomes so as to encompass a broad range of studies that could potentially relate to the subject. However, each included study met at least one of the following conditions, which were also used to classify the study types in the Results section: (1) studies reporting a correlation between data from WDs or their usage and medication adherence or drug usage as outcomes, (2) studies using WDs to detect the act of medication-taking itself, and (3) studies proposing an integrated medication management system that uses WDs to manage medication. For condition 3, studies were only included if they clearly demonstrated the contribution of WDs to medication management, at least through notifications or other functions; studies that simply used WDs to obtain patient vitals or biometric information were excluded. Additionally, studies that conducted these types of investigations as a sub-analysis were also considered for inclusion in this review, even if it was not their main objective.

Organizing the Results

Owing to the relatively small number of existing studies in this field, there is a wide variation in research methodologies, making it challenging to integrate the results into specific metrics. Therefore, in the Results section, we present which of the 3 study types outlined in the "Inclusion and Exclusion Criteria for Papers" section is satisfied by each selected paper. Furthermore, we provide insights into each study's approach in the Discussion section. This scoping review aimed to provide an overview of the research field and not to analyze the effectiveness of interventions. Therefore, we did not critically assess the methodological quality of the included studies.

Results

A flow diagram of the selection procedure is shown in Figure 1. Following the literature search and screening, we ultimately obtained 18 references. A summary of the literature is presented in Table 2.

Figure 1. Flow diagram of study selection.

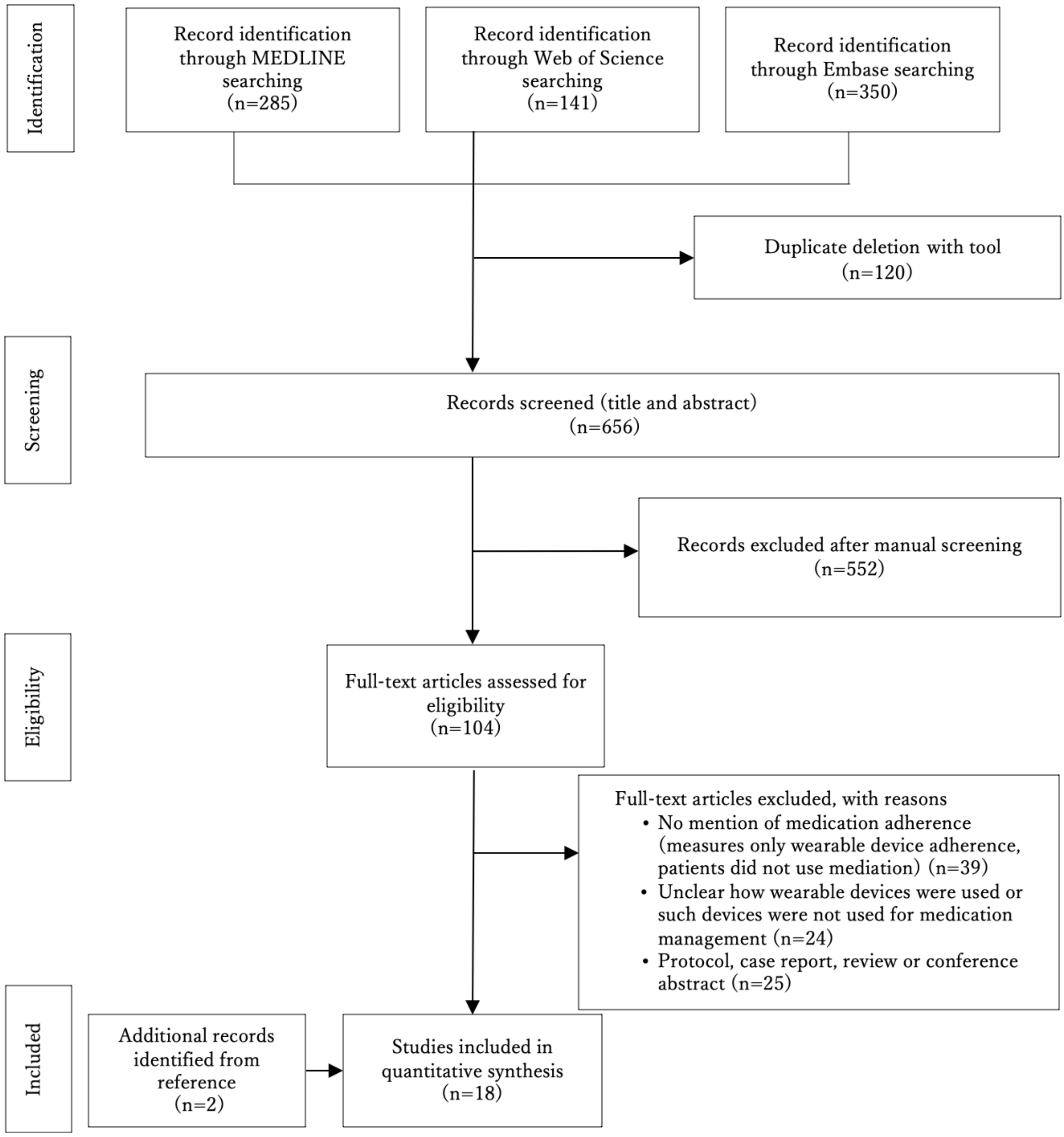


Table . Summary of search results.

Author (published year)	Target population	Findings	Type of wearable device	Sample size, n	Usage of wearable devices	Country	Study type
Agurto et al (2021) [33]	Parkinson disease	It was possible to quantitatively determine the patient medication condition (on/off) due to the speed, acceleration, and symmetry of body movement.	Band (feet, wrist, lumbar, and sternum)	33	Record of acceleration	United States	1, 2
Quisel et al (2019) [31]	Hypertension, diabetes, and dyslipidemia	There is a significant relationship between medication adherence and activity tracking in individuals with chronic illness. Those who use activity tracking have higher medication adherence than those who do not use activity tracking.	Wristband	791 - 10,499	Activity tracking (step count, sleep time)	United States	1
Zhang et al (2020) [34]	Hypertension	Wearable device nonusers have higher medication use scores (*P*<.001) On the other hand, medication use has a negative effect on adherence to the device use (no significant difference).	Wristband	317	Blood pressure monitoring	China	1
Zhang et al (2020) [35]	Hypertension	Wearable device nonusers have higher medication use scores (*P*=.003).	Wristband	212	Blood pressure monitoring	China	1

Author (published year)	Target population	Findings	Type of wearable device	Sample size, n	Usage of wearable devices	Country	Study type
Cochran et al (2021) [36]	Patients with serious mental illness (major depressive disorder, schizophrenia, bipolar 1 disorder)	Higher values of both the previous day's activity rhythm score and activity intensity score characteristics tended to be associated with higher next-day intake rates. Patients with mean activity rhythm scores greater than the patient-level median had higher overall intake rates than those with lower activity rhythm scores (*P*=.004).	Adhesive patch	113	Detection of medication intake, activity tracking (step counts, heart rates)	N/A[a]	1
Belknap et al (2013) [37]	Tuberculosis	The positive detection rate of medication behavior was 95% (95% CI 93.5 - 96.2), and the specificity was 99.7% (95% CI 99.2 - 99.9).	Adhesive patch	30	Detection of medication intake	United States	2
Browne et al (2019) [38]	Tuberculosis	The positive detection accuracy (percentage of correctly identified medications) of wirelessly observed therapy (WOT) was 99.3% (95% CI 98.1 - 100). WOT medication adherence was noninferior to directly observed therapy (DOT) (WOT 95.6% vs DOT 92.7%; *P*=.31).	Adhesive patch	61	Detection of medication intake	United States	2, 3
Lee et al (2021) [39]	General patient (who commonly takes medication)	The medication behavior recognition by the developed model showed an accuracy of 92.7%, a precision of 0.909, and a recall of 0.949.	Wristband (camera attached)	89	Detection of medication intake	South Korea	2

Author (published year)	Target population	Findings	Type of wearable device	Sample size, n	Usage of wearable devices	Country	Study type
Kalantarian et al (2016) [40]	General population[b]	The constructed model was able to accurately classify chewable, saliva swallow, medicine capsules, conversation, and drinking water, with an average accuracy and recall of 90.17% and 88.9%, respectively.	Necklace	N/A[a] (135 instances)	Detection of medication swallow using sensors	N/A	2
Fozoonmayeh et al (2020) [29]	General patient	F_1-scores for the classification of medication activities (tablets, liquid agents) and nonmedication activities (sending SMS text messages, walking, writing, opening, and drinking bottle water) are up to 0.983.	Wristband	24	Tracking acceleration, heart rate, and atmosphere pressure	N/A	2
Spaulding et al (2019) [41]	After acute myocardial infarction	With the integration of the Corrie app and Apple Watch, the participants could be reminded to take drugs and track them directly on their wearable devices.	Wristband	60	Medication record, reminder (notification)[c]	United States	1, 3
da Silva et al (2019) [30]	Hypertension	Sending automatic reminders to patients via commercially available smartphones and smart TVs has reduced the number of missed medication doses and improved treatment compliance.	Wristband	N/A	Reminder (notification)	N/A	3
Levine et al (2019) [42]	Kidney or pancreas or liver transplant recipient (immunosuppressant user)	The use of mobile medical apps in this study did not indicate an increase in medication compliance.	Wristband	108	Reminder (notification)	United States	3

Author (published year)	Target population	Findings	Type of wearable device	Sample size, n	Usage of wearable devices	Country	Study type
DiCarlo et al (2016) [43]	Hypertension	Positive detection accuracy (percentage of correctly identified medications) was 98% (95% CI 96.4-99.1).	Adhesive patch	37	Detection of medication intake	United Kingdom	2, 3
Noble et al (2016) [44]	Hypertension	Wearable device data revealed inappropriate drug use. Of 15 additional patients surveyed, 87% indicated that the device helped improve compliance.	Adhesive patch	54 (15 commercial pharmacies and 39 patients)	Detection of medication intake, activity tracking (rest, activity, and exercise)	United Kingdom	2, 3
Cochran et al (2022) [45]	Schizophrenia	When categorizing engagement with the system as moderate and high engagement, based on medication adherence and device adherence as indicators, the average medication adherence rates were 0.62 for the moderate group and 0.87 for the high engagement group.	Adhesive patch	277	Detection of medication intake, activity tracking (rest, activity, and exercise)	United States	2, 3
Profit et al (2014) [46]	Patients with serious mental illness (major depressive disorder, schizophrenia, bipolar 1 disorder)	In the ingestion detection test, the detection rate of the ingestion sensor by the wearable sensor was 96.6%.	Adhesive patch	29	Detection of medication intake, Activity tracking (rest, activity, and exercise)	N/A	2, 3
Daar et al (2020) [47]	HIV	The system demonstrated its capability to collect real-time ingestion data and automatically send reminder SMS text messages to HIV patients undergoing ARV[d] treatment.	Skin patch (adhesive)	15	Detection of medication	United States	2, 3

[a]N/A: not available.

[b]Although they were not patients, the study was conducted to simulate actual drug-taking behavior.

[c]The document only stated that "Medication adherence is also measured from the smartphone and smartwatch app usage data" and did not provide any details; however, we confirmed the functionality from an external site that provides an overview of the application.

[d]ARV: antiretroviral.

Among the selected literature, 2 (11.1%) papers were published before 2014, 8 (44.4%) papers between 2015 and 2019, and 8 (44.4%) papers between 2020 and 2022. Regarding the geographical distribution of the studies, North America accounted for 8 (44.4%, all from the United States) papers, Europe for 2 (11.1%, both from the United Kingdom) papers, and the Asia-Pacific region for 3 (16.7%, including 2 from China and one from South Korea) papers. However, 5 (27.8%) studies did not specify the study region. In terms of sample size, 6 (33.3%) papers had a sample size of 50 or fewer, 4 (22.2%) papers had between 51 and 100 participants, 2 (11.1%) papers had between 101 and 200 participants, 4 (22.2%) papers had over 200 participants, and 2 (11.1%) papers did not specify the sample size. Additionally, 3 (16.7%) papers did not specify a particular disease as their research subjects, whereas 15 (83.3%) papers focused on specific diseases. Among these, 4 papers addressed neuropsychiatric disorders, 6 (33.3%) addressed lifestyle-related diseases and 3 (16.7%) focused on infectious diseases.

We categorized each study into the following three study types: (1) studies reporting the relationship between WD data or device usage and medication adherence or drug usage as an outcome, (2) studies detecting medication-taking behavior directly using WDs, and (3) studies proposing integrated medication management systems where WDs are used. Below are examples of the study classifications.

Study Type 1: Studies Reporting a Correlation Between Data From WDs or Their Usage and Medication Adherence or Drug Usage as Outcomes

Quisel et al [31] demonstrated that using WDs improved medication adherence among patients with hypertension, diabetes, and hyperlipidemia. Additionally, Cochran et al [36] showed a correlation between activity rhythm scores derived from WD data and medication adherence rates.

Studies classified under study type 1 investigate the relationship between medication adherence outcomes and the use or data of WDs, directly highlighting methodologies to clarify the involvement of WDs in medication adherence. These studies suggest the potential impact of WD use on adherence in real-world settings, as well as the relationship between WD-recorded data and medication adherence.

The studies suggest both positive and negative relationships between WD usage and medication adherence. However, the studies indicating a negative impact on adherence involved new WD distributions at the start of the research, not based on spontaneous usage. While one study suggests high adherence among a large cohort of spontaneous WD users, additional verification with other databases is needed. Studies using WD data reported that features derived from the data were correlated with medication adherence, suggesting the potential to estimate adherence from WD data.

Study Type 2: Studies Using WDs to Detect the Act of Medication-Taking

Browne et al [38] demonstrated that a system combining ingestible sensors and WDs could detect medication-taking behavior with high accuracy. Kalantarian et al [40] developed a method using a necklace-type WD to detect medication-taking behavior through throat movements, achieving an accuracy of 90.17%.

Studies under study type 2 focus on detecting medication-taking behavior itself using WDs. Various methods for detecting medication-taking behavior, such as movement, video, and positional relationships with the medication, have been proposed and verified as practical attempts to contribute to the understanding of medication adherence.

Nearly all studies show that medication-taking behavior can be detected with high accuracy, often exceeding 90%. This is a necessary factor for real-world applicability. However, many of the devices used were proprietary, with only one study using a commercially available WD.

Study Type 3: Studies Proposing an Integrated Medication Management System That Uses WDs to Manage Medication Adherence

Spaulding et al [41] demonstrated that a system combining a smartphone app and WD improved medication adherence in patients post-acute myocardial infarction. This system promoted regular medication intake through reminder functions, contributing to reduced readmission rates.

Studies under study type 3 do not focus solely on WDs but propose systems that manage medication adherence using IoT technology, with WDs serving as one of the interfaces. These studies discuss the development, specific usage, and verification results of medication adherence management systems using WDs, considering real-world applicability.

The studies indicate that WDs are useful interfaces for recording patient adherence. However, most studies used WDs primarily for notification or recording functions, without leveraging the multiple sensors embedded in WDs.

Discussion

Initially, we describe studies that report WD data or associations with device use as an outcome of medication adherence or use (primarily those belonging to study type 1).

Agurto et al [33] investigated the impact of medication on physical activity levels in Parkinson disease patients. By objectively monitoring physical activity, they demonstrated that WDs can distinguish between medicated and non-medicated states. Parkinson disease, characterized by rapid and marked changes in physical activity following dopaminergic medication such as levodopa, is one area where WDs have been actively used [48-50]. However, this study is unique in that it extends the use of wearables to discern medication states.

Quisel et al [31] suggested that higher activity tracking and intensity using WDs are correlated with better medication adherence. Quisel et al [31] and Cochran et al [36] are among the few that have investigated the relationship between activity intensity and adherence. Quisel et al analyzed insurance databases using the proportion of days covered as an adherence measure. However, the proportion of days covered, by only calculating the prescribed days for medication, fails to confirm actual intake, thereby missing instances of nonadherence even when medication is prescribed [51,52]. Another limitation was the exclusion of medication purchases under different insurance plans that were not recorded in the database [31]. These limitations are difficult to overcome in database-based studies, and conducting validation using a different database is necessary to strengthen the results.

Cochran et al [36] scored individuals' activity rhythms or patterns using WD data to analyze their correlation with medication adherence. Although adherence to chronic diseases is expected to be closely related to lifestyle, no studies have objectively measured lifestyle rhythms and analyzed their correlation with medication. The Cochran study showed that more consistent activity rhythms predict better adherence. However, the study's limitations include the use of only 7 days of data and unclear thresholds for activity rhythm categorization, suggesting that further validation is needed for broader application.

Zhang et al conducted 2 parallel studies in 2020 at the same location but with different inclusion criteria: one encompassing all participants and the other focusing on individuals aged over 60 [34,35]. Notably, these studies contrasted with those of Quisel et al, indicating higher medication use scores among individuals not using WDs. However, these studies distributed new devices for the survey, differing from the group that wore the devices voluntarily [31]. Additionally, they focused primarily on device adherence, with medication usage as a secondary inquiry, without mentioning wearable-driven medication management systems. Additionally, this study defined a composite compliance score related to hypertension and reported a positive relationship between the composite compliance score, blood pressure values, and device usage. However, it also reported a negative relationship between medication use and these indicators. It is unlikely that there is a negative correlation between medication use and blood pressure values or hypertension indicators, suggesting the possibility of overlooking important factors such as age.

In type 1 studies, the nature of the target population differs, and the results are contradictory, making it difficult to draw consistent conclusions about the relationship between WD usage and medication adherence at this point. Among populations that regularly use WDs, a positive relationship between the amount of data recorded by the device and medication adherence has been suggested in relatively large samples. Considering the results from populations where WDs were distributed, it is possible that the temperament of those who use WDs reflects their medication adherence, indicating that WD usage itself may not be a factor that improves medication adherence. On the other hand, if appropriate behavioral features are derived from WD data, it may be possible to capture an individual's medication adherence status.

Next, we discussed studies that detected medication-taking behaviors (study type 2). Although the types of data and devices used in these studies vary, all have achieved high accuracy in detection, suggesting that the technical feasibility of medication detection has already been established or is achievable. Some studies have developed both hardware and software independently, while others have only developed software independently. Using custom hardware allows for the collection of more data and the creation of more accurate predictive models tailored to patient outcomes and conditions. However, the costs of development and the lower recognition of these devices could hinder their widespread adoption in the general market.

Noble et al, Belknap et al, Browne et al, DiCarlo et al, Cochran et al, Profit et al, and Daar et al used patch-type devices [37,38,43-47]. These devices detect medication intake using sensors attached to the medication that transmit a signal to a patch worn on the body. Unlike smart pill bottles or detection methods based on arm movements, these patches offer a highly robust method for measuring medication adherence. Among the 7 studies using patch-type devices, 4 specifically reported on the accuracy of medication detection, each achieving an accuracy rate exceeding 95% [37,38,43,46]. This method is highly beneficial for diseases such as tuberculosis and schizophrenia, where continuous medication is crucial; however, the need for a sensor and transmitter for each medication makes it potentially expensive, necessitating a cost-effectiveness analysis for chronic diseases such as hypertension, where immediate adherence is not critical.

Kalantarian et al [40] specifically targeted the action of swallowing to detect medication intake, using a novel approach focused on throat movements. This unique study, which focused on throat movements, recorded a high accuracy of 90.17%. This also suggests that combining this method with smart pill bottles could further improve the accuracy. However, this throat-movement-detecting device, which is not commonly used, may face practical barriers because of its appearance and comfort.

Fozoonmayeh et al [29,39] and Lee et al [39,39] conducted studies to detect medication-taking behavior through arm movements using widely available wristband devices. Fozoonmayeh et al [29] used a commercial smartwatch, which could lower the clinical application barrier owing to its price and availability. Recognizing medication-taking behavior based solely on arm movements is challenging because of similarities with other actions, such as drinking water. However, the F_1-score in this study was 0.983, indicating a highly effective model. Conversely, Lee et al [39] attempted to enhance accuracy by equipping the device with a camera and combining motion detection with image recognition. Although the model's accuracy was lower at 92.7% compared to that of Fozoonmayeh et al [29], image recognition might be necessary to accommodate a variety of medication forms beyond the pills.

Finally, we describe a study proposing an integrated medication management system that uses WDs as part of the system (study type 3). Overall, research in this study supports comprehensive medication management by providing user-friendly interfaces. The primary functions are notifications and reminders, indicating that the role of WDs in medication management is limited. On the other hand, Spaulding et al [41] made it possible to record medication intake directly from a WD display, indicating its role in logging medication. WDs, which provide the most immediate interface for software and systems, have the potential to enhance the efficiency of disease-management systems by expanding their functionality.

These medication management systems are predominantly software-based, and many studies have used commercially available devices [53-55]. However, research such as that by Daar et al [47] involves developing custom devices. Although commercially available devices offer the advantages of being affordable and easily accessible, they have predetermined specifications and data capabilities that may not accommodate all the necessary metrics for certain diseases [11,56]. For instance, Kalantarian et al [40] used a unique necklace-type device to detect the swallowing motion, a form not commonly found in the market. Furthermore, many biochemical markers cannot be measured using commercially available devices, suggesting the need for developing custom devices to realize specific disease management strategies.

Limitations

The limitations of this study include (1) restriction to original articles written in English and published in scientific journals and (2) room for improvement in the categorization of the studies (study type). The field of medication adherence management using WDs is relatively new, and projects in this area are expected to start with small-scale validations. Such studies are not always written in English or published in international journals. During our survey, we screened numerous abstracts from relevant conference presentations and non-English literature that appeared pertinent. However, obtaining complete access to these sources is often challenging, making their inclusion in reviews impractical.

Additionally, the diverse nature of studies on adherence management using WDs led to the categorization of some overlapping research types. Although this can be helpful in understanding the characteristics of the research, it is often duplicative and not a clear-cut classification. While we believe that this classification has minimal impact on the interpretation of the results in this review, defining clearer research directions as the field evolves could potentially facilitate a better understanding of prior studies.

Future Directions

Further research is needed in various regions to understand the relationship between WD usage and medication adherence. On the other hand, the individual data from WDs can objectively capture patient behavior and create features related to adherence, such as daily rhythms and body movements influenced by medication. Currently, there are no reported medication management systems that fully use the various sensors and data WDs possess. However, in the future, by combining the detection of medication-taking actions, it is expected that WD data can be used to estimate individual medication adherence with high accuracy. Exploring better features that can be applied in real-world settings is crucial.

In the studies classified under study type 1 or 2, few have reached the point where they can be used by the general public as commercial products. As mentioned earlier, these studies face many gaps, including the need for basic theory validation, the development of consumer-oriented devices, and cost-effectiveness analysis. Addressing these research gaps with the aim of clinical application is expected to promote the use of WDs and improve medication adherence and patient health outcomes.

Conclusions

Medication management using WDs is currently being implemented based on empirical research, primarily as a simple interface for patient notifications. Technically, the detection of medication-taking behavior has achieved high accuracy, necessitating real-life empirical studies to further leverage this technology. Notably, it has been suggested that specific medication-taking behaviors and daily activity rhythms are related to medication adherence [36]. These findings imply that WDs can predict patients' medication adherence through daily activities and not merely by recognizing isolated medication-taking events. Exploring behavioral data from WDs in future research to clarify the relationship between patient lifestyles and medication practices promises to greatly expand the use of WDs in managing medication adherence.

Acknowledgments

This study was supported by the Japan Science and Technology Agency's Support for Pioneering Research Initiated by the Next Generation (SPRING) program (grant JPMJSP2123).

Authors' Contributions

HI and SH designed this study. HI and HK were responsible for setting the search terms and configuring the search queries, while SH conducted the final review of the search strategy. Screening and integration of results were conducted by HI and HK under the guidance of SI and SH and were reviewed by SI and SH. The integration of results was conducted by HI and HK. HI drafted and completed the manuscript. SH supervised the study. All authors have reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File, 94 KB - humanfactors_v11i1e57652_app1.docx]

Checklist 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist. [DOCX File, 227 KB - humanfactors_v11i1e57652_app2.docx]

References

1. Osterberg L, Blaschke T. Adherence to medication. N Engl J Med 2005 Aug 4;353(5):487-497. [doi: 10.1056/NEJMra050100] [Medline: 16079372]
2. Chakrabarti S. What's in a name? Compliance, adherence and concordance in chronic psychiatric disorders. World J Psychiatry 2014 Jun 22;4(2):30-36. [doi: 10.5498/wjp.v4.i2.30] [Medline: 25019054]
3. World Health Organization. Adherence to Long-Term Therapies: Evidence for Action 2003.
4. Vrijens B, De Geest S, Hughes DA, et al. A new taxonomy for describing and defining adherence to medications. Br J Clin Pharmacol 2012 May;73(5):691-705. [doi: 10.1111/j.1365-2125.2012.04167.x] [Medline: 22486599]
5. Lee YM, Kim RB, Lee HJ, et al. Relationships among medication adherence, lifestyle modification, and health-related quality of life in patients with acute myocardial infarction: a cross-sectional study. Health Qual Life Outcomes 2018 May 22;16(1):100. [doi: 10.1186/s12955-018-0921-z] [Medline: 29788961]
6. Liddelow C, Mullan B, Novoradovskaya E. Exploring medication adherence amongst Australian adults using an extended theory of planned behaviour. Int J Behav Med 2020 Aug;27(4):389-399. [doi: 10.1007/s12529-020-09862-z]
7. Robinson L, Arden MA, Dawson S, Walters SJ, Wildman MJ, Stevenson M. A machine-learning assisted review of the use of habit formation in medication adherence interventions for long-term conditions. Health Psychol Rev 2024 Mar;18(1):1-23. [doi: 10.1080/17437199.2022.2034516] [Medline: 35086431]
8. Conn VS, Ruppar TM. Medication adherence outcomes of 771 intervention trials: systematic review and meta-analysis. Prev Med 2017 Jun;99:269-276. [doi: 10.1016/j.ypmed.2017.03.008] [Medline: 28315760]
9. Yetisen AK, Martinez-Hurtado JL, Ünal B, Khademhosseini A, Butt H. Wearables in medicine. Adv Mater 2018 Jun 11;30(33):e1706910. [doi: 10.1002/adma.201706910] [Medline: 29893068]
10. Lu L, Zhang J, Xie Y, et al. Wearable health devices in health care: narrative systematic review. JMIR Mhealth Uhealth 2020 Nov 9;8(11):e18907. [doi: 10.2196/18907] [Medline: 33164904]
11. Guk K, Han G, Lim J, et al. Evolution of wearable devices with real-time disease monitoring for personalized healthcare. Nanomaterials (Basel) 2019 May 29;9(6):813. [doi: 10.3390/nano9060813] [Medline: 31146479]
12. Sharma A, Badea M, Tiwari S, Marty JL. Wearable biosensors: an alternative and practical approach in healthcare and disease monitoring. Molecules 2021 Feb 1;26(3):748. [doi: 10.3390/molecules26030748] [Medline: 33535493]
13. Han M, Lee YR, Park T, et al. Feasibility and measurement stability of smartwatch-based cuffless blood pressure monitoring: a real-world prospective observational study. Hypertens Res 2023 Apr;46(4):922-931. [doi: 10.1038/s41440-023-01215-z] [Medline: 36781979]
14. Arakawa T. Recent research and developing trends of wearable sensors for detecting blood pressure. Sensors (Basel) 2018 Aug 23;18(9):2772. [doi: 10.3390/s18092772] [Medline: 30142931]
15. Bent B, Cho PJ, Henriquez M, et al. Engineering digital biomarkers of interstitial glucose from noninvasive smartwatches. NPJ Digit Med 2021 Jun 2;4(1):89. [doi: 10.1038/s41746-021-00465-w] [Medline: 34079049]
16. Lee H, Hong YJ, Baik S, Hyeon T, Kim DH. Enzyme-based glucose sensor: from invasive to wearable device. Adv Healthc Mater 2018 Apr;7(8):e1701150. [doi: 10.1002/adhm.201701150] [Medline: 29334198]
17. Kim J, Campbell AS, Wang J. Wearable non-invasive epidermal glucose sensors: a review. Talanta 2018 Jan 15;177:163-170. [doi: 10.1016/j.talanta.2017.08.077] [Medline: 29108571]
18. Prieto-Avalos G, Cruz-Ramos NA, Alor-Hernández G, Sánchez-Cervantes JL, Rodríguez-Mazahua L, Guarneros-Nolasco LR. Wearable devices for physical monitoring of heart: a review. Biosensors (Basel) 2022 May 2;12(5):292. [doi: 10.3390/bios12050292] [Medline: 35624593]
19. Chan PY, Ryan NP, Chen D, McNeil J, Hopper I. Novel wearable and contactless heart rate, respiratory rate, and oxygen saturation monitoring devices: a systematic review and meta-analysis. Anaesthesia 2022 Nov;77(11):1268-1280. [doi: 10.1111/anae.15834] [Medline: 35947876]
20. Sempionatto JR, Lasalde-Ramírez JA, Mahato K, Wang J, Gao W. Wearable chemical sensors for biomarker discovery in the omics era. Nat Rev Chem 2022 Dec;6(12):899-915. [doi: 10.1038/s41570-022-00439-w]
21. ClinicalTrials.gov. URL: https://clinicaltrials.gov/ [accessed 2024-02-14]
22. McDonough DJ, Su X, Gao Z. Health wearable devices for weight and BMI reduction in individuals with overweight/obesity and chronic comorbidities: systematic review and network meta-analysis. Br J Sports Med 2021 Aug;55(16):917-925. [doi: 10.1136/bjsports-2020-103594] [Medline: 33731385]
23. Kwan YH, Yoon S, Tan CS, et al. EMPOWERing Patients With Diabetes Using Profiling and Targeted Feedbacks Delivered Through Smartphone App and Wearable (EMPOWER): protocol for a randomized controlled trial on effectiveness and implementation. Front Public Health 2022;10:805856. [doi: 10.3389/fpubh.2022.805856] [Medline: 35284389]
24. Rovini E, Maremmani C, Cavallo F. How wearable sensors can support Parkinson's disease diagnosis and treatment: a systematic review. Front Neurosci 2017;11:555. [doi: 10.3389/fnins.2017.00555] [Medline: 29056899]

25. Mattison G, Canfell O, Forrester D, et al. The influence of wearables on health care outcomes in chronic disease: systematic review. J Med Internet Res 2022 Jul 1;24(7):e36690. [doi: 10.2196/36690] [Medline: 35776492]
26. Takahashi S, Nakazawa E, Ichinohe S, Akabayashi A, Akabayashi A. Wearable technology for monitoring respiratory rate and SpO2 of COVID-19 patients: a systematic review. Diagnostics (Basel) 2022 Oct 21;12(10):2563. [doi: 10.3390/diagnostics12102563] [Medline: 36292252]
27. Bayoumy K, Gaber M, Elshafeey A, et al. Smart wearable devices in cardiovascular care: where we are and how to move forward. Nat Rev Cardiol 2021 Aug;18(8):581-599. [doi: 10.1038/s41569-021-00522-7] [Medline: 33664502]
28. Mantovani E, Demrozi F, Hertz DL, et al. Wearables, sensors, and smart devices for the detection and monitoring of chemotherapy-induced peripheral neurotoxicity: systematic review and directions for future research. J Peripher Nerv Syst 2022 Dec;27(4):238-258. [doi: 10.1111/jns.12518] [Medline: 36224713]
29. Fozoonmayeh D, Le HV, Wittfoth E, et al. A scalable smartwatch-based medication intake detection system using distributed machine learning. J Med Syst 2020 Feb 28;44(4):76. [doi: 10.1007/s10916-019-1518-8] [Medline: 32112271]
30. da Silva VJ, da Silva Souza V, Guimarães da Cruz R, Mesquita Vidal Martinez de Lucena J, Jazdi N, Ferreira de Lucena Junior V. Commercial devices-based system designed to improve the treatment adherence of hypertensive patients. Sensors (Basel) 2019 Oct 18;19(20):4539. [doi: 10.3390/s19204539] [Medline: 31635394]
31. Quisel T, Foschini L, Zbikowski SM, Juusola JL. The association between medication adherence for chronic conditions and digital health activity tracking: retrospective analysis. J Med Internet Res 2019 Mar 20;21(3):e11486. [doi: 10.2196/11486] [Medline: 30892271]
32. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
33. Agurto C, Heisig S, Abrami A, Ho BK, Caggiano V. Parkinson’s disease medication state and severity assessment based on coordination during walking. PLoS One 2021;16(2):e0244842. [doi: 10.1371/journal.pone.0244842] [Medline: 33596202]
34. Zhang Y, Yang N, Si G, et al. What matters the adherence with BP 24-hr self-monitoring wearable device among hypertensive patients? A population-based survey. Transl Behav Med 2020 Oct 8;10(4):1053-1063. [doi: 10.1093/tbm/ibz069] [Medline: 31167031]
35. Zhang Y, Fang Y, Xu Y, et al. Adherence with blood pressure monitoring wearable device among the elderly with hypertension: the case of rural China. Brain Behav 2020 Jun;10(6):e01599. [doi: 10.1002/brb3.1599] [Medline: 32385965]
36. Cochran JM, Heidary Z, Knights J. Characterization of activity behavior using a digital medicine system and comparison to medication ingestion in patients with serious mental illness. NPJ Digit Med 2021 Apr 6;4(1):63. [doi: 10.1038/s41746-021-00436-1] [Medline: 33824406]
37. Belknap R, Weis S, Brookens A, et al. Feasibility of an ingestible sensor-based system for monitoring adherence to tuberculosis therapy. PLoS One 2013;8(1):e53373. [doi: 10.1371/journal.pone.0053373] [Medline: 23308203]
38. Browne SH, Umlauf A, Tucker AJ, et al. Wirelessly observed therapy compared to directly observed therapy to confirm and support tuberculosis treatment adherence: a randomized controlled trial. PLoS Med 2019 Oct;16(10):e1002891. [doi: 10.1371/journal.pmed.1002891] [Medline: 31584944]
39. Lee H, Youm S. Development of a wearable camera and AI algorithm for medication behavior recognition. Sensors (Basel) 2021 May;21(11):3594. [doi: 10.3390/s21113594]
40. Kalantarian H, Motamed B, Alshurafa N, Sarrafzadeh M. A wearable sensor system for medication adherence prediction. Artif Intell Med 2016 May;69:43-52. [doi: 10.1016/j.artmed.2016.03.004] [Medline: 27068873]
41. Spaulding EM, Marvel FA, Lee MA, et al. Corrie health digital platform for self-management in secondary prevention after acute myocardial infarction. Circ Cardiovasc Qual Outcomes 2019 May;12(5):e005509. [doi: 10.1161/CIRCOUTCOMES.119.005509] [Medline: 31043065]
42. Levine D, Torabi J, Choinski K, Rocca JP, Graham JA. Transplant surgery enters a new era: increasing immunosuppressive medication adherence through mobile apps and smart watches. Am J Surg 2019 Jul;218(1):18-20. [doi: 10.1016/j.amjsurg.2019.02.018] [Medline: 30799019]
43. DiCarlo LA, Weinstein RL, Morimoto CB, et al. Patient-centered home care using digital medicine and telemetric data for hypertension: feasibility and acceptability of objective ambulatory assessment. J Clin Hypertens (Greenwich) 2016 Sep;18(9):901-906. [doi: 10.1111/jch.12787] [Medline: 26890041]
44. Noble K, Brown K, Medina M, et al. Medication adherence and activity patterns underlying uncontrolled hypertension: assessment and recommendations by practicing pharmacists using digital health care. J Am Pharm Assoc 2016 May;56(3):310-315. [doi: 10.1016/j.japh.2016.01.005]
45. Cochran JM, Fang H, Le Gallo C, Peters-Strickland T, Lindenmayer JP, Reuteman-Fowler JC. Participant engagement and symptom improvement: aripiprazole tablets with sensor for the treatment of schizophrenia. Patient Prefer Adherence 2022;16:1805-1817. [doi: 10.2147/PPA.S362889] [Medline: 35923658]
46. Profit D, Rohatagi S, Zhao C, Hatch A, Docherty JP, Peters-Strickland TS. Developing a digital medicine system in psychiatry: ingestion detection rate and latency period. J Clin Psychiatry 2016 Sep;77(9):e1095-e1100. [doi: 10.4088/JCP.16m10643] [Medline: 27379966]

47. Daar ES, Rosen MI, Wang Y, et al. Real-time and wireless assessment of adherence to antiretroviral therapy with co-encapsulated ingestion sensor in HIV-infected patients: a pilot study. Clin Transl Sci 2020 Jan;13(1):189-194. [doi: 10.1111/cts.12701] [Medline: 31536156]
48. Moreau C, Rouaud T, Grabli D, et al. Overview on wearable sensors for the management of Parkinson's disease. NPJ Parkinsons Dis 2023 Nov 2;9(1):153. [doi: 10.1038/s41531-023-00585-y] [Medline: 37919332]
49. Lu R, Xu Y, Li X, et al. Evaluation of wearable sensor devices in Parkinson's disease: a review of current status and future prospects. Parkinsons Dis 2020;2020:4693019. [doi: 10.1155/2020/4693019] [Medline: 33029343]
50. Silva de Lima AL, Evers LJW, Hahn T, et al. Freezing of gait and fall detection in Parkinson's disease using wearable sensors: a systematic review. J Neurol 2017 Aug;264(8):1642-1654. [doi: 10.1007/s00415-017-8424-0] [Medline: 28251357]
51. Bijlsma MJ, Janssen F, Hak E. Estimating time-varying drug adherence using electronic records: extending the proportion of days covered (PDC) method. Pharmacoepidemiol Drug Saf 2016 Mar;25(3):325-332. [doi: 10.1002/pds.3935] [Medline: 26687394]
52. Prieto-Merino D, Mulick A, Armstrong C, et al. Estimating proportion of days covered (PDC) using real-world online medicine suppliers' datasets. J Pharm Policy Pract 2021 Dec 29;14(1):113. [doi: 10.1186/s40545-021-00385-w] [Medline: 34965882]
53. Islam SMS, Mishra V, Siddiqui MU, et al. Smartphone apps for diabetes medication adherence: systematic review. JMIR Diabetes 2022 Jun 21;7(2):e33264. [doi: 10.2196/33264] [Medline: 35727613]
54. Ahmed I, Ahmad NS, Ali S, et al. Medication adherence apps: review and content analysis. JMIR Mhealth Uhealth 2018 Mar 16;6(3):e62. [doi: 10.2196/mhealth.6432] [Medline: 29549075]
55. Aldeer M, Javanmard M, Martin RP. A review of medication adherence monitoring technologies. Appl Syst Innov 2018 Jun;1(2):14. [doi: 10.3390/asi1020014]
56. Contreras-Naranjo JC, Wei Q, Ozcan A. Mobile phone-based microscopy, sensing, and diagnostics. IEEE J Select Topics Quantum Electron 2016 May;22(3):1-14. [doi: 10.1109/JSTQE.2015.2478657]

Abbreviations

MeSH: Medical Subject Headings
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
WD: wearable device

Edited by A Kushniruk, E Borycki; submitted 22.02.24; peer-reviewed by A Austin, B Foroutan, J Moses, S Garner; revised version received 28.08.24; accepted 13.09.24; published 27.11.24.

Please cite as:
Iino H, Kizaki H, Imai S, Hori S
Medication Management Initiatives Using Wearable Devices: Scoping Review
JMIR Hum Factors 2024;11:e57652
URL: https://humanfactors.jmir.org/2024/1/e57652
doi:10.2196/57652

Cocreative Development of Robotic Interaction Systems for Health Care: Scoping Review

Pascal Müller, MSc; Patrick Jahn, Prof Dr

Health Service Research Working Group | Acute Care, Department of Internal Medicine, Faculty of Medicine, University Medicine Halle (Saale), Martin-Luther-University Halle-Wittenberg, Magdeburger Straße 12, Halle (Saale), Germany

Corresponding Author:
Pascal Müller, MSc
Health Service Research Working Group | Acute Care, Department of Internal Medicine, Faculty of Medicine, University Medicine Halle (Saale), Martin-Luther-University Halle-Wittenberg, , Magdeburger Straße 12, Halle (Saale), , Germany

Abstract

Background: Robotic technologies present challenges to health care professionals and are therefore rarely used. Barriers such as lack of controllability and adaptability and complex control functions affect the human-robot relationship. In addition to educational opportunities, the possibility of individual adaptation can improve the usability and practical implementation of robotics. Previous work has focused on developments from a technology-centered perspective and has included user interests too late in the process.

Objective: This study addresses the following research question: What cocreative research approaches are used in the field of nursing robotics to improve the usability, intended use, and goal-directed application of robotic developments for nurses and to support the nursing process?

Methods: This scoping review provides an overview of the topic and the research activities taking place within it. Five databases and the reference lists of the identified publications were searched for studies without further restrictions. Studies were included if they developed and evaluated interaction and control platforms for robotic systems in health care in a cocreative way with end users.

Results: The search resulted in 419 hits, of which 3 publications were included. All publications were feasibility or user studies that were mainly carried out in the European Union. The 3 interaction and control platforms presented were all prototypes and not commercially available. In addition to those in need of care, all studies also included family carers and health care professionals.

Conclusions: Robotic interaction and control platforms in health care are rarely, if ever, developed and evaluated with feasibility or user studies that include prototypes and end users. While the involvement of end users is crucial, this review emphasizes that all stakeholders, including health care professionals, should participate in the development process to ensure a holistic understanding of application needs and a focus on user experiences and practical health care needs. It is emphasized that the active involvement of end users in the development process is critical to effectively meeting the needs of the target group.

Trial Registration: Deutsches Register Klinischer Studien DRKS00034195; https://drks.de/search/de/trial/DRKS00034195

(JMIR Hum Factors 2024;11:e58046) doi:10.2196/58046

KEYWORDS

human-robot interaction; cocreation; robotics; user-centered design; health care

Introduction

The narrative of the digital transformation of health care confronts health care professionals with the challenge of using robotic systems. The challenge is to adapt the technologies to the different and individual needs of patients. Standardized robotic functions, such as those designed for industrial robots, regularly reach the limits of their usability in care situations [1].

For a human-robot relationship to be highly usable, it must be meaningful and effective to the user. The medical journalist Nicole Janke [2] suggests that the main barriers to the use of robotics in health care are the lack of controllability, the lack of adaptability, and the complexity of control functions for changing users, contexts of use, and suitability for the user. The current inflexibility is one of the reasons for the rather low penetration of already available robotic systems in everyday life and, especially, in care. On the one hand, their use will be improved if health care professionals are introduced to the applications in a structured way through tailor-made teaching programs [3]. On the other hand, functionality will be increased if members of these nontechnical professions are given the opportunity to make certain adjustments to individual care situations themselves, thus improving situation-specific usability.

Implementation science is currently shifting from linear and safe development in controlled laboratory environments to more iterative, participatory, and complex models where interventions are developed and evaluated directly in the later field of application [4,5]. Participatory design approaches such as cocreation or design-based research [6] can be a solution to achieve usability and user acceptance. Cocreation is a collaborative approach that involves end users and relevant stakeholders in all phases of a project, from needs analysis and problem definition to the evaluation of prototypes and the final phase of a project [7].

Previous work has addressed cocreative research in technological development from a more general, theoretical perspective, in the context of raising awareness among target groups and identifying needs from a technology-centered perspective. Reference to user interests often occurs only in the testing phase of a finished technology [8-10]. Involving end users as early as possible in the development process can be seen as a way to increase acceptance and have a positive impact not only on patient satisfaction but also on the quality of care [11,12]. In addition, cocreation can increase the success of implementations of evidence-based interventions and policies through equal and deep involvement of end users [13,14]. As a result, social determinants and contextual factors responsible for the feasibility and acceptability of interventions are influenced at the earliest stage of the development cycle [15].

Currently, there are no established methods that address the adaptation of the cocreative development process to the health care or nursing context to address challenges such as collaboration or power structures [8]. To our knowledge, there also have been no systematic reviews focusing on robotic interaction and control platforms in health care. Therefore, this study addresses the following research question: What cocreative research approaches are used in the field of nursing robotics to improve the usability, intended use, and goal-directed application of robotic developments for nurses and to support the nursing process? The following subquestions can be formulated: (1) How is the cocreative process of robotic technology development designed in the context of nursing? (2) How can changes in usability, intended use, and goal-directed application be measured over the course of the development process?

Methods

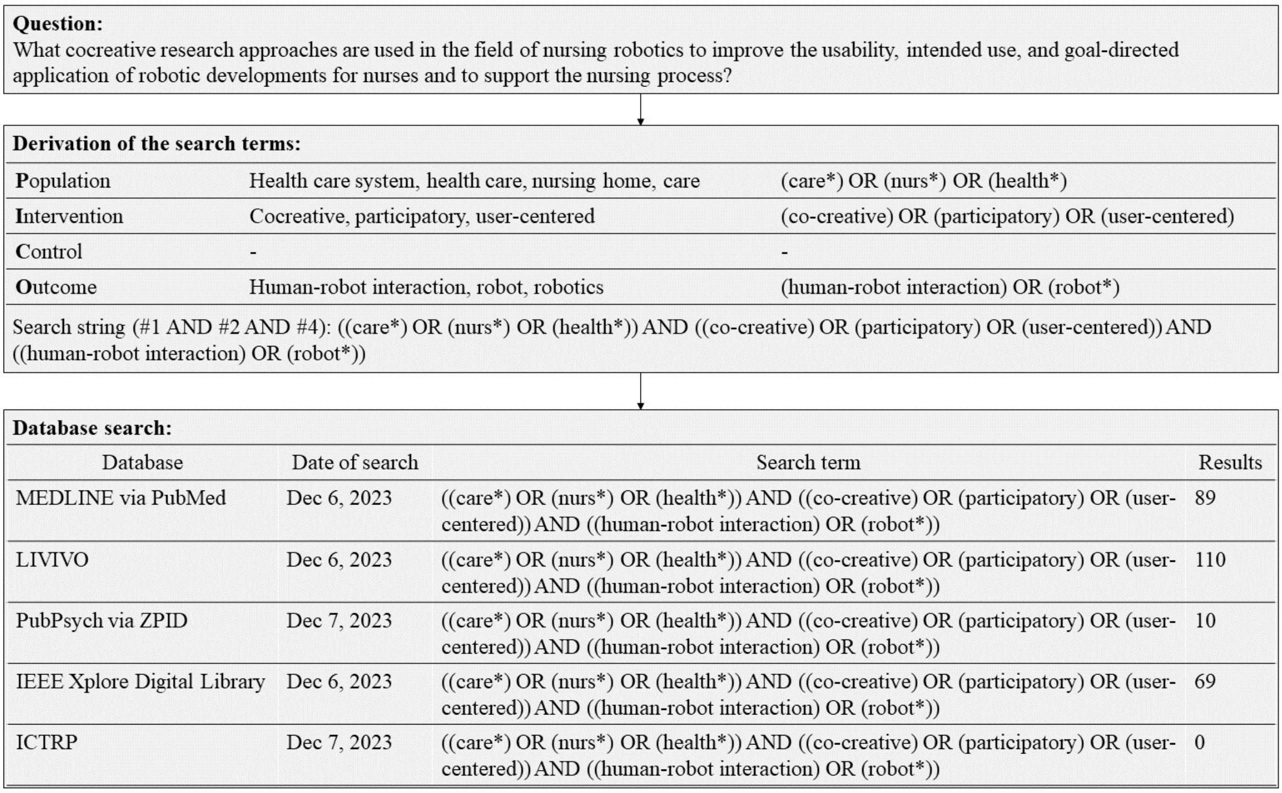
A scoping review was carried out to provide an overview of the research field, outlining the extent and nature of research activity, mapping approaches and key concepts, and identifying research gaps [16]. The content and structure of the report are based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [17].

Search Strategy

A systematic search was conducted of MEDLINE (via PubMed), LIVIVO, PubPsych (via ZPID), and IEEE Xplore Digital Library (via the IEEE website). Trial registrations were searched via the International Clinical Trials Registry Platform (ICTRP). The search terms were derived using the population, intervention, control, outcome (PICO) scheme. Figure 1 illustrates the derivation of the search terms and the database search.

The reference lists of the publications identified by the search were scanned for additional relevant publications. Only publications that reported on robotic interaction and control platforms and described a cocreative development process were included.

Figure 1. Derivation of the search terms and the database search. ICTRP: International Clinical Trials Registry Platform.



Question:
What cocreative research approaches are used in the field of nursing robotics to improve the usability, intended use, and goal-directed application of robotic developments for nurses and to support the nursing process?

Derivation of the search terms:

Population	Health care system, health care, nursing home, care	(care*) OR (nurs*) OR (health*)
Intervention	Cocreative, participatory, user-centered	(co-creative) OR (participatory) OR (user-centered)
Control	-	-
Outcome	Human-robot interaction, robot, robotics	(human-robot interaction) OR (robot*)

Search string (#1 AND #2 AND #4): ((care*) OR (nurs*) OR (health*)) AND ((co-creative) OR (participatory) OR (user-centered)) AND ((human-robot interaction) OR (robot*))

Database search:

Database	Date of search	Search term	Results
MEDLINE via PubMed	Dec 6, 2023	((care*) OR (nurs*) OR (health*)) AND ((co-creative) OR (participatory) OR (user-centered)) AND ((human-robot interaction) OR (robot*))	89
LIVIVO	Dec 6, 2023	((care*) OR (nurs*) OR (health*)) AND ((co-creative) OR (participatory) OR (user-centered)) AND ((human-robot interaction) OR (robot*))	110
PubPsych via ZPID	Dec 7, 2023	((care*) OR (nurs*) OR (health*)) AND ((co-creative) OR (participatory) OR (user-centered)) AND ((human-robot interaction) OR (robot*))	10
IEEE Xplore Digital Library	Dec 6, 2023	((care*) OR (nurs*) OR (health*)) AND ((co-creative) OR (participatory) OR (user-centered)) AND ((human-robot interaction) OR (robot*))	69
ICTRP	Dec 7, 2023	((care*) OR (nurs*) OR (health*)) AND ((co-creative) OR (participatory) OR (user-centered)) AND ((human-robot interaction) OR (robot*))	0

Study Selection

The criteria for the selection of publications are described in Textbox 1. According to the question of this scoping review, robotic interaction and control systems that were developed in a nursing context with end users (nurses) and evaluated in practice were included.

The term *robotic system* used in this paper is based on the ISO 8373:2012 definition. According to this, a robot performs useful tasks autonomously, in the sense that it is able to perform these tasks on the basis of its sensor data without human intervention.

Despite the great progress in robotics, it remains a challenge to synthesize a variety of interaction scenarios (eg, speech, image, text, or movement) in a natural way. Research in human-robot interaction includes both multimodal input signals from humans to robots and multimodal output signals from robots to humans [18]. The aim is to improve the user experience, reduce annoying processes, and promote adoption. The latest research approaches in the field of care also need to be considered in order to translate advances in robotics into practice by developing a natural and adaptive style of interaction [19].

The relevance of a robotic intervention to health care is determined by its structure and services. For example, health services can be provided in the outpatient setting by general practitioners or specialists in nonmedical professions, in the inpatient setting, and as rehabilitation services. In addition to the treatment of diseases, prevention and health promotion are also a focus of the health care system [20,21].

As the end users are explicitly defined as nurses, robotic interventions that are more established in the medical context (eg, surgery) and are subject to different frameworks were excluded. The preliminary research identified a great deal of development of robotic systems and application scenarios for nursing care. However, as the purely technical consideration of application has not yet led to comprehensive implementation, new research approaches such as participation should be brought into focus. Even pure evaluations in cocreative design cannot do justice to the problem and were therefore excluded from this scoping review. The language of publication was not restricted to avoid further reducing the search results of relevant publications.

Textbox 1. Inclusion and exclusion criteria of the search.

Inclusion criteria

- Robotic systems, defined according to ISO 8373:2012
- Health care
- Intervention: interaction and control platform for robots
- Outcome: cocreative development and evaluation

Exclusion criteria

- Cocreative assessment only
- Development of robotic systems
- Intended use: medical procedures

Collection and Analysis of Data

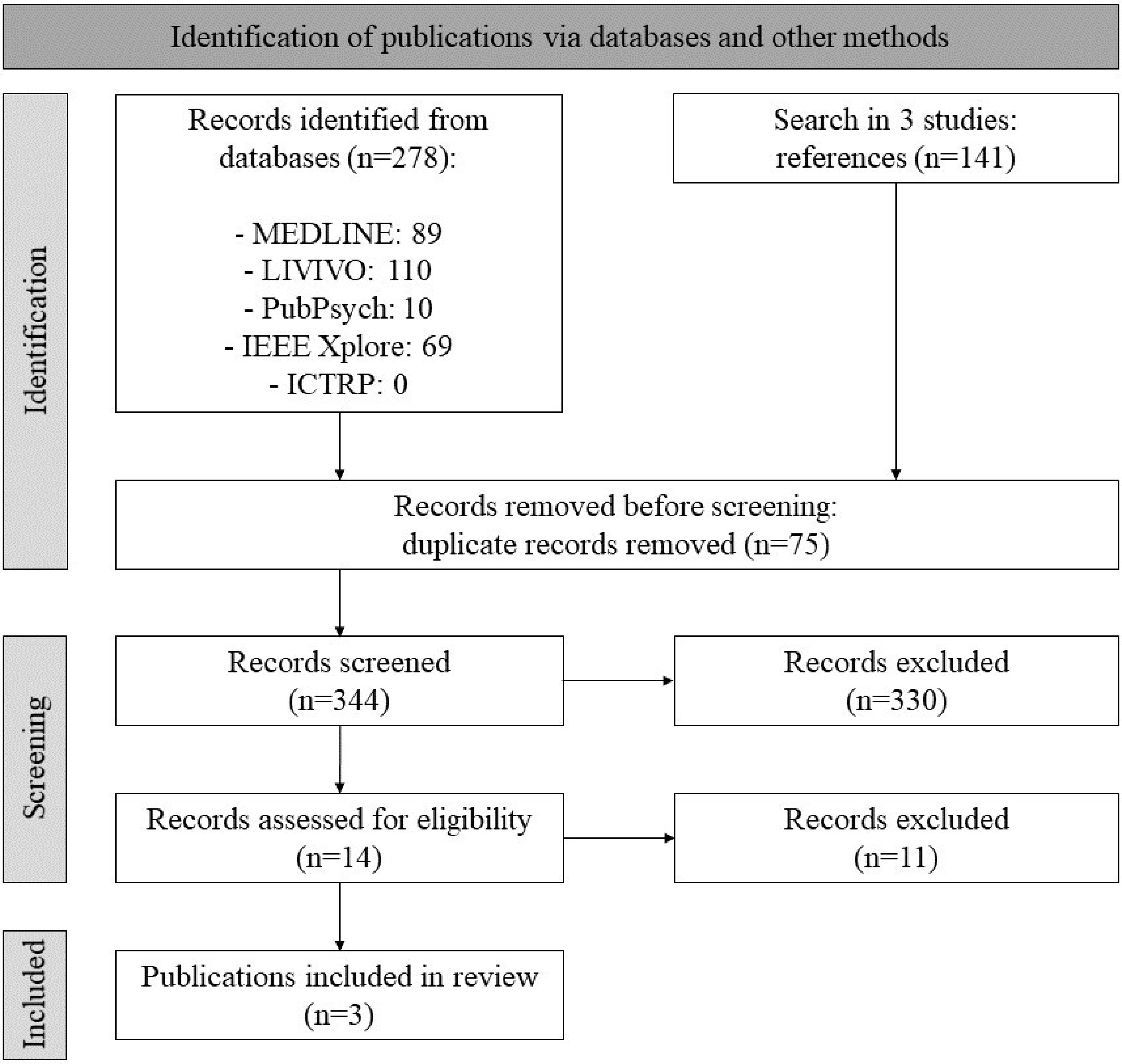
The characteristics of the included studies are summarized and assessed in Multimedia Appendix 1 [22-24]. The summary and preparation of the data can be described as "data-driven thematic analysis," in which prominent and recurring themes in the literature are identified, summarized under thematic headings, and subsumed into a higher-order theoretical structure through induction and interpretation [25,26].

Results

Research and Study Selection

The databases were searched manually by PM in December 2023. There were a total of 419 hits, from which 75 duplicates were removed. The remaining 344 hits were reduced to 14 full texts after title and abstract screening. After checking the full texts for eligibility, 3 studies were included. The screening process was repeated with the references of the 3 included studies. However, no further relevant sources could be identified, so 3 studies were finally included (Figure 2). Uncertainties regarding inclusion were discussed with PJ. The bibliographic data of the included publications and the full table of extracted data can be found in Multimedia Appendix 1 [22-24].

Figure 2. Flowchart of the study selection. ICTRP: International Clinical Trials Registry Platform.



Characteristics of the Included Studies

All publications included were in English. The studies were conducted in 6 countries; 4 of these countries were in the European Union (EU-28), which was represented by 2 publications [22,23]. Both studies were multinational and were partly conducted in Switzerland and Germany. The third included study was conducted in New Zealand [24].

All included publications were feasibility or user studies [22-24]. All had a cocreative study design. This involves users and stakeholders in the development of the robotic interaction and control platform.

One publication did not provide information about ethics committee approval and did not mention that informed consent was obtained from participants or their carers [23].

Characteristics of the Study Participants

In addition to involving end users, cocreative development approaches also require consideration of relevant stakeholders who may interact with the end product and have a legitimate interest in its design. Health care is not an individual service provided by a single professional group, but requires interdisciplinary cooperation between different professions. In the spirit of a stakeholder analysis, the participants in the included publications are identified below.

The 3 included studies reported a total of 299 participants (Table 1), with additional carers including informal carers in one study and both formal and informal carers in another study. In one publication, the term *carer* could not be assigned to formal or informal care. In addition, one study included family members, facility managers, general practitioners, pharmacists, sociologists, geriatricians, psychologists, and computer specialists. One study involved professional teleassistants in a service center. The studies were designed in such a way that the researchers involved the carers in a cocreative way in the development of application scenarios and interaction concepts. People older than 65 years were included in 2 studies, while in 1 study the average age of the participants was only reported as 80.5 years. Children and adolescents were not reported. One study did not report the health status of the participants. The other 2 reported on people with mild to moderate cognitive impairment. One study defined intact hearing and vision as inclusion criteria (Table 1).

Table . Characteristics of the included participants (n=299). The studies included a mean of 99.67 (SD 100.5; range 16-241) participants.

Characteristics	Studies, n	References
Age group (years)		
≥65	2	Margaritini et al [22], Mast et al [23]
18 to 64	0	None
<18	0	None
N/A[a]	1	Tiwari et al [24]
Health status		
Without cognitive impairments	0	None
Cognitive impairments	2	Margaritini et al [22], Tiwari et al [24]
Without hearing or visual impairments	1	Margaritini et al [22]
N/A[a]	1	Mast et al [23]
Other persons involved in care		
None	0	None
Formal caregivers	0	None
Informal caregivers	1	Mast et al [23]
Formal and informal caregivers	1	Margaritini et al [22]
“Caregiver” (not further specified)	1	Tiwari et al [24]
Health care specialists	1	Tiwari et al [24]
Subject-matter experts	2	Mast et al [23], Tiwari et al [24]

[a]N/A: not available.

Characteristics of the Robotic Interaction and Control Platforms

The 3 studies reported on different robotic interaction and control platforms (Textbox 2). All were developed as prototypes and were not commercially available. One study was particularly noteworthy for claiming that the platform that was developed worked for all available service robots. It took into account the dynamic autonomy of the robot and the different digital skills of the target groups. Machine learning could be used to individualize the robot and extend its range of functions [23].

The capabilities and functions of the robotic systems are retrospectively summarized in 4 themes derived from the studies after review (Table 2). Carrying or manipulating objects was described in 1 case. For example, a water bottle was picked up and brought to the user [23]. Reminder functions were described in 2 studies, including for taking medication or for upcoming appointments [22,24]. Monitoring of people was described in 2 studies, including recording of well-being and sleep quality or emergency detection and assessment [22,24]. Communication with third parties via text, audio, or video was described in all scenarios.

Two studies developed an interaction concept consisting of differentiated user interfaces with an adapted range of functions for different end users (people in need of care, carers, and health care professionals) [22,23]. The user interfaces for health care professionals were described as web-based applications in all publications.

Interaction via screens of different sizes using touch gestures was described in all 3 studies. A combination of audio output and written visualization was also chosen as output in all 3 studies. In 2 studies, facial or audio recognition of the target person was used to start the interaction [22,24]. The following requirements were defined for the user interface: a flat menu structures with step-by-step interactions [23]; no foreign or technical language [23]; large and high-contrast colors, buttons, and fonts [22-24]; and the possibility to adjust the volume [22,24].

Textbox 2. Brief description of the robotic interaction and control platforms.

GUARDIAN platform [22]

- Senior App: enables direct feedback to caregivers from the interaction
- Caregiver App: provides the ability to remotely monitor the older person's well-being and activities
- Misty II: serves as a stress sensor for older people

Interaction and control platform for the Care-O-bot 3 [23]

- Care-O-bot 3: development platform
- User Interface for Local Elderly User (UI-LOC): enables autonomous activities
- User Interface for Remote Caregivers (UI-CG): enables semiautonomous navigation, scene-based autonomous manipulation without predefined action sequences, and training of objects and action sequences
- User Interface for Professional Teleassistants (UI-PRO): enables semiautonomous telemanipulation, object training, error handling, and emergency management

Robogen [24]

- Web-based application to support the medication process in home care

Table . Functions of the robotic systems.

Functions	Margaritini et al [22]	Mast et al [23]	Tiwari et al [24]	Studies, n
Carry or manipulate objects		✓		1
Reminders (eg, medications, appointments)	✓		✓	2
Monitoring (eg, well-being, emergencies)	✓		✓	2
Communication (eg, video telephony)	✓	✓	✓	3
Functions, n	3	2	3	

Cocreative Study Design

Margaritini and colleagues [22] divided the cocreation process into 2 phases. In the first, 3-month phase, people with care needs and their carers (formal and informal) used an initial version of the platform. The aim of this phase was to evaluate the response of older people to the use of this new technology. As a result, technical change requests from older people were identified in order to improve the overall experience with the platform and, in particular, the usability of the graphical interface. At the beginning of the study, participants were introduced to the platform by the research team (consisting of psychologists, biomedical engineers, and physiotherapists) and helped to set it up. Monthly evaluation visits were then carried out to collect quantitative and qualitative data. The second phase followed the same participation pattern. The platform was also tested over a period of 3 months.

Mast and colleagues [23] refer to user-centered design in their research approach. A total of 6 user studies were conducted at different stages of the project. Again, 2 main phases can be identified. In the first phase (preparatory studies), mainly user studies and analytical studies were carried out. A needs analysis was carried out with older people and care staff using a mixed methods design. The focus was on the requirements for robotic systems and their integration into the daily life of end users, as well as on the technical requirements. Iterative design was the second phase. After the technical development, the existing prototypes were evaluated with older people, informal carers, and professional teleassistants. For this purpose, a usability study was carried out to evaluate the different user interfaces with representatives of the target group. The results were then used to revise the platform.

The third study, by Tiwari and colleagues [24], had 3 iterative cycles. The first cycle, which defined the framework concepts, involved developing an understanding of the underlying process that the platform was intended to map. This was done by observing nurses at work and then evaluating the information collected. In addition, interviews were conducted with people in need of care, relatives, facility management, nursing staff, doctors, and pharmacists. In the second cycle (the design of the application), prototypes of the application were discussed with computer scientists, sociologists, geriatricians, psychologists, nurses, and doctors. In the final phase of testing and refinement, a usability study was carried out with people in need of care.

The evaluation methods used by the studies (Table 3) collected both quantitative and qualitative data on requirements for robotic interventions, interactions and controls, usability, ease of use, and acceptance, using established methods such as focus groups, think-aloud, and questionnaires. Observational data were collected using video, photographs, and transcripts. Interviews were conducted without information about the use of guidelines. Recordings of the user interfaces of the robotic systems were also used for evaluation. Carers, health professionals, and experts were involved in the evaluation in addition to older people. All publications reported on established evaluation tools.

The interventions were reviewed for their characteristics. Three studies were conducted in the participants' home environment [22-24]. In 1 study, the interaction between the participants and the robotic systems followed a schedule set by the researchers [24]. In the other 2 studies, participants were free to interact with the system [22,23].

The duration of the intervention, in terms of participants' exposure to the robotic system or use per person, was reported in all publications. One study reported 2 intervention cycles of 3 months each [22]. The other 2 studies reported durations of 1 hour [23] and 2 hours [24].

Table . Evaluation methods.

Evaluation method		References
Health questionnaires		
	Short Form Health Survey	Margaritini et al [22]
	Mini-Mental State Examination	Margaritini et al [22]
	Generalized Anxiety Disorder Scale	Margaritini et al [22]
Outcome-related measurements		
	EQ-5D-5L	Margaritini et al [22]
	Social Connectedness Scale	Margaritini et al [22]
	Zarit Burden Interview	Margaritini et al [22]
Evaluation of technology		
	Unified Theory of Acceptance and Use of Technology	Margaritini et al [22]
	Technology Acceptance Model	Margaritini et al [22]
	AttrakDiff questionnaire	Mast et al [23]
Data collection methods		
	Think aloud	Tiwari et al [24]
	Self-created, modified, or unspecified questionnaires	Margaritini et al [22], Mast et al [23], Tiwari et al [24]
	Interviews (persons in need of care, caregivers, health care professionals, experts)	Margaritini et al [22], Mast et al [23], Tiwari et al [24]
	Video or observation logs	Mast et al [23], Tiwari et al [24]
	Robot data (touchscreen)	Tiwari et al [24]
	Focus group	Margaritini et al [22], Mast et al [23], Tiwari et al [24]
	Ethnographic study	Mast et al [23]
	Interaction analysis	Mast et al [23]
	Cognitive status (unspecified)	Tiwari et al [24]

Outcomes

The results of the tests were not described in detail in the trials. One publication merely presented the study protocol for its development [22]. The other 2 studies focused on the presentation of the platforms and the incorporation of the results into development; as subprojects of larger studies, only selected results were presented [23,24]. Technical aspects such as feasibility, usability, suitability, functionality, and specific requests for future features were recorded as outcomes. The user perspective was considered in all 3 studies, with usefulness and interaction with the robot being of interest. Two studies described the preceding requirements analysis, which specified functions and requirements that were crucial for acceptance by the target group [23,24]. An evaluation of the cocreative collaboration between the participants was not mentioned in any of the publications.

Discussion

This scoping review on the state of the art in cocreative development of robotic interaction and control platforms in health care shows that only a few publications have dealt with the interaction and control of robots by end users in a cocreative way. The studies have all been feasibility or user studies of prototypes with target groups including people in need of care, carers, or health care professionals. Only technical aspects such as usability or functionality were described. None of them evaluated the cocreative study design.

The identified studies all focused on home care. This takes into account the expected increasing shortage of skilled care workers. In order to reduce the workload of carers and at the same time enable people to live self-determined and independent lives for as long as possible, their own home is the ideal place [27-29]. There are no research reports on residential care.

The applicability of robotic systems in health care requires interaction and control mechanisms that can be easily adapted to the individual preferences of health care professionals without the need for engineers or programmers. In addition, this interaction and control should be based on familiar concepts such as touch-based screen inputs or audio commands. The aim should be to use robotic interventions to create ethically justifiable and socially acceptable added value that primarily supports and relieves health care professionals in their activities and accompanies those in need of care in their daily lives with dignity. The aim should not be simply to compensate for system deficits and staff shortages [30].

In the development process of such interaction and control platforms, it is important to distance oneself from purely technical solutions and instead rely on the active participation of the end users in order to reduce ethical concerns and avoid developments that do not meet the needs of the target group. All 3 studies included in this review considered the application needs and requirements of the end users in terms of "user-centered design." To this end, a needs analysis was first carried out to understand the underlying processes and requirements of the platform. The focus was on the user interface, that is, the controls and existing technical functions. After technical implementation in the form of prototyping, these were discussed with the end users. Improvements were then made to the platform based on the data collected, and the platform was evaluated again with the target group. In this way, it was possible to design a user-oriented end application that, although not commercially available, at least addressed the acceptance problem as a possible cause of the low penetration of robotics in health care [31].

The available studies have also shown that the involvement of potential end users alone is not sufficient. All stakeholders need to be involved in order to obtain a holistic picture of the intended application. In addition to care recipients and their carers (formal and informal), implementation often has an impact on other professions, such as doctors or therapists. All legitimate stakeholders should therefore be considered and involved in the technical development.

The approach of the included studies represents a change of perspective from a purely technical consideration of functions and control elements to an orientation toward the experiences and needs of end users and practical care. The underlying development process is oriented toward users and their everyday lives as well as the care process, as required by Roland Berger GmbH [32]. The generalization of the care process into standardized procedures is a challenging and complex scenario for robotic interventions due to the human component, that is, the interests of the end user. When aspects such as individualized interactions are added, it becomes impossible for technical developers to meet the requirements and needs of the target group without the involvement of health care professionals in the development process [33,34]. Health care professionals are an important part of the research. With their understanding of diseases and their impact on the lives of those in need of care, as well as their own research expertise, they are asked as potential end users to influence technical developments toward their needs and requirements for interaction and control [35]. The generally skeptical attitude of health care professionals toward technical applications [33,34] explains why the results of this review identified only a few publications on this question, which also referred exclusively to prototypes.

The literature suggests that technical feasibility studies in cocreative design have used both quantitative and qualitative research methods. Quantitative analyses were mainly used to objectively measure the success of the study, while qualitative methods were used selectively and for a specific topic (eg, needs analysis or evaluation of prototypes). The studies did not discuss evaluation of the methodology or presentation of results and processes. Therefore, no comparison with other development concepts is possible. It remains questionable how the stakeholders involved perceived the collaboration and whether the eventual acceptance of the platforms was higher than with previously developed platforms. An evaluation of the attitudes of the health care professionals involved in the studies toward technological interventions could provide additional valuable insights for the implementation of technologies in health care.

The scoping review method used here is only intended to provide an overview of the research field. It does not consider the effectiveness of the studies or derive recommendations for clinical practice [36]. So far, only feasibility studies of robotic interaction and control platforms are available. A systematic assessment of the included studies using a critical appraisal tool was not carried out. This was not part of the research interest.

The operationalization of the research question and the development of the search terms are justified and can be used as a starting point for further literature searches. The process is transparent and fully documented (Figure 1). Due to the resources available in the project, only one reviewer was involved in the selection and assessment of the studies. Despite the care taken, bias cannot be excluded.

Acknowledgments

This work was created as part of the TDG (Translationsregion für digitale Gesundheitsversorgung) project EduXBot (Educational Exploration Robot Application Platform), funded by the Federal Ministry of Education and Research via WIR! – Wandel durch

Innovation in der Region (funding number FKZ 03WIR3118B). We acknowledge the financial support of the Open Access Publication Fund of the Martin-Luther-University Halle-Wittenberg.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data on included publications.
[XLSX File, 61 KB - humanfactors_v11i1e58046_app1.xlsx]

Checklist 1
Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) checklist.
[PDF File, 510 KB - humanfactors_v11i1e58046_app2.pdf]

References

1. Janda V. Usability ist keine eigenschaft von technik. In: Schubert C, Schulz-Schaeffer I, editors. Berliner Schlüssel Zur Techniksoziologie: Springer Fachmedien; 2019:347-374. [doi: 10.1007/978-3-658-22257-4_13]
2. Janke N. Robotik in der Pflege. Verband der Ersatzkassen. 2018. URL: https://www.vdek.com/magazin/ausgaben/2018-0304/robotik.html [accessed 2024-08-28]
3. Geist L, Immenschuh U, Jahn P, et al. Identifikation von lernfördernden maßnahmen zur einführung von digitalen und assistiven technologien (DAT) in prozesse der pflegerischen versorgung: eine qualitative studie. HB Science 2022;13(3-4):152-161. [doi: 10.1007/s16024-022-00372-4]
4. Verloigne M, Altenburg T, Cardon G, et al. Making co-creation a trustworthy methodology for closing the implementation gap between knowledge and action in health promotion: the Health CASCADE project. Perspect Public Health 2023 Jul;143(4):196-198. [doi: 10.1177/17579139221136718] [Medline: 37589328]
5. Gibbons M. Mode 2 society and the emergence of context-sensitive science. Sci Pub Pol 2000 Jun 1;27(3):159-163. [doi: 10.3152/147154300781782011]
6. Armstrong M, Dopp C, Welsh J. Design-based research. In: Kimmons R, Caskurlu S, editors. The Student’s Guide to Learning Design and Research: EdTech Books; 2020:37-41. [doi: 10.59668/10]
7. Vargas C, Whelan J, Brimblecombe J, Allender S. Co-creation, co-design, co-production for public health - a perspective on definition and distinctions. Pub Health Res Pract 2022 Jun 15;32(2):3222211. [doi: 10.17061/phrp3222211] [Medline: 35702744]
8. Dalko K, Kraft B, Jahn P, Schildmann J, Hofstetter S. Cocreation of assistive technologies for patients with long COVID: qualitative analysis of a literature review on the challenges of patient involvement in health and nursing sciences. J Med Internet Res 2023 Aug 15;25:e46297. [doi: 10.2196/46297] [Medline: 37581906]
9. Kehl C. Wege zu verantwortungsvoller forschung und entwicklung im bereich der pflegrobotik: die ambivalente rolle der ethik. In: Bendel O, editor. Pflegeroboter: Springer Gabler; 2018:141-160. [doi: 10.1007/978-3-658-22698-5_8]
10. Maibaum A, Bischof A, Hergesell J, Lipp B. A critique of robotics in health care. AI Soc 2022 Jun;37(2):467-477. [doi: 10.1007/s00146-021-01206-z]
11. Lyon AR, Munson SA, Renn BN, et al. Use of human-centered design to improve implementation of evidence-based psychotherapies in low-resource communities: protocol for studies applying a framework to assess usability. JMIR Res Protoc 2019 Oct 9;8(10):e14990. [doi: 10.2196/14990] [Medline: 31599736]
12. Dopp AR, Parisi KE, Munson SA, Lyon AR. Aligning implementation and user-centered design strategies to enhance the impact of health services: results from a concept mapping study. Impl Sci Commun 2020;1:17. [doi: 10.1186/s43058-020-00020-w] [Medline: 32885179]
13. Flaspohler PD, Meehan C, Maras MA, Keller KE. Ready, willing, and able: developing a support system to promote implementation of school-based prevention programs. Am J Community Psychol 2012 Dec;50(3-4):428-444. [doi: 10.1007/s10464-012-9520-z] [Medline: 22618024]
14. Greenhalgh T, Jackson C, Shaw S, Janamian T. Achieving research impact through co-creation in community-based health services: literature review and case study. Milbank Q 2016 Jun;94(2):392-429. [doi: 10.1111/1468-0009.12197] [Medline: 27265562]
15. Longworth GR, Goh K, Agnello DM, et al. A review of implementation and evaluation frameworks for public health interventions to inform co-creation: a Health CASCADE study. Health Res Policy Syst 2024 Mar 28;22(1):39. [doi: 10.1186/s12961-024-01126-6] [Medline: 38549162]
16. Arksey H, O’Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]

17. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred Reporting Items for Systematic reviews and Meta-Analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097. [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
18. Su H, Qi W, Chen J, Yang C, Sandoval J, Laribi MA. Recent advancements in multimodal human-robot interaction. Front Neurorobot 2023;17:1084000. [doi: 10.3389/fnbot.2023.1084000] [Medline: 37250671]
19. Fang B, Wei X, Sun F, Huang H, Yu Y, Liu H. Skill learning for human-robot interaction using wearable device. Tinshhua Sci Technol 2019;24(6):654-662. [doi: 10.26599/TST.2018.9010096]
20. Das Deutsche gesundheitssystem. Bundesministerium für Gesundheit. 2020. URL: https://tinyurl.com/27dp3j7z [accessed 2023-12-22]
21. Gesundheitsversorgung in Deutschland. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). 2018. URL: https://www.gesundheitsinformation.de/das-deutsche-gesundheitssystem.html [accessed 2023-12-22]
22. Margaritini A, Benadduci M, Amabili G, et al. The social robot companion to support homecare nurses: the guardian study protocol. Contemp Clin Trials Commun 2022 Dec;30:101024. [doi: 10.1016/j.conctc.2022.101024] [Medline: 36340699]
23. Mast M, Burmester M, Kruger K, et al. User-centered design of a dynamic-autonomy remote interaction concept for manipulation-capable robots to assist elderly people in the home. J Hum Robot Interact 2012:96-118. [doi: 10.5898/JHRI.1.1.Mast]
24. Tiwari P, Warren J, Day K. Empowering older patients to engage in self care: designing an interactive robotic device. AMIA Annu Symp Proc 2011;2011:1402-1411. [Medline: 22195203]
25. Dixon-Woods M, Agarwal S, Jones D, Young B, Sutton A. Synthesising qualitative and quantitative evidence: a review of possible methods. J Health Serv Res Policy 2005 Jan;10(1):45-53. [doi: 10.1177/135581960501000110] [Medline: 15667704]
26. Noblit GW, Hare RD. Meta-Ethnography: Synthesizing Qualitative Studies: SAGE Publications; 1988. [doi: 10.4135/9781412985000]
27. Klie T. Expertise, akzeptanz und anteilnahme: voraussetzungen für einen gelingenderen alltag mit demenz - ergebnisse des DAK pflegereports 2017. In: Storm A, editor. Pflegereport 2017 Gutes Leben Mit Demenz: Daten, Erfahrungen Und Praxis: Medhochzwei Verlag GmbH; 2017:2-15.
28. Afentakis A, Maier T. Projektionen des personalbedarfs und -angebots in pflegeberufen bis 2025. In: Wirtschaft Und Statistik: Statistisches Bundesamt; 2010:990-1002.
29. Mestheneos E. Ageing in place in the European Union. Glob Age Iss Act 2011;7(2):17-24 [FREE Full text]
30. Zegelin A, Meyer G. Roboter gegen personalengpässe in der pflege? Pflege 2018 Apr;31(2):61-62. [doi: 10.1024/1012-5302/a000607] [Medline: 29583073]
31. Hülsken-Giesler M. Technik und neue technologien in der pflege. In: Brandenburg H, Dorschner S, Bekel G, Fenchel V, Georg J, editors. Pflegewissenschaft 1 Lehr- Und Arbeitsbuch Zur Einführung in Das Wissenschaftliche Denken in Der Pflege, 3rd edition: Hogrefe Verlag; 2015:262-279.
32. ePflege: informations- und kommunikationstechnologie für die pflege. Roland Berger. 2017. URL: https://www.rolandberger.com/publications/publication_pdf/roland_berger_epflege_abschlussbericht.pdf [accessed 2024-02-01]
33. Papadopoulos I, Koulouglioti C, Lazzarino R, Ali S. Enablers and barriers to the implementation of socially assistive humanoid robots in health and social care: a systematic review. BMJ Open 2020 Jan 9;10(1):e033096. [doi: 10.1136/bmjopen-2019-033096] [Medline: 31924639]
34. Huang R, Li H, Suomi R, Li C, Peltoniemi T. Intelligent physical robots in health care: systematic literature review. J Med Internet Res 2023 Jan 18;25:e39786. [doi: 10.2196/39786] [Medline: 36652280]
35. Paulicke D, Helbig K, Voigt J, Stoevesandt D, Jahn P. Multimodale und interprofessionelle weiterbildungsangebote für ärzte und pflegefachkräfte zur integration technologischer rund robotischer assistenz zur versorgungssicherung für ältere menschen mit demenz in Sachsen-Anhalt. Presented at: 16. Deutscher Kongress für Versorgungsforschung (DKVF); Oct 4-6, 2017; Berlin, Germany.
36. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Info Libr J 2009 Jun;26(2):91-108. [doi: 10.1111/j.1471-1842.2009.00848.x] [Medline: 19490148]

Abbreviations

EU: European Union
ICTRP: International Clinical Trials Registry Platform
PICO: population, intervention, control, outcome
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by A Parush; submitted 04.03.24; peer-reviewed by A Poulsen, T Bock, X Cheng; revised version received 03.06.24; accepted 17.07.24; published 12.09.24.

Please cite as:
Müller P, Jahn P
Cocreative Development of Robotic Interaction Systems for Health Care: Scoping Review
JMIR Hum Factors 2024;11:e58046
URL: https://humanfactors.jmir.org/2024/1/e58046
doi:10.2196/58046

Original Paper

Safety in Teletriage by Nurses and Physicians in the United States and Israel: Narrative Review and Qualitative Study

Motti Haimi[1,2,3], MD, MHA, PhD; Sheila Quilter Wheeler[4], BA, MSc

[1]Rappaport Faculty of Medicine, Technion – Israel Institute of Technology, Haifa, Israel

[2]Health Systems Management Department, The Max Stern Yezreel Valley College, Emek Yezreel, Israel

[3]Meuhedet Healthcare Services – North District, Tel Aviv, Israel

[4]TeleTriage Systems, San Anselmo, CA, United States

Corresponding Author:
Motti Haimi, MD, MHA, PhD
Rappaport Faculty of Medicine
Technion – Israel Institute of Technology
Efron 1, Bat Galim
Haifa, 3200003
Israel
Phone: 972 504557767
Email: morx@netvision.net.il

Abstract

Background: The safety of telemedicine in general and telephone triage (teletriage) safety in particular have been a focus of concern since the 1970s. Today, telehealth, now subsuming teletriage, has a basic structure and process intended to promote safety. However, inadequate telehealth systems may also compromise patient safety. The COVID-19 pandemic accelerated rapid but uneven telehealth growth, both technologically and professionally. Within 5-10 years, the field will likely be more technologically advanced; however, these advances may still outpace professional standards. The need for an evidence-based system is crucial and urgent.

Objective: Our aim was to explore ways that developed teletriage systems produce safe outcomes by examining key system components and questioning long-held assumptions.

Methods: We examined safety by performing a narrative review of the literature using key terms concerning patient safety in teletriage. In addition, we conducted system analysis of 2 typical formal systems, physician led and nurse led, in Israel and the United States, respectively, and evaluated those systems' respective approaches to safety. Additionally, we conducted in-depth interviews with representative physicians and 1 nurse using a qualitative approach.

Results: The review of literature indicated that research on various aspects of telehealth and teletriage safety is still sparse and of variable quality, producing conflicting and inconsistent results. Researchers, possibly unfamiliar with this complicated field, use an array of poorly defined terms and appear to design studies based on unfounded assumptions. The interviews with health care professionals demonstrated several challenges encountered during teletriage, mainly making diagnosis from a distance, treating unfamiliar patients, a stressful atmosphere, working alone, and technological difficulties. However, they reported using several measures that help them make accurate diagnoses and reasonable decisions, thus keeping patient safety, such as using their expertise and intuition, using structured protocols, and considering nonmedical factors and patient preferences (shared decision-making).

Conclusions: Remote encounters about acute, worrisome symptoms are time sensitive, requiring decision-making under conditions of uncertainty and urgency. Patient safety and safe professional practice are extremely important in the field of teletriage, which has a high potential for error. This underregulated subspecialty lacks adequate development and substantive research on system safety. Research may commingle terminology and widely different, ill-defined groups of decision makers with wide variation in decision-making skills, clinical training, experience, and job qualifications, thereby confounding results. The rapid pace of telehealth's technological growth creates urgency in identifying safe systems to guide developers and clinicians about needed improvements.

(JMIR Hum Factors 2024;11:e50676) doi:10.2196/50676

KEYWORDS

telephone triage; teletriage; telehealth; telemedicine; safety; system error; human error; triage; outcome; patient safety

Introduction

Definitions and Terminology

Telemedicine refers to the electronic transmission of medical data from one source to another to promote clinical health. A wide range of services and applications, including 2-way video, email, smartphones, and other communications technology, are included in telemedicine. With the aid of these technologies, patients and caregivers who are geographically separated can communicate and receive treatment, consultation, follow-up, counseling, and health education, as well as engage in medical intervention, monitoring, and remote hospitalization [1,2]. The biggest benefits of telemedicine, aside from cost savings, are expanding patient access to treatment, expanding the availability of medical services, and improving clinicians' efficacy [3].

The delivery and facilitation of health and health-related services, such as medical care, provider and patient education, health information services, and self-care, using telecommunications and digital communication technology is known as *telehealth*.

Although telehealth and telemedicine are frequently used synonymously, the term "telehealth" is used as an umbrella term to refer to all aspects and activities of health care and the health care system that are carried out via telecommunications technology, as opposed to the more specific term "telemedicine," which only refers to the practice of medicine remotely [4,5].

Triage is the process of classifying and prioritizing symptoms. Based on quality, and in the context of health services, triage refers to the process of ranking patients according to their need for care. Using triage services can lower health care costs by preventing patients from making needless and expensive trips to emergency departments (EDs) and by assisting them with self-care and informal care while the doctor is away [6,7].

Telephone triage (assessment and triage of symptoms by telephone) predates telehealth by about 50 years. In the past 5-10 years, the broad industry of telehealth has subsumed telephone triage, which has quickly evolved into *teletriage* to include a wide range of high-tech features (video, biotelemetry, and patient wearables) to enhance remote, brief, but urgent encounters [6,8].

Teletriage is an unscheduled, brief (2-10 minutes), urgent encounter (by telephone only) initiated by patients seeking an estimate of symptom urgency and triage by a clinician to get an urgent on-site evaluation and definitive diagnosis [6,8].

Televisits (via video technology) are now a common substitute for a face-to-face medical appointment and may be 20-30 minutes in length.

Definitions and terms related to telehealth and teletriage are included in Multimedia Appendix 1.

Teletriage: History and Characteristics

Wheeler et al [8] defined teletriage as the complex process of remotely assessing acute, worrisome symptoms to estimate their urgency and to render clinical advice and triage for further evaluation and diagnosis, as appropriate. The goal is to ensure the safe, timely, and appropriate disposition of patient symptoms remotely. This service is accomplished with remote encounters by telephone or real-time video (including biotelemetry). A disposition is a directive from clinician to patient about when and where to be further evaluated and treated. It may also include a risk estimation statement, such as "your symptoms sound urgent," to both inform and motivate patients to comply with the medical advice [8].

Historically, the need for teletriage became an issue when health maintenance organizations (HMOs) realized that they could be more cost-effective by conserving on-site appointments for the sickest patients, which is a form of triage and a way to control access [9].

The overarching goal of medical care (systems and processes) is to use valid, reliable components and experienced trained clinicians to produce safe outcomes. Since the 1970s, clinicians have informally performed teletriage in ambulatory care settings ranging from urgent care and EDs, physician offices, clinics, and student health centers to disease management and ambulatory surgery. Beginning in the midnineties, teletriage and the telehealth industry began developing early systems [10].

One description of teletriage [10] is that it is a time-sensitive, complex, human-technology hybrid process of remote medical decision-making. Currently and in the future, a range of technologies will provide a range of information. On the continuum of care, teletriage can now be acknowledged as the entry point to clinical care. It legitimately qualifies as "prehospital care."

By discussing treatment alternatives and the need for care, teletriage aims to identify the most appropriate degree of care that is needed. These alternatives could involve self-care or informal care, normal or emergency doctor visits, prompt referral to the ED or emergency clinics, or ambulance dispatch, depending on the data collected during evaluations. To support self-care and informal care, teletriage services may also entail providing information and help for difficult medical decisions, as well as managing symptoms. In a variety of medical facilities, nurses or doctors conduct teletriage [11,12].

The teletriage system, which is primarily run by nurses, determines the level of medical urgency and the type of health care that is needed when patients are contacted by telephone. This system is crucial for delivering affordable, effective, and secure health care [13].

The decision-making process is difficult and stressful in the emergency teletriage scenario because decisions must be made quickly and are dependent on nonvisible, unreliable, and incomplete information and nonvisible indicators. Additionally, patients' capacities for clear symptom communication differ, particularly when young patients are involved. The lack of precise criteria for making the decision further increases the difficulty of the decision-making process [14,15].

Over the past few years, numerous Western nations and sizable corporations have started to offer primary health care services after regular business hours. In 2020, the COVID-19 pandemic accelerated telehealth growth exponentially. Almost overnight, telehealth rapidly became an established, essential service [16].

Currently, many US health plans provide advice lines. These services are advertised as a benefit of patients' health plans. Advice lines offer clinicians' advice for patients who have concerns about acute or worrisome symptoms who are calling from home. A telephone call to an after-hours advice line is typically the patients' first attempt to gain access—a medical consultation for a symptom that patients interpret as urgent. However, standards are still lacking for clinical decision makers, their experience, qualifications, clinical training, and practice [8,17,18].

Once considered embryonic, telehealth now appears to be in an adolescent phase. It is rapidly and erratically growing, and technology is outpacing clinical standards. Telehealth appears to be undergoing an identity crisis [9].

Opposing forces—technology, cost containment, and safe clinical practice—now struggle to claim control of the field, one so new that regulation cannot keep pace. There are inherent risks in the clinical task—remote, rapid, clinical decision-making using software that serves technological interests but may not serve clinical safety [18].

In the United States, these forces are quite evident, as health care needs to save money may be at odds with patients' needs for access. Furthermore, health care institutions may attempt to limit patients' use of ED, urgent care, office, and clinic services to be cost-effective or to use less costly (and less qualified) staff in the telehealth process, thereby reducing safety [10,19].

In the United States, evidence-based electronic guidelines have not yet emerged. No telehealth-based professional organization yet exists. Some agencies have developed regulations, including the Utilization Review Accreditation Commission (URAC) [20], the American Academy of Ambulatory Care Nursing [21], the American College of Emergency Physicians (ACEP) [22], the American Nurses Association (ANA) [23], the Emergency Nurses Association [24], and the North American Nursing Diagnosis Association (NANDA) [25]. Inadequately designed technology can lead to unintended consequences, while field testing may not be adequate [26,27]. The ACEP [22] has developed descriptions and broad classifications of emergent to acute symptom patterns for on-site triage. There is a need for a similar classification for teletriage.

Clearly, we are in early days in telehealth research, with the need to define meaningful measures for safe outcomes (Table 1).

Table 1. Safe outcomes.

Outcome	Description
AR[a]	A timely, safe disposition: right place, right time, for the right reason. ARs are considered safe.
OR[b]	A referral deemed (retrospectively) by some to be unnecessary at the time and place initially recommended. ORs are likely safe but not cost-effective.
UR[c]	A referral to a lower level of care than is safe or timely, often resulting in a delay in care. URs have the potential to cause or result in patient harm [8].

[a]AR: appropriate referral.

[b]OR: overreferral.

[c]UR: underreferral.

Controversies have emerged in relation to referrals (outcomes). Both appropriate referrals (ARs) and overreferrals (ORs) are considered safe, but ORs are not cost-effective—a less desirable outcome. Some experts suspect that doctors of medicine (MDs) are reluctant to define safety or to criticize other MDs'/researchers' work.

Without a consensus on safe outcomes that are evidence based, it will be difficult for the industry to make meaningful progress toward the goal of safety. Research on meaningful safe outcomes is needed. We chose to discuss teletriage safety for several reasons: patient safety and safe practice are important topics, and teletriage has a high potential for error.

Teletriage involves making medical decisions under conditions of uncertainty and urgency. Teletriage has also conflicting goals: the institutions' need to control costs (especially in the United States), while also ensuring patient access to care in a safe, timely manner. Furthermore, this underregulated subspecialty lacks adequate research on system safety.

Our review and analysis present a glimpse of current safety through analyzing 2 developed representative systems in 2 countries: Israel (Clalit Health Services) and the United States (Redwood Healthcare Plan [RHP]). We examined each system to learn how developed each system is and to explore the elements that might influence safe practice and patient outcomes.

This might be the first study to review and compare 2 formal teletriage systems. Both authors have performed triage in formal systems and taught and consulted in telehealth for a combined 50 years. Teletriage was the focus of this study. We focused on urgent, time-sensitive calls from patients or their families regarding acute or worrisome symptoms. We believe the telehealth industry and telemedicine can benefit from our findings.

Methods

Study Design

This study included 2 parts:

- A narrative review of the literature: studies describing nurses' and physicians' teletriage systems from the United States and Israel
- Qualitative assessment, including interviews with several physicians (Israel) and 1 nurse (the United States)

Part 1: Narrative Review

Both key authors of this paper have practiced in the field of telehealth; thus, they have a reality-based perspective on the subspecialty.

We restricted our review to system features (structure, process, and outcomes) to provide a more orderly review in this variable and broad field. In this narrative review, we discussed the various facets of and challenges in teletriage, with special focus on the United States and Israel, which serve as representatives for teletriage for nurses and physicians, respectively.

Search Terms

Using the following key search terms, we searched PubMed, Medline, and Google Scholar for papers relevant to this review: Telephone Triage AND Teletriage AND Telehealth AND Telemedicine AND Telecare (and Tele-Triage); + Safety, + Systems, + Physician-led, + Nurse-led. + System Error, + Human Error.

Selection Criteria

It was essential that we study developed systems, because even today, US telehealth practice is still typically unregulated—variously devoid of complete or evidence-based components, such as guidelines, formal documentation, qualified staff, clinical training, and standards—in many office, clinic, ED, and ambulatory care settings.

For our critical analysis, we focused on the best examples of current practice: large, formal clinical call centers or HMOs. These services in the United States provide 24–7 clinical call coverage.

We narrowed our review of the literature to studies that focused either directly or generally on teletriage safety. See Multimedia Appendix 2 for criteria for selecting papers. Only English language publications that were published in scholarly journals or organizations between 1970 and 2023 were included. All types of papers were considered, including original papers, reports of randomized clinical trials, observational studies, and editorials or essays by key opinion leaders.

A summary of early research (1977 to the 1990s) focused mostly on the physician practice of teletriage. A recent review (the 1990s to 2022) summarized and critiqued teletriage safety research.

Part 2: Qualitative Assessment

Using a qualitative approach, we conducted interviews with 15 representative physicians who worked in a pediatric teletriage service (Clalit Health Services) in Israel. In addition, we interviewed 1 nurse who worked in a nurse teltriage service in the United States.

To obtain their subjective perspectives on maintaining patient safety in this setting, the physicians were asked about factors that may have impacted their reaching a "correct" diagnosis and deciding on reasonable and appropriate treatment.

To gather detailed and accurate information that would accurately reflect the participants' subjective experiences, we used a semistructured qualitative study (SSQS) technique in this study. Participants' replies were evaluated and analyzed thematically when themes were found.

The use of open-ended questions, which gave the study its qualitative quality, allowed participants to candidly discuss the challenges they encounter in teletriage settings and the strategies they use to ensure patient safety.

The research complied with the Standards for Qualitative Research (SRQR) items [28]. We examined the responses using qualitative content analysis, which is a systematic procedure for collecting and analyzing qualitative data. Using a consistent set of codes to group texts with comparable content and creating themes and subcategories within themes from participant replies, this technique aims to "answer questions such as what, why, and how, and the common patterns in the data are searched for" [29].

Ethical Considerations

Informed consent was obtained from the physicians and nurse participating in the qualitative section of the study. All necessary approvals for this study were obtained from the Ethics Committees of Clalit Health Services and the University of Haifa (approval numbers 0031-16COM2 and 458/16, respectively).

Results

Telemedicine and Teletriage Growth Surge During the COVID-19 Pandemic

Telemedicine, or the use of digital and remote medical technologies to connect patients and caregivers, has become the hottest and most talked-about area of technology, thanks to the COVID-19 pandemic. The influence of the pandemic on the area of telemedicine worldwide is best summarized by the *New York Times* headline "10 Years of Change in One Week: Telemedicine on Fast Track" dated April 20, 2020.

COVID-19 plagued the world for most of 2020, posing a serious threat to public health. Although many health organizations were primarily focused on combating the immediate effects of COVID-19, maintaining basic and vital therapeutic services was equally important. Initial responses in many nations included clinic closures and the suspension of all noncritical medical services [30,31].

Telemedicine provides ongoing medical care, while maintaining strict social distancing. To reduce their exposure to others and still obtain medical care, patients at risk may benefit from staying at home. As a result, it is not surprising that health care systems worldwide are turning to telemedicine, which has led to an exponential surge in its use as opposed to a previously slow uptake of the novel practice [32,33]. Thus, because of the COVID-19 pandemic, teletriage services have been implemented more frequently [34].

The benefits of teletriage during the COVID-19 pandemic have been described in recent studies; these studies show that this technique removes face-to-face contact, lowers the danger of exposure for medical personnel and other patients, and conserves scarce resources. Results suggest that more investigation is needed to ascertain how teletriage affects clinical outcomes, expenditures, and the use of follow-up care [35,36].

Although the COVID-19 pandemic has fueled the awareness and growth of technology and televisits, which are a convenience and infection control, the COVID-19 period has not made teletriage systems safer. It has made technology proliferate explosively.

Teletriage: First Point of Access to Care

Patients call advice lines for a reason. They want to know whether their symptoms are urgent. Clinical decision makers assess the symptoms, estimate the urgency, triage the symptoms, and advise when, where, and why the patient should be seen. Teletriage is designed specifically for this purpose—estimating symptom urgency and triage to ensure timely access to care. On the continuum of care, teletriage can now be acknowledged as the entry point to clinical care. It legitimately qualifies as "prehospital care" [8].

The primary function of teletriage is the assessment and management of symptoms by telephone, which also calls for expert judgment, clinical evaluation, and proactive information gathering from the patient [6,37].

According to researchers, nurses estimate and rule out symptom urgency to determine a disposition by using pattern recognition. "Telephonic medical diagnosis of patients' problems" is what telephone medicine, as practiced by doctors, is defined as [15,38].

Teletriage System Safety

The task in teletriage is to safely assess symptoms, estimate the urgency, and triage the symptoms presented remotely and then advise a disposition (time and place) for them to be further evaluated. The goal is to "make good decisions under conditions of uncertainty and urgency" to avoid the risk of delay in care, diagnosis, or treatment. Compared to in-person consultations, teletriage is a complex activity that entails certain inherent dangers because there is no visual contact and no nonverbal communication [39-41].

While performing teletriage, nurses must rely on audio signals rather than visual ones, although patients can speak about their symptoms using different terms. The ability of clinicians to communicate effectively is crucial, but there are also several other abilities that must be present, including the ability to recognize verbal cues, concentrate on obtaining a focused history, and understand the importance of having proper documentation [14,42,43].

Other characteristics of after-hours care that could pose risks include a high patient call volume, a variety of clinical conditions presented, the likelihood of urgent conditions being present, unknown patients, knowledge gaps regarding patients' medical histories, and the potential for information transfer discontinuity. Concerns have been raised because teletriage might compromise patient safety [44-48].

Regarding the reliability and safety of teletriage services, several recent studies have produced contradictory findings. Some studies were pessimistic, reporting that patient safety is frequently jeopardized by teletriage decisions [49]; service providers do not always forward the case to the on-call physician, when necessary [50]; and only a small number of diagnostic and therapeutic choices made during teletriage consultations offer the same level of health care as in-person conversations [51]. Inadequate visual cues that help doctors identify patients in acute condition were indicated as patient safety hazards in a study using teletriage [40].

However, more reassuring findings have been reported by other studies on the safety of teletriage systems. For instance, Blank et al [52] reviewed studies in which telephone counsel was contrasted with professional advice that is thought to be acceptable in that circumstance (ie, the "gold standard" of professional advice). The accuracy/appropriateness rate was 44%-98% in this review, with a median of 75%. Most decisions were appropriate according to a different study [14].

Concerning teletriage system effectiveness, the evidence also points to a variety of outcomes. According to certain studies, teletriage interventions, particularly for parents of small children and for older patients with chronic diseases, significantly reduce the number of emergency visits and readmissions [53,54]. Additionally, patients have stated that teletriage services have gained their trust and satisfaction. One study, however, found that a significant portion of patients who were directed to the ED using teletriage may have been treated elsewhere [55].

Based on a summary of several systematic reviews, when considered as a whole, the available research does not offer conclusive answers to queries concerning the standard of care delivered, the equity of access, costs, or outcomes in teletriage settings [18].

Growth alone in a new subspecialty will not guarantee safety. Developing a safe system is essential to any subspecialty, especially teletriage and telehealth. Defining the new subspecialty is one of the first challenges and sets the stage for transparency and, later, safety [14].

Even with the use of video and other technologies, remote symptom assessment is a uniquely risky task. Fraught with uncertainty, and many unknowns, teletriage is extremely time driven and time sensitive. A delay in care can be lethal if a required follow-up evaluation and treatment are not performed in a timely manner. In addition, teletriage is still in an underdeveloped state and lacks a reliable system. Finally, nurses and physicians perform this decision-making task under surprisingly difficult conditions [14,43-46].

Human factors in teletriage that challenge and possibly impair clinicians' decision-making process are detailed in Textbox 1.

Textbox 1. Telehealth risks (human factors).

- Inability to see patients (technology dependent)
- Ability to see but not to touch or gather patient vital signs (technology dependent)
- Extreme brevity of patient encounters (5-15 minutes)
- Incomplete or inaccurate information provided by patients
- Extensive sensory deprivation (endured by clinicians; technology dependent)
- Physical and cognitive demands imposed by high call volumes
- Potential for decision fatigue due to call volume and repetitive nature of the task [26,56]
- Clinicians often not knowing the patient, their education level, or their likelihood of compliance with advice
- A lack of structure (standardization of process and structure)
- Institutional pressures on clinicians to act as a gatekeeper rather than an access facilitator
- User-unfriendly electronic and paper guidelines

One way to avoid the risk of delays in care is to create a system. The Donabedian model [57-59] provides a framework for examining and evaluating health service quality. According to Donabedian [57-59], information about the quality of care can be drawn from 3 categories: structure, process, and outcomes.

Like other subspecialties, teletriage requires certain components to support safety. These components include standards (policies and procedures); sufficient numbers of qualified, experienced clinicians; specialized clinical training in medical decision-making; evidence-based, transparent, user-friendly guidelines; and electronic medical records (EMRs), audiotapes, or written documentation.

System Components: Evidence of a Duty to Care

Not surprisingly, in malpractice cases (when an error has occurred and a patient has been injured or has died), expert witnesses for the plaintiff always request tangible evidence of the system [9]:

- Guidelines used in the call (paper or electronic, eg, computerized decision support system [CDSS])
- Qualified experienced clinicians: résumés of nurses who managed the call, adequate numbers of clinicians
- Standards or policies and procedures, including job descriptions and qualifications
- Call center standards
- Actual call documentation: EMRs, paper form, or transcription of audiotaped calls
- Clinical teletriage training program materials

System Error

System error is thought to be the worst form of medical error [26]. Determining the effect of safety requires an examination of the problem of system error, defined as a failure of systems, processes, or conditions that are intended to prevent errors from occurring and that might lead people to make mistakes [60].

The Institute of Medicine (IOM) [60] has broadly defined system error as the “wrong match of plan” or the “failure to use any plan” to prevent error. For example, IOM research shows that the after-hours time, when no system in place, is especially risky in the United States [60]. In telehealth, complete systems (process or structure) are a first step toward reducing system error. Complete does not imply evidence-based or quality systems, however [8].

Malpractice in Teletriage

When a patient is harmed through unsafe telepractice, a malpractice case ensues. The plaintiff’s expert witnesses request evidence of care for that event: all documents that provide evidence of an adequate system, as described before. Institutions that can produce evidence of care are more able to demonstrate fulfillment of the duty of due care.

Physician teletriage malpractice may be related to the lack of a basic, complete teletriage system [16,49,61-65]. Nurse teletriage malpractice may be related to both the lack of a complete system or practicing in a complete system made up of faulty components [6,8,44,66-68].

What Are Meaningful Outcome Measures?

“We don’t look for patterns of our recurrent mistakes, or devise and refine potential solutions for them. But we could, and that is the ultimate point” [69].

We know what error and near misses look like. However, we have not yet clearly defined what constitutes safe practice and outcomes. Many researchers define telehealth safety variably, based on medical consensus on a study-by-study basis. Research continues to focus on nonessential elements of the process or structure (ie, communications, type of practitioner, patient compliance, and satisfaction).

The unfortunate outcomes described in malpractice [70-73] serve as fragments of the larger picture—system error, the essential and underrecognized problem.

Historically, medicine and nursing adhere to the key obligation “First, do no harm.” Nonmaleficence, which is derived from the maxim, is one of the principal precepts of bioethics—a fundamental principle worldwide.

Currently, professional organizations, such as the ANA [23] and the American Medical Association (AMA), typically set standards to guide medical decision-making, ethical practice, and patient safety. Formal systems—evidence-based structures and methods and guidelines—support clinicians' safe practice and promote safe outcomes. Such system components are evolving slowly.

Safety Studies on US Teletriage

Research on the safety of teletriage systems, whether practiced by registered nurses (RNs) or MDs, is scarce [54].

Safety Research in the United States

Early studies examining the system structure and process provide a basis to inform research on system error. Although safety is often a topic of telehealth research, to the best of our knowledge, system error is still underresearched.

It is likely that the proprietary nature of telehealth technologies interferes with research on system safety. Telehealth trends make it difficult to achieve system transparency. The field urgently needs evidence-based CDSSs, EMRs, and other new technologies, such as features that provide feedback on outcomes to clinicians for the purpose of learning from their mistakes or successes.

In addition, CDSS, computerized decision-making system (CDMS) and EMR components, so fundamental to the clinical decision-making process, make it essential that these technologies be demonstrably and verifiably safe and effective. Questions remain about the safety of guideline technologies [74].

Early Research (1977-1990)

Early studies on teletriage focused on physician practice. Predictably, key demographic groups of frequent calls included infants and children, the elderly, and women. Topics also included categories of symptomatic calls and urgent situations: the sudden, rapid death of children, calls to the ED and poison centers, postpartum concerns, suicidal callers, and cases resulting in malpractice [9].

The first studies on remote telephone encounters often focused on problems that plagued physicians: strategies for reducing inappropriate after-hours calls, follow-up postdischarge calls, characteristics and perceptions of after-hours callers and high users ("frequent flyers"), call patterns, and dissatisfaction in pediatric practice. In general, US physicians were dissatisfied with the task of teletriage [9].

Research by Perrin and Goodman [75] marked the beginning of a change in how teletriage was practiced in the United States. The study compared the teletriage practices of pediatric nurse practitioners (PNPs) with those of pediatricians. Researchers found that PNPs are as safe and proficient as physicians, although PNPs take slightly more time to manage calls.

Research in 1990-2000

Research later focused on nurses' safety: communication, close calls, malpractice claims, access, chest pain, the influence of after-hours calls, and clinical and nonclinical decision makers. Later, the first teletriage training manual for nurses was published [9]. Lephrohon and Patel [14] showed how nurses practicing teletriage made decisions, describing pattern recognition and estimation of urgency as key decision-making strategies.

Research in 2000-2023

In the 2000s, rudimentary systems emerged [8]. Research highlighted the field's disorganization and lack of professional development [76]. Patient safety research was inconsistent and of variable quality, often commingling widely different clinical and nonclinical decision makers, intermingling terminology, and making unquestioned assumptions. Evidence-based studies were sparse.

A recently published systematic review [77] assessed the effectiveness of teletriage as one of these technologies during the COVID-19 pandemic. Studies investigating teletriage's effect on patient safety, clinical outcomes, and patient satisfaction were included. The authors concluded that teletriage interventions reduce unnecessary visits, improve clinical outcomes, reduce mortality and injuries, increase patient satisfaction, reduce health care provider workload, improve access to primary care consultation, and increase patient safety and satisfaction.

In Multimedia Appendix 3, we describe a developed teletriage center in the United States and include an interview with a qualified nurse working in this call center. Throughout the interview, she describes her personal feelings and reflections. Table 2 describes the required education, key system components, decision-making strategies, and goals of both Israeli physicians and US nurses.

Table 2. Decision maker comparison: Israeli MDs[a] and US RNs[b].[c]

Decision maker	Minimum qualifications	System components	Decision-making strategies	Task objectives
Physician (autonomous, licensed clinician)	Doctorate level: 7 years of science-based clinical education and training + pediatrics specialty training for 4.5 years	Documentation: • Regulation: state medical board clinical training, guidelines	• Diagnosis • Clinical judgment • Critical thinking	• Make a medical diagnosis. • Identify and verify emergencies and urgencies.
Licensed nurse (autonomous, licensed clinician supported by a medically developed CDSS[d])	Associate of arts (AA)/bachelor of science (BS)/master of science (MS)/doctor of nursing practice (DNP): 2-7 years science-based clinical education and training	≥3 components • Guidelines: CDSS • Documentation: EMRs[e]/audiotaping, clinical teletriage training • Practice standards: American Academy of Ambulatory Care Nursing (AAACN) • Call center standards: URAC[f] • Regulation: Board of Registered Nursing	• Pattern recognition • Clinical judgment • Contextual information • Nursing process • Critical thinking	• Identify and verify emergencies and urgencies. • Estimate symptom urgency. • Rule out symptom urgency. • Interpret patient responses.

[a]MD: doctor of medicine.

[b]RN: registered nurse.

[c]Partially adapted from Wheeler [8], with permission.

[d]CDSS: computerized decision support system.

[e]EMR: electronic medical record.

[f]URAC: Utilization Review Accreditation Commission.

Telemedicine and Teletriage in Israel

In Israel, most of the health care and social assistance is public, including health care, welfare, child support, and old age and disability benefits. The national mandatory statutory health insurance system used in Israel is based on the Bismarck model. Both designated and ordinary taxes are used to pay for it. All citizens are required to join 1 of the 4 health plans (also known as mutualities or sick funds). The health plans provide both insurance for their members and a public basket of services, either through operating their own services or entering into contracts with service providers [78-80]. All 4 health plans are fully computerized, and all doctors and most other health care providers use EMRs that either are directly linked to the central medical record of the health plan through the internet or comprise its whole internal system. Between all community services, there is practically complete clinical data sharing. Highly developed decision support systems help with these.

Each health plan has highly advanced personal health records that allow members to access their own medical data online. These data entail prescription drug purchases and visits to the doctor, as well as imaging, laboratory, and other diagnostic test findings. Most of this is presently available online and via a smartphone in at least 2 of the health plans. Based on medical data and protocols created by the health plans, these plans currently provide proactive warnings and reminders for their members. The doctors at Maccabi, the second-largest health plan, can view their computerized medical information using a smartphone [78].

In Israel, physicians typically provide for all telehealth services, referred to as telemedicine. The physician practice of medicine or telemedicine is a range of remote high-tech remote encounters. The Ministry of Health (MOH) in Israel has regulations that apply to telemedicine services. Telemedicine standards were released in 2012 and have since been revised, as necessary, for different medical specialties. The MOH [79] provided an update in 2019 that details requirements for providing medical care remotely.

Although the worldwide pandemic has significantly accelerated what appears to be the next digital medical revolution, Israel has long recognized the enormous potential of telemedicine and has made it a national priority by allocating significant resources, establishing pertinent regulations, and promoting partnerships between health organizations, research institutions, start-up businesses, and independent researchers.

"Digital Health as an Engine of Growth" is a national priority program that Israel declared in March 2018. By using the information and communication technologies that are readily available to the entire Israeli population, the Israeli MOH [80] has stated that it is its mission to "bring about a leap in the health system that will enable it to become sustainable, advanced, innovative, renewed, and constantly improving." In other words, the opportunity to further implement and expand a variety of telemedicine solutions is created by the worldwide acceleration of technology development and the digital revolution. The realization of the significance of digital health for the efficiency of the health care system and the requirement to offer strategic, systemic, and all-encompassing solutions for the foreseeable future are embodied in this national priority program [80].

Israel benefits greatly from a mix of human resources, a sizable number of businesses engaged in the development of digital medicine, and a sizable investment in research and development (R&D). It is a leader in communications and cyber innovation, which is essential to the creation of cutting-edge digital medicine that will be used worldwide.

The conditions for the successful implementation of telemedicine in Israel are encouraging: the population has individual identification numbers, digital medical records are stored in sizable databases, all people have access to medical insurance, the standard of medicine is high, and communication technologies are of high quality and are widely available throughout the country [81].

In Israel, all health plans operate telemedicine services in one form or another. For administrative requirements with the clinic and the attending personal physician, they all permit online services. With each of them, the attending physician can also be reached via telephone or video call during clinic hours and sometimes even after hours.

Additionally, several of the health plans offer online pediatric and family services that primarily act as medical triage after working hours, throughout the evenings, nights, and weekends. The patients can use telephone or video calls and occasionally even submit images during the online consultations [82,83].

Some health plans have also begun using the TytoCare test device, which enables online physical assessment. During a digital visit, the equipment checks the patient's heart rate, respiration, temperature, ears, throat, and skin lesions using a variety of medical devices. A few Israeli hospitals have already begun to offer telemedicine consultations, particularly for presurgery evaluations, follow-up care, genetic and dietitian consultations, and even remote rehabilitation.

The quality of the telemedicine service provided and its safety are now the 2 most important factors to consider. Some telemedicine promotion initiatives during the pandemic seem to be predicated on the idea that a sizable part of outpatient visits may be effectively managed remotely, and patients can be prioritized for telemedicine services without endangering their safety or the standard of care [84].

An Israeli study [85] emphasized the growth in telemedicine usage during the first COVID-19 lockdown in Israel, as well as the anticipated partial fall in usage following the pandemic's end. As of May 2020, most Israeli pediatricians recommended that once the pandemic has passed, they return to in-person consultations and base their therapeutic judgments on frontal data rather than on data obtained through telemedicine contacts [85].

There are not many studies on the safety of telemedicine or teletriage services conducted in Israel. Haimi et al [84] examined the level of safety of a pediatric telemedicine service, paying particular attention to the accuracy of the diagnoses and the reasonability of judgments made by the online doctors. This service serves as a time-sensitive teletriage of spontaneous calls from parents about acute, worrisome symptoms of their children that require triage (symptom sorting). The study showed high levels of diagnosis accuracy (98.5%) and decision reasonableness (92%).

In addition to the literature review, using a qualitative study, we interviewed 15 physicians who had worked at the Clalit Pediatricians Online Service (a teletriage service) over the past 5 years [82-84,86]. Using a semistructured interview protocol form, we questioned the physicians about the difficulties and obstacles they face in the teletriage setting that may affect their capability in maintaining patient safety. In addition, they were asked about their perceptions of their capacity to uphold patient safety in this teletriage environment and, in particular, regarding elements that impacted their capacity to make reasonable decisions, determine the best course of action, and diagnose accurately, while upholding patient safety.

The physicians described several difficulties they face in the teletriage setting that may impact their ability to maintain patient safety [84]. The main factor was the difficulty to make a diagnosis from a distance due to the physician's inability to perform a physical examination in the telemedicine setting. Additional factors were treating unfamiliar patients, working alone, working under stressful conditions, having technological difficulties, and having a moral conflict between their desire to please and provide parents with good service on the one hand and the wish to maintain good medical practices on the other. While describing the challenges they face the teletriage setting, the physicians described various techniques and tools that they use to ensure patient safety.

Using a thematic analysis, we used the participants' replies to determine themes. These themes were compared with the original transcriptions to determine whether they accurately reflected the original data, guaranteeing a constant flow. The following themes were gleaned from the interviews with the 15 physicians:

- Use of intuition: Many physicians claimed to have used their intuition during the diagnostic process and frequently in relation to parents.

  You learn to rely on your intuition … whether you feel that the parents understand what you are saying, or that in this case, your instructions won't help.

  There is adversity, especially regarding certain decisions—I am sometimes hesitant about what to do, since I'm alone, especially at nights, and have to rely a lot on my intuition.

- Expertise: Most medical professionals believe that their clinical expertise in pediatrics in general and in telemedicine in particular aids in their diagnostic and decision-making processes. The more experience a medical professional has in telemedicine, the more confident they feel.

  During my first few days at work, I was afraid I would miss things or that there would be problems. After a while, however, I began to work with more confidence and less stress.

  There are some difficult aspects. At first, I felt insecure, but over time I gained experience (even the ability to diagnose better than the face-to-face doctor)! Like diagnosing a child with diabetic ketoacidosis …

- Using protocols: Many physicians said they use protocols and rules of thumb when making decisions. Most also use the protocols that are generated for special circumstances.

They believed this assisted them in maintaining patient safety. They were also conscious of potential biases in their thinking.

> *I use protocols. For example, head injuries among babies under the age of six months, or a high fever among babies younger than one month old. These make it easier to make a decision.*
>
> *I use some rules of thumb. For instance, if a young boy is able to jump around, then he does not have appendicitis.*

- Making shared decisions with parents: A few medical professionals reported talking to the parents of their patients about their opinions on the diagnostic process and potential treatment options.

> *I used to share my decision-making process with the parents. If there were several options, I would let the parents decide. In such a case, I depend on them.*
>
> *I usually share, but I do not consult. I give my opinion and explain it, and only then do I wait for feedback.*

- Using nonmedical factors: Most of the physicians agreed that they consider nonmedical considerations, in addition to medical factors when making decisions. Their opinions of the parents, particularly their level of comprehension, anxiety level, health literacy level, and the assurance that the parents will act appropriately if the child's illness worsens, are the most important considerations. The family's ability to access medical care was another crucial nonmedical element.

> *In addition to medical factors, the parents' tone of voice and level of stress may affect my decision, even if it seems to be a simple diagnosis … Language is also a factor. For example, new immigrants do not always understand me, and I am therefore more prone to sending them to the ED …*
>
> *Aside from the medical condition, the patient's place of residency is also important. Living far from a medical care facility is a factor, and I will be more likely to consider an ED referral. In such cases, I also ask more questions about the availability of the doctor nearby.*
>
> *You have to trust the parents' information and rely on them to follow the instructions correctly. If I feel that the chances of me being understood are poor (due to a lack of understanding or oversophistication on the part of the parents), I will refer them to the ED more easily.*

- Additional techniques: The physicians schedule video conversations with the parents in cases of diagnostic doubt, ask them to send digital images, or schedule a follow-up call a few hours later.

> *If I needed additional information, I would arrange a video call or a follow-up call at a later time. Rarely would I consult with a senior physician.*
>
> *Despite the difficulty making the decision, pictures and videos often compensate for the lack of a physical examination … In one case, I managed to correctly diagnose a child with intussusception!*

Despite the difficulties and obstacles mentioned by online doctors [79], many of the physicians surveyed in this study reported having generally positive experiences with their telephone assessments and feeling confident in their ability to conduct thorough assessments and reach the right treatment decisions.

The key conclusions, with examples and comparisons between the 2 systems, are shown in Table 3.

Table 3. Key conclusions derived from the findings.

Key topics	Findings
Specialized clinical training for teletriage tasks	
RHP[a]	The RHP does not provide formal specialized teletriage training for nurses. However, it requires formal training for its electronic algorithms. Physicians present lectures on various specialties for the nurses.
Clalit Health Services	Teletriage training for pediatricians is not available. The authors believe training would aid pediatricians in making safer decisions during online consultations.
Conclusion for both systems	Judging from the interviews with nurses and physicians, it appears that both systems' clinical training is not adequate and formal training would be beneficial. Clinical training for any new subspecialty is an essential safety measure. Research has shown that clinical preparation has the potential to build confidence, improve performance, and reduce error, while improving morale [70-72,75,82,83].
Electronic algorithms and protocols	
RHP	With rare exception, the RHP requires nurses to follow and heavily rely on electronic algorithms in decision making. This raises the question of whether the RHP's electronic algorithms function more as a CDMS[b] than as a CDSS[c] [73]. The nurse interviewed (Ms Finley) stated that the overreliance on algorithms discourages nurses' critical thinking and dampens her initiative to perform a more thorough preliminary symptom assessment and to promote interpersonal interactions.
Clalit Health Services	The Clalit system provides several written protocols for certain clinical scenarios, and physicians are encouraged but not required to use them. In our qualitative interviews, many physicians said they used protocols and rules of thumb when making decisions.
Conclusion for both systems	For both nurses and physicians, guidelines are a key decision support tool. In addition, guideline quality (validity and reliability) requires evidence-based research—long overdue in this risk-prone field.
Documentation	
RHP	The RHP system provides 2 methods for call documentation, an audiotape recording and an electronic paper trail—a record of the patient-clinician encounter derived from a given guideline. However, the documentation output is limited to a patient's yes/no responses to the algorithmic questions. The result is an anonymized history with few details or context specific to a given patient [26]. Finley stated that physicians who later evaluate patients on-site do not have a good sense of why the patients were advised to be seen urgently. The RHP later developed a new policy allowing nurses to use a free-text area to document a brief symptom history using standard questions to elicit more specific details and context. Quality assurance is further bolstered by audiotaping all calls for follow-up review.
Clalit Health Services	The Clalit system requires physicians to document calls, completed in the child's medical file. As a result, the personal physician can view the online consultation during business hours. However, the language used in the documentation is completely up to the individual physician.
Conclusion for both systems	The RHP "paper trail" appears safer and more complete. However, the documented output appears to introduce confusion into on-site follow-up encounters. Clalit Health Services' lack of standardized language requirement may interfere with communication and continuity of care—a professional principle. Both systems are inadequate and increase miscommunication—one of the most common, recurrent error in this field.
Clinical call center standards (policies and procedures): clinicians' knowledge and experience	
RHP	According to Finley's interview, the RHP appears to have no job requirements or job descriptions and according to its policy may hire inexperienced nurse graduates. New nursing graduates are a poor match for the medical decision-making task, which according to many experts, requires a minimum 5-year bedside experience.
Clalit Health Services	The Clalit system hires only certified pediatricians, even though their level of experience as pediatricians in general and as online physicians may vary greatly.
Conclusion for both systems	Experience is critical in decision-making. Both groups could benefit from improved standards for required experience and job qualifications.
Clinical call center standards (policies and procedures): call length (teletriage meeting duration)	
RHP	Although it is the customary role of management to develop call center policies and procedures (standards), at the RHP, staff nurses have developed a minimal number of standards. One is a maximum call length, while another is a closing reminder to callers to call back if symptoms worsen or change.
Clalit Health Services	The Clalit system does not place any constraints on session length. However, since physicians are paid "per consultation," it may be an incentive to process calls quickly, although using the best medical decision.

Key topics	Findings
Clinical call center standards (policies and procedures): patient outcome feedback	
RHP	RHP nurses operate in a vacuum regarding patient outcomes (follow-up diagnosis). Outcomes provide feedback and are a measure of patient safety. Feedback about one's decisions is essential to improved practice and one of the strongest risk management measures available [87]. The rationale for not providing feedback to nurses is based on the Health Insurance Portability and Accountability Act (HIPAA). This federal law does not prevent US physicians' access to patient outcomes, however.
Clalit Health Services	Clalit physicians have complete access to the outcomes of their calls. Learning of their mistakes or successes may improve their practice and safety.
Conclusion for both systems	Ignorance about outcomes of one's decisions has never been shown to improve practice. Feedback mechanisms, known as planned error recovery, not only allow practitioners to learn the final diagnosis and thus improve their practice but also may improve guideline design and quality.

[a]RHP: Redwood Healthcare Plan.

[b]CDMS: computerized decision-making system.

[c]CDSS: computerized decision support system.

Discussion

Principal Findings

This narrative review and analysis presented a glimpse into current teletriage safety by analyzing 2 established and representative systems in 2 countries: Israel and the United States. We examined each system to learn how developed each is, perform a comparative analysis of both systems' safety, and explore the elements that might influence safe practice and patient outcomes.

In the initial stage, we carried out a thorough analysis of papers pertaining to patient safety in teletriage scenarios. Current research yields conflicting results regarding the dependability and security of teletriage systems. Although some critics claim that teletriage decisions frequently endanger patient safety [40,49-51], other research claims that using teletriage systems results in better safety outcomes [14,52].

We also analyzed a clinical call center of a large national US HMO based on the responses of a representative advice nurse to an interview (Multimedia Appendix 3), highlighting areas of risk that may contribute to system error [17]. We found that this representative system is still underdeveloped and lacks certain risk management elements. We based our conclusions on the interview, recent research, legal and risk management requirements related to the duty of due care, medical and nursing traditions, and existing subspecialty structures and processes.

In addition, we performed a qualitative study in which we interviewed 15 Israeli physicians working in a pediatric teletriage service in Israel, asking them about factors that affect their ability to maintain patient safety, while providing an accurate diagnosis, making appropriate decisions, and choosing the best course of action [83]. The physicians discussed the challenges they encounter in the telemedicine/teletriage context and the many strategies they use to arrive at the best diagnosis and course of care, protecting patient safety. These strategies include using their experience and intuition, using protocols generated for special clinical scenarios, making shared decisions with the patients (or their parents in the case of children), applying nonmedical criteria to aid in decision-making in situations where the medical data are ambiguous, and using more sophisticated tools (eg, video chats) when additional details are required. Many of the physicians surveyed in this study reported having generally positive experiences with their telephone assessments and feeling confident in their ability to conduct thorough assessments and make the best treatment decisions, despite the challenges and blockages described [82].

This study may be the first to examine and compare 2 official telehealth systems. For a combined 45-50 years, the 2 authors have performed triage in formal systems, taught, and provided consultation in the field of telehealth.

Teletriage, as stressed in this research, is the process of evaluating and prioritizing symptoms using telecommunication technologies. The main goal of teletriage is to assess and manage symptoms by telephone, which necessitates the use of professional judgment, clinical assessment, and proactive patient information gathering. The purpose of teletriage is to determine whether the needed on-site evaluation should take place and, if so, the venue and time. Teletriage involves clinical decision-making under remote and uncertain conditions. An overarching goal of teletriage safety is to avoid delays in care or diagnosis, which can cause patient harm.

Clinicians typically estimate the urgency of acute symptoms remotely and advise a disposition (triage level) for further medical diagnosis and treatment, as appropriate. The growth of teletriage services has accelerated due to the COVID-19 outbreak.

All types of health care delivery must consider safety, but with teletriage, this is both more crucial and challenging because acute symptoms may be time sensitive. Delay in care and diagnosis can result in harm to patients. Since there is no visible contact or nonverbal communication during teletriage, it is a more complicated activity than in-person consultations and it has certain inherent risks. The rapid pace of telehealth's growth creates urgency in identifying safe systems to guide developers and clinicians about needed improvement. Establishing a system is a key strategy to reduce the possibility of delay in care and diagnosis.

In the United States and internationally, one way to be cost-effective is to use the least paid person who can *safely* do the job—an RN. Internationally, nurses have traditionally performed this task since the late 1980s. Early studies found that nurses are a safe substitute for physicians [14,73]. Thus, although physicians initially performed this task, they later delegated it to nurses.

Health care institutions historically provide standard features to support nurses and to enhance safety (subspecialty clinical training, standards, and documentation). In the case of teletriage, guidelines are typically written by physicians, similar to standing orders. These components provide a structure and process for this subspecialty and underpin safe practice.

An evolving subspecialty, even after 50 years, teletriage appears misunderstood and neglected. System error is thought to be a result of the absence or inadequacy of systems. In malpractice cases, expert witnesses for the patient or their family request evidence of the duty of due care. Typically, this evidence comprises documents: call documentation, guidelines used, clinical training materials, policies, and procedures (standards), including written job descriptions and qualifications.

Clearly, this analysis must acknowledge that contexts of the institutions described here differ in terms of respective health care systems and decision makers' clinical qualifications. The US health care system, and teletriage in particular, is plagued by disparate, competing forces: institutional cost containment, the need for professional standards, and diverging technological goals—the emphasis of speed over safety. This scenario requires better risk management.

Israel has universal health care, which appears to act differently. Physicians' depth and breadth of education and clinical training are superior to those of nurses. The US health care system compensates for this difference by providing more structure in the form of guidelines—typically developed collaboratively by physicians and software engineers. Physicians are not actual users of the guidelines that nurses are required to use.

Another variable is that of the populations served. Clalit pediatricians serve the needs of a diverse but still circumscribed pediatric population, whereas RHP nurses serve a broad, diverse population in terms of age range, symptom presentation, and diversity. This is a large order for nurses to manage and calls for a robust structure and process.

Finally, both RHP and Clalit systems share a common problem: incomplete systems of variable quality. The Clalit system's safety appears to rely on physician decision-making expertise, where standards, guidelines, and training are not that strong. The RHP may appear more complete. Safety may hinge on physician-developed electronic guidelines. Standards and training appear piecemeal or added as an afterthought. Without a meaningful, evidence-based structure and process in teletriage, quality (including safety) is at risk [18,58]. If establishing a system is a strategy to reduce possible error, then both systems could benefit from similar improvements.

Even if expert-level physicians require a less robust system, it appears that both physicians and nurses could benefit from specialized clinical training. In addition, consistent feedback regarding patient outcomes, known as planned error recovery—an essential error reduction strategy—promotes a method to self-check or to double-check another person's work [87].

Teletriage electronic algorithms must be evidence based. These guidelines are typically collaboratively developed by physicians and software developers. Nurses are required to use them, whereas physicians rarely use such tools.

Our narrative review and in-person interviews with physicians and a nurse about their experiences working in teletriage settings yielded several key findings, including the absence of specific formal training for the medical personnel working in teletriage; problematic protocols in particular clinical scenarios that, although not always available for all scenarios, are of low quality and do not allow for flexibility and agility, when needed; problematic documentation (mainly in nurse teletriage); inadequate experience and knowledge of the personnel who must make decisions in the face of uncertainty and urgency; limitations on the duration of calls or compensation based on the number of calls (which incentivizes personnel to conclude sessions promptly); and unsuitable feedback mechanisms that prevent personnel from understanding what transpired with patients and from learning from errors.

Drawing from our individual findings, the essential elements of teletriage are:

- Specialized clinical training for teletriage tasks
- Electronic algorithms and protocols
- Documentation
- Clinical call center standards: clinicians' knowledge and experience, call length (teletriage meeting duration), patient outcome feedback

Limitations

As with any narrative evaluation, selection bias cannot be completely ignored, even if this narrative analysis of the current literature was quite extensive and comprehensive and included a qualitative assessment of physicians and a nurse working in a teletriage setting.

Conclusion

Like other subspecialties, teletriage necessitates several elements to support safety, including qualified, experienced clinicians in sufficient numbers; specialized clinical training in medical decision-making; evidence-based, open, and approachable guidelines; and EMRs, audiotapes or written documentation, and standards (policies and procedures).

Fostering teletriage patient safety can be accomplished by taking the following general steps to improve MD and nurse practice in both Israel's and the United States' clinical call systems:

- Adequate training: Providers must receive adequate training to properly monitor and provide telehealth services. This includes knowledge of the systems being used, as well as familiarity with medical terminology and protocol.
- Regulation of telecommunication devices and systems: Providers must be aware of the regulations and requirements for the telecommunication devices and systems they use. This includes ensuring that the equipment is in good

working order and adheres to all safety and security regulations.
- Appropriate patient population: Telehealth services should only be used to treat patients who are stable and not at risk for an immediate life-threatening event. This will help ensure patient safety and avoid unwanted outcomes.
- Careful monitoring: In appropriate consultations, when needed, providers must carefully monitor patients and document any changes in their condition. This will help ensure that any changes or issues are addressed quickly and appropriately. Typically, nurses do not perform this task; in Israel, this is the role of the physician, for example, by using devices such as TytoCare.
- Quality assurance: Quality assurance protocols must be in place to ensure the accuracy and effectiveness of providers' services. This includes regularly reviewing documentation and providing feedback on any services deemed inadequate.
- Follow-up care: Providers must ensure that any patient receiving telehealth services receives follow-up care. This can include referrals to specialists or any other services needed to address any health concerns. Typically, nurses do not perform this task; in Israel, this is the role of the physician.
- Evidence-based studies of systems and safety: Misguided researchers unfamiliar with the triage task have produced confusing, misleading studies. Research that nibbles around the edges of the problem (patient or clinician satisfaction, clinician stress levels and attitudes, nonclinician practice) fails to address the core problem—system error. The telehealth industry requires long-overdue evidence-based outcome studies that meaningfully demonstrate the structures and processes that inform and strengthen safety.

Acknowledgments

We would like to express our gratitude to the physicians and the one nurse who took part in the interviews. We received no funding for this study.

Data Availability

All data analyzed during this study are included in this published paper and its supplementary information files. Additional data sets generated during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

MH was involved in the conceptualization of the study, methodology, investigation, data curation, interviewing the physicians, analysis, and writing the paper. SQW was involved in data curation, methodology, formal analysis, interviewing the nurse, and writing the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Definitions and terminology.
[DOCX File , 21 KB - humanfactors_v11i1e50676_app1.docx]

Multimedia Appendix 2
Paper selection criteria.
[DOCX File , 18 KB - humanfactors_v11i1e50676_app2.docx]

Multimedia Appendix 3
Case study from the United States.
[DOCX File , 30 KB - humanfactors_v11i1e50676_app3.docx]

References

1. Telemedicine market insights: the ultimate solution for quality care. Grand View Research. URL: https://tinyurl.com/4rhe6uyk [accessed 2024-03-14]
2. Grand View Research. URL: https://tinyurl.com/bp6k8n8r [accessed 2024-03-14]
3. Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: a systematic review. J Telemed Telecare 2018 Jan;24(1):4-12 [FREE Full text] [doi: 10.1177/1357633X16674087] [Medline: 29320966]
4. NEJM Catalyst. What is telehealth? New Engl J Med 2018 Feb 1;4(1):Online [FREE Full text] [doi: 10.1056/CAT.18.0268]
5. Global Market Insights. 2023 Jan. URL: https://tinyurl.com/2w36v6mn [accessed 2024-03-14]

6. Huibers L, Smits M, Renaud V, Giesen P, Wensing M. Safety of telephone triage in out-of-hours care: a systematic review. Scand J Prim Health Care 2011 Dec 29;29(4):198-209 [FREE Full text] [doi: 10.3109/02813432.2011.629150] [Medline: 22126218]
7. Boggan JC, Shoup JP, Whited JD, Van Voorhees E, Gordon AM, Rushton S, et al. Effectiveness of acute care remote triage systems: a systematic review. J Gen Intern Med 2020 Jul 02;35(7):2136-2145 [FREE Full text] [doi: 10.1007/s11606-019-05585-4] [Medline: 31898116]
8. Wheeler SQ, Greenberg ME, Mahlmeister L, Wolfe N. Safety of clinical and non-clinical decision makers in telephone triage: a narrative review. J Telemed Telecare 2015 Sep;21(6):305-322. [doi: 10.1177/1357633X15571650] [Medline: 25761468]
9. Wheeler SQ, Windt J. Telephone Triage: Theory, Practice and Protocol Development. San Anselmo, CA: TeleTriage Systems Publishers; 1993.
10. Wheeler SQ, Mahlmeister L. Telephone Triage Risk Management in Primary Care, Urgent Care and ED Settings. San Anselmo, CA: TeleTriage Systems Publishers; 2020.
11. Baren J, Rothrock S, Brennan J. In: Saunders WB, editor. Pediatric Emergency Medicine. Philadelphia, PA: Elsevier Health Sciences; 2008.
12. Taylor TB. COVID-19: rapid application of technology for emergency department tele-triage. American College of Emergency Physicians. 2020 Oct 6. URL: https://tinyurl.com/3bu2ncm5 [accessed 2024-03-14]
13. Huibers L, Keizer E, Giesen P, Grol R, Wensing M. Nurse telephone triage: good quality associated with appropriate decisions. Fam Pract 2012 Oct;29(5):547-552 [FREE Full text] [doi: 10.1093/fampra/cms005] [Medline: 22327415]
14. Leprohon J, Patel VL. Decision-making strategies for telephone triage in emergency medical services. Med Decis Mak 1995;15(3):240-253. [doi: 10.1177/0272989X9501500307] [Medline: 7564938]
15. Katz HP, Kaltsounis D, Halloran L, Mondor M. Patient safety and telephone medicine : some lessons from closed claim case review. J Gen Intern Med 2008 May;23(5):517-522 [FREE Full text] [doi: 10.1007/s11606-007-0491-y] [Medline: 18228110]
16. Nash DB. Insights into the use of digital tools for people with mental health disorders. Popul Health Manag 2021 Aug;24(S2):S53-S54. [doi: 10.1089/pop.2021.29004.dbn] [Medline: 34382865]
17. Kaminsky E, Röing M, Björkman A, Holmström IK. Telephone nursing in Sweden: a narrative literature review. Nurs Health Sci 2017 Sep 15;19(3):278-286. [doi: 10.1111/nhs.12349] [Medline: 28618087]
18. Lake R, Georgiou A, Li J, Li L, Byrne M, Robinson M, et al. The quality, safety and governance of telephone triage and advice services - an overview of evidence from systematic reviews. BMC Health Serv Res 2017 Aug 30;17(1):614 [FREE Full text] [doi: 10.1186/s12913-017-2564-x] [Medline: 28854916]
19. Wheeler SQ. ED telephone triage: lessons learned from unusual calls. J Emerg Nurs 1989;15(6):481-487. [Medline: 2687540]
20. Health contact center accreditation. Utilization Review Accreditation Commission. 2019. URL: https://www.urac.org/programs/health-call-center-accreditation [accessed 2024-03-14]
21. American Academy of Ambulatory Care Nursing. In: Anglea T, Murray C, Mastal M, Clelland S, editors. Scope and Standards of Practice for Professional Telehealth Nursing, 6th Edition. Pitman, NJ: American Academy of Ambulatory Care Nursing; 2018.
22. Framework for protecting patients when emergency care is out-of-network. American College of Emergency Physicians. 2019. URL: https://tinyurl.com/nhjxvdem [accessed 2024-03-14]
23. Principles of nursing documentation. American Nurses Association. 2010. URL: https://tinyurl.com/46mdkyrm [accessed 2024-03-14]
24. Emergency Nurses Association. Emergency Nurses Association position statement. Telephone advice. J Emerg Nurs 1991 Oct;17(5):52A. [Medline: 1921090]
25. Glossary of terms. NANDA International. URL: http://www.nanda.org/nanda-i-resources/glossary-of-terms/ [accessed 2024-03-14]
26. Wachter R. The Digital Doctor, Hope, Hype and Harm at the Dawn of Medicine's Computer Age. New York, NY: McGraw Hill; 2015.
27. Institute of Medicine (US) Committee on Standards for Developing Trustworthy Clinical Practice Guidelines. In: Graham R, Mancher M, Wolman DM, Greenfield S, Steinberg E, editors. Clinical Practice Guidelines We Can Trust. Washington, DC: National Academies Press; 2011.
28. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
29. Cho JY, Lee E. Reducing confusion about grounded theory and qualitative content analysis: similarities and differences. Qual Rep 2014;19(32):1-20 [FREE Full text] [doi: 10.46743/2160-3715/2014.1028]
30. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]

31. Ting DSW, Carin L, Dzau V, Wong TY. Digital technology and COVID-19. Nat Med 2020 Apr 27;26(4):459-461 [FREE Full text] [doi: 10.1038/s41591-020-0824-5] [Medline: 32284618]
32. Elkbuli A, Ehrlich H, McKenney M. The effective use of telemedicine to save lives and maintain structure in a healthcare system: current response to COVID-19. Am J Emerg Med 2021 Jun;44:468-469 [FREE Full text] [doi: 10.1016/j.ajem.2020.04.003] [Medline: 32303410]
33. Bashshur R, Doarn C, Frenk J, Kvedar J, Woolliscroft J. Telemedicine and the COVID-19 pandemic, lessons for the future. Telemed J E Health 2020 May;26(5):571-573 [FREE Full text] [doi: 10.1089/tmj.2020.29040.rb] [Medline: 32275485]
34. Smith AC, Thomas E, Snoswell CL, Haydon H, Mehrotra A, Clemensen J, et al. Telehealth for global emergencies: implications for coronavirus disease 2019 (COVID-19). J Telemed Telecare 2020 Mar 20;26(5):309-313. [doi: 10.1177/1357633x20916567]
35. Christianson J, Christianson E. White paper: using telehealth in the emergency department to minimize risk to health care providers and conserve resources during the COVID-19 response. InDoc, PA. 2020. URL: ttps://tinyurl.com/yc3h8nbe [accessed 2024-03-14]
36. Desborough J, Hall Dykgraaf S, de Toca L, Davis S, Roberts L, Kelaher C, et al. Australia's national COVID-19 primary care response. Med J Aust 2020 Aug 04;213(3):104-106.e1 [FREE Full text] [doi: 10.5694/mja2.50693] [Medline: 32623740]
37. Wheeler SQ. Telephone Triage Protocols for Adult Populations. New York, NY: McGraw Hill Publishers; 2009.
38. Reisman AB, Brown KE. Preventing communication errors in telephone medicine. J Gen Intern Med 2005 Oct;20(10):959-963 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.0199.x] [Medline: 16191150]
39. Patel A, Dale J, Crouch R. Satisfaction with telephone advice from an accident and emergency department: identifying areas for service improvement. Qual Health Care 1997 Sep 01;6(3):140-145 [FREE Full text] [doi: 10.1136/qshc.6.3.140] [Medline: 10173771]
40. Salk ED, Schriger DL, Hubbell KA, Schwartz BL. Effect of visual cues, vital signs, and protocols on triage: a prospective randomized crossover trial. Ann Emerg Med 1998 Dec;32(6):655-664. [doi: 10.1016/s0196-0644(98)70063-0] [Medline: 9832660]
41. Smith S. Helping parents cope with crying babies: decision-making and interaction at NHS Direct. J Adv Nurs 2010 Feb 20;66(2):381-391. [doi: 10.1111/j.1365-2648.2009.05195.x] [Medline: 20423421]
42. Holmström I. Decision aid software programs in telenursing: not used as intended? Experiences of Swedish telenurses. Nurs Health Sci 2007 Mar 13;9(1):23-28. [doi: 10.1111/j.1442-2018.2007.00299.x] [Medline: 17300541]
43. Car J, Freeman GK, Partridge MR, Sheikh A. Improving quality and safety of telephone based delivery of care: teaching telephone consultation skills. Qual Saf Health Care 2004 Feb 01;13(1):2-3 [FREE Full text] [doi: 10.1136/qshc.2003.009241] [Medline: 14757786]
44. Giesen P, Ferwerda R, Tijssen R, Mokkink H, Drijver R, van den Bosch W, et al. Safety of telephone triage in general practitioner cooperatives: do triage nurses correctly estimate urgency? Qual Saf Health Care 2007 Jun 01;16(3):181-184 [FREE Full text] [doi: 10.1136/qshc.2006.018846] [Medline: 17545343]
45. Smits M, Huibers L, Kerssemeijer B, de Feijter E, Wensing M, Giesen P. Patient safety in out-of-hours primary care: a review of patient records. BMC Health Serv Res 2010 Dec 10;10(1):335 [FREE Full text] [doi: 10.1186/1472-6963-10-335] [Medline: 21143949]
46. Derkx HP, Rethans JE, Muijtjens AM, Maiburg BH, Winkens R, van Rooij HG, et al. Quality of clinical aspects of call handling at Dutch out of hours centres: cross sectional national study. BMJ 2008 Sep 12;337(sep12 1):a1264-a1264 [FREE Full text] [doi: 10.1136/bmj.a1264] [Medline: 18790814]
47. Fourny M, Lucas A, Belle L, Debaty G, Casez P, Bouvaist H, et al. Inappropriate dispatcher decision for emergency medical service users with acute myocardial infarction. Am J Emerg Med 2011 Jan;29(1):37-42. [doi: 10.1016/j.ajem.2009.07.008] [Medline: 20825772]
48. Moriarty H, McLeod D, Dowell A. Mystery shopping in health service evaluation. Br J Gen Pract 2003 Dec;53(497):942-946 [FREE Full text] [Medline: 14960218]
49. Killip S, Ireson CL, Love MM, Fleming ST, Katirai W, Sandford K. Patient safety in after-hours telephone medicine. Fam Med 2007 Jun;39(6):404-409 [FREE Full text] [Medline: 17549649]
50. Hildebrandt DE, Westfall JM, Smith PC. After-hours telephone triage affects patient safety. J Fam Pract 2003 Mar;52(3):222-227. [Medline: 12620177]
51. Hersh W, Helfand M, Wallace J, Kraemer D, Patterson P, Shapiro S, et al. A systematic review of the efficacy of telemedicine for making diagnostic and management decisions. J Telemed Telecare 2002;8(4):197-209. [doi: 10.1258/135763302320272167] [Medline: 12217102]
52. Blank L, Coster J, O'Cathain A, Knowles E, Tosh J, Turner J, et al. The appropriateness of, and compliance with, telephone triage decisions: a systematic review and narrative synthesis. J Adv Nurs 2012 Dec 07;68(12):2610-2621. [doi: 10.1111/j.1365-2648.2012.06052.x] [Medline: 22676805]
53. Nataraj J, Stempien J, Netherton S, Wahba MY, Oyedokun TO. Emergency department referrals from a provincial medical call centre: is it more than just 1-800-go-to-emerg? Can J Emerg Med 2020 Mar 24;22(2):241-244. [doi: 10.1017/cem.2019.420] [Medline: 31645231]

54. Sax DR, Vinson DR, Yamin CK, Huang J, Falck TM, Bhargava R, et al. Tele-triage outcomes for patients with chest pain: comparing physicians and registered nurses. Health Aff (Millwood) 2018 Dec;37(12):1997-2004. [doi: 10.1377/hlthaff.2018.05079] [Medline: 30633669]
55. Turbitt E, Freed GL. Use of a telenursing triage service by Victorian parents attending the emergency department for their child's lower urgency condition. Emerg Med Australas 2015 Dec 03;27(6):558-562. [doi: 10.1111/1742-6723.12477] [Medline: 26337783]
56. Baumeister RF, Tierney J. Willpower: Rediscovering the Greatest Human Strength. London, UK: Penguin Press; 2011.
57. Donabedian A. Evaluating the quality of medical care. Milbank Memorial Fund Q 1966 Jul;44(3):166. [doi: 10.2307/3348969]
58. Donabedian A. Quality assurance. Structure, process and outcome. Nurs Stand 1992;7(11 Suppl QA):4-5. [Medline: 1489693]
59. Berwick D, Fox DM. "Evaluating the quality of medical care": Donabedian's classic article 50 years later. Milbank Q 2016 Jun;94(2):237-241 [FREE Full text] [doi: 10.1111/1468-0009.12189] [Medline: 27265554]
60. Institute of Medicine (US) Committee on Quality of Health Care in America. In: Kohn LT, Corrigan JM, Donaldson MS, editors. To Err is Human: Building a Safer Health System. Washington, DC: National Academies Press; 2000.
61. Scarfone RJ, Luberti AA, Mistry RD. Outcomes of children referred to an emergency department by an after-hours call center. Pediatr Emerg Care 2004 Jun;20(6):367-372. [doi: 10.1097/01.pec.0000133610.42699.41] [Medline: 15179144]
62. Hildebrandt DE, Westfall JM, Fernald DH, Pace WD. Harm resulting from inappropriate telephone triage in primary care. J Am Board Fam Med 2006 Sep 01;19(5):437-442 [FREE Full text] [doi: 10.3122/jabfm.19.5.437] [Medline: 16951292]
63. Wetterneck TB, Karsh BT. Human factors applications to understanding and using close calls to improve health care. In: Wu A, editor. The Value of Close Calls in Improving Patient Safety: Learning How to Avoid and Mitigate Patient Harm. Oak Brook, IL: The Joint Commission Resources; 2011:39-53.
64. Makary MA, Daniel M. Medical error—the third leading cause of death in the US. BMJ 2016 May 03;353:i2139 [FREE Full text] [doi: 10.1136/bmj.i2139] [Medline: 27143499]
65. Walker J. EHR: a MedMal minefield for practitioners? JLNC 2018;29(2):9-13 [FREE Full text] [doi: 10.30710/JLNC.29.2.2018.9]
66. Hirsh D, Simon H, Massey R, Thornton L, Simon J. The host hospital 24-hour underreferral rate: an automated measure of call-center safety. Pediatrics 2007 Jun;119(6):1139-1144. [doi: 10.1542/peds.2006-1986] [Medline: 17545381]
67. Mahlmeister L. Human factors and errors in perinatal care. J Perinat Neonat Nurs 2015;24(1):12-21. [doi: 10.1097/jpn.0b013e3181cdb6e0]
68. Mahlmeister L, Van Mullem C. The process of triage in perinatal settings: clinical and legal issues. J Perinat Neonat Nurs 2000 Mar;13(4):13-30. [doi: 10.1097/00005237-200003000-00003] [Medline: 11075083]
69. Gawande A. The Checklist Manifesto: How to Get Things Right. New York, NY: Picador; 2010.
70. Ernesäter A, Engström M, Holmström I, Winblad U. Incident reporting in nurse-led national telephone triage in Sweden: the reported errors reveal a pattern that needs to be broken. J Telemed Telecare 2010 May 10;16(5):243-247. [doi: 10.1258/jtt.2009.090813] [Medline: 20457800]
71. Ernesäter A, Winblad U, Engström M, Holmström IK. Malpractice claims regarding calls to Swedish telephone advice nursing: what went wrong and why? J Telemed Telecare 2012 Aug 24;18(7):379-383. [doi: 10.1258/jtt.2012.120416] [Medline: 22923361]
72. Budryk Z. Malpractice claims against nurses on the rise. FIERCE Healthcare. 2016 Feb 26. URL: https://tinyurl.com/yn8t8764 [accessed 2024-03-14]
73. Oken RL. What have we learned from pediatric malpractice cases? AAP News. 2018. URL: https://publications.aap.org/aapnews/news/14845 [accessed 2024-03-14]
74. Greatbatch D, Hanlon G, Goode J, O'Cathain A, Strangleman T, Luff D. Telephone triage, expert systems and clinical expertise. Sociol Health Illn 2005 Sep;27(6):802-830 [FREE Full text] [doi: 10.1111/j.1467-9566.2005.00475.x] [Medline: 16283900]
75. Perrin EC, Goodman HC. Telephone management of acute pediatric illnesses. N Engl J Med 1978 Jan 19;298(3):130-135. [doi: 10.1056/nejm197801192980304]
76. Carrasqueiro S, Oliveira M, Encarnação P. Evaluation of telephone triage and advice services: a systematic review on methods, metrics and results. Stud Health Technol Inform 2011;169:407-411. [Medline: 21893782]
77. Farzandipour M, Nabovati E, Sharif R. The effectiveness of tele-triage during the COVID-19 pandemic: a systematic review and narrative synthesis. J Telemed Telecare 2023 Jan 23:1357633X221150278 [FREE Full text] [doi: 10.1177/1357633X221150278] [Medline: 36683438]
78. International health care system profiles. The Commonwealth Fund. 2019. URL: https://tinyurl.com/ycym3fcb [accessed 2024-03-14]
79. Discussed: criteria for operating a remote health service. Ministry of Health Israel. URL: https://www.health.gov.il/hozer/mk06_2019.pdf [accessed 2024-03-14]
80. National plan for digital health. Ministry of Health Israel. URL: https://www.health.gov.il/About/projects/DigitalHealth/Pages/default.aspx [accessed 2024-03-14]

81. Gamus A, Chodick G. Telemedicine after COVID-19: the Israeli perspective. Isr Med Assoc J 2020 Aug;22(8):467-469 [FREE Full text] [Medline: 33236577]
82. Haimi M, Brammli-Greenberg S, Waisman Y, Baron-Epel O. Physicians' experiences, attitudes and challenges in a pediatric telemedicine service. Pediatr Res 2018 Nov;84(5):650-656. [doi: 10.1038/s41390-018-0117-6] [Medline: 30120402]
83. Haimi M, Brammli-Greenberg S, Waisman Y, Stein N, Baron-Epel O. The role of non-medical factors in physicians' decision-making process in a pediatric telemedicine service. Health Informat J 2020 Jun;26(2):1152-1176 [FREE Full text] [doi: 10.1177/1460458219870660] [Medline: 31566448]
84. Haimi M, Brammli-Greenberg S, Baron-Epel O, Waisman Y. Assessing patient safety in a pediatric telemedicine setting: a multi-methods study. BMC Med Inform Decis Mak 2020 Apr 03;20(1):63 [FREE Full text] [doi: 10.1186/s12911-020-1074-7] [Medline: 32245469]
85. Grossman Z, Chodick G, Reingold SM, Chapnick G, Ashkenazi S. The future of telemedicine visits after COVID-19: perceptions of primary care pediatricians. Isr J Health Policy Res 2020 Oct 20;9(1):53 [FREE Full text] [doi: 10.1186/s13584-020-00414-0] [Medline: 33081834]
86. Nevet A, Bitton Y, Wolf L, Waisman Y. [Telemedicine: a novel service in pediatric emergency care]. Harefuah 2016 Jul;155(7):410-413. [Medline: 28514123]
87. Human factors analysis in patient safety systems. The Joint Commission. 2015 Apr. URL: https://tinyurl.com/38zjxcts [accessed 2024-03-14]

Abbreviations

ACEP: American College of Emergency Physicians
ANA: American Nurses Association
AR: appropriate referral
CDMS: computerized decision-making system
CDSS: computerized decision support system
ED: emergency department
EMR: electronic medical record
HMO: health maintenance organization
MD: doctor of medicine
MOH: Ministry of Health
OR: overreferral
PNP: pediatric nurse practitioner
RHP: Redwood Healthcare Plan
RN: registered nurse
UR: underreferral
URAC: Utilization Review Accreditation Commission

Edited by A Kushniruk; submitted 09.07.23; peer-reviewed by G Chodick, A Benis, T Aslanidis; comments to author 07.11.23; revised version received 25.11.23; accepted 28.02.24; published 25.03.24.

Please cite as:
Haimi M, Wheeler SQ
Safety in Teletriage by Nurses and Physicians in the United States and Israel: Narrative Review and Qualitative Study
JMIR Hum Factors 2024;11:e50676
URL: https://humanfactors.jmir.org/2024/1/e50676
doi:10.2196/50676
PMID:38526526

Review

Use of Creative Frameworks in Health Care to Solve Data and Information Problems: Scoping Review

Elisabeth Veronica Mess[1], MA; Frank Kramer[2], PhD; Julia Krumme[1], PhD; Nico Kanelakis[3], Dipl Pol (univ); Alexandra Teynor[1], PhD

[1]Institute for Agile Software Development, Technical University of Applied Sciences Augsburg, Augsburg, Germany

[2]Chair of IT Infrastructure for Translational Medical Research, University of Augsburg, Augsburg, Germany

[3]Technical University of Applied Sciences Augsburg, Augsburg, Germany

Corresponding Author:
Elisabeth Veronica Mess, MA
Institute for Agile Software Development
Technical University of Applied Sciences Augsburg
An der Hochschule 1
Augsburg, 86161
Germany
Phone: 49 +49 821 5586 36
Email: elisabethveronica.mess@hs-augsburg.de

Abstract

Background: Digitization is vital for data management, especially in health care. However, problems still hinder health care stakeholders in their daily work while collecting, processing, and providing health data or information. Data are missing, incorrect, cannot be collected, or information is inadequately presented. These problems can be seen as data or information problems. A proven way to elicit requirements for (software) systems is by using creative frameworks (eg, user-centered design, design thinking, lean UX [user experience], or service design) or creative methods (eg, mind mapping, storyboarding, 6 thinking hats, or interaction room). However, to what extent they are used to solve data or information-related problems in health care is unclear.

Objective: The primary objective of this scoping review is to investigate the use of creative frameworks in addressing data and information problems in health care.

Methods: Following JBI guidelines and the PRISMA-ScR framework, this paper analyzes selected papers, answering whether creative frameworks addressed health care data or information problems. Focusing on data problems (elicitation or collection, processing) and information problems (provision or visualization), the review examined German and English papers published between 2018 and 2022 using keywords related to “data,” “design,” and “user-centered.” The database SCOPUS was used.

Results: Of the 898 query results, only 23 papers described a data or information problem and a creative method to solve it. These were included in the follow-up analysis and divided into different problem categories: data collection (n=7), data processing (n=1), information visualization (n=11), and mixed problems meaning data and information problem present (n=4). The analysis showed that most identified problems fall into the information visualization category. This could indicate that creative frameworks are particularly suitable for solving information or visualization problems and less for other, more abstract areas such as data problems. The results also showed that most researchers applied a creative framework after they knew what specific (data or information) problem they had (n=21). Only a minority chose a creative framework to identify a problem and realize it was a data or information problem (n=2). In response to these findings, the paper discusses the need for a new approach that addresses health care data and information challenges by promoting collaboration, iterative feedback, and user-centered development.

Conclusions: Although the potential of creative frameworks is undisputed, applying these in solving data and information problems is a minority. To harness this potential, a suitable method needs to be developed to support health care system stakeholders. This method could be the User-Centered Data Approach.

(JMIR Hum Factors 2024;11:e55182) doi:10.2196/55182

KEYWORDS

creative frameworks; data and information problems; data collection; data processing; data provision; health care; information visualization; interdisciplinary teams; user-centered design; user-centered data design; user-centric development

Introduction

Background

Despite advancing digitization in the health care sector, problems still hinder health care stakeholders in their daily work while collecting, processing, and providing health data or information. These problems can be seen as data or information problems.

This scoping review aims to review if creative frameworks were used to solve data or information problems in health care.

A secondary objective was to understand how these creative frameworks were applied in detail to understand their possible impact on solving data and information problems.

Definition of Data and Information

The correlation between data and information can be visualized with the information pyramid [1], whereas data are necessary to get information that leads to further possible knowledge and wisdom (Figure 1).

Figure 1. Data Information Knowledge and Wisdom Hierarchy (DIKW).

Data are defined as "facts and statistics collected for reference or analysis [and] the quantities, characters, or symbols on which a computer performs operations" [2]. Data gain value when analyzed and structured into information, which is crucial for extracting insights and making informed decisions. The word information is defined by [3] as "notice, news or advice communicated by word or writing." Claude Shannon, the father of modern information theory, quantifies information as a measure of uncertainty or surprise in a message, introducing the concept of entropy as in [4].

Over the past decades, replacing "information" with "data" has become common. The word data now includes entities and meaning [5]. However, this can cause misunderstandings, so it is important to consider the specific discipline's framework [6].

In general, data can be categorized as structured, unstructured, or semistructured. Structured data are organized and formatted in a predefined manner, with each element assigned a specific type, often found in databases or spreadsheets (eg, CSV and XML) [7]. In health care, this type of data includes electronic health records (patient demographic information, diagnosis codes, and medication lists) [8].

Unstructured data lacks a predefined structure, including text, images, audio, video, and social media posts, making it harder to process automatically [9,10]. In health care, this data can be descriptions or medical history, mainly text in free written form. Semistructured data combine elements of both, with some organizational structure but not a rigid schema (eg, JSON files, XML documents with optional tags, or a combination of checkboxes and descriptions).

The Importance of Data and Information in Health Care

Medical decision-making is one area that benefits crucially from leveraging data and information. Health care providers can gain valuable insights into patients' diseases or treatment outcomes [11]. Also, it can improve the patient's health status or satisfaction (cf. [12,13]). It can also significantly influence the quality of patient care and patient safety. By continuously analyzing, for example, clinical data and patient feedback, it is possible to identify areas needing improvement [14]. Another aspect is positively adjusting operational activities by analyzing patient flow, operational processes, and workflows [15].

Data and Information Problems in Health Care

Despite the importance of data and information, health care professionals and practitioners must often make highly calculated and accurate decisions with limited information, resources, and knowledge [16-20]. That puts a more significant burden on the staff and prevents them from harnessing the potential value of data and information. These problems can be seen as data and information problems.

A data problem is present when the necessary data for further processing, analysis, and provision are unavailable or cannot be collected (sufficiently). That can involve anomalies, discrepancies, or limitations within the data itself or difficulties related to data processing, integration, quality assurance, or ethical considerations.

An information problem can arise when the information is not provided sufficiently or is unrepresentative. This points to a challenge when seeking, evaluating, synthesizing, or using information sources and can encompass difficulties related to information scarcity, retrieval, relevance assessment, credibility assessment, information organization, or ethical considerations.

It is important to stress that (regardless of these definitions) the complexity of these problems constantly changes depending on the context and the target group. Due to that, a "one-fits-all" concept or solution is unrealistic.

Creative Frameworks in Software Development in Health Care

Diverse approaches are used in various fields to stimulate innovative thinking, generate novel ideas, and foster problem-solving. Some of these well-known creative approaches are design thinking, user-centered design (UCD) or human-centered design (HCD), lean UX [user experience], lean startup, and service design. These creative approaches are frameworks that provide a structured approach or guidelines that define the overall context, principles, and objectives for fostering creativity. They often offer an overarching perspective and may encompass multiple creative methods.

In contrast, an individual creative method refers to a specific technique or tool. They are typically more focused compared to frameworks. Creative methods include mind mapping, storyboarding, 6 thinking hats, and scamper or interaction room.

In the context of health care, the approaches design thinking, UCD, and HCD are often successfully used to develop user-centered software or hardware to support health care professionals' daily practice [21-24]. This may be due in no small part to the fact that specifically human-centered design is anchored in ISO (International Organization for Standardization) standards, which aim to ensure consistency, quality, safety, efficiency, and interoperability in different disciplines.

An example is ISO 9241, focusing on the Ergonomics of human-system interaction. In the ISO 9241:2019, Part 210 focuses on a human-centered design for interactive systems. The HCD process is divided into different activities and sub-activities and focuses on their responsibilities for human-centered quality in software development [25].

Creative Frameworks and Data or Information Problems in Health Care

Data and information problems can vary significantly in their complexity—from anomalies or limitations in the data to information overload. Depending on the context, further complex combinations of problems can arise and pose challenges for all relevant target groups.

Many of these problems cannot be solved by one discipline alone. Interdisciplinary and human-centered exchange is necessary to find the best possible solution.

Creative frameworks are ideal here; they provide a structure and help find creative, human-centered solutions in an interdisciplinary team.

An initial keyword search [26] conducted by the authors showed that data and information problems in health care were partially solved by using creative frameworks. A total of 100 papers were screened for two acceptance criteria: (1) data or information problem present and (2) a creative framework used to solve it. The screening resulted in 4 papers that met the criteria. However, these results were not representative (small sampling size, incomplete data), so it was impossible to deduce the extent to which creative frameworks are disseminated and applied in detail in the context of data and information problems and whether this output is just a coincidence. After all, no other publication was found that deals with this focus, which led to conducting a scoping review to obtain further information.

Methods

Overview

In this work, a scoping review was conducted to determine if creative frameworks were used to solve data or information problems in health care. A secondary objective was to understand how creative frameworks were applied in detail to understand their possible impact on solving data and information problems.

The scoping review was conducted using the guidelines of the Joanna Briggs Institute (JBI) [27] and the PRISMA-ScR (Preferred Reporting Items for Systematic Analyses and Meta-Analyses extension for Scoping Reviews); see Figures 2 and 3. It provides researchers with a preferred reporting item and a checklist, as in [28]. Papers meeting the criteria for inclusion were chosen with a developed decision flow diagram (see Figure 2).

Figure 2. Overview of the decision flow diagram with the different scenarios.

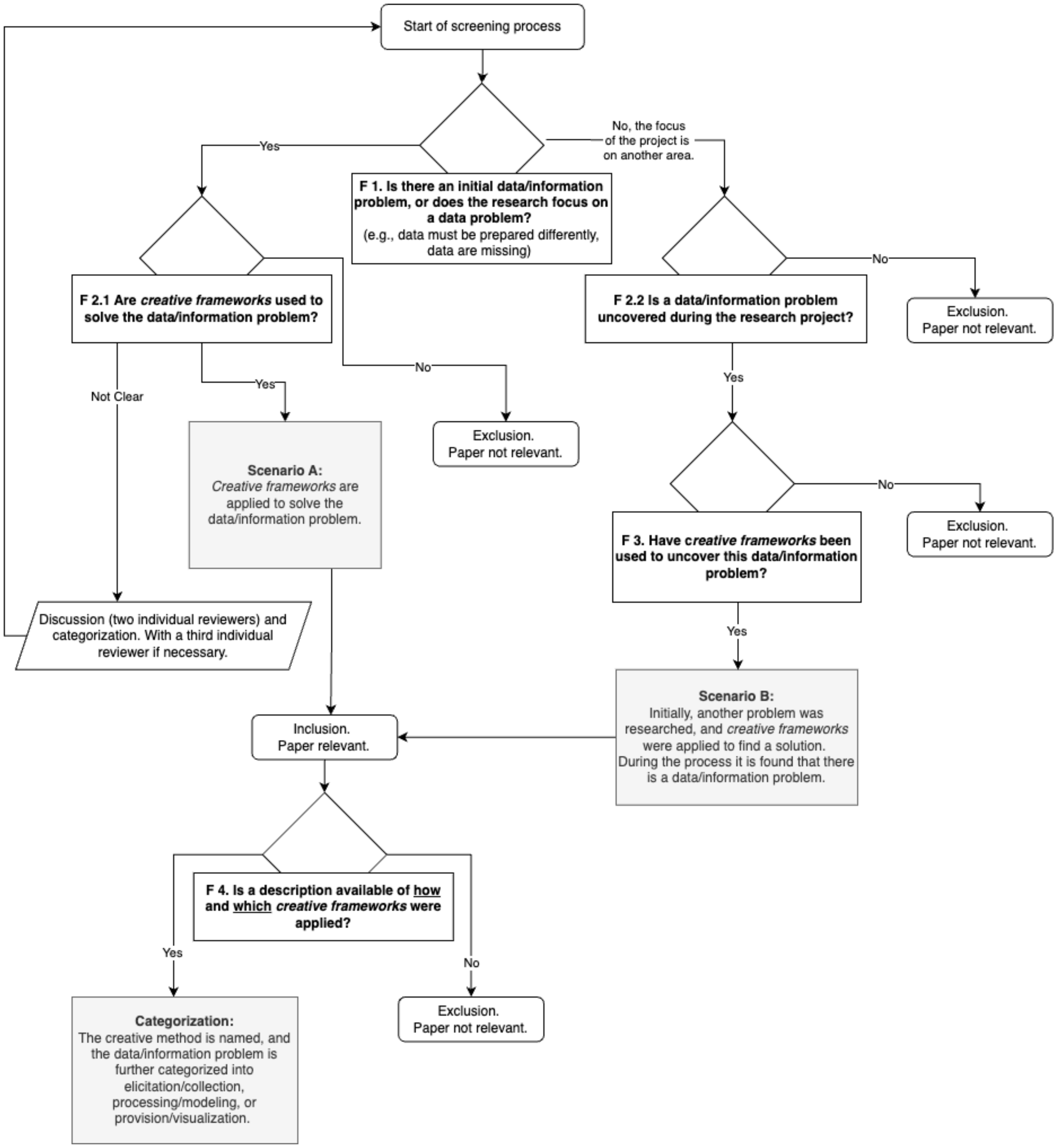
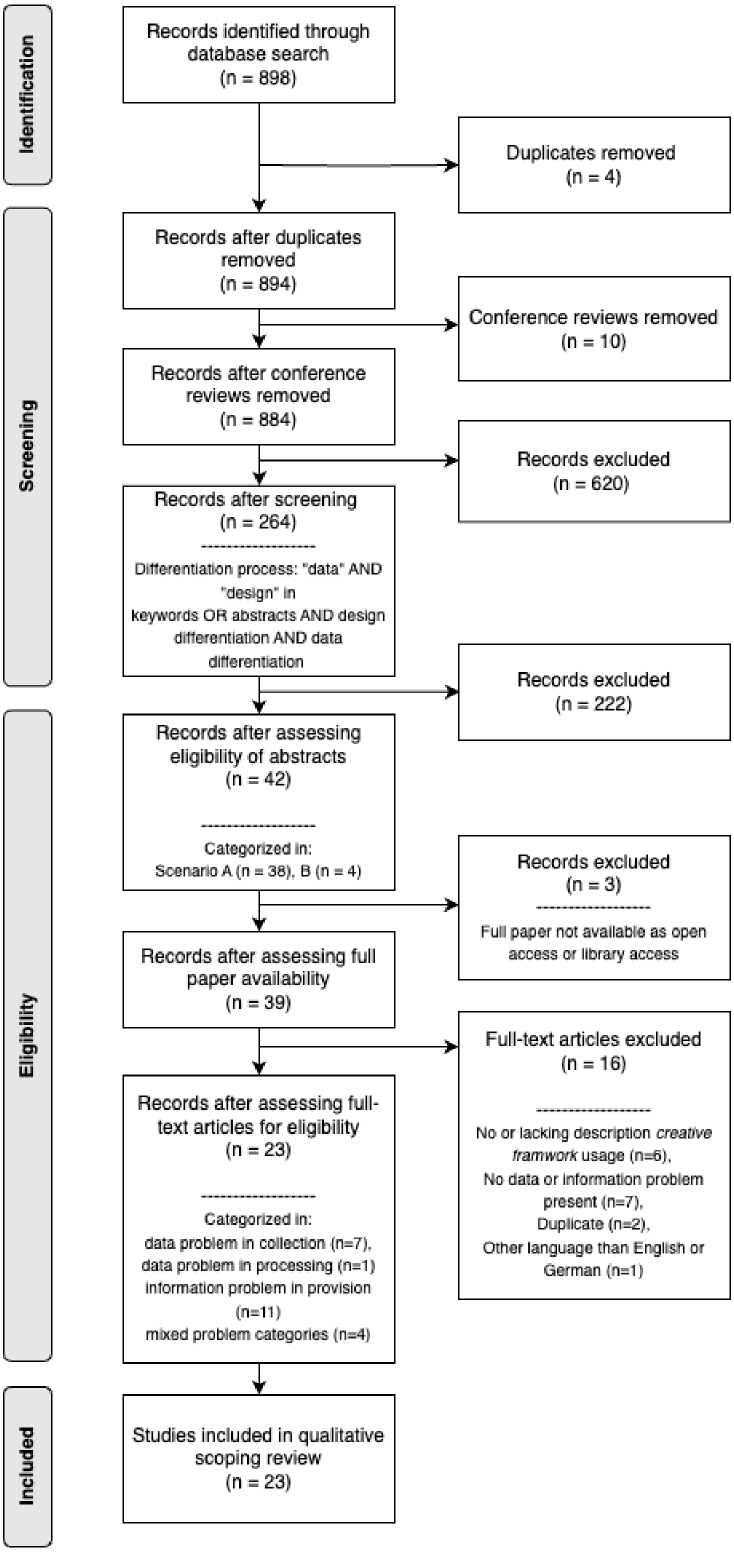


Figure 3. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) flow diagram, including the different rounds and scenario decisions conducted.



Research Question

The following research question was explored: Have creative frameworks been used in solving "data or information problems" in medicine or health care, for example, the collection, processing, or provision of data?

Search Strategy

SCOPUS, ACM, and IEEE libraries were initially considered suitable because they are known for their extensive coverage of academic literature (journals, conference proceedings, reports), particularly in computer science. Additionally, the published papers are peer reviewed, and they provide advanced search features.

During initial keyword testing, it was found that the libraries ACM and IEEE are unsuitable for this scoping review as they have either keyword or content group limitations. Due to the number of different keyword spellings (see Tables I and II) and the need for a parallel search in the document title, abstract, and keywords, SCOPUS was chosen.

Studies published in German and English from 2018 to 2022 were included. The search was performed on July 6, 2022.

Search String and Source Selection

The main search strings were »data« AND »design« in the categories of Computer Science, Medicine, Nursing, and Health Professions. Due to the more common definition of data, which refers to a data-based definition of information as mentioned in the Introduction, the keyword information was not included. However, further analysis of the query results will interpret the word data and distinguish between data and information.

Concerning the selection of creative frameworks, the authors decided to focus on UCD, HCD, and design thinking because they have already proven successful in software development. Additional keywords like service design, lean UX, and lean startup were tested for relevance. Service design returned a significant query result with approximately 200 papers; thus, it was included in the search string. Lean UX and lean startup delivered only a handful, and the keywords were referenced neither in the title nor abstract nor keywords, so these 2 creative frameworks were excluded from the search string.

The remaining 4 keywords for creative frameworks were searched using the title, abstracts, and authors' keywords. As these main keywords are insufficient to differentiate when and how creative frameworks have been used (collection, processing, provision or visualization), further subkeywords were defined. That was primarily due to different spelling and general writing options for UCD and HCD, as seen in Multimedia Appendix 1. The overall number of keywords and substrings was 32 (including main and differentiation keywords). Definitions of the keywords are listed in a separate appendix.

All retrieved papers were matched against a decision flow diagram (see Figure 2). This step was included to guarantee that all perspectives and approaches to using creative frameworks in data and information problem-solving were included. Therefore, the papers were further categorized in this diagram with different implementation scenarios of creative frameworks (creative frameworks used yes/no, data or information problem solving with creative frameworks yes/no).

Study or Source of Evidence Selection

All identified papers were collected within an EndNote database and shared with all reviewers after the search. The differentiation of the paper focused on screening the document title, abstract, and keywords. This step was conducted individually with smart groups by each reviewer. In EndNote, smart groups can be populated automatically by predetermined criteria. It can be described as "Search in this group in the title, abstract, keywords for keywords X, Y, Z."

Duplicates and conference reviews were removed within the screening process.

Potentially relevant sources were retrieved in full while checking the eligibility. Any disagreements between the reviewers at each stage of the selection process were resolved through discussion or with an additional reviewer.

The search results and study inclusion process are presented in the PRISMA-ScR flow diagram, as shown in Figure 3.

Results

Overview

The literature search retrieved 898 items from SCOPUS. After the removal of duplicates (n=4) and conference reviews (n=10), titles, abstracts, and keywords of 884 papers were screened according to the inclusion and exclusion criteria (description scenarios see Table 1 and decision flow diagram see Figure 2).

The overall process can be viewed with the PRISMA ScR flow diagram (see Figure 3). Afterward, further research results are presented.

Table 1. Description of 2 defined scenarios of the decision flow diagram.

Scenarios	Descriptions
Scenario A	• Creative frameworks are applied directly to solve a data or information problem. • A data problem was evident at the beginning of the study, and a creative method was used to solve a data or information problem. That means the awareness of the data problem was high, and choosing creative frameworks to solve the problem was a conscious decision.
Scenario B	• Creative frameworks are applied indirectly to solve a data or information problem. • That means the awareness of a data or information problem came later in the research process, and a creative method was conducted first.

Findings

A total of 42 abstracts and articles were assessed for eligibility, and 39 papers could be retrieved as full texts. Two reviewers screened the remaining 39 full papers individually and categorized them into data problems (elicitation or collection, processing), information problems (provision/visualization), and mixed problems (data and information problems).

During this process, 23 papers were included (see Table 2), and 16 were excluded (see Table 3). Any disagreements between the 2 reviewers were discussed, and a third reviewer was consulted in cases of doubt. A detailed overview of the categorization of all 39 papers can be found in Multimedia Appendix 2.

Table 2. Summary of available full-text papers included.

Problem category	Number of papers
Data elicitation or collection	7
Data processing or modeling	1
Data, information provision, or visualization	11
Mixed problems (data and information problem present)	4
Total amount of papers included	23

Table 3. Summary of available full-text papers excluded.

Duplicate	2
Other languages used than English or German	1
No data or information problems described	7
Insufficient or missing description of the creative framework	6
Total amount of papers excluded	16

Key Finding 1—General Minority

At the beginning of the scoping review, 898 papers were identified. In the end, 23 papers met all inclusion criteria. With these papers, the research question, “Have creative frameworks (UCD, HCD, design thinking, or service design) been used in solving ‘data/information problems’, for example, in the collection, processing, and/or provision of data?” can be answered with yes but it is a minority. Only 23 papers used creative frameworks to solve data and information problems in health care.

At this point, the question arises as to why. Is the application of creative frameworks in this context insufficient, or was the focus on solving specific data and information problems with creative frameworks not given? Or are there other factors contributing to this, like the lack of insufficient description and exclusion (see Key Finding 2)? Since creative frameworks have proven successful in developing health care software and hardware in the health care sector, transferring these frameworks and techniques to solve data and information problems does not seem so far-fetched.

Key Finding 2—Lacking Descriptions of Problem Statements and Method Description

From all query results, 264 had the keywords »data« AND »design« AND 1 design differentiation (e.g., design thinking) AND 1 data differentiation (e.g., data collection).

To better understand what, how, and why the creative frameworks were applied, the papers (n=264) were further screened for 1. data/information problem present and 2. Creative framework used. In the first step of this activity, the reviewers focused on the title, abstract, and keywords and searched for a problem statement and the mention of a creative framework (that goes beyond mentioning it in the keywords). If a problem was mentioned, it was checked to see whether it was a data or information problem.

It is important to stress that the word “problem” does not mean that every paper needs to focus on a major or new problem and try to find a solution. In some cases, the problem may be rather “small” and very specific, for example, the UI design no longer works for the target group and needs to be adapted. In other cases, the problem is more general and indicates, for example, a lack of sufficient data for further research. In total, 222 papers had to be excluded because they did not contain a problem statement that referred to a data or information problem or because no creative framework was mentioned. In these cases, it is difficult to completely understand the authors’ thoughts and decision-making process. Further clarification with the authors could help, but it is beyond the scope of this review.

From the remaining 39 papers that were checked for eligibility (full paper read), only 23 could be included (see Multimedia Appendix 3 for further details). In many contributions, the data or information problem was not described in detail (n=7) and only briefly mentioned in the abstract. The description of how creative frameworks were applied was insufficient or missing in 6 papers. That could point to the fact that the authors did not feel a more detailed description was necessary or that they maybe had difficulty applying the creative framework. Further details are provided in Multimedia Appendix 4 showing 6 examples of included (n=2) and excluded (n=3) papers.

Key Finding 3—Initial Clarity About Problems Needed

During the screening and development of the decision framework for inclusion criteria, it was found that there might be different types of awareness when and how a creative framework is used to solve a problem. Either “We have a data or information problem, and we use creative frameworks to solve it” or “We have a problem, and we use creative frameworks to solve the problem or we use it to find out more about it. As a result, we found a data or information problem.”

Most papers (n=21) identified a data or information problem first (see Scenario A) and later decided to apply creative framework to solve it. In only 2 cases (n=2) did the researchers address a different type of problem and identified a data or information problem by applying a creative framework (referring to Scenario B). This might point to the fact that the problem must be apparent first before deciding to use a creative framework to solve it.

Key Finding 4—Creative Frameworks Most Applied in Case of Information Problems

Most of the included papers (n=11) dealt with an information problem, and 8 dealt with a data problem. Some papers (n=4) dealt with a combination of data and information problems and were categorized as mixed problems (problem in data collection, data processing and information provision n=1, problem in data collection and data processing n=2, and problem in data collection and data provision n=1).

That could indicate that creative frameworks are particularly suitable for solving provision or visualization problems and less for other, more abstract areas such as data problems. Also, combining several creative frameworks to solve a data/information problem leads to a preferred visual (data) output.

However, the remaining 8 papers show that it is possible to use a creative frameworks for a more abstract problem. This could mean that creative frameworks have not yet been identified as suitable.

Key Finding 5—UCD is Most Frequently Used as a Creative Method

In the process of testing keywords for relevance, UCD, HCD, and design thinking were considered the most relevant as they have already proven to be successful in software and hardware development. After testing additional keywords such as service design, lean UX, and lean startup for relevance, only service design yielded a significant number of query results, approximately 200 papers. Therefore, it was incorporated into the search string. However, lean UX and lean startup produced only a handful of results, and these keywords were not referenced in the papers' title, abstract, or keywords. Consequently, these 2 creative frameworks were excluded from the search string.

Most of the final papers used only UCD to solve the problem at hand (n=16); the other remaining papers used either design thinking (n=2), HCD (n=1), other creative frameworks such as Rapid Contextual Framework (n=1), Design Science (n=1), or Design Study Methodology (n=2). Only a few papers used more than one creative method (n=2). That indicates that UCD is, compared to design thinking, HCD, or service design, the most common, best-known, or best-understood method to apply in health care settings.

Concerning the papers that used UCD (n=16), they can be divided into data collection problem (n=5), data processing problem (n=1), information problem (n=7), mixed data, and information problem (n=2).

This indicates that UCD is specifically used and probably suited for data collection problems and information problems. It is unclear if it is less suitable for data processing problems, as we do not know if it could be successfully applied or if researchers might know of the possibility of doing so.

Key Finding 6—no Unified Spelling

As shown at the beginning of Tables I and II (in Multimedia Appendix 1), many keywords (n=32) were used for this qualitative scoping review. Different spellings had to be considered, especially for the keyword »user-centered design« and further differentiations. That prolonged the overall process of the scoping review as it had to be continuously checked that all differentiations were included during the screening and eligibility process. Overall, this made it challenging to proceed efficiently and effectively. It is still possible that relevant papers might not be part of this work because not all relevant keywords and their spelling variants have been included.

Closing the Gap: A New Approach Is Needed

The results of the initial literature research and scoping review demonstrate that creative frameworks have been used to address challenges in health care, such as information overload or data collection hindrances. However, this specific approach tends to be a minority.

Having identified an application gap, we propose a new idea—a Human-Centered Data Approach. With this, we aim to propose a structured way to minimize data and information problems in the health care setting.

The main goal is to analyze and integrate the users' data and information needs and include these already at the earliest stage of development possible. This could influence the quality and quantity of data and the user's overall value.

First, because the user's awareness of data and information increases: "What are data or information, what importance do they have for me, and what information do I need?" Second, it is more likely that the applicability and understandability of the data and information for the user will increase.

It could also be possible to categorize data quality, quantity, or value according to stakeholders. A guideline could be part of the new approach to optimizing the data and, hence, information and include the user throughout the development process, for example, concepts or solutions.

Furthermore, developing this new approach within an interdisciplinary team is also essential. Different perspectives are needed to develop this new approach and conduct it correctly. Furthermore, to interpret potentials and risks that come with user-centered data (from the perspective of ethics: with more data value to the user, a loss could have even more severe (personal) consequences). Therefore, computer science, design, ethics, society stakeholders, and policy disciplines would be mandatory.

A sound basis for the approach could be to use an already existing and successful method, like HCD, UCD, or design thinking, and apply it to the context of data and information problems. That shall be done with an interdisciplinary team. The presentation of this will be part of the following publication.

Discussion

Overall, this scoping review aimed to review if creative frameworks were used to solve data or information problems in health care. A secondary objective was to understand how these creative frameworks were applied in detail to understand their possible impact on solving data and information problems.

Identifying a relatively small number of papers (23 out of 898) that meet the inclusion criteria raises intriguing questions about applying creative frameworks in health care problem-solving. This small number may be due to the combination of data/information problems and creative frameworks, but this might also be because many papers had to be excluded due to a lack of clear data-/information problem description and method application (n=222), which were inclusion criteria. Authors may need to recognize the significance of providing detailed accounts of problem definitions and applying creative techniques to facilitate knowledge dissemination and the reproducibility of research outcomes.

Nevertheless, while it is established that methods using creative methods such as UCD, HCD, or design thinking have been successful in health care software and hardware development, their limited use in addressing data and information problems suggests that there may be barriers or unexplored opportunities. That prompts us to consider whether health care has yet to fully harness creative frameworks' potential in data-related challenges.

Most papers dealing with information problems (11 out of 23) hint at the suitability of creative frameworks for solving information provision and visualization challenges. This observation suggests that creative frameworks might be particularly suitable for addressing issues that require a visual representation of data. However, based on the findings, it is impossible to conclude that one specific creative framework is best suited for a specific problem. To do so, a higher number of results is needed.

Despite this, the presence of papers that tackle more abstract data problems (8 out of 23) shows that creative frameworks are not confined to information provision but can also be applied effectively to other complex data-related issues. This contrariness highlights the need for a more comprehensive understanding of when and how to apply creative frameworks across various health care contexts.

The differentiation between scenarios where researchers identify a data/information problem first or use creative frameworks to uncover such problems underscores the importance of problem clarity in adopting creative frameworks. Most cases (21 out of 23) had clarity about their problem statement in the beginning. However, cases where creative frameworks reveal hidden data or information problems (2 out of 23), indicate that creative frameworks can also be employed as exploratory tools to uncover latent issues. This nuanced perspective emphasizes the versatility of creative frameworks in health care research.

UCD, as the most used creative method (in 16 out of 23 papers), may suggest that it is the most familiar and widely accepted approach in health care settings. Its preference might also be attributed to its established reputation and applicability in health care contexts due to its references to ISO standards.

Another issue was the inconsistent spelling and terminology of UCD and HCD in the literature, which raises a practical concern for researchers conducting qualitative scoping reviews. The diversity in keywords and spellings could hinder the retrieval of relevant papers and add complexity to the literature search process. That emphasizes the importance of standardization in terminology and the need for researchers to employ comprehensive search strategies encompassing multiple spellings and variations to examine the available literature thoroughly.

With this work, we wanted to research if creative frameworks have been used to solve data or information problems in health care. We believe we have found an answer to this: yes, but it is a minority. We also wanted to highlight the need for a new approach to use the potential of creative frameworks to solve these problems. Additionally, we want to highlight that interdisciplinary discussion and evaluation, which includes the target group(s) in the earliest and all steps, is necessary to gain valuable data and information.

Limitations

Concerning the scoping review, the selection of full papers for inclusion in this study depends on the authors' ability to provide a clear problem statement and a precise presentation of their methodology. It is possible that some of the excluded papers may have conducted research that could align with the acceptance criteria but failed to meet these criteria during the screening process due to their lack of description.

Furthermore, studies with significant findings are more likely to be published, while those with null or negative results may not be accessible, leading to an incomplete picture. Also, studies that SCOPUS does not index are not part of this review but can still hold valuable results.

Additionally, it is important to stress that selecting studies and extracting data in scoping reviews involves some subjectivity. Researchers must make judgment calls about study inclusion and data extraction, which can introduce bias and influence the overall outcome.

In fast-evolving fields like computer science, it is possible that this research does not cover all the relevant research despite the vast time span, as the overall screening process takes a longer time and is resource-intensive.

Lastly, the review heavily depends on available literature. Three articles could not be screened for eligibility because they were neither available as open access nor through online academic libraries.

Conclusions

In conclusion, this scoping review offers valuable insights into using creative frameworks in health care research. The findings underscore the potential benefits of applying creative frameworks to address data and information problems while highlighting the need for greater clarity in problem definition, a more comprehensive exploration of creative methodologies, and improved reporting practices.

As health care evolves, embracing creative problem-solving approaches may prove instrumental in tackling complex data and information challenges.

We believe that the adaption or further development of UCD to the context of data and information problems could positively influence tackling data and information challenges. Ultimately, a new approach could initiate interdisciplinary dialogue, allowing researchers, practitioners, and stakeholders to reshape health care data management and improve patient care, informed decision-making, and the overall efficiency of health care systems.

Nevertheless, at this point, it is yet unclear if this will succeed or not. However, it could help to foster interdisciplinary discussion at the earliest development stage of software or hardware solutions.

Acknowledgments

This research is part of the project CARE REGIO, funded by the Bavarian State Ministry of Health, Care and Prevention.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of 12 keywords for data and 17 keywords for design (primarily due to different notations).
[DOCX File , 14 KB - humanfactors_v11i1e55182_app1.docx]

Multimedia Appendix 2
Six search keywords are defined and described in detail.
[DOCX File , 54 KB - humanfactors_v11i1e55182_app2.docx]

Multimedia Appendix 3
Overview of all eligible papers (n=39).
[DOCX File , 170 KB - humanfactors_v11i1e55182_app3.docx]

Multimedia Appendix 4
Detailed examples (n=6) of included and excluded papers.
[DOCX File , 60 KB - humanfactors_v11i1e55182_app4.docx]

Multimedia Appendix 5
PRISMA-ScR checklist.
[PDF File (Adobe PDF File), 652 KB - humanfactors_v11i1e55182_app5.pdf]

References

1. Rowley J. The wisdom hierarchy: representations of the DIKW hierarchy. J Inf Sci 2007;33(2):163-180. [doi: 10.1177/0165551506070706]
2. Oxford Dictionaries. Paperback Oxford English Dictionary. United Kingdom: Oxford University Press; 2012.
3. Webster N. A Dictionary of the English Language: Intended to Exhibit. London: Black, Young, and Young; 1831.
4. Shannon CE. A mathematical theory of communication. Bell System Technical Journal 1948;27:623-656. [doi: 10.1109/9780470544242.ch1]
5. Floridi L. Information, A Very Short Introduction. New York: Oxford University Press; 2010.
6. Capurro R, Hjørland B. The concept of information. Annual Review Info Sci & Tec 2005;37(1):343-411. [doi: 10.1002/aris.1440370109]
7. Gandomi A, Haider M. Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management 2015 Apr;35(2):137-144. [doi: 10.1016/j.ijinfomgt.2014.10.007]
8. Theis S, Rasche PWV, Bröhl C, Wille M, Mertens A. Task-data taxonomy for health data visualizations: web-based survey with experts and older adults. JMIR Med Inform 2018;6(3):e39 [FREE Full text] [doi: 10.2196/medinform.9394] [Medline: 29986844]
9. Inmon HW, Nesavich A. Tapping Into Unstructured Data: Integrating Unstructured Data and Textual Analytics Into Business Intelligence. United States: Pearson Education; 2007.
10. Isson JP. How to improve customer acquisition, customer retention, and fraud detection and prevention. In: Unstructured Data Analytics. New Jersey: John Wiley & Sons, Inc; 2018.

11. Dicuonzo G, Galeone G, Shini M, Massari A. Towards the use of big data in healthcare: a literature review. Healthcare (Basel) 2022;10(7):1232 [FREE Full text] [doi: 10.3390/healthcare10071232] [Medline: 35885759]
12. Krimsky WS, Mroz IB, McIlwaine JK, Surgenor SD, Christian D, Corwin HL, et al. A model for increasing patient safety in the intensive care unit: increasing the implementation rates of proven safety measures. Qual Saf Health Care 2009;18(1):74-80. [doi: 10.1136/qshc.2007.024844] [Medline: 19204137]
13. Leppla L, Hobelsberger S, Rockstein D, Werlitz V, Pschenitza S, Heidegger P, SMILe study team. Implementation science meets software development to create eHealth components for an integrated care model for allogeneic stem cell transplantation facilitated by eHealth: the SMILe study as an example. J Nurs Scholarsh 2021;53(1):35-45. [doi: 10.1111/jnu.12621] [Medline: 33348461]
14. Hughes RG. Hughes, patient safety and quality: an evidence-based handbook for nurses. Advances in Patient Safety 2008.
15. Subrahmanya SVG, Shetty DK, Patil V, Hameed BMZ, Paul R, Smriti K, et al. The role of data science in healthcare advancements: applications, benefits, and future prospects. Ir J Med Sci 2022;191(4):1473-1483 [FREE Full text] [doi: 10.1007/s11845-021-02730-z] [Medline: 34398394]
16. Schwartz WB, Gorry GA, Kassirer JP, Essig A. Decision analysis and clinical judgment. Am J Med 1973;55(3):459-472. [doi: 10.1016/0002-9343(73)90203-9] [Medline: 4743347]
17. Eddy DM. Clinical policies and the quality of clinical practice. N Engl J Med 1982;307(6):343-347. [doi: 10.1056/NEJM198208053070604] [Medline: 7088099]
18. Campbell EJM. The diagnosing mind. Lancet 1987;1(8537):849-851. [doi: 10.1016/s0140-6736(87)91620-5] [Medline: 2882245]
19. Etzioni A. Humble decision making. Harv Bus Rev 1989;67(4):122-126. [Medline: 10293838]
20. Anju SS, Sravani B, Madala SR. Publicly verifiable vibrant digital medical information systems. J Phys Conf Ser 2021;2089(1):012074. [doi: 10.1088/1742-6596/2089/1/012074]
21. Kuge J, Grundgeiger T, Schlosser P, Sanderson P, Happel O. Design and evaluation of a head-worn display application for multi-patient monitoring. : Nowhere and Everywhere, Association for Computing Machinery, Inc; 2021 Presented at: DIS '21: Proceedings of the 2021 ACM Designing Interactive Systems Conference; 2021 June 28; New York p. 879-890. [doi: 10.1145/3461778.3462011]
22. Reese TJ, Segall N, Del Fiol G, Tonna JE, Kawamoto K, Weir C, et al. Iterative heuristic design of temporal graphic displays with clinical domain experts. J Clin Monit Comput 2021;35(5):1119-1131 [FREE Full text] [doi: 10.1007/s10877-020-00571-2] [Medline: 32743757]
23. Li M, Makulec A, Nutley T. Applying User-Centered Design to Data Use Challenges: What We Learned. North Carolina: MEASURE Evaluation; 2017.
24. Macharia P, Katz D, Otieno G, Sharma M, Wamuti B. Implementing open data kit integrating barcode-based client identification in HIV testing and linkage to care: a user-centered design. 2021 Presented at: IEEE International Humanitarian Technology Conference (IHTC); 2021 Dec 02; United Kingdom. [doi: 10.1109/ihtc53077.2021.9698914]
25. ISO 9241-210. Ergonomics of human-system interaction—Part 210: Human-centred design for interactive systems. Cartagena: International Organization for Standardization; 2019. URL: https://www.iso.org/obp/ui/#iso:std:iso:9241:-210:ed-2:v1:e
26. Mess EV, Krumme J, Kramer F, Teynor A. Data and information problems in health care and how creative methods might solve them. Stud Health Technol Inform 2023;305:335-338. [doi: 10.3233/SHTI230498] [Medline: 37387032]
27. von Elm E, Schreiber G, Haupt CC. Methodische Anleitung für Scoping Reviews (JBI-methodologie). Zeitschrift für Evidenz. Fortbildung und Qualität im Gesundheitswesen 2019;143:1-7. [doi: 10.1016/j.zefq.2019.05.004] [Medline: 31296451]
28. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]

Abbreviations

HCD: human-centered design
ISO: International Organization for Standardization
JBI: Joanna Briggs Institute
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
UCD: user-centered design
UX: user experience

Edited by A Parush; submitted 05.12.23; peer-reviewed by K Bochtler, A Keniston; comments to author 05.03.24; revised version received 26.04.24; accepted 27.06.24; published 13.09.24.

Please cite as:
Mess EV, Kramer F, Krumme J, Kanelakis N, Teynor A
Use of Creative Frameworks in Health Care to Solve Data and Information Problems: Scoping Review
JMIR Hum Factors 2024;11:e55182
URL: https://humanfactors.jmir.org/2024/1/e55182
doi:10.2196/55182
PMID:

Original Paper

Evaluating the Usability of an mHealth App for Empowering Cancer Survivors With Disabilities: Heuristic Evaluation and Usability Testing

Rachel F Adler[1,2], PhD; Kevin Baez[1], BS; Paulina Morales[1], BS; Jocelyn Sotelo[1], BS; David Victorson[3], PhD; Susan Magasi[4], PhD

[1]Department of Computer Science, Northeastern Illinois University, Chicago, IL, United States

[2]School of Information Sciences, University of Illinois Urbana-Champaign, Champaign, IL, United States

[3]Department of Medical Social Sciences, Northwestern University, Evanston, IL, United States

[4]Department of Occupational Therapy, University of Illinois Chicago, Chicago, IL, United States

Corresponding Author:
Rachel F Adler, PhD
School of Information Sciences
University of Illinois Urbana-Champaign
501 E. Daniel St.
Champaign, IL, 61820
United States
Phone: 1 217 244 4192
Email: radler@illinois.edu

Abstract

Background: More than 18 million cancer survivors are living in the United States. The effects of cancer and its treatments can have cognitive, psychological, physical, and social consequences that many survivors find incredibly disabling. Posttreatment support is often unavailable or underused, especially for survivors living with disabilities. This leaves them to deal with new obstacles and struggles on their own, oftentimes feeling lost during this transition. Mobile health (mHealth) interventions have been shown to effectively aid cancer survivors in dealing with many of the aftereffects of cancer and its treatments; these interventions hold immense potential for survivors living with disabilities. We developed a prototype for WeCanManage, an mHealth-delivered self-management intervention to empower cancer survivors living with disabilities through problem-solving, mindfulness, and self-advocacy training.

Objective: Our study conducted a heuristic evaluation of the WeCanManage high-fidelity prototype and assessed its usability among cancer survivors with known disabilities.

Methods: We evaluated the prototype using Nielsen's 10 principles of heuristic evaluation with 22 human-computer interaction university students. On the basis of the heuristic evaluation findings, we modified the prototype and conducted usability testing on 10 cancer survivors with a variety of known disabilities, examining effectiveness, efficiency, usability, and satisfaction, including a completion of the modified System Usability Scale (SUS).

Results: The findings from the heuristic evaluation were mostly favorable, highlighting the need for a help guide, addressing accessibility concerns, and enhancing the navigation experience. After usability testing, the average SUS score was 81, indicating a good-excellent design. The participants in the usability testing sample expressed positive reactions toward the app's design, educational content and videos, and the available means of connecting with others. They identified areas for improvement, such as improving accessibility, simplifying navigation within the community forums, and providing a more convenient method to access the help guide.

Conclusions: Overall, usability testing showed positive results for the design of WeCanManage. The course content and features helped participants feel heard, understood, and less alone.

(JMIR Hum Factors 2024;11:e51522) doi:10.2196/51522

KEYWORDS

mobile health; mHealth; apps; usability; cancer survivors; accessibility; disabilities; cancer; oncology; heuristics; empowerment; advocacy; mindfulness; problem-solving

Introduction

Background

There are an estimated 18.1 million cancer survivors in the United States, and the number is projected to increase to 22.5 million by 2032 [1]. Approximately 40% of cancer survivors experience long-term physical, cognitive, psychological, and social consequences of cancer and its treatment, which can lead to significant disability [2]. These effects can include physical challenges, including but not limited to pain, fatigue, decreased functional mobility, limb loss, lymphedema, speech and swallowing difficulties, emotional challenges (as cancer survivors may experience anxiety or depression), and cognitive challenges (such as "chemo brain") [3-5]. These aftereffects can lead to activity limitations and participation restrictions, which according to contemporary frameworks and legal definitions may be considered as disabilities [6,7]. Yet, even with significant functional impairments, not all cancer survivors self-identify as disabled [8,9]. Regardless of the terminology used, the aftereffects of cancer and their related functional impacts can have a significant negative impact on well-being and health-related quality of life [10]. Survivorship plans and rehabilitation programs, which play a crucial role in restoring survivors' physical and emotional well-being, are frequently underused by cancer survivors [11]. This can be due to obstacles like time, financial constraints, and transportation issues [12], which hinder their accessibility. Mobile health (mHealth) apps can help make rehabilitation services accessible and put them in the hands of those who need them.

mHealth Apps

Mobile technologies—smartphones, tablets, and smartwatches—are increasingly ubiquitous in today's society and can be used almost anywhere [13]. The Pew Research Center reports that 85% of American adults own smartphones, and the ownership is relatively consistent across genders; racial groups; and urban, suburban, and rural users [14]. This leads to an increase in the development of mHealth apps. The COVID-19 pandemic has led to mHealth strategies becoming even more important in cancer care. According to the recommendations of Curigliano et al [15], patients with cancer should be offered mHealth strategies to support symptom management and adoption of healthy behaviors. The number of mHealth apps has increased throughout the years, with around 325,000 apps available in 2017 [16]. Charbonneau et al [17] identified 123 mHealth apps for cancer survivors available in the 2 most important marketplaces (ie, Apple iTunes and Google Play). Typical areas of usage in cancer are disease management support (eg, symptom monitoring, management of side effects, medication reminder and dosing, and access to health information), support of healthy behavior (eg, healthy diet and increased physical activity), or the connection with other patients (eg, social support through peers) [18-20].

Evaluating the Usability of mHealth Apps

It is important to gather qualitative and quantitative data on mHealth apps to determine how satisfied users would be with the product at hand. According to one scoping review, of 133 different eHealth articles that conducted usability testing, 105 used questionnaires, 57 used task completion, 45 used "think aloud," 37 conducted interviews, 18 performed heuristic evaluation, and 13 used focus groups [21]. The System Usability Scale (SUS) was the most frequently used questionnaire with a total of 44 studies. A combination of methods was used in 88 of the studies. Further, cancer was tied as the second most frequently evaluated health condition (n=10), with only mental health being evaluated more often (n=12).

Usability testing is a common effective method for evaluating the usability of mHealth apps. Studies have shown that usability testing is an effective method for examining mHealth apps for diabetes [22,23], depression [22,24], and youth at risk for developing psychosis [25], as well as managing pain [26], heart failure [27], and cancer symptoms [28]. Common questionnaires often included variations on the Mobile Application Rating Scale [25,27] or the SUS [22,24,26]. Additional techniques often employed in usability testing include measuring time per task [26] and using think aloud techniques [29]. In addition to evaluating fully implemented mobile apps, studies have conducted usability testing on prototypes of mHealth apps for supporting mental health [30], chronic kidney disease [29], fall risk detection system for older users [31], HIV [32], and cancer survivors [33-35]. Many studies have conducted heuristic evaluation before usability testing on an mHealth prototype to fix usability issues before bringing it to users [28,29,32,33]. While Nielsen's 10-point usability heuristics [36] are geared toward computer-based applications, most of these are also applicable in mobile app design. The SUS questionnaire was also commonly used in usability testing studies for examining mHealth prototypes [29,31,37].

WeCanManage App

We designed a high-fidelity prototype for WeCanManage, an evidence-informed mHealth self-management intervention, aimed at empowering individuals with tools to effectively manage cancer as a chronic condition. Users are asked to log into the app daily for 5-10 minutes to complete mobile microlearning modules of self-management content. The intervention content is based on extensive literature review and formative interviews with cancer survivors with known disabilities (n=30) and supportive cancer care professionals including social workers, psychologists, occupational and physical therapists, and a physiatrist specializing in cancer rehabilitation (n=5) [9]. A team of survivor scientists, people with lived experiences of cancer and disability, further informed intervention content and focus. Intervention content is presented sequentially as information is scaffolded on itself to promote depth of learning, retention, and application. The content is divided into 4 broad sections: WeCanRelate (fosters a sense of validating and normalizing the survivorship experience), WeCanAdapt (teaches goal direction self-management strategies), WeCanBe (emphasizes mindfulness-based practices), and WeCanSpeakUp (addresses self-advocacy and disability rights). In addition to the instructional content, WeCanManage provides users with 3 circles of support, including one-on-one connections with other users (Connect to Peers [C2P]), community forums (to discuss intervention content and shared experiences with the entire user community), and a library with evidence-informed educational content [38]. We conducted a thorough evaluation of the usability of the high-fidelity prototype for cancer survivors with disabilities, employing both heuristic evaluation and usability testing to assess its effectiveness in addressing the unique needs and challenges of this user group.

Methods

WeCanManage High-Fidelity Prototype

The high-fidelity prototype was created on Marvel [39], a web-based collaborative design platform that provides tools for creating wireframes, designs, and prototypes of interactive applications. We aimed to design WeCanManage specifically for smartphone usage. The prototype of WeCanManage allows users to navigate between the Home, Journey (Courses), C2P, Community (Community Forum), and Library (see Figure 1).

The Course section provides cancer survivors with an educational intervention that works with them on dealing with the long-term effects of their newly acquired disabilities through problem-solving, mindfulness, and self-advocacy. The content is designed to be a 4-week program where the user unlocks a series of microlessons divided into 4 modules (WeCanRelate, WeCanAdapt, WeCanBreathe, and WeCanSpeakUp), which educate users with different methods to deal with the effects of postcancer treatment in their daily life. To prioritize user control and accessibility, the course content is conveyed through mobile microlearning modules, presented in different formats such as readable text, clickable text-based cards, and audio (Figure 2).

At the end of many of the daily sessions, there are interactive engagement activities, such as reflections that feed into the Community Forum and knowledge checks (see Figure 3). The engagement activities are designed to support consolidation of knowledge and application of course content to the user's lived experiences.

The Community and C2P sections offer users a chance to engage with others, fostering networking opportunities and creating a support system with individuals undergoing similar experiences. C2P facilitates connections with others, allowing users to filter by categories like cancer type and disability, while Community features discussion forums for each of the 4 course sections and an open discussion forum. Lastly, the Library section contains additional evidence-informed resources such as articles and factsheets. The various sections of the prototypes were initially created as a low-fidelity prototype through an iterative co-design approach involving both the design teams and cancer survivors, who served as representatives of our targeted audience [40].

Because of its prototype nature, users could navigate all links, but functionalities such as real-time chat with other users and composing reflections or community posts were not operational. To overcome this, we incorporated simulated features in the prototype, triggering them automatically on user interaction. After creating the high-fidelity prototype, we evaluated it through 2 distinct methods: heuristic evaluation and usability testing.

Figure 1. Screenshots of the WeCanManage prototype: (A) Home, (B) Journey, (C) Connect to Peers (C2P), (D) Community, and (E) Library.

Figure 2. Screenshots before heuristic evaluation: (A) card view and (B) learning format after clicking on the Formats icon.

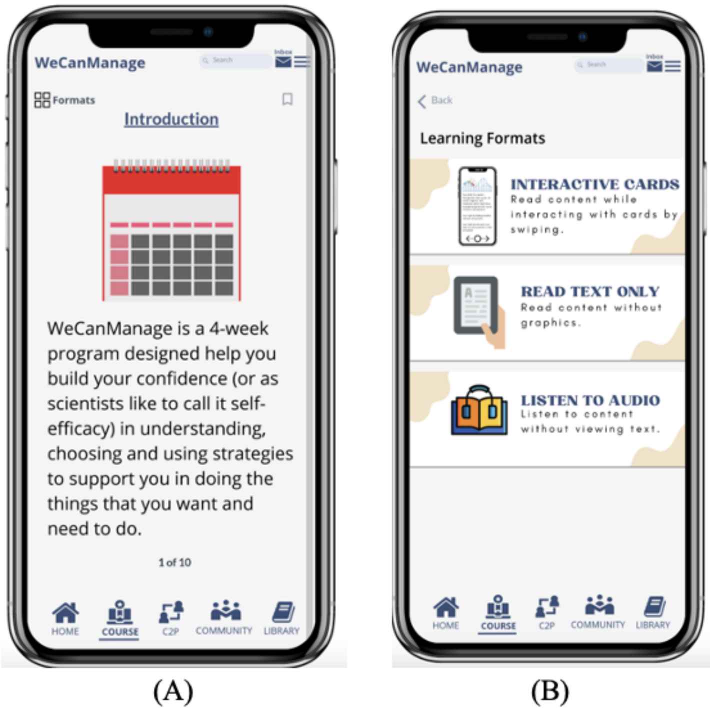
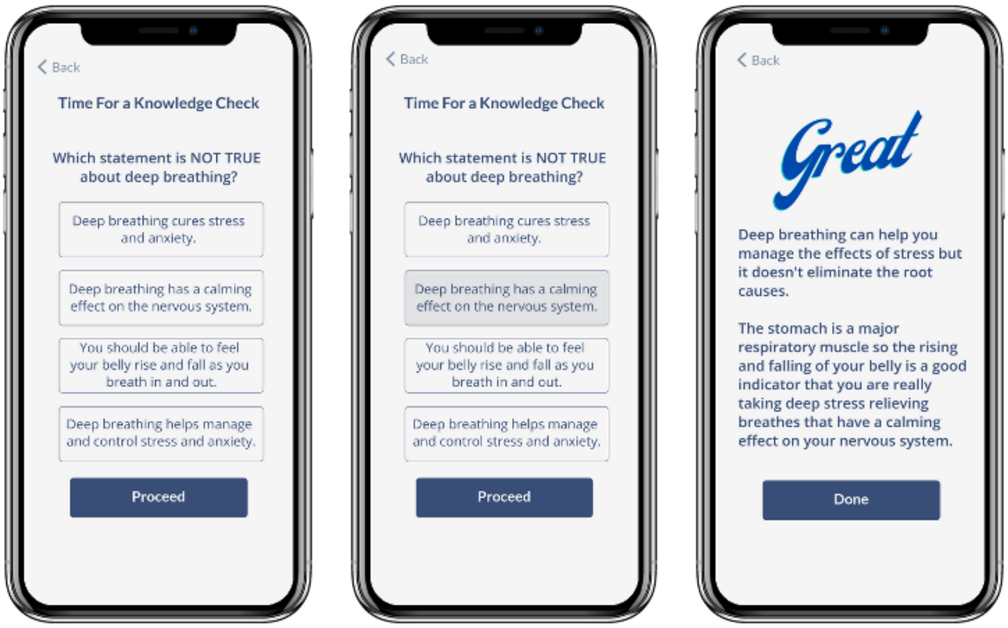


Figure 3. Sample of a knowledge check.



Methodology for Heuristic Evaluation

Nielsen's 10 principles of heuristic evaluation [36] were used for the initial testing of the prototype (Textbox 1). The prototype was given to 22 undergraduate students at a Midwestern university taking a human-computer interaction course in the Spring of 2022 who were trained in conducting heuristic evaluation. No supplemental demographic data were gathered. They were given the WeCanManage prototype during a class period of 1 hour 15 minutes. During the session, students were split into 6 groups, and each group was given 5 tasks to complete using the prototype. We created 3 sets of 5 tasks, and therefore every 2 groups completed the same tasks. The tasks included going through the introduction course module, switching to text and video fields, and filtering the users by a specific disability through the C2P page. Students logged in to classroom computers and accessed Maze, an online testing platform used to monitor assessment details [41], recorded the path taken by students to complete tasks, and presented questions about their experience to help track their progress. At the end of the session, the groups documented violations of the 10 heuristic principles and rated their usability severity on a 0-4 scale, where 0 is not a usability problem and 4 is a usability catastrophe. Furthermore, the student evaluators filled out a questionnaire through Maze providing feedback and thoughts on the prototype's design. The questionnaire covered their likes and dislikes of the design, their impressions of course modules, and the ease of changing the format of the content.

Textbox 1. Ten principles of heuristic evaluation from Nielsen [36].

1. Visibility of system status
2. Match between system and the real world
3. User control and freedom
4. Consistency and standards
5. Error prevention
6. Recognition rather than recall
7. Flexibility and efficiency of use
8. Aesthetic and minimalist design
9. Help users recognize, diagnose, and recover from errors
10. Help and documentation

Methodology for Usability Testing

We modified the prototype based on the feedback from heuristic evaluation and conducted usability testing over Zoom. We used purposive sampling with targeted outreach through cancer survivorship networks, including both clinical and community. To be eligible for participation, individuals had to meet the following inclusion criteria: be 18 years or older; have a history of breast cancer, head and neck cancer, or sarcoma; have completed active treatment; self-identify as a person with a disability; and possess the ability to understand and communicate in English. Participants received a gift card for their time. Sessions lasted approximately 90 minutes. Sessions were recorded and participants shared their screens for data collection. Participants were told to connect to Zoom on a computer or laptop device. Usability testing occurred between September 2022 and February 2023. As we encountered minor issues with the Maze platform during the heuristic evaluation, including audio malfunctions, we transitioned to Ballpark, an extension of Marvel that facilitated usability testing of the prototype. Participants were given 8 tasks to complete (see Textbox 2). They were told that they were on day 6 of the 4-week period. Consequently, they could access content from sessions 1-6, while subsequent sessions remained locked to replicate the user's sequential navigation experience, with new content being unlocked on a daily basis. The first 6 tasks were based on the course sessions and navigating through each course by reading the content cards and doing related engagement activities. Task 2 required participants to switch the viewing mode using the accessibility features (eye symbol) to the text-only mode, while task 6 involved watching a 1 minute 20 second–long mindfulness video, instead of the default card format. The final 2 tasks (tasks 7 and 8) focused on navigating the Community Forum and C2P sections. After each task, participants rated their satisfaction level and the time taken to complete each task using a 7-point Likert scale. On finishing all 8 tasks, participants had the opportunity to freely explore the app using a "think aloud" approach to express their thoughts and experiences.

To evaluate usability, participants completed the modified SUS, a reliable and valid 10-item questionnaire that assesses usability [42,43]. While the SUS has been around since 1986, it has been shown to be effective in evaluating the usability of recent health apps [44]. To calculate SUS scores, 1 is subtracted from the raw score of the odd-numbered items (those items phrased in a positive way), and the raw score of the even-numbered items (those items phrased in a negative way) is subtracted from 5. The total scores are then multiplied by 2.5 to derive the "standardized SUS score," which ranges from 0 to 100. A SUS score of 68 is considered average usability [45], while a score above 80.3 is deemed an A grade, placing it in the top 10% of scores [46] and corresponding to a narrative rating of good-excellent [47]. In addition, we included open-ended questions to gather feedback on participants' preferences and areas for improvement regarding the app. Examples of these questions include "How easy or difficult was it to see all the content on the screen?" and "What did you think of the design of the course modules?"

To assess the effectiveness of the app design, following a similar approach to Adler et al [48], we evaluated task completion by having 2 independent coders review each recording and code whether the participants

- Completed the task quickly on their own (C)
- Completed the task on their own though it took a little longer (L)
- Needed help to complete the task (H)

The coders achieved an agreement percentage of 87.5%. Any discrepancies were resolved through discussion. To assess efficiency, we analyzed the number of misclicks (clicks outside of clickable areas in the prototype) and the time taken to complete each task.

Textbox 2. Eight tasks given to usability testing participants.

Course

1. Go to the Course and click on the WeCanRelate session. Read through all of the cards.
2. Go to the Course and click on the Introduction session. Switch to Text view to read all the cards at once using the eye symbol on the bottom left of the first screen of the module.
3. Go to the Course and click on the Celebrating & Taking Stock session. Read through all the cards and then go to the reflection. Start "typing" your reflection and post it. Do you see your post accurately reflected?
4. Go to the Course and click on the Straight Talk About Symptoms session. Read through the cards and follow the link to the library and the Understanding the Cancer Rehabilitation Team Fact Sheet.
5. Go to the Course and click on the Deep Breathing session. Read through the content and complete the knowledge check. Did you get the correct answer?
6. Go to the Course and click on the Body Awareness session and go through to the end of the module by watching the video.

Community

1. Go to the Community Forum. Create a new post in the Open Discussion forum. Enter a title, select the community tag, enter text, and post your response.

Connect to Peers

1. Find the Connect to Peers (C2P) option and filter to narrow the search to people who are deaf or hard of hearing.

Ethics Approval

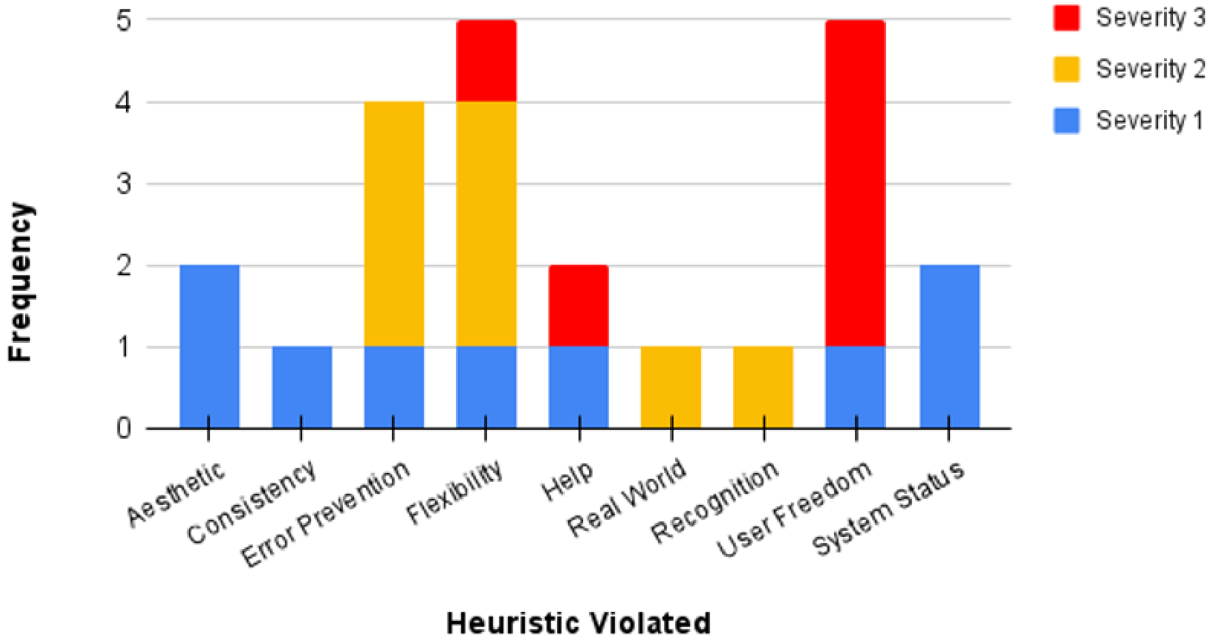
We obtained institutional review board approval from the participating universities in the project (University of Illinois Chicago #2020-1067, Northeastern Illinois University #79, and Northwestern University #NUUIC21CC03).

Results

Results of Heuristic Evaluation

We conducted an analysis of the identified heuristic violations and their severity. The highest severity rating recorded was a 3, as illustrated in Figure 4. The most frequent heuristic violations were related to flexibility, user control, and freedom, followed by error prevention. The issues identified were primarily navigation problems within the prototype, missing back buttons, and font size being too small. Suggestions for improvement were also raised, such as adding an FAQ page, a way to contact the creators or administrators, and including a walk-through or how-to page. Student evaluators expressed appreciation for the images and content, the knowledge check feature, the color scheme, and the layout. They found the app easy to read and navigate. The dislikes expressed included the absence of a help guide and nonfunctional back buttons. Additionally, some groups reported having difficulty finding the format button to switch the mode of learning to text-only or audio.

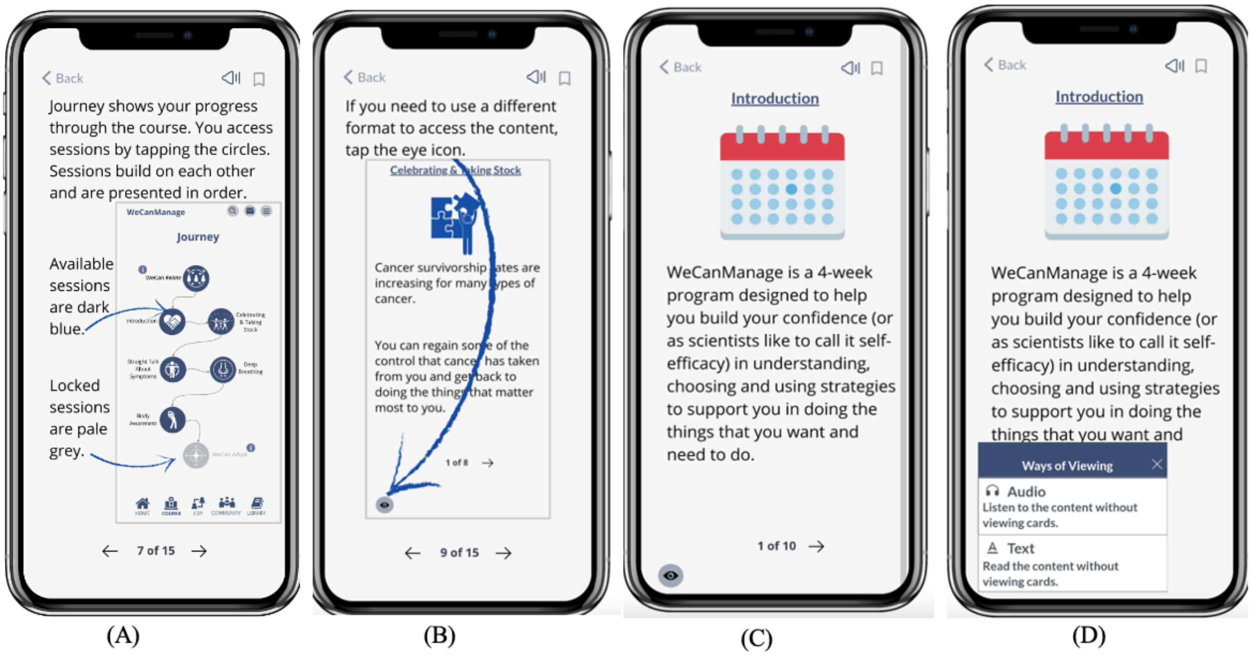
Figure 4. Graph displaying the frequency and severity of heuristic violations.



Modifications Based on Heuristic Evaluation

Drawing from the findings of the heuristic evaluation, we enhanced the prototype by introducing a help guide (Figure 5A and B) and seamlessly integrating it into the first course session. We also revised the method for switching accessibility format features (Figure 5C and D). Furthermore, we increased the font size on multiple screens and improved navigation by implementing additional back buttons for a smoother user experience.

Figure 5. Updated prototype screens after heuristic evaluation. (A,B) Help guide incorporated into the first course session. (C,D) Updates to the accessibility format and switching from card view to audio or text views.



Results of Usability Testing

We had 10 cancer survivors with disabilities (9 female, 1 male; 9 White or Caucasian, 1 Black or African American) who completed usability testing. The average age of the participants was 59 years. Usability scores show that participants had an overall positive reception to the design of the prototype. We had an average SUS score of 81; our prototype's usability is therefore considered good to excellent with a grade of an A and in the top 10%.

We assessed participants' satisfaction levels and the time taken to complete each task. The average scores for these 2 measurements are presented in Table 1. Generally, participants exhibited high satisfaction rates; however, lower numbers were observed for task 2 (finding the eye icon to change the accessibility format), task 7 (creating a post in the Community Forum), and task 8 (using the filter in C2P).

In addition, we evaluated the effectiveness of the app design by categorizing participants' task completion into 3 groups: completed quickly (C), completed with a little more time (L), or required assistance to complete the task (H). Overall, most participants completed their tasks without any issues, with only 17 of 80 cases (21%) needing help to complete them (see Figure 6). During task 1, a slight learning curve was observed as some participants had difficulty locating the correct module, leading to the need for assistance in completing the task. However, this issue was not prevalent in subsequent tasks. Task 2 revealed that some participants encountered challenges while switching the card format to text view using the eye symbol, as they had trouble locating the button. In task 4, some participants faced difficulties clicking on the correct resource within the Library as directed in the learning module. For tasks 7 and 8, several participants struggled to navigate both the Community and C2P sections because certain text and icons were too small or unclear in their function, leading to confusion on what to do.

Likewise, while analyzing efficiency based on the number of misclicks per task, tasks 7 and 8 exhibited notably higher misclick rates (Table 2). The table also presents the actual time taken per task, with task 1 showing higher time than the other tasks. As mentioned earlier, task 1 had a learning curve, but it also involved reading the most cards (15 cards) as we integrated the help guide into the first course session. Therefore, this finding is expected given the additional content to review in task 1.

The prototype's help guide received a positive response, with 8 of 10 participants (80%) rating it as very helpful or extremely helpful. Similarly, 8 of 10 participants (80%) reported finding the eye symbol (to change the course format) easily. In response to open-ended questions, participants expressed their likes and dislikes of the prototype and its design. Many participants shared positive opinions on the design and content of the modules, finding them helpful and insightful. The video located within one of the modules received positive feedback, with some expressing a desire for additional videos. The purpose of the Community section was well liked as participants enjoyed having a place to freely express themselves with other cancer survivors and appreciated the opportunity for users to support each other. The Library resources were found to be informative and useful, covering a wide range of topics.

Our findings were overwhelmingly positive, supported by quotes from participants (some written and some oral):

> *I want to see the whole thing work! I know that this is a prototype, but I want to see more!*
>
> *Great app, it would have been very helpful to me when I was just out [of] treatment.*
>
> *Even though I'm not very comfortable with technology, and that might be because of my age, … I don't think that this would be difficult for me. I think there'd be a real fast learning curve. I felt good and positive when I realized I had learned something, and I could just click on it now without having to think about it.*
>
> *I do like the app. I like that I know I'm not alone feeling this way.*

These participant quotes reflect their enthusiasm and positive experiences with the app, highlighting its potential benefits and ease of use.

On the basis of our session observations and participants' feedback on areas for improvement, we identified several issues:

- Accessibility concerns, including small font sizes and icons, particularly with the navigation arrows on cards, the top navigation bar, and the eye icon.
- Some participants experienced confusion while navigating the Community page when creating new posts.
- Difficulty in locating and using the filter option within the C2P page.
- Participants expressed a desire for an easy way to return to the help guide.
- Feedback indicated a preference for changing the robotic voices used in the audio format for the modules. The prototype used Google US English from voicegenerator.io, but the intention is to have a real person's voice in future implementations.

Addressing these areas for improvement can further enhance the app's usability and user experience.

Table 1. Average satisfaction per task and time per task (out of 7).

Task	Average task satisfaction	Average time satisfaction
1	6.5	6.4
2	5.7	5.9
3	6.6	6.5
4	6.5	6.2
5	6.7	6.3
6	6.8	6.6
7	5.2	5.5
8	5.8	5.7

Figure 6. Graph displaying the frequency of H (required assistance to complete the task), C (completed quickly), and L (completed with a little more time) ratings given to participants as they completed a task.

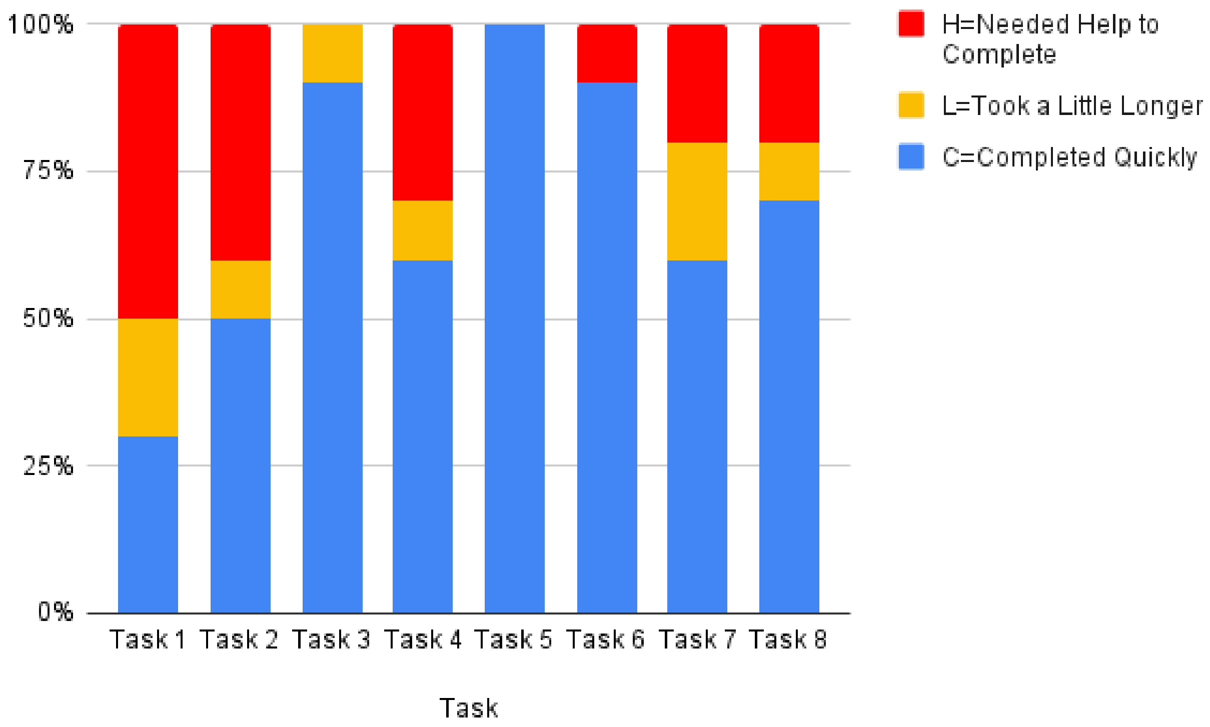


Table 2. Percentage of misclicks and time per task.

Task	Misclicks (%)	Time (minutes)
1	8	3:28
2	4.75	2:28
3	3.30	2:19
4	5.64	2:13
5	0.83	2:15
6	0	1:57
7	19.24	1:34
8	16.38	0:44

Modifications Based on Usability Testing

On the basis of the findings from usability testing, we made several modifications to the prototype. To enhance usability, we increased the sizes of navigation icons, the eye icon, arrows within cards, and the top navigation bar. Throughout the application, we enlarged or bolded fonts for easier reading, including the “create new post” button in the Community section. We redesigned the layout of the Community Forum, increasing text and margins to achieve a cleaner and more concise design. Additionally, we revamped the subscribe button to reduce confusion (see Figures 7 and 8). To improve accessibility, we enlarged the C2P filter. Finally, we added a convenient way to return to the help guide by including it in the hamburger menu icon on the main page. These changes aim to enhance user experience and address the identified issues during usability testing.

Figure 7. Modifications made to the Community before and after usability testing.

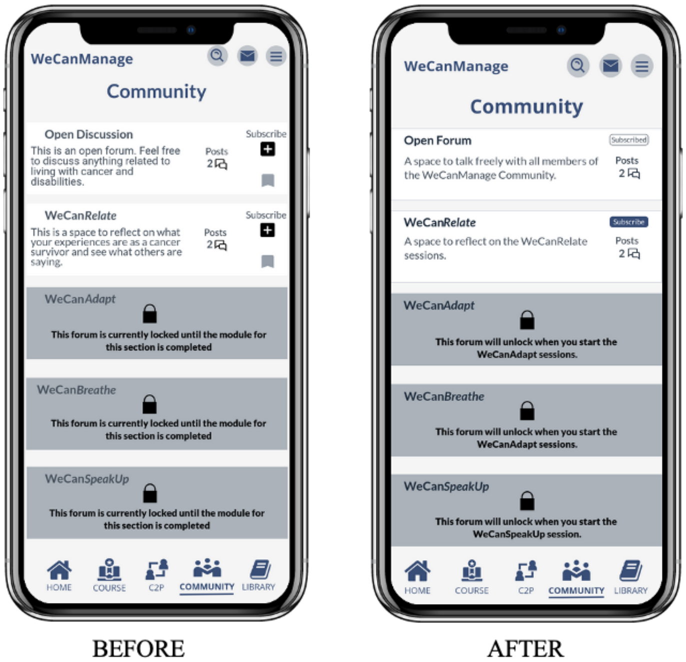
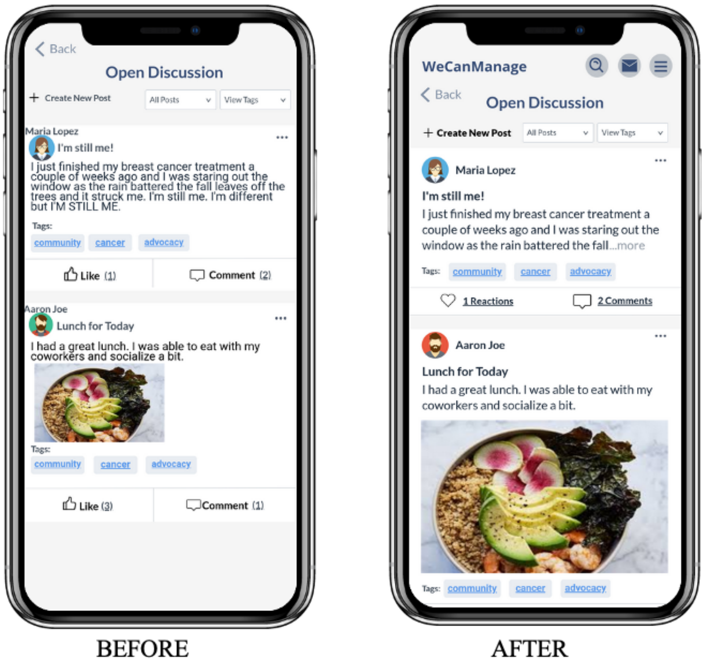


Figure 8. Modifications made to the Open Discussion design before and after usability testing .



Discussion

Principal Findings

Cancer and its treatments can lead to long-term disabilities, significantly impacting a survivor's overall quality of life [10]. Unfortunately, postcancer treatment resources are often limited, further exacerbating the challenges faced by survivors [49,50]. To address this, we developed a high-fidelity prototype for an mHealth app called WeCanManage, aimed at empowering cancer survivors with disabilities to effectively self-manage the long-term effects of cancer treatment. Through conducting the heuristic evaluation, valuable improvements were made, including the incorporation of a helpful guide and the enhancement of accessibility formatting options, ultimately enhancing the overall user experience of the app.

In usability testing, we engaged cancer survivors with disabilities, using multiple methods such as task completion, think aloud strategies, SUS, perceived task satisfaction, and open-ended questions. These methods have been extensively used to evaluate various applications, with the SUS being one of the commonly used questionnaires [21]. The results of usability testing were overwhelmingly positive, with cancer survivors expressing appreciation for the app's content, features, and design. The prototype achieved an impressive SUS score of 81, ranking it in the top 10% of scores and earning an A grade. Moreover, participants reported high satisfaction levels and efficiency, with average scores of 6.2 and 6.1 (out of 7), respectively. Conducting usability testing enabled us to thoroughly assess the app's overall effectiveness, efficiency, satisfaction, and usability. We were able to identify areas for improvement, particularly in terms of accessibility. The insights gained from this testing process have allowed us to refine and enhance the app, ensuring a positive user experience for cancer survivors with disabilities.

In a study by Fuller-Tyszkiewicz et al [24], end users rated an mHealth prototype higher in usability and reported a more positive experience than clinical experts. Interestingly, users did not share the same concerns about the amount and layout of content presented as the experts had anticipated [24]. This discrepancy underscores the significance of testing potential users to tailor the app to their specific needs and preferences. While expert opinions (whether clinical or in design) are valuable, evaluating an app on actual users is ideal.

Implications for Designers and Researchers

One of our primary findings is the importance of accessibility when designing applications for cancer survivors. Our app was specifically designed for cancer survivors with disabilities, and as such, we incorporated customized options to switch the learning style. Users could choose between clicking through content cards and accessing audio or text-only views. This flexibility proved to be helpful, particularly for participants with cognitive issues like "chemo brain," who found it easier to navigate the audio versions of the course sessions. However, during testing, we identified other accessibility concerns related to font sizes and icons. Some users found them too small to see, click on, and navigate effectively. Addressing these issues is essential to ensure an inclusive and user-friendly experience for all app users.

The importance of having a help feature was revealed during heuristic evaluation, and through usability testing, we learned that users expressed a desire for a convenient way to return to the help guide. In response to this feedback, we have now incorporated the option to access the help guide directly from our main menu.

One comment expressed by many of our participants was how lonely the experience of a cancer survivor is. Consistent with findings from other studies that highlight the significance of social features in mHealth apps [51], participants expressed their appreciation for the Community Forum and C2P sections. These features provide a valuable opportunity for them to connect with others facing similar situations, fostering a sense of community and support. Additionally, participants reported that reading the content in the course sessions made them realize that their experiences were shared by others, helping them feel less isolated and reassured that they were not alone in their journey. When asked what they liked about the app, one participant wrote the following: "The information, reliable and trustworthy, … and the realization that I am not alone."

Limitations

Our aim was to achieve a minimum of 12 participants for usability testing, as SUS results are ideally derived from 12 or more participants [52,53]. However, we encountered challenges in recruitment because of technical difficulties, such as some participants lacking access to a laptop or facing issues with Zoom and screen sharing, leading to incomplete usability testing. Additionally, recruitment was hindered by our specific inclusion criteria, which focused on individuals who identified as having a disability. These challenges impacted our ability to reach the desired number of participants for the usability testing phase. Nevertheless, it is worth noting that according to Nielsen [54], 5 participants are typically adequate for identifying usability problems. Thus, we can reasonably infer that our processes have successfully identified the majority of issues, providing a level of confidence in the validity of our findings despite the lower number of participants in the usability testing phase. Additionally, it is worth mentioning that several studies evaluating mHealth prototypes have used the SUS with fewer than 12 participants [29,31,37]. We encountered instances where some participants experienced lingering effects of cancer and its treatment, but they did not self-identify as having a disability, resulting in their exclusion from usability testing. This finding has important implications for the implementation and adoption of WeCanManage, ensuring that cancer survivors experiencing disabling aftereffects can fully benefit from the tool and appreciate its relevance and value in their daily lives and experiences.

Furthermore, as this was a prototype, not all features were fully implemented (eg, the ability to create a post on the forum or direct message a user was mimicked), which may have caused some participants to encounter difficulties in the Community section of the prototype. In addition, during usability testing, participants expressed concerns regarding text and icon sizes. It is important to note that the testing was conducted over Zoom using computers (not mobile devices), and the prototype's size (matching that of a phone) might have posed challenges during interaction, which may not be representative of the real application's experience. Finally, it is worth noting that the age of participants and their level of comfort with technology might have influenced their overall experience [55]. Nevertheless, because these individuals constitute our target user base, it remains essential for us to maintain the app's usability and accessibility to meet their needs.

Conclusions

When creating an mHealth app, it is crucial to evaluate it with the target users in mind, in our case, cancer survivors with disabilities. Usability testing allowed us to identify the design's strengths and areas requiring improvement. The WeCanManage prototype achieved a SUS score of 81, placing it in the top 10% of scores. Our future work will involve feasibility testing of an implemented web-based mobile app of WeCanManage. This will enable us to further refine the application and ensure that it meets the needs and preferences of our target users, enhancing its overall usability and impact.

Acknowledgments

This research was supported by the National Cancer Institute of the National Institutes of Health (U54CA202995, U54CA202997, and U54CA203000). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Additional support was provided by Northeastern Illinois University's Student Center for Science Engagement, Committee on Organized Research, and College of Business and Technology. The authors would also like to acknowledge contributions from the following team members: Melissa Delgado, Daniela Raudales Reyes, Elizabeth Jarvis, Sabrina Cadena, and Bruriah Horowitz.

Conflicts of Interest

None declared.

References

1. Cancer facts and figures 2023. American Cancer Society. 2023. URL: https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/2023-cancer-facts-figures.html [accessed 2024-02-21]
2. National Research Council. In: Stovall E, Hewitt M, Greenfield S, editors. From Cancer Patient to Cancer Survivor: Lost in Transition. Washington, DC: National Academies Press; 2005.
3. Alfano CM, Ganz PA, Rowland JH, Hahn EE. Cancer survivorship and cancer rehabilitation: revitalizing the link. J Clin Oncol 2012;30(9):904-906 [FREE Full text] [doi: 10.1200/JCO.2011.37.1674] [Medline: 22355063]
4. Henderson F, Cross AJ, Baraniak AR. 'A new normal with chemobrain': experiences of the impact of chemotherapy-related cognitive deficits in long-term breast cancer survivors. Health Psychol Open 2019;6(1):2055102919832234 [FREE Full text] [doi: 10.1177/2055102919832234] [Medline: 30873289]
5. Nguyen NP, Sallah S, Karlsson U, Antoine JE. Combined chemotherapy and radiation therapy for head and neck malignancies: quality of life issues. Cancer 2002;94(4):1131-1141 [FREE Full text] [doi: 10.1002/cncr.10257] [Medline: 11920484]
6. International Classification of Functioning, Disability, and Health: ICF. Geneva: World Health Organization; 2001.
7. What is the definition of disability under the ADA? ADA National Network. 2015. URL: https://adata.org/faq/what-definition-disability-under-ada [accessed 2024-02-21]
8. Winters CA, Marshall HK, Victorson DE, Adler RF, Magasi S. Does the Canadian Occupational Performance Measure (COPM) capture the complex experiences of cancer survivors? A mixed methods approach. Occup Ther Health Care 2023:1-17 [FREE Full text] [doi: 10.1080/07380577.2023.2215875] [Medline: 37302410]
9. Magasi S, Marshall HK, Winters C, Victorson D. Cancer survivors' disability experiences and identities: a qualitative exploration to advance cancer equity. Int J Environ Res Public Health 2022;19(5):3112 [FREE Full text] [doi: 10.3390/ijerph19053112] [Medline: 35270802]
10. Newman RM, Alfano CM, Radomski MV, Pergolotti M, Wolf TJ, Sleight AG, et al. Catalyzing research to optimize cancer survivors' participation in work and life roles. OTJR (Thorofare N J) 2019;39(4):189-196 [FREE Full text] [doi: 10.1177/1539449219844749] [Medline: 31046601]
11. Raj VS, Pugh TM, Yaguda SI, Mitchell CH, Mullan SS, Garces NS. The who, what, why, when, where, and how of team-based interdisciplinary cancer rehabilitation. Semin Oncol Nurs 2020;36(1):150974 [FREE Full text] [doi: 10.1016/j.soncn.2019.150974] [Medline: 31955923]
12. Stubblefield MD. The underutilization of rehabilitation to treat physical impairments in breast cancer survivors. PM R 2017;9(9S2):S317-S323 [FREE Full text] [doi: 10.1016/j.pmrj.2017.05.010] [Medline: 28942906]
13. What is mobile technology? IBM. 2023. URL: https://www.ibm.com/topics/mobile-technology [accessed 2023-05-23]
14. Mobile fact sheet. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2023-05-21]
15. Curigliano G, Banerjee S, Cervantes A, Garassino MC, Garrido P, Girard N, et al. Managing cancer patients during the COVID-19 pandemic: an ESMO multidisciplinary expert consensus. Ann Oncol 2020;31(10):1320-1335 [FREE Full text] [doi: 10.1016/j.annonc.2020.07.010] [Medline: 32745693]
16. Pohl M. 325,000 mobile health apps available in 2017—Android now the leading mHealth platform. Research2Guidance. 2017. URL: https://research2guidance.com/325000-mobile-health-apps-available-in-2017/ [accessed 2024-02-21]
17. Charbonneau DH, Hightower S, Katz A, Zhang K, Abrams J, Senft N, et al. Smartphone apps for cancer: a content analysis of the digital health marketplace. Digit Health 2020;6:2055207620905413 [FREE Full text] [doi: 10.1177/2055207620905413] [Medline: 32110428]
18. Boulley GE, Leroy T, Bernetière C, Paquienseguy F, Desfriches-Doria O, Préau M. Digital health interventions to help living with cancer: a systematic review of participants' engagement and psychosocial effects. Psychooncology 2018;27(12):2677-2686 [FREE Full text] [doi: 10.1002/pon.4867] [Medline: 30152074]

19. Silva EH, Lawler S, Langbecker D. The effectiveness of mHealth for self-management in improving pain, psychological distress, fatigue, and sleep in cancer survivors: a systematic review. J Cancer Surviv 2019;13(1):97-107 [FREE Full text] [doi: 10.1007/s11764-018-0730-8] [Medline: 30635865]
20. Zheng C, Chen X, Weng L, Guo L, Xu H, Lin M, et al. Benefits of mobile apps for cancer pain management: systematic review. JMIR Mhealth Uhealth 2020;8(1):e17055 [FREE Full text] [doi: 10.2196/17055] [Medline: 32012088]
21. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: a scoping review. Int J Med Inform 2019;126:95-104 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
22. Hartzler AL, Venkatakrishnan A, Mohan S, Silva M, Lozano P, Ralston JD, et al. Acceptability of a team-based mobile health (mHealth) application for lifestyle self-management in individuals with chronic illnesses. Annu Int Conf IEEE Eng Med Biol Soc 2016;2016:3277-3281 [FREE Full text] [doi: 10.1109/EMBC.2016.7591428] [Medline: 28269007]
23. Martinez W, Threatt AL, Rosenbloom ST, Wallston KA, Hickson GB, Elasy TA. A patient-facing diabetes dashboard embedded in a patient web portal: design sprint and usability testing. JMIR Hum Factors 2018;5(3):e26 [FREE Full text] [doi: 10.2196/humanfactors.9569] [Medline: 30249579]
24. Fuller-Tyszkiewicz M, Richardson B, Klein B, Skouteris H, Christensen H, Austin D, et al. A mobile app-based intervention for depression: end-user and expert usability testing study. JMIR Ment Health 2018;5(3):e54 [FREE Full text] [doi: 10.2196/mental.9445] [Medline: 30139722]
25. Santesteban-Echarri O, Tang J, Fernandes J, Addington J. Development and usability testing of SOMO, a mobile-based application to monitor social functioning for youth at clinical high-risk for psychosis. Digit Psych 2020;1(1):4-19 [FREE Full text] [doi: 10.24989/dp.v1i1.1806]
26. Reynoldson C, Stones C, Allsop M, Gardner P, Bennett MI, Closs SJ, et al. Assessing the quality and usability of smartphone apps for pain self-management. Pain Med 2014;15(6):898-909 [FREE Full text] [doi: 10.1111/pme.12327] [Medline: 24422990]
27. Woods L, Cummings E, Duff J, Walker K. Design thinking for mHealth application co-design to support heart failure self-management. Stud Health Technol Inform 2017;241:97-102. [Medline: 28809190]
28. Fu MR, Axelrod D, Guth AA, Rampertaap K, El-Shammaa N, Hiotis K, et al. mHealth self-care interventions: managing symptoms following breast cancer treatment. Mhealth 2016;2(28):1-18 [FREE Full text] [doi: 10.21037/mhealth.2016.07.03] [Medline: 27493951]
29. Donald M, Beanlands H, Straus SE, Smekal M, Gil S, Elliott MJ, et al. A web-based self-management support prototype for adults with chronic kidney disease (My kidneys My Health): co-design and usability testing. JMIR Form Res 2021;5(2):e22220 [FREE Full text] [doi: 10.2196/22220] [Medline: 33560245]
30. Ospina-Pinillos L, Davenport T, Diaz AM, Navarro-Mancilla A, Scott EM, Hickie IB. Using participatory design methodologies to co-design and culturally adapt the Spanish version of the Mental Health eClinic: qualitative study. J Med Internet Res 2019;21(8):e14127 [FREE Full text] [doi: 10.2196/14127] [Medline: 31376271]
31. Harte R, Quinlan LR, Glynn L, Rodríguez-Molinero A, Baker PM, Scharf T, et al. Human-centered design study: enhancing the usability of a mobile phone app in an integrated falls risk detection system for use by older adult users. JMIR Mhealth Uhealth 2017;5(5):e71 [FREE Full text] [doi: 10.2196/mhealth.7046] [Medline: 28559227]
32. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
33. Iacobelli F, Adler RF, Buitrago D, Buscemi J, Corden ME, Perez-Tamayo A, et al. Designing an mHealth application to bridge health disparities in Latina breast cancer survivors: a community-supported design approach. Design Health (Abingdon) 2018;2(1):58-76 [FREE Full text] [doi: 10.1080/24735132.2018.1452871] [Medline: 30506017]
34. Beatty L, Koczwara B, Butow P, Turner J, Girgis A, Schofield P, et al. Development and usability testing of a web-based psychosocial intervention for women living with metastatic breast cancer: finding my way-advanced. J Cancer Surviv 2021;15(3):403-409 [FREE Full text] [doi: 10.1007/s11764-021-01019-5] [Medline: 33723741]
35. Børøsund E, Mirkovic J, Clark MM, Ehlers SL, Andrykowski MA, Bergland A, et al. A stress management app intervention for cancer survivors: design, development, and usability testing. JMIR Form Res 2018;2(2):e19 [FREE Full text] [doi: 10.2196/formative.9954] [Medline: 30684438]
36. Nielsen J. Heuristic evaluation. In: Nielsen J, Mack RL, editors. Usability Inspection Methods. New York: John Wiley & Sons; 1994.
37. Derks YP, Klaassen R, Westerhof GJ, Bohlmeijer ET, Noordzij ML. Development of an ambulatory biofeedback app to enhance emotional awareness in patients with borderline personality disorder: multicycle usability testing study. JMIR Mhealth Uhealth 2019;7(10):e13479 [FREE Full text] [doi: 10.2196/13479] [Medline: 31617851]
38. Magasi S, Banas J, Horowitz B, Reis JP, The K, Wilson T, et al. WeCanConnect: development of a community-informed mhealth tool for people with disabilities and cancer. Prog Community Health Partnersh 2019;13(5):49-59 [FREE Full text] [doi: 10.1353/cpr.2019.0038] [Medline: 31378735]
39. Rapid prototyping, testing and handoff for modern design teams. Marvel. URL: https://marvelapp.com/ [accessed 2023-09-20]
40. Adler RF, Morales P, Sotelo J, Magasi S. Developing an mHealth app for empowering cancer survivors with disabilities: co-design study. JMIR Form Res 2022;6(7):e37706 [FREE Full text] [doi: 10.2196/37706] [Medline: 35881439]

41. The holistic solution for continuous product discovery. Maze. URL: https://maze.co/ [accessed 2023-09-20]
42. Brooke J. SUS: a quick and dirty usability scale. In: Thomas B, Weerdmeester B, McClelland IL, Jordan J, editors. Usability Evaluation In Industry. London, Bristol, PA: CRC Press; 1996:189-194.
43. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Intl J Human–Computer Interact 2008;24(6):574-594 [FREE Full text] [doi: 10.1080/10447310802205776]
44. Hyzy M, Bond R, Mulvenna M, Bai L, Dix A, Leigh S, et al. System usability scale benchmarking for digital health apps: meta-analysis. JMIR Mhealth Uhealth 2022;10(8):e37290 [FREE Full text] [doi: 10.2196/37290] [Medline: 35980732]
45. Sauro J, Lewis JR. Quantifying the User Experience: Practical Statistics for User Research. Cambridge, MA: Morgan Kaufmann; 2016.
46. Sauro J. Measuring usability with the System Usability Scale (SUS). MeasuringU. 2011. URL: https://measuringu.com/sus/ [accessed 2023-08-01]
47. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J usability Stud 2009;4(3):114-123.
48. Adler RF, Paley A, Li Zhao AL, Pack H, Servantez S, Pah AR, et al. A user-centered approach to developing an AI system analyzing U.S. federal court data. Artif Intell Law 2022;31(3):547-570 [FREE Full text] [doi: 10.1007/s10506-022-09320-z]
49. Barsevick AM, Leader A, Bradley PK, Avery T, Dean LT, DiCarlo M, et al. Post-treatment problems of African American breast cancer survivors. Support Care Cancer 2016;24(12):4979-4986 [FREE Full text] [doi: 10.1007/s00520-016-3359-z] [Medline: 27543161]
50. Ridner SH, Rhoten BA, Radina ME, Adair M, Bush-Foster S, Sinclair V. Breast cancer survivors' perspectives of critical lymphedema self-care support needs. Support Care Cancer 2016;24(6):2743-2750 [FREE Full text] [doi: 10.1007/s00520-016-3089-2] [Medline: 26810422]
51. Jessen S, Mirkovic J, Ruland CM. Creating gameful design in mHealth: a participatory co-design approach. JMIR Mhealth Uhealth 2018;6(12):e11579 [FREE Full text] [doi: 10.2196/11579] [Medline: 30552080]
52. Lewis JR. The system usability scale: past, present, and future. Int J Hum-Comput Interact 2018;34(7):577-590 [FREE Full text]
53. Tullis TS, Stetson JN. A comparison of questionnaires for assessing website usability. 2004 Presented at: Usability Professionals' Association Conference 2004: 13th Annual UPA Conference; June 7-11, 2004; Minneapolis, Minnesota, USA. [doi: 10.1007/978-1-84628-795-4_3]
54. Nielsen J. Why you only need to test with 5 users. Nielsen Norman Group. 2000. URL: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/ [accessed 2024-02-21]
55. Iancu I, Iancu B. Designing mobile technology for elderly. A theoretical overview. Technol Forecast Soc Change 2020;155:119977 [FREE Full text] [doi: 10.1016/j.techfore.2020.119977]

Abbreviations

C2P: Connect to Peers
mHealth: mobile health
SUS: System Usability Scale

Edited by C Jacob; submitted 03.08.23; peer-reviewed by E Baker, C Baxter; comments to author 06.09.23; revised version received 07.11.23; accepted 11.11.23; published 02.04.24.

Please cite as:
Adler RF, Baez K, Morales P, Sotelo J, Victorson D, Magasi S
Evaluating the Usability of an mHealth App for Empowering Cancer Survivors With Disabilities: Heuristic Evaluation and Usability Testing
JMIR Hum Factors 2024;11:e51522
URL: https://humanfactors.jmir.org/2024/1/e51522
doi:10.2196/51522
PMID:38564261

Feasibility of Monitoring Heart and Respiratory Rates Using Nonwearable Devices and Consistency of the Measured Parameters: Pilot Feasibility Study

Kasumi Ikuta, RN, MSN; Miya Aishima, RN, PhD; Maiko Noguchi-Watanabe, RN, PhD; Sakiko Fukui, RN, PhD

Department of Home Health and Palliative Care Nursing, Graduate School of Health Care Sciences, Tokyo Medical and Dental University, Bunkyo-Ku, Japan

Corresponding Author:
Sakiko Fukui, RN, PhD
Department of Home Health and Palliative Care Nursing, Graduate School of Health Care Sciences, Tokyo Medical and Dental University, , Bunkyo-Ku, , Japan

Abstract

Background: As Japan is the world's fastest-aging society with a declining population, it is challenging to secure human resources for care providers. Therefore, the Japanese government is promoting digital transformation and the use of nursing care equipment, including nonwearable devices that monitor heart and respiratory rates. However, the feasibility of monitoring heart and respiratory rates with nonwearable devices and the consistency of the rates measured have not been reported.

Objective: In this study, we focused on a sheet-type nonwearable device (Safety Sheep Sensor) introduced in many nursing homes. We evaluated the feasibility of monitoring heart rate (HR) and respiratory rate (RR) continuously using nonwearable devices and the consistency of the HR and RR measured.

Methods: A sheet-type nonwearable device that measured HR and RR every minute through body vibrations was placed under the mattress of each participant. The participants in study 1 were healthy individuals aged 20 - 60 years (n=21), while those in study 2 were older adults living in multidwelling houses and required nursing care (n=20). The HR was measured using standard methods by the nurse and using the wearable device (Silmee Bar-type Lite sensor), and RR was measured by the nurse. The primary outcome was the mean difference in HR and RR between nonwearable devices and standard methods.

Results: The mean difference in HR was −0.32 (SD 3.12) in study 1 and 0.04 (SD: 3.98) in study 2; both the differences were within the predefined accepted discrepancies (<5 beats/min). The mean difference in RR was −0.98 (SD 3.01) in study 1 and −0.49 (SD 2.40) in study 2; both the differences were within the predefined accepted discrepancies (3 breaths/min).

Conclusions: HR and RR measurements obtained using the nonwearable devices and the standard method were similar. Continuous monitoring of vital signs using nonwearable devices can aid in the early detection of abnormal conditions in older people.

(JMIR Hum Factors 2024;11:e56547) doi:10.2196/56547

KEYWORDS

heart rate; older adults; respiratory rate; nonwearable devices; vital signs

Introduction

The vital signs of older adults change with age [1,2], making it necessary to monitor them at their residences and in nursing homes [3,4]. Continuous monitoring of their heart rate (HR) and respiratory rate (RR) to obtain essential data and immediate assessment of any abnormality by medical personnel have proven successful in reducing hospitalization and mortality rates [5]. However, measuring vital signs once or twice daily for older people living at home and in nursing homes is unrealistic [6]. Moreover, the impact of COVID-19 necessitated remote monitoring of vital signs with fewer visits from medical staff to prevent infections [7]. Therefore, continuous vital sign monitoring devices that can be used without affecting daily activities must be developed.

Wearable devices (WDs), including wrist-worn devices [8], ViSi Mobile (Sotera Wireless, Carlsbad, CA) [9], and health patches (VitalConnect, San Jose, CA) [9], can monitor vital signs without interfering with daily activities. Measuring vital signs continuously with a WD may improve patient comfort and safety [10-12] and reduce nurse workload [10] by reducing the number of nurse-conducted measurements. However, redness and itching are common among older adults using WDs [9] because their skin is particularly fragile and WDs come in direct contact with the skin [13]. Therefore, developing and introducing nonwearable devices (NWDs) for monitoring older patients continuously is desirable [14].

NWDs, which measure vital signs without direct patient contact, reduce older adults' feeling of restraint and awareness of being monitored as well as the workload associated with vital sign assessments for care providers. NWDs such as highly sensitive fiber optic mattress [11,15] and office chairs [16] are being researched and developed. Verification of its use in magnetic resonance imaging as well as in medical and nursing care settings is underway [17] and has attracted widespread attention.

As Japan is the world's fastest-aging society with a declining population [18], it is challenging to secure human resources for care providers [19]. Therefore, the Japanese government is promoting digital transformation and the use of nursing care equipment, including NWDs. Many nursing homes have purchased and introduced NWDs since 2015 when the government started providing assistance payment for such devices as policy guidance [20,21]. Sheet-type NWDs, such as Safety Sheep sensors α (NJI Co., Ltd., Fukushima, Japan), NEMURI SCAN (Paramount Bed Co., Ltd., Tokyo, Japan), and Mimamoleaf (Techno Horizon Co., Ltd., Nagoya, Japan), have been developed and introduced in Japan. We focused on Safety Sheep sensors α, which is a monitoring device that has been introduced in many nursing homes [22], including 3319 nursing homes in Japan as of December 2022. This NWD includes a highly sensitive pressure sensor placed under a mattress and does not come in contact with the individual's skin. It detects body vibrations and calculates the individual's status (lying on bed/moving on bed/getting out of bed), HR, and RR every minute. These data are displayed on a monitor at nursing homes. Care providers check these data and, if they are abnormal, can take immediate action and provide appropriate care.

However, the data from NWDs, including sheet-type NWDs, cannot be used by nursing homes as longitudinal data to evaluate changes in individuals accurately and provide appropriate care [23] because the consistency of the HR and RR measured by such instruments has not been evaluated yet [24]. Moreover, based on the accumulated data, it may be possible to detect deteriorating conditions during end-of-life care and unplanned hospital visits at an early stage, concentrate care on high-risk targets, and improve the quality of care using fewer personnel. Therefore, this study aimed to evaluate the feasibility of continuous HR and RR monitoring and consistency of the rates measured using the NWD.

Methods

Setting and Participants

This was a prospective, observational, and pilot feasibility study. Two studies were conducted to evaluate the feasibility of continuous HR and RR monitoring, and the consistency of the HRs and RRs was measured using an NWD (Safety Sheep sensors α). In this study, the study power was calculated based on a previous study [9], and a sample size of 20 was estimated to obtain sufficient data for analysis.

Study 1: Healthy Participants

Study 1 included 22 healthy participants aged 20 - 60 years who were working at a company. The authors recruited participants from June 1 to June 30, 2022. Participants were excluded if they had been diagnosed with heart or respiratory disease. To maintain the privacy of the participants, their vital signs were measured in a private room in company A. The NWD was placed under the mattress after preparing a bed. Each participant came to the room at separate designated times for measurements.

Study 2: Older Adults Who Needed Nursing Care

Study 2 included 26 older adults aged ≥65 years who were living in multidwelling houses managed by a company and required nursing care. Each multidwelling house had a care worker who was available for 24 hours daily. Thus, older adults who required nursing care could immediately receive nursing care from their care workers. Managers working in multidwelling houses recruited older adults for this study from July 1 to October 30, 2022. Participants diagnosed with cardiac or respiratory disease or those with implanted medical electronic devices, such as pacemakers, were excluded. Each participant was visited at appointed dates and times, and their HR and RR were measured in their rooms.

Data Collection

In this study, data were collected by 2 nurses: one measured the HR and RR, while the other recorded them. They waited for the participant to visit the private room at company A in study 1 but visited the participant's room themselves in study 2. First, the nurses explained the purpose and methods of this study to the participants. Second, the participants lay on a bed with the NWD in place. Third, the nurse who took the measurements attached a WD to the participants' chests and asked them to rest for 5 minutes. The same nurse then measured the RR over 1 minute and reported five sets of measurements to the recording nurse. The recording nurse documented the start time of the measurement and the number of RRs. After a 1-minute break, the HR was measured using the same method. After completing the measurements, the participants were informed that the procedure was complete, and the WD was removed. Finally, the participants were checked for skin abnormalities.

Measurement

Nonwearable Device

The Safety Sheep sensor α (width 800 mm; height 17 mm; depth 150 mm) was used as the NWD (Figure 1) [25]. In the multidwelling house, this NWD was placed in all the residents' beds. The NWD comprises a sensor unit, amplifier/analog-to-digital converter unit, signal processing unit, and communication unit (Figure 2). The HR and RR are determined from heartbeat- and respiration-derived vibration waves transmitted through the device, respectively.

First, in the sensor unit, 6 piezoelectric elements capable of detecting minute vibrations are placed at 10-cm intervals. The vibrations generated by the participants mainly included those derived from heartbeat, respiration, and body movements. The vibration data acquired at the sensor section are converted into digital data at the amplifier/analog-to-digital converter unit. Subsequently, in the signal processing unit, HR and RR are determined from the digitally converted data (Figure 3).

Figure 1. Positioning of the nonwearable devices.

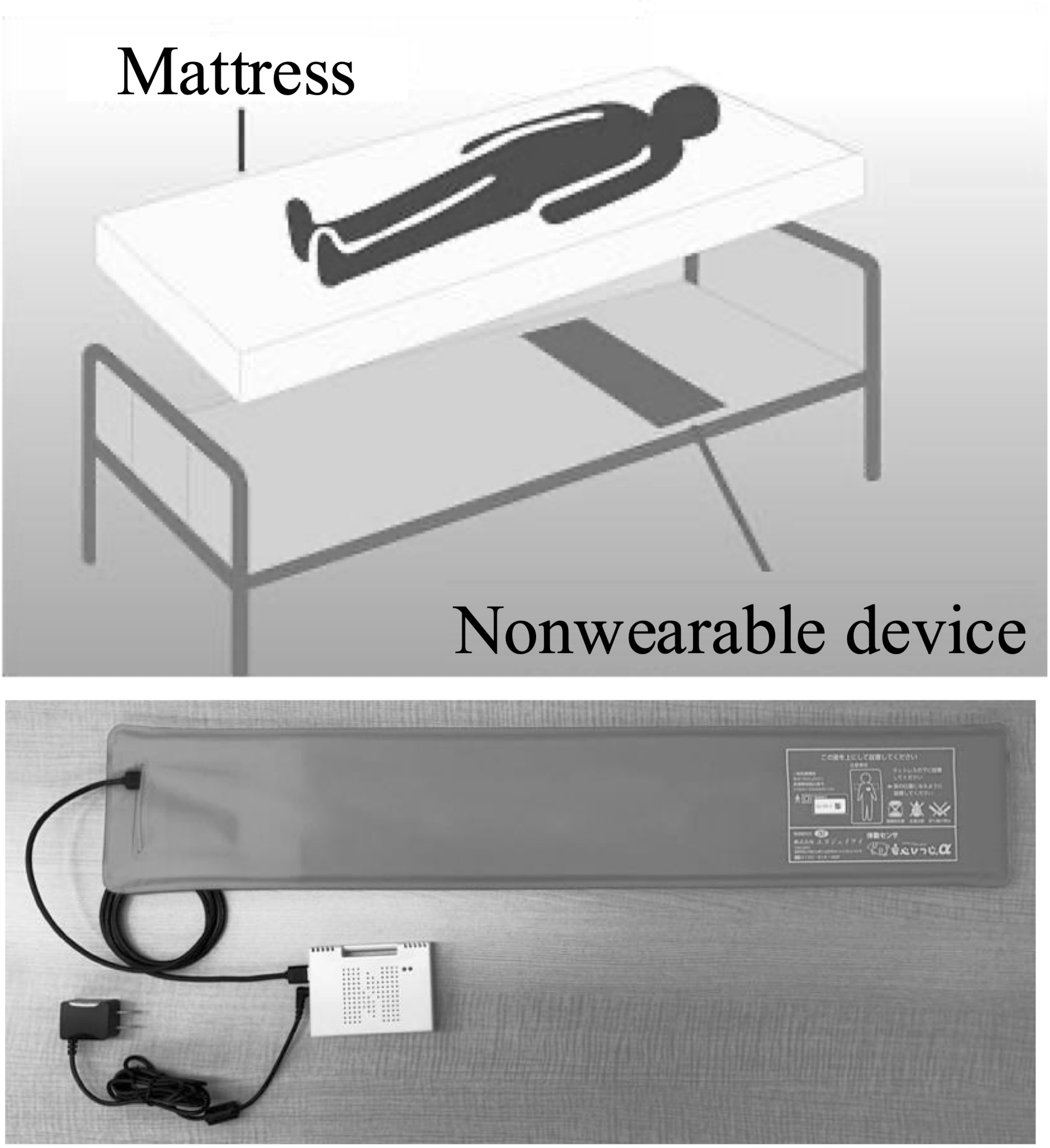


Figure 2. Nonwearable device system.

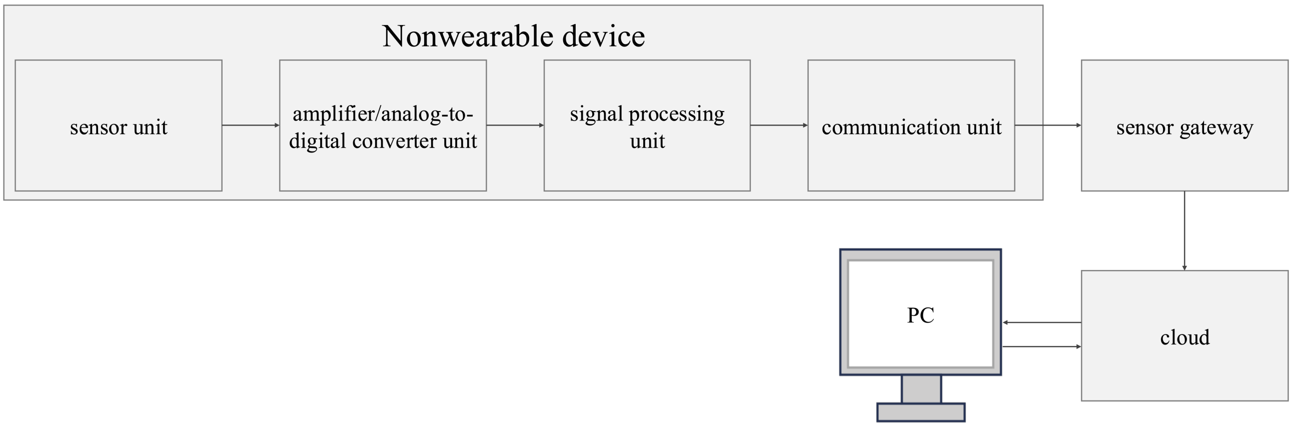
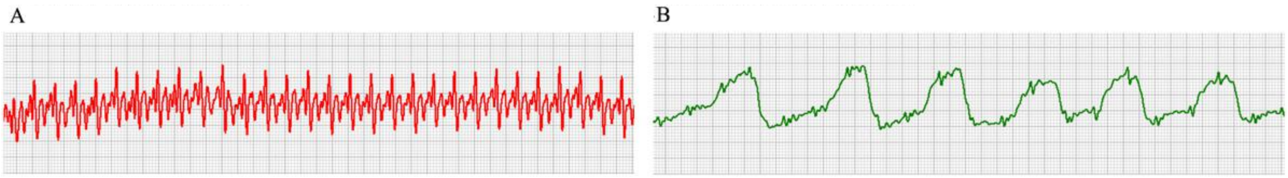


Figure 3. Heartbeat wave (A) and respiratory wave (B).



Heartbeats have relatively high-frequency components and occur at periodic intervals; after passing through a high-pass filter, characteristics derived from heartbeats are extracted, and HR is detected from the intervals between the extracted characteristics. The frequency components of respiration mainly have low-frequency components and occur at relatively regular intervals; after passing through the low-pass filter, the same process used to detect HR is used to detect RR. The frequency components of body movement are mainly low-frequency components, like respiration, but often do not repeat cyclically, and the amplitude is much larger than that of respiration. Very large signals are generated simultaneously on several piezoelectric elements, distinguishing them from respiration. The determined HR and RR are sent via the communication unit and Raspberry Pi to the cloud. Care workers at the nursing homes can access the cloud server using a web application on their personal computers and view the data with a browser (Figure 2).

Measurement of Vital Signs by Nurses

In this study, the standard method used to measure HR and RR was performed by nurses as previously described [9,10,26]. Briefly, 2 nurses with 20 years of experience measured the HR and RR of all participants. The nurse measured the HR by touching the participant's radial artery for 1 minute, and the RR by visually observing the thoracic movement for 1 minute. The nurse avoided touching the mattress while taking the measurements to eliminate interference.

Electrocardiograms and respiratory effort belts are other standard methods for measuring HR and RR; however, they are difficult to use in nursing homes staffed with only a few medical professionals [27,28]. Moreover, these contact devices have many electrodes and are not recommended for use in nonhospital settings as they cause physical restraint and discomfort in older adults [28]. Therefore, in this study, the standard method was used to ensure the safety of older adults and avoid causing discomfort.

Wearable Device

A Silmee Bar-type Lite sensor (TDK Co., Tokyo, Japan; width 64 mm; height 96 mm; depth 28 mm) was used as the WD to measure the HR of participants. This device could simultaneously measure electrocardiogram signals, pulse wave, acceleration, and skin temperature and was set to 125 Hz for the electrocardiograph and offline mode. This device was attached to the participants' chests at 3 cm below the collarbone using a special gel pad, and their electrocardiogram signals were measured. The HR was calculated per minute based on the time between heartbeats, which is the device output from an electrocardiogram.

HR cannot be measured visually but is preferably measured at the location of the heart. Therefore, we measured the HR both by using the WD and by manually palpating the participant's radial artery as mentioned in the previous section.

Collection of Basic Information

In study 1, the following items representing basic participant information were self-reported: sex, age, height, weight, respiratory diseases, and heart diseases. In study 2, the following data were collected from care workers in multidwelling houses: sex, age, height, weight, level of care needed, and diagnosis of dementia. Older adults who required nursing care were classified based on their care needs as levels 1‐5. Each municipality has certified care needs according to the level of nursing care required [29,30]. Care need level 1 requires relatively low assistance, while care need level 5 requires extensive assistance with personal care.

Data Analysis

The primary outcome was the mean difference between the HR and RR measured using the NWD and those measured by the nurses and the WD. We defined HR<5 beats/min and RR<3 breaths/min as the accepted mean difference before starting the study. We also defined the maximum error not leading to a change in care or observation as the criterion based on previous studies [9,31].

HR and RR data from the NWD were each compared with measurements taken by the nurses and WD at the same time point. The measured data were compared at each time point, and measurement errors were calculated. Subsequently, Bland-Altman plots were created to assess the agreement between the data measured using the NWD and those measured using the WD [32,33]. This method has been widely used in comparative studies [32,34]. We calculated the mean difference and 95% limits of agreement (LoAs) between the NWD data and manually measured data. Python (version 3.8.1; Python Software Foundation) and R statistical software (version 3.6.3; R Foundation for Statistical Research) were used for analysis. All statistical tests were 2-sided, and statistical significance was set at P<.05.

Ethical Considerations

This study was conducted in accordance with the principles of the Declaration of Helsinki and approved by the Human Study Ethics Committee of the authors' affiliated university (approval number: M2021-374).

The participants were informed about the study aims and procedure, both verbally and in writing, and consented to participate by signing a consent form. For older adults with difficulty communicating due to dementia or other cognitive impairments, their relatives were contacted, and their proxy consent was obtained. All participants (and their family members, when needed) were informed that participation in the study was voluntary and that they could withdraw at any point. Moreover, we explained that the participants would not be disadvantaged by not participating in the study, discontinuing participation, or withdrawing their cooperation. To protect privacy and confidentiality, the data collection nurse assigned each participant an identification number, and the list linking names and identification numbers was kept by the nurse. The researchers did not have access to this list, ensuring that the data remained anonymized.

HR measurement using the WD required direct contact with the skin; therefore, we attempted to protect the skin by applying a protectant (PureBarrer; Granmate Co., Chiba, Japan) before attaching the device. After the measurements, a release agent (3M Cavilon Remover; 3M, Maplewood, MN) was used for careful removal. No compensation was provided to participants for their involvement in this study.

Results

Demographics

A total of 22 healthy individuals participated in study 1 (Figure 4). The study population included 11 (50%) male participants and 11 (50%) female participants with a mean age of 47 (IQR 32.5 - 57) years (Table 1).

Figure 4. Flow diagram depicting the inclusion and exclusion of study participants.

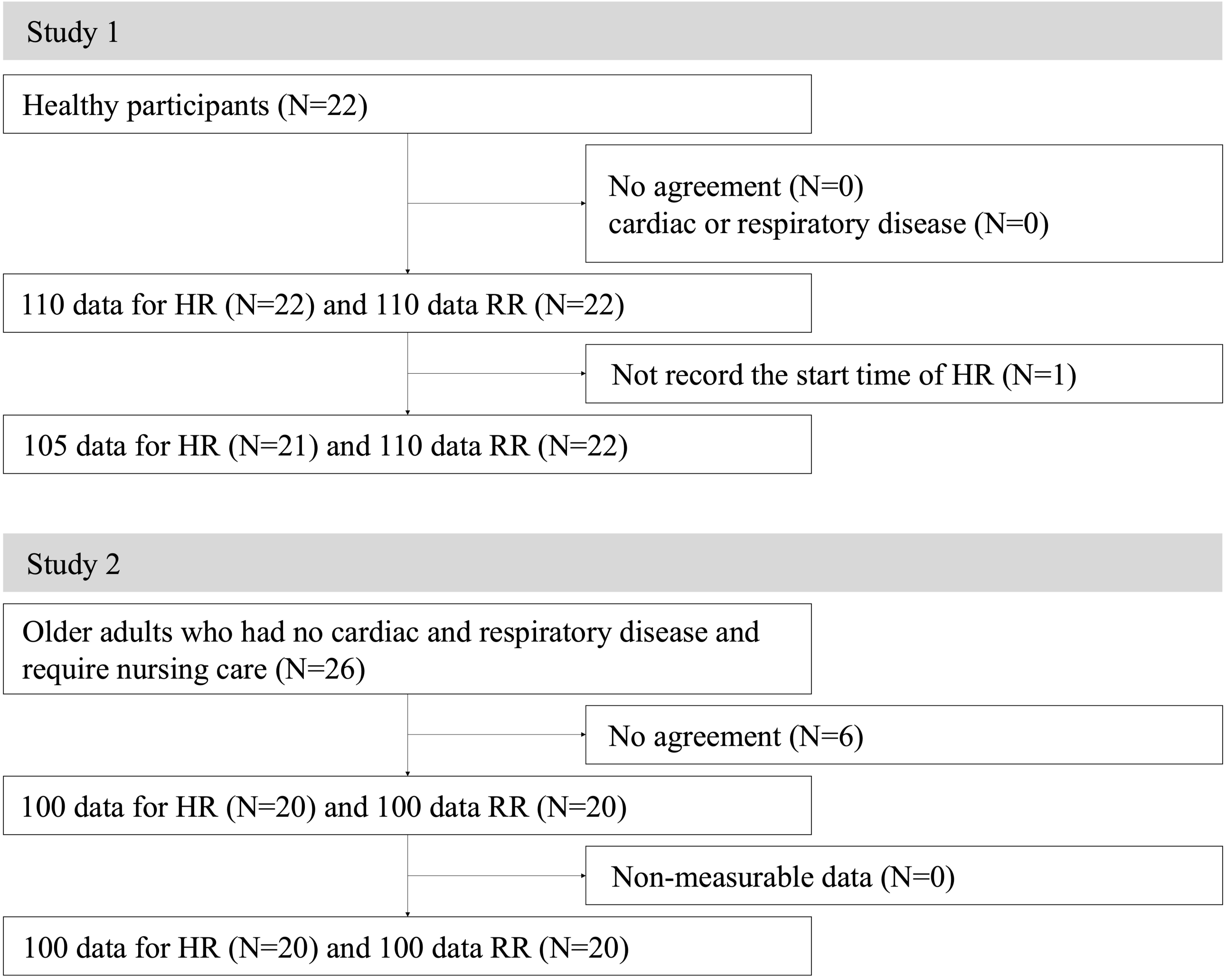


Table . Participants' characteristics for studies 1 and 2. Data are presented as median (IQR) or n (%).

	Value
Study 1 (n=21)	
Male sex, n (%)	11 (50)
Age (years), median (IQR)	47 (32.5 - 57.0)
Body mass index (BMI; kg/m^2), median (IQR)	22.2 (20.6 - 23.9)
Study 2 (n=20)	
Male sex, n (%)	3 (15)
Age (years), median (IQR)	87.5 (85.0 - 90.5)
BMI (kg/m^2), median (IQR)	22.1 (20.1 - 23.6)
Respiratory disease, n (%)	0 (0)
Care need level 1, n (%)	7 (35)
Care need level 2, n (%)	6 (30)
Care need level 3, n (%)	6 (30)
Care need level 4, n (%)	0 (0)
Care need level 5, n (%)	1 (5)
Dementia, n (%)	12 (60)

In study 2, 26 older adults requiring nursing care were selected, of whom 20 participated (Figure 4). Three older adults did not consent, and 3 could not obtain consent from their families. The study population included 3 (15%) male participants and 17 (85%) female participants with a mean age of 87.5 (IQR 85 - 90.5) years (Table 1). Among the participants, 7 (35%) needed care level 1, 6 (30%) needed care level 2, 6 (30%) needed care level 3, and 1 (5%) needed care level 5. None of the participants dropped out during the study periods in studies 1 or 2.

Technical Feasibility

In study 1, 110 HR data points (22 participants) and 110 RR data points (22 participants) were measured. The start time of HR measurement was not recorded for 1 participant owing to human error. Therefore, the NWD, WD, and nurse measurements for this participant could not be merged. We excluded the HR data of the participant (5 data points) and included 105 data points for HR (21 participants) and 110 data points for RR (22 participants).

In study 2, 100 data points each for HR (20 participants) and RR (20 participants) were measured. There were no missing data; therefore, 100 data points each were included for the HR (20 participants) and RR (20 participants).

Heart Rate

The mean differences measured by the NWD, nurses, and WD are shown in Table 2. The mean HR in study 1 measured by the NWD was 66.57 (SD 8.45) beats/min, the nurse measured 66.25 (SD 8.05) beats/min, and the WD recorded 66.57 (SD 8.45) beats/min. The mean HRs in study 2 measured by the NWD, nurse, and WD were 64.22 (SD 6.13), 64.18 (SD 7.20), and 65.21 (SD 7.84) beats/min, respectively.

The Bland-Altman plots are shown in Figure 5 and Table 3. The mean, differences, and LoAs (1.96 SD) were plotted. The y-axis in Figure 5 indicates the measurement differences and LoAs. First, the differences between NWD and nurses for 95% were within the LoAs; however, wide LoAs were observed (study 1 [lower LoA to upper LoA]: −6.86 to 4.90; study 2: −7.72 to 7.80). Second, the differences between NWD and WD for 95% were within the LoAs; however, wide LoAs were observed (study 1: −3.42 to 4.02; study 2: −7.20 to 9.26).

The differences in the HR measurements are presented in Table 4. Approximately 90% of the measurement differences were within 5 measurements in studies 1 and 2.

Table . Means and differences of measurements taken using the NWD, nurse, and WD.

		Measurements, n	NWD[a], mean (SD)	Nurse, mean (SD)	WD[b], mean (SD)	NWD vs nurse, mean difference (SD)	NWD vs WD, mean difference (SD)
Study 1							
	HR[c] (beats/min)	105	68.05 (6.9)	66.25 (8.05)	66.9 (8.28)	−0.32 (3.12)	0.33 (1.86)
	RR[d] (beats/min)	110	13.27 (7.9)	12.55 (3.27)	—[e]	−0.98 (3.01)	—
Study 2							
	HR (beats/min)	100	66.13 (5.21)	64.22 (6.13)	65.21 (7.84)	0.04 (3.98)	1.03 (4.22)
	RR (beats/min)	100	23.22 (1.55)	15.65 (3.22)	—	−0.49 (2.4)	—

[a]NWD: nonwearable device.
[b]WD: wearable device.
[c]HR: heart rate.
[d]RR: respiratory rate.
[e]Not applicable.

Figure 5. Bland-Altman plot. HR: heart rate; NWD: nonwearable device; RR: respiratory rate; WD: wearable device.

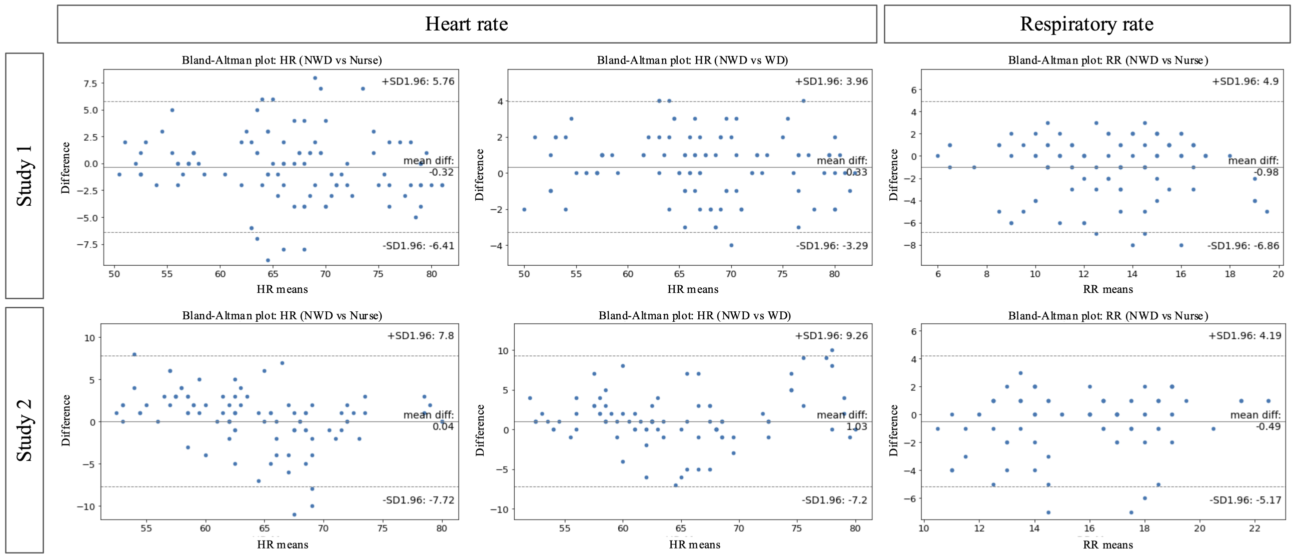


Table . Bland-Altman analysis for HR and RR measured using the NWD.

		Measurements, n	NWD[a] vs nurses			NWD vs WD[b]		
			Mean (SD)	Lower LoA[c]	Upper LoA	Mean (SD)	Lower LoA	Upper LoA
Study 1								
	HR[d] (beats/min)	105	−0.32 (3.12)	−6.86	4.90	0.33 (1.86)	−3.42	4.02
	RR[e] (beats/min)	110	−0.98 (3.01)	−6.41	5.76	—[f]	—	—
Study 2								
	HR (beats/min)	100	0.04 (3.98)	−7.72	7.80	1.03 (4.22)	−7.20	9.26
	RR (beats/min)	100	−0.49 (2.40)	−5.17	4.19	—	—	—

[a]NWD: nonwearable device.
[b]WD: wearable device.
[c]LoA: limit of agreement.
[d]HR: heart rate.
[e]RR: respiratory rate.
[f]Not applicable.

Table . Measurement difference for NWDs in studies 1 and 2.

		Study 1 measurements (HR[a]: n=105; RR[b]: n=110), n (%)		Study 2 measurements (HR and RR: n=100), n (%)	
		NWD[c] vs nurse	NWD vs WD[d]	NWD vs nurse	NWD vs WD
HR (beats/min)					
	≤2	71 (67.62)	88 (83.81)	62 (62)	62 (62)
	>2	23 (21.9)	16 (15.24)	25 (25)	22 (22)
	>5	11 (10.48)	1 (0.95)	9 (9)	12 (12)
	>10	0 (0)	0 (0)	4 (4)	4 (4)
	>15	0 (0)	0 (0)	0 (0)	0 (0)
RR (breaths/min)					
	≤2	81 (73.64)	—[e]	84 (84)	—
	>2	19 (17.27)	—	11 (11)	—
	>5	9 (8.18)	—	5 (5)	—
	>10	0 (0)	—	0 (0)	—
	>15	1 (0.91)	—	0 (0)	—

[a]HR: heart rate.
[b]RR: respiratory rate.
[c]NWD: nonwearable device.
[d]WD: wearable device.
[e]Not applicable.

Respiratory Rate

The mean differences measured by NWD, nurses, and WD are shown in Table 2. The mean RR in study 1 measured by the NWD was 13.54 (SD 3.31) breaths/min, and by the nurse was 12.55 (SD 3.27) breaths/min. The mean RR in study 2 measured by NWD was 16.14 (SD 2.89) breaths/min and by the nurse was 15.65 (SD 3.22) breaths/min.

The Bland-Altman plots are shown in Figure 5 and Table 3. The mean differences and LoAs (SD 1.96) were plotted. The y-axis in Figure 5 indicates the measurement differences and LoAs. The differences between the data obtained from the NWD and nurses for 95% were within the LoAs. However, wide LoAs were observed (study 1 [lower LoA to upper LoA]: −6.41 to 5.76; study 2: −5.17 to 4.19).

The differences in RR measurements are presented in Table 4. Approximately 90% of the measurement differences were within 5 measurements in studies 1 and 2.

Discussion

This study evaluated the feasibility of monitoring HR and RR continuously using an NWD placed under the participant's mattress and measured the consistency of the rates. The consistency of the HR and RR measured was proved by the finding that the mean differences in HR and RR calculated using the NWD in healthy participants and older adults who required nursing care were within the predetermined and accepted cutoffs. Moreover, none of the participants dropped out of the study or complained of physical abnormalities. Therefore, continuous monitoring of vital signs using NWDs is feasible at residences and in nursing homes.

First, no participants dropped out or complained in this study. Although a WD can provide accurate measurements, skin redness and itching have been reported due to skin contact with the device [9]; in contrast, our study participants did not report such skin-related issues. Moreover, if participants are aware that they are being monitored, their RR may tend to be lower than normal [35]. NWDs provide a nonintrusive alternative that allows for continuous monitoring in bed without skin contact or irritation [36]. In this study, an NWD yielded HR and RR values that were close to normal values because the patients were not uncomfortable and were less likely to notice that they were being monitored.

Second, we showed the correlation of HR measured using the NWD. The mean difference in HR measured using the NWD was within the predetermined acceptable range (study 1: −0.32 and 0.33 [nurse and WD, respectively] and study 2: 0.04 and 1.03 [nurse and WD, respectively]). In addition, the HRs measured with the NWD and WD were similar in this study. The mean differences in HR measured using other WDs are reportedly −0.20 (SD 5.54) [9], −1.1 (SD 3.8) [31], and 1.8 (SD 1.8) [37], close to those observed in this study. The device used in these previous studies was a wireless WD with electrodes attached to the anterior chest [9,31]. Although there are differences in measurement methods (ie, direct vs indirect contact with the skin), measurement positions (anterior chest vs posterior back), and device systems (electrical signals of the heart vs waves from body vibrations), the consistency in HR measurements suggests that both NWDs and WDs achieve similar accuracy. WDs can monitor an older adult's activity continuously without much discomfort [36]. Both NWDs and WDs are noninvasive and can be easily integrated into the daily lives of older adults [38]. Therefore, it is important to select the appropriate device based on the specific needs and condition of the patient.

Third, we showed the correlation of RR measured by the NWD. The mean difference in the RR measured using the NWD was within the predetermined acceptable range (study 1: −0.98; study 2: −0.49). The mean differences in RR measured using other WDs are −2.3 (SD 6.8) [23] and 1.19 (SD 3.43) [9], similar to that in this study. We observed that the NWD could measure RR almost as well as the nurses, with a low measurement error (<5 breaths/min). However, a large difference of 17 breaths/min was seen in the measurement error (5 and 22 breaths/min for the nurse and NWD, respectively), consistent with a previous study by Weenk et al [9], who reported a difference of 26 breaths/min. In the present and previous studies, nurses assessed RR by visually observing the chest [9]. The reproducibility of the method is limited by high interobserver variability [39], and a large difference was expected in this study. In addition, the number of events related to the reliability of bradycardic and tachycardic respiration was low in this study. Therefore, their reliability could not be evaluated. The reasons for this include the short measurement time per participant, exclusion of participants with cardiac or respiratory disease, and starting the measurement after the participants had rested for 5 minutes. In the future, it is necessary to conduct long-term measurements, such as overnight measurements, to evaluate events related to the reliability of bradycardic and tachycardic respiration in older adults.

The strength of our study is that we evaluated the consistency of HR and RR measured using an NWD that has already been introduced in many facilities in Japan, and suggested the possibility of improving the nursing home environment. We believe that monitoring the HR and RR of older adults safely and unobtrusively when they are in bed may help in detecting sudden changes and providing suitable care without frequent visits. NWDs reduce the number of vital sign measurements and rounds and the burden on care providers, nurses, and care workers, allowing them to concentrate on care [40]. Given the declining population and limited number of health care workers, devices with consistent HR and RR will help optimize the patient's environment. Therefore, we believe that NWDs can be applied in various facilities.

This study had some limitations. First, the participants were healthy or older adults without cardiac or respiratory disease. Therefore, the consistency in participants with cardiac and respiratory diseases should be examined further. Second, the measurement times for HR and RR in this study were short (5 minutes). When the NWD was installed at the residences and nursing homes, participants were continuously monitored while they were in bed. The disadvantages of continuous measurements, such as those taken overnight, were not considered. Third, the criterion measurement is lacking. In this study, nurse-based measurement was selected as the criterion for HR and RR as described in many studies [9,18,26]. Contact devices, such as electrocardiography, have also been used [8,9]. However, they are known to cause skin issues, such as peeling of the stratum corneum and red spots [41,42]. In this study, older adults (whose skin is typically more fragile) were recruited [13]. Therefore, we decided to use nurse-based measurement to prevent any discomfort and adverse events among the study participants. Ideally, HR and RR should have been measured using both methods. Despite these limitations, the consistency of HR and RR measured using the NWD suggests that this method may be useful for monitoring vital signs at residences and in nursing homes.

This study evaluated the feasibility and consistency of measuring HR and RR using an NWD placed under the mattress. The mean differences in HR and RR measured by the NWD were both within the predefined accepted discrepancies. No physical abnormalities were noted during the measurements. Therefore, we suggest using these devices at residences and in nursing homes. Safe monitoring of vital signs using NWDs is expected in the future for the early detection of abnormal conditions without inconveniencing older adults, care workers, and nurses. We believe that NWDs promote medical digital transformation, which will enable care providers to observe conditions accurately and provide appropriate care through the data obtained from such devices.

Acknowledgments

We would like to thank all the participants in this study. We would also like to thank Ms Kazumi Sato, Ms Kaori Sato, and other staff of NJI Co., Ltd. for their cooperation. This study was supported by the Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research (grant 21H04849 to SF) and NJI Corporation (grant 21AB100084 to SF).

Data Availability

Data from this study are available by contacting our research team.

Authors' Contributions

KI, AM, MNW, and SF contributed to the research idea, study design, and organization of the study. KI contributed to data management and statistical analysis. KI, AM, MNW, and SF contributed to the interpretation. KI, AM, MNW, and SF contributed to data acquisition. Each author has reviewed, discussed, and agreed to their individual contributions ahead of submission.

Conflicts of Interest

SF and MNW received grant support from NJI Co., Ltd., which developed the Safety Sheep Sensor.

References

1. Borský P, Holmannová D, Fiala Z, Borská L, Hruška L, Kučera O. Physiology of ageing. Cas Lek Cesk 2022;161(1):11-16. [Medline: 35354289]
2. Flint B, Tadi P. Physiology, Aging: StatPearls; 2023. [Medline: 32310566]
3. Celler BG, Sparks RS. Home telemonitoring of vital signs--technical challenges and future directions. IEEE J Biomed Health Inform 2015 Jan;19(1):82-91. [doi: 10.1109/JBHI.2014.2351413] [Medline: 25163076]
4. Ashcraft AS, Owen DC. From nursing home to acute care: signs, symptoms, and strategies used to prevent transfer. Geriatr Nurs (Minneap) 2014 Jul;35(4):316-320. [doi: 10.1016/j.gerinurse.2014.06.007]
5. Harris DA, Archbald-Pannone L, Kaur J, et al. Rapid telehealth-centered response to COVID-19 outbreaks in postacute and long-term care facilities. Telemed J E Health 2021 Jan;27(1):102-106. [doi: 10.1089/tmj.2020.0236] [Medline: 32644899]
6. Watkins T, Whisman L, Booker P. Nursing assessment of continuous vital sign surveillance to improve patient safety on the medical/surgical unit. J Clin Nurs 2016 Jan;25(1-2):278-281. [doi: 10.1111/jocn.13102] [Medline: 26531215]
7. Hollander JE, Carr BG. Virtually perfect? Telemedicine for COVID-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
8. Kroll RR, Boyd JG, Maslove DM. Accuracy of a wrist-worn wearable device for monitoring heart rates in hospital inpatients: a prospective observational study. J Med Internet Res 2016 Sep 20;18(9):e253. [doi: 10.2196/jmir.6025] [Medline: 27651304]
9. Weenk M, van Goor H, Frietman B, et al. Continuous monitoring of vital signs using wearable devices on the general ward: pilot study. JMIR Mhealth Uhealth 2017 Jul;5(7):e91. [doi: 10.2196/mhealth.7208]
10. Boatin AA, Wylie BJ, Goldfarb I, et al. Wireless vital sign monitoring in pregnant women: a functionality and acceptability study. Telemed J E Health 2016 Jul;22(7):564-571. [doi: 10.1089/tmj.2015.0173] [Medline: 27172448]
11. Wang S, Ni X, Li L, et al. Noninvasive monitoring of vital signs based on highly sensitive fiber optic mattress. IEEE Sensors J 2020 Jun;20(11):6182-6190. [doi: 10.1109/JSEN.2020.2974313]
12. Wang Z, Yang Z, Dong T. A review of wearable technologies for elderly care that can accurately track indoor position, recognize physical activities and monitor vital signs in real time. Sensors 2017 Feb;17(2):341. [doi: 10.3390/s17020341]
13. Russell-Goldman E, Murphy GF. The pathobiology of skin aging: new insights into an old dilemma. Am J Pathol 2020 Jul;190(7):1356-1369. [doi: 10.1016/j.ajpath.2020.03.007] [Medline: 32246919]
14. Ullal A, Su BY, Enayati M, et al. Non-invasive monitoring of vital signs for older adults using recliner chairs. Health Technol 2021 Jan;11(1):169-184. [doi: 10.1007/s12553-020-00503-9]
15. de Tommasi F, Presti DL, Carassiti M, Schena E, Massaroni C. Smart mattress based on fiber bragg grating sensors for respiratory monitoring: a feasibility test. Presented at: 2021 IEEE International Workshop on Metrology for Industry 4.0 & IoT (MetroInd4.0&IoT); Jun 7 to 9, 2021:; Rome, Italy p. 532-537. [doi: 10.1109/MetroInd4.0IoT51437.2021.9488539]
16. Prata D, Carvalho A, Costa FM, Marques C, Leitao C. Unobtrusive monitoring of the respiratory rate in an office desk chair with FBG sensors. 2021 Presented at: 2021 IEEE International Workshop on Metrology for Industry 4.0 & IoT (MetroInd4.0&IoT); Rome, Italy p. 177-181. [doi: 10.1109/MetroInd4.0IoT51437.2021.9488528]
17. Rohan R, Venkadeshwaran K, Ranjan P. Recent advancements of fiber bragg grating sensors in biomedical application: a review. J Opt 2024 Feb;53(1):282-293. [doi: 10.1007/s12596-023-01134-9]
18. World population ageing. United Nations. 2015. URL: https://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2015_Report.pdf [accessed 2024-10-02]
19. Dai 8 ki kaigo hoken zigyou keikaku ni motoduku kaigosyokuinn no hituyousuu nituite. Ministry of Health, Labour and Welfare. 2021. URL: https://www.mhlw.go.jp/stf/houdou/0000207323_00005.html [accessed 2024-09-30]
20. Kaigo robot no kaihatu fukyu no 488 sokushin. Ministry of Health, Labour and Welfare. 2021. URL: https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000209634.html [accessed 2024-09-30]
21. Kaigo robot to donyu sien 491 tokubetu zigyou (Heisei 27 nendo hosei yosan). Ministry of Health, Labour and Welfare. 2015. URL: https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000112870.html [accessed 2024-09-30]
22. Kaigo robbot no kaihatsu 494 hukyuu no sokushin. Ministry of Health, Labour and Welfare. 2020. URL: https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000209634.html [accessed 2024-09-30]
23. Morley JE, Caplan G, Cesari M, et al. International survey of nursing home research priorities. J Am Med Dir Assoc 2014 May;15(5):309-312. [doi: 10.1016/j.jamda.2014.03.003] [Medline: 24703926]
24. Sim I. Mobile devices and health. N Engl J Med 2019 Sep 5;381(10):956-968. [doi: 10.1056/NEJMra1806949] [Medline: 31483966]
25. A sensor to protect your loved ones. Safety Sheep Sensor. URL: https://www.anshin-hitsuji.jp/Templates/PDF/Pamphlet-English.pdf [accessed 2024-09-30]
26. Steinhubl SR, Feye D, Levine AC, Conkright C, Wegerich SW, Conkright G. Validation of a portable, deployable system for continuous vital sign monitoring using a multiparametric wearable sensor and personalised analytics in an Ebola treatment centre. BMJ Glob Health 2016 Jul;1(1):e000070. [doi: 10.1136/bmjgh-2016-000070]

27. Patel V, Orchanian-Cheff A, Wu R. Evaluating the validity and utility of wearable technology for continuously monitoring patients in a hospital setting: systematic review. JMIR Mhealth Uhealth 2021 Aug;9(8):e17411. [doi: 10.2196/17411]
28. Sun G, Matsui T, Watai Y, et al. Vital-SCOPE: design and evaluation of a smart vital sign monitor for simultaneous measurement of pulse rate, respiratory rate, and body temperature for patient monitoring. J Sens 2018 Feb 1:1-7. [doi: 10.1155/2018/4371872]
29. Takahashi S, Yonekura Y, Takanashi N, Tanno K. Risk factors of long-term care insurance certification in Japan: a scoping review. Int J Environ Res Public Health 2022 Feb 14;19(4):2162. [doi: 10.3390/ijerph19042162] [Medline: 35206350]
30. Ikegami N. Financing long-term care: lessons from Japan. Int J Health Policy Manag 2019 Aug 1;8(8):462-466. [doi: 10.15171/ijhpm.2019.35] [Medline: 31441285]
31. Breteler M, Huizinga E, van Loon K, et al. Reliability of wireless monitoring using a wearable patch sensor in high-risk surgical patients at a step-down unit in the Netherlands: a clinical validation study. BMJ Open 2018 Feb;8(2):e020162. [doi: 10.1136/bmjopen-2017-020162] [Medline: 29487076]
32. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb;1(8476):307-310. [doi: 10.1016/j.ijnurstu] [Medline: 2868172]
33. Bland JM, Altman DG. Applying the right statistics: analyses of measurement studies. Ultrasound Obstet Gyne 2003 Jul;22(1):85-93. [doi: 10.1002/uog.122]
34. Gerke O. Reporting standards for a Bland-Altman agreement analysis: a review of methodological reviews. Diagnostics (Basel) 2020 May 22;10(5):334. [doi: 10.3390/diagnostics10050334] [Medline: 32456091]
35. Hill A, Kelly E, Horswill MS, Watson MO. The effects of awareness and count duration on adult respiratory rate measurements: an experimental study. J Clin Nurs 2018 Feb;27(3-4):546-554. [doi: 10.1111/jocn.13861] [Medline: 28426897]
36. Arandia N, Garate JI, Mabe J. Monitoring of vital signs in the home environment: a review of current technologies and solutions. 2023 Presented at: Proceedings of the 16th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 1: BIOSTEC; Feb 16-18, 2023; Lisbon, Portugal p. 108-115. [doi: 10.5220/0011646700003414]
37. Hahnen C, Freeman CG, Haldar N, et al. Accuracy of vital signs measurements by a smartwatch and a portable health device: validation study. JMIR Mhealth Uhealth 2020 Feb;8(2):e16811. [doi: 10.2196/16811] [Medline: 32049066]
38. Olmedo-Aguirre JO, Reyes-Campos J, Alor-Hernández G, Machorro-Cano I, Rodríguez-Mazahua L, Sánchez-Cervantes JL. Remote healthcare for elderly people using wearables: a review. Biosensors (Basel) 2022 Jan;12(2):73. [doi: 10.3390/bios12020073] [Medline: 35200334]
39. Edmonds ZV, Mower WR, Lovato LM, Lomeli R. The reliability of vital sign measurements. Ann Emerg Med 2002 Mar;39(3):233-237. [doi: 10.1067/mem.2002.122017] [Medline: 11867974]
40. Fulmer T, Reuben DB, Auerbach J, Fick DM, Galambos C, Johnson KS. Actualizing better health and health care for older adults. Health Aff (Millwood) 2021 Feb;40(2):219-225. [doi: 10.1377/hlthaff.2020.01470] [Medline: 33476185]
41. Feig DS, Donovan LE, Corcoy R, et al. Continuous glucose monitoring in pregnant women with type 1 diabetes (CONCEPTT): a multicentre international randomised controlled trial. Lancet 2017 Nov;390:2347-2359. [doi: 10.1016/S0140-6736(17)32400-5] [Medline: 28923465]
42. Ling Y, An T, Yap LW, Zhu B, Gong S, Cheng W. Disruptive, soft, wearable sensors. Adv Mater 2020 May;32. [doi: 10.1002/adma.201904664] [Medline: 31721340]

Abbreviations

HR: heart rate
LoA: limit of agreement
NWD: nonwearable device
RR: respiratory rate
WD: wearable device

Edited by P Santana-Mancilla; submitted 20.01.24; peer-reviewed by A Hassan, J Ramirez, S Kepak; revised version received 25.06.24; accepted 02.07.24; published 08.10.24.

Please cite as:
Ikuta K, Aishima M, Noguchi-Watanabe M, Fukui S
Feasibility of Monitoring Heart and Respiratory Rates Using Nonwearable Devices and Consistency of the Measured Parameters: Pilot Feasibility Study
JMIR Hum Factors 2024;11:e56547
URL: https://humanfactors.jmir.org/2024/1/e56547
doi:10.2196/56547

The Impact of Information Relevancy and Interactivity on Intensivists' Trust in a Machine Learning–Based Bacteremia Prediction System: Simulation Study

Omer Katzburg[1], MSc; Michael Roimi[2], MD; Amit Frenkel[3], MD; Roy Ilan[2], MD; Yuval Bitan[1], PhD

[1]Department of Health Policy and Management, Ben-Gurion University of the Negev, Be'er Sheva, Israel
[2]General Intensive Care Unit, Rambam Medical Center, Haifa, Israel
[3]General Intensive Care Unit, Soroka Medical Center, Be'er Sheva, Israel

Corresponding Author:
Yuval Bitan, PhD
Department of Health Policy and Management, Ben-Gurion University of the Negev, , Be'er Sheva, , Israel

Abstract

Background: The exponential growth in computing power and the increasing digitization of information have substantially advanced the machine learning (ML) research field. However, ML algorithms are often considered "black boxes," and this fosters distrust. In medical domains, in which mistakes can result in fatal outcomes, practitioners may be especially reluctant to trust ML algorithms.

Objective: The aim of this study is to explore the effect of user-interface design features on intensivists' trust in an ML-based clinical decision support system.

Methods: A total of 47 physicians from critical care specialties were presented with 3 patient cases of bacteremia in the setting of an ML-based simulation system. Three conditions of the simulation were tested according to combinations of information relevancy and interactivity. Participants' trust in the system was assessed by their agreement with the system's prediction and a postexperiment questionnaire. Linear regression models were applied to measure the effects.

Results: Participants' agreement with the system's prediction did not differ according to the experimental conditions. However, in the postexperiment questionnaire, higher information relevancy ratings and interactivity ratings were associated with higher perceived trust in the system (*P*<.001 for both). The explicit visual presentation of the features of the ML algorithm on the user interface resulted in lower trust among the participants (*P*=.05).

Conclusions: Information relevancy and interactivity features should be considered in the design of the user interface of ML-based clinical decision support systems to enhance intensivists' trust. This study sheds light on the connection between information relevancy, interactivity, and trust in human-ML interaction, specifically in the intensive care unit environment.

(JMIR Hum Factors 2024;11:e56924) doi:10.2196/56924

KEYWORDS

user-interface design; user-interface designs; user interface; human-automation interaction; human-automation interactions; trust in automation; automation; human-computer interaction; human-computer interactions; human-ML; human-ML interaction; human-ML interactions; decision making; decision support system; clinical decision support; decision support; decision support systems; machine learning; ML; artificial intelligence; AI; machine learning algorithm; machine learning algorithms; digitization; digitization of information

Introduction

Overview

In the intensive care unit (ICU), intensivists make an extremely high number of decisions. For example, McKenzie et al [1] found that approximately 100 decisions are made every morning round. According to Ward et al [2], despite the continual increase in the number of ICUs, the number of intensivists remains about the same, resulting in an extremely high workload. The high rate of decision-making together with the continuous overload prompts the need for decision support tools.

Although machine learning (ML) algorithms and systems serving the medical community are continually increasing, their adoption into routine health care practice is not guaranteed [3]. One reason is the complexity of the algorithms, which often leads to clinicians' lack of trust in such systems [4]. A multidisciplinary approach may enhance trust, by considering the human factor, the technological aspect, and the interaction between them [5]. This study examined 2 human-automation interaction features that emphasize the importance of the human factor in the design of ML-based clinical decision support systems (CDSSs).

Clinical Decision Support Systems

To date, many CDSSs are categorized as "expert systems"—systems that try to imitate the way an ideal physician would think. These systems generate conclusions based on sets of rules [6]. In contrast, ML algorithms approach problems in the opposite way—they generate rules from historical data [6,7]. ML algorithms are currently being developed in almost every field of medicine and, in many instances, are already providing equal or even greater accuracy than physicians (eg, [8-10]). However, though ML CDSSs can enhance the quality of care, the adoption of such systems in all medical fields, and specifically in critical care, remains low [11].

In contrast to expert systems, ML algorithms are complex, and understanding and explaining the reasoning underlying them is often impossible [12]. Thus, ML algorithms are frequently considered black box algorithms. This fosters physicians' distrust and skepticism of ML systems [13] and has been suggested as a major cause of the low rates of adoption and acceptance of these systems within the medical community [14]. Wrong decisions made by intensivists can result in severe and even fatal outcomes. Thus, they may be reluctant to share their decision-making responsibilities with black box CDSSs that they do not understand [11].

Interpretable ML

As ML algorithms are developed to serve humans, human interaction with them must be considered. One approach to move from a "black box" to a "clear box" [15] lies in the growing field of interpretable ML [16-19]. Miller [20] offered an approach that combines artificial intelligence, social science, and human-computer interaction (HCI). He referred to "human–agent interaction" as the intersection of these 3 domains, including it as part of the interpretable ML field. Impressive work has been performed on interpretable ML in the HCI community (eg, [21-24]). Unfortunately, the ML community and the HCI community do not always work together [25]. This results in poor usability of many interpretable ML algorithms [20], yet opens an opportunity for HCI and interaction design researchers to seek means of enhancing trust in ML CDSSs [26].

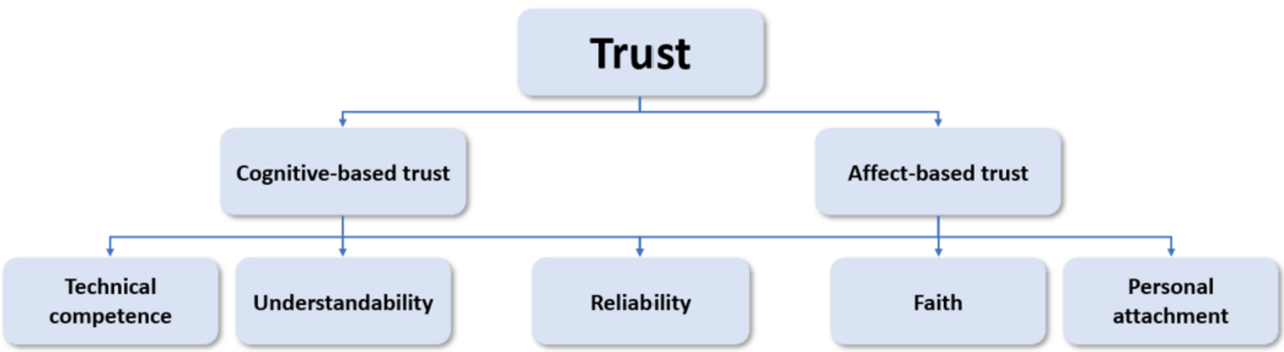
Human-Automation Trust

Parasuraman and Riley [27] defined automation as a technology that executes "a function that was previously carried out by a human." This wide definition covers all kinds of machines, computers, and applications of artificial intelligence. Human-automation trust is a well-studied subject (eg, [28-34]). In the context of human social interactions, trust can be defined as "the willingness to be vulnerable to the actions of another person" [35]. Research has shown that humans perceive computers as social actors and may interact with them as they would with each other [36-38]. The interaction between humans and automated systems, or, in the context of this study, intensivists and black box algorithms, has also been shown to be substantially influenced by trust [31].

Although human-automation trust is being researched by many disciplines, no dominant model or approach has been determined for its measure. However, a well-accepted conclusion is that trust is not a standalone construct, but rather multidimensional [32]. In this study, we used the definition of Lee and See [29] for human-automation trust "an attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability." This definition corresponds well with the interaction between intensivists and ML CDSSs, even though the ICU environment is characterized by high levels of both uncertainty and vulnerability.

According to Madsen and Gregor [30], human-computer trust is comprised of 2 main dimensions—cognition-based trust (CBT) and affect-based trust (ABT). CBT is based on the user's intellectual perceptions of the system's characteristics, while ABT is based on the user's emotional responses to the system. The 2 dimensions can be further subdivided. CBT is comprised of the understandability of the system and the technical competence of the system, whereas ABT is comprised of faith, personal attachment, and reliability. Madsen and Gregor [30] note that reliability was also found to influence CBT, although its influence on ABT is stronger. The researchers suggested a questionnaire for measuring trust, which we implemented in this study (Figure 1).

Figure 1. Madsen and Gregor's [30] human-computer trust model.



Aim

The primary aim of this study was to investigate the influence of elements of the user interface (UI) design on intensivists' trust in ML-based CDSSs ("black-box"–based algorithms). From the many UI elements that can be modified, the 2 that were chosen and compared are information relevancy and interface interactivity.

The literature is abundant regarding information relevancy, interactivity, and trust, as well as the influence of the 2 former factors on the latter. However, to the best of our knowledge, no research has assessed connections between information relevancy, interactivity, and trust in the context of human-ML interaction, specifically in the context of the ICU environment.

Hypothesis 1: Information Relevancy

Information relevancy concerns the degree to which users perceive that the information content of a system meets their needs [39]. This factor was found to positively influence user satisfaction with websites [39,40] and users' trust in health infomediaries [41]. Relevant information has been found to be an attribute that is more crucial for users than usability and convenient use of the system [42]. Considering the above, our hypothesis is as follows:

- *Higher levels of information relevancy will lead to higher levels of trust in the system.* [H1]

Hypothesis 2: Interface Interactivity

Interactivity can be defined in various ways. For this study, we used a common definition by Steuer [43]—"the extent to which users can participate in modifying the format and content of a mediated environment in real time." Interactivity is considered to strongly influence users' experiences during the interaction [44] and is key to the success of e-commerce websites [45-47]. Interactivity was found to increase users' trust in websites in general and specifically in e-commerce, mobile commerce [48,49], and brand loyalty [44]. Although most of the literature on interactivity has focused on e-commerce trust and intentions to use websites, we expected greater interface interactivity to positively influence the interaction between ML CDSSs and intensivists, and to enhance their trust. Considering the above, our hypothesis is as follows:

- *Higher levels of interface interactivity will lead to higher levels of trust in the system.* [H2]

Methods

Overview

To test the hypotheses, a laboratory experiment with 3 conditions was designed. This enabled testing the effects of information relevancy and interactivity on intensivists' trust in a simulated ML-based bacteremia prediction system. Bacteremia is a common phenomenon in ICUs, that clinicians need to identify and respond to [50]. Thus, a decision support system that assists clinicians in identifying this condition can serve as a good reference for generalizing and deriving implications for the UI design of many ML-based CDSSs. Each experimental condition was characterized by a different set of UI. The effects were measured with both a behavioral measure (the participants' decisions that were captured by the simulation software) and a postexperiment questionnaire that captured their perceived understanding of the system.

Participants

The participants were 47 physicians (female: n=14; male: n=33) from critical care specialties of 5 tertiary hospitals in Israel. They were recruited through a convenience sample of on-duty physicians and were free to withdraw from the study at any time. The experiment was conducted for 1 month, between the first and second COVID-19 lockdowns in Israel. All the participants were compensated with a gift card (US $15) and there were no exclusion criteria except for being a critical care physician.

Ethical Considerations

This research complied with the American Psychological Association Code of Ethics and was approved by the institutional review board at Ben-Gurion University of the Negev (21-12-19). Informed consent was obtained from each participant.

Experimental Design

To test the hypotheses, a 2 × 2 (relevant/nonrelevant×interactive/noninteractive) between-subjects fractional factorial experiment was designed. The experiment included 3 conditions (as shown in Table 1). The 15 - 16 participants were randomly assigned to 1 of the 3 conditions; the duration of their performance was not limited. A total of 3 clinical cases of patients who were hospitalized in an ICU with medical conditions implying bacteremia onset were extracted. The presentations of these cases were designed by 3 experienced intensivists to provide accurate context.

Table . The experimental conditions.

	Noninteractive	Interactive
Nonrelevant information	1	—[a]
Relevant information	2	3

[a]Not tested.

Apparatus and Stimuli

A total of 3 UIs that represent 3 medical conditions were designed using Axure RP software (version 9.1; Axure Software Solutions, Inc). The interfaces were imitating an ML bacteremia prediction system. The system, which at the time of the study was still in its development stage, provides prediction and a list of the main features that were significant for the prediction algorithm. The right section of all the interfaces presented similar time-series charts. The charts included trends over time for the 10 clinical measures that are most related to bacteremia prediction. The information that was presented in the left section was manipulated to match the 3 conditions. An example of an interface (condition 2) is shown in Figures 2-4.

Figure 2. The right section shows the time-series chart, and the left section shows the patient's current clinical measures. HR: heart rate; ICU: intensive care unit; MAP: mean arterial pressure; RR: respiratory rate; WBC: white blood cell count.

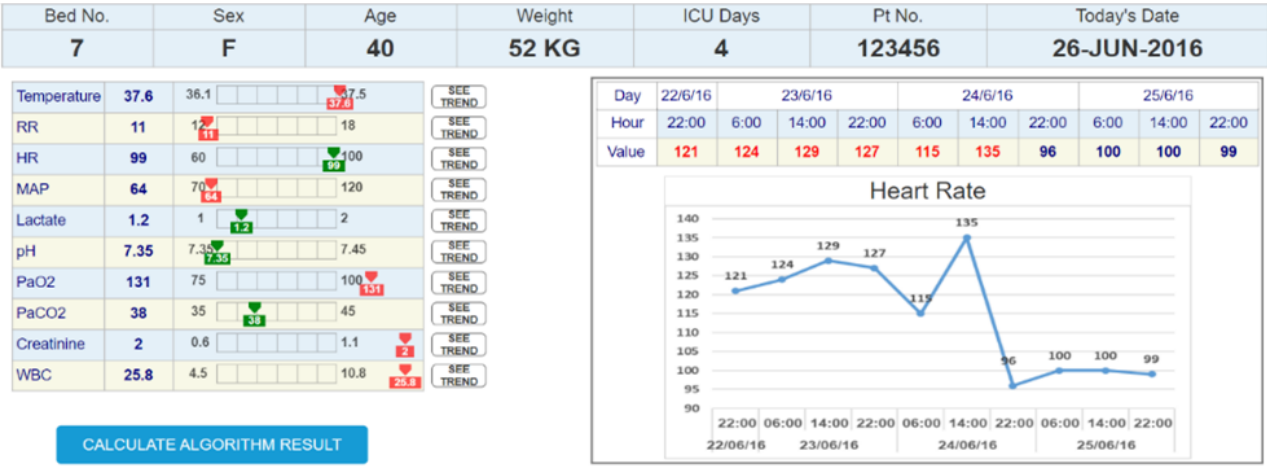


Figure 3. The bacteremia prediction system calculates the result. HR: heart rate; ICU: intensive care unit; MAP: mean arterial pressure; RR: respiratory rate; WBC: white blood cell count.

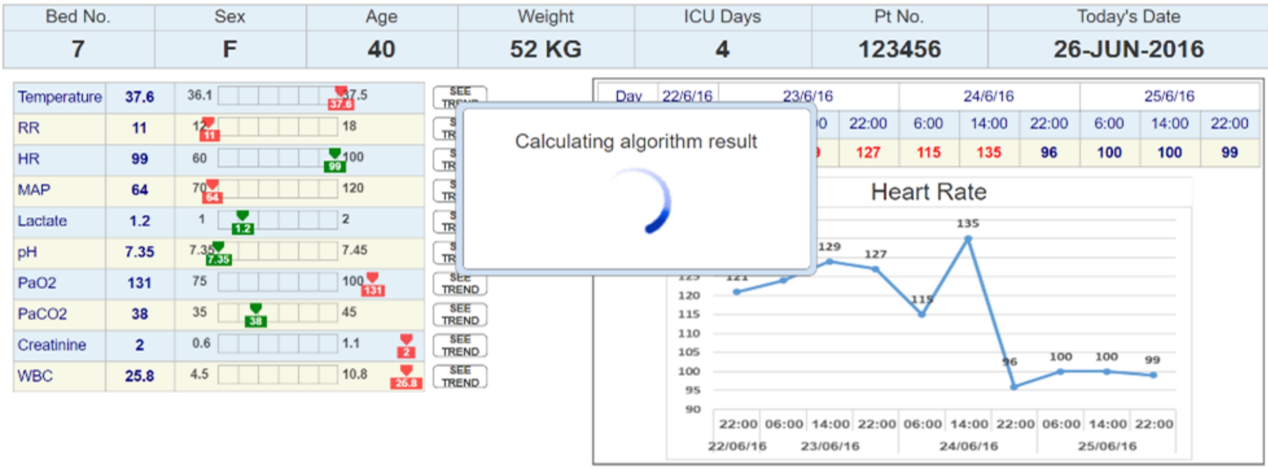
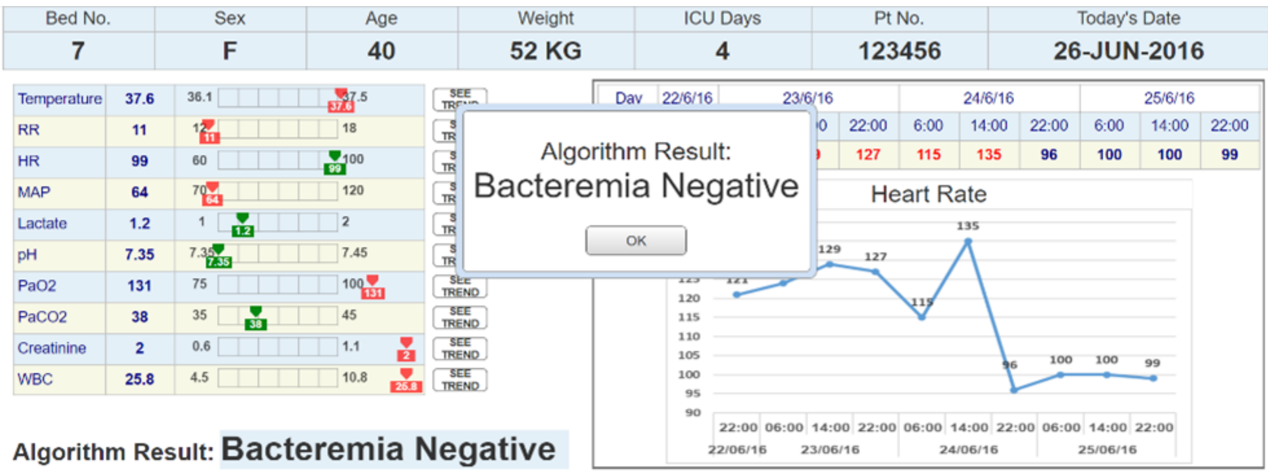


Figure 4. The bacteremia prediction system presents its prediction. HR: heart rate; ICU: intensive care unit; MAP: mean arterial pressure; RR: respiratory rate; WBC: white blood cell count.



The information relevancy level was set by the type of clinical measurements that were presented in a table in the left section of the chart. For the relevant information conditions, the information presented in the table comprised the current values of the same clinical measures that clinicians usually use to assess a patient’s condition. In addition, the normal range of each measure was presented. In the nonrelevant information condition, the information presented in the table comprised the values of the 10 features that were ranked as most important by the bacteremia ML prediction algorithm for making the prediction. Although these features were most significant for the prediction algorithm, they were not usually used by clinicians and, therefore, were considered nonrelevant (see Figure 5).

The interface interaction level was set by the type of interaction that the participants were assigned with the UI. In the interactive condition, the participants were required to enter values of the patient's current clinical measures (the values provided in the written clinical case) before they could explore the other charts and information. Entering and copying values to and from the patient record is a common task clinicians apply in a subset of the IT systems in the ICU. In the noninteractive conditions, the information about the patients appeared right away, and the participants could only explore the information and ask the system for its prediction (see Figure 6).

The fourth combination, nonrelevant information and interactivity, was not tested, as in the nonrelevant information condition, the information that was presented was of the features of the algorithm. Thus, including the algorithm features in the clinical case and entering them into the UI would seem unrealistic.

Figure 5. The relevant and nonrelevant conditions. The top frame shows the relevant information condition with the patient's current clinical measures; the bottom frame shows the nonrelevant information condition with the values of the algorithm's most important features. ArtBPD: arterial line blood pressure; HR: heart rate; ICU: intensive care unit; MAP: mean arterial pressure; PEEP: positive end-expiratory pressure: P/F ratio: PaO_2/FIO_2: oxygen arterial pressure to percentage of inspired oxygen ratio; RR: respiratory rate; WBC: white blood cell count.

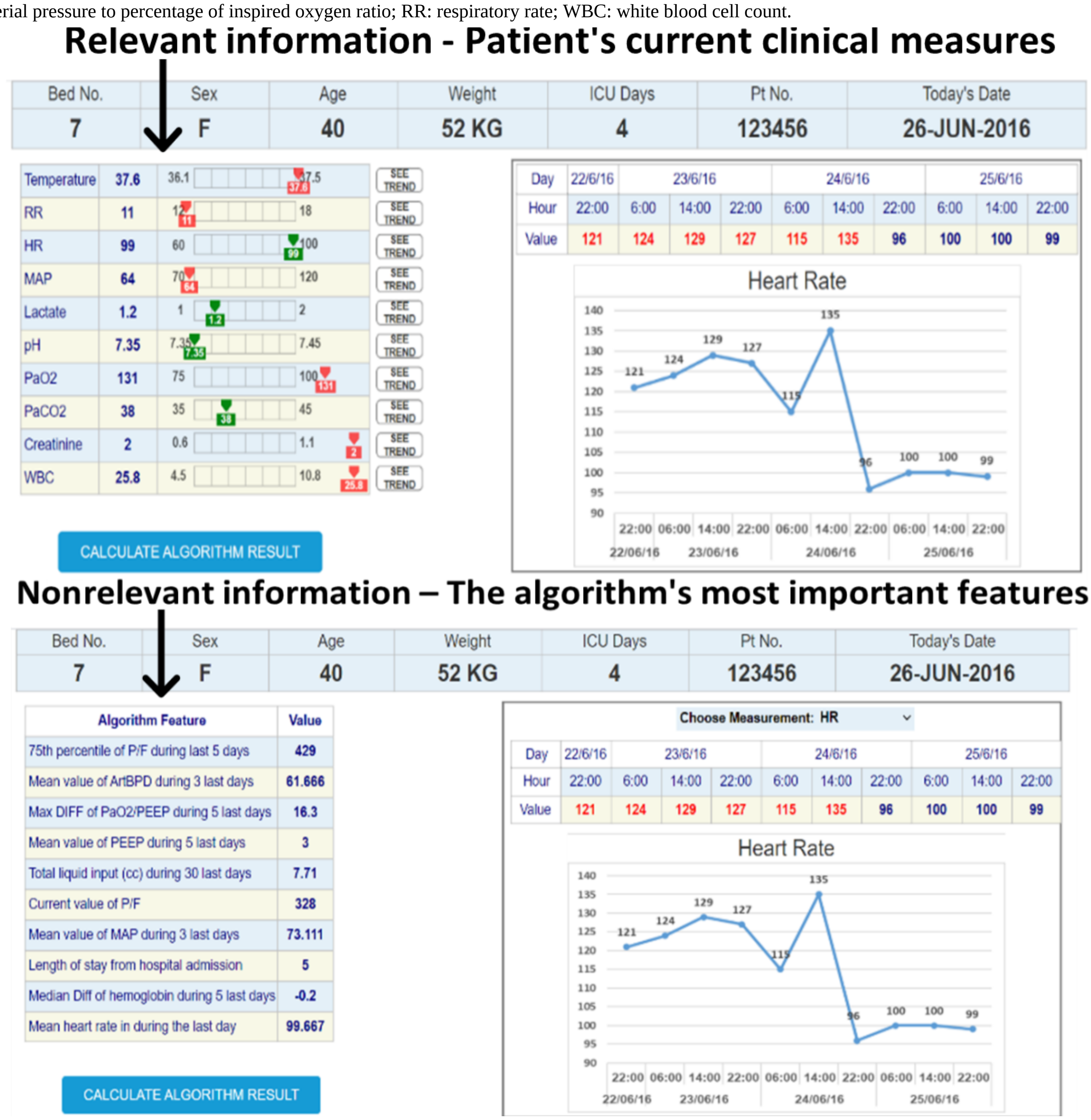
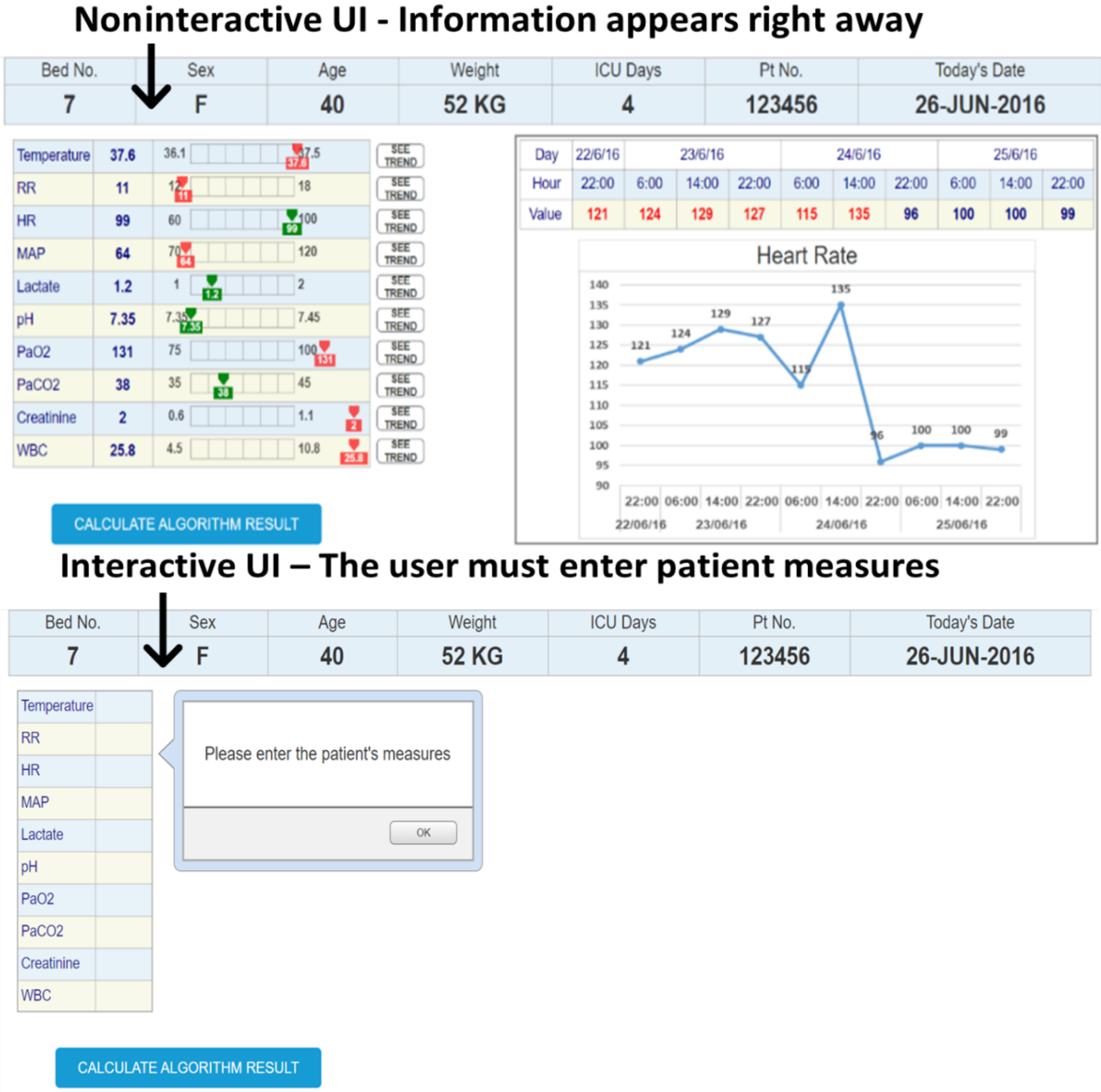


Figure 6. The interactive and noninteractive conditions. The top frame shows the noninteractive condition. The bottom frame shows the interactive condition, in which participants had to actively engage with the UI and provide the patient's current clinical measures, before they could explore the other charts and information. HR: heart rate; ICU: intensive care unit; MAP: mean arterial pressure; RR: respiratory rate; UI: user interface: WBC: white blood cell count.



Procedure

The participants were introduced to the purpose of the study and received an explanation about the ML bacteremia prediction system. They were then introduced to the simulation software, with the UI fitting the condition they were assigned. The participants were asked to first read the clinical case, and only then to explore the UI. After exploring the UI, they could click on the "calculate algorithm result" button to receive the algorithm's prediction. The predictions that were presented to the participants were accurate. Participants in the interactive condition had to enter the values of the patient's current clinical measures before the system calculated the algorithm result. All the participants were asked to handle the information as though they were taking the described patient under their care, and the information provided was all that was available to them.

After the algorithm presented its prediction, the participants could continue to explore the UI and the information presented, and then answer whether they agreed with the algorithm's prediction or not. After answering this question, they proceeded to the same procedure with two additional clinical cases. To avoid order bias, counterbalancing was used. The number of times participants agreed with the system's prediction represents their reaction to the system.

Postexperiment Questionnaires

After completing the 3 clinical cases, the participants answered 2 demographic questions about their experience and gender and 3 questionnaires about their trust in the system, the interactivity of the system, and the information relevancy of the system. The postexperiment questionnaires measured perceived understanding of the system. These consisted of the AIMQ (AIM quality) questionnaire [51] to measure information relevancy, 7 items from an interactivity questionnaire [44] that assessed interactivity, and 14 items from a questionnaire that assessed trust [30]. The latter questionnaire associated the CBT subdimensions of understandability, technical competence, and reliability from the human-computer trust questionnaire. All the questionnaires used a 7-point Likert scale (1=low and 7=high). See Table 2 for the entire list of the variables. To control for possible variance, the gender and years of experience of the participants were recorded. These analyses were performed because studies have shown a significant impact of gender [5,28,52] and years of experience [29,53] on the interaction of humans with automation, and a consequent influence on the development of human-automation trust. The questionnaire questions are presented in Multimedia Appendix 1.

Table . The experiment variables.

Construct	Scale	How it was measured
Years of experience	Continuous	Demographics
Gender	Nominal	Demographics
UI[a] level of information relevancy	Binary	By design
UI level of interactivity	Binary	By design
Information relevancy rating	Discrete (1-7)	AIMQ[b] questionnaire [51]
Interactivity rating	Discrete (1-7)	McMillan and Hwang [44]
Understandability	Discrete (1-7)	HCT[c]; Medsen and Gregor [30]
Technical competence	Discrete (1-7)	HCT; Medsen and Gregor [30]
Reliability	Discrete (1-7)	HCT; Medsen and Gregor [30]
Cognitive-based trust	Discrete (1-7)	HCT; Medsen and Gregor [30]
Agreement with the system	Discrete (0 - 3)	Simulation software

[a]UI: user interface.
[b]AIMQ: AIM quality.
[c]HCT: human-computer trust.

Data Analysis

To measure the participants' immediate reaction to the system, the participants were grouped by the number of times they agreed with the system's prediction. This information was compared with their information relevancy rating. Due to the different group sizes, the Welch test was used to conduct the comparisons.

A linear regression model was used to assess the influence of several variables on trust as a single construct (cognitive-based trust). Although the 2 study hypotheses aimed to identify the main effects of information relevancy and interactivity on trust, variables 1 - 6 (years of experience, gender, UI level of information relevancy, UI level of interactivity, information relevancy rating, and interactivity rating) were included in the model to control for possible variance. Interactions were assessed on gender and years of experience with all the other variables.

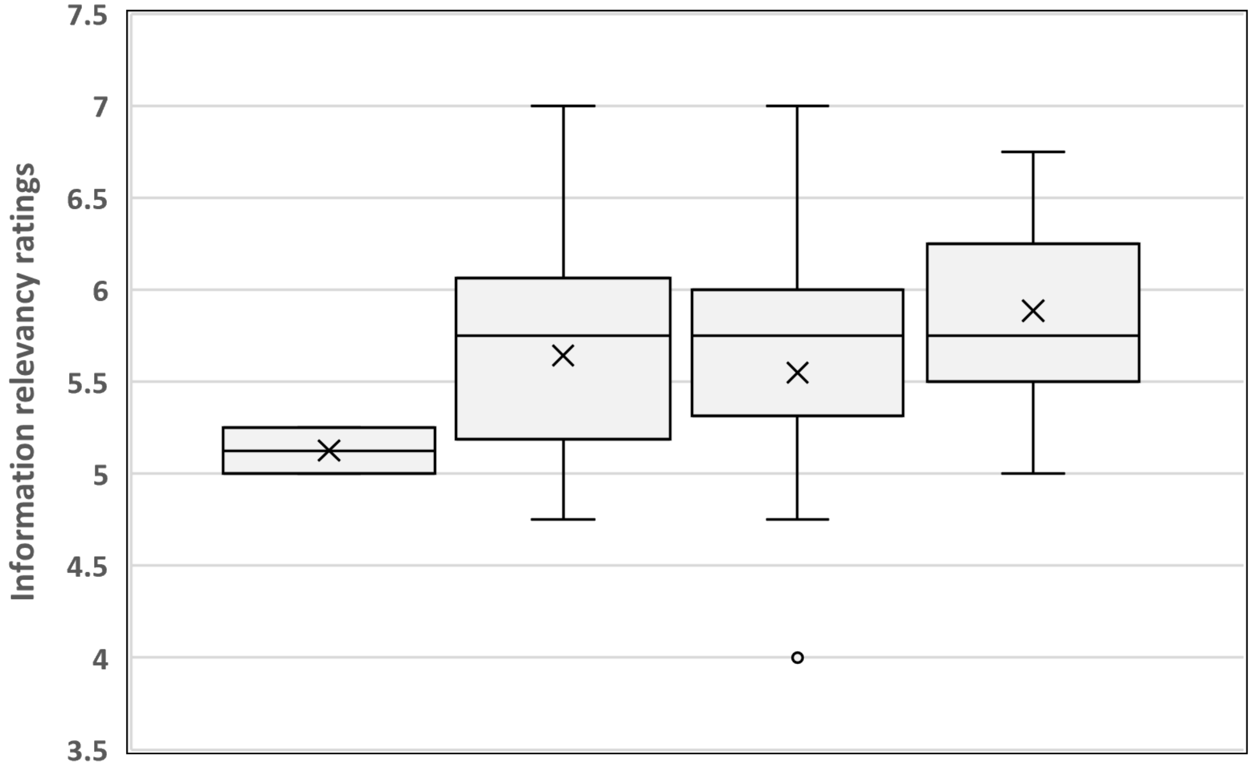
Three linear regression models were used to assess the effect of CBT subdimensions. Variables 1 - 6 were included in the models to control possible variance. Interactions were assessed on gender and years of experience with all the other variables.

Results

Participants' Agreement With the System's Prediction

The conditions of the experiments (variables 3 and 4; Table 2) were not found to be associated with the participants' trust in the system. However, participants' responses to the postexperiment questionnaires reveal significant findings. Overall, the higher the participants rated information relevancy, the more frequently they agreed with the system's prediction. Information relevancy was rated significantly higher among those who agreed 3 times with the system's prediction compared to those who did not agree at all (t_{11}=–3.924, 2-tailed; P=.05). No other comparisons between the groups were significant (see Figure 7). Participants' agreement with the system's prediction did not differ according to their experience, gender, or the interactivity ratings of the system.

Figure 7. Box plot of the information relevancy ratings and the number of times participants agreed with the system's prediction. A significant difference was found between participants who agreed with all the system's prediction and participants who did not agree at all.



Trust as a Single Construct

The significant main effect for the UI level of information relevancy revealed that relevant information resulted in higher perceived trust (β=2.684; *P*=.05). Higher information relevancy ratings (β=.824; *P*<.001) and higher interactivity ratings (β=.613; *P*<.001) were associated with higher perceived trust in the system. A significant interaction between UI level of interactivity and years of experience (β=–.056; *P*=.05) revealed lower trust ratings among experienced participants with higher interactivity ratings. The adjusted R^2 of the regression model was 0.5296.

CBT Subdimensions

A significant main effect was observed for the UI level of information relevancy and technical competence (β=4.5; *P*<.001). In addition, across all the models, significant main effects for information relevancy ratings and interactivity ratings were observed. The statistical measures are summarized in Table 3. No other significant main or interaction effects were observed across the subdimensions.

Table . Statistics for the subdimensions of the cognition-based trust (CBT) dimension.

CBT subdimension	$\beta_{\text{information relevancy ratings}}$	$P_{\text{information relevancy ratings}}$	$\beta_{\text{interactivity ratings}}$	$P_{\text{interactivity ratings}}$
Technical competence	1.18	<.001	.6	<.001
Understandability	.4	<.001	.53	<.001
Reliability	.72	<.001	.53	<.001

Discussion

Principal Findings

Trust is difficult to measure. Participants' agreement with the system's prediction did not differ according to the experimental conditions. However, in the postexperiment questionnaire, higher information relevancy ratings and interactivity ratings were associated with higher perceived trust in the system, and the explicit visual presentation of the features of the ML algorithm on the user interface resulted in lower trust by the participants.

Information Relevancy

The results of our experiment revealed that information relevancy plays an important role in operators' trust in ML-based systems. Two different, but complementary questions were addressed and they are (1) to what extent does relevant information enhance intensivists' trust in ML-based CDSSs? and (2) what type of information do intensivists consider to be relevant? The answer to the first question is derived directly from the results—perceived relevant information is important and affects various aspects of the operators' trust in the system. This finding supports the first hypothesis and corroborates studies from diverse domains, which found that information relevancy substantially influences users' trust in technological systems [41,54,55].

Regarding the second question, discerning the type of information that intensivists consider relevant is more complicated. As hypothesized, providing detailed information about the algorithm's features decreased the participants' trust in the system. A possible explanation for the decreased trust is that the participants found the detailed information about the ML algorithm confusing and irrelevant. Accordingly, the information about the ML algorithm may have supported the participants' belief that they were dealing with a black box algorithm, and this, in turn, may have fostered distrust of the system [13].

Across all the CBT subdimensions assessed (understandability, technical competence, and reliability), the greater the relevancy of the information presented in the UI, according to the participants, the higher their trust. This concurs with the analysis of trust as a standalone construct and thus supports the first hypothesis.

The understandability and reliability ratings were not found to differ significantly between the information relevancy conditions. This suggests that the presentation of ML features did not significantly decrease the participants' ratings of understandability and reliability. However, ratings of technical competence did differ between the information relevancy conditions. This could indicate a stronger effect on trust, in the technical competence subdimension, compared to understandability and reliability.

Interactivity

The participants' trust ratings were not found to differ significantly between conditions. However, trust ratings increased as participants' perception of UI interactivity increased. This finding supports the second hypothesis and is in line with a meta-analysis by Yang and Shen [56], which concluded that perceived interactivity was much more effective than objective interactivity.

Two possibilities arise to explain the gap between participants' perceptions of the interactivity and the actual UI level of interactivity. First, within the 2 interactivity levels, the objective gap between the different conditions may not have been strong enough. The less interactive condition also forced 2-way communication between the participants and the UI. Possibly, the initial user engagement did not add enough interactivity to render a noticeable difference. Alternatively, the participants may not have perceived increased interactivity. Second, although entering and copying values to and from the patient record is a common task clinicians must apply in a subset of the IT systems in the ICU, participants may have considered that manually entering the patient's clinical measures was dull or redundant. This could have reduced participants' opinion of the system and led to lower trust ratings.

Although more interactive perceptions of the UI were associated with higher trust ratings, it is arguable whether extreme levels of interactivity are always preferable. Kalet et al [57] investigated the influence of different interactivity levels in a computer-assisted instruction system on medical students' performances. They found that a mid-range UI level of interactivity maximized improvements in the performance of clinical skills. Yang and Shen [56] found that extremely high levels of website interactivity were less effective than moderate levels. However, pinpointing the exact amount of moderate interactivity, universally or specifically for a domain, is challenging. Furthermore, treating interactivity as a continuous variable and fitting it into a linear regression model could lead to measurement and interpretation errors. According to Yang and Shen [56], interactivity should be considered as a curvilinear variable, with the peak at the center of the curve and not at the edges. When fitting a linear regression model to an interactivity variable, the latter is considered linear, but this is not always the case. This approach may fail to capture the real influence of different levels of interactivity.

Across the 3 CBT dimensions examined (understandability, technical competence, and reliability), the more interactive the UI, according to the participants' perception, the higher their trust. This was precisely the situation when trust was analyzed as a standalone construct. Otherwise, the interactivity levels examined were not found to differ between the CBT dimensions. Notably, a linear regression model was set for each subdimension. Although the results showed that the more interactive the UI, the higher the ratings for each subdimension, moderate levels of interactivity may have had a greater effect on those subdimensions.

Finally, the literature is scant regarding correlations between experience and interactivity, and additional research is needed to elaborate on the significant negative interaction across years of experience and interactivity ratings.

Limitations and Future Research

Some limitations of this study represent opportunities for future research. First, the study design, limited resources, and the period the study was conducted (between the first and second waves of the COVID-19 pandemic) posed limitations on participant recruitment. The limited sample size dictated a design with only 2 levels of each variable. Future research should explore advanced and more realistic UI interactions and different information types. Second, although Madsen and Gregor's [30] approach was used to analyze trust, the ABT dimensions were not explored. Such investigation is needed to obtain a wider view of the relations between trust and its subdimensions, both cognitive-based and affect-based. Third, due to time limitations, the study did not evaluate participants' attitudes and changes in trust in the system over time. Finally, the study was performed in a simulation environment, using a specific interface design, and using case studies rather than real-time data from patients. Investigating clinician collaboration with a variety of interface designs, within real-world information systems used in diverse health care settings could yield a deeper understanding of future interface design.

Conclusions

Developing ML algorithms is only the first step toward improving medical treatment. To increase acceptance and trust of ML-based CDSSs, and expand their use, a broader and more multidisciplinary approach (eg, user-centered design) should be taken. This approach needs to be specifically evaluated in the health care work environment, considering its unique challenges and professional personnel. A better understanding of means to increase intensivists' trust in ML-based CDSSs may open new opportunities for user-centered design and improved decision-making processes in the ICU.

Human factor studies, like this one, highlight the importance of understanding the effect of specific UI features when designing ML-based CDSS and other "artificial intelligence" systems. This study focused on the effects of 2 UI features related to intensivists' trust in ML-based CDSSs. We demonstrated that the level of relevancy of the information that is presented in the UI and the interactivity level of the UI can play major roles when designing ML-based CDSSs. However, to enhance trust in these systems, more UI features should be investigated.

A wide point of view on trust should be maintained. In this study, trust as a standalone construct was influenced significantly by the different information relevancy levels in the tested conditions. Of the CBT subdimensions, only technical competence was influenced in the same way. These findings emphasize the need to analyze trust from different perspectives. For the research community and system designers, this may promote a broad understanding of means to enhance and foster trust in ML-based CDSSs, as well as in other "artificial intelligence" systems.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The questionnaire.
[DOCX File, 20 KB - humanfactors_v11i1e56924_app1.docx]

References

1. McKenzie MS, Auriemma CL, Olenik J, Cooney E, Gabler NB, Halpern SD. An observational study of decision making by medical intensivists. Crit Care Med 2015 Aug;43(8):1660-1668. [doi: 10.1097/CCM.0000000000001084] [Medline: 26035147]
2. Ward NS, Afessa B, Kleinpell R, et al. Intensivist/patient ratios in closed ICUs: a statement from the Society of Critical Care Medicine Taskforce on ICU Staffing. Crit Care Med 2013 Feb;41(2):638-645. [doi: 10.1097/CCM.0b013e3182741478] [Medline: 23263586]
3. Sharma M, Savage C, Nair M, Larsson I, Svedberg P, Nygren JM. Artificial intelligence applications in health care practice: scoping review. J Med Internet Res 2022 Oct 5;24(10):e40238. [doi: 10.2196/40238] [Medline: 36197712]
4. Poon AIF, Sung JJY. Opening the black box of AI-medicine. J Gastroenterol Hepatol 2021 Mar;36(3):581-584. [doi: 10.1111/jgh.15384] [Medline: 33709609]
5. Liberman-Pincu E, Bitan Y. Fule—functionality, usability, look-and-feel and evaluation novel user-centered product design methodology—illustrated in the case of an autonomous medical device. Appl Sci 2021 Feb;11(3):985. [doi: 10.3390/app11030985]
6. Obermeyer Z, Emanuel EJ. Predicting the future - big data, machine learning, and clinical medicine. N Engl J Med 2016 Sep 29;375(13):1216-1219. [doi: 10.1056/NEJMp1606181] [Medline: 27682033]
7. Mohammed M, Pathan ASK. Automatic Defense Against Zero-Day Polymorphic Worms in Communication Networks: Auerbach Publications; 2013.
8. Islam MM, Nasrin T, Walther BA, Wu CC, Yang HC, Li YC. Prediction of sepsis patients using machine learning approach: a meta-analysis. Comput Methods Programs Biomed 2019 Mar;170(January):1-9. [doi: 10.1016/j.cmpb.2018.12.027] [Medline: 30712598]
9. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017 Feb 2;542(7639):115-118. [doi: 10.1038/nature21056] [Medline: 28117445]
10. Beck AH, Sangoi AR, Leung S, et al. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. Sci Transl Med 2011 Nov 9;3(108):108ra113. [doi: 10.1126/scitranslmed.3002564] [Medline: 22072638]
11. Sanchez-Pinto LN, Luo Y, Churpek MM. Big data and data science in critical care. Chest 2018 Nov;154(5):1239-1248. [doi: 10.1016/j.chest.2018.04.037] [Medline: 29752973]
12. Choy G, Khalilzadeh O, Michalski M, et al. Current applications and future impact of machine learning in radiology. Radiology 2018 Aug;288(2):318-328. [doi: 10.1148/radiol.2018171820] [Medline: 29944078]
13. Stewart J, Sprivulis P, Dwivedi G. Artificial intelligence and machine learning in emergency medicine. Emerg Med Australas 2018 Dec;30(6):870-874. [doi: 10.1111/1742-6723.13145] [Medline: 30014578]
14. Wu M, Hughes MC, Parbhoo S, Zazzi M, Roth V, Doshi-Velez F. Beyond sparsity: tree regularization of deep models for interpretability. Presented at: AAAI'18: AAAI Conference on Artificial Intelligence; Feb 2-7, 2018; New Orleans, LA. [doi: 10.1609/aaai.v32i1.11501]

15. Bitan Y, Patterson ES. Unique challenges in user interface design for medical devices that use predictive algorithms. Proc Int Symp Hum Factors Ergon Health Care 2020 Sep;9(1):265-266. [doi: 10.1177/2327857920091004]
16. Narayanan M, Chen E, He J, Kim B, Gershman S, Doshi-Velez F. How do humans understand explanations from machine learning systems? an evaluation of the human-interpretability of explanation. arXiv. Preprint posted online on Feb 2, 2018 URL: https://arxiv.org/abs/1802.00682 [accessed 2024-07-25] [doi: 10.48550/arXiv.1802.00682]
17. Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. arXiv. Preprint posted online on Feb 28, 2017 URL: https://arxiv.org/abs/1702.08608 [accessed 2024-07-25] [doi: 10.48550/arXiv.1702.08608]
18. Du M, Liu N, Hu X. Techniques for interpretable machine learning. Commun ACM 2019 Dec 20;63(1):68-77. [doi: 10.1145/3359786]
19. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 2019 May;1(5):206-215. [doi: 10.1038/s42256-019-0048-x] [Medline: 35603010]
20. Miller T. Explanation in artificial intelligence: insights from the social sciences. Artificial Intelligence 2019 Feb;267:1-38. [doi: 10.1016/j.artint.2018.07.007]
21. Kay M, Kola T, Hullman JR, Munson SA. When (ish) is my bus? User-centered visualizations of uncertainty in everyday, mobile predictive systems. Presented at: CHI'16: CHI Conference on Human Factors in Computing Systems; May 7-12, 2016; San Jose, CA p. 5092-5103. [doi: 10.1145/2858036.2858558]
22. Kulesza T, Stumpf S, Burnett M, Kwan I. Tell me more? The effects of mental model soundness on personalizing an intelligent agent. Presented at: CHI '12: CHI Conference on Human Factors in Computing Systems; May 5-10, 2012; Austen, TX p. 1-10. [doi: 10.1145/2207676.2207678]
23. Stumpf S, Rajaram V, Li L, et al. Interacting meaningfully with machine learning systems: three experiments. Int J Hum Comput Stud 2009 Aug;67(8):639-662. [doi: 10.1016/j.ijhcs.2009.03.004]
24. Yang R, Newman MW. Learning from a learning thermostat: lessons for intelligent systems for the home. Presented at: UbiComp '13: The 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing; Sep 8-12, 2013; Zurich, Switzerland p. 93-102. [doi: 10.1145/2493432.2493489]
25. Xu W, Dainoff MJ, Ge L, Gao Z. Transitioning to human interaction with AI systems: new challenges and opportunities for HCI professionals to enable human-centered AI. Int J Hum-Comput Interact 2023 Feb 7;39(3):494-518. [doi: 10.1080/10447318.2022.2041900]
26. Abdul A, Vermeulen J, Wang D, Lim BY, Kankanhalli M. Trends and trajectories for explainable, accountable and intelligible systems. Presented at: CHI '18: CHI Conference on Human Factors in Computing Systems; Apr 21-26, 2018; Montreal, QC p. 1-18. [doi: 10.1145/3173574.3174156]
27. Parasuraman R, Riley V. Humans and automation: use, misuse, disuse, abuse. Hum Factors 1997 Jun;39(2):230-253. [doi: 10.1518/001872097778543886]
28. Hoff KA, Bashir M. Trust in automation: integrating empirical evidence on factors that influence trust. Hum Factors 2015 May;57(3):407-434. [doi: 10.1177/0018720814547570] [Medline: 25875432]
29. Lee JD, See KA. Trust in automation: designing for appropriate reliance. Hum Factors 2004 Feb;46(1):50-80. [doi: 10.1518/hfes.46.1.50.30392]
30. Madsen M, Gregor S. Measuring human-computer trust. Proc Elev Australas Conf Inf Syst 2000;53:6-8 [FREE Full text]
31. Ghazizadeh M, Lee JD, Boyle LN. Extending the technology acceptance model to assess automation. Cogn Tech Work 2012 Mar;14(1):39-49. [doi: 10.1007/s10111-011-0194-3]
32. Jian JY, Bisantz AM, Drury CG. Foundations for an empirically determined scale of trust in automated systems. Int J Cogn Ergon 2000 Mar;4(1):53-71. [doi: 10.1207/S15327566IJCE0401_04]
33. Hengstler M, Enkel E, Duelli S. Applied artificial intelligence and trust-the case of autonomous vehicles and medical assistance devices. Technol Forecast Soc Change 2016 Apr;105:105-120. [doi: 10.1016/j.techfore.2015.12.014]
34. Sheridan TB. Extending three existing models to analysis of trust in automation: signal detection, statistical parameter estimation, and model-based control. Hum Factors 2019 Nov;61(7):1162-1170. [doi: 10.1177/0018720819829951] [Medline: 30811950]
35. Mayer RC, Davis JH, Schoorman FD. An integrative model of organizational trust. Acad Manag Rev 1995 Jul;20(3):709. [doi: 10.2307/258792]
36. Cassell J, Sullivan J, Prevost S, Churchill E. Embodied Conversational Agents: MIT Press; 2000.
37. Paiva A. Affective interactions: towards a new generation of computer interfaces. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Springer; 1999. [doi: 10.1007/10720296]
38. Pickering MJ, Garrod S. Toward a mechanistic psychology of dialogue. Behav Brain Sci 2004 Apr;27(2):169-190. [doi: 10.1017/s0140525x04000056] [Medline: 15595235]
39. Muylle S, Moenaert R, Despontin M. The conceptualization and empirical validation of web site user satisfaction. Inf Manag 2004 May;41(5):543-560. [doi: 10.1016/S0378-7206(03)00089-2]
40. Park YA, Gretzel U. Success factors for destination marketing web sites: a qualitative meta-analysis. J Travel Res 2007 Aug;46(1):46-63. [doi: 10.1177/0047287507302381]

41. Song J, Zahedi FM. Trust in health infomediaries. Decis Support Syst 2007 Mar;43(2):390-407. [doi: 10.1016/j.dss.2006.11.011]
42. Tsakonas G, Papatheodorou C. Analysing and evaluating usefulness and usability in electronic information services. J Inf Sci 2006 Oct;32(5):400-419. [doi: 10.1177/0165551506065934]
43. Steuer J. Defining virtual reality: dimensions determining telepresence. J Commun 1992 Dec;42(4):73-93. [doi: 10.1111/j.1460-2466.1992.tb00812.x]
44. McMillan SJ, Hwang JS. Measures of perceived interactivity: an exploration of the role of direction of communication, user control, and time in shaping perceptions of interactivity. J Advert 2002 Oct;31(3):29-42. [doi: 10.1080/00913367.2002.10673674]
45. Bezjian-Avery A, Calder B, Iacobucci D. New media interactive advertising vs traditional advertising. J Advert Res 1998;38(4):23-32.
46. Hoffman DL, Novak TP. Marketing in hypermedia computer-mediated environments: conceptual foundations. J Mark 1996 Jul;60(3):50-68. [doi: 10.1177/002224299606000304]
47. Sorrell M, Salama E, Levin M, et al. The future of interactive marketing. Harvard Business Review. 1996. URL: https://hbr.org/1996/11/the-future-of-interactive-marketing [accessed 2024-07-25]
48. Cyr D, Head M, Ivanov A. Perceived Interactivity leading to e-loyalty: development of a model for cognitive–affective user responses. Int J Hum Comput Stud 2009 Oct;67(10):850-869. [doi: 10.1016/j.ijhcs.2009.07.004]
49. Lee T. The impact of perceptions of interactivity on customer trust and transaction intentions in mobile commerce. J Electron Commer Res 2005;6(3):165-180 [FREE Full text]
50. Vallés J, León C, Alvarez-Lerma F. Nosocomial bacteremia in critically ill patients: a multicenter study evaluating epidemiology and prognosis. Spanish collaborative group for infections in intensive care units of Sociedad Espanola de Medicina Intensiva Y Unidades Coronarias (SEMIUC). Clin Infect Dis 1997 Mar;24(3):387-395. [doi: 10.1093/clinids/24.3.387] [Medline: 9114190]
51. Lee YW, Strong DM, Kahn BK, Wang RY. AIMQ: a methodology for information quality assessment. Inf Manage 2002 Dec;40(2):133-146. [doi: 10.1016/S0378-7206(02)00043-5]
52. Schaefer KE, Chen JYC, Szalma JL, Hancock PA. A meta-analysis of factors influencing the development of trust in automation: implications for understanding autonomy in future systems. Hum Factors 2016 May;58(3):377-400. [doi: 10.1177/0018720816634228] [Medline: 27005902]
53. Stanton B, Jensen T. Trust and artificial intelligence. : National Institute of Standards and Technology; 2021 URL: https://www.nist.gov/publications/trust-and-artificial-intelligence-draft [accessed 2024-07-25]
54. Nicolaou AI, McKnight DH. Perceived information quality in data exchanges: effects on risk, trust, and intention to use. Inf Syst Res 2006 Dec;17(4):332-351. [doi: 10.1287/isre.1060.0103]
55. Zhou T. An empirical examination of initial trust in mobile banking. Internet Res 2011 Aug 12;21(5):527-540. [doi: 10.1108/10662241111176353]
56. Yang F, Shen F. Effects of web interactivity: a meta-analysis. Commun Res 2018 Jul;45(5):635-658. [doi: 10.1177/0093650217700748]
57. Kalet AL, Song HS, Sarpel U, et al. Just enough, but not too much Interactivity leads to better clinical skills performance after a computer assisted learning module. Med Teach 2012;34(10):833-839. [doi: 10.3109/0142159X.2012.706727] [Medline: 22917265]

Abbreviations

ABT: affect-based trust
AIMQ: AIM quality
CBT: cognition-based trust
CDSS: clinical decision support system
HCI: human-computer interaction
ICU: intensive care unit
ML: machine learning
UI: user interface

Edited by A Choudhury; submitted 30.01.24; peer-reviewed by A Bamgboje-Ayodele, L Herrle, M Sedlmayr, R Marshall, S Sarbadhikari; revised version received 09.05.24; accepted 24.05.24; published 01.08.24.

Please cite as:
Katzburg O, Roimi M, Frenkel A, Ilan R, Bitan Y
The Impact of Information Relevancy and Interactivity on Intensivists' Trust in a Machine Learning–Based Bacteremia Prediction System: Simulation Study
JMIR Hum Factors 2024;11:e56924
URL: https://humanfactors.jmir.org/2024/1/e56924
doi:10.2196/56924

Perceived Benefit and Satisfaction With a Tablet Computer and an Emergency Smartwatch by Older Adults and Their Relatives: Prospective Real-World Pilot Study

Patrick Wiegel[1,2,3*], PhD; Marina Liselotte Fotteler[2,4*], MSc; Brigitte Kohn[1,3], BSc; Sarah Mayer[1,3], Dr biol hum; Filippo Maria Verri[1,2,3], MD; Dhayana Dallmeier[1,3,5], MD, PhD; Michael Denkinger[1,2,3], MD

[1]Research Unit on Ageing, Agaplesion Bethesda Clinic Ulm, Zollernring 26, Ulm, Germany

[2]Institute for Geriatric Research, Ulm University Medical Center, Ulm, Germany

[3]Geriatric Center Ulm, Ulm, Germany

[4]DigiHealth Institute, Neu-Ulm University of Applied Sciences, Neu-Ulm, Germany

[5]Department of Epidemiology, Boston University School of Public Health, Boston, MA, United States

[*]these authors contributed equally

Corresponding Author:
Patrick Wiegel, PhD
Research Unit on Ageing, Agaplesion Bethesda Clinic Ulm, , Zollernring 26, Ulm, , Germany

Abstract

Background: Assistive technologies (ATs) have the potential to promote the quality of life and independent living of older adults and, further, to relieve the burden of formal and informal caregivers and relatives. Technological developments over the last decades have led to a boost of available ATs. However, evidence on the benefits and satisfaction with ATs in real-world applications remains scarce.

Objective: This prospective, real-world, pilot study tested the perceived benefit and satisfaction with different ATs in the real-world environment.

Methods: Community-dwelling adults aged ≥65 and their relatives tested a tablet computer with a simplified interface or a smartwatch with programmable emergency contacts for 8 weeks in their everyday life. Perceived benefits and satisfaction with ATs were assessed by all older adults and their relatives using different assessment tools before and after the intervention. Outcome measures included the Technology Usage Inventory, Quebec User Evaluation of Satisfaction with Assistive Technology 2.0, and Canadian Occupational Performance Measure.

Results: A total of 17 older adults (tablet computer: n=8, 47% and smartwatch: n=9, 53%) and 16 relatives (tablet computer: n=7, 44% and smartwatch: n=9, 56%) were included in the study. The number of participants that were frail (according to the Clinical Frailty Scale) and received care was higher in the smartwatch group than in the tablet computer group. Older adults of the smartwatch group reported higher technology acceptance (Technology Usage Inventory) and satisfaction (Quebec User Evaluation of Satisfaction with Assistive Technology 2.0) scores than those of the tablet computer group, although the differences were not significant (all $P>.05$). In the tablet computer group, relatives had significantly higher ratings on the item *intention to use* than older adults ($t_{12.3}$=3.3, P=.006). Identified everyday issues with the Canadian Occupational Performance Measure included contact/communication and entertainment/information for the tablet computer, safety and getting help in emergency situations for the smartwatch, and the usability of the AT for both devices. While the performance (t_8=3.5, P=.008) and satisfaction (t_8=3.2, P=.01) in these domains significantly improved in the smartwatch group, changes in the tablet computer group were inconsistent (all $P>.05$).

Conclusions: This study highlights the remaining obstacles for the widespread and effective application of ATs in the everyday life of older adults and their relatives. While the results do not provide evidence for a positive effect regarding communication deficits, perceived benefits could be shown for the area of safety. Future research and technical developments need to consider not only the preferences, problems, and goals of older adults but also their relatives and caregivers to improve the acceptability and effectiveness of ATs.

(JMIR Hum Factors 2024;11:e53811) doi:10.2196/53811

KEYWORDS

assistive technology; older adults; caregiver; benefits; usability; gerontechnology

Introduction

In their global disability action plan from 2015, the World Health Organization (WHO) defined assistive technologies (ATs) as "any item, piece of equipment or product, whether it is acquired commercially, modified or customized, that is used to increase, maintain or improve the functional capabilities of individuals with disability" [1]. For older adults, ATs can positively impact not only functionality but also autonomy, safety, and communication. In particular, ATs might support older adults who wish to remain in their own homes instead of depending on institutional care. In addition, the application of ATs has become a valid option in relieving the formal and informal caregivers of older adults [2].

In recent years, the field of ATs has evolved to include a wide range of devices for different audiences, varying in complexity and price. Examples include mobile health applications [3], wearable devices [4,5], robotic systems [6], virtual reality applications [7], or sensory aids [8]. ATs are used for various purposes such as personal disease management [3], managing fall risks [9], ensuring correct medication [10,11], and preventing social isolation [12,13]. Clearly, there is an abundance of available ATs assisting with age-related challenges. Yet, the effectiveness of ATs for older adults remains inconclusive [3,9,12,14]. Nevertheless, studies investigating the benefits and usability of ATs for older adults in the real-world environment are still lacking [15-17]. A focus on multimorbid and frail populations is of special interest, as one would expect that this cohort might especially benefit from the use of ATs. In addition, the use of ATs by older frail and nonfrail adults often depends on and includes relatives and caregivers who significantly contribute to the (successful) application of ATs. Additionally, this group itself might benefit from the application of ATs through a reduction of care burden [18]. Thus, studies testing the effectiveness of ATs should also consider the effects on caregivers and relatives.

Older adults are less likely than younger adults to use technology, and this depends on a number of factors including sociodemographic factors, attitudinal variables, and cognitive abilities [19]. However, despite the assumption that older adults are afraid and reluctant to use ATs, recent evidence indicates that they are relatively open to the idea of using technologies in their everyday life, independent of their age and subjective health status [20,21]. The factors that influence the intention of older adults to use digital technologies are complex and include environmental, psychological, and social determinants [22]. Among the most important ranked criteria for the selection and use of ATs by older adults are the promotion of independence, affordability, ease of use, and ethical company policies [23]. Thus, the use and effectiveness of ATs ultimately depend on the alignment of the technological developments and user needs.

Recently, emergency buttons or watches and wearable devices with GPS tracking or fall detection for older adults have gained special interest. However, few studies exist investigating the benefit of these devices for older adults. Most existing research focuses on wearables worn for disease monitoring or activity tracking [24]. A recent study with an emergency button (worn as a wristband or necklace) that was connected to the landline did not show any improvement for health-related quality of life and other outcome measures [5]. Focusing on the issues of loneliness and lack of social interaction, software systems for (tablet) computers designed specifically for older adults are entering the market. The systems feature large icons and simple menus. Functionalities include video calls, access to the news, photo albums, or games. However, the benefit of such ATs remains unclear. While 1 study reported a significant improvement of loneliness, social support, and well-being among community-dwelling older adults living alone using a special computer system [13], a large systematic review and meta-analysis of similar interventions found no benefits on psychological outcomes in people with cognitive impairment and dementia [7].

This prospective, real-world, pilot study aims to assess the individual benefits and satisfaction gained from an emergency smartwatch and a senior tablet computer by community-dwelling older adults and their relatives in the real-world environment. The products were selected based on the results of a product competition for companies.

Methods

Study Population

Since the use of ATs often requires or supports the contact and interaction with relatives, this study considered 2 participant groups: older adults and their relatives or a close friend (hereinafter referred to as relatives). The inclusion criteria for older adults were (1) aged ≥65 years; (2) residence in the area of Ulm (distance ≤50 km); (3) an independent or assistive living situation; (4) the ability to speak and read German; and (5) the ability to consent to the study participation (adults with dementia were excluded). In addition, older adults had to have intact vision and hearing (normal or corrected). The relatives had to (1) be a family member or someone close to the adult; (2) speak and read German; (3) give consent to study participation; and (4) have a smartphone, tablet computer, or PC with internet connection. Participants were recruited in Southern Germany using convenience (via analogous and digital recruitment methods such as flyers, social media, and clinic staff) and snowball sampling (asking participants for other potential participants). As this was an exploratory study, no sample size calculation was conducted. The aim was to recruit between 6 to 10 participants per selected device. This number was defined based on feasibility. For 4 different ATs with 5 available devices each, 2 rounds of the study with 40 participants can be performed within 4 months.

Ethical Considerations

Ethical approval was obtained from Ulm University Ethical Committee (Nr. 230/21, 05.07.2021). Before entering the study, individuals received detailed information about the study and provided written informed consent to participate. Data analyses were performed on pseudonymized data. No financial remuneration was provided for participation.

Selection of the ATs

Initially, 4 different AT were chosen for this interventional study. The ATs were selected from the 2020 product competition “Daheim Dank Digital” (at home thanks to digitalization) [25] aimed at startups and established companies in German-speaking countries (Germany, Austria, and Switzerland). Companies were asked to apply with ATs that could compensate age-related deficits and impairments described in 5 predefined use cases. The use cases were designed together with experts from different areas of expertise and focused on (1) nocturnal restlessness and fall risk; (2) loneliness, hearing and vision impairment, and forgetfulness; (3) inactivity and listlessness in daily routine; (4) urinary incontinence and reduced fluid and food intake; and (5) limited mobility, weakness, and loneliness.

A total of 9 companies presented their products, and a scientific jury evaluated the devices based on availability; readiness for use; and impact in at least 1 of the categories of communication, security, or autonomy. These categories were specified as target areas in the overall project. A total of 4 products were selected for this user study based on their availability. The identified devices focused on safety (an intelligent bed exit alarm from the company NevisQ and a smartwatch with an emergency button from the company CareIOT GmbH) and communication (the Media4Care tablet computer from Media4Care GmbH and the Eldertech app, which supports communication and coordination of care within a family, from Eldertech GmbH). At the time of the study, all products were commercially available and had a CE certificate.

Study Design

We performed a prospective, real-world, pilot study of 4 different ATs with community-dwelling adults aged ≥65 years and their relatives. Together with the participants, the investigators decided which product was the most suitable for the older adult. Based on their individual needs, a decision was made for 1 of the targeted areas: communication or safety. Unfortunately, no candidates were identified for the evaluation of the bed exit alarm and the app; therefore, the selection was limited to the tablet computer and the emergency smartwatch. Both devices were used in their versions that were current as of August 2021. Details on the main functionalities of these 2 devices are listed in Table 1. Further technical information on the 2 products can be found in Multimedia Appendix 1.

In an introductory assessment session, all participants and their relatives were informed of the study procedure, goals, and possible risks and consented to participate. Baseline data were captured in a subsequent meeting at the participant’s home or in the study center (SM and BK). Participants were given the selected AT and a short introduction to the device. Participants then tested the AT in their daily life over a period of 8 weeks. In case problems with the use of the AT arose, participants were asked to first seek help in the manual, on the web, or via the company hotline before contacting the study center. This was done to reproduce a real-life setting as accurately as possible.

During the study period, the study team called the participants after 1, 2, 4, and 6 weeks to inquire about the use of the product; help solve possible issues; and remind participants to document errors, problems, and the frequency of use. After the test period of 8 weeks, the participants and their relatives were assessed a second time. This meeting took place at the participant’s home or in the study center. The relatives performed an independent assessment of the AT based on their user experience gathered during the testing period (assisting or interacting with older adults). Thus, the relatives’ ratings represent their own perception of the AT rather than the perception of the usefulness of the AT for the older adult.

Table . Main functionalities of the 2 assistive technologies tested in the study.

Device	Main functionalities and features
Smartwatch	• Emergency button, including the notification of contacts via phone and mail • GPS tracking • Analysis of movement profile and automatic notification of contact persons in case of deviations from the norm
Tablet computer	• Video calls • Messenger service • Entertainment (games, news, podcasts, radio, and music) • Photo album • Touchscreen

The questionnaires were paper based and filled out by 1 of the investigators (SM or BK) during the interview sessions before and after the intervention. Older adults and their relative were interviewed in separate sessions to reduce the interview duration and to avoid interaction, except when older adults required support from the relative. Interviews were performed face-to-face with all older adults. Interviews with relatives were also performed by phone.

Data Collection

At baseline, sociodemographic data were collected including sex, age, living situation (alone, together with partner, or with someone else), education (<10 y or ≥10 y), level of care (administratively assigned level of care measuring a person’s care need and determining their claim for additional support), self-perceived health status (excellent, very good, good, fair, or poor), and self-perceived age. Additionally, information on interest in technology was collected (high, medium, or low).

To assess older adults' ability to perform instrumental activities of daily living (IADL), the Lawton scale was used [26]. The corrected IADL score was calculated taking into account the possibility that the activities have never been performed [27]. To capture additional information about the participants' social situation, the Lubben Social Network Scale-6 was used [28]. With the Clinical Frailty Scale (CFS), frailty and fitness of the older adults (scale from 1="very fit" to 9="terminally ill") were assessed [29]. In addition, both older adults and relatives had to rate their life satisfaction (scale from 0 to 10, with higher values indicating higher life satisfaction).

Frequency of Use

Older adults were asked to document their use frequency of the technology in each of the 8 intervention weeks. The frequency of use was reported on a 5-point scale, that is, it captured whether the technology was used on 7, 5 - 6, 3 - 4, 1 - 2, or 0 days per week.

Technology Usage Inventory

The Technology Usage Inventory (TUI) was administered to the older adults and their relatives to assess the influence of psychological factors on the use and acceptance of technology [30]. The items *curiosity* and *technology anxiety* were asked before the intervention, whereas the items *interest*, *usability*, *usefulness*, *skepticism*, and *accessibility* were asked after the intervention. Answers were given on a 7-point Likert scale (1="strongly disagree" and 7="strongly agree"). In addition, the *intention to use* the AT was asked on a visual analog scale (0="agree" and 10="disagree") after the intervention. For orientation purposes, the item *intention to use* has been reversed in the *Results* section of this paper.

Quebec User Evaluation of Satisfaction With Assistive Technology 2.0

Satisfaction with the technology was evaluated after the intervention using the German version of the Quebec User Evaluation of Satisfaction with Assistive Technology (QUEST) 2.0 [31]. Of the original 12 items of the QUEST 2.0, only the 8 items relating to an AT device (domain device) were considered. Each question was scored on a 5-point scale (1="not satisfied at all" and 5="very satisfied"). The questionnaire was completed by both the older adults and their relative. Additionally, they were asked to choose the 3 items with the highest relevance for them. Overall satisfaction was calculated as the mean score across all answered items. The QUEST 2.0 has been shown to be a valid and reliable assessment for a population of assistive device users [32].

Canadian Occupational Performance Measure

The Canadian Occupational Performance Measure (COPM) is a semiquantitative, client-centered instrument that was used to investigate the perceived satisfaction with the performance of the AT in the specified problem areas of the older adults and their relatives [33]. For this purpose, everyday issues of personal importance related to the use of the AT were determined together with an occupational therapist during the interview before the intervention. For each participant, up to 5 everyday issues were identified, and for each of these issues, the older adults and their relatives had to rate their perceived level of performance and satisfaction on a scale from 1 to 10, with higher scores indicating higher performance or satisfaction. Ratings for the defined issues were then summed and divided by the number of issues to obtain a performance score and a satisfaction score. After testing the AT, the older adults and their relatives repeated the rating on the previously defined everyday issues of personal importance.

Data Analysis and Statistics

For all outcome measures, descriptive statistics were calculated and included the mean (SD). The focus of the analysis was on the outcomes that assessed the usability aspects of the ATs, that is, the TUI and QUEST 2.0. In addition, exploratory statistical inference testing was performed to assess the effects of the ATs on the specified everyday issues of the older adults and their relatives (COPM). The assumption of a normal distribution was tested by the Kolmogorov-Smirnov test of normality. Subsequently, 2-tailed t tests were used to compare intervention groups using the Welch t tests (life satisfaction, TUI, and QUEST 2.0) and preintervention and postintervention data using paired t tests (COPM). Uncorrected P values are presented in the *Results* section. However, to control the false discovery rate, we used the correction of the P values via the Benjamini-Hochberg procedure [34]. If this correction changed the outcome of the statistical test, this is reported in the *Results* section. In addition, effect sizes were calculated using Cohen d. The statistical significance was set at $P<.05$ for all tests. Statistical analyses were performed using RStudio (RStudio Team).

Results

Participants

The study was performed from August 2021 to April 2022. A total of 44 older adults were screened for inclusion in the study. From this sample, 18 older adults and their relatives met the inclusion criteria and were enrolled in this study. There was 1 dropout who stopped study participation due to dissatisfaction with the technology. A total of 17 older adults (tablet computer: n=8, 47% and smartwatch: n=9, 53%) and 16 relatives (tablet computer: n=7, 44% and smartwatch: n=9, 56%) completed the study and were included in the final data analysis. In the tablet computer group, there were 2 cases where the same relative belonged to 2 older adults, and there was 1 case where 2 relatives belonged to the same older adult.

The sociodemographic characteristics of the older adults and their relatives are shown in Tables 2 and 3, respectively. Older adults in both groups were aged >80 years on average but reported a younger self-perceived age. Even though participants in the smartwatch group were older on average, they felt younger than those in the tablet computer group. Data on self-perceived health status, level of care, corrected IADL, and frailty indicated that older adults in the smartwatch group were more dependent and slightly frailer than those in the tablet computer group. A total of 4 participants, all in the smartwatch group, perceived their own health to be poor or fair. However, older adults of the smartwatch group were socially more engaged than those of the tablet computer group. While most older adults (tablet computer: 7/8, 88% and smartwatch: 6/9, 67%) and all relatives indicated at least a medium level of technology interest, one-third (3/9, 33%) of participants in the smartwatch group reported a low level of technology interest. Relatives in the smartwatch group had a higher mean age and included more male persons than the relatives in the tablet computer group.

Table . Sociodemographic characteristics of older adults. Data are presented as frequency and percentage or as mean (SD).

Characteristics		Tablet computer (n=8)	Smartwatch (n=9)
Sex, n (%)			
	Female	5 (62)	5 (56)
	Male	3 (38)	4 (44)
Age (years), mean (SD)		80.1 (8.2)	82.7 (7.9)
Self-perceived age (years), mean (SD)		76.9 (12.4)	72.6 (15.4)
Living situation, n (%)			
	Alone	3 (38)	6 (67)
	With partner	5 (62)	3 (33)
Education (years), n (%)			
	<10	4 (50)	4 (44)
	≥10	4 (50)	5 (56)
Self-perceived health status, n (%)			
	Excellent or very good	2 (25)	1 (11)
	Good	6 (75)	4 (44)
	Fair or poor	0 (0)	4 (44)
Level of care [a] **, n (%)**			
	Yes	2 (25)	5 (56)
	No	6 (75)	4 (44)
Corrected IADL[b], mean (SD)		7.4 (1.4)	5.3 (2.6)
CFS[c] score, mean (SD)		3.4 (1.2)	3.7 (2.1)
Frailty (CFS score>4), n (%)		0 (0)	4 (44)
LSNS-6[d] score, mean (SD)		15.4 (3.5)	21.8 (2.0)
Socially isolated (LSNS-6 score<12), n (%)		1 (12)	0 (0)
Technology interest, n (%)			
	High	4 (50)	4 (44)
	Medium	3 (38)	2 (22)
	Low	1 (12)	3 (33)

[a]Level of care: administratively assigned level of care measuring a person's care need and determining their claim for additional support.
[b]IADL: instrumental activities of daily living (Lawton scale).
[c]CFS: Clinical Frailty Scale.
[d]LSNS-6: Lubben Social Network Scale-6.

Table . Sociodemographic characteristics of relatives. Data are presented as frequency and percentage or as mean (SD).

Characteristics		Tablet computer (n=7)	Smartwatch (n=9)
Sex, n (%)			
	Female	6 (86)	5 (56)
	Male	1 (14)	4 (44)
Age (years), mean (SD)		50.3 (18.2)	64.3 (11.3)
Education (years), n (%)			
	<10	1 (14)	1 (11)
	≥10	6 (86)	8 (89)
Technology interest, n (%)			
	High	3 (43)	5 (56)
	Medium	4 (57)	4 (44)
	Low	0 (0)	0 (0)

Frequency of Use

The frequency of use of the tablet computer and smartwatch in everyday life was heterogeneous across the study period. In the tablet computer group, 5 older adults reported varying use frequencies, which ranged between not using the technology at all (0× per week) and regularly using the technology (7× per week). Two participants of the tablet computer group stopped using the technology after the first week of the intervention. In the smartwatch group, 6 older adults used the technology consistently across all 8 intervention weeks (mostly 7× per week). In the smartwatch group, 2 participants stopped using the technology after the first 1 to 3 weeks. A total of 2 older adults (1 in each group) did not document their frequency of use.

Life Satisfaction

The mean life satisfaction of older adults changed nonsignificantly from 7.9 (SD 3.3) to 7.3 (SD 3.1) in the tablet computer group (t_7=1.1, *P*=.31, *d*=0.29) and from 7.2 (SD 2.1) to 8.1 (SD 1.9) in the smartwatch group (t_8=1.1, *P*=.29, *d*=0.44). Similarly, the mean life satisfaction of relatives changed nonsignificantly from 8.9 (SD 0.9) to 8.0 (SD 2.1) in the tablet computer group (t_6=1.5, *P*=.17, *d*=0.53) and from 7.7 (SD 1.4) to 8.1 (SD 0.7) in the smartwatch group (t_6=0.9, *P*=.41, *d*=0.39). In the smartwatch group, 2 participants only provided the preintervention (n=1) or postintervention (n=1) value and were not considered in the statistical analysis.

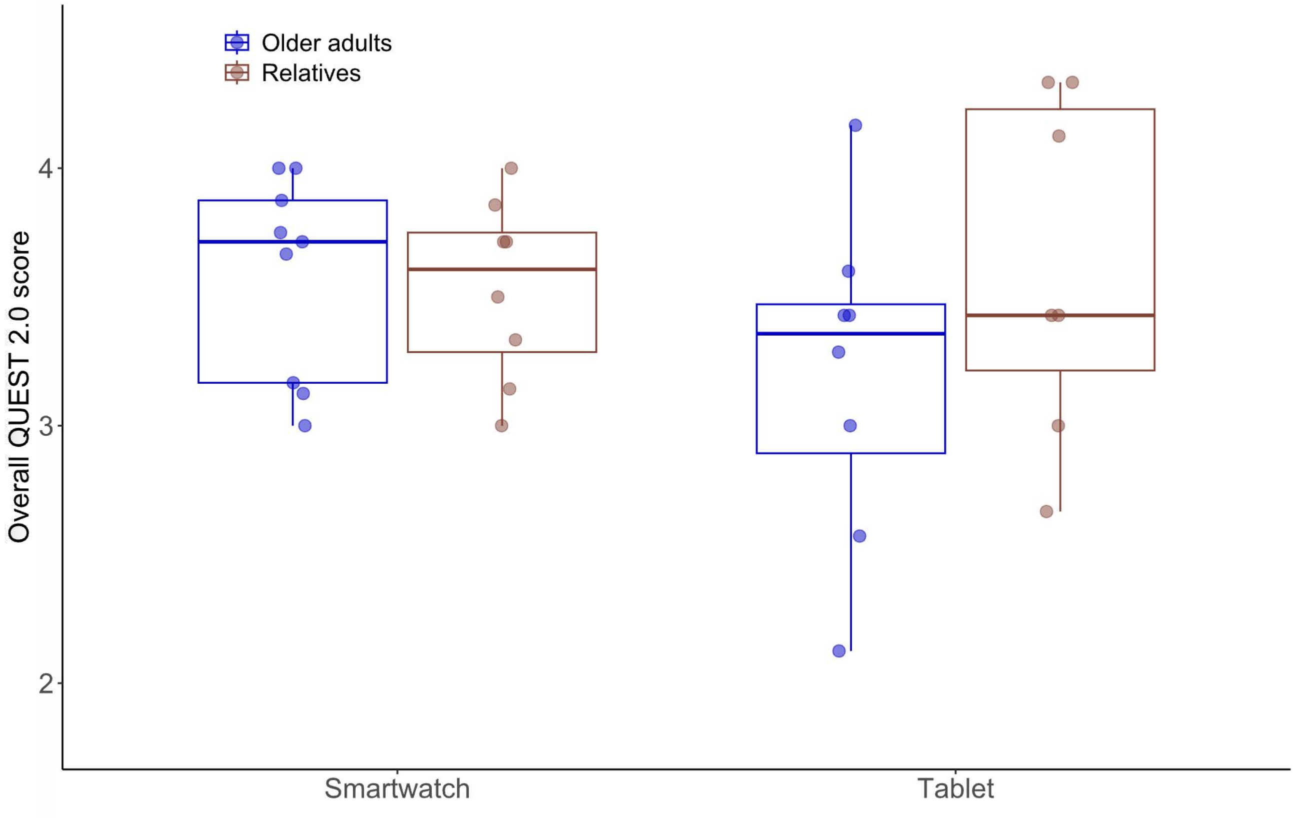
Technology Acceptance

Prior to the intervention, mean ratings for technology acceptance of older adults and their relatives ranged between 5.8 and 6.5 points for *fearfulness* and between 4.4 and 5.1 points for *curiosity* in the tablet computer and smartwatch group, respectively. For both items, the differences between the tablet computer and smartwatch groups were not significant (older adults: *fearfulness P*=.11 and *curiosity P*=.77; relatives: *fearfulness P*=.43 and *curiosity P*=.36). After the intervention, ratings from the older adults and their relatives of the tablet computer and smartwatch groups were similar for the items *interest*, *accessibility*, *usability*, and *skepticism* (older adults: *interest P*=.90, *accessibility P*=.64, *usability P*=.26, and *skepticism P*=.67; relatives: *interest P*=.65, *accessibility P*=.78, *usability P*=.64, and *skepticism P*=.39). This was different for the items *usefulness* and *intention to use*. TUI scores on the item *usefulness* were higher in the smartwatch group (older adults: mean 5.0, SD 1.0; relatives: mean 5.4, SD 1.4) than in the tablet computer group (older adults: mean 3.8, SD 1.5; relatives: mean 4.2, SD 1.6), although this did not reach statistical significance (older adults: $t_{12.3}$=1.9, *P*=.07, *d*=0.98; relatives: $t_{12.1}$=1.6, *P*=.14, *d*=0.83). TUI scores on the item *intention to use* were significantly higher in the smartwatch group (mean 6.1, SD 3.4) than in the tablet computer group (mean 2.7, SD 2.2; $t_{13.6}$=2.4, *P*=.03, *d*=1.14) in older adults (note that after correction for multiple comparisons, the *P* value exceeded .05). Although ratings on *intention to use* from the relatives were also higher in the smartwatch group (mean 7.8, SD 2.7) than in the tablet computer group (mean 5.9, SD 1.5), the differences were not significant ($t_{11.0}$=1.7, *P*=.11, *d*=0.86). Interestingly, in the tablet computer group, relatives' ratings on the item *intention to use* (mean 5.9, SD 1.5) were on average twice as high as those of older adults (mean 2.6, SD 2.2; $t_{12.3}$=3.3, *P*=.006, *d*=1.66). In contrast, the relatives' and older adults' ratings on the item *intention to use* were similar in the smartwatch group (older adults: mean 6.1, SD 3.4; relatives: mean 7.8, SD 2.7; $t_{14.8}$=1.1, *P*=.28, *d*=0.54). Descriptive data on all items of the TUI can be found in Multimedia Appendix 2.

Satisfaction With the Technology

The overall satisfaction score of the QUEST 2.0 across all items was similar for the older adults and relatives of both groups (Figure 1). The mean satisfaction score of older adults was 3.2 (SD 0.6) in the tablet computer group and 3.6 (SD 0.4) in the smartwatch group ($t_{11.4}$=1.5, *P*=.16, *d*=0.75). Similarly, the mean satisfaction across all items rated by the relatives was 3.6 (SD 0.7) in the tablet computer group and 3.5 (SD 0.4) in the smartwatch group ($t_{8.6}$=0.3, *P*=.77, *d*=0.16).

Figure 1. QUEST 2.0 score (scale from 1 to 5) of older adults and their relatives for the tablet computer and smartwatch groups. Higher values represent higher satisfaction. Dots represent data from individual participants. QUEST: Quebec User Evaluation of Satisfaction with Assistive Technology.



The 3 most relevant rated items for both AT groups were *safety and reliability* (n=4), *ease of use* (n=7), and *effectiveness* (n=7). Thus, these items were analyzed separately (Table 4). This analysis suggested that older adults of the smartwatch groups rated all 3 items better than older adults of the tablet computer group. Likewise, the relatives of the smartwatch group rated the items *safety and reliability* and *effectiveness* better than those of the tablet computer group. This was different for the item *ease of use*. This item was rated better by the relatives of the tablet computer group than those of the smartwatch group. In addition, item ratings of relatives were better than those of older adults in the tablet computer group. However, none of the abovementioned comparisons yielded significant test results (older adults: *safety and reliability* P=.06, *ease of use* P=.57, and *effectiveness* P=.40; relatives: *safety and reliability* P=.21, *ease of use* P=.40, and *effectiveness* P=.21).

Table . Results (mean and SD) of selected QUEST[a] 2.0 items (scale from 1 to 5) for the tablet computer and smartwatch groups. Higher values represent higher satisfaction.

QUEST 2.0 items	Tablet computer		Smartwatch	
	Older adults (n=8)	Relatives (n=7)	Older adults (n=9)	Relatives (n=9)
Safety and reliability, mean (SD)	2.8 (1.0)	3.2 (1.5)	3.7 (0.7)	3.5 (1.2)
Ease of use, mean (SD)	3.1 (1.4)	3.9 (1.2)	3.9 (0.8)	3.5 (1.2)
Effectiveness, mean (SD)	3.1 (1.4)	3.2 (1.5)	3.6 (1.5)	3.9 (1.6)

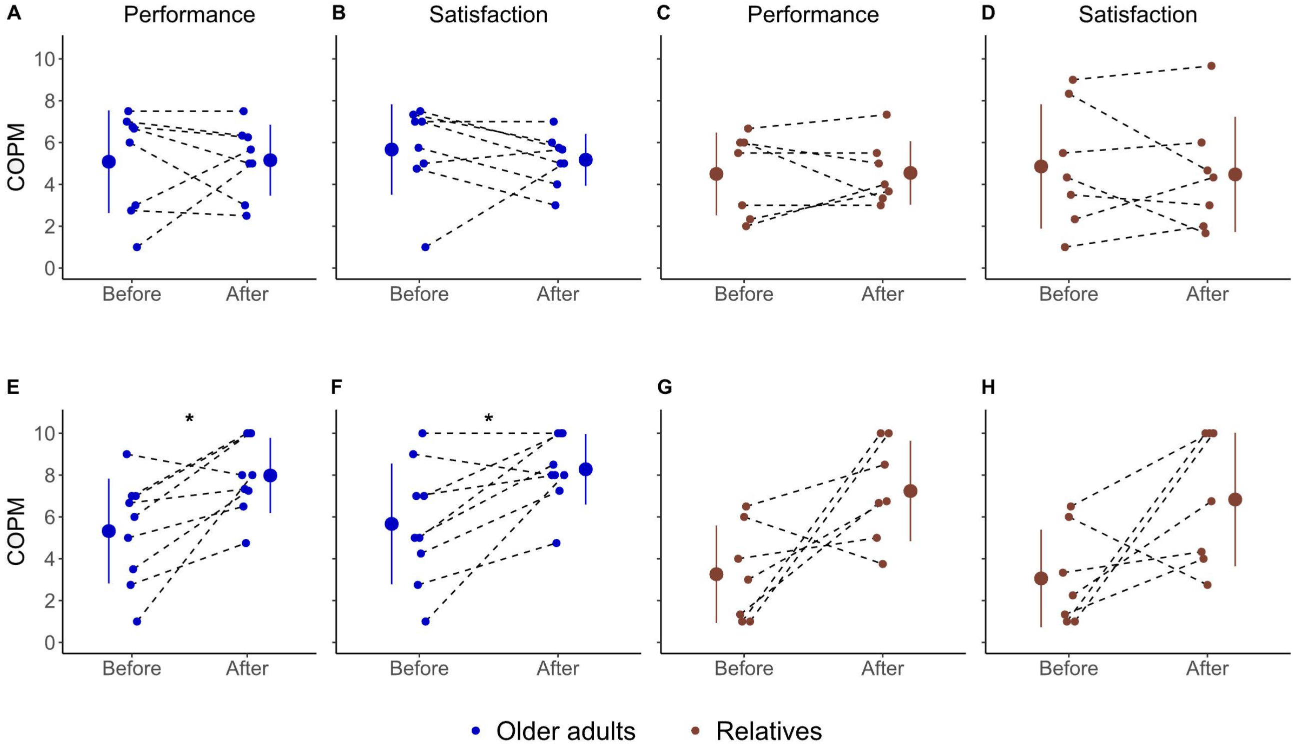
[a]QUEST: Quebec User Evaluation of Satisfaction with Assistive Technology.

Self-Perceived Performance and Satisfaction

The most frequently reported everyday issues identified in the COPM were contact/communication with relatives (tablet computer group), entertainment/information on news (tablet computer group), perception of safety (smartwatch group), getting help in emergency situations (smartwatch group), and usability of the AT (both groups). The number of specified issues ranged between 1 and 4.

In the tablet computer group, the older adults and their relatives reported diverse scores for self-perceived performance and satisfaction, with both improved and reduced scores. Statistical analysis showed that self-perceived performance (older adults: t_7=0.09, P=.93, d=0.03; relatives: t_6=0.08, P=.94, d=0.03) and satisfaction (older adults: t_7=0.6, P=.54, d=0.23; relatives: t_6=0.5, P=.64, d=0.18) did not significantly change in the tablet computer group during the intervention (Figure 2A-D). This was different in the smartwatch group. All but 1 older adult and 1 relative reported improved performance and satisfaction in the defined issues. Statistical analysis showed that self-perceived performance (t_8=3.5, *P*=.008, *d*=1.17) and satisfaction (t_8=3.2, *P*=.01, *d*=1.06) significantly increased in the older adults of the smartwatch group (Figure 2E-F). In the relatives of the smartwatch group, self-perceived performance (t_6=2.5, *P*=.04, *d*=0.96) and satisfaction (t_6=2.3, *P*=.06, *d*=0.87) increased, although this result was not statistically significant (after multiple comparison correction; Figure 2G-H).

Figure 2. COPM score (scale from 0 to 10) of older adults and their relatives for the (A-D) tablet computer and (E-H) smartwatch groups. Higher values represent higher self-perceived performance and satisfaction. Small dots represent data from individual participants. Bigger dots and error bars represent mean and SD. Asterisks (*) indicate significant differences after correction for multiple comparisons. COPM: Canadian Occupational Performance Measure.



Discussion

Principal Findings

This prospective, real-world, pilot study analyzed the perceived benefit and satisfaction with 2 types of AT for older adults and their relatives addressing different user needs (communication related vs safety related). The devices were (1) a tablet computer with a video call function and entertainment content for older adults and (2) a smartwatch with an emergency button and GPS tracking. Both older adults and their relatives in the smartwatch group consistently reported improved outcomes in the domains of life satisfaction and self-perceived performance and positive ratings for technology acceptance and satisfaction. Several participants emphasized feeling more secure when going out while wearing the device. In contrast, experiences with the tablet computer were more diverse, with several participants reporting poorer technology acceptance, satisfaction, and self-perceived performance.

Wearables with sensor-based risk assessment or fall detection and GPS locating can contribute to reducing older adults' fear of going outside and thus preserve participation in the community, autonomy, and mobility [35-37]. However, issues with device aesthetics, reliability, and ease of use can negatively impact device acceptance [35,38,39]. Most available studies that focus on the accuracy of device measurements are conducted within a laboratory setting. Contrary to our study, the main sensor location in existing studies is the waist or lower back [36,37]. However, it has been shown that a sensor location at the wrist (eg, watch or wristband) is an important feature for better acceptance of the device [40]. Another study found the highest satisfaction for a pendant worn on a key chain or around the neck [35]. Both forms have the advantage of being familiar or comfortable—aspects that are essential for technology adoption as stated by Fischer et al [41]. Although many emergency smartwatches for older adults have entered the market in recent years, only a few intervention studies with devices similar to the smartwatch tested here exist. Thus, the benefit and results for older adults and their caregivers remain inconclusive [38,42]. Future interventional studies in a real-world setting that take into account the wearers' wishes and focus on reliability, familiarity, and ease of use are needed.

Tablet computers for older adults are designed to improve participation, provide social support, and reduce loneliness or anxiety. Based on the existing research, the effectiveness of these systems remains inconclusive [5,13,43]. In our study, older adults in the tablet computer group reported on average reduced life satisfaction after the 8-week test period and worse scores for safety and reliability when compared to the smartwatch group. Several participants reported technical failures. Some of the issues mentioned included video calls that did not work, difficulties in receiving pictures from relatives, and defective charging. Considering the high value put on device reliability, these problems are part of the explanation for the negative results of the tablet computer. Additionally, the device is operated via a touchscreen, a technology older adults are mostly unfamiliar with. Consistently, the item ease of use was rated higher by the relatives in the tablet computer group, who belong to a generation already familiar with the use of a touchscreen [43].

Participants received no training prior to using the ATs in their daily lives. This procedure was chosen to create a situation that is as realistic as possible. In most cases, older adults purchasing ATs need to rely on the assistance of their relatives or the company support hotline [35]. In both groups, there were 2 participants who stopped using the ATs after the first few weeks of the intervention period, potentially due to a lack of adequate preparation or lack of support. Indeed, previous research indicates that older adults wish to receive additional training of the technology items they use in the home [44]. Thus, it might be necessary to incorporate training, particularly for more innovative ATs with reduced familiarity and potentially challenging features [38].

This study included 3 older adults characterized as vulnerable or prefrail (CFS score=4) and 4 with some level of frailty (CFS score>4). Physical frailty is a “medical syndrome with multiple causes and contributors that is characterized by diminished strength, endurance, and reduced physiologic function that increases an individual’s vulnerability for developing increased dependency and/or death.” (p. 4) [45]. Frail older adults face specific challenges when it comes to using and deriving benefits from ATs [46]. However, this subgroup is frequently underrepresented in intervention studies, thereby impeding a comprehensive understanding of their requirements [47]. In this study, all vulnerable or frail older adults reported a high likelihood for intention to use the ATs, indicating that this population is overall open to using ATs in their daily lives. Additionally, vulnerable or frail older adults had an average technology satisfaction score of 3.5 (SD 0.5), that is, their satisfaction was between “more or less satisfied” and “quite satisfied.”

As clearly deduced from the introductory WHO definition [1], the term AT is a very broad umbrella term, and it covers an extremely heterogeneous group of products. Each product targets users with specific needs and specific goals. Measuring such diversity is challenging within scientific research. The COPM is unique in the sense that it allows the analysis and quantification of individual user needs and specific AT aims. Thus, it can be adapted to different settings, targets, domains, and participants—irrespective of their age, sex, or other attributes. The tool has been used in other studies investigating the effectiveness of ATs [48,49]. The TUI and QUEST 2.0, however, include predefined items or questions that are not developed specifically for older adults. A recent study suggests that independence, affordability, ease of use, and ethics are the most important AT evaluation criteria for older adults [23]. It is possible that these domains are not adequately represented by the TUI and QUEST 2.0. Assessments focusing on older adults or specifically adapted to the needs and wishes of this group might be better suited in future studies.

Limitations

The 2 ATs studied target different domains (communication vs safety). Thus, the comparability and generalizability of results are limited. However, the used of the COPM allowed an item-specific evaluation according to its properties to help resolve the identified issues. Although the overall interest in the study was high, only 17 older adults and 16 relatives were included in this study. Thus, the statistical power of the study was relatively low, and further studies with larger samples are required to draw robust conclusions. Having the participation of a relative as a prerequisite for enrollment was one of the biggest obstacles for recruitment as they often declined to participate due to time restrictions. However, as many ATs require or support the interaction with relatives, we consider it important to investigate both perspectives, that of the older adults and that of the relatives. A large systematic review on ATs for older adults with dementia analyzed 571 studies and found that most investigations of clinical effectiveness were conducted with small sample sizes of <20 participants [50]. Conducting real-life intervention studies on the effectiveness of ATs requires a significant amount of effort and time. Combined with difficult recruitment among the population aged ≥65 years, this explains the low sample sizes. Unfortunately, we did not find any older adults–caregiver dyads who were willing to test the smart bed rail or the care planning app and able to participate in the study assessments. Both products are possibly of interest for frail older adults with a higher risk for fall or need for care. These individuals might be too sick to participate in a study such as this, live in some form of assisted living facilities or nursing homes, and be more difficult to reach. Older adults with no interest in technology did not participate in the study, causing a certain sample bias. The study was conducted with the device versions (including the software) that were current as of August 2021. The reproduction of the study using newer device versions could potentially change the results.

Conclusion

This prospective, real-world, pilot study confirmed the potential of ATs to support older adults and their relatives, especially for safety-related issues, but also highlighted the remaining obstacles for widespread use. Frail older adults and their relatives, who would potentially benefit the most from ATs, are especially difficult to reach. Thus, future research and technical developments of ATs should take into account the preferences, problems, and goals of older adults. In addition, this study highlights that individualized measures such as the COPM are necessary to identify the needs and assess the user benefits of the ATs in real-world applications.

Acknowledgments

This study was supported through funds from the German Federal Ministry of Education and Research for the project Future City 2030 (grant 13ZS0054A). The devices that were tested were graciously provided by the manufacturers for the duration of the study. Neither the funding party nor the manufacturers influenced the study design, data collection, analysis, interpretation, or the writing of the manuscript. The authors would also like to thank Sybille Beck for her support in performing the interviews during the study.

Authors' Contributions

MD, DD, SM, BK, and MLF conceived and designed the research. BK and SM were responsible for recruitment and data collection. BK, PW, MLF, and FMV analyzed the data. PW and MLF drafted the paper. All authors edited and revised the manuscript. All authors approved the final version of manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Technical Information on tablet and smartwatch.
[PDF File, 107 KB - humanfactors_v11i1e53811_app1.pdf]

Multimedia Appendix 2
Results (mean and SD) of the Technology Usage Inventory (TUI, scale from 1 to 7 for all items except for intention to use (scale from 1 to 10)) for the tablet and smartwatch group. Higher values represent better evaluations.
[PDF File, 99 KB - humanfactors_v11i1e53811_app2.pdf]

References

1. WHO global disability action plan 2014-2021 - better health for all people with disability. World Health Organization. 2015 Nov 30. URL: https://www.who.int/publications/i/item/who-global-disability-action-plan-2014-2021 [accessed 2024-02-16]
2. Agree EM, Freedman VA, Cornman JC, Wolf DA, Marcotte JE. Reconsidering substitution in long-term care: when does assistive technology take the place of personal care? J Gerontol B Psychol Sci Soc Sci 2005 Sep;60(5):S272-S280. [doi: 10.1093/geronb/60.5.s272] [Medline: 16131628]
3. Khoong EC, Olazo K, Rivadeneira NA, et al. Mobile health strategies for blood pressure self-management in urban populations with digital barriers: systematic review and meta-analyses. NPJ Digit Med 2021 Jul 22;4(1):114. [doi: 10.1038/s41746-021-00486-5] [Medline: 34294852]
4. Alharbi M, Straiton N, Smith S, Neubeck L, Gallagher R. Data management and wearables in older adults: a systematic review. Maturitas 2019 Jun;124:100-110. [doi: 10.1016/j.maturitas.2019.03.012] [Medline: 30910279]
5. Morgenstern LB, Adelman EE, Hughes R, Wing JJ, Lisabeth LD. The Women Independently Living Alone with a Medical Alert Device (WILMA) trial. Transl Stroke Res 2015 Oct;6(5):355-360. [doi: 10.1007/s12975-015-0411-0] [Medline: 26031786]
6. Sierra M SD, Garcia A DE, Otálora S, et al. Assessment of a robotic walker in older adults with Parkinson's disease in daily living activities. Front Neurorobot 2021 Dec 14;15:742281. [doi: 10.3389/fnbot.2021.742281] [Medline: 34970132]
7. Domenicucci R, Ferrandes F, Sarlo M, Borella E, Belacchi C. Efficacy of ICT-based interventions in improving psychological outcomes among older adults with MCI and dementia: a systematic review and meta-analysis. Ageing Res Rev 2022 Dec;82:101781. [doi: 10.1016/j.arr.2022.101781] [Medline: 36343879]
8. Moisseiev E, Mannis MJ. Evaluation of a portable artificial vision device among patients with low vision. JAMA Ophthalmol 2016 Jul 1;134(7):748-752. [doi: 10.1001/jamaophthalmol.2016.1000] [Medline: 27148909]
9. Tay JL, Xie HT. Novel interventions significantly reduce falls with fractures: a meta-analysis and systematic review. Geriatr Nurs 2023;52:181-190. [doi: 10.1016/j.gerinurse.2023.06.004] [Medline: 37390566]
10. Choi YYC, Fineberg M, Kassavou A. Effectiveness of remote interventions to improve medication adherence in patients after stroke: a systematic literature review and meta-analysis. Behav Sci (Basel) 2023 Mar 11;13(3):246. [doi: 10.3390/bs13030246] [Medline: 36975271]
11. Kang GCY. Technology-based interventions to improve adherence to antihypertensive medications – an evidence-based review. Digit Health 2022 May 3;8:20552076221089725. [doi: 10.1177/20552076221089725] [Medline: 35531090]
12. Rai HK, Kernaghan D, Schoonmade L, Egan KJ, Pot AM. Digital technologies to prevent social isolation and loneliness in dementia: a systematic review. J Alzheimers Dis 2022;90(2):513-528. [doi: 10.3233/JAD-220438] [Medline: 36120780]

13. Czaja SJ, Boot WR, Charness N, Rogers WA, Sharit J. Improving social support for older adults through technology: findings from the PRISM randomized controlled trial. Gerontologist 2018 May 8;58(3):467-477. [doi: 10.1093/geront/gnw249] [Medline: 28201730]
14. Ollevier A, Aguiar G, Palomino M, Simpelaere IS. How can technology support ageing in place in healthy older adults? a systematic review. Public Health Rev 2020 Nov 23;41(1):26. [doi: 10.1186/s40985-020-00143-4] [Medline: 33292707]
15. Khosravi P, Ghapanchi AH. Investigating the effectiveness of technologies applied to assist seniors: a systematic literature review. Int J Med Inform 2016 Jan;85(1):17-26. [doi: 10.1016/j.ijmedinf.2015.05.014] [Medline: 26216463]
16. Fotteler ML, Mühlbauer V, Brefka S, et al. The effectiveness of assistive technologies for older adults and the influence of frailty: systematic literature review of randomized controlled trials. JMIR Aging 2022 Apr 4;5(2):e31916. [doi: 10.2196/31916] [Medline: 35377322]
17. Marasinghe KM, Chaurasia A, Adil M, Liu QY, Nur TI, Oremus M. The impact of assistive devices on community-dwelling older adults and their informal caregivers: a systematic review. BMC Geriatr 2022 Nov 24;22(1):897. [doi: 10.1186/s12877-022-03557-8] [Medline: 36424533]
18. Madara Marasinghe K. Assistive technologies in reducing caregiver burden among informal caregivers of older adults: a systematic review. Disabil Rehabil Assist Technol 2016;11(5):353-360. [doi: 10.3109/17483107.2015.1087061] [Medline: 26371519]
19. Czaja SJ, Charness N, Fisk AD, et al. Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). Psychol Aging 2006 Jun;21(2):333-352. [doi: 10.1037/0882-7974.21.2.333] [Medline: 16768579]
20. Jaschinski C, Ben Allouch S, Peters O, Cachucho R, van Dijk JAGM. Acceptance of technologies for aging in place: a conceptual model. J Med Internet Res 2021 Mar 31;23(3):e22613. [doi: 10.2196/22613] [Medline: 33787505]
21. Mitzner TL, Boron JB, Fausset CB, et al. Older adults talk technology: technology usage and attitudes. Comput Human Behav 2010 Nov 1;26(6):1710-1721. [doi: 10.1016/j.chb.2010.06.020] [Medline: 20967133]
22. Schroeder T, Dodds L, Georgiou A, Gewald H, Siette J. Older adults and new technology: mapping review of the factors associated with older adults' intention to adopt digital technologies. JMIR Aging 2023 May 16;6:e44564. [doi: 10.2196/44564] [Medline: 37191976]
23. Kabacińska K, Vu K, Tam M, Edwards O, Miller WC, Robillard JM. "Functioning better is doing better": older adults' priorities for the evaluation of assistive technology. Assist Technol 2023 Jul 4;35(4):367-373. [doi: 10.1080/10400435.2022.2113180] [Medline: 35972791]
24. Moore K, O'Shea E, Kenny L, et al. Older adults' experiences with using wearable devices: qualitative systematic review and meta-synthesis. JMIR Mhealth Uhealth 2021 Jun 3;9(6):e23832. [doi: 10.2196/23832] [Medline: 34081020]
25. Der produktwettbewerb für innovative alltagshelfer. Daheim Dank Digital. URL: http://511394120.swh.strato-hosting.eu/Produktwettbewerb-2024/ [accessed 2024-02-16]
26. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist 1969;9(3):179-186. [Medline: 5349366]
27. Dufournet M, Moutet C, Achi S, et al. Proposition of a corrected measure of the Lawton instrumental activities of daily living score. BMC Geriatr 2021 Jan 12;21(1):39. [doi: 10.1186/s12877-020-01995-w] [Medline: 33430781]
28. Lubben J, Blozik E, Gillmann G, et al. Performance of an abbreviated version of the Lubben Social Network Scale among three European community-dwelling older adult populations. Gerontologist 2006 Aug;46(4):503-513. [doi: 10.1093/geront/46.4.503] [Medline: 16921004]
29. Rockwood K, Theou O. Using the Clinical Frailty Scale in allocating scarce health care resources. Can Geriatr J 2020 Sep 1;23(3):210-215. [doi: 10.5770/cgj.23.463] [Medline: 32904824]
30. Kothgassner OD, Felnhofer A, Hauk N, Kastenhofer E, Gomm J, Kryspin-Exner I. TUI: Technology Usage Inventory manual. Forschungsförderungsgesellschaft. 2012. URL: https://www.ffg.at/sites/default/files/allgemeine_downloads/thematische%20programme/programmdokumente/tui_manual.pdf [accessed 2023-07-28]
31. Demers L, Weiss-Lambrou R, Ska B. The Quebec User Evaluation of Satisfaction with Assistive Technology (QUEST 2.0): an overview and recent progress. Technol Disabil 2002 Sep 29;14(3):101-105. [doi: 10.3233/TAD-2002-14304]
32. Demers L, Monette M, Lapierre Y, Arnold DL, Wolfson C. Reliability, validity, and applicability of the Quebec User Evaluation of Satisfaction with Assistive Technology (QUEST 2.0) for adults with multiple sclerosis. Disabil Rehabil 2002;24(1-3):21-30. [doi: 10.1080/09638280110066352] [Medline: 11827151]
33. Law M, Baptiste S, McColl M, Opzoomer A, Polatajko H, Pollock N. The Canadian Occupational Performance Measure: an outcome measure for occupational therapy. Can J Occup Ther 1990 Apr;57(2):82-87. [doi: 10.1177/000841749005700207] [Medline: 10104738]
34. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 1995 Jan;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
35. Williamson B, Aplin T, de Jonge D, Goyne M. Tracking down a solution: exploring the acceptability and value of wearable GPS devices for older persons, individuals with a disability and their support persons. Disabil Rehabil Assist Technol 2017 Nov;12(8):822-831. [doi: 10.1080/17483107.2016.1272140] [Medline: 28100086]

36. Bezold J, Krell-Roesch J, Eckert T, Jekauc D, Woll A. Sensor-based fall risk assessment in older adults with or without cognitive impairment: a systematic review. Eur Rev Aging Phys Act 2021 Jul 9;18(1):15. [doi: 10.1186/s11556-021-00266-w] [Medline: 34243722]
37. Bet P, Castro PC, Ponti MA. Fall detection and fall risk assessment in older person using wearable sensors: a systematic review. Int J Med Inform 2019 Oct;130:103946. [doi: 10.1016/j.ijmedinf.2019.08.006] [Medline: 31450081]
38. Chung J, Brakey HR, Reeder B, Myers O, Demiris G. Community-dwelling older adults' acceptance of smartwatches for health and location tracking. Int J Older People Nurs 2023 Jan;18(1):e12490. [doi: 10.1111/opn.12490] [Medline: 35818900]
39. Hvalič-Touzery S, Šetinc M, Dolničar V. Benefits of a wearable activity tracker with safety features for older adults: an intervention study. Int J Environ Res Public Health 2022 Nov 25;19(23):36497796. [doi: 10.3390/ijerph192315723] [Medline: 36497796]
40. Rice LA, Fliflet A, Frechette M, et al. Insights on an automated fall detection device designed for older adult wheelchair and scooter users: a qualitative study. Disabil Health J 2022 Jan;15(1S):101207. [doi: 10.1016/j.dhjo.2021.101207] [Medline: 34503941]
41. Fischer SH, David D, Crotty BH, Dierks M, Safran C. Acceptance and use of health information technology by community-dwelling elders. Int J Med Inform 2014 Sep;83(9):624-635. [doi: 10.1016/j.ijmedinf.2014.06.005] [Medline: 24996581]
42. Megges H, Freiesleben SD, Rösch C, Knoll N, Wessel L, Peters O. User experience and clinical effectiveness with two wearable global positioning system devices in home dementia care. Alzheimers Dement (N Y) 2018 Nov 9;4:636-644. [doi: 10.1016/j.trci.2018.10.002] [Medline: 30519629]
43. Davison TE, Nayer K, Coxon S, et al. A personalized multimedia device to treat agitated behavior and improve mood in people with dementia: a pilot study. Geriatr Nurs 2016;37(1):25-29. [doi: 10.1016/j.gerinurse.2015.08.013] [Medline: 26412509]
44. Mitzner TL, Fausset CB, Boron JB, et al. Older adults' training preferences for learning to use technology. Proc Hum Factors Ergon Soc Annu Meet 2008 Sep;52(26):2047-2051. [doi: 10.1177/154193120805202603] [Medline: 25309139]
45. Morley JE, Vellas B, van Kan GA, et al. Frailty consensus: a call to action. J Am Med Dir Assoc 2013 Jun;14(6):392-397. [doi: 10.1016/j.jamda.2013.03.022] [Medline: 23764209]
46. Skymne C, Dahlin-Ivanoff S, Claesson L, Eklund K. Getting used to assistive devices: ambivalent experiences by frail elderly persons. Scand J Occup Ther 2012 Mar;19(2):194-203. [doi: 10.3109/11038128.2011.569757] [Medline: 21534712]
47. Fotteler M, Risch B, Gaugisch P, et al. Obstacles to using assistive technology for older adults – results from a focus group analysis. Stud Health Technol Inform 2021 May 27;281:994-998. [doi: 10.3233/SHTI210327] [Medline: 34042822]
48. Gordon S, Telford-Sharp F, Crowe W, Champion S. Effectiveness of a co-designed technology package on perceptions of safety in community-dwelling older adults. Australas J Ageing 2022 Sep;41(3):e257-e265. [doi: 10.1111/ajag.13095] [Medline: 35635481]
49. Kjeken I, Darre S, Smedslund G, Hagen KB, Nossum R. Effect of assistive technology in hand osteoarthritis: a randomised controlled trial. Ann Rheum Dis 2011 Aug;70(8):1447-1452. [doi: 10.1136/ard.2010.148668] [Medline: 21571733]
50. Ienca M, Fabrice J, Elger B, et al. Intelligent assistive technology for Alzheimer's disease and other dementias: a systematic review. J Alzheimers Dis 2017;56(4):1301-1340. [doi: 10.3233/JAD-161037] [Medline: 28222516]

Abbreviations

AT: assistive technology
CFS: Clinical Frailty Scale
COPM: Canadian Occupational Performance Measure
IADL: instrumental activities of daily living
QUEST: Quebec User Evaluation of Satisfaction with Assistive Technology
TUI: Technology Usage Inventory
WHO: World Health Organization

Edited by Y Gong; submitted 19.10.23; peer-reviewed by C Baxter, C Becker; revised version received 08.03.24; accepted 07.06.24; published 02.08.24.

Please cite as:
Wiegel P, Fotteler ML, Kohn B, Mayer S, Verri FM, Dallmeier D, Denkinger M
Perceived Benefit and Satisfaction With a Tablet Computer and an Emergency Smartwatch by Older Adults and Their Relatives: Prospective Real-World Pilot Study
JMIR Hum Factors 2024;11:e53811
URL: https://humanfactors.jmir.org/2024/1/e53811
doi:10.2196/53811

Original Paper

User Perceptions of Visual Clot in a High-Fidelity Simulation Study: Mixed Qualitative-Quantitative Study

Greta Gasciauskaite[1], MD; Clara Castellucci[1], MD; Amos Malorgio[1], MD; Alexandra D Budowski[1], MD; Giovanna Schweiger[1], MD; Michaela Kolbe[2], Dr rer nat; Daniel Fries[1], MD; Bastian Grande[1], MD; Christoph B Nöthiger[1], MD; Donat R Spahn[1], MD; Tadzio R Roche[1], MD; David W Tscholl[1], MD; Samira Akbas[1], MD

[1]Institute of Anesthesiology, University Hospital Zurich, Zurich, Switzerland

[2]Simulation Centre, University Hospital Zurich, Zurich, Switzerland

Corresponding Author:
Samira Akbas, MD
Institute of Anesthesiology
University Hospital Zurich
Raemistrasse 100
Zurich, 8091
Switzerland
Phone: 41 432532242
Email: Samira.Akbas@usz.ch

Abstract

Background: Viscoelastic hemostatic assays, such as rotational thromboelastometry (ROTEM) or thromboelastography, enable prompt diagnosis and accelerate targeted treatment. However, the complex interpretation of the results remains challenging. Visual Clot—a situation awareness-based visualization technology—was developed to assist clinicians in interpreting viscoelastic tests.

Objective: Following a previous high-fidelity simulation study, we analyzed users' perceptions of the technology, to identify its strengths and limitations from clinicians' perspectives.

Methods: This is a mixed qualitative-quantitative study consisting of interviews and a survey. After solving coagulation scenarios using Visual Clot in high-fidelity simulations, we interviewed anesthesia personnel about the perceived advantages and disadvantages of the new tool. We used a template approach to identify dominant themes in interview responses. From these themes, we defined 5 statements, which were then rated on Likert scales in a questionnaire.

Results: We interviewed 77 participants and 23 completed the survey. We identified 9 frequently mentioned topics by analyzing the interview responses. The most common themes were "positive design features," "intuitive and easy to learn," and "lack of a quantitative component." In the survey, 21 respondents agreed that Visual Clot is easy to learn and 16 respondents stated that a combination of Visual Clot and ROTEM would help them manage complex hemostatic situations.

Conclusions: A group of anesthesia care providers found Visual Clot well-designed, intuitive, and easy to learn. Participants highlighted its usefulness in emergencies, especially for clinicians inexperienced in coagulation management. However, the lack of quantitative information is an area for improvement.

(JMIR Hum Factors 2024;11:e47991) doi:10.2196/47991

KEYWORDS

Visual Clot; blood coagulation; blood coagulation test; hemostasis; rotational thromboelastometry; situation awareness; user-centered design; visualization; user; perception; interpretation; thromboelastography; viscoelastic hemostatic; technology; coagulation; quantitative information

Introduction

Rapid hemostatic assessment is essential to targeted coagulation management in acute bleeding [1]. Increasingly used viscoelastic hemostatic assays, such as rotational thromboelastometry (ROTEM) or thromboelastography, enable faster insights into coagulation dysfunction than conventional laboratory tests. Standard coagulation assays are not optimal for managing acute hemorrhages that require rapid therapeutic action, as it often takes more than an hour to obtain the results [2,3]. European and North American transfusion recommendations underline the advantages of viscoelastic hemostatic assays for managing trauma and severe perioperative bleeding, including the reduced need for transfusions, fewer perioperative complications, shorter hospitalization, and lower overall treatment costs [1,4,5]. Its usefulness has been demonstrated in many operative areas such as obstetrics [6,7], pediatric surgery [8], transplantation [9], cardiac surgery [10,11], neurosurgery [12], and burn surgery [13,14]. Viscoelastic hemostatic tests are also paramount in the diagnostic and treatment adjustment of hematological disorders, such as inherited afibrinogenemia, hemophilia, or multiple myeloma [15-18]. However, despite these technologies' widespread use and considerable advantages, their results' interpretation remains challenging and requires well-trained clinical personnel [19-21]. Visual Clot—a situation awareness-based visualization technology—was developed to support health care professionals in interpreting viscoelastic test results by reducing the complexity of their presentation. Based on raw ROTEM data, the results are displayed in real time as a 3D animated model of a blood clot to represent various elements of hemostasis, including platelets, plasmatic factors, and fibrin. It can also effectively illustrate the influence of heparin and hyperfibrinolysis [22] (Multimedia Appendix 1). In a high-fidelity simulation study, anesthesiologists using Visual Clot were 2.2 times more likely to articulate the correct therapeutic approach. In addition, these anesthesiologists had a lower median time to administer the first appropriate targeted coagulation product. Overall, physicians presented with the results of viscoelastic testing using Visual Clot were approximately 56% more likely to provide accurate therapeutic interventions. In the same study, physicians were 3.5 times more likely to feel confident in their decisions when working with Visual Clot compared to traditional ROTEM results [23]. In the first computer-based study analyzing user perceptions of Visual Clot, participants described the technology as well-designed, easy to learn, and intuitive [24]. The guiding principles of the Visual Clot technology that result in enhanced situation awareness include Endsley's user-centered design principles [25], Wittgenstein's philosophy as articulated in Tractatus Logico-Philosophicus [26], and insights from the National Aeronautics and Space Administration (NASA) publication "On Organization of Information: Approach and Early Work" by Degani et al [27]. Endsley's principles emphasize the use of direct visual representations of data to enhance situational awareness, a central principle in Visual Clot's data visualization. Wittgenstein's theory emphasizes the importance of logical representations that meaningfully correspond to the reality they are intended to represent. Visual Clot follows this principle by visually representing elements such as fibrin, platelets, plasmatic factors, hyperfibrinolysis, and bleeding. Following NASA's approach, Visual Clot strives to achieve the highest level of "order and wholeness" by consolidating all essential data into a single display. The primary goal of Visual Clot technology is to provide the care provider with situational information quickly and with minimal cognitive load. In this study, we aimed to capture and analyze perceptions of anesthesia personnel working with Visual Clot in a high-fidelity simulation to identify the strengths and recognize the potential for future improvements.

Methods

Ethical Considerations

The Cantonal Ethics Committee of the Canton of Zurich reviewed the study protocol and issued a declaration of no objection (Business Management System for Ethics Committees Number Req-2021-01112). Furthermore, each participant gave informed consent to use his or her data for research purposes. Participation was voluntary and without financial compensation.

Study Design

We conducted a researcher-initiated single-center mixed qualitative-quantitative study at the University Hospital Zurich, Institute of Anesthesiology, Switzerland. Study participants were anesthesia personnel, including staff anesthesiologists, residents, and nurses. After participating in a high-fidelity simulation study of perioperative bleeding scenarios, where they worked with Visual Clot and ROTEM, we interviewed participants on their perceptions of Visual Clot technology.

As a second step, the same participants received an email invitation to participate in a survey a few weeks later. They rated statements we generated from identified and frequently mentioned themes in interview responses on a Likert scale.

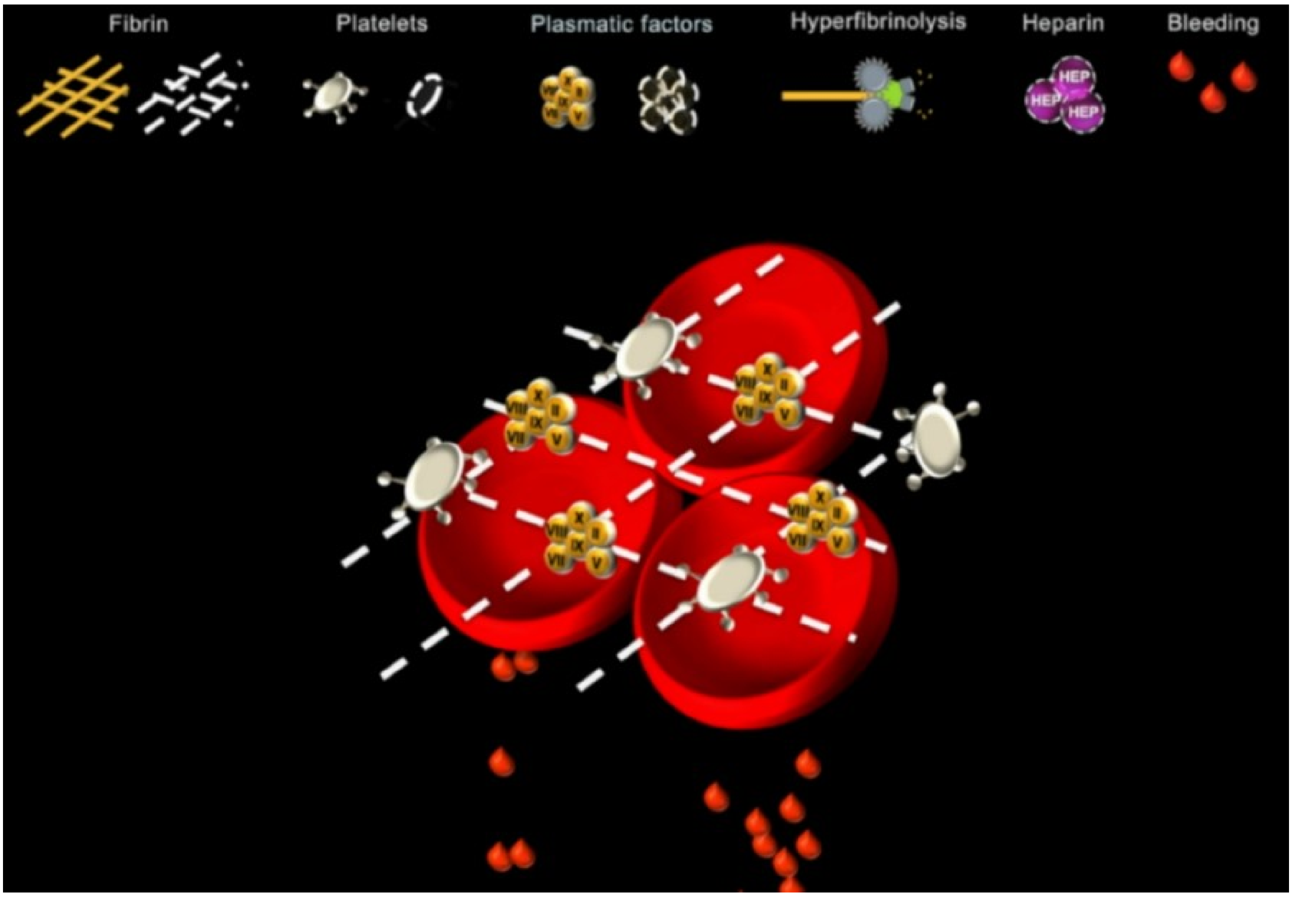
Previous High-Fidelity Simulation Study

In the high-fidelity simulation study [23], anesthesia teams, composed of a staff anesthesiologist, a resident, and an anesthesia nurse, participated in high-fidelity perioperative bleeding scenarios using either Visual Clot or ROTEM. The primary outcome of the study was correct targeted coagulation therapy. Secondary outcomes were time to targeted coagulation therapy, confidence, and workload.

ROTEM is the standard of care for managing acute hemorrhage in the study center, so all participants were familiar with the technology before participating [20]. Some participants had taken part in previous Visual Clot studies and, therefore, were already familiar with the technology [22,24].

Nevertheless, before the simulations began, we gave a 10-minute presentation that reviewed ROTEM and introduced Visual Clot. Multimedia Appendix 1 provides an instructional video of Visual Clot. Participants were invited to ask questions freely before starting work in the simulation environment. Each team solved 1 of 4 different perioperative bleeding scenarios, which were randomly allocated. We ended the scenarios when all necessary therapeutic measures were derived or, at the latest, after 15 minutes. Figure 1 illustrates an example of a Visual Clot printout used in the simulation study.

Figure 1. A Visual Clot result presentation showing a fibrin deficiency. The fibrin in the clot is shown as a dashed line, indicating its absence. The blood drops indicate the presence of a coagulation pathology.



Participant Interviews

After the simulations, we encouraged the participants to freely verbalize their thoughts in a distraction-free environment while the data collectors made field notes. The only suggestion to the participants before the interviews was to verbalize their positive and negative opinions of Visual Clot. The participants could define final adjustments in the collected answers at the end of the interviews.

Survey

In the second step, we formulated 5 statements to summarize the insights gathered during the interviews. The statements were submitted for evaluation on a 5-point Likert scale graded from "strongly agree" to "strongly disagree." An email invitation was sent to all interviewed participants.

Outcomes and Statistical Analyses

Part I: Participant Interviews

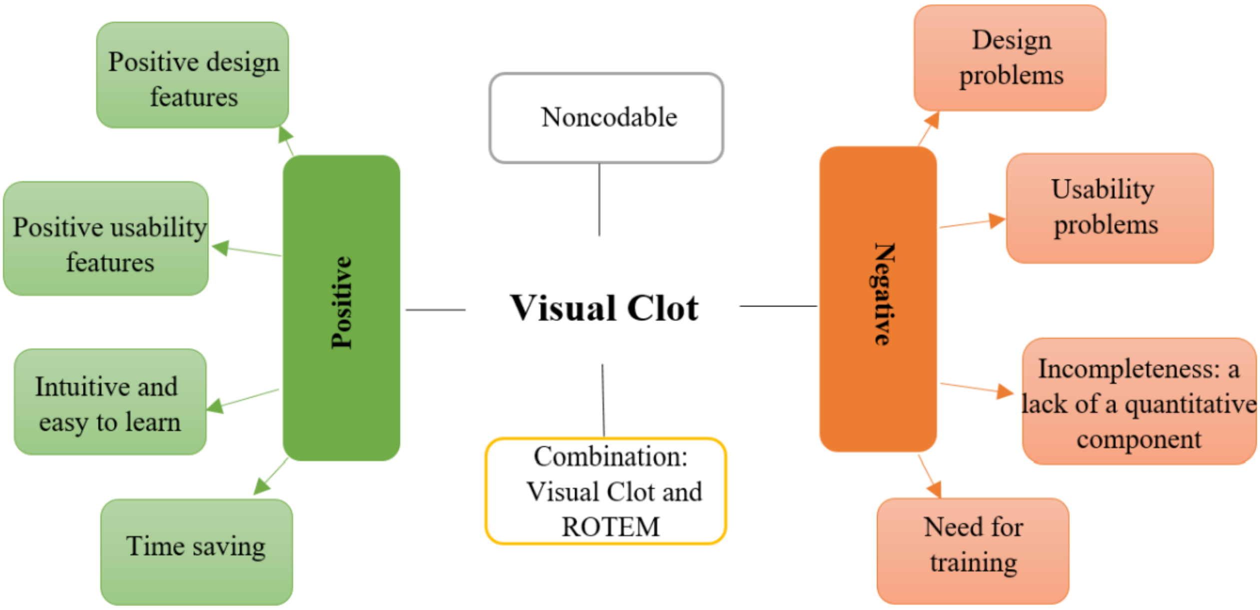
Collected interview responses were translated from original German to English using a translation system DeepL (DeepL GmbH). Multimedia Appendix 2 provides the complete translated field notes.

The most commonly used terms in positive and negative responses were identified using the word count function. Word groups with the same root were united, excluding the frequently used filler words such as "to," "and," or "the" (Table 1). Using a template approach [28] we identified the major themes that dominated participants' answers. As a result, we generated a coding tree (Figure 2). According to the coding template, we assigned statements to the themes. A total of 3 of the study authors, all anesthesiology residents GG, GS, and SA, rated the interview statements separately from each other using the coding tree (Figure 2). If the 3 investigators disagreed after multiple data coding, the final decision was taken in a joint discussion. Interrater reliability was calculated to investigate the consistency of the coding tree's application.

Table 1. The most commonly used positive and negative terms to describe Visual Clot.

Positive terms		Negative terms	
Terms	Frequency, n	Terms	Frequency, n
Easy or easier	28	Missing	13
Good	16	Quantitative or quantification	11
Fast or faster	13	Information	9
See a problem	12	Values	8
Interpret quickly	11	Time	8
Simple	9	Fibrinogen	6
Understand	9	Numbers	6
Visual	8	Hyperfibrinolysis	6
Interpretation	7	Less	5
Intuitive	7	Confusing	4
At a glance	7	Flashing	4
Overview	7	Simplified	3

Figure 2. A coding tree representing the themes describing positive and negative user perceptions. ROTEM: rotational thromboelastometry.



Part II: Survey

The literature states that quantitative data can help generalize and confirm specific observations found in qualitative research [29-32]. For the subsequent survey, we defined 5 statements based on the previously identified themes. The same group of interviewed anesthesiologists was asked to rate them on 5-point Likert scales in a questionnaire created using Google Forms (Alphabet Inc). Participants were informed that the survey takes only a few minutes to complete, participation is voluntary, and no compensation is offered. The translated announcement of the survey invitation is displayed in Multimedia Appendix 3. The data collection was finished 3 weeks after the questionnaire was sent.

Statistical Analysis

The interview data analysis and figures were made using Microsoft Word and Excel (Microsoft Corp). We present the number of statements and their percentage distribution in the identified themes.

To define the interrater reliability of the coding template, we calculated Fleiss' Kappa using R (version 4.0.5; R Foundation for Statistical Computing). We calculated every statement's median and IQR for the survey analysis. We used the Wilcoxon signed rank test to determine the difference between the median and neutral answers. Statistical significance was indicated as $P<.05$.

Results

Study and Participant Characteristics

Detailed information on the study and participants is provided in Table 2. Residents and nurses were the dominant participants in the interviews. The most experienced participant had 33 years of experience in anesthesia. The least experienced had less than 1 year. Residents and nurses also dominated the survey.

Table 2. Study and participant characteristics.

Characteristics	Values
Study characteristics	
Total number of interviewed participants, n	77
Total number of participants completed the survey, n	23
Participant characteristics	
Interview participants	
Staff physicians, n (%)	8 (10.4)
Residents, n (%)	35 (45.5)
Anesthesiology nurses, n (%)	34 (44.2)
Anesthesia experience in years, median (IQR)	8 (3-10)
Number of ROTEM[a] interpretations per year, median (IQR)	26 (5-41)
Survey participants	
Staff physicians, n (%)	7 (30.4)
Residents, n (%)	8 (34.8)
Anesthesiology nurses, n (%)	8 (34.8)

[a]ROTEM: rotational thromboelastometry.

Part I: Qualitative Analysis of Interview Answers

Word Count Analysis

The most frequently used words and word combinations used to describe the advantages of Visual Clot were: easier or easy (26/77, 33.8%), interpret or interpretation (23.4%, 18/77 participants), quick or quickly (19.5%, 15/77 participants), visual, visualize, visualized, or visualization (19.5%, 15/77 participants), good (16.9%, 13/77 participants), faster or fast (15.6%, 12/77 participants). In the group of statements describing the limitations of Visual Clot, the words and word groups most frequently used were: ROTEM (23.4%, 18/77 participants), missing (information or values or numbers; 16.9%, 13/77 participants), quantitative or quantification (16.9%, 13/77 participants). Table 1 visually represents the most commonly used words in positive and negative perceptions.

Coding Tree

Figure 2 shows the generated coding tree, including 2 main domains and 9 themes. The interrater reliability of the tree raters was 0.856 (95% CI 0.831-0.880), indicating almost perfect agreement [33].

Statements Describing Visual Clot

Table 3 demonstrates examples of statements assigned to particular subtopics with participant counts and percentages.

A total of 4 comments were defined as positive but not assigned to any themes. There was 1 such statement in the negative group. A total of 19 comments were not assigned to any theme and were described as noncodable.

Table 3. Statements examples assigned to particular domain and subtopics with participant count and percentages.

Major domain and subtopics	Examples
Positive statements describing Visual Clot (179/319, 56.1%)	
Positive design features (63/295, 21.4%)	• "Very simplified" (Participant 13). • "Visual presentation" (Participant 21). • "Tells figuratively what to do" (Participant 26).
Positive usability features (30/295, 10.2%)	• "It can be perfectly integrated in the clinic" (Participant 24). • "A good tool to get an overview" (Participant 32). • "You can see at-a-glance what is missing" (Participant 70).
Intuitive and easy to learn (61/295, 20.7%)	• "Directly applicable and does not require long training" (Participant 5). • "Very intuitive" (Participant 75). • "Easy to understand" (Participant 76).
Time saving (21/295, 7.1%)	• "2-3 seconds to a quick overview" (Participant 22). • "Immediate detection of the problem" (Participant 5). • "Quick to interpret" (Participant 27).
Negative statements describing Visual Clot (113/319, 35.4%)	
Design problems (23/295, 7.8%)	• "Confusing, blinking" (Participant 57). • "Quality of hyperfibrinolysis difficult to demonstrate" (Participant 22). • "Not everything you see is relevant" (Participant 24).
Usability problems (12/295, 4.1%)	• "Quantity of change is not visible" (Participant 2). • "You have to know the pictures first" (Participant 30). • "No prioritization possible" (Participant 76).
Incompleteness: a lack of a quantitative component (62/295, 21.0%)	• "Also values that are in normal range—is it close to the limit or not?" (Participant 54). • "No graduation 'all or nothing'" (Participant 60). • "Exact quantification not possible" (Participant 4).
Need for training (15/295, 5.1%)	• "Not yet established" (Participant 29). • "Needs to get used to it" (Participant 45). • "Needs habituation (not used by default yet)" (Participant 50).
Combination: Visual Clot and ROTEM[a] (8/295, 2.7%)	
	• "Would be nice to have it together with ROTEM" (Participant 50). • "Ideal solution if could be combined with ROTEM" (Participant 49). • "Combination of both necessary" (Participant 74).

[a]ROTEM: rotational thromboelastometry.

Positive Statements Describing Visual Clot

Positive Design Features

Most comments were made on this topic, emphasizing that a "pictorially summarized" (participant 2) and "visually appealing" (participant 3) data presentation allows one to see "the relevant ROTEM information at-a-glance" (participant 9). Such a design supports health care professionals in making clinical decisions. It is essential in emergencies because the actual coagulation status is immediately visible (participant 14) and it is instantly apparent which hemostasis components are missing (participant 12).

Positive Usability Features

Visual Clot is "a good tool for broad application," stated participant 72. It enables "pre-interpretation of the complex information" (participant 3) and focuses "on the essential" (participant 65). The benefits of the Visual Clot in urgent situations were also highlighted: the technology is "very good for emergencies," stated participant 13.

Intuitive and Easy to Learn

As in the previous study [24], the Visual Clot was also described here as intuitive and easy to learn. "Very intuitive, short time needed to understand it,"—pointed out participant 6. It was underlined that visualizations provided by the Visual Clot are "quickly recognizable even by untrained persons or with little knowledge of coagulation" (participant 13).

Time Saving

The Visual Clot provides an "overview at-a-glance," as participant 76 said. "I immediately saw what was missing," stated participant 16. These features lead to quicker diagnosis—"focus is faster on the problem"—as participant 46 said, and thus to faster initiation of treatment.

Negative Statements Describing Visual Clot

Design Problems

Several ideas that could potentially enhance Visual Clot's design were identified. Participant 47 pointed out that the presentation of platelets and fibrinogen are similar, and thus it is difficult to distinguish. Participant 48 also agreed: "I did not notice that platelets were missing because it was white and dashed like fibrinogen." Some participants found that the Visual Clot is too dynamic—too much movement on the screen, which can lead to distraction and make the interpretation of the results difficult "Even if coagulation status is fine, everything is moving, and you can poorly differentiate what is missing" (participant 53), "moves too much, even if everything is fine—distraction" (participant 55).

Usability Problems

Visual Clot is "confusing at the beginning"—stated participant 31 and added that it is "difficult to use without routine." Visual Clot provides "too much information at once,"—participant 53 pointed out.

Incompleteness: Lack of a Quantitative Component

The central Visual Clot aspect criticized was the technology's incompleteness in terms of lacking a quantitative component. Several participants stated that the Visual Clot is "not precise" (participant 1), which can be explained in the words of participants 53 and 9, respectively, who said that in the Visual Clot "quantitative is missing" and that one "can get more information with the ROTEM."

Need for Training

The main point identified in the participants' opinions on this topic was the lack of experience working with this technology and that it is a very new tool not yet established in clinical practice.

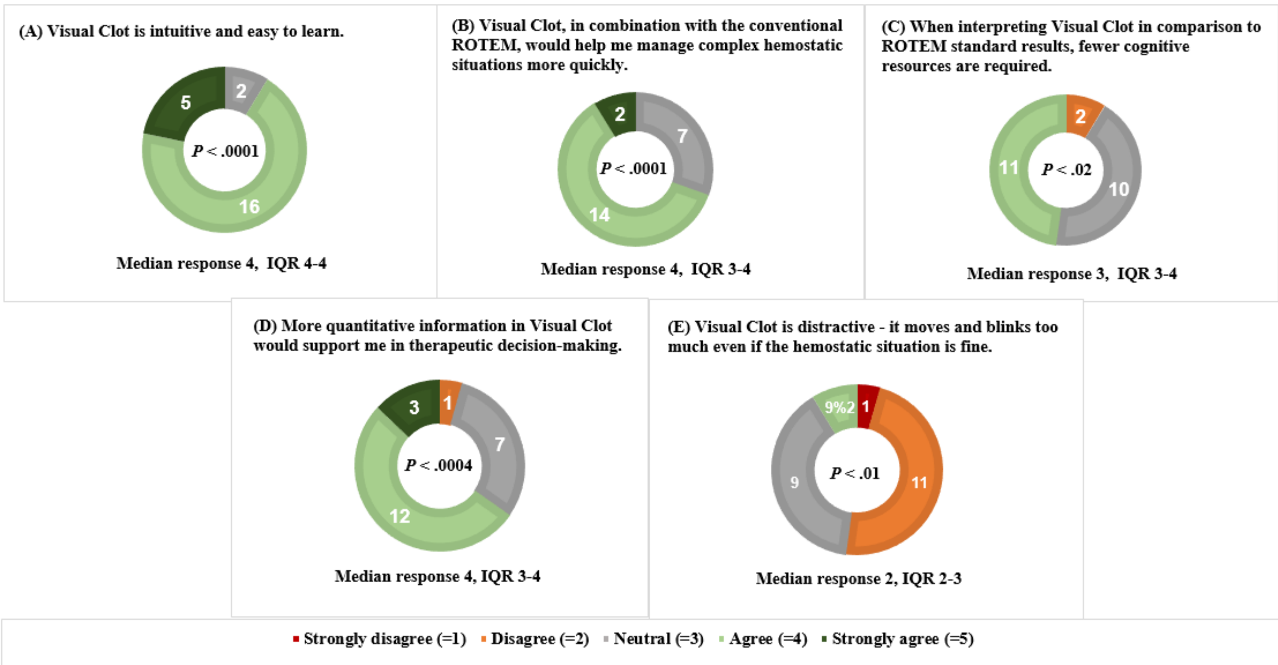
Combination of Visual Clot and ROTEM

Several participants said they could benefit from combining the Visual Clot and ROTEM when interpreting coagulation assays. "A combination of Visual Clot and ROTEM would be perfect," pointed out participant 19, while participant 74 said, "a combination of both is necessary." There was no difference in positive and negative statements based on participants' specialty or level of experience.

Part II: Analysis of Statements Assessed in the Survey

Figure 3 shows the detailed evaluation of the statements rated in the survey.

Figure 3. Pie charts presenting survey results with the number of participants who chose a particular category (N=23). ROTEM: rotational thromboelastometry.



All sample medians differed statistically significantly from neutral (P<.05). The number of participants in the quantitative part of the evaluation differs from the qualitative part because not all participants completed the questionnaire. The results are presented as medians and IQR. P values are provided to indicate a statistically significant difference between the median of the sample and the neutral value.

Discussion

Principal Findings

This mixed qualitative-quantitative study analyzed the perceptions of anesthesiology personnel regarding Visual Clot—a new situation-awareness and user-oriented visualization technology for viscoelastic hemostatic resuscitation—after the high-fidelity simulation study. User perceptions enable us to identify the positive aspects of the technology and reveal the potential for improvement in the future. After computer-based studies, this is the first time that Visual Clot has been evaluated in a high-fidelity simulation study, a validated process for testing a noncertified product in an environment that closely resembles clinical reality [34,35].

The principal findings demonstrate that the design features of Visual Clot have received the most positive comments. As in the previous computer-based Visual Clot study [24], the participants of this high-fidelity simulation study emphasized that the way this technology is designed provides a good overview of the clotting situation and is an additional help in the decision-making process during acute bleeding situations. Further, Visual Clot was described as intuitive and easy to learn. Participants repeatedly mentioned that the results of Visual Clot are quickly recognizable and understandable even by inexperienced clinicians. The main criticism concerned the lack of quantitative information.

Previous Visual Clot studies [22-24,36,37] underline the benefits of additional visualization technology, simplifying standard ROTEM data interpretation [20]. Anesthesia providers using Visual Clot in a high-fidelity simulation study were more likely to correctly administer targeted coagulation therapy and to give the first targeted coagulation product faster. In addition, participants demonstrated greater decision-making confidence with Visual Clot [23]. Moreover, the correctness of the clinical decisions was independent of previous rotational thromboelastometry knowledge and experience.

The superior participants' performance when working with Visual Clot may be explained by its design supporting the strengths of human sensory perception. The Visual Clot was developed to assist care providers in managing highly complex coagulation situations, presenting the data in an awareness-oriented interface design. The main aim of this design is to convey the information as quickly as possible and with the lowest cognitive effort [25].

Principles of situation awareness-oriented and user-centered design enables effective data management and a comprehensive understanding of what is happening and thus help to stay situationally aware. This concept is essential in many domains, including medicine, where managing complex and dynamic situations is fundamental [38,39].

Its definition breaks down into three separate phases (1) the perception of environmental elements in the current situation within a volume of time and space, (2) understanding their meaning, and (3) their projection in the near future. Based on this, the Visual Clot data are visually represented, preprocessed, and simplified. The results of coagulation parameters are divided into 3 categories: too low, normal, or too high. Such information presentation increases diagnostic confidence, but numeric indicators are needed for precise data analysis and targeted treatment initiation. As previously indicated, the lack of quantitative information is reflected in user responses. It also explains the participants' considerations that combining Visual Clot and ROTEM would be helpful in clinical decision-making.

Some other technologies based on situation awareness and user-centered design principles include Philips Visual Patient Avatar (Philips) [40], AlertWatch (AlertWatch Inc), Dynamic Lung Panel and PulmoSight (Hamilton Medical AG), HemoSight and Physiology Screen (Mindray Medical International Limited), and Alarm Status Visualizer (Masimo Corp) [41,42].

This study showed user perceptions regarding the new situation awareness-based, user-oriented technology for thromboelastometry data presentation—Visual Clot. It makes us aware of the user's needs and could help us simplify information processing and decision-making in the future. An integral facet of advancing the technological framework informed by the results of this study lies in the prospect of merging quantitative data into the Visual Clot platform and presenting this merged information in a consolidated interface. This concerted integration promises to align both quantitative and qualitative data to provide a more complete and accurate representation of prevailing conditions. This integration can be achieved in a variety of ways, including the direct overlay of numerical values onto the Visual Clot visualization, or the parallel juxtaposition of a complementary graphical representation alongside the numerical data set.

Strengths and Limitations

This study has several strengths and limitations. The interview part of the study has the typical limitations of qualitative research. The findings of qualitative analysis cannot be extrapolated to larger populations with the same certainty as quantitative results because the findings are on the subjective basis and not tested for statistical significance [43]. However, the quantitative survey helped to provide greater insight into the importance of the main themes identified. Moreover, the interviewed participants were selected according to their availability in the clinical praxis and not randomly.

Furthermore, the number of participants in the survey was lower than in the interviews because not all participants in the simulation study completed the survey. Finally, it is a single-center study performed in a university hospital with high care standards in Europe. User perceptions may vary across diverse clinical settings in different parts of the world.

Conclusions

After previous studies investigating user perceptions of Visual Clot in computer-based simulation studies, this is the first study to analyze the user perceptions of Visual Clot in a high-fidelity simulation—the intermediate step between computer-based simulation studies in a laboratory and real-life use. In this study, Visual Clot appeared to be a well-accepted additional tool supporting health care professionals working with ROTEM. Based on participants' perceptions, user-centered and situation awareness-oriented design, as shown in Visual Clot, can simplify the presentation of complex information and thus make critical decision-making quicker and more efficient. The benefits of this technology have been particularly highlighted in emergencies and even for care providers with little experience in coagulation management. Participants described Visual Clot as intuitive and easy to learn. The lack of a quantitative component has been identified as a significant limitation. These findings highlight the advantages of Visual Clot and its potential for improvement may help further develop this and other situation awareness-based technologies.

Acknowledgments

The authors are thankful to the study participants for their time and effort. The Institute of Anesthesiology of the University Hospital of Zurich, Zurich, Switzerland and the University of Zurich, Zurich, Switzerland funded this study.

Data Availability

The data sets used and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

GG, AM, CC, ADB, GS, MK, BG, CBN, TRR, DWT, and SA contributed to the conceptualization; GG, AM, CC, ADB, GS, MK, BG, CBN, TRR, DWT, and SA contributed to the methodology; GG, GS, and SA performed the formal analysis; AM, CC, ADB, GS, TRR, and DWT contributed to the investigation; GG, GS, and SA performed data curation; GG and DWT contributed to writing—original draft preparation; GG, DRS, CBN, DWT, SA, and DF contributed to writing—review and editing; GG and DWT contributed to visualization; DWT and SA performed supervision; and DWT performed project administration.

Conflicts of Interest

DRS's academic department is receiving grant support from the Swiss National Science Foundation, Berne, Switzerland, the Swiss Society of Anesthesiology and Perioperative Medicine, Berne, Switzerland, the Swiss Foundation for Anesthesia Research, Zurich, Switzerland, Vifor SA, Villars-sur-Glâne, Switzerland and Vifor (International) AG, St. Gallen, Switzerland. DRS is cochair of the ABC-Trauma Faculty, sponsored by unrestricted educational grants from Novo Nordisk Health Care AG, Zurich, Switzerland, CSL Behring GmbH, Marburg, Germany, LFB Biomédicaments, Courtaboeuf Cedex, France and Octapharma AG, Lachen, Switzerland. DRS received honoraria or travel support for consulting or lecturing from Alliance Rouge, Bern, Switzerland, Danube University of Krems, Austria, European Society of Anesthesiology and Intensive Care, Brussels, BE, International Foundation for Patient Blood Management, Basel, Switzerland, Korean Society of Anesthesiologists, Seoul, Korea, Network for the Advancement of Patient Blood Management, Haemostasis and Thrombosis, Paris, France, Society for the Advancement of Blood Management, Mount Royal NJ, Alexion Pharmaceuticals Inc, Boston, MA, AstraZeneca AG, Baar, Switzerland, Bayer AG, Zürich, Switzerland, B. Braun Melsungen AG, Melsungen, Germany, Baxter AG, Glattpark, Switzerland, CSL Behring GmbH, Hattersheim am Main, Germany and Berne, Switzerland, CSL Vifor (Switzerland) Villars-sur-Glâne, Switzerland, CSL Vifor (International), St Gallen, Switzerland, Celgene International II Sàrl, Couvet, Switzerland, Daiichi Sankyo AG, Thalwil, Switzerland, Haemonetics, Braintree, Massachusetts, United States, LFB Biomédicaments, Courtaboeuf Cedex, France, Merck Sharp & Dohme, Kenilworth, New Jersey, United States, Novo Nordisk Health Care AG, Zurich, Switzerland, Octapharma AG, Lachen, Switzerland, Pharmacosmos A/S, Holbaek, Denmark, Pierre Fabre Pharma, Alschwil, Switzerland, Portola Schweiz GmbH, Aarau, Switzerland, Roche Diagnostics International Ltd, Reinach, Switzerland, Sarstedt AG & Co, Sevelen, Switzerland and Nümbrecht, Germany, Shire Switzerland GmbH, Zug, Switzerland, Takeda, Glattpark, Switzerland, Werfen, Bedford, MA, Zuellig Pharma Holdings, Singapore, Singapore. CBN is an inventor of Visual Patient and Visual Patient Predictive technologies, for which the University of Zurich and Koninklijke Philips N.V. hold patents, patent applications, design protections, and trademarks. Joint-development and licensing agreements exist with Philips Medizin Systeme Böblingen GmbH, Böblingen, Germany; Koninklijke Philips N.V., Amsterdam, The Netherlands; Philips Research/Philips Electronics Nederland BV, Eindhoven, The Netherlands; and Philips USA, Cambridge, Massachusetts, United States. Within the framework of these agreements, CBN receives travel support, lecturing and consulting honoraria, and may potentially receive royalties in the event of successful commercialization. CBN is an inventor of Visual Clot technology, with patent applications, design protections, and trademarks held by the University of Zurich. In case of successful commercialization, CBN may receive royalties. CBN is an inventor of Visual Blood technology, for which the University of Zurich holds patent applications and design protections; potential royalties may follow successful commercialization. CBN received travel support, lecturing, and consulting honoraria from Instrumentation Laboratory—Werfen, Bedford, Massachusetts, United States. DWT is the first named inventor of Visual Patient and Visual Patient Predictive technologies, for which the University of Zurich and Koninklijke Philips N.V. hold patents, patent applications, design protections, and trademarks. Joint-development and licensing agreements exist with Philips Medizin Systeme Böblingen GmbH, Böblingen, Germany; Koninklijke Philips N.V., Amsterdam, The Netherlands; Philips Research/Philips Electronics Nederland BV, Eindhoven, The Netherlands; and Philips USA, Cambridge, Massachusetts, United States. Within the framework of these agreements, DWT receives research funding, travel support, lecturing and consulting honoraria, and may potentially receive royalties in the event of successful commercialization. DWT also holds a position on the Philips Patient Safety Advisory Board. DWT is the first named inventor of Visual Clot technology, with patent applications, design protections, and trademarks held by the University of Zurich. In case of successful commercialization, DWT may receive royalties. DWT is the first named inventor of Visual Blood technology, for which the University of Zurich holds patent applications and design protections; potential royalties may follow successful commercialization. Additionally, DWT received travel support, lecturing, and consulting honoraria

from Instrumentation Laboratory—Werfen, Bedford, Massachusetts, United States, the Swiss Foundation for Anaesthesia Research in Zurich, Switzerland, and the International Symposium on Intensive Care and Emergency Medicine in Brussels, Belgium. No other funding or competing interests declared.

Multimedia Appendix 1
Educational Visual Clot video.
[MOV File , 19983 KB - humanfactors_v11i1e47991_app1.mov]

Multimedia Appendix 2
Complete translated field notes of participant interviews.
[PDF File (Adobe PDF File), 669 KB - humanfactors_v11i1e47991_app2.pdf]

Multimedia Appendix 3
Translated announcement of the survey invitation.
[PDF File (Adobe PDF File), 204 KB - humanfactors_v11i1e47991_app3.pdf]

References

1. Spahn DR, Bouillon B, Cerny V, Duranteau J, Filipescu D, Hunt BJ, et al. The European guideline on management of major bleeding and coagulopathy following trauma: fifth edition. Crit Care 2019;23(1):98 [FREE Full text] [doi: 10.1186/s13054-019-2347-3] [Medline: 30917843]
2. Weber CF, Zacharowski K. [Point of care 2.0: coagulation monitoring using Rotem® Sigma and Teg® 6S]. Anasthesiol Intensivmed Notfallmed Schmerzther 2018;53(6):412-424. [doi: 10.1055/s-0043-107755] [Medline: 29945283]
3. Riehl K, Lefering R, Maegele M, Caspers M, Migliorini F, Schenker H, et al. Is ROTEM diagnostic in trauma care associated with lower mortality rates in bleeding patients?-A retrospective analysis of 7461 patients derived from the TraumaRegister DGU®. J Clin Med 2022;11(20):6150 [FREE Full text] [doi: 10.3390/jcm11206150] [Medline: 36294471]
4. American Society of Anesthesiologists Task Force on Perioperative Blood Management. Practice guidelines for perioperative blood management: an updated report by the American Society of Anesthesiologists Task Force on Perioperative blood management*. Anesthesiology 2015;122(2):241-275 [FREE Full text] [doi: 10.1097/ALN.0000000000000463] [Medline: 25545654]
5. Tibi P, McClure RS, Huang J, Baker RA, Fitzgerald D, Mazer CD, et al. STS/SCA/AmSECT/SABM update to the clinical practice guidelines on patient blood management. Ann Thorac Surg 2021;112(3):981-1004 [FREE Full text] [doi: 10.1016/j.athoracsur.2021.03.033] [Medline: 34217505]
6. Collins PW, Lilley G, Bruynseels D, Laurent DBS, Cannings-John R, Precious E, et al. Fibrin-based clot formation as an early and rapid biomarker for progression of postpartum hemorrhage: a prospective study. Blood 2014;124(11):1727-1736 [FREE Full text] [doi: 10.1182/blood-2014-04-567891] [Medline: 25024304]
7. McNamara H, Kenyon C, Smith R, Mallaiah S, Barclay P. Four years' experience of a ROTEM® -guided algorithm for treatment of coagulopathy in obstetric haemorrhage. Anaesthesia 2019;74(8):984-991 [FREE Full text] [doi: 10.1111/anae.14628] [Medline: 30950521]
8. Wikkelsø A, Wetterslev J, Møller AM, Afshari A. Thromboelastography (TEG) or thromboelastometry (ROTEM) to monitor haemostatic treatment versus usual care in adults or children with bleeding. Cochrane Database Syst Rev 2016;2016(8):CD007871 [FREE Full text] [doi: 10.1002/14651858.CD007871.pub3] [Medline: 27552162]
9. Hawkins RB, Raymond SL, Hartjes T, Efron PA, Larson SD, Andreoni KA, et al. Review: the perioperative use of thromboelastography for liver transplant patients. Transplant Proc 2018;50(10):3552-3558. [doi: 10.1016/j.transproceed.2018.07.032] [Medline: 30577236]
10. Weber CF, Görlinger K, Meininger D, Herrmann E, Bingold T, Moritz A, et al. Point-of-care testing: a prospective, randomized clinical trial of efficacy in coagulopathic cardiac surgery patients. Anesthesiology 2012;117(3):531-547 [FREE Full text] [doi: 10.1097/ALN.0b013e318264c644] [Medline: 22914710]
11. Görlinger K, Dirkmann D, Hanke AA, Kamler M, Kottenberg E, Thielmann M, et al. First-line therapy with coagulation factor concentrates combined with point-of-care coagulation testing is associated with decreased allogeneic blood transfusion in cardiovascular surgery: a retrospective, single-center cohort study. Anesthesiology 2011;115(6):1179-1191 [FREE Full text] [doi: 10.1097/ALN.0b013e31823497dd] [Medline: 21970887]
12. Kvint S, Schuster J, Kumar MA. Neurosurgical applications of viscoelastic hemostatic assays. Neurosurg Focus 2017;43(5):E9 [FREE Full text] [doi: 10.3171/2017.8.FOCUS17447] [Medline: 29088950]
13. Welling H, Ostrowski SR, Stensballe J, Vestergaard MR, Partoft S, White J, et al. Management of bleeding in major burn surgery. Burns 2019;45(4):755-762. [doi: 10.1016/j.burns.2018.08.024] [Medline: 30292526]
14. Sahli SD, Pedrazzi N, Braun J, Spahn DR, Kaserer A, Plock JA. Effect of a factor-based coagulation management on blood product use after major burn injury: a retrospective cohort study. Burns 2021;47(7):1486-1494 [FREE Full text] [doi: 10.1016/j.burns.2021.08.015] [Medline: 34465509]

15. Treliński J, Misiewicz M, Robak M, Smolewski P, Chojnowski K. Assessment of rotation thromboelastometry (ROTEM) parameters in patients with multiple myeloma at diagnosis. Thromb Res 2014;133(4):667-670. [doi: 10.1016/j.thromres.2014.01.011] [Medline: 24451990]
16. Young GA, Carmona R, Garcia VC. Thromboelastography and thrombin generation assay in inherited afibrinogenemia. Haemophilia 2018;24(6):e410-e416. [doi: 10.1111/hae.13620] [Medline: 30345592]
17. Chitlur M, Rivard GE, Lillicrap D, Mann K, Shima M, Young G, Factor VIII, Factor IX, and Rare Coagulation Disorders Subcommittee of the Scientific and Standardisation Committee of the International Society on Thrombosis and Haemostasis. Recommendations for performing thromboelastography/thromboelastometry in hemophilia: communication from the SSC of the ISTH. J Thromb Haemost 2014;12(1):103-106 [FREE Full text] [doi: 10.1111/jth.12458] [Medline: 24261669]
18. Young G, Sørensen B, Dargaud Y, Negrier C, Brummel-Ziedins K, Key NS. Thrombin generation and whole blood viscoelastic assays in the management of hemophilia: current state of art and future perspectives. Blood 2013;121(11):1944-1950 [FREE Full text] [doi: 10.1182/blood-2012-08-378935] [Medline: 23319573]
19. Srivastava A, Kelleher A. Point-of-care coagulation testing. Contin Educ Anaesth Crit Care Pain 2013;13(1):12-16 [FREE Full text] [doi: 10.1093/bjaceaccp/mks049]
20. Gasciauskaite G, Malorgio A, Castellucci C, Budowski A, Schweiger G, Kolbe M, et al. User perceptions of ROTEM-guided haemostatic resuscitation: a mixed qualitative-quantitative study. Bioengineering (Basel) 2023;10(3):386 [FREE Full text] [doi: 10.3390/bioengineering10030386] [Medline: 36978777]
21. Dixon A, Underwood S, Schreiber M. Implementing thrombelastography: experiences from a level I trauma institution. Transfusion 2020;60(Suppl 6):S29-S32 [FREE Full text] [doi: 10.1111/trf.16070] [Medline: 33089931]
22. Rössler J, Meybohm P, Spahn DR, Zacharowski K, Braun J, Nöthiger CB, et al. Improving decision making through presentation of viscoelastic tests as a 3D animated blood clot: the visual clot. Anaesthesia 2020;75(8):1059-1069 [FREE Full text] [doi: 10.1111/anae.14985] [Medline: 32030729]
23. Castellucci C, Malorgio A, Budowski AD, Akbas S, Kolbe M, Grande B, et al. Coagulation management of critically bleeding patients with viscoelastic testing presented as a 3D-animated blood clot (the visual clot): randomized controlled high-fidelity simulation study. J Med Internet Res 2023;25:e43895 [FREE Full text] [doi: 10.2196/43895] [Medline: 37824182]
24. Roche TR, Said S, Rössler J, Gozdzik M, Meybohm P, Zacharowski K, et al. Physicians' perceptions of a situation awareness-oriented visualization technology for viscoelastic blood coagulation management (visual clot): mixed methods study. JMIR Serious Games 2020;8(4):e19036 [FREE Full text] [doi: 10.2196/19036] [Medline: 33172834]
25. Endsley MR. Designing for Situation Awareness: An Approach to User-Centered Design, Second Edition. Boca Raton, FL, USA: CRC Press Inc; 2011.
26. Wittgenstein L. Tractatus Logico-Philosophicus. London, UK: Routledge & Kegan Paul; 1922.
27. Degani A, Jorgensen C, Iverson D, Shafto M, Olson L. On Organization of Information: Approach and Early Work. Moffett Field, California: National Aeronautics and Space Administration, Ames Research Center; 2009.
28. Symon G, Cassell C. Qualitative Organizational Research: Core Methods and Current Challenges. Thousand Oaks, CA: SAGE Publications; 2012.
29. Guetterman TC, Fetters MD, Creswell JW. Integrating quantitative and qualitative results in health science mixed methods research through joint displays. Ann Fam Med 2015;13(6):554-561 [FREE Full text] [doi: 10.1370/afm.1865] [Medline: 26553895]
30. Smajic E, Avdic D, Pasic A, Prcic A, Stancic M. Mixed methodology of scientific research in healthcare. Acta Inform Med 2022;30(1):57-60 [FREE Full text] [doi: 10.5455/aim.2022.30.57-60] [Medline: 35800901]
31. Kaur N, Vedel I, El Sherif R, Pluye P. Practical mixed methods strategies used to integrate qualitative and quantitative methods in community-based primary health care research. Fam Pract 2019;36(5):666-671 [FREE Full text] [doi: 10.1093/fampra/cmz010] [Medline: 30888015]
32. Webster CS, McKillop A, Bennett W, Bagg W. A qualitative and semiquantitative exploration of the experience of a rural and regional clinical placement programme. Med Sci Educ 2020;30(2):783-789 [FREE Full text] [doi: 10.1007/s40670-020-00949-6] [Medline: 34457735]
33. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159-174. [Medline: 843571]
34. Merry AF, Hannam JA, Webster CS, Edwards KE, Torrie J, Frampton C, et al. Retesting the hypothesis of a clinical randomized controlled trial in a simulation environment to validate anesthesia simulation in error research (the VASER study). Anesthesiology 2017;126(3):472-481 [FREE Full text] [doi: 10.1097/ALN.0000000000001514] [Medline: 28079566]
35. Kurup V, Matei V, Ray J. Role of in-situ simulation for training in healthcare: opportunities and challenges. Curr Opin Anaesthesiol 2017;30(6):755-760. [doi: 10.1097/ACO.0000000000000514] [Medline: 28968283]
36. Castellucci C, Braun J, Said S, Roche TR, Nöthiger CB, Spahn DR, et al. Faster time to treatment decision of viscoelastic coagulation test results through improved perception with the animated visual clot: a multicenter comparative eye-tracking study. Diagnostics (Basel) 2022;12(5):1269 [FREE Full text] [doi: 10.3390/diagnostics12051269] [Medline: 35626425]

37. Said S, Roche TR, Braun J, Ganter MT, Meybohm P, Herrmann J, et al. Effects of an animated blood clot technology (visual clot) on the decision-making of users inexperienced in viscoelastic testing: multicenter trial. J Med Internet Res 2021;23(5):e27124 [FREE Full text] [doi: 10.2196/27124] [Medline: 33843602]
38. Schulz C, Endsley MR, Kochs EF, Gelb AW, Wagner KJ. Situation awareness in anesthesia: concept and research. Anesthesiology 2013;118(3):729-742 [FREE Full text] [doi: 10.1097/ALN.0b013e318280a40f] [Medline: 23291626]
39. Kang J, Hu J, Yan C, Xing X, Tu S, Zhou F. Development and applications of the anaesthetists' non-technical skills behavioural marker system: protocol for a systematic review. BMJ Open 2022;12(12):e065519 [FREE Full text] [doi: 10.1136/bmjopen-2022-065519] [Medline: 36517093]
40. Patient monitoring: visual patient avatar. Philips. 2023. URL: https://www.philips.co.uk/healthcare/technology/visual-patient-avatar [accessed 2023-12-22]
41. Products. mindray. 2023. URL: https://www.mindray.com/uk/products [accessed 2023-12-22]
42. Gasciauskaite G, Lunkiewicz J, Roche TR, Spahn DR, Nöthiger CB, Tscholl DW. Human-centered visualization technologies for patient monitoring are the future: a narrative review. Crit Care 2023;27(1):254 [FREE Full text] [doi: 10.1186/s13054-023-04544-0] [Medline: 37381008]
43. Atieno OA. An analysis of the strengths and limitation of qualitative and quantitative research paradigms. Probl Educ 21st Century 2009;13:13-38 [FREE Full text]

Abbreviations

NASA: National Aeronautics and Space Administration
ROTEM: rotational thromboelastometry

Edited by A Kushniruk; submitted 07.04.23; peer-reviewed by D Smriti, M Endsley; comments to author 23.09.23; revised version received 17.10.23; accepted 27.11.23; published 11.01.24.

Please cite as:
Gasciauskaite G, Castellucci C, Malorgio A, Budowski AD, Schweiger G, Kolbe M, Fries D, Grande B, Nöthiger CB, Spahn DR, Roche TR, Tscholl DW, Akbas S
User Perceptions of Visual Clot in a High-Fidelity Simulation Study: Mixed Qualitative-Quantitative Study
JMIR Hum Factors 2024;11:e47991
URL: https://humanfactors.jmir.org/2024/1/e47991
doi:10.2196/47991
PMID:38206666

Original Paper

Digital Care Pathway for Patients With Sleep Apnea in Specialized Care: Mixed Methods Study

Jari Haverinen[1,2], MSc, MHS; Terttu Harju[3], MD, PhD; Hanna Mikkonen[4], RNC; Pia Liljamo[5], RNC, PhD; Miia Turpeinen[6], MD, PhD; Jarmo Reponen[2,6], MD, PhD

[1]Finnish Coordinating Center for Health Technology Assessment, Oulu University Hospital, Oulu, Finland
[2]FinnTelemedicum, Research Unit of Health Sciences and Technology, Faculty of Medicine, University of Oulu, Oulu, Finland
[3]Medical Research Center Oulu, Oulu Pulmonary Department, Oulu University Hospital and University of Oulu, Oulu, Finland
[4]The Wellbeing Services County of North Ostrobothnia, Oulu, Finland
[5]Finnish Institute for Health and Welfare, Department of Knowledge Brokers, Data and Analytics Unit, Helsinki, Finland
[6]Medical Research Center Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland

Corresponding Author:
Jari Haverinen, MSc, MHS
Finnish Coordinating Center for Health Technology Assessment
Oulu University Hospital
Kajaanintie 50
Oulu, PL10
Finland
Phone: 358 504095446
Email: jari.haverinen@pohde.fi

Abstract

Background: Sleep apnea is a significant public health disorder in Finland, with a prevalence of 3.7%. Continuous positive airway pressure (CPAP) therapy is the first-line treatment for moderate or severe sleep apnea. From November 18, 2019, all patients who started their CPAP therapy at Oulu University Hospital were attached to a sleep apnea digital care pathway (SA-DCP) and were instructed on its use. Some patients still did not use the SA-DCP although they had started their CPAP therapy.

Objective: We aimed to study health care professionals' (HCPs') perspectives on the SA-DCP and its usefulness for their work; whether the main targets of SA-DCP can be reached: shortening the initial guiding sessions of CPAP therapy, reducing patient calls and contact with HCPs, and improving patients' adherence to CPAP therapy; and patients' perspectives on the SA-DCP and its usefulness to them.

Methods: Overall, 6 HCPs were interviewed in May and June 2021. The survey for SA-DCP users (58/91, 64%) and SA-DCP nonusers (33/91, 36%) was conducted in 2 phases: from May to August 2021 and January to June 2022. CPAP device remote monitoring data were collected from SA-DCP users (80/170, 47.1%) and SA-DCP nonusers (90/170, 52.9%) in May 2021. The registered phone call data were collected during 2019, 2020, and 2021. Feedback on the SA-DCP was collected from 446 patients between February and March 2022.

Results: According to HCPs, introducing the SA-DCP had not yet significantly improved their workload and work practices, but it had brought more flexibility in some communication situations. A larger proportion of SA-DCP users familiarized themselves with prior information about CPAP therapy before the initial guiding session than nonusers (43/58, 74% vs 16/33, 49%; P=.02). Some patients still had not received prior information about CPAP therapy; therefore, most of the sessions were carried out according to their needs. According to the patient survey and remote monitoring data of CPAP devices, adherence to CPAP therapy was high for both SA-DCP users and nonusers. The number of patients' phone calls to HCPs did not decrease during the study. SA-DCP users perceived their abilities to use information and communications technology to be better than nonusers (mean 4.2, SD 0.8 vs mean 3.2, SD 1.2; P<.001).

Conclusions: According to this study, not all the goals set for the introduction of the SA-DCP have been achieved. Despite using the SA-DCP, some patients still wanted to communicate with HCPs by phone. The most significant factors explaining the nonuse of the SA-DCP were lower digital literacy and older age of the patients. In the future, more attention should be paid to these user groups when designing and introducing upcoming digital care pathways.

(JMIR Hum Factors 2024;11:e47809) doi:10.2196/47809

KEYWORDS
health services; telehealth; telemedicine; health personnel; sleep apnea syndromes; mobile phone

Introduction

Background

Sleep apnea is a significant public health disorder in Finland, with a prevalence of 3.7%. The prevalence of sleep apnea worldwide has been increasing in relation to the obesity pandemic [1]. Untreated sleep apnea increases cardiovascular diseases, accidents, likelihood of taking sick leave, and premature mortality [2]. The clinical severity of sleep apnea is defined based on 3 components: daytime sleepiness owing to sleep apnea, the apnea-hypopnea index (AHI), and arterial blood oxygen saturation [2]. Continuous positive airway pressure (CPAP) therapy is the first-line treatment for moderate or severe sleep apnea in addition to conservative therapy (ie, weight loss, avoidance of sleep-disturbing substances, and lifestyle issues) [2]. CPAP therapy is a safe and efficient treatment for sleep apnea, relieving both daytime and nighttime symptoms and improving traffic safety [3,4]. In Finland, the need for CPAP treatment and the number of outpatient visits in both specialized and primary care have increased considerably because of the increased number of patients with sleep apnea [5].

The digitalization of health care has been seen as a potential option for offering treatment to patients regardless of time and place and involving them in their own care [6,7]. In addition, digitalization has the potential to make health care systems more efficient [8]. Despite its potential to improve health care services, digitalization does not automatically guarantee better services [9]. It has been noted as a problem, for example, that digital services are not necessarily aligned with clinician and patient preferences [9]. The challenge is that, in some cases, they complement rather than substitute the current services, and care processes are not always redesigned to achieve the best benefits from digital services [9-12]. Citizens' willingness and ability to use electronic services is also an obstacle to realizing the benefits of health care digitalization [13,14]. Because data breaches cause potentially catastrophic consequences, information security concerns have weakened patients' adoption of digital health services [15,16]. The challenges of the technical implementation of digital services, such as missing functionalities and lack of interoperability with existing information systems, have weakened the willingness of health care professionals (HCPs) to use them [10].

Factors promoting the adoption of digital health services are their perceived benefits for patients and patients' previous positive experiences with electronic services [13,15]. Previous studies showed that digital health interventions can improve patients' adherence to their care [17,18]. For example, Aardoom et al [18] showed that adherence to CPAP therapy in patients with sleep apnea can be improved with digital interventions in the initial months of treatment. Adherence to the use of the digital health service has also been found in some studies to positively affect outcomes [19,20]. Good digital literacy promotes the use of digital health services; studies have found young people have better digital literacy than older age groups [13,21]. As the user base of digital health care services can be very broad, and users can have functional limitations owing to age or illness, the ease of use of these services is important in promoting their use [15,22].

Finland's first phase of health care digitalization involved the digitalizing of HCPs' tools, such as electronic patient records; e-prescribing and digitalization have progressed well [23]. Currently, Finnish citizens are increasingly offered digital health care services and products [22,24]. Several countries, including Finland, have introduced new health technology assessment methods to ensure that digital health provides evidence-based benefits [16,22,25]. Digital care pathways (DCPs) are an example of digital health care services, and today, there are >300 DCPs in use in Finnish specialized care units [26]. One of the main goals of DCPs is to complement or replace traditional health care appointment visits [26]. In addition, DCPs aim to support and help in the self-treatment of long-term illnesses, monitoring, and adaptation to the illness, as well as enable patients to prepare for various health care procedures beforehand [26]. Several DCPs have been studied in Finland from the perspective of HCPs, organizations, and patients [7,10,13,27-31]. One of these DCPs is the sleep apnea DCP (SA-DCP), which was introduced at Oulu University Hospital (OUH) on November 18, 2019 [32]. All patients who start their CPAP therapy in OUH will be attached to the SA-DCP, that is, their patient data will be recorded in it, and they will be instructed on how to log in and use it [32]. When a patient starts on the SA-DCP, they register as an SA-DCP user through strong identification by accepting the terms of use and privacy statement and entering his or her contact information [33].

Objectives

In OUH, the CPAP therapy for patients with sleep apnea begins with an initial guiding session where patients are instructed on using their CPAP device. SA-DCP contains information and instructions about CPAP therapy; therefore, it would be desirable for patients to familiarize themselves with that information in advance [32]. In this way, the initial guiding session of CPAP therapy could be shortened because the basic information about CPAP therapy would not need to be reviewed again during the sessions. The SA-DCP contains reliable information about sleep apnea, its treatment, and CPAP therapy [32]. With the introduction of SA-DCP, it would be desirable to reduce patients' phone calls and other contacts with HCPs when information can be found in the SA-DCP. The SA-DCP also includes electronic messaging between patients and HCPs, which could reduce such calls [32]. The major aim of the SA-DCP is to increase patients' adherence to CPAP therapy. However, there is still a challenge in that some patients with sleep apnea do not log in and use it.

The main aims of the study are as follows:

1. To investigate HCPs' perspectives on the SA-DCP and its usefulness for their work
2. To determine whether the main targets of SA-DCP can be reached: shortening the initial guiding sessions of CPAP therapy, reducing patient calls and contact with HCP, and increasing patients' adherence to CPAP therapy.
3. To examine patients' perspectives on the SA-DCP and its usefulness.

Methods

Study Participants and Data Collection

The study population included HCPs at the OUH and patients who had started their CPAP therapy at the OUH. The patient population consisted of 2 groups. *SA-DCP users* were patients who had registered with the SA-DCP. *SA-DCP nonusers* referred to patients who had not registered with the SA-DCP.

Interviews of HCPs

HCPs of the OUH were contacted via email. Overall, 6 HCPs participated in the interviews from May to June 2021. Of these, 4 (67%) HCPs worked with patients, 1 (17%) was a supervisor, and 1 (17%) connected patients with sleep apnea to the SA-DCP and booked their appointments. The interviews were conducted remotely using a structured questionnaire. The HCPs provided voluntary informed consent for the interview by submitting a signed document. The interviews were then recorded and transcribed.

Survey for Patients With Sleep Apnea

The first part of material collection was conducted between May and August 2021. With the help of OUH HCPs, the survey, along with an invitation to participate and information about it, was sent to SA-DCP nonusers by mail. Respondents could send their responses by prepaid mail or electronically using Webropol Ltd's Webropol survey tool. SA-DPC users were informed about the study through the SA-DCP. They provided their consent and answered the survey using the SA-DCP questionnaire.

The second part of material collection was conducted between January and June 2022. Both SA-DCP users and SA-DCP nonusers were informed about the study with the annual device delivery in an assistive equipment center (AEC). They could send their responses by prepaid mail or answer electronically using Webropol Ltd's Webropol survey tool.

The patients' survey included multiple choice questions, 5-item Likert-type questions (with choices ranging from strongly disagree to strongly agree), and open-ended questions. In total, 33 SA-DCP nonusers and 58 SA-DCP users responded to the survey.

Remote Monitoring Data of CPAP Devices

Information about patients' adherence to CPAP therapy was collected from the remote monitoring data of CPAP devices. The HCP of OUH carried out the material collection manually in May 2021 in connection with 1-year controls of CPAP therapy. The collected information was anonymized and provided to the researchers. In total, CPAP remote monitoring data were collected from 90 SA-DCP nonusers and 80 SA-DCP users.

Registered Data of Phone Calls

The information about the number of patients' phone calls per year to an AEC was collected from Aurora Innovation Ltd's TeleQ program. The registered phone call data were collected during 2019, 2020, and 2021.

SA-DCP Customer Feedback Survey

Patients using the SA-DCP had the opportunity to provide customer feedback using the SA-DCP survey tool. The patients provided informed consent through the SA-DCP that their customer feedback could also be used for research purposes. The customer feedback did not contain any personal information. Feedback on the SA-DCP from 446 patients between February 18 and March 24, 2022, was included in this study.

Statistical Methods

Patients' survey data were analyzed using SPSS software (version 28.0; IBM Corp). Descriptive statistics were applied to calculate the mean and SD for continuous data and frequency and percentage for categorical data. Baseline differences between the groups were explored using a 2-tailed independent sample *t* test for continuous variables and chi-square test for categorical variables. A *P* value <.05 was considered statistically significant for all analyses.

Qualitative Analysis

Qualitative methods were used in this study to analyze the open-ended questions in patient surveys and interviews with HCPs. The collected material was first analyzed using an inductive content analysis method to obtain a comprehensive understanding [34]. Initially, the HCPs' and patients' responses to the open-ended questions were open coded. Subsequently, the analyzed data were grouped into subcategories, and then similar findings were combined into the main categories to enable the final analysis. Finally, the textual data were analyzed using the quantification method [35].

Ethical Considerations

The study followed the guidelines of the Finnish Advisory Board on Research Integrity [36]. According to Finnish Law (488/1999), this study was exempted from review by the institutional review board (ethics committee of Northern Ostrobothnia Hospital District). The respondents were informed of the study. All participants voluntarily participated in the study and provided their informed consent. The results were processed such that no participants were identifiable in the results or quotations of this study. Sensitive personal information was not collected. The data were processed and stored in a secure environment according to the procedures of the University of Oulu.

Results

The Number of New CPAP Therapies, SA-DCP Users, and Phone Calls in the Years Studied

The number of new CPAP therapies in OUH between 2019 and 2021 is presented in Table 1. The percentage of SA-DCP users has increased annually, but there are still patients who do not use the SA-DCP (Table 1). The number of phone calls per year to an AEC is presented in Table 1.

Table 1. New continuous positive airway pressure (CPAP) therapies, patients attached to the sleep apnea digital care pathway (SA-DCP), phone calls to an assistive equipment center (AEC) per year, and percentage of SA-DCP users.

	Year		
	2019	2020	2021
The number of new CPAP therapies	1172	1645	1160
The number of patients attached to the SA-DCP	292[a]	1645	1160
The number of SA-DCP users	130	1006	935
Percentage of SA-DCP users	44.5	61.2	80.6
The number of phone calls per year to an AEC	2784	4068	4020

[a]The SA-DCP was introduced from November 18, 2019, onward.

HCPs' Perspectives on the SA-DCP and its Usefulness for Their Work

On the basis of the interviews with HCPs, the main themes, facilitators, and barriers related to using the SA-DCP are presented in Textbox 1. According to the interviewed HCPs, they were unable to identify significant changes in their workload and working practices following the introduction of the SA-DCP. Only one responder perceived that his workload had slightly increased because the SA-DCP did not support integration with electronic patient record; therefore, patient data had to be transferred manually from one program to another (Textbox 1). However, HCPs reported that in some situations, the SA-DCP brought more flexibility to their work practices regarding patient communication (Textbox 1). For example, it enabled them to respond to patients' DCP messages during nonurgent work times, not only prereserved times. HCPs also reported that the initial guiding session of CPAP therapy went more smoothly for SA-DCP users who had familiarized themselves with the information about CPAP therapy through the SA-DCP (Textbox 1). The interviewed HCPs hoped that patients would make more use of the SA-DCP and its possibilities so its benefits would be better used.

Textbox 1. Themes and perceived barriers and facilitators regarding implementation of the sleep apnea digital care pathway (SA-DCP) according to health care professionals (HCPs).

Use rate of SA-DCP

- Barriers
  - SA-DCP's use rate had been lower than HCPs assumed it would be.
  - Some patients still thought that the only proper contact was personal contact with HCPs.
  - Patients' previous experiences with the need to log in to several digital health care services reduced their motivation to use them.
- Facilitators
  - Reminder text messages about logging into the SA-DCP have been sent to patients since June 2020.

Initial guiding session of continuous positive airway pressure (CPAP) therapy

- Barriers
  - Some SA-DCP users and nonusers still had not familiarized themselves with the prior information about CPAP therapy in advance.
- Facilitators
  - The guidance went more smoothly for SA-DCP users who familiarized themselves with the prior information about CPAP therapy.
  - From the patients' perspective, the instructional videos available in the SA-DCP were perceived as useful and clear.

Patients' communication practices with HCP

- Barriers
  - There were still a lot of phone calls.
  - HCPs also had to be reminded that they should not always call patients in connection with treatment controls but send a message via the SA-DCP.
- Facilitators
  - The SA-DCP gives patients more flexibility to contact HCPs regardless of time and place. For example, the patient may be in a location where they cannot answer the HCP's phone call.
  - HCPs may instruct the patient during a phone call to watch SA-DCP's educational video to get a better understanding of the matter.

Patients' adherence to CPAP therapy

- Barriers
  - There was no clear indication that patients' adherence to CPAP therapy was higher with the introduction of the SA-DCP.
- Facilitators
  - Reports obtained from CPAP devices had increased some patients' adherence to CPAP therapy.

Integration of SA-DCP into existing information and communications technology systems

- Barriers
  - The SA-DCP had to be used in a different web browser than electronic patient record (EPR).
  - Remote monitoring of CPAP devices requires a separate program, and remote monitoring data cannot be viewed via the SA-DCP.
  - Attaching patients to the SA-DCP is laborious and must be done manually by copying patient information from the EPR.
  - Data had to be copied manually from SA-DCP's messages into patients' care plans.

Workload and work practices of HCPs

- Barriers
  - HCPs' workloads did not change with the introduction of the SA-DCP.
- Facilitators

- The SA-DCP brought more flexibility to HCPs' work practices regarding communication with patients.
- The SA-DCP was a good way to deliver the necessary contact information to patients and thereby instruct them to reserve time for the necessary procedures by themselves.

The professionals also brought up ideas for the development of SA-DCP. They hoped that SA-DCP's integration with other information and communications technology (ICT) systems would be improved. One factor that caused a large workload for professionals was arranging appointment times for patients. The time reserved for the patient may not always suit him or her, necessitating a discussion about a more suitable time. If the patient could book appointments through the SA-DCP, it would greatly reduce the professionals' working hours. Two respondents mentioned that in the future, an initial guiding session of CPAP therapy could also be carried out remotely, but this would require that patients for whom this would be suitable should be identified in advance. The professionals hoped that all surveys and measurements made by the patients related to their treatment would be available in an electronic format. The hope was also that the SA-DCP's calendar would automatically remind patients, for example, to renew equipment, giving them more responsibility for managing their own affairs. One respondent wished that instructional videos could be directly linked to SA-DCP's messages so that patients would not have to search for them.

The HCP interviewees perceived digital services in health care as a positive thing. According to them, the services should be easy to use, and the real end users of the services should be included in their development. One respondent believed that patients will use digital health care services more frequently in the future, but such systems are always initially met with resistance. The respondent mentioned that at first, patients in Finland were against e-prescribing and the Patient Data Repository of Kanta Services, but today, such services are commonplace, and people use them smoothly.

Comparison of Characteristics Between SA-DCP Users and SA-DCP Nonusers

According to the patients' survey, there were no statistically significant differences in age, sex, and smoking status between SA-DCP users and nonusers (Table 2). According to the remote monitoring data of CPAP devices, SA-DCP nonusers were older than SA-DCP users (mean 59.1, SD 13.8 vs mean 55.3, SD 10.8; $P<.049$; Table 3). Compared with nonusers, SA-DCP users perceived their own abilities to use ICT to be better (mean 4.2, SD 0.8 vs mean 3.2, SD 1.2; $P<.001$); they used computers, tablets, or smartphones more often (58/58, 100% vs 27/33 81%; overall $P=.002$); and they were more accustomed to using electronic services (mean 4.8, SD 0.5 vs mean 4.1, SD 1.2; $P=.006$; Table 2). There was no statistically significant difference in how regularly SA-DCP users and nonusers used the electronic services (Table 2). SA-DCP users thought that communication about SA-DCP and how to log in had been clear, although SA-DCP nonusers thought that it had not (yes 52/58, 91% vs yes 7/33, 24%; overall $P<.001$; Table 2). Compared with SA-DCP users, SA-DCP nonusers preferred phone calls or physical appointments with HCPs to manage their health-related issues (Table 2). Neither SA-DCP users nor SA-DCP nonusers had any major concerns about the data security and protection of digital health care services (Table 2).

Table 2. Patient responses to the survey.

Characteristics	Use of SA-DCP[a]		*P* value
	Nonusers (n=33)	Users (n=58)	
Age (years), mean (SD)	61.9 (11.6)	57.3 (12.0)	.08
Sex, n (%)			.07
Male	26 (79)	34 (59)	
Female	7 (21)	24 (41)	
Physical training frequency, n (%)			.03
Daily	16 (49)	15 (26)	
Weekly	13 (39)	37 (64)	
Monthly	1 (3)	5 (9)	
Less than monthly	1 (3)	0 (0)	
No physical training	2 (3)	1 (2)	
Smoking, n (%)			>.99
Yes	2 (6)	4 (7)	
No	31 (94)	53 (93)	
Adherence to CPAP[b] therapy (own assessment; Likert scale 1-5), mean (SD)	4.8 (0.5)	4.7 (0.8)	.87
Patients familiar with CPAP therapy before the initial guiding session	16 (49)	43 (74)	.02
The patient became familiar through SA-DCP, n (%)	N/A[c]	29 (67)	—[d]
Average use of the CPAP device per night (hours), mean (SD)	6.3 (1.0)	6.3 (1.3)	.97
Has CPAP therapy helped the patient's sleep apnea?, n (%)			>.99
Yes	28 (85)	49 (85)	
No	0 (0)	0 (0)	
Cannot say	5 (15)	9 (15)	
Information and communication technology skills (own assessment; Likert scale 1-5), mean (SD)	3.2 (1.2)	4.2 (0.8)	<.001
Patient's computer, tablet, or smartphone use, n (%)			.002
Regularly (weekly)	27 (82)	58 (100)	
Randomly (less often than weekly)	4 (12)	0 (0)	
None	2 (6)	0 (0)	
How accustomed is the patient to using electronic services (eg, banking services, appointment services, etc; own assessment)? (Likert scale 1-5), mean (SD)	4.1 (1.2)	4.8 (0.5)	.006
If the patient uses electronic services, how regularly?, n (%)			.55
Daily	20 (69)	44 (76)	
Weekly	7 (24)	11 (19)	
Monthly	1 (3)	3 (5)	
Less often than monthly	1 (3)	0 (0)	
Would the patient choose an electronic service or a phone call as a contact method regarding her or his treatment?, n (%)			<.001
Electronic service	13 (39)	48 (83)	
Phone call	20 (61)	10 (17)	
If the patient could choose either an electronic service (eg, remote consultation) or a physical appointment regarding her or his treatment, which method would she or he prefer?, n (%)			.048
Electronic service	10 (30)	31 (53)	
Physical appointment	23 (70)	27 (46)	

Characteristics	Use of SA-DCP[a]		*P* value
	Nonusers (n=33)	Users (n=58)	
Patient concerns about the data security and protection of digital health care services (Likert scale 1-5), mean (SD)	2.6 (1.2)	2.2 (1.1)	.23
Has communication about SA-DCP and how to log in to it been sufficiently clear?, n (%)			<.001
Yes	7 (24)	52 (91)	
No	22 (76)	5 (9)	
Did SA-DCP increase the patient's adherence to CPAP therapy?, n (%)			
Yes	N/A	42 (72)	
No	N/A	16 (28)	
Did the patient contact HCP[e] during her or his treatment period?, n (%)	17 (52)	33 (57)	.67
Through SA-DCP	N/A	17 (52)	
Through another contact method	17 (100)	16 (48)	
The contact was related to (total), n	26	37	—
Treatment of sleep apnea, n (%)	8 (31)	3 (8)	—
CPAP therapy, n (%)	15 (58)	23 (62)	—
Other issues, n (%)	3 (12)	11 (30)	—
Did patients who contacted HCP get the help they needed?, n (%)			
Yes	16 (94)	31 (94)	.41
Through SA-DCP messaging	N/A	16 (52)	—
Through another contact method	16 (100)	15 (48)	—
Did the patient need to find additional information about his or her treatment without contacting HCPs during the treatment period?, n (%)	9 (27)	23 (40)	.48
The patient got the information she or he needed	8 (89)	21 (91)	>.99
Through SA-DCP	N/A	8 (38)	—
Through another source (internet, patient organizations, etc)	8 (100)	13 (62)	—

[a]SA-DCP: sleep apnea digital care pathway.
[b]CPAP: continuous positive airway pressure.
[c]N/A: not applicable.
[d]Not available.
[e]HCP: health care professional.

Table 3. Remote monitoring data of continuous positive airway pressure (CPAP) devices.

Characteristics	Use of SA-DCP[a]		*P* value
	Nonusers (n=90)	Users (n=80)	
Age (years), mean (SD)	59.1 (13.8)	55.3 (10.8)	.049
Sex, n (%)			.35
Male	48 (53.3)	49 (61.3)	
Female	42 (46.7)	31 (38.8)	
AHI[b] at diagnosis, mean (SD)	32.5 (18.0)	30.6 (18.9)	.51
AHI residual in treatment, mean (SD)	2.4 (2.7)	2.1 (3.9)	.51
Percentage of nights CPAP was used, mean (SD)	92.6 (12.7)	91.5 (18.8)	.66
Hours of CPAP use per night, mean (SD)	6.2 (1.5)	6.1 (1.8)	.69
CPAP device mask leak, mean (SD)	3.1 (6.1)	2.8 (3.2)	.74
CPAP device median pressure, mean (SD)	8.4 (2.2)	7.8 (2.0)	.06

[a]SA-DCP: sleep apnea digital care pathway.
[b]AHI: apnea-hypopnea index.

Patients' Rationales for Using or Not Using SA-DCP

Patients were asked about their rationale for using or not using the SA-DCP (Table 4). SA-DCP users mostly adopted the SA-DCP because they thought that signing up for the SA-DCP was part of their treatment process (42/58, 72%). SA-DCP nonusers did not adopt the SA-DCP mainly because they were unaware of it (15/33, 46%).

Table 4. Patients' rationales for using or not using the sleep apnea digital care pathway (SA-DCP).

Patients' rationales for using or not using the sleep apnea digital care pathway (SA-DCP)	Values, n (%)
Patients' rationales for using the SA-DCP (n=58)	
I thought signing up for SA-DCP was part of my treatment process	42 (72)
It was recommended to me	39 (67)
It allows me to take care of my affairs regardless of time and place	31 (53)
I am very accustomed to using electronic services	25 (42)
I can more easily get information about sleep apnea and its treatment	18 (31)
I prefer to use electronic services for my treatment	15 (26)
I can more easily get information about CPAP[a] therapy	15 (26)
I can take care of things related to my care more safely during the current COVID-19 period	12 (21)
By using SA-DCP, I am more committed to my treatment	8 (14)
Other reasons	2 (3)
Patients' rationales for not using the SA-DCP (n=33)	
I am not aware of SA-DCP	15 (46)
I prefer physical appointments	13 (39)
I prefer phone calls	10 (30)
I am aware of SA-DCP, but I forgot to log in	9 (27)
I do not know how to use electronic services	7 (21)
I do not want to use electronic services	6 (18)
The use of electronic services is generally difficult	6 (18)
I do not receive personal help through electronic services	5 (15)
Other reasons	4 (12)
I am concerned about the data security and protection of electronic services	3 (9)
My sleep apnea treatment and CPAP therapy are balanced, so I do not need to contact health care professionals through any communication channel.	3 (9)
My sleep apnea treatment and CPAP therapy are balanced, so I do not need additional information through any communication channel.	1 (3)
I do not feel the need to log into SA-DCP as part of my CPAP therapy	1 (3)

[a]CPAP: continuous positive airway pressure.

Patients' Prefamiliarization With CPAP Therapy Before the Initial Guiding Session

A larger proportion of SA-DCP users had familiarized themselves with prior information about CPAP therapy before the initial guiding session of CPAP therapy than SA-DCP nonusers (43/58, 74% vs 16/33, 49%; *P*=.02; Table 2). Among the 48 SA-DCP users who familiarized themselves with information about CPAP therapy beforehand, 29 (67%) performed it through the SA-DCP (Table 2). Most SA-DCP nonusers (6/16, 38%) said they had received the preliminary information from a spouse or a relative who had already used a CPAP device. The other sources of information for both groups were the internet (6/59, 10%), private health care providers (3/59, 5%), primary health care units (2/59, 3%), and the Duodecim medical information database (2/59, 3%). SA-DCP users also received information from occupational health care units (2/43, 5%) and the Facebook sleep apnea support group (2/43, 5%). Correspondingly, SA-DCP nonusers received information from specialized care units (2/16, 13%), research articles (1/16, 6%), and AEC (1/16, 6%).

The initial guidance sessions of CPAP therapy were carried out with small groups of patients (4-8 patients at a time). According to HCPs, the initial guiding sessions were smoother for patients who had already familiarized themselves with prior information about CPAP therapy through the SA-DCP (Textbox 1). The problem was that many patients still did not have prior information about CPAP therapy; therefore, most of the initial guiding sessions had to be implemented according to their needs. According to the HCPs, patients found the instructional videos available in the SA-DCP to be useful and clear (Textbox 1). Patients were also instructed to familiarize themselves with them and other information material found on the SA-DCP even after the sessions if they had further questions.

Patients' Information Needs About Sleep Apnea and CPAP Therapy

SA-DCP includes electronic messaging functionality between patients and HCPs and information about sleep apnea, its treatment, and CPAP therapy. According to the survey responses of SA-DCP users, most patients looked for information about the SA-DCP, sleep apnea, self-treatment of sleep apnea, and cleaning and maintenance of the CPAP device (Table 5). The messaging functionality of the SA-DCP and its "frequently asked questions" function were not widely used; only 38% (22/58) of SA-DCP users used them (Table 5).

Table 5. Functionalities of the sleep apnea digital care pathway (SA-DCP) used by patients according to the SA-DCP users survey (N=58).

	Values, n (%)
The number of patients who familiarized themselves with the following information materials	
Welcome to SA-DCP	55 (95)
Sleep apnea	45 (78)
Cleaning and maintenance of the CPAP[a] device	43 (74)
Self-treatment of sleep apnea	42 (72)
CPAP therapy	39 (67)
Preparing for the CPAP therapy initial guiding session	38 (67)
Sleep apnea and driving ability	32 (55)
CPAP therapy in unusual everyday situations	28 (48)
Controls, rehabilitation, and social security	25 (43)
Frequently asked questions	22 (38)
The patient has communicated with the HCP[b] in matters related to his or her treatment through the messaging functionality of SA-DCP	22 (38)

[a]CPAP: continuous positive airway pressure.

[b]HCP: health care professional.

During the treatment period, both SA-DCP users and SA-DCP nonusers sought additional information regarding their treatment without contacting HCPs (23/58, 39% vs 9/33, 27%; *P*=.48; Table 2). Among 23 SA-DCP users who sought more information, 8 (35%) performed it through the SA-DCP. The other reported information sources for SA-DCP users were the internet (9/23, 39%), Facebook sleep apnea support group (1/23, 4%), patient organizations (1/23, 4%), rehabilitation (1/23, 4%), and professional education (1/23, 4%). SA-DCP nonusers received additional information from the following sources: the internet (5/9, 56%), Facebook sleep apnea support group (1/9, 11%), scientific articles (1/9, 11%), and information material provided by private health care services providers (1/9, 11%). Most SA-DCP nonusers and SA-DCP users who sought more information about CPAP therapy and sleep apnea received the information they needed (8/9, 89% vs 21/23, 91%; *P*>.99; Table 2).

Patients' Communication Practices With HCPs

During the treatment period, 52% (17/33) of SA-DCP nonusers and 57% (33/58) of SA-DCP users contacted HCPs (Table 2). Among 33 SA-DCP users who contacted HCPs, 17 (52%) used the SA-DCP, and the rest used other contact methods (Table 2). SA-DCP users who preferred contact methods other than SA-DCP messages were older (mean 63.3, SD 9.4 vs mean 55.7, SD 10.8; *P*<.04). Phone calls were the most important form of contact for SA-DCP users (10/33, 30%). SA-DCP nonusers (9/17, 53%) mostly contacted HCP via phone calls. The next most common contact method for both groups was a physical visit to an AEC or a primary health care unit. The contact mostly concerned CPAP therapy; this was the case for 88% (15/17) of SA-DCP nonusers and 70% (23/33) of SA-DCP users (Table 2). Most SA-DCP nonusers and SA-DCP users who contacted HCPs received the help they needed (16/17, 94% vs 31/33, 94%; *P*=.41; Table 2).

According to the phone call register data, the annual number of phone calls to an AEC was still high even after the introduction of SA-DCP (Table 1). An exact comparison of phone calls to AECs per patient between different years could not be made because the number of new CPAP therapies in the OUH varied between different years, and the number of annual phone calls also showed contacts with HCPs from patients whose CPAP therapies had started in previous years (Table 1). The results of 2019 mainly represent a situation in which the SA-DCP was not yet in use at OUH because it was introduced at the very end of 2019. The results of 2021 represent a situation in which the SA-DCP had been in use at OUH for approximately 2 years. HCPs also indicated the same; there was no significant decrease in the number of phone calls, and there were still many phone calls related to CPAP therapy (Textbox 1). HCPs emphasized that they try to guide patients during phone calls to use the SA-DCP more in matters related to their care. Although the patients' affairs were handled mostly with phone calls, the HCPs thought the instructional videos and informational materials included in the SA-DCP were valuable. It was possible to better explain things to patients with them (Textbox 1). For example, HCPs may instruct the patient during a phone call to watch SA-DCP's educational video to get a better understanding of the matter. The HCPs also emphasized that the SA-DCP is a good way to deliver the necessary contact information to patients, allowing them to reserve time for the necessary procedures themselves (Textbox 1).

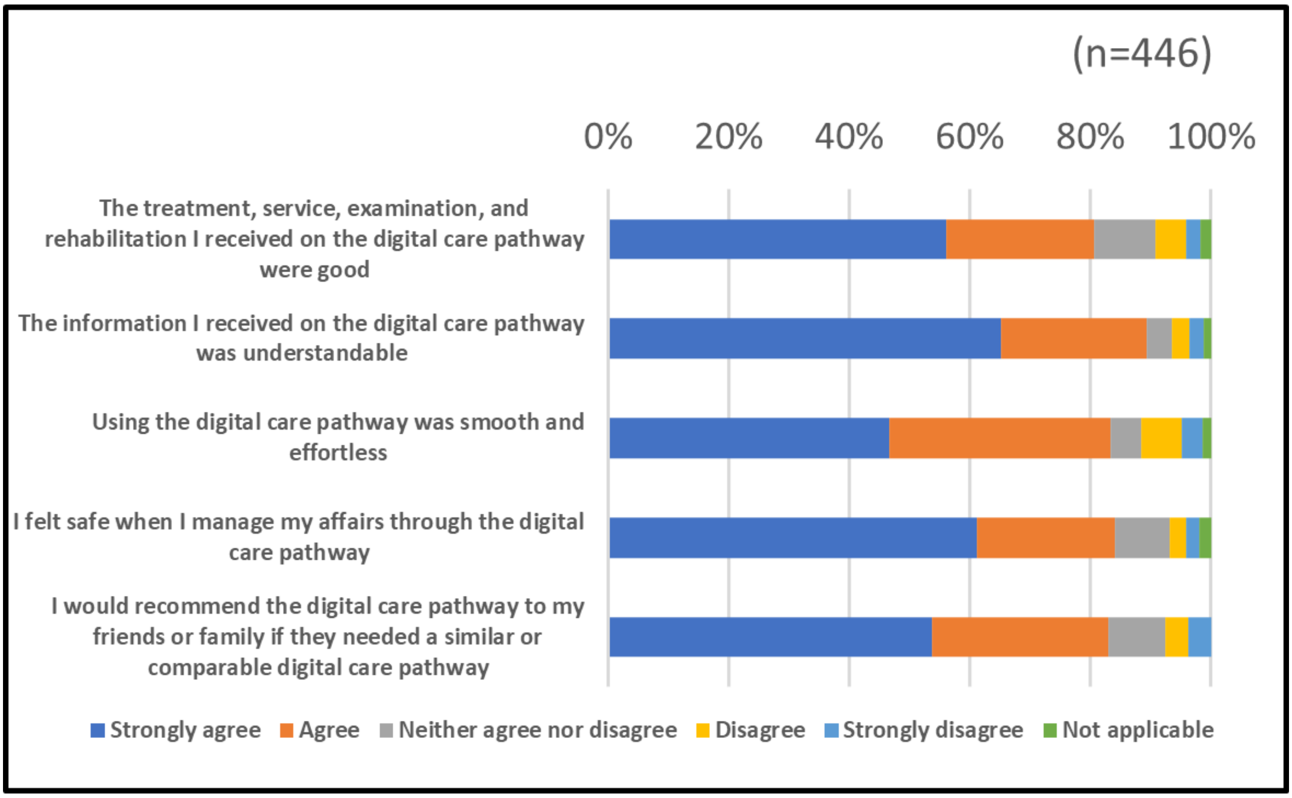
Patient Adherence to CPAP Therapy

According to the patients' responses to the survey and remote monitoring data of CPAP devices, adherence to CPAP therapy was high in both groups (Tables 2 and 3). Both groups used the CPAP device on average for >6 hours per night and on >90% of nights (Tables 2 and 3). On the basis of the patients' own assessments, adherence to CPAP therapy was high in both SA-DCP nonusers (mean 4.8, SD 0.5) and SA-DCP users (mean 4.7, SD 0.8; Table 2). In addition, according to the patients' survey, 72% (42/58) of SA-DCP users reported that SA-DCP had made them more motivated to perform their own CPAP therapy (Table 2). Most patients in both groups believed that CPAP therapy helped them treat sleep apnea (Table 2). The remote monitoring data of CPAP devices showed that CPAP therapy had significantly reduced the number of AHIs for both groups (Table 3).

Patient Feedback About SA-DCP

A total of 446 patients responded to the customer feedback survey; their feedback is shown in Figure 1. Patient feedback on the SA-DCP was generally positive; most of them agreed or strongly agreed with the survey claims (Figure 1). When examining the results, it should be noted that the questions of the patient feedback survey are common to every DCP in the OUH. As the functionalities offered by DCPs vary according to the care chains of different diseases, not all the questions are necessarily valid for every DCP. For example, examinations are not offered through the SA-DCP.

Figure 1. Patient feedback about the sleep apnea digital care pathway.



Of 446 patients, 102 (22.9%) who responded to the survey provided free-form feedback on the SA-DCP. Moreover, 22 patients gave generally positive feedback about the SA-DCP. For the most part, they did not elaborate on their feedback. According to 2 respondents, the possibility to use the services remotely was a good thing, and according to 2 respondents, the SA-DCP was a good and modern service. However, 19 patients thought that they did not need to use the SA-DCP, or that it did not add value to their treatment. Moreover, 11 respondents mentioned that communication and information about the SA-DCP should be improved. According to 9 respondents, the SA-DCP contained good and comprehensive information about sleep apnea and its treatment, as well as CPAP therapy. However, 3 respondents mentioned that although the SA-DCP contained good information, the same information can be found on the internet. With regard to SA-DCP's messaging feature, 5 respondents thought it was a functional solution. Conversely, 9 respondents said that they encountered problems or delays related to messaging and 9 respondents desired new features for the SA-DCP, such as better search functionality. As the information content of SA-DCP was only available in Finnish during the research, some respondents presented English language support as a need for future development. According to 5 responses, SA-DCP's user interface was clear, and its usability was good. In contrast, 4 respondents stated that the user interface could still be improved. Three respondents had technical problems and challenges when using the SA-DCP. Two users reported that the SA-DCP worked well technically. Three respondents said that they would not like to manage their affairs through digital services. Four respondents reported that they had experienced challenges using the SA-DCP, especially in relation to finding their own care path.

Discussion

Principal Findings

This study investigated whether the 3 main goals for introducing the SA-DCP at OUH were achieved. The first aim of introducing the SA-DCP was to shorten the initial guiding sessions of CPAP therapy on the assumption that the patients would have familiarized themselves with prior information about CPAP therapy in advance through the SA-DCP. The second main aim was to reduce the number of patients' phone calls and contacts to HCPs, especially when the information can be found in the SA-DCP. The primary goal of implementing SA-DCP at OUH was to improve patients' adherence to CPAP therapy. However, according to the results of this study, not all the objectives of introducing the SA-DCP were achieved.

On the basis of the HCP's responses to this study, shortening the initial guiding sessions of CPAP therapy had not been fully achieved, although a significantly larger number of SA-DCP users had familiarized themselves with prior information about CPAP therapy compared with SA-DCP nonusers. In this regard, it can be said that SA-DCP has contributed to the better preparation of patients for sessions. The initial guiding sessions were smoother for patients who had already familiarized themselves with prior information regarding CPAP therapy through the SA-DCP. However, many patients still did not have prior information about CPAP therapy; therefore, most sessions had to be implemented according to their needs. Because digital services may require care process changes to get the most out of them, 2 HCPs mentioned that the initial guidance sessions could also be carried out remotely in the future; however, this would require that the patients for whom this procedure would be suitable should be identified in advance [11,12].

Despite previous studies showing that DCPs would make it possible to reduce the number of patient phone calls to HCPs, this did not happen in the case of SA-DCP [37,38]. The annual number of phone calls to an AEC was still high even after the introduction of SA-DCP, according to the phone call register data. As the number of patients' phone calls related to CPAP therapy was still high, HCPs mentioned that it was difficult to assess the actual change in the number of phone calls. However, they perceived that the number of patient calls did not decrease significantly. Previous studies have shown that patients' ability to use electronic services also promotes the use of digital health care services [39,40]. However, Jenssen et al [41] found that despite the regular use of new digital technologies and services such as electronic banking, few of their study participants supported using these tools for communicating with their HCPs. The same behavior pattern can also be observed in the case of SA-DCP. Although SA-DCP users in this study perceived their ability to use ICT to be good and used computers, tablets, or smartphones regularly and were accustomed to using electronic services, only approximately half of them contacted the HCP with SA-DCP messages when needed. Among SA-DCP users, phone calls were the most important other contact method. The notable finding was that SA-DCP users who preferred another contact method were older.

Patient concerns about data security and protection have weakened their willingness to use electronic communication methods in health care [15,42]. On the basis of this study, this would not be an explanatory factor for the low use of SA-DCP messages, as both SA-DCP users and SA-DCP nonusers were not significantly concerned about the data security and protection of digital health care services. Zanaboni and Fagerlund [43] discovered that communicating via electronic tools was less time-consuming from the patient's perspective than communicating via phone calls. However, some participants indicated that the time elapsed to receive a response from the HCP was more important than the time spent using the service itself. Long response times have been seen as one of the most important reasons for patients' dissatisfaction with electronic communication in health care [39,44]. In a Norwegian study, older patients hoped that their electronic messages would be answered the next day at the latest; otherwise, they experienced dissatisfaction with the service [39]. From the patients' point of view, they may perceive that a phone call is a quick and convenient way to handle their health-related matters [45-47]. The fundamental difference is that a phone call involves real-time interactive communication, whereas SA-DCP messages can be defined as asynchronous communication [48]. The patient may ask follow-up questions during the phone call and the HCP can answer them immediately. When using electronic communication tools, there may be delays in answers to questions and possible follow-up questions because of asynchronous communication, as the patient and the HCP may not be dealing with the issue simultaneously [48].

One of the main goals of introducing the SA-DCP was to improve patients' adherence to CPAP therapy. This study showed no statistical difference between SA-DCP users' and nonusers' adherence to CPAP therapy. Adherence to CPAP therapy was high in both groups according to the patients' own estimates and remote monitoring data of CPAP devices. Both groups performed CPAP therapy regularly and reported that it helped them to treat their sleep apnea. In addition, 72% (42/58) of SA-DCP users reported that SA-DCP motivated them to perform their own CPAP therapy. Unfortunately, this study did not ask why the participants felt this. The role of the SA-DCP was to complement CPAP therapy by providing information and an electronic communication channel. It did not include clear mechanisms for influencing patients' behavior related to their own health as digital health interventions typically do, for example, in relation to weight management [49-51]. The CPAP therapy clearly helped the participants in this study to reduce the number of AHIs. Presumably, the biggest motivation for performing CPAP therapy came from alleviating sleep apnea symptoms and not so much from using the SA-DCP; therefore, the SA-DCP was not a significant factor in explaining adherence to CPAP therapy.

This study investigated HCPs' perspectives on the SA-DCP and its usefulness for their work. Although previous studies determined that DCPs could potentially free health care services capacity for other purposes and reduce the workload of HCPs, the results of this study do not support these results in the case of SA-DCP [27,28]. The HCPs who participated in the study were unable to define significant changes in their workload and work practices after the introduction of SA-DCP. The primary aim of HCPs was for patients to use the SA-DCP more so that its benefits could be better used. Previous studies have highlighted that DCPs can promote work flexibility, for example, by enabling HCPs to respond to patients' DCP messages at nonurgent, not only prereserved times [27,31]. The responses of HCPs in this study pointed out the same. With the help of the DCP, patients can access the information it contains before and after contact with HCPs, thus reducing patient follow-up questions [12]. From this perspective, HCPs felt that educational videos and information materials on SA-DCP were beneficial because, through them, the patients could better understand things. From a technical point of view, the SA-DCP's weak integration with existing ICT systems was seen as one of its key shortcomings and an area for future development. The lack of interoperability with existing ICT systems has been found to weaken the willingness of HCPs to use digital health care services and increase their workload [10,52]. According to the interviewed HCPs, lack of integration reduced the fluency of their work, increased the workload of one responder, and can cause risks from the perspective of information protection and patient safety when patient information is copied manually between different programs.

Digital health care services are intended to help patients become more active actors, more adherent to their own care, and change their behavior in a more favorable direction for their health [49-51,53,54]. Promising results have already been achieved, for example, in treating obesity with the help of digital services [30,51]. In the case of SA-DCP, it was hoped that patients would be active and familiarize themselves with the information contained in it about sleep apnea and CPAP therapy. According to the patient survey, most SA-DCP users have done so. Although most SA-DCP users familiarized themselves with the information in SA-DCP, there was no statistically significant difference in the proportion of SA-DCP users and nonusers who sought additional information about their illness or CPAP therapy. From this perspective, it cannot be said that SA-DCP users are more active actors. It has been established that digital health care services can lower the threshold for patients to contact HCPs [37,54,55]. According to this study, there was no statistically significant difference between the percentage of SA-DCP users and SA-DCP nonusers who contacted HCPs during their treatment period. However, this study did not ask how often the patients contacted the HCPs. On the basis of the results of this study, it seems that patients sought additional information about their illness or contacted HCPs when they had a real need, regardless of the information source or communication method.

Most SA-DCP users thought that the treatment they received through the SA-DCP was good; it was fine technically, a safe service, and the information it contained was clear and understandable. However, some patients still did not use SA-DCP, although the relative number of active SA-DCP users increased during the study period. Lack of digital literacy is one of the barriers to promoting the use of digital health care services. Older adults, in particular, tend to have lower digital literacy than the general population [39,56]. Mannheim et al [40] emphasized in their study that older adults are not a homogeneous group in terms of digital literacy and should also be better included when designing digital health care services [40]. On the basis of the patient survey, there was no statistically significant difference in the age of SA-DCP users and SA-DCP nonusers, but based on remote monitoring data from CPAP devices, SA-DCP nonusers were older. According to this study, SA-DCP nonusers perceived their abilities to use ICT to be worse; they used computers, tablets, or smartphones more rarely and were less accustomed to using electronic services than SA-DCP users. SA-DCP nonusers preferred phone calls or physical appointments to manage their health-related issues with HCPs. The results showed a statistically significant difference in how clearly the patients perceived the communication about SA-DCP. Only 24% (7/38) of SA-DCP nonusers considered communication to be clear, and ignorance of the SA-DCP was the most common reason for them not to use the SA-DCP. After the diagnosis of sleep apnea, the patients received an information letter containing information about the disease and its treatment. This letter also included information on the SA-DCP and how to use it. Did SA-DCP nonusers think the SA-DCP was not adequately explained because they did not want to use digital health care services in the first place and preferred to conduct their health-related issues through phone calls or physical visits? They may not have paid attention to the SA-DCP information letter if they do not typically use or are not willing to use digital health care services or if they perceive they have weak skills in using them.

One of the key findings of this study is that the nonuse of SA-DCP and its functionalities among patients with sleep apnea means that its full potential is not being used. This can be seen, for example, in the initial guiding sessions of CPAP therapy, when some patients still come without prior knowledge. Although the number of SA-DCP users increased during the years covered by this study, not all SA-DCP functionalities were significantly used. In particular, this was reflected in the fact that SA-DCP messages were not widely used; therefore, the number of calls to AECs was not reduced. This study found that lower digital literacy and older age were significant factors in explaining the nonuse of the SA-DCP. Older SA-DCP users more often favored other contact methods, such as phone calls, when contacting HCPs during their treatment period. In the future, special attention should be paid to how digital health care services are designed according to the needs of older adults with weak digital literacy. Care processes should be better adapted to the requirements of digital health care services. Clearly, only the traditional information letter about SA-DCP is not sufficient to encourage all patients to adopt it. If there are challenges in deployment, patients could be more actively encouraged to adopt the SA-DCP and offered support. Previous studies have highlighted that the desire of older adults to use digital health care services can be supported by offering guidance and peer support [39,56]. Studies have also emphasized that both professionals and patients should be closely involved in DCP development to obtain the best benefit and that development should be a continuous process [10,29]. With age, various functional limitations, such as diminished eyesight related to diabetes or deteriorated motor skills owing to rheumatism, can increase and thus make it more difficult to use digital services [57,58]. Therefore, special attention should be paid to the usability and accessibility of digital services. The real end users should be involved in the design process, as the interviewed HCPs highlighted [22,25,58].

Limitations

Our study had some limitations. Patients with sleep apnea give up CPAP therapy for different reasons, which can bias this study's data regarding patients' adherence to CPAP therapy. Most patients who responded to the survey had continued CPAP therapy for ≥1 year, and remote monitoring data on CPAP devices were collected in connection with 1-year control. Unfortunately, when the study was carried out, no information was available on the proportion of SA-DCP users and SA-DCP nonusers who had discontinued CPAP therapy. This would have provided additional information about patients' adherence to CPAP therapy. Previous results have highlighted that high attrition rates hinder achieving the full benefits of digital health care services. During the implementation of the study, the SA-DCP did not enable the automatic collection of log data on the activity of patients using the SA-DCP, but through the automatic log data, it was only possible to determine that the patient had used the SA-DCP. Therefore, this study did not examine patients' adherence to SA-DCP use, but only whether they had used the service.

On the basis of the study's results, approximately half of SA-DCP users still contacted HCPs in a way other than through SA-DCP messages, although they reported having good digital literacy. Most SA-DCP nonusers also preferred phone calls to contact HCPs. However, in this study, SA-DCP users and SA-DCP nonusers were not asked why some preferred phone calls to contact HCPs instead of electronic messaging. Future research is needed to better understand this behavior pattern. This study did not ask patients how many times they contacted HCPs; it only investigated whether the patients contacted HCPs during their treatment period. Information on the number of contacts would have provided valuable information on whether using the SA-DCP can lower the threshold for contacting HCPs.

One of the goals of the SA-DCP was to increase patients' adherence to CPAP therapy, and most SA-DCP users felt this was the case. Although the results of the survey and the remote monitoring data of the CPAP devices showed that there was no statistically significant difference in adherence to CPAP therapy between the groups, it would have been beneficial to ask SA-DCP users why most of them felt that the SA-DCP had increased their adherence to CPAP therapy. However, this was not investigated in this study. The sample size of the interviewed HCPs was small in this study. However, the answers to the HCPs were mostly consistent. Most of them thought there were no significant changes to their workload and work practices; there were still many phone calls from patients. At the time of writing, the SA-DCP did not enable the automatic collection of log data about the number of electronic messages. If this information had been available, it would have enabled a better comparison between the volumes of phone calls and SA-DCP messages.

Conclusions

According to this study, not all the goals set for introducing the SA-DCP have been achieved. The HCPs who participated in the study could not define significant changes in their workload and work practices after the introduction of SA-DCP. The SA-DCP has brought more flexibility to HCPs' work practices regarding patient communication. Despite using SA-DPC, some patients still wanted to communicate with HCPs by phone. Adherence to CPAP therapy was high in both SA-DCP users and nonusers. Patients' lower digital literacy and older age were the most significant factors explaining the nonuse of the SA-DCP. In the future, more attention should be paid to how these user groups should be considered in the design and introduction of the DCPs.

Acknowledgments

This research was funded by the Strategic Research Council (336278 and 358417). The authors are grateful to the study participants who generously gave their time to share their views on the sleep apnea digital care pathway. The authors would also like to thank the health care professionals from Oulu University Hospital who helped in the practical implementation of the study.

Authors' Contributions

All authors participated sufficiently in the work to take public responsibility for the appropriate portions of the content. JH, TH, and JR were responsible for study conception and design. JH, HM, PL, and TH performed data acquisition. JH analyzed and interpreted the data. JH, TH, MT, PL, and JR drafted the manuscript.

Conflicts of Interest

PL and HM have been involved in the national and regional development of the digital care pathways of Health Village. All other authors declare no other conflicts of interest.

References

1. Palomäki M, Saaresranta T, Anttalainen U, Partinen M, Keto J, Linna M. Multimorbidity and overall comorbidity of sleep apnoea: a Finnish nationwide study. ERJ Open Res 2022 Apr;8(2):1-12 [FREE Full text] [doi: 10.1183/23120541.00646-2021] [Medline: 35677395]
2. Saaresranta T, Anttalainen U, Brander P, Helin-Salmivaara A, Hiltunen L, Himanen SL, et al. Uniapnea (obstruktiivinen uniapnea aikuisilla) [Update on current care guidelines: obstructive sleep apnoea syndrome in adults]. Lääketieteellinen

Aikakauskirja Duodecim 2010;126(22):2633-2635 [FREE Full text] [doi: 10.1183/9781849840248.013911] [Medline: 21188878]
3. Myllylä M, Anttalainen U, Saaresranta T, Laitinen T. Motor vehicle accidents in CPAP-compliant obstructive sleep apnea patients-a long-term observational study. Sleep Breath 2020 Sep;24(3):1089-1095 [FREE Full text] [doi: 10.1007/s11325-020-02023-2] [Medline: 32060778]
4. Giles TL, Lasserson TJ, Smith BJ, White J, Wright J, Cates CJ. Continuous positive airways pressure for obstructive sleep apnoea in adults. Cochrane Database Syst Rev 2006 Jul 19(3):CD001106. [doi: 10.1002/14651858.CD001106.pub3] [Medline: 16855960]
5. Mattila T, Hasala H, Kreivi HR, Avellan-Hietanen H, Bachour A, Herse F, et al. Changes in the societal burden caused by sleep apnoea in Finland from 1996 to 2018: a national registry study. Lancet Reg Health Eur 2022 May;16:100338 [FREE Full text] [doi: 10.1016/j.lanepe.2022.100338] [Medline: 35252943]
6. Hanlon P, Daines L, Campbell C, McKinstry B, Weller D, Pinnock H. Telehealth interventions to support self-management of long-term conditions: a systematic metareview of diabetes, heart failure, asthma, chronic obstructive pulmonary disease, and cancer. J Med Internet Res 2017 May 17;19(5):e172 [FREE Full text] [doi: 10.2196/jmir.6688] [Medline: 28526671]
7. Liljamo P, Wahlberg J, Mikkonen H, Reponen J. A digital care pathway to access healthcare without time and place restrictions. Stud Health Technol Inform 2021 Dec 15;284:163-165. [doi: 10.3233/SHTI210690] [Medline: 34920495]
8. eHealth. World Health Organization. URL: http://www.emro.who.int/health-topics/ehealth/ [accessed 2023-03-23]
9. O'Reilly-Jacob M, Mohr P, Ellen M, Petersen C, Sarkisian C, Attipoe S, et al. Digital health and low-value care. Healthc (Amst) 2021 Jun;9(2):100533 [FREE Full text] [doi: 10.1016/j.hjdsi.2021.100533] [Medline: 33714891]
10. Kujala S, Ammenwerth E, Kolanen H, Ervast M. Applying and extending the FITT framework to identify the challenges and opportunities of successful eHealth services for patient self-management: qualitative interview study. J Med Internet Res 2020 Aug 12;22(8):e17696 [FREE Full text] [doi: 10.2196/17696] [Medline: 32784175]
11. Koivuluoma M, Haverinen J, Reponen J. Digital health services as part of care processes. FinJeHeW 2022 Nov 07;14(3):284-298. [doi: 10.23996/fjhw.111788]
12. Heijsters F, Santema J, Mullender M, Bouman MB, Bruijne MD, van Nassau F. Stakeholders barriers and facilitators for the implementation of a personalised digital care pathway: a qualitative study. BMJ Open 2022 Nov 23;12(11):e065778 [FREE Full text] [doi: 10.1136/bmjopen-2022-065778] [Medline: 36418140]
13. Saranto K, Kivekäs E, Kuosmanen P, Kinnunen UM. Electronic health services in the patients' daily activities - willingness to use health village services. Stud Health Technol Inform 2018;247:586-590. [Medline: 29678028]
14. Ciere Y, van der Vaart R, van der Meulen-De Jong AE, Maljaars PW, van Buul AR, Koopmans JG, et al. Implementation of an eHealth self-management care path for chronic somatic conditions. Clin eHealth 2019;2:3-11. [doi: 10.1016/j.ceh.2019.04.001]
15. Deng Z, Hong Z, Ren C, Zhang W, Xiang F. What predicts patients' adoption intention toward mHealth services in China: empirical study. JMIR Mhealth Uhealth 2018 Aug 29;6(8):e172 [FREE Full text] [doi: 10.2196/mhealth.9316] [Medline: 30158101]
16. Jääskelä J, Haverinen J, Kaksonen R, Reponen J, Halunen K, Tokola T. Digi-HTA, assessment framework for digital healthcare services: information security and data protection in health technology – initial experiences. FinJeHeW 2022;14(1):19-30 [FREE Full text] [doi: 10.23996/fjhw.111776]
17. Pouls BP, Vriezekolk JE, Bekker CL, Linn AJ, van Onzenoort HA, Vervloet M, et al. Effect of interactive eHealth interventions on improving medication adherence in adults with long-term medication: systematic review. J Med Internet Res 2021 Jan 08;23(1):e18901 [FREE Full text] [doi: 10.2196/18901] [Medline: 33416501]
18. Aardoom JJ, Loheide-Niesmann L, Ossebaard HC, Riper H. Effectiveness of eHealth interventions in improving treatment adherence for adults with obstructive sleep apnea: meta-analytic review. J Med Internet Res 2020 Feb 18;22(2):e16972 [FREE Full text] [doi: 10.2196/16972] [Medline: 32130137]
19. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
20. Manwaring JL, Bryson SW, Goldschmidt AB, Winzelberg AJ, Luce KH, Cunning D, et al. Do adherence variables predict outcome in an online program for the prevention of eating disorders? J Consult Clin Psychol 2008 Apr;76(2):341-346. [doi: 10.1037/0022-006X.76.2.341] [Medline: 18377129]
21. Li S, Cui G, Yin Y, Xu H. Associations between health literacy, digital skill, and eHealth literacy among older Chinese adults: a cross-sectional study. Digit Health 2023;9:20552076231178431 [FREE Full text] [doi: 10.1177/20552076231178431] [Medline: 37256010]
22. Haverinen J, Keränen N, Falkenbach P, Maijala A, Kolehmainen T, Reponen J. Digi-HTA: health technology assessment framework for digital healthcare services. FinJeHeW 2019 Nov 02;11(4):326-341. [doi: 10.23996/fjhw.82538]
23. Haverinen J, Keränen N, Tuovinen T, Ruotanen R, Reponen J. National development and regional differences in eHealth maturity in Finnish public health care: survey study. JMIR Med Inform 2022 Aug 12;10(8):e35612 [FREE Full text] [doi: 10.2196/35612] [Medline: 35969462]

24. Ruotanen R, Kangas M, Tuovinen T, Keränen N, Haverinen J, Reponen J. Finnish e-health services intended for citizens – national and regional development. FinJeHeW 2021 Oct 25;13(3):283-301 [FREE Full text] [doi: 10.23996/fjhw.109778]
25. Haverinen J, Turpeinen M, Falkenbach P, Reponen J. Implementation of a new Digi-HTA process for digital health technologies in Finland. Int J Technol Assess Health Care 2022 Aug 19;38(1):e68. [doi: 10.1017/S0266462322000502] [Medline: 35983625]
26. Digital care pathways. Health Village. 2022. URL: https://www.terveyskyla.fi/en/mypath/digital-care-pathways [accessed 2023-04-01]
27. Tuomikoski K, Liljamo P, Reponen J, Kanste O. The effects of digital care pathways on the healthcare professionals: processes in specialized health care. FinJeHeW 2022 Nov 07;14(3):326-338 [FREE Full text] [doi: 10.23996/fjhw.112648]
28. Väätäinen S, Soini E, Arvonen S, Suvitie P. Cost-benefit predictions of the Virtual Hospital 2.0 digitalized secondary care path for women with endometriosis in Turku area and in Finland. In: Proceedings of the 24th Finnish National Conference on Telemedicine and eHealth. 2019 Nov Presented at: eHealth '19; April 2-3, 2019; Kuopio, Finland. [doi: 10.1016/j.jval.2019.09.1200]
29. Liljamo P, Säilynoja H, Tuomikoski K, Henner A, Koivunen K. Patient-centered development of a digital care pathway for arrhythmia patients. Stud Health Technol Inform 2020 Nov 23;275:127-131. [doi: 10.3233/SHTI200708] [Medline: 33227754]
30. Kupila SK, Venäläinen MS, Suojanen LU, Rosengård-Bärlund M, Ahola AJ, Elo LL, et al. Weight loss trajectories in healthy weight coaching: cohort study. JMIR Form Res 2022 Mar 09;6(3):e26374 [FREE Full text] [doi: 10.2196/26374] [Medline: 35262494]
31. Niemi A, Hupli MI, Koivunen M. The use of electronic communication for patient - professional interaction – nursing staff's point of view. FinJeHeW 2016 Mar;8(4):200-215. [doi: 10.1111/jan.12529] [Medline: 25250739]
32. OYS Uniapnea digihoitopolku. Health Village. 2022. URL: https://www.terveyskyla.fi/omapolku/digihoitopolut#termid=95 [accessed 2023-04-01]
33. Registration and login instructions. Health Village. 2022. URL: https://www.terveyskyla.fi/en/mypath/registration-and-login-instructions [accessed 2023-08-22]
34. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
35. Sarja J. Not all numerical data leads to quantitative research: introduction to qualitative quantification analysis. Int J Quant Qual Res Methods 2020;8(2):19-24 [FREE Full text]
36. Ihmiseen kohdistuvan tutkimuksen eettiset periaatteet ja ihmistieteiden eettinen ennakkoarviointi Suomessa. Tutkimuseettinen Neuvottelukunta. 2019. URL: https://tenk.fi/sites/default/files/2021-01/Ihmistieteiden_eettisen_ennakkoarvioinnin_ohje_2020.pdf [accessed 2023-04-01]
37. Laukka E, Huhtakangas M, Heponiemi T, Kujala S, Kaihlanen A, Gluschkoff K, et al. Health care professionals' experiences of patient-professional communication over patient portals: systematic review of qualitative studies. J Med Internet Res 2020 Dec 08;22(12):e21623 [FREE Full text] [doi: 10.2196/21623] [Medline: 33289674]
38. Miller Jr DP, Latulipe C, Melius KA, Quandt SA, Arcury TA. Primary care providers' views of patient portals: interview study of perceived benefits and consequences. J Med Internet Res 2016 Jan 15;18(1):e8 [FREE Full text] [doi: 10.2196/jmir.4953] [Medline: 26772771]
39. Kristiansen E, Atherton H, Austad B, Bergmo T, Norberg BL, Zanaboni P. Older patients' experiences of access to and use of e-consultations with the general practitioner in Norway: an interview study. Scand J Prim Health Care 2023 Mar;41(1):33-42 [FREE Full text] [doi: 10.1080/02813432.2022.2161307] [Medline: 36592342]
40. Mannheim I, Schwartz E, Xi W, Buttigieg SC, McDonnell-Naughton M, Wouters EJ, et al. Inclusion of older adults in the research and design of digital technology. Int J Environ Res Public Health 2019 Oct 02;16(19):3718 [FREE Full text] [doi: 10.3390/ijerph16193718] [Medline: 31581632]
41. Jenssen BP, Mitra N, Shah A, Wan F, Grande D. Using digital technology to engage and communicate with patients: a survey of patient attitudes. J Gen Intern Med 2016 Jan;31(1):85-92 [FREE Full text] [doi: 10.1007/s11606-015-3517-x] [Medline: 26385117]
42. Ye J, Rust G, Fry-Johnson Y, Strothers H. E-mail in patient-provider communication: a systematic review. Patient Educ Couns 2010 Aug;80(2):266-273 [FREE Full text] [doi: 10.1016/j.pec.2009.09.038] [Medline: 19914022]
43. Zanaboni P, Fagerlund AJ. Patients' use and experiences with e-consultation and other digital health services with their general practitioner in Norway: results from an online survey. BMJ Open 2020 Jun 17;10(6):e034773 [FREE Full text] [doi: 10.1136/bmjopen-2019-034773] [Medline: 32554721]
44. Liederman EM, Morefield CS. Web messaging: a new tool for patient-physician communication. J Am Med Inform Assoc 2003;10(3):260-270 [FREE Full text] [doi: 10.1197/jamia.M1259] [Medline: 12626378]
45. Lin C, Wittevrongel L, Moore L, Beaty BL, Ross SE. An internet-based patient-provider communication system: randomized controlled trial. J Med Internet Res 2005 Aug 05;7(4):e47 [FREE Full text] [doi: 10.2196/jmir.7.4.e47] [Medline: 16236699]
46. Kuusisto H, Keränen T, Oksanen K, Saranto K. Neurological referral patients are satisfied with telemedicine consultation. Suomen lääkärilehti 2013;68(9):667-671.

47. Car J, Sheikh A. Telephone consultations. BMJ 2003 May 03;326(7396):966-969 [FREE Full text] [doi: 10.1136/bmj.326.7396.966] [Medline: 12727771]
48. Mechanic OJ, Persaud Y, Kimball AB. Telehealth systems. In: StatPearls [Internet]. Treasure Island, FL: StatPearls Publishing; 2022.
49. Voorheis P, Bhuiya AR, Kuluski K, Pham Q, Petch J. Making sense of theories, models, and frameworks in digital health behavior change design: qualitative descriptive study. J Med Internet Res 2023 Mar 15;25:e45095 [FREE Full text] [doi: 10.2196/45095] [Medline: 36920442]
50. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
51. Teeriniemi AM, Salonurmi T, Jokelainen T, Vähänikkilä H, Alahäivälä T, Karppinen P, et al. A randomized clinical trial of the effectiveness of a web-based health behaviour change support system and group lifestyle counselling on body weight loss in overweight and obese subjects: 2-year outcomes. J Intern Med 2018 Nov 04;284(5):534-545 [FREE Full text] [doi: 10.1111/joim.12802] [Medline: 29974563]
52. Nazi KM. The personal health record paradox: health care professionals' perspectives and the information ecology of personal health record systems in organizational and clinical settings. J Med Internet Res 2013 Apr 04;15(4):e70 [FREE Full text] [doi: 10.2196/jmir.2443] [Medline: 23557596]
53. Blasiak A, Sapanel Y, Leitman D, Ng WY, De Nicola R, Lee VV, et al. Omnichannel communication to boost patient engagement and behavioral change with digital health interventions. J Med Internet Res 2022 Nov 16;24(11):e41463 [FREE Full text] [doi: 10.2196/41463] [Medline: 36383427]
54. Sasseville M, LeBlanc A, Boucher M, Dugas M, Mbemba G, Tchuente J, et al. Digital health interventions for the management of mental health in people with chronic diseases: a rapid review. BMJ Open 2021 Apr 05;11(4):e044437 [FREE Full text] [doi: 10.1136/bmjopen-2020-044437] [Medline: 33820786]
55. Das A, Faxvaag A, Svanæs D. The impact of an eHealth portal on health care professionals' interaction with patients: qualitative study. J Med Internet Res 2015 Nov 24;17(11):e267 [FREE Full text] [doi: 10.2196/jmir.4950] [Medline: 26601678]
56. Schreurs K, Quan-Haase A, Martin K. Problematizing the digital literacy paradox in the context of older adults’ ICT use: aging, media discourse, and self-determination. Can J Commun 2017 May 23;42(2):359-377. [doi: 10.22230/cjc.2017v42n2a3130]
57. Wildenbos GA, Peute L, Jaspers M. Aging barriers influencing mobile health usability for older adults: a literature based framework (MOLD-US). Int J Med Inform 2018 Jun;114:66-75. [doi: 10.1016/j.ijmedinf.2018.03.012] [Medline: 29673606]
58. Gilbert BJ, Goodman E, Chadda A, Hatfield D, Forman DE, Panch T. The role of mobile health in elderly populations. Curr Geri Rep 2015 Aug 21;4(4):347-352. [doi: 10.1007/s13670-015-0145-6]

Abbreviations

AEC: assistive equipment center
AHI: apnea-hypopnea index
CPAP: continuous positive airway pressure
DCP: digital care pathway
HCP: health care professional
ICT: information and communications technology
OUH: Oulu University Hospital
SA-DCP: sleep apnea digital care pathway

Edited by A Kushniruk; submitted 02.04.23; peer-reviewed by S Kupila, J Roland, Y Sapanel; comments to author 23.07.23; revised version received 31.08.23; accepted 20.01.24; published 22.02.24.

Please cite as:
Haverinen J, Harju T, Mikkonen H, Liljamo P, Turpeinen M, Reponen J
Digital Care Pathway for Patients With Sleep Apnea in Specialized Care: Mixed Methods Study
JMIR Hum Factors 2024;11:e47809
URL: https://humanfactors.jmir.org/2024/1/e47809
doi:10.2196/47809
PMID:38386368

Original Paper

User Experience Evaluation of a Spinal Surgery Robot: Workload, Usability, and Satisfaction Study

Hyeonkyeong Choi[1], BSPH; Seunghee Kim[1], BSPH; Wonseuk Jang[1,2], PhD

[1]Department of Medical Device Engineering and Management, Yonsei University College of Medicine, Seoul, Republic of Korea

[2]Medical Device Usability Research Center, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea

Corresponding Author:
Wonseuk Jang, PhD
Department of Medical Device Engineering and Management
Yonsei University College of Medicine
20 63rd Road, Unju-ro, Gangnam-gu
Seoul, 06229
Republic of Korea
Phone: 82 1091485404
Fax: 82 02 2019 5442
Email: WS.JANG@yuhs.ac

Abstract

Background: Robotic spine surgery has continued to evolve since its US Food and Drug Administration approval in 2004, with products now including real-time video guidance and navigation during surgery. As the market for robotic surgical devices evolves, it is important to consider usability factors.

Objective: The primary objective of this study was to determine the user experience of a surgical-assistive robotic device. The secondary objective was to evaluate workload, usability, the After-Scenario Questionnaire (ASQ), and the System Usability Scale (SUS). In addition, this study compares the workload, usability, and satisfaction survey of the device among different occupational groups using the device.

Methods: Doctors (n=15) and nurses (n=15), the intended users of the surgical assistant robot, participated in the usability evaluation. Participants performed essential scenarios for the surgical assistant robot and provided scenario-specific satisfaction (ASQ), workload (NASA Task Load Index), and usability (SUS) scores.

Results: Both doctors and nurses had task success rates of 85% or higher for each scenario. ASQ results showed that both doctors and nurses were least satisfied with ease of completing the task of registration (group 1: mean 4.73, SD 1.57 and group 2: mean 4.47, SD 1.8), amount of time it took (group 1: mean 4.47, SD 1.63 and group 2: mean 4.40, SD 2.09), and support information satisfaction (group 1: mean 5.13, SD 1.50 and group 2: mean 5.13, SD 1.89). All participants had low workloads, and the overall Task Load Index score had a *P* value of .77, which is greater than .05. The SUS results showed that the overall usability mean for doctors was 64.17 (SD 16.52) and the mean for nurses was 61.67 (SD 19.18), with a *P* value of .84, which is greater than .05, indicating no difference between the 2 groups.

Conclusions: In this study, doctors and nurses evaluated the interaction of the device in a simulated environment, the operating room. By evaluating the use experience and usability of the device with real intended users, we can develop a more effective and convenient user interface.

(JMIR Hum Factors 2024;11:e54425) doi:10.2196/54425

KEYWORDS

robot spine surgery; usability; satisfaction; System Usability Scale; surgical navigation systems; robotics; surgery; neurosurgery

Introduction

Background

Spine surgery is used to treat degenerative diseases and deformities of the spine, with 45 million surgeries performed annually in the United States [1]. The use of robotic-assisted navigation is increasing as the number of patients undergoing lumbar spinal fixation increases [2]. Spine surgery typically involves 7 people in the operating room, with an operator surgeon, a surgical first assistant (who may be a doctor or physician assistant nurse, depending on operating room staffing), and scrub nurse in the sterile area and a circulating nurse and radiologist in the nonsterile area. Nonoperative personnel include an anesthesiologist and an anesthesiologist assistant. The use of robotics in spine surgery is usually reserved for difficult anatomical areas where it is difficult to fix screws blindly. Spinal fusion surgery is the insertion and fixation of pedicle screws into the vertebrae to eliminate pain by preventing movement between vertebrae [3]. It is also used for quick insertion in severe scoliosis, collapsed vertebrae, or long-level fusion in patients with difficult anatomy, usually at the iliac screw, C1, C2, C7, T1, and T2, or for other reasons. This is usually used for kyphosis and scoliosis correction.

Robot Spine Surgery

Surgical navigation systems are used to plan the procedure and guide the surgeon in inserting the screws [1]. Robot spine surgery is popularly used to increase the accuracy of inserting screws in the spine, and the first robot used in spine surgery was the Spine Assist (Mazor Surgical Technologies), which received Food and Drug Administration clearance in 2004 [4]. The third-generation Mazor X system was cleared by the Food and Drug Administration in 2016 and, compared to previous generations, has a robotic arm that is attached to the patient's body and can be viewed through a camera to ensure that the screws are inserted and the robotic arm is moving well [5]. The Mazor X Stealth Edition technology, which adds real-time image guidance and navigation during surgery, was cleared in 2019 and combines the best of both worlds: traditional spinal robotic surgery guidance and real-time software confirmation [5,6]. Figure 1 shows the evaluation device, which consists of a robotic arm, main console, and optional staff console, and is manufactured in South Korea. Like the Mazor X Stealth Edition technology, this product is capable of real-time image guidance and navigation during surgery.

Figure 1. The CUVIS spine robot system: robotic arm (left), main console with navigation optical infrared tracking camera (middle), and the control workstation (right). OTS: optical tracking system.

Usability

According to IEC (International Electrotechnical Commission) 62366-1 [7], usability has the following meaning: a "characteristic of the user interface that facilitates use and thereby establishes effectiveness, efficiency and user satisfaction in the intended use environment." ISO (International Organization for Standardization) 9241-210 [8] defines usability as the "extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use." These documents demonstrate that good usability in medical device design is essential to preventing error-related risks.

Evaluating usability focuses on determining whether the test device is easy for users to use. To evaluate the user experience of the device, we used usability tests and surveys for effectiveness, efficiency, satisfaction, and workload. The usability test was mainly used to identify use errors and efficiency, while the After-Scenario Questionnaire (ASQ) was conducted to evaluate the satisfaction of each task. NASA Task Load Index (NASA TLX) was used to measure the workload of the device, and System Usability Scale (SUS) was used to check the overall system usability of the device. Both ASQ and SUS identified the satisfaction and efficiency of the device, but in this study, ASQ identified the efficiency and satisfaction of each scenario, while SUS evaluated the satisfaction and efficiency of the device in the overall workflow.

Methods

Recruitment

We recruited 30 medical staff from Severance Hospital in South Korea. There were 15 doctors and 15 nurses. The intended users of the device are doctors and operating room nurses. Due to the different tasks that doctors and nurses have to perform when using the device overall, both groups were selected to participate in the usability test. We recruited through recommendations from colleagues and notice on the bulletin board at the Future Medicine Research Center at Gangnam Severance Hospital. Doctors and nurses who have experience using a robotic surgical device or navigation system were selected. For doctors, we selected those with the necessary knowledge of spine surgery, and for nurses, we selected those with experience in the operating room. However, those who had worked in the operating room for less than 1 year were excluded. After confirming these inclusion and exclusion criteria, the screening was conducted.

Testing Procedure

One participant per session participated in the usability test, with the participant completing the assessment in a sequence guided by a facilitator, and an observer in the observation room videotaping the assessment and completing an observation sheet. A total of 2 moderators and 2 observers participated in the evaluation, with one of the moderators acting as a nurse if a doctor participated in the evaluation, and one of the moderators acting as a doctor if a nurse participated in the evaluation. In addition, observers were used to reduce bias by having 2 observers observe the usability test to ensure that one person's opinion was not biased.

The facilitator introduced the participants to the usability test, obtained their informed consent, and trained them on the device for 20 minutes. Participants were allowed to interact with the device as much time as they needed. Afterward, 15 minutes were allowed between the training and evaluation to ensure that the training did not directly influence the evaluation [9]. Doctors and nurses were given different tasks because of the different job duties they do when operating with assessment devices. The tasks given to nurses focus on doctors' instructions from preoperative preparation to surgery, while doctors focus on the surgery itself rather than preparing devices.

Participants then completed the evaluation for 40 minutes, with 8 scenarios (24 tasks) for doctors and 12 scenarios (41 tasks) for nurses. Doctors were asked to complete the following scenarios: preparation for use, preplanning of surgery, fixation of patient marker, scan, registration, verification, revisions of surgical planning, and navigation; nurses were asked to complete the following scenarios: preparation for use, system operation, initializing manipulator, drape, preparation of surgery, scan, registration, verification, planning of surgery, navigation, use of the emergency stop switch, and cleaning up after surgery. Tasks for each scenario are shown in Tables 1 and 2.

Table 1. Use scenarios of doctors.

Use scenario/task number	Task description
Preparation for use	
Task 1	Check the contents related to the emergency button, foot switch, and foot jamming in the user manual.
Task 2	Check the hand jamming label on the robotic arm manipulator.
Preplanning of surgery	
Task 3	Select the spine level as follows: • L4-L, L4-R • L5-L, L5-R
Task 4	After loading the first CT[a] data, check the CT data.
Task 5	Create an implant screw insertion path for target L4-L and L4-R and change the insertion path by moving the screw in the MPR[b] view.
Task 6	Create an implant screw insertion path for target L5-L and L5-R and change the insertion path using the arrow.
Task 7	Check for collision between each screw.
Fixation of patient marker	
Task 8	Attach the patient marker to the patient.
Scan	
Task 9	Attach the registration tool adapter to the end effector.
Task 10	After activating the hand guide function by pressing the AP[c] button, change the position of the end effector according to the guidance on the pop-up window.
Task 11	For C-Arm scan, attach the source calibrator to the end effector in the direction of AP and move the end effector to enable tracking by OTS[d].
Task 12	Check if the ROI[e] includes the calibration marker, and the position and direction of the letters "R," "G," and "J" match the image, and then check pass or fail of registration.
Registration	
Task 13	Perform segmentation to distinguish the surgical target in the image. • L4-L, L4-R • L5-L, L5-R
Task 14	Perform labeling to assign target level information of ROI of 3D image and 2D image segmented for each spine level.
Task 15	Adjust the ROI box so that the ROI of target L4 covers all the L4 vertebra area.
Task 16	Perform 2D and 3D image registration for each spine level.
Verification	
Task 17	After adjusting the CT image to overlay appropriately for target L4, check the registration result using the preview button and select whether to approve it.
Task 18	After selecting whether to approve for target L5, perform image registration again so that the ROI includes all the vertebra area.
Revision of surgical planning	
Task 19	Check the plan on 2D and 3D images, respectively.
Task 20	As a result of planning for the entire target, check whether the robot can move in an area.
Navigation	
Task 21	Insert the screw of target L4-L.
Task 22	Through the [PRE-OP] screen, indicate the values for the insertion depth of the L4-L tapper, the amount of force applied to the end effector, and the patient's movement.
Task 23	Through the [INTRA-OP] screen, indicate the values for the insertion depth of the L4-L screw, the amount of force applied to the end effector, and the patient's movement.
Task 24	Move the end effector to the ready position for screw insertion to the target L4-R.

[a]CT: computed tomography.
[b]MPR: multiplanar reconstruction.
[c]AP: anterior-posterior.
[d]OTS: optical tracking system
[e]ROI: region of interest.

Table 2. Use scenarios of nurses.

Use scenario/task number	Task description
Preparation for use	
Task 1	Check the contents related to the emergency button, foot switch, and foot jamming in the user manual.
Task 2	Check the hand jamming label on the robotic arm manipulator.
Task 3	Check if there are any abnormalities in the exterior of the robot marker frame and robotic arm.
Task 4	Place the main console and staff console in a convenient location during surgery.
Task 5	After checking the device and accessories in the operating room, assemble the marker ball.
Task 6	Assemble the surgical tools such as the tapper's driver and marker.
Task 7	Assemble the surgical tools such as screwdriver and marker.
Task 8	Assemble the clamp to be used to fix the patient marker.
System operation	
Task 9	Connect the power and cables of the robotic arm, main console, and staff console.
Task 10	After connecting the foot switch of the robotic arm, turn on the power of the robotic arm.
Initializing manipulator	
Task 11	After logging in, select the surgical method and imaging device.
Task 12	Check the connection status of the foot switch.
Task 13	After selecting the robot position as "right," initialize the manipulator (required to check movement notification sound and operation LED[a]).
Task 14	Verify that the line laser on the robot marker intersects the area within range (required to check movement notification sound and operation LED).
Drape	
Task 15	Follow the on-screen instructions to drape the patient to prevent infection (proceed in order of manipulator drape, base drape, and robot marker drape).
Task 16	After installing the end effector of the robotic arm, assemble the marker ball where the robot marker drape is installed.
Task 17	Move the manipulator to the ready position.
Preparation for surgery	
Task 18	Move the robotic arm for patient surgery.
Task 19	Check the surgical tools through OTS[b], and if all surgical tools are not checked by the OTS camera, check if they are within the operating area.
Task 20	The robot marker is not being recognized by the OTS camera due to damage to the marker ball. Replace with a new marker ball.
Task 21	Please load the surgical data.
Scan	
Task 22	Check if the ROI[c] includes the calibration marker, and the position and direction of the letters "R," "G," and "J" match the image, and then check pass or fail.
Registration	
Task 23	Perform segmentation to distinguish the surgical target in the image. • L4-L, L4-R • L5-L, L5-R
Task 24	Perform labeling to assign target level information of ROI of the 3D image and 2D image segmented for each spine level.
Task 25	Adjust the box so that the ROI of target L4 covers all of the vertebra area.
Task 26	Perform 2D and 3D image registration for each spine level.
Verification	

Use scenario/task number	Task description
Task 27	Use the slide control at the bottom of the image to check whether the 2D and 3D images match to check the registration result.
Task 28	For target L4, move the CT[d] (DRR[e]) image by using the triangular button to adjust the 2 body images to be similar.
Task 29	For target L4 whose registration result has been adjusted, use the preview button to check the registration result and select whether or not to approve it.
Task 30	After selecting whether or not to approve for target L5, perform image registration again so that the ROI includes all of the vertebra area (target L5: registration failed).
Task 31	After displaying the planned data on the screen through the preview button for each target for which the registration result has been adjusted, check if the registration is completed normally.
Task 32	Depending on the registration result of target L5, select whether or not to approve (target L5: registration completed normally).
Planning of surgery	
Task 33	Check whether the robot can move to the planned position.
Navigation	
Task 34	On the screen, move the end effector of the robotic arm to the planned guide position relative to target L4-L.
Task 35	Move the end effector to the original position for the guide.
Task 36	On the screen, move the end effector of the robotic arm to the planned guide position relative to target L4-R.
Use of the emergency stop switch	
Task 37	(At the moment, the manipulator is positioned too close to the patient.) Press the emergency button.
Task 38	Release the emergency button.
Cleaning up after surgery	
Task 39	Shut down the main console.
Task 40	Shut down the robotic arm.
Task 41	Disconnect the cable.

[a]LED: light emitting diode.

[b]OTS: optical tracking system.

[c]ROI: region of interest.

[d]CT: computed tomography.

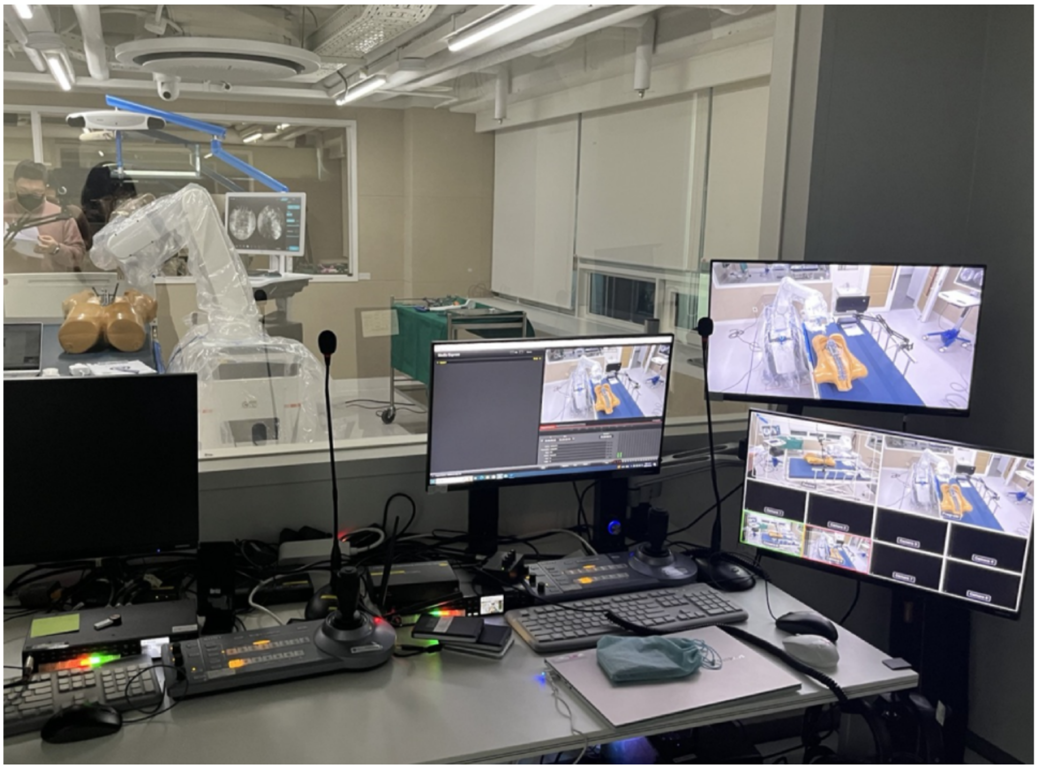
[e]DRR: digitally reconstructed radiograph.

The test environment as shown in Figure 2 is organized to resemble the operating room. Participants used the device following prompts presented on a stand monitor. The test environment was organized similar to an operating room, considering the use environment of the robot spine surgery. An operating room bed, an upper torso dummy, and a patient monitoring device were prepared similar to the actual operating room environment. The temperature and humidity of the evaluation room were measured and recorded right before the evaluation. Similar to a real operating room, the temperature was kept between 20 °C and 24 °C, and the humidity was between 30% and 60%.

The evaluation facilitator guided the participant if they requested assistance with a use scenario, and an observer recorded all participant interactions from outside the test room with a 1-way mirror. The test observation environment setting is shown in Figure 3. The observer used a program from Media Express to record the progress of the usability evaluation. At the end of the evaluation, 3 types of questionnaires were administered.

Figure 2. Test environment: the simulated environment is organized to resemble the operating room in which the evaluator is used.

Figure 3. Test observation environment: we set up monitoring equipment to observe and record the entire evaluation process in real time.



Statistical Analysis

ASQ Measure

After each scenario, the participants completed the ASQ created by Lewis [10] and developed from the ISO 9241-11 standard questionnaire [11]. This is one of the most popular surveys for assessing usability because it is the simplest and its 3 items are easy for participants to understand [11]. As shown in Textbox 1, the ASQ consists of 3 questions, each corresponding to the user's satisfaction with the ease, efficiency according to the time taken to complete the scenario, and validity of the information provided. Participants responded to each question on a 7-point Likert scale [10,12]. Participants rated their satisfaction about the device's usability based on each task scenario [13]. A score of "1" means strongly disagree, and a score of "7" means strongly agree [14]. We found the mean and SD for the 3 questions participants asked ASQ.

Textbox 1. After-Scenario Questionnaire (ASQ).

ASQ1: Overall, I am satisfied with the ease of completing the tasks in this scenario.

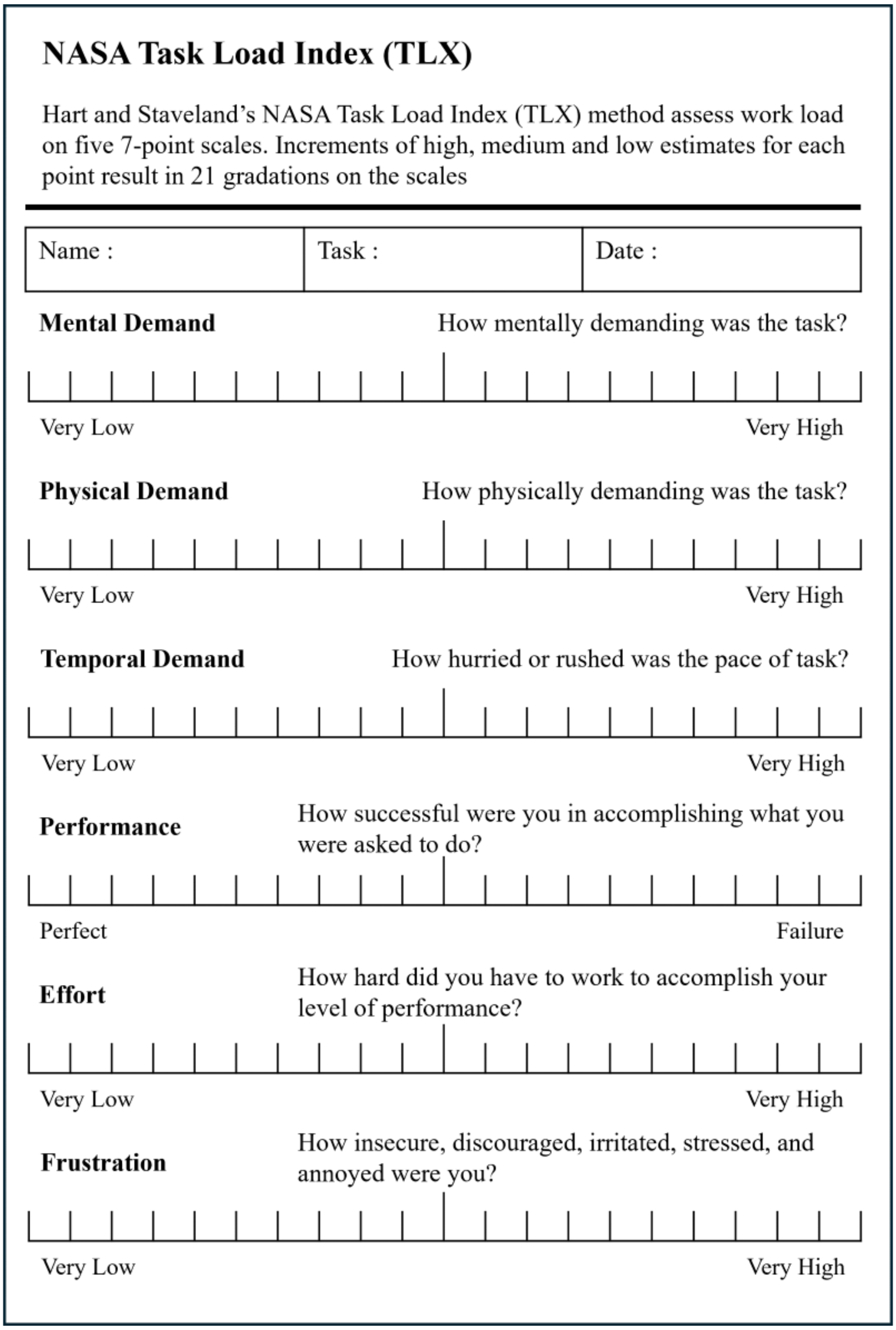
ASQ2: Overall, I am satisfied with the amount of time it took to complete the tasks in this scenario.

ASQ3: Overall, I am satisfied with the support information (digital help, messages, and documentation) when completing the tasks.

NASA TLX Measure

NASA TLX measures cognitive workload, and like usability, workload is a complex construct that determines the amount of physical and mental effort required to use an interface [15,16]. The workload is assessed by the US NASA TLX [15,17]. The most effective way to assess a worker's perceived job difficulty is to ask questions directly to workers who have experienced the job. As shown in Figure 4, the Task Load Index (TLX) uses 6 dimensions to measure workload. The 6 metrics are mental demand, temporal demand, physical demand, performance, effort, and frustration [16,18]. The NASA TLX scores are evaluated by dividing the score into 21 steps, subtracting 1 from the score, and multiplying it by 5 to express it on a scale of 0 to 100 [19,20]. On a scale of 100, when the score is lower, the workload is lower. Less work means a less complex and easier-to-use user interface. On a 100-point scale, the workload can be described as low (0-9), medium (10-29), rather high (30-49), high (50-79), and very high (80-100) [21]. In the NASA TLX, performance assesses satisfaction with task completion, with the lowest number representing perfect and the highest number representing failure [22]. The point system for mental, physical, temporal, effort, and frustration part ranges from very low to very high [15,16,19,22,23].

Figure 4. NASA Task Load Index.



NASA Task Load Index (TLX)

Hart and Staveland's NASA Task Load Index (TLX) method assess work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales

Name :	Task :	Date :

Mental Demand How mentally demanding was the task?

Very Low Very High

Physical Demand How physically demanding was the task?

Very Low Very High

Temporal Demand How hurried or rushed was the pace of task?

Very Low Very High

Performance How successful were you in accomplishing what you were asked to do?

Perfect Failure

Effort How hard did you have to work to accomplish your level of performance?

Very Low Very High

Frustration How insecure, discouraged, irritated, stressed, and annoyed were you?

Very Low Very High

SUS Measure

As shown in Textbox 2, the SUS consists of 10 items that assess the participant's level of agreement with the overall usability of the system, with odd-numbered items being positive and even-numbered items being negative [24]. SUS is the most commonly used usability assessment questionnaire [24]. Participants responded to each item on a 5-point Likert scale [13]. The scale ranges from 1=strongly disagree to 5=strongly agree [25,26]. To calculate the SUS score from the points acquired from the 5-point Likert scale, the following subtractions were used. For odd-numbered items, subtract 1 from the user response, and for even-numbered items, subtract the user responses from 5. With this calculation, the value range changes from 0 to 4. The most positive response is 4. The scores from each converted response were multiplied by 2.5 to a total possible point of 100. The SUS is percentage-based and divided into 5 levels: A (>80.3), B (68-80.3), C (68), D (51-68), and F (<51) [24]. A score of 85 is considered very good usability, and a score of 68-84 is considered good usability [25,26].

Textbox 2. System Usability Scale (SUS) items.

SUS1: I think that I would like to use this system frequently.

SUS2: I found the system unnecessarily complex.

SUS3: I thought the system was easy to use.

SUS4: I think that I would need the support of a technical person to be able to use this system.

SUS5: I found the various functions in this system were well integrated.

SUS6: I thought there was too much inconsistency in this system.

SUS7: I would imagine that most people would learn to use this system very quickly.

SUS8: I found the system very cumbersome to use.

SUS9: I felt very confident using the system.

SUS10: I needed to learn a lot of things before I could get going with this system.

Data Analysis

ASQ, NASA TLX, and SUS results were computed using SPSS (version 22; IBM Corp) [27]. Descriptive statistics were performed on for doctor and nurse characteristics. Doctors and nurses were compared on age, gender, work experience, and use experience with similar devices. For the questionnaire items, values were compared between groups using 2-tailed t tests for normality and Mann-Whitney U tests for nonparametric tests. Figures are presented as the mean and SD, and $P<.05$ was considered significant.

Ethical Considerations

This study was approved by the institutional review board of Yonsei University Health System, Gangnam Severance Hospital (3-2022-0493). All the participants who passed the screening signed an informed consent form. Furthermore, all information collected about the participants was anonymized. This study complied with the Code of Ethics. Participants received monetary compensation for participating in the evaluation.

Results

User Statistics

In total, 15 doctors and 15 nurses, each representing the intended users of the surgical assistant, participated in the evaluation. Participants were recruited from doctors and nurses at Severance Hospital. Table 3 shows the sociodemographic characteristics of participants. Both doctors and nurses were between the ages of 30 and 39 years. For both doctors and nurses, those with different experience levels were recruited, and opinions were collected from all the participants. In particular, those with more experience with surgical devices were able to gather relevant opinions because they were more familiar with the device or the existing surgical methods, while those with less experience focused on whether the device was easy to use without much experience. Doctors' professional experience ranged from 2 to 20 years, with an average of 7.53 (SD 5.45) years of professional experience. The nurses' professional experience ranged from 5 to 26 years, with an average of 12.93 (SD 6.43) years. The surgical assistants had used Medtronic (Medtronic), Stryker (Stryker Corp), and Curexo (Curexo, Inc), with an average of 3 (SD 2.36) years of experience.

Table 3. Sociodemographic characteristics and experience of the test participants (N=30).

Variable	Group 1 (doctors), n (%)	Group 2 (nurses), n (%)
Sociodemographic characteristics		
Age (years)		
20-29	3 (20)	1 (7)
30-39	10 (67)	8 (53)
40-49	1 (7)	6 (40)
50-59	1 (7)	0 (0)
Sex		
Male	14 (93)	6 (40)
Female	1 (7)	9 (60)
Work experience		
More than 1 year, less than 5 years	7 (47)	0 (0)
More than 5 years, less than 10 years	3 (20)	4 (27)
More than 10 years, less than 15 years	3 (20)	6 (40)
More than 15 years, less than 20 years	1 (7)	1 (7)
More than 20 years	1 (7)	4 (27)
Use experience with similar devices		
Device name		
Medtronic	12 (80)	7 (47)
Stryker	1 (7)	0 (0)
Curexo	2 (13)	8 (53)
Use experience		
Less than 1 year	6 (40)	3 (20)
More than 1 year, less than 3 years	5 (33)	8 (53)
More than 3 years, less than 5 years	3 (20)	4 (27)
More than 5 years, less than 10 years	0 (0)	0 (0)
More than 10 years	1 (7)	0 (0)

Task Completion

The 15 doctors performed 24 tasks within 8 large scenarios, while the nurses performed a total of 41 tasks within 12 scenarios. As shown in Table 4, for the doctors, all 8 scenarios had a success rate of 90% or higher, with the lowest success rate for the revising a surgical plan scenario. As shown in Table 5, for nurses, all 11 scenarios except the surgical plan had a success rate of 90% or higher, with the planning of surgery scenario having an 87% success rate.

Table 4. Task completion rate in doctors.

	Task pass rate (%)	Task failure rate (%)
Preparations for use	96.67	3.33
Preplanning of surgery	98.67	1.33
Fixation of patient marker	100	0
Scan	96.67	3.33
Registration	100	0
Verification	96.67	3.33
Revision of surgical planning	93.33	6.67
Navigation	98.33	1.67

Table 5. Task completion rate in nurses.

	Task pass rate (%)	Task failure rate (%)
Preparations for use	97.50	2.5
System operations	96.67	3.33
Initialization manipulator	100	0
Drape	93.33	6.67
Preparation for surgery	91.67	8.33
Scan	100	0
Registration	98.33	1.67
Verification	91.11	8.89
Revision of surgical planning	86.67	13.33
Navigation	100	0
Use of the emergency stop switch	100	0
Cleaning up after surgery	95.56	4.44

Among the scenarios in which the doctors did not successfully complete a task during the assessment, the critical tasks were as follows: in task 11, the participant failed to follow the process of attaching the source calibrator in the opposite direction to track the optical tracking system and did not recognize the correct attachment method. In addition, pressing the anterior-posterior scan button and moving the end effector closer to the dummy proceeded correctly, but before attaching the source calibrator, the optical tracking system process could not be performed because it did not proceed in the existing pop-up window and proceeded to the data acquisition step. In task 18, participants selected the target but were unable to click the "Re-matching" button. To proceed with rematching, a target needs to be selected and pressed, but the Re-matching button could not be clicked because the target was not selected. In task 21, the participant did not recognize whether the robot movement was completed by continuously pressing the foot switch without releasing it. In this case, the participant said that he was unable to perform the task because there was no indication on the screen that the robot's movement was complete, and there was no visual or audible user interface.

The nurse was unable to complete task 28 due to difficulty using the image adjustment feature. Participants were asked to move the computed tomography image and adjust the body image to be similar but could not comprehend how to use the "Adjustment" function or the "Re-matching" function (the user did not recognize the intended function itself). Even when the "Adjustment" function was used, it was observed that the user could not use the "Adjustment" function in the way intended by trying to adjust the overlayed screen itself rather than adjusting the screen by pressing the button. If an accurate match is not made, the manipulator may move to a different location than the user's target location, causing potential harm. In addition, nurses had difficulty using the reassembly feature of task 30. The "Re-matching" function could not be used because the target was released while pressing the "Disapproved" button in the rematching task, or the "Adjustment" function was used rather than using the "Re-matching" function. Participants failed to perform the task because they did not recognize that the "Re-matching" function could only be used by resetting the target that was released when pressing the "Disapproved" button, or that "Re-matching" meant rematching. This caused potential harm by moving the manipulator to a location different from the user's target location. Nurses were unsuccessful in tasks such as registration, verification, and navigation because these tasks are usually performed through doctors' orders. During the scenarios, there were no given orders, forcing the nurses to make their own decisions, which they are not accustomed to.

Overall, 4 doctors said that when creating a screw position in the planning stage, the position is created in a completely different part from the actual location; thus, it would be better if the position could be created closer to the target, and when moving the position, that it would be better to be able to check other position paths at the same time. In total, 7 doctors said it would be better if there was notification or guidance for the arrival of the robot arm at the target so that moving to the guided position can be recognized. In addition, 5 doctors and 8 nurses found that in the overall process of selecting and adjusting the region of interest (ROI) box to the target area, it was inconvenient to select and release the box, and that it was difficult to adjust because of its excessive rotation.

Usability (ASQ)

After the usability evaluation, doctors and nurses were surveyed using the ASQ for each scenario. For both doctors and nurses, the ASQ for registration was divided into 2 parts: first, segmentation and labeling, and second, ROI setting and image matching. As shown in Tables 6 and 7, among the registration items, both doctors and nurses had the lowest scores for the ROI setting and image matching, followed by ease of completing the task (group 1: mean 4.73, SD 1.57 and group 2: mean 4.47, SD 1.89), amount of time it took (group 1: mean 4.47, SD 1.63 and group 2: mean 4.40, SD 2.09), and support information satisfaction (group 1: mean 5.13, SD 1.50 and group 2: mean 5.13, SD 1.89). The doctors' opinions were mainly that it was inconvenient to have to click on the line precisely; thus, the ease of adjustment should be improved. Nurses reported that they were less sensitive to the 360-degree rotation button at the top of the ROI box and had difficulty clarifying the image while adjusting the ROI box.

Table 6. After-Scenario Questionnaire result in group 1 (doctors).

	Ease of completing the task, mean (SD)	Amount of time it took, mean (SD)	Support information satisfaction, mean (SD)
User manual	5.56 (1.12)	5.54 (1.28)	5.63 (1.19)
Preplanning of surgery	5.80 (0.98)	5.53 (1.26)	5.93 (0.77)
Fixation of patient marker	6.00 (0.97)	6.40 (0.61)	—[a]
Scan	6.00 (0.63)	5.60 (0.95)	5.93 (0.77)
Registration 1 (segmentation and labeling)	5.73 (1.18)	5.07 (1.53)	6.00 (0.82)
Registration 2 (ROI[b] setting and image matching)	4.73 (1.57)	4.47 (1.63)	5.13 (1.50)
Verification	5.60 (1.14)	5.40 (1.31)	5.73 (1.00)
Revision of surgical planning	5.80 (0.83)	5.93 (0.68)	5.93 (1.29)
Navigation	5.67 (1.07)	5.93 (0.77)	5.33 (1.45)

[a]Not available; fixation of patient markers was not surveyed because they do not have any on-screen information.

[b]ROI: region of interest.

Table 7. After-Scenario Questionnaire result in group 2 (nurses).

	Ease of completing the task, mean (SD)	Amount of time it took, mean (SD)	Support information satisfaction, mean (SD)
User manual	6.00 (1.23)	5.80 (1.63)	5.97 (1.23)
Preparations for use	6.47 (0.62)	6.07 (1.12)	6.40 (1.02)
System operations	6.27 (0.85)	6.20 (0.98)	6.2 (1.11)
Initialization manipulator	6.13 (1.15)	5.87 (1.45)	5.87 (1.45)
Drape	5.27 (1.84)	5.67 (1.62)	5.80 (1.51)
Preparation for surgery	6.13 (0.88)	6.07 (0.93)	6.27 (0.93)
Scan	6.07 (0.85)	5.87 (0.88)	6.07 (1.06)
Registration 1 (segmentation and labeling)	5.53 (1.50)	4.93 (1.81)	5.67 (1.85)
Registration 2 (ROI[a] setting and image matching)	4.47 (1.89)	4.40 (2.09)	5.13 (1.89)
Verification	4.93 (1.77)	5.20 (1.38)	5.40 (1.74)
Navigation	6.60 (0.61)	6.60 (0.61)	6.60 (0.61)
Use of the emergency stop switch	6.73 (0.44)	6.60 (0.61)	6.73 (0.44)
Cleaning up after surgery	6.47 (0.88)	6.40 (1.02)	6.40 (1.02)

[a]ROI: region of interest.

Tables S1 and S2 in Multimedia Appendix 1 show categorization by use experience with similar devices. The ASQ results did not show significant differences in satisfaction based on use experience and years of experience with robotic surgical systems. For doctors, those with more than 3 years of experience using robotic surgical systems found it easier and faster to perform tasks. For nurses, participants with more experience using similar devices scored higher than those with less than 3 years of experience on the need to prepare before surgery.

Workload (NASA TLX)

Table 8 shows the results of the workload of the assistive robotic surgery devices by occupational group for mental demand, physical demand, temporal demand, performance, effort, frustration, and overall TLX. The Mann-Whitney *U* test comparing the TLX scores of the doctors and nurses, including mental demand (*P*=.81), physical demand (*P*=.90), temporal demand (*P*=.87), performance (*P*=.81), and frustration (*P*=.81) and the independent 2-sample *t* test comparing the TLX scores of effort (*P*=.64) and overall TLX (*P*=.77) showed no significant differences in the scores. Doctors' workload levels were generally in the medium (10-29), and nurses were also in the medium (10-29) except for effort.

Table 8. Result of NASA Task Load Index.

	Group 1[a] (n=15), mean (SD)	Group 2[b] (n=15), mean (SD)	*t* test[c] (*df*)	*U* test[d]	*P* value[e]
Mental demand	30.67 (25.97)	29.33 (29.39)	N/A[f]	106.5	.81
Physical demand	13.00 (15.09)	11.33 (12.17)	N/A	109.5	.90
Temporal demand	15.00 (17.63)	22.33 (30.58)	N/A	108.5	.87
Performance	25.00 (28.09)	28.33 (26.16)	N/A	106	.81
Effort	29.67 (24.60)	34.33 (28.78)	–0.477 (28)	N/A	.64
Frustration	19.00 (22.22)	18.33 (26.70)	N/A	106	.81
Overall Task Load Index	22.06 (18.58)	24.00 (16.96)	–0.299 (28)	N/A	.77

[a]Group 1: doctors.

[b]Group 2: nurses.

[c]Because the data were normally distributed, a independent 2-sample *t* test was used.

[d]Because the data were not normally distributed, the Mann-Whitney *U* test was performed.

[e]*P* values were determined with the independent 2-sample *t* test and Mann-Whitney *U* test for continuous variables.

[f]N/A: not applicable.

Figure 5 shows a boxplot of the NASA TLX results for doctors and nurses. In addition, Figure 6 shows the NASA TLX results for all evaluation participants (doctors and nurses). The box plots show the maximum (45-100), median (5-25) minimum (0), first quartile (0-17.5), and third quartile (12.5-52.5), with the center box showing the median of 50% of the cases. When comparing the workload of the doctors and nurses, there was no significant difference as shown in Table 8, and we can see that 3 categories, physical demand, temporal demand, and frustration, have lower workloads than the others.

Figure 5. Workload results by group: distribution of the NASA TLX scores for doctors and nurses. TLX: Task Load Index.

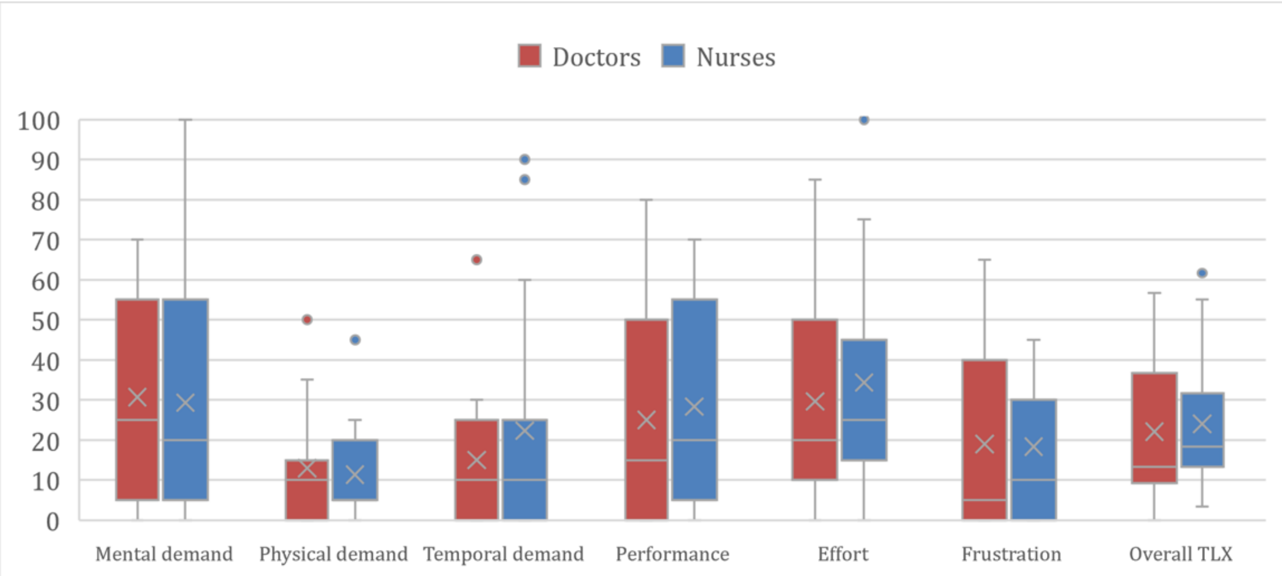
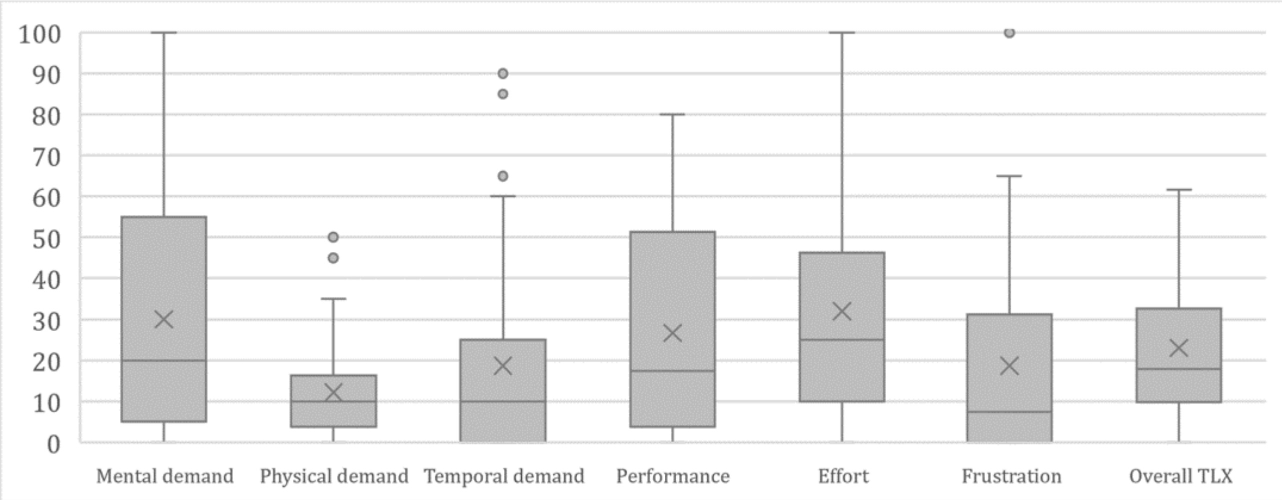


Figure 6. Workload results: distribution of the overall NASA TLX scores. TLX: Task Load Index.



In Tables 9 and 10, NASA TLX scores were compared based on use experience with similar devices. Participants who had been using robotic surgical systems for more than 3 years, both doctors and nurses, reported that the evaluation device required a lot of effort to use. Participants had difficulty using the evaluation device because it was more complex than similar robotic surgical devices. However, when it comes to temporal demand, those who have been using robotic surgical systems for more than 3 years reported that it is not time-consuming. The workflow of the evaluation device is not much different from existing robotic surgical devices, and the graphical user interface is easy to use and can be performed quickly.

Table 9. Comparison of the NASA Task Load Index by use experience of doctors.

	All, mean (SD)	Less than 3 years, mean (SD)	More than 3 years, mean (SD)
Mental demand	30.67 (25.97)	27.00 (28.21)	38.00 (21.68)
Physical demand	13.00 (15.09)	14.00 (16.12)	11.00 (14.32)
Temporal demand	15.00 (17.63)	18.50 (19.87)	8.00 (10.37)
Performance	25.00 (28.09)	22.50 (28.80)	30.00 (29.15)
Effort	29.67 (24.60)	21.50 (19.59)	46.00 (27.48)
Frustration	19.00 (22.22)	15.50 (20.20)	26.00 (26.79)

Table 10. Comparison of the NASA Task Load Index by use experience of nurses.

	All, mean (SD)	Less than 3 years, mean (SD)	More than 3 years, mean (SD)
Mental demand	29.33 (29.39)	25.91 (22.56)	38.75 (46.61)
Physical demand	11.33 (12.17)	12.27 (12.72)	8.75 (11.81)
Temporal demand	22.33 (30.58)	28.64 (33.70)	5.00 (5.77)
Performance	28.33 (26.16)	22.27 (24.73)	45.00 (25.50)
Effort	34.33 (28.78)	30.45 (25.73)	45.00 (38.08)
Frustration	18.33 (26.70)	15.45 (15.40)	26.25 (49.22)

Usability (SUS)

Both doctors and nurses who participated in the usability test completed the SUS questionnaire. Table 11 shows that the mean score of SUS was 64.17 (SD 16.52) for doctors and 61.67 (SD 19.18) for nurses. The nonparametric Mann-Whitney *U* test was used to analyze *U* values and *P* values presented in Table 11.

When comparing the SUS scores of the doctors and nurses (P=.84, greater than .05), we can see that there was no significant difference between the 2 values. Figure 5 is a boxplot comparing the SUS scores of the doctors and nurses, showing a baseline of 68, which is the average SUS score. For the nurses and doctors, this corresponds to a grade of D on the SUS scale.

Table 11. Result of the System Usability Scale (SUS).

	Group 1[a], mean (SD)	Group 2[b], mean (SD)	*U* test[c]	*P* value[d]
SUS1: I think that I would like to use this system frequently.	2.33 (1.14)	2.60 (1.14)	96.5	.51
SUS2: I found the system unnecessarily complex.	2.4 (1.02)	2.67 (1.25)	92.5	.41
SUS3: I thought the system was easy to use.	2.93 (0.77)	2.93 (1.00)	104	.74
SUS4: I think that I would need the support of a technical person to be able to use this system.	2.00 (1.32)	1.00 (0.97)	65	.05
SUS5: I found the various functions in this system were well integrated.	3.00 (0.82)	2.93 (1.06)	109.5	.90
SUS6: I thought there was too much inconsistency in this system.	2.67 (0.94)	2.47 (1.31)	111	.97
SUS7: I would imagine that most people would learn to use this system very quickly.	2.87 (0.88)	3.20 (1.11)	82	.22
SUS8: I found the system very cumbersome to use.	2.53 (1.09)	2.73 (1.06)	100.5	.62
SUS9: I felt very confident using the system.	2.87 (0.81)	2.73 (0.88)	65	.05
SUS10: I needed to learn a lot of things before I could get going with this system.	2.07 (1.18)	2.00 (1.10)	106.5	.81
Overall, SUS score on 0 to 100 normalized scale	64.17 (16.52)	61.67 (19.18)	107.5	.84

[a]Group 1: doctors.

[b]Group 2: nurses.

[c]Because data were not normally distributed, the Mann-Whitney *U* test was performed.

[d]*P* values were determined with an independent 2-sample *t* test and the Mann-Whitney *U* test for continuous variables.

When we compared participants' use experience with the device between those who had used it for more than 3 years and those who had used it for less than 3 years, we found that for doctors, those who had used it for more than 3 years had lower SUS scores than those who had used it for less than 3 years. For doctors, the average SUS score for participants with 3 or more years of experience is 55.5 (SD 19.13), while the average SUS score for those with less than 3 years is 68.5 (SD 13.05). For nurses, similar to doctors, we found that participants who had used a similar device for more than 3 years had lower scores than those who had used it for less than 3 years, at 53.75 versus 64.55 (Table 12).

Table 12. System Usability Scale (SUS) comparison by use experience of a similar device.

	Group 1[a]			Group 2[b]		
	All, mean (SD)	Less than 3 years, mean (SD)	More than 3 years, mean (SD)	All, mean (SD)	Less than 3 years, mean (SD)	More than 3 years, mean (SD)
SUS	64.17 (16.52)	68.5 (13.05)	55.5 (19.13)	61.67 (19.18)	64.55 (15.14)	53.75 (25.77)

[a]Group 1: doctors.

[b]Group 2: nurses.

Discussion

Principal Findings

This study is a summative evaluation to examine the usability of the frameless stereotaxic navigation system (model CS200) that is used as an auxiliary tool for guiding the surgical tool to the target position and posture planned by the user in the incision or percutaneous spinal surgery. Regarding the use scenario, (1) task success, (2) use error, (3) satisfaction (ASQ), (4) workload (NASA TLX), and (5) SUS related to the usability of the test device were evaluated and analyzed. The usability test was conducted by professional medical staff who have completed specialized medical education and obtained professional medical qualifications. The participants in the usability test were doctors and nurses who have experience in using spinal surgery robots or navigation systems in operating rooms.

For doctors, all 8 scenarios had a success rate of at least 93% or higher, and for nurses, all 12 scenarios had a success rate of at least 87% or higher. Doctors had the lowest success rate of 93% on the "revision of surgical planning" scenario. In the ASQ results, the average score for "ease of completing the task" was 5.66 (SD 0.36), the average score for "amount of time it took" was 5.54 (SD 0.52), and the average score for "support information satisfaction" was 5.70 (SD 0.30). When performing the usability evaluation, the "revision of surgical planning" scenario had the lowest success rate; however, the 3 ASQ scores were higher than the average: 5.80 (SD 0.83), 5.93 (SD 0.68), and 5.93 (SD 1.29).

The "image matching" scenario had the lowest score for each item in the satisfaction score, even though it had a 100% success rate. We found that a high success rate on the evaluation task does not necessarily indicate high usability satisfaction. Despite the high task success rate in the usability test, the low satisfaction rate in the questionnaire that evaluated the usability aspects of the device indicates a lack of satisfaction with the device.

Although there were no difficulties in performance, the task of adjusting the ROI within the "registration" scenario was criticized for its difficulty in accurately adjusting the ROI and its lack of usability. Nurses, like doctors, had the lowest success rate of 87% in the "revision of surgical planning" scenario. However, the ASQ survey results for the "registration" scenario, which had the highest success rate of 98%, showed the lowest scores for "ease of completing the task" with a mean of 4.47 (SD 1.89), "amount of time it took" with a mean of 4.40 (SD 2.09), and "support information satisfaction" with a mean of 5.13 (SD 1.89). Similar to the doctors, when adjusting the ROI box, many of the nurses commented that the 360-degree rotation button at the top was not sensitive, and the video was difficult to see clearly. In addition, both doctors and nurses reported that when moving the robot arm using the footswitch during the "navigation" scenario, there was no visual or audible indication of how far the robot arm had moved and whether it had completed its movement, resulting in collisions between the robot arm and the patient stack. If used with real patients, this could lead to a significant risk of patient injury. We believe that the usability of these screens needs to be improved.

When comparing the workloads of the 2 groups who primarily use the assessment tool, we found that there was no difference, and that the workloads of mental demand and performance are high for both groups. Doctors and nurses commonly commented that the process was too complicated and laborious and that they had doubts about the accuracy of the ROI adjustment. In addition, since the robot assists in surgery, we thought it would be a quick process, but we found that the robot needed more time to move than expected, which affected the workload.

The SUS also showed no significant difference between doctors and nurses, with slightly lower-than-average satisfaction scores. The doctor gave the system a low score on the usability scale because it was too time-consuming to use in the actual operating room. Nurses gave low scores because of the time-consuming setup prior to actual surgery. Both doctors and nurses gave low scores, especially on the items that they felt they would need technician support to use the system and that they would need to learn about the system before its use, because many of them had never used an assessment device before and were not familiar with it. In addition, the lack of usability, with no explanation of what to do next on the device and no prompts to prevent errors, contributed to the low scores. At the hospital where participants work, engineers who have no difficulty using similar devices are present to aid doctors and nurses in using the device. Because of their reliance on engineers, many of the participants had difficulty in using the device alone and commented that they needed the engineers' support. This suggests that the device needs to be highly usable with an easily understandable user interface and a screen design that is familiar to medical staff so that they can use the device without engineers' assistance. Overall, when evaluating usability, there was no significant difference between participants who had used similar devices for more than 3 years and those who had used them for less than 3 years, except for ASQ and SUS, which evaluate satisfaction, and NASA TLX, which evaluates the difficulty of the operator's job. It was found that there were differences in job duties when using the equipment that they were familiar with as well as differences in the time taken to perform the tasks. In other items, there seems to be no problem in using the device once they are familiar with it.

Table S3 in Multimedia Appendix 1 shows the improvements made to the device since the usability evaluation. The user interface was improved by quantifying and intuitively displaying data that used to be shown only as graphs. Confusing highlighted buttons that hindered the use of the device were rearranged to decrease the errors made by users.

To further increase the usability of the device, the footswitch needs to be improved to recognize how much movement is required to move the device by displaying the information on the interface. In particular, during the navigation phase, it would increase the usability if a notification or on-screen guidance appeared when the target was reached while moving to the location guided by the robotic arm. In addition, the drape is divided into 3 stages, while other similar devices only have 1 drape, increasing the risk of contamination.

Although the participants have experience using third-party equipment, they had difficulty using this evaluation device for the first time because they were not familiar with it; however, we do not think there will be any major problems once they are accustomed to the evaluation device. In addition, as a robot that guides the position of the screw in the patient's body, it should have a more accurate and simpler workflow, making it more competitive with other products.

Limitations

This study is limited by the fact that our evaluation took place at a single institution, Severance Hospital in South Korea. In South Korea, robotic surgery is not yet widely used, and many people have not used robotic surgical instruments before; thus, they are still unfamiliar with robotic surgical instruments. However, the strength of this study is that we conducted usability tests with doctors and nurses in the operating room, who are the closest users of the new system, the surgical assistant robot.

Conclusions

In group 1, a success rate of 93% or higher was observed in all 24 tasks. In group 2, a success rate of 87% or higher was observed in 38 of 41 tasks. A success rate of 80% was observed in the task related to marker ball view confirmation (task 18), 80% in the task related to the use of the "Adjustment" function (task 28), and 75% in the task related to using "Re-matching" (task 30).

In addition, subjective data such as follow-up questions and surveys were more effective in identifying shortcomings and judging the usability, satisfaction, and effectiveness of the device than quantitative data such as the number of use errors (task completion rate) and satisfaction evaluation scores. In terms of error, participants provided a lot of feedback, including suggestions for mitigating potential risks. Although the task success rate was high, the workload and SUS scores were lower than the baseline, suggesting that improving the device user interface would increase the usability of the system. We recommend that the results of this test can be used in other usability engineering processes to improve the overall usability, satisfaction, completeness, and efficiency.

Authors' Contributions

HC wrote the paper and conducted all user experience evaluation, data collection, statistical analysis, and data interpretation. SK conducted user experience evaluation and data collection, such as comparison of the NASA Task Load Index, After-Scenario Questionnaire, and System Usability Scale by use experience. All authors conducted study design and reviewed the final paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
After-Scenario Questionnaire detailed results for doctors and nurses.
[DOCX File , 3454 KB - humanfactors_v11i1e54425_app1.docx]

References

1. Balicki M, Kyne S, Toporek G, Holthuizen R, Homan R, Popovic A, et al. Design and control of an image-guided robot for spine surgery in a hybrid OR. Int J Med Robot 2020;16(4):e2108. [doi: 10.1002/rcs.2108] [Medline: 32270913]
2. Vadalà G, De Salvatore S, Ambrosio L, Russo F, Papalia R, Denaro V. Robotic spine surgery and augmented reality systems: a state of the art. Neurospine 2020;17(1):88-100 [FREE Full text] [doi: 10.14245/ns.2040060.030] [Medline: 32252158]
3. Lee JW, Kim SM, Kim YS, Chung WK. Automated surgical planning system for spinal fusion surgery with three-dimensional pedicle model. J Inst Control Robot Syst 2011;17(8):807-813 [FREE Full text] [doi: 10.5302/j.icros.2011.17.8.807]
4. Wang TY, Park C, Dalton T, Rajkumar S, McCray E, Owolo E, et al. Robotic navigation in spine surgery: where are we now and where are we going? J Clin Neurosci 2021;94:298-304. [doi: 10.1016/j.jocn.2021.10.034] [Medline: 34863454]
5. Perfetti DC, Kisinde S, Rogers-LaVanne MP, Satin AM, Lieberman IH. Robotic spine surgery: past, present, and future. Spine (Phila Pa 1976) 2022;47(13):909-921. [doi: 10.1097/BRS.0000000000004357] [Medline: 35472043]
6. Mao JZ, Agyei JO, Khan A, Hess RM, Jowdy PK, Mullin JP, et al. Technologic evolution of navigation and robotics in spine surgery: a historical perspective. World Neurosurg 2021;145:159-167. [doi: 10.1016/j.wneu.2020.08.224] [Medline: 32916361]
7. IEC 62366-1:2015. Medical devices: Part 1: Application of usability engineering to medical devices. International Electrotechnical Commission. 2015. URL: https://www.iso.org/standard/63179.html [accessed 2024-03-15]
8. ISO 9241-210:2010. Ergonomics of human-system interaction: Part 210: Human-centred design for interactive systems. International Organization for Standardization. 2010. URL: https://www.iso.org/standard/77520.html [accessed 2024-03-15]
9. Wiklund M, Birmingham L, Larsen SA. Writing Human Factors Plans and Reports for Medical Technology Development. Arlington, VA: Association for the Advancement of Medical Instrumentation; 2018.
10. Lewis JR. Psychometric evaluation of an after-scenario questionnaire for computer usability studies: the ASQ. ACM SIGCHI Bull 1991;23(1):78-81 [FREE Full text] [doi: 10.1145/122672.122692]
11. Muhammad HAB, Hamid K, Iqbal MW, Khurram SK, Shahzad F, Shaheryar M. Usability impact of adaptive culture in smart phones. 2022 Presented at: VIPERC2022: 1st International Virtual Conference on Visual Pattern Extraction and Recognition for Cultural Heritage Understanding; September 12, 2022; Pescara, Italy URL: https://www.researchgate.net/profile/Khalid-Hamid-7/publication/365683592_Usability_Impact_of_Adaptive_Culture_in_Smart_Phones/links/637e4e032f4bca7fd0850d20/Usability-Impact-of-Adaptive-Culture-in-Smart-Phones.pdf
12. Lewis JR. An after-scenario questionnaire for usability studies: psychometric evaluation over three trials. ACM SIGCHI Bull 1991;23(4):79 [FREE Full text] [doi: 10.1145/126729.1056077]
13. Vanterpool KB, Gacki-Smith J, Kuramitsu B, Downey M, Nordstrom MJ, Luken M, et al. A patient-centered website (within reach) to foster informed decision-making about upper extremity vascularized composite allotransplantation: development and usability study. JMIR Form Res 2023;7:e44144 [FREE Full text] [doi: 10.2196/44144] [Medline: 36749618]
14. Yang YF, Chao CJ, Lin YL, Chang CK. Usability testing of Japanese captions segmentation system to scaffold beginners to comprehend Japanese videos. Int J Cyber Soc Educ 2013;6(1):1-14. [doi: 10.7903/ijcse.1083]
15. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Adv Psychol 1988;52:139-183. [doi: 10.1016/s0166-4115(08)62386-9]
16. Hart SG. Nasa-Task Load Index (NASA-TLX); 20 years later. Proc Hum Factors Ergon Soc Annu Meet 2016;50(9):904-908. [doi: 10.1177/154193120605000909]

17. Jiang M, Liu S, Gao J, Feng Q, Zhang Q. A usability study of 3 radiotherapy systems: a comparative evaluation based on expert evaluation and user experience. Med Sci Monit 2019;25:578-589 [FREE Full text] [doi: 10.12659/MSM.913160] [Medline: 30661077]
18. Idrisov B, Hallgren KA, Michaels A, Soth S, Darnton J, Grekin P, et al. Workload, usability, and engagement with a mobile app supporting video observation of methadone take-home dosing: usability study. JMIR Hum Factors 2023;10:e42654 [FREE Full text] [doi: 10.2196/42654] [Medline: 37440298]
19. Hart SG. NASA Task Load Index (TLX). NASA Ames Research Center. 1986. URL: https://ntrs.nasa.gov/citations/20000021487 [accessed 2024-03-12]
20. Patel T, Ivo J, Faisal S, McDougall A, Carducci J, Pritchard S, et al. A prospective study of usability and workload of electronic medication adherence products by older adults, caregivers, and health care providers. J Med Internet Res 2020;22(6):e18073 [FREE Full text] [doi: 10.2196/18073] [Medline: 32348292]
21. Sugarindra M, Suryoputro MR, Permana AI. Mental workload measurement in operator control room using NASA-TLX. IOP Conf Ser Mater Sci Eng 2017;277:012022 [FREE Full text] [doi: 10.1088/1757-899x/277/1/012022]
22. Yurko YY, Scerbo MW, Prabhu AS, Acker CE, Stefanidis D. Higher mental workload is associated with poorer laparoscopic performance as measured by the NASA-TLX tool. Simul Healthc 2010;5(5):267-271 [FREE Full text] [doi: 10.1097/SIH.0b013e3181e3f329] [Medline: 21330808]
23. Kim Y, Son J, Jang W. Usability study on patient monitoring systems: an evaluation of a user interface based on user experience and preference. Med Sci Monit 2023;29:e938570 [FREE Full text] [doi: 10.12659/MSM.938570] [Medline: 37270667]
24. Lewis JR. The system usability scale: past, present, and future. Int J Hum Comput Interact 2018;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
25. Del Rocio Sevilla-Gonzalez M, Loaeza LM, Lazaro-Carrera LS, Ramirez BB, Rodríguez AV, Peralta-Pedrero ML, et al. Spanish version of the system usability scale for the assessment of electronic tools: development and validation. JMIR Hum Factors 2020;7(4):e21161 [FREE Full text] [doi: 10.2196/21161] [Medline: 33325828]
26. Marzuki MFM, Yaacob NA, Yaacob NM. Translation, cross-cultural adaptation, and validation of the Malay version of the system usability scale questionnaire for the assessment of mobile apps. JMIR Hum Factors 2018;5(2):e10308 [FREE Full text] [doi: 10.2196/10308] [Medline: 29759955]
27. Chaniaud N, Métayer N, Megalakaki O, Loup-Escande E. Effect of prior health knowledge on the usability of two home medical devices: usability study. JMIR Mhealth Uhealth 2020;8(9):e17983 [FREE Full text] [doi: 10.2196/17983] [Medline: 32955454]

Abbreviations

ASQ: After-Scenario Questionnaire
IEC: International Electrotechnical Commission
ISO: International Organization for Standardization
NASA TLX: NASA Task Load Index
ROI: region of interest
SUS: System Usability Scale
TLX: Task Load Index

Edited by A Kushniruk; submitted 27.11.23; peer-reviewed by K Chen, X Cheng; comments to author 17.01.24; revised version received 07.02.24; accepted 26.02.24; published 01.04.24.

Please cite as:
Choi H, Kim S, Jang W
User Experience Evaluation of a Spinal Surgery Robot: Workload, Usability, and Satisfaction Study
JMIR Hum Factors 2024;11:e54425
URL: https://humanfactors.jmir.org/2024/1/e54425
doi:10.2196/54425
PMID:38432688

Short Paper

Assessing Differences in mHealth Usability and App Experiences Among Young African American Women: Secondary Analysis of a Randomized Controlled Trial

Claudia A Opper[1], BA; Felicia A Browne[1,2], MPH, ScD; Brittni N Howard[1], BA; William A Zule[1], MPH, DrPH; Wendee M Wechsberg[1,2,3], MS, PhD

[1]RTI International, Research Triangle Park, NC, United States

[2]Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, United States

[3]Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Durham, NC, United States

Corresponding Author:
Felicia A Browne, MPH, ScD
RTI International
3040 East Cornwallis Road
PO Box 12194
Research Triangle Park, NC, 27709-2194
United States
Phone: 1 919 541 6596
Email: fbrowne@rti.org

Abstract

Background: In North Carolina, HIV continues to disproportionately affect young African American women. Although mobile health (mHealth) technology appears to be a tool capable of making public health information more accessible for key populations, previous technology use and social determinants may impact users' mHealth experiences.

Objective: The objective of this study was to evaluate mHealth usability, assessing differences based on previous technology use and social determinants among a sample of African American women in emerging adulthood.

Methods: As part of a National Institute on Drug Abuse–funded randomized controlled trial with African American women (aged 18-25 years), counties were assigned to receive an evidence-based HIV risk reduction intervention through mHealth and participants were asked to complete usability surveys at 6- and 12-month follow-ups. Participants' first survey responses were analyzed through 2-tailed *t* tests and linear regression models to examine associations with previous technology use and social determinants (P<.05).

Results: The mean System Usability Scale (SUS) score was 69.2 (SD 17.9; n=159), which was higher than the threshold of acceptability (68.0). Participants who had previously used a tablet indicated higher usability compared to participants without previous use (mean 72.9, SD 18.1 vs mean 57.6, SD 11.4; P<.001), and participants with previous smartphone use also reported higher usability compared to participants without previous use (mean 71.9, SD 18.3 vs mean 58.0, SD 10.7; P<.001). Differences in SUS scores were observed among those reporting homelessness (mean 58.3, SD 19.0 vs mean 70.8, SD 17.2; P=.01), unemployment (mean 65.9, SD 17.2 vs mean 71.6, SD 18.1; P=.04), or current school enrollment (mean 73.2, SD 18.5 vs mean 65.4, SD 16.5; P=.006). Statistically significant associations were not observed for food insecurity (mean 67.3, SD 18.6 vs mean 69.9, SD 17.7; P=.45).

Conclusions: Although above-average usability was observed overall, these findings demonstrate differences in mHealth usability based on past and current life experiences. As mHealth interventions become more prevalent, these findings may have important implications for ensuring that mHealth apps improve the reach of evidence-based interventions.

Trial Registration: ClinicalTrials.gov NCT02965014; https://clinicaltrials.gov/study/NCT02965014

International Registered Report Identifier (IRRID): RR2-10.1186/s12889-018-5796-8

(JMIR Hum Factors 2024;11:e51518) doi:10.2196/51518

KEYWORDS

HIV; Black women; mobile apps; social determinants of health; prevention; substance use; usability

Introduction

In 1981, the first report was published identifying the disease that was later known as AIDS, marking the official beginning of the HIV epidemic [1]. In that same year, IBM's first PC was sold to the public [2]. As the HIV epidemic persists, there may be an opportunity to embrace the digital age we are living in and leverage technological solutions as we work toward the shared goal of ending the HIV epidemic.

HIV incidence rates remain disproportionate based on race in the United States, as rates for Black or African American women are 10.9 times higher than rates for White women [3]. Further, the highest rates of HIV diagnoses occur in the US South [3]. Many interventions have been developed for Black or African American women [4-8], but barriers such as lack of transportation, limited childcare access, concerns over privacy, and community-level stigma impede access to HIV testing and prevention services [9,10].

Most Americans have smart mobile devices [11]. These devices show promise in diminishing barriers and connecting key populations with public health information through increasingly convenient and private pathways. Research has demonstrated that being a woman, young, and African American are characteristics associated with being more likely to prefer mobile health (mHealth) when given a choice or to use mobile devices to seek health information [12,13], supporting mHealth interventions as potentially effective tools for this key population.

However, individuals may experience and engage with mHealth interventions differently. As the prevalence of mHealth apps continues to increase [14], there is a need to understand mHealth usability. In a scoping review of electronic health applications from 2014-2017, the rate of new health applications available outpaced the rate of published usability studies [15]. The authors explained that while most digital health apps are developed commercially, the results of commercial usability studies are not typically published [15]. Because of the limited reported data for mHealth usability, this study examined the usability of an mHealth HIV prevention intervention among young African American women in the US South.

A previous study adapted a best-evidence, women-focused HIV behavioral intervention for young African American women (the Young Women's CoOp), which demonstrated efficacy in reducing sexual risk [16]. In preparation for a trial to test intervention delivery, an mHealth version of the Young Women's Coop was developed [17]. This analysis of usability scores for the Young Women's CoOp mHealth app examines whether previous technology experience and social determinants are associated with mHealth experiences.

Methods

Overview

Analyses in this paper encompass an assessment of the usability of the Young Women's CoOp intervention that was adapted to an mHealth platform. The parent study reached 652 young African American women (aged 18-25 years) in North Carolina who reported recent condomless sex with a male partner and substance use. A complete description of the parent study's eligibility criteria and procedures can be found in the study protocol paper [18].

In a 3-arm randomized trial implementing a cross-over design, 3 counties were assigned to receive the in-person delivery of the Young Women's CoOp intervention, the mHealth delivery of the Young Women's CoOp intervention, or standard HIV counseling and testing. Among the enrolled sample, 197 women were in the counties assigned to receive the mHealth delivery, which consisted of a 1-on-1 mHealth orientation and an Android tablet preloaded with the app that contained the 2-session intervention. Following the orientation, a tablet was provided to each participant to take with them. The study team requested that tablets be returned at the 6-month follow-up appointment.

The usability of the mHealth intervention was evaluated using a modified version of the 10-item System Usability Scale (SUS) [19]. SUS scores range from 0 to 100, with scores above 68 considered above average [20]. To account for participants missing follow-up appointments, mHealth participants completed the usability survey as part of an audio-computer–assisted self-interview (ACASI) at both 6-month and 12-month follow-ups. For participants who completed the usability survey at both follow-ups, only their first (6-month) survey response was considered. ACASI was administered in person in an attempt to engage and collect responses from mHealth users who may have had difficulty using the tablet or lost the tablet and who may have had challenges completing a tablet-hosted survey.

Social determinants (homelessness, unemployment, food insecurity, and school enrollment) were measured at study enrollment. Social determinant variables were either assessed as dichotomous questions (homelessness and school enrollment) or recoded into dichotomous variables (unemployment and food insecurity). Homelessness, unemployment, and school enrollment assessed an individual's current state and food insecurity asked about one's household. Descriptive statistics and 2-tailed *t* tests were used to assess bivariate associations between social determinants and usability scores. Linear regression was conducted to examine these associations while controlling for previous tablet use and previous smartphone use. Analyses were conducted in Stata 17 (StataCorp) using the threshold of $P<.05$ for statistical significance.

Ethical Considerations

The full study received approval from the Office of Research Protection's Institutional Review Board at RTI International (IRB ID number: 13836). Further, committees from Wake County Human Services and Durham County Department of Public Health, along with administration from the Guilford County Department of Public Health, granted study approval. All participants provided written informed consent. Several procedures were instituted to protect the privacy and confidentiality of study participants, including all staff members involved in data collection and analysis signing and abiding by Staff Agreements of Confidentiality. Additionally, each participant was assigned a unique alphanumeric participant identifier to limit study data being connected to identifying information, such as name and contact information. Study participants were compensated for completing the baseline appointment with US $50 in gift cards, the 6-month follow-up appointment with US $70 in gift cards, and the 12-month follow-up appointment with US $100 in gift cards.

Results

The overall mean SUS score was 69.2 (SD 17.9; n=159). Less than 12% (n=19) of participants did not have experience with a tablet or smartphone before the study. Variability of SUS scores by previous technology use and social determinants is shown in Table 1.

Participants who had previous tablet use reported higher SUS scores on average than participants who had not previously used a tablet (72.9, SD 18.1 vs 57.6, SD 11.4; *P*<.001). Similarly, participants who had previously used a smartphone had a higher mean SUS score than participants who had not (71.9, SD 18.3 vs 58.0, SD 10.7; *P*<.001).

Additionally, the mean SUS scores were under the acceptable threshold for participants reporting food insecurity, homelessness, unemployment, or no current school enrollment. Statistically significant differences in mean SUS scores were observed among those reporting homelessness (58.3, SD 19.0 vs 70.8, SD 17.2; *P*=.01), unemployment (65.9, SD 17.2 vs 71.6, SD 18.1; *P*=.04), or current school enrollment (73.2, SD 18.5 vs 65.4, SD 16.5; *P*=.006). Statistically significant associations were not observed in the SUS score based on food insecurity (67.3, SD 18.6 vs 69.9, SD 17.7; *P*=.45). When accounting for previous mobile technology experience in each model, homelessness and current school enrollment were statistically significant, but unemployment and food insecurity were not statistically significant (Table 2).

Table 1. Bivariate associations between System Usability Scale (SUS) score and previous technology use and social determinants of health.

	Frequency, n (%)	SUS score, mean (SD)	*P* value
Previous tablet use			<.001
Yes	121 (76.1)	72.9 (18.1)	
No	38 (23.9)	57.6 (11.4)	
Previous smartphone use			<.001
Yes	128 (80.5)	71.9 (18.3)	
No	31 (19.5)	58.0 (10.7)	
Homelessness			.01
Yes	20 (12.6)	58.3 (19.0)	
No	139 (87.4)	70.8 (17.2)	
Unemployment			.04
Yes	67 (42.1)	65.9 (17.2)	
No	92 (57.9)	71.6 (18.1)	
Food insecurity			.45
Yes	41 (25.8)	67.3 (18.6)	
No	118 (74.2)	69.9 (17.7)	
In school			.006
Yes	77 (48.4)	73.2 (18.5)	
No	82 (51.6)	65.4 (16.5)	

Table 2. Associations between System Usability Scale (SUS) score and social determinants of health, adjusting for previous tablet use and smartphone use.

Independent variables	Coefficient (95% CI)	*P* value
Homelessness	–9.0 (–16.8 to –1.1)	.03
Unemployment	–4.9 (–10.1 to 0.3)	.07
Food insecurity	–1.5 (–7.5 to 4.4)	.61
In school	6.2 (1.1 to 11.3)	.02

Discussion

As mHealth continues to become more prevalent, these findings show an overall above-average usability for the mHealth adaptation of the Young Women's CoOp intervention. Notably, there were differences in SUS scores based on a participant's past experiences with technology. Although less than 12% of the study sample had not previously used a smartphone or tablet, this proportion was higher than what may be expected for this age group based on national survey data for young adults [11]. This finding suggests the importance of considering ways the digital divide and other factors may impact familiarity with mobile technology when designing mHealth interventions for this key population. In this study, before participants were given tablets with the mHealth intervention, they received a brief orientation to the app. Given the lower reported usability among those who lacked experience with mobile technology, these findings suggest examining further whether providing more guidance during app orientation could improve app usability for those with limited previous experience. Additionally, ongoing and other technology support (eg, a chat feature where trained staff can provide technology support to users within the app) may be strategies to explore with guidance from the intended end users to see if they improve mHealth experiences for individuals with less mobile technology experience.

Given privacy and stigma-related barriers that may hinder access to in-person HIV prevention programs and services for young African American women [9,10], mHealth could be an attractive solution. However, our findings exemplify that not only previous experience with technology but also diversity in participants' life circumstances, such as homelessness and school enrollment, can be associated with usability. Though the format may appear well-positioned for young adults, it is imperative to consider how a surge in the use of mHealth may miss the opportunity to maximally address existing health disparities if some users encounter barriers when operating the mHealth app that undermine their experiences. Additional guidance and support will be essential for those with factors associated with lower usability.

This study should be considered in relation to a few limitations. All participants completed the intervention using a study-issued Android tablet. Noting how a device's model or operating system may affect usage, some user experiences may have been shaped by the device specifications. In future usability studies, it may be valuable to have participants use their own devices to minimize the chance that device unfamiliarity affects assessments of app usability. Further, participants were asked to assess usability at 6- and 12-month follow-ups. Though participants still had access to the app before returning the device at their follow-up appointment, there was potential for recall bias as usability may have been assessed months after a participant's last app interaction. Additionally, it should be noted that all experiences with the app and data collection occurred before the COVID-19 pandemic. With a greater shift to digital formats for health, education, social, and other services throughout the pandemic, access to mobile devices and familiarity with receiving information through mobile technology may have increased since this study.

Despite these limitations, the study prompts important considerations as the health sector embraces digital technology. The overall above-average usability score signals the potential value of using mHealth as a delivery method in the public health toolkit to further expand the reach of evidence-based interventions to those who may need it the most.

Acknowledgments

The authors would like to thank the participants and the study collaborators at the Durham County Department of Public Health, the Guilford County Department of Public Health, and Wake County Human Services.

This study was funded by the National Institute on Drug Abuse (grant R01 DA041009). The content is solely the work of the authors and does not necessarily reflect the views of the National Institute on Drug Abuse.

Conflicts of Interest

None declared.

References

1. Global HIV/AIDS timeline. KFF. URL: https://www.kff.org/global-health-policy/timeline/global-hivaids-timeline/ [accessed 2022-12-19]
2. Timeline of computer history. Computer History Museum. URL: https://www.computerhistory.org/timeline/1981/ [accessed 2023-01-09]
3. HIV surveillance report, 2020. Vol. 33. Centers for Disease Control and Prevention. 2022. URL: http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html [accessed 2023-07-25]
4. Wechsberg WM, Lam WKK, Zule WA, Bobashev G. Efficacy of a woman-focused intervention to reduce HIV risk and increase self-sufficiency among African American crack abusers. Am J Public Health 2004;94(7):1165-1173 [FREE Full text] [doi: 10.2105/ajph.94.7.1165] [Medline: 15226138]
5. DiClemente RJ, Wingood GM, Harrington KF, Lang DL, Davies SL, Hook EW, et al. Efficacy of an HIV prevention intervention for African American adolescent girls: a randomized controlled trial. JAMA 2004;292(2):171-179 [FREE Full text] [doi: 10.1001/jama.292.2.171] [Medline: 15249566]

6. Gilbert L, Goddard-Eckrich D, Chang M, Hunt T, Wu E, Johnson K, et al. Effectiveness of a culturally tailored HIV and sexually transmitted infection prevention intervention for Black women in community supervision programs: a randomized clinical trial. JAMA Netw Open 2021;4(4):e215226 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.5226] [Medline: 33835175]
7. Jemmott LS, Jemmott JB, O'Leary A. Effects on sexual risk behavior and STD rate of brief HIV/STD prevention interventions for African American women in primary care settings. Am J Public Health 2007;97(6):1034-1040 [FREE Full text] [doi: 10.2105/AJPH.2003.020271] [Medline: 17463391]
8. Donenberg GR, Kendall AD, Emerson E, Fletcher FE, Bray BC, McCabe K. IMARA: a mother-daughter group randomized controlled trial to reduce sexually transmitted infections in Black/African-American adolescents. PLoS One 2020;15(11):e0239650 [FREE Full text] [doi: 10.1371/journal.pone.0239650] [Medline: 33137103]
9. Mathews A, Farley S, Conserve DF, Knight K, Le'Marus A, Blumberg M, et al. "Meet people where they are": a qualitative study of community barriers and facilitators to HIV testing and HIV self-testing among African Americans in urban and rural areas in North Carolina. BMC Public Health 2020;20(1):494 [FREE Full text] [doi: 10.1186/s12889-020-08582-z] [Medline: 32295568]
10. Rimmler S, Golin C, Coleman J, Welgus H, Shaughnessy S, Taraskiewicz L, et al. Structural barriers to HIV prevention and services: perspectives of African American women in low-income communities. Health Educ Behav 2022;49(6):1022-1032 [FREE Full text] [doi: 10.1177/10901981221109138] [Medline: 35856333]
11. Mobile fact sheet. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2023-01-15]
12. Khairat S, Liu S, Zaman T, Edson B, Gianforcaro R. Factors determining patients' choice between mobile health and telemedicine: predictive analytics assessment. JMIR mHealth and uHealth 2019;7(6):e13772 [FREE Full text] [doi: 10.2196/13772] [Medline: 31199332]
13. Fox S, Duggan M. Mobile health 2012. Pew Research Center. 2012. URL: https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/2012/PIP_MobileHealth2012_FINAL.pdf [accessed 2023-01-09]
14. Bhuyan SS, Lu N, Chandak A, Kim H, Wyant D, Bhatt J, et al. Use of mobile health applications for health-seeking behavior among US adults. J Med Syst 2016;40(6):153. [doi: 10.1007/s10916-016-0492-7] [Medline: 27147516]
15. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: a scoping review. Int J Med Inform 2019;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
16. Wechsberg WM, Browne FA, Zule WA, Novak SP, Doherty IA, Kline TL, et al. Efficacy of the Young Women's CoOp: an HIV risk-reduction intervention for substance-using African-American female adolescents in the South. J Child Adolesc Subst Abuse 2017;26(3):205-218 [FREE Full text] [doi: 10.1080/1067828X.2016.1260511] [Medline: 28845096]
17. Watkins RL, Browne FA, Kizakevich PN, Howard BN, Turner LB, Eckhoff R, et al. An evidence-based HIV risk-reduction intervention for young African American women in the US South using mHealth: adaptation and development study. JMIR Form Res 2022;6(5):e34041 [FREE Full text] [doi: 10.2196/34041] [Medline: 35532978]
18. Browne FA, Wechsberg WM, Kizakevich PN, Zule WA, Bonner CP, Madison AN, et al. mHealth versus face-to-face: study protocol for a randomized trial to test a gender-focused intervention for young African American women at risk for HIV in North Carolina. BMC Public Health 2018;18(1):982 [FREE Full text] [doi: 10.1186/s12889-018-5796-8] [Medline: 30081868]
19. System Usability Scale (SUS). US General Services Administration. URL: https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html [accessed 2022-12-19]
20. Brooke J. SUS: a 'quick and dirty' usability scale. Usability Eval Ind 1996:189-194 [FREE Full text]

Abbreviations

ACASI: audio-computer–assisted self-interview
mHealth: mobile health
SUS: System Usability Scale

Edited by A Kushniruk; submitted 02.08.23; peer-reviewed by D Dunsmuir, C Sun; comments to author 26.11.23; revised version received 17.01.24; accepted 20.02.24; published 16.04.24.

Please cite as:
Opper CA, Browne FA, Howard BN, Zule WA, Wechsberg WM
Assessing Differences in mHealth Usability and App Experiences Among Young African American Women: Secondary Analysis of a Randomized Controlled Trial
JMIR Hum Factors 2024;11:e51518
URL: https://humanfactors.jmir.org/2024/1/e51518
doi:10.2196/51518
PMID:38625721

Original Paper

Interactive Electronic Pegboard for Enhancing Manual Dexterity and Cognitive Abilities: Instrument Usability Study

Shih-Ying Chien[1,2], PhD; Alice MK Wong[2], MD; Ching-Yi Wu[2,3,4], ScD; Sara L Beckman[5], PhD

[1]Department of Industrial Design, Chang Gung University, Taoyuan, Taiwan

[2]Department of Physical Medicine and Rehabilitation, Chang Gung Medical Foundation, Taoyuan, Taiwan

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

[4]Healthy Aging Research Center, Chang Gung University, Taoyuan, Taiwan

[5]Haas School of Business, University of California, Berkeley, Berkeley, CA, United States

Corresponding Author:
Shih-Ying Chien, PhD
Department of Industrial Design
Chang Gung University
No.259, Wen-Hwa 1st Road, Kweishan
Taoyuan, 333
Taiwan
Phone: 886 3 2118800 ext 5669
Email: shihying@mail.cgu.edu.tw

Abstract

Background: Strokes pose a substantial health burden, impacting 1 in 6 people globally. One-tenth of patients will endure a second, often more severe, stroke within a year. Alarmingly, a younger demographic is being affected due to recent lifestyle changes. As fine motor and cognitive issues arise, patient disability as well as the strain on caregivers and health care resources is exacerbated. Contemporary occupational therapy assesses manual dexterity and cognitive functions through object manipulation and pen-and-paper recordings. However, these assessments are typically isolated, which makes it challenging for therapists to comprehensively evaluate specific patient conditions. Furthermore, the reliance on one-on-one training and assessment approaches on manual documentation is inefficient and prone to transcription errors.

Objective: This study examines the feasibility of using an interactive electronic pegboard for stroke rehabilitation in clinical settings.

Methods: A total of 10 patients with a history of stroke and 10 healthy older individuals were recruited. With a limit of 10 minutes, both groups of participants underwent a series of challenges involving tasks related to manual operation, shape recognition, and color discrimination. All participants underwent the Box and Block Test and the Purdue Pegboard Test to assess manual dexterity, as well as an array of cognitive assessments, including the Trail Making Test and the Mini-Mental Status Examination, which served as a basis to quantify participants' attention, executive functioning, and cognitive abilities.

Results: The findings validate the potential application of an interactive electronic pegboard for stroke rehabilitation in clinical contexts. Significant statistical differences ($P<.01$) were observed across all assessed variables, including age, Box and Block Test results, Purdue Pegboard Test outcomes, Trail Making Test-A scores, and Mini-Mental Status Examination performance, between patients with a history of stroke and their healthy older counterparts. Functional and task testing, along with questionnaire interviews, revealed that patients with a history of stroke demonstrated prolonged completion times and slightly inferior performance. Nonetheless, most patients perceived the prototype as user-friendly and engaging. Thus, in the context of patient rehabilitation interventions or the evaluation of patient cognition, physical functioning, or manual dexterity assessments, the developed pegboard could potentially serve as a valuable tool for hand function, attention, and cognitive rehabilitation, thereby mitigating the burden on health care professionals.

Conclusions: Health care professionals can use digital electronic pegboards not only as a precise one-on-one training tool but also as a flexible system that can be configured for online or offline, single-player or multiplayer use. Through data analysis, a more informed examination of patients' cognitive and functional issues can be conducted. Importantly, patient records will be fully retained throughout practices, exercises, or tests, and by leveraging the characteristics of big data, patients can receive the most accurate rehabilitation prescriptions, thereby assisting them in obtaining optimal care.

(JMIR Hum Factors 2024;11:e56357) doi:10.2196/56357

KEYWORDS

interactive electronic pegboard; stroke; hand dexterity; cognitive rehabilitation; system

Introduction

Worldwide, the aging population continues to increase, with several attendant problems [1,2]. The process of aging entails repercussions that extend beyond mere physiological conditions and encompasses a diverse spectrum of complications [3]. Older individuals are confronted with economic, psychological, and societal predicaments stemming from physical aging [4,5]. Hypertension is a critical condition intricately connected with the occurrence of strokes [6,7]. In 2020, a total of 7.08 million individuals globally died due to cerebrovascular disorders [8]. Encouragingly, continuous advancements in medical technology have increased the survival rate for patients with a history of stroke to 62% [9]. Nonetheless, even in cases of survival, 90% of patients experience residual effects, making rehabilitation approaches pivotal [10,11].

Stemming from damage to cerebral tissue, cerebral stroke gives rise to a variety of distinct neurological symptoms contingent upon the site of injury [12,13]. This often culminates in motor, sensory, and cognitive impairments among patients with a history of stroke, in which reduced attentional focus and memory deficits are common [14,15]. The aftermath of a stroke can have negative effects on patients' daily lives, occupational status, and social involvement [16,17]. To enhance physical mobility, manual proficiency, and cognitive aptitude, occupational therapy is the gold standard for elevating overall function [18]. Rehabilitation procedures are initiated once a patient's vital signs stabilize [19]. Clinical evidence indicates that due to significant individual variations among patients with a history of stroke, including age, rehabilitation needs vary [20-22]. Furthermore, older adults predominantly seek to restore ambulatory capacities, whereas younger individuals emphasize intricate fine motor rehabilitation exercises due to occupational demands [23].

In clinical practice, therapists often use calibrated instruments to evaluate and document patients' manual dexterity and cognitive recovery capabilities in one-on-one settings [24-26] using standard methodologies, such as the Purdue Pegboard Test (PPT) and the Nine-Hole Peg Test [24,27]. However, the use of a countdown timer to measure tasks within specific time frames has been validated as an effective means to infer attention, cognition, and manual dexterity capabilities in clinical contexts [28,29]. However, there remain substantial challenges, including human resource depletion, increased time expenditures, difficulties in effective disease progression tracking, and recording errors. Acharya et al [24] emphasized the inherent delay, particularly in response time, with the traditional interactive training method. Compared with the current setup, there is a consistent observation of higher measured timing. Taking the commercially available Neofect Smart Pegboard as an example, it serves as an electronic pegboard [30]. While it offers several advantages, it does not feature long-term tracking and precise prescriptions for individual patients. Furthermore, using the electronic prototype of the Grooved Pegboard Test proposed by Al-Naami et al [31] in 2021 as an example, its operational efficiency shows no significant differences compared with the traditional method of manually recording rehabilitation outcomes. This experiment validates the feasibility of an electronic pegboard test to measure hand-time dexterity with impaired hand functionality, indicating comparable or even superior effectiveness when compared with the conventional manual recording approach.

In this study, we used electronic sensing techniques integrated with Wi-Fi and tablet devices to achieve a higher level of precision in evaluating tasks and time of completion [32]. This digitized approach facilitates accurate documentation of the intricacies associated with each practice and assessment, thereby enhancing the overall precision of the rehabilitation process [33,34]. The principal objective of this study was to subject the prototype to initial evaluation and testing involving patients with a history of stroke and healthy older individuals. We aimed to determine the appropriateness of the set difficulty levels, time constraints, and speed of the prototype.

Methods

Overview

The experimental apparatus consisted of an iPad, 5 color-sensitive building blocks, and 3 variations of task casings. The system's underlying sensing mechanism relied on the modulation of capacitance values resulting from the interaction between the sensing electrodes of the panel and the human body. The conductive building blocks generated stimulation signals that served as surrogate agents for fingers.

A schematic representation is shown in Figure 1. Paired with the distinctive visual patterns on the back of each building block, these visual patterns upon contact with the iPad screen were detected and recognized through pressure sensing. This design was aimed at assessing the responsiveness, visual acuity, and color perception abilities of the participants during rehabilitation interactions (Figure 2).

All rehabilitation tasks and exercises integrated time calculation and countdown functions. Patients were given the option to choose between "independent practice" and "interactive practice" modes. During independent practice, after the "start" button was pressed and the countdown timer initiated, randomized questions were presented. Each practice session was preconfigured for a duration of 10 minutes, and completion and error rates were captured.

For interactive practice, therapists preset practice durations and modify difficulty levels (rehabilitation prescriptions). Through a Wi-Fi connection, therapists administer questions to make an online assessment of patients' abilities. After patients perform the tasks, both patients and therapists receive practice and rehabilitation reports, with all exercise records automatically stored in the cloud. A flowchart illustrating the operation of the proposed pegboard is shown in Figure 3.

The proposed system encompasses 3 distinct modes, each presenting varying levels of complexity.

Figure 1. Design of the proposed system.

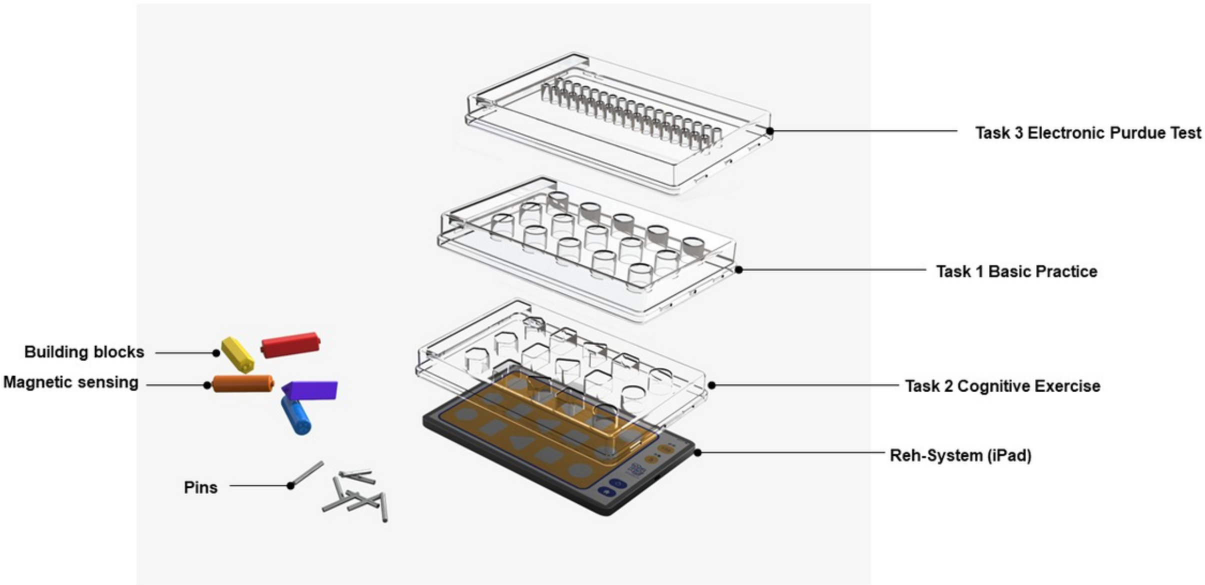


Figure 2. Interactive sensor blocks and tablet interface scenario.

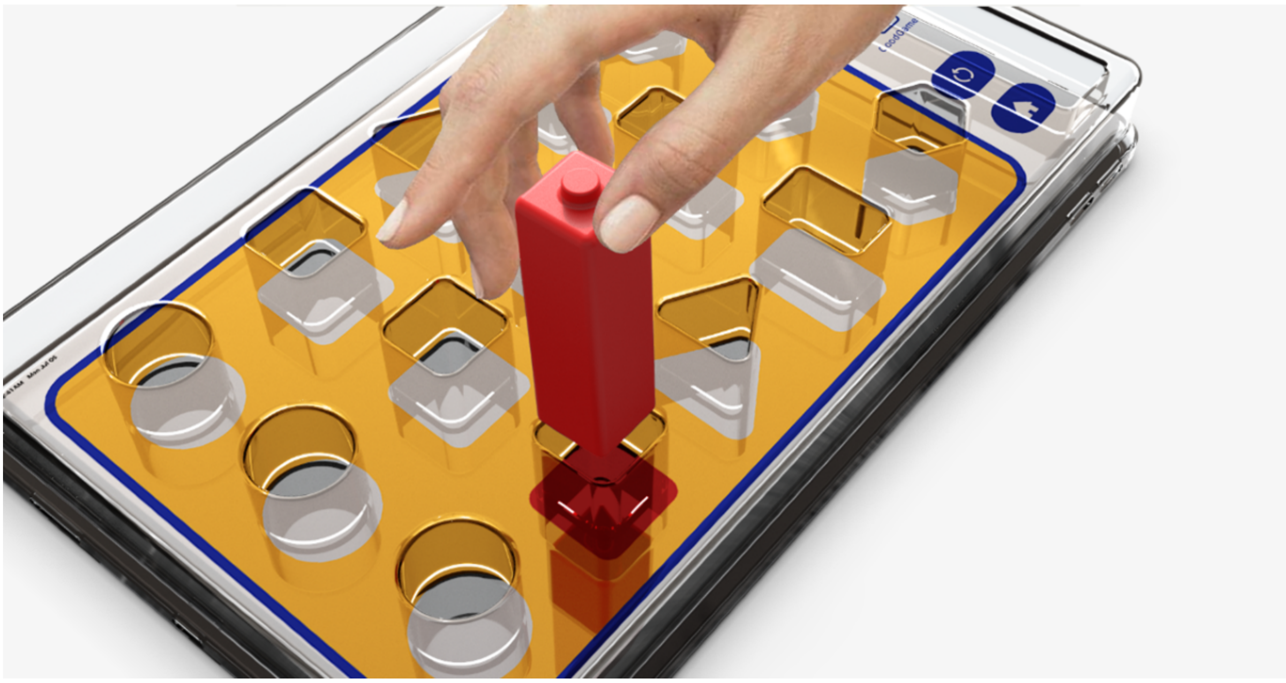
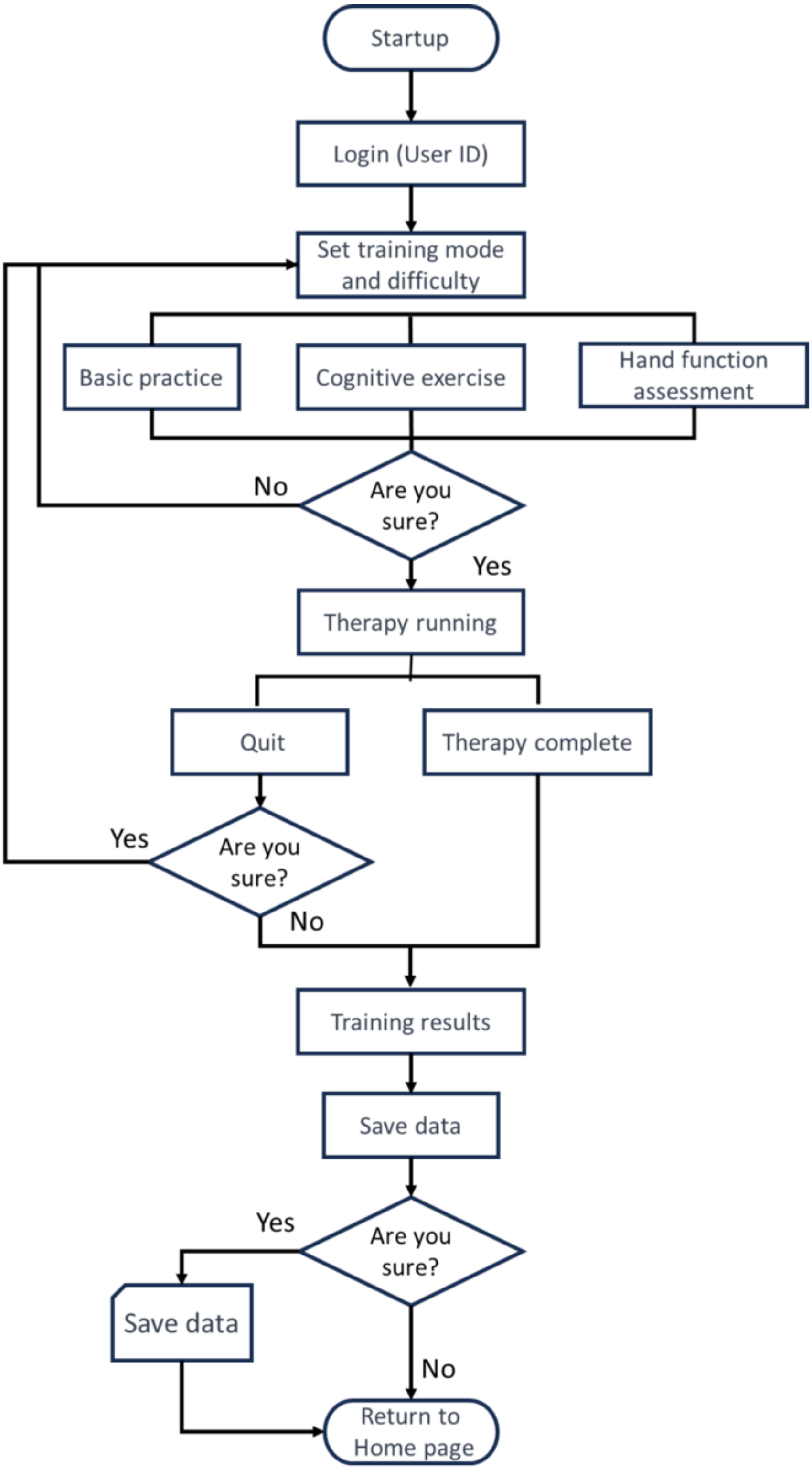


Figure 3. Flowchart of the proposed system operation.



Basic Practice

In basic practice (BP)–1 mode, users are assigned the task of associating the building blocks with the corresponding positions guided by reminder lights on the tablet screen (Figure 4A). Users earn points when they correctly insert the blocks into the panel's corresponding positions. In the intermediate level (BP-2), users place the blocks in the corresponding positions as indicated by the lights of the screen within the given time frame while concurrently considering the variety of colors presented. The advanced level (BP-3) introduces a speed variable to increase the complexity of the task.

Figure 4. Three modes of the proposed system: (A) basic practice; (B) cognitive exercise; and (C) Electronic Purdue Test.

Cognitive Exercise

In cognitive exercise (CE)-1 mode, patients are required to distinguish the shapes of the building blocks and place them according to the patterns displayed on the screen. Users earn points when the blocks are correctly positioned. Upon advancing to the CE-2 level, users need to not only identify the corresponding shaped blocks but also distinguish the colors indicated by the lights. Points are awarded only when both the shape and the color are correct. In addition, this mode introduces varying levels of complexity related to color discrimination and speed (Figure 4B).

Electronic Purdue Test

In the design of the electronic Purdue Test (EPT) level, adherence to the principles of the PPT was paramount. The illuminated signals were meticulously crafted to guide patients in sequentially inserting pegs into corresponding holes (Figure 4C). The assessment consists of distinct 30-second trials for the right, left, and both hands, with individual scores recorded and aggregated. In addition, a 60-second bilateral combination test is administered once. This comprehensive set of evaluations is repeated 3 times. The platform automatically calculates the average score, which serves as the test score. Unlike traditional training, the device guides patients to place pegs into corresponding positions through the use of light signals, which remain illuminated until the pegs are properly placed.

Given the inherent variability in individual patient capabilities, prescribed treatments should differ. To facilitate precision health care, the system incorporates a user login mechanism to generate personalized digital rehabilitation plans and records. The proposed design comprises 3 different modes and 3 difficulty levels of exercises. Preestablished exercises encompass directional movements (including upward, downward, leftward, and rightward motions). During the initial stages of the rehabilitation regimen, the system uses a mechanism of stochastic question generation. As the system accumulates practice data, it systematically discerns and assimilates the individual requirements of each patient, thereby tailoring subsequent questions to enhance areas of observed weakness.

Research Aims

This research had the following aims: (1) we sought to investigate the suitability of the time and difficulty settings for both patients with a history of stroke and healthy users, (2) we explored the correlation between hand function and cognitive abilities, and (3) we conducted a usability questionnaire for the proposed system.

Participant Recruitment

A total of 20 older adults aged between 65 and 80 years were recruited: 10 (50%) patients with a history of stroke and 10 (50%) individuals with no history of strokes. Prior to their inclusion, all participants provided signed informed consent. All participants exhibited right-handed dominance. The inclusion criteria were delineated based on the following: (1) capacity to independently maintain a seated position for a duration exceeding 20 minutes, unaided by external assistance; (2) possession of fundamental communication skills; and (3) relatively uncomplicated functional performance in the assessment of daily life activities. As a safeguard against potential trial-related risks, individuals with a history of recurrent stroke, severe muscular atrophy, or pronounced physical frailty were excluded.

Experimental Procedure

All participants underwent the Box and Block Test (BBT) and the PPT to assess manual dexterity, as well as an array of cognitive assessments, including the Trail Making Test (TMT) and the Mini-Mental Status Examination (MMSE), which served as a basis to quantify participants' attention, executive functioning, and cognitive abilities. Figure 5 shows an image of the proposed system in use.

The experimental session was conducted on a one-on-one basis. The prototype was positioned before each participant, and the 15 pegs, consisting of 5 distinctive colors integral to the interactive system, were methodically arranged adjacent to the central apparatus. Participants sequentially underwent 3 testing modes (BP, CE, and EPT), using their right hand exclusively. Except for the EPT, which followed the PPT criteria, the length of the BP and CE tests was 10 minutes each. The test outcomes encompassed the number of correct responses, completion time, and rehabilitation reports, all of which were concurrently displayed, stored within the apparatus, and uploaded to the cloud platform.

To gain a thorough understanding of users' interactions with the proposed system, we used the System Usability Scale [25] with a 5-point Likert scale to reveal users' perceptions regarding aspects of system acceptance, design appeal, and perceived task difficulty.

Figure 5. Image of the proposed system in use.

Statistical Analysis

This study used SPSS software (version 20; IBM Corp) for statistical analyses. Data analysis involved a comparative assessment between patients with a history of stroke and healthy older individuals, exploring both demographic characteristics and scores obtained from the proposed system, with the Wilcoxon rank sum test used for statistical analysis. To elucidate the potential associations between participants' manual dexterity and cognitive faculties among patients with a history of stroke, this study also used the Spearman rank correlation coefficient, with statistical significance set at *P*<.05.

Ethical Considerations

A total of 20 participants were recruited for this study. All participants provided informed consent by signing a consent form, and the study was conducted in accordance with institutional review board (IRB) regulations using anonymized data, with personal information removed and replaced by codes. No participants withdrew from the study during the research period. To ensure the validity and fairness of the experiment, no monetary or material benefits will be provided during the trial period, in accordance with the IRB application statement. Participants are expected to provide genuine feedback on the product developed in this project based on their intuitive reactions.

The research was conducted within the Department of Physical Medicine and Rehabilitation at Chang Gung Hospital, Taiwan and received approval from the Research Ethics Committee for Human Subject Protection of Chang Gung Medical Foundation (IRB: 202301197A3).

Results

This study included a total of 20 participants (10 patients with a history of stroke and 10 healthy participants). With a limit of 10 minutes, both groups of participants underwent a series of challenges involving tasks related to manual operation, shape recognition, and color discrimination. The statistical analysis revealed statistically significant discrepancies between patients with a history of stroke and healthy participants across all variables (*P*<.05). These differences were evident in all assessed parameters, indicating the potential of the equipment to serve as an assessment tool for both motor and cognitive abilities in both healthy individuals and patients with a history of stroke, with additional training and testing capabilities (Table S1 in Multimedia Appendix 1).

Among older participants who had not experienced a stroke, performance in tasks involving the dominant hand (right hand) during the BBT and the PPT as well as cognitive performance in the TMT was notably superior to those who had experienced a stroke (Figure S1A and S1B in Multimedia Appendix 1). However, no significant differences were observed between the 2 groups in the MMSE test, which assessed memory abilities (Table S2 in Multimedia Appendix 1). Furthermore, in the MMSE test assessing memory abilities, both groups of subjects showed significant differences (*P*<.01) (Table S2 in Multimedia Appendix 1).

Statistical analysis revealed significant negative correlations between performance in the BP-1 or CE-1 task and dexterity tests (P<.01). In addition, there were significant correlations between multicolors (BP-2 or CE-2) and dexterity or cognitive tests and a significant negative correlation between scores in the EPT and cognitive performance on the TMT-A and MMSE tests (P<.05) (Table S3 in Multimedia Appendix 1). The number of correct answers was used as the score in BP and CE; the time required was used as the score in BP, CE, and EPT in the single and multicolor tests. In terms of usability, 40% (4/10) of patients with a history of stroke and 60% (6/10) of healthy participants deemed the prototype user-friendly (Figure S2A in Multimedia Appendix 1).

Furthermore, all healthy individuals (100%) and majority of patients with a history of stroke (90%, 9/10) found the proposed system highly engaging. During the more demanding CE training, more than 80% (8/10) and 50% (5/10) of healthy participants and patients with a history of stroke, respectively, considered the system both challenging and stimulating. Participants additionally expressed a positive disposition toward the EPT and provided overall positive feedback (Figure S2B in Multimedia Appendix 1).

In assessing task difficulty, nearly 80% (8/10) and 60% (6/10) of healthy participants and patients with a history of stroke, respectively, perceived the BP training tasks as easy and straightforward. However, as participants advanced to the more challenging CT training, there was a noticeable increase in the perceived complexity of tasks, in which only 40% (4/10) and 20% (2/10) of healthy participants and patients with a history of stroke, respectively, found this phase easy. Moreover, 60% (6/10) of healthy participants found the EPT straightforward, while 40% (4/10) of patients with a history of stroke indicated a moderate level of challenge associated with the EPT training (Figure S2C in Multimedia Appendix 1).

Discussion

Principal Findings

Across the 5 tasks investigated in this study, the stroke rehabilitation group exhibited significantly lower scores in the use of the proposed system compared to the healthy participants. In accordance with previous research, advancing age and disease manifest changes and declines in hand function, muscle strength, agility, and cognitive abilities [35,36]. This was evident in the use of the proposed system.

Previous studies have highlighted the repercussions of cerebral damage on patients with a history of stroke, such as compromised cognitive, motor, sensory, and functional capabilities, as well as pain, balance issues, visual challenges, and restricted engagement in activities [37,38]. Sudden cognitive deterioration occurs as a result of these conditions, with 5% of patients with a history of stroke exhibiting dementia symptoms [39,40]. Thus, the augmentation of hand function rehabilitation for patients is imperative. Our findings underscore a close interrelation between manual dexterity and cognitive aptitude [41]. In future research, we intend to explore the use of the proposed system paired with auditory and visual cues to ascertain its potential to guide and enhance visual acuity, attention, and cognitive capabilities among cohorts of different ages and individuals afflicted with cerebral impairments.

The BP and CE tests encompassed factors of both single-color and multicolor conditions. Participants encountered operating difficulties as cognitive demands increased. The CE assessment included recognition of color and object shape, where performance consistently declined across all participants. This observation aligns with previous findings indicating a decline in change detection accuracy with an increase in cognitive load [26,27]. Consequently, as the number of colors increased, the attentional burden on the patients correspondingly increased. This suggests that a graded system with different levels of difficulty might be useful.

During the progression of single-color BP-1 and CE-1 sessions, negative correlations were observed in dexterity tests. In the context of multicolor BP and CE sessions, a distinct and significant correlation was found between the use of multicolors and performance levels on the TMT-A test. The test results suggest that when participants engaged in color recognition and discrimination tasks, the attentional demands for single-color and multicolor tasks differed. In other words, multicolor exercises presented an increased cognitive challenge, affecting manual dexterity and attention switching.

The majority of participants found the proposed system highly user-friendly, in part because the size of the system resembled traditional training pegboards, which maintained familiarity and reduced the need for adaptation. As therapists were not required to manually record participants' actions or time them with stopwatches, this design was advantageous for both users and evaluators. In the various tasks, approximately 60% (12/20) to 70% (14/20) of all participants found BP training to be interesting, while 40% (8/20) to 60% (12/20) of the participants considered the EPT tasks engaging. Regarding the difficulty level, 60% (12/20) to 70% (14/20) of the participants perceived BP training as relatively easy; however, as operating constraints increased (such as color and shape elements), the participants commonly reported that tasks became more challenging and demanding. This finding is consistent with prior research indicating that increasing the difficulty to match users' current abilities enhanced their confidence, maintained attention and engagement in tasks, and promoted a more positive and enjoyable acceptance of new challenges. In this preliminary experiment, neither patients with a history of stroke nor healthy participants were able to complete the tasks within the allotted time, and none of the participants achieved a perfect score. This is likely attributable to the time constraints imposed by the experimental design or the capabilities of the users. Therefore, future studies will include basing task difficulty settings on user performance, similar to leveling up in a video game. The gamification of rehabilitation, in addition to fostering effective interactivity, is facilitated by the incorporation of voice and music assistance. This approach contributes to enhancing the enjoyment of rehabilitation, transforming it from a tedious and uninteresting process. Furthermore, it effectively redirects patients' attention away from pain, thereby augmenting the overall appeal of the rehabilitation process.

The proposed system is equipped for practice, training, and assessment. The majority of similar products on the market predominantly focus on training manual dexterity and do not offer timing and recording functions [28-30]. Certain designs acknowledge the significance of cognitive training and use shape as a cognitive judgment criterion; however, these designs lack elements that enhance the attention of rehabilitation patients, such as auditory cues, visual stimuli, or color-guided prompts. This deficiency in interactive mechanisms often results in users struggling to sustain or commit to rehabilitation efforts. Finally, in terms of assessment functionality, contemporary clinical practice still relies on manual documentation and human intervention for upper limb assessments. The proposed system not only incorporates timing and counting features during upper limb assessments but also introduces guiding and competitive elements, positioning patients to achieve better recovery outcomes. Currently, the system is converting data from each patient's rehabilitation sessions into charts. This aids both patients and health care professionals in gaining a more comprehensive understanding of the rehabilitation and recovery status. This system is thus highly advantageous compared with current commercial products [32-34,42]. The interactive electronic pegboard integrates the merits of existing market offerings and further introduces automated assessment and scoring mechanisms for accurately placed pegs. By surpassing the limitations inherent in conventional fixed training paradigms, this system systematically and comprehensively records the training progress of each case. As data accumulate via a learning model, the system develops a profound understanding of user-specific requirements and consequently extrapolates optimal and customized training regimens tailored to individual users, representing a major step toward precise rehabilitation goals. Finally, the proposed system exhibits greater versatility in its training curriculum and offers increased variability and flexibility. Rehabilitation with digital tools will no doubt significantly enhance users' interest and attention.

Our preliminary investigation indicates that the proposed system is beneficial in the training, assessment, and testing of patients with a history of stroke. The outcomes showcase positive responses concerning hand function training and cognitive ability assessment among patients with a history of stroke. However, to ascertain reliability and validity, a greater number of participants, including diverse age groups and individuals with cerebral impairments, should be recruited in future investigations [35].

Patients with a history of stroke often grapple with diminished motivation for rehabilitation and a lack of immediate feedback, hindering their ability to maintain consistent participation in rehabilitation regimens [36-38]. While the platform devised in this study uses auditory and visual cues to encourage perseverance in rehabilitation, there remains room for improvement in configuring different challenge levels and real-time feedback mechanisms based on varying patient capacities [39-41]. This is a pivotal objective for refinement and enhancement.

Conclusions

The primary objective of this research was to bridge the gap between clinical requirements and product development through customized rehabilitation training based on individual differences. Through the analysis and assessment of data and providing personalized training modes tailored to specific differences, we aimed to predict patients' hand dexterity and cognitive functional abilities. We thus developed the interactive electronic pegboard, a novel software- and hardware-integrated system for stroke rehabilitation, with the purpose of evaluating dexterity and cognitive functions through various task types and multidemonstration patterns. Preliminary findings indicate the efficacy of the system for training and assessment. The ultimate goal of this research is to develop an intelligent system capable of delivering individualized optimized rehabilitation regimens based on the varying needs of users.

Acknowledgments

The authors express their gratitude to all participants in this project for their generous contribution of time, sharing of usage experiences, and insights, which greatly enriched this study. Furthermore, they extend their sincere appreciation to Professor Ching-Yi Wu from the Department of Occupational Therapy at Chang Gung University for providing valuable suggestions and crucial guidance in the experimental design. Finally, they especially thank the Rehabilitation Department at Chang Gung Memorial Hospital for their clinical support throughout this research endeavor. This study received support from the National Science and Technology Council through grant 111-2222-E-182-002-MY2, as well as from the Chang Gung University Innovative Prototype Project with grant EMRPD1M0101.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

SYC was involved in product development and study design, conducted the interviews and prototype testing, collected and analyzed the data, and drafted the manuscript. AMKW and CYW provided clinical guidance. SLB contributed to the manuscript by offering suggestions and participating in revisions.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed participant data and analyses.
[DOCX File , 245 KB - humanfactors_v11i1e56357_app1.docx]

References

1. Bray F, Jemal A, Grey N, Ferlay J, Forman D. Global cancer transitions according to the human development index (2008-2030): a population-based study. Lancet Oncol 2012;13(8):790-801. [doi: 10.1016/S1470-2045(12)70211-5] [Medline: 22658655]
2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71(3):209-249 [FREE Full text] [doi: 10.3322/caac.21660] [Medline: 33538338]
3. McClintock MK, Dale W, Laumann EO, Waite L. Empirical redefinition of comprehensive health and well-being in the older adults of the United States. Proc Natl Acad Sci U S A 2016;113(22):E3071-E3080 [FREE Full text] [doi: 10.1073/pnas.1514968113] [Medline: 27185911]
4. van den Akker M, Buntinx F, Knottnerus JA. Comorbidity or multimorbidity: what's in a name? a review of literature. Eur J Gen Pract 2009;2(2):65-70 [FREE Full text] [doi: 10.3109/13814789609162146]
5. Stein KD, Syrjala KL, Andrykowski MA. Physical and psychological long-term and late effects of cancer. Cancer 2008;112(11 Suppl):2577-2592 [FREE Full text] [doi: 10.1002/cncr.23448] [Medline: 18428205]
6. MacMahon S, Peto R, Cutler J, Collins R, Sorlie P, Neaton J, et al. Blood pressure, stroke, and coronary heart disease. part 1, prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. Lancet 1990;335(8692):765-774. [doi: 10.1016/0140-6736(90)90878-9] [Medline: 1969518]
7. Gorelick PB, Scuteri A, Black SE, Decarli C, Greenberg SM, Iadecola C, et al. Vascular contributions to cognitive impairment and dementia: a statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2011;42(9):2672-2713 [FREE Full text] [doi: 10.1161/STR.0b013e3182299496] [Medline: 21778438]
8. Manji RA, Arora RC, Singal RK, Hiebert BM, Menkis AH. Early rehospitalization after prolonged intensive care unit stay post cardiac surgery: outcomes and modifiable risk factors. J Am Heart Assoc 2017;6(2):e004072 [FREE Full text] [doi: 10.1161/JAHA.116.004072] [Medline: 28174166]
9. Filsoufi F, Rahmanian PB, Castillo JG, Bronster D, Adams DH. Incidence, topography, predictors and long-term survival after stroke in patients undergoing coronary artery bypass grafting. Ann Thorac Surg 2008;85(3):862-870. [doi: 10.1016/j.athoracsur.2007.10.060] [Medline: 18291158]
10. Hoskins WJ, McGuire WP, Brady MF, Homesley HD, Creasman WT, Berman M, et al. The effect of diameter of largest residual disease on survival after primary cytoreductive surgery in patients with suboptimal residual epithelial ovarian carcinoma. Am J Obstet Gynecol 1994;170(4):974-979. [doi: 10.1016/s0002-9378(94)70090-7] [Medline: 8166218]
11. Wasserstrom WR, Glass JP, Posner JB. Diagnosis and treatment of leptomeningeal metastases from solid tumors: experience with 90 patients. Cancer 1982;49(4):759-772 [FREE Full text] [doi: 10.1002/1097-0142(19820215)49:4<759::aid-cncr2820490427>3.0.co;2-7] [Medline: 6895713]
12. Gaetz M. The neurophysiology of brain injury. Clin Neurophysiol 2004;115(1):4-18. [doi: 10.1016/s1388-2457(03)00258-x] [Medline: 14706464]
13. Schmahmann JD, Smith EE, Eichler FS, Filley CM. Cerebral white matter: neuroanatomy, clinical neurology, and neurobehavioral correlates. Ann N Y Acad Sci 2008;1142(1):266-309 [FREE Full text] [doi: 10.1196/annals.1444.017] [Medline: 18990132]
14. Mahncke HW, Bronstone A, Merzenich MM. Brain plasticity and functional losses in the aged: scientific bases for a novel intervention. Prog Brain Res 2006;157:81-109. [doi: 10.1016/S0079-6123(06)57006-2] [Medline: 17046669]
15. Prokopenko SV, Mozheyko EY, Petrova MM, Koryagina TD, Kaskaeva DS, Chernykh TV, et al. Correction of post-stroke cognitive impairments using computer programs. J Neurol Sci 2013;325(1-2):148-153. [doi: 10.1016/j.jns.2012.12.024] [Medline: 23312291]
16. Hertzog C, Kramer AF, Wilson RS, Lindenberger U. Enrichment effects on adult cognitive development: can the functional capacity of older adults be preserved and enhanced? Psychol Sci Public Interest 2008;9(1):1-65 [FREE Full text] [doi: 10.1111/j.1539-6053.2009.01034.x] [Medline: 26162004]
17. Liberman RP. Recovery From Disability: Manual of Psychiatric Rehabilitation. Washington, DC: American Psychiatric Publishing; 2009.
18. Pendleton HM, Schultz-Krohn W. Pedretti's Occupational Therapy: Practice Skills for Physical Dysfunction. St. Louis, MO: Elsevier Health Sciences; 2017.

19. Duncan PW, Zorowitz R, Bates B, Choi JY, Glasberg JJ, Graham GD, et al. Management of adult stroke rehabilitation care: a clinical practice guideline. Stroke 2005;36(9):e100-e143 [FREE Full text] [doi: 10.1161/01.STR.0000180861.54180.FF] [Medline: 16120836]
20. Quinn TJ, Paolucci S, Sunnerhagen KS, Sivenius J, Walker MF, Toni D, et al. Evidence-based stroke rehabilitation: an expanded guidance document from the European Stroke Organisation (ESO) guidelines for management of ischaemic stroke and transient ischaemic attack 2008. J Rehabil Med 2009;41(2):99-111 [FREE Full text] [doi: 10.2340/16501977-0301] [Medline: 19225703]
21. Heinemann AW, Linacre JM, Wright BD, Hamilton BB, Granger C. Relationships between impairment and physical disability as measured by the functional independence measure. Arch Phys Med Rehabil 1993;74(6):566-573. [doi: 10.1016/0003-9993(93)90153-2] [Medline: 8503745]
22. Winstein CJ, Stein J, Arena R, Bates B, Cherney LR, Cramer SC, et al. Guidelines for adult stroke rehabilitation and recovery: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2016;47(6):e98-e169 [FREE Full text] [doi: 10.1161/STR.0000000000000098] [Medline: 27145936]
23. Majmudar S, Wu J, Paganoni S. Rehabilitation in amyotrophic lateral sclerosis: why it matters. Muscle Nerve 2014;50(1):4-13 [FREE Full text] [doi: 10.1002/mus.24202] [Medline: 24510737]
24. Acharya KA, Bhat S, Kanthi M, Rao BK. Fine motor assessment in upper extremity using custom-made electronic pegboard test. J Med Signals Sens 2022;12(1):76-83 [FREE Full text] [doi: 10.4103/jmss.JMSS_58_20] [Medline: 35265469]
25. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput 2008;24(6):574-594 [FREE Full text] [doi: 10.1080/10447310802205776]
26. Vrij A, Mann SA, Fisher RP, Leal S, Milne R, Bull R. Increasing cognitive load to facilitate lie detection: the benefit of recalling an event in reverse order. Law Hum Behav 2008;32(3):253-265. [doi: 10.1007/s10979-007-9103-y] [Medline: 17694424]
27. Setz C, Arnrich B, Schumm J, La Marca R, Tröster G, Ehlert U. Discriminating stress from cognitive load using a wearable EDA device. IEEE Trans Inf Technol Biomed 2010;14(2):410-417 [FREE Full text] [doi: 10.1109/TITB.2009.2036164] [Medline: 19906598]
28. Curcio G, Ferrara M, De Gennaro L. Sleep loss, learning capacity and academic performance. Sleep Med Rev 2006;10(5):323-337. [doi: 10.1016/j.smrv.2005.11.001] [Medline: 16564189]
29. Schmidt RA, Lee TD, Winstein CJ, Wulf G, Zelaznik HN. Motor control and Learning: A Behavioral Emphasis. Champaign, IL: Human Kinetics; 2018.
30. Ericsson KA, Charness N. Expert performance: its structure and acquisition. Am Psychol 1994;49(8):725-747 [FREE Full text] [doi: 10.1037//0003-066x.49.8.725]
31. Al-Naami B, Al-Naimat F, Almalty A, Visconti P, Al-Hinnawi A. A prototype of an electronic pegboard test to measure hand-time dexterity with impaired hand functionality. ASI 2021 Dec 22;5(1):2. [doi: 10.3390/asi5010002]
32. Lee JS, Kim SW. Effects of Korean computer-based cognitive rehabilitation program on the memory in healthy elderly. JIAPTR 2018;9(4):1591-1595 [FREE Full text] [doi: 10.20540/JIAPTR.2018.9.4.1591]
33. Ham Y, Shin JH. Efficiency and usability of a modified pegboard incorporating computerized technology for upper limb rehabilitation in patients with stroke. Top Stroke Rehabil 2023;30(4):333-341 [FREE Full text] [doi: 10.1080/10749357.2022.2058293] [Medline: 35348031]
34. Okahashi S, Sakamoto K, Hashiya F, Kumasaka K, Yamaguchi T, Seiyama A, et al. Development of an electric pegboard (e-Peg) for hand dexterity improvement and cognitive rehabilitation: a preliminary study. Adv Biomed Eng 2023;12:81-90 [FREE Full text] [doi: 10.14326/abe.12.81]
35. Weintraub S, Salmon D, Mercaldo N, Ferris S, Graff-Radford NR, Chui H, et al. The Alzheimer's Disease Centers' Uniform Data Set (UDS): the neuropsychologic test battery. Alzheimer Dis Assoc Disord 2009;23(2):91-101 [FREE Full text] [doi: 10.1097/WAD.0b013e318191c7dd] [Medline: 19474567]
36. Taylor J, Taylor S. Psychological Approaches to Sports Injury Rehabilitation. Gaithersburg, MD: Aspen Publishers; 1997.
37. Sohlberg MM, Mateer CA. Cognitive Rehabilitation: An Integrative Neuropsychological Approach. New York, NY: Guilford Press; 2001.
38. Borghese NA, Pirovano M, Lanzi PL, Wüest S, de Bruin ED. Computational intelligence and game design for effective at-home stroke rehabilitation. Games Health J 2013;2(2):81-88 [FREE Full text] [doi: 10.1089/g4h.2012.0073] [Medline: 24761321]
39. Sung M, Marci C, Pentland A. Wearable feedback systems for rehabilitation. J Neuroeng Rehabil 2005;2:17 [FREE Full text] [doi: 10.1186/1743-0003-2-17] [Medline: 15987514]
40. Sveistrup H. Motor rehabilitation using virtual reality. J Neuroeng Rehabil 2004;1(1):10 [FREE Full text] [doi: 10.1186/1743-0003-1-10] [Medline: 15679945]
41. Novak D. Promoting motivation during robot-assisted rehabilitation. In: Colombo R, Sanguineti V, editors. Rehabilitation Robotics. St. Louis, MO: Elsevier Science; 2018:149-158.
42. Shin JH, Kim MY, Lee JY, Jeon YJ, Kim S, Lee S, et al. Effects of virtual reality-based rehabilitation on distal upper extremity function and health-related quality of life: a single-blinded, randomized controlled trial. J Neuroeng Rehabil 2016;13(1):17 [FREE Full text] [doi: 10.1186/s12984-016-0125-x] [Medline: 26911438]

Abbreviations

BBT: Box and Block Test
BP: basic practice
CE: cognitive exercise
EPT: Electronic Purdue Test
MMSE: Mini-Mental Status Examination
PPT: Purdue Pegboard Test
TMT: Trail Making Test

Edited by Y Gong; submitted 14.01.24; peer-reviewed by YW Shau, C Baxter; comments to author 04.03.24; revised version received 22.03.24; accepted 04.04.24; published 21.06.24.

Please cite as:
Chien SY, Wong AMK, Wu CY, Beckman SL
Interactive Electronic Pegboard for Enhancing Manual Dexterity and Cognitive Abilities: Instrument Usability Study
JMIR Hum Factors 2024;11:e56357
URL: https://humanfactors.jmir.org/2024/1/e56357
doi:10.2196/56357
PMID:38904991

Original Paper

Effects of User Experience in Automated Information Processing on Perceived Usefulness of Digital Contact-Tracing Apps: Cross-Sectional Survey Study

Tim Schrills[1], BSc, MSc; Lilian Kojan[1], BSc, MSc; Marthe Gruner[1], BSc, MSc; André Calero Valdez[1], Prof Dr; Thomas Franke[1], Prof Dr

[1]Institute for Multimedia and Interactive Systems, Universität zu Lübeck, Lübeck, Germany

Corresponding Author:
Tim Schrills, BSc, MSc
Institute for Multimedia and Interactive Systems
Universität zu Lübeck
Ratzeburger Allee 160
Lübeck, 23560
Germany
Phone: 49 451 3101 ext 5135
Fax: 49 451 3101 5104
Email: Tim.schrills@uni-luebeck.de

Abstract

Background: In pandemic situations, digital contact tracing (DCT) can be an effective way to assess one's risk of infection and inform others in case of infection. DCT apps can support the information gathering and analysis processes of users aiming to trace contacts. However, users' use intention and use of DCT information may depend on the perceived benefits of contact tracing. While existing research has examined acceptance in DCT, automation-related user experience factors have been overlooked.

Objective: We pursued three goals: (1) to analyze how automation-related user experience (ie, perceived trustworthiness, traceability, and usefulness) relates to user behavior toward a DCT app, (2) to contextualize these effects with health behavior factors (ie, threat appraisal and moral obligation), and (3) to collect qualitative data on user demands for improved DCT communication.

Methods: Survey data were collected from 317 users of a nationwide-distributed DCT app during the COVID-19 pandemic after it had been in app stores for >1 year using a web-based convenience sample. We assessed automation-related user experience. In addition, we assessed threat appraisal and moral obligation regarding DCT use to estimate a partial least squares structural equation model predicting use intention. To provide practical steps to improve the user experience, we surveyed users' needs for improved communication of information via the app and analyzed their responses using thematic analysis.

Results: Data validity and perceived usefulness showed a significant correlation of r=0.38 (P<.001), goal congruity and perceived usefulness correlated at r=0.47 (P<.001), and result diagnosticity and perceived usefulness had a strong correlation of r=0.56 (P<.001). In addition, a correlation of r=0.35 (P<.001) was observed between Subjective Information Processing Awareness and perceived usefulness, suggesting that automation-related changes might influence the perceived utility of DCT. Finally, a moderate positive correlation of r=0.47 (P<.001) was found between perceived usefulness and use intention, highlighting the connection between user experience variables and use intention. Partial least squares structural equation modeling explained 55.6% of the variance in use intention, with the strongest direct predictor being perceived trustworthiness (β=.54; P<.001) followed by moral obligation (β=.22; P<.001). Based on the qualitative data, users mainly demanded more detailed information about contacts (eg, place and time of contact). They also wanted to share information (eg, whether they wore a mask) to improve the accuracy and diagnosticity of risk calculation.

Conclusions: The perceived result diagnosticity of DCT apps is crucial for perceived trustworthiness and use intention. By designing for high diagnosticity for the user, DCT apps could improve their support in the action regulation of users, resulting in higher perceived trustworthiness and use in pandemic situations. In general, automation-related user experience has greater importance for use intention than general health behavior or experience.

(JMIR Hum Factors 2024;11:e53940) doi:10.2196/53940

KEYWORDS

COVID-19; contact tracing; user experience; trust; health information processing

Introduction

Background

During pandemic situations, efficiently acquiring, storing, and evaluating information on physical contacts can be crucial for both individuals and public health agencies aiming to curb infection dynamics [1]. Manual tracing of such contacts is practically impossible, leading to a growing development and research of digital tools supporting such efforts, commonly referred to as digital contact tracing (DCT) apps [2]. By allowing for automation, DCT tools effectively allow for contact tracing. They aim to allow individual users to assess their own risk status with minimal effort and offer support in daily action regulation, such as in decision situations, regarding isolation or notification of previous contacts [3]. If used correctly, DCT can aid in breaking chains of infection and thereby support curbing pandemic spread. For example, in Germany, a DCT called *Corona-Warn-App* (CWA) [4] was developed on behalf of the Federal Ministry of Health, and it was downloaded >40 million times [5].

However, the extent to which individuals use DCT can vary vastly [6]. Previous research has shown that it is crucial whether users perceive a DCT app as beneficial to guide them in pandemic contexts [7]. This core factor is in line with existing models of health behavior (eg, the influential Health Belief Model [HBM] [8]). Within the HBM, perceived benefit is outlined as a central determinant for the implementation of health behavior [7]. When investigating health-related technology, the HBM is frequently connected with models of technology acceptance [9]. As part of these models, the perceived usefulness or performance of technology is similarly postulated as a central variable for use intention. In this paper, we refer to the term *usefulness* as it is better suited than *benefits* to describe the effects of a specific technology. Thereby, we refer to usefulness as "the degree to which a person believes that using a particular system would enhance their [...] performance" [10].

Examining psychological processes revolving around the perception of DCT usefulness is a crucial research topic to understand the adoption and efficient implementation of DCT. Extensive research has shown the importance of the perceived usefulness of DCT for different applications and in different countries [11-15]. All in all, extending existing theoretical approaches such as the HBM by focusing on user experience variables in DCT allows for clear guidelines on improving DCT design and uptake.

The usefulness that a user can experience from DCT results from the automation it provides. DCT takes over tasks that would otherwise need to be done manually (eg, recording contacts, estimating distance and exposure to contacts, and calculating risk based on the vaccination status of contacts). Therefore, it can be defined as an automated system. In general, automation can be defined as a system's ability to "offload, assist, or replace human performance at corresponding stages of human information processing" [16]. The human action that DCT seeks to automate is the continuous recording and analysis of contact data to monitor an individual's risk of infection. While there is a large body of research on automation, its adverse biases, and its impact on human performance [17-19], less research focuses on the psychological processes involved when users evaluate the usefulness of automated contact tracing.

Parasuraman et al [20] define 4 evaluation criteria on how automation can affect human performance: situation awareness [21], trust (cf complacency and trust [22]), skill degradation [23], and workload [24]. When users want to make situation-adequate decisions, they benefit from improved situation awareness. Situation awareness, in turn, can be improved by DCT. As long as the information or recommendations provided by DCT apps are perceived as trustworthy, users may use them to determine the right course of action. Accordingly, a DCT's ability to support situation awareness as well as trust formation (refer to the study by Hoff and Bashir [25]) may lead to perceived usefulness. On the other hand, in the context of DCT apps, one cannot assume that users are potentially losing a previously existing skill through automation; DCT app users are not able to stop sick individuals or themselves. Along the same line, DCT app users profit from automation as it reduces manual work in contact tracing. Therefore, we propose to examine users' experience of situation awareness and trustworthiness when using DCT apps.

While research has demonstrated that usefulness strongly impacts use intention [26], factors unrelated to the specific DCT app might affect whether people intend to use the system. The HBM positions threat appraisal as another factor directly influencing use intention [7]. While using a DCT app changes neither the susceptibility nor the severity (in comparison, refer to the study by Costa [27]) related to an infection, it is still plausible that users with higher threat appraisal are more interested in their own risk status and, therefore, more likely to use a DCT app (eg, to be able to detect and react to an infection as early as possible). Therefore, threat appraisal may influence use intention independent of the specified design of DCT apps. In addition, recent research has also shown that the theoretical framework of the HBM does profit from incorporating prosocial aspects of decisions [28,29] (ie, using a DCT app may provide a sense of moral obligation to others). Even though individuals with immunity may perceive a lower personal threat, they may feel a personal obligation to track and inform contacts. Overall, to fully investigate the influence of the perceived usefulness of a DCT system on the use intention, a comparison with system-nonspecific factors (ie, threat appraisal) and personal moral obligation should be made. To the best of our knowledge, no previous study has focused on examining the perception of automation-related usefulness while addressing threat appraisal and moral obligation as system-independent factors influencing use intention.

Research Objective

The objective of this research was to examine how automation-related user experience affects the perceived usefulness of contact tracing as well as use intention of DCT apps and how user experience could be improved. To do so, our approach consisted of multiple methods. The first was quantitatively assessing and analyzing the impact of automation-related user experience (ie, experienced system traceability and perceived trustworthiness) as well as system knowledge on the intention of using a DCT app. The second was contextualizing the effects of automation-related user experience measures with factors related to health protection behavior (ie, threat appraisal and moral obligation). The third was a qualitative analysis of user demands for improved information communication between users and the DCT app. Therefore, the key contribution of this research is a better understanding of how system characteristics lead to perceived usefulness of DCT and how optimal DCT apps can increase use intention through automation-related user experience. Thus, this research supports the human-centered design of DCT apps.

To address these research objectives, 317 users of the CWA DCT system were surveyed about their experience with the app through a web-based questionnaire. A partial least squares structural equation model (PLS-SEM) was used to quantitatively describe the relationships among psychological factors regarding DCT use. This approach was supplemented by a thematic analysis of qualitative user requests on desired communication of information between users and the system.

Related Research

Use Intention of DCT

DCT describes software applications that support documenting information of physical contact or proximity between people (cf [30]). This includes both the (partially) automated acquisition of contact information and the analysis of this information (eg, to determine an individual's risk of infection [31]). In pandemic situations, users might have the goal to avoid contributing to the further spread of the pandemic disease and, thus, face a control task. This means that users need to constantly self-regulate their actions in relation to their environment (eg, how many people around them are infected). While users strive to achieve this goal, they are constantly facing a changing environment (ie, exposure to infected persons). To maintain control, they need to constantly acquire and analyze information and decide, for example, whether they want to isolate themselves. Such actions taken by users have a profound impact on the trajectory of their individual situation—they potentially curtail further contacts and, thereby, change the future information acquisition process. In this process, DCT constitutes a crucial tool for behavioral control as the information provided functions both as feedback for previous behavior and as an indicator for future behavior.

Although DCT applications, especially on mobile devices, first generated high interest during the COVID-19 pandemic [32], they had already been used previously (refer to, eg, the study by Sacks et al [33]). Due to their wide applicability and potential role in public health systems during the COVID-19 pandemic, research on user behavior toward DCT has increased. Here, diverging acceptance models (such as the Unified Theory of Acceptance and Use of Technology and the technology acceptance model) have been evaluated to understand DCT use intention (eg, the study by Velicia-Martin et al [34]).

As indicated at the outset, previous research on DCT app use has leveraged not only acceptance models but also more general models of health behavior such as the HBM or the Theory of Planned Behavior [35]. Such models have been successfully used in research on the uptake and maintenance of other pandemic protective behaviors. In that context, there is consistent evidence of the importance of factors related to the behavior itself, such as perceived usefulness; factors related to perceived risk, such as threat appraisal; and social and normative factors [11,36]. However, in the DCT context, results are mixed. While there is broad support for the importance of factors such as use intention [35] and perceived usefulness [7], evidence of the role of the other factors is less consistent. For example, Tomczyk et al [35] found evidence of the role of both subjective norms and threat appraisal. In contrast, Walrave et al [7] did not include normative factors in their study and found no significant relationship between threat appraisal and DCT adoption. In a different approach to conceptualizing norms, Zabel et al [37] found a strong association between DCT adoption and moral intensity, a construct that derives the perceived obligation for DCT adoption from a range of beliefs, including beliefs about both usefulness and risk. This not only mirrors findings on the association between moral obligation and other pandemic protective behaviors, but as the community benefit of DCT might outweigh the individual benefit, it also appears to be a promising avenue for exploring the relationship between norms and DCT use. Accordingly, it remains an important task of DCT research to understand the relative influence and interplay of both factors such as perceived usefulness, and factors such as threat appraisal or moral obligation on use intention.

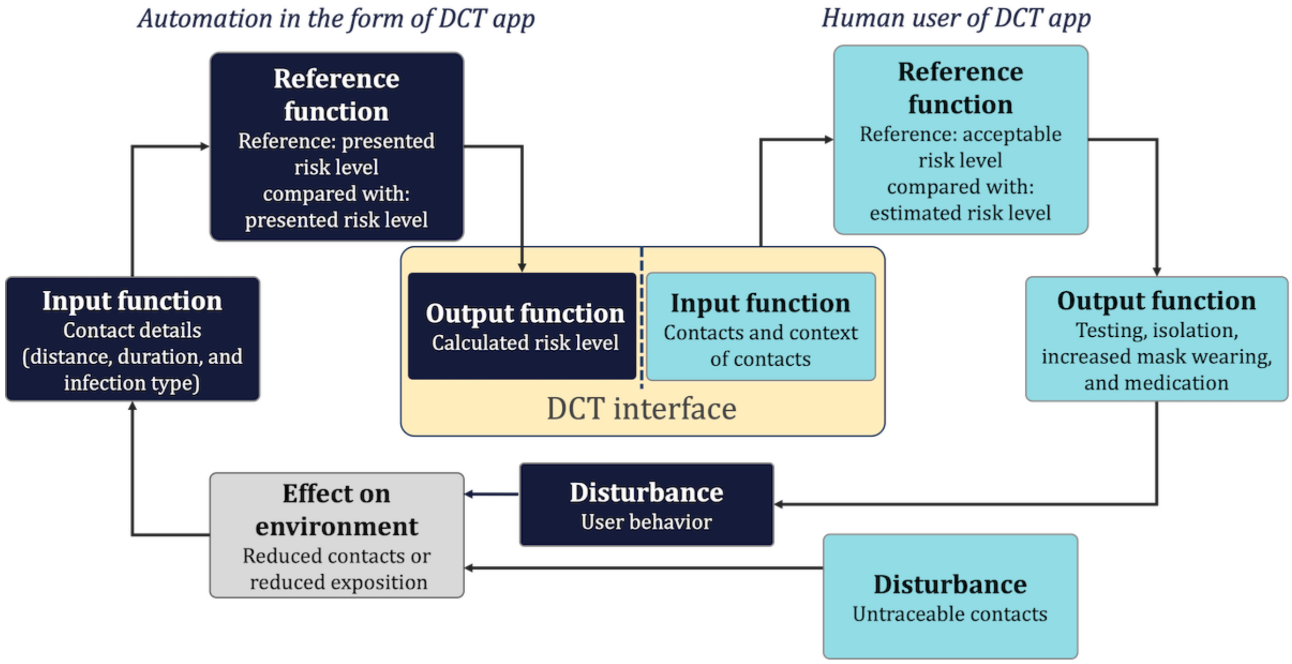
One reason for the ambiguity of existing results can be the variability of operationalizations—trust, for example, is highlighted in multiple studies as decisive for DCT use intention [7,35,37]. However, the conceptualization of trust can be challenging and context-dependent [38]. In DCT, for example, trust could influence one's belief regarding how effectively DCT can support the individual in avoiding an infection. On the other hand, trust can be related to the data security of private information (refer to, eg, the study by Altmann et al [39]). Therefore, a context-sensitive and theory-based conceptualization of trust is necessary to operationalize it adequately.

Breaking Down Automation-Related User Experience in DCT

In a pandemic context, the goal of users can be characterized as behavior that avoids both becoming infected and spreading infection to others. Still, they may desire to meet other people or use public transport and, therefore, are continuously adapting their behavior based on how they perceive the risk situation (ie, for simplification, a perceived risk level; refer to the study by Wilde [40]). This risk level refers to the probability of being infected by, for example, a virus. Acquisition of information on the current risk level is supported by DCT and becomes critical information for comparison, prompting actions to reduce risk.

Contact tracing involves data gathering but also decision-making processes that influence individual and collective health outcomes. It integrates continuous information processing and, therefore, can be viewed through the theoretical lens of control-theoretical conceptions of human-machine systems. The control loop model of action regulation in contact tracing can be extended to accommodate for DCT as automation (ie, a system) that takes over tasks in the acquisition, analysis, and decision selection of contact information [20]. However, maintaining an acceptable risk level [40] is not a singular, finite process but a continuous one. Accordingly, we propose to model information acquisition, analysis, and decision selection as parts of an action regulation consisting of an input function, a reference function, and an output function. As depicted in Figure 1, both human and machine information processing can be modeled within a conceptual control loop to reflect continuous information processing. The conceptual control loop model (Figure 1) illustrates the integration of human and automation activities into a joint action regulation.

Figure 1. Conceptual control loop model of joint human-machine action regulation in digital contact tracing (DCT). The assessment of the machine processing steps (input, reference, and output) is central to the perceived trustworthiness (perceived data validity, perceived goal congruity, and perceived result diagnosticity) of the system.

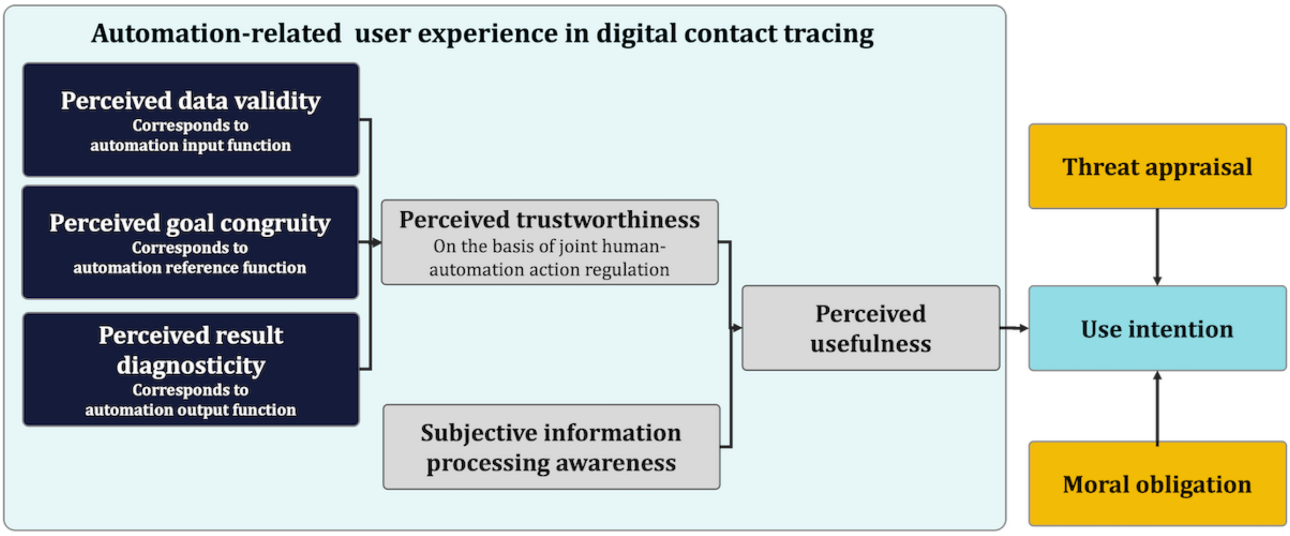


Based on the model presented in Figure 1, we assumed that users' interaction with DCT apps is based on their evaluation of automated input, reference, and output functions. They assess the correctness of the data that the DCT system uses (*input function*), the data's congruence with the users' goals (*reference function*), and the utility of the data's communicated results (*output function*). Any lack of transparency in their joint action regulation can diminish perceived trustworthiness as well as hamper situation awareness. For instance, if the system fails to capture necessary data accurately or align with personal goals such as identifying the source of infection versus alerting those potentially infected, perceived trustworthiness may decline. Accordingly, parallel to similar phenomena in other automation contexts that do not reveal which information is used as part of the input function, an out-of-the-loop unfamiliarity might cause decreasing situation awareness [20]. Furthermore, the user experience may suffer if the system's output, such as an imprecise infection risk description, is insufficient for users to decide the next course of action, therefore impeding the perceived usefulness.

In addition, users' perception of the system is dependent on their expectations of information processing (cf [41]; ie, how the DCT system processes contact-related data). For example, whether a DCT app processes others' vaccination status will only matter to users who are interested in that information, and disclosing that the app processes vaccination information will only impact the system perception of those users. As such, to understand the formation of perceived usefulness, users' subjective situation awareness is more important than their factual situation awareness. However, as introduced by Schrills and Franke [42], subjective evaluation of a user's ability to "perceive, understand and predict a system's information processing," described as subjective information processing awareness, can serve as a construct to assess users' perception of an automation's effect on situation awareness. However, users' perception of their information processing awareness might not be reflected in the accuracy of their knowledge about the system's information processing.

The previous concepts of perceived data validity, goal congruity, result diagnosticity, trustworthiness, subjective information processing awareness, and perceived usefulness can be subsumed as automation-related user experience. Automation-related user experience, following the 9241 standard from the International Organization for Standardization, can be defined as *the perception and response of a person resulting from using or anticipating the use of automated systems*. On the basis of our proposed conception of automation-related user experience, we conceptualized a model of factors of use intention in DCT centered on perceived usefulness of automation as depicted in Figure 2. In addition, threat appraisal and moral obligation as factors independent of DCT use are integrated as measures to evaluate the influence of automation-related user experience on use intention comparatively. Threat appraisal and moral obligation are not connected with properties of the DCT app; that is, they influence whether a user wants to demonstrate behavior to trace contacts but not how useful a specific app is perceived to be.

Figure 2. Research model on automation-related user experience and the effect on use intention of digital contact-tracing apps.



This Study

On the basis of the presented research model, the objective of this study was to investigate how automation-related user experience affects the perceived usefulness of contact tracing as well as the use intention of DCT and how user experience could be improved. We aimed to contribute to research on DCT adoption and use by examining possible pathways to enhance use intention via user experience. On the basis of the proposed research model, we analyzed the following hypotheses: (1) perceived trustworthiness correlates positively with perceived usefulness (hypothesis 1), (2) subjective information processing awareness correlates positively with perceived usefulness (hypothesis 2), and (3) perceived usefulness correlates positively with use intention (hypothesis 3).

In addition, we examined the relationship among all the aforementioned variables in a structural equation modeling (SEM), where we tested automation-related variables as well as variables not related to the specific DCT system: (1) threat appraisal is positively related to use intention (hypothesis 4) and (2) moral obligation is positively related to use intention (hypothesis 5).

Accordingly, the research model depicted in Figure 2 serves as a basis for an SEM analysis that integrated both automation-related user experience and automation-independent variables (threat appraisal and moral obligation).

We supplemented our quantitative findings with qualitative data on the requirements for improved information processing, providing a deeper insight into users' interactions with the app. This mixed methods approach allowed us to uncover underlying patterns and themes that cannot be identified through quantitative data alone, providing a more comprehensive understanding of the user experience.

Methods

Participants

Participants were recruited via social networks (Twitter [subsequently rebranded X] and Facebook), where an image and a link to the study were shared showing a picture of the CWA and asking for participation (ie, our sample was self-selected). The recruitment strategy specifically targeted individuals who had experience using the CWA. Eligibility for the study required participants to be aged ≥18 years and have at least fluent German skills. The study was conducted on the web, with data collection taking place via a web-based questionnaire between June 1, 2022, and July 31, 2022, using LimeSurvey (LimeSurvey GmbH) [43]. We decided not to inquire further about demographic variables to maintain high levels of privacy due to the context of the study (tracking apps).

A total of 317 participants were included in the study (refer to the Data Exclusion section for further details). As user diversity can have a significant impact on the individual user experience and the perceived trustworthiness, we assessed the affinity for technology interaction (ATI) [44]. ATI describes the individual tendency to actively engage in intensive technology interaction. The ATI was measured using a scale validated in various large samples. Our sample ranged from 1 to 6, with an average value of 4.19 (SD 1.26) which was somewhat higher than the value of 3.5 that Franke et al [44] assumed for the general population based on quota sampling. This corresponds with the self-selection of the sample; we can assume that users who installed the CWA may have, in general, a higher level of ATI than the general population.

Ethical Considerations

This study was registered (under 2022-413) at the Ethics Committee of the University of Lübeck. Before participating in the study, individuals received detailed information about the study and provided written consent to partake. For anonymity, no additional demographic data of the users were queried. No financial remuneration was provided for participation.

Scales and Procedure

Overview

To capture the psychological concepts described previously, multiple scales were developed and presented to participants after they provided informed consent. Except for those for experienced system traceability [42], all items were generated by the researchers based on theoretical considerations and discussed within a team of 3 experts in human-machine interaction.

All items used a 6-point Likert response scale (*completely disagree*=1, *largely disagree*=2, *slightly disagree*=3, *slightly agree*=4, *largely agree*=5, and *completely agree*=6), with the only exception being the semantic differential used for perceived usefulness. For all variables except knowledge, a mean score of all items of the scale was calculated and used for further analysis. All the original items were in German and are presented in this manuscript in English.

Use Intention

Use intention was captured using a 3-item scale focusing on participants' intention and future commitment to use the CWA during the pandemic (Multimedia Appendix 1).

Threat Appraisal

A 4-item scale was used aiming to comprehend the participants' perceived risk and concerns related to a possible infection (Multimedia Appendix 1).

Experienced System Traceability

Experienced system traceability was assessed using the 6-item Subjective Information Processing Awareness scale [42] measuring the perceived transparency, understandability, and predictability of information collection and processing by the system (Multimedia Appendix 1).

Moral Obligation

Moral obligation was evaluated using a 3-item scale capturing the participants' sense of responsibility and ethical obligation toward using the CWA (Multimedia Appendix 1).

Perceived Trustworthiness

Perceived trustworthiness was measured across 3 subscales, each addressing the trustworthiness of input, reference, and output in the cybernetic control loop (Multimedia Appendix 1).

Perceived Usefulness

Perceived usefulness was assessed using a semantic differential scale with labels indicative of the perceived efficiency, precision, safety, complexity, and reliability of the system when cooperating with it (for instructions and labels, refer to Multimedia Appendix 1).

Statistical Analysis

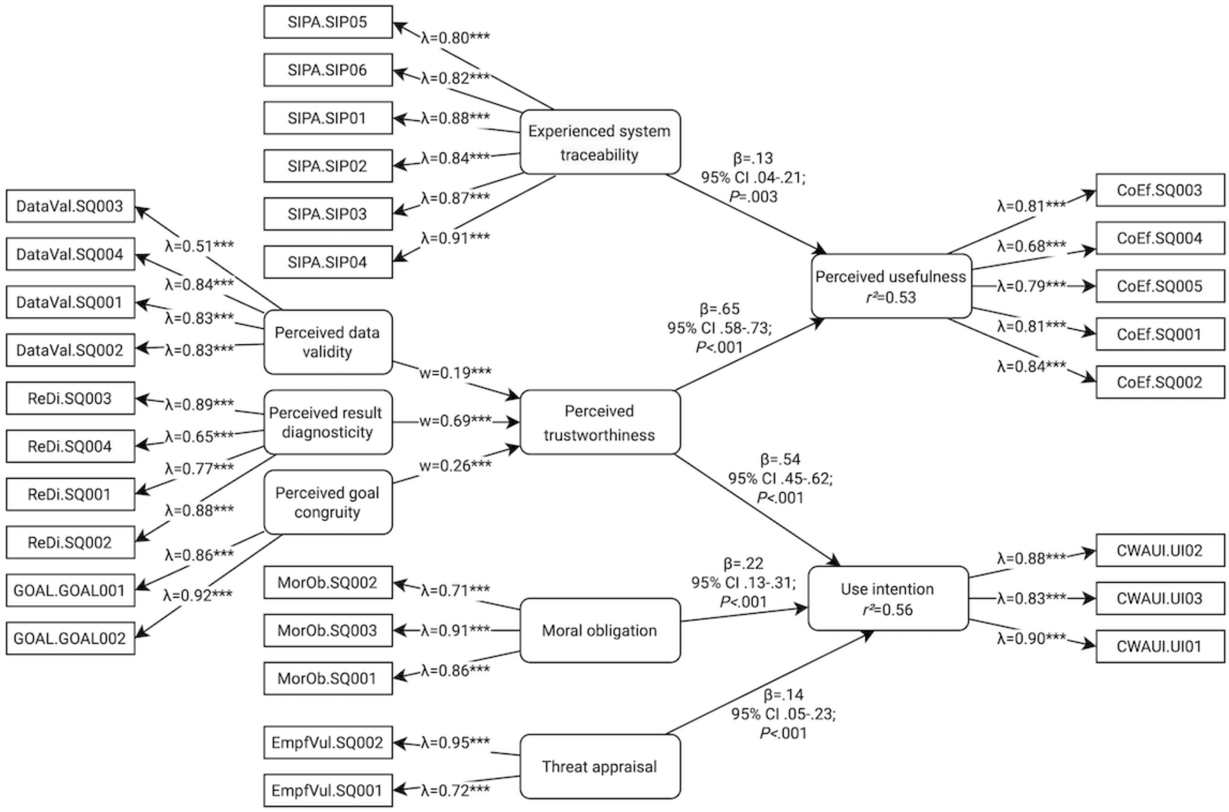
Overview

The data collected in this study were analyzed using R (version 4.31; R Foundation for Statistical Computing) [45]. Initially, the normal distribution of the data was tested to ensure that assumptions of normality were met. Given that the data did not follow a normal distribution, nonparametric tests such as the Welch 2-tailed *t* test were applied to determine statistical significance. In addition, considering the multiple comparisons performed in the calculation of correlations, a Bonferroni correction was used to control for the risk of type I error. Corrected *P* values are reported. The analysis was based on the preregistration, which can be found under <omitted for blinded review>.

PLS-SEM is a statistical modeling method combining aspects of regression and factor analysis. It allows for the simultaneous estimation of the relationship between indicators (ie, manifest variables) and constructs (ie, the latent variables formed from the manifest variables) and the relationship between the constructs themselves. These parts of the models are called the measurement model and structural model [46]. PLS-SEM is robust to nonparametric data, can work with small samples, and is especially suited for exploratory research [47], making it a great fit for this study. We followed the extensive iterative process of model assessment described in the work by Hair [46]. Our iterative approach is documented in Multimedia Appendix 2.

The hypothesized PLS-SEM contains all paths depicted in Figure 3. In addition, we tested whether the paths from perceived trustworthiness, system knowledge, and experienced system traceability to use intention were all mediated by perceived usefulness or whether there were also direct effects.

Figure 3. Partial least squares structural equation model after multiple iterations for the proposed research model. Rounded corners indicate constructs based on our research model; rectangular shapes denote indicators that were measured directly in the survey.



All constructs except perceived trustworthiness were specified as mode-A constructs. The respective indicators are described in the Scales and Procedure section. Perceived trustworthiness was specified as a mode-B higher-order construct consisting of perceived data validity, perceived result diagnosticity, and perceived goal congruity. We report explained variance using R^2, path coefficients using β with *P* values and 95% CIs, and effect sizes using the Cohen f^2.

Power

For the PLS-SEM, a retrospective power analysis using the inverse square root method revealed that, given our sample size (N=317), the smallest path coefficient, and a 5% significance level, we achieved a statistical power of 72% [48].

Data Exclusion

Before the statistical analysis, the data set with 370 responses was carefully reviewed for any inconsistencies, missing data, and outliers. Cases with incomplete or implausible responses (53/370, 14.3% in total) were identified and excluded from the analysis to maintain the integrity of the data set.

Qualitative Data Analysis

To obtain a deeper insight into users' demand for information provision and preservation in the interaction with the CWA, qualitative data were collected via open-ended questions (ie, *what information would you like to get from the system?* [Automation to human; question 1] and *What information would you like to feed to the system?* [Human to automation; question 2]).

As a widely used tool, thematic analysis aims to support the systematic identification, analysis, and reporting of patterns (ie, themes) in qualitative reporting data. Both inductive and deductive approaches were applied using theoretical assumptions as the basis for creating the themes, which were then adapted based on the data collected [49]. The data were coded using MAXQDA (version 20; VERBI GmbH [50]). For a structured and reliable analysis approach, a coding scheme with clear definitions of codes and example coding was developed in multiple iterations (Multimedia Appendix 3). For the evaluation, two perspectives of information needs between humans and automation should be covered: (1) human to automation and (2) automation to human. In total, 2 coders coded the data based on the developed scheme. An intercoder reliability of κ=0.90 (for automation-to-human information demands) and κ=0.87 (for human-to-automation information demands) was achieved. Hence, the level of agreement was strong in both cases [51].

Coded themes for information needs in both automation to human and human to automation included contact or risk information, pandemic-related information, app-related information, and assumptions for perceived information processing. Subcodes were created to enhance coding accuracy (Multimedia Appendix 3) but were not analyzed in detail as the focus remained on the top-level codes. Codes that could not be assigned to one of the themes were assigned to the category *others*. As several participants commented, for example, on the suspected reasons for the limitation of information processing, another category was added (ie, assumed reasons for perceived information processing) to avoid losing these data. Both the categories *others* and *assumed reasons for perceived information processing* were not evaluated for this study.

Missing answers to the questions asked and specific statements that there was no demand for information were assigned the code *none*. This code was assigned only once per person and statement. Thus, in the end, it was possible to clearly distinguish how many of the 317 respondents indicated information needs and how many did not. Ultimately, automation-to-human information demand statements from 45.4% (144/317) of the participants and human-to-automation information demand statements from 27.1% (86/317) of the participants were analyzed (Table 1).

Table 1. Number of respondents that indicated information demands versus no information demands.

Variable	Response distribution, n (%)	
	Respondents (n=317)	Responses (n=377)
Demands		
Information demand (A2H[a])	144 (45.4)	257 (68.2)
Information demand (H2A[b])	86 (27.1)	120 (31.8)
No demands		
Information demand (A2H)	173 (54.5)	120 (31.8)
Information demand (H2A)	231 (72.9)	257 (68.2)

[a]A2H: automation to human.

[b]H2A: human to automation.

Results

Overview

For hypothesis 1, the analysis revealed moderate positive correlations for all factors of perceived trustworthiness. The correlation between data validity and perceived usefulness was significant, with a coefficient of r=0.38 and P<.001. The correlation between goal congruity and perceived usefulness showed a coefficient of r=0.47 and P<.001, indicating a moderate positive linear relationship. Result diagnosticity and perceived usefulness exhibited a strong positive correlation, with a coefficient of r=0.56 and P<.001. In general, all measures of perceived trustworthiness and perceived usefulness exhibited a positive relationship, supporting hypothesis 1.

For hypothesis 2, a correlation coefficient of r=0.35 (P<.001) was observed, suggesting a moderate positive linear relationship between subjective information processing awareness and perceived usefulness; a positive relationship between SIPA and perceived usefulness (hypothesis 2) was supported by the data. This indicates that automation-related phenomena such as changes in situation awareness might influence the perceived usefulness of DCT.

For hypothesis 3, the correlation coefficient between perceived usefulness and use intention was r=0.47 and P<.001, indicating a moderate positive correlation. Hence, our results support the hypothesis (hypothesis 3) that perceived usefulness is positively related to use intention (hypothesis 3). In combination with our previous results, this indicates strong relationships between user experience variables and use intention.

In summary, all variables showed statistically significant correlations with perceived usefulness. These correlations ranged from moderate to strong positive relationships. These results strengthen our assumption that perceived usefulness of DCT is strongly related to automation-related user experience.

SEM Approach

The final PLS-SEM is depicted in Figure 3. The explained variance for use intention was R^2=0.56. It was directly predicted by perceived trustworthiness (β=.54, 95% CI .45-.62; P<.001; f^2=0.44), moral obligation (β=.22, 95% CI .13-.31; P<.001; f^2=0.07), and threat appraisal (β=.14, 95% CI .05-.23; P<.001; f^2=0.04). Thus, there was a large effect for perceived trustworthiness and a small effect for the other constructs. Still, hypotheses 4 and 5 were supported.

Within the perceived trustworthiness higher-order construct, the highest weight was assigned to perceived result diagnosticity (w=0.69; P<.001), implying that this subconstruct contributes most to perceived trustworthiness, followed by perceived goal congruity (w=0.26; P<.001) and perceived data validity (w=0.19; P<.001).

We did not find evidence for a mediating effect of perceived usefulness on the paths from perceived trustworthiness, system knowledge, and experienced system traceability to use intention. However, we did find direct effects of perceived trustworthiness (β=.65, 95% CI .58-.73; P<.001; f^2=0.65) and experienced system traceability (β=.13, .04-.21; P=.003; f^2=0.02) on perceived usefulness (R^2=0.53).

Qualitative Analysis

Overview

Two directions of information flow were analyzed to assess the information demands of CWA users: (1) human to automation—information that users want to provide to the system and (2) automation to human—information that users want to receive from the system. In total, 3 overarching themes were explored and analyzed in more detail (Textbox 1).

Textbox 1. Analyzed themes and description of each theme. The detailed coding scheme can be found in Multimedia Appendix 3.

Contact- or risk-related information

- Time-related information: information regarding the period of the contact, the duration of the contact, the time passed since the contact, and the period during which contact tracing was possible
- Location-related information: information related to the place of contact, direct or indirect contact, and indoor or outdoor contact
- Exposition-related information: information about the masking status in the contact situation and the distance between the persons in contact
- Action-related information: information on possible and suggested courses of action after contact
- Information related to the warning person: information concerning the time when the warning person tested positive, the time when the warning person became infected, the warning person's first symptoms, the warning person's vaccination status, and the infected person's virus variant

Pandemic-related information

- Statistics: information related to statistical content on the pandemic in terms of the number of defects or infections

App-related information

- Number of users: information about the number of users of the Corona-Warn-App
- General calculation-related information: information on reasons for changing risk calculation and the system parameters used for calculations
- Certainty about the result: information related to the certainty of the results calculated by the system
- Integration of tests (self- and externally administered): information about the possibility to enter or delete test results on the app
- Linking with private data: information on the possibility of linking app functions with private data

Descriptive Data

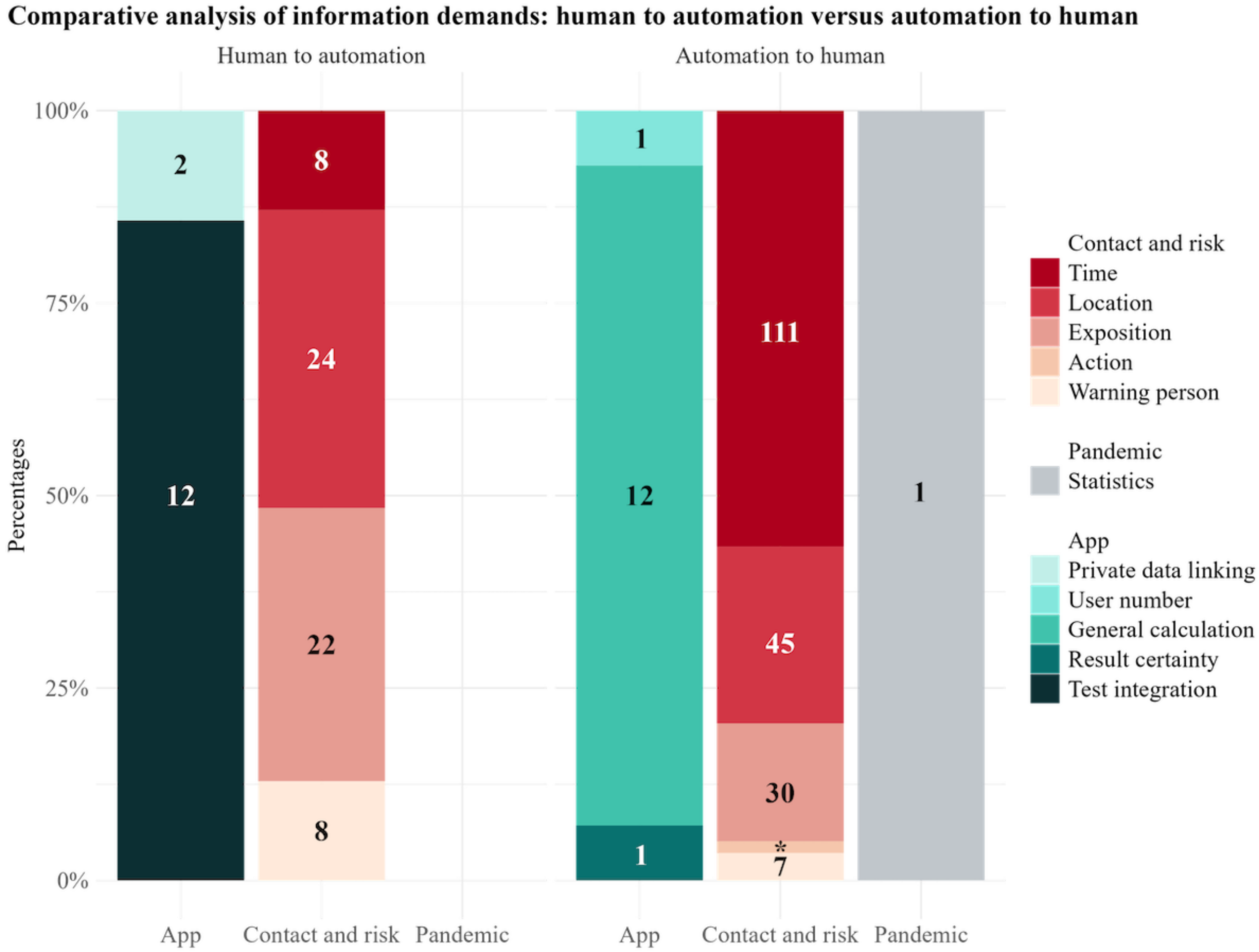
Overview

The overall number of statements amounted to 211 in automation to human and 76 in human to automation. Within these 2 categories, the themes were distributed unevenly. Information regarding contact and risk accounted for most statements in both categories (automation to human: 196/211, 92.9% of statements; human to automation: 62/76, 82% of statements). The remaining statements were (almost) exclusively distributed among app-related information (automation to human: 14/211, 6.6% of statements; human to automation: 14/76, 18% of statements) as barely any needs were stated for pandemic-related information (automation to human: 1/211, 0.5% of statements; human to automation: 0 statements).

Regarding the subcodes, the distribution also varied between both themes (Figure 4). For contact- and risk-related information, the information related to time, location, and exposure accounted for the largest proportion of demands within this theme in both categories. However, the distribution of statement proportions differed clearly between automation to human and human to automation. Time-related information was demanded most in automation to human (111/196, 56.6% of statements) but least in human to automation (8/62, 13% of statements). Demands for location-related information did not differ greatly between automation to human (45/196, 23% of statements) and human to automation (24/62, 39% of statements), nor did exposition-related information (automation to human: 30/196, 15.3% of statements; human to automation: 22/62, 35% of statements).

In terms of *app-related information*, the demands for information about the system's general calculation (automation to human: 12/14, 86% of statements; human to automation: 0% of statements) and the integration of tests (automation to human: 0% of statements; human to automation: 12/14, 86% of statements) differed in particular between the categories. The remaining subcodes hardly received any consideration. In both categories (automation to human and human to automation), almost no statements regarding *pandemic-related information* were made.

Figure 4. Relative demands regarding information from automation to human (left) and from human to automation (right). The numbers in the column sections indicate the number of statements under each code.



Human to Automation

In human to automation, certain claims emerged with particular frequency in the demand for contact- and risk-related and app-related information. Information demands on contact and risk mainly focused on time- and exposure-related information. For example, the interest in informing the app of one's location and whether one was in an enclosed space or outdoors was present:

> *Tell the app something about the specific location (enclosed space, fresh air).*

Exposition-related information demands mainly focused on informing the app when one wore or had worn a mask:

> *The wearing of a mouth-nose covering should be entered and thus taken into account in the risk calculation.*

Regarding the demand for the integration of app-related information, the participants predominantly highlighted the integration of self-administered or externally administered tests:

> *That I am Corona positive without having done a Polymera-Chain Reaction (PCR) test. (Perhaps with indication that the result is not PCR verified).*

Automation to Human

In the automation to human category, contact- and risk-related and app-related information were queried with similar frequency. The contact- and risk-related information in this category most often referred to time-related information with a request for the time of the risk encounter. However, the desired preciseness of the temporal data differed (exact time vs more approximate time: "When was the encounter? (At least as a time frame, e.g., between 8-12 o'clock)" vs "The specific time [...] of a risk encounter would be helpful"). The location of the risk encounter was another type of information that participants commonly solicited. Most asked for information about a rather specific location ("At which location did a contact take place?"); few seemed to be interested in the characteristics of the location ("Indoors or outdoors?").

Exposure-related information demanded from the system included the number of devices or persons present at the time of exposure ("[...] with how many devices was the contact?"), the distance to the warning person ("At what distance was the encounter?"), and the masking status. In particular, masking status included the person's own status of having worn a mask or whether the other person was wearing a mask at the time of the risk encounter ("Was I wearing a mask? Was the other person wearing a mask?").

App-related information demands mainly focused on the parameters of the calculation ("What factors led to this result?") and reasons for a status change ("How exactly the risk determination works, i.e., how distance and time to a positively tested person actually have to be, in order for me to receive a notification and for the status to be changed").

Discussion

Principal Findings

The objective of this study was to understand automation-related user experience, its connection to perceived usefulness, and the use intention of DCT. Our data showed that perceived trustworthiness is a critical factor in understanding use intention as well as the perceived usefulness of DCT apps. Interestingly, users' experience of a system as supportive in their action regulation affects their use intention more strongly than external factors such as threat appraisal or moral obligation. In addition, our qualitative analysis revealed that users mainly want to communicate with the system about information that is relevant to their decision-making. For instance, providing more precise information about masking status when in contact with other people could assist a user in making an immediate decision regarding isolation. Overall, our findings suggest a strong relationship between the diagnosticity of automated information processing and use intention.

Practical Implications

As a first major implication, the high effect of result diagnosticity on perceived trustworthiness demonstrates the importance of human-centered information processing in (partially) automated health applications. Within the interconnected human-machine information processing loops (Figure 1), the machine provides information as part of the human input function. As discussed by Miller [52], intelligent systems such as DCT should aim to improve users' ability to access and use (processed) information rather than to present and justify a particular outcome. In DCT app design, the integration of DCT information into a joint human-automation action regulation should be prioritized. Accordingly, when developing evaluative systems [52] that support the evaluation of alternatives rather than suggesting specific actions, it is important to consider what evaluative process a user needs to undertake. While previous research has already identified the need for actionable information [53], the information presented by DCT apps needs to be understood in the context of human action regulation and the influence of automated systems in human action regulation. A possible solution to support diagnosticity in DCT is so-called proactive contact tracing [54], which integrates more information sources and can potentially enrich DCT results.

Second, the results indicate a strong user need for information to be provided in sufficient detail. An interface optimized for communicating information could enable users to make their own assessment of the situation. In many DCT apps, users request the ability to retrieve information about possible contacts, such as time, location, or even the person involved [55]. Our study showed similar results (eg, a high demand for detailed information about the [exact] time of detected contacts). Again, the demand for more detailed information relates to the diagnosticity of the information provided by the system. If users are only given information about their current risk of infection, they cannot evaluate the validity of this information, potentially leading them to ignore it. They would require additional context-related information about potential contacts, such as whether the individuals were wearing masks or were located in an enclosed room, to make informed decisions about their behavior. Our results demonstrate that use intention is strongly connected to the perceived diagnosticity of the DCT app. On the basis of our qualitative findings, we can assume that the diagnosticity of DCT users depends on the level of detail they receive about possible contacts. Accordingly, the provision of details that support users' information processing is even more important for their use intention than threat appraisal or moral obligation. In accordance with psychological research on motivation [56], supporting users' intrinsic motivation for diagnostic information could lead to better adherence regarding DCT apps than, for instance, exposing them to extrinsic motivators that increase threat appraisal (eg, describing the consequences of infection [57]).

Third, in contradiction to users' demand for detailed information on contacts, a major concern in DCT is privacy [55]. While it is often argued that too much detail conflicts with privacy, it is important to find ways to improve the diagnosticity of information as this determines the use intention. Possible solutions include differential privacy, which allows for sufficient detail for increased diagnosticity while keeping personal data confidential. In addition, many users requested features that do not compromise the privacy of others, such as the ability to inform the system about masking status. Thus, allowing users to refine the input received by the DCT app may increase the perceived diagnosticity of the results. The integration of masking status can be seen as a measure to improve the accuracy of the apps in determining risk levels, ultimately increasing the use intention.

Overall, our results suggest that focusing on the diagnosticity of the information presented in DCT apps could result in improvement in users' health behavior. During the COVID-19 pandemic, users reported that they were unsure about the correct or best action to take to contain the pandemic or could not correctly assess the risk of certain situations [58]. However, this certainty is particularly important when it comes to health decisions. With sufficient diagnostic accuracy, DCT apps may be able to better reduce this uncertainty and, thus, become a crucial component in the management of pandemics in the long term, also positively affecting users' willingness to provide data on a social level. It is also crucial that DCT apps do not follow the *recommend and defend* principle [52], which could lead to a long-term reduction in motivation, but instead provide information that supports individual decisions. If compliance with effective pandemic control measures can be increased as a result, it will be possible to respond more effectively to future pandemics.

Theoretical and Methodological Implications

In our data, the perceived trustworthiness of a DCT app had a greater influence on use intention than threat appraisal or moral obligation. Furthermore, while previous studies [26] have relied on perceived usefulness, our findings in the PLS-SEM do not suggest that it mediates the relationship between perceived trustworthiness and use intention. However, usefulness can be seen as an ambiguous concept without a specific connection to the design of DCT apps. In this way, focusing on perceived usefulness could hinder approaches to improve DCT by adopting DCT app design and functionality. In contrast, a lack of perceived result diagnosticity indicates to developers that the information provided by a DCT app needs to be adapted to have an impact on joint action regulation. Our research suggests that designers of automated systems should specify the potential actions that users can take and identify decision points at which users may require diagnostic information, such as whether to proceed with a specific action. In addition, highlighting the role of diagnosticity indicates how models of technology in medical systems should be developed. Existing models (such as the technology acceptance model) do not specify to what extent a system's usefulness depends on perceived diagnosticity. Our research demonstrates that behavioral models focusing on information-based decisions are needed to address automated technology in health, for example, DCT.

However, one can argue that the difference between perceived result diagnosticity and perceived usefulness is arbitrary; in a joint human-automation action regulation, the diagnosticity of information seems to be equal to perceived usefulness. However, by directly addressing perceived result diagnosticity as a central variable of automation-related user experience, empirical research can identify paths to improve action regulation support of DCT without previously defining what is useful about a system or not. When a DCT app can deliver information that users can use to regulate their actions, users report a higher intention to use it. Therefore, applying result diagnosticity as a variable in human-automation research is a methodological contribution supporting future research in intelligent automation.

On the basis of our findings, future research on DCT needs to determine how to improve the diagnosticity of DCT apps. This paper introduced a conceptual control loop model of joint human-machine action regulation, which can support research approaches in optimizing perceived diagnosticity as a central variable for automation-related user experience. Addressing the joint action regulation in DCT and health behavior is crucial to understand how the information provided by DCT apps can be integrated into human information processing and how DCT apps influence the human output function. Information that improves the evaluation of individual contacts, such as contact location, masking status, or vaccination level, could improve perceived trustworthiness and use intention of DCT apps. By demonstrating how information processing between human users and DCT apps is integrated, our research supports a shift from viewing human users as receivers of machine results to viewing them as actors using DCT information.

All in all, our findings regarding the significance of diagnosticity have implications for the design of automated information processing in a broader context. Users did not primarily prioritize data validity or goal congruence; instead, their focus lay in determining whether they could trust the system to provide information that would assist their own decision-making process. This may be a general trend in automated information processing.

Limitations and Further Research

All participants of this study were users of the CWA. However, as Walrave et al [59] describe, many citizens in Germany did not use DCT apps, for example, because they did not want to share their data or did not think they were effective. Thus, the findings presented on the impact of perceived diagnosticity may not be applicable to citizens who did not use the app at all. These individuals may have chosen not to use the app for reasons beyond those discussed in this paper. The perceived diagnosticity of a DCT app is only relevant for use intention when potential users are interested in determining their individual risk level or making decisions based on their estimated risk level. That is, our sample may bias the results and underestimate factors relevant to nonusers. For example, nonusers might reject the app because they do not trust the provider of the system. Accordingly, the results of our study may support improving DCT for existing users but not convincing nonusers to use DCT. Further studies need to address nonusers and examine how automation-related user experience affects their decision not to use DCT.

In addition, users may have misconceptions about the factors contributing to the risk of infection and may expect the system to provide irrelevant information that does not aid in making an informed decision. Accordingly, they might report a low perceived diagnosticity while the information provided in the app offers sufficient diagnosticity. The accuracy of one's mental model [60] may influence the perception of actual diagnostic information as nondiagnostic (for a discussion of diagnosticity, refer to the study by Garcia-Marques et al [61]). To tackle false models of diagnosticity, DCT apps should support users in correcting their mental model, for example, by explaining how they can use the provided information. This could be done by simulating decision situations with and without DCT information, offering users the experience of diagnosticity.

Improving the perceived diagnosticity could be beneficial for use intention but could negatively affect perceived data privacy [55]. For example, a function that allows users to communicate when they are wearing a mask could be abused to track specific contacts, therefore revealing potential infections of other users. Data privacy is a critical concern in DCT use [59]. Therefore, current DCT apps are designed to protect the data of other users at the cost of the diagnosticity of information. This research did aim to understand the effect of user experience in automated DCT but did not include how users evaluate potential risks of data privacy violations or approaches to address them (cf [62]). Future research should identify how to balance the desired level of perceived result diagnosticity and data privacy concerns. For example, in direct communication, users who reveal information about their web-based status can see the web-based status of others, allowing them to choose which balance between diagnosticity and data protection they desire. The same function could be implemented in DCT apps to support automation-related user experience. Allowing users to choose their level of diagnosticity themselves allows them also to control how DCT apps influence their decision-making, thus strengthening user autonomy.

Finally, this study had a cross-sectional design that did not assess how automation-related user experience and use intention regarding a DCT app may change over time. Previous research has demonstrated that automation-related user experience can change over time (eg, because users adapt to the system or they improve how they use the system). Future research on automation-related user experience in DCT apps needs to include a longitudinal study design to capture effects of behavior change and users' perception.

Conclusions

In conclusion, this research highlights the relevance of automation-related user experience in DCT and its role in enabling the effective action regulation of DCT users. Here, providing detailed and diagnostic information is crucial for users to make informed assessments of their situation and actions. The presented quantitative results echo the qualitatively assessed user demand for more detailed information about potential contacts, such as time, location, and context (eg, mask use and indoor or outdoor setting).

Interestingly, our data suggest that other factors not directly related to the app, such as moral obligation and threat appraisal, are less relevant compared to automation-related user experience, especially to the perceived diagnosticity of the information provided by DCT apps. The presented results are also more specific than those of previous studies that relied on perceived usefulness. Our research model did not suggest that perceived usefulness mediates the relationship between perceived trustworthiness and use intention. Instead, we propose that DCT designers should focus on providing diagnostic information at critical decision points.

However, privacy remains a major concern in DCT. While it is often argued that too much detail conflicts with privacy, it is crucial to find ways to improve the diagnosticity of information without compromising privacy. Solutions could include differential privacy or features that do not compromise the privacy of others, such as the ability to inform the system about masking status.

The main impact of our results on the design of DCT apps and health policy is that DCT apps need to provide sufficient diagnosticity to be perceived as useful. This means that (1) the possible actions of users need to be understood before the design of the DCT algorithm and apps and (2) the presented information needs to support them in choosing the correct action. Focusing on the diagnosticity of the information presented in DCT apps could, in turn, also influence user performance. During the COVID-19 pandemic, a significant percentage of users reported uncertainty about the best actions to take or could not correctly assess the risk of certain decisions. Therefore, improving diagnostics could contribute to better and safer decisions.

In summary, our study underscores the importance of balancing detailed and diagnostic information with privacy concerns in DCT apps. As we move forward in this digital age, it is crucial to continue exploring ways to optimize DCT while respecting user privacy.

Acknowledgments

This research was funded by the Federal Ministry of Education and Research of Germany within the framework of the Cooperative and Communicating AI project (project 01GP1908).

Authors' Contributions

TS contributed to conceptualization, methodology, investigation, resources, data curation, data analysis, writing—original draft, visualization, and funding acquisition. LK contributed to data curation, data analysis, and writing—review and editing. MG contributed to investigation, data curation, writing—review and editing, and visualization. ACV contributed to writing—review and editing and supervision. TF contributed to conceptualization, writing—review and editing, and supervision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Scales for the cross-sectional survey study on user experience.
[PDF File (Adobe PDF File), 68 KB - humanfactors_v11i1e53940_app1.pdf]

Multimedia Appendix 2
Documentation of the iterative process of partial least squares structural equation modeling.
[PDF File (Adobe PDF File), 418 KB - humanfactors_v11i1e53940_app2.pdf]

Multimedia Appendix 3
Coding scheme.
[PDF File (Adobe PDF File), 125 KB - humanfactors_v11i1e53940_app3.pdf]

References

1. Barrat A, Cattuto C, Kivelä M, Lehmann S, Saramäki J. Effect of manual and digital contact tracing on COVID-19 outbreaks: a study on empirical contact data. J R Soc Interface 2021 May 05;18(178):20201000 [FREE Full text] [doi: 10.1098/rsif.2020.1000] [Medline: 33947224]
2. Shahroz M, Ahmad F, Younis MS, Ahmad N, Kamel Boulos MN, Vinuesa R, et al. COVID-19 digital contact tracing applications and techniques: a review post initial deployments. Transp Eng 2021 Sep;5:100072. [doi: 10.1016/j.treng.2021.100072]
3. O'Connell J, Abbas M, Beecham S, Buckley J, Chochlov M, Fitzgerald B, et al. Best practice guidance for digital contact tracing apps: a cross-disciplinary review of the literature. JMIR Mhealth Uhealth 2021 Jun 07;9(6):e27753 [FREE Full text] [doi: 10.2196/27753] [Medline: 34003764]
4. Corona-Warn-App. GitHub. 2020. URL: https://github.com/corona-warn-app [accessed 2023-10-16]
5. Corona-Warn-App: 45 millionen downloads. Presse- und Informationsamt der Bundesregierung. 2022. URL: https://www.bundesregierung.de/breg-de/themen/coronavirus/cwa-45-mio-downloads-1994916 [accessed 2023-10-23]
6. Li T, Cobb C, Yang JJ, Baviskar S, Agarwal Y, Li B, et al. What makes people install a COVID-19 contact-tracing app? Understanding the influence of app design and individual difference on contact-tracing app adoption intention. Pervasive Mob Comput 2021 Aug;75:101439 [FREE Full text] [doi: 10.1016/j.pmcj.2021.101439] [Medline: 36569467]
7. Walrave M, Waeterloos C, Ponnet K. Adoption of a contact tracing app for containing COVID-19: a health belief model approach. JMIR Public Health Surveill 2020 Sep 01;6(3):e20572 [FREE Full text] [doi: 10.2196/20572] [Medline: 32755882]
8. Janz NK, Becker MH. The health belief model: a decade later. Health Educ Q 1984 Jan 01;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]
9. Ahadzadeh AS, Pahlevan Sharif S, Ong FS, Khong KW. Integrating health belief model and technology acceptance model: an investigation of health-related internet use. J Med Internet Res 2015 Feb 19;17(2):e45 [FREE Full text] [doi: 10.2196/jmir.3564] [Medline: 25700481]
10. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
11. Huang Z, Guo H, Lim HY, Chow A. Determinants of the acceptance and adoption of a digital contact tracing tool during the COVID-19 pandemic in Singapore. Epidemiol Infect 2022 Mar 02;150:e54. [doi: 10.1017/s0950268822000401]
12. Oyibo K, Morita PP. Factors influencing the willingness to download contact tracing apps among the American population. In: Proceedings of the Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization. 2023 Presented at: UMAP '23; June 26-29, 2023; Limassol, Cyprus p. 147-156 URL: https://dl.acm.org/doi/abs/10.1145/3563359.3596983 [doi: 10.1145/3563359.3596983]
13. Oyibo K, Sahu KS, Oetomo A, Morita PP. Factors influencing the adoption of contact tracing applications: systematic review and recommendations. Front Digit Health 2022;4:862466 [FREE Full text] [doi: 10.3389/fdgth.2022.862466] [Medline: 35592459]
14. Vogt F, Haire B, Selvey L, Katelaris AL, Kaldor J. Effectiveness evaluation of digital contact tracing for COVID-19 in New South Wales, Australia. Lancet Public Health 2022 Mar;7(3):e250-e258 [FREE Full text] [doi: 10.1016/S2468-2667(22)00010-X] [Medline: 35131045]
15. Prakash AV, Das S, Pillai KR. Understanding digital contact tracing app continuance: insights from India. Health Policy Technol 2021 Dec;10(4):100573. [doi: 10.1016/j.hlpt.2021.100573]
16. Onnasch L, Wickens CD, Li H, Manzey D. Human performance consequences of stages and levels of automation: an integrated meta-analysis. Hum Factors 2014 May 20;56(3):476-488. [doi: 10.1177/0018720813501549] [Medline: 24930170]
17. Bowden V, Griffiths N, Strickland L, Loft S. Detecting a single automation failure: the impact of expected (but not experienced) automation reliability. Hum Factors 2023 Jun;65(4):533-545. [doi: 10.1177/00187208211037188] [Medline: 34375538]
18. Clegg BA, Vieane AZ, Wickens CD, Gutzwiller RS, Sebok AL. The effects of automation-induced complacency on fault diagnosis and management performance in process control. Proc Hum Factors Ergon Soc Annu Meet 2014 Oct 17;58(1):844-848. [doi: 10.1177/1541931214581178]
19. Parasuraman R, Manzey DH. Complacency and bias in human use of automation: an attentional integration. Hum Factors 2010 Jun 20;52(3):381-410. [doi: 10.1177/0018720810376055] [Medline: 21077562]
20. Parasuraman R, Sheridan TB, Wickens CD. A model for types and levels of human interaction with automation. IEEE Trans Syst Man Cybern A Syst Hum 2000 May;30(3):286-297. [doi: 10.1109/3468.844354] [Medline: 11760769]
21. Endsley MR. Automation and situation awareness. In: Parasuraman R, Mouloua M, editors. Automation and Human Performance: Theory and Applications. Boca Raton, FL: CRC Press; 1996:163-181.
22. Bahner JE, Hüper AD, Manzey D. Misuse of automated decision aids: complacency, automation bias and the impact of training experience. Int J Hum Comput Stud 2008 Sep;66(9):688-699. [doi: 10.1016/j.ijhcs.2008.06.001]
23. Volz K, Yang E, Dudley R, Lynch E, Dropps M, Dorneich MC. An evaluation of cognitive skill degradation in information automation. Proc Hum Factors Ergon Soc Annu Meet 2016 Sep 15;60(1):191-195. [doi: 10.1177/1541931213601043]

24. Röttger S, Bali K, Manzey D. Impact of automated decision aids on performance, operator behaviour and workload in a simulated supervisory control task. Ergonomics 2009 May 28;52(5):512-523. [doi: 10.1080/00140130802379129] [Medline: 19296323]
25. Hoff KA, Bashir M. Trust in automation: integrating empirical evidence on factors that influence trust. Hum Factors 2015 May;57(3):407-434. [doi: 10.1177/0018720814547570] [Medline: 25875432]
26. Krüger N, Behne A, Beinke JH, Stibe A, Teuteberg F. Exploring user acceptance determinants of COVID-19-tracing apps to manage the pandemic. Int J Technol Hum Interact 2022;18(1):27. [doi: 10.4018/ijthi.293197]
27. Costa MF. Health belief model for coronavirus infection risk determinants. Rev Saude Publica 2020 May 07;54:47 [FREE Full text] [doi: 10.11606/s1518-8787.2020054002494] [Medline: 32491096]
28. Kojan L, Burbach L, Ziefle M, Calero Valdez A. Perceptions of behaviour efficacy, not perceptions of threat, are drivers of COVID-19 protective behaviour in Germany. Humanit Soc Sci Commun 2022 Mar 24;9(1):97. [doi: 10.1057/S41599-022-01098-4]
29. Chu H, Liu S. Integrating health behavior theories to predict American's intention to receive a COVID-19 vaccine. Patient Educ Couns 2021 Aug;104(8):1878-1886 [FREE Full text] [doi: 10.1016/j.pec.2021.02.031] [Medline: 33632632]
30. Kleinman RA, Merkel C. Digital contact tracing for COVID-19. CMAJ 2020 Jun 15;192(24):E653-E656 [FREE Full text] [doi: 10.1503/cmaj.200922] [Medline: 32461324]
31. Maccari L, Cagno V. Do we need a contact tracing app? Comput Commun 2021 Jan 15;166:9-18 [FREE Full text] [doi: 10.1016/j.comcom.2020.11.007] [Medline: 33235399]
32. Ahmed N, Michelin RA, Xue W, Ruj S, Malaney R, Kanhere SS, et al. A survey of COVID-19 contact tracing apps. IEEE Access 2020;8:134577-134601. [doi: 10.1109/access.2020.3010226]
33. Sacks JA, Zehe E, Redick C, Bah A, Cowger K, Camara M, et al. Introduction of mobile health tools to support Ebola surveillance and contact tracing in Guinea. Glob Health Sci Pract 2015 Nov 12;3(4):646-659 [FREE Full text] [doi: 10.9745/GHSP-D-15-00207] [Medline: 26681710]
34. Velicia-Martin F, Cabrera-Sanchez JP, Gil-Cordero E, Palos-Sanchez PR. Researching COVID-19 tracing app acceptance: incorporating theory from the technological acceptance model. PeerJ Comput Sci 2021;7:e316 [FREE Full text] [doi: 10.7717/peerj-cs.316] [Medline: 33816983]
35. Tomczyk S, Barth S, Schmidt S, Muehlan H. Utilizing health behavior change and technology acceptance models to predict the adoption of COVID-19 contact tracing apps: cross-sectional survey study. J Med Internet Res 2021 May 19;23(5):e25447 [FREE Full text] [doi: 10.2196/25447] [Medline: 33882016]
36. Shushtari ZJ, Salimi Y, Ahmadi S, Rajabi-Gilan N, Shirazikhah M, Biglarian A, et al. Social determinants of adherence to COVID-19 preventive guidelines: a comprehensive review. Osong Public Health Res Perspect 2021 Dec;12(6):346-360 [FREE Full text] [doi: 10.24171/j.phrp.2021.0180] [Medline: 34965686]
37. Zabel S, Schlaile MP, Otto S. Breaking the chain with individual gain? Investigating the moral intensity of COVID-19 digital contact tracing. Comput Human Behav 2023 Jun;143:107699 [FREE Full text] [doi: 10.1016/j.chb.2023.107699] [Medline: 36818428]
38. Johnson M, Bradshaw JM. The role of interdependence in trust. In: Nam CS, Lyons JB, editors. Trust in Human-Robot Interaction. Cambridge, MA: Academic Press; 2021:379-403.
39. Altmann S, Milsom L, Zillessen H, Blasone R, Gerdon F, Bach R, et al. Acceptability of app-based contact tracing for COVID-19: cross-country survey study. JMIR Mhealth Uhealth 2020 Aug 28;8(8):e19857 [FREE Full text] [doi: 10.2196/19857] [Medline: 32759102]
40. Wilde GJ. Risk homeostasis theory: an overview. Inj Prev 1998 Jun 01;4(2):89-91 [FREE Full text] [doi: 10.1136/ip.4.2.89] [Medline: 9666358]
41. Roscoe RD, Wilson J, Johnson AC, Mayra CR. Presentation, expectations, and experience: sources of student perceptions of automated writing evaluation. Comput Human Behav 2017 May;70:207-221. [doi: 10.1016/j.chb.2016.12.076]
42. Schrills T, Franke T. How do users experience traceability of AI systems? Examining subjective information processing awareness in automated insulin delivery (AID) systems. ACM Trans Interact Intell Syst 2023;13(4):1-34 [FREE Full text] [doi: 10.31234/osf.io/3v9b8]
43. LimeSurvey: an open source survey tool. LimeSurvey GmbH. URL: http://www.limesurvey.org [accessed 2024-04-29]
44. Franke T, Attig C, Wessel D. A personal resource for technology interaction: development and validation of the affinity for technology interaction (ATI) scale. Int J Hum Comput Interact 2018 Mar 30;35(6):456-467. [doi: 10.1080/10447318.2018.1456150]
45. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2021. URL: https://www.R-project.org/ [accessed 2024-04-29]
46. Hair JF, Hult GT, Ringle CM, Sarstedt M. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). 2nd edition. Thousand Oaks, CA: Sage Publications; 2017.
47. Hair JF, Risher JJ, Sarstedt M, Ringle CM. When to use and how to report the results of PLS-SEM. Eur Bus Rev 2019 Jan 14;31(1):2-24. [doi: 10.1108/ebr-11-2018-0203]
48. Kock N, Hadaya P. Minimum sample size estimation in PLS‐SEM: the inverse square root and gamma‐exponential methods. Inf Syst J 2018;28(1):227-261. [doi: 10.1111/isj.12131]

49. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
50. All-in-one tool for qualitative data analysis & mixed methods. MAXQDA - Distribution by VERBI GmbH. URL: https://www.maxqda.com [accessed 2022-08-17]
51. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
52. Miller T. Explainable AI is dead, long live explainable AI!: hypothesis-driven decision support using evaluative AI. In: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency. 2023 Presented at: FAccT '23; June 12-15, 2023; Chicago, IL p. 333-342 URL: https://dl.acm.org/doi/10.1145/3593013.3594001 [doi: 10.1145/3593013.3594001]
53. Caballero A, Leath K, Watson J. COVID-19 consumer health information needs improvement to be readable and actionable by high-risk populations. Front Commun 2020 Aug 7;5 [FREE Full text] [doi: 10.3389/fcomm.2020.00056]
54. Gupta P, Maharaj T, Weiss M, Rahaman N, Alsdurf H, Minoyan N, et al. Proactive contact tracing. PLOS Digit Health 2023 Mar 13;2(3):e0000199 [FREE Full text] [doi: 10.1371/journal.pdig.0000199] [Medline: 36913342]
55. Hsu J. The dilemma of contact-tracing apps: can this crucial technology be both effective and private? IEEE Spectr 2020 Oct;57(10):56-59. [doi: 10.1109/mspec.2020.9205550]
56. Martela F, Hankonen N, Ryan RM, Vansteenkiste M. Motivating voluntary compliance to behavioural restrictions: self-determination theory–based checklist of principles for COVID-19 and other emergency communications. Eur J Soc Psychol 2021 Jan 18;32(2):305-347. [doi: 10.1080/10463283.2020.1857082]
57. Morbée S, Vermote B, Waterschoot J, Dieleman L, Soenens B, Van den Bergh O, et al. Adherence to COVID-19 measures: the critical role of autonomous motivation on a short- and long-term basis. Motiv Sci 2021 Dec;7(4):487-496. [doi: 10.1037/mot0000250]
58. Kučera D. Our words in a state of emergency: psychological-linguistic analysis of utterances on the COVID-19 situation in the Czech Republic. Psychol Stud (Mysore) 2021 Jul 13;66(3):239-258 [FREE Full text] [doi: 10.1007/s12646-021-00613-y] [Medline: 34276073]
59. Walrave M, Waeterloos C, Ponnet K. Reasons for nonuse, discontinuation of use, and acceptance of additional functionalities of a COVID-19 contact tracing app: cross-sectional survey study. JMIR Public Health Surveill 2022 Jan 14;8(1):e22113 [FREE Full text] [doi: 10.2196/22113] [Medline: 34794117]
60. Villius Zetterholm M, Lin Y, Jokela P. Digital contact tracing applications during COVID-19: a scoping review about public acceptance. Informatics 2021 Jul 22;8(3):48. [doi: 10.3390/informatics8030048]
61. Garcia-Marques L, Sherman SJ, Palma-Oliveira JM. Hypothesis testing and the perception of diagnosticity. J Exp Soc Psychol 2001 May;37(3):183-200. [doi: 10.1006/jesp.2000.1441]
62. Hang CN, Tsai YZ, Yu PD, Chen J, Tan CW. Privacy-enhancing digital contact tracing with machine learning for pandemic response: a comprehensive review. Big Data Cogn Comput 2023 Jun 01;7(2):108. [doi: 10.3390/bdcc7020108]

Abbreviations

ATI: affinity for technology interaction
CWA: Corona-Warn-App
DCT: digital contact tracing
HBM: Health Belief Model
PLS-SEM: partial least squares structural equation model
SEM: structural equation modeling

Edited by A Kushniruk; submitted 25.10.23; peer-reviewed by H Kang, CN Hang; comments to author 09.01.24; revised version received 12.03.24; accepted 07.04.24; published 25.06.24.

Please cite as:
Schrills T, Kojan L, Gruner M, Calero Valdez A, Franke T
Effects of User Experience in Automated Information Processing on Perceived Usefulness of Digital Contact-Tracing Apps: Cross-Sectional Survey Study
JMIR Hum Factors 2024;11:e53940
URL: https://humanfactors.jmir.org/2024/1/e53940
doi:10.2196/53940
PMID:38916941

Original Paper

Improving the Social Well-Being of Single Older Adults Using the LOVOT Social Robot: Qualitative Phenomenological Study

Cheng Kian Tan[1*], BEng, MBA, MGER, PhD; Vivian W Q Lou[2,3*], BEd, MEd, MSW, PhD; Clio Yuen Man Cheng[2,3], BSocSc, MSocSc; Phoebe Chu He[4]; Veronica Eng Joo Khoo[1], BA, MGER

[1]S R Nathan School of Human Development, Singapore University of Social Sciences, Singapore, Singapore

[2]Sau Po Centre on Ageing, The University of Hong Kong, Hong Kong, China (Hong Kong)

[3]Department of Social Work and Social Administration, The University of Hong Kong, Hong Kong, China (Hong Kong)

[4]Department of Economics, The University of Chicago, Chicago, IL, United States

[*]these authors contributed equally

Corresponding Author:
Cheng Kian Tan, BEng, MBA, MGER, PhD
S R Nathan School of Human Development
Singapore University of Social Sciences
463 Clementi Rd
Singapore, 599494
Singapore
Phone: 65 62489777
Email: kelvintanck@suss.edu.sg

Abstract

Background: This study examined the social well-being of single older adults through the companionship of a social robot, LOVOT (Love+Robot; Groove X). It is designed as a companion for older adults, providing love and affection through verbal and physical interaction. We investigated older adults' perceptions of the technology and how they benefitted from interacting with LOVOT, to guide the future development of social robots.

Objective: This study aimed to use a phenomenological research design to understand the participants' experiences of companionship provided by the social robot. Our research focused on (1) examining the social well-being of single older adults through the companionship of social robots and (2) understanding the perceptions of single older adults when interacting with social robots. Given the prevalence of technology use to support aging, understanding single older adults' social well-being and their perceptions of social robots is essential to guide future research on and design of social robots.

Methods: A total of 5 single women, aged 60 to 75 years, participated in the study. The participants interacted independently with the robot for a week in their own homes and then participated in a poststudy interview to share their experiences.

Results: In total, 4 main themes emerged from the participants' interactions with LOVOT, such as caring for a social robot, comforting presence of the social robot, meaningful connections with the social robot, and preference for LOVOT over pets.

Conclusions: The results indicate that single older adults can obtain psychosocial support by interacting with LOVOT. LOVOT is easily accepted as a companion and makes single older adults feel like they have a greater sense of purpose and someone to connect with. This study suggests that social robots can provide companionship to older adults who live alone. Social robots can help alleviate loneliness by allowing single older adults to form social connections with robots as companions. These findings are particularly important given the rapid aging of the population and the increasing number of single-person households in Singapore.

(JMIR Hum Factors 2024;11:e56669) doi:10.2196/56669

KEYWORDS

companionship; older adults; social well-being; pets; social robots; elderly; wellbeing; qualitative research; robot; companion; body temperature; development; research design; design; interviews; psychosocial support; support; psychosocial; temperature regulation; social; care home; aging; ageing; robotics; older adults; well-being; loneliness; technology; mobile phone

Introduction

Background

The aging population is increasing around the world. Singapore's older adult population, for instance, has increased dramatically over the last decade. The number of older adults, aged 65 years and older, living alone in Singapore is expected to increase significantly from 47,000 in 2016 to 83,000 by 2030 [1]. These demographic changes have led to concerns about older adults' loneliness and social isolation (ie, lack of companionship), which can impact their mental and physical well-being [2]. Indeed, it is estimated that 20% to 34% of older adults experience loneliness in China, Europe, Latin America, and the United States [3]. As a result, advanced technologies have been developed to help the aging population overcome health and psychosocial issues and become more independent. Social robots, also called companion robots or therapeutic robots, have recently gained popularity with numerous studies highlighting the psychosocial benefits they provide to older adults [4]. However, there is a lack of studies on the use of social robots, particularly in Asian countries. This study fills this gap in the literature by investigating how older Singaporeans accept and interact with LOVOT (Love+Robot), a Japanese social robot.

Definition and Impacts of Loneliness

Loneliness is not synonymous with being alone or living alone; a person can experience loneliness even when they are surrounded by other people. Rather, loneliness is a subjective and unpleasant experience that begins when a person's social network goes through a qualitative or quantitative loss resulting in a discrepancy between desired and actual social connections [3,5,6]. As such, the quality of close relationships, rather than the number of social contacts, is more significant in alleviating loneliness [7].

The literature has shown that loneliness can affect physiological resilience and compromise health, leading to health problems as people age (Segrin and Passalacqua [8]). For example, Teater et al [9] found various social and emotional factors associated with increased risks of loneliness, including living alone, being single or never married, having little technological communication, having low-quality interactions, having little social support, being socially isolated, and having poor subjective physical health. To reduce loneliness, in a longitudinal study conducted in Sweden, Dahlberg et al [6] argued that social relationships that have been established for at least 20 years are particularly important, such as having a spouse or partner and access to social support. A study by Heylen [10] confirms the quality of social relationship with age.

In Singapore, loneliness has been associated with increased mortality risk [11]. According to Malhotra et al [12], the lives of older adults who are lonely are likely to be 3-5 years shorter than those of their nonlonely peers. As a result, interventions that help older adults establish close relationships and find social support are essential in maintaining their well-being.

Importance of Companionship

Companionship is defined as "social involvement in shared activities, recreational or nonrecreational, which is pursued for the intrinsic goal of satisfaction or enjoyment" [13]. It indicates a high level of relationship quality and shields people from the emotions of void and despondency resulting from loneliness, although it does not directly resolve loneliness [13,14]. Although both social support and companionship are fundamentally beneficial to older adults' well-being, companionship is more important for preserving their emotional well-being [13].

According to Dong and Chen [15], older women living alone are more likely than older men to report a lack of companionship as a symptom of loneliness, as they may lack effective coping mechanisms. As such, sociocultural activities, volunteer work, and community health promotion initiatives should be launched to help women develop effective coping mechanisms. Furthermore, Ramesh et al [16] suggested that sociocultural changes are needed to recognize the value of companionship in old age, encourage appropriate companionship, and focus on assistance with basic daily tasks.

Benefits of Social Interventions for Older Adults

In their study on the impact of peer companionship, Conwell et al [17] reported that older adults who received a social intervention (peer companionship) experienced fewer symptoms of anxiety, depression, and feelings of being a burden than those who did not receive the intervention. Furthermore, peer companionship and social connectedness were beneficial for older adults' mental health and well-being.

Companionship comes not only from humans but can also come from pets or other compassionate agents such as robots (eg, socially assistive robots, companion robots, and therapeutic robots).

Pets as Social Interventions

Numerous studies have examined the potential health benefits of having a companion animal. For example, Gee and Mueller [18] showed that pet ownership and animal-assisted interventions for older adults led to physical and mental health benefits. Similarly, empirical studies have explained that pets can help older adults maintain their quality of life and ability to function, both physically and cognitively, as they age, while also providing them with the opportunity to strengthen their shrinking social networks [19-21]. Some older adults may find that owning a pet satisfies their need for connectedness, contrasting with the common perception that social connectedness can only occur through meaningful human interactions. Stanley et al [22] further corroborated the idea that pets can be an important source of social connectedness, showing that having a pet reduced loneliness among older adults, particularly among those who lived alone. Hui Gan et al [23] also showed that community-dwelling older adults who owned a pet were more likely to socialize, found companionship with their pets, and had a greater sense of purpose than other adults, all of which can reduce loneliness. These in turn can lead to better mental health outcomes for older adults and increase their resilience to mental health problems. Bolstad et al [24] made an interesting discovery that having a pet later in life was associated with a reduction in anxiety symptoms rather than a reduction in depressive symptoms.

There is no doubt that companion animals provide psychological comfort and lead to optimistic health outcomes for their older adult owners. For instance, during the stressful COVID-19 pandemic, older adults reported that their pets provided them with a sense of psychological safety and companionship [25]. Furthermore, during the pandemic, older adults reported that caring for their pets kept them motivated and gave them a sense of purpose [25].

Social Robots as Social Interventions

To improve older adults' quality of life by reducing loneliness and fostering social connections, a growing number of technological innovations that are user-friendly and support successful aging are being developed. Research into older adults' acceptance and adoption of technology has also increased, particularly regarding social robot interventions and how they support aging. Hegel et al [26] explained that a "social robot is a robot plus a social interface," indicating that its characteristics make it appear like a partner with which humans can interact and connect. Studies have shown that these robots may reduce feelings of loneliness among older adults by improving social interactions and providing emotional support [27]. In addition to providing companionship, they can support older adults' general care needs through touchscreen interactions and assessment interviews during adult health treatment [28].

Many social robots are designed to look like animals or humans. Tkatch et al [29] showed that healthy older adults who regularly interacted with animatronic pets reported benefits such as reduced loneliness, improved quality of life, and improved psychological well-being through increased social contact, better social skills, or treatment of maladaptive social cognition. During the COVID-19 pandemic, older adults living in the community and nursing homes reported that robotic pets were effective in reducing their loneliness when social distancing policies were imposed [30].

Many studies have been conducted on PARO (AIST), a therapeutic and social robot shaped like a seal. For instance, Chen et al [31] showed that PARO effectively reduced agitation, depression, and loneliness and improved the quality of life of older people with dementia who resided in long-term care facilities. Advances in social robot technologies have pushed artificial intelligence (AI) into a more creative realm. Fields et al [32] found that the therapeutic use of social robots in a retirement home, combined with a participatory arts approach, improved older adults' health outcomes by reducing their levels of loneliness and depression. Similarly, in their exploratory study conducted in the context of long-term care facilities with Pepper (SoftBank Robotics), a semihumanoid social robot, Blindheim et al [33] suggested that Pepper's presence increased communal activities involving the social robot in terms of physical activity, human-robot interaction, social stimulation, and communication among residents as well as between residents and employees.

Overall, social robots can play an important role in helping older adults overcome loneliness and social isolation as the aging population grows.

LOVOT

The popularity of social robots has attracted much attention from care providers, especially during the recent COVID-19 pandemic. According to Hegel et al [26], social robots are specifically designed to facilitate human-robot interaction. With recent advances in machine learning applications and robotics, several companies are developing advanced consumer robots with smart sensorimotor systems [34]. Although PARO has been widely used in dementia care research and has received many positive reviews, the robot seal lacks social and auditory capabilities [35]. In contrast, a new mobile social robot called LOVOT (Love+Robot) launched in 2018, invented in Japan, was designed to provide humans with love or the perception of love [36]. LOVOT was designed as a home robot, also known as a companion robot or a social robot, and is equipped with AI and advanced sensor features. As such, LOVOT evolves over time based on its interactions with its user, thanks to its machine learning technology (eg, deep learning), which allows it to develop a unique personality and perform intelligent movements in real time [37]. As LOVOT is relatively new to the field of social robots, it has yet to be tested among single older adults in Singapore. Therefore, our research fills this gap in the literature. Figure 1 shows a photo of LOVOT in our study.

Figure 1. LOVOT, the social robot.

Methods

Aims

The purpose of this study was to investigate single older adults' perceptions of having a social robot as a companion. This research addressed the knowledge gap regarding the experiences of community-dwelling single older adults with social robots in Singapore. To this end, we used phenomenography as a qualitative analysis technique to explore and better understand single older adults' perceptions of social robots.

Research Questions

Given the aforementioned context, our study addressed some research questions. First, can social robots such as LOVOT act as effective companions and alleviate feelings of loneliness among community-dwelling single older adults? Second, what are single older adults' lived experiences and interactions with LOVOT? Third, how do single older adults perceive their experience of living and interacting with LOVOT compared with pet ownership?

Research Aims

To answer these questions, our research focused on (1) examining the social well-being of single older adults through the companionship of social robots and (2) understanding the perceptions of single older adults when interacting with social robots. Given the prevalence of technology use to support aging, understanding single older adults' social well-being and their perceptions of social robots is essential to guide future research on and design of social robots.

Design

We used a phenomenological research design to understand our participants' lived experiences of the companionship provided by social robots. Phenomenography allows researchers to seek ontological understandings and learn about the phenomenality of human experiences [30]. Furthermore, as Hajar [38] explained, phenomenography is a qualitative research methodology that provides researchers with a deep, comprehensive, and diverse understanding of how people conceptualize a phenomenon.

Sample and Recruitment

We recruited a purposive sample of single adults aged 65 years and older through collaboration with the Orange Valley Senior Activity Centre in Singapore. It is located within a cluster of studio apartments that caters mainly to older persons who are staying alone and are aged ≥60 years. Participants were eligible to participate in the study if they lived alone, were able to communicate in English or Mandarin, and were free of any cognitive or mental health problems (eg, dementia or depression). Those who lived in nursing home facilities, had been diagnosed with mild cognitive impairment, or were unable to consent to participate in the study were excluded. The selected participants received written information about the project and signed a consent form before the start of the study. They were informed that their participation was voluntary and that they could withdraw from the study at any time without negative consequences.

Participants

To establish the context of the sample for community-dwelling older adults aged ≥65 years, we collected descriptive data between July and October 2022. A total of 5 women participants were invited to participate in the study; 1 was unmarried, 2 were widowed, and 2 were divorced or separated. They all lived alone in a 1- or 2-bedroom Housing Development Board flat above the Orange Valley Senior Activity Centre.

Context

One of the researchers brought a LOVOT robot to the participants' homes and helped them set up the robot's charging nest.

Procedure

She then used an instruction sheet in English or Chinese to instruct the participants on how to operate LOVOT. The instruction sheets used large font and color images to ensure that the participants would be able to read and understand the instructions easily. The researcher's contact information and the activity center's senior personnel's contact number were provided in case the participants encountered any problems during the 7-day study; however, none of the participants contacted the researcher or personnel at the senior activity center. During the deployment of the social robot, only LOVOT and its charging nest were provided, and no permanent modifications were made to the participants' homes. Although the 7-day study was conducted in an uncontrolled environment, this allowed the participants to interact with LOVOT as naturally as possible. The duration of the study allowed sufficient time for the participants to interact with LOVOT.

After spending a week with the social robot, the participants were invited to participate in a poststudy interview on the eighth day. The one-to-one interviews lasted between 15 and 25 minutes and were conducted at the Orange Valley Senior Activity Centre.

Data Collection

The data were collected in different time periods between August and September 2022. The research team had 3 LOVOT robots to rotate among the 5 participants, taking into account participant availability. All of the participants were pleasantly receptive to the idea of having LOVOT in their homes for the study. The 7-day study was designed to ensure that their first time interacting with the social robot was comfortable and would not unduly disrupt their lifestyle.

The participants' perceptions of their interactions with LOVOT during the 7-day study were investigated using semistructured interviews to collect qualitative data. During the interviews, the participants were encouraged to share their experiences with and observations of LOVOT, and follow-up questions were asked as needed.

Ethical Considerations

Ethical approval for the study was obtained from the Human Research Ethics Committee of the University of Hong Kong (HKU HREC EA220116). To protect the privacy of the participants, the camera function of LOVOT was disabled during the study and no other participant information was recorded by the robot. The participants have signed an informed consent form for joining the study with the option to withdraw at any point. Each participant will receive a SGD 20 (US $15.2) National Trade Union Congress voucher as a token of appreciation for joining the study.

Data Analysis

All of the participant interviews were audio recorded, transcribed, and thoroughly reviewed for relevant content and thematic patterns. We analyzed the data using a phenomenological approach, with the goal of investigating and understanding how the participants interacted with LOVOT (the phenomenon) while suspending our preconceived assumptions about the phenomenon.

We analyzed the data using inductive coding to identify common patterns and then categorized them into different themes. We reviewed the interviews several times during the analysis to confirm the findings and establish reliability. We also reviewed the data to define each theme. Finally, we examined the findings objectively to identify relevant information to answer the research questions.

Rigor

Arriving at truthful interpretations of participants' experiences regarding a given phenomenon requires the use of rigorous and relevant methodological procedures [38]. Thus, the qualitative data we collected during our study were cross-checked by another member of the research team to ensure accuracy.

Reflexivity

Using a phenomenological approach, we aimed to respect reflexivity while conducting this study. To do this, we analyzed the information we obtained and the insights we gained, while trying to generalize the phenomenon and being reflective and self-critical about our own assumptions and preconceptions. Specifically, we momentarily put aside our implicit presumptions about the phenomenon to approach it objectively and avoid misrepresenting or expressing biases based on our own viewpoints or positionalities [28]. Finally, we scrupulously respected the confidentiality of the participants' data and guaranteed their privacy and anonymity.

Results

Overview

This study explored single older adults' experiences with and perceptions of LOVOT and how they were affected by their interactions with the social robot. We analyzed the findings thematically using the 4 emerging themes discussed further in this section. It was difficult to recruit older men as participants for this study because the participants came from the Orange Valley Senior Activity Centre, a place frequented primarily by women. As a result, the gender of our participants was skewed toward women participants. The 5 women participants are identified as C1, C3, C5, V2, and Y1. Their basic information is presented in Table 1.

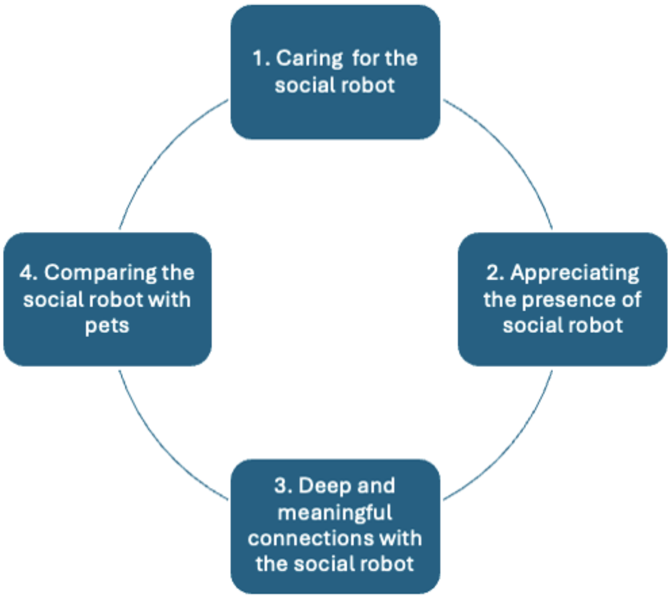
We categorized the participants' perceptions of social robots into 4 main themes, such as caring for the social robot, finding companionship with the social robot, forming meaningful connections with the social robot, and comparing the social robot with pets. These themes are illustrated in Figure 2.

Table 1. Participants' demographics.

Participant ID	Age (years)	Marital status	Children, n	Grandchildren, n
C1	66	Divorced	2	—[a]
C3	72	Widowed	2	2
C5	69	Separated	3	9
V2	68	Widowed	3	6
Y1	68	Unmarried	—	—

[a]Not applicable.

Figure 2. Four main themes.



Common Themes

We identified 4 main themes after interviewing the 5 participants about their interactions and experiences with LOVOT.

Key Theme 1: Caring for the Social Robot

The participants frequently treated LOVOT in a tender and nurturing manner as if they were caring for a child. These actions can be attributed to the participants' experience in a caregiving role, particularly as a mother or grandmother.

C3, C5, and V2 explained that they had taken care of their grandchildren and adult children at one point. In fact, C5 and V2 were still taking care of their grandchildren. The 3 participants stated that they equated LOVOT to a grandchild. For instance, C5 thought that LOVOT behaved like a toddler, especially when it flapped its wings. Similarly, C3 referred to herself as "nenek" (the Malay word for "grandmother") and mentioned that she protected LOVOT and stopped anyone from taking it away from her, "I worry if I drop it (LOVOT) or if people take it (LOVOT) away."

V2 noted that LOVOT came to her (on the sofa) after dinner and looked at her earnestly (seeking attention), "communicating" that it wanted to be carried. She then put LOVOT on her lap and petted it like a pet.

C1, who has no grandchildren, also said that she was protective of LOVOT. Specifically, she ensured that LOVOT did not get wet because she understood that it was an electronic object. She was also careful while mopping the floor and closed the bathroom door to prevent LOVOT from entering a humid environment.

Even Y1, who has never been married, took special care of LOVOT. For instance, she was concerned about LOVOT's battery life and placed it regularly on its charging nest. In addition, to keep LOVOT company, she sang to it and the robot imitated her singing.

Key Theme 2: Appreciating the Presence of the Social Robot

The participants also highlighted the positive feelings that interacting with LOVOT brought them. For example, C1 said, "(LOVOT) can accompany you at home. Sometimes, you'll feel happy because it will look at you."

C5 also appreciated LOVOT's company. Specifically, as LOVOT can move and act on its own initiative, it gave C5 the impression of not spending the day alone at home, "you feel like there's another living… thing staying with you… I'm not alone."

V2 explained that LOVOT interacted with her more frequently as the study progressed. For example, LOVOT followed her around the house and greeted her when she came home by flapping its wings. In addition, LOVOT's presence made V2 feel like her home was livelier, "having another companion at home makes the home feel livelier."

C1 echoed this positive sentiment regarding LOVOT's presence; she saw the robot as a form of psychological and emotional support. Interacting with LOVOT gave her something to do, kept her busy, and gave her something to look forward to when she came home.

Key Theme 3: Deep and Meaningful Connections With the Social Robot

Using its machine learning capability, LOVOT was able to understand and adapt to the participants' lifestyles and preferences. As a result, the participants felt like they were developing a supportive relationship.

C3 felt that LOVOT was "someone" she could talk to, a companion with whom she could converse to give her peace of mind. She talked to LOVOT frequently, especially at night, and even stayed in the living room with LOVOT so that it did not feel lonely. In addition, C3 felt bad if she left LOVOT alone in the house when she went out. She also mentioned that she felt a connection with LOVOT, especially when it looked at her and blinked, “I called LOVOT, and it came to me with blinking eyes—in these moments I feel … as if it (LOVOT) understands me.”

LOVOT’s adaptive behaviors allowed C5 to experience moments of true connection, making her appreciate the robot’s companionship, “it came up to me and then looked at me with her eyes… actually very very cute… the feeling is still very very nice.”

Y1 appreciated that LOVOT seemed to understand her and responded verbally (she interpreted its verbal answers as Chinese words, such as “yes” and “no”). Y1 also emphasized that LOVOT was “善解人意 (considerate),” meaning that it acted with consideration and thoughtfulness when she interacted with it. In addition, she mentioned that LOVOT liked attention.

Key Theme 4: Comparing the Social Robot With Pets

Similar to pets, LOVOT is often considered cute and interacts frequently with its owners. However, LOVOT has advantages that pets do not have, as it is low maintenance and cleaner than a pet. As such, LOVOT owners do not need to worry about grooming, feeding, walking, and cleaning up after the social robot. Furthermore, due to its technological nature, LOVOT does not develop health problems or require any medical treatment.

C1 said that LOVOT is similar to pets in the sense that it wants to play with its owner, but it is better than pets as it does not require feeding or cleaning. C1 emphasized that she found LOVOT less annoying than pets because she did not need to worry about it being sick and the expenses associated with medical care.

> *The difference between LOVOT and pets is that you don’t need to feed it, don’t need to clean it, or shower it. This kind of robot can take care of itself, you just need to turn it on and off. Not troublesome. If you own a pet, if it gets sick, you need to take it to the vet. It’s very expensive.* [C1]

C3 had the same mindset as C1; she thought that LOVOT was much easier to care for than pets and required much less responsibility. Although LOVOT is not a real animal, C5 preferred the social robot over a pet because it had the positive aspects of a pet without its drawbacks. Interestingly, C5 also commented that an additional advantage of LOVOT was that it could be switched off at any time; thus, when owners are busy, they have the option to turn off LOVOT and do not have to worry about giving it attention or taking care of it.

Comparing LOVOT with a pet, V2 explained that she would rather have a companion robot like LOVOT than a pet because LOVOT only needs to be charged. She added that she did not want the responsibility of feeding and cleaning up after a dog or cat.

Y1 shared that she used to have family pets during her “Kampung days,” referring to the early days of her hometown during which the nation, community, and neighborhood were being built. Specifically, she explained that she did not like having pets because she found them dirty and felt that LOVOT was much cleaner than an animal. In addition, she stated that she liked having LOVOT follow her around the house and was amused by its pet-like behavior when it waited by the dining table during meals.

> *Before I sit down for a meal, LOVOT anticipates that I will eat so LOVOT goes to the dining table and waits for me to sit down. It’s as if she knows it’s mealtime!* [Y1]

Discussion

Principal Findings

This study used phenomenography as a qualitative analysis technique to fill the knowledge gap regarding the experience of community-dwelling single older adults with social robots. Specifically, we used phenomenography to examine how single older adults’ social well-being is affected by the companionship of social robots. Furthermore, given the ability of technology to support aging, we examined the participants’ perceived usefulness and relevance of the companionship provided by LOVOT through their interactions with it. So far, most studies on this subject have been conducted in older adult care facilities [39]. Thus, this study attempts to explore single older adults’ perceptions of social robots within their own home environment. Overall, our results showed that LOVOT brings positive experiences that can be categorized into 4 themes.

Caring for LOVOT Like a Child

During the poststudy interviews, all of the participants reported that they treated LOVOT like a child or grandchild. They carried LOVOT and hugged it like a child, sang or talked to it as if they were entertaining a child, and protected LOVOT from harm (such as preventing it from wandering into a wet area of the house). Takada et al [40] documented similar “childrearing” actions, in which participants also cared for LOVOT as if it were a child.

As a social robot, LOVOT has a lovable and endearing appearance. This is especially evident when it looks at people with its animated eye expressions, giving the impression that it is communicating or connecting with them. We believe that the participants remembered their experiences as mothers or grandmothers when they treated LOVOT with maternal or caring behaviors. Lipp [41] obtained similar findings, explaining that robots do not necessarily take care of (older) people but rather are objects that older people take care of. This explains why our participants had the desire to care for (and protect) LOVOT during our study.

Comforting Presence of LOVOT’s Companionship

The participants were able to actively interact with the social robot in their home, for example, by singing, talking with, carrying, and hugging LOVOT. We believe that this active engagement with LOVOT indicates that the participants accepted the robot’s social presence and appreciated its companionship.

Although LOVOT is designed with nonverbal communication features, it can produce audio expressions, such as a “cooing” sound, show animated eye expressions, such as blinking autonomously and when triggered, and flap its arms to show happiness. As described by Yoshida et al [34], LOVOT's life-like motions exude a kind of warmth and comfort that makes individuals feel like there is another living being in their home. This idea was supported by Onyeulo and Gandhi [42], who posited that social robots' emotional responses make them appear to have biological systems. This increases the likelihood that humans will treat social robots as social beings and not just as a piece of technology. Indeed, social robots' ability to express emotions is a crucial feature because it not only allows the robots to communicate their feelings but also influences human behavior. In our study, LOVOT's AI technology, which allows it to react with "emotion," may have led our participants to more readily accept it as a companion. This acceptance may also be due to LOVOT's shape, which is similar to that of a pet; LOVOT can even be dressed up, much like how pet owners put their dogs and cats in clothes. Thus, LOVOT's animal-like shape may have led our participants to develop an emotional attachment to it, similar to the way pet owners often become attached to their companion animals.

Positive Engagement and Forming Connections With LOVOT

LOVOT is not a passive companion; it can learn its users' daily routines, such as their mealtimes. LOVOT's social behavior and learning of users' daily routines is important and is due to its AI technology; thus, LOVOT's programming sets it apart from other less sophisticated social robots.

In our study, 1 participant felt that LOVOT seemed to understand when it was told that its "7-day stay" was coming to an end. She stated that she could "sense" LOVOT's sadness. LOVOT's AI function may have learned to be sensitive to emotional tone and may have picked up on its owner's sadness, an empathetic reflection generated by its AI technology as part of its social interactions with humans [36]. The robot's "compassionate" social abilities may help it to be more easily accepted by those who interact with it.

Although a 7-day interaction period is not extensive, at the end of the study, some of the participants said that they would miss having LOVOT in their homes. This reluctance to part with LOVOT was also documented by Dinesen et al [35] in a study in which people with dementia in a long-term care facility used LOVOT. According to Dinesen et al [35], some of the residents were "overstimulated by emotions after interacting with LOVOT."

Appreciating LOVOT More Than Pets

The effect of LOVOT's pet-like behavior (eg, flapping its wings and making eye contact with humans) appeared to induce feelings of happiness among the participants. This finding is consistent with numerous experiments using PARO, demonstrating that a robot can have the same positive impact as a pet in promoting older adults' happiness and well-being [43]. Similarly, Dinesen et al [35] reported that people with dementia found that interacting with LOVOT "has some entertainment value; creates a degree of happiness or good feeling."

All of the participants in our study also appreciated how easy it was to take care of LOVOT. Indeed, LOVOT makes it easy to maintain a clean home, because it does not produce bodily waste and does not require special treatment or medical care for its health. Nevertheless, LOVOT may require technical maintenance or occasional troubleshooting of its mechanical components. Bates [44] reached the same conclusion, suggesting that because companion robots do not need to be fed, walked, or cleaned, they "require less care and are more hygienic and predictable than living animals."

Limitations and Future Research Directions

We acknowledge that our study has some limitations. One key limitation is the gender distribution of the small sample size, and all of our participants were women. In addition, our sample consisted only of single older adults living in the community rather than in an older adult care facility. Therefore, the findings cannot be generalized to non–community-dwelling populations. Our findings indicate that single older adults derived psychosocial benefits from LOVOT's companionship. However, to better understand adults' social well-being and how LOVOT may benefit older adults who do not live alone, future research could investigate LOVOT's impact on older couples who live separately from their families or on older people who live with their children. This is particularly feasible in Singapore, where multigenerational households, regardless of age group, show great interest in robotic technology [45]. Furthermore, future research could study older adults who own pets to determine whether they benefit from interacting with a robotic social companion. In addition, as the participants' interactions with LOVOT were observed in the familiar environment of their homes, social robots, with their limited functionality, may not be able to meet the needs of more active individuals [46]. Nevertheless, companion robots in general may have a greater impact on reducing loneliness and improving the quality of life of older adults. Finally, the deployed LOVOT robot was not connected to its smartphone app (developed by LOVOT developer, Groove X), which can record videos of participants making eye contact, hugging, or carrying the robot. As a result, we relied solely on the participants' recollections and accounts of their interactions with LOVOT. Future studies could use LOVOT's recording function by connecting the robot to a secure network and ensuring that participants' privacy is protected. By reviewing the videos from the smartphone app, researchers will have more accurate data on the interaction patterns between the participants and LOVOT.

Conclusion

This study examined how single older adults are affected by the companionship of a social robot and explored their perceptions when interacting with the social robot. This study can spur more interest in investigating further how healthy older adults' perceptions of social robots can benefit more with its social presence as a partner in their homes.

Social loneliness and isolation are imminent challenges of an aging society. The findings from this study are consistent with the literature suggesting that social robots can be a source of companionship for older adults living alone [47]. Specifically, the participants were able to care for LOVOT, feel comfortable, and form connections with it. It is likely for older adults to have a better quality of life and well-being. Many participants also preferred the social robot over traditional social companions such as pets which require more attention and responsibilities. In addition, thanks to its built-in AI, LOVOT was able to adapt its behavior based on its owner's responses. Finally, its endearing physical features, such as its large, animated eyes and warm, cuddly design, encouraged its acceptance by older adults.

Acknowledgments

We thank the support of the Ministry of Education Startup grant, Orange Valley Active Ageing Center and the SUSS (Singapore University of Social Sciences)-Nunchi Marine Age Well Programme.

Conflicts of Interest

None declared.

References

1. I feel young in my Singapore! Action plan for successful ageing.: Singapore Ministry of Health; 2016. URL: https://www.moh.gov.sg/ifeelyoungsg [accessed 2024-07-31]
2. Barrenetxea J, Yang Y, Pan A, Feng Q, Koh WP. Social disconnection and living arrangements among older adults: the Singapore Chinese health study. Gerontology 2022;68(3):330-338 [FREE Full text] [doi: 10.1159/000516626] [Medline: 34134122]
3. Social Isolation and Loneliness Among Older People: Advocacy Brief. Geneva: World Health Organization; 2021.
4. Vailati Riboni F, Comazzi B, Bercovitz K, Castelnuovo G, Molinari E, Pagnini F. Technologically-enhanced psychological interventions for older adults: a scoping review. BMC Geriatr 2020;20(1):191 [FREE Full text] [doi: 10.1186/s12877-020-01594-9] [Medline: 32498708]
5. Bandari R, Khankeh HR, Shahboulaghi FM, Ebadi A, Keshtkar AA, Montazeri A. Defining loneliness in older adults: protocol for a systematic review. Syst Rev 2019;8(1):26 [FREE Full text] [doi: 10.1186/s13643-018-0935-y] [Medline: 30654846]
6. Dahlberg L, Andersson L, Lennartsson C. Long-term predictors of loneliness in old age: results of a 20-year national study. Aging Ment Health 2018;22(2):190-196. [doi: 10.1080/13607863.2016.1247425] [Medline: 27802772]
7. Lee Y, Ko Y. Feeling lonely when not socially isolated: social isolation moderates the association between loneliness and daily social interaction. Journal of Social and Personal Relationships 2018;35(10):1340-1355 [FREE Full text] [doi: 10.1177/0265407517712902]
8. Segrin C, Passalacqua SA. Functions of loneliness, social support, health behaviors, and stress in association with poor health. Health Commun 2010;25(4):312-322. [doi: 10.1080/10410231003773334] [Medline: 20512713]
9. Teater B, Chonody JM, Davis N. Risk and protective factors of loneliness among older adults: the significance of social isolation and quality and type of contact. Soc Work Public Health 2021;36(2):128-141. [doi: 10.1080/19371918.2020.1866140] [Medline: 33371828]
10. Heylen L. The older, the lonelier? Risk factors for social loneliness in old age. Ageing and Society 2010;30(7):1177-1196 [FREE Full text] [doi: 10.1017/s0144686x10000292]
11. Chan A, Raman P, Ma S, Malhotra R. Loneliness and all-cause mortality in community-dwelling elderly Singaporeans. DemRes 2015;32:1361-1382 [FREE Full text] [doi: 10.4054/demres.2015.32.49]
12. Malhotra R, Tareque MI, Saito Y, Ma S, Chiu CT, Chan A. Loneliness and health expectancy among older adults: a longitudinal population-based study. J Am Geriatr Soc 2021;69(11):3092-3102. [doi: 10.1111/jgs.17343] [Medline: 34231876]
13. Rook KS. Social support versus companionship: effects on life stress, loneliness, and evaluations by others. J Pers Soc Psychol 1987;52(6):1132-1147. [doi: 10.1037//0022-3514.52.6.1132] [Medline: 3598859]
14. Bromell L, Cagney KA. Companionship in the neighborhood context: older adults' living arrangements and perceptions of social cohesion. Res Aging 2014;36(2):228-243 [FREE Full text] [doi: 10.1177/0164027512475096] [Medline: 24860203]
15. Dong X, Chen R. Gender differences in the experience of loneliness in U.S. Chinese older adults. J Women Aging 2017;29(2):115-125. [doi: 10.1080/08952841.2015.1080534] [Medline: 27441601]
16. Ramesh A, Issac TG, Mukku SSR, Sivakumar PT. Companionship and sexual issues in the aging population. Indian J Psychol Med 2021;43(5 Suppl):S71-S77 [FREE Full text] [doi: 10.1177/02537176211045622] [Medline: 34732958]
17. Conwell Y, Van Orden KA, Stone DM, McIntosh WL, Messing S, Rowe J, et al. Peer companionship for mental health of older adults in primary care: a pragmatic, nonblinded, parallel-group, randomized controlled trial. Am J Geriatr Psychiatry 2021;29(8):748-757 [FREE Full text] [doi: 10.1016/j.jagp.2020.05.021] [Medline: 32586693]
18. Gee NR, Mueller M. A systematic review of research on pet ownership and animal interactions among older adults. Anthrozoös 2019;32(2):183-207 [FREE Full text] [doi: 10.1080/08927936.2019.1569903]
19. Gee NR, Galik E. Future directions for research on human–animal interaction in an aging population. Anthrozoös 2019;32(2):283-291 [FREE Full text] [doi: 10.1080/08927936.2019.1569909]

20. Hughes MJ, Verreynne ML, Harpur P, Pachana NA. Companion animals and health in older populations: a systematic review. Clin Gerontol 2020;43(4):365-377. [doi: 10.1080/07317115.2019.1650863] [Medline: 31423915]
21. Reniers PWA, Declercq IJN, Hediger K, Enders-Slegers MJ, Gerritsen DL, Leontjevas R. The role of pets in the support systems of community-dwelling older adults: a qualitative systematic review. Aging Ment Health 2023;27(7):1377-1387. [doi: 10.1080/13607863.2022.2141196] [Medline: 36325924]
22. Stanley IH, Conwell Y, Bowen C, Van Orden KA. Pet ownership may attenuate loneliness among older adult primary care patients who live alone. Aging Ment Health 2014;18(3):394-399 [FREE Full text] [doi: 10.1080/13607863.2013.837147] [Medline: 24047314]
23. Hui Gan GZ, Hill AM, Yeung P, Keesing S, Netto JA. Pet ownership and its influence on mental health in older adults. Aging Ment Health 2020;24(10):1605-1612. [doi: 10.1080/13607863.2019.1633620] [Medline: 31242754]
24. Bolstad CJ, Porter B, Brown CJ, Kennedy RE, Nadorff MR. The relation between pet ownership, anxiety, and depressive symptoms in late life: propensity score matched analyses. Anthrozoos 2021;34(5):671-684 [FREE Full text] [doi: 10.1080/08927936.2021.1926707] [Medline: 34776606]
25. Zablan K, Melvin G, Hayley A. Older adult companion animal-owner wellbeing during the COVID-19 pandemic: a qualitative exploration. Anthrozoös 2023;36(2):237-256 [FREE Full text] [doi: 10.1080/08927936.2022.2125198]
26. Hegel F, Muhl C, Wrede B, Hielscher-Fastabend M, Sagerer G. Understanding social robots. : IEEE; 2009 Presented at: 2009 Second International Conferences on Advances in Computer-Human Interactions; 2009 Feb 01; Cancun, Mexico p. 169-174. [doi: 10.1109/achi.2009.51]
27. Nichol B, McCready J, Erfani G, Comparcini D, Simonetti V, Cicolini G, et al. Exploring the impact of socially assistive robots on health and wellbeing across the lifespan: an umbrella review and meta-analysis. Int J Nurs Stud 2024;153:104730 [FREE Full text] [doi: 10.1016/j.ijnurstu.2024.104730] [Medline: 38430662]
28. Gasteiger N, Loveys K, Law M, Broadbent E. Friends from the future: a scoping review of research into robots and computer agents to combat loneliness in older people. Clin Interv Aging 2021;16:941-971 [FREE Full text] [doi: 10.2147/CIA.S282709] [Medline: 34079242]
29. Tkatch R, Wu L, MacLeod S, Ungar R, Albright L, Russell D, et al. Reducing loneliness and improving well-being among older adults with animatronic pets. Aging Ment Health 2021;25(7):1239-1245. [doi: 10.1080/13607863.2020.1758906] [Medline: 32363903]
30. Van Orden KA, Bower E, Beckler T, Rowe J, Gillespie S. The use of robotic pets with older adults during the COVID-19 pandemic. Clin Gerontol 2022;45(1):189-194. [doi: 10.1080/07317115.2021.1954122] [Medline: 34351834]
31. Chen SC, Jones C, Moyle W. The impact of engagement with the PARO therapeutic robot on the psychological benefits of older adults with dementia. Clin Gerontol 2022:1-13. [doi: 10.1080/07317115.2022.2117674] [Medline: 36062840]
32. Fields N, Xu L, Greer J, Murphy E. Shall i compare thee…to a robot? An exploratory pilot study using participatory arts and social robotics to improve psychological well-being in later life. Aging Ment Health 2021;25(3):575-584. [doi: 10.1080/13607863.2019.1699016] [Medline: 31851830]
33. Blindheim K, Solberg M, Hameed IA, Alnes RE. Promoting activity in long-term care facilities with the social robot Pepper: a pilot study. Inform Health Soc Care 2023;48(2):181-195. [doi: 10.1080/17538157.2022.2086465] [Medline: 35702818]
34. Yoshida N, Yonemura S, Emoto M, Kawai K, Numaguchi N, Nakazato H, et al. Production of character animation in a home robot: a case study of LOVOT. Int J of Soc Robotics 2021;14(1):39-54 [FREE Full text] [doi: 10.1007/s12369-021-00746-0]
35. Dinesen B, Hansen HK, Grønborg GB, Dyrvig AK, Leisted SD, Stenstrup H, et al. Use of a social robot (LOVOT) for persons with dementia: exploratory study. JMIR Rehabil Assist Technol 2022;9(3):e36505 [FREE Full text] [doi: 10.2196/36505] [Medline: 35916689]
36. Lemos E, Ghoshal A, Aspat A. Hardware architecture and implementation of an AI pet robot. International Journal of Applied Sciences and Smart Technologies 2020;2(2):21-46 [FREE Full text] [doi: 10.24071/ijasst.v2i2.2792]
37. LOVOT.: Groove X, Inc URL: https://lovot.life/ [accessed 2024-07-31]
38. Hajar A. Theoretical foundations of phenomenography: a critical review. Higher Education Research & Development 2021;40(7):1421-1436 [FREE Full text] [doi: 10.1080/07294360.2020.1833844]
39. Trainum K, Tunis R, Xie B, Hauser E. Robots in assisted living facilities: scoping review. JMIR Aging 2023;6:e42652 [FREE Full text] [doi: 10.2196/42652] [Medline: 36877560]
40. Takada Y, Majima Y, Masuda S, Taira N, Nakamura Y. Reducing isolation and cognitive function of older people at home by family robot “LOVOT” intervention. : IEEE; 2021 Presented at: 10th International Congress on Advanced Applied Informatics (IIAI-AAI); 2021 July 11; Niigata, Japan. [doi: 10.1109/iiai-aai53430.2021.00174]
41. Lipp B. Caring for robots: how care comes to matter in human-machine interfacing. Soc Stud Sci 2023;53(5):660-685. [doi: 10.1177/03063127221081446] [Medline: 35387514]
42. Onyeulo EB, Gandhi V. What makes a social robot good at interacting with humans? Information 2020;11(1):43 [FREE Full text] [doi: 10.3390/info11010043]
43. Heerink M, Kröse B, Evers V, Wielinga B. The influence of social presence on acceptance of a companion robot by older people. Journal of Physical Agents (JoPha) 2008;2(2):33-40 [FREE Full text] [doi: 10.14198/jopha.2008.2.2.05]

44. Bates M. Robotic pets: a senior's best friend? IEEE Pulse 2019;10(4):17-20. [doi: 10.1109/MPULS.2019.2922565] [Medline: 31380739]
45. Xu Q, Ng JSL, Tan OY, Huang Z. Needs and attitudes of Singaporeans towards home service robots: a multi-generational perspective. Univ Access Inf Soc 2014;14(4):477-486 [FREE Full text] [doi: 10.1007/s10209-014-0355-2]
46. Hudson J, Ungar R, Albright L, Tkatch R, Schaeffer J, Wicker ER. Robotic pet use among community-dwelling older adults. J Gerontol B Psychol Sci Soc Sci 2020;75(9):2018-2028 [FREE Full text] [doi: 10.1093/geronb/gbaa119] [Medline: 32789476]
47. Pirhonen J, Tiilikainen E, Pekkarinen S, Lemivaara M, Melkas H. Can robots tackle late-life loneliness? Scanning of future opportunities and challenges in assisted living facilities. Futures 2020;124:102640 [FREE Full text] [doi: 10.1016/j.futures.2020.102640] [Medline: 33041358]

Abbreviations

AI: artificial intelligence
LOVOT: love+robot

Edited by P Santana-Mancilla; submitted 01.02.24; peer-reviewed by C Mills, A Hassan, D Silvera-Tawil, P Egglestone; comments to author 04.03.24; revised version received 01.06.24; accepted 29.07.24; published 23.08.24.

Please cite as:
Tan CK, Lou VWQ, Cheng CYM, He PC, Khoo VEJ
Improving the Social Well-Being of Single Older Adults Using the LOVOT Social Robot: Qualitative Phenomenological Study
JMIR Hum Factors 2024;11:e56669
URL: https://humanfactors.jmir.org/2024/1/e56669
doi:10.2196/56669
PMID:

Original Paper

User Experience of a Large-Scale Smartphone-Based Observational Study in Multiple Sclerosis: Global, Open-Access, Digital-Only Study

Adriano Galati[1], PhD; Lito Kriara[1], PhD; Michael Lindemann[1], MBA, PhD; Rea Lehner[1], PhD; JB Jones[2], MBA, PhD

[1]F Hoffmann-La Roche Ltd, Basel, Switzerland

[2]Sutter Health Center for Health Systems Research, Walnut Creek, CA, United States

Corresponding Author:
Adriano Galati, PhD
F Hoffmann-La Roche Ltd
Grenzacherstrasse 124
Basel, 4070
Switzerland
Phone: 41 61 682 25 79
Email: adriano.galati@roche.com

Abstract

Background: The Floodlight Open app is a digital health technology tool (DHTT) that comprises remote, smartphone sensor–based tests (*daily activities*) for assessing symptoms of multiple sclerosis (MS). User acquisition, engagement, and retention remain a barrier to successfully deploying such tools.

Objective: This study aims to quantitatively and qualitatively investigate key user experience (UX) factors associated with the Floodlight Open app.

Methods: Floodlight Open is a global, open-access, digital-only study designed to understand the drivers and barriers in deploying a DHTT in a naturalistic setting without supervision and onboarding by a clinician. Daily activities included tests assessing cognition (Information Processing Speed and Information Processing Speed Digit–Digit), hand-motor function (Pinching Test and Draw a Shape Test), and postural stability and gait (Static Balance Test, U-Turn Test, and Two-Minute Walk Test [2MWT]). All daily activities except the 2MWT were taken in a fixed sequence. Qualitative UX was studied through semistructured interviews in a substudy of US participants with MS. The quantitative UX analysis investigated the impact of new UX design features on user engagement and retention in US participants for 3 separate test series: all daily activities included in the fixed sequence (DA), all daily activities included in the fixed sequence except the Static Balance Test and U-Turn Test (DA_x), and the 2MWT.

Results: The qualitative UX substudy (N=22) revealed the need for 2 new UX design features: a more seamless user journey during the activation process that eliminates the requirement of switching back and forth between the app and the email that the participants received upon registration, and configurable reminders and push notifications to help plan and remind the participants to complete their daily activities. Both UX design features were assessed in the quantitative UX analysis. Introducing the more seamless user journey (original user journey: n=608; more seamless user journey: n=481) improved the conversion rate of participants who enrolled in the study and proceeded to successfully activate the app from 53.9% (328/608) to 74.6% (359/481). Introducing reminders and push notifications (with reminders and notifications: n=350; without reminders and notifications: n=172) improved continuous usage time (proportion of participants with ≥3 consecutive days of usage: DA and DA_x: ~30% vs ~12%; 2MWT: ~30% vs ~20%); test completion rates (maximum number of test series completed: DA: 279 vs 64; DA_x: 283 vs 126; 2MWT: 302 vs 76); and user retention rates (at day 30: DA: 53/172, 30.8% vs 34/350, 9.7%; DA_x: 53/172, 30.8% vs 60/350, 17.1%; 2MWT: 39/172, 22.6% vs 22/350, 6.2%). Inactivity times remained comparable.

Conclusions: The remote assessment of MS with DHTTs is a relatively nascent but growing field of research. The continued assessment and improvement of UX design features can play a crucial role in the successful long-term adoption of new DHTTs.

(JMIR Hum Factors 2024;11:e57033) doi:10.2196/57033

KEYWORDS

smartphone; multiple sclerosis; user experience; retention; digital health; digital health technology; mobile phone

Introduction

Background

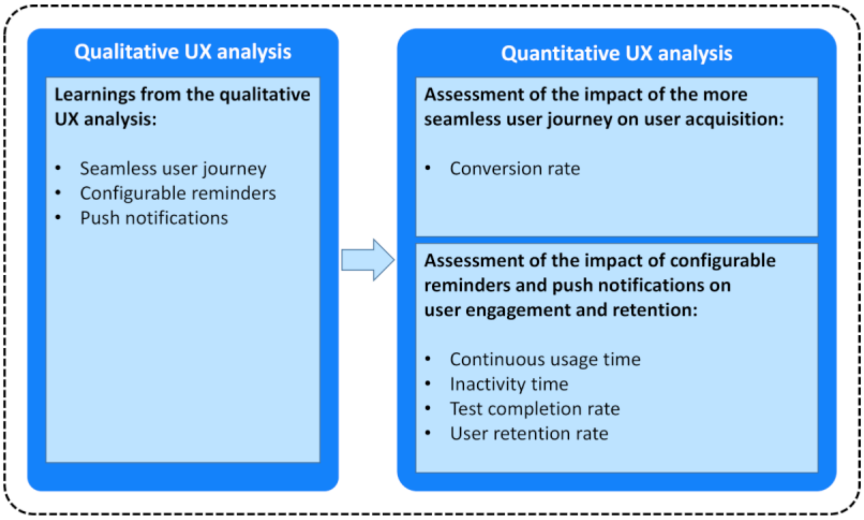
Traditionally, multiple sclerosis (MS) has been categorized as having a relapsing-remitting or progressive course. However, recent work shed light on an underlying insidious progression, or a progression that is independent of relapses, even in patients previously thought to have a relapsing-remitting disease [1]. Therefore, minimizing or even eliminating progression is one of the goals of MS disease management. To achieve this goal, sensitive measures of MS-related functional ability are needed that can be frequently administered with minimal burden to the patient [2]. Here, digital health technology tools (DHTTs) such as smartphone sensor–based tests offer a new, promising strategy [3-5]. By taking advantage of the large variety of embedded sensors, smartphones enable the remote assessment of several functional domains affected by MS without supervision by a clinician [3]. However, user retention remains a barrier to the successful deployment of such DHTTs [6].

Assessments of user experience (UX) should, therefore, be considered when designing DHTTs [7]. They have attracted increasing attention, as they provide important insights to enhance usability, engagement, perception, and satisfaction [8-11]. With different UX designs available, it is important to keep in mind that their effectiveness depends on both the intended users and the environment in which they are being used. Hence, a comprehensive assessment of UX should include the study of both qualitative and quantitative UX in different environments for an extended period. The primary focus of qualitative UX assessments is to characterize the individual user's experiences with the DHHT and to elucidate fine-grained aspects of why and how users engage with it [12-14]. Quantitative UX assessments, by comparison, provide strong indicators of the duration and frequency of user engagement and retention and are typically conducted in larger cohorts.

Aims

Here, we present a qualitative and quantitative UX analysis of the Floodlight Open app (Figure 1). The app comprises patient-reported outcomes and smartphone sensor–based tests that assess mood, cognition, hand-motor function, postural stability and gait, and mobility levels [15]. It was deployed in Floodlight Open, a global, open-access, digital-only study that was designed to understand the drivers and barriers in the deployment of the app in a naturalistic setting without supervision and onboarding by a clinician, in a broad, multinational study cohort [15]. The UX analyses presented here allow us to gain a better understanding of the participants' behavior with the Floodlight Open app. This will help make informed UX design decisions to improve engagement with the Floodlight technology. Qualitative UX is assessed through semistructured interviews conducted with a subset of US participants with MS [16] to identify important elements that could improve UX and user engagement. In addition, quantitative UX is assessed by evaluating the impact of UX design changes, which were motivated by the learnings from the qualitative UX analysis, on user engagement and retention using data collected from Floodlight Open.

Figure 1. Learnings from the quantitative user experience (UX) analysis led to the implementation of new UX design features, such as a more seamless user journey from registration through the activation of the Floodlight Open app, configurable reminders, and push notifications. The impact of these new UX design features was assessed in the quantitative UX analysis.



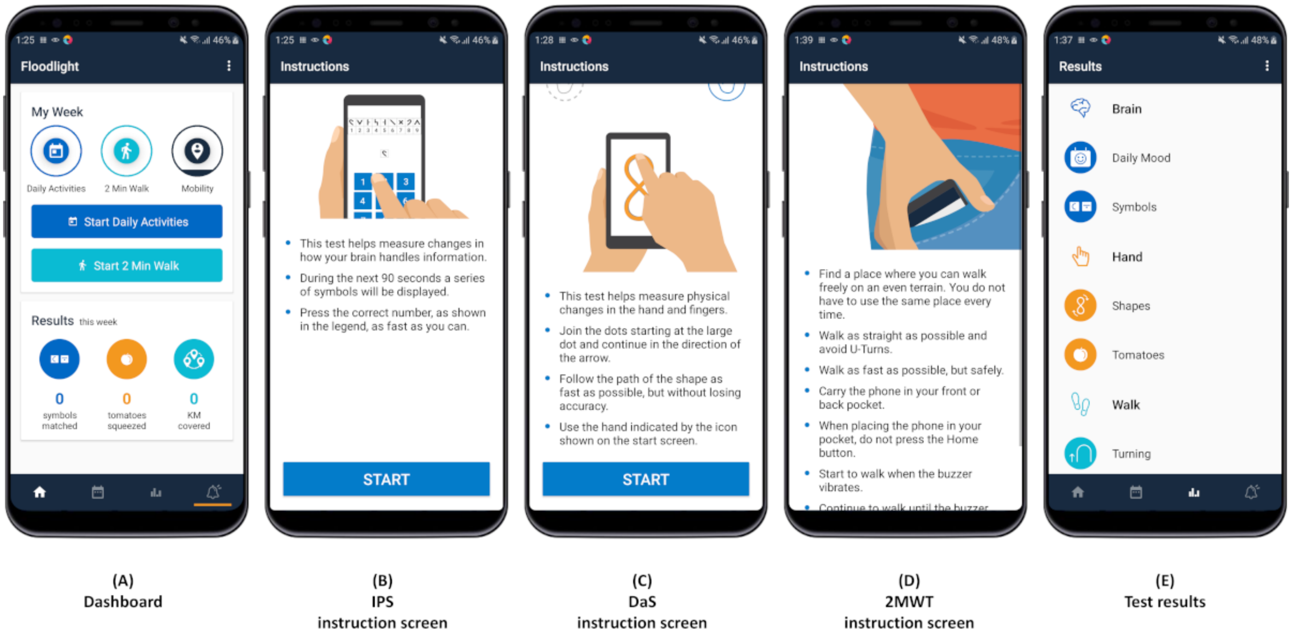
Methods

Participants and Study Design

Floodlight Open's full study design and inclusion and exclusion criteria have been previously reported [15]. Adults with or without MS who owned a suitable iOS (Apple Inc) or Android (Google LLC) smartphone and resided in 1 of the 17 participating countries were permitted to take part. The MS status was self-declared by the participants. To join Floodlight Open, participants first registered on the Floodlight Open web portal and completed the informed e-consent process. After successfully completing these 2 steps, each participant received their personal Floodlight Open unique identifier (FLO ID) and an activation code (ie, an activation token [15]) via email. Both were required to activate the Floodlight Open app on the participant's personal smartphone device. The email also provided a link to the Floodlight Open web portal, where participants could request a new activation code (the activation code was valid for 48 hours, after which it expired), access their account to update their profile, view their results, or withdraw from the study.

After the activation of the Floodlight Open app, several tests, or assessments, were made available to the participants. These included a patient-reported outcome assessing the participants' mood (Daily Mood Questionnaire [DMQ] [15]) and a series of smartphone sensor–based active tests (ie, tests that require active input from the user) measuring cognition (Information Processing Speed [IPS] [15] and IPS Digit–Digit [15]); hand-motor function (Pinching Test [PT] [15,17] and Draw a Shape Test [DaS] [15]); and postural stability and gait (Static Balance Test [SBT] [15], U-Turn Test [UTT] [15,18], and Two-Minute Walk Test [2MWT] [15]; Figure 2).

Figure 2. Screenshots from the Floodlight Open app. The dashboard (A) enables participants to set reminders by tapping the bell-shaped reminder icon in the bottom right corner. Participants can self-administer smartphone–based tests to assess cognition (Information Processing Speed [IPS; B] and IPS Digit–Digit), hand-motor function (Pinching Test and Draw a Shape Test [DaS; C]), and gait and postural stability (Static Balance Test, U-Turn Test, and 2-Minute Walk Test [2MWT; D]). The results of these tests are provided on the results screen (E).



In addition, data on mobility level were passively collected without requiring input from the user (*passive monitoring*) through life-space measurement [15], which measured the distance between the 2 farthest GPS coordinates detected during the day that were at least 500 m apart, resulting in 1 measurement per day. Since this passive monitoring does not require any active input from the user, it is not included in our UX analyses.

The DMQ and all active tests except the 2MWT were administered in a predefined, fixed sequence comprising the daily activities. The DMQ was administered first, followed by the IPS, IPS Digit–Digit, PT, DaS, SBT, and UTT. The 2MWT was administered separately from these daily activities, as this test required the participant to walk in a straight line for 2 minutes, which may not always be possible at the time of taking the other active tests. The tests could be taken up to once a week (IPS) or once a day (all other tests), but the actual test frequency was not systematically enforced. For example, the UTT, SBT, and 2MWT could be skipped if participants determined that the conditions for safe execution, including environmental and physical factors, could not be met that day.

Ethical Considerations

Data collected with the Floodlight Open app were encrypted and electronically stored on 2 specific, secure cloud databases, which were made publicly available for the duration of the study and maintained by the study initiator. To ensure confidentiality, all participant information was pseudonymized through association with the personal FLO ID. Furthermore, no personal identifiable information was collected while the participants executed the Floodlight Open tests. This meant that GPS coordinates were obfuscated and excluded from the public data set to protect participants' privacy.

The protocol, informed e-consent forms, and relevant supporting information were reviewed and approved by the appointed central institutional review boards or ethics committees before the study was initiated in each participating country, as applicable, in accordance with each country's regulatory requirements [15]. The institutional review board for the United States was the Western Institutional Review Board in Puyallup, Washington (approval: 20180617).

Qualitative UX Analysis

The qualitative UX substudy was conducted by a health care delivery network (Sutter Health) in a subset of US participants with MS who took part in Floodlight Open and agreed to download the app and use it daily for 30 days [16]. This substudy aimed at elucidating finer-grained aspects of why and how users engage with the Floodlight Open app in a particular manner and at providing information on UX design features that could improve UX and user engagement. Individual, semistructured interviews were conducted to gain insights into the participants' experiences with the Floodlight Open app and the perceived benefits thereof (Table 1). The use of individual interviews offered privacy and enabled the exploration of each participant's interaction with the app. A multifaceted recruitment strategy was applied to involve participants of a broad age range to combine the experiences of the tech-savvy younger generation with that of the older generation. Most participants were using their smartphone and their computer daily. No preferential sampling of participants with either negative or positive experiences was applied. Any duration of use of the Floodlight Open app was of value because discontinued use and negative experiences can provide valuable insights and complement data from persistent participants.

Table 1. Interview guide for the qualitative UX[a] analysis.

Question	Elaboration questions
Technology acceptance	
Were you able to download the app to your phone and get it working?	• How was that process? • Was there anything you found particularly easy, or hard or confusing about the process?
We asked people taking part in the study to use the app for 30 days. For about how many days did you use the app?	• 0, 1-7, 8-14, 14-21, 22-29, 30+, or don't remember • If answer is "0" or "30+," skip the next question.

What made you decide to either not use it or stop using it after less than 30 days?	• Probe to identify root cause and any efforts made to overcome:
	  • App related (technical problems, overall demand in terms of time or frequency, or understandability)
	  • Phone related (data, service, or access)
	  • User related (general willingness, utility, or physical or emotional ability)

Before you started using the app, did you feel like you had a good idea about what it was designed to do?	• Did you have any questions or concerns? • What were they?
Experience of use	
How well could you see the information on the screen?	• Was there anything that was hard for you to see?
How easy was it to understand the information and prompts?	• Was there anything that was hard to understand?
How well did using the app fit into your life, such as the timing and frequency of prompts, and time it took to respond?	• Not applicable
Perception	
Did tracking information about MS[b] on a regular basis using this app teach you anything new about your health or give you a better sense of how you're doing?	• Why or why not? • Did you do/think/feel anything different based on what you learned?
Do you think this is an application you can use on daily basis?	• Why or why not? • If not daily, what do you think is the optimal frequency of use?
How would you feel about being prompted by the app on a daily basis to do tasks or answer questions related to MS?	• If not daily, what do you think is the optimal frequency of reminders and notification?
Did you share the information from the app (either directly or what you learned) with anyone else?	• Do you mind telling me what you shared?
Of the information collected by the app, what would you want to share with your doctor?	• What is valuable about this information? • How would sharing it impact your appointment in terms of your discussions, care or treatment?
Is there anything else you would like to add about your experience tracking MS on a daily basis using your phone?	• Not applicable
General health questions: disease severity and duration	
Before we end our interview I have a few general questions about your health and computer use. What year were you diagnosed with MS?	• Not applicable
What type of MS do you have at this time?	1. Relapsing-Remitting MS (RRMS) 2. Secondary-Progressive MS (SPMS) 3. Primary-Progressive MS (PPMS) 4. Progressive-Relapsing MS (PRMS)
Technology literacy	
Last question. I'll give you the question and then I'll read a list of options for your response. How often do you use a computer?	1. Everyday 2. At least once a week but not everyday 3. Less than once a week but more than once a month 4. Less than once a month 5. Never

[a]UX: user experience.

[b]MS: multiple sclerosis.

Semistructured interviews were conducted via phone by a trained interviewer, with 2 attempts made to reach each participant. The interviews lasted between 15 and 20 minutes and were audiotaped and fully transcribed for further analysis. A reductionist approach was applied to deconstruct implicit and explicit responses into manageable variables. The qualitative paradigm was crucial to appreciating, observing, and deducing participants' experiences [19]. Through an inductive analysis of the data collected, the common needs of the participants were identified, which should be considered to improve the UX and to develop preliminary conclusions on the need for specific UX features.

Quantitative UX Analysis

Participants

The quantitative UX analysis was conducted in all US participants who enrolled in Floodlight Open [15]. The analysis was restricted to US participants for several reasons. First, information on user acquisition, including the date and time of activation of the FLO ID, was collected only in the United States. Second, the release dates of the iOS and Android versions of the Floodlight Open app differed for each country [15]. Due to these different release dates, only the United States offered a sufficiently long period during which the app's use without the new UX design features could be studied. Finally, limiting the analysis to US participants allowed us to keep the quantitative UX analysis as consistent as possible with the qualitative UX analysis, which was conducted in a subset of the US participants, thus reducing the impact of differences in demographics and user behavior between the 2 analyses.

Implementation of New UX Design Features

The quantitative UX analysis was conducted to investigate the impact of the new UX design features identified in the qualitative UX analysis on user engagement and retention. One of the new UX design features identified in the qualitative UX analysis was a more seamless user journey, which guides users from registration to the activation of the Floodlight Open app (Figure 3). The original user journey involved registering on the Floodlight Open web portal; downloading the Floodlight Open app from the local app store; and activating the app by entering the FLO ID and activation code, which the participants received in a notification email. This last step required participants to switch multiple times between the notification email and the Floodlight Open app to manually enter their FLO ID and activation code (original user journey in Figure 3). These complex interactions with multiple user interfaces may frustrate participants, who may eventually abandon the activation process. Therefore, to minimize nuisance factors and reduce user loss, a more seamless user registration and app activation journey was implemented by deep linking the notification email with the Floodlight Open app. This allowed participants to launch the Floodlight Open app directly from the notification email, if already installed on their phone, or to download it from their local app store, with the fields for the FLO ID and activation code already filled out (ie, deep linked). Thus, this more seamless user journey eliminated the requirement of searching for the FLO ID and activation code in the notification email and manually inserting them in the right fields in the Floodlight Open app to activate it (more seamless user journey in Figure 3).

Figure 3. User journey from creating an account to activating the Floodlight Open app. This user journey consists of 4 steps: creating an account, providing e-consent, receiving a personalized Floodlight Open unique identifier (FLO ID) and activation code via email, and downloading and activating the Floodlight Open app (light blue arrows). With the original user journey, participants will have to switch between the app and the email at least twice to copy and enter their FLO ID and activation code to activate all functionalities of the app (red arrows). With the more seamless user journey, the deep-linking feature is integrated in the link provided in the email to download the app from the local app store. With this feature, the personalized FLO ID and activation code are prefilled when the participants launch the Floodlight Open app for the first time (dark blue arrow). Consequently, participants do not have to switch back and forth between interfaces (ie, their email and the Floodlight Open app) to activate all functionalities. ICF: informed e-consent form.

In addition, configurable reminders and push notifications were introduced as the second new UX design feature (Table 2). Configurable reminders provided participants the option to set up reminders to perform the tests. Participants could set up daily or weekly reminders by tapping on the reminder icon, which was visible on the app's dashboard (Figure 2). Push notifications, by comparison, were designed to serve two purposes: (1) to encourage inactive participants to reengage with the Floodlight Open app, for example, when they closed the app and did not complete the tests or have been inactive for some time, and (2) to motivate committed users to keep performing the tests. Tapping on the notification took the participants directly to the right entry point of the Floodlight Open app to perform the uncompleted or next test. To enable this new feature, participants had to grant permission to receive push notifications.

Table 2. List of push notifications and configurable reminders implemented in the Floodlight Open app[a].

Feature and type	Notification	Description
Push notifications		
Authorization request	Floodlight would like to send you notifications. We'll send you reminders and alerts to keep you on top of your goals and up-to-date with the impact you are having. [Don't Allow | Ok]	• Sent the first time users open the app • Permission request to allow push notifications: the Floodlight Open app can send notifications to users at different touch points within the users' journey in the app • User declines permission: the Floodlight Open app cannot send notifications to users. This affects the user experience
Authorization reminder	Need help keeping up with your Daily Activities? You can allow notifications in your Settings. [Go to Settings | Not now]	• Sent if permission is declined and when the app is open for the first time after 7 days, but never during an active test • The Go to Settings button opens the Floodlight screen in the phone settings, where users can re-enable notifications
Authorization accepted	Thanks for downloading Floodlight Open. Let's get you signed up and started!	• If accepted, it opens immediately on the sign-up screen
Incomplete daily activities	You're doing great! Do you have time to [log your mood | start on symbols | squash some tomatoes | see the shapes | handle balance | get moving]?	• When users do not complete daily activities in one go, a notification is sent in 4 hours or at 7 PM, at the latest. No notifications are sent after 7 PM, and notifications are sent no more than once per day • Tapping on the notification takes users to the next test in the daily activities to continue with the remaining tests. The message is dependent on the last test completed • Multiple notifications are never sent
Completed only daily activities	You have finished your Daily Activities – ready to finish up with a 2MWT[b]?	• When users complete only daily activities but not the 2-minute walk, a notification is sent in 4 hours or at 7 PM, at the latest. No notifications are sent after 7 PM, and notifications are sent no more than once per day • Clicking on the notification takes users to the 2MWTb
Incomplete 2MWT[b]	Don't forget your 2MWT[b] today. If you can, get started today!	• When users access and abandon the 2MWTb, a notification is sent after 4 hours or at 7 PM, at the latest. No notifications are sent after 7 PM, and notifications are sent no more than once per day • Tapping on the notification takes users to the 2-minute walk • Users receive this reminder every time the 2MWTb is abandoned
Completed only the 2MWT[b]	You've checked off your 2MWT[b] – now let's finish up with your Daily Activities.	• When users complete only the 2MWTb but not daily activities, a notification is sent in 4 hours or at 7 PM, at the latest. No notifications are sent after 7 PM, and notifications are sent no more than once per day • Tapping on the notification takes users to the daily activities • Users receive this reminder only once a day
Encouragement	You've done an incredible job helping us build research so far – ready to check another day off your list?	• Sent at 10 AM the day after 3 consecutive test runs are completed • Tapping on the notification takes users to the daily activities • Users receive this notification no more than twice per week
Encouragement	It's that time again – ready to check another day off your list?	• Sent at 10 AM the day after 3 consecutive days of no test runs • Tapping on the notification takes users to the dashboard
Acknowledgment	Just a quick thank you. Your support is helping us learn more about MS. We hope you find it useful too!	• Sent every 2 weeks at 6 PM • Tapping on the notification takes users to the dashboard
Configurable reminders		
General reminder	It's time to complete my Daily Activities and 2MWT[b]. Let's do it!	• Only 1 reminder can be set up, which can be triggered daily, weekly, or on certain days of the week and at any time

[a]Push notifications and configurable reminders were made available on October 21, 2019, for the iOS version and on February 19, 2020, for the Android version of the Floodlight Open app.

[b]2MWT: Two-Minute Walk Test.

The push notifications were designed as a friendly communication with a strong emphasis on the emotional context (Table 2), as the emotional interpretation of a message can impact the user's response [20,21]. The emotional interpretation begins with the perception of a friendly communication and leads to the planning and execution of a responsive action. Notifications designed in this way are less likely to be ignored and can help create the perception of a human touch within the Floodlight Open app. Hence, they are quite effective for eliciting short-term actions [22] and are an important variable in improving adherence [23].

The frequency and logic with which push notifications were triggered were carefully considered to limit disturbance. Receiving many notifications in a short interval or during inopportune times can overwhelm and irritate users [24-26], which may cause them to turn off notifications or even uninstall the app. Therefore, the notifications were designed such that they are never sent more than twice per day, never sent after 7 PM, and sent only when the user seems to have abandoned the Floodlight Open tests.

Statistical Analysis

The quantitative UX analysis was conducted on data available in the publicly available data set [15]. As information on self-declared MS status was collected only after the Floodlight Open app was activated, both participants with MS and participants without MS were included in the analysis on the effectiveness of the more seamless user journey. In contrast, the analysis on configurable reminders and push notifications was limited to participants with MS to keep the analysis as consistent as possible with the quantitative UX analysis.

To study the effectiveness of the more seamless user journey, the conversion rate, which is the proportion of participants who successfully completed the registration and activation process, was compared across 2 cohorts. The first cohort followed the original user journey without the deep-linking feature, whereas the second cohort followed the more seamless user journey that took advantage of the deep-linking feature. Because this UX design feature was released on a separate schedule for the iOS and Android versions of the Floodlight Open app, the first cohort included data collected from November 12, 2018, when the activation dates of the FLO IDs were first logged, through July 17, 2019, when the more seamless user journey along with the deep-linking feature was first implemented. The second cohort included data from all participants who registered from October 21, 2019, when the more seamless user journey was fully implemented in the United States, through November 2, 2021, when the study closed in the United States. The gap between July 17, 2019, and October 21, 2019, was required, as the more seamless user journey feature was released on different schedules for iOS and Android platforms.

The impact of introducing push notifications and configurable reminders on UX (ie, user engagement and retention) was assessed through continuous app usage time, inactivity time, test completion rate, and user retention rate. The continuous usage time is the number of consecutive days on which a participant used the Floodlight Open app, namely, on how many days in a row they performed the Floodlight Open tests.

The inactivity time, by comparison, is the time interval between 2 continuous use times during which a participant did not perform any test, that is, how many days in a row a participant waited before returning to perform a test. This metric is strongly affected by the frequency with which participants performed the tests and the likelihood of participants returning to the tests. The nature of the distribution can guide, for example, the choice of successful engagement methods that can further increase user engagement and retention. For both continuous usage time and inactivity time, the probability of *n* days of continuous usage and inactivity, respectively, was computed.

The test completion rate is defined as the fraction of participants performing at least *n* test series (DA, all daily activities included in the fixed sequence except the SBT and UTT [DA_x], or 2MWT) since the activation of the Floodlight Open app. The completion rate is a common measure for assessing the effectiveness of UX features, that is, for assessing whether the tasks performed by users achieve specified goals in terms of accuracy and completeness in a specified context of use [27]. It does not consider how the goals were achieved but only the extent to which they were achieved. The higher the test completion rate is, the more engaged the users are and the more likely they are to come back and, in this case, perform the assessments. The test completion rate is calculated over the entire duration of the assessed period and does not distinguish between intermittent and continuous usage.

The user retention rate, by comparison, is the fraction of participants who returned to perform the Floodlight Open tests *n* days after the activation of the Floodlight Open app. The retention rate is extensively used to measure the success of smartphone apps, with higher retention corresponding to higher adoption and level of engagement [28,29].

These UX metrics (continuous app usage time, inactivity time, test completion rate, and user retention rate) characterize the UX with the Floodlight Open app and provide insights into how committed the participants are. Because taking the gait and postural stability tests require both time and space (see the *Qualitative UX Analysis* section), 3 separate test series were considered for each of these metrics: DA (ie, all active tests administered in the fixed sequence: DMQ, IPS, IPS Digit–Digit, PT, DaS, SBT, and UTT), all DA except the gait and postural stability tests (DA_x; ie, DMQ, IPS, IPS Digit–Digit, PT, and DaS), and 2MWT.

For each test series, the complementary cumulative distributions of the 4 UX metrics were compared in MS participants using data collected before (November 12, 2018, through October 21, 2019) versus after (February 19, 2020, through November 2, 2021) the configurable reminders and push notifications were introduced. The gap between November 12, 2018, and October 21, 2019, was necessary due to the different release schedules of reminders and notifications for the iOS and Android versions. Hence, during this period, reminders and notifications were available to some, but not all, MS participants. It is possible that individual participants were included in both observation periods if they took the Floodlight Open tests during both periods. Because the different durations of these 2 periods might impact the continuous usage times and the inactivity times, both UX metrics were assessed over the first 343 days of data collected from each participant.

Results

Participants

Until November 2, 2021, when the study closed in the United States, the Floodlight Open app was downloaded 5225 times, including 4240 (81.15%) times on iOS devices and 985 (18.85%) times on Android devices, across the 17 participating countries. The baseline demographics of the US participants with MS included in the qualitative and quantitative UX analyses on reminders and notifications are presented in Table 3.

Table 3. Baseline demographics of US participants with MS[a] included in the qualitative UX[b] analysis and the quantitative UX analysis.

Variable	Qualitative UX analysis (N=22[c])	Quantitative UX analysis[d], n (%)					
		FLO[e] app without configurable reminders and push notifications[f] (n=350)		FLO app with configurable reminders and push notifications[g] (n=172)		All[h] (N=498)	
		Female (n=264)	Male (n=86)	Female (n=123)	Male (n=49)	Female (n=368)	Male (n=130)
Age (y), mean (SD; range)	50 (9.9; 18-74)	47.53 (11.4; 19-71)	49.63 (13.2; 19-84)	48.69 (12.29; 22-73)	50.73 (10.57; 30-79)	47.6 (11.6; 19-73)	49.7 (12.7;19-84)
Age distribution (y), n (%)							
18-24	0 (0)	4 (1.5)	1 (1.2)	4 (3.3)	0 (0)	6 (1.6)	1 (0.8)
25-34	0 (0)	32 (12.1)	12 (13.9)	10 (8.1)	3 (6.1)	42 (11.4)	16 (12.3)
35-44	3 (13.6)	64 (24.2)	19 (22.1)	32 (26)	11 (22.5)	99 (26.9)	29 (22.3)
45-54	7 (31.8)	90 (34.1)	22 (25.6)	31 (25.2)	19 (38.8)	111 (30.2)	38 (29.2)
55-64	7 (31.8)	53 (20.1)	24 (27.9)	31 (25.2)	12 (24.5)	78 (21.2)	34 (26.2)
65-74	5 (22.7)	21 (8)	7 (8.1)	15 (12.2)	3 (6.1)	32 (8.7)	10 (7.7)
≥75	0 (0)	0 (0)	1 (1.2)	0 (0)	1 (2)	0 (0)	2 (1.5)

[a]MS: multiple sclerosis.

[b]UX: user experience.

[c]Of the 22 participants included in the qualitative user experience (UX) analysis, 19 (86.4%) were female.

[d]Some participants may be included both in the cohort with configurable reminders and notifications and in the cohort without configurable reminders and notifications if they provided data during both periods. Demographic information was collected only after the successful activation of the Floodlight Open app and is, therefore, not available for the participants included in the assessment of the more seamless user journey.

[e]FLO: Floodlight Open.

[f]Includes all US participants with MS who participated in the study from November 12, 2018, through October 20, 2019.

[g]Includes all US participants with MS who participated in the study from February 19, 2020, through November 2, 2021.

[h]Includes all US participants with MS who participated in the study from November 12, 2018, through November 2, 2021, and were included in the analysis of the impact of introducing configurable reminders and push notifications.

Qualitative UX Analysis

A total of 22 US participants were enrolled in the qualitative substudy, of which 15 (68%) were interviewed. Between 7 and 19 participants are considered adequate for qualitative research [30].

The qualitative UX analysis reveals key points for improving the UX with the Floodlight Open app (Table 4). While many participants did not face any difficulties downloading the Floodlight Open app and were willing to perform the tests daily, some were unaware or confused about the activation process despite instructions being provided on the study's web portal (see *Need for a More Seamless User Journey* section in Table 4). Other participants reported that they did not complete the activation process due to distractions or the lack of reminders to complete this step. Here, a more seamless journey from registering on the web portal to activating the Floodlight Open app could improve the UX. In addition, participants who successfully activated the Floodlight Open app self-reported that they used it for 2 to 30 days and agreed that screen contents, prompts, and information were easy to see and understand. However, they also reported that they often forgot to perform the tests or felt it was not always feasible to perform all the tests in one setting. Therefore, configurable reminders and push notifications would have been beneficial (see bottom section in Table 4). Such UX design features could be used to set up reminders to perform the tests or to stop the current test and plan to complete the tests later. Furthermore, participants were not always available to perform tests that required them to stand or walk:

> *I can only proceed so far because I'm in the car waiting for my kid; I can't stand and balance...or do the walking back and forth and the balancing stuff.*
>
> *The part that is difficult to get done is the 2-minute walk. To find the time...is sometimes hard for someone like me who is more active.*
>
> *Certain tasks...were a little too long. How many more lines do I need to draw? How many more tomatoes can I pinch?*

Table 4. Learnings from the qualitative UX[a] analysis.

Learnings from the qualitative UX analysis	Comments made by participants that support the learnings
Need for a more seamless user journey	• "I see a register button and an activate button, and I don't know if I am registered, or what I need to do to activate it. I remember getting to this page and thinking there's something I have to do, some code or something to be able to activate it." • "I let it go too long before I tried to use it, and so the registration/activation process was not clear to me at that point and I never followed up on it."
Need for configurable reminders and push notifications	• "I just don't know. I think maybe it was something that was on my phone, and I was like, 'What is this? You know?'" • "I would say life got in the way." • "I just forgot. I got distracted. It just went out of my brain and never came back. What I should have done in retrospect is put a reminder popup in my phone, like to pop up every day. I would much rather the app remind me. The less I have to touch my phone, the better my life is." • "I have been terrible at using the app. I've used it twice. I think one of the problems I had with the app is it doesn't remind me to do it. If my phone doesn't tell me to do something, it probably doesn't happen. I would have been much more inclined to do it if it had reminded me to do it." • "Things were crazy with work. I've had other things on my mind. I didn't see anything pop up reminding me to use it." • "I didn't read the instructions. I had to guess what to do with the matching shapes. At first I thought I was being timed so I was rushing."

[a]UX: user experience.

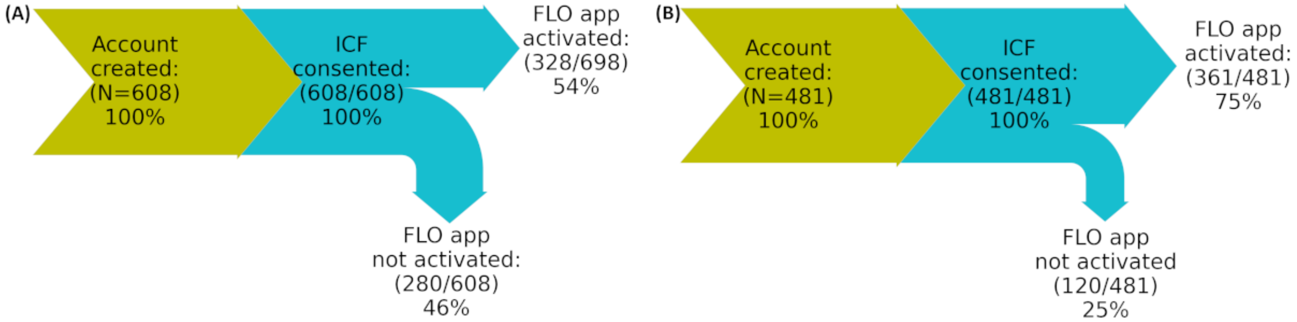
Consequently, in the subsequent section on quantitative UX analysis, UX with the Floodlight Open app was investigated with respect to DA, DA_x, and 2MWT.

Quantitative UX Analysis

User Journey

Of the 1089 US participants, 608 (55.8%) with or without MS (self-declared MS disease status was not available for this analysis, as this information was collected only after the successful activation of the app) followed the original user journey without the deep-linking feature, and 481 (44.2%) participants with or without MS followed the newer, more seamless user journey with the deep-linking feature. The flow diagrams in Figure 4 show the conversion rates from creating an account to activating the app for both cohorts. In both cohorts, all participants who created an account completed the registration process, provided informed consent, and enrolled in the study. However, Figure 4 shows that introducing the new user journey improved the conversion rate of the participants who succeeded in activating the app (359/481, 74.6% with the more seamless user journey vs 328/608, 53.9% with the original user journey).

Figure 4. Conversion rate diagrams. These diagrams depict the user journey from registration on the Floodlight Open web portal (creating a Floodlight Open account) to the activation of the Floodlight Open app (A) before and (B) after implementing the more seamless user journey. FLO: Floodlight Open; ICF: informed e-consent form.



Configurable Reminders and Push Notifications

The analysis on configurable reminders and push notifications included 172 participants with MS to whom this UX design feature was available and 350 participants with MS to whom this feature was not available (Table 3). The cumulative distributions of the continuous usage times of participants with MS using the Floodlight Open app with reminders and notifications and those using the Floodlight Open app without reminders and notifications are plotted in a log-log scale for DA, DA_x, and 2MWT in Figure 5A. The introduction of reminders and notifications increased the continuous app usage times. The longest continuous usage times with reminders and notifications versus those without reminders and notifications were 103 versus 13 days for DA, 224 versus 61 days for DA_x, and 65 versus 12 days for 2MWT. Similarly, reminders and notifications increased the proportion of participants using the app for at least 3 consecutive days (DA and DA_x: ~30% vs ~12%; 2MWT: ~30% vs ~20%).

Figure 5. Complementary cumulative distributions of the continuous app use time (A), inactivity time (B), test completion rate (C), and user retention rate (D) of participants with multiple sclerosis (MS) using the Floodlight Open app with reminders and notifications (wR&N; blue) and without reminders and notifications (woR&N; green) for all daily activities performed in the fixed sequence (DA), all daily activities performed in the fixed sequence without Static Balance Test and U-Turn Test (DAx), and Two-Minute Walk Test (2MWT). Continuous app usage time, test completion rate, and user retention rate improved with reminders and notifications after an initial onboarding phase, whereas inactivity time was comparable between the 2 cohorts.

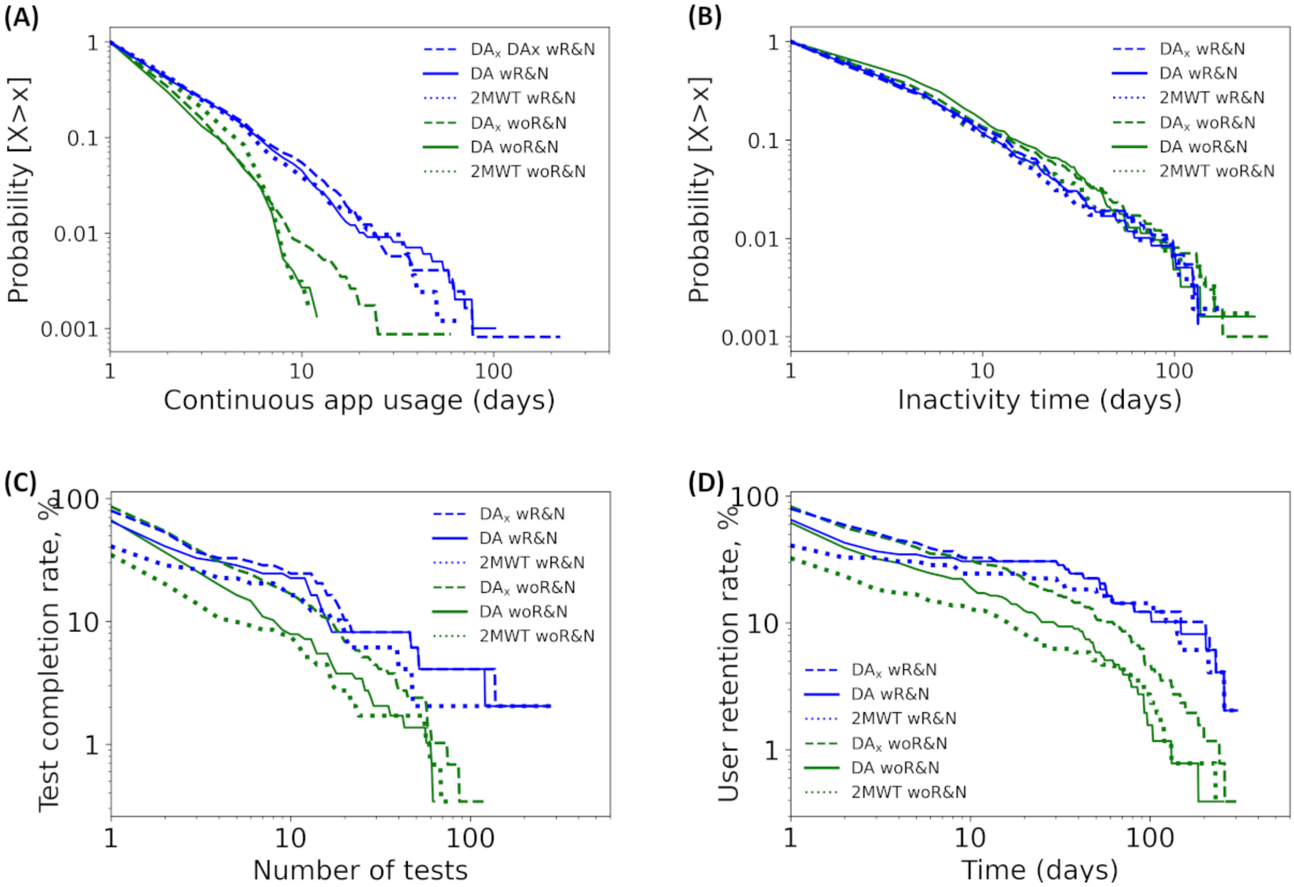


Figure 5B presents the inactivity time distributions from participants using the Floodlight Open app with configurable reminders and push notifications and those using the Floodlight Open app without configurable reminders and push notifications on a log-log scale for DA, DA_x, and 2MWT. These distributions were similar for participants with reminders and notifications and those without reminders and notifications for DA, DA_x, and 2MWT, exhibiting an approximate power law with a coefficient of 0.96 in both cohorts. The longest inactivity times with reminders and notifications versus without reminders and notifications were 145 versus 263 days for DA, 138 versus 309 days for DA_x, and 181 versus 263 days for 2MWT.

Figure 5C presents the test completion rate distributions for DA, DA_x, and 2MWT. These distributions show that the introduction of reminders and notifications resulted in higher test completion rates, except during the initial onboarding phase when DA and DA_x completion rates were similar in both cohorts. During this onboarding phase, day-1 test completion rates with and without reminders and notifications were 66% (113/172) and 66.9% (234/350), respectively, for DA and 81.3% (140/172) and 86% (301/350), respectively, for DA_x. These test completion rates are expected to be <100% because not all participants who activated the Floodlight Open app proceeded to perform at least 1 DA or DA_x series. However, the effectiveness of introducing reminders and notifications became evident after this initial onboarding phase (ie, after the first DA series and after the third DA_x series; Figure 5C). When considering the 2MWT instead, participants with reminders and notifications showed higher test completion rates than participants without reminders and notifications (41.5% [72/172] with reminders and notifications vs 34.8% [122/350] without reminders and notifications) on the first day itself. Across all 3 test series, the largest number of test series completed increased with the introduction of reminders and notifications (279 vs 64 for DA, 283 vs 126 for DA_x, and 302 vs 76 for 2MWT).

Figure 5D presents the user retention rate distributions. During the initial onboarding phase, user retention rates were comparable between participants with reminders and notifications and those without reminders and notifications. On day 1, DA user retention rates were 65.7% (113/172) and 61.7% (216/350), respectively, and DA_x user retention rates were 79.6% (137/172) and 82.8% (290/350), respectively (Table 4). DA_x user retention rates continued to be comparable between both cohorts for approximately the first 10 days. After this onboarding phase, however, the effectiveness of reminders and notifications in improving user retention rate became evident. Interestingly, after the high drop-off observed during the first few days after the activation of the app (3 days for DA and DA_x and 2 days for 2MWT), all user retention rate distributions obtained with reminders and notifications plateaued up to approximately day 30. On day 30, user retention rates were considerably higher in participants with versus participants without reminders and notifications (53/172, 30.8% vs 34/350, 9.7% for DA; 53/172, 30.8% vs 60/350, 17.1% for DA_x; and 39/172, 22.6% vs 22/350, 6.2% for 2MWT; Table 5).

Table 5. User retention rates of the Floodlight Open app with and without reminders and notifications from participants with MS[a] and average user retention rates for medical apps and health and fitness apps reported by Rosenfelder [31] on days 1, 3, 7, 14, and 30.

Apps, configuration, and series	User retention rate: Day 1	Day 3	Day 7	Day 14	Day 30
Floodlight Open app, % (n/N)					
Without configurable reminders and push notifications					
DA[b]	61.7 (216/350)	32 (112/350)	23.4 (82/350)	16.3 (57/350)	9.7 (34/350)
DA_x[c]	82.8 (290/350)	48.8 (171/350)	34.5 (121/350)	28 (98/350)	17.1 (60/350)
2MWT[d]	32.34 (113/350)	18.2 (64/350)	14.2 (50/350)	11.3 (39/350)	6.2 (22/350)
With configurable reminders and push notifications					
DA	65.7 (113/172)	36.6 (63/172)	32.6 (56/172)	30.8 (53/172)	30.8 (53/172)
DA_x	79.6 (137/172)	51.1 (88/172)	36.6 (63/172)	30.8 (53/172)	30.8 (53/172)
2MWT	40.7 (70/172)	32.6 (56/172)	28.5 (49/172)	24.4 (42/172)	22.6 (39/172)
Rosenfelder [31], 2020, %[e]					
Medical apps					
Active	20	11.52	9.24	7.19	5.46
Health and fitness apps					
Active	18.37	10.56	7.77	5.56	3.6

[a]MS: multiple sclerosis.

[b]DA: all daily activities that were administered in a predefined, fixed sequence.

[c]DA_x: all daily activities that were administered in a predefined, fixed sequence, except the Static Balance Test and U-Turn Test.

[d]2MWT: Two-Minute Walk Test.

[e]The absolute number of uses retained at each time point were not reported in Rosenfelder [31], 2020.

Discussion

Principal Findings

Smartphones enable out-of-clinic assessments of chronic neurological diseases. Despite the rapidly increasing number of mobile health care apps available for consumers' self-care, there is a paucity of research into the UX of DHTTs for MS. In this paper, we present our qualitative and quantitative UX analyses of the Floodlight Open app and demonstrate that the adoption of key UX design features markedly improved user acquisition, engagement, and retention. Understanding how the presence or absence of specific UX design features affects participants' experiences offers important guidance for the refinement of an existing or the design of a new DHTT. Therefore, UX design features are a significant consideration when designing and evaluating such tools.

Assessing MS symptoms over time with the Floodlight Open app requires specific user stimulation and guidance through seamless and flexible user journeys enhanced by engaging features to improve the overall UX. By considering participants' feedback collected in the quantitative UX substudy, we enhanced the onboarding process by implementing a more seamless user registration and app activation journey and improved the UX through configurable reminders and push notifications. We derived several distributions from real-world data that describe key aspects of participants' behavior in their environment. Such measurements and their statistical analysis might be used to help describe realistic user behavior and design health care apps similar to our Floodlight Open app. Our findings allow us to gain a better understanding of users' experience in such settings and to make informed design decisions for self-monitoring mobile apps that support people with MS.

Users' first experience with the registration to the Floodlight Open study and the activation of the study app is important to increase user acquisition. Furthermore, the first few days after activating the app are similarly critical for long-term engagement and retention. For instance, user acquisition, engagement, and retention can all be improved by guiding users throughout the Floodlight Open study registration, app activation process, and daily activities, reminding them about the daily activities to be performed and allowing them to plan when to perform the daily activities. Our results demonstrate that a seamless UX is essential to minimize the rate at which users drop out along the end-to-end journey. With a seamless user journey, users are led along the registration-activation journey and are not distracted by unrelated content or activities, resulting in a higher conversion rate from creating an account to activating the Floodlight Open app.

Configurable reminders and push notifications can help keep the participants engaged with mobile health care apps, such as the Floodlight Open app, and help retain them beyond the initial onboarding phase. Our findings show that continuous use times, test completion rates, and user retention rates are generally higher with configurable reminders and push notifications than without configurable reminders and push notifications, whereas inactivity time is comparable between the 2 cohorts. This improvement in user engagement is most evident for 2MWT from the first day and for DA and DA_x after the initial onboarding phase, during which the participants are still exploring the Floodlight Open app. Therefore, to fully optimize the UX, the app should ask the users to accept and receive reminders and notifications only after this initial onboarding phase [32]. Interrupting users with notifications too soon after the activation of the app may create frustration and may ultimately lead them to abandon the app. The app should, therefore, minimize the mental and physical interaction efforts during the first few days of use.

Comparing the different test series, we noted slightly worse user retention for 2MWT than for either DA or DA_x. Taking the 2MWT requires the greatest effort, which is in line with some dissatisfaction reported in a previous proof-of-concept study [33]. It is possible that the study participants did want to take the 2MWT but did not find a suitable opportunity to take it each day. In fact, participants in Floodlight Open were found to persistently take the 2MWT at least once per week [15], which suggests that a more flexible or less frequent assessment schedule could be beneficial.

Comparison With Prior Work

Although several other smartphone apps are available for people with MS (for a review on MS apps, refer to the study by Howard et al [34]), to our knowledge, this is the first analysis to assess the impact of UX design features on user engagement and retention. Nonetheless, previous analyses have shown that user engagement and retention remain two of the barriers to the long-term successful deployment of DHTTs (for reviews, refer to works of Pratap et al [6] and Amagai et al [35]). Several UX design features have been suggested to improve the overall UX, including a more seamless user journey, configurable reminders, and push notifications [23,35-38]. However, differences in the definitions used to assess user retention (eg, user retention definitions based on the user simply opening or interacting with the DHHT [31] vs completing a test or series of tests [6]), limited statistical power due to small sample sizes [37], and differences in the observed study period [37] may make direct comparisons of user retention across different DHHTs challenging.

To compare like with like, we computed the user retention with the Floodlight Open app for the same observation period as that reported in the study by Rosenfelder [31] for medical as well as health and lifestyle apps. Both data sets indicate that the largest reduction of users occurs from days 1 to 3, whereas the reduction in user retention is low between days 7 and 30. This suggests that a significant proportion of users who will not use the Floodlight Open app (or other health care apps) in the long term will stop using the app within the first 3 days and that this dropout is likely to be observed within the first 7 days. These results also suggest that participants who use the app for >3 days are likely to use it for at least 30 days. A similar dropout during the first 7 days has also been previously reported for the whole study cohort of Floodlight Open [15].

Of note, our user retention rates are approximately 2- to 6-fold higher than those reported by Rosenfelder [31] (Table 5) despite using a more conservative definition of user retention (completing a test series vs simply interacting with or opening the app). It is conceivable that not all apps included in the report by Rosenfelder [31] feature the same UX design features as the Floodlight Open app does. However, favorable user retention rates were observed with the Floodlight Open app even when the app did not have configurable reminders or push notifications. This suggests that differences in the perceived benefit of the app might have also contributed to the favorable retention profile, although other reasons cannot be excluded [39]. Despite our favorable user retention and user engagement findings, a previous analysis showed that improvement in smartphone sensor data collection could be achieved through passive data collection methods [15].

Limitations

A few limitations are noted. First, while participants could turn off the configurable reminders and push notifications, this action was not logged. Instead, we compared participants who enrolled before with those who enrolled after these features were implemented. For the former cohort, we included only the data collected up to the implementation of reminders and notifications in our analyses to be able to assess the impact of reminders and notifications. Second, our results were influenced by the duration of the study and by incoming participants who activated the Floodlight Open app on different dates over the duration of the study and might not be part of the study for the same number of days. Besides, for events approaching the duration of the study, there is an artificially lower likelihood of observation, and events lasting longer than the study cannot be observed. Finally, the use of the Floodlight Open app in a clinical study other than the Floodlight Open observational study may boost user engagement and retention due to additional motivational incentives. Such clinical referrals have been shown to improve user retention [6]. However, information on which participants participated in such studies is not available, and without it, accounting for it is not possible.

Conclusions

We presented the qualitative and quantitative UX analyses of the Floodlight Open app, a DHTT for MS. Learnings from the qualitative UX analysis led to the implementation of a more seamless user journey and the introduction of configurable reminders and push notifications. These US design features improved user acquisition, user engagement, and user retention. The continued assessment of UX and improvement of UX design features are critical steps in optimizing the long-term adoption of the Floodlight Open technology and similar DHTTs.

Acknowledgments

The authors would like to thank all the Roche Floodlight Open team members, in particular, Luca Capezzuto, as well as all the participants and researchers who expressed interest in this study. This research was funded by F. Hoffmann-La Roche Ltd, Basel, Switzerland. Writing and editorial assistance for this manuscript was provided by Sven Holm, PhD, contractor for F. Hoffmann-La Roche Ltd. Up-to-date details on Roche’s Global Policy on Sharing of Clinical Study Information and how to request access to related clinical study documents are provided [40].

Data Availability

Request for the data underlying this publication requires a detailed, hypothesis-driven statistical analysis plan that is collaboratively developed by the requester and company subject matter experts. Such requests should be directed to dbm.datarequest@roche.com for consideration. Anonymized records for individual patients across >1 data source external to Roche cannot, and should not, be linked due to a potential increase in the risk of patient reidentification.

Conflicts of Interest

AG, LK, and RL are employees of F Hoffmann-La Roche Ltd. ML is a consultant for F Hoffmann-La Roche Ltd via Inovigate. JBJ is an employee of Sutter Health’s Center for Health Systems Research. He has received research grants from Genentech and AstraZeneca.

References

1. Kappos L, Wolinsky JS, Giovannoni G, Arnold DL, Wang Q, Bernasconi C, et al. Contribution of relapse-independent progression vs relapse-associated worsening to overall confirmed disability accumulation in typical relapsing multiple sclerosis in a pooled analysis of 2 randomized clinical trials. JAMA Neurol 2020 Sep 01;77(9):1132-1140 [FREE Full text] [doi: 10.1001/jamaneurol.2020.1568] [Medline: 32511687]
2. Hobart J, Bowen A, Pepper G, Crofts H, Eberhard L, Berger T, et al. International consensus on quality standards for brain health-focused care in multiple sclerosis. Mult Scler 2019 Nov;25(13):1809-1818 [FREE Full text] [doi: 10.1177/1352458518809326] [Medline: 30381987]
3. Montalban X, Graves J, Midaglia L, Mulero P, Julian L, Baker M, et al. A smartphone sensor-based digital outcome assessment of multiple sclerosis. Mult Scler 2022 Apr;28(4):654-664 [FREE Full text] [doi: 10.1177/13524585211028561] [Medline: 34259588]
4. van der Walt A, Butzkueven H, Shin RK, Midaglia L, Capezzuto L, Lindemann M, et al. Developing a digital solution for remote assessment in multiple sclerosis: from concept to software as a medical device. Brain Sci 2021 Sep 21;11(9):1247 [FREE Full text] [doi: 10.3390/brainsci11091247] [Medline: 34573267]
5. Gromisch ES, Turner AP, Haselkorn JK, Lo AC, Agresta T. Mobile health (mHealth) usage, barriers, and technological considerations in persons with multiple sclerosis: a literature review. JAMIA Open 2020 Dec 15;4(3):ooaa067 [FREE Full text] [doi: 10.1093/jamiaopen/ooaa067] [Medline: 34514349]
6. Pratap A, Neto EC, Snyder P, Stepnowsky C, Elhadad N, Grant D, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020 Feb 17;3:21 [FREE Full text] [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
7. Law EL, Roto V, Hassenzahl M, Vermeeren AP, Kort J. Understanding, scoping and defining user experience: a survey approach. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2009 Presented at: CHI '09; April 4-9, 2009; Boston, MA. [doi: 10.1145/1518701.1518813]
8. Bitkina OV, Kim HK, Park J. Usability and user experience of medical devices: an overview of the current state, analysis methodologies, and future challenges. Int J Ind Ergon 2020 Mar;76:102932. [doi: 10.1016/j.ergon.2020.102932]
9. Park J, Han SH. Defining user value: a case study of a smartphone. Int J Ind Ergon 2013 Jul;43(4):274-282. [doi: 10.1016/j.ergon.2013.04.005]

10. Park J, Han SH, Kim HK, Cho Y, Park W. Developing elements of user experience for mobile phones and services: survey, interview, and observation approaches. Hum Factors Ergon Manuf Serv Ind 2011 Oct 20;23(4):279-293. [doi: 10.1002/hfm.20316]
11. Park J, Han SH, Kim HK, Oh S, Moon H. Modeling user experience: a case study on a mobile device. Int J Ind Ergon 2013 Mar;43(2):187-196. [doi: 10.1016/j.ergon.2013.01.005]
12. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
13. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
14. Smith W, Ploderer B, Wadley G, Webber S, Borland R. Trajectories of engagement and disengagement with a story-based smoking cessation app. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI '17; May 6-11, 2017; Denver, CO. [doi: 10.1145/3025453.3026054]
15. Oh J, Capezzuto L, Kriara L, Schjodt-Eriksen J, van Beek J, Bernasconi C, et al. Use of smartphone-based remote assessments of multiple sclerosis in Floodlight Open, a global, prospective, open-access study. Sci Rep 2024 Jan 02;14(1):122 [FREE Full text] [doi: 10.1038/s41598-023-49299-4] [Medline: 38168498]
16. Jones J, Yan X, Liu J, Popova I, Mitchell D, Lindemann M, et al. Early experience engaging patients in the use of a mobile application for tracking multiple sclerosis. Mult Scler J 2020. [doi: 10.1016/j.neurol.2021.02.360]
17. Graves JS, Elantkowski M, Zhang YP, Dondelinger F, Lipsmeier F, Bernasconi C, et al. Assessment of upper extremity function in multiple sclerosis: feasibility of a digital pinching test. JMIR Form Res 2023 Oct 02;7:e46521 [FREE Full text] [doi: 10.2196/46521] [Medline: 37782540]
18. Cheng WY, Bourke AK, Lipsmeier F, Bernasconi C, Belachew S, Gossens C, et al. U-turn speed is a valid and reliable smartphone-based measure of multiple sclerosis-related gait and balance impairment. Gait Posture 2021 Feb;84:120-126 [FREE Full text] [doi: 10.1016/j.gaitpost.2020.11.025] [Medline: 33310432]
19. Hesser-Biber SN, Leavy PL. The Practice of Qualitative Research. Thousand Oaks, CA: SAGE Publications, Inc; 2011.
20. Gloyd DM. Positive user experience and medical adherence. In: Proceedings of the 2003 International Conference on Designing Pleasurable Products and Interfaces. 2003 Presented at: DPPI '03; June 23-26, 2003; Pittsburgh, PA. [doi: 10.1145/782901.782902]
21. Leventhal H, Safer MA, Panagis DM. The impact of communications on the self-regulation of health beliefs, decisions, and behavior. Health Educ Q 1983;10(1):3-29. [doi: 10.1177/109019818301000101] [Medline: 6629788]
22. Lipkus IM, Hollands JG. The visual communication of risk. J Natl Cancer Inst Monogr 1999(25):149-163. [doi: 10.1093/oxfordjournals.jncimonographs.a024191] [Medline: 10854471]
23. Morrison LG, Hargood C, Pejovic V, Geraghty AW, Lloyd S, Goodman N, et al. The effect of timing and frequency of push notifications on usage of a smartphone-based stress management intervention: an exploratory trial. PLoS One 2017 Jan 3;12(1):e0169162 [FREE Full text] [doi: 10.1371/journal.pone.0169162] [Medline: 28046034]
24. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
25. Whelan ME, Orme MW, Kingsnorth AP, Sherar LB, Denton FL, Esliger DW. Examining the use of glucose and physical activity self-monitoring technologies in individuals at moderate to high risk of developing type 2 diabetes: randomized trial. JMIR Mhealth Uhealth 2019 Oct 28;7(10):e14195 [FREE Full text] [doi: 10.2196/14195] [Medline: 31661077]
26. Klasnja P, Harrison BL, LeGrand L, LaMarca A, Froehlich J, Hudson SE. Using wearable sensors and real time inference to understand human recall of routine activities. In: Proceedings of the 10th International Conference on Ubiquitous Computing. 2008 Presented at: UbiComp '08; September 21-24, 2008; Seoul, Korea. [doi: 10.1145/1409635.1409656]
27. ISO/IEC 25022:2016 systems and software engineering — systems and software quality requirements and evaluation (SQuaRE) — measurement of quality in use. International Organization for Standardization. 2016. URL: https://www.iso.org/standard/35746.html [accessed 2024-05-29]
28. Sigg S, Lagerspetz E, Peltonen E, Nurmi P, Tarkoma S. Exploiting usage to predict instantaneous app popularity: trend filters and retention rates. ACM Trans Web 2019 Apr 02;13(2):1-25. [doi: 10.1145/3199677]
29. Zuniga A, Flores H, Lagerspetz E, Tarkoma S, Manner J, Hui P, et al. Tortoise or hare? Quantifying the effects of performance on mobile app retention. In: Proceedings of the World Wide Web Conference. 2019 Presented at: WWW '19; May 13-17, 2019; San Francisco, CA. [doi: 10.1145/3308558.3313428]
30. Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: a systematic review of empirical tests. Soc Sci Med 2022 Jan;292:114523 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114523] [Medline: 34785096]
31. Rosenfelder S. 2020 app retention benchmarks. AppsFlyer. URL: https://www.appsflyer.com/infograms/app-retention-benchmarks-2020/ [accessed 2022-01-05]
32. Gamberini L, Petrucci G, Spoto A, Spagnolli A. Embedded persuasive strategies to obtain visitors' data: comparing reward and reciprocity in an amateur, knowledge-based website. In: Proceedings of the Second International Conference on

Persuasive Technology. 2007 Presented at: PERSUASIVE 2007; April 26-27, 2007; Palo Alto, CA. [doi: 10.1007/978-3-540-77006-0_24]

33. Midaglia L, Mulero P, Montalban X, Graves J, Hauser SL, Julian L, et al. Adherence and satisfaction of smartphone- and smartwatch-based remote active testing and passive monitoring in people with multiple sclerosis: nonrandomized interventional feasibility study. J Med Internet Res 2019 Aug 30;21(8):e14863 [FREE Full text] [doi: 10.2196/14863] [Medline: 31471961]
34. Howard Z, Win KT, Guan V. Mobile apps used for people living with multiple sclerosis: a scoping review. Mult Scler Relat Disord 2023 May;73:104628 [FREE Full text] [doi: 10.1016/j.msard.2023.104628] [Medline: 37003008]
35. Amagai S, Pila S, Kaat AJ, Nowinski CJ, Gershon RC. Challenges in participant engagement and retention using mobile health apps: literature review. J Med Internet Res 2022 May 26;24(4):e35120 [FREE Full text] [doi: 10.2196/35120] [Medline: 35471414]
36. Wang S, Sporrel K, van Hoof H, Simons M, de Boer RD, Ettema D, et al. Reinforcement learning to send reminders at right moments in smartphone exercise application: a feasibility study. Int J Environ Res Public Health 2021 Jun 04;18(11):6059 [FREE Full text] [doi: 10.3390/ijerph18116059] [Medline: 34199880]
37. Bremer W, Sarker A. Recruitment and retention in mobile application-based intervention studies: a critical synopsis of challenges and opportunities. Inform Health Soc Care 2023 Apr 03;48(2):139-152. [doi: 10.1080/17538157.2022.2082297] [Medline: 35656732]
38. Polhemus A, Simblett S, Dawe Lane E, Elliott B, Jilka S, Negbenose E, et al. Experiences of health tracking in mobile apps for multiple sclerosis: a qualitative content analysis of user reviews. Mult Scler Relat Disord 2023 Jan;69:104435. [doi: 10.1016/j.msard.2022.104435] [Medline: 36493561]
39. Giunti G, Kool J, Rivera Romero O, Dorronzoro Zubiete E. Exploring the specific needs of persons with multiple sclerosis for mHealth solutions for physical activity: mixed-methods study. JMIR Mhealth Uhealth 2018 Feb 09;6(2):e37 [FREE Full text] [doi: 10.2196/mhealth.8996] [Medline: 29426814]
40. Our commitment to transparency of clinical study information. Roche. URL: https://go.roche.com/data_sharing [accessed 2024-02-02]

Abbreviations

2MWT: Two-Minute Walk Test
DA: all daily activities that were administered in a predefined, fixed sequence
DaS: Draw a Shape Test
DAx: all daily activities that were administered in a predefined, fixed sequence excluding the Static Balance Test and U-Turn Test
DHTT: digital health technology tool
DMQ: Daily Mood Questionnaire
FLO ID: personal Floodlight Open unique identifier
IPS: Information Processing Speed
MS: multiple sclerosis
PT: Pinching Test
SBT: Static Balance Test
UTT: U-Turn Test
UX: user experience

Edited by K Adapa; submitted 08.02.24; peer-reviewed by G Figueredo, C Ochoa-Zezzatti; comments to author 29.04.24; revised version received 17.06.24; accepted 02.07.24; published 11.09.24.

Please cite as:
Galati A, Kriara L, Lindemann M, Lehner R, Jones JB
User Experience of a Large-Scale Smartphone-Based Observational Study in Multiple Sclerosis: Global, Open-Access, Digital-Only Study
JMIR Hum Factors 2024;11:e57033
URL: https://humanfactors.jmir.org/2024/1/e57033
doi:10.2196/57033
PMID:

Original Paper

A Compressive Armchair (OTO) to Perform Deep Pressure Therapy in Children With Autism Spectrum Disorder: User-Centered Design and Feasibility Study

Thomas Gargot[1,2,3], MD, PhD; Amandine Vachaud[1,2]; Clémence Gilard[1,2]; Alexia Audrain[4]; Marie Gomot[2,3], PhD; Marco Guidotti[1,2], MD, PhD; Frédéric Briend[2,3], PhD; Joëlle Malvy[1,2], MD; Frédérique Bonnet Brilhault[1,2,3], MD, PhD

[1]Child and Adolescent Psychiatry Department, University Hospital Tours, Tours, France

[2]EXcellence Center in Autism and Neurodevelopmental Disorders—Tours ExAC-T, Tours, France

[3]Université de Tours, INSERM, Imaging Brain & Neuropsychiatry iBraiN U1253, 37032, Tours, France

[4]Labaa Company, Nantes, France

Corresponding Author:
Thomas Gargot, MD, PhD
Child and Adolescent Psychiatry Department
University Hospital Tours
2 Bd Tonnelle
Tours, 37000
France
Phone: 33 0234389435
Email: gargot@univ-tours.fr

Abstract

Background: Deep pressure therapy (DPT) is widely used to reduce anxiety in children with autism spectrum disorder (ASD), but evidence of its efficacy is limited.

Objective: This study aims to design a usable, nonstigmatizing compressive armchair that can be easily controlled, electronically, by the user.

Methods: A user-centered approach was used to assess the usability of the device. Testing was carried out in a day hospital for children with ASD in France, with a convenience sample of children with severe forms of ASD and intellectual deficiency (N=39). The Witteman design guideline was used. The System Usability Scale and time of use were reported.

Results: The final product is a compressive armchair designed to be user centered, with 4 different cells that can be inflated to induce tailored pressure on the body. The pressure level is recorded electronically. Usability was between good and excellent. The device was used by 39 children, once or twice weekly, over a period of 31 months. Each session lasted between 3 and 20 minutes. The armchair takes up less space than a hug machine. Performing sessions with the chair is feasible.

Conclusions: First clinical impressions show a decrease in anxiety, improved emotional regulation, and improved attention. DPT is widely used in occupational therapy and frequently requested by parents, but efficacy studies are too scarce to make evidence-based recommendations for its use. The results presented here support further controlled efficacy studies of DPT in the treatment of anxiety in children with ASD.

(JMIR Hum Factors 2024;11:e55754) doi:10.2196/55754

KEYWORDS

deep pressure therapy; proprioception; compression; autism spectrum disorder

Introduction

Background

Autism spectrum disorder (ASD) is defined by (1) persistent deficits in social communication and social interaction across multiple contexts and (2) restricted, repetitive patterns of behavior, interests, or activities [1]. According to the American Psychiatric Association [1], the prevalence of autism is 1%.

Sensory difficulties are frequently found in individuals with ASD sensitivity [2], in particular somatosensory system difficulties such as aberrant skin sensitivity [3,4] (including pressure detection) and proprioception. These sensory anomalies could underlie the pathophysiological processes that lead to impaired social development [5].

Proprioception is the sensory registration of the ongoing spatial configuration of the body. It includes the position of the body segments in space, the force and the speed of movement, and the integration of gravity and body balance. Proprioception impacts behavioral regulation and motor control [6]. Blanche et al [7] showed that children with ASD present proprioceptive processing difficulties that are different from those of children with other developmental disabilities and their typically developing counterparts. However, Morris et al [8] and Fuentes et al [9] did not confirm these proprioceptive difficulties in experimental paradigms. It is possible that the deficits rely mostly on multisensory integration [10].

Sensory Integration Theory

Ayres et al [6] introduced sensory integration theory to explain sensory processing issues in children with ASD. The focus of sensory-based intervention is to maintain an optimal level of arousal between hypo- and hyperstimulation, allowing the individual to respond to the environment in an adaptive manner [11,12].

Several techniques or devices can be used in sensory-based interventions, in particular deep pressure therapy (DPT); for instance, Wilbargers suggested a deep pressure and proprioceptive protocol [13]. Systematic reviews show that sensory integration therapies, which use play activities and sensory-enhanced interactions, have positive effects, but the quality of the studies is not sufficient to confirm these results [11,14]. Nevertheless, the American Occupational Therapy Association recommends the use of sensory strategies for individuals with ASD [12,14].

In a web survey involving 552 parents of children with ASD, Green et al [15] showed that sensory integration was the third most used treatment after speech therapy and visual schedules and before behavioral methods. In a survey involving 152 parents of children with ASD, Peña et al [16] reported high acceptability of sensory-based methods. These interventions were considered to be “very important” or “important.” Main barriers to the use of sensory-based methods were the lack of recommendations and difficulty in using or difficulty in accessing this kind of intervention [16]. Among these techniques, DPT was of particular interest.

Devices and Strategies for DPT

Different devices and strategies can be used to deliver DPT to patients with ASD (Table 1).

Table 1. Comparison of several devices used to induce deep pressure therapy in children with autism spectrum disorder.

Type of device	Principle	Level of evidence	N[a]	Time of use	Measure of efficacy	Control	Efficacy	Acceptance	Cost	Autonomy of the patient	References
Weighted blankets	Weight	Systematic review, population-based observational study	Observational: 1785; interventional: 160	>8 h daily	Sleep, STAI[b], electrodermal activity, pulse rate	Nothing or light plastic chain blanket	Conflicting evidence	+++[c]	+	+++	[17-25]
Therapeutic body wrap	Tightening	One RCT[d]	48	45 min; 2 times/wk	Aberrant Behavior Checklist, irritability	Dry vs wet sheet therapeutic body wrap	+ but no waiting list comparative arm	+/–	+	– – –	[26-29]
Shape memory vest	Tightening	None, prototypes	None	Unknown	None	None	Unknown	Unknown	++	Unknown	[30,31]
Compression vest	Pressure by inflation	SCRD[e]	3	20 min, daily (unclear) for 22-50 d	Stereotypies	Fully deflated vest or no vest	No efficacy	+++ ?	++	+++	[32,33]
Manual squeezing	Manual squeezing	SCRD	8	5-15 min, until 3 times/d for 3 mo	Visual analog scales (calmness, engaged, responsivity, happy, communicative)	None	+/–	+++	+	–	[34]
Hug or squeeze machine	Compression by a plate	RCT	12	20 min; 1 time/wk for 6 wk	The Conners Parent Rating Scale, electrodermal activity	Not receiving deep pressure in the disengaged hug machine	++	+	+++ (reusable)	–	[35-40]
Compressive garments	Tightening	Observational study	14	>1-16 h daily for 6 wk	Aberrant Behavior Checklist, sensory integration (Dunn Sensory. Profile), postural sway, motor performance	None	+ but no comparative arm	+++	+++ (tailored)	+/–	[6]
Sitting hug machine	Compression by a plate	SCRD	2	Not reported	Stereotypical behaviors	None	+ but no comparative arm	+++	++ (reusable)	+++	[35]
Compressive chair	Compression by inflated cushions	None, prototype	None	Unknown	None	None	Unknown	+++	++ (reusable)	+++	—[f]

[a]N: number of included studies in a systematic review.

[b]STAI: State Trait Anxiety Inventory-10.

[c]+++ (very) to - - - (not at all).

[d]RCT: randomized control trial.

[e]SCRD: single case research design.

[f]Not applicable.

Hug or Squeeze Machine

Krauss [35] used a pressure apparatus consisting of 2 stacked air mattresses to deliver DPT to 23 typically developing college students during an examination period. Heart rate and self-reported anxiety were measured using the State-Trait Anxiety Inventory. The control group did not receive DPT. There was no objective difference between the 2 conditions. Subjectively, participants in the deep pressure group reported a relaxing effect. The baseline level of anxiety in this population was low. The author did not exclude that the confinement alone may have induced subjective feelings of relaxation.

Grandin [41] developed a hug device to allow self-administration of lateral body pressure for individuals with high levels of anxiety. Edelson et al [36] tested this device on children with ASD (n=12). A control group received deep pressure via a disengaged hug machine. Participants had 20-minute sessions every week for 6 weeks. Arousal and anxiety was measured using the Conners Parent Rating Scale and electrodermal activity. The hug device decreased anxiety according to both behavioral and physiological measures. As pressure can be controlled by the individual, this device may be useful for children with marked anxiety. No side effects were reported.

The ergonomics of the device are an important issue [36,42]. Lo and Huang [38] interviewed professionals and patients, concluding that it was important to improve the ergonomic design of the device and its acceptability. For instance, it is necessary to lie down or to squat, which can be difficult for some children. The system is very bulky. The controller is outside of the machine and cannot be activated autonomously. It is not possible to choose the part of the body that the individual or professional wants to squeeze.

Lo and Huang [38] suggested a sitting hug machine that is more compact, controllable by the patient or the therapist, and can apply pressure selectively to either the shoulders or bottom part of the body. Stereotypies decreased during intervention for 2 children.

Afif et al [39] designed a portable, inflatable hug machine. It was tested on 5 children with ASD. They measured heart rate variability. They found that the inflatable wrap model decreased heart rate but could not find this effect with a manual pull [40].

This paper reports the design of a hug machine that aimed to (1) improve controllability of the pressure by the professional, allowing replicability and making the device useful for both care and research; (2) improve controllability of the pressure by the children or adults with ASD, allowing different pressure on the bottom and top of the body; (3) use pressure instead of restraint; and (4) be more attractive and less stigmatizing. This paper describes the design method and the device itself and evaluates the usability of the device.

Methods

Overview

In a population with special needs such as ASD with intellectual deficiency (ID), gathering children or adults with ASD feedback could be complicated by communication and social difficulties. It is important to have a tailored, user-centered strategy to improve acceptability and usability before assessing efficacy. Because many of the children were not verbal due to associated ID, we collected feedback from the professionals who guided and observed the children during the care. We also asked feedback from 1 adult with ASD. The first uses of the device were video recorded to tailor the design of the device to the needs of the patient and the therapist.

Design Goal and Process

Overview

The seat was designed by Alexia Audrain [43], a furniture maker, to address the needs of individuals with ASD for DPT (Figure 1). The project was carried out in partnership with the medical-educational institute of Blain, France, over a period of 1.5 years. After the review of existing technologies, we chose a sitting position that is natural and relaxing but allows one to be active. During the sessions, it allows the professional to keep an eye on the individual with ASD and facilitate the communication to better understand his needs. To assess the ergonomics of the shape of the chair, we tested several inclinations with a Sacco or a bean bag chair (a large fabric bag filled with polystyrene beans; Figure 2A). This allowed us to test several postures and choose one.

The educators, psychomotrician, and the director of the center identified the requirements of the device and gave feedback during the design process. The prototype consisted of inflatable cushions and was used to understand the amount of pressure required to verify the principle of action required to apply side pressure on the body and define the main technical characteristics (Figure 2B).

The sociomedical team (around 10 professionals) and a convenience sample of 30 children with ASD tested the prototype. The designer observed the behaviors of the professionals and children during the test and asked questions about the experience of using the device.

Based on the test results, different models of a compressive chair were designed and sketched in 3 dimensions to validate the form, the materials, and the colors of the prototype before construction. The first model was made in June 2019 and presented to medical-educational institute professionals, children, and the graduation committee. After this, the model was presented to a general audience and another medical educational institute (specialized educators, speech therapist, occupational therapist, and psychologist), Saint-Jean-de-Boiseau near Nantes, France, and was tested with 5 children. The device was also used with a nonverbal adult (30 y) with ASD.

Figure 1. Timeline of the design of the OTO chair. ASD: autism spectrum disorder; ID: intellectual deficiency; SUS: System Usability Scale.

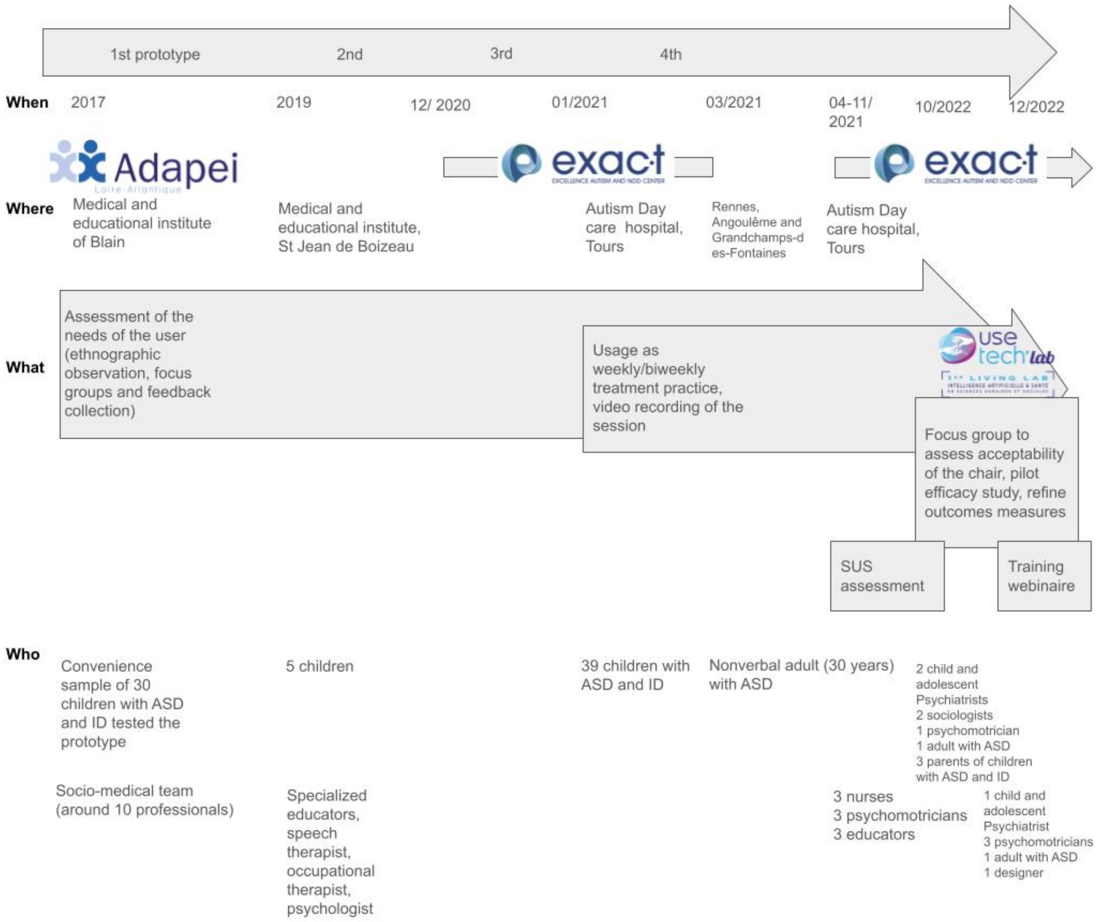
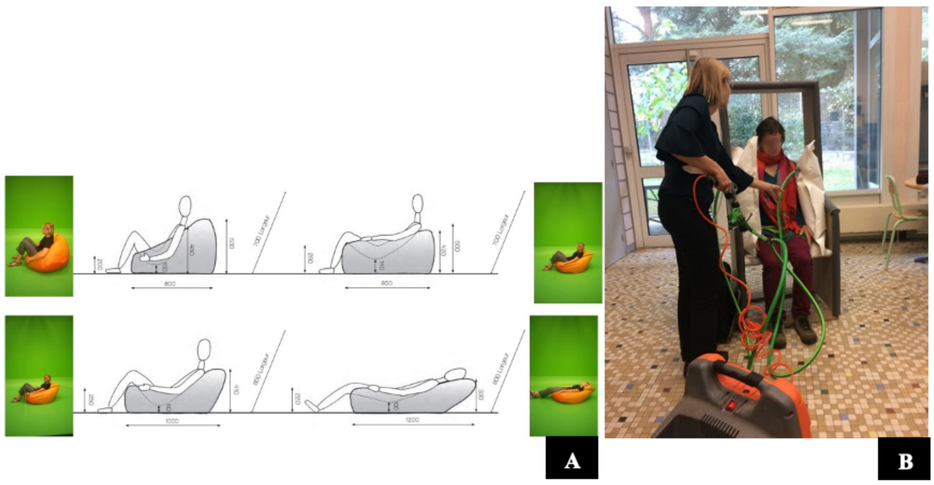


Figure 2. (A) Prototype of the OTO armchair and (B) evaluation of the ergonomics of the chair.



Testing Within the Hospital and Modification

In December 2020, OTO, a squeezing armchair, was presented to and tested by the Tours hospital team before introducing it for use in the hospital. The hospital team gave suggestions to enhance the experience of people with ASD.

New functionalities were added before the seat was installed in a day care hospital:

- Access to pressure measurement in real time for the professional
- Logging of pressure measurement, action used, and timestamp
- Seat operation monitoring and safety protocol in case of failure
- Remote control connected to the outside of the seat

Integration in the Ward

The device was transferred to the autism day hospital of the Excellence Center of Neurodevelopmental Disorders in the university hospital of Tours, France, in January 2021. After obtaining consent from parents, 39 children with ASD and ID used the device in everyday care (weekly or biweekly). The initial sessions were filmed, with pressure logs recorded by the device, in order to better understand its use in real life. The improvements made after these sessions include the following: cushions were tailored to the morphology of participants, deflation of the upper and lower cells was dissociated, a control panel was used to set the pressure limit, and the feedback light was deactivated because it was too stimulating for children. The noise accompanying inflation and deflation was reduced.

A second group of modifications were made after 2 months and implemented in March 2021:

- A range of different sizes of back cushion
- New cells with a different design for the upper and lower cells to adjust the squeezing effect
- New valves, pipes, and pump to reduce noise
- New remote design with 4 buttons to reduce air pressure independently in the upper or lower cells, and a light that can be deactivated
- A tablet connected to the seat that lets the professional define the maximum pressure

Between April and November, the seat was used in the hospital and tested for 2 months in services in Rennes, Angoulème, and Grandchamps-des-Fontaines. Based on the request of additional features from professionals in Tours and feedback from the 3 other services, some adjustments were made:

- Enhanced experience: easier on/off process, enhanced pressure measurement
- Noise reduction with valve modification and firmware adjustment
- Control panel enhancement, addition of a second remote
- Safety: continuous monitoring of program execution, error handling, and codification

Description of the End Product

OTO is a squeezing armchair that uses inflatable cells to induce deep pressure on the legs and the trunk (Figure 3; see video [44]). The pressure is progressive, measurable, and homogeneous and can be tailored for each child. Four different inflatable cells allow for modulation by varying pressures on either the shoulders, arm and trunk, or the hip and thighs.

The pressure can be controlled by the children via a remote with simple pictograms, which improves autonomy and predictability for the child. The control panel allows the children to set the maximum pressure level for the upper and lower cells. The maximum default pressure is 60 mm Hg for the upper cells and 80 mm Hg for the lower cells. This corresponds to the pressure a swimmer perceives 1 m under water. The seat records the use of the device with accompanying time logs.

The sitting position makes the device less bulky and allows more freedom of movement for the children, who do not need to lay down as they would in a hug machine. It allows the child to easily leave the armchair if uncomfortable. The footrest can be used to rest the legs, as a step for smaller children to enhance stability, and allow the health care provider to have the same height as the children and to maintain eye contact.

Pastel colors were used to limit sensory stimulation. Edges were avoided to make the armchair safer and more attractive. The device was developed to look like a cocoon to induce a feeling of privacy, to limit the stigmatization of its use, and to limit outside noise or light stimulation. The structure is made from beech wood with a metallic structure for the base. All cloth is removable and washable.

Figure 3. (A) OTO, a compressive armchair to induce deep pressure in children with ASD (final product) and (B) presentation of the different components of the OTO chair. ASD: autism spectrum disorder.

Measures of Clinical Impressions

Clinical impressions were collected via a logbook with visual analog scales and free report in a day care hospital.

Technology Readiness Level

The technology readiness level allows us to assess the level of maturity of a technology [45].

Human-Centered Design for Personal Health Tools

The Human-Centered Design for Personal Health Tools (UCD-11, a validated scale based on a systematic review of the design and development processes of 348 personal health tools) was used to assess the design process [46].

Time of Use

Beyond this design, it is important to assess the time of use of the device to assess acceptability in children and professionals. Here, it was measured by the chair.

Measure of Usability

Usability is very important when a system is used with children with special needs and should be assessed in real life [47]. The System Usability Scale (SUS) measured how psychometricians, nurses, and educators perceived the usability of the system [48]. The SUS has been used previously in autism and measures perceived usability from the perspective of professionals, rather than patients themselves [49]. It was measured in a day care hospital.

Feasibility of an Efficacy Study

We used the National Institute of Health and Care Research definition reported in first figure in the paper by Eldridge et al [50].

Ethical Considerations

As a design study, this type of research is not a clinical research, thus it is not subject to public health code and associated ethical rules, according to French regulations (Code de la santé publique - Article R1121-1; Legifrance).

Results

Usage

Recruitment was done between January and July 2023 in the ASD day hospital in Tours. A total of 39 children aged between 3 and 12 years with ASD and ID were included. Four children had to stop using the chair because of difficulties in tolerating noise, being in an enclosed space, or continuing anxiety despite habituation. These children had difficulties in sensory modulation and motor and emotional regulation. The system was adapted to decrease the inflation and deflation sound of the device.

The children used the remote heterogeneously. Some children controlled the remotes by themselves, and others relied on the professional. Different levels of pressure were used with different rhythms of pressure and deflation.

First Clinical Impression

First clinical impressions suggest an increase in pleasure and body relaxation, a decrease in anxiety, better postural stability, and better social contact (gaze and touch). When children came back to the group, it appeared that their attention and emotional regulation had improved.

However, in this population, there was marked intra- and interperson variability. A formal evaluation in a clinical trial with a larger sample and controlled procedures is required to formally assess the efficacy of the device.

Technology Readiness

The technology readiness level was between 8 and 9, meaning that the system can be used for several patients. Minor bugs and user-interface issues with the control panel need to be rectified.

Human-Centered Design for Personal Health Tools

We used several methods following the recommendations of UCD-11 (Textbox 1).

Textbox 1. Application of the Human-Centered Design for Personal Health Tools (UCD-11) during the design of the OTO chair.

1. Were potential end users (eg, patients, caregivers, family and friends, and surrogates) involved in any steps to help understand users (eg, who they are and in what context might they use the tool) and their needs?

- We performed (1) ethnographic observation of existing practices; (2) informal needs assessment; (3) contextual inquiry, (4) literature review summarized here, and (5) a training webinar to discuss with individuals with autism spectrum disorder (ASD) their sensory issues, needs, and how the OTO chair could be used in everyday life. A protocol to assess the efficacy was developed through a focus group led by sociologists. Several scales to measure outcomes were suggested during this focus group.

2. Were potential end users involved in any steps of designing, developing, and/or refining a prototype?

- After ethnographic observation of existing practices, psychometricians gave feedback after sessions with patients to help develop and refine the prototype.

3. Were potential end users involved in any steps intended to evaluate prototypes or a final version of the tool?

- We assessed usability among professionals (psychometricians, nurses, and educators) involved in guiding children during use of the OTO chair. The professionals gave feedback after the use of prototypes and again after the use of the final product.

4. Were potential end users asked their opinions of the tool in any way?

- We are finalizing focus groups with children, parents, and professionals (reported elsewhere) to assess the sensory issues of children with ASD and the use of different tools and techniques to tackle sensory issues in ASD.

5. Were potential end users observed using the tool in any way?

- The sessions were filmed to allow the designer to tailor the design of the chair to the needs of the children.

6. Did the development process have 3 or more iterative cycles?

- Four iterations were done.

7. Were changes between iterative cycles explicitly reported in any way?

- Major changes are reported in this paper.

8. Were health professionals asked their opinion of the tool at any point?

- Child and adolescent psychiatrists, psychomotricians, nurses, and educators provided feedback. These professionals are the most likely to use the OTO chair with children. We gathered feedback on the usability of the chair and observed them using the tool.

9. Were health professionals consulted before the first prototype was developed?

- Ethnographic evaluations were carried out with professionals.

10. Were health professionals consulted between initial and final prototypes?

- After the prototype was developed, health professionals gave feedback, which was used to finalize the design of the product.

11. Was an expert panel involved?

- The armchair received several prizes from several committees:
  - The Canopé (€5000, being US $5560), a national innovation competition organized by Forinvest and the Superior School of Wood, specialized in wood technology.
  - The St Pierre Foundation health innovation prize (€25,000, being US $28,000). The St Pierre Foundation specializes in children's health, and the award is decided by a panel of medical professionals.
  - James Dyson award for design. Awarded by the James Dyson Foundation and decided by a panel of engineers.
  - Startup and innovation day prize, 2022. This prize recognizes innovative startups.
  - Crédit Mutuel 4S Semeur d'innovation 2023.
  - Caisse d'épargne mon projet innovant 2021.
  - French Tech Tremplin for innovative companies. Launched by people underrepresented in the tech industry.
  - Handitech trophy, awarded by the French Ministry of Health and French Ministry of Digital Technology.

Time of Use

In a day care hospital, the first sessions were habituation sessions of around 5 minutes. The system was used for 3-20 minutes weekly or biweekly, with 272 hours of total use in the ward. The system was always used with a therapist and never alone and planned around the schedule of the children.

Usability of the Product

The SUS was carried out with 9 professionals (3 psychomotricians, 3 nurses, and 3 educators) in July 2022 in a day care hospital. A mean score of 81 out of 100 was obtained (indicating between good and excellent usability), corresponding to a B score in a scale from A (best) to F (worst) [51].

Feasibility of an Efficacy Study

The feasibility of an efficacy study was determined as follows:

1. SD of the outcome measure, which is needed in some cases to estimate the sample size: We could not measure clinical data because of regulations; thus, it is not possible to estimate the sample size. The literature review found that results from a similar device tested on 12 patients supported our clinical impression.
2. Willingness of participants to be randomized: During a focus group in October 2022, parents of children using the chair and 1 autistic adult confirmed their interest in the device and their willingness to participate in an efficacy trial.
3. Willingness of clinicians to recruit participants: Clinicians in a day care hospital and 5 other centers expressed willingness to participate in an efficacy study.
4. Number of eligible patients: With a prevalence of 1%, ASD is quite frequently diagnosed. A lot of children with ASD also have sensory issues. The number of eligible patients seems large enough to conduct an efficacy study.
5. Characteristics of the proposed outcome measure: In the same focus group, in October 2022, the suggested outcome measure (Child Behavior Checklist) was not considered suitable as it was not considered specific enough. Other suggested outcome measures were considered relevant.
6. Adherence and compliance rates: There was good compliance with the device, with only a few dropouts during the early phase of design when the system was too loud for some participants. Most of the questionnaires are already used in clinical practice; others seemed acceptable by the focus group.
7. Availability of data needed: Most of the children in the center have a proper diagnosis. If recruitment is done in other centers, the absence of use of Autism Diagnostic Interview-Revised (ADI-R) and Autism Diagnostic Observation Schedule (ADOS) and poor experience of clinical research could be a limitation.
8. Time needed to collect and analyze data: Getting all the administrative authorization for a medical device in at-risk population (children with ID) and the prospective organization of an efficacy study, planned to run over 12 weeks, may make time management complicated.

Discussion

Overview

The process of development of an ergonomic compressive chair to induce deep pressure in children with ASD is described. The design was user centered according to the methodology of Witteman et al [46]. The system is considered usable by professionals according to the SUS.

User-Centeredness

This device was accepted by the clinicians and patients and their family. The system was primarily used by psychomotor and occupational therapists, but use by nurses and educators was also possible. It did not require the support of a technician. The goal was to increase the acceptability of the device and autonomy of the participant, as well as decrease the stigmatization associated with ASD and its care.

Armchair Use, Child Profile, and Time of Use

According to the experience of the psychomotricians and analysis of video footage by an independent clinician, usability was better for older children (>8 y). There were no side effects reported. Children could leave the armchair easily. If they experienced discomfort, the therapist was able to deflate the cushion.

The system was used in a small room with limited visual and auditory stimulation. Sometimes, professionals suggested children to use a neck cushion to improve relaxation.

Time of use shows that the device was included in everyday care and suggests that it would be routinely adopted in practice.

Future Plans

Acceptability

Ongoing focus groups and simulations with children, parents, and professionals will examine perspectives on sensory issues in ASD and the acceptability of different devices proposed for DPT and sensory therapy. This will provide more formal data on the perspectives of different users on the sensory peculiarities and needs of people with ASD, as well as solutions and the role of this compressive armchair as a therapeutic approach.

Randomized Controlled Trials

This device will enable further evaluation of DPT. In future, we plan to properly characterize and report the data of the individuals including precise diagnostic information (ADI-R and ADOS), their sensory profile and proprioception deficits [16], score on the Echelle des Particularités Sensori-psychomotrices dans l'Autisme (EPSA) scale [52], and underlying pathophysiological processes (eg, heart rate variability and electrodermal activity) using a wearable monitoring device to measure physiological data. To improve acceptability in children with most anxiety, it seems that the presentation of the device should be progressive.

Recruitment for a controlled efficacy study of DPT in ASD seems promising. We have received requests from teams of the original study to be involved in testing the device and expressing willingness to be involved in an efficacy study. Despite complexity of administrative authorizations and time management of a prospective study, such a study seems feasible.

Limitations

Usability Testing

There is a consensus on the need to improve usability of devices in ASD but not on the methods used to measure usability [47]. The SUS is widely used to measure usability, but it can be difficult to use it in individuals with ASD. Thus, it can be amended for use in persons with autism or be filled in by the professional accompanying them [47,53,54]. Usually, the SUS questionnaire is filled in by the person using the system. Here, and in previous studies such as Zhong et al [4], usability was reported by the therapist. Amended versions of current usability tests or the development of alternative means of assessment would improve the assessment of usability.

We think that the time of use and feedback from experts (UCD-11, item 11) reported here and the simulations and focus groups with patients and their parents that we plan to report later are complementary methods that are in favor of a good usability.

Measures of Efficacy

This study does not allow any firm conclusions to be drawn about the efficacy of the device in reducing anxiety in ASD. However, this study showed that an efficacy study is feasible [50]. In future efficacy studies, it would be important to report the precise clinical profiles of children and their pressure needs. Because of the design and preliminary nature of the study, French regulations do not allow the reporting of clinical data.

Implication for Occupational Therapy Practice

This device has the potential to facilitate the design of well-conducted studies to better understand the rationale behind using and the efficacy of DPT in ASD.

Conclusions

We describe the design process, end product, and user feedback after the use of a compressive chair to conduct DPT in children with ASD. This device would allow the children or adults with ASD to better control the pressure and facilitate high-quality studies to understand the rationale behind using (role of proprioception) and the efficacy of DPT in reducing anxiety in children with ASD.

Acknowledgments

We would like to thank John Bost Research Foundation, which funded this study, and Eulalie Arnaud and Suzanne Heron for their careful reading of the text.

Conflicts of Interest

The device described in this paper is the intellectual property of Alexia Audrain.

References

1. Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). Washington, DC: American Psychiatric Association; 2013.
2. Kojovic N, Ben Hadid L, Franchini M, Schaer M. Sensory processing issues and their association with social difficulties in children with autism spectrum disorders. J Clin Med 2019;8(10):1508 [FREE Full text] [doi: 10.3390/jcm8101508] [Medline: 31547076]
3. Asmika A, Oktafiani LDA, Kusworini K, Sujuti H, Andarini S. Autistic children are more responsive to tactile sensory stimulus. Iran J Child Neurol 2018;12(4):37-44 [FREE Full text] [Medline: 30279707]
4. Zhong X, Wang L, Xu L, Lian J, Chen J, Gong X, et al. Disturbance of skin sensation and autism spectrum disorder: a bidirectional Mendelian randomization study. Brain Behav 2023;13(11):e3238 [FREE Full text] [doi: 10.1002/brb3.3238] [Medline: 37670485]
5. Bonnet-Brilhault F, Tuller L, Prévost P, Malvy J, Zebib R, Ferré S, et al. A strategic plan to identify key neurophysiological mechanisms and brain circuits in autism. J Chem Neuroanat 2018;89:69-72. [doi: 10.1016/j.jchemneu.2017.11.007] [Medline: 29128349]
6. Guinchat V, Vlamynck E, Diaz L, Chambon C, Pouzenc J, Cravero C, et al. Compressive garments in individuals with autism and severe proprioceptive dysfunction: a retrospective exploratory case series. Children (Basel) 2020;7(7):77 [FREE Full text] [doi: 10.3390/children7070077] [Medline: 32668622]
7. Blanche EI, Reinoso G, Chang MC, Bodison S. Proprioceptive processing difficulties among children with autism spectrum disorders and developmental disabilities. Am J Occup Ther 2012;66(5):621-624 [FREE Full text] [doi: 10.5014/ajot.2012.004234] [Medline: 22917129]
8. Morris SL, Foster CJ, Parsons R, Falkmer M, Falkmer T, Rosalie SM. Differences in the use of vision and proprioception for postural control in autism spectrum disorder. Neuroscience 2015;307:273-280. [doi: 10.1016/j.neuroscience.2015.08.040] [Medline: 26314635]
9. Fuentes C, Mostofsky SH, Bastian AJ. No proprioceptive deficits in autism despite movement-related sensory and execution impairments. J Autism Dev Disord 2011;41(10):1352-1361 [FREE Full text] [doi: 10.1007/s10803-010-1161-1] [Medline: 21165765]

10. Robertson CE, Baron-Cohen S. Sensory perception in autism. Nat Rev Neurosci 2017;18(11):671-684. [doi: 10.1038/nrn.2017.112] [Medline: 28951611]
11. Case-Smith J, Weaver LL, Fristad MA. A systematic review of sensory processing interventions for children with autism spectrum disorders. Autism 2015;19(2):133-148. [doi: 10.1177/1362361313517762] [Medline: 24477447]
12. Watling R, Hauer S. Effectiveness of ayres sensory integration® and sensory-based interventions for people with autism spectrum disorder: a systematic review. Am J Occup Ther 2015;69(5):6905180030p1-6905180030p12. [doi: 10.5014/ajot.2015.018051] [Medline: 26356655]
13. Lancaster S, Zachry A, Duck A, Harris A, Page E, Sanders J. Delivery of the Wilbarger protocol: a survey of pediatric occupational therapy practitioners. J Occup Ther Sch Early Interv 2016;9(3):281-289 [FREE Full text] [doi: 10.1080/19411243.2016.1169243]
14. Watling R, Miller Kuhaneck H, Parham LD, Schaaf R. Occupational Therapy Practice Guidelines for Children and Youth with Challenges in Sensory Integration and Sensory Processing. AOTA Press: Bethesda, MD; 2018.
15. Green VA, Pituch KA, Itchon J, Choi A, O'Reilly M, Sigafoos J. Internet survey of treatments used by parents of children with autism. Res Dev Disabil 2006;27(1):70-84. [doi: 10.1016/j.ridd.2004.12.002] [Medline: 15919178]
16. Peña M, Ng Y, Ripat J, Anagnostou E. Brief report: parent perspectives on sensory-based interventions for children with autism spectrum disorder. J Autism Dev Disord 2021;51(6):2109-2114 [FREE Full text] [doi: 10.1007/s10803-020-04644-8] [Medline: 32852638]
17. Charleson JL. Effectiveness of weighted blankets as an intervention for sleep problems in children with autism [Thesis]. University of Canterbury. 2024. URL: http://ir.canterbury.ac.nz/bitstream/10092/9515/1/Thesis_fulltext.pdf [accessed 2024-10-21]
18. Gringras P, Green D, Wright B, Rush C, Sparrowhawk M, Pratt K, et al. Weighted blankets and sleep in autistic children--a randomized controlled trial. Pediatrics 2014 Aug;134(2):298-306. [doi: 10.1542/peds.2013-4285] [Medline: 25022743]
19. Eron K, Kohnert L, Watters A, Logan C, Weisner-Rose M, Mehler PS. Weighted blanket use: a systematic review. Am J Occup Ther 2020;74(2):7402205010p1-7402205010p14. [doi: 10.5014/ajot.2020.037358] [Medline: 32204779]
20. Bolic Baric V, Skuthälla S, Pettersson M, Gustafsson PA, Kjellberg A. The effectiveness of weighted blankets on sleep and everyday activities - a retrospective follow-up study of children and adults with attention deficit hyperactivity disorder and/or autism spectrum disorder. Scand J Occup Ther 2023 Nov;30(8):1357-1367 [FREE Full text] [doi: 10.1080/11038128.2021.1939414] [Medline: 34184958]
21. Ekholm B, Spulber S, Adler M. A randomized controlled study of weighted chain blankets for insomnia in psychiatric disorders. J Clin Sleep Med 2020 Sep 15;16(9):1567-1577 [FREE Full text] [doi: 10.5664/jcsm.8636] [Medline: 32536366]
22. Steingrímsson S, Odéus E, Cederlund M, Franzén S, Helgesson C, Nyström K, et al. Weighted blanket and sleep medication use among adults with psychiatric diagnosis - a population-based register study. Nord J Psychiatry 2022 Jan;76(1):29-36. [doi: 10.1080/08039488.2021.1931713] [Medline: 34106812]
23. Mullen B, Champagne T, Krishnamurty S, Dickson D, Gao R. Exploring the safety and therapeutic effects of deep pressure stimulation using a weighted blanket. Occup Ther Ment Health 2008 Mar 06;24(1):65-89. [doi: 10.1300/J004v24n01_05]
24. Becklund AL, Rapp-McCall L, Nudo J. Using weighted blankets in an inpatient mental health hospital to decrease anxiety. J Integr Med 2021 Mar;19(2):129-134. [doi: 10.1016/j.joim.2020.11.004] [Medline: 33317955]
25. Biswas TT, Infirri RS, Hagman S, Berglin L. An assistive sleeping bag for children with autism spectrum disorder. Fash Text 2018 Jul 09;5(1):18. [doi: 10.1186/s40691-018-0133-5]
26. Opsommer E, Dubois J, Bangerter G, Panchaud R, Martin D, Skuza K. Therapeutic body wraps in Swiss public adult acute inpatient wards. a retrospective descriptive cohort study. J Psychiatr Ment Health Nurs 2016 Apr;23(3-4):207-216. [doi: 10.1111/jpm.12304] [Medline: 27126065]
27. Delion P, Labreuche J, Deplanque D, Cohen D, Duhamel A, Lallié C, et al. Therapeutic body wraps (TBW) for treatment of severe injurious behaviour in children with autism spectrum disorder (ASD): A 3-month randomized controlled feasibility study. PLoS One 2018 Jun 29;13(6):e0198726 [FREE Full text] [doi: 10.1371/journal.pone.0198726] [Medline: 29958284]
28. Amaral D, Rogers SJ, Baron-Cohen S, Bourgeron T, Caffo E, Fombonne E, et al. Against le packing: a consensus statement. J Am Acad Child Adolesc Psychiatry 2011 Feb;50(2):191-192. [doi: 10.1016/j.jaac.2010.11.018] [Medline: 21241956]
29. Chamak B. Packing : quand des parents témoignent [Article in French]. Neuropsychiatrie de l'Enfance et de l'Adolescence 2020 Nov;68(7):339-346. [doi: 10.1016/j.neurenf.2020.08.001]
30. Duvall JC, Schleif N, Dunne LE, Holschuh B. Active "hugging" vest for deep touch pressure therapy. 2016 Presented at: 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp 2016; September 12-16, 2016; Heidelberg, Germany p. 458-463. [doi: 10.1145/2968219.2971344]
31. Duvall JC, Schleif N, Dunne LE, Holschuh B. Dynamic compression garments for sensory processing disorder treatment using integrated active materials. J Med Device 2019 Jun;13(2):021001. [doi: 10.1115/1.4042599]
32. Reynolds S, Lane SJ, Mullen B. Effects of deep pressure stimulation on physiological arousal. Am J Occup Ther 2015;69(3):6903350010p1-6903350010p5. [doi: 10.5014/ajot.2015.015560] [Medline: 25871605]
33. Watkins N, Sparling E. The effectiveness of the snug vest on stereotypic behaviors in children diagnosed with an autism spectrum disorder. Behav Modif 2014 May;38(3):412-427. [doi: 10.1177/0145445514532128] [Medline: 24778434]

34. Bestbier L, Williams TI.). The immediate effects of deep pressure on young people with autism and severe intellectual difficulties: demonstrating individual differences. Occup Ther Int 2017;2017:7534972 [FREE Full text] [doi: 10.1155/2017/7534972] [Medline: 29097980]
35. Krauss KE. The effects of deep pressure touch on anxiety. Am J Occup Ther 1987;41(6):366-373. [doi: 10.5014/ajot.41.6.366] [Medline: 3688151]
36. Edelson SM, Edelson MG, Kerr DC, Grandin T. Behavioral and physiological effects of deep pressure on children with autism: a pilot study evaluating the efficacy of grandin's hug machine. Am J Occup Ther 1999;53(2):145-152. [doi: 10.5014/ajot.53.2.145] [Medline: 10200837]
37. Minoura M, Tani I, Ishii T, Gunji YP. Squeezed and released self: using a squeeze machine to degrade the peri-personal space (PPS) boundary. Psychol Conscious 2021 Sep;8(3):248-257. [doi: 10.1037/cns0000223]
38. Lo JS, Huang SC. Creative design of sitting hug machine in the treatment of students with autism. MATEC Web Conf 2018;213(1):01009 [FREE Full text] [doi: 10.1051/matecconf/201821301009]
39. Afif IY, Maula MI, Aliyafi MB, Aji AL, Winarni TI, Jamari J. Design of hug machine portable seat for autistic children in public transport application. IOP Conf Ser: Mater Sci Eng 2021;1096(1):012034 [FREE Full text]
40. Maula MI, Aji A, Aliyafi M, Afif I, Ammarullah M, Winarni T, et al. The subjective comfort test of autism hug machine portable seat. J Intellect Disabl Diagn Treat 2021;9(3):182-188 [FREE Full text] [doi: 10.6000/2292-2598.2021.09.02.4]
41. Grandin T. Calming effects of deep touch pressure in patients with autistic disorder, college students, and animals. J Child Adolesc Psychopharmacol 1992;2(1):63-72. [doi: 10.1089/cap.1992.2.63] [Medline: 19630623]
42. Lane SJ, Mailloux Z, Schoen S, Bundy A, May-Benson TA, Parham LD, et al. Neural Foundations of Ayres Sensory Integration®. Brain Sci 2019;9(7):153 [FREE Full text] [doi: 10.3390/brainsci9070153] [Medline: 31261689]
43. OTO chair, Alexia Audrain. LABAA. URL: https://www.oto-chair.com/ [accessed 2024-10-03]
44. OTO chair demonstration, Alexia Audrain. LABAA. URL: https://www.canva.com/design/DAFh3Z7wlZs/lBsbLYfidBARmp_8_sRH8g/watch [accessed 2024-10-03]
45. Defense Acquisition Guidebook. VA, USA: DOD; 2006.
46. Witteman H, Vaisson G, Provencher T, Dansokho SC, Colquhoun H, Dugas M, et al. Development and validation of UCD-11: an 11-item measure of user-centered design for patient-centered tools. OSF Preprints 2019:1-21 [FREE Full text]
47. Aguiar YPC, Galy E, Godde A, Trémaud M, Tardif C. AutismGuide: a usability guidelines to design software solutions for users with autism spectrum disorder. Behav Inform Technol 2020;41(6):1132-1150. [doi: 10.1080/0144929x.2020.1856927]
48. Brooke J. SUS—a quick and dirty usability scale. In: Usability Evaluation in Industry. Boca Raton, FL: CRC Press; 1996.
49. Weiss PL, Gal E, Zancanaro M, Giusti L, Cobb S, Millen L, et al. Usability of technology supported social competence training for children on the autism spectrum. 2011 Presented at: 2011 International Conference on Virtual Rehabilitation; 2011 June 27-29; Zurich, Switzerland p. 1-8.
50. Eldridge SM, Lancaster GA, Campbell MJ, Thabane L, Hopewell S, Coleman CL, et al. Defining feasibility and pilot studies in preparation for randomised controlled trials: development of a conceptual framework. PLoS One 2016;11(3):e0150205 [FREE Full text] [doi: 10.1371/journal.pone.0150205] [Medline: 26978655]
51. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123.
52. Le Menn-Tripi C, Vachaud A, Defas N, Malvy J, Roux S, Bonnet-Brilhault F. L'évaluation sensori-psychomotrice dans l'autisme : un nouvel outil d'aide au diagnostic fonctionnel sensory-psychomotor evaluation in autism: a new tool for functional diagnosis. L'Encéphale 2019;45(4):312-319. [doi: 10.1016/j.encep.2018.12.003]
53. Gentile V, Adjorlu A, Serafin S, Rocchesso D, Sorce S. Touch or touchless?: evaluating usability of interactive displays for persons with autistic spectrum disorders. 2019 Presented at: Proceedings of the 8th ACM International Symposium on Pervasive Displays; 2019 June 12; Palermo, Italy p. 1-7 URL: https://doi.org/10.1145/3321335.3324946
54. Parish-Morris J, Solórzano R, Ravindran V, Sazawal V, Turnacioglu S, Zitter A, et al. Immersive virtual reality to improve police interaction skills in adolescents and adults with autism spectrum disorder: preliminary results of a phase i feasibility and safety trial. Annu Rev CyberTherapy Telemed 2018;16:50-56.

Abbreviations

ADI-R: Autism Diagnostic Interview-Revised
ADOS: Autism Diagnostic Observation Schedule
ASD: autism spectrum disorder
DPT: deep pressure therapy
EPSA: Echelle des Particularités Sensori-psychomotrices dans l'Autisme
ID: intellectual deficiency
SUS: System Usability Scale
UCD-11: Human-Centered Design for Personal Health Tools

Edited by A Parush; submitted 22.12.23; peer-reviewed by E Blanche, J Brooke; comments to author 30.01.24; revised version received 23.04.24; accepted 16.07.24; published 05.11.24.

Please cite as:
Gargot T, Vachaud A, Gilard C, Audrain A, Gomot M, Guidotti M, Briend F, Malvy J, Bonnet Brilhault F
A Compressive Armchair (OTO) to Perform Deep Pressure Therapy in Children With Autism Spectrum Disorder: User-Centered Design and Feasibility Study
JMIR Hum Factors 2024;11:e55754
URL: https://humanfactors.jmir.org/2024/1/e55754
doi:[10.2196/55754](https://doi.org/10.2196/55754)
PMID:

The Use of Mobile Health Care Among Medical Professionals in the Sichuan-Chongqing Region: Cross-Sectional Survey Study

Yan Tang[1], MD; Juan Yang[2], MD; Ni Wang[3], MD; Xin Wang[1], MBBS; Wenli Hu[1], MBBS

[1]Department of Neurology, Chongqing Red Cross Hospital (People’s Hospital of Jiangbei District), Chongqing, China

[2]Department of Neurology, The Third Affiliated Hospital of Chongqing Medical University, No. 1 Shuanghu Branch Road, Huixing Street, Yubei District, Chongqing, China

[3]Department of Neurology, JiuLongPo Distict People 's Hospital, Chongqing, China

Corresponding Author:
Juan Yang, MD
Department of Neurology, The Third Affiliated Hospital of Chongqing Medical University, , No. 1 Shuanghu Branch Road, Huixing Street, Yubei District, Chongqing, , China

Abstract

Background: The emergence and integration of mobile health care technology have fundamentally transformed the health care industry, providing unprecedented opportunities to improve health care services and professional practice. Despite its immense potential, the adoption of mobile health care technology among health care professionals remains uneven, particularly in resource-limited regions.

Objective: This study aims to explore the use and influencing factors of mobile health care among health care professionals in the Sichuan-Chongqing region of China and make recommendations.

Methods: Convenience sampling was used in a cross-sectional study conducted from November 8 to November 14, 2023, to survey frontline clinical health care professionals at 5 district-level secondary public hospitals in the Sichuan-Chongqing region. A web-based questionnaire was used to investigate the use of mobile health care and its influencing factors among the participants. Descriptive analysis and logistic regression analysis were used in the study.

Results: A total of 550 valid questionnaires were completed. Among the surveyed health care professionals, only 18.7% (103/550) used mobile health care, with a satisfaction rate of only 50.5% (52/103). Around 81.3% (447/550) did not use any form of mobile health care. The age group of 30 - 39 years was found to be a significant factor influencing the use of mobile health care by health care professionals (*P*=.03). The main reasons for not using mobile health care among health care professionals were lack of appropriate technical training and support (266/447, 59.5%), lack of suitable management-specific apps (204/447, 45.6%), and concerns about increased workload (180/447, 40.3%). There were significant differences in the single-factor analysis of the reasons for the nonuse of mobile health care among health care professionals from different specialties (*P*=.04). Logistic regression analysis indicated that age was the only significant factor influencing the use of mobile health care by health care professionals (*P*=.04).

Conclusions: The utilization rate of mobile health care among health care professionals in the Sichuan-Chongqing region is low. Age is a significant factor that influences whether health care professionals use mobile health care. Providing appropriate technical training and support may help improve the enthusiasm of health care professionals in using mobile health care.

(JMIR Hum Factors 2024;11:e59153) doi:10.2196/59153

KEYWORDS

health care professionals; mobile health care; technical training; cross-sectional survey; utilization; mobile; usage; China; web-based questionnaire; logistic regression; training; support

Introduction

Currently, with the rapid development of internet technology both domestically and internationally, mobile health, mobile internet, health management, and medical system informatization have become hot research topics in China in recent years [1]. Internet health care refers to the integration of medical information and interconnectedness, encompassing various forms of health care services such as medical information, electronic health records, disease risk assessment, web-based disease consultation, telemedicine, and rehabilitation. Mobile health care, on the other hand, combines internet health care with mobile apps, focusing on medical and health-related apps. As a new health care model actively encouraged by the government, mobile health care can better meet the increasing health care demands of the people and contribute to the realization of the “Healthy China” strategic goals.

In recent years, the aging process of the population in China has been accelerating, leading to an increase in the number of chronic disease patients. The application of mobile health care technology in the field of patient management is increasingly showing its key role, especially in improving work efficiency and patient satisfaction [2]. Traditional health care models face challenges due to factors like geography, economy, transportation, and facilities, complicating patient education and disease management for health care professionals. In contrast, mobile health care eliminates these barriers, enabling health care professionals to continuously track patient conditions throughout the entire disease management cycle, thus demonstrating its wide applicability and significant promotional value. In addition, the application of mobile health care technology not only enhances the accessibility and continuity of health care services but also provides patients with more personalized and convenient medical experiences, further promoting the innovation and progress of health care service models. However, in China, especially in rural areas, the grassroots medical and health information network is still not well established. Large level-III hospitals focus on the diagnosis and treatment of difficult and critical cases, leaving little energy for patient follow-up management. Therefore, we believe that using mobile health care for patient management deserves more attention and implementation from secondary general hospitals.

Previous research on mobile health care in China has primarily focused on the construction of telemedicine platforms and the development and application of mobile health care information systems [3,4], or on patient use patterns [5-7]. However, it has largely overlooked the fact that, within the realm of mobile health care, the acceptance and use by health care professionals are crucial determinants of patient engagement. Therefore, investigating both the adoption patterns and the factors influencing health care professionals' use of mobile health care technologies is essential.

From the perspective of the growth rate of the mobile health care market in China, the national growth rate has slowed down, and users in major cities have gradually become accustomed to using mobile health care. In contrast, in the southwestern region, particularly the Sichuan-Chongqing area—a key medical hub—the use of mobile health care by professionals not only affects the local service quality but also impacts the broader regional health care landscape. Despite this importance, studies focusing on this demographic within the region are absent. To address this gap, this study surveys the current use and influencing factors of mobile health care among health care professionals in the Sichuan-Chongqing area and provides recommendations. This study is not just about mapping the current use of these technologies. It is about understanding the factors that encourage or deter health care professionals from adopting mobile health tools in their daily practice. By focusing on this underexplored area, our findings aim to inform more than just policy—it is about enabling better health outcomes and shaping a future where health care reaches everyone efficiently and effectively in the region.

Methods

Study Design and Participants

This was a cross-sectional survey study conducted using convenience sampling. From November 8 to November 14, 2023, a web-based questionnaire survey was conducted among frontline health care professionals. This study investigated 5 district-level secondary public hospitals based on geographic location and population density (million-level population density): 3 located in central urban areas, specifically Chongqing Jiangbei District People's Hospital, Chongqing Jiulongpo District Second People's Hospital, and Sichuan Chengdu Chenghua District Third People's Hospital; and 2 situated in more distant districts, namely Chongqing Shapingba District Chenjiqiao Hospital and Sichuan Anyue County Third People's Hospital. Prior to the survey, all participants were informed that their participation in this survey was voluntary, and they were informed of the purpose and significance of this survey research. The questionnaire survey was completed anonymously using the platform QuestionStar. All data were anonymized.

Data Collection

This study used a self-designed questionnaire survey, which included the following aspects: (1) general information about the participants: sex, age, education background, professional title, occupation, and major; (2) whether they use mobile health care; (3) use of mobile health care: the name of the tool used, its functions, and frequency of use; (4) factors influencing the users: work efficiency, patient compliance, health outcomes, advantages, challenges, satisfaction, and areas for improvement; and (5) reasons for nonusers: lack of appropriate training and technical support, concerns about patient data privacy and security issues, concerns about increased workload due to web-based apps, lack of suitable apps for specific management purposes, negative or unwilling patient response, and lack of awareness of the potential benefits of web-based apps.

Statistical Analysis

All data were organized and analyzed using SPSS software (version 18.0; IBM Corporation). Continuous and normally distributed metric data were presented as means and SDs, while nonnormally distributed metric data were presented as medians and IQRs. Categorical data were presented as frequencies and percentages. A chi-square test was used to analyze the correlation between the general characteristics of the participants and the use of mobile health care, as well as the reasons for nonuse. Multivariable logistic regression analysis was conducted to identify the factors influencing the use of mobile health care. A significance level of P<.05 was considered statistically significant. The reliability of the self-designed questionnaire was assessed using the Cronbach α.

Ethical Considerations

The study was discussed by the Ethics Committee of Chongqing Red Cross Hospital (People's Hospital of Jiangbei District), which determined that the research involves the use of human information data but does not cause physical harm, nor does it involve sensitive personal information or commercial interests. Consequently, this study was exempted from ethical review [8].

Results

Background Characteristics

The participants in this survey were identified as frontline clinical doctors or nurses, interns, and residents. A small sample questionnaire test was conducted first to assess the reliability of the self-made questionnaire, which showed good reliability (Cronbach α=.92). The final formal questionnaire was then determined, and a large sample was collected. A total of 556 health care professionals completed the survey, with 550 valid questionnaires, resulting in a valid response rate of 98.9%. The age of the participants ranged from 18 to 59 years, with an average age of 33.24 (SD 7.02) years. Females accounted for 77.1% (424/550) of the participants. A total of 82% (451/550) of the medical personnel had received undergraduate or higher education, 30.2% (166/550) were nurses, and 64.4% (354/550) were doctors (Table 1).

Table . Background characteristics of respondents.

Variable	Values
Sex, n (%)	
Male	126 (22.9)
Female	424 (77.1)
Age (years), mean (SD)	33.2 (7.0)
Age (years), n (%)	
<30	171 (31.1)
30 - 39	294 (53.5)
>40	85 (15.5)
Education background, n (%)	
Junior college or below	99 (18.0)
Undergraduate	280 (50.9)
Postgraduate or above	171 (31.1)
Professional title, n (%)	
Junior and below	214 (38.9)
Intermediate	257 (46.7)
Senior	79 (14.4)
Occupation, n (%)	
Physician	354 (64.4)
Nurse	166 (30.2)
Technician	18 (3.3)
Others	12 (2.2)
Major, n (%)	
Clinical department	445 (80.9)
Ancillary departments	50 (9.1)
Public health and others	55 (10.0)

Specific Use of Mobile Health Care

According to the survey, only 18.7% (103/550) of health care professionals use mobile health care. The most commonly used platforms include “PDA, Yi Doctor, Yi Nurse, and WeChat,” accounting for 44.7% (46/103) of the total. There are also various other mobile health care platforms such as “Micro Doctor, Good Doctor, 317 Nurse, JD Doctor, Creative Doctor, Doctor’s Palm, Doctor’s Circle,” and so on. The main function used is patient record management, accounting for 78.6% (81/103). The use frequency is at least once a day, accounting for 57.3% (59/103). The majority of respondents believe that using mobile health care can improve work efficiency, improve patient compliance, and enhance patient health. However, the satisfaction rate is only 50.5% (52/103), with the main areas for improvement being data security and training support (Table 2).

Table . Use of mobile health care.

Use status	Values, n (%)
Function used	
Patient record management	81 (78.6)
Appointment management	37 (35.9)
Prescription management	44 (42.7)
Communication and consultation	49 (47.6)
Disease monitoring and tracking	55 (53.4)
Other	13 (12.6)
Use frequency	
≥1 time per day	59 (57.3)
≥1 time per week	23 (22.3)
≥1 time per month	6 (5.8)
Occasionally	15 (14.6)
Impact on work efficiency	
Improve efficiency	80 (77.7)
No significant impact	17 (16.5)
Decrease efficiency	2 (1.9)
Unclear	4 (3.9)
Effect on patient compliance	
Yes	64 (62.1)
No	16 (15.5)
Uncertain	23 (22.3)
Effect on patient health	
Yes	67 (65)
No	13 (12.6)
Uncertain	23 (22.3)
Advantages	
Improve self-management ability	64 (62.1)
Save time	76 (73.8)
Provide real-time data	76 (73.8)
Increase communication frequency	66 (64.1)
Challenges	
Privacy and security	78 (75.7)
Technical barriers	67 (65.0)
Patient response	50 (48.5)
Satisfaction	
Satisfied	52 (50.5)
Average	47 (45.6)
Dissatisfied	4 (3.9)
Areas for improvement	
Training support	73 (70.9)
Data security	74 (71.8)
User-friendliness	70 (68.0)

Use status	Values, n (%)
Increase functionality	64 (62.1)

Univariate Analysis for Whether to Use Mobile Health Care

Around 81.3% (447/550) of surveyed health care professionals did not use any mobile health care. Age, particularly the 30 - 39 years age group, was found to be a significant factor influencing whether health care professionals use mobile health care (*P*=.03). Other factors such as gender, education level, professional title, occupation, and specialty did not have a significant impact on the use of mobile health care among health care professionals, with no statistically significant differences observed (Table 3).

Table . Univariate analysis of factors influencing the use of mobile health care among medical personnel.

Factor	Used (n=103)	Not used (n=447)	Chi-square (*df*)	*P* value
Sex, n (%)			0.024 (1)	.88
Male	23 (4.2)	103 (18.7)		
Female	80 (14.5)	344 (62.5)		
Age (years), n (%)			6.791 (2)	.03
<30	21 (3.8)	150 (27.3)		
30 - 39	64 (11.6)	230 (41.8)		
≥40	18 (3.3)	67 (12.2)		
Education background, n (%)			4.47 (2)	.11
Junior college or below	16 (2.9)	83 (15.1)		
Undergraduate	62 (11.3)	218 (39.6)		
Postgraduate or above	25 (4.5)	146 (26.5)		
Professional title, n (%)			4.067 (2)	.13
Junior and below	32 (5.8)	182 (33.1)		
Intermediate	57 (10.4)	200 (36.4)		
Senior	14 (2.5)	65 (11.8)		
Occupation, n (%)			1.73 (2)	.42
Physician	63 (11.5)	291 (52.9)		
Nurse	36 (6.5)	130 (23.6)		
Technician and others	4 (0.7)	26 (4.7)		
Major, n (%)			4.638 (2)	.10
Clinical department	90 (16.4)	355 (64.5)		
Ancillary departments	4 (0.7)	46 (8.4)		
Public health and others	9 (1.6)	46 (8.4)		

Multivariable Logistic Regression Analysis for Whether to Use Mobile Health Care

Multivariable logistic regression analysis was conducted with health care professionals' use of mobile health care as the dependent variable (1 for use, 0 for nonuse) and gender, age, education, professional title, occupation, and major as independent variables. The results showed that age was the only significant factor influencing the use of mobile health care by health care professionals (*P*=.04) (Table 4).

Table . Multivariable logistic regression analysis of factors influencing health care professionals' use of mobile health care.

Independent variables	B[a]	SE	Wald chi-square (*df*)	OR[b] (95% CI)	*P* value
Sex	0.053	0.272	0.038 (1)	1.054 (0.619 - 1.796)	.85
Age	0.5	0.237	4.439 (1)	1.648 (1.035 - 2.624)	.04
Education background	−0.278	0.201	1.91 (1)	0.757 (0.511 - 1.123)	.17
Professional title	−0.089	0.195	0.207 (1)	0.915 (0.625 - 1.340)	.65
Occupation	−0.03	0.191	0.025 (1)	0.971 (0.668 - 1.410)	.88
Major	−0.105	0.066	2.482 (1)	0.901 (0.791 - 1.026)	.12

[a]B: regression coefficient.

[b]OR: odds ratio.

Analysis of Reasons for Not Using Mobile Health Care

The common reasons identified through the presurvey for healthcare professionals not using mobile health are as follows: lack of appropriate technical training and support (266/447, 59.5%), lack of suitable management for specific applications (204/447, 45.6%), concerns about increased workload (180/447, 40.3%), concerns about patient data privacy and security (165/447, 36.9%), ambiguity regarding the potential benefits of web-based applications (164/447, 36.7%), negative reactions or unwillingness from patients (135/447, 30.2%), and other reasons (74/447, 16.6%).

Main Reasons for Health Care Professionals Not Using Mobile Health Care

Single-factor analysis of factors influencing health care professionals' nonuse of mobile health care showed no significant differences in reasons for nonuse based on sex, age, educational background, professional title, and occupation. However, there were certain differences in reasons for nonuse among different majors (*P*=.04). The main reasons for nonuse among clinical department personnel were lack of appropriate technical training and support, lack of suitable management apps for specific types, and concerns about increased workload. The main reasons for nonuse among ancillary department personnel were lack of appropriate technical training and support, lack of awareness of potential benefits of web-based apps, and lack of understanding of potential benefits of web-based apps. The main reasons for nonuse among public health and other department personnel were lack of appropriate technical training and support, lack of suitable management apps for specific types, concerns about patient data privacy and security, and lack of awareness of potential benefits of web-based apps (Table 5).

Table . Single-factor analysis of reasons for health care professionals' nonuse of mobile health care.

	Lack of appropriate technical training and support	Concerns about patient data privacy and security	Concerns about increased workload	Lack of suitable management-specific apps	Patients have negative reactions or are unwilling	Not clear about the potential benefits of web-based apps	Others	*P* value
Sex, n (%)								.75
Male	60 (58.3)	40 (38.8)	45 (43.7)	44 (42.7)	26 (25.2)	41 (39.8)	13 (12.6)	
Female	206 (59.9)	125 (36.3)	135 (39.2)	160 (46.5)	109 (31.7)	123 (35.8)	61 (17.7)	
Age, n (%)								.96
<30	81 (54.0)	52 (34.7)	59 (39.3)	56 (37.3)	42 (28.0)	56 (37.3)	28 (18.7)	
30 - 39	143 (62.2)	86 (37.4)	94 (40.9)	117 (50.9)	76 (33.0)	84 (36.5)	37 (16.1)	
≥40	42 (62.7)	27 (40.3)	27 (40.3)	31 (46.3)	17 (25.4)	24 (35.8)	9 (13.4)	
Education background, n (%)								.91
Junior college or below	46 (55.4)	31 (37.3)	31 (37.3)	34 (41.0)	28 (33.7)	30 (36.1)	17 (20.5)	
Undergraduate	131 (60.1)	82 (37.6)	90 (41.3)	107 (49.1)	66 (30.3)	90 (41.3)	40 (18.3)	
Postgraduate or above	89 (61.0)	52 (35.6)	59 (40.4)	63 (43.2)	41 (28.1)	44 (30.1)	17 (11.6)	
Professional title, n (%)								.99
Junior and below	102 (56.0)	69 (37.9)	78 (42.9)	80 (44.0)	52 (28.6)	70 (38.5)	32 (17.6)	
Intermediate	127 (63.5)	72 (36.0)	79 (39.5)	93 (46.5)	66 (33.0)	70 (35.0)	32 (16.0)	
Senior	37 (56.9)	24 (36.9)	23 (35.4)	31 (47.7)	17 (26.2)	24 (36.9)	10 (15.4)	
Occupation, n (%)								.09
Physician	173 (59.5)	112 (38.5)	118 (40.5)	139 (47.8)	87 (29.9)	99 (34.0)	41 (14.1)	
Nurse	84 (64.6)	47 (36.2)	56 (43.1)	60 (46.2)	43 (33.1)	55 (42.3)	22 (16.9)	
Technician	9 (34.6)	6 (23.1)	6 (23.1)	5 (19.2)	5 (19.2)	10 (38.5)	11 (42.3)	
Major, n (%)								.04
Clinical department	221 (62.3)	132 (37.2)	148 (41.7)	170 (47.9)	114 (32.1)	130 (36.6)	46 (13.0)	
Ancillary departments	19 (41.3)	14 (30.4)	16 (34.8)	17 (37.0)	9 (19.6)	17 (37.0)	15 (32.6)	
Public health and others	26 (56.5)	19 (41.3)	16 (34.8)	17 (37.0)	12 (26.1)	17 (37.0)	13 (28.3)	

Discussion

Principal Findings

The global COVID-19 pandemic has greatly increased the use of mobile health care in many countries, especially among populations at high risk, such as those with chronic underlying conditions [9]. The demand for mobile health care has surged, but this study found that the use of mobile health care among frontline health care professionals in the Sichuan-Chongqing region is generally poor, with a use rate of only 18.7% (103/550). Similarly, a study by Zhu et al [10] found that only 12.6% of health care professionals have used smart apps. Furthermore, we found that the satisfaction rate among health care professionals who use mobile health care is only 50.5% (52/103), indicating a need for significant improvement in training support and data security. Among nonusers, there are some differences in reasons among different professions, but the main reasons include the lack of appropriate technical training and support, as well as the lack of specific management apps. Therefore, the most prominent issue for surveyed health care professionals, regardless of whether they use mobile health care or not, is the lack of technical training and support.

It is necessary for health care institutions to provide relevant training for health care professionals in using mobile health care. Research has shown that 95% of respondents had not received any training before using mobile health [11], and leadership support and training for health care professionals are crucial in providing mobile health services [12,13]. In this study, the most commonly used mobile health care by health care professionals were “PDA, Yi Doctor, Yi Nurse, and WeChat,” all of which benefited from the unified requirements of the hospital leadership. Other health care professionals also used apps such as “Good Doctor, WeDoctor, 317 Nurse” on their own. A total of 75.7% (78/103) of respondents believed that there are limitations and risks in using personal mobile devices, and the privacy and data security of patients need to be protected. This is similar to the findings of Rowe-Setz et al [14]. Therefore, the popularization of mobile health care requires leadership to lead the development of relevant policies and guidelines, provide practical training for employees, and ensure the smooth and efficient use of technical devices. The content should include specific coverage of the following aspects: (1) basic operations of mobile health care, data security, privacy protection, and how it can help doctors enhance their value and improve their professional skills; (2) provision of real-life cases and user feedback to demonstrate how mobile health care can help health care professionals save time and energy, and enhance job satisfaction; (3) ensuring good integration of mobile apps with existing hospital information systems to avoid duplicate data entry and operational redundancy, thereby reducing workload; (4) establishment of a technical support hotline or web-based platform to promptly address issues and concerns encountered by health care professionals during app use; and (5) in addition to regular training courses for health care professionals, establishment of various guidelines to support mobile health care activities. When mobile health care services are regularly scheduled, they become part of daily work, just like any other familiar operational aspect in the workplace, thereby improving the technical capabilities of professionals. Implementing the above suggestions may help health care professionals overcome barriers to using mobile health care, enhance their willingness and ability to adopt mobile technologies, and promote the digital transformation of health care services and the improvement of health care quality.

Although 81.3% (447/550) of health care professionals surveyed did not use any mobile health services, this does not mean that they lack the willingness to use them. Research has shown that health care professionals have a high acceptance of mobile health services [15]. The lack of use despite the intention is based on the consideration of costs and benefits, which are referred to as perceived benefits. It is also mentioned that building a personal brand for doctors has a positive impact on their willingness to use mobile health care [16,17]. However, there are some health care professionals who hold negative and skeptical attitudes and do not want to acquire the new skills required for mobile health care. Leaders need to try to influence their attitudes and the atmosphere of the workplace to make their use of mobile health care more positive.

The proliferation of mobile health care apps has provided users with more choices, but the wide variety and varying quality of these apps have increased the difficulty for health care professionals to select the right app for managing specific diseases. There have been studies exploring the feasibility of using mobile health care for chronic disease management [18,19]. Therefore, it is important to focus on developing personalized mobile apps tailored to specific medical fields such as hypertension, diabetes, and stroke for effective chronic disease management. The design of mobile health care programs should prioritize the involvement of health care professionals [20], enabling them to better manage patients; promote a shift from doctor-patient relationships to partnerships; and facilitate interdisciplinary team collaboration, patient education, and information sharing across different health care institutions for the same disease. This approach aims to meet the specific needs of health care professionals and achieve the goal of comprehensive management with a patient-centered approach for long-term follow-up care.

Pay attention to seed users and fully tap into the potential user base. Among Chinese mobile phone users, young users account for over 90%. The results of this study show that age, especially the 30 - 39 years age group, is a significant factor influencing whether health care professionals use mobile health care. The middle-aged and young user group is large in size, has a strong demand for medical and health services, is more receptive to new things, and has a higher interest in smart apps [21,22]. This suggests that focusing on health care professionals in the 30 - 39 years age group as seed users for targeted attention and promotion may achieve more significant results.

Limitations

In this study, a convenience sampling method was used, limited to health care professionals in secondary public hospitals in the Sichuan-Chongqing region, which may not fully reflect the situation of health care professionals in other regions or hospitals of different levels. In future research, random sampling or stratified sampling methods could be considered to improve the representativeness of the sample. Additionally, efforts should be made to expand the scope of the study to include more regions and different types of hospitals in order to obtain more generalizable research results.

Conclusion

This survey reveals that frontline health care professionals in the Sichuan-Chongqing region have low use of mobile health care, primarily due to a lack of training support and concerns about data security. Health care professionals have the intention to use it but need to address the cost-benefit issue. It is recommended that leadership provide training support, establish policy guidelines, influence attitudes, and create a positive atmosphere. Emphasis should be placed on the development of personalized applications in the field of chronic disease management, with a focus on the participation of young health care professionals. In summary, efforts should be concentrated on improving training, addressing security concerns, and strengthening leadership guidance to promote the application of mobile health care in the Sichuan-Chongqing region.

Acknowledgments

Thanks are due to Jun Qian and Jian Wang for their valuable assistance with the work. This study was supported by the Chongqing Science and Health joint medical research project (2022MSXM121).

Authors' Contributions

YT, NW, XW, and WLH carried out the studies, participated in collecting data, and drafted the manuscript. YT and JY performed the statistical analysis and participated in its design. All authors participated in the acquisition, analysis, or interpretation of data and drafting of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Rezende LC, Ribeiro EG, Parreiras LC, et al. Telehealth and telemedicine in the management of adult patients after hospitalization for COPD exacerbation: a scoping review. J Bras Pneumol 2023;49(3):e20220067. [doi: 10.36416/1806-3756/e20220067] [Medline: 37132694]
2. Antonacci G, Benevento E, Bonavitacola S, et al. Healthcare professional and manager perceptions on drivers, benefits, and challenges of telemedicine: results from a cross-sectional survey in the Italian NHS. BMC Health Serv Res 2023 Oct 18;23(1):1115. [doi: 10.1186/s12913-023-10100-x] [Medline: 37853448]
3. Noorbergen TJ, Adam MTP, Teubner T, Collins CE. Using co-design in mobile health system development: a qualitative study with experts in co-design and mobile health system development. JMIR Mhealth Uhealth 2021 Nov 10;9(11):e27896. [doi: 10.2196/27896] [Medline: 34757323]
4. Chen X, Wu X, Zhang Q, et al. The construction and operational models of internet hospitals in China: a hospital-based survey study. BMC Health Serv Res 2023 Jun 21;23(1):669. [doi: 10.1186/s12913-023-09675-2] [Medline: 37344831]
5. Krishnaswami A, Beavers C, Dorsch MP, et al. Gerotechnology for older adults with cardiovascular diseases: JACC state-of-the-art review. J Am Coll Cardiol 2020 Dec 1;76(22):2650-2670. [doi: 10.1016/j.jacc.2020.09.606] [Medline: 33243384]
6. Gimbel RW, Rennert LM, Crawford P, et al. Enhancing patient activation and self-management activities in patients with type 2 diabetes using the US department of defense mobile health care environment: feasibility study. J Med Internet Res 2020;22(5):e17968. [doi: 10.2196/17968]
7. Walle AD, Ferede TA, Shibabaw AA, et al. Willingness of diabetes mellitus patients to use mHealth applications and its associated factors for self-care management in a low-income country: an input for digital health implementation. BMJ Health Care Inform 2023 May;30(1):e100761. [doi: 10.1136/bmjhci-2023-100761]
8. Measures for the ethical review of life science and medical research involving human beings. Department of Science and Technology Education of China. URL: http://www.nhc.gov.cn/qjjys/s7946/202302/c3374c180dc5489d85f95df5b46afaf5.shtml#\ [accessed 2024-09-11]
9. Vergara J, Parish A, Smallheer B. Telehealth: opportunities in geriatric patient care during COVID-19. Geriatr Nurs 2020;41(5):657-658. [doi: 10.1016/j.gerinurse.2020.08.013] [Medline: 32912647]
10. Zhu J, Weng H, Ou P, Li L. Use and acceptance of smart elderly care apps among Chinese medical staff and older individuals: web-based hybrid survey study. JMIR Form Res 2023 Jun 13;7:e41919. [doi: 10.2196/41919] [Medline: 37310777]
11. Elawady A, Khalil A, Assaf O, Toure S, Cassidy C. Telemedicine during COVID-19: a survey of health care professionals' perceptions. Monaldi Arch Chest Dis 2020 Sep 22;90(4):32959627. [doi: 10.4081/monaldi.2020.1528] [Medline: 32959627]
12. Myllymäki S, Laukka E, Kanste O. Health and social care frontline leaders' perceptions of competence management in telemedicine in Finland: an interview study. J Nurs Manag 2022 Oct;30(7):2724-2732. [doi: 10.1111/jonm.13740] [Medline: 35852809]
13. Odendaal WA, Anstey Watkins J, Leon N, et al. Health workers' perceptions and experiences of using mHealth technologies to deliver primary healthcare services: a qualitative evidence synthesis. Cochrane Database Syst Rev 2020 Mar 26;3(3):CD011942. [doi: 10.1002/14651858.CD011942.pub2] [Medline: 32216074]
14. Rowe-Setz G, Behringer W, Roland D, Lyttle MD, Jahn HK. Fifteen minute consultation: when can I use a medical app? Arch Dis Child Educ Pract Ed 2021 Oct;106(5):258-263. [doi: 10.1136/archdischild-2020-319448] [Medline: 32948634]
15. Barteit S, Neuhann F, Bärnighausen T, et al. Technology acceptance and information system success of a mobile electronic platform for nonphysician clinical students in Zambia: prospective, nonrandomized intervention study. J Med Internet Res 2019 Oct 9;21(10):e14748. [doi: 10.2196/14748] [Medline: 31599731]
16. Zhai Y, Song X, Chen Y, Lu W. A study of mobile medical app user satisfaction incorporating theme analysis and review sentiment tendencies. Int J Environ Res Public Health 2022 Jun 17;19(12):7466. [doi: 10.3390/ijerph19127466] [Medline: 35742713]
17. Kim HS. Towards telemedicine adoption in Korea: 10 practical recommendations for physicians. J Korean Med Sci 2021 May 3;36(17):e103. [doi: 10.3346/jkms.2021.36.e103] [Medline: 33942575]

18. Steinman L, Heang H, van Pelt M, et al. Facilitators and barriers to chronic disease self-management and mobile health interventions for people living with diabetes and hypertension in Cambodia: qualitative study. JMIR Mhealth Uhealth 2020 Apr 24;8(4):e13536. [doi: 10.2196/13536] [Medline: 32329737]
19. Quach S, Michaelchuk W, Benoit A, et al. Mobile heath applications for self-management in chronic lung disease: a systematic review. Netw Model Anal Health Inform Bioinform 2023;12(1):25. [doi: 10.1007/s13721-023-00419-0] [Medline: 37305790]
20. Pokhrel P, Karmacharya R, Taylor Salisbury T, et al. Perception of healthcare workers on mobile app-based clinical guideline for the detection and treatment of mental health problems in primary care: a qualitative study in Nepal. BMC Med Inform Decis Mak 2021 Jan 19;21(1):21. [doi: 10.1186/s12911-021-01386-0] [Medline: 33468120]
21. Chowdhury SZ, Stevens S, Wu C, et al. An age-old problem or an old-age problem? a UK survey of attitudes, historical use and recommendations by healthcare professionals to use healthcare apps. BMC Geriatr 2023 Feb 24;23(1):110. [doi: 10.1186/s12877-023-03772-x] [Medline: 36823564]
22. Jabour AM, Rehman W, Idrees S, Thanganadar H, Hira K, Alarifi MA. The adoption of mobile health applications among university students in health colleges. J Multidiscip Healthc 2021;14:1267-1273. [doi: 10.2147/JMDH.S310539] [Medline: 34103927]

Edited by A Kushniruk, Y Quintana; submitted 03.04.24; peer-reviewed by SUD Khan, Z Guo; revised version received 18.08.24; accepted 05.09.24; published 24.10.24.

Please cite as:
Tang Y, Yang J, Wang N, Wang X, Hu W
The Use of Mobile Health Care Among Medical Professionals in the Sichuan-Chongqing Region: Cross-Sectional Survey Study
JMIR Hum Factors 2024;11:e59153
URL: https://humanfactors.jmir.org/2024/1/e59153
doi:10.2196/59153

Building a Client Resource and Communication Platform for Community-Based Organizations to Address Health and Social Needs: Co-Design Study

Courtney Lyles[1], PhD; Beth Berrean[2], MBA; Ana Buenaventura[2], MBA; Svetlana Milter[2], MEng; Dayana Daniel Hernandez[2], BE; Urmimala Sarkar[3], MPH, MD; Christian Gutierrez[3], BA; Nynikka Palmer[3], MPH, DrPH; William Brown III[4], MA, DrPH, PhD

[1]UC Davis Center for Healthcare Policy and Research, UC Davis School of Medicine, University of California, Davis, Sacramento, CA, United States

[2]Technology Services Unit, School of Medicine, University of California, San Francisco, San Francisco, CA, United States

[3]Department of General Internal Medicine, Zuckerberg San Francisco General Hospital, University of California, San Francisco, San Francisco, CA, United States

[4]Division of Prevention Science, Department of Medicine, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Courtney Lyles, PhD
UC Davis Center for Healthcare Policy and Research, UC Davis School of Medicine, University of California, Davis, , Sacramento, CA, , United States

Abstract

Background: Connecting individuals to existing community resources is critical to addressing social needs and improving population health. While there is much ongoing informatics work embedding social needs screening and referrals into health care systems and their electronic health records, there has been less focus on the digital ecosystem and needs of community-based organizations (CBOs) providing or connecting individuals to these resources.

Objective: We used human-centered design to develop a digital platform for CBOs, focused on identification of health and social resources and communication with their clients.

Methods: Centered in the Develop phase of the design process, we conducted in-depth interviews in 2 phases with community-based organizational leadership and staff to create and iterate on the platform. We elicited and mapped participant feedback to theory-informed domains from the Technology Acceptance Model, such as Usefulness and Ease of Use, to build the final product and summarized all major design decisions as the platform development proceeded.

Results: Overall, we completed 22 interviews with 18 community-based organizational leadership and staff in 2 consecutive Develop phases. After coding of the interview transcripts, there were 4 major themes related to usability, relevance, and external factors impacting use. Specifically, CBOs expressed an interest in a customer relationship management software to manage their client interactions and communications, and they needed specific additional features to address the scope of their everyday work, namely (1) digital and SMS text messaging communication with clients and (2) easy ways to identify relevant community resources based on diverse client needs and various program eligibility criteria. Finally, clear implementation needs emerged, such as digital training and support for staff using new platforms. The final platform, titled “Mapping to Enhance the Vitality of Engaged Neighborhoods (MAVEN),” was completed in the Salesforce environment in 2022, and it included features and functions directly mapped to the design process.

Conclusions: Engaging community organizations in user-centered design of a health and social resource platform was essential to tapping into their deep expertise in serving local communities and neighborhoods. Design methods informed by behavioral theory can be similarly employed in other informatics research. Moving forward, much more work will be necessary to support the implementation of platforms specific to CBOs’ needs, especially given the resources, training, and customization needed in these settings.

(JMIR Hum Factors 2024;11:e53939) doi:10.2196/53939

KEYWORDS

mHealth; mobile health; eHealth; electronic health; application; digital health; digital ecosystem; informatics; community-based; community; co-design; human-centered design; community health; population health; technology; innovation; operations; social needs; health resources; qualitative analysis

Introduction

Research supports expanding screening and referral for social needs within health care systems [1]. Approximately 20% of health outcomes are associated with medical care, while the remaining influences are social, community, and structural factors—sometimes referred to as social needs and the broader social determinants of health [2]. Health systems need to work collaboratively with community-based organizations (CBOs) to address unmet social needs that affect health and health care outcomes [3].

Much of the research to date has focused on health care system electronic health record (EHR)–based platforms that can send and receive information from community organizations about social needs resource referrals, such as new EHR screening tools to assess social needs domains like food or housing insecurity, as well as EHR-integrated platforms that can recommend and connect patients to relevant programs in the community. There is emerging evidence that these EHR-based platforms can improve screening and referrals, but downstream impacts on health care and health outcomes are less clear [4]. However, there is a need for better digital platforms explicitly *within* CBOs that identify or connect clients to social and community resources to promote health. Even though CBOs communicate with community members about health and social needs and connect clients with other community resources, they have much different workflows than clinicians and staff in health care settings and they do not have access to existing EHR platforms. Digital design focused outside of the health care system is essential to contributing to a future broader ecosystem of referrals and connections between community organizations, social services agencies, public health agencies, and health care settings [5].

We employed human-centered design methods to build a platform to support CBOs in understanding client social resource needs and connecting them with the most relevant resources in neighborhoods within San Francisco. We outline here our design methods and the concrete platform build decisions that emerged from our work with community organizations, as both the methodological approaches and the design features from this work are relevant for other organizations and projects focused on social and health needs.

Methods

Study Design

As a part of a 4-year National Institutes of Health project titled "Mapping to Enhance the Vitality of Engaged Neighborhoods (MAVEN)," we built and then pilot-tested a digital platform for CBOs serving communities within the San Francisco Bay Area. While this project focused on specific health and social resource referrals in San Francisco, the design methods as well as the final platform characteristics are relevant for many other communities working on the same questions and processes for CBOs. This work was informed by the Double Diamond design framework throughout all phases of the study [6]. We have previously published the Discovery and Define phases of this project, in which we conducted extensive qualitative work with community members and community leaders to identify the core audience and design principles for this work. In brief, in the Discovery phase, we highlighted CBOs as the pivotal audience for digital platform development (given their central role in resource provision and trusted roles within their communities) [7]. Then, the Define phase highlighted the need to build a broad and scalable digital platform to support resource needs in multiple domains and for diverse sets of clients or populations served [8]. In the next Develop phase of work outlined here, we focused on the final build of the MAVEN digital platform.

While not all design studies are theoretically oriented, we combined the rigorous design methods with a well-established theory to additionally bolster the rigor of the work and further enhance the generalizability of the study. In particular, we used the Technology Acceptance Model to frame our interview guides and analysis [9]. This theory summarizes core concepts of technology acceptance and use into major domains, such as Ease of Use and Usefulness, which have been validated in many previous studies assessing how and why technologies are adopted [9]. Thus, the addition of theory to our design work allowed us to ensure that participant feedback was consistently documented and categorized against validated constructs throughout the study.

Building the MAVEN Platform: the Develop Phase

In the Develop phase, we completed 2 iterative rounds of design with CBOs in the San Francisco Bay Area (interview guide available in Multimedia Appendix 1). First, we completed 8 interviews with CBOs in late 2020 and early 2021 to understand their existing workflows using digital or online platforms to track and communicate with their clientele. At the end of these open-ended discussions, we also asked community leader participants to look at example digital resources to explore the features and relevance of these type of platforms in their everyday work. For example, community leaders reviewed websites with resources (such as on SF Department of Children, Youth and Their Families), providing overall opinions about usability as well as the types of information and content they might want to see displayed [10]. This phase was primarily open-ended and hypothetical, given that the platforms shown to participants were not always relevant for their own work. The interviews were conducted and recorded via Zoom (Zoom Video Communications, Inc), with the audio then converted into text via professional transcription.

After this round of feedback was completed, we created a minimal viable product (MVP) of the MAVEN social resource tool for CBOs and conducted additional co-design interviews. An MVP is a standard step in the design processes in which the team creates the simplest working prototype that can be tested, which is critical before more intensive digital building [11]. The MVP was created in spring 2021, and we completed an additional 13 think aloud interviews with CBO leaders to test the MVP in May and June 2021 [12]. These video interviews were conducted and recorded via Zoom, and the audio was then professionally transcribed. The think aloud process asked each participant to navigate the MVP on their own sequentially through each task or feature in the platform (eg, logging in, adding clients, sending a communication to a client), with prompts at every task that elicited technical errors or roadblocks as well as participants' overall opinions and reactions to the MVP.

Data Analysis

We used qualitative descriptive methods to complete our analyses within the Develop phase of the work, combining the participant feedback across all iterations of the platform build [13]. The text of transcribed interviews was analyzed using Dedoose (SocioCultural Research Associates) qualitative analysis software, and the second-round interview transcripts were also compared side by side with videos of the MVP think aloud procedures. Using open coding, we first identified overall patterns of feedback from participants, holding regular team meetings to establish consensus on the major usability and content feedback categories that emerged (ie, specific feature needs and preferences, technical barriers to use). Next, we mapped these categories onto Technology Acceptance Model domains to examine how the overall discussion categories were related to documented theoretical constructs such as relevance or usefulness as well as ease of use of the platform.

Finally, we made overall design decisions informed by these qualitative findings to incorporate into the final MAVEN platform. More specifically, we used the data from both rounds of feedback to prioritize new functionality that needed to be added to the platform, as well as improvements or changes in the MVP features that would improve relevance or usability. These design changes are summarized here, with the final MAVEN tool completed in the spring of 2022.

Ethical Considerations

This study was approved by the University of California San Francisco (institutional review board no. 18-25696). The participants provided written consent for the interviews to be audio and video recorded, and the transcripts were deidentified prior to the analysis. Each participant was compensated US $75 per interview.

Results

Study Sample

We completed 22 in-depth interviews (9 in the first phase and 13 in the second phase of the study), with 18 unique CBO leaders and staff in total (4 participants were interviewed in both phases of work). The 18 participants in the Develop phase of the MAVEN project work ranged from individual consultants working on multiple health campaigns in their local neighborhoods to mid-sized nonprofit organizations running programs and provisioning services in their communities. The full summary of participants is shown in Table 1.

Table . Summary of community-based organizations staff or leader participants.

Phase	Organization type	Participant title or role	Location
1	Nonprofit	Director of community partnerships and program evaluation for HIV programs	San Francisco, CA
1	Nonprofit	Agent and disability resource specialist	Mission Neighborhood, San Francisco, CA
1	Nonprofit	Information and assistance specialist	Mission Neighborhood, San Francisco, CA
1	Nonprofit	Program director, women's cancer program	Tenderloin Neighborhood, San Francisco, CA
1	Nonprofit	Program coordinator, HIV/AIDS programs	San Francisco, CA
1, 2	Community advocate	Health educator and activist	San Francisco Bay Area, CA Los Angeles, CA
1, 2	Nonprofit	Senior services manager	Mission neighborhood, San Francisco, CA
1, 2	Local planning council; nonprofit	Member; senior director of programs	Tenderloin neighborhood, San Francisco, CA
1, 2	Nonprofit	Executive director	San Francisco, CA
2	Local government housing or community organization	Director, social/health resources	Tenderloin neighborhood, San Francisco, CA
2	Regional advisory group; local government organization	Advisory member; HIV case manager	Alaska, Hawaii, California
2	Nonprofit	Program director, senior services programming	San Francisco, CA
2	Local government housing or community organization	Program director, digital programming	Tenderloin Neighborhood, San Francisco, CA
2	Community advocate	Community health/HIV activist	San Francisco, CA
2	Local government Public health department	Community health outreach/HIV activist	Alameda, CA
2	Federally qualified health center	Health educator	Marin City, CA
2	Nonprofit	Wellness manager	Bay View Neighborhood, San Francisco, CA
2	Public research university system	Health equity strategist/consultant	San Francisco, CA

Qualitative Results

Analyzing all in-depth interviews during the Develop phase, we identified 4 major themes from community leader participants, which are mapped to the Technology Acceptance Model domains of Usefulness, Ease of Use, and External Factors influencing use.

Overall Usefulness

In the domain of usefulness, there was consensus that access to technology platforms with the ability to track clients across programs and communicate with clients more seamlessly was a priority for CBOs. Several of the organizations' current workflows involved using out-of-the-box and free programs for their daily work of tracking clients, finding and referring people to resources, and communicating with clients. While existing platforms allow digital communication with clients and tracking client contact information or program participation, the standard work was often nonintegrated and potentially duplicative. For example, one participant stated:

> *If you want to refer somebody or some participant to other services,…you can Google it and just find out the exact address and telephone that they can call. I think it's the best tool that we can have now to just give more information and referrals.*

It was clear from these participants that better functionality beyond free platforms would improve their daily work at CBOs, such as by standardizing the search and communication workflows into a single place.

More specifically, there was an understanding among community leaders that existing consumer or customer resource management (CRM) platforms, such as Salesforce (Salesforce, Inc) and other similar CRM software, were important to be able to complete multiple client functions together—such as tracking clients and communicating with clients (individually or in groups) without toggling between multiple different platforms. For example, one participant stated:

> *I love [seeing]…a lot of information and a lot of resources at the same time. So, I don't have to look for another screen and start looking for another information for the clients.*

Thus, any CRM platform that combined database management for clients alongside communication tools was mentioned as more ideal for CBO workflows. Specific to the selection of a CRM software, participants most often mentioned Salesforce as the platform of interest (either currently in use or a wish for use in the near future): "I think at least 50% [of organizations we partner with] are using some portion of Salesforce [with] their clients."

Relatedly, the usefulness of up-to-date lists of community-based resources was a top priority for almost all participants. For example, one community leader stated, "I think that [technology] would be a good way to maybe have a resource list of frequently used CBOs [and resources they provide]." Similarly, another participant stated:

> *I think being able to see what [resource] is available—and I think if this stuff is updated in real time, it's super helpful…. Having specific contacts [at each organization] for different opportunities[/resources] is really key just because I think that's often something that people spend a lot of time trying to determine.*

However, it was clear that this registry of community resources would require upkeep and trust in the resources presented. For example, a participant mentioned:

> *[A registry of] existing community resources—that's a lot to keep up-to-date. And if you just don't have it, then you're not disappointing people [clients].*

A final useful function of a platform for CBOs centered around communication with clients. Comments such as these solidified the need to have integrated and customizable communication with clients as a core function of the platform: "You…have subgroups of different clients depending on what the [program] is, and then you would just create a new [communication] for whatever thing you were doing," and "I feel like, just knowing colleagues at other places, a lot of people are paying money for text and email blasts."

Overall Ease of Use

Next, there were a number of usability issues that emerged during the interviews.

In addition, when interacting with the early MVP of the MAVEN platform, there were several usability comments made about the layout and the terminology, especially in the context of community-oriented work. One participant stated it as:

> *I just want to make sure that [you know that] some of these words that you're using are a little technical…. Maybe finding other words…that will describe something that's more simple and easier to find.*

And another participant mentioned:

> *Since it's our first time, it's going to be difficult just to find everything, but once you get to know it, I think just it's going to take less than a week just to see and working with the database or with the website.*

While the overall usability ratings of the MVP were favorable, there were several areas of improvements that community leaders mentioned that were specific to increasing the usability of a platform that served diverse clients and were necessary in many CBOs' workflows. First, participants identified a need to have easier ways to find recommended resources for clients, as exemplified in comments such as:

> *[Make it] easy to find if it's in alphabetical order or has some kind of a search engine that I could put in 'Help with income taxes' or 'Grocery shopping' or whatever the [client] needs.*

Similarly, another participant stated that it is important to ensure that referrals to resources match their participants' eligibility and identity:

> *[To make sure] you could search by eligibility requirement, whether it's senior citizen, or HIV-positive, or somebody who identifies as LGBTQ.*

Next, participants identified clear, equity-focused improvements to the platform, especially around language accessibility and sociodemographic representation of fields (particularly careful collection of sexual orientation and gender identity data for LGBTQ+ populations). For example, one participant stated, "When I talk about cultural sensitivity, like I said, it's not just a language also. If it's from LGBTQIA community or different cultures, it's very, very important. Age also is another issue," and another stated, "I know that many agencies have different [languages], for example, Chinese and English, also, Russian and English, Spanish and English…. So, it should be in multiple languages."

External Factor: Varying Implementation Contexts Within CBOs

Finally, there were many external factors that emerged as critical to use of the MAVEN platform. These comments generally focused on the implementation context for using such a digital platform.

First, there was a clear need to focus on digital skills and necessary training among CBO staff. For example, participants stated, "We know Salesforce has really high potential, but we don't have the time to learn it all…. So, we don't use this fancy thing because it was just too overwhelming to use, to learn it and so we're just comfortable with what we know," and "I feel like there's certainly a spectrum of computer literacy across my team, and so I think getting folks that are less technologically adept onboard for using something like this could be a little bit challenging, but I don't think it would be that hard."

Participants also discussed that no platform would be successful without leadership support as well as very detailed training among frontline staff. For example, participants stated, "I feel like it's important to have a collective buy-in or an agreement or some kind of commitment to using a certain resource to sort of get it off the ground," and "Oftentimes, there are lots of tools that you will get sent that will be given to somebody but I think they're only as good as the people that are using them." Furthermore, there were common comments about on-the-ground support for staff members at CBOs such as, "I was thinking more training for community leaders in how to really effectively use this tool."

In addition, there were more nuanced conversations with CBOs about their current workflows with clients that needed to be considered in order to get uptake of a new system. In particular, because CBO staff used in-person and phone outreach often for clients, any digital tool would require consideration of how to blend such high-touch, face-to-face communication with the lower-touch messaging and reminders that digital platforms can facilitate. Quotes such as these exemplified this idea:

> *So it's about having a conversation with the client and understanding how they best connect the services and also what barriers may be in the past they have faced when trying to connect with other providers and sort of strategizing ways to circumvent those challenges.*

Comments from CBOs about their workflows highlighted that even a perfect digital solution that easily facilitated messaging may supplement but not replace high-touch conversations with clients about their preferences for resources.

Finally, some of the conversation about implementation centered around 2 additional points that related to the overall uptake and dissemination of any platform. The security and privacy of information in any client database was critical, especially given the sensitive information disclosed by clients and the known predatory behavior of hackers and other criminal behavior. This was exemplified in the following quote: "Organizations may take advantage of…elderly patients." Moreover, the implementation considerations for wider spread and uptake of digital platforms for health and social needs also centered on the practical work of maintenance, which cannot be overlooked. For example, one participant stated:

> *There is a lot of movement that occurs in public service organizations that having the most up to date data or contacts could be a concern and a lot of work on us to continue to update.*

In other words, the ability to use existing resource lists that are actively updated and monitored for relevance is essential for implementation and will not be solved by a technical solution alone.

Final Design Decisions and Build

Given the qualitative findings across both the open-ended and MVP testing of the MAVEN platform, we made multiple design decisions that reflected the community-based organizational leadership input and preferences. Overall, we chose to design within Salesforce, activate several features within the platform, and then integrate outside tools to provide information and tools within the platform. Table 2 summarized core findings matched to concrete features or changes within the MAVEN platform.

Table . Major design needs from co-design feedback.

Design needs identified by participants	Subsequent feature or design decision
Platform need: Community leaders used or wanted CRM[a] platform to manage their clients and workflows.	Selection of Salesforce platform for MAVEN[b] build.
Feature need: Community leaders had knowledge of social or health resources but strongly desired a trusted and easy-to-use directory to improve client resource referrals.	Review of several options for an up-to-date list of community resources; selection of San Francisco Service Guide existing resource list (imported into MAVEN via API[c]).
Feature need: Simple and seamless communication was a priority to reach and stay connected to clients.	Activation of Twilio within Salesforce Campaigns to allow for 2-way texting functionality.
Feature need: Matching of available resources to specific client needs and identities was critical.	Selection of ServiceMatch feature from Salesforce, which uses an algorithm to recommend "best" resource by resource type or category (eg, food), zip code, and eligibility criteria (eg, language of service provision).
Usability improvement: Improvement of language and data collection fields within digital platforms needed to better represent client identities.	While using existing CRM platform, we expanded client intake fields to be better representative of the community; exploration of designing in multiple languages (but unable to do so in this Develop period).
Implementation consideration: Workflows and staffing capacity within CBOs[d] varied greatly and needed to be prioritized.	Implementation support and training identified as a high priority during rollout.

[a]CRM: customer resource management.

[b]MAVEN: Mapping to Enhance the Vitality of Engaged Neighborhoods.

[c]API: application programming interface.

[d]CBO: community-based organization.

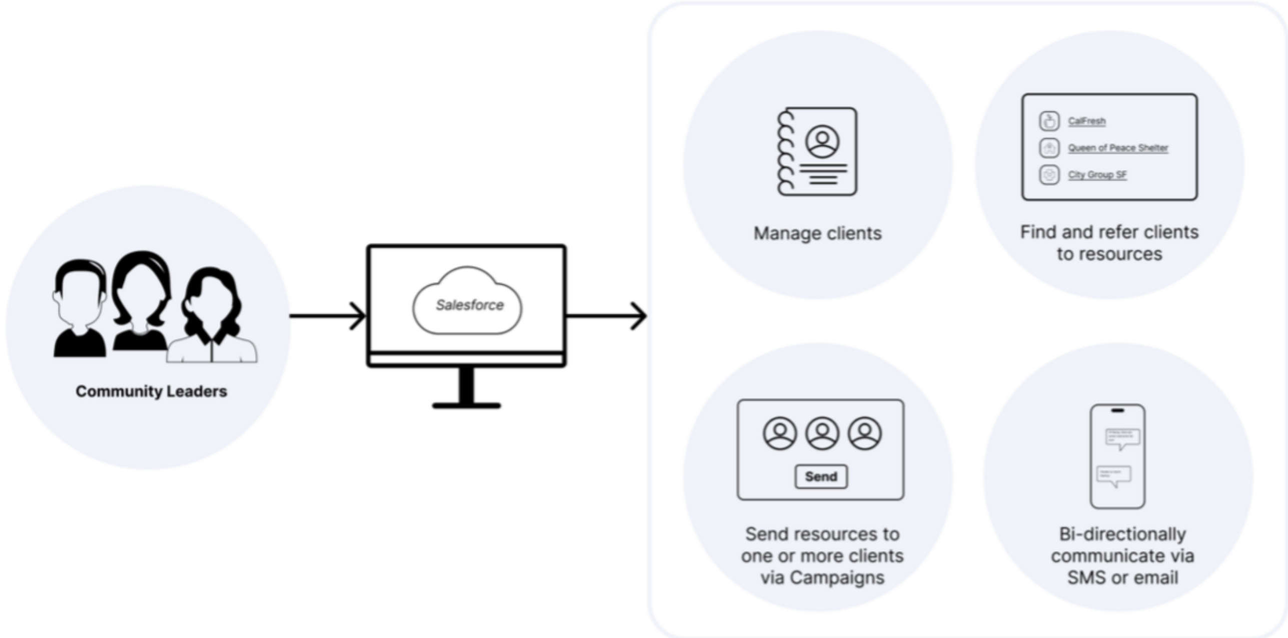
We ended up implementing these design decisions in stages within our work. In the first MVP build of the MAVEN platform (Figure 1), we identified the core platform and communication functionality needed:

1. Salesforce environment, as this was a platform used in about half of existing community organizations to track client and program information (but only accessible or viewable

within their own organization). This selection was regularly expressed as a desired platform by most CBOs in the study.

2. Enabling the Campaign feature within the Salesforce environment, which allows group SMS and email communication within the platform (on top of existing client-tracking features). Thus, community leaders could login, create new contacts or clients, and also create new messages with any content they developed or inserted.
3. Adding Twilio-for-Salesforce–managed package to seamlessly allow 2-way SMS text messaging via the platform.

Figure 1. MAVEN minimal viable product. MAVEN: Mapping to Enhance the Vitality of Engaged Neighborhoods; SF: San Francisco.

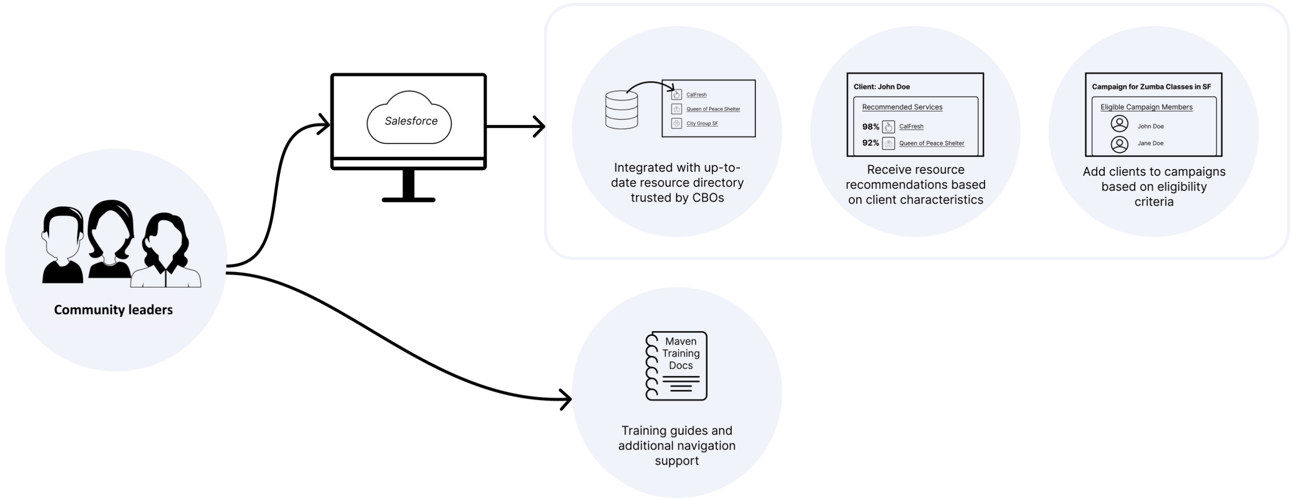


After the second phase of testing an MVP of the MAVEN platform, we identified additional platform features as new final components of the tool (Figure 2):

1. Integration with an existing resource directory that was trusted among community organizations. After review of several existing lists of community resources in San Francisco, we chose a freely available and regularly updated resource list called the "San Francisco Service Guide" (managed by a nonprofit called "ShelterTech") [14]. This registry of resources included categories of resources: locations, hours, contact information for each resource, and some free-text eligibility information for the resource (eg, available to those aged 65 years or older via a senior center) [14]. We connected SF Service Guide via application programming interface into Salesforce and built an automated refresh process to update the list of records on a weekly basis. The underlying SF Service Guide repository was regularly updated every 90 days by volunteers at ShelterTech.
2. Ability to receive "recommendations" for a client resource based on category of resource, location, and eligibility criteria. We chose a freely available and open-source tool within the Salesforce environment called ServiceMatch, which was available via Salesforce.org Impact Labs [15]. We customized the existing code for ServiceMatch within MAVEN, adjusting the algorithm to be able to use the existing resource list information to make recommendations of the most relevant resources based on client characteristics (eg, eligibility of resource based on age, location [ie, client zip code], and resource type [ie, search for food vs housing]). Finally, connecting ServiceMatch with Twilio allows community leaders to refer these relevant resources to clients via SMS in addition to email and printout with 1 click.
3. Extension of existing Campaigns Salesforce tool that automatically adds clients to the campaign based on the resource eligibility criteria and allows referral of multiple resources to 1 or more clients. Thus, the community leaders can create a campaign and add multiple resources to the campaign, and the system automatically selects clients based on the unified eligibility criteria across all campaign resources selected. Broadcast messaging service was used to send bulk SMS messages to clients included in the campaign.
4. Rollout of platform was completed with the addition of implementation support. Finally, we created detailed user guides and training materials to support CBOs in using a platform that was often very new to their daily workflow and addressed some of the major navigation questions that were confusing in the platform.

The final MAVEN product combined elements from multiple places that did not previously exist: the Campaigns feature (with additional communication functionality), plus curated resources directly from SF Resource Guide, plus searching of resources by type, location, and eligibility via ServiceMatch.

Figure 2. MAVEN phase 2. CBO: community-based organization; MAVEN: Mapping to Enhance the Vitality of Engaged Neighborhoods; SF: San Francisco.



Discussion

Principal Findings

This paper outlines a co-design process to create and iterate a digital platform for CBOs, focused on finding relevant health and social resources and easily connecting clients to these resources. Overall, human-centered design methods elucidated core needs and preferences of CBOs for health and social resource screening and referrals. This is an audience that has often been overlooked in the clinical informatics literature given that CBOs work outside of the EHR and have entirely different workflows with their clients as compared with a health care environment [16,17]. Specifically, we outlined multiple platform features and design decisions that matched the needs expressed during in-depth qualitative interviews, applying a rigorous and theory-informed approach to the design.

Our findings are likely to have relevance outside of the San Francisco context because of the design choices made in this study. First, many CBOs across the country are looking to use CRM software and specifically Salesforce to manage client communication, which was also reflected in our study findings. Therefore, there is the ability for other groups to use the existing Salesforce functionality identified in the MAVEN study (such as ServiceMatch and the Campaigns feature with Twilio integration). In addition, we chose to integrate an existing resource list via application programming interface into the MAVEN platform, which can be swapped out for any other resource list based on the local context in other communities. Importantly, the co-design process in this study to build the MAVEN platform led not only to the enhancement of the tool but also to specific attention and focus on the context for using the tool, such as use of existing operating platforms, workflow considerations, staffing and workforce capacity, and flexibility of the tool to meet the diverse needs of clients served [18,19].

Moreover, this study is a useful example of the Develop phase of the design work, which emphasizes real-world implementation and potential constraints, which are a critical complement to the open brainstorming and ideation of the Discovery phase of design [20]. Moving forward in this field, it will also be critical for future studies to focus on human-centered design to advance health equity. Equity-focused design methods are spreading, and these methods are essential for building platforms that prioritize the needs of marginalized communities and address multiple levels of influences on health outcomes [21,22].

Limitations

This study had several limitations, such as the completion of the work in a single geographic area (the San Francisco Bay Area) and the relatively small number of organizations participating in the design. However, both the design methods and the findings that highlight implementation considerations are widely generalizable to many other cities and municipalities (in addition to Salesforce build considerations). Finally, the budget of this research project also limited the number of phases of iterations and the final set of features activated. Similarly, because of budget, we were not able to transition use of the MAVEN platform from the University of California San Francisco (UCSF) Salesforce license (where the study testing occurred) to independent licenses at CBOs, which remains a barrier for longer-term uptake.

Conclusions

Moving forward, there is a strong need for more community-engaged research and informatics co-design to support the development and implementation of platforms that assist CBOs in their daily work. Understanding individual social needs within health care delivery systems is essential for health promotion, but solutions that require delivering health and social resources within local neighborhoods will not be solved by health care systems alone. In order to partner with CBOs and follow their expertise, we also cannot ignore the resources, capacity, and skills or training needs in these organizations (as our study and many others have highlighted), as real impact will be diminished without investment in CBOs based on their priorities.

Authors' Contributions

CL, BB, AB, and WB conceptualized the study, interpreted data, and drafted/edited the manuscript. SM, DDH, CG, NP, and US interpreted data and critically edited the manuscript.

Conflicts of Interest

US holds current research funding from the National Cancer Institute of the National Institutes of Health, California Healthcare Foundation, the Patient-Centered Outcomes Research Institute, and the Agency for Healthcare Research and Quality. She has received prior grant funding from the Gordon and Betty Moore Foundation, the Blue Shield of California Foundation, HopeLab, the US Food and Drug Administration, and the Commonwealth Fund. She received gift funding from The Doctors Company Foundation. She holds contract funding from InquisitHealth and RecoverX. US serves as a scientific/expert advisor for HealthTech 4 Medicaid (volunteer). She is a member of the American Medical Association's Equity and Innovation Advisory Group (honoraria) and is on the Board of Directors of the Collaborative for Accountability and Improvement (volunteer). She is an advisor for Waymark (shares) and for Ceteri Capital I GP, LLC (shares). She has been a clinical advisor for Omada Health (honoraria) and an advisory board member for Doximity (honoraria, stock).

Multimedia Appendix 1

Semistructured interview questions for community-based organizational leaders.

[DOCX File, 15 KB - humanfactors_v11i1e53939_app1.docx]

References

1. Thompson T, McQueen A, Croston M, et al. Social needs and health-related outcomes among Medicaid beneficiaries. Health Educ Behav 2019 Jun;46(3):436-444. [doi: 10.1177/1090198118822724] [Medline: 30654655]
2. Magnan S. Social determinants of health 101 for health care: five plus five. NAM Perspect 2017 Oct;7(10). [doi: 10.31478/201710c]
3. Park Y, Quinn JW, Hurvitz PM, et al. Addressing patient's unmet social needs: disparities in access to social services in the United States from 1990 to 2014, a national times series study. BMC Health Serv Res 2022 Mar 19;22(1):367. [doi: 10.1186/s12913-022-07749-1] [Medline: 35305617]
4. Yan AF, Chen Z, Wang Y, et al. Effectiveness of social needs screening and interventions in clinical settings on utilization, cost, and clinical outcomes: a systematic review. Health Equity 2022;6(1):454-475. [doi: 10.1089/heq.2022.0010] [Medline: 35801145]
5. Brewster AL, Yuan CT, Tan AX, Tangoren CG, Curry LA. Collaboration in health care and social service networks for older adults. Med Care 2019;57(5):327-333. [doi: 10.1097/MLR.0000000000001097]
6. Rosala M. Discovery: definition. Nielsen Norman Group. 2020 Mar 15. URL: https://www.nngroup.com/articles/discovery-phase/ [accessed 2023-09-06]
7. Nguyen KH, Fields JD, Cemballi AG, et al. The role of community-based organizations in improving chronic care for safety-net populations. J Am Board Fam Med 2021;34(4):698-708. [doi: 10.3122/jabfm.2021.04.200591] [Medline: 34312263]
8. Nguyen KH, Cemballi AG, Fields JD, Brown W, Pantell MS, Lyles CR. Applying a socioecological framework to chronic disease management: implications for social informatics interventions in safety-net healthcare settings. JAMIA Open 2022 Mar 25;5(1):ooac014. [doi: 10.1093/jamiaopen/ooac014] [Medline: 35571359]
9. Marangunić N, Granić A. Technology acceptance model: a literature review from 1986 to 2013. Univ Access Inf Soc 2015 Mar;14(1):81-95. [doi: 10.1007/s10209-014-0348-1]
10. Find a program. SF Department of Children, Youth & Their Families. URL: https://www.dcyf.org/find-a-program [accessed 2023-10-24]
11. Lenarduzzi V, Taibi D. MVP explained: a systematic mapping study on the definitions of minimal viable product. Presented at: 2016 42th Euromicro Conference on Software Engineering and Advanced Applications (SEAA); Aug 31 to Sep 2, 2016:; Limassol, Cyprus p. 112-119. [doi: 10.1109/SEAA.2016.56]
12. Jaspers MWM, Steen T, van den Bos C, Geenen M. The think aloud method: a guide to user interface design. Int J Med Inform 2004 Nov;73(11-12):781-795. [doi: 10.1016/j.ijmedinf.2004.08.003] [Medline: 15491929]
13. Doyle L, McCabe C, Keogh B, Brady A, McCann M. An overview of the qualitative descriptive design within nursing research. J Res Nurs 2020 Aug;25(5):443-455. [doi: 10.1177/1744987119880234] [Medline: 34394658]
14. Find essential services in San Francisco. SF Service Guide. URL: https://sfserviceguide.org [accessed 2023-10-20]
15. Impact Labs: Service Match. Salesforce AppExchange. URL: https://appexchange.salesforce.com/appxListingDetail?listingId=a0N4V00000G62EOUAZ [accessed 2023-10-20]
16. Matheson GO, Pacione C, Shultz RK, Klügl M. Leveraging human-centered design in chronic disease prevention. Am J Prev Med 2015 Apr;48(4):472-479. [doi: 10.1016/j.amepre.2014.10.014] [Medline: 25700655]
17. Nijagal MA, Patel D, Lyles C, et al. Using human centered design to identify opportunities for reducing inequities in perinatal care. BMC Health Serv Res 2021 Jul 20;21(1):714. [doi: 10.1186/s12913-021-06609-8] [Medline: 34284758]

18. Harte R, Glynn L, Rodríguez-Molinero A, et al. A human-centered design methodology to enhance the usability, human factors, and user experience of connected health systems: a three-phase methodology. JMIR Hum Factors 2017 Mar 16;4(1):e5443. [doi: 10.2196/humanfactors.5443] [Medline: 28302594]
19. Chen E, Neta G, Roberts MC. Complementary approaches to problem solving in healthcare and public health: implementation science and human-centered design. Transl Behav Med 2021 May 25;11(5):1115-1121. [doi: 10.1093/tbm/ibaa079] [Medline: 32986098]
20. Zikmund-Fisher BJ, Tuepker A, Metcalf EE, Strange W, Teo AR. Applying user-centered design in the development of nudges for a pragmatic trial to reduce no-shows among veterans. Patient Educ Couns 2022 Jun;105(6):1620-1627. [doi: 10.1016/j.pec.2021.10.024] [Medline: 34756639]
21. Harrington C, Erete S, Piper AM. Deconstructing community-based collaborative design: towards more equitable participatory design engagements. Proc ACM Hum-Comput Interact 2019 Nov 7;3(CSCW):1-25. [doi: 10.1145/3359318]
22. Chen E, Leos C, Kowitt SD, Moracco KE. Enhancing community-based participatory research through human-centered design strategies. Health Promot Pract 2020 Jan;21(1):37-48. [doi: 10.1177/1524839919850557] [Medline: 31131633]

Abbreviations

CBO: community-based organization
CRM: customer resource management
EHR: electronic health record
MAVEN: Mapping to Enhance the Vitality of Engaged Neighborhoods
MVP: minimal viable product

Edited by C Jacob; submitted 07.11.23; peer-reviewed by N Rossi, R Gold, W Adegu; revised version received 24.05.24; accepted 25.05.24; published 16.08.24.

Please cite as:
Lyles C, Berrean B, Buenaventura A, Milter S, Hernandez DD, Sarkar U, Gutierrez C, Palmer N, Brown III W
Building a Client Resource and Communication Platform for Community-Based Organizations to Address Health and Social Needs: Co-Design Study
JMIR Hum Factors 2024;11:e53939
URL: https://humanfactors.jmir.org/2024/1/e53939
doi:10.2196/53939

Original Paper

Health Care Professionals' Perspectives Before and After Use of eDialogue for Team-Based Digital Communication Across Settings: Qualitative Study

Lili Worre Høpfner Jensen[1*], BSN, MSc; Ole Rahbek[1*], MD, PhD; Rikke Emilie Kildahl Lauritsen[1*], BPT, MSc; Søren Kold[1*], MD, PhD; Birthe Dinesen[2*], MSc, PhD

[1]Interdisciplinary Othopaedics, Orthopaedic Surgery Department, Aalborg University Hospital, Aalborg, Denmark

[2]Laboratory for Welfare Technologies-Digital Health and Rehabilitation, ExerciseTech, Department of Health Science and Technology, Aalborg University, Aalborg, Denmark

[*]all authors contributed equally

Corresponding Author:
Lili Worre Høpfner Jensen, BSN, MSc
Interdisciplinary Othopaedics
Orthopaedic Surgery Department
Aalborg University Hospital
Hobrovej 18-22
Aalborg, 9000
Denmark
Phone: 45 60229406
Email: lili.jensen@rn.dk

Abstract

Background: Orthopedic surgical treatment is a transversal task that requires the active involvement of patients, relatives, and health care professionals (HCPs) across various settings. However, after hospital discharge, communication is challenged and undertaken primarily by phone. New digital communication solutions have the potential to create a space for seamless and patient-centered dialogue across discipline and sector boundaries. When evaluating new communication solutions, knowledge about HCPs' needs and perspectives of use must be explored, as it is they who are responsible for implementing changes in practice.

Objective: This study aimed to (1) investigate HCPs' perceptions of current communication pathways (phase 1) and (2) explore their experiences of using a simple messenger-like solution (eDialogue) for team-based digital communication across settings (phase 2).

Methods: We used a triangulation of qualitative data collection techniques, including document analysis, observations, focus groups, and individual interviews of HCPs before (n=28) and after (n=12) their use of eDialogue. Data collection and analysis were inspired by the Consolidated Framework for Implementation Research (CFIR) to specifically understand facilitators and barriers to implementation as perceived by HCPs.

Results: HCPs perceive current communication pathways as insufficient for both patients and themselves. Phone calls are disruptive, and there is a lack of direct communication modalities when communication crosses sector boundaries. HCPs experienced the use of eDialogue as a quick and easy way for timely interdisciplinary interaction with patients and other HCPs across settings; however, concerns were raised about time consumption.

Conclusions: eDialogue can provide needed support for interdisciplinary and cross-sectoral patient-centered communication. However, future studies of this solution should address its impact and the use of resources.

(JMIR Hum Factors 2024;11:e53391) doi:10.2196/53391

KEYWORDS

CFIR; Consolidated Framework for Implementation Research; digital communication; hospital discharge; implementation science; interdisciplinary communication; orthopedic surgery; patient-provider communication; postoperative care; qualitative research; text messaging

Introduction

Treatment of patients undergoing orthopedic surgery is a cross-disciplinary task formed in partnership with the patient. Communication and collaboration between the patient and different professional groups across various settings are key to achieving quality in patient trajectories and clinical outcomes [1-3]. While hospitalization times are decreasing, an increasing part of the postoperative period takes place in the patient's home and with support from municipal health care professionals (HCPs) [4,5]. However, they are largely dependent on contact with hospital staff when problems related to treatment and care arise.

In Denmark, the current means of communication between patients undergoing orthopedic surgery and HCPs across sectors is primarily by phone, but the synchronicity of this is inflexible and time-consuming. Moreover, HCPs across sectors communicate through different electronic systems, but without including patients in the dialogues. New communication strategies must aim to provide seamless communication paths that reach beyond the existing silos of the health care system and include patients as partners [6].

Digital patient platforms are being introduced in Denmark [7,8] as well as internationally [9,10]. Patients can receive digital patient education, see test results, and answer questionnaires used by clinicians to tailor treatment plans. In some cases, patients are given the opportunity to send texts in a secured chat to HCPs at the hospital before and after hospitalization in addition to phone calls. Internationally, secure messaging is reported as the most used feature on patient platforms [9]. Even though questions are not limited to nursing tasks, answering the messages is often delegated to nurses in outpatient clinics or wards at the hospital [7,8]. This leads to duplicate work for the nurses, who will act as intermediaries or gatekeepers for the questions that patients might have, in the same way as secretaries are gatekeepers for patient-initiated phone calls. Moreover, HCPs from the municipality are not involved in these digital encounters. Even though the surgeon at the hospital holds the primary responsibility for the orthopedic treatment [11], there are no direct communication modalities available between the patient, surgeon, and HCPs across sectors in the postoperative period. A team-based approach to the use of digital communication, involving the patient and all HCPs in their care team, may improve postdischarge communication and support patients more optimally after surgery and discharge. Our focus for this study was on communication pathways both involving patient-to-provider communication as well as provider-to-provider communication, as this is interwoven and interdependent in clinical practice.

In an exploratory qualitative study, we tested a simple messenger-like solution for team-based digital communication between patients and HCPs across sectors (eDialogue), and the perspectives of patients and their use of the solution have been reported in another study. However, when testing new communication pathways in health care, it is pivotal to explore the perspectives of all end users to identify their needs, motivations, and barriers to use at an early preimplementation stage [12]. Therefore, this study aimed to (1) investigate HCPs' perceptions of current communication pathways with orthopedic surgery patients and collaborating HCPs across sectors, as well as their expectations for eDialogue (phase 1), and (2) explore their experiences of using eDialogue for team-based communication (phase 2).

Methods

Study Design

We used a triangulation of qualitative data collection techniques to understand contextual factors and what opportunities and challenges exist before (phase 1) and after (phase 2) the use of eDialogue. This included document analysis, observations [13], semistructured focus groups [14], and individual interviews [15]. Reporting this study followed the Consolidated Criteria for Reporting Qualitative Research (COREQ) checklist [16].

Theoretical Framework

Conducting this study, we were inspired by the metatheoretical framework and terminology described by Damschroder et al [17]: the Consolidated Framework for Implementation Research (CFIR). The CFIR is widely used in health services research and specifically adapted to understand facilitators and barriers to implementation, even at an early preimplementation stage [17,18]. CFIR is centered around five key domains related to implementation, including (1) the intervention, (2) the inner setting, (3) the outer setting, (4) the individuals involved in the intervention, and (5) the processes conducted to implement the intervention [17]. To each domain belong underlying constructs, which describe factors that can either motivate or hinder implementation [17]. Selected CFIR domains and constructs guided our data collection by informing the interview guides and the observation protocol in combination with exploratory questions. In an inductive-deductive approach, CFIR domains and constructs were used to structure data analysis and the reporting of our findings, while still being open to emerging themes. By using CFIR, we aimed to promote structured knowledge building for future implementation strategies that may encourage the adoption of eDialogue in clinical practice.

Participants and Setting

The study originated from the orthopedic surgery department at Aalborg University Hospital, which is a tertiary hospital in Denmark. The Danish health care system is mainly financed by general taxes and is therefore provided free of charge to individuals. It operates across 3 administrative and political levels, which are the state (national level), the regions (regional level), and the municipalities (local level). Hospital care is provided by the 5 regions of Denmark, and primary care and social services, such as rehabilitation outside hospitals, home nursing, and physiotherapy, are provided by the 98 municipalities of Denmark. Even though there is cofinancing and close collaboration between the regional and local levels, HCPs are employed in different organizations and use different electronic health records. There are defined care pathways for patients in need of treatment and care across settings that outline the tasks of the HCPs employed at the different levels, just as there is legislation that the HCPs must follow. However, major challenges exist in communication and collaboration across settings, especially related to patients in transitions of care from hospital to home.

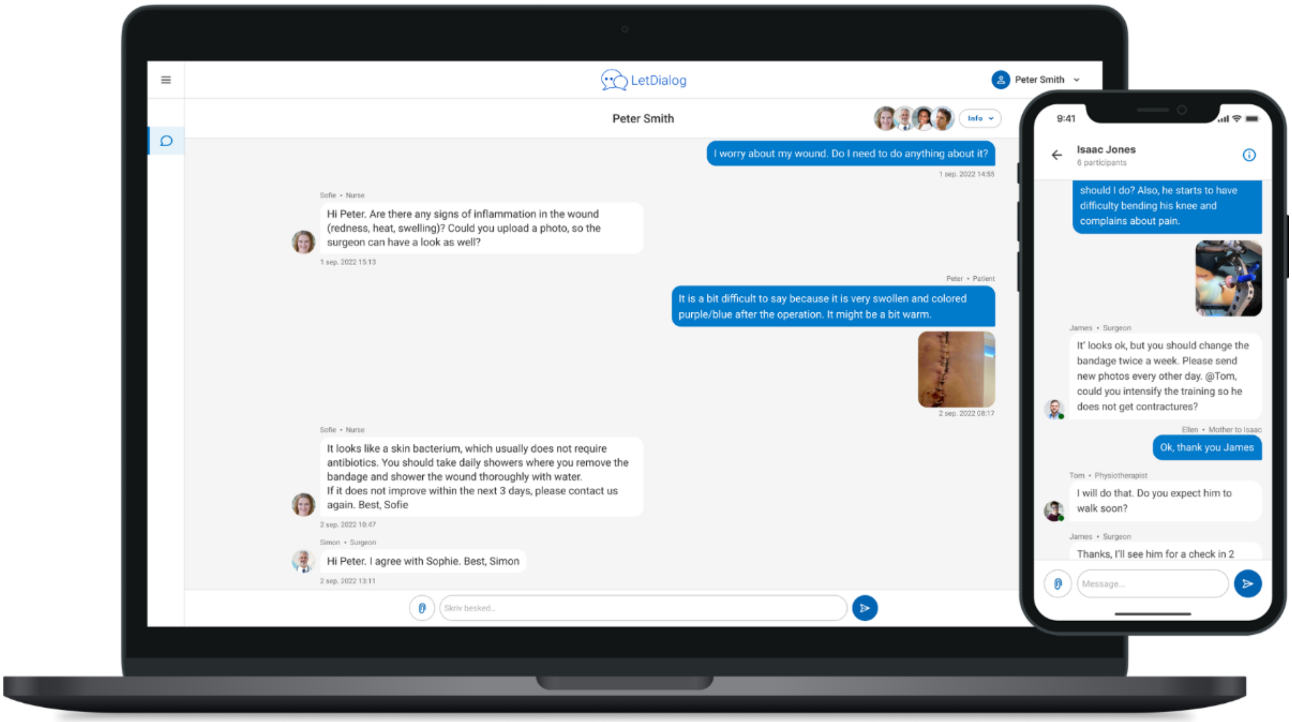
Phase 1: Before eDialogue

In phase 1, orthopedic surgeons, secretaries, nurses, and physiotherapists from Aalborg University Hospital and home care nurses and physiotherapists from the Aalborg municipality were recruited for preintervention focus groups (n=6) to investigate their perceptions of current communication pathways and their expectations for eDialogue. Inclusion criteria were HCPs working with orthopedic patients from 2 different subspecialties that were recruited to test and explore eDialogue. These were patients undergoing either deformity correction surgery involving complex prolonged treatment with hospitalization or anterior cruciate ligament reconstruction performed as day surgery (ie, discharged on the day of surgery). HCPs were recruited from different units at the hospital, including the outpatient clinic, the ward, and the physiotherapy department, and from different districts of the Aalborg municipality. Exclusion criteria were HCPs who had sparse knowledge of orthopedic treatment and care; for example, personnel hired within the past year. We purposely strove to include HCPs from various vocational roles to achieve a detailed understanding of the clinical trajectory and interdisciplinary communication with patients undergoing orthopedic surgery. Inclusion persisted until data saturation was reached for the interviews, that is, no new themes occurred [15].

Intervention: eDialogue

Team-based digital communication between patients and HCPs across settings was facilitated through a technical General Data Protection Regulation-compliant solution assessed by an app for a smartphone or through a website (Figure 1). The technical solution is already in use in some municipalities in Denmark in the field of social education [19], but has never been used to facilitate communication in health care or across sectors. The solution was chosen by the research team before the study based on the simple and intuitive interfaces and discussed with patients undergoing orthopedic surgery and HCPs in an initial workshop before this study.

Figure 1. The figure shows screenshots of digital dialogues between patients and health care professionals (HCPs) across settings from the study. Access was either by app on a smartphone or by web, using a simple messenger-like user interface.



Patients from 2 orthopedic surgery subspecialties were recruited consecutively for this study and offered to use eDialogue for 2 months after they had been discharged with their team of HCPs across settings.

Just as patients were helped to create an account using a digital signature (NemID), HCPs were guided to become users of eDialogue. Most HCPs accessed it through the website, but some preferred access through the app on their smartphones. Finger touch or face recognition could be used for login if access was through the app. During registration, all participants were given a short introduction to how to use eDialogue, including how to send texts and photos and get notifications of new posts.

It was explained to HCPs that they were expected to provide answers to patients' questions with a maximum response time of 24 hours on weekdays. In each individual case, patients decided which of the HCPs in their team of care they wanted to join the digital dialogue, and the HCPs were contacted and invited to join by the primary author (LWHJ). All communication was asynchronous, using text messages and photos; thus, no video calls could be made through the solution. Patients had access for 2 months after hospital discharge. Upon request and agreement with their team of HCPs, access could be extended beyond the study period. The digital dialogues were stored in a secure cloud-based solution [19], and a data processor agreement was made before the study.

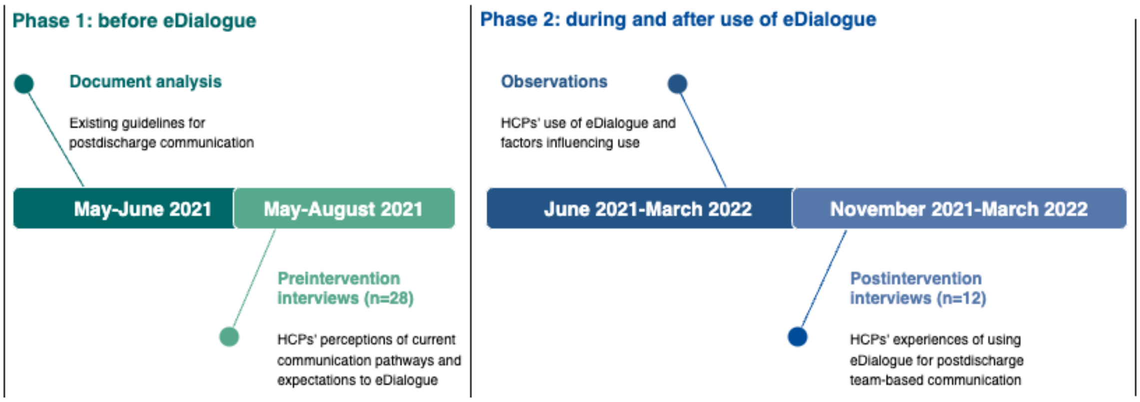
Phase 2: During and After Use of eDialogue

HCPs were recruited for interviews after their use of eDialogue with patients and other HCPs across disciplines and sectors. The inclusion criteria were involvement in eDialogue with ≥3 patients. There were no exclusion criteria.

Data Collection

Data collection was structured according to the two phases to achieve thorough insight into HCPs' perceptions of current communication pathways and their expectations of eDialogue before use (phase 1), and to explore their experiences with access to eDialogue (phase 2). Figure 2 illustrates the triangulation of data collection techniques across the 2 phases of this study.

Figure 2. The figure illustrates the timeline and data collection for phases 1 and 2 of this study. In phase 1, document analysis and preintervention interviews with health care professionals (HCPs) were performed. In phase 2, observations were conducted continuously as eDialogue was used, and, ultimately, postintervention interviews were performed.



Phase 1: Document Analysis and Preintervention Interviews

An initial document analysis of existing guidelines for communication between patients undergoing orthopedic surgery and HCPs across sectors was carried out with the aim of gaining insight into the current context for communication. First, we identified relevant practical documents by searching different Danish web pages related to the political and regulatory guidelines on transitions of care from hospital to home and strategies for using information technologies in health care, for example, the Ministry of Health, the Local Government of Denmark, and the Danish Society for Patient Safety. We also searched the local web page of Aalborg University Hospital for clinical practice guidelines describing the procedures that HCPs must follow when patients or municipal providers contact them regarding discharged patients. Second, we applied a snowball strategy, using references from the initial search. We did not formally analyze the documents, but we used knowledge of the context to understand the framework under which HCPs must work and to qualify the interview guide.

This was followed by focus groups with HCPs across the hospital and municipality (n=28). The aim of the focus groups (n=6) was to explore HCPs' perceptions of current communication pathways and their expectations of eDialogue before use.

The interview guide was inspired by the CFIR Interview Guide Tool [20], including exploratory questions to provide space for emerging reflections. The interview guide was tested on 2 HCPs from the hospital and discussed among the authors until agreement was reached. Minor additions were made before the first focus group.

All preintervention interviews were conducted as semistructured focus groups, dividing HCPs according to their vocational roles and setting (hospital or municipality). HCPs interviewed were surgeons at the hospital (n=5), secretaries from the hospital (n=3), nurses from the hospital ward (n=5), nurses from the outpatient clinic (n=3), home care nurses from the municipality (n=3), physiotherapists from the hospital (n=5), and physiotherapists from the municipality (n=4). Using preexisting groups as focus groups was based on the assumption that it would make participants discuss and compare their reflections in depth in the same context and without an underlying power structure that could occur if professions were mixed [14]. At the beginning of each interview, background variables such as gender, vocational role, and years of experience with patients undergoing orthopedic surgery were collected.

All interviews with HCPs from the hospital were conducted face-to-face by the first author (LWHJ). For the first 2 focus groups, a project nurse was present to register observations during the interviews and to take notes to qualify and supplement the interview. Focus groups with HCPs from the municipality were performed remotely by video, as data collection occurred during the coronavirus outbreak and most HCPs outside of the hospital were not physically located in the same place. The interviewer summarized key points during and at the end of each focus group to facilitate further reflection and to make sure her interpretation corresponded with what the HCPs had said [15]. Field notes were made at the end of each focus group so as to remember details of the context, group interaction, and nonverbal communication [15]. The focus groups lasted an average of 1 hour (between 45 and 90 minutes).

Phase 2: Observations and Postintervention Interviews

In total, eDialogue was used with 31 patients and with the involvement of 24 different HCPs. When the last patient had

had access to eDialogue with their team of HCPs for 2 months, a convenience sample of participating HCPs across the hospital and municipality were interviewed (n=12), including surgeons from the hospital (n=5), physiotherapists from the hospital (n=2), and from the municipality (n=5). We performed 7 individual interviews with physiotherapists across hospitals and municipalities and 1 focus group with 5 surgeons. The aim of the interviews was to explore their experiences with eDialogue. All interviews were conducted by LWHJ, audio recorded, and followed a predefined semistructured interview guide inspired by the CFIR Interview Guide Tool [20] and additional exploratory questions. Interviews with HCPs from the hospital were performed face-to-face, and interviews with HCPs from the municipality were conducted remotely based on the participants' wishes.

During the study period, we observed the use of eDialogue by HCPs and documented this in Word (Microsoft Corporation) files. The aim was to observe issues related to HCPs' use of eDialogue that were reported to the project group or observed in dialogues (an administrator from the project group was present in all dialogues to observe if eDialogue was used in acute situations). HCPs were encouraged to contact the first author if they experienced any problems with eDialogue or had concerns or questions during use, and these were documented as well. Data collected through observations were used to qualify the follow-up interviews in phase 2 and were also imported to NVivo (QSR International) for analysis in conjunction with interview data.

Data Analysis

Data were analyzed for phase 1 and then phase 2, respectively. Interviews were audio recorded using a digital voice recorder (DM-450; Olympus) and transcribed verbatim immediately afterward. Word files with the transcriptions were imported to NVivo for data analysis (NVivo 12, version 20.6.2) [21]. Inspired by Brinkmann and Kvale [15], using an inductive-deductive approach, we performed thematic analysis focusing on meaning (Textbox 1).

Textbox 1. Steps of the thematic analysis using an inductive-deductive approach.

Meaning coding

- Full transcripts were read several times by both LWHJ and REKL.
- To define the initial coding template and to achieve intersubjectivity, the first 4 interviews of each phase were coded by LWHJ and REKL individually before meeting to compare and discuss codes until mutual agreement was achieved. When the coding template was defined, LWHJ applied the same codes to the entire data set. The approach to this step was inductive, thus reflective of the issues raised in the data set.

Meaning condensation

- Theme development was undertaken with a more deductive approach, where domains and constructs from the Consolidated Framework for Implementation Research (CFIR) were used to organize the codes and inform theme development to specifically focus on facilitators and barriers to eDialogue. However, in developing themes, we were open to emerging themes that did not fit the CFIR domains and constructs. Codes and themes were reread and revised by LWHJ in collaboration with REKL and BD in several iterations.

Meaning interpretation

- Definitions and narrative descriptions of themes were made. Data extracts were selected to be presented in the manuscript.
- The final analysis and description of the findings were written.

Data analysis was conducted separately for phase 1 and phase 2 following the 3 steps of meaning coding, meaning condensation, and meaning interpretation. In phase 2, we added notes from observations to the data set to achieve an in-depth understanding of the context in which HCPs had used eDialogue and any problems occurring during use.

Ethical Considerations

The Ethics Committee of Northern Jutland was contacted before the start of the study. They decided by email on March 18, 2021, that the study did not require approval (journal number 2021-000438), as the intervention would not have consequences for diagnostics or treatment. We registered the study at the Regional Committee on Health Research (ID 2021-057). The study followed the Helsinki Declaration, and all participants received both oral and written information as well as thorough guidance in the use of eDialogue. To take into account patients' possible use of eDialogue in emergency situations, an administrator was present in all digital dialogues.

Results

Participant Characteristics

In phase 1, a total of 28 HCPs were recruited across vocational roles and hospital and municipal settings (Table 1). All surgeons, nurses, physiotherapists, and secretaries from the clinical orthopedic surgery subspecialties at the hospital, from which the patients were recruited (deformity correction or anterior cruciate ligament injury), were invited to participate in interviews. However, 2 surgeons, 1 nurse from the outpatient clinic, 1 nurse from the municipality, and 3 secretaries were not able to. Nurses from the ward were purposefully selected based on years of experience and a pragmatic approach to who would be able to participate in interviews during their work hours. On average, HCPs had 11 (range 1-30) years of experience with patients undergoing orthopedic surgery.

Table 1. Vocational roles of health care professionals who were interviewed in phases 1 and 2.

Vocational role	Phase 1 (N=28), n	Involved in eDialogue (n=24), n	Phase 2 (n=12), n
Orthopedic surgeon, hospital	5	7	5
Nurse, outpatient clinic, hospital	3	1	N/A[a]
Nurse, ward, hospital	5	N/A	N/A
Physiotherapist, hospital	5	5	2
Secretary, hospital	3	N/A	N/A
Physiotherapist, municipality	4	11	5
Nurse, municipality	3	N/A	N/A

[a]N/A: not applicable.

In phase 2, a total of 12 HCPs were included for interviews, of whom 8 had also participated in focus groups in phase 1. The HCPs recruited at this stage were a sample of those who had experiences with communication in eDialogue (Table 1). Of whom, 8 HCPs interviewed for phase 2 had also participated in focus groups in phase 1. In total, 24 HCPs across the hospital and municipality were involved in eDialogue. However, we prioritized including those who had been set up to communicate in eDialogue with ≥3 patients. One nurse from the outpatient clinic had been involved in 3 dialogues but was not able to participate due to being absent at the time of the interviews. No nurses from the ward or the municipality were users of eDialogue and thus were not interviewed in phase 2. Secretaries were not interviewed in phase 2, as we decided not to include them in eDialogue at this point.

On average, there were 3.3 (range 2-4) HCPs per patient in the dialogues. All patients were at least connected with the orthopedic surgeon, and 25 of 31 patients had their municipal or hospital-based physiotherapist involved as well.

Themes and Subthemes Identified in Phases 1 and 2

In Table 2, the findings of the analysis of phases 1 and 2 are presented together in main themes organized by the CFIR domains and constructs and additional subthemes. This is to display the before-and-after perspectives of HCPs. Following the table, we elaborate on subthemes in narrative text according to phases 1 and 2 and by using selected quotes from interviews.

The main themes are organized by CFIR domains and constructs, and subthemes elaborate on these for phases 1 and 2, respectively. Emerging themes occurred in both phases that did not match any of the CFIR constructs, and they are therefore described under additional emerging themes.

Table 2. Themes and subthemes from phases 1 and 2 organized by the Consolidated Framework for Implementation Research (CFIR) and additional emerging themes.

Main themes: CFIR domains and constructs	Subthemes: Phase 1: before eDialogue	Subthemes: Phase 2: after eDialogue
Intervention characteristics		
Relative advantage	• Contradictory expectations for using eDialogue versus phone call • A lifeline and reassurance for both patients and HCPs[a] • Hidden work can become visible	• Quick and easy to interact in eDialogue • Photos in eDialogue improve the quality of communication
Adaptability	• N/A[b]	• Development of individual strategies and workflows for the use of eDialogue
Outer setting		
Needs and resources	• Patients are messengers of information between HCPs	• Timely and effective interdisciplinary communication with patients across settings
Inner setting		
Tension for change	• Feeling like an insufficient intermediary • Phone calls are disruptive, yet necessary	• N/A
Relative priority	• Experiences of technology fatigue	• N/A
Compatibility	• N/A	• Divergent perceptions of how well eDialogue meets needs
Available resources	• N/A	• Concerns about resource consumption • Need for clarification regarding financial incentives
Characteristics of individuals		
Self-efficacy	• N/A	• To express oneself in writing
Additional emerging themes		
Previous experiences with digital communication	• Email and SMS text messaging are already used with patients and for interdisciplinary communication; however, standardization is lacking	• N/A
Reflection and learning	• N/A	• A new space for studying patients' needs

[a]HCP: health care professional.

[b]N/A: not applicable.

Phase 1: Current Communication Pathways and Expectations for eDialogue

Intervention Characteristics

Even though the majority of HCPs expected eDialogue to provide optimized interdisciplinary communication, prevent conflicting recommendations to patients, and provide easier access for patients, there were contradictory expectations for the use of eDialogue. On one hand, HCPs had concerns about whether answering messages would require more of their time and go beyond working hours, but on the other hand, they thought it would be easier to answer in eDialogue than by phone. Concerns also centered around whether using text as a communication medium would be adequate for all patients and if misunderstandings would occur due to wrong interpretations. HCPs were especially worried about whether they would pick up on complications to the same extent as they do by phone.

> *I can't hear the patient's voice answering back, and if they have understood my answers (…) however, it depends on the complexity of their questions, whether it's just how many repetitions was it, or something that could be more serious.* [Physiotherapist, municipality, preintervention interview]

Being able to send photos in eDialogue was expected to be an important feature that might offset the challenges of using text for communication. Even though some HCPs had reservations and conflicting opinions about eDialogue before its use, they all agreed that it would be a reassurance and a lifeline for both patients and HCPs across settings. Additionally, a physiotherapist from the hospital reflected on how eDialogue might bring "hidden tasks" to light.

> *What I thought at first would be negative like, oh then we have to do that too, will probably actually be reversed, so that the hidden tasks, we solve by calling and writing notes and emails and things like that, becomes more visible and can be accounted for during work hours.* [Physiotherapist, hospital, preintervention interview]

Outer Setting

The analysis revealed that HCPs experience current communication pathways in the postoperative period to be challenging, both by phone and existing electronic systems. This leads to workarounds, such as HCPs giving patients oral instructions or written notes to bring to other HCPs to ensure timely information. However, using patients as messengers of information between HCPs is perceived as insufficient yet necessary in current communication pathways. A physiotherapist from the municipality described how current systems do not support the patient's trajectory across sectors.

> *There are watertight shutters between the communication systems, i.e., what they write in the medical record at the hospital and what I write here. The surgeon at the hospital can't see that, and (…) I can't see his note.* [Physiotherapist, municipality, preintervention interview]

Inner Setting

Across professional roles and settings, HCPs expressed a need for change to enable easier sharing of knowledge and communication. This was especially the case for complex and long-term orthopedic surgical treatments where multiple HCPs are involved. Knowing each other across settings, for example, by being former colleagues, was a mediating factor for communication between HCPs. However, it was not perceived as sustainable.

Getting in contact with each other and patients by phone is considered time consuming due to the synchronicity of phone calls. A nurse from the inpatient department described how phone calls would sometimes be left until the next day if questions required the involvement of another HCP. This left the nurses feeling like inadequate intermediaries and could be a risk to patient safety. Similar experiences were described by physiotherapists, who often found themselves being asked about issues outside of their competencies; for example, questions about wounds and medication.

HCPs from the hospital described how phone calls are disruptive to their work processes, even though they understood the need for them. In addition to inquiries from patients, they receive phone calls from a wide range of HCPs in hospitals, municipalities, and private settings. Although secretaries act as gatekeepers, nurses from the outpatient clinic, and in the inpatient department in particular, handle many phone calls daily.

> *It's constant, isn't it? (...) it takes my attention away from the dialogue, the communication and the relationship that I'm in the middle of. Then you're like, oh sorry, this phone call is actually more important than you are (the patient they are with).* [Nurse, outpatient clinic, hospital, preintervention interview]

Addressing eDialogue as a novel communication solution to support team-based communication between patients and HCPs across settings, most HCPs were positive about the change it might bring. However, they expressed some degree of technology fatigue that made them skeptical of yet another system without integration into existing systems.

Emerging Theme

HCPs described previous experiences with using digital communication with patients, usually by email or SMS text messaging. Most often, it is used as a way to provide psychological reassurance to patients or to solve specific complex problems, where the HCPs have specialist knowledge. Even FaceTime was described as being used once with a patient to inspect a wound from a distance. However, the disadvantages of the current nonsystematic use of digital communication with patients were reflected. Concerns were raised regarding using a private phone number and the risk of introducing data security breaches. Also, giving some patients the opportunity for direct digital contact and others not was perceived as problematic. Thus, if used inconsistently, it may lead to inequality in patients' access to health care.

Furthermore, HCPs described how they use email or SMS text messaging to communicate with each other, for example, to share thoughts on treatments or rehabilitation. They do this as a workaround to traditional communication pathways or because it is perceived as less disturbing to each other. Thus, the use of digital text-based communication is not uncommon for HCPs in this study. However, it is not standardized or even articulated among colleagues or management.

Phase 2: HCPs' Perspectives of eDialogue After Use

Intervention Characteristics

All HCPs agreed that the technical solution for eDialogue was very intuitive and did not need a thorough introduction, as opposed to other solutions with more features. Most HCPs articulated that questions were quick and easy to handle during work hours. Especially the asynchrony of the contact and the use of photos improved the quality of communication and their experiences of eDialogue for patient communication.

> *The big advantage of this, is that they can send a photo (…). If it wasn't a possibility, I think there would be a lot of writing about something that we couldn't really clarify, and then we would still have to call them in (for an extra check). Being able to send a photo, that's really crucial for this to work.* [Surgeon, hospital, postintervention interview]

The analysis demonstrated that HCPs developed individual strategies for answering questions in eDialogue. Notifications were automatically sent to participants when there were new messages in the system, but there were no integrated reminders to follow up if the messages were not read within 24 hours, and this led to the development of individual workflows.

> *(The notification) on email, when there is a new message, I will not delete it until I have answered. That way, it helps me keep track.* [Surgeon, hospital, postintervention interview]

eDialogue was mainly used by patients as a place to ask postdischarge questions to HCPs. In general, most questions from patients were answered by surgeons and physiotherapists from the hospital. Municipal physiotherapists described being hesitant to involve themselves actively in answering, as they experienced hospital staff being quick to answer the patients. However, they emphasized that they used the information given to the patient by hospital staff in their subsequent contact with patients. This "indirect" use was perceived as valuable to them.

> *It has been very rewarding to just follow the dialogue, even though I was not active in it. The fact that the patient can just send a photo and ask 'what does this look like?', then he is immediately calmed down. It's rather smart, and also that I know of it right away.* [Physiotherapist, municipality, postintervention interview]

Outer Setting

HCPs stressed that the team-based approach made it easier to share timely information with the patient and other HCPs, and thereby it created more effective communication pathways. Physiotherapists highlighted how their previous perceptions of being an insufficient intermediary between the patient and other HCPs were changed when communication could take place directly in eDialogue.

> *It was actually really nice that he (the patient) just took it directly with the surgeon. Because I can have doubts (…) and you don't want to burden the surgeon by calling.* [Physiotherapist, municipality, postintervention interviews]

Inner Setting

Even though HCPs acknowledged the impact that eDialogue had for patients, there were discrepancies in their perception of how it was used in this study, and it affected their acceptance of the solution. For example, some HCPs thought that the team-based approach was not necessary for all patients involved or that they lacked a secretary for administrative tasks. As such, they highlighted that some questions might be better answered in other ways, for example, by providing better patient education or by including other HCPs in the dialogue.

> *I think it is difficult to say that the patients' questions are not relevant because they must be since they ask them, but who should answer them, and how quickly should they have an answer, can be discussed.* [Physiotherapist, hospital, postintervention interview]

However, when using eDialogue with patients for complex orthopedic treatments, HCPs expressed that the team-based approach was very valuable to the patients and their workflows.

> *I think it was good. They (patients) feel that there is a team around them, and I get the feeling that I'm not the only one being responsible. Also, I don't have to spend time calling the physiotherapist to say 'Hey, can't you just look at this?' when he's already in the dialogue.* [Surgeon, hospital, postintervention interview]

HCPs strongly experienced that access to eDialogue provided reassurance for patients. However, in consideration of the sparse health care resources, it was a general opinion that eDialogue should only be offered to patients for complex treatments. This provoked an ethical discussion of how HCPs could distinguish between who should be offered the solution and who should not. HCPs highlighted that an assessment of effects should be addressed, both in terms of resource consumption and patient outcomes.

> *One of my concerns with systems like this is that if we have to use it with all patients (…), then I think it could become a burden. And also, I think it will be difficult to say, well, it's only for some patients, because why them?* [Surgeon, hospital, postintervention interview]

HCPs agreed that clarification is needed regarding financial incentives before implementing eDialogue. Along with concerns about resources to answer the questions, this was a perceived barrier to use.

> *I think the barriers are time and finances (...) there is, of course, someone who looks at what I produce. And I think it should be some kind of service that should be visible (to others), if we have to evaluate a photo or send back a response (through eDialogue).* [Physiotherapist, municipality, postintervention interview]

Characteristics of Individuals

In all interviews, HCPs had concerns about whether they expressed themselves clearly enough in writing and how their "tone of voice" would be perceived by patients when formulated in texts. In reflection, they emphasized that the same concerns could arise when talking to patients on the phone.

> *Regardless of whether it's something you say to them or something you text them, it's just as important that you use words they can understand, and I actually often think it's a little easier when you text because you have time to think about it.* [Surgeon, hospital, postintervention interview]

There were clear differences in how HCPs expressed themselves in the texts, and this was discussed in one of the focus groups, where a surgeon had been involved in another surgeon's dialogue due to vacation.

> *I think he (the other surgeon) is very kind in his feedback. I actually noticed that, you (addressed to the other surgeon) have formulated yourself in such a very friendly way, in contrast to what I did to start with. I made it very short, like I might normally answer a text message with a friend (…). I had to remind myself that they don't know me (…) it might be important to pay attention to that.* [Surgeon, hospital, postintervention interview]

Emerging Theme

Both surgeons and physiotherapists described that using eDialogue created interdisciplinary reflection and learning about patients' needs after discharge, and that frequently asked questions could be used to improve future patient education.

> *It gives feedback in relation to the material we use and the way we inform patients now. It might actually be very nice for all of us to know this.* [Physiotherapist, hospital, postintervention interview]

Ultimately, HCPs pointed out that they could learn from each other by reading each other's answers to patients.

Discussion

Principal Findings and Comparison With Previous Work

This study first investigated HCPs' perceptions of current communication pathways with patients and other HCPs involved in the patient's trajectory after orthopedic surgery and discharge, along with their expectations for eDialogue before its use (phase 1). Following initial document analysis, we included a wide range of HCPs across vocational roles and settings in focus groups to obtain an in-depth understanding of their needs and attitudes toward eDialogue. These perspectives are important to capture, as individual and contextual factors as well as initial perceptions of eDialogue may motivate or hinder use [17]. The findings of phase 1 showed that, on the one hand, HCPs perceived a significant tension for change. Current communication pathways are perceived as insufficient, phone calls are disruptive, and patients unfortunately become messengers of information between HCPs across settings. On the other hand, HCPs expressed conflicting attitudes toward eDialogue in advance of its use. Positive or negative attitudes were not limited to certain vocational roles but were expressed in all groups and also as an internal dilemma inherent to the individual. However, there were clear expectations for eDialogue to support patients in the postoperative period and consensus that it may provide optimized interdisciplinary and cross-sectoral communication. At the same time, HCPs experienced some degree of technology fatigue and significant worry that eDialogue would be time-consuming for them to handle.

Second, we explored HCPs' experiences of using eDialogue for team-based digital communication through observations and postintervention interviews (phase 2). Knowing that, even with highly developed plans for execution, undiscovered factors can undermine implementation efforts in the real world [17,18], we searched to identify facilitators and barriers to implementation from the perspectives of key users at an early stage. Findings from phase 2 showed that HCPs experienced eDialogue as a quick and easy way to interact with patients and other HCPs and that eDialogue could support timely and effective interdisciplinary communication across settings. As such, the positive perceptions of the importance of eDialogue described in the preintervention interviews were maintained. Similarly, the use of photos was expected to be important in preintervention interviews, and in postintervention interviews, photos were even suggested as being a significant quality-enhancing element compared to traditional phone calls. Similar findings have been described in other studies of digital communication in health care [22-24].

In interviews in phase 1, HCPs described that they had concerns about communicating with patients in texts because they feared overlooking an important complication or that the patient would misunderstand their written responses. In phase 2, HCPs still expressed concerns about whether they expressed themselves clearly enough. However, they pointed out that the same risks can be present in phone consultations. This perspective is supported by a recent study of telephone consultations in Denmark. Jensen et al [25] found that communication in consultations concerning back pain preceding out-of-hospital cardiac arrest was influenced by the communicative preconditions of the call-taker, thereby addressing the fact that a meaning-constitution is undertaken in the interaction between the patient and the call-taker, not always reflecting the actual problem. To learn from this, HCPs involved with patients through eDialogue and other digital communication solutions must be aware that communicative interaction is always an interpretative task for the receiver of a message. Even though the HCPs' concern might decrease as they gain more experience communicating in writing, their self-efficacy should be supported by formulating clear recommendations, training, and supervision.

Across the interviews of phases 1 and 2, HCPs expressed concerns regarding resource consumption; this was particularly evident among hospital staff. While acknowledging patients' need for easier access to communication with HCPs after discharge, HCPs questioned if the team-based approach was necessary for patients undergoing less complex orthopedic treatments. Nevertheless, there was consensus that eDialogue can support patients in complex and long-term treatments and that a needs assessment to learn who will benefit the most from eDialogue should be made before its implementation so as to best match resources with actual needs. Other studies investigating the use of team-based digital communication have primarily focused on patients with cancer or chronic diseases [26-29]. Patients undergoing orthopedic surgery for complex and long-term treatment suffer similar challenges in health care communication [30], and therefore it is also relevant to develop and test solutions for this group. By using eDialogue for a smaller patient group, the workload caused by the implementation of the solution will decrease.

eDialogue was a solution where both patients and HCPs across settings could communicate freely in the postdischarge period. However, the primary communication in eDialogue was between the patient and HCPs at the hospital. Municipal physiotherapists used eDialogue more indirectly as a way to keep up to date with the patients' progress. As such, findings revealed how physiotherapists in the municipality and patients together would formulate questions to send to the hospital staff. Taking into account this shared use of eDialogue, usage data defining the proportion of messages sent between patients and HCPs and between HCPs across settings would not be representative of their actual use. Moreover, HCPs adapted eDialogue to their contexts and developed individual strategies for providing timely answers. Some strategies were developed because the technical solution lacked better adaptation to the context, for example, an improved notification system, whereas other strategies were based on individual preferences in handling digital communication. All cases emphasize the importance of uncovering the HCPs' context and needs and ensuring that new technology supports them in their work processes so that inappropriate use of new solutions does not end up adding new workarounds and thus hindering the optimal outcome of the technology.

Limitations

This study was inspired by the CFIR to guide data collection and analysis [17,20]. The systematic identification and mapping of what was perceived as important to HCPs to the CFIR domains and constructs was helpful in providing an overview of the multifaceted and conflicting attitudes and experiences of eDialogue. However, we did not apply the CFIR as exhaustively as recommended [17,18], and we may thereby have missed important aspects that could have emerged. Using an inductive-deductive approach in data analysis, however, allowed us to still be explorative, which suited the early phase of the intervention described in this study.

In phase 1, we included a wide range of HCPs involved in the patients' trajectory and communicative circles after surgery and discharge to shed light on their perspectives on current communication pathways. Including HCPs from different settings was a strength to this study, however, the small subgroups of HCPs from the same setting may jeopardize data saturation [15]. However, the theme of the interviews, exclusively focusing on communication, is narrow and may thereby outweigh this issue. For preintervention interviews, data saturation was reached; however, it can be discussed whether data saturation was reached fully for the interviews in phase 2. Observations of HCPs' use of eDialogue, including technical or collaborative issues that were encountered during use, accounted for this and were included in the data analysis for phase 2.

Furthermore, we could have included management and decision makers in the focus group to gain a deeper understanding of the political and managerial context of the use of eDialogue across sectors. However, this was not attempted in this study as we wished to focus on the end users' perspectives.

Our findings derive from a single hospital and a municipal region in Denmark. Therefore, they may not reflect the experiences of HCPs from other parts of the country, where different digital communication solutions have been implemented. Only 1 nurse participated in communication in eDialogue, and thus the experiences of this group of HCPs are not reflected in our findings. Unfortunately, at the start of this study, the coronavirus outbreak was at its peak, and many nurses from the hospital were reassigned to newly opened COVID-19 departments. At the same time, there was a trade union strike among nurses in Denmark, which resulted in the cessation of work for a period of time for many nurses from the municipality. These circumstances put greater work pressure on the nurses, and we continued the study without their active involvement in the dialogues.

Last but not least, some of the research team members behind this study are clinicians and were involved in the decision to test eDialogue. We have tried to overcome this issue by including research team members with little knowledge of the patients and processes in orthopedic surgery. Thus, 2 independent researchers coded and condensed data (LWHJ and REKL) in close discussion with BD, where REKL did not have preliminary knowledge of the context.

Conclusions

HCPs describe current communication pathways as complicated. Phone calls are disruptive to work processes, and the lack of direct communication modalities between patients and HCPs across settings in the postoperative period makes patients become messengers of information between HCPs. To overcome these challenges, HCPs use off-the-shelf digital communication solutions as a workaround; however, use is neither standardized nor data secure. HCPs were open to using eDialogue, although they had reservations, which were partly confirmed and unconfirmed in their subsequent use of eDialogue. Especially, concerns regarding resource consumption were highlighted, and HCPs suggested the solution is particularly valuable in complex and prolonged treatments. The use of eDialogue offers a potentially valuable strategy for future integration of communication across health care settings, breaking down existing silos and taking into account the whole care team and the patient. This study provides knowledge for future strategies for implementing such solutions in orthopedic surgery and other clinical domains.

Acknowledgments

We would like to thank all patients and parents, health care providers, and management staff from departments across hospitals and municipalities in Northern Jutland, Denmark, who agreed to participate in this study. We would also like to thank the project nurse, who was an observer in focus groups, and the students from Aalborg University, who contributed to transcribing large amounts of interview data. This study was partly funded by the Research Fund of Danish Health Region North (journal number 2021-0019).

Conflicts of Interest

None declared.

References

1. Herndon JH, Pollick KJ. Continuing concerns, new challenges, and next steps in physician-patient communication. J Bone Joint Surg Am 2002;84(2):309-315. [doi: 10.2106/00004623-200202000-00019] [Medline: 11861738]
2. Naylor MD, Aiken LH, Kurtzman ET, Olds DM, Hirschman KB. The care span: the importance of transitional care in achieving health reform. Health Aff (Millwood) 2011;30(4):746-754. [doi: 10.1377/hlthaff.2011.0041] [Medline: 21471497]
3. Forster AJ, Murff HJ, Peterson JF, Gandhi TK, Bates DW. The incidence and severity of adverse events affecting patients after discharge from the hospital. Ann Intern Med 2003;138(3):161-167. [doi: 10.7326/0003-4819-138-3-200302040-00007] [Medline: 12558354]
4. Oslock WM, Satiani B, Way DP, Tamer RM, Maurer J, Hawley JD, et al. A contemporary reassessment of the US surgical workforce through 2050 predicts continued shortages and increased productivity demands. Am J Surg 2022;223(1):28-35. [doi: 10.1016/j.amjsurg.2021.07.033] [Medline: 34376275]
5. Wainwright T, Middleton R. An orthopaedic enhanced recovery pathway. Curr Anaesth Crit Care 2010;21(3):114-120 [FREE Full text] [doi: 10.1016/j.cacc.2010.01.003]
6. Pereno A, Eriksson D. A multi-stakeholder perspective on sustainable healthcare: from 2030 onwards. Futures 2020;122:102605 [FREE Full text] [doi: 10.1016/j.futures.2020.102605] [Medline: 32834076]
7. Danbjørg DB, Wagner L, Kristensen BR, Clemensen J. Intervention among new parents followed up by an interview study exploring their experiences of telemedicine after early postnatal discharge. Midwifery 2015;31(6):574-581. [doi: 10.1016/j.midw.2015.02.007] [Medline: 25765743]
8. Holm KG, Brødsgaard A, Zachariassen G, Smith AC, Clemensen J. Parent perspectives of neonatal tele-homecare: a qualitative study. J Telemed Telecare 2019;25(4):221-229. [doi: 10.1177/1357633X18765059] [Medline: 29792079]
9. Cronin RM, Davis SE, Shenson JA, Chen Q, Rosenbloom ST, Jackson GP. Growth of secure messaging through a patient portal as a form of outpatient interaction across clinical specialties. Appl Clin Inform 2015;6(2):288-304 [FREE Full text] [doi: 10.4338/ACI-2014-12-RA-0117] [Medline: 26171076]
10. Baer D. Patient-physician e-mail communication: the kaiser permanente experience. J Oncol Pract 2011;7(4):230-233 [FREE Full text] [doi: 10.1200/JOP.2011.000323] [Medline: 22043186]
11. Healthcare in Denmark: An Overview. Ministry of Health. 2017. URL: https://sum.dk/Media/C/A/Healthcare-in%20denmark%20an%20overview%20english-V16-dec.pdf [accessed 2024-02-14]
12. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
13. DeWalt KM, DeWalt BR. Participant Observation: A Guide for Fieldworkers, Second Edition. Lanham, Md: AltaMira Press; 2011.
14. Morgan DL. Focus Groups as Qualitative Research, Second Edition. Thousand Oaks, CA: SAGE Publications; 1997.
15. Brinkmann S, Kvale S. Interviews: Learning the Craft of Qualitative Research Interviewing, Third Edition. København: SAGE Publications; 2015.
16. Tong A, Sainsbury P, Craig J. Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
17. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
18. Kirk MA, Kelley C, Yankey N, Birken SA, Abadie B, Damschroder L. A systematic review of the use of the consolidated framework for implementation research. Implement Sci 2016;11:72 [FREE Full text] [doi: 10.1186/s13012-016-0437-z] [Medline: 27189233]
19. First Agenda. LetDialog—Sikker SMS. URL: https://letdialog.dk/ [accessed 2023-02-02]
20. The CFIR interview guide tool. CFIR Research Team-Center for Clinical Management Research. URL: https://cfirguide.org/tools/ [accessed 2023-02-02]
21. NVivo Qualitative data analysis software. QSR International. 1999. URL: https://qsrinternational.com/nvivo/nvivo-products/ [accessed 2023-02-02]
22. Lindkvist RM, Sjöström-Strand A, Landgren K, Johnsson BA, Stenström P, Hallström IK. "In a way we took the hospital home"-a descriptive mixed-methods study of parents' usage and experiences of eHealth for self-management after hospital discharge due to pediatric surgery or preterm birth. Int J Environ Res Public Health 2021;18(12):6480 [FREE Full text] [doi: 10.3390/ijerph18126480] [Medline: 34203985]
23. DeAntonio JH, Kang HS, Cockrell HC, Rothstein W, Oiticica C, Lanning DA. Utilization of a handheld telemedicine device in postoperative pediatric surgical care. J Pediatr Surg 2019;54(5):1005-1008. [doi: 10.1016/j.jpedsurg.2019.01.032] [Medline: 30782441]
24. Gund A, Sjöqvist BA, Wigert H, Hentz E, Lindecrantz K, Bry K. A randomized controlled study about the use of eHealth in the home health care of premature infants. BMC Med Inform Decis Mak 2013;13:22 [FREE Full text] [doi: 10.1186/1472-6947-13-22] [Medline: 23394465]

25. Jensen B, Vardinghus-Nielsen H, Mills EHA, Møller AL, Gnesin F, Zylyftari N, et al. "Like a rainy weather inside of me": qualitative content analysis of telephone consultations concerning back pain preceding out-of-hospital cardiac arrest. Int Emerg Nurs 2022;64:101200 [FREE Full text] [doi: 10.1016/j.ienj.2022.101200] [Medline: 35926318]
26. Ekstedt M, Børøsund E, Svenningsen IK, Ruland CM. Reducing errors through a web-based self-management support system. Stud Health Technol Inform 2014;201:328-334. [Medline: 24943563]
27. Voruganti T, Grunfeld E, Jamieson T, Kurahashi AM, Lokuge B, Krzyzanowska MK, et al. My team of care study: a pilot randomized controlled trial of a web-based communication tool for collaborative care in patients with advanced cancer. J Med Internet Res 2017;19(7):e219 [FREE Full text] [doi: 10.2196/jmir.7421] [Medline: 28720558]
28. Husain A, Cohen E, Dubrowski R, Jamieson T, Kurahashi AM, Lokuge B, et al. A clinical communication tool (Loop) for team-based care in pediatric and adult care settings: hybrid mixed methods implementation study. J Med Internet Res 2021;23(3):e25505 [FREE Full text] [doi: 10.2196/25505] [Medline: 33656445]
29. Gulmans J, Vollenbroek-Hutten M, van Gemert-Pijnen L, van Harten W. A web-based communication system for integrated care in cerebral palsy: experienced contribution to parent-professional communication. Int J Integr Care 2012;12:e9 [FREE Full text] [doi: 10.5334/ijic.672] [Medline: 22977435]
30. Balakrishnan AS, Palmer NR, Odisho AY. "Can you hear me now?": postoperative patient-initiated communication with providers. Am J Manag Care 2020;26(10):e333-e341 [FREE Full text] [doi: 10.37765/ajmc.2020.88507] [Medline: 33094946]

Abbreviations

CFIR: Consolidated Framework for Implementation Research
COREQ: Consolidated Criteria for Reporting Qualitative Research
HCP: health care professional

Edited by G Eysenbach; submitted 09.10.23; peer-reviewed by A Mahnke, T Tang; comments to author 02.11.23; revised version received 21.11.23; accepted 31.01.24; published 08.03.24.

Please cite as:
Jensen LWH, Rahbek O, Lauritsen REK, Kold S, Dinesen B
Health Care Professionals' Perspectives Before and After Use of eDialogue for Team-Based Digital Communication Across Settings: Qualitative Study
JMIR Hum Factors 2024;11:e53391
URL: https://humanfactors.jmir.org/2024/1/e53391
doi:10.2196/53391
PMID:38457798

Original Paper

Inefficient Processes and Associated Factors in Primary Care Nursing: System Configuration Analysis

Willi L Tarver[1,2], MLiS, DrPH; April Savoy[2,3,4], PhD; Himalaya Patel[2], PhD; Michael Weiner[2,4,5], MD, MPH; Richard J Holden[4,5,6], MS, PhD

[1]Division of Cancer Prevention and Control, Department of Internal Medicine, The Ohio State University, Columbus, OH, United States

[2]Health Systems Research Center for Health Information and Communication (13-416), Richard L Roudebush Veterans Affairs Medical Center, United States Department of Veterans Affairs, Indianapolis, IN, United States

[3]School of Industrial Engineering, Purdue University, Indianapolis, IN, United States

[4]Regenstrief Institute, Inc, Indianapolis, IN, United States

[5]School of Medicine, Indiana University, Indianapolis, IN, United States

[6]School of Public Health, Indiana University, Bloomington, IN, United States

Corresponding Author:
April Savoy, PhD
School of Industrial Engineering
Purdue University
799 W. Michigan St. ET 201
Indianapolis, IN, 46202
United States
Phone: 1 3172782194
Email: asavoy@purdue.edu

Abstract

Background: Industrywide, primary care nurses' work is increasing in complexity and team orientation. Mobile health information technologies (HITs) designed to aid nurses with indirect care tasks, including charting, have had mixed success. Failed introductions of HIT may be explained by insufficient integration into nurses' work processes, owing to an incomplete or incorrect understanding of the underlying work systems. Despite this need for context, published evidence has focused more on inpatient settings than on primary care.

Objective: This study aims to characterize nurses' and health technicians' perceptions of process inefficiencies in the primary care setting and identify related work system factors.

Methods: Guided by the Systems Engineering Initiative for Patient Safety (SEIPS) 2.0 model, we conducted an exploratory work system analysis with a convenience sample of primary care nurses and health technicians. Semistructured contextual interviews were conducted in 2 sets of primary care clinics in the Midwestern United States, one in an urban tertiary care center and the other in a rural community-based outpatient facility. Using directed qualitative content analysis of transcripts, we identified tasks participants perceived as frequent, redundant, or difficult, related processes, and recommendations for improvement. In addition, we conducted configuration analyses to identify associations between process inefficiencies and work system factors.

Results: We interviewed a convenience sample of 20 primary care nurses and 2 health technicians, averaging approximately 12 years of experience in their current role. Across sites, participants perceived 2 processes, managing patient calls and clinic walk-in visits, as inefficient. Among work system factors, participants described organizational and technological factors associated with inefficiencies. For example, new organization policies to decrease patient waiting invoked frequent, repetitive, and difficult tasks, including chart review and check-in using tablet computers. Participants reported that issues with policy implementation and technology usability contributed to process inefficiencies. Organizational and technological factors were also perceived among participants as the most adaptable. Suggested technology changes included new tools for walk-in triage and patient self-reporting of symptoms.

Conclusions: In response to changes to organizational policy and technology, without compensative changes elsewhere in their primary care work system, participants reported process adaptations. These adaptations indicate inefficient work processes. Understanding how the implementation of organizational policies affects other factors in the primary care work system may improve the quality of such implementations and, in turn, increase the effectiveness and efficiency of primary care nurse processes.

Furthermore, the design and implementation of HIT interventions should consider influential work system factors and their effects on work processes.

(JMIR Hum Factors 2024;11:e49691) doi:10.2196/49691

KEYWORDS

health information technology; mobile devices; nursing and nursing systems; outpatient care; SEIPS 2.0; work-system analysis

Introduction

To meet the rising demand for primary care services [1], the role of primary care nurses is becoming more complex and team-based [2]. Additional industry-wide changes that further complicate primary care nurses' roles include greater autonomy in care management [3] and growing telehealth duties (eg, managing patients using videoconferencing and remote patient home monitoring) [4], in addition to in-person care. Mobile health information technologies (HITs) such as laptops, tablets, and smartphones, particularly among nurses, have been used to facilitate flexibility in documentation, communication, and other tasks that would typically take nurses' attention away from their patients [5-7]. However, inefficient HITs may increase nurses' work burden and lead to unexpected changes in their roles and the dynamics of their care teams [8].

A better understanding of the work system for nurses and their needs can inform the design, development, and successful implementation of these technologies [9]. The few identified systems-level studies involving primary care nurses have demonstrated the usefulness of such a perspective [10,11]; however, to our knowledge, research specifically on primary care nurses' work systems is sparse. To better understand nursing tasks and processes and the work system specific to the primary care context, accounting for the various factors affecting nurses' work is needed [12]. Incorporating work system factors, including people, technology, tasks, organizations, and environments [13], may lead to improvements in the design and implementation of HITs. Furthermore, while each of these constructs is uniquely important, assessing them individually fails to capture how these constructs interact with each other. Accurately eliciting and identifying the needs of this population and understanding the work system specific to the primary care context requires accounting for the various factors affecting nurses' work [12], accounting that may be served well by a systems-level human factors perspective.

The objective of this study is to (1) characterize nurses' perceptions of process inefficiencies in the primary care setting and (2) describe related work-system factors. The Systems Engineering Initiative for Patient Safety (SEIPS) 2.0 model provides a user-centered, systems-level view of work system structure, processes, and outcomes in health care and their relationships [13]. The SEIPS 2.0 model posits that the sociotechnical work system produces work processes which shape outcomes [13]. Understanding how these factors interact has important implications for the nurses' workflow, and the implementation of interventions designed to aid them in completing tasks may influence their beneficial or adverse effects on clinical care. In addition, as more HITs are being introduced into the primary care setting [14,15], our findings will serve as an important step in understanding how to best design and implement these technologies to support primary care nurses.

Methods

Study Design

This was an exploratory study of the work system of primary care nurses and health technicians. Guided by SEIPS 2.0 [13], we conducted contextual interviews, observed work activities, used directed content analysis methods to identify findings, and used configural diagramming to organize and report the findings.

Setting

Our study focused on ambulatory care settings. Recruitment and data collection occurred at 2 sets of primary care clinics in the Midwestern United States. Site 1 was an urban, tertiary care medical center in a large city; site 2 was a rural, community-based outpatient facility in a small town. At both sites, nurses and health technicians regularly interact with an electronic health record (EHR) system to complete nursing processes. At site 1, some clinics distributed laptop computers to nurses, while all other staff used desktop computers in staff workrooms and examination rooms. Site 1 was a participant in the US Department of Veterans Affairs (VA) Mobile Health Provider Program launched in 2014; through this program, over 12,000 Apple iPads have been distributed at more than 60 VA sites, though device usage has been reportedly less than expected [16]. At site 2, nurses and health technicians had open workrooms and used ruggedized portable computers (Panasonic Toughbook CF-H2). Docks for mobile devices were installed in workrooms and examination rooms.

Recruitment

Convenience sampling was used to identify nurses and health technicians at the primary care clinics. In the clinics of this health care system, health technicians work under the supervision of registered nurses to maintain the documents and records used in primary care nursing processes. A list of eligible primary care staff providing care in the clinic was obtained. Staff members were contacted by email to solicit participation. Primary care nurses and health technicians were subsequently contacted in person to gain their consent to engage in the interview process.

Conceptual Model

SEIPS states that a person (eg, health care professional) performs tasks in the clinical care setting that require various tools and technologies (eg, HITs). The use of these tools and technologies to perform these clinical tasks occurs within a physical internal environment governed by organizational conditions as well as a broader external environment [13]. These components make up the work system, interact with each other, and influence each other. Variations in how these components interact can be associated with workflow and health outcomes. Furthermore, SEIPS 2.0 introduces concepts of configurations and adaptations [13]. For example, with each of the work system factors that can interact with one another, the concept of configuration acknowledges that not all of these components are relevant to each process or situation. More specifically, configuration pertains to the subset of components and their interactions that are actually relevant to a particular process or situation [13]. According to SEIPS 2.0, adaptations refer to the changes that have been attempted to decrease the gap between actual and ideal performance [17].

Contextual Interviews

We conducted semistructured contextual interviews among primary care nurses and health technicians. This method of interviewing allows researchers to observe and ask clarifying questions to participants while they are working [18]. Participants assume the role of the expert and are able to demonstrate tasks while working, which may also prompt the discussion of tasks that they may not consider important to the topic during a traditional interview. Researchers, on the other hand, assume the role of a student or apprentice, trying to understand the work process to identify ways of improving it or implementing interventions to address any underlying problems or challenges [19].

A semistructured interview guide (Multimedia Appendix 1) was created by the research team based on SEIPS 2.0 [13]. Interview topics included (1) nurses' perspectives on process inefficiencies in primary care; (2) tasks that were considered frequent, repetitive, difficult, and related to inefficient processes; (3) the types of information needed to complete tasks; (4) the tools and technology needed to complete tasks; (5) organizational factors or policies that affect primary care nurses' abilities to complete tasks; and (6) the use of mobile applications. Terms and their associated definitions were provided to participants before the interview to establish a shared understanding (more details in Table 1). To match the study's focus on HIT, the task scope was limited to clinical and administrative information.

Table 1. Terms and definitions used during interviews with primary nurses and health technicians on process inefficiencies.

Term	Definition
Information-intensive tasks	Require reading, writing, or sharing information (eg, chart review).
Frequent tasks	Performed often or for each patient (eg, looking up patients' contact information or reviewing discharge summaries).
Repetitive tasks	Tasks done repeatedly that should only be done once or not at all (eg, repetitious logins or clicks to access required information).
Difficult tasks	Tasks requiring large amounts of concentration to complete (eg, reviewing labs or determining trends in vitals).

A total of 6 nonclinical researchers (4 with previous interviewing experience, including coauthor HP and 2 volunteers) conducted interviews using a prepared interview guide (Multimedia Appendix 1). Early interviews were led by the 4 experienced staff members, with 1 or 2 other researchers serving as notetakers. These early interviews served as training for the volunteers. In later interviews, researcher roles were rotated to limit the influence of any single interviewer.

Each session was led by 1 interviewer and 1 note-taker. Interviews lasted approximately 45 minutes and were conducted either in the general practice setting of the participants or in a private room. Each interview was audio-recorded and transcribed. Transcripts were done through a contracted professional service. Staff research assistants corrected major transcription errors and removed personal identifiers.

Analysis of Contextual Interviews

Initial qualitative analysis was done by 4 staff research assistants. All had previously served as interviewers. Each transcript was coded fully by 2 staff research assistants. Segments of data (eg, a phrase, sentence, or group of sentences) were coded iteratively. In the first iteration of coding, we identified work system components. We then identified processes performed by primary care nurses and whether they were perceived as frequent, repetitive, or difficult. Furthermore, 3 analysts (WT, AS, and HP) conducted a directed content analysis guided by the SEIPS 2.0 model to identify work system configurations related to frequent, repetitive, and difficult tasks for specific processes [20,21].

Configuration Analysis

The interactions of the coded work system components and processes were used to identify work system configurations. Interactions are defined as segments of data assigned to 2 or more coded components. We reviewed key findings with emphasis on the participant quotes and descriptions of processes and tasks to identify the tasks identified as most influential for each process. Next, we independently defined work system configurations for each frequently reported process and met to discuss and resolve discrepancies. We created configural diagrams [13] of work system elements related strongly to the identified inefficient processes. We then identified misalignments among work system factors. We define misalignment as a mismatch between human and nonhuman (eg, environment, policies, etc) factors that may lead to a breakdown of processes. Based on findings from the configurational analysis, we characterized adaptations as workarounds (ie, current adaptations) or recommendations (ie, future adaptations).

Ethical Considerations

This study complied with the American Psychological Association Code of Ethics and was approved by the Research and Development Committee at the Richard L. Roudebush

Veterans Affairs Medical Center and the Indiana University institutional review board (protocol #1611241830).

Results

Among the 51 eligible staff members who were contacted, 20 nurses and 2 health technicians participated in this study and completed interviews (Table 2). Most participants were female (19/22, 86%); most participants were White (17/22, 77%). Participants had a mean of 8.3 (SD 8.7, range 1-36) years of experience with their current health care employer and 11.9 (SD 8.9, range 1-30) years in their present role.

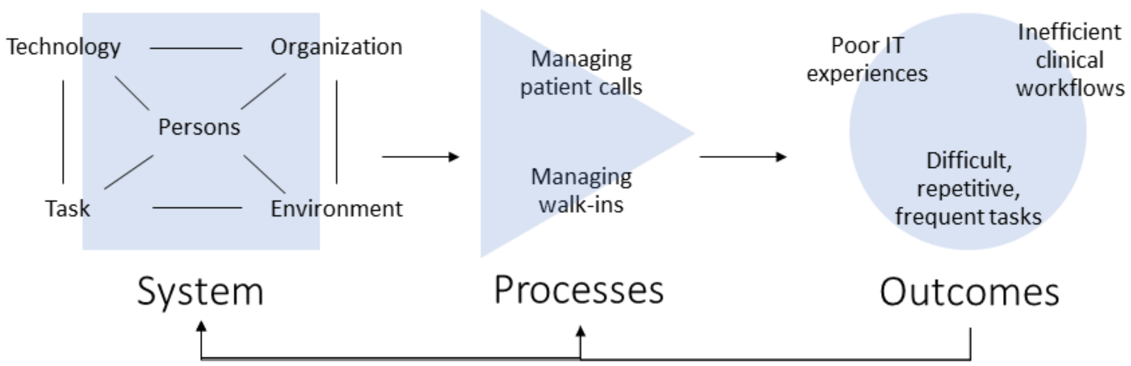
Based on our analysis, Figure 1 depicts the relationship among the nursing work system, processes, and perceived outcomes (Figure 1). Perceived inefficient workflows were associated with managing patient calls (ie, patient response calls) and walk-in patient processes. Related to these processes, participants reported managing notifications, documentation, and chart review as the most frequent, repetitive, and difficult tasks. These discussions highlighted both opportunities and potential barriers to the implementation of potential adaptations to HIT and policies for primary care nursing.

In the following sections, we report on the processes, their influential tasks, and the work system configurations, showing the most relevant components in each process.

Table 2. Demographics of participants at 2 primary care sites, one in an urban medical center (Site 1) and the other in a small community-based outpatient clinic (Site 2).

Characteristic	Both sites (N=22)	Site 1 (n=16)	Site 2 (n=6)
Role, n (%)			
Registered nurse	14 (64)	10 (63)	4 (67)
Licensed practical nurse	6 (27)	4 (25)	2 (33)
Health technician	2 (9)	2 (12)	0 (0)
Race, n (%)			
White	17 (77)	11 (69)	6 (100)
Black	2 (9)	2 (19)	0 (0)
Asian or Pacific Islander	3 (14)	3 (12)	0 (0)
Gender, n (%)			
Female	19 (86)	14 (88)	5 (83)
Male	3 (14)	2 (12)	1 (17)
Years in the role, mean (SD)	11.9 (8.9)	10.6 (8.6)	15.3 (8.8)
Years with current health care employer, mean (SD)	8.3 (8.7)	9.1 (9.8)	6.2 (3.3)

Figure 1. Application of SEIPS (Systems Engineering Initiative for Patient Safety) 2.0 to primary care nursing work systems, processes, and outcomes. Interviews with nurses and health technicians focused on inefficient processes and underlying work systems.



Inefficient Nursing Processes

Managing patient calls and walk-in visits was perceived by participants as inefficient. For patient calls, contributors to inefficiency included repeated attempts to return patient calls and the EHR's inability to track follow-ups. For walk-in processes, contributors to inefficiency included paper-based check-in and incomplete, self-reported patient information. We describe these in more detail below.

Managing Patient Calls

Some participants reported that they received 20 to 50 telephone calls daily from their site's patient call center, which assists patients with various issues, such as scheduling multiple appointments and answering questions related to their care, medication, or paperwork. Patients also contact the call center to return missed calls. Other inbound calls came from patients sharing their frustrations, which participants reported as potentially time-consuming. Participants also received alerts from their EHR system that patients had called the call center. Along with using the EHR system to process alerts, participants also referred to the EHR when calling patients.

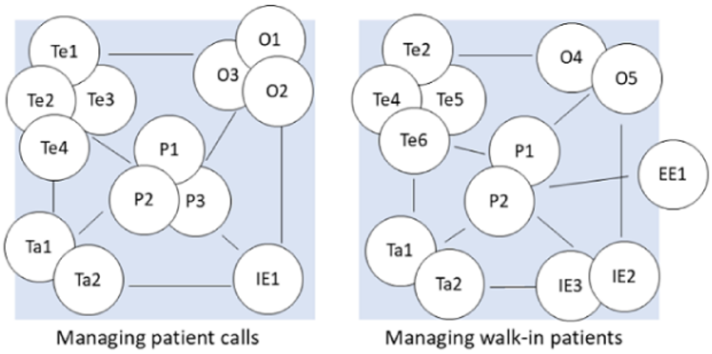
Managing Walk-Ins

Walk-in patients are patients who arrive at the clinic without an appointment. Participants noted that on a typical day, many of their patients were walk-ins. Nurses are required to see these patients regardless of how many patients they have already scheduled, and walk-in visits can often be time-consuming depending on the patient's needs. Managing walk-ins involved triaging and scheduling patients. Triaging patients includes the gathering of important clinical data (eg, vital signs) and preparing the patient to see their primary care provider. Triaging time can vary, depending on the patient and the number of clinical reminders that need to be completed. Walk-in patients often enter the clinic for reasons that do not require an on-site physician evaluation, such as needing a medication refill or having minor aches and pains.

Configuration: Misaligned Work System Factors

Using SEIPS 2.0, we explored the work system configuration of the 2 aforementioned nurse processes, managing patient-response calls and walk-in patients. Participants reported these processes comprised the most frequent, repetitive, and difficult tasks. For each process, we explored the corresponding work system configurations and relevant factors. In Figure 2, we depict the configuration diagrams for both processes.

Figure 2. Application of SEIPS (Systems Engineering Initiative for Patient Safety) 2.0 to 2 primary care nursing work processes: managing patient calls and managing walk-in patients. Factor numbers are for identification and do not imply any order. Ta: Task; Te: Tools and technology; P: People; O: Organization; IE: Internal environment; EE: External environment.



Factors	Managing patient calls	Managing walk-in patients
Person	[P1] Nurses' interpersonal skills varied [P2] Patients needed to feel understood [P3] Some work was sent inappropriately or unnecessarily	[P1] Nurses' interpersonal skills varied [P2] Patients needed to feel understood
External environment		[EE1] Records from other institutions needed more time to retrieve and review
Internal environment	[IE1] Workrooms were shared	[IE2] Exam room availability varied [IE3] Patient's physical presence decreased attention to chart review
Task	[Ta1] Complex patient cases needed more time [Ta2] Day-to-day load varied	[Ta1] Complex patient cases needed more time [Ta2] Day-to-day load varied
Technology	[Te1] EHR notifications were difficult to use [Te2] EHR notes were difficult to access or use [Te3] Electronic references were difficult to access [Te4] EHR data entry was difficult	[Te2] EHR notes were difficult to access or use [Te4] EHR data entry was difficult [Te5] Instant messaging access was low [Te6] Outside records were difficult to access
Organization	[O1] Policy required multiple attempts to schedule with documentation [O2] Calls routed to nurses were not properly routed and increased call volume [O3] Nurse–patient calls increased in duration and content beyond scope	[O4] Walk-in visits began before completing chart reviews [O5] Preparatory work and associated documentation (eg, patient's weight) were required before the patient could be seen by the primary care provider

The following subsections further elaborate on each of the work system factors.

Person

Attitudes toward mobile HIT were shaped by experience with other technologies, including computers on wheels. Some participants viewed the paper as a facilitator of their work-related tasks and favored paper use over mobile technology or applications. Issues such as lack of time, susceptibility to interruption, and inability to take data to patients in nonexamination room spaces encouraged the use of the person's memory skills or notes written on a piece of paper to transfer necessary patient information between various clinical spaces. Other issues related to other staff members, for example, a recurring issue from participants was that the call center does not properly triage patient calls made to the call center.

Environment: Internal and External

Notable environmental differences occurred across the 2 sites, although only a few tasks were affected by these differences. For example, triaging walk-ins varied among clinics. When nurses and health technicians had their own rooms, patients were triaged and seen by the physicians in that room. However, for nurses on 1 team, triaging was completed in 1 room, and then patients could be transferred into the physician's examination room. Thus, for nurses on this team, triaging required finding rooms that were open and contained the right patient education pamphlets.

Participants described the "walk-in pickup," which involves meeting walk-in patients in waiting areas to assess the presentation of symptoms in order to triage the patients quickly and effectively. This method was complicated by initial conversations occurring in nonprivate locations, which limited the level of detail that could be discussed, and by the inability to access medical records electronically during these conversations.

Tasks

Across the 2 processes, participants denoted how complex patient cases needed more time and how the day-to-day load varied. Regarding managing patient calls, participants described frequent tasks, as the call center directs all patient calls to primary care nurses regardless of individual patient needs. Furthermore, participants considered certain tasks that they completed on these calls to be repetitive, including patient education. Some participants indicated that they were largely repeating the information they had provided to patients during previous appointments, which patients had forgotten or were struggling to explain to family members. In addition, many calls come from patients trying to return nurses' missed calls; the call center team is unable to relay any information or schedule appointments for these patients. These instances often resulted in multiple unnecessary back-and-forth phone call exchanges. Participants defined this task as difficult due to the extensive amount of time it took to prioritize the list of calls and related alerts.

Participants also perceived chart review for these patients to be particularly challenging, typically consisting of a review of recent notes, laboratory results, orders, imaging reports, medication lists, or other information to familiarize staff members with a patient's background and recent medical history. Reviews were considered at the following 2 levels. (1) Flash review, which refers to the quick review of medical records in response to the initial question, "I wonder why they're coming?" (2) In-depth review, which refers to the more detailed, investigative review of medical records, typically required for care management. The in-depth review may involve "piecing together" information from a chronological series of notes to discover the narrative patient history and the synthesis of information gathered from different types of notes, tabs, or displays within the EHR. Both types of reviews are complicated when staff members are not familiar with patients (eg, new patients and those from other staff members' panels) and when interruptions occur frequently.

Tools and Technology

Across sites, nurses described problematic aspects of the existing mobile technology. Some participants at site 2 characterized the ruggedized portable PCs as being bulky and useless. Although portable PCs eliminated the need to log in to multiple desktop PCs, mobile devices lost access to the EHR when connecting to or disconnecting from docking stations. Each transition caused the PC to swap between wired and wireless network connections. The connection swap, in turn, disconnected the user from the EHR, resulting in the loss of any unsaved data entered by the user. Instead of providing continuous information interaction and access to the EHR as expected, the portable PCs were usable mainly when docked. This limitation disappointed nurses; mobile devices were used more as luggable desktop computers than for facilitating efficient data collection and communication as expected. Furthermore, at site 1, where mobile technology was distributed and could be used voluntarily, no participants reported it useful for clinical tasks.

Organization

Furthermore, 2 organizational policies influenced the use and usefulness of mobile HIT across the sites. The first was the patient flow policy to limit unnecessary room changes during a health care visit. At site 1, patients move between stations when nurses need to measure vital signs and collect specimens. These stations were often near the nurses' desks, which decreased the need for mobile HIT. At site 2, a policy aimed at reducing patient flow was implemented that required various health care providers, including nurses, to meet the patient in the examination room. This increased the need for nurses to review and document information (eg, vital signs) in examination rooms and hallways away from their desks. In addition, an organizational policy was in place at site 1 that made the use of mobile HIT voluntary. At site 2, the use of mobile HIT was mandated; it was this mandate that accounted for the sustained use of mobile HIT.

We identified work system configurations that contributed to participants' perceptions of inefficient processes. Table 3 displays recurring topics and sample excerpts of configuration details anchored by processes' frequent, repetitive, and difficult tasks. The process of managing patient calls was associated with several tasks that were identified as frequent, repetitive, and difficult. For example, handling patient calls that were inappropriately assigned was identified as both frequent and repetitive. The relevant factors for this process included person factors (ie, patients' need to feel understood, which may be difficult when the patient is assigned to the wrong person), organization factors (ie, this may result in calls increasing in duration and content beyond the scope of the recipient's work), and task factors (ie, the day-to-day load for handling inappropriately assigned calls can vary).

Table 3. Primary care nurses' and health technicians' work system configuration details anchored by processes' frequent, repetitive, and difficult tasks.

Process and task	Representative quote[a]	SEIPS[b] 2.0 work system configuration
Managing patient calls		
Handling patient calls that were inappropriately assigned (Frequent)	"They're calling about an issue that is out of my control, they're very upset, they didn't want their methadone to go away, their methadone is gone, and they want to talk about 30 minutes. That's a real time-waster for me because, you know, that's, I realize they want an outlet, and they're frustrated, but for me, I can't help them, and they're just getting more worked up. Things like, a lot of calls like that where there's really just nothing to do and all they really want to do is talk to someone, but I'm not getting anything done."	• [P[c]2] Patients needed to feel understood • [O[d]3] Nurse–patient calls increased in duration and content beyond scope • [Ta[e]2] Day-to-day load varied
Handling patient calls that were inappropriately assigned (Repetitive)	"The patient calls and the call center relays the message to me if it needs to come to Primary Care. Sometimes it doesn't need to come to Primary Care. Sometimes what usually happens though is I will call the patient back to speak with the patient. I can't get a hold of the patient so then they're calling the call center again. Well, at that point I have a patient in my office. I can't take the phone call. We're doing that all day long. We're playing phone tag constantly. Just very redundant."	• [P1] Nurses' interpersonal skills varied • [O3] Nurse–patient calls increased in duration and content beyond scope • [Ta2] Day-to-day load varied
Managing notifications (Repetitive)	"Oh, there are a lot of view alerts where they send them to multiple people, and a lot of times, they could just come to me. As soon as I see it, I'll do it. But there again, now 2 or 3 other people have to look at it and see if it's been done. That's kind of redundant."	• [P3] Some work was sent inappropriately or unnecessarily • [Te[f]1] EHR[g] notifications were difficult to use • [Ta2] Day-to-day load varied
Scheduling appointments with patients when they are not in the office (Repetitive)	"They want us to continue trying to contact these patients and to me, it just seems redundant. If the patient, wanted an appointment, they would've called and scheduled you know, so it seems like a waste of time when I've got 40 people that I'm trying to call and I have to put a note here, I have to chart it over here, I have to delete the recall."	• [O1] Policy required multiple attempts to schedule with documentation • [Te4] EHR data entry was difficult • [Ta2] Day-to-day load varied
Managing walk-ins		
Checking in walk-ins (Frequent)	"There's somewhere anywhere from 8 to 12 patients scheduled a day and we've got to get them ready to see the provider which can take some time. It can take anywhere from 10 minutes, or it might take 30 minutes to check in some of these patients so that takes up the majority of the time."	• [Te4] EHR data entry was difficult • [Ta1] Complex patient cases needed more time • [Ta2] Day-to-day load varied
Quick patient assessment and triage (Frequent)	"Because I may only get 2 or I may get 7 but they take up a great deal of time and they walk in and you need to stop whatever you're doing and go out to them right then."	• [O4] Walk-in visits began before completing chart reviews • [Ta2] Day-to-day load varied
Chart review of external medical records (Repetitive)	"It's just time consuming to kind of sit and look at all that, especially if you have a patient that maybe came from the outside VA...they might say you know "well I had all this done at the other VA and so kind of trying to pull all the VistA Web [health information exchange service] stuff and look at that is kind of, definitely that's time consuming."	• [O5] Preparatory work and associated documentation (eg, patient's weight) were required before the patient could be seen by the primary care provider • [Te6] Outside records were difficult to access • [EE[h]1] Records from other institutions needed more time to retrieve and review • [Ta2] Day-to-day load varied
Chart review for new patients with little time before patient visit (Difficult)	"Sometimes the walk-ins can be difficult because it might not be your patient if you're covering for someone. You don't know the patient at all and you're trying to piece it all together, because you can't just go to the walk-in doctor and say they're here, they say their head hurts. That's not going to fly, you know, so it's kind of a lot of maybe 15-20 minutes in their chart, and if the patient is in there while you're doing it, that's kind of distracting, because they don't understand that you don't know them and you need to review their chart, so they just take off with their current situation assuming you know their background often, so yeah...."	• [P2] Patients needed to feel understood • [O4] Walk-in visits began before completing chart reviews • [Te2] EHR notes were difficult to access or use • [IE[i]3] Patient's physical presence decreased attention to chart review • [Ta1] Complex patient cases needed more time
Managing patient calls; managing walk-ins		

Process and task	Representative quote[a]	SEIPS[b] 2.0 work system configuration
Chart review before the patient visit (Difficult)	"At least if it's a phone call, I can see the message and why they're calling. I can review the chart. If I need help before calling them back, I can go ahead and get that. When you're sitting in front of somebody and they say well, I'm coughing and I have a headache. I'm having to do all of this in real time."	• [P1] Nurses' interpersonal skills varied • [O4] Walk-in visits began before completing chart reviews • [IE3] Patient's physical presence decreased attention to chart review • [Te2] EHR notes were difficult to access or use • [Ta1] Complex patient cases needed more time
Patient education and chart review (Difficult)	"It takes a lot of time because it's a lot of educating the patients and you know looking back, what was done before, going through labs, meds. You've got to go through side effects of meds and assess everything completely. So, those can be a little bit time consuming… "	• [P1] Nurses' interpersonal skills varied • [Ta1] Complex patient cases needed more time • [Te2] EHR notes were difficult to access or use

[a]Relevant work system elements are listed for the representative quotation. The corresponding task may include factors not listed here.

[b]SEIPS: Systems Engineering Initiative for Patient Safety.

[c]P: People.

[d]O: Organization.

[e]Ta: Task.

[f]Te: Tools and technology.

[g]EHR: electronic health record.

[h]EE: External environment.

[i]IE: Internal environment.

Adaptations: Perceived Adaptability and Recommendations

Work system components "Tools and technology" and "Organization" were associated with the most misaligned factor configurations (Table 3). In addition, these factors were perceived to be the most adaptable among participants. Some reported adaptations were workarounds, while other adaptations were recommendations for unmet needs. Participants did not describe or discuss recommendations related to the remaining work-system components (People, Environments, and Tasks).

Workarounds

Primary systemwide adaptations created process inefficiencies, leading to secondary localized adaptations in the form of user workarounds. At site 2, participants reported that the policies for patient-centered flow and dockable PCs, taken together, limited their EHR review and charting to the times that their PCs were docked. Their interim storage needs were met using paper notes, which also addressed the risk of data loss from EHR disconnections. Paper notes were shared with physicians, who could then review and add to the notes before the notes were entered into the EHR after the visit.

Site 1's walk-in policy, combined with the physical layout of the clinic, was also linked to the use of paper notes. One site 1 clinic created a paper intake form for patients to self-report the reason for their visit and the reason for walking in instead of alternatives (eg, making an appointment or refilling medications through the patient portal). For health technicians, paper notes indicated double documentation: for example, the patient's weight would be written while in the corridor by the weight scale, then reentered into the EHR afterward. Using a different workaround, the "walk-in pickup" described in the previous section, participants balanced their need for intake information with their need to manage each patient's expectations about when they would be seen. This approach necessitated a chart review in which staff looked for information to aid triaging, including the time since the patient's last visit and the frequency and nature of the patient's previous walk-in visits.

At site 1, to work around inefficiencies with patient calls, 1 participant reported using their appointment scheduler software to track outbound follow-up calls about lab results. However, these appointments appeared in the patient-facing portal, which sometimes confused patients who were not expecting such calls.

Recommendations

Recommendations included changes to "tools and technology." Among the participants' suggestions to improve walk-in management were new tools for symptom self-reporting and triage. To improve the check-in process for walk-in patients, 1 recommendation from a participant was a patient-facing technology for collecting patients' descriptions of their health issues (eg, symptoms of congestive heart failure). Another participant's suggestion was a personalized display of the expected waiting time, encouraging patients with less serious ailments to consider scheduling an appointment or requesting information about alternative ways to address their medical concerns. For themselves, participants sought a method to view relevant patient trends, minimizing the need for rushed chart review. Many participants wanted call center staff to offer greater mediation between themselves and patients. Currently, back-and-forth communication is needed to understand a patient's concerns or issues. More and better training resources could be made available.

Discussion

Principal Findings

In this study, we described primary care nurses' work system configurations associated with inefficient processes, misaligned work system factors, and adaptations to guide future interventions. Managing patient calls and managing walk-in patients were inefficient processes. The results from our work system analysis defined nursing tasks associated with each process that was described as frequent, repetitive, or difficult among primary care nurses and health technicians. In addition, we applied SEIPS 2.0's configuration concept [13] to illustrate subsets of work system factors that were associated with workflow inefficiencies. With the model's adaptation concept, we characterized the propagating, negative impacts of changes to a work system component. Some adaptations were made by health care workers in the form of workarounds, with varying success, while other adaptations were recommended.

SEIPS 2.0 Application to Primary Care Nursing Processes

Although SEIPS 2.0 has been used to identify tasks associated with decreased work ability among inpatient nurses [22,23], to our knowledge, this is among the first applications of SEIPS 2.0 specifically to primary care nursing processes. The original SEIPS framework [24] has been identified as a means of describing and evaluating processes in primary care by taking into account the complex, interconnected socio-technical aspects found in the health care system [25]. Lagisetty et al [26] organized their systematic review of primary care opioid use disorder interventions using SEIPS 2.0's concepts of work system factors, processes, and outcomes [13]. Robertson et al [11] used SEIPS 2.0 [13] to identify barriers and facilitators to integrating practice guidelines to reduce under-5 mortality in a primary care clinic in Malawi; their recommendations addressed mostly organizational factors. More recently, Werner et al [20] used SEIPS 2.0 configuration diagrams [13] to illustrate work barriers and facilitators within work system configurations for older adults' transitions between emergency department to home. McCormack et al [27] used SEIPS 2.0 to identify facilitators and barriers to referrals between primary and specialty care services.

Misaligned Work System Factors

The misalignment of tasks, organization, and technology factors was described among multiple configured sets of work system factors. Patient flow policies introduced more nomadic or mobile aspects and interruptions to primary care nursing processes. Other studies have defined telephone calls and conversations as the most common sources of interruptions for nurses and health technicians [28]. In this study, we refer to these as unscheduled tasks. In workflows, increased movement of primary care nurses increased the amount of missed patient phone calls, alerts about missed patient phone calls, and associated patient voicemails. Furthermore, increased presence in hallways and rooms appeared to yield more impromptu conversations. The implemented mobile HIT to support this patient flow was limited by poor connectivity to the network making data transfer difficult, and possible data loss when undocking mobile devices. Experiencing high levels of process discontinuity predisposes health care staff to make errors [29,30]. Processes with unscheduled tasks introduced work fragmentation or a break in continuous work activity. Despite demonstrated resilience against interruptions [31], unscheduled tasks increased nurses' and health technicians' cognitive load and decreased their ability to recall information needed to complete task switching effectively [32]. Minimizing unnecessary interruptions is particularly important in the health care context, where failing to complete tasks can have adverse effects on health outcomes and patient safety.

Adaptations: Feedback and Indicators of Gaps

Based on SEIPS 2.0, the inefficiencies described in our study can be indicators of gaps in performance, quality, or patient safety [13]. Whether initiated or recommended by nurses, the adaptations identified in this study were mostly reactive to new policies recently implemented. These adaptations indicated gaps in HIT performance and quality of nursing processes, which may be linked to patient, nurse, or organizational outcomes [33]. Previous studies have denoted that traditional or routine clinical quality indicators do not always include measures for HIT-related aspects of workflow, such as usability [27,34]. Poor usability of HIT is associated with various types of adaptations, often referred to as workarounds [35,36]. For example, the new policies and HIT (ie, planned adaptations) were implemented to improve patient flow and access to primary care services. Yet, nurses described the generation of new gaps or inefficiencies that propagated throughout the work system, which we characterized as misalignments. Based on their experiences, nurses were able to identify inefficiencies, implement workarounds (adaptations to the patient flow adaptation), and suggest recommendations for future adaptations. While other interpretations of workarounds vary [33,37], SEIPS 2.0 represents these types of adaptations as feedback. Without this type of feedback, monitoring such dynamic complex work systems, important indicators and gaps would not be recognized, and efforts for continuous improvement would be hindered.

Future Design and Implementation Strategies

Studies like this are foundational to future interventional research. Future design and implementation of interventions for these and similar nursing processes are warranted as health care use increases and diversifies with same-day and virtual visits, which have similar aspects and common tasks with the inefficient processes identified in this study. Due to the interconnectivity of work systems, adaptations focused on a single component or subset of components will affect other components. Whether the adaptation is planned or unplanned, the lack of consideration for all the components in the work system increases the potential for negative, unintended consequences. Without work system analyses, the implementation of interventions can negatively affect clinicians, patients, and organizational outcomes. Similar problems in other clinician groups have been addressed by incorporating system-level design and implementation of HIT [38]. Yet, varying institutions have unique needs or processes that demand different solutions. For example, the distribution of mobile devices without apps that are tailored to our participants was insufficient to meet their needs. Therefore, understanding the personalized challenges of an institution warrants a systems-level analysis. Our findings provide a necessary first step in the development and integration of future health information technologies that improve the efficiency of health care delivery by supporting frequent, difficult, and repetitive tasks for nurses. In addition, this research shows that for interventions, including health information technology implementation, to be successful, implementers must also account for work system factors such as organizational policies. Based on our findings, the major opportunities for adaptations are related to workflow policies and supporting health information technologies for primary care nurses.

Since the completion of our study, the VA's Office of Connected Care is supporting an increasing number of provider- and patient-facing mobile apps, including task-specific apps [39]. Providers currently have access to a variety of task-specific apps for mobile computing through the VA App Store [39].

Strengths and Limitations

A strength of this study is that it uses a human factors approach to identify major contributors to inefficiencies at the systems level. This study also has several limitations. A limitation of this study is that the observations and interviews were performed in clinics belonging to 1 integrated health care institution. Convenience sampling may have introduced biases in participants' reporting of work processes and barriers. Our findings may not entirely transfer to other health care systems and settings. Therefore, more attention should be given to aiding in the design and development of user-centered apps in different settings with different work system configurations. Similarly, while we identified notable differences between the 2 sites used for this study, we did not further assess how those differences contributed to our results. Finally, this study focused only on the potential of mobile HIT as a solution to process inefficiencies. As a result, this may have elicited confirmation bias.

Conclusion

Nurses and health technicians perceived that the implementation of new policies and technologies contributed to inefficiencies in nursing workflows across ambulatory settings. A system analysis was an effective method for identifying configured subsets of work system factors associated with perceived gaps in nursing processes. Furthermore, the configuration and adaptation concepts in the SEIPS 2.0 framework aided in the characterization of adaptations to inform future research and interventions. To identify both potential consequences across work system components and nursing processes, system analyses are warranted for the design, implementation, and evaluation of organizational policies or HIT.

Acknowledgments

This work was funded by the Indiana Institute for Medical Research. This publication was made possible with support from the National Institutes of Health, National Center for Advancing Translational Sciences, Clinical and Translational Sciences Award (KL2TR002530 and UL1TR002529). WLT was supported by the VA Advanced Fellowship Program in Medical Informatics of the Office of Academic Affiliations, US Department of Veterans Affairs. AS and HP were supported by the US Department of Veterans Affairs, Veterans Health Administration, Health Systems Research CIN 13-416. MW was Chief of Health Services Research and Development at the Richard L. Roudebush Veterans Affairs Medical Center in Indianapolis, Indiana. The views expressed in this article are those of the authors and do not necessarily represent the views of the US Department of Veterans Affairs or the United States Government. Generative artificial intelligence software was not used in any portion of the manuscript writing.

Data Availability

The data sets generated during and/or analyzed during this study are not publicly available due to the presence of personally identifiable information but are available from the corresponding author on reasonable request.

Authors' Contributions

AS proposed the study and secured funding. AS and HP collected and analyzed data. WLT, AS, and HP drafted the manuscript. HP created the figures. All authors interpreted the findings, made critical revisions, and approved the published manuscript; all authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflicts of Interest

MW reports the following stock holdings: Allscripts Healthcare Solutions, Inc; Centene Corp; DXC Technology Co; General Electric Co; International Business Machines Corp; Kyndryl Holdings, Inc; Micro Focus International PLC; Microsoft Corp; Oracle Corp; PerkinElmer, Inc; Qualcomm, Inc; Walgreens Boots Alliance, Inc; Zimmer Biomet Holdings, Inc; Senseonics Holdings, Inc; Teladoc Health, Inc; and Varex Imaging Corp.

Multimedia Appendix 1

Interview guide.
[DOCX File , 15 KB - humanfactors_v11i1e49691_app1.docx]

References

1. Association of American Medical Colleges. The complexities of physician supply and demand: projections from 2018 to 2033. 2020. URL: https://web.archive.org/web/20240302004510/https://www.aamc.org/system/files/2020-06/stratcomm-aamc-physician-workforce-projections-june-2020.pdf [accessed 2024-01-10]
2. Norful A, Martsolf G, de Jacq K, Poghosyan L. Utilization of registered nurses in primary care teams: a systematic review. Int J Nurs Stud 2017;74:15-23 [FREE Full text] [doi: 10.1016/j.ijnurstu.2017.05.013] [Medline: 28595110]
3. Flinter M, Hsu C, Cromp D, Ladden MD, Wagner EH. Registered nurses in primary care: emerging new roles and contributions to team-based care in high-performing practices. J Ambul Care Manage 2017;40(4):287-296 [FREE Full text] [doi: 10.1097/JAC.0000000000000193] [Medline: 28323721]
4. Rutledge CM, Gustin T. Preparing nurses for roles in telehealth: now is the time!. Online J Issues Nurs 2021;26(1):1-13 [FREE Full text] [doi: 10.3912/ojin.vol26no01man03]
5. Nezamdoust S, Abdekhoda M, Ranjbaran F, Azami-Aghdash S. Adopting mobile health applications by nurses: a scoping review. J Res Nurs 2022;27(5):480-491 [FREE Full text] [doi: 10.1177/17449871221077080] [Medline: 36131693]
6. Argentin Y, Gluck F, Lovis C, Ehrler F. Individual nomad clinical assistant: supporting nurses at the point of care. Stud Health Technol Inform 2016;225:654-655 [FREE Full text] [Medline: 27332295]
7. Silva AMDA, Mascarenhas VHA, Araújo SNM, Machado RDS, Santos AMRD, Andrade EMLR. Mobile technologies in the nursing area. Rev Bras Enferm 2018;71(5):2570-2578 [FREE Full text] [doi: 10.1590/0034-7167-2017-0513] [Medline: 30304191]
8. Zuzelo PR, Gettis C, Hansell AW, Thomas L. Describing the influence of technologies on registered nurses' work. Clin Nurse Spec 2008;22(3):132-140. [doi: 10.1097/01.NUR.0000311693.92662.14] [Medline: 18438162]
9. Huter K, Krick T, Domhoff D, Seibert K, Wolf-Ostermann K, Rothgang H. Effectiveness of digital technologies to support nursing care: results of a scoping review. J Multidiscip Healthc 2020;13:1905-1926 [FREE Full text] [doi: 10.2147/JMDH.S286193] [Medline: 33328736]
10. Wang W, Zhang J, Nicholas S, Yang H, Maitland E. Organisation-level and individual-level predictors of nurse-reported quality of care in primary care: a multilevel study in China. Trop Med Int Health 2023;28(4):308-314 [FREE Full text] [doi: 10.1111/tmi.13861] [Medline: 36756803]
11. Robertson SK, Manson K, Fioratou E. IMCI and ETAT integration at a primary healthcare facility in Malawi: a human factors approach. BMC Health Serv Res 2018;18(1):1014 [FREE Full text] [doi: 10.1186/s12913-018-3803-5] [Medline: 30594185]
12. Lin TC. Mobile nursing information system utilization: the task-technology fit perspective. Comput Inform Nurs 2014;32(3):129-137. [doi: 10.1097/CIN.0000000000000039] [Medline: 24419090]
13. Holden RJ, Carayon P, Gurses AP, Hoonakker P, Hundt AS, Ozok AA, et al. SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. Ergonomics 2013;56(11):1669-1686 [FREE Full text] [doi: 10.1080/00140139.2013.838643] [Medline: 24088063]
14. World Health Organization. Digital technologies: shaping the future of primary health care. 2018. URL: https://www.who.int/docs/default-source/primary-health-care-conference/digital-technologies.pdf [accessed 2024-09-12]
15. Tahsin F, Armas A, Kirakalaprathapan A, Kadu M, Sritharan J, Steele Gray C. Information and communications technologies enabling integrated primary care for patients with complex care needs: scoping review. J Med Internet Res 2023;25:e44035 [FREE Full text] [doi: 10.2196/44035] [Medline: 37074779]
16. Saleem JJ, Savoy A, Etherton G, Herout J. Investigating the need for clinicians to use tablet computers with a newly envisioned electronic health record. Int J Med Inform 2018;110:25-30. [doi: 10.1016/j.ijmedinf.2017.11.013] [Medline: 29331252]
17. Hollnagel E, Woods DD. Joint Cognitive Systems: Foundations of Cognitive Systems Engineering. Boca Raton, US: CRC Press; 2005.
18. Karen H, Sandra J. Contextual inquiry: a participatory technique for system design. In: Participatory Design: Principles and Practices. Boca Raton, US: CRC Press; 2017:177-210.
19. Beyer HR, Holtzblatt K. Apprenticing with the customer. Commun ACM 1995;38(5):45-52. [doi: 10.1145/203356.203365]
20. Werner NE, Rutkowski R, Graske A, Finta MK, Sellers CR, Seshadri S, et al. Exploring SEIPS 2.0 as a model for analyzing care transitions across work systems. Appl Ergon 2020;88:103141 [FREE Full text] [doi: 10.1016/j.apergo.2020.103141] [Medline: 32421635]
21. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
22. Danesh MK, Garosi E, Mazloumi A, Najafi S. Identifying factors influencing cardiac care nurses' work ability within the framework of the SEIPS model. Work 2020;66(3):569-577. [doi: 10.3233/WOR-203199] [Medline: 32623418]

23. Muthukrishnan R, Maqbool Ahmad J. Ergonomic risk factors and risk exposure level of nursing tasks: association with work-related musculoskeletal disorders in nurses. European Journal of Physiotherapy 2020;23(4):248-253. [doi: 10.1080/21679169.2020.1715473]
24. Carayon P, Schoofs Hundt A, Karsh B, Gurses AP, Alvarado CJ, Smith M, et al. Work system design for patient safety: the SEIPS model. Qual Saf Health Care 2006;15 Suppl 1(Suppl 1):i50-i58 [FREE Full text] [doi: 10.1136/qshc.2005.015842] [Medline: 17142610]
25. Wooldridge AR, Carayon P, Hundt AS, Hoonakker PLT. SEIPS-based process modeling in primary care. Appl Ergon 2017;60:240-254 [FREE Full text] [doi: 10.1016/j.apergo.2016.11.010] [Medline: 28166883]
26. Lagisetty P, Klasa K, Bush C, Heisler M, Chopra V, Bohnert A. Primary care models for treating opioid use disorders: what actually works? A systematic review. PLoS One 2017;12(10):e0186315 [FREE Full text] [doi: 10.1371/journal.pone.0186315] [Medline: 29040331]
27. McCormack C, Savoy A, Mathew A, Gilmore AM, Cottingham E, Sangani A, et al. Characterizing barriers to closing cross-institutional referral loops: workflow and information flow analysis. Appl Ergon 2024;115:104177. [doi: 10.1016/j.apergo.2023.104177] [Medline: 38016271]
28. Brixey JJ, Robinson DJ, Tang Z, Johnson TR, Zhang J, Turley JP. Interruptions in workflow for RNs in a level one trauma center. AMIA Annu Symp Proc 2005;2005:86-90 [FREE Full text] [Medline: 16779007]
29. Kalisch BJ, Aebersold M. Interruptions and multitasking in nursing care. Jt Comm J Qual Patient Saf 2010;36(3):126-132. [doi: 10.1016/s1553-7250(10)36021-1] [Medline: 20235414]
30. Laxmisan A, Hakimzada F, Sayan OR, Green RA, Zhang J, Patel VL. The multitasking clinician: decision-making and cognitive demand during and after team handoffs in emergency care. Int J Med Inform 2007;76(11-12):801-811. [doi: 10.1016/j.ijmedinf.2006.09.019] [Medline: 17059892]
31. Zide JS, Mills MJ, Shahani-Denning C, Sweetapple C. Work interruptions resiliency: toward an improved understanding of employee efficiency. JOEPP 2017;4(1):39-58. [doi: 10.1108/joepp-04-2016-0031]
32. Woloshynowych M, Davis R, Brown R, Vincent C. Communication patterns in a UK emergency department. Ann Emerg Med 2007;50(4):407-413. [doi: 10.1016/j.annemergmed.2007.08.001] [Medline: 17881316]
33. Blijleven V, Koelemeijer K, Wetzels M, Jaspers M. Workarounds emerging from electronic health record system usage: consequences for patient safety, effectiveness of care, and efficiency of care. JMIR Hum Factors 2017;4(4):e27 [FREE Full text] [doi: 10.2196/humanfactors.7978] [Medline: 28982645]
34. Madandola OO, Bjarnadottir RI, Yao Y, Ansell M, Dos Santos F, Cho H, et al. The relationship between electronic health records user interface features and data quality of patient clinical information: an integrative review. J Am Med Inform Assoc 2023;31(1):240-255. [doi: 10.1093/jamia/ocad188] [Medline: 37740937]
35. Friedman A, Crosson JC, Howard J, Clark EC, Pellerano M, Karsh B, et al. A typology of electronic health record workarounds in small-to-medium size primary care practices. J Am Med Inform Assoc 2014;21(e1):e78-e83 [FREE Full text] [doi: 10.1136/amiajnl-2013-001686] [Medline: 23904322]
36. Patterson ES. Workarounds to intended use of health information technology: a narrative review of the human factors engineering literature. Hum Factors 2018;60(3):281-292. [doi: 10.1177/0018720818762546] [Medline: 29533682]
37. Flanagan ME, Saleem JJ, Millitello LG, Russ AL, Doebbeling BN. Paper- and computer-based workarounds to electronic health record use at three benchmark institutions. J Am Med Inform Assoc 2013;20(e1):e59-e66 [FREE Full text] [doi: 10.1136/amiajnl-2012-000982] [Medline: 23492593]
38. Jewer J. Investigating a work system approach to implement an emergency department surge management system: case study. J Med Internet Res 2022;24(8):e37472 [FREE Full text] [doi: 10.2196/37472] [Medline: 36006684]
39. Welcome to VA mobile. U.S. Department of Veterans Affairs. 2024. URL: https://mobile.va.gov/ [accessed 2024-01-11]

Abbreviations

EHR: electronic health record
HIT: health information technology
SEIPS: Systems Engineering Initiative for Patient Safety
VA: US Department of Veterans Affairs

Edited by A Kushniruk; submitted 06.06.23; peer-reviewed by S Oh, L Chen; comments to author 23.06.23; revised version received 27.08.23; accepted 04.08.24; published 30.09.24.

Please cite as:
Tarver WL, Savoy A, Patel H, Weiner M, Holden RJ
Inefficient Processes and Associated Factors in Primary Care Nursing: System Configuration Analysis
JMIR Hum Factors 2024;11:e49691
URL: https://humanfactors.jmir.org/2024/1/e49691
doi:10.2196/49691
PMID:

Original Paper

Patients' Expectations of Doctors' Clinical Competencies in the Digital Health Care Era: Qualitative Semistructured Interview Study Among Patients

Humairah Zainal[1*], PhD; Xin Xiao Hui[1*], MA; Julian Thumboo[1,2,3*], MBBS; Warren Fong[2,3,4*], MBBS; Fong Kok Yong[2,3*], MBBS

[1]Health Services Research Unit, Singapore General Hospital, Singapore, Singapore
[2]Department of Rheumatology and Immunology, Singapore General Hospital, Singapore, Singapore
[3]Duke-NUS Medical School, National University of Singapore, Singapore, Singapore
[4]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[*]all authors contributed equally

Corresponding Author:
Humairah Zainal, PhD
Health Services Research Unit
Singapore General Hospital
10 Hospital Boulevard
Singapore, 168582
Singapore
Phone: 65 6908 8949
Email: humairah.zainal@sgh.com.sg

Abstract

Background: Digital technologies have impacted health care delivery globally, and are increasingly being deployed in clinical practice. However, there is limited research on patients' expectations of doctors' clinical competencies when using digital health care technologies (DHTs) in medical care. Understanding these expectations can reveal competency gaps, enhance patient confidence, and contribute to digital innovation initiatives.

Objective: This study explores patients' perceptions of doctors' use of DHTs in clinical care. Using Singapore as a case study, it examines patients' expectations regarding doctors' communication, diagnosis, and treatment skills when using telemedicine, health apps, wearable devices, electronic health records, and artificial intelligence.

Methods: Findings were drawn from individual semistructured interviews with patients from outpatient clinics. Participants were recruited using purposive sampling. Data were analyzed qualitatively using thematic analysis.

Results: Twenty-five participants from different backgrounds and with various chronic conditions participated in the study. They expected doctors to be adept in handling medical data from apps and wearable devices. For telemedicine, participants expected a level of assessment of their medical conditions akin to in-person consultations. In addition, they valued doctors recognizing when a physical examination was necessary. Interestingly, eye contact was appreciated but deemed nonessential by participants across all age bands when electronic health records were used, as they valued the doctor's efficiency more than eye contact. Nonetheless, participants emphasized the need for empathy throughout the clinical encounter regardless of DHT use. Furthermore, younger participants had a greater expectation for DHT use among doctors compared to older ones, who preferred DHTs as a complement rather than a replacement for clinical skills. The former expected doctors to be knowledgeable about the algorithms, principles, and purposes of DHTs such as artificial intelligence technologies to better assist them in diagnosis and treatment.

Conclusions: By identifying patients' expectations of doctors amid increasing health care digitalization, this study highlights that while basic clinical skills remain crucial in the digital age, the role of clinicians needs to evolve with the introduction of DHTs. It has also provided insights into how DHTs can be integrated effectively into clinical settings, aligning with patients' expectations and preferences. Overall, the findings offer a framework for high-income countries to harness DHTs in enhancing health care delivery in the digital era.

(JMIR Hum Factors 2024;11:e51972) doi:10.2196/51972

KEYWORDS
digital health; clinical competence; patient engagement; qualitative research; Singapore; mobile phone

Introduction

Background

Digital health care, which can be defined as the use of advanced technologies to replace or complement existing health care services and practices, is becoming increasingly prevalent in clinical work [1]. Digital health care technologies (DHTs) such as electronic health records (EHRs), telemedicine, wearable devices, mobile health, and other digital tools have been beneficial to health care delivery [2]. The widespread adoption of EHRs has enabled the streamlining of patient data, making it easily accessible to health care providers and improving care coordination [3]. The growth of telemedicine has facilitated consultations between health care providers and patients living in remote and underserved areas [4,5]. Furthermore, the development of wearable sensors and smartphone apps has allowed for the continuous monitoring of neurodegenerative disorders and the detection of related disease symptoms, respectively, among other benefits [6]. The integration of automation and artificial intelligence (AI) in health care has also enhanced various medical procedures including improving diagnosis and overall work efficiency [1,7]. These technologies have proven useful for patients especially where the barriers to adoption are low [1,7].

Singapore, a high-income country in Southeast Asia, makes a compelling case study for investigating how end users perceive the use of DHTs. This is attributed to a myriad of reasons: its high rate of technological advancement, strong government support for digitalization, efficient health care system, and a diverse population. Singapore is renowned for its advanced technological infrastructure and digital capabilities. Despite being the smallest country in Southeast Asia [8], Singapore is ranked first in the world in internet availability [9]. In addition, there is a high proliferation of digital technology and mobile phones in the country [10]. In 2020, 88% of its population was using smartphones compared to 78% globally [9]. A survey has also shown that the potential for digital uptake is high, and Singaporeans are open to incorporating digital tools in their health routines especially if it is beneficial to them [11]. The Singapore government has also invested heavily into transforming Singapore into a Smart Nation, an initiative that leverages technology and data to improve economic competitiveness and enhance the lives of its citizens [12]. In the realm of health care, it has rolled out digital initiatives to fulfill its population's health care needs [9]. These include introducing telehealth, as well as implementing robotics and assistive technology solutions, to help older adults and those with mobility issues [13]. AI technologies have also been used in specialties such as radiology and ophthalmology and are expected to be a transformational force in Singapore's health care [14]. Its well-established health care system makes it an ideal environment to study the integration of DHTs into health care delivery. Notwithstanding its small size of 734 km^2, Singapore is also home to a diverse population with varying demographics, including different cultural influences, age groups, and socioeconomic backgrounds [8]. These factors make Singapore a worthwhile case study to explore patient perceptions and expectations of their doctors' DHT competencies.

Despite the proliferation of DHTs in many high-income countries such as Singapore, there is a lack of research that captures the perspectives of patients on the clinical competencies that doctors need to be equipped with when DHTs are being deployed. Clinical competencies in this study refer to the knowledge, skills, and attitudes that medical doctors should possess when assessing, diagnosing, treating, and caring for their patients [15,16]. Existing research on patient and public involvement in digital health has either taken a broad approach [17,18] or is focused on specific areas, such as digital health consent in the context of clinical care [19]. It is important to examine patients' expectations of doctors' competencies in DHTs in their clinical care for several reasons. First, it would help identify any existing gaps in competencies. Second, patients' confidence in the health care system would depend in part on doctors' using DHTs competently in the diagnosis and management of their medical conditions. Third, studies and reviews have shown that greater efforts to involve patients in digital planning and implementation need to be made from the outset so that they can contribute meaningfully to digital innovation initiatives [17,18].

Objective

This paper aims to address the aforementioned gap by assessing how patients perceive the use of DHTs by doctors in the context of clinical care. Using Singapore as a case study, it examines patients' expectations of their doctors when telemedicine, health apps, wearable devices, EHRs, and AI technologies are used. Specifically, it evaluates their views on how doctors should communicate, diagnose, and treat their conditions when deploying these DHTs. In view of the lack of research done in non-Western contexts on patient involvement in digital health studies [17,19], Singapore represents an important case study for exploring this topic.

Methods

Setting and Sample

We adopted a social constructivist approach to understanding the subjective experiences and diverse interpretations of individuals in their social context. This approach facilitates the understanding of the meaning of a text as an "interaction between the preconceptions of the reader and the intentions of the producer" [20]. Hence, it raises the possibility of an agreement between various individual meaning constructions [20]. A qualitative study involving individual semistructured interviews with patients from outpatient clinics of a public health care cluster in Singapore was performed.

Data were collected from June 2022 to June 2023 through individual interviews with patients. For maximum variation, purposive sampling was used to recruit patients who were seeking treatment for various medical conditions. This type of sampling has been widely adopted in many health sciences research studies, as it enables researchers to select participants with specific characteristics or conditions relevant to the study [21-23]. Accordingly, 1 to 2 patients with at least 1 of the conditions that constituted chronic disease burden in Singapore were recruited [24]. These were the medical conditions that were commonly reported in Singapore and that would have a significant impact on health [24]. They included cancer, cardiovascular disease, diabetes and kidney diseases, neurological diseases, and autoimmune diseases. Amid the trend of an aging population with comorbidities in high-income countries [25], it is imperative to examine the type of clinical competence that may fulfill the health care needs of patients with multiple conditions. In addition, participants from different age groups, spanning between 20 and 39 years, 40 and 59 years, and ≥60 years, were interviewed. This was done with the aim of exploring whether patients from different generational cohorts had varying expectations regarding their doctors' competencies and if so, what those expectations were. Furthermore, we recruited patients from various ethnic groups to assess if cultural factors had any influence on what patients anticipate of their doctors.

We first conducted pilot interviews with 5 patients from the outpatient clinics of a health care cluster who met the aforementioned criteria to ensure that the interview questions were clear and relevant and that the interview process flowed smoothly. We then carried out more interviews until data saturation was reached. Potential participants were first identified by WF before research fellow HZ sent email invitations for the interview. Interviews were then conducted and recorded over Zoom (Zoom Video Communications) unless participants requested an in-person interview. Acceptance of invitations for a Zoom interview served as consent for participation. In the reporting of findings, we followed the Standards for Reporting Qualitative Research proposed by O'Brien et al [26] (Table 1).

Table 1. Standards for reporting qualitative research.

Number	Topic	Item
Title and abstract		
S1	Title	• "Patients' Expectations of Doctors' Clinical Competencies in the Digital Health Care Era: Qualitative Semistructured Interview Study Among Patients"
S2	Abstract	• Please refer to the Abstract in the main text.
Introduction		
S3	Problem formulation	• The study aims to explore patients' expectations of doctors' clinical competencies when using DHTs[a] in medical care, using Singapore as a case study.
S4	Research questions	• What are some of the patients' expectations of and concerns with doctors' clinical competencies when DHTs such as telemedicine, health apps, wearable devices, EHRs[b], and AI[c] are being deployed? What are their views on how doctors should communicate, diagnose, and treat their conditions when deploying these DHTs in clinical care?
Methods		
S5	Qualitative approach and research paradigm	• This manuscript is based on an ethnographic study involving individual semistructured interviews with patients from the outpatient clinics of a health care cluster in Singapore. • It adopts a social constructivist approach and analyzes the data inductively using the 6-step process given by Braun and Clarke [27].
S6	Researcher characteristics and reflexivity	• Please refer to the Data Collection section for elaboration.
S7	Context	• The interviews were conducted after the COVID-19 cases showed a declining trend in Singapore. At the peak of the COVID-19 pandemic, health care settings witnessed an acceleration in the adoption of digital technologies in health care, such as the use of AI to detect the severity of pneumonia in COVID-19 patients, telemedicine for remote consultation, and robots to deliver medication in hospital wards.
S8	Sampling strategy	• After identifying the illnesses that constituted chronic disease burden in Singapore, we conducted pilot interviews with 5 patients from the outpatient clinic of a health care cluster before determining the required sample size. The latter was done by identifying 1 to 2 patients who were having at least 1 of the illnesses from each of the following age bands: 20-39 years, 40-59 years, and ≥60 years.
S9	Ethical issues pertaining to human participants	• Waiver for ethical approval was granted by SingHealth Centralised Institutional Review Board (reference 2020/2880).
S10	Data collection methods	• Please refer to the Methods section in the main text for details.
S11	Data collection instruments and technologies	• 22 interviews were conducted and recorded over Zoom, while 3 were done in person (upon participants' request). The latter interviews were audio recorded. Each interview lasted approximately 40 minutes.
S12	Units of study	• 25 patients who were experiencing any one of the illnesses that constitute chronic disease burden in Singapore took part in the once-off interview.
S13	Data processing	• The interviews were transcribed verbatim by a transcriber, and the transcripts reviewed by HZ and FKY to ensure transcription accuracy. To protect participants' anonymity, we assigned code identifiers beginning with "P" to each of them, an abbreviation for "patients."
S14	Data analysis	• The researchers adopted an inductive thematic analysis approach when evaluating the data to draw common and shared meanings among participants. Coding frameworks and themes were developed iteratively using the 6-step process proposed by Braun and Clarke [27].

Number	Topic	Item
S15	Techniques to enhance trustworthiness	• Collecting data from patients with different medical conditions • Comparing the findings with those of studies conducted in other high-income countries that are facing similar pace of digital transformations in health care • Comparing the findings with up-to-date published data on patients' expectations of doctors and on the views of stakeholders from the health care industry toward the digital competencies needed for current and future clinical practice.
Results and findings		
S16	Synthesis and interpretation	• Refer to the Principal Findings section.
S17	Links to empirical data	• Refer to illustrative quotes in the Results section. Full data are available upon reasonable request to authors.
Discussion		
S18	Integration with prior work, implications, transferability, and contributions to the field	• Refer to the Discussion section.
S19	Limitations	• Refer to the text under Strengths and Limitations.
Others		
S20	Conflicts of interest	• None was reported by the authors.
S21	Funding	• SingHealth Duke-NUS Medicine Academic Clinical Programme under Seah Cheng Siang Distinguished Professorship in Medicine.

[a]DHT: digital health care technology.

[b]EHR: electronic health record.

[c]AI: artificial intelligence.

Data Collection

The interview guide was developed based on the framework given by Kallio et al [28] for developing a qualitative semistructured interview guide. This involved a 5-step process, namely, identifying the prerequisites to use a semistructured interview, retrieving and using previous knowledge in the literature and empirical data based on consultations with experts, formulating the preliminary interview guide with the research team, pilot-testing the interview with 5 patients, and finalizing the full interview guide for data collection [28]. Data from patients with different medical conditions and social backgrounds were then collected to ensure a broad representation of findings. At the beginning of the discussion, the interviewer reiterated the participants' right to withdraw from the study at any time during the interview. In addition, the interviewer informed the participant that the data collected up to the point of withdrawal would be retained and analyzed to enable a comprehensive evaluation of the findings.

The interview questions sought patients' views on their experiences with DHTs, if any; their expectations of their doctors when it comes to treating and diagnosing their illnesses, particularly when DHTs were used; and their suggestions on how different stakeholders in the health care system could improve the way patients were treated and diagnosed with the aid of DHTs (Table 2). The DHTs that we selected were those that were reported to be popular among Singaporeans and that were expected to transform the health care delivery of high-income countries such as Singapore [11,29]. These included health care–related mobile apps. In Singapore, the app that is commonly used by Singaporeans is the HealthHub app. Launched in 2015, it is a one-stop online health information and services portal and mobile app that allows Singaporeans to access their health records, laboratory results, and other health-related information; manage future medical appointments; and pay medical bills, among other functions [9]. The other DHTs were wearable devices, EHRs, AI, and telemedicine, which we defined as the use of information and communications technologies to deliver health care services from a distance [30].

In reviewing the issues of reflexivity, we considered how our assumptions and prejudices might influence our research [31]. In so doing, we discovered that there was a tendency to assume that participants of higher educational qualifications would be more familiar with the terms related to DHTs. Hence, the assumption is that there would be less need for the interviewer, HZ, to explain the terms in detail. However, the pilot interviews proved otherwise. To overcome this bias, HZ defined the key terms found in the interview guide for all participants for subsequent interviews.

Table 2. Interview questions.

Number	Questions
1	• We are living in an age where digital technologies are prevalent in our everyday lives including in our health care experience. New forms of technology such as Artificial Intelligence or AI and robotics have influenced health care treatment and diagnosis. • Do you expect your doctor's roles to change with the advent of digital technologies? If yes, in what ways? • What are your expectations of a doctor when it comes to (1) treating your illness or conditions, (2) diagnosis, and (3) overall patient care?
2	• Do you have any experience of using digital platforms when consulting your doctor? These include video consultations over Zoom, WhatsApp or any other teleconferencing platforms. • If yes, please share with me your experience, particularly with regard to the quality of care you received. • What role do you expect your doctor to play during a teleconsultation? • Do you prefer face-to-face or virtual consultation with your doctor? Why? There are certain limitations of using virtual consultation, such as the absence of physical examination. Do limitations like this matter to you?
3	• Do you have any experience of using other digital tools for health care, particularly a medical monitoring device or wearable device? • If yes, share with me your experience. Did you face any inconveniences or challenges when using these devices? How did you overcome them? • What role do you expect your doctor to play when you are using these devices?
4	• Have you used any health care-related smartphone applications such as "HealthHub," to seek health care services? (HealthHub is a one-stop web portal with an accompanying digital application for obtaining information about health conditions, assessing health records, managing medical appointments, and paying medical bills, among other functions). • If yes, share with me your experience. Did you face any inconveniences when using the app? • If no, are you open to using such apps for health care purposes?
5	• In your opinion, how important is it for doctors to have the ability to interact with their patients and caregivers/family members, including explaining medical terms and giving medical advice clearly? • How important is it for doctors to have good inter-personal skills such as showing respect, care and compassion for their patients? • How important is it for doctors to maintain eye contact with you while he/she is keying in your medical records in the system? Why?
6	• I would like to know your thoughts on what the following stakeholders can do in order to improve patients' experience when digital technologies are being used: • Doctors • Hospitals and polyclinics (public primary health care clinics) • Medical schools (particularly in terms of training future medical graduates in digital technologies)
7	• Do you have any other concerns if your doctor were to use digital technologies to treat you or to diagnose your medical condition?
8	• Do you have any other comments on the skills or knowledge that doctors need to have when treating, diagnosing, assessing, and caring for you especially amid rapid technological advances?

Data Analysis

HZ and FKY read the transcripts independently and adopted an inductive thematic analysis approach when evaluating the data. This allowed us to draw common and shared meanings among the participants [32]. It also enabled the flexibility to accommodate new insights about the data [33]. Coding frameworks and themes were developed iteratively using the 6-step process proposed by Braun and Clarke [27] in which we first familiarized ourselves with the data by reading the transcripts in their entirety before generating relevant codes and combining them into themes. We then reviewed the themes before determining those that we deemed significant to the research questions and finally reporting the findings [27].

We also compared the findings with all relevant up-to-date local and global literature that report on patients' expectations of doctors, as well as views of stakeholders from the health care industry toward the competencies needed for current and future clinical practice [34-39]. Any coding discrepancies were resolved through consensus between HZ and FKY and through seeking the opinions of other coauthors.

Ethics Approval

Waiver for ethical approval was granted by SingHealth Centralised Institutional Review Board (reference 2020/2880).

Results

A total of 25 patients of different genders; from different ethnicities, socioeconomic backgrounds, educational levels, age bands; and with various medical conditions participated in our study. Their demographics are described in Table 3. In summary, 14 (56%) were male and 11 (44%) were female. Only 8 (32%)

out of 25 patients had used telemedicine, while 17 (68%) patients had used a medical monitoring or wearable device. As many as 18 (72%) participants had used the HealthHub app. Generally, the participants were receptive to the use of DHTs as long as their safety and personal medical data were not compromised.

With regard to wearable devices and mobile apps such as HealthHub, most participants (n=18, 72%) regardless of gender, ethnicity, age, socioeconomic background, and educational qualifications expected their doctors to be competent and more proactive in engaging them with the medical data found in these devices. They also expected doctors to handle the data in an ethical manner, in a way that abided by personal data protection laws. Specifically, participants such as P20 would like their doctors to offer them guidance on using medical devices correctly, review the results of laboratory tests with them, and explain any anomalies that may be present in their medical data (Table 4). There were also participants such as P5 who expected doctors to collaborate with patients in their health care journey through advising them on the type of apps that were safe to use instead of taking on paternalistic doctor-patient roles (Table 4). Others, such as P24, expected doctors to comply with the guidelines on data protection. Amid increased cybersecurity risks, she was concerned that her personal information might get leaked when doctors use mobile apps such as WhatsApp to share photos of patients' medical scans and other medical information with their colleagues (Table 4).

Regarding telemedicine, participants expected a level of assessment of their medical conditions that was similar to a physical consultation. Specifically, participants such as P22 expected their doctors to deliver comprehensive quality of care that would not compromise their safety when consulting patients on digital platforms. For example, P22's doctor had mistakenly prescribed him a lower dosage of medicine than what he needed during a Zoom consultation, which necessitated a visit to the clinic to obtain the correct dosage (Table 4). Participants were also appreciative if doctors were able to discern that their condition required a more detailed examination and that an in-person consultation was needed. For example, P24 expected doctors to know when to refer patients with severe medical symptoms to specialists for further medical assessment (Table 4). Moreover, participants would like doctors to perform a holistic examination of their conditions so as not to miss the symptoms of illnesses that could have otherwise been detected through physical examination. Participants such as P2 called for doctors to look out for signs that might reveal illnesses other than the one that the patient was seeking treatment for (Table 4). According to P2, doctors should know how to detect these through a thorough assessment of the patient's body language and facial expression. In general, participants preferred teleconsultation for acute illnesses such as coughs and colds. Physical consultation was deemed necessary when seeking treatment for their chronic conditions.

When it comes to participants' expectations of doctors' communication skills when using EHRs, eye contact was appreciated but deemed nonessential by most participants (n=15, 60%). Generally, participants did not find it necessary for doctors to maintain regular eye contact with them, as they valued the doctors' efficiency in carrying out their clinical work more than eye contact. To them, it was important for doctors to be competent and focused when documenting clinical records in order to avoid committing errors (Table 4). The older patients also shared that the rapport they already established with the doctor rendered sustained eye contact redundant (Table 4).

Although the participants did not regard regular eye contact during clinical documentation as essential, a significant proportion (n=21, 84%) opined that doctors should still display empathy in other phases of the clinical encounter regardless of whether DHTs were used. To participants such as P19, a patient with kidney disease, empathy should be conveyed even more in the digital age when DHTs were being used. P19 was concerned that doctors might not exhibit as much empathy if they were to use DHTs. This sentiment arose from her experience where even in the absence of DHTs, her doctor had not shown empathy toward her; the doctor had dismissed her pain and merely prescribed her medication after she shared about her skin condition (Table 4). A similar sentiment was shared by P25, a patient with paraplegia, who emphasized that empathetic communication, above any other skills, should be central to a doctor's bedside manners. This expectation came about following her brief conversation with her doctor, who informed her that she could no longer walk. According to P25, the doctor did not even show any empathy or take the time to explain how he derived at his diagnosis (Table 4).

In addition, most participants from the age group of 20 to 39 years saw a greater immediate need for doctors to use digital technologies such as AI and machine learning for more accurate diagnosis and treatment than those in the older age groups who were having similar conditions. Participants from the former age group expected doctors to be trained in the algorithms behind these technologies, as well as the underlying principles and purposes of different DHTs in order to better assist them in diagnosis and treatment. For example, P15, a 28-year-old patient, who had encountered difficulties with venipuncture, shared that having a technological device would help doctors locate her veins and draw her blood without inflicting any pain on her. This method was preferred to the existing one where she had to be poked multiple times by her doctors based on trial and error. Similarly, P21, a 27-year-old patient, believed that AI technologies would help streamline certain aspects of the health care process, making diagnosis and treatment more time efficient. This expectation came about after her doctors took 1 year to diagnose her with rheumatoid arthritis. Previously, 4 doctors who attended to her had dismissed her condition, as the symptoms did not fit the regular markers of rheumatoid arthritis.

Table 3. Demographics of participants (N=25).

Characteristics	Participants, n (%)
Gender	
Male	14 (56)
Female	11 (44)
Age (y), range	
20-39	8 (32)
40-59	10 (40)
≥60	7 (28)
Ethnicity	
Chinese	12 (48)
Indian	4 (16)
Malay	8 (32)
Other ethnicities	1 (4)
Educational qualification	
Secondary	5 (20)
Postsecondary: A level, diploma, ITE[a], or other postsecondary qualification	8 (32)
Bachelor’s degree	8 (32)
Master’s degree	3 (12)
PhD	1 (4)
Housing type	
1-2 room HDB[b]	3 (12)
3-room HDB	0 (0)
4-room HDB	8 (32)
5-room HDB	4 (16)
HDB maisonette	2 (8)
EC[c]	1 (4)
Private condominium	4 (16)
Landed property	3 (12)
Primary medical conditions	
Cancer	4 (16)
Cardiovascular diseases	3 (12)
Diabetes	3 (12)
Gout	3 (12)
Kidney disease	3 (12)
Neurological disease	3 (12)
Osteoarthritis	3 (12)
Rheumatoid arthritis	3 (12)
Duration of primary medical condition (years)	
1-5	7 (28)
6-10	8 (32)
11-15	2 (8)
16-20	2 (8)

Characteristics	Participants, n (%)
>20	6 (24)

[a]ITE: Institute of Technical Education.

[b]HDB: Housing & Development Board.

[c]EC: executive condominium.

Table 4. Illustrative quotes from interviews with participants.

Themes and subthemes	Quotes from participants
Active and ethical engagement with medical data in mobile apps and wearable devices	
Guidance from and collaboration with doctors in the use of mobile apps	"I expect doctors to collaborate with patients and guide them on which app is safe to use since there are so many of them out there...One of the points I keep making at various conferences is that, patients and doctors are not really collaborators in Asia, we are more like in a donor-and-a-beneficiary relationship, where the doctor is giving something and we are just taking it. We are not rising up as a partner in care. We need to take some responsibility, you know, contribute, understand, and then work together." [P5 aged 55 years, stroke survivor]
Doctors' knowledge of mobile apps	"I've used HealthHub app to check appointments and the results of my blood test, which will usually be out about two hours after the test. I will check my potassium level et cetera after taking blood at the hospital and before consultation. I expect the doctor to know how to use HealthHub too because it can help to speed up consultation with the doctor since both of us would have seen the health data already even before the consultation. There was once when I saw a young doctor for consultation and told her, 'I already know my potassium level, et cetera from the blood test, and she was like, how did you know?' So, I told her it's all in my phone. She might not know it because the app was just launched back then." [P15 aged 28 years, kidney and systemic lupus erythematosus]
Doctors' engagement with patients' medical data found in mobile apps and wearable devices during consultation	"When using the device, I expect doctors to help me how to use these stuffs correctly, review the results and explain anomalies." [P20 aged 39 years, diabetes, osteoarthritis, and rheumatoid arthritis]
Doctors' integrity when handling patients' data	(On concern about data privacy): "My concern is with data security. Doctors should not send photos of scans to their colleagues on WhatsApp for a quick consult. It's convenient but it's not ideal because sometimes your WhatsApp account can be synced to your Google photos. So, everything just automatically gets backed up in their personal account, unless they go and delete it. If they don't send the scan with our IC (Identification Card) number, it's not so bad. So, it's important to protect confidential information." [P24 aged 39 years, thyroid cancer and rheumatoid arthritis]
Doctors' awareness of cybersecurity risks	(On concern about data privacy): "My concern is with data security. Doctors should not send photos of scans to their colleagues on WhatsApp for a quick consult. It's convenient but it's not ideal because sometimes your WhatsApp account can be synced to your Google photos. So, everything just automatically gets backed up in their personal account, unless they go and delete it. If they don't send the scan with our IC [identification card] number, it's not so bad. So, it's important to protect confidential information." [P24 aged 39 years, thyroid cancer and rheumatoid arthritis]
Comprehensive and uncompromised quality of care	
Undivided attention	"I don't think my expectations of the doctor would be the same when it comes to teleconsult. Yes, it's kind of a replication of the clinical setting where we are talking to each other. But now, the doctor has to look at me and talk, right? Whereas when in the clinic, the doctor can be distracted by the computer, or saying something to the nurse or passing a note to whoever walks in. There are a lot of things happening in the clinical setting whereas in a zoom meeting, you are literally talking to that one person, so I'd expect there to be 100% attention." [P1 aged 54 years, cancer]
Pitfalls and limitations of teleconsultation	"In terms of preference, I still prefer face-to-face so that I can get appropriate care and treatment because when I do it over Zoom, the doctor cannot monitor my blood pressure and breathing. I've been given a lower dosage of medicine before when doing consultation over zoom. When that happened, I still needed to do a face-to-face consult anyway to take a higher dose of medicine." [P22 aged 35 years, asthma and mitral valve prolapse]
Knowledge of medical conditions that require further assessment	"I would expect the doctor to know when it's time to escalate the situation to a specialist. So, if the symptoms are severe, they should refer the patient to urgent care." [P24 aged 39 years, thyroid cancer and rheumatoid arthritis]
Holistic assessment	"Before going into condition-specific assessment, I'd expect the doctor to check the patient's overall health and mobility issues. Maybe for older patients, can get them to stand up, walk a few paces because based on their movements, doctors can detect a lot of other things. So, they shouldn't just focus on their specialty, but examine the patients on other things as well, like their tone of voice, facial expression, and look out for signs that may tell their emotional issues, mental issues. So, doctors should have a protocol for these things before they go into the specifics." [P2 aged 64 years, cancer]
Efficiency outweighs eye contact for patients when doctors use EHRs[a]	
Doctors' efficiency is valued by patients more than eye contact	"I don't expect them to maintain eye contact the whole time because I understand that they need to see the computer, our medical conditions, and everything. It's more important for them to know what's happening to us. Maybe, eye contact is more important in the ward, but in the clinic, I don't really mind if the doctor does not have eye contact with me." [P13 aged 42 years, diabetes]

Themes and subthemes	Quotes from participants
Eye contact is appreciated but deemed nonessential	“Most doctors don’t have eye contact with the patient...when you know your doctor well and have confidence in the doctor, I don’t think it is necessary, though it would be good to have.” [P14 aged 75 years, gout]
Empathetic communication regardless of DHT[b] use	
Patients value human touch regardless of DHT use	“Even as doctors use digital technology, they still need to have that human touch. Based on my experience, I had a skin condition where there was a lump. When I went for consultation, the doctor did not even touch or see it. She just prescribed me antibiotics and asked me to come back for a one-week appointment. So, I think they should show more concern towards the patient, listen to our complaints and problems. Don’t just prescribe medicine and send the patient off. I was so upset. I don’t want other patients to go through the same experience because I was in pain and she didn’t even see to it. So, if without technology, they can already ignore the patient, what more if technology is present?” [P19 aged 56 years, kidney disease]
Effective communication and bedside manners regardless of DHT use	“Even if doctors were to adapt digital technology in the health care setting, at the end of the day, it boils down to whether they can deliver a message or communicate a diagnosis to the patient empathetically or not...when relaying a message, doctors shouldn’t just approach the patient at the bed, and say, ‘Hey, I’d like to tell you that you are diagnosed with this condition,’ then just walk away without elaborating on the statement. It’s not helpful for the patient who is trying to process it mentally and who is not well not-versed in the condition. So, doctors should explain to the patient on how they derive at the diagnosis. When I was officially diagnosed with this disease, I couldn’t move my legs, my toes, I couldn’t move anything at all. The doctor just opened the curtain and told me, ‘I don’t think you can walk anymore,’ and just walked away. I was alone at that time, and just woke up from a one-and-a-half month coma, so I broke down there and then, non-stop. There was just so much emotions. It seems like when people go up the corporate ladder, they tend to neglect the humanistic aspect, the empathy.” [P25 aged 26 years, tuberculosis meningitis and pulmonary tuberculosis]
Competence in using DHTs for clinical procedure and diagnosis	
Competence in using DHTs for venipuncture (younger participants)	“My veins are very fine, so I hope that in the future, there will be a machine or technology that can help locate my veins and poke them without having to poke a lot of times. Humans have to do it based on trial and error, and it’s very painful. Sometimes, my doctors have to poke me four or five times just to get the vein. So, if there is a machine that can poke once only and make sure there is blood in the veins, that would be cool. Otherwise, patients like me will suffer. Sometimes, when the senior doctors ask the junior doctors to try and poke me, they will poke multiple times. It is very painful!” [P15 aged 28 years, kidney and systemic lupus erythematosus]
Competence in using DHTs to make accurate diagnosis (younger participants)	“Digital technologies are good; they can be predictive. Some doctors tend to stick to the books, so, they may not be able to find out the conditions as easily as they will when aided with digital technologies. When I told my doctors that I suspected I have rheumatoid arthritis, they said it’s not rheumatoid because my conditions don’t fit the definitive terms and criteria, like pain in the usual wear and tear areas. About four doctors including GPs and specialists did not think it was rheumatoid. But eventually, it was only when I had pain in my toes, on top of the usual markers, then the specialist diagnosed me with rheumatoid arthritis. So, using digital technologies to diagnose conditions will be helpful...machines are less likely to miss errors; they can pick them up better than doctors.” [P21 aged 27 years, rheumatoid arthritis]
Accuracy and precision (younger participants)	“Digital technologies are good; they can be predictive. Some doctors tend to stick to the books, so, they may not be able to find out the conditions as easily as they will when aided with digital technologies. When I told my doctors that I suspected I have rheumatoid arthritis, they said it’s not rheumatoid because my conditions don’t fit the definitive terms and criteria, like pain in the usual wear and tear areas. About four doctors including GPs and specialists did not think it was rheumatoid. But eventually, it was only when I had pain in my toes, on top of the usual markers, then the specialist diagnosed me with rheumatoid arthritis. So, using digital technologies to diagnose conditions will be helpful...machines are less likely to miss errors; they can pick them up better than doctors.” [P21 aged 27 years, rheumatoid arthritis]
Time efficiency (younger participants)	“Digital technologies are good; they can be predictive. Some doctors tend to stick to the books, so, they may not be able to find out the conditions as easily as they will when aided with digital technologies. When I told my doctors that I suspected I have rheumatoid arthritis, they said it’s not rheumatoid because my conditions don’t fit the definitive terms and criteria, like pain in the usual wear and tear areas. About four doctors including GPs and specialists did not think it was rheumatoid. But eventually, it was only when I had pain in my toes, on top of the usual markers, then the specialist diagnosed me with rheumatoid arthritis. So, using digital technologies to diagnose conditions will be helpful...machines are less likely to miss errors; they can pick them up better than doctors.” [P21 aged 27 years, rheumatoid arthritis]

Themes and subthemes	Quotes from participants
Concern on the reliability of technology (older participants)	"Retirees like me pick up things much slower than the younger generation, who use technology more regularly. I am open to technology, but the thing about technology is that if the technology goes bust, then all the data in the device will be gone. All my appointments will be lost. If you were to lose electricity, or the technology breaks down, then you may not get back the data. I am also open to robots diagnose my conditions or treat me, but I am concerned about what the robots will come up with. So, we shouldn't be too reliant on technology...when it comes to diagnosis, I'll be quite skeptical of course if a robot were to diagnose me. Eventually, you still need a doctor to oversee the diagnosis the technology has come up with." [P14 aged 75 years, gout]
Concern of trust on the reliability of diagnosis if it is based on DHTs alone (older participants)	"I'm not open to technology because technologies like AI do not have emotions like humans do. Even if it helps to improve the accuracy of diagnosis, I'm still not open to it because what if there is no electricity? You can't just depend on technology. It can trip anytime. If there's a trip, everything will be gone and doctors won't know what to do." [P19 aged 56 years, kidney disease]

[a]EHR: electronic health record.

[b]DHT: digital health technology.

Compared to the younger participants, those from the older generation saw less need for the use of DHTs. They were concerned about the reliability of diagnosis, the perceived absence of human touch, and the risk of losing their medical data in the event of a technological breakdown if they were to be treated with DHTs. This was expressed by participants such as P14, a 75-year-old patient, and P19, a 56-year-old patient (Table 4). P14 and P19 continued to favor the diagnosis made by doctors over that made by DHTs due to their greater trust in the expertise of medical professionals. They would also opt for diagnosis and treatment advised by doctors over what is suggested DHTs, as the former was understood to offer emotional connection and understanding, which were lacking in the latter. Moreover, older participants were generally skeptical about the reliability of technology. Relatedly, they had concerns about the doctors' ability to recover their medical data should there be a system failure or technical malfunction.

Overall, the findings indicated that age, compared to other social determinants, was more influential in differentiating the participants' health care experiences and expectations, at least in this study. By gathering the views of the patients who were seeking medical care from outpatient clinics and who had sought treatment in the wards, this study sheds light on the importance of considering their health care experience and expectations to ensure that their needs were not being ignored. Interviewing the patients who had various medical conditions also highlights the type of digital technology that would be useful for specific health care purposes.

Discussion

Principal Findings

Overall, the participants expected doctors to be competent in using different DHTs. With regard to apps and wearables, they would like doctors to integrate data from these devices into doctor-patient communication and to handle their data ethically. When it comes to telemedicine, they expected doctors to deliver comprehensive and high-quality care that does not compromise their safety. According to them, doctors should be knowledgeable enough to identify the type of medical conditions that were appropriate for online consultation and those that needed further assessment. In addition, while participants did not perceive regular eye contact when using EHRs as essential, they still valued the doctor's display of empathy in other phases of the clinical encounter, regardless of DHT use. The younger participants also expected doctors to be trained in the algorithms, principles, and purposes of DHTs such as AI to better assist them in the diagnosis and treatment processes. By interviewing patients, we have obtained the perspectives of an important yet often overlooked stakeholder in the health care system.

To our knowledge, this is the first study to assess patient perspectives of the existing gaps in doctors' clinical competencies when deploying DHTs. Our findings are unique, as the medical conditions among the study participants were varied. Previous studies have either explored DHT deployment among adults with specific single chronic conditions, such as hypertension, diabetes, and chronic heart failure or are not focused on patients' expectations of doctors' clinical skills [12,40-42]. Interviewing participants with various chronic conditions has offered us insights into their preferred mode of consultation; while teleconsultation was preferred for acute illnesses, physical consultation was still deemed necessary for chronic conditions. It also proved that the need to equip doctors with DHT competencies for better diagnosis and treatment was not specialty dependent but was something that needed to be implemented across the health care sector.

One similarity between the findings of this study and those of others is that DHTs are not the main driver of health-seeking behavior among older adults [12,40]. In general, they prefer receiving traditional health care services from medical professionals to exclusively relying on DHTs. However, unlike other studies indicating that older adults with lower educational qualifications (primary or secondary level) are less receptive and less likely to use DHTs [12,40], our research suggests that participants' limited receptiveness is not necessarily correlated with their educational qualifications. Those with postsecondary education and above also expressed reluctance to use DHTs. Our results therefore offer a nuanced perspective on patients' attitudes toward technology.

By conducting qualitative interviews with patients, this study has uncovered diverse views and informed future studies about the need to avoid associating perceived attitudes with specific social identities at the outset of research. This may only run the risk of perpetuating stereotypes about people belonging to certain identities. As the findings have shown, patients' expectations and concerns need not be differentiated by their race, gender, or socioeconomic status. Rather, factors such as age and type of medical conditions tend to be more salient, proving the value of our constructivist and inductive approaches.

Unlike past qualitative studies, the lack of sustained eye contact between the doctor and patient during clinical documentation did not seem to matter to the participants of this study [43-46]. This could be attributed to the transactional relationship between the doctor and patient that characterizes many clinical encounters in Singapore. This type of care recognizes that the patient has a specific need, diagnoses and treats the condition, controls the risk factor, and makes a referral [47]. The quality of care tends to be determined by the ability of the doctor to abide by a set of prescribed guidelines that makes him or her a "good" doctor [47]. This differs from relationship-based care, which tends to focus on the quality of interaction between the doctor and patient [47]. The omnipresence and relative accessibility of doctors in Singapore compared to those in other countries may explain this taken-for-granted aspect of relationship [48]. Furthermore, in a fast-paced country like Singapore where efficiency and accuracy in the management of medical conditions are highly valued [48,49], the lack of eye contact seems like a characteristic that patients are willing to forego. Nonetheless, these should not compromise the humanistic and empathetic practice of medicine. As our participants such as P2 have expressed, appropriate clinical inquiry is still deemed important when DHTs are used. This is reiterated by the existing studies, which prove that attentiveness to the computer screen rather than eye contact with the patient during clinical encounters does adversely affect doctors' psychosocial inquiry, emotional patient responsiveness, and full patient disclosure [46,48].

Some of the findings of this study have also been reported by other countries. These include the importance of identifying the cases that are appropriate for teleconsultation. Past reviews and studies have also reported the need for doctors to be equipped with the knowledge of selecting patients for teleconsultation so as not to compromise their safety [50-52]. However, many of these studies are based on individual medical conditions that are deemed stable. Future research should also consider the selection and identification procedures of patients with comorbidities and complex conditions since these are becoming increasingly common worldwide [53].

The value of incorporating training in AI technologies early in medical education has also been recognized in a large body of works. Examples of AI competencies include knowing the limitations, risks, and medicolegal aspects of AI [54-56]. In addition, concerns about privacy and data intrusion when DHTs are used are also commonly reported by other studies [7,12]. To illustrate, a study conducted with residents, patient representatives, and health care providers at a health facility catchment in Sydney, Australia, reported that both groups expressed concerns with safety issues such as data safety and privacy and the risk of hacking when telemedicine is used [7].

Other works have also recognized the value of integrating data from DHTs into the interaction with patients [6,57]. For example, the study by Loos and Davidson [57] that assessed doctors' views on the potential integration of wearables into patient care showed how effective doctor-patient communication aided by wearables serves a few purposes. These include forging strong interpersonal relationships between the doctor and patient and getting accurate information from patients to inform diagnosis, treatment decisions, and overall management of health [57]. However, as shown by the scoping review conducted by Hilty et al [58], there is currently a lag in the clinical, technological, and administrative workflow at the international level with regard to incorporating sensors, wearables, and remote patient monitoring in clinical care [58]. Hence, standardized frameworks of competencies need to be developed for doctors to effectively engage patients with these devices.

Moreover, our findings reiterate those of other studies when it comes to how age affects the use of DHTs. Specifically, participants who belong to the advanced age bands did not perceive an urgent need for their doctors to deploy DHTs for their medical conditions, unlike the younger ones. This is exemplified by the qualitative study by Low et al [12] of how older adults in Singapore negotiate DHTs in their everyday lives. In this study, the authors discovered not only a low uptake of DHTs among those aged 50 to 65 years but also the lack of perceived immediate need for them to use these technologies despite their willingness to adopt them. This is mainly attributed to the lack of technology-centeredness in their health-seeking behavior [12]. Another study corroborates this by demonstrating how adults from *Generation X*, defined as those born between 1943 and 1960, required technology training more than the *millennial generation*, referring to those born between 1982 and 2000. This is in view of the Generation X's lower comfort level with using DHTs, highlighting their lesser dependence on and reluctance to use DHTs [1]. Hence, initiatives to increase the awareness and acceptance of DHTs among the population are needed if they were to be implemented nationwide.

Recommendations

The findings have reiterated the importance of exploring the views of patients so that their needs and the competencies of doctors can be more aligned in this digital age. This section offers recommendations on how each of the participants' expectations and concerns can be addressed accordingly.

Engaging Data in Apps and Wearables Effectively With Patients

To fulfill patients' expectations of having their doctors communicate the meaning of their medical data with them, protected time is needed for doctors during consultations. On the basis of the participants' sharing, time constraints and a hectic work environment seem to be major barriers that prevent doctors from communicating effectively with patients. Hence, having protected time would allow them sufficient time to offer the necessary advice to their patients on the use of DHTs and provide adequate explanations of the data. Health care institutions can work with designers of technology to put in place a system that works like a voice-to-text tool where communication between the doctor and patient can be captured and transcribed in real time. This would help reduce doctors' time on clinical documentation, allowing them more time to interact with patients and respond appropriately to any of their concerns.

In view of the evolving model of care where the traditional model of the doctor-patient relationship is being replaced by shared decision-making between the doctor and patient, doctors should also act as expert partners in patients' health care journey. In this regard, it is important to consider the study by Mesko et al [59], which highlights that driving behavioral change for patients entails not only just technological shifts but also a consideration of humanistic elements. Specifically, more opportunities need to be created to educate patients about DHTs [59]. For example, doctors can advise their patients on how they can use DHTs in a safe and effective manner during consultations. These include identifying and downloading legitimate apps and teaching them how to access their medical data. This would not only expand patients' knowledge of DHTs but also enhance the doctor's commitment to supporting their overall well-being.

Determining Suitable Medical Conditions for Teleconsultations

To train doctors in holistic assessment of patients' conditions during teleconsultation, as well as in determining conditions suitable for this type of consultation, a comprehensive set of guidelines needs to be developed at the international level. Reviews on telemedicine have shown that there is a lack of guidelines and standards for implementing telemedicine both at the international and national levels [60,61]. Although many telemedicine reports have highlighted the need to develop these guidelines, only a few exist in practice [60]. Current telemedicine guidelines for clinical practice worldwide have been limited to specific specialties such as psychiatry, dermatology, pathology, and radiology [60]. A policy review on telemedicine in the Southeast Asia region of the World Health Organization also reveals significant variations in the adoption and implementation of telemedicine guidelines among 11 Southeast Asia region countries of the World Health Organization [57,61]. Among the 11 countries, only 5 have developed telemedicine guidelines including India, Bangladesh, Thailand, Indonesia, and Nepal [61].

In Singapore, the country's Medical Council and professional bodies have developed codes and guidelines to regulate the use of telemedicine [62,63]. However, the considerations for delivering care via telemedicine are outlined only broadly. Specifically, the reasonableness of conducting teleconsultations is determined by "the clinical context, the clinical objectives and the compatibility of technology to meet those objectives" [62]. The delivery of care using telemedicine is based on general Clinical Practice Guidelines [62]. In addition, there are no specific guidelines to determine the type of cases that are eligible for telemedicine consultation. Currently, the diagnosis, prescription of medicine, and issuance of medical certificates via telemedicine depend on the "professional judgement of the relevant doctor" and the "specific facts and circumstances of each presenting case" [63].

To ensure the standardization of current work practices and guidelines, some authors have highlighted the important role of international telemedicine organizations [60]. For example, organizations such as the American Telemedicine Association and the United Kingdom's Telemedicine and eHealth Forum of the Royal Society of Medicine can take the lead by defining telemedicine guidelines under the direction of clinicians with relevant telemedicine expertise [60]. These clinicians should come from different medical specialties so that the health care needs of patients with comorbidities can be fulfilled.

Practicing Empathetic Communication When Using DHTs

To ensure that empathetic communication is not compromised when doctors are using DHTs, they can be trained to practice this skill alongside the adoption of DHTs. For example, during virtual consultations, doctors can practice active listening when communicating with patients. This involves giving patients their full attention and conveying attentiveness through nonverbal cues such as facial expressions and body language. These skills can help build a connection with patients.

A systematic review on how compassion is discussed in relation to AI technologies in health care shows that there are different ways in which AI can promote compassion. These include enhancement of the empathetic awareness of patients' suffering, empathetic response and relational behavior, communication skills, health coaching, therapeutic interventions, moral development learning, clinical knowledge and assessment, health care quality assessment, therapeutic bond and alliance, and provision of health information and advice [64]. However, most of these studies discuss how AI technologies can be used to train health care professionals and trainees to deliver clinical care using virtual platforms and patients. Further research on how DHTs such as EHRs, apps, and wearables can be leveraged during clinical consultations with real patients needs to be conducted.

At present, the framework proposed by Loos and Davidson [57] of the competencies for clinicians and trainees when using sensors, wearables, and remote monitoring devices may serve as a guide in determining the necessary competencies. This framework is organized according to 3 learner levels, namely, novice or advanced beginner, competent or proficient, and advanced or expert levels and is based on the domains outlined by the Accreditation Council of Graduate Medical Education (ACGME) [57]. Beyond ways to embody interpersonal and good communication skills, it also outlines how clinicians and trainees can deliver quality care through their knowledge of these DHTs, for example, learning the types of diseases that can be treated with wearables (ACGME milestone levels 1-2); developing a plan to review, communicate, and deliberate on data (ACGME milestone levels 3-4); and researching new ways to improve care using DHTs such as AI (ACGME milestone level 5). Standardized evaluation measures may help to determine the effectiveness of such frameworks.

Using AI Technologies Competently and Ethically

To train doctors to be competent in DHTs such as AI, professional bodies can offer courses and training programs to doctors through continuing medical education, as has been done by the Digital Medicine Society in partnership with Rocky Vista College of Medicine and the American College of Osteopathic

Family Medicine in the United States [65]. Such programs can help educate and encourage health care professionals to embrace digital medicine and train them in the benefits, risks, and pitfalls of AI. Another way to encourage the adoption of DHTs among health care professionals is to share published data on studies that involve large patient cohorts and that report on the accuracy of well-documented DHTs.

To address the concerns about possible intrusions of personal data, a regulatory framework that understands the workings of technological innovations and their potential lapses needs to be introduced in order to prevent the leakage of sensitive information. In addition, tighter laws on data breaches need to be legislated to improve public trust. In the United States, laws such as Genetic Non-Discrimination Act serve to defend patients from unauthorized third-party access to data [59]. More such laws need to be devised and implemented to increase public trust in DHTs. Doctors and medical trainees also need to be trained in handling medical data ethically in order to safeguard patients' confidential information. The proliferation of medical data breaches in recent years caused by cyberattacks and ransomware attacks in Singapore and around the world necessitates tighter laws to protect patients' medical data.

Strengths and Limitations

This qualitative study informs us about the expectations and concerns of patients with DHTs. By interviewing participants from different social and economic backgrounds, as well as medical conditions, the sample achieved diversity in narratives, and the study benefited from the rich data. Seeking the opinions of patients from outpatient clinics, most of whom had been hospitalized for the same conditions, also allows us to uncover their experiences in both the inpatient and outpatient settings. In addition, conducting qualitative interviews enables us to gain in-depth insights into patients' experiences and expectations and place their narratives at the forefront of our research.

A perceived limitation of this study is its small sample size. This may limit the generalizability, validity, and reliability of the study. With a small sample size, it is challenging to generalize the findings to a larger population. The findings may only be applicable to the individuals included in the study and may not represent broader groups [21]. Furthermore, a small sample size may not adequately capture the diversity within the population of interest [21]. Consequently, the findings may lack depth and breadth, leading to a limited understanding of the phenomenon under investigation. In addition, small sample sizes increase the risk of selection bias in the recruitment strategy. Researchers may inadvertently select participants who are more accessible or willing to participate, thus leading to a biased sample that does not accurately represent the population.

Nonetheless, as with other qualitative studies that adopt interviewing techniques, an in-depth analysis of participants' narratives offers a contextualized understanding of their expectations and concerns. Furthermore, although not entirely generalizable, the findings from this study would bear important implications for digital health training programs, initiatives, and frameworks in other developed countries. As a scoping review on digital health competency frameworks for health care professionals has shown, frameworks for training doctors and medical trainees in relevant digital competencies are still lacking [66].

Another limitation lies in the lack of language diversity among participants, all of whom happened to be English-speaking. Their English language proficiency is likely attributed to the widespread use of English as the official language of communication in Singapore. Having research participants come from a homogenous linguistic background may restrict the ability to apply the findings to diverse communities where other languages are spoken. Since language is intricately linked to culture and educational background [8], this may also exclude the influence of cultural and class factors on patient expectations. Despite the language limitation, participants came from diverse ethnic, educational, and socioeconomic backgrounds.

In addition, we acknowledge that there are other digital technologies that are not discussed in this study, given its focus on the type of technologies that have been introduced for clinical practice in Singapore. Therefore, future research should also examine how more recent AI technologies such as ChatGPT could potentially transform health care from the perspectives of different stakeholders including patients and health care professionals. ChatGPT is an AI-powered conversational agent developed by OpenAI. It is based on the GPT architecture. Designed to understand and generate humanlike text based on the input it receives, it has the potential to revolutionize health care delivery and services in several ways. These include promoting healthier lifestyle habits through personalized health coaching and interventions and serving as a clinical decision support tool for health care providers. The latter may include offering real-time access to evidence-based guidelines, medical literature, and treatment protocols. Such technologies raise the question of how medical diagnosis, treatment plans, and patient management may evolve in the future.

Conclusions

This study has explored the clinical encounters of patients with chronic illnesses amid the increasing digitalization of health care. Evaluating their perspectives proves crucial, as they can be considered a key facilitator for technology implementation and enhancement in health care settings. By identifying their expectations of doctors' clinical competencies, this study has shown that traditional and basic clinical skills such as effective communication, clinical reasoning, history taking, physical examination, and procedural skills should neither be neglected nor compromised in the digital age. Nonetheless, the role of clinicians needs to evolve with the introduction of DHTs. For example, the way in which DHTs are transforming the traditionally paternalistic rapport between the doctor and patient needs to be considered. With patients now holding the power to assess the standard of health care received and being more involved in the decision-making processes of their own health [59], an expansion of doctors' clinical competencies is needed in order to consider these shifts.

In addition, the findings of this study might inform the health care practices and policies in other high-income countries. These include prioritizing the integration of digital health care education into medical training programs to equip doctors with the necessary competencies. Examples of initiatives may include updating the curricula of medical schools, offering continuing professional development opportunities, and fostering a culture of lifelong learning among health care professionals. In addition, governments and health care systems should allocate resources toward implementing DHTs that align with patient expectations and preferences. They could also implement patient engagement and education initiatives to increase the awareness and acceptance of DHTs among the population. Moreover, policy makers should develop regulations and guidelines to govern the use of DHTs in health care delivery including ensuring data privacy and security and addressing issues related to liability. Overall, the findings from this research could serve as a road map for other high-income countries to leverage the potential of DHTs in enhancing health care delivery in the digital age.

Acknowledgments

This study was funded by SingHealth Duke-NUS Medicine Academic Clinical Programme under Seah Cheng Siang Distinguished Professorship in Medicine. The authors would like to thank the participants of this study for their invaluable insights and the reviewers of this journal for their helpful comments. They are also grateful to the following individuals for their assistance in some of the recruitment efforts: Woon Ting Hui from Singapore General Hospital; Gordon Ho from Society for the Physically Disabled, a local charity set up to help people with disabilities; Zarina Yusof from National Kidney Foundation; and Syawal Yusoff from BERITAMediacorp.

Conflicts of Interest

None declared.

References

1. Alkire (née Nasr) L, O'Connor GE, Myrden S, Köcher S. Patient experience in the digital age: an investigation into the effect of generational cohorts. J Retail Consum Serv 2020 Nov;57:102221. [doi: 10.1016/j.jretconser.2020.102221]
2. Digital health center of excellence. US Food and Drug Administration. 2024 Jul 25. URL: https://www.fda.gov/medical-devices/digital-health-center-excellence [accessed 2024-08-05]
3. Welcher CM, Hersh W, Takesue B, Stagg Elliott V, Hawkins RE. Barriers to medical students' electronic health record access can impede their preparedness for practice. Acad Med 2018 Jan;93(1):48-53 [FREE Full text] [doi: 10.1097/ACM.0000000000001829] [Medline: 28746069]
4. Rienits H, Teuss G, Bonney A. Teaching telehealth consultation skills. Clin Teach 2016 Apr;13(2):119-123. [doi: 10.1111/tct.12378] [Medline: 26016733]
5. Wootton R. Telemedicine and isolated communities: a UK perspective. J Telemed Telecare 1999;5 Suppl 2:S27-S34. [doi: 10.1258/1357633991933495] [Medline: 10628015]
6. Reichmann H, Klingelhoefer L, Bendig J. The use of wearables for the diagnosis and treatment of Parkinson's disease. J Neural Transm (Vienna) 2023 Jun;130(6):783-791 [FREE Full text] [doi: 10.1007/s00702-022-02575-5] [Medline: 36609737]
7. Clay-Williams R, Hibbert P, Carrigan A, Roberts N, Austin E, Fajardo Pulido D, et al. The diversity of providers' and consumers' views of virtual versus inpatient care provision: a qualitative study. BMC Health Serv Res 2023 Jul 04;23(1):724 [FREE Full text] [doi: 10.1186/s12913-023-09715-x] [Medline: 37400807]
8. Population and population structure. Department of Statistics Singapore. URL: https://www.singstat.gov.sg/find-data/search-by-theme/population/population-and-population-structure/latest-data [accessed 2024-04-26]
9. Visaria A, Aithal S, Malhotra R. Digital technology use, in general and for health purposes, by older adults in Singapore. Aging Health Res 2023 Mar;3(1):100117. [doi: 10.1016/j.ahr.2023.100117]
10. Lwin MO, Sheldenkar A, Panchapakesan C. A digital mobile community app for caregivers in Singapore: predevelopment and usability study. JMIR Nurs 2021 May 26;4(2):e25679 [FREE Full text] [doi: 10.2196/25679] [Medline: 34345796]
11. Improving healthcare experience with technologies. Accenture. 2022 Jan 27. URL: https://www.accenture.com/sg-en/insights/health/consumer-survey-2022 [accessed 2024-08-05]
12. Low ST, Sakhardande PG, Lai YF, Long AD, Kaur-Gill S. Attitudes and perceptions toward healthcare technology adoption among older adults in Singapore: a qualitative study. Front Public Health 2021 Feb 15;9:588590 [FREE Full text] [doi: 10.3389/fpubh.2021.588590] [Medline: 33659232]
13. Smart health initiatives. Smart Nation and Digital Government Office. URL: https://www.smartnation.gov.sg/initiatives/health/ [accessed 2024-01-08]
14. Liu S, Ko QS, Heng KQ, Ngiam KY, Feng M. Healthcare transformation in Singapore with artificial intelligence. Front Digit Health 2020 Nov 17;2:592121 [FREE Full text] [doi: 10.3389/fdgth.2020.592121] [Medline: 34713061]
15. Michels ME, Evans DE, Blok GA. What is a clinical skill? Searching for order in chaos through a modified Delphi process. Med Teach 2012;34(8):e573-e581. [doi: 10.3109/0142159X.2012.669218] [Medline: 22530719]
16. Otaki J, Nagata-Kobayashi S, Takayashiki A, Ono M, Fukushi M, Matsumura S. The most requested factors in clinical skills exams for evaluating novice physicians: an internet-based survey of the general public in Japan. BMC Med Educ 2013 May 24;13:74 [FREE Full text] [doi: 10.1186/1472-6920-13-74] [Medline: 23706105]

17. Baines R, Bradwell H, Edwards K, Stevens S, Prime S, Tredinnick-Rowe J, et al. Meaningful patient and public involvement in digital health innovation, implementation and evaluation: a systematic review. Health Expect 2022 Aug;25(4):1232-1245 [FREE Full text] [doi: 10.1111/hex.13506] [Medline: 35526274]
18. Voorheis P, Petch J, Pham Q, Kuluski K. Maximizing the value of patient and public involvement in the digital health co-design process: a qualitative descriptive study with design leaders and patient-public partners. PLOS Digit Health 2023 Oct 25;2(10):e0000213 [FREE Full text] [doi: 10.1371/journal.pdig.0000213] [Medline: 37878566]
19. Kassam I, Ilkina D, Kemp J, Roble H, Carter-Langford A, Shen N. Patient perspectives and preferences for consent in the digital health context: state-of-the-art literature review. J Med Internet Res 2023 Feb 10;25:e42507 [FREE Full text] [doi: 10.2196/42507] [Medline: 36763409]
20. Mayring P. Qualitative content analysis: theoretical foundation, basic procedures and software solution. Social Science Open Access Repository. 2014. URL: https://www.ssoar.info/ssoar/handle/document/39517 [accessed 2024-08-05]
21. Sandelowski M. Sample size in qualitative research. Res Nurs Health 1995 Apr;18(2):179-183. [doi: 10.1002/nur.4770180211] [Medline: 7899572]
22. Morse JM. Qualitative Nursing Research: A Contemporary. Thousand Oaks, CA: SAGE Publications; 1991.
23. Patton MQ. Qualitative Research & Evaluation Methods: Integrating Theory and Practice. Thousand Oaks, CA: SAGE Publications; 2014.
24. Epidemiology and Disease Control Division, Ministry of Health, Singapore, Institute for Health Metrics and Evaluation. The burden of disease in Singapore, 1990–2017: an overview of the global burden of disease study 2017 results. Institute for Health Metrics and Evaluation. 2019. URL: https://www.healthdata.org/sites/default/files/files/policy_report/2019/GBD_2017_Singapore_Report.pdf [accessed 2024-08-05]
25. Chen C, Lim JT, Chia NC, Wang L, Tysinger B, Zissimopolous J, et al. The long-term impact of functional disability on hospitalization spending in Singapore. J Econ Ageing 2019;14:100193 [FREE Full text] [doi: 10.1016/j.jeoa.2019.02.002] [Medline: 31857943]
26. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
27. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
28. Kallio H, Pietilä AM, Johnson M, Kangasniemi M. Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. J Adv Nurs 2016 Dec;72(12):2954-2965. [doi: 10.1111/jan.13031] [Medline: 27221824]
29. Gopal G, Suter-Crazzolara C, Toldo L, Eberhardt W. Digital transformation in healthcare - architectures of present and future information technologies. Clin Chem Lab Med 2019 Feb 25;57(3):328-335 [FREE Full text] [doi: 10.1515/cclm-2018-0658] [Medline: 30530878]
30. Consolidated telemedicine implementation guide. World Health Organization. 2022 Nov 03. URL: https://www.who.int/publications/i/item/9789240059184 [accessed 2024-08-05]
31. Evans K, King D. Studying Society. Milton Park, UK: Routledge; 2005.
32. Kiger ME, Varpio L. Thematic analysis of qualitative data: AMEE guide no. 131. Med Teach 2020 Aug;42(8):846-854. [doi: 10.1080/0142159X.2020.1755030] [Medline: 32356468]
33. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000 Aug;23(4):334-340. [doi: 10.1002/1098-240x(200008)23:4<334::aid-nur9>3.0.co;2-g] [Medline: 10940958]
34. Zainal H, Xiaohui X, Thumboo J, Yong FK. Exploring the views of Singapore junior doctors on medical curricula for the digital age: a case study. PLoS One 2023 Mar 02;18(3):e0281108 [FREE Full text] [doi: 10.1371/journal.pone.0281108] [Medline: 36862708]
35. Zainal H, Xin X, Thumboo J, Fong KY. Medical school curriculum in the digital age: perspectives of clinical educators and teachers. BMC Med Educ 2022 Jun 03;22(1):428 [FREE Full text] [doi: 10.1186/s12909-022-03454-z] [Medline: 35659212]
36. Kamal AH, Zakaria OM, Majzoub RA, Nasir EW. Artificial intelligence in orthopedics: a qualitative exploration of the surgeon perspective. Medicine (Baltimore) 2023 Jun 16;102(24):e34071 [FREE Full text] [doi: 10.1097/MD.0000000000034071] [Medline: 37327255]
37. Tanaka M, Matsumura S, Bito S. Roles and competencies of doctors in artificial intelligence implementation: qualitative analysis through physician interviews. JMIR Form Res 2023 May 18;7:e46020 [FREE Full text] [doi: 10.2196/46020] [Medline: 37200074]
38. Wagner G, Raymond L, Paré G. Understanding prospective physicians' intention to use artificial intelligence in their future medical practice: configurational analysis. JMIR Med Educ 2023 Mar 22;9:e45631 [FREE Full text] [doi: 10.2196/45631] [Medline: 36947121]
39. Zainal H, Xiaohui X, Thumboo J, Kok Yong F. Digital competencies for Singapore's national medical school curriculum: a qualitative study. Med Educ Online 2023 Dec;28(1):2211820 [FREE Full text] [doi: 10.1080/10872981.2023.2211820] [Medline: 37186901]

40. Lai WX, Visaria A, Østbye T, Malhotra R. Prevalence and correlates of use of digital technology for managing hypertension among older adults. J Hum Hypertens 2023 Jan;37(1):80-87 [FREE Full text] [doi: 10.1038/s41371-022-00654-4] [Medline: 35140353]
41. Kaufman N, Mel E. Using digital health technology to prevent and treat diabetes. Diabetes Technol Ther 2020 Feb;22(S1):S63-S78. [doi: 10.1089/dia.2020.2505] [Medline: 32069146]
42. Chrysohoou C, Tsamadias V, Kariori M, Baroutidou A, Gerovassilis G, Akalestos A, et al. Evaluation of the usability of the digital platform navigator KardioUp for the journey of patients with chronic heart failure. Hellenic J Cardiol 2024;75:32-40 [FREE Full text] [doi: 10.1016/j.hjc.2023.06.001] [Medline: 37295668]
43. Ventres W, Kooienga S, Vuckovic N, Marlin R, Nygren P, Stewart V. Physicians, patients, and the electronic health record: an ethnographic analysis. Ann Fam Med 2006;4(2):124-131 [FREE Full text] [doi: 10.1370/afm.425] [Medline: 16569715]
44. Frankel R, Altschuler A, George S, Kinsman J, Jimison H, Robertson NR, et al. Effects of exam-room computing on clinician-patient communication: a longitudinal qualitative study. J Gen Intern Med 2005 Aug;20(8):677-682 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.0163.x] [Medline: 16050873]
45. Montague E, Asan O. Dynamic modeling of patient and physician eye gaze to understand the effects of electronic health records on doctor-patient communication and attention. Int J Med Inform 2014 Mar;83(3):225-234 [FREE Full text] [doi: 10.1016/j.ijmedinf.2013.11.003] [Medline: 24380671]
46. Margalit RS, Roter D, Dunevant MA, Larson S, Reis S. Electronic medical record use and physician-patient communication: an observational study of Israeli primary care encounters. Patient Educ Couns 2006 Apr;61(1):134-141. [doi: 10.1016/j.pec.2005.03.004] [Medline: 16533682]
47. Salisbury H. Helen Salisbury: is transactional care enough? BMJ 2020 Jan 28;368:m226. [doi: 10.1136/bmj.m226] [Medline: 31992553]
48. Zainal H, Quah JH, Smith HE. Experiences of doctors working in Singapore's public primary healthcare clinics. Health Soc Care Community 2022 Sep;30(5):e1948-e1958. [doi: 10.1111/hsc.13626] [Medline: 34725885]
49. Tan KH, Lim BL, Foo Z, Tang JY, Sim M, Lee PT, et al. Prevalence of burnout among healthcare professionals in Singapore. Ann Acad Med Singap 2022 Jul;51(7):409-416 [FREE Full text] [doi: 10.47102/annals-acadmedsg.2021338] [Medline: 35906940]
50. Sarpourian F, Ahmadi Marzaleh M, Fatemi Aghda SA, Zare Z. Application of telemedicine in the ambulance for stroke patients: a systematic review. Prehosp Disaster Med 2023 Oct 25;38(6):774-779. [doi: 10.1017/s1049023x23006519]
51. Chari DA, Wu MJ, Crowson MG, Kozin ED, Rauch SD. Telemedicine algorithm for the management of dizzy patients. Otolaryngol Head Neck Surg 2020 Nov;163(5):857-859. [doi: 10.1177/0194599820935859] [Medline: 32600170]
52. Shih J, Portnoy J. Tips for seeing patients via telemedicine. Curr Allergy Asthma Rep 2018 Aug 15;18(10):50. [doi: 10.1007/s11882-018-0807-5] [Medline: 30112587]
53. Khan M. Funding life-extension research. Cold Spring Harb Perspect Med 2023 Apr 10;13(12):a041208. [doi: 10.1101/cshperspect.a041208] [Medline: 37037599]
54. Foadi N, Koop C, Mikuteit M, Paulmann V, Steffens S, Behrends M. Defining learning outcomes as a prerequisite of implementing a longitudinal and transdisciplinary curriculum with regard to digital competencies at Hannover Medical School. J Med Educ Curric Dev 2021 Jul 21;8:23821205211028347 [FREE Full text] [doi: 10.1177/23821205211028347] [Medline: 34368455]
55. Valikodath NG, Cole E, Ting DS, Campbell JP, Pasquale LR, Chiang MF, et al. Impact of artificial intelligence on medical education in ophthalmology. Transl Vis Sci Technol 2021 Jun 01;10(7):14 [FREE Full text] [doi: 10.1167/tvst.10.7.14] [Medline: 34125146]
56. Puri P, Jha S. Artificial intelligence, automation, and medical education: lessons from economic history. J Am Coll Radiol 2021 Sep;18(9):1345-1347. [doi: 10.1016/j.jacr.2021.05.002] [Medline: 34089666]
57. Loos JR, Davidson EJ. Wearable health monitors and physician-patient communication: the physician's perspective. In: Proceedings of the 49th Hawaii International Conference on System Sciences. 2016 Presented at: HICSS 2016; January 5-8, 2016; Koloa, HI. [doi: 10.1109/hicss.2016.422]
58. Hilty DM, Armstrong CM, Edwards-Stewart A, Gentry MT, Luxton DD, Krupinski EA. Sensor, wearable, and remote patient monitoring competencies for clinical care and training: scoping review. J Technol Behav Sci 2021;6(2):252-277 [FREE Full text] [doi: 10.1007/s41347-020-00190-3] [Medline: 33501372]
59. Meskó B, Drobni Z, Bényei É, Gergely B, Győrffy Z. Digital health is a cultural transformation of traditional healthcare. Mhealth 2017 Sep 14;3:38 [FREE Full text] [doi: 10.21037/mhealth.2017.08.07] [Medline: 29184890]
60. Loane M, Wootton R. A review of guidelines and standards for telemedicine. J Telemed Telecare 2002;8(2):63-71. [doi: 10.1258/1357633021937479] [Medline: 11972941]
61. Sharma P, Sethi MI, Liem A, Bhatti HB, Pandey V, Nair A. A review of telemedicine guidelines in the South-East Asia region. Telemed Rep 2023 Sep 25;4(1):271-278 [FREE Full text] [doi: 10.1089/tmr.2023.0040] [Medline: 37753247]
62. National telemedicine guidelines. Ministry of Health Singapore. 2015 Jan. URL: https://www.moh.gov.sg/docs/librariesprovider5/resources-statistics/guidelines/moh-cir-06_2015_30jan15_telemedicine-guidelines-rev.pdf [accessed 2024-01-15]

63. Paheenthararajah K, Bodulovic G, de Morpurgo M. Telehealth around the world: a global guide. DLA Piper. 2020 Nov 19. URL: https://www.dlapiper.com/en/insights/publications/2020/11/telehealth-around-the-world-global-guide [accessed 2024-08-05]
64. Morrow E, Zidaru T, Ross F, Mason C, Patel KD, Ream M, et al. Artificial intelligence technologies and compassion in healthcare: a systematic scoping review. Front Psychol 2023 Jan 17;13:971044 [FREE Full text] [doi: 10.3389/fpsyg.2022.971044] [Medline: 36733854]
65. Digital Medicine Society. Health education leaders launch continuing medical education for clinicians in digital medicine. PR Newswire. 2022 Apr 20. URL: https://www.prnewswire.com/news-releases/health-education-leaders-launch-continuing-medical-education-for-clinicians-in-digital-medicine-301528403.html [accessed 2024-08-05]
66. Nazeha N, Pavagadhi D, Kyaw BM, Car J, Jimenez G, Tudor Car LT. A digitally competent health workforce: scoping review of educational frameworks. J Med Internet Res 2020 Nov 05;22(11):e22706 [FREE Full text] [doi: 10.2196/22706] [Medline: 33151152]

Abbreviations

ACGME: Accreditation Council of Graduate Medical Education
AI: artificial intelligence
DHT: digital health care technology
EHR: electronic health record

Edited by A Kushniruk; submitted 18.08.23; peer-reviewed by J Aulenkamp, T Aslanidis, Y Chu; comments to author 13.12.23; revised version received 31.01.24; accepted 05.05.24; published 27.08.24.

Please cite as:
Zainal H, Hui XX, Thumboo J, Fong W, Yong FK
Patients' Expectations of Doctors' Clinical Competencies in the Digital Health Care Era: Qualitative Semistructured Interview Study Among Patients
JMIR Hum Factors 2024;11:e51972
URL: https://humanfactors.jmir.org/2024/1/e51972
doi:10.2196/51972
PMID:

Original Paper

Evaluating Factors Affecting Knowledge Sharing Among Health Care Professionals in the Medical Imaging Departments of 2 Cancer Centers: Concurrent Mixed Methods Study

Maryam Almashmoum[1,2], BA, MSc; James Cunningham[1], MSc, PhD; John Ainsworth[1], MSc, PhD

[1]Division of Informatics Imaging and Data Sciences, School of Health Sciences, Faculty of Biology, Medicine, and Health, The University of Manchester, Manchester, United Kingdom

[2]Nuclear Medicine Department, Faisal Sultan Bin Eissa, Kuwait Cancer Control Center, Kuwait City, Kuwait

Corresponding Author:
Maryam Almashmoum, BA, MSc
Division of Informatics Imaging and Data Sciences, School of Health Sciences
Faculty of Biology, Medicine, and Health
The University of Manchester
First Floor, Core Technology Facility
46 Grafton Street
Manchester, M13 9NT
United Kingdom
Phone: 44 07949062010
Email: maryam.almashmoum@postgrad.manchester.ac.uk

Abstract

Background: Knowledge sharing is a crucial part of any knowledge management implementation. It refers to sharing skills and experience among team members in an organization. In a health care setting, sharing knowledge, whether tacit or explicit, is important and can lead to better health care services. In medical imaging departments, knowledge sharing can be of particular importance. There are several factors that affect knowledge-sharing practices in medical imaging departments: individual, departmental, and technological. Evaluating the importance of these factors and understanding their use can help with improving knowledge-sharing practices in medical imaging departments.

Objective: We aimed to assess the level of motivation, identify current knowledge-sharing tools, and evaluate factors affecting knowledge sharing in the medical imaging departments of 2 cancer centers, The Christie, United Kingdom, and the Kuwait Cancer Control Center (KCCC).

Methods: A concurrent mixed methods study was conducted through nonprobability sampling techniques between February 1, 2023, and July 30, 2023. Semistructured interviews were used to validate the results of the quantitative analysis. Data were collected using an electronic questionnaire that was distributed among health care professionals in both cancer centers using Qualtrics. Semistructured interviews were conducted online using Microsoft Teams. The quantitative data were analyzed using the Qualtrics MX software to report the results for each question, whereas the qualitative data were analyzed using a thematic approach with codes classified through NVivo.

Results: In total, 56 respondents from the KCCC and 29 from The Christie participated, with a 100% response rate (56/56, 100% and 29/29, 100%, respectively) based on the Qualtrics survey tool. A total of 59% (17/29) of health care professionals from The Christie shared their knowledge using emails and face-to-face communication as their main tools on a daily basis, and 57% (32/56) of health care professionals from the KCCC used face-to-face communication for knowledge sharing. The mean Likert-scale score of all the components that assessed the factors that affected knowledge-sharing behaviors fell between “somewhat agree” and “strongly agree” in both centers, excepting extrinsic motivation, which was rated as “neither agree nor disagree.” This was similar to the results related to incentives. It was shown that 52% (15/29) of health care professionals at The Christie had no incentives to encourage knowledge-sharing practices. Therefore, establishing clear policies to manage incentives is important to increase knowledge-sharing practices.

Conclusions: This study offered an evaluation of factors that affect knowledge sharing in 2 cancer centers. Most health care professionals were aware of the importance of knowledge-sharing practices in enhancing health care services. Several challenges were identified, such as time constraints, a lack of staff, and the language barrier, which limit knowledge-sharing practices.

Therefore, establishing a clear policy for knowledge sharing is vital to practicing knowledge-sharing behaviors and facing any challenges that limit this practice.

(JMIR Hum Factors 2024;11:e53780) doi:10.2196/53780

KEYWORDS

knowledge management; knowledge sharing; medical imaging departments; cancer centers; The Christie; Kuwait Cancer Control Center; concurrent mixed methods; factors; challenges; definition; mechanisms; practices

Introduction

Background

Knowledge consists of a combination of facts, ideas, experiences, and information that is gained through experience and practice [1]. Knowledge management is the organizational capability to identify, transfer, convert, and share knowledge to attain institutional success. In health care, knowledge is an important asset in following the best medical practices. In recent years, health care intuitions have had a clear mission to establish a strong knowledge management system to use their knowledge in a good way by sharing health care professionals' knowledge with others [2]. Knowledge management is directly related to good performance [3]. The primary aim of knowledge management in a health care setting is to create a culture of knowledge sharing among health care professionals, allowing them to carry out hospital tasks in an efficient way, leading to an increase in successful patient outcomes and a reduction in repetitive medical errors [4,5]. Knowledge management consists of 4 processing steps: knowledge creation, knowledge capture, knowledge sharing, and knowledge application [6]. According to a 2016 report by the World Health Organization, there are many challenges that health care institutions face, such as long wait times for patient services, a lack of health care professionals, and inadequate information and communications technology (ICT) infrastructures in health care centers [1,7]. Good knowledge management practice starts with understanding each processing step and trying to identify the challenges and solutions for each of them [5,6].

Knowledge sharing in health care institutions has a positive impact on institutional performance [8]. Effective knowledge sharing has several benefits related to increasing successful patient outcomes, such as innovation, critical thinking, problem-solving, reducing medical errors, avoiding repetitive medical errors, increasing performance, and gaining competitive advantages [6,9-11]. Furthermore, improving knowledge-sharing practices among health care professionals allows them to learn and use resources efficiently [7]. Health care institutions have complex organizational structures [5]. They employ a variety of multidisciplinary professionals who communicate and share knowledge among each other as a part of day-to-day medical practice. Effective knowledge-sharing practices among health care professionals contribute to a positive overall knowledge-sharing culture [5]. Ipe [12] defined knowledge sharing as "the act of making knowledge available to others within the organizations." Knowledge sharing is the way in which health care professionals share their tacit and explicit knowledge [13]. Tacit knowledge is defined as any thoughts and ideas that exist in the human mind [14]. It is difficult to capture and can be shared through interaction with others [15]. Explicit knowledge, in contrast, exits in the documents, policies, and manuals that departments have, and it is easy to capture [16].

Medical imaging departments consist of several divisions: radiology, nuclear medicine, radiotherapy, and physical radiation. The names of the divisions and their number differ from one hospital to another. Medical imaging departments can be considered the backbone of a hospital due to the importance of the tasks that are performed there by specialized health care professionals [17,18]. The number of health care professionals working in medical imaging departments has increased in recent years due to the expansion of the duties carried out in these departments [19].

Factors Affecting Knowledge Sharing

There are several factors that affect knowledge-sharing practices [20-31]. These factors can differ from one institution to another depending on the nature of each environment [20-31]. Technological factors often differ between institutions due to the use of specialized technologies (eg, the use of a picture archiving and communication system [PACS] in medical imaging departments), whereas individual factors are often the same across institutions as they relate to human nature [32-34]. Departmental factors often have commonalities across institutions related to their dependence on the use of resources for enhancing knowledge sharing among workers [35,36]. Evaluating knowledge-sharing behaviors among employees in medical imaging departments at cancer centers depends on identifying the level of awareness of the importance of knowledge sharing and the factors that affect knowledge-sharing practices [37]. These factors are divided into facilitators and barriers. On the basis of a previous study by Almashmoum et al [38], we can divide the facilitators of knowledge sharing in medical imaging departments in cancer centers into 3 categories—individual, departmental, and technological facilitators—and the barriers that affect knowledge sharing into 4 categories—individual, departmental, technological, and geographical. Table 1 shows these factors and their components [38]. On the basis of these factors, we sought to identify awareness of knowledge-sharing behaviors, evaluate the factors affecting knowledge sharing in the medical imaging departments of 2 cancer centers (The Christie and the Kuwait Cancer Control Center [KCCC]), and compare results across these 2 centers.

Table 1. Factors that affect knowledge-sharing practices in medical imaging departments [38].

	Factors
Facilitators	
Individual facilitators	• Trust • Positive attitudes • Awareness • Experience • Intrinsic motivation • Personality • Self-esteem • Self-efficacy
Departmental facilitators	• Multidisciplinary team and community of oncologists • Leadership • Culture • Teamwork • Extrinsic motivation • Learning and training • Physician rounds • Departmental arrangements
Technological facilitators	• Information and communications technology (picture archiving and communication system, social media, intranet, extranet, telemedicine, and teleradiology) • Network • Digital library
Barriers	
Financial barriers	• Cost
Administrative barriers	• Language • Time • Shortage of staff • Lack of transparency • Lack of experience
Technological barriers	• Low-speed network • Upgrade system • Lack of equipment
Geographical barriers	• Geographical distance

An Overview of The Christie and KCCC

Cancer centers are tertiary care units performing diagnostic scans and therapeutic treatments, blood testing, histology, stem cell laboratory tests, and palliative care related to cancer. The Christie is the largest single cancer center in Europe. It is located in Manchester, United Kingdom; serves 3.2 million people across the Greater Manchester region; and provides treatment for >60,000 patients with cancer per year [39].

The KCCC was established in 1968 and is one of the largest centers in Kuwait that provides comprehensive care for patients with cancer. Its main mission is to focus on improving physician and nursing education. More than 2000 patients with cancer are treated annually at the KCCC [40].

Objectives

This study had several objectives. The purpose of this study was to evaluate knowledge-sharing practices among health care professionals at The Christie and KCCC, as well as identifying the current knowledge-sharing mechanisms and the facilitators and barriers that affect knowledge sharing among health care professionals in the medical imaging departments of the aforementioned 2 cancer centers. From this, we aimed to construct a new definition of knowledge sharing as it relates to medical imaging departments. Finally, we aimed to draw conclusions on how to improve knowledge-sharing practices and their effects on the quality of patient services.

Research Questions

There are several challenges in communication among health care professionals, such as lack of awareness of the importance of knowledge-sharing behaviors and factors that affect knowledge-sharing practices. This affects knowledge-sharing behaviors and, therefore, knowledge management implementation, which is considered important in any health care institution to perform tasks for patients. There are several questions raised based on that observation:

1. What is the level of motivation for knowledge sharing among employees in health care institutions?

2. What current knowledge-sharing tools exist in health care institutions?
3. What are the challenges faced by health care professionals related to knowledge sharing?
4. How can knowledge-sharing practices in medical imaging departments be improved?
5. What are the factors that affect knowledge sharing among health care professionals?
6. What is the definition of knowledge sharing in medical imaging departments?

Methods

Research Design and Sampling Techniques

This study was performed using a concurrent cross-sectional triangulation mixed methods design, which combined online semistructured interviews with health care professionals who worked in medical imaging departments and an electronic survey that was distributed among health care professionals who worked at The Christie and KCCC to evaluate knowledge-sharing practices and identify factors that affect knowledge-sharing behaviors. This study was conducted between February 2023 and July 2023. The sampling techniques that were used for the selection of health care professionals for the questionnaire and semistructured interviews were self-selection sampling and snowball sampling, respectively. A self-selection approach was applied to select participants who indicated a desire to take part in the research [41]. Self-selection sampling was used for the questionnaires and snowball sampling was used for the semistructured interview because this study evaluated knowledge-sharing practices in medical imaging departments. The questionnaire was distributed among professionals using a WhatsApp group (Meta Platforms) at the KCCC and on the internal page of The Christie. Snowball sampling focused on a group of people who had the same specialties to participate in the interviews. There were several specialties in this department, such as physicians, technologists, senior managers, the head of the department, nurses, and radiotherapists.

Ethical Considerations

This study followed the University of Manchester's ethical guidelines. The ethics committee of The Christie determined that the study did not require ethical approval based on the official decision tool of the University of Manchester because the study was conducted with professionals and did not require sensitive questions, vulnerable groups, or risk of disclosures of anonymized information. Whereas the KCCC did require ethical approval especially for them, which it provided (3797). All respondents were health care professionals who signed a consent form to participate in the questionnaires and audio-recorded semistructured interviews. The consent form explained the purpose of the study as well as any other information that the participants might require. In addition, the health care professionals' personal information was kept anonymous and confidential.

Quantitative Methods and Data Analysis

The purpose of this questionnaire was to evaluate the level of awareness of the importance of knowledge-sharing practices in medical imaging departments and the factors that affect knowledge-sharing behaviors. The questions were derived from previous related studies on knowledge sharing and modified to fit the overall research aim and answer the research questions [42-44]. The questionnaire items were written in English. They consisted of both nominal and ordinal scales. The entire questionnaire that was used in this study can be found in Multimedia Appendix 1. It was divided into 6 sections, as shown in Textbox 1. It consisted of 66 questions, with an additional open-ended question at the end. A pretest study was conducted among 10 academic lecturers and PhD students at the University of Manchester. The purpose of the pretest study was to make sure that the questionnaire items could be understood in health care institutions. After that, the electronic questionnaires were distributed among health care professionals at the KCCC using WhatsApp and among health care professionals working at The Christie through a posting on the hospital's intranet. After the collection of the quantitative data, an analysis was performed using the Qualtrics XM software package (Qualtrics International Inc).

Textbox 1. The sections of the questionnaire.

- Section 1: an overview of knowledge sharing
- Section 2: the consent form (5 statements)
- Section 3: demographic profile of the health care professionals (7 multiple-choice questions)
- Section 4: questions about knowledge-sharing practices and background (4 multiple-choice questions)
- Section 5: questions that examined knowledge-sharing factors (55 questions on a 7-point Likert scale from "strongly disagree" to "strongly agree")
- Section 6: 1 open-ended question

Qualitative Methods and Data Analysis

The semistructured interviews were conducted on the web using Microsoft Teams at the same time as the distribution of the questionnaires. The interviews started with brief introductory remarks about knowledge sharing followed by questions that related to the health professionals' experiences and definitions of knowledge sharing. After that, the questions were designed to assess factors and practices related to knowledge sharing. Multimedia Appendix 1 presents the consent form and the interview questions. The semistructured interviews were conducted with health care professionals who were working in the medical imaging departments of both cancer centers. Each interview took approximately 25 to 45 minutes. Invitations were sent electronically via WhatsApp for KCCC participants and posted on the internal page of The Christie. All the participants signed a consent form related to the interviews and audio recording. Thematic analysis was used to analyze the semistructured interview transcripts. The coding and creation of themes was conducted using the NVivo software (Lumivero).

Results

Quantitative Analysis

Demographic Characteristics and Response Rate

A total of 77 health care professionals from The Christie participated in this study with a 100% (77/77) response rate based on the statistical analysis using the Qualtrics XM software. In total, 38% (29/77) of the respondents answered all the survey questions. In addition, all of them worked in the medical imaging department. The response rate from the KCCC was 100% (145/145), of whom 48% (70/145) from all departments completed all survey questions. A total of 80% (56/70) of the health care professionals who participated were from the medical imaging department at the KCCC. The demographic characteristics for both centers are shown in Table 2.

Table 2. Demographic characteristics.

	KCCC[a] (n=56), n (%)	The Christie (n=29), n (%)
Sex		
Male	26 (46)	7 (24)
Female	29 (52)	20 (69)
Prefer not to say	0 (0)	2 (7)
Age group (y)		
<20	0 (0)	1 (3)
20-30	3 (5)	8 (28)
30-40	19 (34)	7 (24)
40-50	25 (45)	8 (28)
50-60	7 (12)	3 (10)
>60	2 (4)	2 (7)
Educational level		
Diploma	7 (12)	4 (14)
First degree (bachelor's)	26 (46)	12 (41)
Master's degree	11 (20)	8 (28)
Doctorate	9 (16)	0 (0)
Other	3 (5)	5 (17)
Work experience (y)		
<10	18 (32)	22 (76)
10-20	29 (52)	6 (21)
20-30	7 (12)	1 (3)
>30	2 (4)	0 (0)

[a]KCCC: Kuwait Cancer Control Center.

Knowledge-Sharing Practices

Knowledge sharing is defined as sharing ideas, thoughts, and experiences among health care professionals to create new knowledge. Sharing knowledge requires several facilitators to accelerate knowledge-sharing behaviors, for example, in morning meeting sessions, multidisciplinary team meetings, and conferences. These practices can improve patient outcomes and minimize medical errors. In addition, they can help make such shared knowledge reusable for all health care professionals. The results at The Christie revealed that 59% (17/29) of health care professionals participated in knowledge-sharing activities available in their department on a daily basis and only 14% (4/29) of health care professionals did not participate in any of those activities. At the KCCC, the results showed that 57% (32/56) of health care professionals participated on a daily basis in the knowledge-sharing activities that were available in their department. On the other hand, only 4% (2/56) of health care professionals never participated in those activities.

The Mechanisms of Knowledge-Sharing Practices

Sharing knowledge among health care professionals requires different mechanisms. Those mechanisms were classified into either physical or online tools, for example, face-to-face communication, phone calls, social media, emails, and Microsoft Teams. The results showed that 83% (24/29) of the health care professionals at The Christie used email and face-to-face communication to share their knowledge, whereas 86% (48/56) of the health care professionals at the KCCC used face-to-face communication as the main tool to share their knowledge. In total, 59% (17/29) of the health care professionals at The Christie used Microsoft Teams as a tool to share knowledge. At the KCCC, the results showed that 48% (27/56) of the health care professionals also used phone calls and social media to share their knowledge.

Level of Motivation

We examined the motivational level by exploring the level of willingness to share knowledge among health care professionals. Most of the health care professionals in both cancer centers had a high motivational level to practice knowledge-sharing behaviors with their peers at the workplace. The findings showed that 38% (21/56) of the health care professionals at the KCCC were highly motivated compared with 48% (14/29) of the health care professionals at The Christie, as shown in Table 3. This percentage indicated that the level of motivation in both centers among health care professionals was high. Regarding incentives and policies, Table 4 shows that 79% (44/56) of the health care professionals who worked at the KCCC indicated that there were incentives to encourage knowledge-sharing practices, whereas half (15/29, 52%) of the health care professionals who worked at The Christie indicated that there were no incentives to encourage knowledge-sharing practices. This comparison showed that the health care professionals at The Christie had a high motivational level but their department did not have incentives and a clear policy to encourage knowledge-sharing practices, which could affect their engagement in these practices.

Table 3. Motivational level of the health care professionals at The Christie and Kuwait Cancer Control Center (KCCC).

	Very low, n (%)	Low, n (%)	Medium, n (%)	High, n (%)	Very high, n (%)
The Christie (n=29)	2 (7)	0 (0)	6 (21)	14 (48)	7 (24)
KCCC (n=56)	0 (0)	5 (9)	14 (25)	21 (38)	16 (29)

Table 4. Comparison of incentives or polices in place to encourage knowledge sharing between The Christie and the Kuwait Cancer Control Center (KCCC).

	Yes, n (%)	No, n (%)
The Christie (n=29)	15 (52)	14 (48)
KCCC (n=56)	44 (79)	12 (21)

Factors Affecting Knowledge-Sharing Practices

The questionnaire evaluated the factors that affect knowledge sharing in medical imaging departments. On the basis of prior work, we identified 19 factors that affect knowledge-sharing behaviors among health care professionals working in medical imaging departments [38]. These factors are divided into 3 categories: individual, departmental, and technological. Individual factors comprise 8 components (trust, positive attitudes, awareness, experience, personality, intrinsic motivation, self-esteem, and self-efficacy). Departmental factors comprise 8 components (community of practice, leadership, culture, teamwork, extrinsic motivation, learning and training, physician rounds, and departmental arrangements). Technological factors comprise 3 components (ICT, network, and digital technology). These factors were measured on a 7-point Likert scale using an equivalent interval of 6/7=0.86. Multimedia Appendix 2 shows the mean score for each component. The mean score was classified as follows: “strongly disagree” for values within the range of 1.00 to 1.86, “disagree” for values within the range of 1.86 to 2.72, “somewhat disagree” for values within the range of 2.72 to 3.58, “neither agree or disagree” for values within the range of 3.58 to 4.44, “somewhat agree” for values within the range of 4.44 to 5.3, “agree” for values within the range of 5.30 to 6.16, and “strongly agree” for values within the range of 6.16 to 7 [28]. All the Likert scale results can be found in Multimedia Appendices 3 and 4 for The Christie and KCCC, respectively.

At the KCCC, the values for all the components of the factors that affect knowledge-sharing behaviors fell between “agree” and “strongly agree.” At The Christie, the values fell between “somewhat agree” and “strongly agree.”

Individual factors are important for enhancing knowledge-sharing practices among health care professionals. Trust plays a vital role in knowledge-sharing practices. It creates a strong relationship among health care professionals. In medical imaging departments, trust between the senior manager and health care professionals and among health care professionals is vital to share knowledge smoothly and provide high-quality health care services. The mean scores for trust were 5.7 and 5.51 in both centers, which corresponds to “agree.” Thus, more than half of the health care professionals had a level of trust that could improve knowledge-sharing behavior by building trust relationships among each other. Awareness of the importance of knowledge sharing helps increase knowledge-sharing behaviors in daily work. In addition, knowledge sharing helps health care professionals gain new knowledge, fosters learning, and prevents repetitive mistakes. The results showed that the mean score for awareness was 6.42 at the KCCC, which corresponds to "strongly agree," versus 5.48 at The Christie, which corresponds to "agree." Therefore, health care professionals in both centers had a clear awareness of the importance of knowledge sharing in improving their skills and health services. Health care professionals in both cancer centers believed that positive attitudes help enhance knowledge-sharing behaviors. The results showed that the mean scores for positive attitudes in both cancer centers were 6.4 and 6.16, which correspond to "strongly agree." Health care professionals having good experience increases knowledge-sharing behaviors. The mean scores for experience were 6.2 at the KCCC and 5.77 at The Christie, which correspond to "strongly agree" and "agree," respectively. Therefore, health care professionals in both centers believed in the importance of experience in enhancing knowledge sharing. In addition, they had enough experience that could help them share it with others. Health care professionals in both centers had an extroverted personality, which opens to others and allows them to share knowledge with others. As shown in Multimedia Appendix 2, the mean scores for personality were 6.15 for KCCC and 6.18 for The Christie, which correspond to "agree" at the KCCC and "strongly agree" at The Christie. Hence, the results showed that most of them have self-confidence and felt open to new ideas when they practiced knowledge sharing. Self-esteem and self-efficacy are the main components of the individual factors. The mean scores for these components in both cancer centers were close to "agree," which means that health care professionals have self-efficacy and self-esteem regarding their ability to successfully share knowledge with their peers. Finally, regarding intrinsic motivation, the mean scores were 5.96 at the KCCC and 5.89 at The Christie, which means that most of the health care professionals had intrinsic motivation that allowed them to share their knowledge.

Health care institutions are considered knowledge-based environments due to the large amount of knowledge, either tacit or explicit, that needs to be managed. To maximize the benefit of that knowledge, each department in a health care institution has a responsibility to share knowledge among their professionals. Extrinsic motivation is important to enhance knowledge-sharing practices through providing acknowledgment, appreciation, incentives, and bonuses to health care professionals. The results showed that the mean score for extrinsic motivation was 5.66, which corresponds to "agree," at the KCCC. Thus, health care professionals who worked at the KCCC received motivation from their senior managers. In contrast, the mean score was 4.37 at The Christie, which corresponds to "neither agree nor disagree." This implies an absence of departmental encouragement of knowledge-sharing behaviors, as the respondents believed in the importance of extrinsic motivation in enhancing knowledge sharing. Health care professionals in both cancer centers believed that the leadership plays a significant role in providing encouragement, improving knowledge-sharing activities, and minimizing conflict. The mean scores for leadership in both cancer centers were close to "agree," which means that the leadership in both cancer centers had a positive impact in enhancing knowledge sharing. Working and learning as a team is important to enhance knowledge-sharing practices. Health care professionals in both centers believed in the importance of teamwork for increasing knowledge-sharing practices. The mean scores for teamwork in both cancer centers were close to "strongly agree," which means that they worked as a team to increase productivity. Both cancer centers had a culture of communication to enhance knowledge sharing, with mean scores close to "strongly agree." The community of practice consists of several communities that specialize in making decisions in specific cancer centers. In addition, multidisciplinary team meetings are one of the most prominent types of meetings in cancer centers. The mean scores for community of practice in both cancer centers corresponded to "agree," which means that, in both centers, there are several specialized meetings that play important roles in making decisions by enhancing knowledge-sharing behaviors. Learning and training activities, such as workshops, lectures, and conferences, play a vital role in enhancing knowledge sharing. At the KCCC, the mean score for this component was 5.79, which corresponds to "agree." Thus, health care professionals participated widely in several learning and training activities to enhance their skills. In contrast, at The Christie, the mean score for this component was 5.25, which corresponds to "somewhat agree." Thus, health care professionals received limited learning and training support to increase their skills. At the KCCC, there were enough empty rooms and open space to enhance knowledge sharing, with a mean score of 6.1 for departmental arrangements, which corresponds to "agree," compared with not enough space at The Christie, with a mean score of 5.29 for departmental arrangements, which corresponds to "somewhat agree." Daily physician rounds play a vital role in developing health care professionals' skills. The mean score for this component was 5.64 at the KCCC, which corresponds to "agree," and 4.82 at The Christie, which corresponds to "somewhat agree." Therefore, health care professionals at the KCCC believed in the importance of physician rounds in enhancing knowledge sharing more than professionals at The Christie.

Technological factors are considered a dynamo of knowledge-sharing practices. High-speed networks play a significant role in enhancing knowledge-sharing practices. Both cancer centers had a high-speed network and believed in its importance, with mean scores of 5.33 at the KCCC and 5.56 at The Christie, which correspond to "agree." ICT is crucial to support sharing information. Intranet, extranet, PACS, and social media are considered types of ICT. In addition, ICT requires skills to use it and maintenance to address and report any related problems. At the KCCC, the mean score for ICT was 5.62, which corresponds to "agree." Thus, in the medical imaging departments, there were enough ICT infrastructures that they used to share their knowledge, and health care professionals were trained well to use those technologies and report any problems they faced. At The Christie, the mean score for ICT was 5.20, which corresponds to "somewhat agree." Thus, in the medical imaging departments, health care professionals used ICT to share their knowledge, and they had enough skills to use it. However, they did not frequently use social media to share their knowledge compared to health care professionals at the KCCC. To access to updated articles and resources, digital libraries were a valuable tool that allowed health care professionals to gain new information to share their knowledge with each other. The mean scores for this component were 5.91 at the KCCC and 5.62 at The Christie, which correspond to "agree." Thus, health care professionals in both cancer centers believed in the importance of digital libraries in enhancing knowledge sharing.

Qualitative Analysis

Overview of Analysis

Semistructured interviews were used to gather the background experience of health care professionals who worked in the medical imaging department (heads of department, technologists, nurses, physicians, and radiotherapists). The data were used to validate the quantitative data. A total of 13 health care professionals participated in this part of the study. Of the 13 participants, 11 (85%) were from the KCCC, and only 2 (15%) were from The Christie. The outcomes of the online semistructured interviews shed light on 3 themes that related to knowledge-sharing practices, such as definitions of knowledge sharing, factors, and challenges to knowledge-sharing behaviors.

Theme 1: Definition of Knowledge Sharing in Medical Imaging Departments

Despite the fact that the term *knowledge sharing* is not new, most of the health care professionals asked for clarification of the term before giving their definition based on their experience in the department. In general, most of them gave the proper definition of knowledge sharing in the medical imaging department; Multimedia Appendix 5 shows their definitions. On the basis of their views, the general key points to structure the definition of knowledge sharing are as follows:

1. All health care professionals have a specific amount of knowledge derived from their studies and experience.
2. Knowledge sharing involves sharing information and updated protocols and circulars among health care professionals.
3. Knowledge sharing takes places between colleagues or among a wider team in one department or with professionals from another hospital.
4. Knowledge sharing occurs through various activities, such as lectures, workshops, meetings, and conferences.
5. There is no benefit from keeping knowledge to oneself. It remains inactive until it is shared.
6. Knowledge sharing among health care professionals helps patients obtain a more accurate diagnosis.

Theme 2: Factors Affecting Knowledge Sharing

The findings of a deep analysis of the qualitative data were consistently validated by the quantitative data. The factors that affect knowledge sharing could be classified as facilitators that enhance knowledge sharing and as barriers when those factors have a negative impact on knowledge-sharing behavior. Facilitators are classified into 3 categories: individual factors, departmental factors, and technological factors.

Subtheme 2.1: Individual Factors

Participants stated that they were aware of the importance of knowledge-sharing practices in maximizing health services in various specialties. In addition, they were aware of the importance of sharing knowledge with other peers to apply it in various situations. Participant C indicated the following:

> *Knowledge sharing is very important amongst clinicians of various specialities and various field.*

Most health care professionals had a variety of skills and experience that they had gained throughout their careers. Their skills and experience allowed them to share their knowledge with other peers who had less experience. However, some of them felt shy about sharing their knowledge with others in a large group. In addition, sharing knowledge among health care professionals depends on the personality of those doing the sharing. Participant B said the following:

> *And lots of knowledge, but those kinds of people also they do not want to share it in public among larger group of audience or larger group of attendees. Things can be accomplished by dealing with getting the information from a group in a way that they do not feel uncomfortable in sharing their ideas.*

Subtheme 2.2: Departmental Factors

According to the findings of this data analysis, there are several factors related to departmental factors. Knowledge-sharing practices increased with the ability of the department to foster a culture of communication that allowed health care professionals to share their knowledge. Participant C indicated the importance of creating a culture to support sharing knowledge:

> *So definitely there is a good positive environment of knowledge sharing and knowledge building also...sharing information maybe I think it changes the culture of the team if we can inspire people to learn and share more.*

In addition, working as a team helps enhance knowledge-sharing practices because health care professionals work in groups to perform specific tasks and procedures. The leadership plays an important role in enhancing knowledge sharing by establishing clear policies for sharing knowledge and allowing health care professionals to participate in several activities. Those activities fall under the component of learning and training. Most respondents indicated that there were several activities available, such as presenting and attending lectures, attending conferences locally and internationally, participating in training sessions and workshops, creating posters, attending seminars, and engaging in continuing education. In addition, respondent C suggested participating in journaling sessions to keep up to date on information and maximize their knowledge, allowing knowledge-sharing with others to improve treatment plans. Extrinsic motivation is divided into 2 categories: physical and emotional incentives. In this regard, participants B and F indicated that excellent evaluations and other signs of appreciation, such as receiving certificates or having their names added to papers, posters, or lectures prepared by professionals, were an effective means of encouragement. A community of practice involves specialized meetings among health care professionals, including multidisciplinary team meetings. These high-level knowledge-sharing meetings can involve decisions on treatment plans for patients with cancer. All health care professionals in different disciplines play a certain role in interpreting the final treatment plan. Therefore, attending these meetings helps them develop their skills by gaining new knowledge. However, a few of the respondents from KCCC indicated that they did not have any idea about those meetings and they were not involved in them at all.

Subtheme 2.3: Technological Factors

Technological factors are considered a dynamo of knowledge-sharing practices. The results revealed that there were several types of ICT infrastructure in both departments, such as PACS, the bleep system, hospital information system, registration information system, and social media apps, in addition to online communication tools such as Zoom and Microsoft Teams. The use of online tools became prevalent after the COVID-19 pandemic to maintain communication among health care professionals by setting online meetings, videoconferences, and online circular discussions for better patient services and to protect their lives. As a consequence, most of the respondents preferred online tools over face-to-face communication for several reasons, such as availability, time saving, and removing geographical barriers. On the other hand, most of the participants indicated that using online tools requires positive attitudes toward technology, maintenance, and a high-speed network to keep the tools active for sharing knowledge. Participants C and K indicated the following:

> *I would say sometimes because of some Internet issues or some Internet connection, yes. So it becomes difficult. Sometimes it takes time for the reports to be automatically uploaded into HIS.*

Theme 3: The Challenges to Knowledge-Sharing Behaviors

The in-depth analysis of the participants' views brought about some of the challenges that health care professionals face in practicing knowledge sharing. They mentioned that a lack of staff is one of the challenges that managers face, which results in a lot of duties in daily work, influencing knowledge sharing. This prevents them from practicing knowledge-sharing activities. Participant B indicated the following:

> *I think this is actually one of the problem in the cancer Control Center is that you don't have we don't have enough stuff for the number of the service that we are doing so it it's you know one can imagine that three or five physician do a delivery of probably 50 or 60 clinical service a day for the whole 365 days a year and we are expected to give our time to knowledge share as reserving one or two hours per week knowledge sharing when we are expected to finalize the clinical service, this is the problem.*

In addition, time constraints were a main challenge to knowledge sharing that participant C mentioned:

> *We wish we had more time for knowledge sharing because definitely it is very useful, but maybe we're not able to do it as often or as for a longer period of time because of the time constraint. Yeah. Our challenge I would say one is the time that we would want, definitely we don't have. I feel that we don't have enough time to sort of have more of discussing of cases. Interacting with more with each other because many times when I have faced that that I want an opinion but there's just no time for me or for the other person to actually look at the case and go into details and try to get some information and then. So, we tend to sort of maybe cut down on the sessions and use that time to report our cases. So, I feel that we should have more dedicated time. can make sure that the message delivered to him that he will ask his colleague or there is a minute. Of course, we can share it with him.*

In addition to these challenges, there were challenges related to a person's attitudes. For instance, some were less interested in sharing their knowledge because of a lack of awareness of the importance of knowledge sharing. Some lacked awareness regarding who was responsible for sharing knowledge and were not aware of the benefits of sharing their knowledge. In addition, according to participant L, "It's a cooperation. I said yes, this is one challenge and then also communication with the patient."

Discussion

Principal Findings

Health care institutions are knowledge-based environments where knowledge-sharing practices are an important step to achieve good knowledge management. This can help these departments reach their intended goals, mission, vision, and objectives for the delivery of high-quality health services. The objectives of this study were to evaluate the factors that affect knowledge sharing in the medical imaging departments of 2 cancer centers, identify the current tools for knowledge-sharing practices, structure a new definition of knowledge sharing, assess the challenges, and identify the areas for improvement in knowledge-sharing behaviors. On the basis of the respondents' views and thoughts on knowledge sharing during the semistructured interviews, we can define knowledge sharing as follows: "All health care professionals have a certain amount of knowledge built up through their work in a specific field. That knowledge (either tacit or explicit) will remain inactive until they share it with their peers or the wider team through meetings, lectures, and workshops and confirm that knowledge, or recreate new knowledge to better help patients."

The results showed that the level of motivation of the health care professionals to share their knowledge in their daily work was high in both cancer centers. There are several studies that suggest the same results regarding the high level of motivation toward knowledge sharing [20,37]. There were several mechanisms for sharing knowledge among health care professionals. The findings illustrated that the current mechanisms of knowledge sharing at The Christie were face-to-face communication and email, each with an equal percentage of participants (24/29, 83%), compared with the use of face-to-face communication as a main tool to share knowledge at the KCCC (48/56, 86%). This agrees with research showing that radiologists prefer face-to-face communication to share their thoughts and interests [45]. In addition, using social media as a tool for sharing knowledge among health care professionals was a common practice at the KCCC, for instance, using WhatsApp as a main tool for internal circulation of information, announcements, and updates regarding protocol. Most respondents expressed that using social media as a online tool is faster than using traditional tools. In terms of availability, the information will remain in an app that can be accessed at any time because most health care professionals are busy with cases and do not have enough time to attend meetings face-to-face. Informing other people about key points of the meetings using those apps helps enhance knowledge-sharing practices among health care professionals. However, the use of social media apps in health care institutions still appears to be more prevalent in the West compared to the East [37,46-48]. During the COVID-19 pandemic, the use of online tools for sharing knowledge came to play a vital role in keeping knowledge circulating among health care professionals and making communication with others safe [49]. There is evidence to suggest that these practices led to an increase in positive patient outcomes and health services during the crisis [49,50]. Using hybrid tools contributes to the ability to share knowledge in a way that suits health care professionals [50].

The health care professionals in this study responded “strongly agree” or “somewhat agree” to questions on the importance of the examined knowledge-sharing practices. Therefore, this study found that health care professionals in both cancer centers believed in the importance of factors that affect knowledge sharing in enhancing knowledge-sharing practices in their daily work. Several studies support this finding [20-31]. However, the mean score for extrinsic motivation at The Christie was 4.37, which corresponds to “neither agree or disagree,” suggesting that the department may not have given enough encouragement for knowledge-sharing practices. This is reflected in the answers to the questions on encouragement and incentives at The Christie. More than half (15/29, 52%) of the respondents at The Christie indicated that there were no incentives to encourage knowledge-sharing practices. Therefore, health care professionals might intend to avoid participating in knowledge-sharing practices, which has a negative impact on practicing knowledge sharing in general.

In addition to those factors, there were several challenges that affected knowledge-sharing practices addressed by respondents in the semistructured interviews at both cancer centers, such as time constraints and attitudes toward knowledge sharing. In addition, at the KCCC, the respondents addressed language as the main challenge in sharing knowledge because the environment consists of international workers, which means that they communicate with their colleagues using their second language. Previous studies have shown that language is the main barrier that limits knowledge-sharing practices [33,34,37,51]. Fatahi et al [52] have illustrated that language is the first route for communication. Therefore, using a universal language that allows all health care professionals to communicate with their peers is important to enhance knowledge-sharing practices.

To improve knowledge-sharing practices, senior managers suggested in the semistructured interviews that creating a clear policy to share knowledge in the department is important in enhancing knowledge-sharing practices, starting with increasing awareness of the importance of knowledge sharing, followed by encouraging participation in various learning and training activities; improving attitudes to using technology; and, finally, providing encouragement via physical incentives (eg, bonuses and promotions) and emotional incentives (eg, excellent evaluations, certificates of appreciation, or adding an individual’s name to a research paper or lecture). Moreover, hiring more staff helps enhance knowledge sharing by reducing the workload and giving staff members time to share their knowledge and participate in several activities.

Limitations

There were several limitations to this study that need to be addressed. The sample from The Christie was small in both parts of the study compared to the sample from the KCCC. There was a lack of staff, which limited their participation in the questionnaires and interviews due to work pressures. This study focused only on medical imaging departments and was limited to only those who worked in those departments. Therefore, this study could not evaluate the level of motivation and factors that affect knowledge sharing in all departments in both centers. Future work needs to assess the level of knowledge-sharing practices. A better approach may be to create a maturity model for knowledge sharing to assess the level of maturity and help managers put a clear plan and policy in place to manage knowledge-sharing practices.

Conclusions

This concurrent mixed methods study provides an evaluation of the factors that affect knowledge sharing in medical imaging departments. In addition, it structured a definition of knowledge sharing in medical imaging departments. On the basis of the questionnaires, health care professionals used face-to-face communication as a main mechanism to share knowledge within these departments at The Christie and KCCC (24/29, 83% and 48/56, 86%, respectively). Therefore, using knowledge-sharing mechanisms within departments for the purpose of enhancing knowledge-sharing practices is vital to speed up the knowledge-sharing process. In addition, health care professionals in the medical imaging departments in both centers had a good personality, positive attitudes, trust, a high self-esteem, and self-efficacy. Therefore, the intention to share knowledge was high in both cancer centers based on the individual factors. In the medical imaging departments, there were health care professionals that led the process of knowledge sharing by creating a culture of communication, setting several meetings, and giving everyone an equal opportunity to participate in learning and training activities. However, regarding extrinsic motivation, half (15/29, 52%) of the respondents at The Christie indicated that there was a lack of incentives, which was reflected in the low mean score for this component. Therefore, establishing a clear plan and providing incentives has a positive impact on knowledge-sharing practices. ICT infrastructures were available in both cancer centers, with a high-speed network to run those technologies. Most of the respondents in the semistructured interviews addressed several challenges of knowledge sharing, such as language barriers and a lack of time, staff, and a clear plan for knowledge sharing. Finally, this study provides managers with an evaluation of the factors that affect knowledge sharing in both cancer centers and allows them to address the challenges and improve them.

Acknowledgments

This study was completed as part of doctoral studies funded by the Ministry of Health, Kuwait, and the University of Manchester, United Kingdom.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The questionnaires items, consent form for the interview, and the questions of the interviews.
[DOCX File , 49 KB - humanfactors_v11i1e53780_app1.docx]

Multimedia Appendix 2
Results for the factors that affect knowledge sharing in both cancer centers.
[DOCX File , 29 KB - humanfactors_v11i1e53780_app2.docx]

Multimedia Appendix 3
Evaluation of factors at The Christie.
[DOCX File , 74 KB - humanfactors_v11i1e53780_app3.docx]

Multimedia Appendix 4
Evaluation of factors at the Kuwait Cancer Control Center.
[DOCX File , 71 KB - humanfactors_v11i1e53780_app4.docx]

Multimedia Appendix 5
Definition of knowledge sharing in medical imaging departments based on the respondents' views.
[DOCX File , 18 KB - humanfactors_v11i1e53780_app5.docx]

References

1. Andualem M, Kebede G, Kumie A. Information needs and seeking behaviour among health professionals working at public hospital and health centres in Bahir Dar, Ethiopia. BMC Health Serv Res 2013 Dec 27;13(1):534 [FREE Full text] [doi: 10.1186/1472-6963-13-534] [Medline: 24373296]
2. Sousa MJ, Dal Mas F, Garcia-Perez A, Cobianchi L. Knowledge in transition in healthcare. Eur J Investig Health Psychol Educ 2020 Jul 17;10(3):733-748 [FREE Full text] [doi: 10.3390/ejihpe10030054] [Medline: 34542508]
3. Bontis N. Managing organizational knowledge by diagnosing intellectual capital: framing and advancing the state of the field. In: Malhotra Y, editor. Knowledge Management and Business Model Innovation. Hershey, PA: IGI Global; 2001:267-297.
4. Yan QY, Xiang F, Shi XX, Zhu Q. Implementation of knowledge management in Chinese hospitals. Curr Med Sci 2018 Apr 30;38(2):372-378. [doi: 10.1007/s11596-018-1888-z] [Medline: 30074199]
5. Candra S, Putrama IK. Applied healthcare knowledge management for hospital in clinical aspect. Telkomnika 2018;16(4). [doi: 10.12928/telkomnika.v16i4.5673]
6. Alavi M, Leidner D. Knowledge management systems: issues, challenges, and benefits. Commun Assoc Inf Syst 1999;1. [doi: 10.17705/1cais.00107]
7. Gordon AN, Hinson RE. Towards a sustainable framework for computer based health information systems (CHIS) for least developed countries (LDCs). Int J Health Care Qual Assur 2007;20(6):532-544. [doi: 10.1108/09526860710819468] [Medline: 18030970]
8. Chen D, Soong SJ, Grimes GJ, Orthner HF. Wireless local area network in a prehospital environment. BMC Med Inform Decis Mak 2004 Aug 31;4(1):12 [FREE Full text] [doi: 10.1186/1472-6947-4-12] [Medline: 15339336]
9. Chien SY, Tsai CH. Dynamic capability, knowledge, learning, and firm performance. J Organ Change Manag 2012;25(3):434-444. [doi: 10.1108/09534811211228148]
10. Lee S, Kim BG, Kim H. An integrated view of knowledge management for performance. J Knowl Manag 2012;16(2):183-203. [doi: 10.1108/13673271211218807]

11. Cardinal LB, Alessandri TM, Turner SF. Knowledge codifiability, resources, and science?based innovation. J Knowl Manag 2001;5(2):195-204. [doi: 10.1108/13673270110393266]
12. Ipe M. Knowledge sharing in organizations: a conceptual framework. Hum Resour Dev Rev 2003 Dec 01;2(4):337-359. [doi: 10.1177/1534484303257985]
13. Nonaka I, von Krogh G, Voelpel S. Organizational knowledge creation theory: evolutionary paths and future advances. Organ Stud 2006 Aug 01;27(8):1179-1208. [doi: 10.1177/0170840606066312]
14. Kothari A, Rudman D, Dobbins M, Rouse M, Sibbald S, Edwards N. The use of tacit and explicit knowledge in public health: a qualitative study. Implement Sci 2012 Mar 20;7:20 [FREE Full text] [doi: 10.1186/1748-5908-7-20] [Medline: 22433980]
15. Grant RM. Toward a knowledge ‐ based theory of the firm. Strateg Manag J 2014 May 20;17(S2):109-122. [doi: 10.1002/smj.4250171110]
16. Spender JC. Making knowledge the basis of a dynamic theory of the firm. Strateg Manag J 2014 May 20;17(S2):45-62. [doi: 10.1002/smj.4250171106]
17. Fletcher A, Worthington D. What is a 'generic' hospital model?--a comparison of 'generic' and 'specific' hospital models of emergency patient flows. Health Care Manag Sci 2009 Dec 9;12(4):374-391. [doi: 10.1007/s10729-009-9108-9] [Medline: 20058527]
18. Jimenez YA, Lewis SJ. Infection prevention and control in the medical imaging environment: a scoping review. Insights Imaging 2023 Jul 14;14(1):121 [FREE Full text] [doi: 10.1186/s13244-023-01470-1] [Medline: 37450068]
19. Abidi SS. Knowledge management in healthcare: towards 'knowledge-driven' decision-support services. Int J Med Inform 2001 Sep;63(1-2):5-18. [doi: 10.1016/s1386-5056(01)00167-8] [Medline: 11518661]
20. Lema B. A framework to support knowledge sharing practice among health care professionals at Yekatit 12 hospital medical college. Addis Ababa University. 2017 Jun. URL: https://tinyurl.com/49a7cpwf [accessed 2024-10-23]
21. Lee HS, Hong SA. Factors affecting hospital employees' knowledge sharing intention and behavior, and innovation behavior. Osong Public Health Res Perspect 2014 Jun;5(3):148-155 [FREE Full text] [doi: 10.1016/j.phrp.2014.04.006] [Medline: 25180147]
22. Tabrizi NM. Models for describing knowledge sharing practices: the case study of UK hospitals. The University of Manchester. 2017 Aug 1. URL: https://tinyurl.com/5xvcr53k [accessed 2024-10-23]
23. Gider Ö, Ocak S, Top M. Perceptions of physicians about knowledge sharing barriers in Turkish health care system. J Med Syst 2015 May 3;39(5):42. [doi: 10.1007/s10916-015-0230-6] [Medline: 25732078]
24. Kim SJ, Park M. Leadership, knowledge sharing, and creativity: the key factors in nurses' innovative behaviors. J Nurs Adm 2015 Dec;45(12):615-621. [doi: 10.1097/NNA.0000000000000274] [Medline: 26565640]
25. Shahmoradi L, Safdari R, Piri Z, Dehghani Mahmodabadi A, Shahmoradi S, Farzaneh Nejad A. Knowledge sharing as a powerful base for management: barriers and solutions. Health Care Manag 2017;36(2):176-183. [doi: 10.1097/hcm.0000000000000150]
26. Yip K. Exploring barriers to knowledge sharing: a case study of a virtual community of practice in a Swedish multinational corporation. Linnaeus University. 2011. URL: https://tinyurl.com/5n8x3er2 [accessed 2024-10-23]
27. Ariati N, Sensuse D, Wuri Handayani P. Factors affecting knowledge sharing capability of doctors in Palembang. J Phys Conf Ser 2020 Apr 01;1500(1):012092 [FREE Full text] [doi: 10.1088/1742-6596/1500/1/012092]
28. MamoMulate E, Gojeh LA. Current status and factors affecting knowledge sharing practices among health professionals in Hiwot Fana Specialized University Hospital in Ethiopia. Univers J Manag 2020 Jul;8(4):160-174. [doi: 10.13189/ujm.2020.080405]
29. Demsash AW, Chakilu B, Mazengia A. Knowledge sharing practice and its associated factors among healthcare providers at University of Gondar Comprehensive Specialized Hospital, North West Ethiopian: cross-sectional study. Research Square Preprint posted online on April 12, 2021 [FREE Full text] [doi: 10.21203/rs.3.rs-407759/v1]
30. Almeida MH, Ramos A, Sousa MJ, Santos CM, Fontes AP. Organizational factors affecting knowledge sharing capabilities: a study with Portuguese health professionals. In: Proceedings of the European Conference on Knowledge Management. 2020 Presented at: ECKM 2020; December 2-4, 2020; Online URL: https://tinyurl.com/59ry937s [doi: 10.34190/ekm.20.153]
31. Pun JK, Matthiessen CM, Murray KA, Slade D. Factors affecting communication in emergency departments: doctors and nurses' perceptions of communication in a trilingual ED in Hong Kong. Int J Emerg Med 2015 Dec 15;8(1):48 [FREE Full text] [doi: 10.1186/s12245-015-0095-y] [Medline: 26667242]
32. Khajouei R, Jahromi ME, Ameri A. Challenges of implementing picture archiving and communication system in multiple hospitals: perspectives of involved staff and users. J Med Syst 2019 May 15;43(7):182. [doi: 10.1007/s10916-019-1319-0] [Medline: 31093803]
33. Adeyelure TS, Kalema BM, Motlanthe BL. An empirical study of knowledge sharing: a case of South African healthcare system. Knowl Manag E Learn Int J 2019;11(1). [doi: 10.34105/j.kmel.2019.11.007]
34. Lam J, Ng B, Shen S, Wong C. How is interprofessional collaboration applied by radiation therapists in the radiation therapy department in British Columbia? J Med Imaging Radiat Sci 2015 Sep;46(3S):S43-51.e2. [doi: 10.1016/j.jmir.2015.04.016] [Medline: 31052107]

35. Sharma V, Stranieri A, Burstein F, Warren J, Daly S, Patterson L, et al. Group decision making in health care: a case study of multidisciplinary meetings. J Decis Syst 2016 Jun 16;25(sup1):476-485. [doi: 10.1080/12460125.2016.1187388]
36. Kilsdonk MJ, Siesling S, Otter R, van Harten WH. Two decades of external peer review of cancer care in general hospitals; the Dutch experience. Cancer Med 2016 Mar 29;5(3):478-485 [FREE Full text] [doi: 10.1002/cam4.612] [Medline: 26714788]
37. Al Mashmoum MJ, Hamade SN. Knowledge sharing among employees of Kuwait Cancer Control Center. Int J Knowl Manag Pract 2019;7(1):15-25 [FREE Full text]
38. Almashmoum M, Cunningham J, Alkhaldi O, Anisworth J. Factors that affect knowledge-sharing behaviors in medical imaging departments in cancer centers: systematic review. JMIR Hum Factors 2023 Jul 12;10:e44327 [FREE Full text] [doi: 10.2196/44327] [Medline: 37436810]
39. A brief history of The Christie. The Christie NHS Foundation Trust. URL: https://tinyurl.com/2uyfnzvr [accessed 2024-10-23]
40. Kuwait Cancer Control Center. Wikipedia. URL: https://en.wikipedia.org/wiki/Kuwait_Cancer_Control_Center [accessed 2024-11-04]
41. Bradley M. Working knowledge: how organizations manage what they know. New Library World 2000 Nov;101(6):282-287. [doi: 10.1108/nlw.2000.101.6.282.4]
42. Alhalhouli ZT, Hassan ZB, Abualkishik AM. An updated model to enhance knowledge sharing among stakeholders in Jordanian hospitals using social networks. World Appl Sci J 2013;25(2):270-280 [FREE Full text]
43. Noor NM, Juhana S. Factors influencing employee knowledge sharing capabilities in electronic government agencies in Malaysia. Int J Comput Sci Issues 2011;8(4):106-114.
44. Lee HS. Knowledge management enablers and process in hospital organizations. Osong Public Health Res Perspect 2017 Feb 28;8(1):26-33 [FREE Full text] [doi: 10.24171/j.phrp.2017.8.1.04] [Medline: 28443221]
45. Dorow PF, Medeiros C, Camozzato TS, Silva C, Vargas FC, Huhn A. Knowledge sharing process between radiologist. Int J Adv Res 2018 Jul 31;6(7):431-436. [doi: 10.21474/ijar01/7380]
46. Taba ST, Hossain L, Willis K, Lewis S. Social networks and expertise development for Australian breast radiologists. BMC Health Serv Res 2017 Feb 11;17(1):131 [FREE Full text] [doi: 10.1186/s12913-016-1938-9] [Medline: 28189143]
47. Alanzi T, Al-Habib DK. The use of social media by healthcare quality personnel in Saudi Arabia. J Environ Public Health 2020 May 22;2020:1417478-1417479 [FREE Full text] [doi: 10.1155/2020/1417478] [Medline: 32565836]
48. Al-Safadi L. The effects of real-time interactive multimedia teleradiology system. Biomed Res Int 2016;2016:4126841 [FREE Full text] [doi: 10.1155/2016/4126841] [Medline: 27294118]
49. Tønnessen Ø, Dhir A, Flåten BT. Digital knowledge sharing and creative performance: work from home during the COVID-19 pandemic. Technol Forecast Soc Change 2021 Sep;170:120866 [FREE Full text] [doi: 10.1016/j.techfore.2021.120866] [Medline: 35068596]
50. Clement KD, Zimmermann EF, Bhatt NR, Light A, Gao C, Kulkarni M, et al. Communication tools in the COVID-19 era and beyond which can optimise professional practice and patient care. BMJ Innov 2020 Sep 08;7(1):217-223. [doi: 10.1136/bmjinnov-2020-000465]
51. Lee YH, Lu TE, Yang CC, Chang G. A multilevel approach on empowering leadership and safety behavior in the medical industry: the mediating effects of knowledge sharing and safety climate. Safety Sci 2019 Aug;117:1-9. [doi: 10.1016/j.ssci.2019.03.022]
52. Fatahi N, Krupic F, Hellström M. Difficulties and possibilities in communication between referring clinicians and radiologists: perspective of clinicians. J Multidiscip Healthc 2019 Jul;Volume 12:555-564. [doi: 10.2147/jmdh.s207649]

Abbreviations

ICT: information and communications technology
KCCC: Kuwait Cancer Control Center
PACS: picture archiving and communication system

Edited by A Kushniruk, E Borycki; submitted 18.10.23; peer-reviewed by L Nemati-Anaraki, T Haygood; comments to author 20.05.24; revised version received 30.08.24; accepted 13.09.24; published 13.11.24.

Please cite as:
Almashmoum M, Cunningham J, Ainsworth J
Evaluating Factors Affecting Knowledge Sharing Among Health Care Professionals in the Medical Imaging Departments of 2 Cancer Centers: Concurrent Mixed Methods Study
JMIR Hum Factors 2024;11:e53780
URL: https://humanfactors.jmir.org/2024/1/e53780
doi:10.2196/53780
PMID:

Italian Version of the mHealth App Usability Questionnaire (Ita-MAUQ): Translation and Validation Study in People With Multiple Sclerosis

Jessica Podda[1], PhD; Erica Grange[1], OT; Alessia Susini[1], MPsy; Andrea Tacchino[1], PhD; Federica Di Antonio[1], MPhys; Ludovico Pedullà[1], PhD; Giampaolo Brichetto[1,2], MD, PhD; Michela Ponzio[1], PhD

[1]Scientific Research Area, Italian Multiple Sclerosis Foundation, Genoa, Italy

[2]Rehabilitation Service of Genoa, Italian Multiple Sclerosis Society, Genoa, Italy

Corresponding Author:
Jessica Podda, PhD
Scientific Research Area, Italian Multiple Sclerosis Foundation, , Genoa, , Italy

Abstract

Background: Telemedicine and mobile health (mHealth) apps have emerged as powerful tools in health care, offering convenient access to services and empowering participants in managing their health. Among populations with chronic and progressive disease such as multiple sclerosis (MS), mHealth apps hold promise for enhancing self-management and care. To be used in clinical practice, the validity and usability of mHealth tools should be tested. The most commonly used method for assessing the usability of electronic technologies are questionnaires.

Objective: This study aimed to translate and validate the English version of the mHealth App Usability Questionnaire into Italian (ita-MAUQ) in a sample of people with MS.

Methods: The 18-item mHealth App Usability Questionnaire was forward- and back-translated from English into Italian by an expert panel, following scientific guidelines for translation and cross-cultural adaptation. The ita-MAUQ (patient version for stand-alone apps) comprises 3 subscales, which are ease of use, interface and satisfaction, and usefulness. After interacting with DIGICOG-MS (Digital Assessment of Cognitive Impairment in Multiple Sclerosis), a novel mHealth app for cognitive self-assessment in MS, people completed the ita-MAUQ and the System Usability Scale, included to test construct validity of the translated questionnaire. Confirmatory factor analysis, internal consistency, test-retest reliability, and construct validity were assessed. Known-groups validity was examined based on disability levels as indicated by the Expanded Disability Status Scale (EDSS) score and gender.

Results: In total, 116 people with MS (female n=74; mean age 47.2, SD 14 years; mean EDSS 3.32, SD 1.72) were enrolled. The ita-MAUQ demonstrated acceptable model fit, good internal consistency (Cronbach α=0.92), and moderate test-retest reliability (intraclass coefficient correlation 0.84). Spearman coefficients revealed significant correlations between the ita-MAUQ total score; the ease of use (5 items), interface and satisfaction (7 items), and usefulness subscales; and the System Usability Scale (all *P* values <.05). Known-group analysis found no difference between people with MS with mild and moderate EDSS (all *P* values >.05), suggesting that ambulation ability, mainly detected by the EDSS, did not affect the ita-MAUQ scores. Interestingly, a statistical difference between female and male participants concerning the ease of use ita-MAUQ subscale was found (*P*=.02).

Conclusions: The ita-MAUQ demonstrated high reliability and validity and it might be used to evaluate the usability, utility, and acceptability of mHealth apps in people with MS.

(JMIR Hum Factors 2024;11:e58079) doi:10.2196/58079

KEYWORDS

mHealth; multiple sclerosis; cognitive assessment; questionnaire validation; usability; mHealth app; mHealth application; validation study; MAUQ; app usability; telemedicine; disability; usability questionnaire; mobile health

Introduction

Telemedicine has enabled convenient and effective visits; reduced unnecessary testing and referrals; maintained good perception of care; reduced travel costs and caregiver burden; and, not least, helped health care providers to manage an ever-increasing volume of information and relationships [1,2]. Growing use of smartphones and tablets has made mobile health (mHealth) apps promising tools for empowering and engaging people in the self-management of their own health [3]. mHealth tools create opportunities to deliver new forms of health care and to expand services without the need to increase the existing workforce. For example, medical apps can be used within various domains such as wellness management, behavior change, health data collection, disease management, self-diagnosis, and rehabilitation as well as act as an electronic patient portal and medication reminder [4,5], leading to greater time spent at home and fewer medical visits at the center [6]. Furthermore, they represent useful solutions for participants with chronic and progressive diseases, such as multiple sclerosis (MS), that require continuous assistance and care.

mHealth apps promise to offer alternative methods for enhanced real-time data capture to screen for, monitor, and treat the heterogeneous symptoms in MS, thus favoring a substantial transformation in traditional paradigms of medicine [7-13]. However, as the number of mHealth apps increases, the demand for scientific evaluation of these solutions is strongly recommended as well [14]. Despite the growing popularity of mHealth apps, the amount of usability reports does not correlate with the number of published digital health implementation studies [15]. For instance, while Salimzadeh and colleagues [3] found 104 MS-related apps in iTunes (Apple Inc) and Google Play (Google LLC), they noted that there was no corresponding evidence regarding the usability and utility of these solutions in people with MS. To be used in clinical practice, the validity and usability of mHealth tools should be tested. mHealth apps must be designed to ensure good usability, and they must be easy to use and able to reach their goals efficiently. As indicated by Wilson and Lankton [16], perceived ease of use and usefulness affect people's intention to adopt mHealth devices. Generally, a mobile app is considered to have good usability when (1) it is efficient, (2) users have a positive opinion about the app, (3) it is easy to learn, (4) it is easy to remember even after users have not used it for a while, and (5) it has a low error rate [17].

The mHealth apps can be grouped according to the nature of the interaction between patients and health care providers in the app: interactive and stand-alone mHealth apps. In interactive mHealth apps, users can send and receive information from their health care providers or patients via the app in a synchronous or asynchronous modality. In stand-alone mHealth apps, users enter, collect, or store health information about themselves or other people, which are not directly sent to the user's health care providers [18].

The most commonly used method for assessing the usability of electronic technologies are questionnaires. General and technology-independent questionnaires such as the System Usability Scale (SUS) [19] and the Post-Study System Usability Questionnaire [20] are usually used in usability studies of mHealth apps [15]. However, these questionnaires were created for general software systems and cannot reliably identify mHealth specific problems that may arise, for example, in health self-management or accessing health care services.

In this context, a new specific usability scale for evaluating the validity of mHealth apps was developed, the mHealth App Usability Questionnaire (MAUQ) [18]. The English version of MAUQ has been translated to various languages such as Malay [21], Chinese [22], Spanish [23], German [24] and French [25]. However, no literature was found reporting a translated version of the questionnaire in Italian, although it was used in 1 study on app usability [26]. Thus, this study aimed to translate and validate the English version of the mHealth App Usability Questionnaire into Italian (ita-MAUQ) in a sample of people with MS.

We specifically hypothesized that the ita-MAUQ would retain acceptable model fit in confirmatory factor analysis, acceptable levels of internal consistency, and test-retest reliability. We also hypothesized an acceptable construct validity, defined based on relations between the ita-MAUQ and another standardized usability scale, and differences in known-groups.

Methods

mHealth App Usability Questionnaire

The MAUQ was first developed by Zhou and colleagues [18]. MAUQ is designed for different users (patients or health care providers) and different interaction modes (interactive or stand-alone). For study purposes, the patient version for stand-alone mHealth apps was used. It consists of 18 items divided into 3 subscales: ease of use (5 items; MAUQ_E), interface and satisfaction (7 items; MAUQ_I), and usefulness (6 items; MAUQ_U). The overall Cronbach α coefficients of the original questionnaire were 0.85 for MAUQ_E, 0.90 for MAUQ_I, and 0.72 for MAUQ_U, which indicated strong internal consistency of the questionnaire [18]. The questionnaire uses a 7-point Likert scoring system: 1 (strongly disagree), 2 (disagree), 3 (somewhat disagree), 4 (neither agree nor disagree), 5 (somewhat agree), 6 (agree), and 7 (strongly agree). The authors point out that there are no licensing fees for using the questionnaire, and it is not necessary to request permission before using it. The questionnaire is freely accessible on the website.

MAUQ Translation and Cross-Cultural Adaptation

The original MAUQ questionnaire for stand-alone mHealth apps (patient version) was translated into the Italian language using a guideline for the translation, adaptation, and validation of instruments or scales for cross-cultural health care research [27]. The questionnaire was first translated by native Italian speakers proficient in English. The forward-translated versions of the instrument were initially compared by a third independent translator regarding ambiguities and discrepancies of words, sentences, and meanings to generate a preliminary initial translated version of the questionnaire. The back translation was verified by translating the Italian version to English by 2 native English speaker translators with a high Italian proficiency. A multidisciplinary panel with 3 health care professionals with expertise in MS (an occupational therapist, a physiotherapist, and a psychologist), 2 researchers with expertise in scale validation, and 1 expert in app development compared the translated versions with the original questionnaire and made modifications to make the questionnaire more understandable to Italian people with MS. The final ita-MAUQ questionnaire can be viewed in Multimedia Appendix 1.

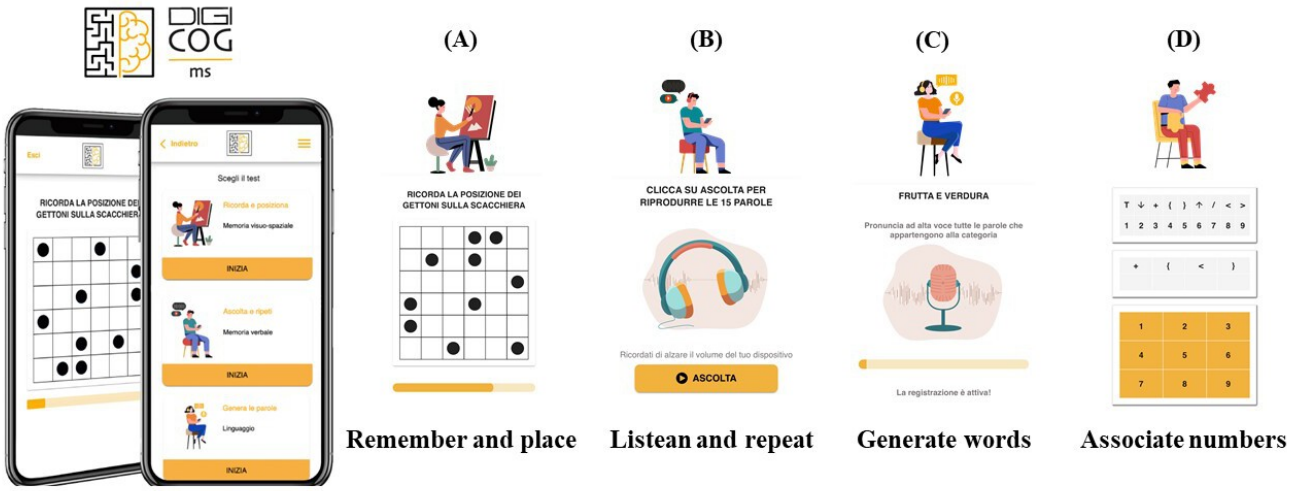
mHealth App Used for Validation of the Ita-MAUQ

In addition to motor and sensory difficulties, cognitive impairment, known as an invisible symptom, affects up to 65% of people with MS. Documented in all MS courses, with more severe deficits in progressive forms, both secondary progressive and primary progressive, compared to relapsing-remitting MS [28], cognitive impairment is recognized as one of the most disturbing disorders in MS, negatively affecting the quality of life and independence of people with MS. Attention, information processing speed, learning and memory, and executive functions seem to be the most commonly affected cognitive domains [29]. Consistent evidence indicates that cognitive functions in people with MS can be grouped into cognitive phenotypes, that is, subgroups of people with MS with a similar pattern of cognitive functioning [30-32]. The validation of the ita-MAUQ was conducted using DIGICOG-MS (Digital Assessment of Cognitive Impairment in Multiple Sclerosis), a smartphone- and tablet-based app for self-assessment of cognitive impairment in people with MS [33] (please see Figure 1 for an overview of the mHealth app). DIGICOG-MS (intellectual property of Italian Multiple Sclerosis Foundation; Italian Society of Authors and Publishers Registration ID: D000018162, 27-12-2022) includes 4 digital tests designed to evaluate the most affected cognitive domains in MS as visuospatial memory, verbal memory, semantic fluency, and information processing speed [34,35]:

- *Remember and place* assesses visuospatial episodic memory. A 36-square grid with 10 black checkers is displayed on the screen for 10 seconds. After the time elapses, the pattern disappears, and participants must reproduce it on a blank checkerboard. This replicates the 10/36 Spatial Recall Test [35], in which a 6 × 6 checkerboard with 10 pieces arranged in a particular pattern is shown to the participant for 10 seconds. Both tests (digital and traditional) include 3 consecutive trials, and the score consists of the total number of correct responses for the 3 trials.
- *Listen and repeat* was developed as an electronic version of the Rey Verbal Learning Test [36] that evaluates verbal memory. Participants listen to a prerecorded list of 15 common nouns and are asked to recall as many words as possible 5 times. Responses are recorded and then scored by the neuropsychologist. In the traditional test, words are read aloud to the participant who is asked to repeat as many words as possible in any order. All pronounced nouns in each of the 5 learning trials are transcribed by the neuropsychologist. For both versions of the test, the total score consists of the number of words recalled across the 5 trials.
- *Generate words* is a digital adaptation of the Word List Generation [35,37] and measures semantic verbal fluency. Participants generate a list of words, typically constrained by a specific semantic category, in 90 seconds. Recordings of pronounced words are processed by the neuropsychologist for scoring. In the traditional test, all words generated within the given semantic category are transcribed by the neuropsychologist. The total score is based on the number of correct words produced.

In a study by Podda and colleagues [33], correlation analysis was performed to determine the strength of the association between digital (ie, remember and place, listen and repeat, generate words, and associate numbers) and traditional (ie, 10/36 Spatial Recall Test, Rey Verbal Learning Test, Word List Generation, and Symbol Digit Modalities Test) tests. Overall, the findings revealed strong correlations between digital and traditional paper-based tests across all cognitive domains, with correlation coefficients (*r*) ranging from 0.58 to 0.78. Test-retest reliability was excellent for verbal memory and information processing speed (intraclass correlation coefficients [ICCs] ≥0.95) and good for visuospatial memory and semantic fluency (ICCs≥0.83).

Figure 1. Overview of DIGICOG-MS, the mobile health for cognitive assessment of people with multiple sclerosis. The 4 digital tests implemented in DIGICOG-MS that measure visuospatial memory (**A**), verbal memory (**B**), semantic fluency (**C**), and information processing speed (**D**). DIGICOG-MS: Digital Assessment of Cognitive Impairment in Multiple Sclerosis.



Study Participants

This study's participants were people with MS enrolled by the Italian MS Foundation and followed as outpatients at the Italian Multiple Sclerosis Society Rehabilitation Service in Genoa (Italy). Eligibility criteria were to be aged 18 years or older; have a confirmed MS diagnosis following the McDonald criteria [38]; have any disease course (relapsing-remitting MS, secondary progressive MS, and primary progressive MS); have not relapsed in the last 3 months; have an Expanded Disability

Status Scale (EDSS) score [39] ≤7.5; and have adequate visual, hearing, and motor capabilities to work on a tablet. Exclusion criteria were a Montreal Cognitive Assessment score <18, neurological and major psychiatric illness, past serious head trauma, and alcohol or drug abuse. Data were collected from January 2023 to November 2023.

Study Procedure

Participants received a brief explanation of this study's process and goal at the Rehabilitation Service of the Italian Multiple Sclerosis Society in Genoa (Italy). People with MS were invited to use DIGICOG-MS. Particularly, they were required first to perform the 4 digital cognitive tests and then to explore other functionalities of the app (ie, how to log in, insert, and modify personal information; check for historical results of a specific test; consult the app tutorial; and ask for support). While the digital cognitive assessment was supervised by a neuropsychologist, people with MS were invited to navigate through DIGICOG-MS autonomously, and no specific instructions on how to use other functionalities of the app were given. After finishing the tasks, people with MS completed the ita-MAUQ and SUS [19,40]. SUS was included to test construct validity of the translated questionnaire. The 10-item SUS evaluates users' personal perceptions about how to use a given system or device, ranging from "strongly disagree" (1 point) to "strongly agree" (5 points). The SUS score is calculated by adding the individual scores and then multiplying that sum by 2.5. Thus, the SUS score ranges from zero (lowest usability) to 100 (highest usability), with a value of 68 considered above average. Although the SUS has already been translated and validated in many languages [41], individual problems that people may have when using mHealth apps are not specifically identified by the SUS.

Statistical Analysis

Categorical data were summarized by numbers and percentages, while numerical data were indicated by the mean and SD. Degree of education was coded as less than or equal to 12 years (primary school), between 13 and 15 years (high school), and equal to or more than 16 years (university).

Since confirmatory factor analysis (CFA) has emerged as a pivotal technique in such contexts, offering a comprehensive method for comparing the hypothesized measurement model structure with the observed one [42], it was conducted on the 18-item ita-MAUQ using the original 3 higher-order factors structure (ie, MAUQ_E, MAUQ_I, and MAUQ_U) [18]. CFA was used instead of "discovering" or exploring potential relationships between variables, as in exploratory factor analysis, because it is designed to test a predefined model based on consistent theoretical expectations [43]. Goodness-of-fit was tested with the ratio between chi-square (χ^2) and *df* (χ^2/df; good if ≤3), root mean square error of approximation (good if ≤0.08) and comparative fit index >0.9 [44]. We reported these statistics using the Satorra-Bentler adjustment because the scale item distribution was nonnormal [45]; the covariance of error terms was considered to improve the model fit. In the data analysis, missing data were replaced with a value of 4.

The internal consistency of the ita-MAUQ was assessed by calculating Cronbach α coefficient and average interitem correlation. The statistically acceptable Cronbach α coefficient should be >0.7 [46], and average interitem correlations should be between 0.30 and 0.70 [47].

The ICC (2-way analysis of variance random effect model for agreement) was calculated to assess the test-retest reliability. A very small sample size is required for estimating the desired value of ICC (especially when a researcher aims to estimate a very high value of ICC). Using power analysis calculations to test reliability between 2 different observations as described by Bujang and Baharum [48], a minimum of 22 participants is needed to have an acceptable ICC value ≥0.5 (α=.05, power=80%, n=2). The ICC was calculated on subscales and total scores, which are expected to remain stable. An ICC value of 0.70 was recommended as a minimum standard for reliability [49].

Ceiling and floor effects were calculated for the overall ita-MAUQ and its subscales. The floor and ceiling effects were defined as the percentages of respondents who reported the lowest score and the highest score, respectively. Floor and ceiling effects were considered present if >15% of participants achieved either the lowest or highest scores in ita-MAUQ and its subscales [50].

To examine the construct validity, Spearman correlations coefficients (ρ), used for nonnormally distributed data, were calculated between the ita-MAUQ overall score and subscale scores, and the SUS. Spearman coefficients were considered low for ρ<0.30, moderate for ρ 0.30-0.59, and high for ρ≥0.60 [51].

Known-groups validity evaluates whether an instrument can discriminate between known groups of people with MS that are expected to score differently on the measure of interest (ie, EDSS and gender). Here, it was assessed by comparing, with the Mann-Whitney *U* test, the total score and subscale scores of participants' groups with different levels of disability. Groups were defined using an EDSS cutoff value of 3.5, discriminating between people with MS with a mild and moderate disability: able (EDSS≤3.5) or unable (EDSS>3.5) to walk without aid or rest for more than 500 m. Furthermore, since previous research investigating gender differences in users' acceptance for website usability highlights gender as a key variable in understanding usage behavior in information and communication technology [5,52,53], ita-MAUQ total score and subscale scores were also compared with groups divided by participants' gender.

The *P* values <.05 were considered statistically significant. Data were analyzed using Stata (version 17; StataCorp).

Ethical Considerations

This study was approved by Regional Ethics Committee of Azienda Ospedaliera "San Martino" of Genoa (Italy N. 240/2022DB id 12354) and conducted according to the Declaration of Helsinki [54]. Before entering this study, participants had to read, complete, and sign an informed consent. Data collected were stored in an anonymized format to properly protect the privacy and confidentiality of participants, ensuring that no participant can be identified from the data provided.

Participants have been informed that data collection could be used only for research purposes. They did not receive any compensation for taking part in the study

Results

In total, 116 people with MS (female: n=74; mean age 47.2, SD 14 years mean EDSS 3.32, SD 1.72) were enrolled (see Table 1).

Results from CFA of the 18 items-MAUQ order structure, defined by Zhou and colleagues [18], indicated acceptable fit for a 3D scale: χ^2/df=2.1, root mean square error of approximation 0.068, and comparative fit index 0.92.

Table . Demographic and clinical sample characteristics (N=116).

Characteristics		Value
Gender, n (%)		
	Male	42 (36.2)
	Female	74 (63.8)
Age (years)		
	Mean (SD)	47.2 (14)
	Range	19 - 70
Years of education, mean (SD)		13.3 (2.6)
Education, n (%)		
	Primary school	16 (13.8)
	High school	80 (69)
	University	20 (17.2)
MS [a] duration (years)		
	Mean (SD)	11.2 (9.6)
	Range	0 - 32
MS course, n (%)		
	RR[b]	91 (78.5)
	SP[c]	15 (12.9)
	PP[d]	10 (8.6)
EDSS [e]		
	Mean (SD)	3.32 (1.72)
	Range	1 - 7.5
EDSS, n (%)		
	Mild disability, score ≤3.5	67 (57.8)
	Moderate disability, score >3.5	49 (42.2)

[a]MS: multiple sclerosis.

[b]RR: relapsing-remitting multiple sclerosis.

[c]SP: secondary progressive multiple sclerosis.

[d]PP: primary progressive multiple sclerosis.

[e]EDSS: Expanded Disability Status Scale.

The internal consistency of the overall ita-MAUQ and each subscale was good. The Cronbach α for the ita-MAUQ was 0.92, and those for the 3 subscales were 0.78 (ita-MAUQ_E), 0.89 (ita-MAUQ_I), and 0.87 (ita-MAUQ_U). Similarly, the average interitem correlation was between 0.41 and 0.53. These results align with published satisfactory thresholds for scale reliability [23,24].

To assess the test-retest reliability, 25 participants were required to complete the ita-MAUQ, 2 weeks apart. ICC was 0.84 (95% CI 0.66 - 0.92) for ita-MAUQ total score, 0.66 (95% CI 0.36 - 0.84) for ita-MAUQ_E, 0.88 (95% CI 0.75 - 0.94) for ita-MAUQ_I, and 0.67 (95% CI 0.38 - 0.84) for ita-MAUQ_U, showing good or moderate temporal stability.

There were no floor or ceiling effects for the ita-MAUQ total score (0% scored 18 and 6.9% scored 126). The floor effect for ita-MAUQ_E and ita-MAUQ_I subscales was again null, while for ita-MAUQ_U, only 1 participant had the worst possible score of 6 (0.9%). The ceiling effect for ita-MAUQ subscales was 38.8%, 36.2%, and 9.5%, respectively. No other previous study reported the measure of ceiling and floor effects of this instrument [22-24], and therefore a comparison cannot be established.

Spearman coefficients revealed significant correlations between ita-MAUQ total score; ita-MAUQ_E, ita-MAUQ_I, ita-MAUQ_U subscales; and SUS (85.43, SD 14.3; all *P* values <.05). Table 2 shows the results of the construct validity analysis.

Known-group analysis found no difference between people with MS with mild and moderate EDSS (all *P* values >.05), suggesting that ambulation ability, mainly detected by the EDSS, did not impact the ita-MAUQ scores (Table 3). Interestingly, statistical differences between female and male participants concerning the ita-MAUQ_E was found (*P*=.02; Table 4).

Table . Spearman correlation (ρ) between the ita-MAUQ total[a] score; ita-MAUQ_E[b], ita-MAUQ_I[c], and ita-MAUQ_U[d] subscales; and SUS[e] total score.

Variable		Value, mean (SD)	SUS	ita-MAUQ_E	ita-MAUQ_I	ita-MAUQ_U	ita-MAUQ_tot
SUS		85.43 (14.3)					
	r		—[f]				
	P value		—				
ita-MAUQ_E		31.7 (4.1)					
	r		0.54	—			
	P value		<.001	—			
ita-MAUQ_I		44.3 (6.3)					
	r		0.60	0.76	—		
	P value		<.001	<.001	—		
ita-MAUQ_U		31.8 (6.8)					
	r		0.26	0.38	0.48	—	
	P value		.005	<.001	<.001	—	
ita-MAUQ_tot		107.8 (14.5)					
	r		0.53	0.78	0.85	0.80	—
	P value		<.001	<.001	<.001	<.001	—

[a]ita-MAUQ_tot: Italian version of the mHealth App Usability Questionnaire.

[b]ita-MAUQ_E: ease of use subscale of the Italian version of the mHealth App Usability Questionnaire.

[c]ita-MAUQ_I: interface and satisfaction subscale of the Italian version of the mHealth App Usability Questionnaire.

[d]ita-MAUQ_U: usefulness subscale of the Italian version of the mHealth App Usability Questionnaire.

[e]SUS: System Usability Scale.

[f]Not applicable.

Table . Comparison of ita-MAUQ total[a] and subscale scores between people with multiple sclerosis with different disability levels.

Variable	ita-MAUQ_E[b], mean (SD)	ita-MAUQ_I[c], mean (SD)	ita-MAUQ_U[d], mean (SD)	ita-MAUQ_tot, mean (SD)
EDSS[e]≤3.5	31.9 (3.6)	44.1 (6.2)	31.0 (6.9)	107.0 (13.9)
EDSS>3.5	31.4 (4.7)	44.7 (6.7)	32.9 (6.6)	109.0 (15.5)
P value	.76	.28	.10	.24

[a]ita-MAUQ_tot: Italian version of the mHealth App Usability Questionnaire.

[b]ita-MAUQ_E: ease of use subscale of the Italian version of the mHealth App Usability Questionnaire.

[c]ita-MAUQ_I: interface and satisfaction subscale of the Italian version of the mHealth App Usability Questionnaire.

[d]ita-MAUQ_U: usefulness subscale of the Italian version of the mHealth App Usability Questionnaire.

[e]EDSS: Expanded Disability Status Scale.

Table . Comparison of ita-MAUQ total[a] and subscale scores between people with multiple sclerosis by gender.

Variable	ita-MAUQ_E[b] mean (SD)	ita-MAUQ_I[c] mean (SD)	ita-MAUQ_U[d] mean (SD)	ita-MAUQ_tot mean (SD)
Female	32.4 (3.6)	45.1 (6.0)	32.0 (7.2)	109.5 (14.3)
Male	30.5 (4.6)	43.1 (6.9)	31.4 (6.1)	104.9 (14.8)
P value	.02	.09	.51	.06

[a]ita-MAUQ_tot: Italian version of the mHealth App Usability Questionnaire.

[b]ita-MAUQ_E: ease of use subscale of the Italian version of the mHealth App Usability Questionnaire.

[c]ita-MAUQ_I: interface and satisfaction subscale of the Italian version of the mHealth App Usability Questionnaire.

[d]ita-MAUQ_U: usefulness subscale of the Italian version of the mHealth App Usability Questionnaire.

Discussion

Digital solutions as mHealth apps promise to offer alternative methods for enhanced real-time data capture to screen for, monitor, and treat symptoms in MS. These solutions may fundamentally shift traditional paradigms of medicine [7-12]. To be used in clinical practice, the validity and usability of mHealth tools should be tested. Questionnaires are the well-known methods for usability testing, but developing a new one might require concerted effort by the members of a research team, extra cost, and a lot of time [55]. Thus, adaptation of established, appropriate, and available questionnaires with documented validity in other languages is recommended [55].

Thus, the aim of this study was to translate and validate the English version of the ita-MAUQ in a sample of people with MS. Overall, findings demonstrated that the novel translated questionnaire ita-MAUQ is a reliable and valid measurement tool to assess the usability of mHealth apps for people with MS. In this context, people with MS self-administered the ita-MAUQ after interacting with DIGICOG-MS, a novel mHealth app for cognitive self-assessment in MS.

Results indicated that the ita-MAUQ had good internal consistency and stability, as indicated by the Cronbach α coefficient of 0.92. This is in line with the original version of the MAUQ [18] and with other translations of the same questionnaire [23,24,56].

Worldwide, Spearman coefficients between ita-MAUQ total score, subscales, and SUS were statistically significant, proving good criterion and construct validity. However, correlation between the SUS and ita-MAUQ_U was found to be low (0.26). This was in line with the study by Zhou et al [18], in which correlation between MAUQ_U and the SUS was 0.383, reflecting that MAUQ_U is mainly about the usefulness of apps for health care, which is an aspect not covered by the SUS.

Compared to another previous study on MAUQ that did not perform test-retest [24], results revealed good or moderate temporal stability. Given that mHealth apps may allow continuous health care services over time, identifying valid methods to test whether a digital tool is reliable in different measurements is crucial.

The ceiling effect of the ita-MAUQ subscores could indicate that the proposed mHealth app was indeed easy to use for many people with MS. Even though the interaction with DIGICOG-MS was supervised by a neuropsychologist to provide adequate responses to any questions from participants, they were invited to navigate through the mHealth app autonomously, since no specific instructions on how to use other app functionalities were given. In this way, we interpreted this ceiling effect positively as a successful interaction.

In general, the known-groups validity of the instrument was shown by the comparison of ita-MAUQ total and subscale scores between people with MS with different disability levels. The results indicate no significant difference in people with MS with mild and moderate EDSS. Since EDSS mainly measured ambulation capacity, this suggests that such mHealth apps could be found usable for both people that need assistance during ambulation and those who are able to walk without any aid or support.

Concerning gender, earlier studies found that perceived ease of use and usefulness technology may differ by gender [5,52]. Our results are in line with a previous study that demonstrated that gender was associated with higher usability scores in female than male participants [57]. Interestingly, male and female people with MS had different scores in ita-MAUQ_E, suggesting that women are more likely than men to be influenced by effort expectancy and facilitating conditions [53]. These results can help developers to enhance the usability of their services for all users with different personal and clinical characteristics, since men and women still have different traits and societal roles, which may affect their perceptions and usage of technologies.

Our study has several limitations. First, as a single-center study, participants' characteristics may limit the interpretation of our results. The study sample may be considered representative of those clinic-attending people with MS followed as outpatients in rehabilitation centers (ie, middle-age or older adults and with a longer disease duration) [58]. Thus, results may not generalize to other populations of participants with MS (eg, young and neo-diagnosed people). Second, in this validation study, we used DIGICOG-MS that was designed and developed to assess a specific symptom in MS, that is cognitive impairment. Given the high frequency of cognitive impairment in people with MS, it is reasonable to conclude that study participants who have experienced such a disturbing symptom tended to appreciate the mHealth app more compared to people with other neurological diseases. People with MS with cognitive impairment might find such digital tools particularly beneficial due to their potential to offer structured cognitive training and monitoring, which can enhance their daily functioning and quality of life [59]. Thus, it cannot be ruled out that the results of the ita-MAUQ validation would have been different with a clinical population with dissimilar characteristics.

Third, this study was conducted in a controlled clinical setting, allowing participants to familiarize themselves with both the novel technology and usability questionnaires' items. While this approach may have some limitations in terms of generalizability to real-life scenarios and may influence participants' engagement and perceived usability, having a facilitator available to assist, if needed, people with MS that could have problems with reading and interpreting the questionnaire items ensured they understand each question before responding. Furthermore, in this study we overlooked other key factors that may influence usability, such as assessing the importance of providing adequate training and continuous support to users or the role of previous experience with a similar technology; investigating how interface design that considering layout, navigation, and accessibility features makes the app user-friendly for participants with cognitive impairments; evaluating the concerns and preferences of users regarding data security and privacy, which are crucial for building trust and ensuring compliance with health regulations data security and privacy. Here, we did not collect additional feedback after the completion of both digital assessment and usability questionnaires from people with MS. Further study should include feedback sections where participants can indicate if they found any items difficult to understand, allowing for continuous improvement of the novel tools. Worldwide, incorporating these factors into future research can lead to the development of more effective, user-friendly, and impactful mHealth apps for people with MS.

In conclusion, the ita-MAUQ demonstrated high reliability and validity, and it might be used to evaluate the usability, utility, and acceptability of mHealth apps in people with MS. This finding is in line with previous validation of the MAUQ in different languages as Malay [21], Chinese [22], Spanish [23], German [24], and French [25], further confirming the cross-cultural validity, reliability, and adaptability of the MAUQ.

Acknowledgments

We would like to thank Maria Madera and Giulia Bignone for their help in participants' enrollment and all the people with MS that took part in our study. This work has been funded as Progetto Ricerca Speciale by Italian Multiple Sclerosis Foundation (2021/Special/005).

Data Availability

All data produced in this study are available upon reasonable request to the corresponding author.

The questionnaire can be used in clinical trials and is available free of charge for non-commercial parties, provided the authors are cited, or by contacting the corresponding author.

Authors' Contributions

JP conceived this study, interpreted results, and drafted this paper. MP performed statistical analysis, created tables, and contributed to the initial draft of this paper. AS performed data collection. FDA, AT, LP, and EG revised this paper. All authors read and approved the final paper. This paper was written entirely by humans. All authors are responsible and accountable for the originality, accuracy, and integrity of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The Italian version of the mHealth App Usability Questionnaire (ita-MAUQ; for stand-alone mHealth apps used by patients).
[DOCX File, 23 KB - humanfactors_v11i1e58079_app1.docx]

References

1. Schwamm LH. Telehealth: seven strategies to successfully implement disruptive technology and transform health care. Health Aff 2014 Feb;33(2):200-206. [doi: 10.1377/hlthaff.2013.1021]
2. Lavorgna L, Brigo F, Moccia M, et al. e-Health and multiple sclerosis: an update. Mult Scler J 2018 Nov;24(13):1657-1664. [doi: 10.1177/1352458518799629] [Medline: 30231004]
3. Salimzadeh Z, Damanabi S, Kalankesh LR, Ferdousi R. Mobile applications for multiple sclerosis: a focus on self-management. Acta Inform Med 2019 Mar;27(1):12-18. [doi: 10.5455/aim.2019.27.12-18] [Medline: 31213737]
4. Kao CK, Liebovitz DM. Consumer mobile health apps: current state, barriers, and future directions. PM R 2017 May;9(5S):S106-S115. [doi: 10.1016/j.pmrj.2017.02.018] [Medline: 28527495]
5. Zhang X, Guo X, Lai KH, Guo F, Li C. Understanding gender differences in m-Health adoption: a modified theory of reasoned action model. Telemed J E Health 2014 Jan;20(1):39-46. [doi: 10.1089/tmj.2013.0092] [Medline: 24161004]

6. van Elburg FRT, Klaver NS, Nieboer AP, Askari M. Gender differences regarding intention to use mHealth applications in the Dutch elderly population: a cross-sectional study. BMC Geriatr 2022 May 24;22(1):449. [doi: 10.1186/s12877-022-03130-3] [Medline: 35610577]
7. Beier M, Alschuler K, Amtmann D, Hughes A, Madathil R, Ehde D. ICAMS: assessing the reliability of a Brief International Cognitive Assessment for Multiple Sclerosis (BICAMS) tablet application. Int J MS Care 2020;22(2):67-74. [doi: 10.7224/1537-2073.2018-108] [Medline: 32410901]
8. van Dongen L, Westerik B, van der Hiele K, et al. Introducing multiple screener: an unsupervised digital screening tool for cognitive deficits in MS. Mult Scler Relat Disord 2020 Feb;38:101479. [doi: 10.1016/j.msard.2019.101479] [Medline: 31760365]
9. Montalban X, Graves J, Midaglia L, et al. A smartphone sensor-based digital outcome assessment of multiple sclerosis. Mult Scler 2022 Apr;28(4):654-664. [doi: 10.1177/13524585211028561] [Medline: 34259588]
10. Hochstrasser C, Rieder S, Jufer-Riedi U, et al. Computerized symbol digit modalities test in a swiss pediatric cohort part 1: validation. Front Psychol 2021 Apr 22;12:631536. [doi: 10.3389/fpsyg.2021.631536] [Medline: 33967898]
11. Costabile T, Signoriello E, Lauro F, et al. Validation of an iPad version of the Brief International Cognitive Assessment for Multiple Sclerosis (BICAMS). Mult Scler Relat Disord 2023 Jun;74:104723. [doi: 10.1016/j.msard.2023.104723] [Medline: 37086633]
12. Hsu WY, Rowles W, Anguera JA, et al. Assessing cognitive function in multiple sclerosis with digital tools: observational study. J Med Internet Res 2021 Dec 30;23(12):e25748. [doi: 10.2196/25748] [Medline: 34967751]
13. Brichetto G, Pedullà L, Podda J, Tacchino A. Beyond center-based testing: understanding and improving functioning with wearable technology in MS. Mult Scler J 2019 Sep;25(10):1402-1411. [doi: 10.1177/1352458519857075] [Medline: 31502913]
14. McKay FH, Cheng C, Wright A, Shill J, Stephens H, Uccellini M. Evaluating mobile phone applications for health behaviour change: a systematic review. J Telemed Telecare 2018 Jan;24(1):22-30. [doi: 10.1177/1357633X16673538] [Medline: 27760883]
15. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: a scoping review. Int J Med Inform 2019 Jun;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
16. Wilson EV, Lankton NK. Modeling patients' acceptance of provider-delivered e-Health. J Am Med Inform Assoc 2004;11(4):241-248. [doi: 10.1197/jamia.M1475] [Medline: 15064290]
17. Zhou L, Bao J, Parmanto B. Systematic review protocol to assess the effectiveness of usability questionnaires in mHealth app studies. JMIR Res Protoc 2017 Aug 1;6(8):e151. [doi: 10.2196/resprot.7826] [Medline: 28765101]
18. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): development and validation study. JMIR mHealth uHealth 2019 Apr 11;7(4):e11500. [doi: 10.2196/11500] [Medline: 30973342]
19. Brooke J. SUS: a quick and dirty usability. In: Usability Evaluation in Industry: Taylor & Francis; 1996:189-194 URL: https://digital.ahrq.gov/sites/default/files/docs/survey/systemusabilityscale%2528sus%2529_comp%255B1%255D.pdf [accessed 2024-08-20]
20. Lewis JR. Psychometric evaluation of the PSSUQ using data from five years of usability studies. Int J Human Comp Interaction 2002 Sep 1;14(3):463-488. [doi: 10.1207/S15327590IJHC143&4_11]
21. Zubairi AM, Kassim NLA. Classical and Rasch analyses of dichotomously scored reading comprehension test items. Malaysian J ELT Res 2006 Mar;2:1-20 [FREE Full text]
22. Shan Y, Ji M, Xie W, et al. Chinese version of the Mobile Health App Usability Questionnaire: translation, adaptation, and validation study. JMIR Form Res 2022 Jul 6;6(7):e37933. [doi: 10.2196/37933] [Medline: 35793132]
23. Quifer-Rada P, Aguilar-Camprubí L, Gómez-Sebastià I, Padró-Arocas A, Mena-Tudela D. Spanish version of the mHealth App Usability Questionnaire (MAUQ) and adaptation to breastfeeding support apps. Int J Med Inform 2023 Jun;174:105062. [doi: 10.1016/j.ijmedinf.2023.105062] [Medline: 37037124]
24. Tacke T, Nohl-Deryk P, Lingwal N, et al. The German version of the mHealth App Usability Questionnaire (GER-MAUQ): translation and validation study in patients with cardiovascular disease. D Health 2024 Jan 31;10:20552076231225168. [doi: 10.1177/20552076231225168] [Medline: 38303970]
25. Gagnon J, Probst S, Chartrand J, Lalonde M. mHealth App Usability Questionnaire for stand-alone mHealth apps used by health care providers: Canadian French translation, cross-cultural adaptation, and validation (part 1). JMIR Form Res 2024 Feb 13;8:e50839. [doi: 10.2196/50839] [Medline: 38349710]
26. Renati R, Bonfiglio NS, Rollo D. Dealing with loved ones' addiction: development of an app to cope with caregivers' stress. Int J Environ Res Public Health 2022 Nov 30;19(23):15950. [doi: 10.3390/ijerph192315950] [Medline: 36498025]
27. Sousa VD, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. J Eval Clin Pract 2011 Apr;17(2):268-274. [doi: 10.1111/j.1365-2753.2010.01434.x] [Medline: 20874835]
28. Amato MP, Prestipino E, Bellinvia A. Identifying risk factors for cognitive issues in multiple sclerosis. Expert Rev Neurother 2019 Apr;19(4):333-347. [doi: 10.1080/14737175.2019.1590199] [Medline: 30829076]

29. Benedict RHB, Amato MP, DeLuca J, Geurts JJG. Cognitive impairment in multiple sclerosis: clinical management, MRI, and therapeutic avenues. Lancet Neurol 2020 Oct;19(10):860-871. [doi: 10.1016/S1474-4422(20)30277-5] [Medline: 32949546]
30. De Meo E, Portaccio E, Giorgio A, et al. Identifying the distinct cognitive phenotypes in multiple sclerosis. JAMA Neurol 2021 Apr 1;78(4):414-425. [doi: 10.1001/jamaneurol.2020.4920] [Medline: 33393981]
31. Leavitt VM, Tosto G, Riley CS. Cognitive phenotypes in multiple sclerosis. J Neurol 2018 Mar;265(3):562-566. [doi: 10.1007/s00415-018-8747-5] [Medline: 29356970]
32. Podda J, Ponzio M, Pedullà L, et al. Predominant cognitive phenotypes in multiple sclerosis: insights from patient-centered outcomes. Mult Scler Relat Disord 2021 Jun;51:102919. [doi: 10.1016/j.msard.2021.102919] [Medline: 33799285]
33. Podda J, Tacchino A, Ponzio M, et al. Mobile health app (DIGICOG-MS®) for self-assessment of cognitive impairment in people with multiple sclerosis: instrument validation and usability study. JMIR Form Res 2024 Jun 20;8:e56074. [doi: 10.2196/56074] [Medline: 38900535]
34. Goretti B, Patti F, Cilia S, et al. The Rao's Brief Repeatable Battery version B: normative values with age, education and gender corrections in an Italian population. Neurol Sci 2014 Jan;35(1):79-82. [doi: 10.1007/s10072-013-1558-7] [Medline: 24101117]
35. Amato MP, Portaccio E, Goretti B, et al. The Rao's Brief Repeatable Battery and Stroop Test: normative values with age, education and gender corrections in an Italian population. Mult Scler J 2006 Dec;12(6):787-793. [doi: 10.1177/1352458506070933] [Medline: 17263008]
36. Carlesimo GA, Caltagirone C, Gainotti G. The mental deterioration battery: normative data, diagnostic reliability and qualitative analyses of cognitive impairment. The group for the standardization of the mental deterioration battery. Eur Neurol 1996;36(6):378-384. [doi: 10.1159/000117297] [Medline: 8954307]
37. Goretti B, Niccolai C, Hakiki B, et al. The Brief International Cognitive Assessment for Multiple Sclerosis (BICAMS): normative values with gender, age and education corrections in the Italian population. BMC Neurol 2014 Sep 10;14:171. [doi: 10.1186/s12883-014-0171-6] [Medline: 25204350]
38. Polman CH, Reingold SC, Banwell B, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann Neurol 2011 Feb;69(2):292-302. [doi: 10.1002/ana.22366] [Medline: 21387374]
39. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 1983 Nov;33(11):1444-1452. [doi: 10.1212/wnl.33.11.1444] [Medline: 6685237]
40. 5 ways to interpret a SUS score. MeasuringU. 2018. URL: https://measuringu.com/interpret-sus-score/ [accessed 2024-08-17]
41. Gao M, Kortum P, Oswald FL. Multi-language toolkit for the System Usability Scale. Int J Hum Comput Interact 2020 Dec 13;36(20):1883-1901. [doi: 10.1080/10447318.2020.1801173]
42. Rios J, Wells C. Evidencia de validez basada en la estructura interna. Psicothema 2014;26:108-116 [FREE Full text]
43. Rogers P. Best practices for your confirmatory factor analysis: a JASP and lavaan tutorial. Behav Res Methods 2024 Mar 13. [doi: 10.3758/s13428-024-02375-7] [Medline: 38480677]
44. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model Multidiscip J 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
45. Satorra A, Bentler PM. Corrections to test statistics and standard errors in covariance structure analysis. In: Latent Variables Analysis: Applications for Developmental Research: Sage Publications; 1994:399-419 URL: https://psycnet.apa.org/record/1996-97111-016 [accessed 2024-08-17]
46. Streiner DL, Norman GR, Cairney J. Health Measurement Scales: A Practical Guide to Their Development and Use: Oxford University Press; 2015. [doi: 10.1093/med/9780192869487.001.0001]
47. Nunnally JC, Bernstein IH. Psychometric Theory: McGraw-Hill; 1994. URL: https://books.google.co.in/books/about/Psychometric_Theory.html?id=r0fuAAAAMAAJ&redir_esc=y [accessed 2024-08-20]
48. Bujang MA, Baharum N. A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: a review. Arch Orofac Sci 2017;12:1-11 [FREE Full text]
49. de Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol 2006 Oct;59(10):1033-1039. [doi: 10.1016/j.jclinepi.2005.10.015] [Medline: 16980142]
50. Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007 Jan;60(1):34-42. [doi: 10.1016/j.jclinepi.2006.03.012] [Medline: 17161752]
51. Cohen J. Set correlation and contingency tables. Appl Psychol Meas 1988 Dec;12(4):425-434. [doi: 10.1177/014662168801200410]
52. Huang Z, Mou J. Gender differences in user perception of usability and performance of online travel agency websites. Technol Soc 2021 Aug;66:101671. [doi: 10.1016/j.techsoc.2021.101671]
53. Nunes A, Limpo T, Castro SL. Acceptance of mobile health applications: examining key determinants and moderators. Front Psychol 2019 Dec 10;10:2791. [doi: 10.3389/fpsyg.2019.02791] [Medline: 31920836]
54. World Medical Association General Assembly. Declaration of Helsinki. Ethical principles for medical research involving human subjects. World Med J 2008;54 [FREE Full text]

55. Mohamad Marzuki MF, Yaacob NA, Yaacob NM. Translation, cross-cultural adaptation, and validation of the Malay version of the System Usability Scale Questionnaire for the assessment of mobile apps. JMIR Hum Factors 2018 May 14;5(2):e10308. [doi: 10.2196/10308] [Medline: 29759955]
56. Mustafa N, Safii NS, Jaffar A, et al. Malay version of the mHealth App Usability Questionnaire (M-MAUQ): translation, adaptation, and validation study. JMIR Mhealth Uhealth 2021 Feb 4;9(2):e24457. [doi: 10.2196/24457] [Medline: 33538704]
57. Albrink K, Schröder D, Joos C, Müller F, Noack EM. Usability of an app for medical history taking in general practice from the patients' perspective: cross-sectional study. JMIR Hum Factors 2024 Jan 5;11:e47755. [doi: 10.2196/47755] [Medline: 38180798]
58. Tacchino A, Brichetto G, Zaratin P, Battaglia MA, Ponzio M. Multiple sclerosis and rehabilitation: an overview of the different rehabilitation settings. Neurol Sci 2017 Dec;38(12):2131-2138. [doi: 10.1007/s10072-017-3110-7] [Medline: 28914385]
59. Waskowiak PT, de Jong BA, Uitdehaag BMJ, et al. Don't be late! Timely identification of cognitive impairment in people with multiple sclerosis: a study protocol. BMC Neurol 2024 Jan 13;24(1):26. [doi: 10.1186/s12883-023-03495-x] [Medline: 38218777]

Abbreviations

CFA: confirmatory factor analysis
DIGICOG-MS: Digital Assessment of Cognitive Impairment in Multiple Sclerosis
EDSS: Expanded Disability Status Scale
ICC: intraclass correlation coefficient
ita-MAUQ: Italian version of the mHealth App Usability Questionnaire
ita-MAUQ_E: ease of use subscale of the Italian version of the mHealth App Usability Questionnaire
ita-MAUQ_I: interface and satisfaction subscale of the Italian version of the mHealth App Usability Questionnaire
ita-MAUQ_U: usefulness subscale of the Italian version of the mHealth App Usability Questionnaire
MAUQ: mHealth App Usability Questionnaire
MAUQ_E: ease of use subscale
MAUQ_I: interface and satisfaction subscale
MAUQ_U: usefulness subscale
mHealth: mobile health
MS: multiple sclerosis
SUS: System Usability Scale

Edited by C Jacob; submitted 05.03.24; peer-reviewed by A Müller, O Ekmekci, P Moorthy; revised version received 04.06.24; accepted 07.06.24; published 30.09.24.

Please cite as:
Podda J, Grange E, Susini A, Tacchino A, Di Antonio F, Pedullà L, Brichetto G, Ponzio M
Italian Version of the mHealth App Usability Questionnaire (Ita-MAUQ): Translation and Validation Study in People With Multiple Sclerosis
JMIR Hum Factors 2024;11:e58079
URL: https://humanfactors.jmir.org/2024/1/e58079
doi:10.2196/58079

Original Paper

Exploring the Role of Complexity in Health Care Technology Bottom-Up Innovations: Multiple-Case Study Using the Nonadoption, Abandonment, Scale-Up, Spread, and Sustainability Complexity Assessment Tool

Ulla Hellstrand Tang[1,2*], PhD; Frida Smith[3,4*], PhD; Ulla Leyla Karilampi[5], PhD; Andreas Gremyr[5,6], PhD

[1]Department of Prosthetics and Orthotics, Sahlgrenska University Hospital, Gothenburg, Sweden

[2]Department of Orthopaedics, Institute of Clinical Sciences, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden

[3]Regional Cancer Centre West, Gothenburg, Sweden

[4]Department of Technology Management and Economics, Collaborative Plattform for Healthcare Improvement, Chalmers University of Technology, Gothenburg, Sweden

[5]Department of Schizophrenia Spectrum Disorders, Sahlgrenska University Hospital, Gothenburg, Sweden

[6]Jönköping Academy for Improvement of Health and Welfare, School of Health and Welfare, Jönköping University, Jönköping, Sweden

[*]these authors contributed equally

Corresponding Author:
Ulla Hellstrand Tang, PhD
Department of Prosthetics and Orthotics, Sahlgrenska University Hospital
Falkenbergsgatan 3
Gothenburg, SE-412 85
Sweden
Phone: 46 706397913
Fax: 46 31408162
Email: ulla.tang@vgregion.se

Abstract

Background: New digital technology presents new challenges to health care on multiple levels. There are calls for further research that considers the complex factors related to digital innovations in complex health care settings to bridge the gap when moving from linear, logistic research to embracing and testing the concept of complexity. The nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) framework was developed to help study complexity in digital innovations.

Objective: This study aims to investigate the role of complexity in the development and deployment of innovations by retrospectively assessing challenges to 4 digital health care innovations initiated from the bottom up.

Methods: A multicase retrospective, deductive, and explorative analysis using the NASSS complexity assessment tool LONG was conducted. In total, 4 bottom-up innovations developed in Region Västra Götaland in Sweden were explored and compared to identify unique and shared complexity-related challenges.

Results: The analysis resulted in joint insights and individual learning. Overall, the complexity was mostly found outside the actual innovation; more specifically, it related to the organization’s readiness to integrate new innovations, how to manage and maintain innovations, and how to finance them. The NASSS framework sheds light on various perspectives that can either facilitate or hinder the adoption, scale-up, and spread of technological innovations. In the domain of condition or diagnosis, a well-informed understanding of the complexity related to the condition or illness (diabetes, cancer, bipolar disorders, and schizophrenia disorders) is of great importance for the innovation. The value proposition needs to be clearly described early to enable an understanding of costs and outcomes. The questions in the NASSS complexity assessment tool LONG were sometimes difficult to comprehend, not only from a language perspective but also due to a lack of understanding of the surrounding organization’s system and its setting.

Conclusions: Even when bottom-up innovations arise within the same support organization, the complexity can vary based on the developmental phase and the unique characteristics of each project. Identifying, defining, and understanding complexity may not solve the issues but substantially improves the prospects for successful deployment. Successful innovation within complex organizations necessitates an adaptive leadership and structures to surmount cultural resistance and organizational impediments.

A rigid, linear, and stepwise approach risks disregarding interconnected variables and dependencies, leading to suboptimal outcomes. Success lies in embracing the complexity with its uncertainty, nurturing creativity, and adopting a nonlinear methodology that accommodates the iterative nature of innovation processes within complex organizations.

(JMIR Hum Factors 2024;11:e50889) doi:10.2196/50889

KEYWORDS

digital; bottom-up innovation; complexity; eHealth; health care; nonadoption, abandonment, scale-up, spread, and sustainability complexity assessment tool; NASSS-CAT; mobile phone

Introduction

Why Is it so Difficult to Develop and Spread New Innovative Technologies in Health Care?

There has been an increasing focus on innovation and the role of new technologies (eg, electronic health records, smartphones, and health applications) in health care. However, developing new technologies comes with significant challenges. Studies show that technology projects in health care, particularly large and complex projects, have a high rate of failure and seldom produce the anticipated results [1-5]. Bottom-up innovations in health care are innovations for service delivery that have been developed "from the ground up," often focusing on preventive patient-centered care, typically driven forward by small interdisciplinary groups of professionals and patients [6-9]. As a result, they may not be captured by existing metrics, thus being "invisible" to senior management and policy makers [10].

The challenges when it comes to developing and making use of innovations, such as spreading or implementing new ways of working [11], have been described by many, often as a "knowledge translation or production problem" [12]. Braithwaite et al [13] compared the traditionally dominant linear and causal thinking that characterizes early implementation science and the evidence-based medicine paradigm with features such as those in systems thinking. The linear approach applies simple, orderly processes with cumulative sequences of stages to produce results building on a knowledge of the way things work, making use of predictable relationships between causes and effects. This has helped generate many successes in the past, not least in health care, but it tends to increase rigidity and fail when applied in more complex and messy systems where things change dynamically and are therefore unpredictable [13]. Instead, systems thinking and the related complexity science recognize system characteristics in building an understanding of how best to move forward.

To drive change in a predictable "simple" system where causes and effects are known, a linear, stepwise approach has a greater chance of success. However, complex systems are not only dynamic but also are often described as adaptive (as in complex adaptive systems) in that they are constituted of agents and artifacts that communicate and learn from each other and the surrounding environment, creating opportunities to learn from experience, self-organize, and evolve, making them less predictable systems [14].

Even though the linear approach has previously dominated implementation and development initiatives in health care, many researchers point to the necessity to apply systems thinking and complexity science when developing health care through the innovative use of new technologies, as exemplified in the study by Greenhalgh and Papoutsi [2]. As the concept of stepwise, linear cause and effect is not sufficient when studying complex systems that evolve in ways that are impossible to predict, it is relevant to use the knowledge of complex systems when understanding and studying health services [15]. Complex systems are defined by (1) intricate intertwined processes, (2) interconnectivity between systems, (3) interconnectivity between levels within systems, and (4) interconnectivity between actors and elements, giving complex systems different properties from those of less complex systems [1]. In short, a complex system does not work linearly but dynamically, with fundamentally different logics [16], and needs to be addressed and understood accordingly during innovation and implementation. If not, there is a risk that new technology and innovations will further increase the complexity rather than actually supporting the needs and demands for an improved health care system [17].

The Challenges

A total of 4 bottom-up innovators found that there was a need to gain insights into the complexity involved in developing and executing bottom-up innovations in a complex health care organization. All 4 innovators had met with hindrances preventing them from moving forward with their innovations. It was necessary to pause and retrospectively try to comprehend the underlying reasons for the stagnation in the 4 cases in question.

Project representatives, all health care professionals, joined forces to identify challenges by assessing project complexity to increase an understanding of the role of complexity and find ways to explore and assess it. As they all worked within the same regional system, it was crucial to involve regional stakeholders (support functions) during the learning process.

The aim of this study was to investigate the role of complexity in the development and deployment of innovations by retrospectively assessing the challenges to 4 digital health care innovations initiated from the bottom up.

Methods

This section describes the theoretical framework that underlies our methodological approach, the settings, and the 4 cases under study, as well as the procedure.

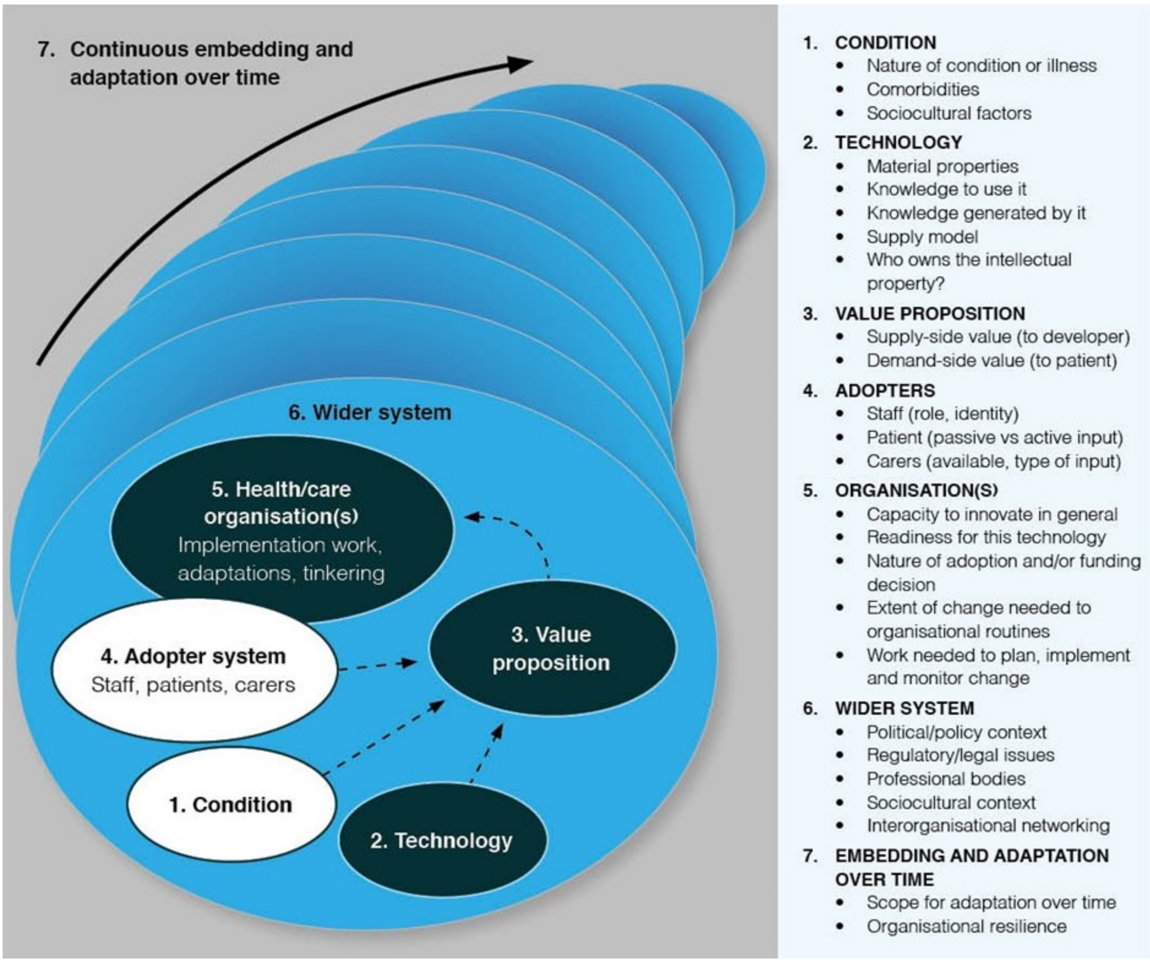
Theoretical Framework

An impressive amount of knowledge related to the diffusion of innovations and their implementation in health care by the start of the new millennium is summarized in the extensive review by Greenhalgh et al [11] from 2004. It builds partly on the ideas by Rogers [18] that innovations have characteristics that will affect their diffusion, as well as affect other domains (eg, the readiness of the system for change, the implementation process, the adopter, and the external wider [sociopolitical] context). As innovations in health care were increasingly associated with new technologies, a new review was conducted by Robert et al [19] in 2010, adding more recent data and focusing on the adoption and assimilation of new technologies into health care.

This, along with the high failure rate of health care technology innovation projects, inspired Greenhalgh and colleagues to deepen their knowledge of the diffusion of innovations, with an emphasis on health technology projects. Building on previous work, reviewing the literature, and using empirical studies of technology implementation, they elaborated on and explained domains of importance. This resulted in the nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) framework (Figure 1 [20], published under Creative Commons Attribution 4.0 International License, CC BY).

The NASSS framework was developed into a complexity assessment tool (NASSS-complexity assessment tool [NASSS-CAT]) [20] to help assess the complexity of health technology projects before, during, or after they were finished.

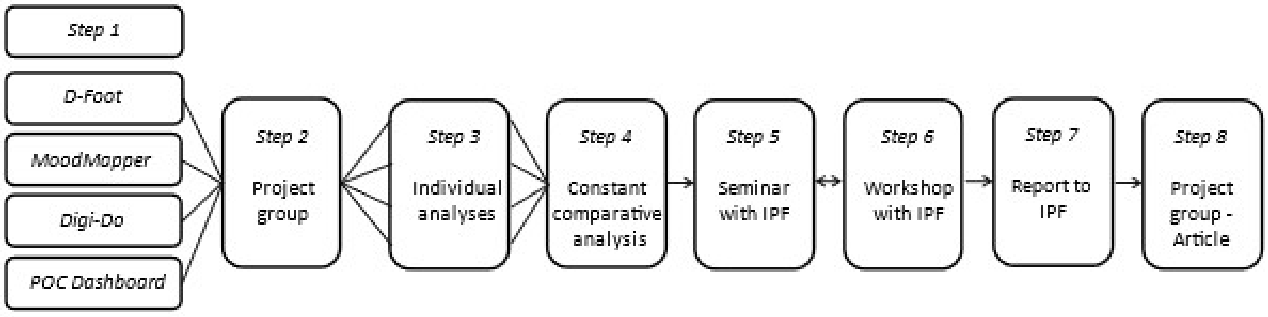
Figure 1. The nonadoption, abandonment, scale-up, spread, and sustainability complexity assessment tool with its 7 domains.



Design and Methodological Approach

A multicase retrospective, deductive, and explorative analysis using the NASSS-CAT LONG was conducted [15,21]. The process of analysis is shown in Figure 2 and is described at the end of the *Methods* section. The complexities of 4 bottom-up innovations developed in Region Västra Götaland (VGR) in Sweden were explored. The NASSS-CAT LONG consists of 2 parts divided into 7 domains (Figure 1). First, one is asked to describe the project and its potential messiness in their own words. Writing this narrative can help surface interdependencies and tricky issues of the project, hence revealing complexity. Second, one answers the questions related to the domain to help them estimate key areas of complexity. One can define whether the question is complex or not complex, whether they do not know, or whether it is not applicable. The total score of orange boxes ticked tells one how complex a certain domain is for their project. In part 2, one is guided through prompts to help them plan for and manage complexity by *reducing* it where possible and *responding* to it if or where it cannot be reduced. The questions can be answered by different people who will provide the needed insights into the domain and the project under evaluation.

Figure 2. Flowchart illustrating the study process when exploring the role of complexity in health care technology projects. This figure presents the different steps (1-8) in the study. Details can be found in the main text. IPF: Innovation Platform.



Setting

The 4 bottom-up innovators were from different parts of the organization in the VGR [22]. All of them had unique experiences of their own departments of the VGR and the surrounding supporting systems, such as the IT departments, purchasing departments, and legal offices. Although all 4 worked in specialized care, the care flow for each of the relevant diagnoses spanned specialized care, primary care, and municipal care.

The VGR is a region on the western coast of Sweden with a total population of 1.8 million. For several years, there has been a push for increasing the development of innovative solutions to the challenges faced by, for example, the public health care system. This has been implemented through the formation of an Innovation Platform (IPF) that develops processes and supportive structures as well as approving funds to help innovative ideas thrive. The mission of the IPF is to contribute to a sustainable innovation system that promotes innovation in health care and ensures that collaboration between academia and business fulfills the needs of patients within the health care system.

In total, 4 bottom-up innovators in the VGR pioneering eHealth and clinical research united in 2019 after realizing that there were barriers to their separate innovations due to an organizational lack of an innovation framework and to inexperience in developing, testing, implementing, and maintaining digital innovations. One author, familiar with the NASSS framework, encouraged the others to explore complexity in innovation. Together, they adapted the NASSS-CAT to identify unique and shared complexities in their innovations. Concurrently, at the same network meeting, representatives from the IPF wanted to be a part of the study, exchanging insights on how to identify and manage complexities in the innovation process.

Ethical Considerations

Because no personal data were collected, no ethics approval was needed. All participants agreed, orally, to take part. No sensitive personal information was collected, and no patients participated in the workshops. The IPF, the regional support resource for innovations in health care, read and commented on the Swedish report before publication.

The 4 Bottom-Up Innovations

Overview

Each innovation is presented in the following sections and in Multimedia Appendix 1 [15,20,23-38]. The cases are heterogeneous with regard to intended user, phase in the innovation process, and place of implementation (locally, regionally, and nationally). Despite their differences, they all existed in the same environment, framed by the regulations of the VGR and its support system for innovation.

Case 1: The D-Foot

The lifetime risk of developing diabetic foot ulcers (DFUs) is as high as 34% for patients with diabetes [23]. It is a burden for the patient and for health care with regard to costs. With prompt prevention, the prevalence of DFUs can be halved [24].

The D-Foot is a digital decision support system designed for preventing DFUs. It conducts early screening and provides treatment recommendations based on a risk grade (ranging from 1=no risk to 4=ongoing foot ulcers) [25,26]. The risk grade is automatically generated through a series of structured foot assessments and patient surveys [27]. A printable report of foot assessments, risk grade, and recommendations is generated. The innovation's reliability and usability have previously been reported and assessed as good [27,28]. The innovator's intention was that the D-Foot would serve as a tool in the national effort to implement a person-centered and seamless care chain for preventing foot ulcers in people living with diabetes [25,26].

The seamless care chain consists of (1) an annual foot examination, (2) a podiatry intervention, (3) the provision of appropriate footwear for at-risk patients, and (4) treatment in a multidisciplinary team for patients with active DFUs [25,39]. The D-Foot was developed as an easy-to-use digital tool to support foot examinations for individuals diagnosed with diabetes, primarily targeting prosthetic and orthotic specialist care [27,28,40]. The goal was to implement the D-Foot nationally, expecting early prevention of DFUs, improved quality of life for affected individuals [29,41], and reduced health care costs [42].

Version 1.0 of the D-Foot software was developed from 2011 to 2016 by an expert group comprising certified prosthetists and orthotists, patient representatives, and orthopedic surgeons in the VGR [27]. Initially, it underwent regional testing with positive results. Thereafter, continuous improvements have been made based on users' comments [28]. Not yet executed is the request from users for integration between the D-Foot and the major medical record system [28].

Case 2: The MoodMapper

Bipolar disorder, often diagnosed in early adulthood, typically necessitates lifelong treatment. It leads to undesirable mood swings affecting daily functioning. Mood episodes vary from extreme “highs” (manic episodes) to severe “lows” (depressive episodes) lasting for days or weeks. Even with proper treatment, mood fluctuations can occur. Collaborative communication between patients and health care providers enhances treatment effectiveness. Moreover, the early detection of behavior changes is of the utmost importance in the successful treatment of bipolar disorder.

The aim of this innovation was to determine whether smartphone use data are a reliable source for studying changes in the digital behavioral patterns of individuals with bipolar disorder by exploring correlations between different parameters of smartphone use data.

The MoodMapper is a mobile app that, through real-time data collection, can provide valuable insights into a patient’s smartphone use. The ambition was, after pilot-testing, to study and evaluate the connection between mobile-generated passive data and documented changes in the patient’s mental well-being, with the goal of making it easier for both patients and health care professionals to monitor the progression of the patient’s condition and make decisions regarding prevention and care.

Case 3: The Digi-Do

Radiation therapy (RT) is a common treatment after breast cancer surgery. The high-technology environment and unfamiliar nature of RT can affect the patient’s experience of the treatment. Misconceptions or a lack of knowledge related to RT processes can increase levels of anxiety and enhance feelings of being unprepared at the beginning of the treatment. Moreover, the waiting time is often fairly long. Cancer care involves several, often independent clinics. Even if the clinical pathway is clearly described, transitions and information exchange can be problematic. RT is only provided at the university hospital in the region, with long distances and long waiting times for many patients.

The Digi-Do tool consists of two separate mobile apps: (1) an app providing a guided digital tour of the RT department, where the patient can familiarize themselves with the department by using virtual reality glasses; and (2) an app with additional information, including questions and answers, practical information, and short animated films about the RT process. The design of both apps was developed in a co-design process with patients and staff [43]. The primary aim of the researcher or innovator was to evaluate whether a digital information tool with virtual reality technology and preparatory information was able to reduce distress and enhance the self-efficacy and health literacy of patients with breast cancer before, during, and after RT. A secondary aim was to explore whether the digital information tool increased patient flow while maintaining or improving the quality of care [44].

Case 4: A Point-of-Care Dashboard for Schizophrenia Care (the PoC Dashboard)

The Department of Schizophrenia Spectrum Disorders at Sahlgrenska University Hospital delivers specialized care for people with psychotic disorders in the metropolitan Gothenburg area (with a population of approximately 600,000 people) in Sweden. Schizophrenia is the most common diagnosis among the approximately 3000 patients who receive care at the department’s 7 outpatient units. Approximately 20% of these patients also need acute inpatient care at 1 of the department’s 4 wards each year.

With the aim of supporting patient coproduction of health, a digital dashboard was developed to be jointly reviewed at the point of care by patients and case managers and psychiatrists to support evaluation and planning, outcome questionnaires, and patients’ care plans [45]. The dashboard was developed between 2016 and 2018 and was piloted at 2 outpatient units with approximately 400 patients for 18 months. The dashboard is one of several connected applications and displays for visualizing data fed by multiple systems to support, for example, planning, management, and triage and includes a unit-level overview of quality indicators to identify patients at risk. The dashboard project also served as a case in the development of the NASSS-CAT [20].

Study Procedure

This study followed an iterative process, including analyses, discussions, and seminars (Figure 2).

As we used the NASSS-CAT, the analysis was deemed to be of an exploratory, deductive nature. First, an individual assessment of each case was made, and then the 4 cases were compared to find similarities. However, while using the NASSS-CAT in each of the 4 cases, the innovators had discussions about how to interpret the questions in the 7 domains. An additional method, namely, constant comparative analysis (CCA), was chosen as it is appropriate in collaborative projects to facilitate and identify agreements and disagreements [46,47] (Multimedia Appendix 2). After agreeing on how to interpret the questions, members of the IPF were invited to complete the analysis in a workshop.

In total, 4 bottom-up innovators had individually experienced complexities during their respective innovation processes from 2010 to 2019.

1. The 4 innovators got together and started the study in January 2020 by learning how to use the NASSS framework based on the work by Greenhalgh et al [15]. Support was available as one of the authors had been involved in the development of the NASSS-CAT [20].
2. The innovators identified complexities in their individual projects using the NASSS-CAT [20] in 2020.
3. During >30 one-hour meetings using the NASSS-CAT and taking minutes, the innovators identified, compared, and discussed similarities and differences regarding complexities in their respective innovations. A CCA was included in the process and is described in Multimedia Appendix 2.

4. A seminar was held in April 2020 with one of the bottom-up innovators and the IPF presenting the concept of complexity and the NASSS framework.
5. In October 2020, another seminar was held with the 4 innovators and staff members from the IPF to discuss the domains of the NASSS-CAT, illustrated by examples and findings from the assessment of the 4 bottom-up innovations. The participants from the IPF discussed and reflected on experiences of complexities. The seminar was recorded and summarized in a report in collaboration with a representative from the IPF [48].
6. The authors were commissioned to write a report (in Swedish) for the IPF exploring and summarizing the NASSS framework and complexity with examples from the bottom-up innovations [48].
7. The insights gained into the role of complexities from the entire aforementioned process were discussed and summarized and are presented in this paper.

Results

In this retrospective exploration of the role of complexity in 4 bottom-up health care innovations in a Swedish region, both similarities and differences emerged among the 4 cases when using the NASSS-CAT (Multimedia Appendix 3). The findings for each domain are described in the following sections.

Complexity Domain 1: The Illness or Condition

This domain has no or low complexity when the illness is well known and an assessment can result in a well-defined diagnosis and when there is, furthermore, knowledge and know-how regarding how to treat the condition successfully. Complexity can be related to conditions with less known causes, a high prevalence of multimorbidity, and challenging sociocultural factors (eg, language barriers). In our study, the cases that addressed mental illness (PoC Dashboard and MoodMapper) and diabetes (D-Foot) had more complexity related to the actual illnesses than the case aiming to prepare women before RT for breast cancer (Digi-Do) [49,50]. Both bipolar disorder and schizophrenia are strongly connected with comorbidity and lifestyle-related conditions, and even though national guidelines exist, there is no simple pathway to treat those conditions. The third case (D-Foot) involved diabetes, also an illness defined as complex due to the patient being treated in various institutions and with several lifestyle factors influencing the outcome of the treatment [39].

Complexity Domain 2: The Technology (or Other Innovations)

An innovation is less complex if it is well known, ready, and easy to use and has a clear supply model, well-defined ownership regarding its intellectual properties, and low or no dependency on other systems. For the actual technologies under study, similarities regarding complexities revolved around interdependencies with other IT systems, ranging from local to regional and even national systems. Even if the technology already existed (D-Foot) or if new software was developed to create a better overview of data in several existing systems (PoC Dashboard), it was difficult to develop the innovation so that it enabled adoption beyond the local settings. Regulations regarding software used as a medical device [30] sometimes prevailed over the simple adaptations for different target groups. “Fireproof” walls exist between organizations (eg, municipal care, primary care, and specialist care), and different versions of the regional information systems, being related to ownership, budget, and management, make it less clear whether and how new technologies can be bought, adapted, and used in a local setting. In contrast, the Digi-Do app does not require any interaction with existing IT systems and was not deemed to be a medical technical device. The MoodMapper app, on the other hand, is complex as it aims to interact with both patients and health care staff, requiring interaction with medical electronic health records as well as ensuring a very high level of security to safeguard the patients’ integrity [51]. The need for supply chains included both purchasing and procurement and clinical implementation, the latter involving questions regarding intellectual properties (is the owner the bottom-up innovator or is it the region?), ownership management (which regulation steers the region when managing a medical device owned by the region?), and updates and maintenance of the eHealth tools (which department in the VGR is responsible for updates and maintenance of the innovations?). All the cases had run into or expected to run into severe complexity when planning to launch their innovations. It was clear that complexity regarding supply chains had not been considered by either the innovator or the VGR when intending to expand from a local level to a regional or national one. An example of the complex challenges related to the spread and maintenance of one of the eHealth tools, the D-Foot, is presented in the following paragraphs.

Regulations regarding funding and ownership made it difficult to implement a supply chain outside the local region as each of Sweden’s 21 regions has its own procurement processes. Since the start, the D-Foot project had been aiming for national spread. In 2017, the IPF approved funding with the aim of testing, Conformité Européenne (CE) marking, and thereafter implementing the D-Foot first in the VGR at the department of prosthetics and orthotics and then nationally. At the same time, several departments of prosthetics and orthotics in other regions were interested in using the D-Foot as soon as the CE marking was finalized. However, in June 2017, an official at the VGR decided that the region was only able to allow the D-Foot to be used within the region (Article 5.5 in the Regulation [European Union] 2017/745) [30]. Following this decision, national spread was impossible. The bottom-up innovator continued to have a dialogue with the IPF seeking a solution for national spread. In 2020, an opportunity for national spread arose by registering the D-Foot as a national medical information system (NMI) at the Medical Products Agency. An NMI is an information system developed for joint use at nationwide, regional, or municipal level in Sweden.

Thus, the D-Foot transitioned from being a “self-manufactured medical device” to becoming an NMI registered with the Swedish Medical Products Agency in 2020. However, the NMI registration was withdrawn by the VGR in 2021 due to new regulations from the Swedish Medical Products Agency [52]. The D-Foot remains a separate software program not integrated into the standard medical record system in the VGR. As a result, one option remained for national spread, namely, to CE mark the D-Foot, a procedure that was not as yet allowed or tested in the region.

Complexity Domain 3: The Value Proposition (Costs and Benefits of the Technology)

Complexity in this domain arises when determining the value provided by the innovation to developers, users (patients, staff, and health care systems), and the broader health care ecosystem. Despite their origin as bottom-up innovations aimed at improving care, the complexities of demonstrating supply-side value in terms of business models and monetary benefits were challenging. Key questions included defining improved value, decision-making processes, the inclusion of nonmonetary values, and the extent of evaluation required: what is regarded as improved value? Who decides? Is it only monetary or other types of value as well? How much does the innovation need to be evaluated and how?

The Digi-Do has a defined regional vision and has faced challenges in quantifying value, especially regarding soft values such as reduced distress and increased health literacy and self-efficacy. The Digi-Do aims to optimize the use of waiting time before RT, adding value by reducing waiting times and queues. Furthermore, the innovation aims to create value for patients by delivering information in a novel, accessible format, potentially improving health literacy even for those with language difficulties or cognitive impairments. It also extends benefits to the patient's social network, enhancing support and knowledge and reducing distress among family and friends affected by the patient's cancer diagnosis. By using the often idle waiting time for meaningful preparation, the innovation may foster a sense of control and inclusion, diminishing distress and worry. Well-prepared patients may navigate the system more efficiently, potentially reducing waiting times for information dissemination.

The evaluation of the outcomes in this specific project is still ongoing through an unpublished randomized controlled trial [44], but so far, the qualitative results show a high level of acceptance of and positivity toward the tool. Nevertheless, there needs to be a discussion about how to endorse a more pragmatic evaluation of both effectiveness and process outcomes [53].

If successful, this approach could be adapted for other health care domains, although commercialization is not the project's primary goal. Measuring soft values has proved challenging as they might not directly impact traditional health care outcomes. The other cases faced similar difficulties in pinpointing the exact stages in which costs and values could be calculated.

Enhancing foot health can improve the quality of life of patients and reduce health care costs associated with treating DFUs and amputations. Objective risk assessment by using the D-Foot precedes interventions, aligning with the vision of providing equal, high-quality care to citizens. Early interventions in the prevention process (D-Foot) might require more resources within primary care but were expected to be cost-effective in the long term due to a reduction in specialist care following fewer ulcer treatments and amputations [31,42]. In terms of quality of life and cost reduction, the value proposition needs further evaluation over a longer period relying on data related to care costs for at-risk patient groups. The D-Foot database contains valuable information on risk groups and foot status, serving as a data source for audits and evaluations to optimize foot care. It could also function as a quality registry, potentially becoming the new diabetic foot register in Sweden.

The MoodMapper aims to provide a more objective risk analysis and early interventions, potentially preventing hospitalization. In the examples of innovations (MoodMapper and PoC Dashboard) designed to prevent relapses in severe mental illness by coordinating data or even by asking patients to send and react to data, the need for hospital care could be reduced. However, this area is as yet unexplored.

The value proposition of the PoC Dashboard remains uncertain. Case managers and patients find the technology useful based on preliminary data. Local testing and piloting suggest perceived effectiveness, although the degree of cost-effectiveness is still unknown. The dashboard streamlines administrative tasks for staff, offering an overview of patients' progress and risks while facilitating collaborative care planning. However, the technology's potential as a commercial product is uncertain, mainly because it is integrated with older systems. Additional uncertainties involve the IT department's role in dashboard maintenance and associated costs.

Complexity Domain 4: The Intended Adopters of the Innovation and Technology

Complexity in this domain is higher when adopting the innovation, necessitating changes in routines, roles, and identities. Innovations that support existing routines with minimal disruption are associated with lower complexity, and all 4 cases required either behavior changes by patients or modifications to work routines for health care staff. For example, the PoC Dashboard simplified patient overview and reduced administrative work for staff, thus positively impacting daily tasks [54].

However, transferring the D-Foot to primary care posed challenges as different health care professionals (podiatrists, nurses, and physicians) with varying roles and routines questioned its added value. The MoodMapper required patients to trust the handling of their behavioral data, which could be challenging for those with symptoms of paranoia.

Complexity Domain 5: The Organization Implementing the Technology

Complexity in this domain pertains to the efforts required to plan, implement, and monitor the innovation's adoption, as well as to the organization's overall capacity for innovation. Challenges included a lack of clear pathways for support, making it necessary to find the right individuals at the right levels for consultations. Different organizational levels faced varying complexities, and despite a desire for innovation, built-in regulations sometimes hindered dissemination. For instance, regulatory obstacles prevented the national spread of the D-Foot.

Complexity Domain 6: The External Context for Innovation

Complexity in this domain is influenced by the political, sociotechnical, and regulatory context, as well as by stakeholder groups and interorganizational networking. In Sweden, despite a national Vision for eHealth by 2025 [55], the existence of 21 independent regions creates complexity in decision-making for local, regional, and national development and for the implementation of digital health care innovations. For instance, there is a national initiative from the government to improve cancer care, but the regions are self-governed in terms of budget and implementation. This means that, even if the regional cancer center had a national assignment to improve cancer care generally and the RT process in the local region specifically, it has no mandate to implement the Digi-Do without the approval of the RT department at each separate regional hospital.

Furthermore, if an innovation needs to be integrated with the IT systems, such as in the other 3 cases, national initiatives can be ruled out by regional procurement, management supply chains, and European regulations regarding medical devices [30]. If regional support and the management of supply chains only permit regional use, there will be no dissemination, and thus, bottom-up innovations risk becoming only local or, at worst, experiencing the "death of innovations" after the initial project phase.

Complexity Domain 7: Emergence Over Time

Complexities were identified in all 4 cases (Multimedia Appendix 3). When summarizing the complexities from domains 1 to 6, all the authors concluded that the complexities were likely to increase in the coming 3 to 5 years, probably due to advances in technology, unexpected events such as pandemics, international conflicts, and new regulations and standards. In the coming years, a new regional medical record system, Millennium, is planned to be implemented. For small bottom-up innovators, it is not yet clear how the implementation of Millennium will affect their innovations [37].

Discussion

Principal Findings

Overview

In this study, we conducted a retrospective, deductive, and exploratory analysis of 4 cases using the NASSS-CAT LONG. We intended to explore whether there were shared or individual challenges related to bottom-up innovation projects in the same health care region. The analysis itself was complex, but it resulted in both common and individual learning, and all but one case have moved forward, partly due to new insights gained that have made progress possible. By applying the NASSS-CAT in various projects, the authors learned several lessons, the most important of which are described in the following sections. After that, we discuss and reflect on the methodology and the need or suggestions for further research. Finally, we briefly present how the cases have developed since the analysis.

The Innovation Versus the System

As proposed by Rogers [18], the properties of the innovation affect its probability of diffusion within and beyond the organization. Through this study, we have become aware of the need to understand the "system" in which we are working to develop and adopt innovations and make effective use of those innovations. The NASSS framework sheds light on various perspectives that can either facilitate or hinder the adoption, scale-up, and spread of technological innovations. Before our projects, none of us had fully considered all these perspectives. During this study, complexity was found and highlighted, involving many issues related to the organization or system rather than the specific innovation itself. Multiple regulations must be considered, and regional procurement [56], management of supply chains, and European regulations regarding medical devices [30] can hinder the spread of innovation. A lack of necessary interorganizational networking further complicates matters.

Linear Logic Versus Dynamic Complex Processes

By applying the NASSS framework, we discovered how the innovation process and the training we had all had in evidence-based medicine and the research process were geared toward a linear process rather than embracing complexity. We also discovered that the complexity was mostly found outside the actual innovation and related more to the system that the innovation was supposed to live in and to regulations and legislation. More specifically, it related to the organization's willingness to integrate new innovations and to questions regarding how to manage, maintain, and finance innovations.

Developing and deploying new bottom-up innovations in health care involves multiple logics [57]. Initially, we attempted to approach this in a traditional linear fashion with sequential steps from idea to widespread adoption. However, we quickly realized that this linear approach did not align with the reality of navigating the complexities of health care innovation. Instead of a straightforward innovation journey, it often felt like traversing a dense jungle, making it challenging, if not impossible, to gain a comprehensive overview of the landscape, identify opportunities, and predict the appropriate course of action.

Complex environments often require creative and dynamic thinking; in contrast, a linear approach may stifle the ability to respond to unexpected challenges or opportunities. Innovation is inherently uncertain and unpredictable [58]. It often involves trial and error, experimentation, and the willingness to explore unconventional ideas. A rigid stepwise approach may not accommodate the iterative and nonlinear nature of the innovation process. Complex organizations involve numerous interconnected variables and dependencies. A linear approach may overlook these interconnections, leading to suboptimal solutions or unintended consequences. Innovation often requires a holistic understanding of the organization's ecosystem. This understanding is hindered if established cultures in the complex organizations are resistant to change [57]. A nonlinear approach may face resistance from employees or departments unwilling to deviate from established norms.

Successful innovation requires addressing cultural and organizational barriers, which may not fit neatly into a linear plan. Finally, complex organizations require adaptive leadership that can navigate ambiguity and inspire a culture of continuous improvement [59]. These are important findings as innovation, particularly in the realm of new technologies, is often seen as a potential solution to address the challenges facing health care. Calls for innovation and new ways of working have come from various sources, including governments, health care organizations, and life sciences clusters. However, the high failure rate of health care technology projects suggests that there may be deficiencies in the structure, resources, and knowledge needed for success [60]. Furthermore, there is a risk of simplifying the complex innovation process by building a support system that is linear. The linear and stepwise approach (first do this, then do that) is counterproductive. While a linear approach may work in certain situations, the nature of innovation in complex organizations demands a more flexible, adaptive, and nonlinear methodology. Embracing uncertainty, fostering creativity, and adapting to change are critical elements that a rigid stepwise approach may not adequately address in the context of complex organizational innovation [58].

Value

For all 4 cases, questions arose related to value and costs. Will there be an initial or a recurrent cost for the product, or will the cost be related to a new service that entails new tasks for staff? There is an advantage in specifying both costs and values, as well as the effect of the innovation on other resources, early in the innovation process. Therefore, health-economy analyses are needed, but they are difficult to design and perform as some innovations focus on increasing soft values that are difficult to translate into monetary variables.

Indeed, evaluating the values—different kinds of values and on different levels—of health care innovation is complex. While clinical testing can demonstrate its usefulness to end users, it is often difficult to determine whether the outcomes involve soft values (eg, reduced distress and improved health literacy and self-efficacy) or hard, monetary values [14]. Furthermore, the distribution of costs and value resulting from an innovation can be intricate, making it hard to assess. Questions arise about the initial and recurrent costs and whether they relate to the product or to new services that require additional staff tasks. Early in the innovation process, there is a need to specify both costs and values, be they monetary or qualitative. Clearly describing and anchoring a value proposition, whether it involves soft or hard values, with stakeholders early in the process is crucial for understanding costs and outcomes. However, finding effective ways to evaluate an innovation before it is ready for large-scale testing can be challenging. Similarly, value and costs stemming from an innovation can be distributed across the organization or organizations in ways that are difficult to assess. Calculating the health costs of improving care processes that involve many actors in a complex organization such as the VGR is complicated [61]. More pragmatic evaluations of both effectiveness and process outcomes are needed and can help show the effect from different angles [53].

For all cases dealt with in this study, the value proposition in terms of quality of life and cost reduction needs further evaluation over a longer period relying on data related to care costs for at-risk patient groups. The D-Foot database contains valuable information on risk groups and foot status, serving as a data source for audits and evaluations to optimize foot care. It could also function as a quality registry, a new diabetic foot register in Sweden. In the MoodMapper, the users comprise patients; their clinical teams; and, occasionally, relatives or caregivers. A published study highlights the value of implementing and receiving psychological relapse prevention for these groups, leading to improved understanding of bipolar disorder [62] that might, in turn, lead to enhanced working relationships and better condition management. However, the evidence is not consistent, and further studies are needed [61]. Moreover, for patients with bipolar disorders, having some of their behavioral patterns (such as step count and estimated sleep) automatically monitored meant that there needed to be a great deal of trust in how data are handled, something that might be difficult for patients experiencing symptoms of paranoia.

Co-Design and Coproduction

Involving users both directly and indirectly at an early stage of the development process is highly beneficial, particularly because what benefits one person may pose challenges for another, thereby creating complexity. Although there are several examples of how coproduction is useful in the innovation process, the existence of complexity must not be neglected in the co-design.

Bottom-up innovations in care encompass a wide spectrum of patient-centric approaches, empowering individuals and communities to actively participate in projects aiming to support well-being. These innovations, driven by the challenges that health care faces, range from self-management tools [62-64] and patient support networks to community-driven health programs [6-8,10,59-61,65-67]. They appear with different approaches, such as lean production [9] and Six Sigma [68]. Coproduction can enhance the 3 Rs in research—reach, rigor, and relevance [69]—by ensuring that the right needs are addressed and that the innovation is practical for both patients and staff.

Enthusiastic innovators and staff should be engaged early in the process, along with representatives from patient organizations or individuals with relevant experience. There is a strong movement toward involving patients in health care improvement, and genuine engagement is necessary for truly bottom-up innovation involvement [70] as it can lead to more radical solutions or suggestions when used correctly [71]. If a technological innovation is too demanding or unfamiliar for users, it is unlikely to be accepted. Piloting with stakeholders is crucial for assessing practicality [59], and using input from stakeholders in the right phase can increase the possibility of finding radical suggestions, as well as saving time for both parties (developers and patients) [71]. We support the idea that coproduction incorporating the multifaceted aspects of complexity is necessary in the evaluation of success in the implementation of bottom-up innovations [4].

Methodological Considerations

Performing a retrospective, deductive analysis as a case study [21] with 4 cases with differences regarding where they were in the innovation process and with different technical solutions was challenging, but it provided multiple valuable insights. The authors found that, before using the NASSS-CAT, users need to be familiar with the NASSS framework [15]. The NASSS-CAT appeared deceptively easy at first, but it was more difficult to use and more time-consuming than expected. A need for a way to track how we could jointly understand and agree on the meaning of the NASSS-CAT by using CCA became apparent during the work, leading to a common language being agreed upon and a consensus being reached on how to interpret the terminology used in the NASSS-CAT. During the CCA, discussions about how to interpret the questions in the 7 domains of the NASSS-CAT took place in cycles, and thus, it was a continuous learning process. As intended by the method [46,47], finding disagreements and negotiating led to a higher degree of understanding not just of the instrument but also of the concept of complexity. The 4 innovators contributed multiple perspectives based on their own cases and discussed their different understandings of the narratives, the domain questions, and the subquestions. As the authors used a nontranslated version of the NASSS-CAT and are native speakers of Swedish and not English, the CCA helped them understand the questions in the NASSS-CAT. Therefore, the use of CCA statements and negotiations on how to interpret the questions in the NASSS-CAT facilitated the analysis and helped create a common language within the group.

The NASSS framework was developed through a detailed review of the existing literature and clinical cases [15,20], but to our knowledge, the tools (NASSS-CAT) have so far been sparsely tested for their ability to unveil complexity in bottom-up projects in public health care. Going from commonly used methods for quality improvement (eg, using the Plan-Do-Study-Act method [72] to incorporate complexity assessment) shows promising results. A recent study used the framework and tool combined with the Plan-Do-Study-Act cycles of improvement to plan and evaluate digital services for patients in Sweden [73]. Similar to our retrospective analysis, that study identified several elements of complexity, explaining a gap among the capacity of adopters, the organization, the wider system, and how intended users valued the service. This gap hindered the innovations from integrating new services into routine care effectively [73]. Similarly to us, these authors found the tool and framework helpful in that they allowed for deeper insights into the project compared to only following method, approach, or cycles or other tools or models for innovation. It seems that, even if complexity is revealed early in the process, this still does not solve the problems. However, if people working with innovation or in supporting innovation become more aware of complex elements, issues might be easier to anticipate or even deal with earlier. Such awareness can thereby help explain obstacles and prevent failure, hence enabling more successful innovation projects in health care, as presented by Greenhalgh et al [15].

Strengths and Limitations

The strength of this study lies in the 4 different cases representing both somatic and psychiatric care and the innovators' long experience in both health care and eHealth. The diversity of innovations presented and the different departments that each of the innovators worked in contribute to a broad overview of shared experiences. None of the innovators had worked together before this study. The fact that all cases came from the same region with the same support function strengthens our results by showing that (1) knowledge of complexity needs to be improved in such systems and (2) the project itself contains complexity in different domains even if we found several common problems. Therefore, the study increased the understanding of the role of complexity, not only in the studied bottom-up innovations but also in the system in which the innovations took place, through prolonged engagement [50]. This study was strengthened by the support of one of the authors, who was involved in the development of the NASSS-CAT [20], but despite this, it appears that adaptation to the setting (geographic and cultural) is crucial.

The retrospective NASSS-CAT analysis of 3 of the 4 cases was mainly performed by the respective innovators without direct input from stakeholders involved in each of the cases. This meant that only 1 perspective from the many actors involved in each of the projects was put forward. The rationale for this was that each innovator had already faced and, therefore, was acquainted with the diverse complexities addressed in all 7 domains. However, other perspectives might have further improved the analysis. The PoC Dashboard project was assessed regarding complexity in a workshop with stakeholders and discussed with management [54].

Even though the NASSS-CAT tools have been used previously [74], more testing in clinical bottom-up innovation cases is needed to scrutinize their utility in a Swedish setting and to learn from the experiences originating from 4 different cases that used the tools.

Use and Usefulness of the NASSS-CAT

In this section, experiences of the use and utility of the NASSS-CAT are presented. At the start, the 4 bottom-up innovators were naïve and expected the NASSS-CAT [20] to be easy to comprehend and use as they identified complexities in their own innovations. As mentioned previously, by using the CCA interpretations from multiple perspectives (the 4 cases), a shared understanding and language regarding how to interpret the questions in the NASSS-CAT was established. The results from the CCA revealed that each of the 4 innovators needed to clarify or consider a number of points in their own NASSS-CAT analysis while assessing complexities in each of the domains. The most important issues were as follows:

1. To define the time frame and the scope that the innovator is assessing.
2. To define the intended users and adopters at the time of the studied project.
3. To rethink the way in which the value proposition can be measured.

4. To consider that questions in the NASSS-CAT regarding ownership; supply chains; and use and spread at the local, regional, national, and international level belong to both "Domain 5: organization" and "Domain 2: technology."

Moving Forward in Supporting Bottom-Up Innovation

This study explored insights from the NASSS framework, revealing that the adoption and dissemination of technological innovations are influenced by organizational and systemic factors rather than by the innovations themselves. The success of bottom-up innovators in navigating complexities emphasizes common challenges across innovations. The NASSS framework has illuminated various perspectives that can either facilitate or impede the adoption, scale-up, and dissemination of technological innovations.

The 4 bottom-up innovators managed to navigate through the complexities within the innovative system, uncovering overarching challenges that unified their respective innovations. However, it is essential to recognize that the NASSS-CAT cannot be used as a linear checklist. Existing support systems, while aiming to foster innovation, may unintentionally follow a linear approach rather than embracing frameworks suitable for complex interventions, such as the Medical Research Council guidance [75]. To better support health care innovators, a "midway filter system" is needed, which offers profound insights into innovation within complex systems. Implementing such a filter between top-down and bottom-up approaches would facilitate bidirectional knowledge transfer. It would enable clinical insights, ideas, and innovations to be discussed in harmony with the regulatory framework, ultimately leading to improved and equitable health care as envisioned by Tierney et al [10], who found that localized, regional, and flexible innovations can shape care in the future [10]. Incentives to connect bottom-up initiatives with a top-down vision at a national level in building systems for digital innovation and health IT are presented by Sheik et al [67] from the United Kingdom. We share their vision to improve usability and interoperability and integrate bottom-up with top-down resources.

Our study, similarly to the research by Batalden and Davidoff [76], discusses the complexities of integrating grassroots idea innovations into established health care systems. Batalden and Davidoff [76] highlight the need for organizational changes and a shift in culture to recognize the value of patient-driven innovations and effectively incorporate them into clinical practice.

Future studies should consider a translation project of the NASSS framework from English into Swedish. This would facilitate the framework's use in a Swedish context, similar to the translations of health-related quality of life questionnaires, which follow guidelines to ensure validity in terms of language and culture [53]. In addition, an evaluation is recommended alongside updates to the NASSS-CAT. Some subquestions may benefit from further splitting, such as assessing the likelihood of technology obsolescence or the measurement of alternative ways to evaluate innovation. It is crucial to involve relevant stakeholders in these changes. Cultural adaptation should receive significant emphasis to provide a language that is relevant to the Swedish context. We also suggest that future studies explore similarities and differences regarding the existence of complexities when bottom-up innovations are developed and implemented in other regions. Finally, we consider making a follow-up prospective evaluation of our 4 innovations. By doing this, we can possibly review the impact of this study on the long-term outcomes of each innovation using the NASSS-CAT LONG.

The Progress of the 4 Cases

The insights gained from our exploration of the existence of complexities in innovation processes led to some of the presented innovations being appreciated in the VGR. The Digi-Do and the D-Foot have gradually, during the study, been acknowledged as important in building future care with digital tools. The VGR has granted the innovator of the D-Foot the legal rights to be spread nationally and internationally and to be implemented and scaled up to prevent DFUs through early screening.

The Digi-Do has been evaluated, and the results show a positive effect on the users, indicating reduced levels of distress and an improved sense of preparedness [43,77]. Hence, the difficulties in evaluating soft values have been successfully dealt with. Since the analysis presented in this paper, the intellectual properties have been transferred to the VGR together with the RT department, and updated versions of the Digi-Do are underway.

Learnings from the complexity assessment of the PoC Dashboard [54] helped address the challenges differently by going for a simpler technical solution with less dependencies on other information systems and focusing on core features such as supporting patients and health care professionals in the planning and evaluation of care. This was done by adapting Dialog+, which is both a tool to measure and monitor patient-reported outcome measures and patient-reported experience measures and a solutions-focused methodology, to fit Swedish psychiatric care [78]. It has since then been piloted and tested for >4 different patient groups in mental health care settings and is being implemented as part of routine psychosis care. The MoodMapper is not an active innovation project in Sweden. However, it is used internationally in research to map behavior changes in mental disorders [79,80].

Conclusions

The NASSS framework increased the bottom-up innovators' understanding of the role of complexity in their innovations. The analysis provided valuable insights by identifying and bringing attention to complexities, particularly within the broader system, albeit requiring a deep understanding. This study enriched our comprehension of the pervasive role of complexity in bottom-up innovations within public health care and shed light on the practical utility of the NASSS-CAT. Early use of a validated tool aids in identifying complexities and pinpointing the domains in which these complexities exist. Importantly, even when bottom-up innovations arise within the same support organization, the complexity can vary based on the developmental phase and the unique characteristics of each project. Identifying, defining, and understanding complexity may not solve the issues but substantially improves the prospects for successful innovation implementation provided the right expertise is available to support the process.

Successful innovation within complex organizational structures necessitates a comprehensive understanding and an adaptive leadership to surmount cultural resistance and organizational impediments. A rigid, linear, and stepwise approach risks disregarding interconnected variables and dependencies, leading to suboptimal outcomes. Success lies in embracing the complexity with its uncertainty, nurturing creativity, and adopting a nonlinear methodology that accommodates the iterative nature of innovation processes within complex organizations.

Acknowledgments

The authors are most grateful for the collaboration with the Innovation Platform in Region Västra Götaland. They would also like to acknowledge the Regional Cancer Centre West, the Department of Prosthetics and Orthotics at Sahlgrenska University Hospital, and the Department of Schizophrenia Spectrum Disorders at Sahlgrenska University Hospital for their support.

Data Availability

The data sets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Description of each of the 4 cases of bottom-up innovations in Swedish health care, including a presentation of how the innovations relate to the 7 domains of the nonadoption, abandonment, scale-up, spread, and sustainability complexity assessment tool.
[DOCX File , 25 KB - humanfactors_v11i1e50889_app1.docx]

Multimedia Appendix 2
A constant comparative analysis of how to interpret the 7 domains of the nonadoption, abandonment, scale-up, spread, and sustainability complexity assessment tool.
[DOCX File , 22 KB - humanfactors_v11i1e50889_app2.docx]

Multimedia Appendix 3
Complexities in the 4 bottom-up innovations assessed using the nonadoption, abandonment, scale-up, spread, and sustainability complexity assessment tool.
[DOCX File , 24 KB - humanfactors_v11i1e50889_app3.docx]

References

1. Chan S. Complex adaptive systems. Massachusetts Institute of Technology. 2001. URL: https://web.mit.edu/esd.83/www/notebook/Complex%20Adaptive%20Systems.pdf [accessed 2024-04-08]
2. Greenhalgh T, Papoutsi C. Studying complexity in health services research: desperately seeking an overdue paradigm shift. BMC Med 2018 Jun 20;16(1):95 [FREE Full text] [doi: 10.1186/s12916-018-1089-4] [Medline: 29921272]
3. Tsoukas H. Don't simplify, complexify: from disjunctive to conjunctive theorizing in organization and management studies. J Manage Stud 2016 Jul 11;54(2):132-153. [doi: 10.1111/joms.12219]
4. Jacob C, Lindeque J, Klein A, Ivory C, Heuss S, Peter MK. Assessing the quality and impact of eHealth tools: systematic literature review and narrative synthesis. JMIR Hum Factors 2023 Mar 23;10:e45143 [FREE Full text] [doi: 10.2196/45143] [Medline: 36843321]
5. Eldh AC, Sverker A, Bendtsen P, Nilsson E. Health care professionals' experience of a digital tool for patient exchange, anamnesis, and triage in primary care: qualitative study. JMIR Hum Factors 2020 Dec 14;7(4):e21698 [FREE Full text] [doi: 10.2196/21698] [Medline: 33315014]
6. Klundert JV, de Korne D, Yuan S, Wang F, van Wijngaarden J. 'Hybrid' top down bottom up health system innovation in rural China: a qualitative analysis. PLoS One 2020;15(10):e0239307 [FREE Full text] [doi: 10.1371/journal.pone.0239307] [Medline: 33027287]
7. Lambert AS, Ces S, Malembaka EB, Van Durme T, Declercq A, Macq J. Evaluation of bottom-up interventions targeting community-dwelling frail older people in Belgium: methodological challenges and lessons for future comparative effectiveness studies. BMC Health Serv Res 2019 Jun 24;19(1):416 [FREE Full text] [doi: 10.1186/s12913-019-4240-9] [Medline: 31234857]

8. Orgill M, Gilson L, Chitha W, Michel J, Erasmus E, Marchal B, et al. A qualitative study of the dissemination and diffusion of innovations: bottom up experiences of senior managers in three health districts in South Africa. Int J Equity Health 2019 Mar 29;18(1):53 [FREE Full text] [doi: 10.1186/s12939-019-0952-z] [Medline: 30925878]
9. van Rossum L, Aij KH, Simons FE, van der Eng N, Ten Have WD. Lean healthcare from a change management perspective. J Health Organ Manag 2016 May 16;30(3):475-493 [FREE Full text] [doi: 10.1108/JHOM-06-2014-0090] [Medline: 27119398]
10. Tierney E, Hannigan A, Kinneen L, May C, O'Sullivan M, King R, et al. Interdisciplinary team working in the Irish primary healthcare system: analysis of 'invisible' bottom up innovations using Normalisation Process Theory. Health Policy 2019 Nov;123(11):1083-1092 [FREE Full text] [doi: 10.1016/j.healthpol.2019.09.002] [Medline: 31575445]
11. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629 [FREE Full text] [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
12. Van de Ven A, Polley DE, Garud R, Venkataraman S. The Innovation Journey. Oxford, UK: Oxford University Press; Feb 21, 2008.
13. Braithwaite J, Glasziou P, Westbrook J. The three numbers you need to know about healthcare: the 60-30-10 challenge. BMC Med 2020 May 04;18(1):102 [FREE Full text] [doi: 10.1186/s12916-020-01563-4] [Medline: 32362273]
14. Sturmberg JP, Bircher J. Better and fulfilling healthcare at lower costs: the need to manage health systems as complex adaptive systems. F1000Res 2019 Jun 5;8:789 [FREE Full text] [doi: 10.12688/f1000research.19414.1] [Medline: 31839925]
15. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
16. Cohn S, Clinch M, Bunn C, Stronge P. Entangled complexity: why complex interventions are just not complicated enough. J Health Serv Res Policy 2013 Jan 01;18(1):40-43. [doi: 10.1258/jhsrp.2012.012036] [Medline: 23393041]
17. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Analysing the role of complexity in explaining the fortunes of technology programmes: empirical application of the NASSS framework. BMC Med 2018 May 14;16(1):66 [FREE Full text] [doi: 10.1186/s12916-018-1050-6] [Medline: 29754584]
18. Rogers EM. Diffusion of innovations: modifications of a model for telecommunications. In: Stoetzer MW, Mahler A, editors. Die Diffusion von Innovationen in der Telekommunikation. Berlin, Heidelberg: Springer; 1995.
19. Robert G, Greenhalgh T, MacFarlane F, Peacock R. Adopting and assimilating new non-pharmaceutical technologies into health care: a systematic review. J Health Serv Res Policy 2010 Oct;15(4):243-250. [doi: 10.1258/jhsrp.2010.009137] [Medline: 20592046]
20. Greenhalgh T, Maylor H, Shaw S, Wherton J, Papoutsi C, Betton V, et al. The NASSS-CAT tools for understanding, guiding, monitoring, and researching technology implementation projects in health and social care: protocol for an evaluation study in real-world settings. JMIR Res Protoc 2020 May 13;9(5):e16861 [FREE Full text] [doi: 10.2196/16861] [Medline: 32401224]
21. Yin RK. Choices in designing qualitative research studies. In: Yin RK, editor. Qualitative Research from Start to Finish. New York, NY: The Guilford Press; 2003:359-386.
22. Västra Götalandsregionen homepage. Västra Götalandsregionen. URL: https://www.vgregion.se/ [accessed 2024-01-29]
23. Armstrong DG, Tan TW, Boulton AJ, Bus SA. Diabetic foot ulcers: a review. JAMA 2023 Jul 03;330(1):62-75. [doi: 10.1001/jama.2023.10578] [Medline: 37395769]
24. Ibrahim A, Jude E, Langton K, Martinez-De Jesus FR, Harkless LB, Gawish H, et al. IDF clinical practice recommendations on the diabetic foot – 2017. International Diabetes Federation. 2017. URL: https://idf.org/media/uploads/2023/05/attachments-61.pdf [accessed 2024-04-08]
25. Fotundersökning vid diabetes. Sveriges Kommuner och Landsting. 2018. URL: https://swenurse.se/download/18.32612fa517b3be241ddb892d/1629654183409/fotundersokning-vid-diabetes.pdf [accessed 2024-04-08]
26. Personcentrerat och sammanhållet vårdförlopp diabetes med hög risk för fotsår. Nationellt Programområde Endokrina Sjukdomar. 2022 Jun 3. URL: https://d2flujgsl7escs.cloudfront.net/external/Vardforlopp_Diabetes.pdf [accessed 2024-04-08]
27. Hellstrand Tang U, Zügner R, Lisovskaja V, Karlsson J, Hagberg K, Tranberg R, et al. The D-Foot, for prosthetists and orthotists, a new eHealth tool useful in useful in risk classification and foot assessment in diabetes. Foot Ankle Online J 2017 Aug;10(2):4 [FREE Full text] [doi: 10.3827/faoj.2017.1002.0004]
28. Zügner R, Jarl G, Sundberg L, Tang UH. Experiences of using a digital tool, the D-foot, in the screening of risk factors for diabetic foot ulcers. J Foot Ankle Res 2022 Dec 13;15(1):90 [FREE Full text] [doi: 10.1186/s13047-022-00594-9] [Medline: 36514099]
29. Ribu L, Hanestad BR, Moum T, Birkeland K, Rustoen T. A comparison of the health-related quality of life in patients with diabetic foot ulcers, with a diabetes group and a nondiabetes group from the general population. Qual Life Res 2007 Mar;16(2):179-189. [doi: 10.1007/s11136-006-0031-y] [Medline: 17091370]
30. Regulation (EU) 2017/745 of the European parliament and of the council. Official Journal of the European Union. 2017 Apr 5. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32017R0745 [accessed 2024-04-08]

31. Ragnarson Tennvall G, Apelqvist J. Health-economic consequences of diabetic foot lesions. Clin Infect Dis 2004 Aug 01;39 Suppl 2:S132-S139. [doi: 10.1086/383275] [Medline: 15306992]
32. Hellstrand S, Sundberg L, Karlsson J, Zügner R, Tranberg R, Hellstrand Tang U. The value of time consumed in healthcare – prevention of diabetic foot ulcers. Heliyon Preprint posted online November 1, 2023 [FREE Full text] [doi: 10.2139/ssrn.4610118]
33. Hellstrand S, Hellstrand Tang U. A new approach to quantifying the sustainability effects of healthcare: applied to the diabetic foot. Foot Ankle Online J 2019;12(3):5. [doi: 10.3827/faoj.2019.1203.0005]
34. Pontin E, Peters S, Lobban F, Rogers A, Morriss RK. Enhanced relapse prevention for bipolar disorder: a qualitative investigation of value perceived for service users and care coordinators. Implement Sci 2009 Feb 09;4(1):4 [FREE Full text] [doi: 10.1186/1748-5908-4-4] [Medline: 19203373]
35. Swedish Medical Products Agency. Manufacture of medical devices 2024. URL: https://www.lakemedelsverket.se/sv [accessed 2024-04-08]
36. Regional utvecklingsstrategi för Västra Götaland 2021 – 2030. Västra Götalandsregionen. 2021. URL: https://tinyurl.com/266rt7be [accessed 2023-12-26]
37. Millennium. Västra Götalandsregionen. URL: https://www.vgregion.se/halsa-och-vard/vardgivarwebben/vardskiftet/millennium/ [accessed 2024-01-29]
38. Manufacture of medical devices. Swedish Medical Products Agency. 2020. URL: https://www.lakemedelsverket.se/en/medical-devices/manufacture#hsecondarybody3 [accessed 2024-04-02]
39. Nationella riktlinjer för diabetesvård. Socialstyrelsen. URL: https://www.socialstyrelsen.se/kunskapsstod-och-regler/regler-och-riktlinjer/nationella-riktlinjer/riktlinjer-och-utvarderingar/diabetes/ [accessed 2024-04-08]
40. Schaper NC, van Netten JJ, Apelqvist J, Bus SA, Fitridge R, Game F, et al. Practical guidelines on the prevention and management of diabetes-related foot disease (IWGDF 2023 update). Diabetes Metab Res Rev 2024 Mar;40(3):e3657. [doi: 10.1002/dmrr.3657] [Medline: 37243927]
41. Ragnarson Tennvall G, Apelqvist J. Health-related quality of life in patients with diabetes mellitus and foot ulcers. J Diabetes Complications 2000;14(5):235-241. [doi: 10.1016/s1056-8727(00)00133-1] [Medline: 11113684]
42. Ragnarson Tennvall G, Apelqvist J. Prevention of diabetes-related foot ulcers and amputations: a cost-utility analysis based on Markov model simulations. Diabetologia 2001 Nov;44(11):2077-2087. [doi: 10.1007/s001250100013] [Medline: 11719840]
43. Grynne A, Browall M, Fristedt S, Ahlberg K, Smith F. Integrating perspectives of patients, healthcare professionals, system developers and academics in the co-design of a digital information tool. PLoS One 2021 Jul 16;16(7):e0253448 [FREE Full text] [doi: 10.1371/journal.pone.0253448] [Medline: 34270564]
44. Fristedt S, Smith F, Grynne A, Browall M. Digi-Do: a digital information tool to support patients with breast cancer before, during, and after start of radiotherapy treatment: an RCT study protocol. BMC Med Inform Decis Mak 2021 Feb 25;21(1):76 [FREE Full text] [doi: 10.1186/s12911-021-01448-3] [Medline: 33632215]
45. Gremyr A, Andersson Gäre B, Greenhalgh T, Malm U, Thor J, Andersson AC. Using complexity assessment to inform the development and deployment of a digital dashboard for schizophrenia care: case study. J Med Internet Res 2020 Apr 23;22(4):e15521 [FREE Full text] [doi: 10.2196/15521] [Medline: 32324143]
46. Steen M. Virtues in participatory design: cooperation, curiosity, creativity, empowerment and reflexivity. Sci Eng Ethics 2013 Sep 18;19(3):945-962. [doi: 10.1007/s11948-012-9380-9] [Medline: 22806218]
47. Smith F, Wallengren C, Öhlén J. Participatory design in education materials in a health care context. Action Res 2016 May 03;15(3):310-336. [doi: 10.1177/1476750316646832]
48. Innovation i komplexa organisationer. Västra Götalandsregionen. 2021 Feb 18. URL: https://www.vgregion.se/siteassets/ovriga_webbplatser/innovationsplattformen/innovation-i-komplexa-organisationer.pdf [accessed 2024-04-08]
49. Nationellt vårdprogram bröstcancer. Kunskapsbank. URL: https://kunskapsbanken.cancercentrum.se/diagnoser/brostcancer/vardprogram/ [accessed 2024-04-08]
50. International guidelines on breast cancer care. European Commission. URL: https://healthcare-quality.jrc.ec.europa.eu/ecibc/international-guidelines [accessed 2024-04-08]
51. Branca Vergano L, Monesi M, Vicenti G, Bizzoca D, Solarino G, Moretti B. Posterior approaches in malleolar fracture: when, why and how. J Biol Regul Homeost Agents 2020;34(3 Suppl. 2):89-95. ADVANCES IN MUSCULOSKELETAL DISEASES AND INFECTIONS. [Medline: 32856446]
52. Nationella medicinska informationssystem - NMI. Läkemedelsverket. 2019 Oct 30. URL: https://www.lakemedelsverket.se/sv/medicinteknik/tillverka/nationella-medicinska-informationssystem#hmainbody1 [accessed 2023-12-26]
53. Ventura F, Brovall M, Smith F. Beyond effectiveness evaluation: contributing to the discussion on complexity of digital health interventions with examples from cancer care. Front Public Health 2022 Jul 22;10:883315 [FREE Full text] [doi: 10.3389/fpubh.2022.883315] [Medline: 35968430]
54. Gremyr A, Holmberg C, Thor J, Malm U, Gäre BA, Andersson AC. How a point-of-care dashboard facilitates co-production of health care and health for and with individuals with psychotic disorders: a mixed-methods case study. BMC Health Serv Res 2022 Dec 30;22(1):1599 [FREE Full text] [doi: 10.1186/s12913-022-08992-2] [Medline: 36585696]
55. Vision e-hälsa 2025. E-hälsa. URL: https://ehalsa2025.se/ [accessed 2024-04-08]

56. Lag (2007:1091) om offentlig upphandling. Sveriges Riksdag. URL: https://www.riksdagen.se/sv/dokument-och-lagar/dokument/svensk-forfattningssamling/lag-20071091-om-offentlig-upphandling_sfs-2007-1091/ [accessed 2024-04-08]
57. Bell H, Rees D, Huxtable-Thomas L, Rich N, Miller E, Thomas R. Innovation adoption research in healthcare: understanding context and embracing complexity. In: Proceedings of the 18th European Conference on Innovation and Entrepreneurship. 2023 Presented at: ECIE 2023; September 21-22, 2023; Porto, Portugal. [doi: 10.34190/ecie.18.1.1865]
58. Chaves BG, Briand C, Bouabida K. Innovation in healthcare organizations: concepts and challenges to consider. Int J Health Res Innov 2021;9(1):1-14. [doi: 10.47260/ijhri/911]
59. Dugstad J, Eide T, Nilsen ER, Eide H. Towards successful digital transformation through co-creation: a longitudinal study of a four-year implementation of digital monitoring technology in residential care for persons with dementia. BMC Health Serv Res 2019 Jun 10;19(1):366 [FREE Full text] [doi: 10.1186/s12913-019-4191-1] [Medline: 31182093]
60. Schliemann T, Danielsen C, Virtanen T, Vuokko R, Hardardottir GA, Alsaker MA, et al. eHealth standardisation in the Nordic countries: technical and partially semantics standardisation as a strategic means for realising national policies in eHealth. Nordic Council of Ministers. 2019. URL: https://norden.diva-portal.org/smash/get/diva2:1340369/FULLTEXT01.pdf [accessed 2024-04-08]
61. Steenkamer B, Drewes H, Putters K, van Oers H, Baan C. Reorganizing and integrating public health, health care, social care and wider public services: a theory-based framework for collaborative adaptive health networks to achieve the triple aim. J Health Serv Res Policy 2020 Jul;25(3):187-201. [doi: 10.1177/1355819620907359] [Medline: 32178546]
62. Mondéjar AG, Silva-Calpa GF, Raposo AB, Mograbi DC. BraPolar: an m-health application for remote monitoring of people with bipolar disorder. In: Proceedings of the IEEE 7th International Conference on Serious Games and Applications for Health (SeGAH). 2019 Presented at: SeGAH 2019; August 5-7, 2019; Kyoto, Japan. [doi: 10.1109/segah.2019.8882469]
63. Faurholt-Jepsen M, Kessing LV. Monitoring and treatment in patients with bipolar disorder using smartphones-new perspectives for improved quality in patient care. Psychiatry Res 2022 Nov;317:114844. [doi: 10.1016/j.psychres.2022.114844] [Medline: 36115167]
64. Shan R, Sarkar S, Martin SS. Digital health technology and mobile devices for the management of diabetes mellitus: state of the art. Diabetologia 2019 Jun;62(6):877-887. [doi: 10.1007/s00125-019-4864-7] [Medline: 30963188]
65. Coulter A, Locock L, Ziebland S, Calabrese J. Collecting data on patient experience is not enough: they must be used to improve care. BMJ 2014 Mar 26;348:g2225. [doi: 10.1136/bmj.g2225] [Medline: 24671966]
66. Coulter A, Oldham J. Person-centred care: what is it and how do we get there? Future Hosp J 2016 Jun;3(2):114-116 [FREE Full text] [doi: 10.7861/futurehosp.3-2-114] [Medline: 31098200]
67. Sheikh A, Anderson M, Albala S, Casadei B, Franklin BD, Richards M, et al. Health information technology and digital innovation for national learning health and care systems. Lancet Digit Health 2021 Jun;3(6):e383-e396 [FREE Full text] [doi: 10.1016/S2589-7500(21)00005-4] [Medline: 33967002]
68. Niñerola A, Sánchez-Rebull MV, Hernández-Lara AB. Quality improvement in healthcare: Six Sigma systematic review. Health Policy 2020 Apr;124(4):438-445. [doi: 10.1016/j.healthpol.2020.01.002] [Medline: 32156468]
69. Balazs CL, Morello-Frosch R. The three R's: how community based participatory research strengthens the rigor, relevance and reach of science. Environ Justice 2013 Feb;6(1):9-16 [FREE Full text] [doi: 10.1089/env.2012.0017] [Medline: 24260590]
70. Ocloo J, Matthews R. From tokenism to empowerment: progressing patient and public involvement in healthcare improvement. BMJ Qual Saf 2016 Aug 18;25(8):626-632 [FREE Full text] [doi: 10.1136/bmjqs-2015-004839] [Medline: 26993640]
71. Gremyr I, Elg M, Smith F, Gustavsson S. Exploring the phase for highest impact on radicality: a cross-sectional study of patient involvement in quality improvement in Swedish healthcare. BMJ Open 2018 Nov 08;8(11):e021958 [FREE Full text] [doi: 10.1136/bmjopen-2018-021958] [Medline: 30413500]
72. Taylor MJ, McNicholas C, Nicolay C, Darzi A, Bell D, Reed JE. Systematic review of the application of the plan-do-study-act method to improve quality in healthcare. BMJ Qual Saf 2014 Apr 11;23(4):290-298 [FREE Full text] [doi: 10.1136/bmjqs-2013-001862] [Medline: 24025320]
73. Sacchi C, Andersson K, Roczniewska M, Luckhaus JL, Malmqvist M, Rodmalm LP, et al. Mind the gap: analysis of two pilot projects of a home telehealth service for persons with complex conditions in a Swedish hospital. BMC Health Serv Res 2023 May 09;23(1):463 [FREE Full text] [doi: 10.1186/s12913-023-09409-4] [Medline: 37161458]
74. Greenhalgh T. How to improve success of technology projects in health and social care. Public Health Res Pract 2018 Sep 27;28(3):2831815 [FREE Full text] [doi: 10.17061/phrp2831815] [Medline: 30406256]
75. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021 Sep 30;374:n2061 [FREE Full text] [doi: 10.1136/bmj.n2061] [Medline: 34593508]
76. Batalden PB, Davidoff F. What is "quality improvement" and how can it transform healthcare? Qual Saf Health Care 2007 Feb;16(1):2-3 [FREE Full text] [doi: 10.1136/qshc.2006.022046] [Medline: 17301192]
77. Grynne A, Wångdahl J, Fristedt S, Smith F, Browall M. Women's experience of the health information process involving a digital information tool before commencing radiation therapy for breast cancer: a deductive interview study. BMC Health Serv Res 2023 Aug 09;23(1):842 [FREE Full text] [doi: 10.1186/s12913-023-09837-2] [Medline: 37559113]

78. Lindenfalk B, Gremyr A, Lundmark M, Jacobsson T. Digitally mediated schizophrenia care - a Swedish case of translating, designing and expert evaluation of dialog. Stud Health Technol Inform 2022 Jun 06;290:882-886. [doi: 10.3233/SHTI220206] [Medline: 35673145]
79. Moukhtarian TR, Reinhard I, Morillas-Romero A, Ryckaert C, Mowlem F, Bozhilova N, et al. Wandering minds in attention-deficit/hyperactivity disorder and borderline personality disorder. Eur Neuropsychopharmacol 2020 Sep;38:98-109. [doi: 10.1016/j.euroneuro.2020.07.005] [Medline: 32703662]
80. Moukhtarian TR, Reinhard I, Moran P, Ryckaert C, Skirrow C, Ebner-Priemer U, et al. Comparable emotional dynamics in women with ADHD and borderline personality disorder. Borderline Personal Disord Emot Dysregul 2021 Feb 12;8(1):6 [FREE Full text] [doi: 10.1186/s40479-021-00144-y] [Medline: 33579385]

Abbreviations

CCA: constant comparative analysis
CE: Conformité Européenne
DFU: diabetic foot ulcer
IPF: Innovation Platform
NASSS: nonadoption, abandonment, scale-up, spread, and sustainability
NASSS-CAT: nonadoption, abandonment, scale-up, spread, and sustainability complexity assessment tool
NMI: national medical information system
RT: radiation therapy
VGR: Region Västra Götaland

Edited by A Kushniruk; submitted 18.07.23; peer-reviewed by M Aanestad, J Wells; comments to author 20.11.23; revised version received 24.02.24; accepted 02.03.24; published 26.04.24.

Please cite as:
Hellstrand Tang U, Smith F, Karilampi UL, Gremyr A
Exploring the Role of Complexity in Health Care Technology Bottom-Up Innovations: Multiple-Case Study Using the Nonadoption, Abandonment, Scale-Up, Spread, and Sustainability Complexity Assessment Tool
JMIR Hum Factors 2024;11:e50889
URL: https://humanfactors.jmir.org/2024/1/e50889
doi:10.2196/50889
PMID:38669076

Original Paper

Provider Adoption of mHealth in Rural Patient Care: Web-Based Survey Study

Bryan P Weichelt[1*], MBA, PhD; Rick Burke[1*], MPH; Burney Kieke[2*], MS; Matt Pilz[1*], BS; Neel Shimpi[3*], PhD

[1]National Farm Medicine Center, Marshfield Clinic Research Institute, Marshfield, WI, United States

[2]Center for Clinical Epidemiology & Population Health, Marshfield Clinic Research Institute, Marshfield, WI, United States

[3]American Dental Association, Chicago, IL, United States

[*]all authors contributed equally

Corresponding Author:
Bryan P Weichelt, MBA, PhD
National Farm Medicine Center
Marshfield Clinic Research Institute
1000 N Oak Ave
Marshfield, WI, 54449
United States
Phone: 1 715 221 7276
Email: weichelt.bryan@marshfieldresearch.org

Abstract

Background: Physicians and patient-facing caregivers have increasingly used mobile health (mHealth) technologies in the past several years, accelerating during the COVID-19 pandemic. However, barriers and feedback surrounding adoption remain relatively understudied and varied across health systems, particularly in rural areas.

Objective: This study aims to identify provider adoption, attitudes, and barriers toward mHealth in a large, multisite, rural US health care system. We investigated (1) mHealth apps that providers use for their own benefit and (2) mHealth apps that a provider uses in conjunction with a patient.

Methods: We surveyed all patient-seeing providers within the Marshfield Clinic Health System with a brief, 16-item, web-based survey assessing attitudes toward mHealth, adoption of these technologies, and perceived barriers faced by providers, their peers, and the institution. Survey results were summarized via descriptive statistics, with log-binomial regression and accompanying pairwise analyses, using Kruskal-Wallis and Jonckheere-Terpstra tests for significance, respectively. Respondents were grouped by reported clinical role and specialty.

Results: We received a 38% (n/N=916/2410) response rate, with 60.7% (n=556) of those sufficiently complete for analyses. Roughly 54.1% (n=301) of respondents reported mHealth use, primarily around decision-making and supplemental information, with use differing based on provider role and years of experience. Self-reported barriers to using mHealth included a lack of knowledge and time to study mHealth technologies. Providers also reported concerns about patients' internet access and the complexity of mHealth apps to adequately use mHealth technologies. Providers believed the health system's barriers were largely privacy, confidentiality, and legal review concerns.

Conclusions: These findings echo similar studies in other health systems, surrounding providers' lack of time and concerns over privacy and confidentiality of patient data. Providers emphasized concerns over the complexity of these technologies for their patients and concerns over patients' internet access to fully use mHealth in their delivery of care.

(JMIR Hum Factors 2024;11:e55443) doi:10.2196/55443

KEYWORDS

mHealth; clinician; physician; rural; patient; mobile; health care; adoption; attitude; attitudes; opinion; perception; perceptions; perspective; perspectives; acceptance; mobile health; app; apps; provider; providers; physicians; survey; surveys; barrier; barriers; digital health

Introduction

Increased technological and medical advancements have naturally led to the intersection of these 2 fields of study, commonly known as mobile health (mHealth) [1]. mHealth has been defined as "mobile computing, medical sensor, and communication technologies for healthcare" and has seen increasing adoption in recent years, particularly following the shift to virtual health delivery during the COVID-19 pandemic [1-3]. However, given privacy concerns, institutional hesitancy, and the wide array of programs and devices available, adoption and use of mHealth have been mixed [4].

The 2 prevailing mobile platforms include Android and iOS, which collectively comprise more than 99% of mobile use today on phones, tablets, and wearable devices [5]. Traditionally, the development of mobile apps required maintaining separate codebases and expertise per platform. Advances in web technologies, coupled with the establishment of these 2 universal operating systems, have ushered in new and dynamically evolving cross-platform solutions. Vastly expanded broadband cellular networks have simultaneously led to a surge in mobile accessibility, with 95% of the globe reaching 3G coverage as of 2022 and 80% obtaining 4G or faster speeds, including throughout rural regions [6].

To streamline development resources while maximizing user reach, cross-platform frameworks have become dominant across all sectors of mobile apps, including mHealth [7]. These libraries leverage existing technologies, often derivatives of web languages such as HTML, cascading style sheets, and JavaScript, while seamlessly interfacing with the native capabilities of each platform. Two of the most popular architectures for modern cross-platform apps include React Native (Meta Platforms) and Flutter (Google) [8]. Similarly, 2 popular hybrid solutions include Cordova and Capacitor, which can efficiently embed existing websites into native web views to achieve familiar native app behavior.

Emerging app-connected wearables—smartwatches, eyewear, earwear, and clothing—synergize with cross-platform apps to offer new ways of interacting with consumers and patients. Adaptation of wearable technologies has boomed in recent years, with a projected 1 billion circulated wearables in 2022 compared with 325 million in 2016 [9]. Android and iOS smartwatches offer core health initiatives, with sensors and apps that can automatically analyze heart health, blood oxygen, sleep cycles, and fitness. This innovation is rapidly superseding traditional life alert functionality with recent developments including fall detection, automatic location-aware emergency dialing, predictive warnings of heart arrhythmia, and reported seamless syncing of medical records [10].

Many promising new use cases of wearable tech in the mHealth industry are emerging after years of research and pilot studies. For example, after 12 years of testing, Apple has proofed a noninvasive continuous blood glucose monitor system that uses silicon photonics and optical spectroscopy, which is expected to be miniaturized into a common watch-sized wearable within 3 years for consumer use [11]. Manufacturers have also continued exploring augmented reality medical applications for eyewear and are working on adaptive lens adjustment technologies, which would dynamically adjust to one's eyesight with no prescription lens required. Other types of wearable devices are continuously being tested, including ones that can monitor saliva or tear gland fluids to detect eye or oral diseases, among other medical conditions [12].

These advances in technologies and applications have moved at incredible speeds, most often ahead of health systems' and providers' organizational abilities or individual preparedness to adopt, test, and implement for their own use or use with patients. Nonetheless, health care providers can and do leverage available advances in medical technologies for the benefit of the patient, and we would fully expect that mHealth apps and wearable technologies are no exception.

To better understand the current environment of mHealth adoption and barriers among rural providers and patients, we sought to further explore two key topics in this study: (1) apps that providers use for their own benefit and (2) apps that a provider uses in conjunction with a patient.

Khatun et al originally described a conceptual model for mhealth readiness through the lens of a health workforce in rural Bangladesh [13]. The model was later advanced and refined by Weichelt et al in 2019, furthering discussions of the interplay of rural patients, clinicians, and their organizations in mHealth adoption [14,15]. This prior research found that the organization plays a role in impacting providers' and patients' adoption of mHealth; however, we need to first gain a deeper understanding of providers' current levels of adoption and familiarity and awareness with these new technologies.

This line of research, beginning with an assessment and inventory of mHealth adoption, is essential for the future of health care delivery. Marshfield Clinic Health System (MCHS) is a predominantly rural health system with patients scattered across northern Wisconsin, the Upper Peninsula of Michigan, and beyond [16]. While well positioned to test and deploy new and innovative technologies in the broad field of mHealth, leadership first needs to gain a deeper understanding of the system's provider and patient needs, desires, and current use. Therefore, we conducted a survey of all patient-seeing providers within MCHS to identify mHealth adoption, attitudes, and perceived barriers to use.

Methods

Data Collection

In July of 2020, we emailed a survey to 2410 MCHS providers via an information systems–supplied "MCHS Providers" email list. The survey was designed to assess providers' motivators and barriers to the adoption of mHealth technologies in patient care. The survey was open and available from July 21 to August 31 (6 weeks), with 2 reminders sent every 2 weeks.

Instrument design and line of inquiry leveraged previous work by this research team, including the previously published conceptual model for assessing necessary conditions for rural health care's mHealth readiness, with an emphasis on clinician-perceived barriers. Providers were asked about mHealth use, both personal and with patients, as well as personal, perceived colleague, patient, and institutional barriers to mHealth adoption. Providers were also asked about the perceived COVID-19 impact on mHealth use and anticipated future mHealth use after the pandemic subsided.

Incentives

Participants were presented with the option of selecting one of five local nonprofits to receive a US $10 donation for their voluntary participation in the survey. We distributed our full budgeted allotment of US $2800 as chosen by the research participants. No other incentives were offered during the study.

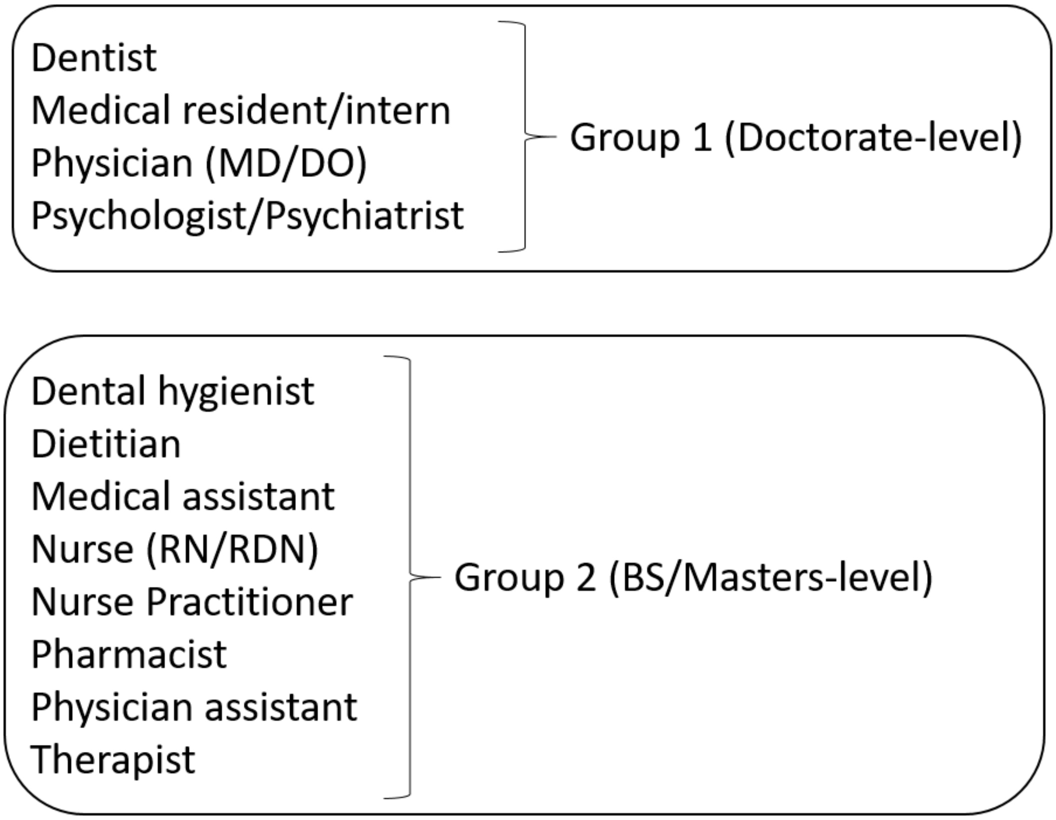
Ethical Considerations

The project submission was evaluated by the Marshfield Clinic institutional review board. It was determined that the activity as described does not meet the definition of human participant research, and no further institutional review board action was needed.

Analyses

Due to the small number of responses in some niche roles and specialties, participants' roles were grouped into 2 categories based on education and degree level (Figure 1). Provider specialty was also compartmentalized into 9 categories, mirroring the distribution of specialties across MCHS.

Figure 1. Grouping of provider roles. RDN: Registered Dietitian Nutritionist; RN: registered nurse.



We used log-binomial regression to analyze survey questions with dichotomous (yes or no) responses [17]. Specifically, we fit univariable models where the dependent variable was the dichotomous response and the independent variable was either provider role, provider specialty, or number of years in practice (7 categories: 0-5, 6-10, 11-15, 16-20, 21-25, 26-30, and >30 years). We assessed the overall statistical significance of the independent variable and proceeded with pairwise comparisons versus a referent category when warranted (ie, when the *P* value for the overall effect was ≤.05). BS and MS-level providers, family medicine, and 16-20 years in practice were the referent categories in the calculations. Since provider roles had 2 categories, corresponding pairwise comparisons were unnecessary (redundant).

A similar general strategy was used for Likert-scaled survey questions (ie, assessment of the overall statistical significance of the independent variable followed by pairwise comparisons vs a referent category when warranted). The Kruskal-Wallis test was used to evaluate overall significance and the Jonckheere-Terpstra test for pairwise comparisons [18,19]. The same independent variables were examined.

Results

Overview

We received a total of 916 responses (38% response rate), of which 556 (60.7%) responses were sufficiently complete to be included in the statistical analyses. Of these responses, 301 (54.1%) participants reported using health-related apps on their phone or tablet. The most common purposes for these apps were use as informational resources (234/301, 77.7%) and for decision-making (180/301, 59.8%). Providers who used mHealth with patients (202/556, 36.3%) reported doing so primarily for exercise and activity monitoring (105/202, 52%), and enhancing patients' experiences via the My Marshfield Clinic app (105/202, 52%), an in-house app that allows patients to schedule appointments, view lab results, message providers, etc. Those who do not use health-related apps (255/556, 45.9%) stated their primary reasons as inadequate information available on the use of such apps (59/255, 23.1%), not having enough time to use the apps (55/255, 21.6%), or being unsure of the organization's attitudes toward mHealth (59/255, 23.1%). The most common barriers to mHealth adoption cited by providers were a lack of both knowledge about mHealth technologies (293/556, 52.7%) and time (201/556, 36.2%), as well as being unsure of their patients' access to reliable internet services (171/556, 30.8%). These same concerns arose when we asked respondents which barriers they thought other providers had surrounding mHealth (319/556, 57.4%; 283/556, 50.9%; and 163/556, 29.3%). Perceived organizational barriers to clinicians using mHealth in their practices were primarily concerns related to confidentiality (313/556, 56.3%) and mHealth technologies being too complicated for patients (288/556, 51.8%). Overall, however, providers had a favorable view of mHealth, with a majority stating they either intend to continue using mHealth following the COVID-19 pandemic or would look further into mHealth technologies.

Provider Demographics

We had a broad representation of specialties and experience levels in our responses. The most common specialties were family medicine, surgery, pediatrics, and physical and occupational therapy. The survey respondents averaged 19 years of experience practicing medicine.

Clinician Adoption

Clinician adoption of mHealth varied by role and specialty. Doctoral-level providers reported higher mHealth use on their own devices compared with other providers, with 65% (n/N=154/237) and 46% (n/N=138/300) adoption (*P*<.001), respectively. Among mHealth users, doctoral-level providers used these apps as an informational resource at a higher rate (131/154, 85.1% vs 98/138, 71%; *P*=.005). Compared with midtenure providers (16-20 years of experience, 51/81, 63% mHealth adoption), mHealth adoption levels were reported to be lower among more experienced providers (34/72, 47.2% for 21-25 years of experience; 29/64, 45.3% for 26-30 years; and 43/84, 51.2% for >30 years), and similar among less experienced providers (52/95, 54.7% for 0-5 years of experience; 43/73, 58.9% for 6-10 years; and 39/64, 60.9% for 11-15 years). mHealth use with patients was similar between doctoral-level and other providers (84/237, 35.4% vs 110/300, 36.7%; *P*=.77) and across the range of years of experience (33/95, 34.7% for 0-5 years of experience; 27/73, 37% for 6-10 years; 22/64, 34.4% for 11-15 years; 29/81, 35.8% for 16-20 years; 34/72, 47.2% for 21-25 years; 24/64, 37.5% for 26-30 years; and 29/84, 34.5% for >30 years; *P*=.63). Notably, compared with family medicine with 48.4% (n/N=31/64) mHealth use with patients, 3 specialties reported use of ≤30% (Cancer Care and Research, 6/27, 22.2%, *P*=.04 vs family medicine; Cardiology, 7/29, 24.1%, *P*=.05; Surgery, 14/60, 23.3%, *P*=.007), while psychiatry and psychology reported 78.3% (n/N=18/23) adoption, significantly higher than family medicine (*P*=.005). No important differences in reported mHealth use with patients were observed regarding diet and nutrition tracking, weight management, dental reminders, direct communication with the patient's care team, and medication reminders. Not surprisingly, psychiatry and psychology reported use of mHealth more frequently for mood and depressive symptom monitoring (12/18, 66.7% vs ≤7% for all other specialties that responded to this question [no responses in Cancer Care and Research, Cardiology, OB/GYN, Physical and Occupational Therapy, and Surgery; 1/24, 4.2% in Pediatrics; and 4/65, 6.2% in other specialties], *P*=.001 vs family medicine, 2/31, 6.4%) and sleep tracking (7/18, 38.9% vs ≤16.7% for all other specialties that responded to this question [no responses in Cardiology, OB/GYN and Physical and Occupational Therapy; 1/6, 16.7% in Cancer Care and Research; 2/24, 8.3% in Pediatrics; 1/14, 7.1% in Surgery; and 4/65, 6.2% in other specialties], *P*=.02 vs family medicine, 2/31, 6.4%). Physical and occupational therapy, cardiology, and psychiatry and psychology reported substantially higher mHealth use with patients for informational and educational purposes (12/17, 70.6%; 4/7, 57.1%; and 10/18, 55.6%, with *P*=.002, .048, and .02 vs family medicine, 7/31, 22.6%).

Clinicians' Perceived Barriers Category 1—Personal (Clinician)

Overall, providers reported lack of knowledge about mHealth technologies (293/556, 52.7% for themselves; 319/556, 57.4% in their perceptions regarding other clinicians) and lack of time (201/556, 36.2% and 283/556, 50.9%) as the primary personal barriers. Insufficient levels of patient internet access were also a commonly cited concern (171/556, 30.8% and 163/556, 29.3%). We found relatively few differences between provider roles and specialties regarding personal barriers to mHealth adoption. Doctoral-level providers cited a greater number of financial barriers surrounding a lack of value in mHealth technologies (29/237, 12.2% vs 12/300, 4%, *P*<.001 for themselves; 44/237, 18.6% vs 26/300, 8.7%, *P*=.001 in their perceptions regarding other clinicians), insufficient reimbursement options (27/237, 11.4% vs 14/300, 4.7%; *P*=.005 for themselves), and mHealth technologies not being worth the cost of adoption (22/237, 9.3% vs 7/300, 2.3%; *P*=.001 for themselves). With respect to their perceptions regarding other clinicians, cancer care and research providers reported a lack of communication between providers at a substantially higher rate than all other specialties (12/27, 44.4% vs 11.7%-33.3% [5/29, 17.2% in Cardiology; 9/27, 33.3% in OB/GYN; 9/47, 19.1% in Pediatrics; 6/42, 14.3% in Physical and Occupational Therapy; 5/23, 21.7% in Psychiatry and Psychology; 7/60, 11.7% in Surgery; and 45/214, 21% in other specialties] *P*=.02 vs family medicine [13/64, 20.3%]). Furthermore, regarding their perceptions of other clinicians, OB and GYN and pediatrics providers reported a lack of knowledge about mHealth technologies at rates that exceeded all other specialties (21/27, 77.8% and 37/47, 78.7% vs 45%-71.4% [13/27, 48.1% in Cancer Care and Research; 20/29, 69% in Cardiology; 30/42, 71.4% in Physical and Occupational Therapy; 13/23, 56.5% in Psychiatry and Psychology; 27/60, 45% in Surgery; and 110/214, 51.4% in other specialties], *P*=.03 and .01 vs family medicine [36/64, 56.3%]). Interestingly, the only self-perceived barrier that was modified by years of experience was the lack of reliable internet access (*P*=.02 for the overall effect). With the exception of relatively new providers (0-5 years of experience; 22/95, 23.2% of these providers reported this concern), providers in age groups with ≤20 years of experience (31/73, 42.5% with 6-10 years of experience; 24/64, 37.5% with 11-15 years; and 33/81, 40.7% with 16-20 years) reported higher rates of this concern than those in age groups with >20 years of experience (15/72, 20.8% with 21-25 years of experience; 18/64, 28.1% with 26-30 years; and 23/84, 27.4% with >30 years).

Clinicians' Perceived Barriers Category 2—Patient

Survey respondents reported substantial perceived patient concerns relating to mHealth technologies being too complicated (371/556, 66.7%), lack of access to mHealth technologies (327/556, 58.8%), poor delivery mechanisms (eg, cell service or internet coverage, 252/556, 45.3%), and privacy concerns (207/556, 37.2%). These perceptions did not differ meaningfully by provider type, specialty, or years of experience, with the exception that privacy concerns were more prevalent in doctoral-level providers (105/237, 44.3% vs 93/300, 31%; *P*=.002).

Clinicians' Perceived Barriers Category 3—Organizational

The most prevalent organizational barriers perceived by providers were concerns related to confidentiality (313/556, 56.3%) and that mHealth technologies were too complicated for patients (288/556, 51.8%). Confidentiality concerns differed meaningfully by provider type (149/237, 62.9% and 154/300, 51.3% for doctoral-level vs other providers, *P*=.007), specialty (*P*<.001 for the overall specialty effect; 37/47, 78.7% vs 36/64, 56.3% for pediatrics vs family medicine, *P*=.01), and years of experience (*P*=.03 for the overall effect; no specific trend across age groups). Privacy concerns (168/556, 30.2% prevalence) varied only by years of experience (*P*=.006 for the overall effect; no specific trend across age groups).

COVID-19 and Anticipated mHealth Adoption

When providers were asked to what degree (1) the COVID-19 pandemic impacted their mHealth adoption and (2) they intend to look further into mHealth following the resumption of normal MCHS activities, meaningful differences were detected only between provider specialties (*P*=.02 and .001 for the overall specialty effects, respectively). These differences were driven by psychiatry and psychology providers, who reported higher scores (10-point Likert scale, where 1=not at all and 10=a great deal) on both survey questions (*P*=.002 and .002 vs family medicine; Figures 2 and 3).

Figure 2. Response distributions for psychiatry and psychology and family medicine for the question "To what degree has the COVID-19 pandemic impacted your mHealth adoption?". mHealth: mobile health.

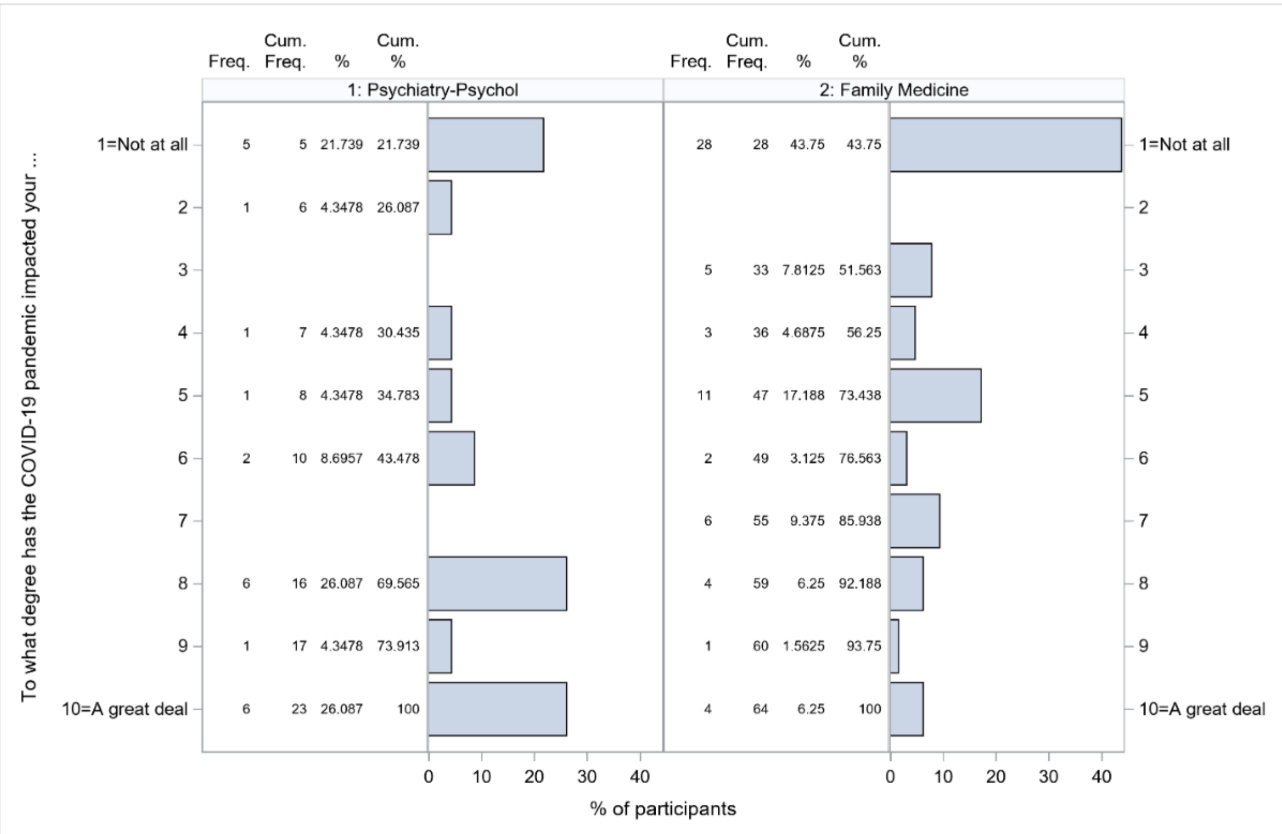
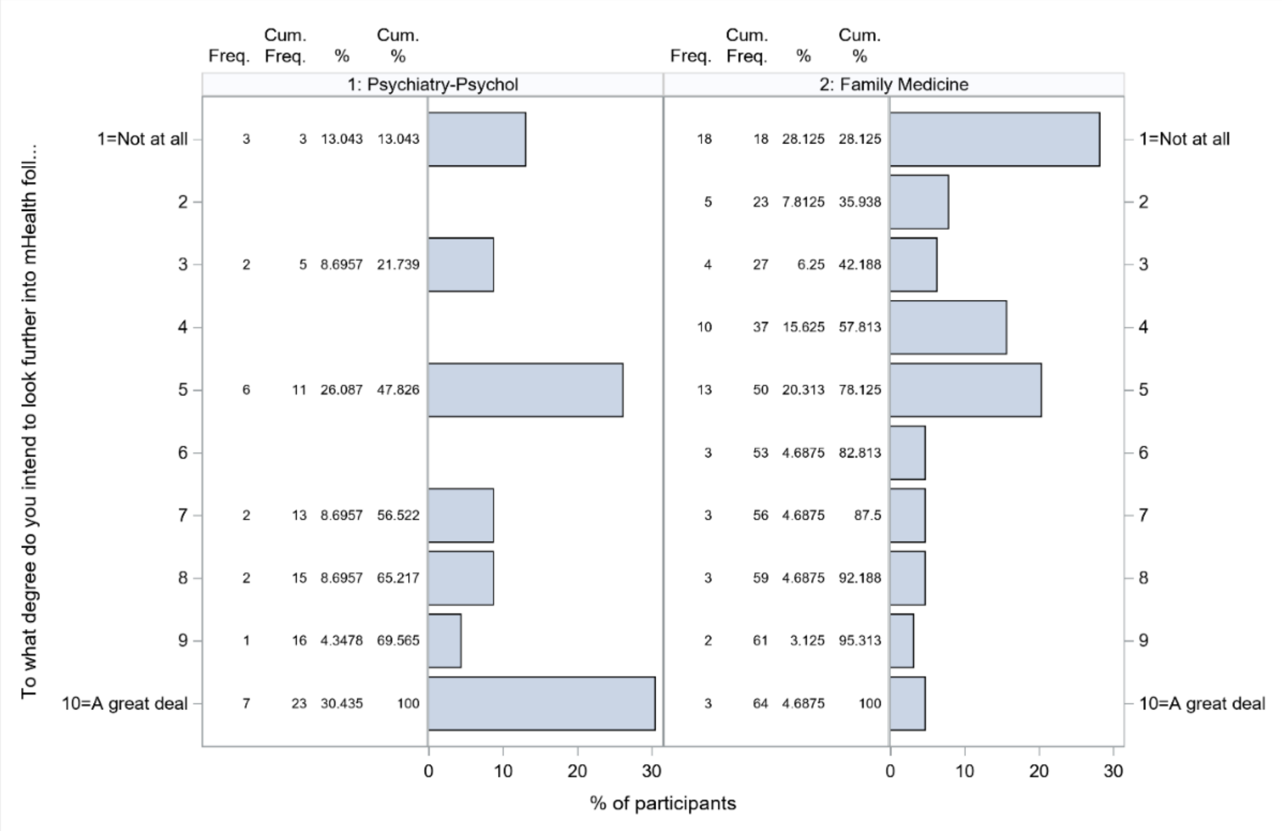


Figure 3. Response distributions for psychiatry and psychology and family medicine for the question “To what degree do you intend to look further into mHealth following the resumption of normal MCHS activities?” MCHS: Marshfield Clinic Health System; mHealth: mobile health.



Discussion

Principal Findings

The varied responses and rate of mHealth use across provider roles and specialties emphasize the variety and task-specific role mHealth can have in a health system. Some specialties, such as psychology and psychiatry, showed high rates of adoption for specific tasks such as mood and sleep tracking; however, no other specialties reported substantially greater mHealth adoption compared with the reference to family medicine. If looking to increase mHealth use across a health system, leadership should consider identifying specific tasks or poorly performing metrics that mHealth could potentially improve upon.

Our survey grouped potential barriers into 3 levels (provider, patient, and organizational) in line with past qualitative findings [13]. Providers’ self-barriers encompassed themes of lack of knowledge about mHealth technologies, lack of time, and lack of patients’ access to the internet. Commonly reported barriers relating to both patients and the organization were mHealth technologies being too complicated and concerns related to privacy. The predominant organizational barrier was confidentiality concerns, whereas lack of access (cell phone coverage, internet, and mHealth technology in general) was a frequently perceived barrier for patients.

It is understandable that health care providers feel overwhelmed, with the top barrier to mHealth use being a lack of time and information. These technologies evolve at incredible speeds. How might one stay abreast of the scientific and technological advances of mHealth technologies? Even in the peer-reviewed literature, which can take months or years to publish, we witness an overwhelming ocean of information. At the time of this writing, a Google Scholar search of “mHealth” papers since 2020 (January 1, 2020, to April 5, 2024) yielded nearly 25,000 results and nearly 8000 results since January 1, 2024.

The pace of emerging and simultaneously retiring technologies remains a substantial barrier across many mHealth studies. A typical full-scale trial to evaluate a mHealth initiative lasts more than 5 years from recruitment, during which time many changes within the pertinent technologies will occur or be superseded entirely [20]. Consequentially, many trials are reduced in scope, hindering true evaluation and understanding of the prospect’s long-term value.

Years of technological ambition surrounding deep machine learning and voluminous data sets reached fruition in the 2020s with the advent of widely accessible artificial intelligence (AI) apps. These AI-powered breakthroughs have impacted nearly every industry, including medicine, in ways that are still in the infancy of exploration. By digitally processing millions of training samples, including imagery, transcripts, audio recordings, and academic papers, sophisticated computer algorithms have reached new potential in data analysis and user reactivity [21]. What once required thousands of hours and access to prohibitively expensive data centers to compute is now within a finger’s reach from any consumer phone or computer.

Leveraging computer-assisted workflows to automate tasks is not a new concept in the medical world. Health care organizations have spent decades exploring increasingly advanced forms of speech recognition software to facilitate medical transcriptions, among other areas of automation [22]. The latest groundbreaking strides in these efforts come in the form of OpenAI's ChatGPT and associated tool sets [23]. A recent study hypothesized more than 130 different ways ChatGPT could positively benefit both patients and doctors in the foreseeable future, including education, prediction support, prevention of medical errors, record-keeping, and continual clinical assistance [24].

However, many analysts warn that such tool sets—when used in isolation without a human consultant—can yield bad data or other repercussions not yet realized. The most prevalent example of these dangers is how AI modeling is prone to hallucinations, in which the chatbot may return seemingly factual and confident responses but uses nonexistent citations or made-up passages due to anomalies in its training data and other limitations [25,26].

Limitations

The response rate to our mHealth survey was 38% (n/N=916/2410), with a further completion rate of 60.7% (n/N=556/916). It is possible that response bias was present, potentially skewing toward clinicians who have an interest in mHealth technologies. While this response rate is moderately high compared with other surveys of providers, the topic of mHealth being mentioned foremost in the survey invitation may have resulted in an overestimation of mHealth use and intentions.

Notably, MCHS, along with many other health care organizations at the time, was struggling due to the COVID-19 pandemic during our survey timeframe, with rolling temporary furloughs throughout the health system. This limited our possible response rate and created uncertainty in the accuracy and complete capture of our sample. However, our survey was open for 6 weeks with multiple reminder emails sent out, theoretically limiting this effect. Nevertheless, biases in responses may remain due to the work environment and shifting priorities surrounding the COVID-19 pandemic. A future resurveying of providers would help characterize these possible impacts.

Conclusions

Health systems should continue to evaluate mHealth adoption, and more formally and proactively investigate innovative solutions. Consulting with patient safety and legal departments regarding the use of mHealth apps is crucial, as quality clinical outcomes are not often in correlation with popularity ratings on app stores [27]. If a mHealth tool is deemed to be a valuable tool for a hospital or health system, leadership should work toward identifying specific options and methods to address health outcomes and work toward simple and concise implementations to improve adoption and patient outcomes. The American Medical Association provides occasional reports and guidelines surrounding mHealth best practices, but does not have an official lobbying body, with more focus on telehealth [28-30]. The US Department of Health and Human Services provides resources for mHealth developers; however, these are primarily focused on privacy and confidentiality and are of little relevance to providers [31].

This study is arguably a foundational and necessary step in assessing a health system's status and potential for mHealth adoption. Further research and continued partnership with advisors and stakeholders will be needed if the health system hopes to more formally integrate mHealth technologies into rural health care.

Acknowledgments

The authors thank Marie Fleisner for her assistance with editing and formatting the manuscript. The authors also thank the clinicians who participated in this study and the stakeholders who helped and supported them throughout the early stages of this project, stating that their insight and guidance have been a tremendous help. The authors especially thank the project's advisory panel: Kori Krueger, MD; Kelsie-Marie Offenwanger, PsyD; Chris Meyer; John Welch; Peter Shave, MBA; Joe Vechinski, MBA; Ryan Cournoyer, RN, MBA; Jeremy Pomeroy, PhD; Casper Bendixsen, PhD; and Michelle Hamland, DNP, RN, RN-BC.

Conflicts of Interest

None declared.

References

1. Park YT. Emerging new era of mobile health technologies. Healthc Inform Res 2016;22(4):253-254 [FREE Full text] [doi: 10.4258/hir.2016.22.4.253] [Medline: 27895955]
2. Alam MZ, Proteek SM, Hoque I. A systematic literature review on mHealth related research during the COVID-19 outbreak. Health Educ 2023;123(1):19-40 [FREE Full text] [doi: 10.1108/he-08-2022-0067]
3. Alam MMD, Alam MZ, Rahman SA, Taghizadeh SK. Factors influencing mHealth adoption and its impact on mental well-being during COVID-19 pandemic: a SEM-ANN approach. J Biomed Inform 2021;116:103722 [FREE Full text] [doi: 10.1016/j.jbi.2021.103722] [Medline: 33705856]
4. Gagnon MP, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016;23(1):212-220 [FREE Full text] [doi: 10.1093/jamia/ocv052] [Medline: 26078410]
5. Market share of mobile operating systems worldwide from 2009 to 2023, by quarter. Statista. 2023. URL: https://www.statista.com/statistics/272698/global-market-share-held-by-mobile-operating-systems-since-2009 [accessed 2024-04-05]

6. Network coverage outlook and mobility report. Ericsson. 2023. URL: https://www.ericsson.com/en/reports-and-papers/mobility-report/dataforecasts/network-coverage [accessed 2024-04-05]
7. Govoruhina A, Nikiforova A. Digital health shopping assistant with react native: a simple technological solution to a complex health problem. 2022 Presented at: 2022 International Conference on Intelligent Data Science Technologies and Applications (IDSTA); September 05-07, 2022; San Antonio, TX, USA. [doi: 10.1109/idsta55301.2022.9923047]
8. Fentaw AE. Cross platform mobile application development: a comparison study of react native vs flutter. JYX Digital Repository. 2020. URL: https://jyx.jyu.fi/handle/123456789/70969# [accessed 2024-05-09]
9. Number of connected wearable devices worldwide from 2016 to 2022. Statista. 2023. URL: https://www.statista.com/statistics/487291/global-connected-wearable-devices/ [accessed 2024-04-05]
10. Niwa M. mHealth as a component of next-generation health care. In: Kodama K, Sengoku S, editors. Mobile Health (mHealth): Rethinking Innovation Management to Harmonize AI and Social Design. Singapore: Springer Nature; 2022:189-209.
11. Non-invasive blood glucose monitoring closer as Apple hits major milestone. Diabaetes.co.uk. 2023. URL: https://www.diabetes.co.uk/news/2023/mar/non-invasive-blood-glucose-monitoring-closer-as-apple-hits-major-milestone.html [accessed 2023-03-30]
12. Sheng A, Lin L, Zhu J, Zhuang J, Li J, Chang L, et al. Micro/nanodevices for assessment and treatment in stomatology and ophthalmology. Microsyst Nanoeng 2021;7:11 [FREE Full text] [doi: 10.1038/s41378-021-00238-1] [Medline: 33532080]
13. Khatun F, Heywood AE, Ray PK, Hanifi SM, Bhuiya A, Liaw S. Determinants of readiness to adopt mHealth in a rural community of Bangladesh. Int J Med Inform 2015 Oct;84(10):847-856. [doi: 10.1016/j.ijmedinf.2015.06.008] [Medline: 26194141]
14. Weichelt B, Bendixsen C, Patrick T. A model for assessing necessary conditions for rural health care's mobile health readiness: qualitative assessment of clinician-perceived barriers. JMIR Mhealth Uhealth 2019;7(11):e11915 [FREE Full text] [doi: 10.2196/11915] [Medline: 31702564]
15. Shimpi N, Glurich I, Maybury C, Wang MQ, Hashimoto K, Acharya A, et al. Knowledge, attitudes, behaviors of women related to pregnancy, and early childhood caries prevention: a cross-sectional pilot study. J Prim Care Community Health 2021;12:21501327211013302 [FREE Full text] [doi: 10.1177/21501327211013302] [Medline: 33949227]
16. About marshfield clinic health system. Marshfield Clinic Health System. 2023. URL: https://www.marshfieldclinic.org/about-us [accessed 2023-09-07]
17. McNutt LA, Wu C, Xue X, Hafner JP. Estimating the relative risk in cohort studies and clinical trials of common outcomes. Am J Epidemiol 2003;157(10):940-943 [FREE Full text] [doi: 10.1093/aje/kwg074] [Medline: 12746247]
18. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. J Am Stat Assoc 1952;47(260):583-621 [FREE Full text] [doi: 10.1080/01621459.1952.10483441]
19. Hollander M, Wolfe DA. Nonparametric Statistical Methods. Hoboken, New Jersey: Wiley; 1999.
20. Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, et al. Mobile health technology evaluation: the mHealth evidence workshop. Am J Prev Med 2013;45(2):228-236 [FREE Full text] [doi: 10.1016/j.amepre.2013.03.017] [Medline: 23867031]
21. Glock P, von Alemann S. The paradigm shift in AI: from human labor to humane creativity. In: Waltl B, Schindler D, Jacob K, Strathausen R, editors. Liquid Legal – Humanization and the Law. Singapore: Springer International Publishing; 2022:215-230.
22. Borowitz SM. Computer-based speech recognition as a replacement for medical transcription. Pediatr Res 1999;45(7):120 [FREE Full text] [doi: 10.1203/00006450-199904020-00715]
23. Hughes A. ChatGPT: everything you need to know about OpenAI's GPT-4 tool. BBC Science Focus Magazine. 2023. URL: https://www.sciencefocus.com/future-technology/gpt-3 [accessed 2024-05-09]
24. Javaid M, Haleem A, Singh RP. ChatGPT for healthcare services: an emerging stage for an innovative perspective. BenchCouncil Trans Benchmarks Stand Eval 2023;3(1):100105 [FREE Full text] [doi: 10.1016/j.tbench.2023.100105]
25. Athaluri SA, Manthena SV, Kesapragada VSRKM, Yarlagadda V, Dave T, Duddumpudi RTS. Exploring the boundaries of reality: investigating the phenomenon of artificial intelligence hallucination in scientific writing through ChatGPT references. Cureus 2023;15(4):e37432 [FREE Full text] [doi: 10.7759/cureus.37432] [Medline: 37182055]
26. Brereton D. Bing AI can't be trusted. DKB Blog. 2023. URL: https://dkb.blog/p/bing-ai-cant-be-trusted [accessed 2023-05-25]
27. Wyatt JC. How can clinicians, specialty societies and others evaluate and improve the quality of apps for patient use? BMC Med 2018;16(1):225 [FREE Full text] [doi: 10.1186/s12916-018-1211-7] [Medline: 30501638]
28. AMA adopts principles to promote safe, effective mHealth applications. American Medical Association. 2016. URL: https://www.ama-assn.org/press-center/press-releases/ama-adopts-principles-promote-safe-effective-mhealth-applications [accessed 2024-04-05]
29. AMA's advocacy efforts push for permanent telehealth advancements. American Medical Association. 2022. URL: https://www.ama-assn.org/practice-management/digital/amas-advocacy-efforts-push-permanent-telehealth-advancements [accessed 2024-05-09]
30. Comprehensive telehealth reform is critical to the future of health care. American Medical Association. 2023. URL: https://www.ama-assn.org/amaone/ama-recovery-plan-america-s-physicians-supporting-telehealth [accessed 2024-04-05]

31. Resources for mobile health apps developers. US Department of Health and Human Services. 2022. URL: https://www.hhs.gov/hipaa/for-professionals/special-topics/health-apps/index.html [accessed 2024-04-04]

Abbreviations

AI: artificial intelligence
MCHS: Marshfield Clinic Health System
mHealth: mobile health

Edited by C Jacob; submitted 13.12.23; peer-reviewed by M Niwa, D Singh; comments to author 22.01.24; revised version received 05.04.24; accepted 08.04.24; published 24.06.24.

Please cite as:
Weichelt BP, Burke R, Kieke B, Pilz M, Shimpi N
Provider Adoption of mHealth in Rural Patient Care: Web-Based Survey Study
JMIR Hum Factors 2024;11:e55443
URL: https://humanfactors.jmir.org/2024/1/e55443
doi:10.2196/55443
PMID:

Original Paper

Evaluation of a Digital Previsit Tool for Identifying Stroke-Related Health Problems Before a Follow-Up Visit (Part 1): Survey Study

Petra Pohl[1*], PhD; Karoline Klerfors[2], MSc; Emma K Kjörk[2,3,4*], PhD

[1]Department of Health and Rehabilitation, Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

[2]Department of Clinical Neuroscience, Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

[3]Centre for Person-Centred Care (GPCC), University of Gothenburg, Gothenburg, Sweden

[4]Sahlgrenska University Hospital, Gothenburg, Sweden

[*]these authors contributed equally

Corresponding Author:
Emma K Kjörk, PhD
Department of Clinical Neuroscience
Institute of Neuroscience and Physiology
Sahlgrenska Academy, University of Gothenburg
Rehabiliteringsmedicin, Vita stråket 12, vån 4
Sahlgrenska universitetssjukhuset
Gothenburg, 41345
Sweden
Phone: 46 31 3422391
Email: emma.kjork@neuro.gu.se

Abstract

Background: Stroke may lead to various disabilities, and a structured follow-up visit is strongly recommended within a few months after an event. To facilitate this visit, the digital previsit tool “Strokehealth” was developed for patients to fill out in advance. The concept Strokehälsa (or Strokehealth) was initially developed in-house as a Windows application, later incorporated in 1177.se.

Objective: The study’s primary objective was to use a patient satisfaction survey to evaluate the digital previsit tool Strokehealth when used before a follow-up visit, with a focus on feasibility and relevance from the perspective of people with stroke. Our secondary objective was to explore the extent to which the previsit tool identified stroke-related health problems.

Methods: Between November 2020 and June 2021, a web-based survey was sent to patients who were scheduled for a follow-up visit after discharge from a stroke unit and had recently filled in the previsit tool. The survey covered demographic characteristics, internet habits, and satisfaction rated using 5 response options. Descriptive statistics were used to present data from both the previsit tool and the survey. We also compared the characteristics of those who completed the previsit tool and those who did not, using nonparametric statistics. Free-text responses were thematically analyzed.

Results: All patients filling out the previsit tool (80/171; age: median 67, range 32-91 years) were community-dwelling. Most had experienced a mild stroke and reported a median of 2 stroke-related health problems (range 0-8), and they were significantly younger than nonresponders (P<.001). The survey evaluating the previsit tool was completed by 73% (58/80; 39 men). The majority (48/58, 83%) reported using the internet daily. Most respondents (56/58, 97%) were either satisfied (n=15) or very satisfied (n=41) with how well the previsit tool captured their health problems. The highest level of dissatisfaction was related to the response options in Strokehealth (n=5). Based on the free-text answers to the survey, we developed 4 themes. First, Strokehealth was perceived to provide a structure that ensured that issues would be emphasized and considered. Second, user-friendliness and accessibility were viewed as acceptable, although respondents suggested improvements. Third, participants raised awareness about being approached digitally for communication and highlighted the importance of how to be approached. Fourth, their experiences with Strokehealth were influenced by their perceptions of the explanatory texts, the response options, and the possibility of elaborating on their answers in free text.

Conclusions: People with stroke considered the freely available previsit tool Strokehealth feasible for preparing in advance for a follow-up visit. Despite high satisfaction with how well the tool captured their health problems, participants indicated that additional free-text responses and revised information could enhance usability. Improvements need to be considered in parallel with qualitative data to ensure that the tool meets patient needs.

Trial Registration: Researchweb 275135; https://www.researchweb.org/is/vgr/project/275135

(JMIR Hum Factors 2024;11:e55852) doi:10.2196/55852

KEYWORDS

e-health; stroke; Strokehälsa; follow-up; previsit; person-centred care; health literacy; digital tool; shared decision-making; survey; mobile phone

Introduction

Stroke affects more than 101 million people worldwide [1] and can lead to a range of physical, cognitive, and emotional disabilities [2]. In its newly launched "Package of interventions for rehabilitation for stroke" within its "Rehabilitation 2030 initiative," the World Health Organization makes clear that people living with stroke need lifelong access to rehabilitation services because of continuing health problems [3]. Furthermore, there is a general trend toward shorter hospital stays after a stroke. Consequently, it is vital to identify individuals at risk of secondary health issues to provide them with prevention, treatment, and support in developing self-management strategies. Continuous health care interactions are required to achieve these outcomes [4].

Depending on the problem area, patients should be assessed accordingly to access targeted interventions including education, advice, and support for self-management [3]. Postdischarge stroke care, however, is often fragmented without standardized routines for long-term follow-up [4], and people frequently have difficulties accessing health care [5]. According to the Stroke Action Plan for Europe, structured follow-up visits should be offered to all patients within 6 months after a stroke to identify stroke-related health problems, address rehabilitation needs, and support adaptation to life after a stroke [4].

As part of the structures of care for people with long-term conditions, patients must be duly prepared and sufficiently informed before a visit to a proactive health care team [6]. In addition, a person-centered approach that involves patients in shared decision-making contributes to a positive impact on health outcomes and patient satisfaction [7]. A person-centered approach entails acknowledgment within health care services that individuals can collaborate with health care professionals and actively engage in the decision-making process, which nurtures the patient's sense of empowerment [8].

To support better-structured follow-up visits after a stroke, the dialogue tool "Post-Stroke Checklist" was developed for health care professionals to use during outpatient visits [9]. The checklist comprises 11 questions that can aid in identifying common health problems (eg, mobility, cognition, and life after stroke) and guide health care professionals on appropriate actions, including recommendations for referrals [9]. Satisfaction with the dialogue tool has generally been high, but health care professionals have noted challenges in managing the checklist within the allotted timeframe [10,11], and patients have requested the ability to prepare in advance [11]. This can be achieved with previsit tools that can enhance patient experience, patient engagement, and practice efficiency [12]. In response to this, the previsit tool Strokehälsa ("Strokehealth") was developed based on the questions from the Post-Stroke Checklist [13]. The aim was to capture health problems and provide patients with information about common consequences after stroke and time to reflect [13]. Initially designed as a digital tool to enhance accessibility and usefulness, Strokehealth is now also available in a picture-supported version and in paper format, freely accessible in multiple languages. Other digital previsit tools usually concentrate solely on gathering self-reported data [12], or they may be more comprehensive, such as the stroke-related previsit tool "Rehabkompassen" [14]. Both Strokehealth and Rehabkompassen were initially developed in-house as a Windows application and later incorporated in the Swedish national patient portal.

Development and use of digital health service tools are crucial to involving patients in proactive management of their health [15]. Furthermore, as the use of diverse technologies increases within the health care domain, more patient engagement and participation will be required [16]. In general, digital tools yield positive impacts on patient empowerment, self-management, communication, and patient engagement [16-18]. In those with long-term conditions such as stroke, however, digital health solutions must be user-friendly in terms of eHealth literacy needs (ie, ability to comprehend health information and actively engage with eHealth services [19]), as these patients report related difficulties more than the general population [20]. The development of these tools thus must incorporate consideration of patient-related factors, including cognitive ability to process information, need for a sense of security and control, and intrinsic motivation to engage with digital health care services [19]. The patient perspective is a high priority in eHealth [17] because personal motivation for using digital health services is key to gaining eHealth literacy [19]. In keeping with these precepts, development of the previsit tool, Strokehealth, incorporated a comprehensive participatory approach, involving people with first-hand experience with stroke and health care professionals in the co-design process [13].

Initially, a prototype of Strokehealth was created, and subsequent iterations were tested on purposively selected patients with stroke and relevant health care professionals [13]. Based on user feedback, version 1.0 incorporated 11 features from the Post-Stroke Checklist and 3 additional questions pertaining to oral health, eating or swallowing problems, and other challenges after stroke. The final question offered a free-text option for users to add anything else they wished to share before the visit. In addition, explanatory texts were attached to all questions, and an advisory text was added.

Version 1.0 was launched at the Swedish national patient portal 1177.se. Personal accounts are created in the patient portal using a social security number and an electronic ID, enabling notifications for activities such as form submissions. We chose this secure platform because it provides a sustainable solution and enhances accessibility for patients [13].

The aim was to develop an easy-to-use previsit tool perceived as meaningful for users, with optimized conditions for implementation in keeping with service-design principles [21]. Strokehealth already has shown a potential to enable people with stroke to prepare for visits, to capture care needs, and to provide patients with valuable information related to stroke [13]. Its validity in a real-world setting, however, remains to be established. Real world feasibility testing is needed to evaluate the tool’s usability and efficacy in capturing health problems and to facilitate its further development. The main study aim, thus, was to evaluate the previsit tool, Strokehealth, as used by people with stroke before a scheduled follow-up visit, with a focus on feasibility and relevance from the patient’s perspective. A second aim was to explore the extent to which Strokehealth could identify stroke-related health problems.

Methods

Study Design

This study is part of a larger research project aimed at developing and evaluating the digital previsit tool Strokehealth in an article series with quantitative and qualitative methods, in line with the established framework for developing and evaluating complex interventions [22]. In total, 2 separate data collection procedures were used in this study (part 1)—one from the previsit tool and the other from a subsequently administered closed web-based patient satisfaction survey. The CHERRIES (Checklist for Reporting Results of Internet E-Surveys) checklist was followed in this study (Multimedia Appendix 1) [23]. Web-based surveys provide an efficient way to track user views and allow for the analysis of large volumes of information [24].

Ethical Considerations

The study was approved by the Swedish Ethical Review Authority (2017/556-17, 2020-03324, and 2021-06723-02). Patients received information about the research project after logging into the patient portal and were asked whether they agreed to participate in the study. Informed consent was obtained from patients as part of the process. It was not possible to continue filling out the survey if this question remained unanswered. The initial information clearly stated that completing the survey was voluntary, and that participation would not affect their medical care. No incentives were offered.

Study Context

Between November 2020 and June 2021, consecutive patients discharged from a stroke unit and scheduled for a follow-up visit with a stroke team member in primary health care received a digital message instructing them to log in to the patient portal and complete the previsit tool, Strokehealth, before their appointment (within 1-2 weeks). After submitting Strokehealth, patients received a second digital message prompting them to log in to the portal once again, where they were encouraged to fill out the patient satisfaction survey on the same platform (Multimedia Appendix 2). The log-in requirement prevented duplicate entries from the same user so that only unique visitors were registered.

Sampling and Participants

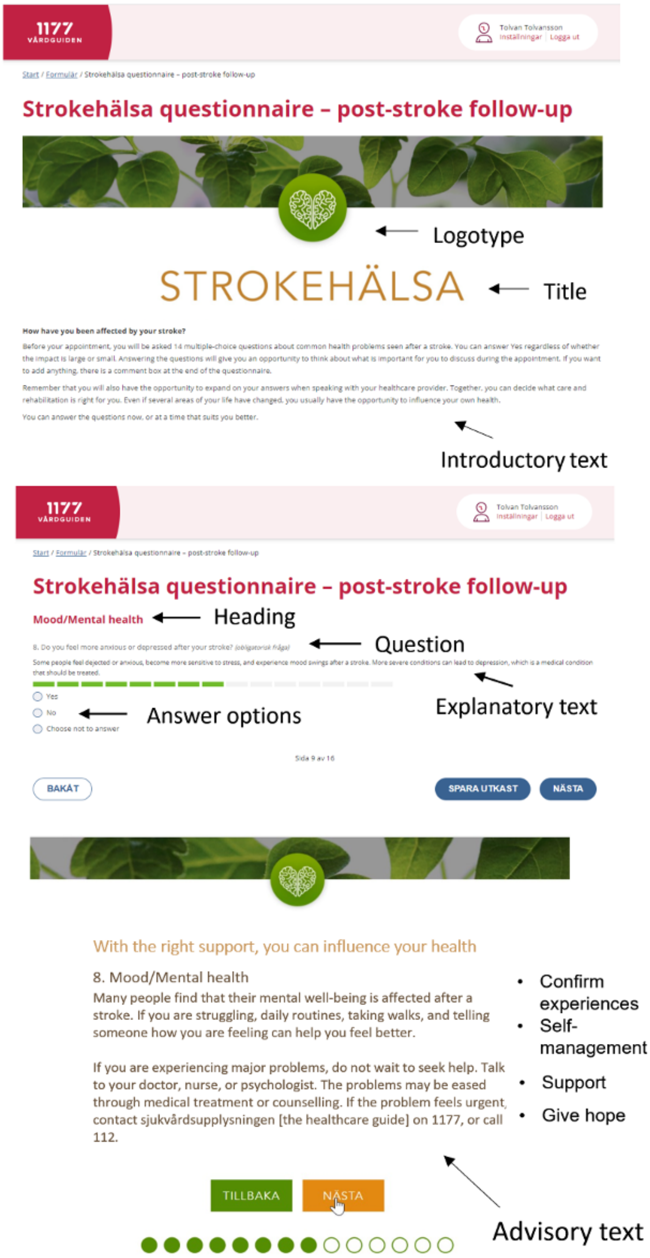
A convenience sampling was used with the aim of gathering data from a minimum of 50 patients [25]. Furthermore, 4 publicly-funded health care units were invited to participate (3 hospital-based stroke units and 1 primary health care unit). After inclusion in the study, the primary health care unit withdrew from participation because of reorganizations resulting from the COVID-19 pandemic, so ultimately only the 3 hospital units were included. One hospital (unit 1) was situated in an urban area, and 2 (unit 2 and unit 3) were situated in 2 middle-sized cities serving more rural areas. In addition, 1 stroke-specialized nurse at each unit was selected to participate. The follow-up visit was scheduled between 3 weeks and 3 months after the stroke event, depending on the standard routines within each unit.

Data Collection From the Previsit Tool Strokehealth

After receiving the digital message about Strokehealth, patients could choose to fill out the tool directly or to do so later. Help from next-of-kin was allowed but asked to be noted in the response. Each of the 3 nurses could monitor response status, and if no response was obtained, they could send a total of 2 reminders.

Strokehealth version 1.0 (Figure 1) began with a brief introductory text. Within this text, patients were encouraged to read additional information on stroke prevention and self-management strategies, with a clickable web link leading to advisory texts. Following the introductory text, patients answered 14 questions related to various health areas. These questions provided three response options, “yes,” “no,” and “choose not to answer,” depending on whether respondents indicated a health problem or not. Finally, at the conclusion of Strokehealth, patients were given the opportunity to provide a free-text response to the question: “Is there anything else you want to add before your visit?”

Figure 1. A screenshot of the elements in the previsit tool Strokehälsa (English version; reproduced from Kjörk et al [13], which is published under Creative Commons Attribution 4.0 International License [26]).



Data Collection From the Patient Satisfaction Survey

The survey was constructed in accordance with the technology acceptance model, with a focus on ease of use and perceived usefulness and acceptability [27] regarding Strokehealth. Aspects of ease of use were explored, for example, related to the need for support from others when using Strokehealth, satisfaction with navigating in the tool, and access to the advisory text (web link). Perceived usefulness was explored, for example, related to whether their health problems were captured and their satisfaction with using Strokehealth before a visit. Acceptability was explored, for example, regarding their satisfaction with the layout, answer options, and if they would recommend its use. The usability and technical functionality of the survey were tested in collaboration with a co-designing partner patient before the survey was fielded.

The survey was visible as a 3-page survey containing 15 items with fixed response options and 4 items for free-text answers. The items were divided into different focus areas and each page included 3, 4, and 12 items, respectively, with the ability to scroll before viewing the next page. The items covered demographic data (age, sex, living conditions, education, and source of income), internet use habits, devices used to complete Strokehealth, perceptions about the usability of Strokehealth, and any stroke-related health problem that respondents felt was not addressed. For the last 2 questions, the 5 response options "very satisfied," "satisfied," "dissatisfied," "very dissatisfied," and "don't know" were presented with no subsequent adaptive questioning. The 4 items for free-text answers were "Did you miss any health-related problems?"; "Name three advantages of Strokehealth"; "Name three disadvantages of Strokehealth"; and "Do you have any suggestions for improvement?" Before a completed survey was submitted, respondents could view a summary of the responses and change any response if they wished. Once the survey was completed and submitted, respondents could not enter the survey interface again. Patients who filled out Strokehealth but did not complete the subsequent survey could receive a reminder.

Additional Data Collection

To describe the diagnoses and other characteristics correctly, clinical data were retrieved from the national stroke quality registry Riksstroke. Information included prestroke living conditions, stroke characteristics, and length of hospital stay. If data were missing, complementary information was collected from the medical records (although not all information could be retrieved due to missing data). Furthermore, on a few occasions, it became evident that some respondents did not fully understand one of the questions in the survey. Clarification was then obtained through personal interviews (conducted by EKK).

Data Analysis

Data were anonymized before all analyses to ensure patient privacy protection. Descriptive statistics were performed using IBM SPSS Statistics (version 24). Categorical values are presented by frequencies and proportion and quantitative variables as medians with ranges or IQRs. With respect to Strokehealth, the time interval from patient notification to registration of a response, as well as the responses themselves, were compiled based on the 3 units. We also conducted an analysis comparing data for those who did and did not respond to Strokehealth. For the nonresponse analysis, we used 2 statistical tests, which were the chi-square test of independence for comparison of 2 categorical variables and the Mann-Whitney *U* test for comparing a categorical with a continuous variable.

Free-text answers were analyzed using a qualitative thematic analysis as outlined by Braun and Clarke [28]. First, we (EKK and PP) analyzed the free-text answers from Strokehealth in an inductive manifest manner with a realist approach to create descriptive themes. In addition, we (KK and EKK) analyzed the free-text answers from the survey using an inductive and latent approach, and the underpinning philosophy was based on a constructionistic approach. The analysis started with the authors reading through the responses several times to become familiar with the data and then discussing underlying meanings and patterns, followed by manual coding. The data were then grouped based on potential themes. The authors reviewed themes several times after discussing them.

Results

Participant Characteristics and Responding Approaches

In total, 171 people with stroke were consecutively recruited and received the digital previsit tool, Strokehealth. Of these, 40% (68/171) were women, and 60% (103/171) were men. Finally, 47% (80/171) completed Strokehealth, with the largest proportions at unit 1 (34/80, 42%) and unit 2 (33/80, 41%). Patients from unit 1 were younger compared with the other 2 units (Figure 2 and Table 1). The subsequent survey was completed by 73% (58/80; Figure 2). One person started but did not complete the survey, and their responses were not included in the analyses. Most respondents to both Strokehealth (45/80, 56%) and the survey (39/80, 67%) were men. All completed the multiple-choice responses, but only about half responded to one or more of the free-text questions. The most common level of education among respondents was a university degree, and pension compensation was the most prevalent source of income (Table 1).

Figure 2. Flowchart of participants and dropouts divided based on the unit where they had their follow-up visit.

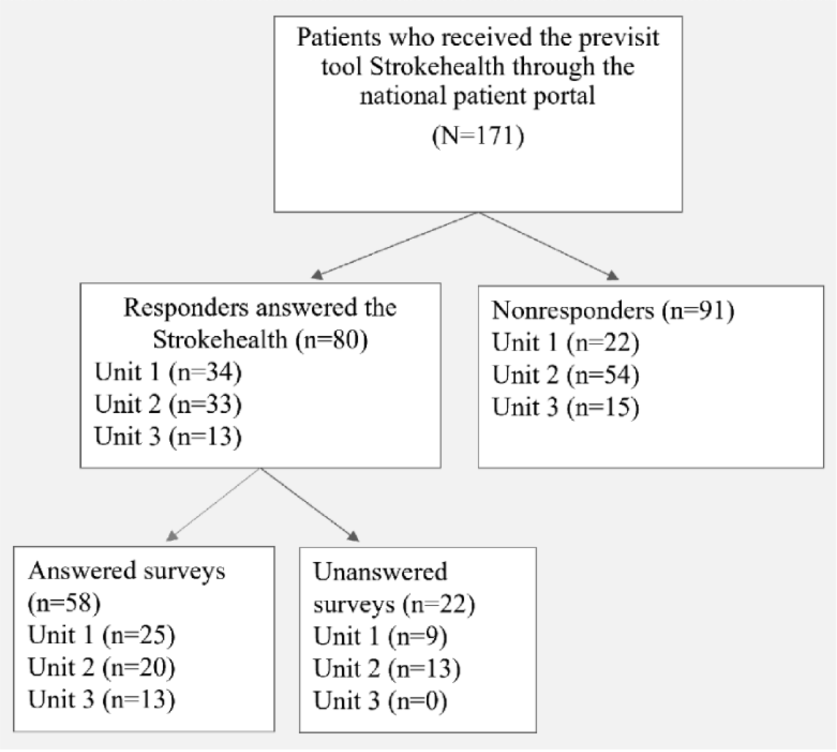


Table 1. Participant characteristics.

Characteristics	Patients answering the previsit tool (n=80)	Respondents of the subsequent survey (n=58)
Age in years, median (range[a])		
All patients	67 (32-91)	67 (43-91)
Unit 1	58 (32-85)	57 (43-85)
Unit 2	71 (44-91)	70 (53-91)
Unit 3	73 (51-87)	73 (51-87)
Sex, n (%)		
Male	45 (56)	39 (67)
Education (highest level), n (%)		
Mandatory	—[b]	13 (22)
High School	—	17 (29)
University	—	25 (43)
Other	—	3 (5)
Source of income at inclusion, n (%)		
Work	—	24 (41)
Sick leave	—	3 (5)
Retirement	—	29 (50)
Studies	—	1 (2)
Other	—	1 (2)
Prestroke living conditions[c], n (%)		
Without assisted care in own home	80 (100)	58 (100)
Living alone	22 (28)	13 (24)
Independent	71 (89)	51 (96)
Stroke characteristics (onset), n (%)		
Cerebral infarct	65 (81)	51(87)
Intracerebral hemorrhage	4 (5)	4 (7)
Other cerebrovascular events[d]	5 (6)	3 (5)
Previous stroke	10 (12)	6 (10)
Stroke severity[c], NIHSS[e], median (range[a])	1 (0-13)	1 (0-13)
Stroke-related outcomes		
Length of hospital stay in days, median (range[a])	5 (1-35)	5 (1-35)
Discharged to own home, n (%)	80 (100)	58 (100)

[a]Range: minimum-maximum.

[b]Not applicable.

[c]Missing data: Prestroke living conditions (n=5), stroke severity (n=7).

[d]Other cerebrovascular events included transient ischemic attack (n= 3), subarachnoid bleeding (n=1), and sinus thrombosis (n=1).

[e]NIHSS: National Institutes of Health Stroke Scale; measured ≤24 hours of admission, normal values 0-42. Values are presented as numbers and valid percentages unless stated otherwise.

The time interval from patient notification of Strokehealth to registration of a patient response varied greatly, with a median response time of 13 hours (range 9 minutes to 14 days). Internet use was high, with most respondents (48/58, 82%) reporting using the internet several times a day, and others (6/58, 10%) using it a few times a week, a few (2/58, 3%) times each month, or a few times each year (2/58, 3%). The most commonly used device was a smartphone (29/58, 50%), followed by a computer (24/58, 41%) and a tablet such as an iPad (Apple Inc; 5/58, 9%).

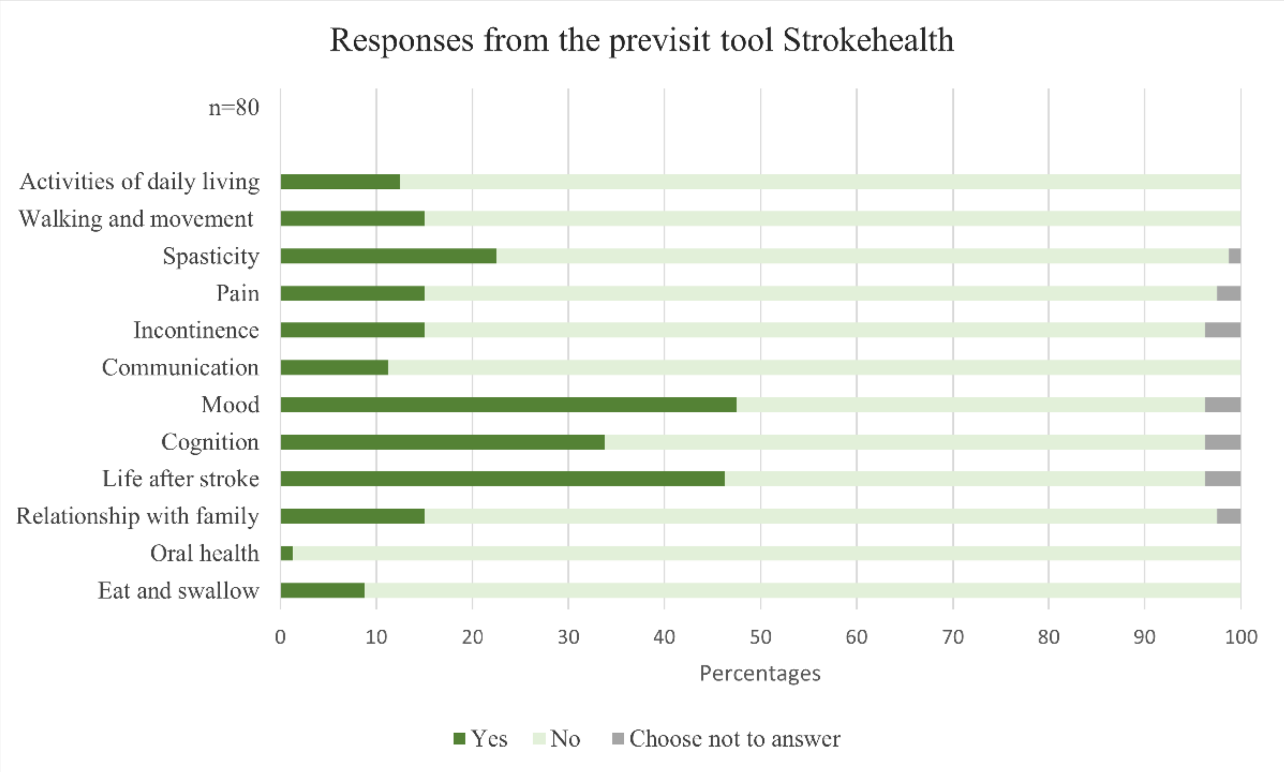
Comparing Patients Responding to the Previsit Tool Strokehealth Versus Nonrespondents

Respondents and nonrespondents differed significantly in age (*P*<.001), with a median age of 67 (IQR 56-75) years among respondents compared with 77 (IQR 69-83) years among nonrespondents. Gender proportions did not differ between the 2 groups (*P*=.32). In comparison with the included patients whose data are given in Table 1, nonrespondents (91/171, 53%) had a higher National Institutes of Health Stroke Scale score (stroke severity), with a median of 2 (IQR 1-6) versus 1 (IQR 0-3), and more often had a history of a previous stroke (17/87, 20%, vs 10/80, 12% for respondents). The proportion with cerebral infarct was 91% (82/90), compared with 81% (65/80) among respondents. Furthermore, 97% (84/87) were discharged to their own home, whereas 100% (80/80) of respondents were discharged home.

Stroke-Related Health Problems Identified Within the Previsit Tool Strokehealth

Among those completing Strokehealth (80/171, the most reported health problems were as follows: mood, with 48% (38/80) experiencing feelings of anxiety or depression after their stroke; life after stroke, with 46% (37/80) noticing difficulties in carrying out tasks they deemed important; and cognition, with 34% (27/80) facing challenges in thinking, concentrating, and remembering. Conversely, oral health was the least frequently reported area, with only 1 respondent indicating a related issue (Figure 3).

Figure 3. Response displayed by percentage of respondents to the different health areas brought up in the previsit tool Strokehealth.



The initial question on Strokehealth, about the patient's interest in receiving information on stroke prevention, received the highest proportion of affirmative responses (70/80, 88%). Patients reported an average of 2 health problems per person (median 2, IQR 0-3.8). When considering different care units, patients at unit 1 had a lower average number of problems (median 1.5, IQR 0-3) compared with patients at unit 2 (median 2, IQR 1-5) and unit 3 (median 2, IQR 0-5.5).

The free-text option in the previsit tool ("Is there anything else you want to add before your visit?") was used by 45% (36/80) of patients. The thematic qualitative analysis generated 3 categories (Textbox 1).

Textbox 1. Categories created by thematic qualitative analysis based on answers in free-text in the previsit tool Strokehealth.

Health- and risk-related worries:

- questions about medication side effects;
- medical records and requests for advice or practical support;
- concerns about health, including surgery risks; and
- specific questions centered on recommendations for returning to work, medical certificates, and use of a ladder.

Health problems and their daily impact:

- descriptions of health issues and perceived impact on daily life activities;
- health problems noted in relation to language, vision, taste, weakness, mobility, sensory perceptions, dizziness, balance, headaches, irritability, and sensitivity to sound and light; and
- impact on daily life related to isolation because of driving limitations and difficulties with house cleaning, baking, taking walks, shopping, and an inability to last the entire day.

Explanation of personal circumstances:

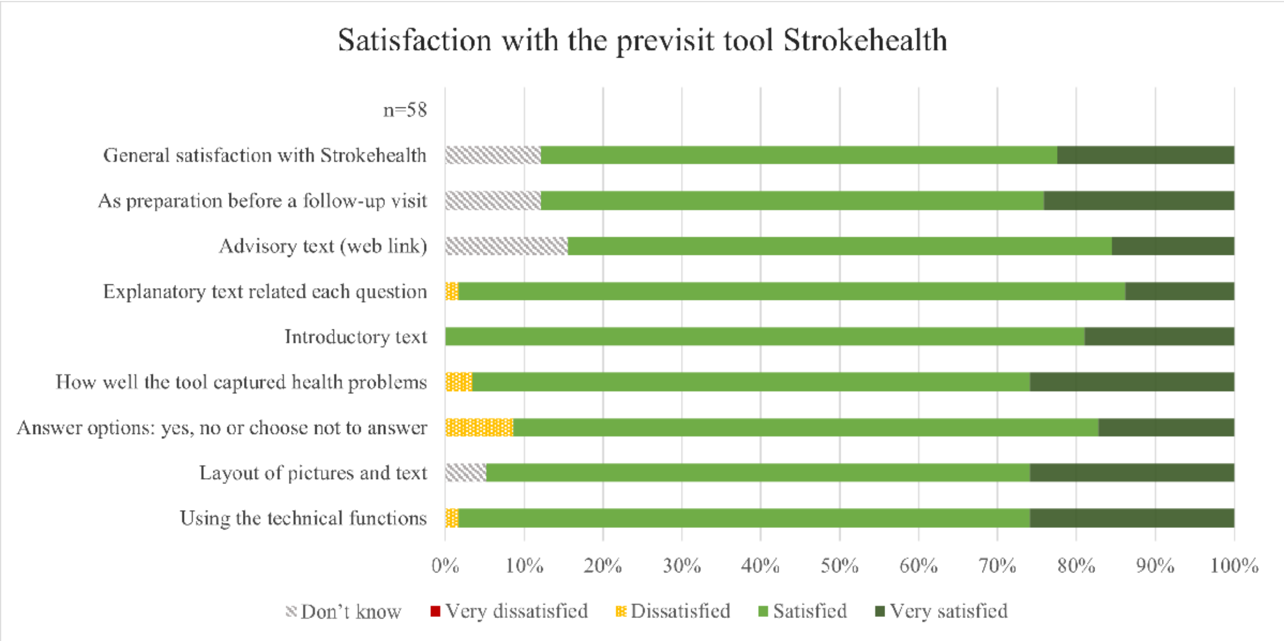
- factors that may have influenced their responses or general comments,
- additional diagnoses mentioned such as multiple sclerosis, and
- expressed a desire to provide more nuanced responses to certain questions during the visit.

Satisfaction With the Previsit Tool

The majority of respondents (56/58, 97%) expressed satisfaction or high satisfaction with how well the previsit tool captured their health problems after stroke. In the free-text option, a few respondents identified some missing health areas in the previsit tool. These included vision (1/58), fatigue (2/58), mental health (1/58), and the location of pain, if present (1/58). The remaining responses to this question (6/58) pertained to the design of the previsit tool rather than to a missing health area.

Overall, the feedback and satisfaction with Strokehealth were positive (Figure 4). None of the 58 individuals reported being very dissatisfied. Similarly, the number of respondents expressing dissatisfaction with any aspect of Strokehealth was minimal. The majority of respondents expressed either satisfaction or high satisfaction on the items addressing their experience with Strokehealth (Figure 4).

Figure 4. Satisfaction among respondents to the survey regarding different aspects of the previsit tool Strokehealth. Answers are displayed by percentage (n=58). Full formulations of the survey questions can be found in Multimedia Appendix 2.



A question that received the highest number of "Don't know" responses (9/58) pertained to satisfaction with the advisory text about self-care and available support resources. This information was provided through a separate link, which some respondents did not see (10/58). The highest level of dissatisfaction (5/58) related to the layout of the response options in Strokehealth ("Yes," "No," or "Choose not to answer").

Most respondents (43/58, 74%) expressed an intention to recommend Strokehealth to others who had experienced a stroke, but 26% (15/58) responded with "Don't know." The qualitative interviews (conducted by EKK) clarified that some respondents did not fully understand the question or found it challenging to respond on behalf of someone else. Of note, none of the 58 survey respondents stated that they would not recommend Strokehealth to others who had experienced a stroke.

Survey: Qualitative Text Analysis

In total, there were 20, 19, 17, and 10 written responses, respectively, to the 4 free-text questions on the survey. After analyzing these answers, the authors developed 4 themes that are described further with illustrative quotations.

Structure That Ensures That Issues Will Be Emphasized

Strokehealth was perceived as a valuable tool for structured follow-up care. Respondents expressed that it served as a supportive mechanism, capturing important aspects that might otherwise have gone unnoticed. In addition, respondents believed that Strokehealth facilitated increased involvement in their own care. The tool was reported to provide comprehensive and informative content, offering enhanced insights into their condition and available support. This sentiment is echoed in the following quotes: "Some things that had not been mentioned previously by health professionals were emphasized" and "good and clear debrief." Furthermore, responses such as "good to prepare oneself in peace and quiet before the conversation" or "It [Strokehealth] helps me to reflect on my situation" indicated that Strokehealth was perceived as a structured tool for reflection and preparation.

Importance of User-Friendliness and Accessibility

Statements such as "Good that it is available on the web and easy to access the national patient portal" or "Easy to fill out" indicated that the previsit tool was perceived as user-friendly and readily accessible. However, 1 participant expressed the desire for "Better information than an email that it [Strokehealth] was available to fill out," suggesting that some individuals may have been unaware of Strokehealth. To enhance usability, 1 participant suggested the ability to reopen and make alterations after completing the tool. In terms of usefulness, patient respondents thought that Strokehealth was a time-saving resource for health professionals. Nevertheless, suggestions for improvement included the desire for prompt feedback on completed responses. Furthermore, one participant's comment, "Define more what 'strokehälsa' is," may reflect either a lack of sufficient background information about Strokehealth or a misunderstanding of its identity as a digital tool.

A Digital Approach as a Means of Communication

Using a digital tool such as Strokehealth can contribute to patients' feeling more acknowledged and heard. One of the patients stated that 1 advantage of Strokehealth was that "the questions are asked at all and [I] am given the opportunity to be answered." Another expressed "It feels good to get to answer questions about the stroke." One of the respondents remarked, "This is the first contact I've had with medical care since I was discharged from the stroke unit. I wonder if this is the best contact?" Similarly, another respondent stated, "… Forms, either on paper or digital, create an impersonal impression in an often painful situation. I suppose the spectrum of how this is received by the patients is wide, depending on the consequences of the stroke." Thus, attitudes toward the use of digital tools and the sense of inclusion or exclusion within the digital context played a significant role in shaping the perception of Strokehealth.

Experience of Answering Influenced by Response Options

Respondents frequently mentioned the answer options and a need to explain further and elaborate on their answers. They expressed dissatisfaction with the limited response options of "Yes," "No," and "Choose not to answer" for problem areas, as they felt that these options did not adequately capture the complexity of their individual situations. As 1 respondent noted, "The answers to the questions were a bit too much 'all or nothing', there weren't enough alternatives in between." Participants also highlighted the absence of opportunities to provide explanations or elaborate on their answers, stating that the single free-text question at the end of the tool was insufficient. One participant commented, "If the response was 'No', one wanted to be able to explain it in direct connection to that question." Furthermore, patients complained about the limited word count allowed for the sole free-text response. This limitation also was visible in practice, in that several free-text replies ended abruptly in the middle of a sentence. In addition, respondents found some of the explanatory texts unclear or inconsistent with the questions, leading to confusion and the need to make assumptions about their intended meaning. Furthermore, respondents identified specific missing questions in the tool, such as inquiries about pain localization, while acknowledging the challenge of encompassing every individual's unique problem within a standardized format.

Discussion

Principal Findings and Comparison With Previous Work

The digital previsit tool Strokehealth was designed for patients with stroke to complete before a follow-up visit to support a focused discussion with the stroke team member during the visit. The current findings from this real-world feasibility testing indicate that Strokehealth is a user-friendly and useful tool, essentially confirming previous findings [13]. Furthermore, Strokehealth effectively captured stroke-related health problems and prepared patients satisfactorily for the visit with the health care professional. However, data collected from Strokehealth and the subsequent survey also highlight important aspects to consider in the continuing co-design process to ensure that the tool meets patient needs.

Respondents considered that Strokehealth satisfactorily facilitated the process of identifying stroke-related health problems, even though the vast majority had experienced what was assessed to be a mild stroke. In addition, these findings support that subtle symptoms such as cognitive impairments are common even after clinical recovery from stroke [29]. This study thus confirms the previously recognized potential benefits of a tool like Strokehealth [13]. All the health problems noted in Strokehealth had at least 1 response, demonstrating the relevance of the original Post-Stroke Checklist [9], as well as the 3 additional questions. However, patients also mentioned other health issues in Strokehealth and the survey. These issues are not specified as questions in Strokehealth but can be regarded as indirectly assessed (eg, fatigue, vision, or headache) in a manner analogous to the checklist [10,11]. Furthermore, the high proportion (36/80, 45%) using the free-text option at the end of Strokehealth illustrates that any perceived health issues can be identified in some way with this tool. The free-text option enables each patient's unique needs to be captured, in keeping with person-centered care [7,8]. Accordingly, the current findings indicate that the number of questions and response options in Strokehealth work satisfactorily in combination with free-text options to identify a patient's health problems. However, our qualitative findings indicate that the level of answer options, information, and free text needs to be considered (unpublished data) together with the results presented here.

Of note, digital tools such as Strokehealth should be considered an integrated part of the health service [16,21,30]. Our results indicate that Strokehealth has the potential to empower patients and increase their engagement in the follow-up process. In contrast to conventional health care practices in which patients are summoned for prescheduled visits, Strokehealth gives patients the chance to be informed about common consequences after stroke and to reflect on them in advance. The invitation to complete Strokehealth thus can be seen as a starting point for a shared decision-making process, enhancing patient-provider communication during the visit [30] by motivating patients to think about their life after stroke. However, the theme "digital approach as means of communication" raises awareness about patient expectations. Although Strokehealth provided additional support compared with traditional care, some people expected more from their first contact with health care. The principle of "digital first" has become an increasingly common strategy to facilitate proactive support or triage before a physical in-person visit [31], but it may not be consonant with current patient expectations. A patient's experience with health care relies to a large extent on a health care system's ability to meet patient expectations [32]. This reliance underscores that the development and use of Strokehealth need to be handled as an integrated part of overall follow-up [21,30] and adapted for the local context [22] (eg, provide information regarding follow-up routines before discharge) to better meet patient expectations.

The fact that 47% (80/171) of the patients filled in Strokehealth in advance is encouraging and shows that they perceive Strokehealth as acceptable. Also encouraging is that most respondents used a smartphone when filling in the form, which suggests that Strokehealth contributes to more accessible health care. The patient is no longer restricted to health care facilities or their own home to take part in health services; instead, they can prepare for a visit at the time and place of their choosing. However, the impact of contextual factors on effectiveness, acceptability [22], and eHealth literacy [19,33] is important to consider. A patient's ability to engage with digital health services is influenced by how well the system meets patient needs and not only a patient's ability to understand and use health information [16,19]. Easy access prompts a design in which the patient is asked to fill in Strokehealth in a time and place that supports reflection. Since a few patients submitted Strokehealth very quickly (within 9 minutes), there may be a need to add clarification in the introductory information, for example, a sentence encouraging patients to choose an appropriate time to answer. Understanding the contextual factors influencing these response experiences can provide valuable insights into user's behavior and preferences, which in turn can inform continuous design and increase accessibility to health care services [21,22]. The initial attempt with Strokehealth [13] to maintain an easy-to-use tool guided by theoretical frameworks (eg, technology acceptance model) [27] with a person-centered focus [7] will also guide the ongoing co-design process.

Strokehealth was perceived to be feasible in a group of people who were almost all using the internet daily. In Sweden, 7 out of 10 people who were older than 75 years use the internet, and many retired people are as accustomed as younger people to doing so [34]. Although patients who completed Strokehealth were statistically significantly younger than those who did not, the number of elderly people using digital services is increasing [34,35]. With this in mind, Strokehealth was co-designed with stakeholders to meet future demands. Others have emphasized the importance of close collaboration with all stakeholders when developing effective digital health services [21]. However, several aspects still must be considered in enhancing user friendliness for the broad range of people with stroke. Among factors to consider in the design process are the ability to process information, the need to feel secure, and patients' motivations to be engaged in their own care [19]. Although next-of-kin were involved as support in some cases in this study, this involvement might not have been necessary if potential barriers were better addressed. Not all people are motivated or able to use digital health services, and for this reason, Strokehealth is available in multiple languages, in a picture-supported version, and in paper format [36]. Altogether, offering Strokehealth in different modalities is aimed at overcoming certain accessibility challenges to better meet the needs of the broad range of people with stroke.

Overall, this study contributes to the growing body of evidence supporting the effectiveness and implementation of digital tools in health care [21,37,38]. It is now recommended within the Swedish health care system that Strokehealth is administered before a planned follow-up visit and is thus expected to aid health care teams in adopting a previsit planning approach. Consequently, the preparatory opportunities that Strokehealth provides and the time during the actual visit can be used more effectively. Moreover, with increased knowledge, the hope is that patients can reduce their reliance on health care support and rely more on self-management strategies. User-friendly tools indeed have been shown to increase engagement and adherence to self-management strategies [17,39], which is also one of the main purposes of Strokehealth. By building on existing research, future studies can further explore and refine the role of digital tools, such as Strokehealth, in improving the quality of stroke care. More in-depth qualitative investigation with patients as well as with health care professionals is needed to explore the experiences of using Strokehealth, including among diverse social groups and people with communication difficulties or complex needs.

Strengths and Limitations

A major strength of this feasibility testing is the evaluation of Strokehealth in a clinical context with real world patients before a scheduled follow-up visit. Furthermore, respondents represented people across a range of ages, sexes, education levels, and different health care settings in urban and rural areas. Consequently, the generalizability of the tool's usability is enhanced. However, there are some limitations. First, despite consecutive sampling, the follow-up routines at each site influenced the sample, leading to a lack of patients with more severe stroke or a higher level of dependency in daily life, and to a range of time intervals since the stroke. Nevertheless, this representation is in line with the stroke population in Sweden, where most patients are classified as having had a mild stroke [40]. Second, during inclusion, researchers did not verify the diagnosis, and some people invited to follow-up had a potential transient ischemic attack that was only recognized later through checks of the health registry or charts. In balance against this limitation, another strength of the study is the comprehensive description of clinical and demographic data, which in turn provides knowledge about the potential target group for Strokehealth. Third, the influence of recall bias cannot be ruled out for the responses to Strokehealth or the survey. Given the number of free-text answers, though, it seems that most respondents had a clear opinion. A final strength is the high response rate of the survey. Since most strokes are mild in nature, as confirmed in our study, the current version of Strokehealth would be considered a user-friendly tool for most people with stroke.

Conclusions

This real-world feasibility study shows that the previsit tool, Strokehealth, is feasible when used before a follow-up visit after a stroke. Satisfaction with the tool's ability to capture health problems was high among patients, with a majority having experienced a mild stroke and being regular users of the internet. However, during subsequent co-design processes, features such as free-text options and information need to be considered in parallel with qualitative data. Further research is needed to explore the use and benefits for a broader range of users, including people with communication difficulties, and to gain health professionals' perspectives.

Acknowledgments

We thank people with stroke for their participation and their willingness to share their experiences. We also thank the health care professionals for handling the previsit tool besides their ordinary routines.

This study was funded in part by grants from the Swedish state under an agreement among the Swedish government and the county councils, the ALF agreement (ALFGBG-965653); the Local Research and Development Board for Gothenburg and Södra Bohuslän (VGFOUGSB-76712); Sahlgrenska Hospitals Foundation (SU-892141, SU-793481); the Swedish Stroke Association; an unconditional grant from Allergan; the Hjalmar Svenssons Foundation; Göteborg Foundation for Neurological Research; Rune and Ulla Amlövs Foundation; Per-Olof Ahls Foundation; John and Brit Wennerströms Foundation; the Innovation foundation at region Västra Götaland (VGRINN-968305); and the Greta and Einar Askers Foundation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CHERRIES (Checklist for Reporting Results of Internet E-Surveys) checklist.
[DOCX File , 22 KB - humanfactors_v11i1e55852_app1.docx]

Multimedia Appendix 2
Patient satisfaction survey translated into English.
[DOCX File , 22 KB - humanfactors_v11i1e55852_app2.docx]

References

1. GBD 2019 Stroke Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the global burden of disease study 2019. Lancet Neurol 2021;20(10):795-820 [FREE Full text] [doi: 10.1016/S1474-4422(21)00252-0] [Medline: 34487721]
2. Crichton SL, Bray BD, McKevitt C, Rudd AG, Wolfe CDA. Patient outcomes up to 15 years after stroke: survival, disability, quality of life, cognition and mental health. J Neurol Neurosurg Psychiatry 2016;87(10):1091-1098. [doi: 10.1136/jnnp-2016-313361] [Medline: 27451353]
3. Package of interventions for rehabilitation. Module 3. Neurological conditions. Geneva: World Health Organization; 2023. URL: http://www.who.int/publications/i/item/9789240071131 [accessed 2023-11-03]

4. Norrving B, Barrick J, Davalos A, Dichgans M, Cordonnier C, Guekht A, et al. Action plan for stroke in Europe 2018-2030. Eur Stroke J 2018;3(4):309-336 [FREE Full text] [doi: 10.1177/2396987318808719] [Medline: 31236480]
5. Pindus DM, Mullis R, Lim L, Wellwood I, Rundell AV, Abd Aziz NA, et al. Stroke survivors' and informal caregivers' experiences of primary care and community healthcare services - a systematic review and meta-ethnography. PLoS One 2018;13(2):e0192533 [FREE Full text] [doi: 10.1371/journal.pone.0192533] [Medline: 29466383]
6. Epping-Jordan JE, Pruitt SD, Bengoa R, Wagner EH. Improving the quality of health care for chronic conditions. Qual Saf Health Care 2004;13(4):299-305 [FREE Full text] [doi: 10.1136/qhc.13.4.299] [Medline: 15289634]
7. Ekman I, Swedberg K, Taft C, Lindseth A, Norberg A, Brink E, et al. Person-centered care--ready for prime time. Eur J Cardiovasc Nurs 2011;10(4):248-251. [doi: 10.1016/j.ejcnurse.2011.06.008] [Medline: 21764386]
8. Martín-Sanz MB, Salazar-de-la-Guerra RM, Cuenca-Zaldivar JN, Salcedo-Perez-Juana M, Garcia-Bravo C, Palacios-Ceña D. Person-centred care in individuals with stroke: a qualitative study using in-depth interviews. Ann Med 2022;54(1):2167-2180 [FREE Full text] [doi: 10.1080/07853890.2022.2105393] [Medline: 35930376]
9. Philp I, Brainin M, Walker MF, Ward AB, Gillard P, Shields AL, Global Stroke Community Advisory Panel. Development of a poststroke checklist to standardize follow-up care for stroke survivors. J Stroke Cerebrovasc Dis 2013;22(7):e173-e180 [FREE Full text] [doi: 10.1016/j.jstrokecerebrovasdis.2012.10.016] [Medline: 23265778]
10. Ward AB, Chen C, Norrving B, Gillard P, Walker MF, Blackburn S, Global Stroke Community Advisory Panel (GSCAP). Evaluation of the post stroke checklist: a pilot study in the United Kingdom and Singapore. Int J Stroke 2014;9 Suppl A100:76-84 [FREE Full text] [doi: 10.1111/ijs.12291] [Medline: 25088427]
11. Kjörk EK, Carlsson G, Sunnerhagen KS, Lundgren-Nilsson A. Experiences using the poststroke checklist in Sweden with a focus on feasibility and relevance: a mixed-method design. BMJ Open 2019;9(5):e028218 [FREE Full text] [doi: 10.1136/bmjopen-2018-028218] [Medline: 31072862]
12. Stacey D, Lewis KB, Smith M, Carley M, Volk R, Douglas EE, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2024;1(1):CD001431. [doi: 10.1002/14651858.CD001431.pub6] [Medline: 38284415]
13. Kjörk EK, Sunnerhagen KS, Lundgren-Nilsson A, Andersson AK, Carlsson G. Development of a digital tool for people with a long-term condition using stroke as a case example: participatory design approach. JMIR Hum Factors 2022;9(2):e35478 [FREE Full text] [doi: 10.2196/35478] [Medline: 35657650]
14. Hu X, Jonzén K, Lindahl OA, Karlsson M, Norström F, Lundström E, et al. Digital graphic follow-up tool (rehabkompassen) for identifying rehabilitation needs among people after stroke: randomized clinical feasibility study. JMIR Hum Factors 2022;9(3):e38704 [FREE Full text] [doi: 10.2196/38704] [Medline: 35904867]
15. Hibbard JH. Patient activation and the use of information to support informed health decisions. Patient Educ Couns 2017;100(1):5-7. [doi: 10.1016/j.pec.2016.07.006] [Medline: 27432014]
16. Rosenlund M, Kinnunen UM, Saranto K. The use of digital health services among patients and citizens living at home: scoping review. J Med Internet Res 2023;25:e44711 [FREE Full text] [doi: 10.2196/44711] [Medline: 36972122]
17. Madanian S, Nakarada-Kordic I, Reay S, Chetty T. Patients' perspectives on digital health tools. PEC Innov 2023;2:100171 [FREE Full text] [doi: 10.1016/j.pecinn.2023.100171] [Medline: 37384154]
18. Spreadbury JH, Young A, Kipps CM. A comprehensive literature search of digital health technology use in neurological conditions: review of digital tools to promote self-management and support. J Med Internet Res 2022;24(7):e31929. [doi: 10.2196/31929] [Medline: 35900822]
19. Norgaard O, Furstrand D, Klokker L, Karnoe A, Batterham R, Kayser L. The e-health literacy framework: a conceptual framework for characterizing e-health users and their interaction with e-health systems. Knowledge Management & E-Learning 2015;7(4):522-540. [doi: 10.34105/j.kmel.2015.07.035]
20. Friis K, Lasgaard M, Osborne RH, Maindal HT. Gaps in understanding health and engagement with healthcare providers across common long-term conditions: a population survey of health literacy in 29,473 Danish citizens. BMJ Open 2016;6(1):e009627 [FREE Full text] [doi: 10.1136/bmjopen-2015-009627] [Medline: 26769783]
21. Shaw J, Agarwal P, Desveaux L, Palma DC, Stamenova V, Jamieson T, et al. Beyond "implementation": digital health innovation and service design. NPJ Digit Med 2018;1:48 [FREE Full text] [doi: 10.1038/s41746-018-0059-8] [Medline: 31304327]
22. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of medical research council guidance. BMJ 2021;374:n2061 [FREE Full text] [doi: 10.1136/bmj.n2061] [Medline: 34593508]
23. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet E-surveys (CHERRIES). J Med Internet Res 2004;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
24. Howard JS, Toonstra JL, Meade AR, Whale Conley CE, Mattacola CG. Feasibility of conducting a web-based survey of patient-reported outcomes and rehabilitation progress. Digit Health 2016;2:2055207616644844 [FREE Full text] [doi: 10.1177/2055207616644844] [Medline: 29942553]
25. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976) 2000;25(24):3186-3191. [doi: 10.1097/00007632-200012150-00014] [Medline: 11124735]

26. Attribution 4.0 International (CC BY 4.0). Creative Commons. URL: http://creativecommons.org/licenses/by/4.0/ [accessed 2024-08-17]
27. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. Mis Quarterly 1989;13(3):319. [doi: 10.2307/249008]
28. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
29. Jokinen H, Melkas S, Ylikoski R, Pohjasvaara T, Kaste M, Erkinjuntti T, et al. Post-stroke cognitive impairment is common even after successful clinical recovery. Eur J Neurol 2015;22(9):1288-1294. [doi: 10.1111/ene.12743] [Medline: 26040251]
30. Griffioen I, Melles M, Stiggelbout A, Snelders D. The potential of service design for improving the implementation of shared decision-making. Design for Health 2017;1(2):194-209. [doi: 10.1080/24735132.2017.1386944]
31. Healthcare guide 1177. 1177 Vårdguiden. 2023. URL: http://www.1177.se/om-1177 [accessed 2023-11-30]
32. Bowling A, Rowe G, McKee M. Patients' experiences of their healthcare in relation to their expectations and satisfaction: a population survey. J R Soc Med 2013;106(4):143-149 [FREE Full text] [doi: 10.1258/jrsm.2012.120147] [Medline: 23564898]
33. Muscat DM, Shepherd HL, Nutbeam D, Trevena L, McCaffery KJ. Health literacy and shared decision-making: exploring the relationship to enable meaningful patient engagement in healthcare. J Gen Intern Med 2021;36(2):521-524 [FREE Full text] [doi: 10.1007/s11606-020-05912-0] [Medline: 32472490]
34. Svenskarna och Internet 2022 - en årlig Studie av Svenska Folkets Internetvanor.: The Swedish Internet Foundation URL: https://svenskarnaochinternet.se/rapporter/svenskarna-och-internet-2023 [accessed 2023-11-30]
35. Arcury TA, Sandberg JC, Melius KP, Quandt SA, Leng X, Latulipe C, et al. Older adult internet use and eHealth literacy. J Appl Gerontol 2020;39(2):141-150 [FREE Full text] [doi: 10.1177/0733464818807468] [Medline: 30353776]
36. Strokehälsa. University of Gothenburg. 2023. URL: https://www.gu.se/en/neuroscience-physiology/strokehalsatm [accessed 2023-11-30]
37. Denecke K, Romero OR, Petersen C, Benham-Hutchins M, Cabrer M, Davies S, et al. Defining and scoping participatory health informatics: an eDelphi study. Methods Inf Med 2023;62(3-04):90-99 [FREE Full text] [doi: 10.1055/a-2035-3008] [Medline: 36787885]
38. Digital health. World Health Organization. URL: https://www.who.int/health-topics/digital-health [accessed 2023-12-04]
39. Lorig KR, Sobel DS, Ritter PL, Laurent D, Hobbs M. Effect of a self-management program on patients with chronic disease. Eff Clin Pract 2001;4(6):256-262. [Medline: 11769298]
40. Reinholdsson M, Palstam A, Sunnerhagen KS. Prestroke physical activity could influence acute stroke severity (part of PAPSIGOT). Neurology 2018;91(16):e1461-e1467 [FREE Full text] [doi: 10.1212/WNL.0000000000006354] [Medline: 30232251]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys

Edited by A Kushniruk, Y Senathirajah; submitted 08.01.24; peer-reviewed by J Moll, A Oostdyk, S Davies; comments to author 03.03.24; revised version received 28.03.24; accepted 04.08.24; published 03.09.24.

Please cite as:
Pohl P, Klerfors K, Kjörk EK
Evaluation of a Digital Previsit Tool for Identifying Stroke-Related Health Problems Before a Follow-Up Visit (Part 1): Survey Study
JMIR Hum Factors 2024;11:e55852
URL: https://humanfactors.jmir.org/2024/1/e55852
doi:10.2196/55852
PMID:

Original Paper

German Version of the Telehealth Usability Questionnaire and Derived Short Questionnaires for Usability and Perceived Usefulness in Health Care Assessment in Telehealth and Digital Therapeutics: Instrument Validation Study

Jannik Zimmermann[1*], MSc; Harriet Morf[2,3*], MBA, MD; Florian Scharf[4], Prof Dr; Johannes Knitza[5,6], MHBA, PD Dr; Heidi Moeller[1], Prof Dr; Felix Muehlensiepen[6,7], MPH, Dr rer medic; Michaela Nathrath[8], Prof Dr; Till Orlemann[9], MD; Thomas Voelker[8], MD; Merlin Deckers[8], MD

[1]Department of Psychology – Theory and Methodology of Counseling, University of Kassel, Kassel, Germany

[2]Department of Internal Medicine 3, Rheumatology and Immunology, University Hospital Erlangen, Friedrich-Alexander University (FAU), Erlangen-Nuremberg, Erlangen, Germany

[3]Deutsches Zentrum Immuntherapie (DZI), University Hospital Erlangen, Friedrich-Alexander University (FAU), Erlangen-Nuremberg, Erlangen, Germany

[4]Department of Psychology – Psychological Research Methods, University of Kassel, Kassel, Germany

[5]Institute for Digital Medicine, University Hospital of Giessen and Marburg, Philipps University Marburg, Marburg, Germany

[6]AGEIS, Université Grenoble Alpes, Grenoble, France

[7]Center for Health Services Research, Faculty of Health Sciences, Brandenburg Medical School Theodor Fontane, Ruedersdorf bei Berlin, Germany

[8]Palliative Care Team for Children Kassel, Kassel, Germany

[9]Department of Internal Medicine 1, University Hospital Erlangen, Friedrich-Alexander University (FAU), Erlangen-Nuremberg, Erlangen, Germany

[*]these authors contributed equally

Corresponding Author:
Merlin Deckers, MD
Palliative Care Team for Children Kassel
Mönchebergstr. 41-43
Kassel, 34125
Germany
Phone: 49 56198017558
Email: merlin.deckers@kleine-riesen-nordhessen.de

Abstract

Background: The exponential growth of telehealth is revolutionizing health care delivery, but its evaluation has not matched the pace of its uptake. Various forms of assessment, from single-item to more extensive questionnaires, have been used to assess telehealth and digital therapeutics and their usability. The most frequently used questionnaire is the “Telehealth Usability Questionnaire” (TUQ). The use of the TUQ is limited by its restricted availability in languages other than English and its feasibility.

Objective: The aims of this study were to create a translated German TUQ version and to derive a short questionnaire for patients—“Telehealth Usability and Perceived Usefulness Short Questionnaire for patients” (TUUSQ).

Methods: As a first step, the original 21-item TUQ was forward and back-translated twice. In the second step, 13 TUQ items were selected for their suitability for the general evaluation of telehealth on the basis of expert opinion. These 13 items were surveyed between July 2022 and September 2023 in 4 studies with patients and family members of palliative care, as well as patients with chronic autoimmune diseases, evaluating 13 health care apps, including digital therapeutics and a telehealth system (n_1=128, n_2=220, n_3=30, and n_4=12). Psychometric exploratory factor analysis was conducted.

Results: The analysis revealed that a parsimonious factor structure with 2 factors (“perceived usefulness in health care” and “usability”) is sufficient to describe the patient’s perception. Consequently, the questionnaire could be shortened to 6 items without compromising its informativeness.

Conclusions: We provide a linguistically precise German version of the TUQ for assessing the usability and perceived usefulness of telehealth. Beyond that, we supply a highly feasible shortened version that is versatile for general use in telehealth, mobile health, and digital therapeutics, which distinguishes between the 2 factors “perceived usefulness in health care” and “usability” in patients.

Trial Registration: German Clinical Trials Register DRKS00030546; https://drks.de/search/de/trial/DRKS00030546

(JMIR Hum Factors 2024;11:e57771) doi:10.2196/57771

KEYWORDS

mHealth; mobile health; telehealth; usability; questionnaire validation; technology acceptance model; validity; questionnaire translation; Net Promoter Scale; NPS; usefulness; autoimmune chronic diseases; questionnaire; German; digital therapeutics; therapeutics; feasibility

Introduction

Telehealth, Mobile Health, and Digital Therapeutics

Telehealth is an umbrella term defined as “the provision of healthcare remotely by means of telecommunications technology,” whereas mobile health (mHealth) is an overlapping definition for “the use of mobile devices so that patients can solicit services electronically, use apps to verify information, and manage or monitor treatment or problems or other health-related issues” [1,2].

The exponential growth of telehealth and mHealth is revolutionizing health care delivery, because they have the potential to remove geographical barriers, increase access to medical services, and improve overall care quality [3]. Particularly, during the COVID-19 pandemic, there was a significant increase in the use of telehealth and mHealth, but the evaluation and its methodology have not matched the pace of its uptake [4,5]. Patients can thus receive support throughout the entire patient pathway, including app-supported diagnoses, therapy, and monitoring. Approved digital therapeutics (German: Digitale Gesundheitsanwendungen) are apps to improve treatment; their costs are covered by the statutory health insurance system in Germany. In the following discussion, we use “telehealth” to encompass the terms “mHealth” and “digital therapeutics.”

To achieve the greatest possible benefit from telehealth, usability is the key factor, especially with patients who have a cognitive limitation, are incapacitated by their disease, or are children [6-8]. This means that even evidence-based technology is not particularly effective for patient outcomes if it is difficult to use. This could be due to the technology itself or varying levels of eHealth literacy among users. According to Norman and Skinner [9], eHealth literacy is “the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem.”

Furthermore, measuring usability also protects patients from errors or experiencing harm. For example, if a certain telehealth system saves or displays medical data incorrectly and this leads to incorrect treatment, this inadequate usability can also disadvantage patients [10].

Definition of Usability

The International Organization for Standardization (ISO) norm 9241-11 defines usability as [11] “the extent to which a system, product or service can be used by specified users to achieve specific goals with effectiveness, efficiency and satisfaction in a specified context of use.”

Yet, many researchers have used additional attributes to assess usability [12]. A systematic literature study by Weichbroth showed that in descending order of priority, learnability, memorability, cognitive load, errors, simplicity, and ease of use were additional attributes used to assess usability in mobile settings [13]. However, this literature study excluded publications from medical and health subject areas. Sousa et al [14] conducted a systematic review of usability questionnaires for eHealth and showed that many existing usability questionnaires share these attributes but generally lacked effectiveness, cognitive load, simplicity, and ease of use. Interestingly, Sousa and Lopez [14] reported that the majority of usability questionnaires not only assessed usability but also the perceived usefulness in health care. The questionnaires included questions aiming to assess whether telehealth was helpful in fulfilling health care needs, which is not part of the usability definitions listed. The usability attributes aim only to assess the app’s efficiency, for example, “duration spent on each screen” or the app’s effectiveness, for example, “number of steps required to complete a task” [13].

Usability Questionnaires

On the one hand, authors of various studies used the single-item “Net Promoter Scale” (NPS) and the derived Net Promoter Score to evaluate telehealth [15-18]. However, the psychometric correlates of the Net Promoter Score are not clear, and it is also thought to measure satisfaction and acceptance [15]. We believe that the NPS seems to be a valuable instrument because it offers a straightforward, quantifiable, and very short measure of the user experience and is easy for patients to understand and respond to accurately. Its numerical scale facilitates clear aggregation and analysis of data, allowing for effective comparison over time and across patient groups. The categorization into promoters, passives, and detractors appears to provide actionable insights for further improvements and its widespread use across industries, including health care [19].

On the other hand, up to 38-item questionnaires were used to measure usability [20]. We have, therefore, decided to include the NPS in our study to assess its association with known usability attributes.

A closer look at usability questionnaires reveals a need for development. First, different questionnaires exist side by side, sometimes measuring only different facets of usability or usability-related constructs [14]. Second, many questionnaires have little empirical evidence regarding their psychometric properties or, third, they can only assess the usability of a specific or single technology [14]. Fourth, most of the questionnaires are only available in English. Accordingly, there are hardly any validated and appropriate questionnaires for usability studies in the German language [21]. As far as we know, only the German translation of the System Usability Scale—the origin of all usability questionnaires—seems to be available for wider use so far [22]. However, none of the 4 existing German versions is convincing [23]. The first 3 points also become clear when you look at the small number of questionnaires available in German.

Some questionnaires are available in German capture usability-related constructs, but do not focus directly on usability (eg, Mobile App Rating Scale-German [MARS-G] [24] and the User Experience Questionnaire [25]). The AttrakDiff questionnaire measures usability as merely 1 dimension among several others [26]. More specifically for the telehealth area, Altmann et al [27] published a German version and a short version of the "Telemedicine Perception Questionnaire" (TMPQ) in 2022. The original questionnaire includes 17 items designed to evaluate the patients' impressions of home telecare, as well as to assess its potential risks and benefits. Thus, the TMPQ does not measure usability per se (see first point) and is limited to evaluating older patients receiving video consultations from a nurse (see third point). Moreover, the validation of the questionnaire is limited to only 32 and 10 participants in its validation study [28] (see second point). The German translation could be shortened after subgroup analysis to a short version with 5 items. The German version of the TMPQ showed sufficient reliability (Cronbach α=0.76) in Altmann's study with 32 participants compared to the original study (Cronbach α=0.8). For the brief version, reliability was still acceptable with Cronbach α of 0.72.

There is another small number of questionnaires available in German whose area of application is very limited (eg, ISONORM 9241/110 on desktop apps [29]). Specific to the telehealth area, different authors in this field offered German translations of the mHealth App Usability Questionnaire (MAUQ) [30-32] (see third point). Moorthy et al [30] validated their translated MAUQ in a specific sample of 133 patients with cancer but, presumably due to the small sample size, the factor structure of the translated questionnaire was not further investigated. Kopka et al [31] provided a German version and a German short version in a sample of 148 patients using a symptom checker app in an emergency department in a randomized controlled trial. They showed that the original factor structure did not fit the data well, but no further investigation of the factor structure was conducted. In their validation study (n=53; see second point), Tacke et al [32] showed a strong positive correlation between their MAUQ translation and the System Usability Scale (SUS). However, the factor structure was not examined due to the sample size.

Considering these shortcomings, we still see a need for an appropriate questionnaire available in German. Our aim is to ensure that this German questionnaire is suitable for assessing the general usability of both telehealth and video consultations. In addition, the factor structure is to be evaluated on the basis of a sufficiently large dataset.

Telehealth Usability Questionnaire

The Telehealth Usability Questionnaire (TUQ) by Parmanto et al [33] in 2016 measures all usability attributes except memorability, allows the evaluation of video consultations, and is the most used usability questionnaire [34,35]. The TUQ uses preexisting items from other questionnaires and is freely available following the Creative Commons license 4.0. The TUQ is recommended by other authors and by frameworks for assessing telehealth [14,17,33-35]. We, therefore, translated and validated the TUQ in German (see Methods for further information on the TUQ).

Bibiloni et al [36] published an exploratory factor analysis (EFA; 150 questionnaires) of the TUQ [33] relating to video consultation. They found that 2 factors were sufficient to model the observed data. After the questionnaire was adapted to 12 items by an expert team, a confirmatory factor analysis (269 questionnaires) was performed. Both factors could be measured with good reliability, but they were highly positively correlated. Despite adapting the items and shortening the questionnaire to 12 items, the main problem with the questionnaire was that no good differentiation between usability and perceived usefulness in health care could be achieved. Although the high factor correlation raised the question of whether respondents differentiate between these 2 aspects, a 1-factor model showed a clearly worse fit than the 2-factor model. A limit to the application of the short questionnaire of Bibiloni et al [36] was that the inclusion criteria merely required a single instance of a video consultation and thus did not allow for the general evaluation of apps in telehealth. Besides, as the factor analysis of Bibiloni et al [36] confirmed, the TUQ measures usability, as well as perceived usefulness in health care, and the name "Telehealth Usability and Usefulness Questionnaire" would be more appropriate.

The primary objective of this study was to develop and validate a German language version of the TUQ and compare it to the NPS. This adaptation aims to make the TUQ readily accessible and broadly applicable for evaluating telehealth usability within German-speaking populations. The second aim was to reduce the number of items in order to optimize the feasibility for use in general field studies in telehealth.

Methods

Stages of the Study

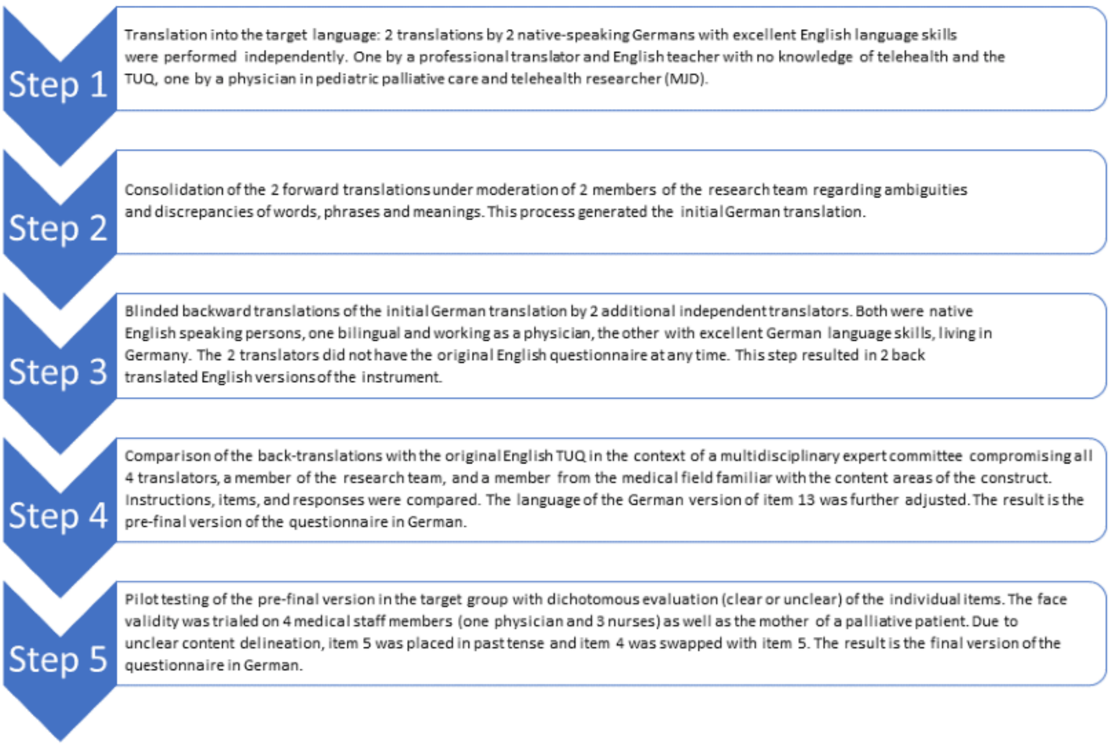
Stage I (April 2022-January 2023) consists of translation, adaptation, and pilot-testing of face validity as a method of construct validity. Stage II (July 2022-September 2023) consists of the development of a short-scale—psychometric testing and final item selection. The reporting of this study has been structured according to the recommendations of Streiner and Kottner [37] for reporting the results of studies of instrument and scale development and testing.

Stage I: Translation, Adaptation, and Pilot User Testing

High-quality translations can only be produced if measurement instruments are linguistically replicated and culturally adapted as rigorously as possible. Sousa and Rojjanasrirat [38] presented a guideline on “Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research.” In agreement with the lead author, Bambang Parmanto, we (MD and JZ) conducted translation and cross-cultural adaptation with a multidisciplinary expert committee. The members of the expert committee were selected on the basis of their knowledge and experience in the area of application of the questionnaire and their language skills, as reflected in their previous research activities and clinical experience. Translations were carried out by native speakers with proven expertise in the field of application and a high level of language skills in the target language of the translation. The recruitment took place within the authors and through the authors’ network.

To include users’ opinions and views on the German version of the TUQ, we conducted pilot-testing with medical staff, as well as a relative of palliative care children (Figure 1). Several staff members with a high level of clinical experience and patients in palliative care and their relatives were asked to participate in the pilot-testing. The aim of the test was to check whether the items were clear, understandable, and comprehensible. We recruited via the network of authors. The 3 nurses, 1 physician, and 1 relative who participated in the test were asked in an interview to indicate whether the item was clear or not (clear or unclear) and to briefly formulate how they understood the item (thinking aloud method). As a result, we made minor linguistic adjustments.

Figure 1. Stage I: the translation process and pilot user testing. TUQ: Telehealth Usability Questionnaire.



Stage II: Development of a Short Scale: Psychometric Testing in Different Target Populations and Final Item Selection

Design and Setting

A prospective observational cohort study was conducted on 2 sites—site 1 has an assessment of 13 digital therapeutics, some of which allow direct contact to the physicians of the gastroenterology and rheumatology outpatient clinic at the University Hospital Erlangen. Site 2 has an assessment of the video consultation and auscultation features of a telemedical system with patients and their parents receiving pediatric palliative home care (PPHC) in the German state of Hesse [39,40]. Both sites used modified versions of the TUQ. The survey period was between July 2022 and September 2023.

Participants and Patients

All patients who took part in the survey on site 1 were part of 1 of 3 studies conducted by the outpatient clinic at Erlangen University Hospital and were prescribed 1 of the 13 digital therapeutics in the period from January to September 2023 (see Table S1 in Multimedia Appendix 1). All patients who took part in the survey on site 2 were patients with ongoing PPHC. All patients had to first sign the written informed consent. Inclusion criteria for site 1 were a minimum age of 18 years, diagnosis of a rheumatological disease or inflammatory bowel disease, and a prescribed digital therapeutic. The inclusion criteria for site 2 were ongoing PPHC. The exclusion criteria for site 2 were younger than 18 years of age for patients and family members and lack of mental incapacity. Baseline data on demographic characteristics, disease status, and type of digital therapeutics were collected.

Data Collection, Procedures, Measurement, and Scales

Baseline data on demographic characteristics, disease status, and usability measured were collected using questionnaires. We calculated the mean and SD for all results. Patients were asked to complete a shortened version of the TUQ before and after the consultation in the ambulance (site 1) or after the use of the telemedical system (site 2). For these studies, 13 TUQ items were selected for their suitability for the general evaluation of telehealth on the basis of expert opinion (see Multimedia Appendix 2 for Cronbach α and Multimedia Appendix 3 for reasons of exclusion). In addition, the single-item NPS [41] was completed on site 1. The survey was completed partly on site or internet-based by email using the REDCap (Research Electronic Data Capture; Vanderbilt University) data collection system (site 1) or the Unipark data collection system (site 2). Participants were asked “How likely are you to recommend this app to other patients?” (original item: “How likely are you to recommend this service?”) and could respond on an 11-point scale ranging from 0 (“Very unlikely”) to 10 (“Very likely”). Depending on the results, patients could be divided into 3 groups—“promoters” (rating 9 or 10), “neutral” (rating 7 or 8), or “detractors” (rating 0 to 6).

The TUQ contains six domains with a total of 21 items, which are thought to represent each individual usability factor (see also Table 1): (1) usefulness (n=3 items), (2) ease of use and learnability (n=3), (3) interface quality (n=4), (4) interaction quality (n=4), (5) reliability (n=3), and (6) satisfaction and future use (n=4).

Table 1. English and German versions of the TUQ[a] for patients[b].

Item	English	German
Usefulness (German: Nützlichkeit)		
1	Telehealth improves my access to healthcare services.	Die App verbessert meinen Zugang zur Gesundheitsversorgung.
2	Telehealth saves me time traveling to a hospital or specialist clinic.	Durch die App spare ich Zeit in ein Krankenhaus oder zu einem niedergelassenen Arzt oder zu einer niedergelassenen Ärztin zu fahren.
3	Telehealth provides for my healthcare needs.	Die App kann mich bei meinen gesundheitlichen Anliegen unterstützen.
Ease of use and learnability (German: Benutzerfreundlichkeit und Erlernbarkeit)		
4	It was simple to use this system.	Die App lässt sich einfach bedienen.
5	It was easy to learn to use the system.	Die Bedienung der App war leicht zu erlernen.
6	I believe I could become productive quickly using this system.	Ich glaube, ich könnte die App schnell erfolgreich einsetzen.
Interface quality (German: Qualität der Benutzeroberfläche)		
7	The way I interact with this system is pleasant.	Die Benutzeroberfläche der App ist angenehm gestaltet.
8	I like using the system.	Ich bediene die Benutzeroberfläche der App gerne.
9	The system is simple and easy to understand.	Die Benutzeroberfläche der App ist einfach und leicht zu verstehen.
10	This system is able to do everything I would want it to be able to do.	Die Bedienung der Benutzeroberfläche ermöglicht alles, was ich von ihr erwarte.
Interaction quality (German: Qualität der Interaktion)		
11	I could easily talk to the clinician using the telehealth system.	Es war einfach, über die App mit dem Gesundheitspersonal zu sprechen.
12	I could hear the clinician clearly using the telehealth system.	Über die App konnte ich das Gesundheitspersonal klar und deutlich hören.
13	I felt I was able to express myself effectively.	Ich hatte den Eindruck, das Gesundheitspersonal hat mein Anliegen verstanden.
14	Using the telehealth system, I could see the clinician as well as if we met in person.	Über die App konnte ich das Gesundheitspersonal genauso gut sehen wie bei einem persönlichen Treffen.
Reliability (German: Verlässlichkeit)		
15	I think the visits provided over the telehealth system are the same as in-person visits.	Für mich sind Kontakte über die App gleichwertig mit Hausbesuchen.
16	Whenever I made a mistake using the system, I could recover easily and quickly.	Wann immer ich einen Fehler bei der Verwendung der App gemacht habe, konnte ich diesen schnell und einfach beheben.
17	The system gave error messages that clearly told me how to fix problems.	Die Fehlermeldungen der App sind eindeutig und hilfreich beim Lösen von Problemen.
Satisfaction and future use (German: Zufriedenheit und künftige Nutzungsabsicht)		
18	I feel comfortable communicating with the clinician using the telehealth system.	Ich fühle mich wohl, wenn ich über die App mit dem Gesundheitspersonal kommuniziere.
19	Telehealth is an acceptable way to receive health care services.	Es ist akzeptabel, Gesundheitsversorgung über die App zu erhalten.
20	I would use telehealth services again.	Ich würde die App wieder benutzen.
21	Overall, I am satisfied with this telehealth system.	Insgesamt bin ich zufrieden mit der App.

[a]Telehealth Usability Questionnaire.

[b]The German version shown was used for assessing patients receiving PPHC. However, as for the English TUQ, items can be adapted to different health care settings. Pilot-testing resulted in the following change: item 4 was put in the past tense instead of the present tense as shown in the table. Additionally, item 4 was swapped with item 5 (not shown). This version and the translation of the TUQ for health care professionals are free to use following the Creative Commons license 4.0, see Multimedia Appendix 4.

With respect to our initial definition above, the first domain, "usefulness," measures perceived usefulness in health care. The "efficiency" attribute, proposed by Nielsen [12], aims to assess the "level of attainable productivity of the user after he has learned the system." We see this attribute covered by TUQ item number 6—"I believe I could become productive quickly using this system" [33,34].

All other attributes are covered in the corresponding domains; however, Nielsen's [12] "memorability" attribute of usability does not seem to be covered in the TUQ. The TUQ has a Likert scale of 1 "strongly disagree" to 7 "strongly agree" as a response option; there are no reverse-scored items. The development study reports good reliability of the usability factors (usefulness: Cronbach α=0.85; ease of use: Cronbach α=0.93; effectiveness: Cronbach α=0.87; reliability: Cronbach α=0.81; satisfaction: Cronbach α=0.92) [33].

Statistical Analysis

Descriptive statistics were performed using SPSS (version 27.0; IBM Corp). EFAs were conducted in R (version 4.3.2; R Core Team [42,43]) using the packages lavaan, semTools, and psych [44,45]. The number of factors was determined using parallel analysis [46]. EFA models were estimated using full information and maximum likelihood information to account for occasional missing values (<5% per case and item). The initial factor solutions were rotated using an oblique Geomin rotation with 100 random starts for the gradient projection algorithm [47]. The Geomin parameter was set to 0.001, strongly favoring solutions with lower cross-loadings [48-51]. Based on the initial factor solution, a short scale was developed by removing items in order to achieve a simple structure and refine the substantive interpretation of the factors.

Ethical Considerations

The study was approved by the institutional review board of the Medical Faculty of the University of Erlangen-Nuremberg, Germany (22-425-Bm; January 25, 2023); the University of Kassel, Germany (202213; April 28, 2022); and University of Giessen (AZ 64/22; September 16, 2022). This study is registered in the German Clinical Trials Register (DRKS00030546). Participation in the survey was voluntary. All patients gave their written informed consent before study inclusion. All patients who participated in this study were coded with a consecutive number in a pseudonymization procedure. The data collected were stored and analyzed in a password-protected database. Only previously defined and authorized persons had access to this data. Patients had the option of withdrawing their participation in the study at any time, whereby all personal data were irrevocably deleted. The study was conducted in accordance with the ethical guidelines of the Declaration of Helsinki.

Results

Stage I: Translation, Cross-Cultural Adaptation, and Validity

The TUQ was translated in a step-by-step protocol shown in Figure 1. The expert committee discussed several minor cultural and linguistic differences, and the original developer approved all the adjustments. Original TUQ item 4 "It was simple to use this system" was swapped with item 5 "It was easy to learn to use the system," as multiple pilot user testers were irritated by this pair of questions.

The TUQ questionnaire is designed to assess different types of telehealth and, depending on which technological application is to be assessed, its wording can be adapted accordingly. There was a need to clarify the terms "telehealth," "system," and "telehealth system," which were all replaced by the term "app" to make the questionnaire applicable to apps. See Table 1 for the complete translation.

Stage II: Development of a Short Scale: Psychometric Testing and Final-Item Selection

Patient Characteristics

In total, data from 390 patients were collected in Germany. A total of 41.2% (160/390) were male and the mean age was 41.79 (SD 13.55) years (see Multimedia Appendix 1). All patients used digital therapeutics, except for group 4, which comprised children, adolescents, and young adults with life-limiting illnesses living at home using a telemedical system for video consultation and auscultation [39,40].

Descriptive Statistics, Factor Analysis TUQ, and Construction of a TUQ Short Version for Patients (Telehealth Usability and Perceived Usefulness Short Questionnaire for Patients)

Multimedia Appendix 5 provides an overview of the descriptive statistics including the correlations of the items surveyed. Parallel analysis suggested a 2-factor solution, that is, substantially fewer factors than the 6 factors originally proposed for the questionnaire. We also explored 1 and 3-factor solutions but deemed a 2-factor solution most plausible from a substantive point of view. More specifically, the 3-factor solution was defined by a dominant factor that encompassed most of the items, a smaller factor that separated the items of the reliability subscale, and a minor factor that isolated the item "Telehealth saves me time traveling to a hospital or specialist clinic" (see Table 1, item 2). The 1-factor solution blended all aspects together but with 0 loading for item 2. In contrast, the 2-factor solution resulted in 2 equally strong factors. These two factors represented distinct aspects and they are (1) perceived usefulness in health care which refers to the app's effectiveness in health care, including anticipated future use and (2) usability—pertaining to the user's experience while operating the app. Both factors explained a substantial proportion of the observed variance (39.8% and 37.2%, respectively) and were positively correlated (r=0.59, 95% CI 0.56-0.63). The overall fit of the initial model was good (comparative fit index [CFI]=0.95; standardized root-mean-square residual [SRMR]=0.03; root-mean-square error of approximation [RMSEA]=0.13; $X^2(53)=385.07$; P<.001). Further details on the factor loadings and communalities can be found in Multimedia Appendix 6. However, the initial factor loading solution was characterized by many cross-loadings. We removed all items with high cross-loadings from the model in order to sharpen the interpretation, arriving at a short version with 3 items per factor (all considerations are reported in Multimedia

Appendix 6). The final version of the scale showed an excellent fit (CFI=0.99; SRMR=0.00; RMSEA=0.02; $X^2(53)=4.86$; *P*=.30). The correlation between the factors was 0.80 (95% CI 0.75-0.85), and the factors explained 41.6% and 40.3% of the total observed variances, respectively. The standardized factor loadings of the final shortened version are displayed in Table 2. The reliability of the factors as estimated by McDonald ω was good for both factors (usability: 0.86 and perceived usefulness in health care 0.89) [52,53]. Cronbach α for the factor's usability (0.92) and perceived usefulness in health care (0.93) indicates excellent internal consistency.

Table 2. Telehealth Usability and Perceived Usefulness Short Questionnaire for Patients (TUUSQ) factor loadings (N=390)[a].

Item number	Factor loading F1, (95% CI)	Factor loading F2, (95% CI)	Communalities	Attributes
Perceived usefulness in health care				
Item number 1: The app improves my access to healthcare services.	0.97 (0.89 to 1.04)[b]	–0.1 (–0.19 to –0.02)	0.79	Usefulness
Item number 2: The app provides for my healthcare needs.	0.91 (0.86 to 0.96)[b]	0.04 (–0.01 to 0.08)	0.88	Usefulness
Item number 3: I would use the app again.	0.79 (–0.11 to –0.01)[b]	0.14 (0.96 to 1.03)	0.82	Future intention of use
Usability				
Item number 4: It was simple to use the app	–0.06 (–0.07 to 0.23)	0.99 (0.74 to 1.01)[b]	0.91	Ease of use
Item number 5: The way I interact with the app is pleasant.	0.08 (–0.07 to 0.26)	0.87 (0.58 to 0.89)[b]	0.87	User interface quality
Item number 6: Whenever I made a mistake using the app, I could recover easily and quickly.	0.09 (0.02 to 0.25)	0.73 (0.68 to 0.90)[b]	0.65	Reliability

[a]English Telehealth Usability Questionnaire (TUQ) items according to the German version referring to "the app." Factor loadings for the proposed Telehealth Usability and Perceived Usefulness Short Questionnaire for patients (TUUSQ). Factor 1=perceived usefulness in health care and factor 2=usability. The extraction method was oblique Geomin rotation. Adapted from Parmanto et al [33].

[b]Factor loadings above 0.30.

The factors identified in the TUQ show very weak correlations with the NPS—perceived usefulness in health care and NPS—*r*=0.11 (95% CI –0.00 to 0.21); usability with NPS: *r*=–0.11 (95% CI –0.22 to –0.01). Furthermore, the NPS shows no correlation to the majority of the TUQ items and very weak correlations to 6 TUQ items (see Multimedia Appendix 7).

The low strength of these correlations makes it very unlikely that the construct measured by the NPS is similar to the constructs measured by the TUQ (see also Multimedia Appendix 7). To double-check possible relationships between the NPS and the TUQ items, we determined the Net Promoter Score on the basis of the NPS [15] and calculated the Kendall Tau-b coefficients (2-sided). A total of 9 significant negative correlations with small effect size [54] resulted—between items 3 and 11 with a range of *r*=–0.10 to *r*=–0.19.

Discussion

Stage I: Complete TUQ Now Available in German in a High-Quality Translation

The complete TUQ was translated into German and cross-culturally adapted. It is comprehensible and equivalent to the English version [33] (see Table 1).

Stage II: Development of the Telehealth Usability and Perceived Usefulness Short Questionnaire for Patients

We identified 2 factors ("perceived usefulness in health care" and "usability") in the TUQ which are sufficient to describe the patient's perception. The NPS does not allow an assessment of usability attributes. The TUQ could be shortened to 6 items without compromising its informativeness as discussed.

Factor Structure

The TUQ shows a 2-factor structure—on the one hand "usability," on the other hand "perceived usefulness in health care." Both factors correlate highly positively (*r*=0.59, 95% CI 0.56-0.63). The same factors were also shown as the main factors of the Spanish version of the TUQ by the working group of Bibiloni et al [36] in 2020. A factor analysis of the Thai version of the TUQ also resulted in a 2-factor model, comparable both to our study and Bibiloni et al [36,55]. The 2 factors were "accessibility" and "utility." To the best of our knowledge, no further data on the TUQ factor structure were found in other studies [34]. In line with this research, and despite the high correlation of the factors, we deemed the 2-factor solution plausible for 2 reasons—first, it is conceivable that a health care app is technically well-designed but does not fulfill its health-related purpose for the patients. Hence, these aspects are distinguishable, and a 2D questionnaire encourages respondents to think about these aspects separately. In that sense, the high correlation may be an artifact of the context in which patients with highly specialized health care apps were asked to fill in the questionnaire. Second, previous research [56] has shown that factor analysis tends to err in the direction of extracting too few rather than too many factors, thus, the results underline the theoretical notion that there are 2 distinguishable factors in the TUQ.

We removed TUQ items (see Table 1; items 2, 5, 6, 8, 17, 19, and 21) largely due to high cross-loadings in order to sharpen the interpretation, arriving at a short version with 3 items per factor (see Multimedia Appendix 8). As the TUQ, similar to many other usability questionnaires, also contains items measuring the perceived usefulness, we, therefore, propose a clear title stating the dual purpose of the short questionnaire, thus "Telehealth Usability and Perceived Usefulness Short Questionnaire for patients" (TUUSQ). Bibiloni et al [36] proposed a short version with 12 items to assess usability in telehealth focusing on video consultations. This short version shares 4 items with the TUUSQ (see Table 3).

Table 3. History of selected TUQ[a] items[b].

Item number in TUQ	1	2	3	4	5	7	11	13	14	16	18	19	20	21
Item included in TUUSQ[c]	✓		✓	✓		✓				✓			✓	
Item included by Bibiloni et al [36]	✓	✓		✓	✓		✓	✓	✓	✓	✓	✓	✓	✓
Source of item PSSUQ[d]				✓	✓	✓				✓	✓	✓		✓
Source of item TSQ[e]	✓	✓	✓				✓	✓	✓		✓	✓	✓	✓
Source of item TAM[f]					✓									

[a]TUQ: Telehealth Usability Questionnaire.

[b]The source of the selected TUQ items is shown, as these were originally developed by the authors of the PSSUQ, TSQ, and TAM questionnaires and afterward included in the TUQ [57-59]. TUUSQ items 1, 2, and 3 originate from the TSQ, items 4, 5, and 6 from the PSSUQ questionnaire. Items selected by Bibiloni et al [36] for assessing video consultations also originate from the TAM questionnaire [36].

[c]TUUSQ: Telehealth Usability and Perceived Usefulness Short Questionnaire for patients.

[d]PSSUQ: Post Study System Usability Questionnaire [57].

[e]TSQ: Telemedicine Satisfaction Questionnaire [58].

[f]TAM: Technology Acceptance Model [59].

Factor Perceived Usefulness in Health Care

The factor that we named "perceived usefulness in health care" was included in the TUUSQ with 2 items relating to health care access and support. In addition, 1 item regarding the intention of future use (Cronbach α=0.79) loads this factor. We interpret this finding as the future intention of use is highly associated with perceived usefulness in health care and not with good usability experience. Other studies in other contexts also demonstrated that satisfaction and future intention of use show a high correlation [60-63].

Factor Usability

The second factor, "usability," contains 1 item each regarding ease of use, reliability, and interface quality. The TUUSQ thus lacks the TUQ items addressing the usability attributes as defined by Nielsen [12], that is, "efficiency" (item 6), "efficacy" (item 2), and "satisfaction" (item 21). However, as our data show no benefit in adding further items, and that the feasibility of a 6-item questionnaire is very good; therefore, we advocate this short version. Interestingly, in contrast to the usability attributes, the TUQ item addressing interface quality showed the second-highest factor loading for the factor usability in our study. One of the first usability questionnaires, the "Post Study System Usability Questionnaire" (PSSUQ) showed a 3-factor model with the factors "interface quality," "system usefulness," and "information quality" [57]. Saeed et al [64] also showed that, in the context of telehealth home monitoring, the quality of the user interface is of utmost importance for patient usability. Weichbroth [13] reported that among others less commonly assessed usability attributes for the mobile setting include navigation, operability, attractiveness, aesthetics, accessibility, and interaction [13]. Possibly our German translation of the TUQ item 7 "The way I interact with this app is pleasant" which means literally "the app's user interface has a pleasant design" also relates to these attributes (see Table 1). This finding was also present in usability studies in other health care settings [65].

The original TUQ contains no items regarding memorability. We decided against adding an item to assess memorability as our study design as it does not allow testing for memorability, that is, ease of reusability of the applications after long periods of disuse. Of course, the usefulness of telehealth should be memorable and practical in the longer term, but this question was not part of this study. Further studies on this with a more extended use of telehealth should be carried out in the future.

NPS in Health Care

Interestingly, the NPS did not show moderate to strong correlations with either the 2 identified factors—perceived usefulness in health care and NPS: 0.105 (95% CI –0.00 to 0.21); usability with NPS: –0.11 (95% CI –0.22 to –0.01) or with any of the individual TUQ items. These results show that the underlying construct of NPS is not associated with any of the patient's usability or perceived usefulness in health care attributes covered by the TUQ in our studies. As the TUQ contains an item to assess satisfaction (item 21), our studies support the findings of Krol et al [15] that the NPS is not associated with patient satisfaction. The construct that is measured by the NPS in health care continues to remain elusive [66]. Future research assessing health care quality should include the NPS in conjunction with larger surveys [66]. This is necessary since single-item measures are notoriously unreliable—potentially explaining this null finding—and should not be used as a critical variable in high-stakes settings. Moreover, possibly a correlation to other attributes might reveal the underlying construct of the NPS and pave the way to evidence-based use in health care. Finally, it should be noted that Adams et al [66] suggest limiting the use of the NPS to certain health care settings, for example, where patients have a choice of provider. This was not the case in our study.

Origin and Quality of TUUSQ Items

The items selected from the TUQ for the TUUSQ originate from the PSSUQ and Telehealth Satisfaction Questionnaire (TSQ) questionnaires (see Table 3 [57,58]). All TUUSQ items addressing usability originate from the PSSUQ. The PSSUQ was reviewed by Sousa and Lopez [14] in 2017 and assessed as one of the best available usability questionnaires, although the TUQ was not included in this study [33]. The PSSUQ's items were generated using an empirical study and showed very good internal consistency (Cronbach α=0.97). However, the quality assessment of validity, reliability, user-centeredness, sample size, and feasibility by Sousa and Rojjansrirat [38] yielded a medium-quality score due to low sample size and lack of reported user-centeredness during item generation. The PSSUQ is also sensitive to user-group and system differences [57].

The TUUSQ items assessing perceived usefulness in health care all originate from the TSQ. The TSQ shows a 3-factor model [58] and all items used by the TUQ and thus TUUSQ belong to the factor called "quality of care provided." No review of the psychometric properties of the TSQ is available. The reported sample size was low and lacked reported user-centeredness but showed good internal consistency (Cronbach α=0.93).

Limitations

The TUQ was translated in a way that has proven itself in research. Nevertheless, individual words may seem inappropriate for some target groups (eg, item 15—"Hausbesuche" for "in-person visits"). Depending on the context, "Vor-Ort-Termine," for example, may seem more appropriate here. We would like to point out that such adjustments can have an impact on the quality of the questionnaire.

The development of the TUUSQ short questionnaire was for the main part only examined patients who received digital therapeutics or used a telehealth system. Further studies are needed to evaluate whether the short questionnaire can be successfully applied more widely in the area of telehealth. Furthermore, the applicability of the TUUSQ may be limited in other cultural contexts, as this study only included patients from Germany. This limitation also applies to the original instruments (PSSUQ and TSQ) used as references. Cultural and contextual differences between the original settings of these instruments and this study could affect their relevance and accuracy in different populations. Thus, we encourage further studies using our translated version to analyze the factor structure in other datasets to investigate the generalizability of our conclusions.

Conclusions

The TUUSQ offers a short and highly feasible questionnaire for assessing and distinguishing the perceived usefulness and usability of telehealth. The TUUSQ contains solely generalizable items which allow for its use in many different telehealth contexts, as advocated by Sousa and Lopez [14].

Based on our results and the binary factor TUUSQ structure and the recommendations of Sousa and Rojjansrirat [38], we propose the following sequential approach for the assessment of telehealth apps—first, assess whether the app addresses a relevant health care need for patients. If not, and patients report no improved access to health care and health care support, or future intention of use is not reported, the scope of the app should be reevaluated. If perceived usefulness in health care is given, the TUUSQ provides software developers with concrete information on the app's usability giving feedback on ease of use, reliability, and the quality of the user interface.

Following our study, the TUUSQ can be used in Danish, German, Portuguese, Slovene, Thai, and Urdu, as validated TUQ translations are available for these languages [33,55,67,68]. The TUUSQ is free for commercial and noncommercial use following the Creative Commons license 4.0 [69]. If, besides perceived usefulness in health care and usability, the quality of the video connection during a video consultation, as well as the suitability of this medium are of interest to the researchers, we recommend using the TUQ short version as proposed by Bibiloni et al [36].

Acknowledgments

The authors thank the patients, families, collaborators, and pediatric palliative care teams for their participation in this study. The authors thank Madita Mannheims, whose contributions to stage I went toward her obtaining her masters degree in psychology, and all the translators for their valuable time and input. This research was funded by the Hessian Ministry for Social Affairs and Integration grant (18z4350-0001/2021/001).

Authors' Contributions

MD, JZ, HM, and FS wrote the draft paper. FS, JZ, and MD performed the statistical analysis. Translations were conducted by MD and JZ. Studies were conducted by TO (study 1), JK (study 2), HM (study 3), and MD and JZ (study 4). All authors (MD, JZ, HM, FS, JK, TO, HM, MN, and TV) reviewed the draft and provided comments for change. All authors approved the final paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of participating patients in a prospective observational cohort study at the University Hospital Erlangen and receiving pediatric palliative home care (PPHC) in the German state of Hessen from July 2022 to September 2023 (mean, SD or n, %).
[DOCX File , 17 KB - humanfactors_v11i1e57771_app1.docx]

Multimedia Appendix 2
Internal consistency of the Telehealth Usability Questionnaire (TUQ) subscales of the shortened version of the TUQ.
[DOCX File , 15 KB - humanfactors_v11i1e57771_app2.docx]

Multimedia Appendix 3
Exclusion reasons for certain Telehealth Usability Questionnaire (TUQ) items. TUQ Items marked with asterisks were included in the shortened version.
[DOCX File , 21 KB - humanfactors_v11i1e57771_app3.docx]

Multimedia Appendix 4
German Telehealth Usability Questionnaire version for health care professionals.
[DOCX File , 37 KB - humanfactors_v11i1e57771_app4.docx]

Multimedia Appendix 5 [DOCX File , 24 KB - humanfactors_v11i1e57771_app5.docx]

Multimedia Appendix 6
Standardized factor loadings and communalities (Comm.) after Geomin rotation for a 2-factor model using the 13 adapted items from the Telehealth Usability Questionnaire (TUQ) for which data were available incl. considerations for removing single items. English TUQ items according to the German version refer to “the app.” N=390. Factor 1= perceived unsefulness in health care and factor 2= usability. The factor correlation for this solution was 0.59. The fit was good overall (SRMR =0.31, CFI =0.95, RMSEA =0.13). Factor loadings above 0.30 are in bold. Adapted from [33]. Items marked with asterisks were included in the short version.
[DOCX File , 24 KB - humanfactors_v11i1e57771_app6.docx]

Multimedia Appendix 7
Correlations for NPS and 13 Telehealth Usability Questionnaire (TUQ) items. Pearson correlations. Shortened version of the TUQ [33], see Table S3. NPS: Net Promoter Scale.
[DOCX File , 15 KB - humanfactors_v11i1e57771_app7.docx]

Multimedia Appendix 8
German Telehealth Usability and Perceived Usefulness Short Questionnaire version for patients.
[DOCX File , 117 KB - humanfactors_v11i1e57771_app8.docx]

References

1. Discover how our language data could power. OxfordLanguages. URL: https://languages.oup.com/, [accessed 2024-01-08]
2. da Fonseca MH, Kovaleski F, Picinin CT, Pedroso B, Rubbo P. E-Health practices and technologies: a systematic review from 2014 to 2019. Healthcare (Basel) 2021;9(9):1192 [FREE Full text] [doi: 10.3390/healthcare9091192] [Medline: 34574966]
3. Bestsennyy O, Gilbert G, Harris A, Rost J. Telehealth: a quarter-trillion-dollar post-COVID-19 reality? McKinsey & Company. URL: https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/telehealth-a-quarter-trillion-dollar-post-covid-19-reality [accessed 2021-07-14]

4. Seifert A, Batsis JA, Smith AC. Telemedicine in long-term care facilities during and beyond COVID-19: challenges caused by the digital divide. Front Public Health 2020;8:601595 [FREE Full text] [doi: 10.3389/fpubh.2020.601595] [Medline: 33194999]
5. Li MM, Rising KL, Goldberg EM. Transitioning to telehealth? A guide to evaluating outcomes. Health Policy Technol 2022;11(3):100623 [FREE Full text] [doi: 10.1016/j.hlpt.2022.100623] [Medline: 35369128]
6. Schlomann A, Seifert A, Zank S, Rietz C. Assistive technology and mobile ICT usage among oldest-old cohorts: comparison of the oldest-old in private homes and in long-term care facilities. Res Aging 2020;42(5-6):163-173. [doi: 10.1177/0164027520911286] [Medline: 32167019]
7. Corbett JA, Opladen JM, Bisognano JD. Telemedicine can revolutionize the treatment of chronic disease. Int J Cardiol Hypertens 2020;7:100051 [FREE Full text] [doi: 10.1016/j.ijchy.2020.100051] [Medline: 33330846]
8. Kodjebacheva GD, Tang C, Groesbeck F, Walker L, Woodworth J, Schindler-Ruwisch J. Telehealth use in pediatric care during the COVID-19 pandemic: a qualitative study on the perspectives of caregivers. Children (Basel) 2023;10(2):311 [FREE Full text] [doi: 10.3390/children10020311] [Medline: 36832439]
9. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
10. Koppel R, Metlay JP, Cohen A, Abaluck B, Localio AR, Kimmel SE, et al. Role of computerized physician order entry systems in facilitating medication errors. JAMA 2005;293(10):1197-1203. [doi: 10.1001/jama.293.10.1197] [Medline: 15755942]
11. Abran A, Khelifi A, Suryn A, Seffah A. Usability meanings and interpretations in ISO standards. Software Qual J 2003;11(4):325-338.
12. Nielsen J. Usability Engineering. Boston: Academic Press; 1993.
13. Weichbroth P. Usability of mobile applications: a systematic literature study. IEEE Access 2020;8:55563-55577. [doi: 10.1109/access.2020.2981892]
14. Sousa VEC, Dunn Lopez K. Towards usable E-Health. A systematic review of usability questionnaires. Appl Clin Inform 2017;8(2):470-490 [FREE Full text] [doi: 10.4338/ACI-2016-10-R-0170] [Medline: 28487932]
15. Krol MW, de Boer D, Delnoij DM, Rademakers JJDJM. The net promoter score--an asset to patient experience surveys? Health Expect 2015;18(6):3099-3109 [FREE Full text] [doi: 10.1111/hex.12297] [Medline: 25345554]
16. von Rohr S, Knitza J, Grahammer M, Schmalzing M, Kuhn S, Schett G, et al. Student-led clinics and ePROs to accelerate diagnosis and treatment of patients with axial spondyloarthritis: results from a prospective pilot study. Rheumatol Int 2023;43(10):1905-1911 [FREE Full text] [doi: 10.1007/s00296-023-05392-5] [Medline: 37486433]
17. Labinsky H, Gupta L, Raimondo MG, Schett G, Knitza J. Real-world usage of digital health applications (DiGA) in rheumatology: results from a German patient survey. Rheumatol Int 2023;43(4):713-719 [FREE Full text] [doi: 10.1007/s00296-022-05261-7] [Medline: 36543961]
18. Labinsky H, Ukalovic D, Hartmann F, Runft V, Wichmann A, Jakubcik J, et al. An AI-powered clinical decision support system to predict flares in rheumatoid arthritis: a pilot study. Diagnostics (Basel) 2023;13(1):148 [FREE Full text] [doi: 10.3390/diagnostics13010148] [Medline: 36611439]
19. Boylan AM, Turk A, van Velthoven MH, Powell J. Online patient feedback as a measure of quality in primary care: a multimethod study using correlation and qualitative analysis. BMJ Open 2020;10(2):e031820 [FREE Full text] [doi: 10.1136/bmjopen-2019-031820] [Medline: 32114461]
20. Wilkinson A, Forbes A, Bloomfield J, Fincham Gee C. An exploration of four web-based open and flexible learning modules in post-registration nurse education. Int J Nurs Stud 2004;41(4):411-424. [doi: 10.1016/j.ijnurstu.2003.11.001] [Medline: 15050852]
21. Fragebogenmatrix. German UPA. URL: https://germanupa.de/wissen/fragebogenmatrix?field_einsatzfelder_target_id=All&field_gemessene_dimensionen_target_id%5B%5D=174, [accessed 2024-02-01]
22. Gao M, Kortum P, Oswald FL. Multi-language toolkit for the system usability scale. Int J Hum-Comput Interact 2020;36(20):1883-1901. [doi: 10.1080/10447318.2020.1801173]
23. Brix TJ, Janssen A, Storck M, Varghese J. Comparison of German translations of the system usability scale - which to take? Stud Health Technol Inform 2023;307:96-101. [doi: 10.3233/SHTI230699] [Medline: 37697842]
24. Messner E, Terhorst Y, Barke A, Baumeister H, Stoyanov S, Hides L, et al. The German version of the mobile app rating scale (MARS-G): development and validation study. JMIR Mhealth Uhealth 2020;8(3):e14479 [FREE Full text] [doi: 10.2196/14479] [Medline: 32217504]
25. User experience questionnaire. UEQ. URL: https://www.ueq-online.org [accessed 2024-02-08]
26. Hassenzahl M, Burmester M, Koller F. AttrakDiff: ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualität. In: Mensch & Computer 2003. Vieweg+Teubner Verlag. Berichte des German Chapter of the ACM, vol 57. Berlin: Vieweg+Teubner Verlag; 2023.
27. Altmann P, Ivkic D, Ponleitner M, Leutmezer F, Willinger U, Schmoeger M, et al. Individual perception of telehealth: validation of a german translation of the telemedicine perception questionnaire and a derived short version. Int J Environ Res Public Health 2022;19(2):902 [FREE Full text] [doi: 10.3390/ijerph19020902] [Medline: 35055724]

28. Demiris G, Speedie S, Finkelstein S. A questionnaire for the assessment of patients' impressions of the risks and benefits of home telecare. J Telemed Telecare 2000;6(5):278-284. [doi: 10.1258/1357633001935914] [Medline: 11070589]
29. Prümper J. Der benutzungsfragebogen ISONORM 9241/10: ergebnisse zur reliabilität und validität. In: Software-Ergonomie '97. Berichte des German Chapter of the ACM, vol 49. Berlin: Vieweg+Teubner Verlag; 1997.
30. Moorthy P, Weinert L, Harms BC, Anders C, Siegel F. German version of the mHealth app usability questionnaire in a cohort of patients with cancer: translation and validation study. JMIR Hum Factors 2023;10:e51090 [FREE Full text] [doi: 10.2196/51090] [Medline: 37910144]
31. Kopka M, Slagman A, Schorr C, Krampe H, Altendorf M, Balzer F, et al. German mHealth app usability questionnaire (G-MAUQ) and short version (G-MAUQ-S): translation and validation study. Smart Health 2024;34:100517 [FREE Full text] [doi: 10.1016/j.smhl.2024.100517]
32. Tacke T, Nohl-Deryk P, Lingwal N, Reimer LM, Starnecker F, Güthlin C, et al. The German version of the mHealth app usability questionnaire (GER-MAUQ): translation and validation study in patients with cardiovascular disease. Digit Health 2024;10:20552076231225168 [FREE Full text] [doi: 10.1177/20552076231225168] [Medline: 38303970]
33. Parmanto B, Lewis AN, Graham KM, Bertolet MH. Development of the telehealth usability questionnaire (TUQ). Int J Telerehabil 2016;8(1):3-10 [FREE Full text] [doi: 10.5195/ijt.2016.6196] [Medline: 27563386]
34. Hajesmaeel-Gohari S, Bahaadinbeigy K. The most used questionnaires for evaluating telemedicine services. BMC Med Inform Decis Mak 2021;21(1):36 [FREE Full text] [doi: 10.1186/s12911-021-01407-y] [Medline: 33531013]
35. Langbecker D, Caffery LJ, Gillespie N, Smith AC. Using survey methods in telehealth research: a practical guide. J Telemed Telecare 2017;23(9):770-779. [doi: 10.1177/1357633X17721814] [Medline: 28728502]
36. Bibiloni N, Torre AC, Angles MV, Terrasa SA, Vazquez Peña FR, Sommer J, et al. [Validation of a Spanish questionnaire on telemedicine usability]. Medicina (B Aires) 2020;80(6):649-653 [FREE Full text] [Medline: 33254109]
37. Streiner DL, Kottner J. Recommendations for reporting the results of studies of instrument and scale development and testing. J Adv Nurs 2014;70(9):1970-1979. [doi: 10.1111/jan.12402] [Medline: 24684713]
38. Sousa VD, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. J Eval Clin Pract 2011;17(2):268-274. [doi: 10.1111/j.1365-2753.2010.01434.x] [Medline: 20874835]
39. Zimmermann J, Heilmann ML, Fisch-Jessen M, Hauch H, Kruempelmann S, Moeller H, et al. Telehealth needs and concerns of stakeholders in pediatric palliative home care. Children (Basel) 2023;10(8):1315 [FREE Full text] [doi: 10.3390/children10081315] [Medline: 37628315]
40. Ohligs M, Stocklassa S, Rossaint R, Czaplik M, Follmann A. Employment of telemedicine in nursing homes: clinical requirement analysis, system development and first test results. Clin Interv Aging 2020;15:1427-1437 [FREE Full text] [doi: 10.2147/CIA.S260098] [Medline: 32884251]
41. Hamilton DF, Lane JV, Gaston P, Patton JT, Macdonald DJ, Simpson AHRW, et al. Assessing treatment outcomes using a single question: the net promoter score. Bone Joint J 2014;96-B(5):622-628 [FREE Full text] [doi: 10.1302/0301-620X.96B5.32434] [Medline: 24788496]
42. The R Core Team. R: a language and environment for statistical computing. MSOR connections. 2014. URL: https://cran.r-project.org/doc/manuals/r-release/fullrefman.pdf [accessed 2024-10-04]
43. The R project for statistical computing. R. URL: https://www.r-project.org/ [accessed 2023-09-08]
44. lavaan: latent variable analysis. R. URL: https://cran.r-project.org/web/packages/lavaan/citation.html, [accessed 2023-09-08]
45. Useful tools for structural equation modeling. R. URL: http://cran.nexr.com/web/packages/semTools/citation.html [accessed 2023-08-09]
46. Horn JL. A rationale and test for the number of factors in factor analysis. Psychometrika 1965;30:179-185 [FREE Full text] [doi: 10.1007/bf02289447]
47. Yates A. Multivariate Exploratory Data Analysis: A Perspective on Exploratory Factor Analysis. New York: State University of New York Press; 1987.
48. Marsh HW, Liem GAD, Martin AJ, Morin AJS, Nagengast B. Methodological measurement fruitfulness of exploratory structural equation modeling (ESEM): new approaches to key substantive issues in motivation and engagement. J Psychoeduc Assess 2011;29(4):322-346 [FREE Full text] [doi: 10.1177/0734282911406657]
49. Marsh HW, Lüdtke O, Muthén B, Asparouhov T, Morin AJS, Trautwein U, et al. A new look at the big five factor structure through exploratory structural equation modeling. Psychol Assess 2010;22(3):471-491. [doi: 10.1037/a0019227] [Medline: 20822261]
50. Marsh HW, Muthén B, Asparouhov T, Lüdtke O, Robitzsch A, Morin AJS, et al. Exploratory structural equation modeling, integrating CFA and EFA: application to students' evaluations of university teaching. Struct Equ Modeling: Multidiscip Res J 2009;16(3):439-476 [FREE Full text] [doi: 10.1080/10705510903008220]
51. Scharf F, Nestler S. A comparison of simple structure rotation criteria in temporal exploratory factor analysis for event-related potential data. Eur J Res Methods Behav Soc Sci 2019;15(Suppl 1):43-60 [FREE Full text]
52. McDonald RP. Test Theory: A Unified Treatment. United States: Lawrence Erlbaum Associates Publishers; 1999.
53. McNeish D. Thanks coefficient alpha, we'll take it from here. Psychol Methods 2018;23(3):412-433. [doi: 10.1037/met0000144] [Medline: 28557467]

54. Cohen JE. Statistical Power Analysis for the Behavioral Sciences (2nd ed.). New York: Routledge; 1988.
55. Hirunwiwatkul P, Pongpanich P, Tulvatana W, Jariyakosol S, Phuenpathom W, Krittanupong S, et al. Evaluation of psychometric properties of Thai version telehealth usability questionnaire (T-TUQ). Int J Telerehabil 2023;15(2):e6577 [FREE Full text] [doi: 10.5195/ijt.2023.6577] [Medline: 38162944]
56. Beauducel A. Problem with parallel analysis in data sets with oblique simple structure. Methods Psychol Res 2001:6.
57. Lewis JR. Psychometric evaluation of the PSSUQ using data from five years of usability studies. Int J Hum-Comput Interact 2002;14(3-4):463-488. [doi: 10.1080/10447318.2002.9669130]
58. Yip MP, Chang AM, Chan J, MacKenzie AE. Development of the telemedicine satisfaction questionnaire to evaluate patient satisfaction with telemedicine: a preliminary study. J Telemed Telecare 2003;9(1):46-50. [doi: 10.1258/135763303321159693] [Medline: 12641893]
59. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989;13(3):319-340. [doi: 10.2307/249008]
60. Ejdys J, Kozlowska J. Acceptance of e-learning at university level during the COVID-19 pandemic situation-teachers' and students' perspective. Pol J Manag Stud 2021;23(2):106-129 [FREE Full text]
61. Ejdys J. Factors influencing satisfaction and future intention to use e-learning at the university level. Foresight STI Governance 2022;16(2):52-64. [doi: 10.17323/2500-2597.2022.2.52.64]
62. Zhu Y, Zhao Z, Guo J, Wang Y, Zhang C, Zheng J, et al. Understanding use intention of mHealth applications based on the unified theory of acceptance and use of technology 2 (UTAUT-2) model in China. Int J Environ Res Public Health 2023;20(4):3139 [FREE Full text] [doi: 10.3390/ijerph20043139] [Medline: 36833830]
63. Alalwan AA, Dwivedi YK, Rana NP. Factors influencing adoption of mobile banking by Jordanian bank customers: extending UTAUT2 with trust. Int J Inf Manag 2017;37(3):99-110 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2017.01.002]
64. Saeed N, Manzoor M, Khosravi P. An exploration of usability issues in telecare monitoring systems and possible solutions: a systematic literature review. Disabil Rehabil Assist Technol 2020;15(3):271-281. [doi: 10.1080/17483107.2019.1578998] [Medline: 30794009]
65. Peikari HR, Shah MH, Zakaria MS, Yasin NM, Elhissi A. The impacts of second generation e-prescribing usability on community pharmacists outcomes. Res Social Adm Pharm 2015;11(3):339-351. [doi: 10.1016/j.sapharm.2014.08.011] [Medline: 25262599]
66. Adams C, Walpola R, Schembri AM, Harrison R. The ultimate question? Evaluating the use of net promoter score in healthcare: a systematic review. Health Expect 2022;25(5):2328-2339 [FREE Full text] [doi: 10.1111/hex.13577] [Medline: 35985676]
67. Bender C, Nielsen MG, Hejlesen O, Hangaard S. Assessing the use of telemedicine among people with diabetes: a Danish translation and cross-cultural adaptation of the telemedicine usability questionnaire. 2020 Presented at: Proceedings of the 18th Scandinavian Conference on Health Informatics; 2022 Aug 01; Norway p. 14-20.
68. Santos MR, Malaguti C, Cabral LA, Soares AL, Neves LHG, Sena LDA, et al. The Brazilian version of the telehealth usability questionnaire (telehealth usability questionnaire Brazil): translation, cross-cultural adaptation, and psychometric properties. Rev Assoc Med Bras (1992) 2023;69(12):e20230228 [FREE Full text] [doi: 10.1590/1806-9282.20230228] [Medline: 37971117]
69. Attribution 4.0 international. Creative Commons. URL: https://creativecommons.org/licenses/by/4.0/ [accessed 2024-02-05]

Abbreviations

CFI: comparative fit index
EFA: exploratory factor analysis
ISO: International Organization for Standardization
MARS-G: Mobile App Rating Scale-German
MAUQ: mHealth App Usability Questionnaire
mHealth: mobile health
NPS: Net Promoter Scale
PPHC: pediatric palliative home care
PSSUQ: Post Study System Usability Questionnaire
REDCap: Research Electronic Data Capture
RMSEA: root-mean-square error of approximation
SRMR: standardized root-mean-square residual
SUS: System Usability Scale
TMPQ: Telemedicine Perception Questionnaire
TSQ: Telehealth Satisfaction Questionnaire
TUQ: Telehealth Usability Questionnaire
TUUSQ: Telehealth Usability and Perceived Usefulness Short Questionnaire for patients

Edited by A Kushniruk, E Borycki; submitted 26.02.24; peer-reviewed by A Hassan, R Knudsen, J Mistry, M Uncovska; comments to author 06.05.24; revised version received 30.06.24; accepted 13.09.24; published 21.11.24.

Please cite as:
Zimmermann J, Morf H, Scharf F, Knitza J, Moeller H, Muehlensiepen F, Nathrath M, Orlemann T, Voelker T, Deckers M
German Version of the Telehealth Usability Questionnaire and Derived Short Questionnaires for Usability and Perceived Usefulness in Health Care Assessment in Telehealth and Digital Therapeutics: Instrument Validation Study
JMIR Hum Factors 2024;11:e57771
URL: https://humanfactors.jmir.org/2024/1/e57771
doi:10.2196/57771
PMID:

Assessment of Acceptability, Usage, and Impact on Caregivers of Children With Autism's Stress and Mindfulness: Multiple-Method Feasibility Study of the 5Minutes4Myself App's Mindfulness Module

Elizabeth Larson[1], OTR, PhD; Rebecca L Mattie[2], OTR; Sophia A Riffkin[3], OTR

[1]Occupational Therapy Program, Department of Kinesiology, University of Wisconsin-Madison, 2180 MSC, 1300 University Avenue, Madison, WI, United States

[2]Honeybee OT, LLC, Private Practice, Onalaska, WI, United States

[3]University of Wisconsin-Madison, Madison, WI, United States

Corresponding Author:
Elizabeth Larson, OTR, PhD
Occupational Therapy Program, Department of Kinesiology, University of Wisconsin-Madison, , 2180 MSC, 1300 University Avenue, Madison, WI, , United States

Abstract

Background: Caregiver wellness programs need to be easily accessible to address caregivers' constraints to participation.

Objective: We aimed to assess the feasibility of 5Minutes4Myself app's mindfulness module (usability, usage, and impact on caregivers' levels of mindfulness and perceived stress).

Methods: Before and after participation in the 5Minutes4Myself program, 15 participants were asked to complete the Perceived Stress Scale (PSS) and Five Facet Mindfulness Questionnaire (FFMQ). Data on the usage of app-delivered meditations were collected electronically via the app, and app usability was rated on the Modified System Usability Scale. Analyses assessed participants' frequency of use of app-delivered meditations, app usability, and changes in participants' stress and mindfulness post intervention.

Results: Overall, participants completed 10.9 minutes of mindfulness meditations per week and rated the app 76.7, indicating above-average usability. Related samples *t* tests (2-tailed) found that group PSS (t_{10}=1.20, *P*=.26) and FFMQ (t_{10}=−1.57, *P*=.15) pre- or postintervention mean scores were not significantly different. However, a visualization of pre- and post-PSS and mindfulness scores suggested there was a group of responders who had decreased stress with increased mindfulness. This was confirmed via an individual change analysis. The effect size of the FFMQ scores (*d*=0.47) suggests there may be treatment effects with a larger sample. A hierarchical multiple regression analysis examined the degree mindfulness impacted perceived stress; 20% of the variance in participants' perceived stress could be attributed to increases in self-rated mindfulness (*P*=.04) when controlling for preintervention stress levels.

Conclusions: Caregivers found the app highly usable and on average used low-dose levels of mindfulness meditations (10 min/wk). For responders, increased mindfulness was related to stress reduction to population-based levels.

Trial Registration: ClinicalTrials.gov NCT03771001; https://clinicaltrials.gov/study/NCT03771001

(JMIR Hum Factors 2024;11:e54171) doi:10.2196/54171

KEYWORDS

autism; caregiver; activities; mindfulness; mobile application; stress; wellness; app; application; usage; children; developmental disability; usability; acceptability; meditation; wellness application

Introduction

Wellness programs are critically needed for the many parent caregivers of children with autism spectrum disorder (ASD) who experience increased levels of stress, poorer health outcomes, and decreased quality of life compared to parents of children with other developmental disabilities or typical development [1-6]. Caring for a child with ASD has been found to have a significant impact on the caregivers' overall well-being [7,8]. More than half of mothers of children with ASD report mental health problems, a substantially higher percentage than the general population [5]. Time-intensive care demands that remain high over time, challenging child behaviors, and social stigma limiting participation in community activities increase the need for wellness promotion while at the same time make participation more difficult [5,8,9]. Caregivers often put their child and family's needs before their own self-nurturing and wellness-promoting activities [10]. Ideally, wellness programs for caregivers should be tailored to be widely available to the most caregivers and address common caregiving constraints that could limit participation such as lack of time and difficulties with attending regular programming and using effective evidence-based interventions.

Evidence suggests that mindfulness practices bolster wellness for caregivers. These programs emphasize acceptance of life circumstances and loving-kindness toward oneself and others despite challenges [11]. Multiple delivery formats, including individual, group, and self-guided mindfulness trainings, have been shown to be more effective than other stress reducing interventions for caregivers of children with ASD and other developmental disabilities [12-16]. Researchers have focused on designing mindfulness program features that support these parents' ability to participate consistently over time to provide much needed mental health and wellness benefits. Rayan and Ahmad [17] modified a mindfulness-based intervention (MBI) program for parents in Jordan [13]. The modified program was shortened to 5-weeks and included 33 hours of training, phone follow-up sessions, and text message reminders of assignments and support. Their quasi-experimental study (n=104) found caregivers of children with ASD who participated in the MBI intervention had significantly reduced stress compared to a control group. With their modified MBI, attrition rates decreased [13,17].

However, likely due to the intensity and time demands of caregiving and the occurrence of unexpected life events, some caregivers find it difficult to attend in-person mindfulness programs that require weekly in-person time commitment and travel to community sites [12,18,19]. Harnessing the usefulness of mindfulness practices for parents may require developing an alternative to the longer in-person mindfulness programs that may overcome some of the lifestyle and time barriers these caregivers experience.

Mobile platforms offer greater flexibility for delivering self-guided meditations compared to in-person or phone options. Mobile phone texts have thus far been used in this population to extend interventions delivered in in-person training sessions [17]. Tailored delivery to times and places that fit caregivers' lifestyles, may make mindfulness intervention more convenient and accessible and cost-effective [20,21]. Digital interventions that allow personalization, goal setting, and accountability may promote caregivers' participation in wellness promotion and reduce social health inequalities [22,23]. Given the significant need and current evidence for the usefulness of mindfulness approaches, we developed, through a collaborative recursive process with and for caregivers, a mobile app to create a more user-friendly and accessible mindfulness module for the 5Minutes4Myself caregiver wellness program [24].

This analysis used feasibility study data and was intended to examine the app usability and who may have benefitted from program participation by investigating whether there were changes in stress and mindfulness, specifically focusing on this new element of app-delivered mindfulness meditations. We aimed to examine the app's usability, caregivers' usage of mindfulness content, and the effect of mindfulness meditations on participants' postparticipation levels of mindfulness and perceived stress. We recognized that other components of the 5Minutes4Myself wellness program may also have influenced perceived stress levels. We expected that (1) the app as designed with and by caregivers would be rated as acceptable, (2) the caregiver-tailoring and habit-building features of the app would encourage regular use of meditations at least twice weekly over 4 months' time, and (3) using the micromindfulness meditations of the modified 5Minutes4Myself wellness program would increase caregivers' mindfulness and reduce self-rated stress.

Methods

Study Design

This study examined the acceptability and usability of the smartphone app and its mindfulness meditations using a quasi-experimental mixed-methods design with pre- and postintervention assessments.

App Development

The 5Minutes4Myself wellness program was developed with and for caregivers of children with ASDs. We designed a collaborative research-informed approach. First in a focus group, we shared the best available evidence for wellness promotion and occupational science principles for lifestyle change with caregivers. Occupational sciences' core construct is that what we do each day (ie, how we "occupy" our time) matters, and how we balance our participation in a range of daily activities or occupations with different experiences can support or undermine our quality of life [24]. To create changes in our wellness requires leveraging and sustaining strategies to revise our often entrenched and comfortable daily routines. The evidence-based program principles and elements presented to caregivers focused on "doing" (eg, activities that promote pleasure or social connection) and lifestyle change strategies to achieve changes, "thinking" (eg, positive psychology practices) and "goal visioning" [24]. Second, we elicited caregivers' feedback on these principles and elements, confirmed caregivers' preferences and developed additional goal-visioning activities as requested. Third, we delivered the program. It included a lifestyle consultation to develop goals using motivational interviewing (MI), coaching over 4 months to pursue wellness goals, and "doing" content delivered via an iPad (eg, knitting tutorials, yoga, meditation, and activity trackers). Last, we evaluated the program in a postparticipation focus group.

After completing the program, the key change participants requested was the development of a mobile phone app to deliver mindfulness and other content that would be available to be used anytime and anywhere [24]. Further refinements caregivers requested to meet their needs and limited capacities to participate in a wellness program were to eliminate "doing" content, which was not well used on iPads, and add guilt-free goal accountability check-ins, goal profiles to remind them of their selected wellness goals and rewards to reinforce their efforts. Due to their frequent fatigue and limited technology experience, caregivers asked for simple app navigation such as prompts that quickly redirected them to the next mindfulness meditation in the series, without having to remember where they left off, and intuitive navigation between screens, and calming color schemes. Given the request for a smartphone app, we polled the caregivers on their current technology and internet access. Due to their financial circumstances not all caregivers had smartphones, and many had older versions, and not all had home access to the internet. We recognized these would be issues that needed to be addressed in the app design.

We worked with a local software start-up company to develop both an iOS and Android app version similar in content and design, and with an expert mindfulness-based stress reduction (MBSR) trainer or clinical psychologist to develop mindfulness content. We provided the software developers with several design and function parameters tailored for our caregiving participants: aesthetically pleasing interfaces, simple navigation, offline availability of content and reminders, usability on older phones, easy personalization of goal and notification reminders, dashboard tracking of mindfulness meditation usage, request coaching help function, goal check-ins, guilt-free check-ins, and rewards for full or partial goal success. The software development team used Test Flight (Apple Inc) to deliver beta versions to our research team for testing and provided a link to a data dashboard that hosted meditation usage data and goal check-in submissions. The psychologist developed and recorded the mindfulness meditations that were specifically tailored for a caregiver. For example, loving-kindness meditations often included directing nonjudgmental and accepting thoughts toward themselves and their child. Recorded meditations were 4‐21 minutes long (Table 1). The meditation series began with an introduction to mindfulness, and then alternated between meditations emphasizing awareness of the present, being aware of the physical body, and loving-kindness. For ease of navigation, the mindfulness meditations were presented in a vertical stream with a glowing dot highlighting the next meditation to be listened to (Figure 1).

Table . 5Minutes4Myself app-delivered meditations.

Title	Content	Length (min)
Introduction of mindfulness	Introduction to program and instructor	11
In this moment, pause	Being present	3
Kind witnessing of the present	Body scan	7
Receiving and letting go	Loving-kindness	5
Breath as a teacher	Being present	6
Awake with awareness and presence	Body scan	18
Resting in the breath	Loving-kindness	10
Simply choose to pause	Being present	6
Dropping into relaxation	Body scan	22
Release, letting go	Loving-kindness	16
Resting in the moment	Breath awareness	17
Kind witnessing of the present	Body scan	8
Resting in the breath	Breath awareness	10
Deep breathing	Breath awareness	12
Awake with awareness and presence	Body scan	18
Releasing, letting go	Loving-kindness	16
Dropping into relaxation	Body scan	22

Figure 1. Mindfulness meditation stream screen with glow highlighting current progress.

In total, 2 additional cohorts were delivered the 5Minutes4Myself program and confirmed the usefulness of these app elements (goal profile, self-timed reminders to do wellness activities, goal check-ins, goal achievement rewards, mindfulness meditations, and design or graphics) and requested navigation changes (ability to go forward as well as backward in the meditation stream) and moving goal check-ins to an opening screen [24]. Revisions were made to the app between these 2 cohorts to address app instability and remove the request for coaching support feature that was not used.

Participants

Recruitment

Participants were recruited through email solicitation, posters, flyers, and brochures in the community and at conferences for families of children with ASD. Recruitment also occurred through social media advertising to ASD-specific groups.

Interested individuals were screened by phone interview to assure they met inclusion criteria. Informed consent was obtained from all individual participants prior to participation at focus groups or in individual meetings. In total, 15 individuals in cohorts participated in 4-month cycles, over 1 year's time. This sample was considered sufficient for proof-of-concept to assess the feasibility of the mobile app.

Inclusion and Exclusion Criteria

Caregivers were included if they were the primary caregiver of at least one child with a diagnosis of ASD aged between 8 - 21 years. We selected this age grouping with the assumption that families had established care routines and thus the caregiver had more capacity to alter their daily routines for wellness promotion. Caregivers had to have a desire to participate in a lifestyle wellness program, including a mindfulness component, and be willing to commit sufficient time to participate in the program. Exclusion criteria included (1) parenting a child younger than 8 or older than 21 years of age, or (2) a caregiver diagnosis of significant mental illness such as schizophrenia, bipolar disorder, schizoaffective disorder, pervasive developmental disorder, obsessive-compulsive disorder, panic disorder, posttraumatic stress disorder, or an eating disorder. We included participants who had or were being treated for depression since clinical or subclinical levels are highly prevalent in this caregiving group [25]. In total, 5 reported being currently treated for depression. Further, 40% (6/15) of the caregivers scored 16 or greater on the Center for Epidemiologic Studies Depression Scale Revised, suggesting they were at risk for depression (mean 17, range 8 - 43). A recent meta-analysis suggests that the use of stand-alone mindfulness practices, without the additional readings and instruction typically provided in MBSR programs, has been found to reduce depression and anxiety [26] thus the program could have benefits for caregivers with these diagnoses.

Measures

Overview

Ratings of the app usability, app usage, and pre- and postintervention measures of self-reported stress and mindfulness were used to examine app acceptability and the effects of the 5Minutes4Myself smartphone mobile app micromindfulness meditations on caregivers' mindfulness and stress. Measures used included the Modified System Usability Scale (MSUS), Perceived Stress Scale (PSS), Five Facet Mindfulness Questionnaire (FFMQ), and app meditation usage data. In addition, participants completed a survey that asked: their age, self-identified race or ethnicity, marital status (single, married, separated, or divorced), education (high school, associate, undergraduate, or graduate), average family income, children's names and ages, caregiving or work (full-time caregiver, part-time work, or full-time work), medical conditions and medicines taken. This information was gathered to understand the family resources and caregivers' health. Participants self-identified as White.

Modified System Usability Scale

The MSUS, a 10-item survey that yields a single score of overall usability, has been widely used to assess technology quality and functionality [27,28]. It assesses effectiveness, efficiency, and user satisfaction [26-28]. While items are rated on a 5-point Likert scale, they are weighted so that the final scores range from 0 - 100 [26]. It is reliable, with α=.91-.92 and with good concurrent validity (α=.806) [26]. Scores of 68 are considered above average usability, and scores of 80.3 or greater are in the top 10% of MSUS scores [28]. Participants are asked to rate 10 statements such as "I thought the system was easy to use" or "I felt very confident using the system" from 1=strongly disagree to 5=strongly agree.

App Usage

To assure participants were using the meditations, data were collected via the app on the time spent listening to meditations and the number of meditations completed during the program. This allowed us to examine the dosage of mindfulness with accuracy.

About the PSS

The PSS is a reliable and valid self-report measure of stress (weak to moderate association criterion validity r=0.11-0.67, internal consistency α=.60-.91, and test-retest reliability r=0.55-0.86) [29-31]. The PSS items ask participants to rate the degree to which they have experienced feelings and thoughts of unpredictability, uncontrollability, and overload in the last month. PSS captures both daily challenges and major events within the prior 30 days. Further, 10 questions measure the degree an individual appraises their life situations as stressful using a 5-point Likert scale from "0" (never) to "4" (very often). For example, participants rated statements such as "In the last month, how often have you felt difficulties were piling up so high that you could not overcome them?" or "In the last month, how often have you felt nervous and stressed?" Higher total scores on the PSS indicate more stress [29]. For this study, the PSS measured the self-reported level of stress experienced by the participants in the month before beginning and during the last month of participating in the 5Minutes4Myself wellness program. The 10-item version was used.

About the FFMQ

The FFMQ is an adequately valid, reliable, and internally consistent self-report measure of mindfulness practice in an individual's life (α=.73-.91, n=376) [32-34]. The 39-item measure addresses 5 characteristics that have been identified as integral to mindfulness practice including: describing, observing, acting with awareness, nonjudging of inner experience, and nonreactivity to inner experiences. Questions are rated on a 5-point Likert scale ranging from "1" (never or very rarely) to "5" (very often or always). Items include "I find it difficult to stay focused on what's happening in the present," "I tell myself that I shouldn't be thinking the way I'm thinking," and "When I have distressing thoughts or images, I judge myself as good or bad, depending on what the thought/image is about." Higher total scores represent higher levels of mindfulness [32]. In the context of this study, FFMQ measured the degree of participants' mindfulness before and during the last month of participating in 5Minutes4Myself wellness program which included using the smartphone mobile app mindfulness podcasts.

Procedures

5Minutes4Myself Wellness Program

All cohorts completed equivalent procedures. The 5Minutes4Myself wellness program began with each cohort attending a community-building focus group where participants discussed what worked and did not work in their lives and completed health and well-being surveys using the mobile phone app. Participants unable to attend focus groups completed surveys using the mobile app at another time, or on paper and mailed in results (this was necessary due to an Android version app failure for downloading preintervention surveys). Next, each participant was assigned a coach trained in MI; the fidelity of MI use is reported elsewhere [35]. MI is a practice in which use of collaboration, evocation, and privileging participant autonomy are essential to eliciting participant-desired lifestyle change and supporting goal attainment [36]. Participants worked with coaches to select goals meaningful to them in the initial lifestyle consultation and created a plan for instituting the goals into her or his daily life over the 4-month program. Participants identified 3 to 5 goals, which included 1 goal of regular weekly mindfulness practice. The mindfulness goal was required of participants for this feasibility study since it was a key element of the application being developed. Coaches meet at least monthly with participants, either in person or on the phone, to discuss their goal progress and revise strategies as needed.

Postintervention focus groups were conducted after 4 month’s participation. Caregivers again discussed what worked or did not work in their lives, were queried about their view of the program elements, and completed the same surveys as pre intervention as well as the MSUS. For this analysis, we used data from the MSUS, PSS, and FFMQ.

Smartphone Mobile App

Participants were trained to use the 5Minutes4Myself mobile app to access mindfulness meditations, to create their goal profile, to program goal reminders at preferred times, and to report weekly progress on goals via an electronic check-in. It was loaded onto participants’ personal phones, either in Android or iOS format or if needed onto a loaned Android smartphone capable of operating in a wireless environment (no phone service was provided). The audio-recorded meditations were voiced by the licensed psychologist who developed them and ranged in length from 4 - 21 minutes, sequenced to increase in length over time. These began with a brief introduction to mindfulness concepts and were sequenced and available 1 per week; the focus of each meditation rotated through breath awareness, loving-kindness, and body awareness (body scans). Participants were free to choose when and where they listened to the offloaded meditations, available without an internet connection. Participants chose how often they wanted to be reminded of each of their goals: weekly, using a “guilt-free” check-in, participants for each goal included the mindfulness goal. After completing the designated week’s meditation, the mobile app advanced to the next week’s meditation; participants could revisit past meditations but could not progress to new ones until the following week. This was designed to facilitate ease of use and require no tracking on the users’ part. After the first cohort used the app, we revised the app to address bugs and improve features for following cohorts.

Data Collection, Management, and Analyses

Data gathered via the 5Minutes4Myself app were transferred from a secure third-party server from the application development company to the research team and removed from company servers. The software development personnel who managed the data completed the human participant research training. Caregivers with any missing data on stress or mindfulness measures were excluded from this analysis (n=4). Quantitative data were analyzed using statistical software IBM SPSS (version 24.0, IBM Corp). Descriptive statistics including means and SDs were calculated for participant demographics, and for perceived stress and mindfulness practice pre and post intervention. To contextualize the level of stress experienced, participants’ perceived stress scores were compared to a population-based mean for the PSS for their age group [29].

To explore mental health outcomes of the feasibility study, related samples *t* tests (2-tailed) were used to evaluate whether participants’ stress scores as a group (PSS) were statistically different after participating in the wellness program. A second related samples *t* test evaluated scores of mindfulness (FFMQ) before and at the end of the intervention to determine whether they were significantly different for the group. Next, to understand who benefitted from the program, we visualized the trends in these 2 measures from pre and post intervention. Responders were defined as having decreased PSS scores post intervention. In addition, to examine individual level change, an analysis was conducted using Norman and colleagues’ [37,38] approach that considers clinically meaningful change as at or above half a SD of the initial group mean. After standard mean differences are calculated (d=DiSx where D_i is individual change, S_x is SD of group at baseline) they are grouped into “improved” ($d\geq0.50$), “no change” ($-0.5<d>0.5$), and “worsened” ($d\leq-0.50$). Lastly, a hierarchical multiple regression was conducted to examine whether the pre-post change in mindfulness (change in FFMQ scores) explained variance in postparticipation perceived stress scores while controlling for preintervention stress levels with an α level of .05.

Ethical Considerations

The study was reviewed and approved by the University of Wisconsin-Madison Educational and Social/Behavioral Institutional Review Board (2015 - 1004). Participants were provided an informed consent form to review, given the opportunity to ask questions, and informed of their ability to opt out at any time. Participants who signed the consent form were invited to participate. In the compiled dataset, participants were assigned numbers; a participant key was created and stored on a password-protected computer. No compensation was provided.

Results

Characteristics of Participants

In total, 15 participants participated in the 5Minutes4Myself wellness program. While not all participants completed all 3 of the coaching check-ins, program completion was deemed to be participation in the initial focus group or an introductory session, completion of the lifestyle consultation, and at least 2 of 3 monthly coaching sessions. Using this metric, 13 caregivers completed the 4-month 5Minutes4Myself program for a participation rate of 87% (13/15; 2/15, 13% dropout). Further, 2 participants were excluded from the analysis due to incomplete postintervention survey responses, which were due to a family emergency and scheduling conflict with the final focus group. Neither person responded to additional requests to complete the postintervention surveys. Additionally, 11 participants completed both pre- and postintervention surveys used in this analysis. Sample characteristics and pre- and postintervention survey scores for mindfulness (FFMQ) and stress (PSS) are presented in Table 2. In this convenience sample of participants, most self-identified as White women, were married, reported caregiving full-time, were raising 2 children (one or more with an autism spectrum or other developmental disorder) with graduate school education.

Table . Sample characteristics and pre-postintervention scores[a] (N=11).

Characteristics	Values
Age (years), mean (range, SD)	47.85 (36 - 65, 8.46)
Sex, n (%)	
Male	1 (9)
Female	10 (91)
Race, n (%)	
White	11 (100)
Family income, median (range; IQR)	US $87,499 (<US $25,000 - $100,000+; US $50,000-$95,000)
Education, n (%)	
Graduate	8 (73)
Undergraduate	0 (0)
Associates	1 (9)
High school	1 (9)
Marital status, n (%)	
Single	1 (9)
Married	10 (91)
Caregiver status, n (%)	
Full-time caregiver[b]	7 (64)
Employment status, n (%)	
Part-time employee	3 (27)
Full-time employee	2 (18)
Number of children, mean (range, SD)	2.21 (1 - 5, 1.12)
Age of child with autism spectrum disorder (years), mean (range, SD)	14.1 (9 - 20, 3.69)
Baseline stress (PSS[c]), mean (range, SD)	23 (16 - 29, 4.45)
Stress post intervention (PSS), mean (range, SD)	21 (13 - 36, 8.23)
Baseline mindfulness (FFMQ[d]), mean (range, SD)	119.91 (90 - 146, 15.49)
Postintervention mindfulness (FFMQ), mean (range, SD)	127 (93 - 153, 17.71)

[a]One participant provided only partial demographic information.

[b]Full-time caregiver indicates major or sole responsibility for child's care around school hours.

[c]PSS: Perceived Stress Scale.

[d]FFMQ: Five Facet Mindfulness Questionnaire.

App Quality and Usability

Overall, the mean caregiver rating of the app was 78.6 (range 55 - 100) on the MSUS (>65 is above average usability and >80 is the top 10% of all products tested using the MSUS) [39]. Usability ratings improved from the version used by the first cohort (mean 67.5, SD 22.6) to the second version used by the remaining cohorts (mean 86.2, SD 13.5). Further, 3 of the 11 caregivers rated it below the 65 threshold; all these participants were in the first cohort using the first version of the app. As is common in app development, several issues led to the app being unavailable at times to all cohorts during their program. Specifically, software updates to the mobile app were performed during the intervention period and required resetting of goal reminders or resulted in limited availability of meditations.

App Usage and Frequency of Meditations

We aimed to support caregivers' adoption of the app-delivered program by providing reminder notifications and coaching support, to encourage meditations at least twice weekly over 4 months' time. Further, 1 participant met these criteria, and another nearly met it, but the remainder meditated twice weekly only 5 or less than 5 weeks' time over the course of the program. On average, they listened for 10.2 (SD 3.8) minutes per week to 20 meditations. Shorter tracks (~5 min in length) were most frequently listened to. The user data suggest that participants, as a group, attempted to listen more often (439 attempts) than they completed podcasts (314 complete or within 30 s of completion), pointing to potential app instability and preference for shorter meditations.

Levels of Stress and Mindfulness

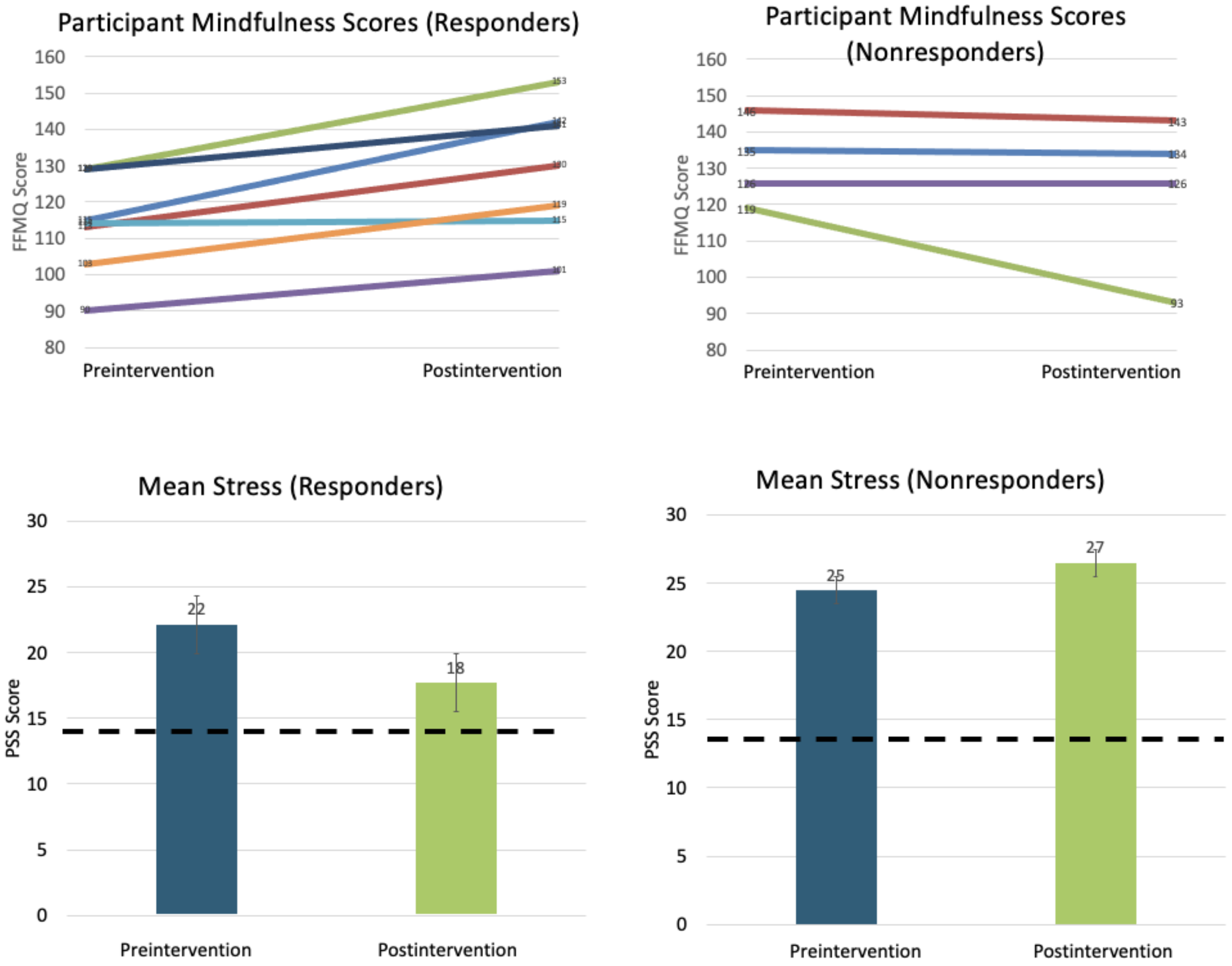
Overview

As a group, participants scored 23 (SD 4.45) on the preintervention measure of stress and 21 (SD 8.23) post intervention. As a reference, the normative mean for this age group, based on a Harris Poll of 2387 respondents, was 12.6 (SD 6.1) on the PSS 10 item [30]. Our reported values are similar to baseline PSS 10 item scores for Bazzano and colleagues' [18] sample of caregivers (n=66, PSS mean 22.6, SD 6.2), and Conner and colleagues' [2] (n=67, PSS mean 20.73, SD 7.03). These PSS scores suggest that overall this was a highly stressed group, and these findings are similar to the level of stress found in other studies of caregivers of children with autism or disabilities. At baseline, 3 of 11 participants were over 2 SDs above the normative population mean; the remaining 7 participants were over 1 SD above the mean for population stress levels on PSS [30]. After completion of the 5Minutes4Myself wellness program, more than half of the participants (6 of 11) were within 1 SD of mean for population stress levels [30].

In terms of mindfulness, the preintervention group mindfulness mean was 119.91 (SD 15.49) and ranged from 90 to 146. Post intervention, the mean FFMQ scores were 127 (SD 17.71) with a similar range of 93 to 153. No population norms are available for the FFMQ for comparison to allow us to reflect on the meaningfulness of the participants' scores.

To assist in visualizing the data, participants were grouped into mindfulness responders and nonresponders. Responders were identified as individuals whose mindfulness scores improved over the intervention period whereas nonresponders' mindfulness scores either stayed the same or decreased. Figure 2 displays the individual mindfulness scores (FFMQ) at pre and post intervention and mean stress scores (PSS) for these 2 groups. The small error bars suggest the 2 groups of caregivers are similar to each other in terms of stress.

Figure 2. This figure groups scores into responders and nonresponders based on their mindfulness score (FFMQ). Responders' mindfulness scores on the FFMQ increased from pre intervention to post intervention, as hypothesized for 7 participants. Nonresponders' mindfulness scores decreased from pre intervention to post intervention for 4 participants. Each line represents 1 participant's mindfulness score (FFMQ) from pre intervention to post intervention. Corresponding group mean stress (PSS) scores are displayed below. The dotted line indicates a population-based mean for the PSS for this age group [29]. FFMQ: Five Facet of Mindfulness Questionnaire; PSS: Perceived Stress Scale.



In terms of individual change in the PSS, individual change calculations ranged from −1.57 to 2.02 (reverse coded so that positive numbers indicated reduced stress). When grouped according to Norman and colleagues' [38] guidelines, 4 participants experienced worse stress and 7 had reduced stress post participation. This affirms the responder visualization of the data.

Changes in Stress and Mindfulness Pre and Post Intervention

A related samples *t* test evaluated whether participants' PSS mean stress scores were significantly different before and after participating in the wellness program. Preliminary analyses of normality, random selection, and homogeneity of variances were conducted and ensured no assumptions were violated. The group PSS postintervention mean was not statistically and significantly different than the preintervention group mean (t_{10}=1.20, *P*=.26). The effect size *d* of 0.36 indicates a small to medium effect size. Given the small sample used for this feasibility study, these findings do not clearly indicate whether this change in PSS score was a meaningful difference for all or some of our participants [40].

For mindfulness, a related samples *t* test evaluated whether participants' FFMQ scores differed following participation in the program. The group mean FFMQ was not significantly different following participation in the program (t_{10}=−1.57, *P*=.15). However, in this case, the effect size was larger (*d*=0.47), suggesting that a treatment effect might be evident with a larger sample size.

Postintervention Stress as Predicted by Change in Mindfulness

Given that 1 effect size suggested some potential treatment effect and the trends noted in responders, a hierarchical multiple regression was used to assess the central question of how change in mindfulness (FFMQ) predicted levels of stress post intervention (PSS), after controlling for preintervention stress levels. Preliminary analyses were conducted and ensured all assumptions of multiple regression were met including linear relationship between variables, normal distribution of residuals, an independent sample, no multicollinearity, and homoscedasticity, despite the small sample size (n=11). The results of this regression are provided in Table 3. These results suggest that 20% of the variance in stress, beyond the 50% of variance accounted for by prestress levels, can be attributed to pre-post changes in the FFMQ (or increases in mindfulness; *P*=.04). Therefore, change in the FFMQ was a significant predictor of postintervention stress levels for these caregivers of children with ASDs.

Table . Hierarchical regression analysis of predictors of stress.

Predictor variables	Model 1	Model 2
Preintervention stress (Perceived Stress Scale), B[a] (SE)	1.29[b] (0.395)	0.86[b] (0.36)
Change in mindfulness (Five Facet Mindfulness Questionnaire), B (SE)	—[c]	−0.28[b] (0.11)
R^2/adjusted R^2	0.544/0.493	0.743/0.679
F change in R^2	10.73[b]	6.20[b]

[a]B: unstandardized β.

[b]$P<.05$

[c]Not applicable.

Discussion

Principal Results

This study examined the usability and usage of an app designed to deliver mindfulness meditations and the effects of the mindfulness training on stress experienced by caregivers of children with ASD. The mobile app, as designed and then revised, was found to be above average in usability, especially the second version. Although caregivers did not use the mindfulness meditations as frequently as intended, they did use them on average for 10 minutes per week, a microdose compared to other mindfulness programs for caregivers.

While group differences in pre- and poststress and mindfulness scores were not significant in this analysis, 1 effect size offered tentative support for the impact of the 5Minutes4Myself program on mindfulness. As Cohen [41] suggests, statistical significance and calculations of effect size are useful only in so far as they can be used to interpret the meaningfulness of the score change to the participant [39]. This seems to be supported by the individual level analysis, where 7 of 11 caregivers showed improvement or decreased stress post participation. In addition, this study linked increased mindfulness to a reduction in stress; increases in mindfulness accounted for 20% of the individual variance in postintervention stress scores.

Limitations

The pre- or postintervention group changes in stress and mindfulness were statistically insignificant. While the individual change analysis did show benefits for some caregivers, this analysis does not identify if any specific characteristics indicate who might be more likely to benefit from the program. Further, the analysis relied on a small sample (n=11). Future analyses of larger samples may help us to better understand the effects of this wellness program on stress and the contribution of the mobile-mindfulness component.

While these preliminary findings support the use of a mobile app in the 5Minutes4Myself wellness program, the developing technology was a significant barrier for some caregivers. Complications with application were a limitation of this study and accounted for lost postsurvey data for 2 participants.

We also did not conduct follow-up measures for this feasibility study. Long-term follow-up measures are needed to examine benefits of participation in the 5Minutes4Myself wellness program within the first- and second-year post participation. Further exploration of the 5Minutes4Myself wellness program will offer greater insight into how to best serve caregivers of children with ASD to support stress reduction and positive well-being outcomes.

Comparison With Prior Work

When compared to other studies, our caregivers showed similar percentages in reduction of stress (9% mean decrease) and increase in mindfulness (10% increase) as studies assessing larger populations of caregivers who reported statistically significant findings. For example, 2 traditional and more time-intensive MBSR programs found that caregivers' increased mindfulness was associated with a 7.4% reduction in self-rated perceived stress [19] and that parents and teachers of children with special needs had a 10% increase in mindfulness (as calculated from reported data) and an 8% reduction in stress [18].

Given that increases in mindfulness accounted for 20% of the individual variance in postintervention stress scores, this highlights the importance of mindfulness, be it a small dose, in diminishing stress. Benn and colleagues' [19] work corroborated that increased mindfulness was key to stress reduction. Using the same stress and mindfulness measures as in our study to assess outcomes of an 8-week MBSR-style training program for parents of children with special needs, Benn and colleagues [19] conducted a mediations analysis. Their analysis identified increased mindfulness as the mediator for positive changes in health and well-being outcomes.

Consistent with previous research that used much larger "doses" of mindfulness, such as those provided in MBSR-based programs [1,12,19], the results of this analysis suggest that a micromindfulness approach is a viable intervention for stress reduction for caregivers of children with ASD. Carmody and

Baer's [42] analysis seems to confirm that shorter mindfulness practices are indeed viable alternatives to the time-intense MBSR programs. They found that in-class hours and treatment effect sizes were not significantly related for their population, suggesting low-dose practices, such as what we used in this program, can be an effective alternative for time-challenged individuals. Lunsky and colleagues [43] also noted that while most of the parents of children with ASD in their study attended only 4 of 6 sessions and reported brief and irregular home practice, these parents still derived psychological benefit with less distress post intervention compared to a control group. Berghoff and colleagues [44] examined dose for college students assigned to either 10- or 20-minute practices of mindfulness for 14 days; they found that significant reductions in stress were independent of the actual time practiced. The total dose for most students was only an hour over 2 weeks. While our caregivers did not listen as often as proposed, using a mobile app for this population appeared to be a feasible delivery system for brief mindfulness training and shorter meditations were more practical for this group. Our approach offered caregivers the flexibility to choose when and where they listened to brief (4 - 21 min) mindfulness podcasts while having regular guided practices such as class sessions.

Use of a mobile-delivery system for mindfulness practice opens the opportunity for an intervention that is client-centered and can be more readily incorporated into a participant's daily routine. Hollis and colleagues [45] suggest that mobile apps have the potential to transform service provision to clients by providing content, monitoring, and access to providers. Yet, findings from a large-scale study using a newly developed mindfulness application (VGZ Mindfulness Coach; VGZ Health Insurer N.V) appear to suggest flexible provision of easily accessible mindfulness content is not sufficient to encourage participation; only 58% of their participants used the application [46]. Engaging clients and clinicians to collaboratively design to meet unmet needs and to assure the usability of the technology is essential to not only develop strongly evidence-based interventions but also assure their adoption by the targeted population [45,47,48].

Most mental health programs for caregivers of children with autism have been delivered solely in person, which may partially account for the high attrition rates ranging from 17% - 45% [49]. We choose to use an app to deliver content to increase the accessibility of this program for caregivers and decrease attrition, which in our small sample was 13%. Preliminary data suggest this mindfulness training provided stress reduction for some of the participating caregivers. In line with Latulippe and colleagues' [23] findings, we found that in designing an eHealth option for caregivers we needed to attend to user characteristics including (1) limits in financial and personal resources, (2) available access to internet, and (3) users' comfort in using and managing technology. In addition, for our app it was important to attend to caregivers' reported reduced problem-solving capacities which were likely due to the ongoing stress and the vigilance of caregiving. Using a collaborative participant-centered approach to app design helped ensure we created a highly usable application that was confirmed via caregiver ratings of above-average usability on the MSUS [10,27,39].

Conclusions

To our knowledge, this program is the first of its kind to use a mobile app to deliver a wellness program, that included mindfulness meditations, to caregivers of children with autism. The unique contribution of this wellness program, beyond these initial promising findings of increasing mindfulness and reducing stress, is its intentional design to easily fit into caregivers' challenging and time-demanding daily lives. The 5Minutes4Myself app can be used at anytime and anywhere without the need for traditional in-person mindfulness training. As this study suggests, use of the mobile-mindfulness component may be a feasible intervention for caregivers that allows them to participate in an evidence-based intervention for stress reduction at a time and place that is convenient and possible within their everyday lives.

Acknowledgments

This study was funded by grants from the American Occupational Therapy Foundation and the Virginia Horne Henry Fund, University of Wisconsin-Madison.

Conflicts of Interest

None declared.

References

1. Cachia RL, Anderson A, Moore DWJ. Mindfulness, stress and well-being in parents of children with autism spectrum disorder: a systematic review. J Child Fam Stud 2016 Jan;25(1):1-14. [doi: 10.1007/s10826-015-0193-8]
2. Conner CM, White SW. Stress in mothers of children with autism: trait mindfulness as a protective factor. Res Autism Spectr Disord 2014 Jun;8(6):617-624. [doi: 10.1016/j.rasd.2014.02.001]
3. Leonard BJ, Johnson AL, Brust JD. Caregivers of children with disabilities: a comparison of those managing "OK" and those needing more help. Child Health Care 1993;22(2):93-105. [doi: 10.1207/s15326888chc2202_2] [Medline: 10171474]
4. Peer JW, Hillman SB. Stress and resilience for parents of children with intellectual and developmental disabilities: a review of key factors and recommendations for practitioners. Pol Pract Intel Disabi 2014 Jun;11(2):92-98. [doi: 10.1111/jppi.12072]

5. Sawyer MG, Bittman M, La Greca AM, Crettenden AD, Harchak TF, Martin J. Time demands of caring for children with autism: what are the implications for maternal mental health? J Autism Dev Disord 2010 May;40(5):620-628. [doi: 10.1007/s10803-009-0912-3] [Medline: 19949845]
6. Zaidman-Zait A, Mirenda P, Duku E, et al. Impact of personal and social resources on parenting stress in mothers of children with autism spectrum disorder. Autism 2017 Feb;21(2):155-166. [doi: 10.1177/1362361316633033] [Medline: 27091948]
7. Vasilopoulou E, Nisbet J. The quality of life of parents of children with autism spectrum disorder: a systematic review. Res Autism Spectr Disord 2016 Mar;23:36-49. [doi: 10.1016/j.rasd.2015.11.008]
8. Werner S, Shulman C. Subjective well-being among family caregivers of individuals with developmental disabilities: the role of affiliate stigma and psychosocial moderating variables. Res Dev Disabil 2013 Nov;34(11):4103-4114. [doi: 10.1016/j.ridd.2013.08.029]
9. Estes A, Munson J, Dawson G, Koehler E, Zhou XH, Abbott R. Parenting stress and psychological functioning among mothers of preschool children with autism and developmental delay. Autism 2009 Jul;13(4):375-387. [doi: 10.1177/1362361309105658] [Medline: 19535467]
10. Larson E, Miller-Bishoff T. Family routines within the ecological niche: an analysis of the psychological well-being of U.S. caregivers of children with disabilities. Front Psychol 2014;5:1-14. [doi: 10.3389/fpsyg.2014.00495]
11. Hofmann SG, Grossman P, Hinton DE. Loving-kindness and compassion meditation: potential for psychological interventions. Clin Psychol Rev 2011 Nov;31(7):1126-1132. [doi: 10.1016/j.cpr.2011.07.003] [Medline: 21840289]
12. Dykens EM, Fisher MH, Taylor JL, Lambert W, Miodrag N. Reducing distress in mothers of children with autism and other disabilities: a randomized trial. Pediatrics 2014 Aug;134(2):e454-e463. [doi: 10.1542/peds.2013-3164] [Medline: 25049350]
13. Ferraioli SJ, Harris SL. Comparative effects of mindfulness and skills-based parent training programs for parents of children with autism: feasibility and preliminary outcome data. Mindfulness (N Y) 2013 Jun;4(2):89-101. [doi: 10.1007/s12671-012-0099-0]
14. Neece CL. Mindfulness ‐ based stress reduction for parents of young children with developmental delays: implications for parental mental health and child behavior problems. Res Intellect Disabil 2014 Mar;27(2):174-186. [doi: 10.1111/jar.12064]
15. Petcharat M, Liehr P. Mindfulness training for parents of children with special needs: guidance for nurses in mental health practice. Child Adoles Psych Nursing 2017 Feb;30(1):35-46. [doi: 10.1111/jcap.12169]
16. Whitebird RR, Kreitzer MJ, Crain AL, Lewis BA, Hanson LR, Enstad CJ. Mindfulness-based stress reduction for family caregivers: a randomized controlled trial. Gerontol 2013 Aug;53(4):676-686. [doi: 10.1093/geront/gns126] [Medline: 23070934]
17. Rayan A, Ahmad M. Effectiveness of mindfulness-based intervention on perceived stress, anxiety, and depression among parents of children with autism spectrum disorder. Mindfulness (NY) 2017 Jun;8(3):677-690. [doi: 10.1007/s12671-016-0595-8]
18. Bazzano A, Wolfe C, Zylowska L, et al. Mindfulness based stress reduction (MBSR) for parents and caregivers of individuals with developmental disabilities: a community-based approach. J Child Fam Stud 2015 Feb;24(2):298-308. [doi: 10.1007/s10826-013-9836-9]
19. Benn R, Akiva T, Arel S, Roeser RW. Mindfulness training effects for parents and educators of children with special needs. Dev Psychol 2012;48(5):1476-1487. [doi: 10.1037/a0027537]
20. Tomori K, Uezu S, Kinjo S, Ogahara K, Nagatani R, Higashi T. Utilization of the iPad application: Aid for Decision-making in Occupation Choice. Occup Ther Int 2012 Jun;19(2):88-97. [doi: 10.1002/oti.325] [Medline: 21936012]
21. Luxton DD, Hansen RN, Stanfill K. Mobile app self-care versus in-office care for stress reduction: a cost minimization analysis. J Telemed Telecare 2014 Dec;20(8):431-435. [doi: 10.1177/1357633X14555616] [Medline: 25316037]
22. Chatterjee A, Prinz A, Gerdes M, Martinez S. Digital interventions on healthy lifestyle management: systematic review. J Med Internet Res 2021 Nov 17;23(11):e26931. [doi: 10.2196/26931] [Medline: 34787575]
23. Latulippe K, Hamel C, Giroux D. Social health inequalities and eHealth: a literature review with qualitative synthesis of theoretical and empirical studies. J Med Internet Res 2017 Apr 27;19(4):e136. [doi: 10.2196/jmir.6731] [Medline: 28450271]
24. Larson EA. 5Minutes4Myself: development of a wellness program for caregivers of children with autism. Can J Occup Ther 2024 Mar;91(1):17-28. [doi: 10.1177/00084174231172036] [Medline: 37106578]
25. Bitsika V, Sharpley CF. Stress, anxiety and depression among parents of children with autism spectrum disorder. Aust J guid couns 2004 Dec;14(2):151-161. [doi: 10.1017/S1037291100002466]
26. Blanck P, Perleth S, Heidenreich T, et al. Effects of mindfulness exercises as stand-alone intervention on symptoms of anxiety and depression: systematic review and meta-analysis. Behav Res Ther 2018 Mar;102:25-35. [doi: 10.1016/j.brat.2017.12.002]
27. Brooke J. SUS: a ‘quick and dirty’ usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability Evaluation in Industry: Taylor and Francis; 1996:189-194.
28. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
29. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [doi: 10.2307/2136404] [Medline: 6668417]

30. Cohen S, Williamson G. Perceived stress in a probability sample of the United States. In: Spacapan S, Oskamp O, editors. The Social Psychology of Health: Sage; 1988.
31. Lee EH. Review of the psychometric evidence of the Perceived Stress Scale. Asian Nurs Res (Korean Soc Nurs Sci) 2012 Dec;6(4):121-127. [doi: 10.1016/j.anr.2012.08.004]
32. Baer RA, Smith GT, Hopkins J, Krietemeyer J, Toney L. Using self-report assessment methods to explore facets of mindfulness. Assessment 2006 Mar;13(1):27-45. [doi: 10.1177/1073191105283504] [Medline: 16443717]
33. Baer RA, Smith GT, Lykins E, et al. Construct validity of the Five Facet Mindfulness Questionnaire in meditating and nonmeditating samples. Assessment 2008 Sep;15(3):329-342. [doi: 10.1177/1073191107313003]
34. Bohlmeijer E, ten Klooster PM, Fledderus M, Veehof M, Baer R. Psychometric properties of the Five Facet Mindfulness Questionnaire in depressed adults and development of a short form. Assessment 2011 Sep;18(3):308-320. [doi: 10.1177/1073191111408231] [Medline: 21586480]
35. Larson EA, Jeglum N, Shmays C. Mixed-method evaluation of fidelity of motivational interviewing-based coaching in 5Minutes4Myself wellness program for caregivers of children with autism. PEC Innov 2023 Dec;2:100149. [doi: 10.1016/j.pecinn.2023.100149] [Medline: 37214523]
36. Martins RK, McNeil DW. Review of motivational interviewing in promoting health behaviors. Clin Psychol Rev 2009 Jun;29(4):283-293. [doi: 10.1016/j.cpr.2009.02.001]
37. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care 2003 May;41(5):582-592. [doi: 10.1097/01.MLR.0000062554.74615.4C] [Medline: 12719681]
38. Norman GR, Sloan JA, Wyrwich KW. The truly remarkable universality of half a standard deviation: confirmation through another look. Expert Rev Pharmacoecon Outcomes Res 2004 Oct;4(5):581-585. [doi: 10.1586/14737167.4.5.581] [Medline: 19807551]
39. Sauro J. Measuring usability with the System Usability Scale (SUS). MeasuringU. 2011 Feb 3. URL: http://www.measuringu.com/sus.php [accessed 2024-10-24]
40. Ellis PD. Thresholds for interpreting effect sizes. The Hong Kong Polytechnic University. 2009. URL: http://www.polyu.edu.hk/mm/effectsizefaqs/thresholds_for_interpreting_effect_sizes2.html [accessed 2018-06-04]
41. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155] [Medline: 19565683]
42. Carmody J, Baer RA. How long does a mindfulness ‐ based stress reduction program need to be? A review of class contact hours and effect sizes for psychological distress. J Clin Psychol 2009 Jun;65(6):627-638. [doi: 10.1002/jclp.20555]
43. Lunsky Y, Hastings RP, Weiss JA, M Palucka A, Hutton S, White K. Comparative effects of mindfulness and support and information group interventions for parents of adults with autism spectrum disorder and other developmental disabilities. J Autism Dev Disord 2017 Jun;47(6):1769-1779. [doi: 10.1007/s10803-017-3099-z] [Medline: 28374207]
44. Berghoff CR, Wheeless LE, Ritzert TR, Wooley CM, Forsyth JP. Mindfulness meditation adherence in a college sample: comparison of a 10-min versus 20-min 2-week daily practice. Mindfulness (NY) 2017 Dec;8(6):1513-1521. [doi: 10.1007/s12671-017-0717-y]
45. Hollis C, Morriss R, Martin J, et al. Technological innovations in mental healthcare: harnessing the digital revolution. Br J Psychiatry 2015 Apr;206(4):263-265. [doi: 10.1192/bjp.bp.113.142612] [Medline: 25833865]
46. van Emmerik AAP, Berings F, Lancee J. Efficacy of a mindfulness-based mobile application: a randomized waiting-list controlled trial. Mindfulness (N Y) 2018 Feb;9(1):187-198. [doi: 10.1007/s12671-017-0761-7]
47. Egan KJ, Clark P, Deen Z, et al. Understanding current needs and future expectations of informal caregivers for technology to support health and well-being: national survey study. J Aging 2022;5(1):e15413. [doi: 10.2196/15413]
48. Merchán-Baeza JA, Borralleras Andreu C, Minobes-Molina E, Grau Carrión S, Romero-Mas M, Ramon-Aribau A. Co-created technological solutions for caregivers in health care: systematic review. J Med Internet Res 2023 May 1;25:e41260. [doi: 10.2196/41260] [Medline: 37126384]
49. Da Paz NS, Wallander JL. Interventions that target improvements in mental health for parents of children with autism spectrum disorders: a narrative review. Clin Psychol Rev 2017 Feb;51:1-14. [doi: 10.1016/j.cpr.2016.10.006]

Abbreviations

ASD: autism spectrum disorder
FFMQ: Five Facet Mindfulness Questionnaire
MBI: mindfulness-based intervention
MBSR: mindfulness-based stress reduction
MI: motivational interviewing
MSUS: Modified System Usability Scale
PSS: Perceived Stress Scale

Edited by C Jacob; submitted 31.10.23; peer-reviewed by C Slade, P Washington; revised version received 03.06.24; accepted 04.06.24; published 31.10.24.

Please cite as:
Larson E, Mattie RL, Riffkin SA
Assessment of Acceptability, Usage, and Impact on Caregivers of Children With Autism's Stress and Mindfulness: Multiple-Method Feasibility Study of the 5Minutes4Myself App's Mindfulness Module
JMIR Hum Factors 2024;11:e54171
URL: https://humanfactors.jmir.org/2024/1/e54171
doi:10.2196/54171

Original Paper

Leveraging Generative AI Tools to Support the Development of Digital Solutions in Health Care Research: Case Study

Danissa V Rodriguez[1], PhD; Katharine Lawrence[1,2], MPH, MD; Javier Gonzalez[2], BS; Beatrix Brandfield-Harvey[1], BA; Lynn Xu[1], MPH; Sumaiya Tasneem[1], MPH; Defne L Levine[1], MPH; Devin Mann[1,2], MS, MD

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

[2]Medical Center Information Technology, Department of Health Informatics, New York University Langone Health, New York, NY, United States

Corresponding Author:
Danissa V Rodriguez, PhD
Department of Population Health
New York University Grossman School of Medicine
227 East 30th Street
6th Floor
New York, NY, 10016
United States
Phone: 1 646 501 2684
Email: danissa.rodriguez@nyulangone.org

Abstract

Background: Generative artificial intelligence has the potential to revolutionize health technology product development by improving coding quality, efficiency, documentation, quality assessment and review, and troubleshooting.

Objective: This paper explores the application of a commercially available generative artificial intelligence tool (ChatGPT) to the development of a digital health behavior change intervention designed to support patient engagement in a commercial digital diabetes prevention program.

Methods: We examined the capacity, advantages, and limitations of ChatGPT to support digital product idea conceptualization, intervention content development, and the software engineering process, including software requirement generation, software design, and code production. In total, 11 evaluators, each with at least 10 years of experience in fields of study ranging from medicine and implementation science to computer science, participated in the output review process (ChatGPT vs human-generated output). All had familiarity or prior exposure to the original personalized automatic messaging system intervention. The evaluators rated the ChatGPT-produced outputs in terms of understandability, usability, novelty, relevance, completeness, and efficiency.

Results: Most metrics received positive scores. We identified that ChatGPT can (1) support developers to achieve high-quality products faster and (2) facilitate nontechnical communication and system understanding between technical and nontechnical team members around the development goal of rapid and easy-to-build computational solutions for medical technologies.

Conclusions: ChatGPT can serve as a usable facilitator for researchers engaging in the software development life cycle, from product conceptualization to feature identification and user story development to code generation.

Trial Registration: ClinicalTrials.gov NCT04049500; https://clinicaltrials.gov/ct2/show/NCT04049500

(JMIR Hum Factors 2024;11:e52885) doi:10.2196/52885

KEYWORDS

digital health; GenAI; generative; artificial intelligence; ChatGPT; software engineering; mHealth; mobile health; app; apps; application; applications; diabetes; diabetic; diabetes prevention; digital prescription; software; engagement; behaviour change; behavior change; developer; developers; LLM; LLMs; language model; language models; NLP; natural language processing

Introduction

Health care has undergone a digital transformation, resulting in a growing reliance on software engineering for medical use cases, including health care research. However, little guidance exists for health researchers on how to effectively develop digital health interventions [1]; in particular, software development challenges that include expertise gaps in coding, custom development needs, high costs, and time constraints result in multilevel barriers to designing and deploying a usable, scalable, and sustainable digital health product [1].

Generative artificial intelligence (GenAI) technologies such as ChatGPT can potentially support researchers in health technology endeavors by providing foundational frameworks and processes for the software development life cycle [2]. These systems can help reduce time and enhance precision for technology-based research projects by supporting both nonprogrammers and experienced programmers in code development, troubleshooting, and cleaning [2]. Moreover, the ability to use GenAI to generate content from different perspectives (expert or nonexpert) can facilitate and improve communication between technical and nontechnical team members of multidisciplinary teams. For example, a nontechnical team member can write their ideas in natural text and then use GenAI to request assistance in creating discussion points to communicate to a technical team audience. GenAI tools may also help health technology researchers refine research questions, identify appropriate theoretical frameworks and models, and leverage popular implementation strategies such as design thinking to build effective, theory-grounded, and evidence-based digital health interventions. ChatGPT (OpenAI, Microsoft Corporation) has already demonstrated feasibility as a support tool for clinical decision support development in health care [3], and more broadly as a coding copilot in programming and engineering [4,5].

This study explores the use of ChatGPT to recreate a personalized automatic messaging system (PAMS), which was developed as part of a digital health research initiative to support patient engagement with a commercial digital diabetes prevention program (dDPP). We examine the capacity, advantages, and limitations of ChatGPT to support product ideation and conceptualization, intervention content development, and the software engineering process including software requirement generation, software design, and code production. This paper provides insights to support the GenAI-assisted development of computational tools that are usable, reliable, extensible, and in line with the standards of modern coding practices. The framework includes prompts for both the intervention conceptualization as well as the main phases of the software development process.

Methods

Settings and Intervention Development Context

In previous work [6], we described the development of PAMS, a novel integrated multicomponent communications platform, to promote patient-provider communication and patient engagement in a commercial dDPP (Noom; Noom, Inc). The PAMS intervention included early prototyping and user testing, a technical development phase, and a randomized controlled trial. The core content and user experience features of PAMS were identified, prototyped, and evaluated using the well-established design thinking "discover, define, design, and test" approach to iteratively gather information, define, design, and refine the engagement intervention [7]. Stakeholders included: patients with prediabetes and their support network (eg, caregivers and partners), primary care providers, health technologists, programmers and computer scientists, behavioral change theorists and subject matter experts, the research administrative team, and dDPP product developers and coaches. The main components of this PAMS intervention include (1) a theory-driven behavior change messaging library, (2) a personalized automated message system delivery platform (SMS text messaging–based), and (3) EHR-integrated data visualizations. The PAMS messaging library uses an integrated framework that combines established theoretical models for behavior change with human-centered design strategies to maximize the evidence-based conditions for behavior change and the user acceptance and use of a digital health product. The technical development of PAMS followed an agile software development approach based on incremental 2-week sprint cycles consisting of requirement planning, design, development, and testing of a specific set of functional features. In this paper, we will recreate this development process using GenAI (ChatGPT).

ChatGPT-PAMS Experiment Design

To evaluate the effectiveness of using GenAI to support the development of digital tools in medical settings, our experiment is based on recreating PAMS using GenAI (ChatGPT) and evaluating human-generated vs ChatGPT-generated documentation. To accurately capture the ideation and development process, our multidisciplinary team reviewed all documentation and processes used in the early stages of PAMS conceptualization, including supporting theoretical models, content and features, and technical development. We then recreated these processes via a series of prompts for ChatGPT-4 to assist with the generation of theory, content, user stories, requirement documents, design diagrams, and the code for a subset of the requirements. Outputs from ChatGPT were reviewed and compared to human-generated documentation by 11 evaluating team members. Evaluators consisted of clinicians, behavioral scientists, programmers, and research staff working in digital health and technology for behavior change research. Collectively, they represent more than 50 years of clinical, research, design, and computer science experience. The evaluators independently rated the quality of various aspects of information provided by ChatGPT on a Likert scale, where higher ratings indicated greater quality of information (*1: very poor; 2: poor; 3: acceptable; 4: good; 5: very good; N/A: not applicable*). Aspects of evaluation included: understandability (Does this output make sense given the context of the study and prompts?), novelty (Were new ideas generated?) [3], usability (Does this create a usable output?), relevance (Does this create a useful output?), efficiency (Would having these outputs have saved time?), and potential for bias (What unintended consequences might arise from these outputs?) [6]. Evaluators were also asked to give an overall score on the quality of the ChatGPT output (Overall, how good would you say this output is?). Post review, a group debrief was conducted, using a semistructured interview guide to facilitate discussion regarding perceptions of outputs and rationale for ratings.

Ethical Considerations

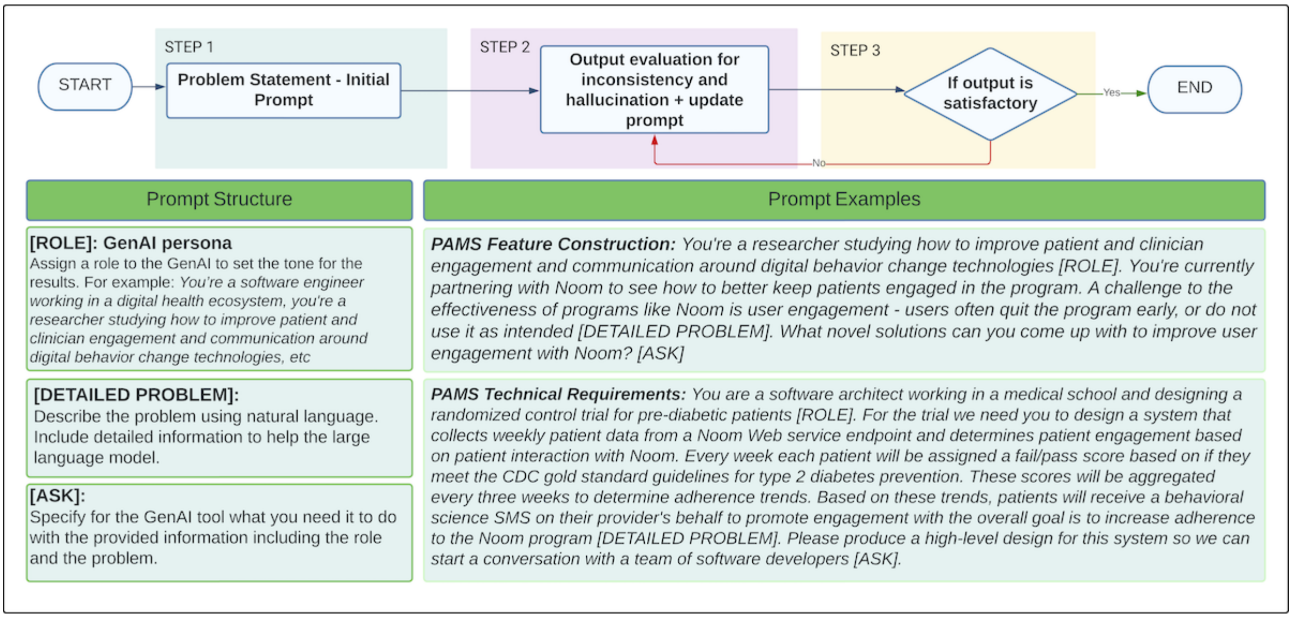
Ethical considerations helped guide the initial development of research methods and reduce potential risks for participants in the original study implementation with the PAMS intervention [7]. Recreating the technical development of a system previously built as part of the dDPP randomized controlled trial (NCT04049500) has not introduced any new risks to patients. Patients were not involved in this research examining the use of GenAI in the development of digital health care solutions. No patient data was used in the prompt generation phase.

Regarding ethical considerations for the methods used in this paper, as an attempt to mitigate evaluator biases, we worked with a diverse team of evaluators who were aware of the initial study but were not necessarily involved in the technical development. Additionally, we understand the limitations and concerns of the use of ChatGPT including possible hallucinations and incorrect answers. Thus, we emphasize the need for human expertise to identify correct and incorrect outputs and have flagged this as a study consideration. When developing the GenAI-based solution, we used the same considerations for data security, patient usability, accessibility, and data privacy used in the original human-developed solution.

Prompt Generation Framework

Prompt engineering focuses on the skill of designing and creating effective prompts that guide ChatGPT to produce the best possible output for your task. We followed existing literature [8-11] combined with our expertise and experimentation to provide a framework that yields the best result when developing a digital solution like PAMS (Figure 1).

Figure 1. ChatGPT prompt structure and prompt examples. CDC: Centers for Disease Control and Prevention; GenAI: generative artificial intelligence; PAMS: personalized automatic messaging system.



Results

PAMS Concept and User Experience Generation

Overview

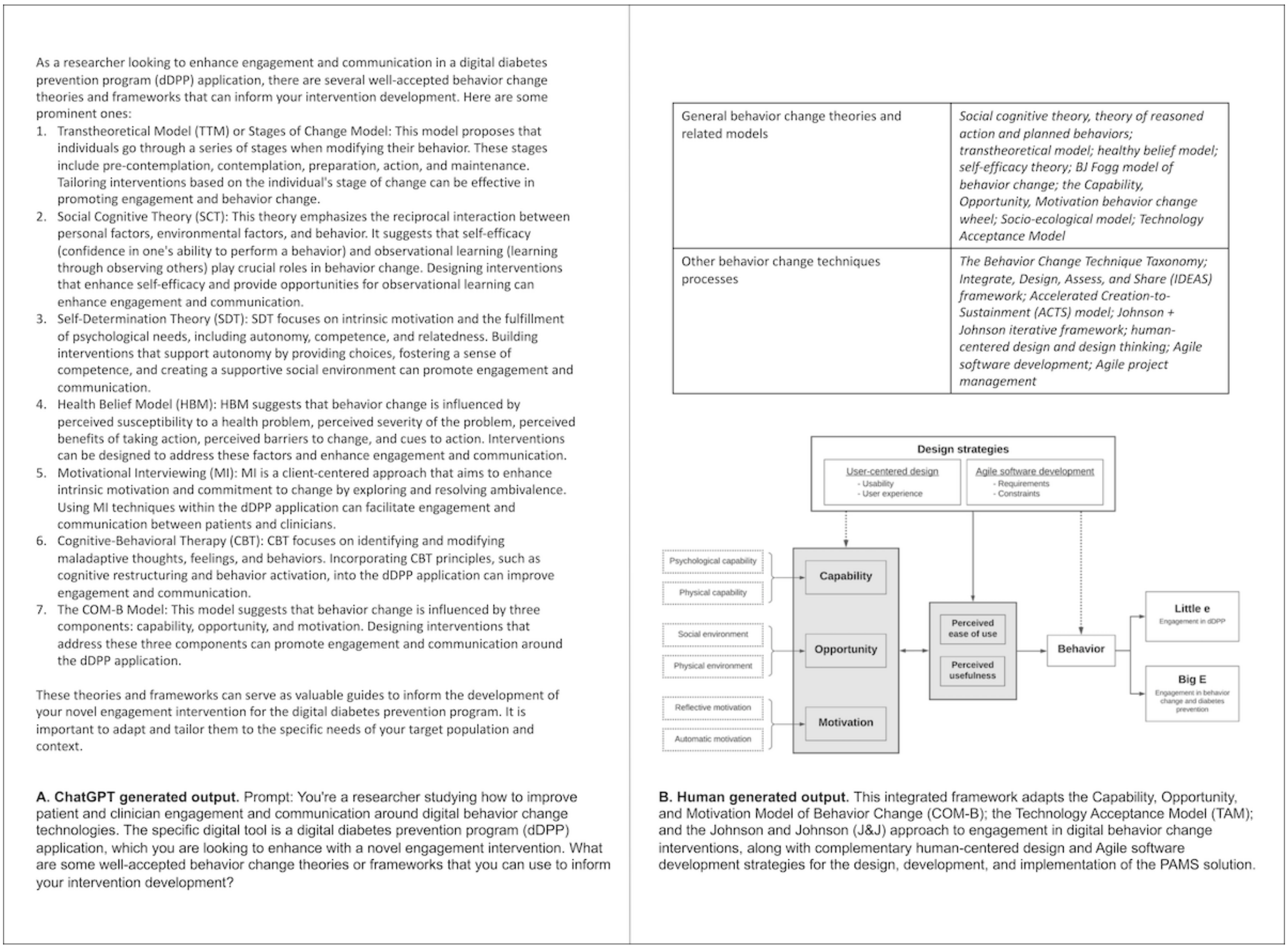
Core components of the PAMS intervention were conceptualized and designed via an underlying behavior change theory, design principles and personas, and a message content library.

Underpinning Behavior Change Theory and Approach

Human-Generated Solution

Leveraging behavior change literature review and interviews with behavior change theory content experts (n=4), the research team initially identified ten unique behavior change theories and six process models that were considered to be an appropriate fit for the aims of the overall intervention. A unique model was developed that captured (1) the relevant underlying behavior change theory, (2) implementation strategies, and (3) unique contexts of the technology environment (Figure 2A).

Figure 2. Underpinning behavior change theory and approach outcome of ChatGPT vs human-generated output. ACTS: Accelerated Creation to Sustainment; BJ: Brian Jeffrey; CBT: cognitive behavioral therapy; COM-B: capability, opportunity, and motivation model of behavior change; dDPP: digital diabetes prevention program; HBM: health belief model; IDEAS: Integrate, Design, Assess, and Share; J&J: Johnson and Johnson; MI: motivational interviewing; SCT: social cognitive theory; SDT: self-determination theory; TTM: transtheoretical model.



GenAI Solution

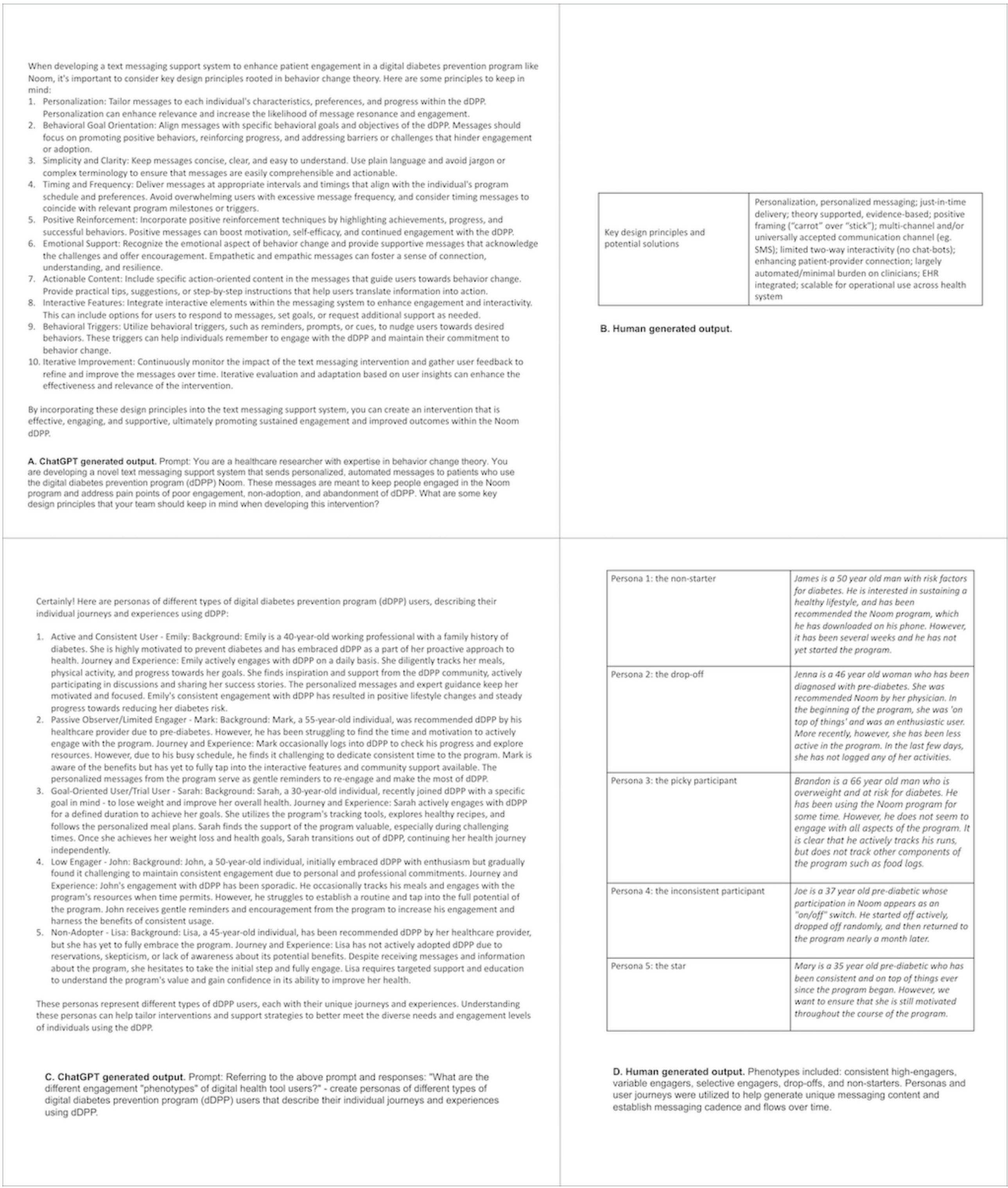
When prompted, ChatGPT identified seven relevant well-accepted behavior change theories and frameworks to inform a “dDPP support intervention” (Figure 2B). It did not provide information on the transtheoretical domains framework, or the taxonomy of behavior change techniques, but when prompted on these, identified both as acceptable strategies for use.

User Experience: Design Principles, Personas, and Messaging Content

Human-Generated Solution

The research team used a human-centered design approach to identify key design principles, defined as the set of considerations that form the basis of the PAMS product (Figure 3B). These were developed from insights gathered via a review of relevant digital behavior change research, consultation with content and theoretical experts in digital health and implementation science (n=3), and two group interviews (n=9). From these insights, five relevant fictional personas were designed to capture the various phenotypes of user engagement with the commercial dDPP, along with unique user journeys developed to describe their projected engagement with the program over time (Figure 3D). Overall, over 193 unique messages were developed, each grounded by a relevant behavior change technique and tailored to an individual phenotype’s user journey. These elements were continuously revisited and refined during the testing phases of the dDPP research. This included a 6-month near-live user testing phase consisting of nine patients engaging with various iterations of the PAMS prototype, and a 12-month live single-arm pilot phase consisting of 25 patients using PAMS-beta with the commercial dDPP platform.

Figure 3. User experience: design principles, personas, and messaging content outcome of ChatGPT vs human-generated output. dDPP: digital diabetes prevention program; EHR: electronic health record.



When developing a text messaging support system to enhance patient engagement in a digital diabetes prevention program like Noom, it's important to consider key design principles rooted in behavior change theory. Here are some principles to keep in mind:

1. Personalization: Tailor messages to each individual's characteristics, preferences, and progress within the dDPP. Personalization can enhance relevance and increase the likelihood of message resonance and engagement.
2. Behavioral Goal Orientation: Align messages with specific behavioral goals and objectives of the dDPP. Messages should focus on promoting positive behaviors, reinforcing progress, and addressing barriers or challenges that hinder engagement or adoption.
3. Simplicity and Clarity: Keep messages concise, clear, and easy to understand. Use plain language and avoid jargon or complex terminology to ensure that messages are easily comprehensible and actionable.
4. Timing and Frequency: Deliver messages at appropriate intervals and timings that align with the individual's program schedule and preferences. Avoid overwhelming users with excessive message frequency, and consider timing messages to coincide with relevant program milestones or triggers.
5. Positive Reinforcement: Incorporate positive reinforcement techniques by highlighting achievements, progress, and successful behaviors. Positive messages can boost motivation, self-efficacy, and continued engagement with the dDPP.
6. Emotional Support: Recognize the emotional aspect of behavior change and provide supportive messages that acknowledge the challenges and offer encouragement. Empathetic and empathic messages can foster a sense of connection, understanding, and resilience.
7. Actionable Content: Include specific action-oriented content in the messages that guide users towards behavior change. Provide practical tips, suggestions, or step-by-step instructions that help users translate information into action.
8. Interactive Features: Integrate interactive elements within the messaging system to enhance engagement and interactivity. This can include options for users to respond to messages, set goals, or request additional support as needed.
9. Behavioral Triggers: Utilize behavioral triggers, such as reminders, prompts, or cues, to nudge users towards desired behaviors. These triggers can help individuals remember to engage with the dDPP and maintain their commitment to behavior change.
10. Iterative Improvement: Continuously monitor the impact of the text messaging intervention and gather user feedback to refine and improve the messages over time. Iterative evaluation and adaptation based on user insights can enhance the effectiveness and relevance of the intervention.

By incorporating these design principles into the text messaging support system, you can create an intervention that is effective, engaging, and supportive, ultimately promoting sustained engagement and improved outcomes within the Noom dDPP.

A. ChatGPT generated output. Prompt: You are a healthcare researcher with expertise in behavior change theory. You are developing a novel text messaging support system that sends personalized, automated messages to patients who use the digital diabetes prevention program (dDPP) Noom. These messages are meant to keep people engaged in the Noom program and address pain points of poor engagement, non-adoption, and abandonment of dDPP. What are some key design principles that your team should keep in mind when developing this intervention?

Key design principles and potential solutions	Personalization, personalized messaging; just-in-time delivery; theory supported, evidence-based; positive framing (“carrot” over “stick”); multi-channel and/or universally accepted communication channel (eg. SMS); limited two-way interactivity (no chat-bots); enhancing patient-provider connection; largely automated/minimal burden on clinicians; EHR integrated; scalable for operational use across health system

B. Human generated output.

Certainly! Here are personas of different types of digital diabetes prevention program (dDPP) users, describing their individual journeys and experiences using dDPP:

1. Active and Consistent User - Emily: Background: Emily is a 40-year-old working professional with a family history of diabetes. She is highly motivated to prevent diabetes and has embraced dDPP as a part of her proactive approach to health. Journey and Experience: Emily actively engages with dDPP on a daily basis. She diligently tracks her meals, physical activity, and progress towards her goals. She finds inspiration and support from the dDPP community, actively participating in discussions and sharing her success stories. The personalized messages and expert guidance keep her motivated and focused. Emily's consistent engagement with dDPP has resulted in positive lifestyle changes and steady progress towards reducing her diabetes risk.
2. Passive Observer/Limited Engager - Mark: Background: Mark, a 55-year-old individual, was recommended dDPP by his healthcare provider due to pre-diabetes. However, he has been struggling to find the time and motivation to actively engage with the program. Journey and Experience: Mark occasionally logs into dDPP to check his progress and explore resources. However, due to his busy schedule, he finds it challenging to dedicate consistent time to the program. Mark is aware of the benefits but has yet to fully tap into the interactive features and community support available. The personalized messages from the program serve as gentle reminders to re-engage and make the most of dDPP.
3. Goal-Oriented User/Trial User - Sarah: Background: Sarah, a 30-year-old individual, recently joined dDPP with a specific goal in mind - to lose weight and improve her overall health. Journey and Experience: Sarah actively engages with dDPP for a defined duration to achieve her goals. She utilizes the program's tracking tools, explores healthy recipes, and follows the personalized meal plans. Sarah finds the support of the program valuable, especially during challenging times. Once she achieves her weight loss and health goals, Sarah transitions out of dDPP, continuing her health journey independently.
4. Low Engager - John: Background: John, a 50-year-old individual, initially embraced dDPP with enthusiasm but gradually found it challenging to maintain consistent engagement due to personal and professional commitments. Journey and Experience: John's engagement with dDPP has been sporadic. He occasionally tracks his meals and engages with the program's resources when time permits. However, he struggles to establish a routine and tap into the full potential of the program. John receives gentle reminders and encouragement from the program to increase his engagement and harness the benefits of consistent usage.
5. Non-Adopter - Lisa: Background: Lisa, a 45-year-old individual, has been recommended dDPP by her healthcare provider, but she has yet to fully embrace the program. Journey and Experience: Lisa has not actively adopted dDPP due to reservations, skepticism, or lack of awareness about its potential benefits. Despite receiving messages and information about the program, she hesitates to take the initial step and fully engage. Lisa requires targeted support and education to understand the program's value and gain confidence in its ability to improve her health.

These personas represent different types of dDPP users, each with their unique journeys and experiences. Understanding these personas can help tailor interventions and support strategies to better meet the diverse needs and engagement levels of individuals using the dDPP.

C. ChatGPT generated output. Prompt: Referring to the above prompt and responses: "What are the different engagement "phenotypes" of digital health tool users?" - create personas of different types of digital diabetes prevention program (dDPP) users that describe their individual journeys and experiences using dDPP.

Persona 1: the non-starter	*James is a 50 year old man with risk factors for diabetes. He is interested in sustaining a healthy lifestyle, and has been recommended the Noom program, which he has downloaded on his phone. However, it has been several weeks and he has not yet started the program.*
Persona 2: the drop-off	*Jenna is a 46 year old woman who has been diagnosed with pre-diabetes. She was recommended Noom by her physician. In the beginning of the program, she was 'on top of things' and was an enthusiastic user. More recently, however, she has been less active in the program. In the last few days, she has not logged any of her activities.*
Persona 3: the picky participant	*Brandon is a 66 year old man who is overweight and at risk for diabetes. He has been using the Noom program for some time. However, he does not seem to engage with all aspects of the program. It is clear that he actively tracks his runs, but does not track other components of the program such as food logs.*
Persona 4: the inconsistent participant	*Joe is a 37 year old pre-diabetic whose participation in Noom appears as an "on/off" switch. He started off actively, dropped off randomly, and then returned to the program nearly a month later.*
Persona 5: the star	*Mary is a 35 year old pre-diabetic who has been consistent and on top of things ever since the program began. However, we want to ensure that she is still motivated throughout the course of the program.*

D. Human generated output. Phenotypes included: consistent high-engagers, variable engagers, selective engagers, drop-offs, and non-starters. Personas and user journeys were utilized to help generate unique messaging content and establish messaging cadence and flows over time.

GenAI Solution

ChatGPT was prompted from multiple perspectives (researcher, clinician, and patient) to identify key design principles (Figure 3A) and sample solutions for the PAMS intervention. It also provided common engagement phenotypes for digital health tool users, based on patterns of use, frequency, duration, and “other elements.” Of note, nonadopters were not identified within the initial round of phenotypes. ChatGPT also developed personas for each of the identified engagement phenotypes, including persona names, backgrounds, and individual journeys. ChatGPT was able to produce five to ten unique messages targeted toward each phenotype and to adapt these messages based on various additional prompts. The user types or personas generated by ChatGPT are consistent with the human-generated users and cover all the phenotypes identified in our previous research (eg, mapping to a specific behavior change technique and reflecting a key design principle; Figure 3C).

PAMS Technical Development

Overview

The technical development includes a PAMS requirements document and architectural design and code.

Technical Requirements (User Stories)

Human-Generated Solution

Following the data collection and intervention design period, we created, as a team, a series of user stories (Figure 4B) which were followed by system requirements to describe the intended use cases, features, and challenges of the proposed PAMS software. Initial system requirements represent the "minimum viable product" that was developed, piloted, and further refined (Figure 4D). Our development team followed software engineering principles to generate the requirements document.

Figure 4. User stories and technical requirements outcome of ChatGPT vs human-generated outputs. API: application programming interface; CDC: Centers for Disease Control and Prevention; EHR: electronic health record; PAMS: personalized automatic messaging system; REDCap: Research Electronic Data Capture; UX: user interface.

1. As a patient, I want to be enrolled in the Noom app and provided with a Fitbit and a smart scale so that I can participate in the randomized control trial and track my progress towards preventing type 2 diabetes.
2. As a patient, I want my data to be collected from the Noom web service endpoint on a weekly basis so that my engagement with the Noom program can be evaluated.
3. As a patient, I want to receive a fail/pass score every week based on my adherence to the CDC gold standard guidelines for type 2 diabetes prevention so that I can track my progress and adjust my behavior accordingly.
4. As a patient, I want to receive a personalized behavioral science SMS every three weeks based on my adherence trends to the Noom program so that I can receive suggestions for improving my engagement and preventing the onset of type 2 diabetes.
5. As a healthcare provider, I want to be able to access reports and analytics based on patient data so that I can monitor patient engagement and adherence to the Noom program.
6. As a researcher, I want to be able to access patient data to analyze the effectiveness of the program and draw conclusions about its efficacy in preventing type 2 diabetes.
7. As a system administrator, I want to ensure that patient data is stored securely in a database and is compliant with data protection regulations to maintain patient privacy and confidentiality.

A. ChatGPT generated output. Prompt: You are a software engineer working in a medical school and designing a system to support a randomized control trial for pre-diabetic patients aimed to increase patient adherence to the digital prescription, NOOM. In the trial, patients are consented and enrolled in the Noom app. Additionally, they are provided with a Fitbit and an smart scale to support data collection. You need you to design a system that weekly collects patient data from a Noom Web service endpoint and runs an internal algorithm based on gold standard use to determine patient engagement based on patient interaction with Noom. Every week each patient will be assigned a fail/pass score based on if they meet the CDC gold standard guidelines for type 2 diabetes prevention. These scores will be aggregated every three weeks to determine adherence trends. Based on these trends, patients will receive a behavioral science SMS on their provider's behalf to promote engagement with the overall goal to increase adherence to the Noom program. Please produce user stories for this system so we can start a conversation with a team of software developers

Prompt for refined version below: Please refine these user stories considering that the roles are patient, provider, and research team. The role of the research team is to keep monitoring the progress if the project and get access to data such as messages going out and in , and patients data to update decisions.

As a dDPP user/patient…
I want easy ways to access Noom
I want Noom to utilize my existing technology preferences (eg. smartphone, tablet, email)
I want to be motivated to use Noom regularly
I want to know how I am doing in Noom/get regular feedback on how I'm doing in Noom
I want to be able to connect with others using Noom/compare myself to others
I want my doctor to know how I am doing in Noom
I want to see the ongoing benefits of Noom to my health
I want tailored support while I use Noom
I want to be able to connect my existing health devices to Noom
I do not want Noom to make me feel bad about myself
I want support that is positive and encouraging

As a primary care doctor…
I want to know what digital health tools my patients are using
I want to have easy access to Noom patient data that does not interfere with my current workflows
I want my patients Noom data to be integrated into my existing EHR data
I want to be able to make informed decisions based on my patients Noom data
I want to be able to educate my patients on their health based on Noom
I do not want Noom to negatively impact my workflows

As a dDPP researcher…
I want to be able to identify eligible patients
I want to be able to contact eligible patients using existing technology (eg MyChart)
I want to be able to track our patients Noom use
I want to be able to have detailed patient engagement data from NOOM
I want to remind patients to fill out necessary surveys and onboarding.
I want to know (if) the participants have filled out the relevant study surveys, so I can nudge them if they haven't

B. Human generated output.

Functional Requirements

Retrieve Data Points
- The system shall fetch the following data points from Noom's API: Physical Activity data, including exercise time and steps. Weigh-ins data. Articles read data. Social Activity data, including messages sent to coaches and group posts. The system shall ensure that the fetched data is properly structured and accessible for further processing.

Calculate Compound Metrics
- The system shall combine the "patient daily exercise time" and "daily steps taken" data points into a single metric representing Physical Activity.
- The system shall combine the "messages sent to Noom coaches" and "group posts to Noom's user boards" data points into a single metric representing Social Activity.

Determine Weekly Adherence
- The system shall calculate adherence for each week based on the following thresholds:
- For Physical Activity: Check if the total exercise time for the week is at least 1500 minutes or the total steps taken meet the recommended goal. For weigh-ins: Check if there is at least one weigh-in recorded for the week. For articles read: Check if there is at least one article read during the week. For Social Activity: Check if there is at least one message sent to a coach or one group post made during the week.

Determine Three-Week Adherence
- The system shall analyze the weekly adherence data over the past three weeks to evaluate the overall engagement.
- The system shall calculate the percentage of weeks in which the patient met the adherence criteria for each behavior metric.
- The system shall determine if the patient has met the engagement criteria for each behavior metric over the three-week period.

Generate Nudges
- Based on the calculated adherence for the past three weeks, the system shall categorize the patient's engagement level as "Passing" or "Not Passing."
- The system shall prepare nudge messages tailored to the patient's engagement level over the past three weeks. The nudge messages shall include personalized nudges related to each behavior metric based on the patient's performance.

Schedule Nudge Delivery
- The system shall determine the timing for sending the nudges, following the requirement of sending messages once every three weeks.
- The system shall set up a scheduled task or job to send the nudges to the patient's mobile phone number via SMS.

Monitoring and Evaluation
- The system shall continuously monitor and evaluate the effectiveness of the nudges in promoting patient engagement.
- The system shall collect and analyze data on patient responses, engagement levels, and behavior change outcomes. Based on user feedback and observed outcomes, the system shall iterate on the algorithm and nudge content to improve effectiveness.

Non-Functional Requirements

Performance
- The system shall fetch and process data from Noom's API in a timely manner to ensure real-time nudges generation.
- The nudges delivery task or job shall be scheduled and executed efficiently without delays.

Reliability
- The system shall be robust and resilient, handling potential errors or exceptions during data retrieval, processing, and nudges delivery.
- The nudges delivery mechanism shall have fail-safe measures to ensure message delivery even in case

C. ChatGPT generated output. Prompt: Thanks for this. As a software engineer, give me a well-structured requirements document using the user stories you have already generated in the previous iteration.

The system should be able to connect to the Noom API for seamless data transfer.
The system shall be able to store Noom data into internal data bases
The system should be able to calculate an weekly adherence score using Individual noom patient's data.
The system shall be able to calculate "trends" using 3 weeks of data.
The system shall be able to classify patients using calculated trends.
The system should be able to comment to the twilio API for messaging functionality
The system should be able to send messages to patients based on the adherence score
The system shall be able to automatically trigger the message functionality
The system shall be able to send messages on patient's individual dates such as (3 weeks into the study, birthday, etc)
The system shall be able to send reminders for messages to fill surveys and discuss dDPP study during doctors' visits.
The system shall be able to reply to patient messages
The system shall be able to collect and store message replies in internal databases.
The system should be able to connect to the REDCAP API for secure data transfer and storage
The system shall be able to connect with the EHR (EPIC)
The system shall be able to display the provider UX in EPIC for seamless integration.
The system shall be able to provide an interactive UX for providers to interact with the system.
The UX shall show Noom data, and all doctors requested information

D. Human generated output. Main requirements involve the capacity to query, store, and safely manipulate patients' data; calculate engagement levels; categorize patients' according to engagement level; read, store, and manipulate a behavioral science-based message pool; send personalized messages based on patients' engagement, and generate a reporting dashboard for providers. Requirements include applying high-security data manipulation standards. Data consumed and generated by PAMS needs to be recorded for future evaluation and further analysis of the effect of messages in achieving engagement.

GenAI Solution

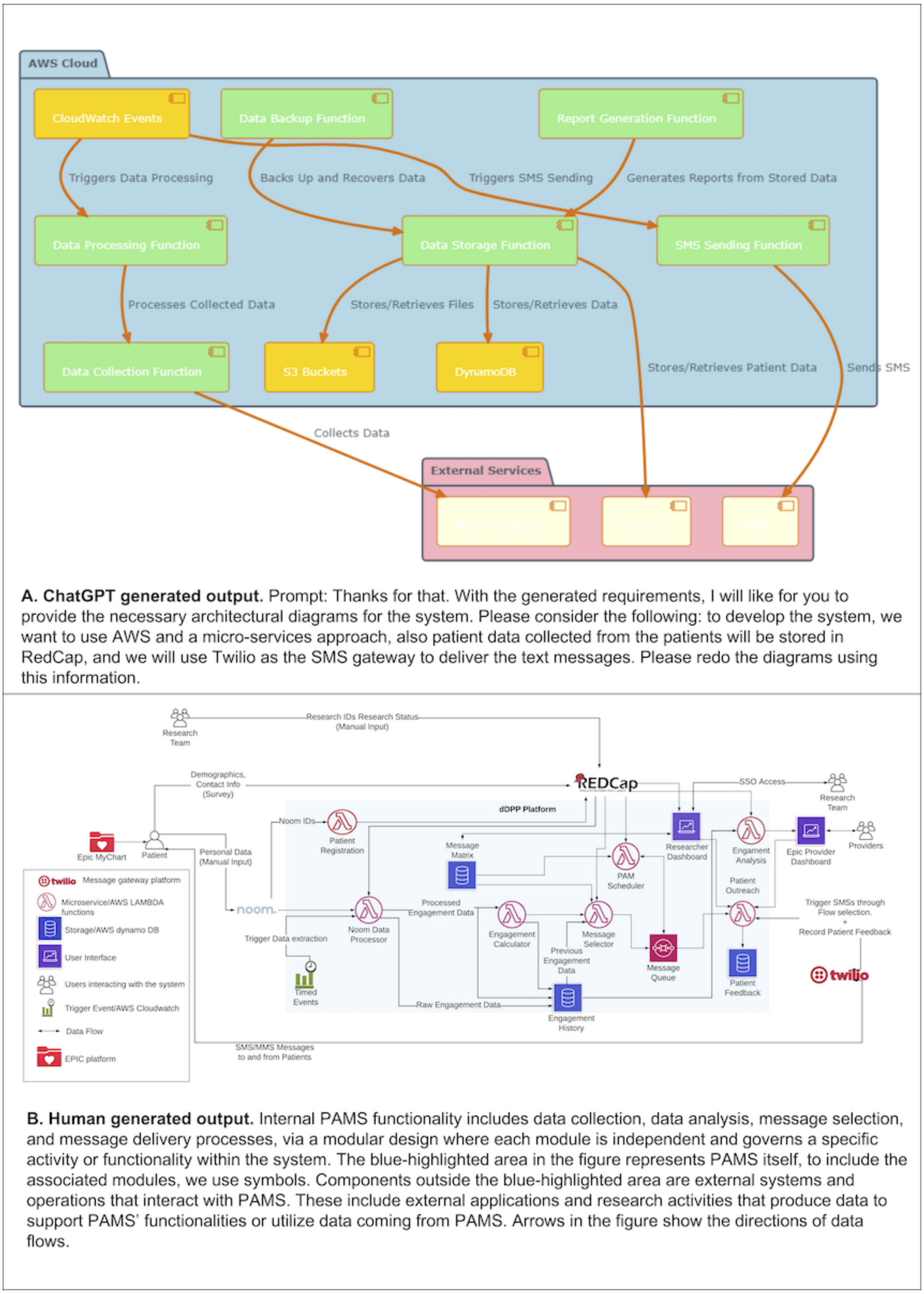
We used the output of the "feature construction phase" to inform the GenAI output for requirements. During the initial stages of the prompting phase, we refrained from suggesting solutions, allowing ChatGPT to generate potential solutions autonomously. We reviewed and evaluated these outputs, eliminating impractical or incompatible solution paths that did not align with the intentions or capabilities of our team. Once we reached a satisfactory outcome but faced uncertainty regarding the next steps, we instructed ChatGPT to assume a different "personality" (eg, software architect) and used the previous outputs as a foundation for the new role's initial prompts. Throughout this process, we encouraged each "personality" to seek clarifications by asking questions and provided feedback without biasing toward any predetermined solution. We repeated this process at least four times for each personality type, engaging in a back-and-forth roleplay with multiple personalities (researcher, architect, and developer), transitioning to a different personality when it became evident that the current one could no longer progress without additional feedback (Figures 4A and 4C).

Architectural Design

Human-Generated Solution

After the requirement phase, our software development team developed the PAMS architectural diagram, which is a graphical representation of the system that includes (1) a set of components (eg, a database and computational modules) that will perform a function required by the system; (2) the set of connectors that will help in coordination, communication, and cooperation between the components; and (3) conditions for how components can be integrated to form the system (Figure 5B).

Figure 5. Architectural design for PAMS. ChatGPT vs human-generated output. AWS: Amazon Web Services; dDPP: digital diabetes prevention program; PAMS: personalized automatic messaging system; REDCap: Research Electronic Data Capture.



A. ChatGPT generated output. Prompt: Thanks for that. With the generated requirements, I will like for you to provide the necessary architectural diagrams for the system. Please consider the following: to develop the system, we want to use AWS and a micro-services approach, also patient data collected from the patients will be stored in RedCap, and we will use Twilio as the SMS gateway to deliver the text messages. Please redo the diagrams using this information.

B. Human generated output. Internal PAMS functionality includes data collection, data analysis, message selection, and message delivery processes, via a modular design where each module is independent and governs a specific activity or functionality within the system. The blue-highlighted area in the figure represents PAMS itself, to include the associated modules, we use symbols. Components outside the blue-highlighted area are external systems and operations that interact with PAMS. These include external applications and research activities that produce data to support PAMS' functionalities or utilize data coming from PAMS. Arrows in the figure show the directions of data flows.

GenAI Solution

For the GenAI-generated architectural design, we leveraged the outputs of the requirement phase and the available ChatGPT plugins to designate the GenAI model as a software engineer and proceeded to develop an architectural diagram. During this process, we engaged in iterative prompting and provided explicit instructions to ChatGPT, specifying the use of Amazon Web Services (AWS) for development, integration of external systems such as Twilio (Twilio Inc) and REDCap (Research Electronic Data Capture; Vanderbilt University), and the adoption of a microservice approach to facilitate the efforts of our development team (Figure 5A).

Code

Human-Generated Solution

PAMS components include several lambda functions that execute its engagement or adherence algorithm, messaging, and data manipulation functionalities. Most of the functions are coded and developed using Python (Python Software Foundation) and Scala (École Polytechnique Fédérale Lausanne) as programming languages. AWS was used for the development of PAMS [12]. Our developers followed our microservice approach design using an event-driven model [13,14]. The main components of PAMS are AWS lambda functions which are triggered by different events such as updates to S3 buckets, modifications on DynamoDB (AWS) tables, or CloudWatch (AWS) events. External interactions of PAMS use application programming interface calls, which secure effective data transfer (Figure 6B).

Figure 6. Code for the function that calculates patient adherence and engagement trends. ChatGPT vs human-generated outputs.

A. ChatGPT generated output. Prompt: You are an experienced software developer tasked with writing code for the business logic. Write code in Scala for the function responsible for calculating engagement trends for patients every three weeks. Assume all other functionality is being worked on independently and mock calls to services and data sources as needed.

B. Human generated output. The code was generated by our developer following traditional software engineering principles and standardized code development styles.

GenAI Solution

To facilitate the generation of the coded solution using ChatGPT, we assigned the role of a software engineer to the model and specifically requested it to generate Scala code for a specific functionality, namely the "calculate engagement trends" function. Consistent with the iterative nature of the GenAI-based software development process, we engaged in a back-and-forth interaction with ChatGPT, iterating over the prompt and its output while providing expert guidance to ensure optimal results. While allowing ChatGPT to generate free text, we evaluated each output for accuracy and adherence to the desired specifications (Figure 6A).

Internal Review of Human Vs GenAI Outputs

The 11 evaluators participated in the output review process. All had familiarity or prior exposure to the original PAMS intervention. Overall, evaluators rated the ChatGPT-produced outputs as positive for the theoretical background and design phase in terms of understandability, usability, novelty, relevance, and efficiency. For these two components, the question about completeness showed the most variability with divided opinion among "agree" and "disagree" and the bias was mostly categorized as "neither agree nor disagree." For the first part of the technical development (user stories and requirement documents), most of the raters found the ChatGPT output positive in terms of understandability, usability, and relevance. In terms of completeness and novelty, requirements were better rated than the user stories which represent an interesting output since requirements are derived from the user stories. We hypothesize that our raters were expecting better user stories, but once these were defined, they considered ChatGPT to be effective at turning these into the requirements. In terms of bias, similar to the theoretical background and design phase, the most popular answer was "neither agree nor disagree." For the more technical pieces of the development that required software engineering knowledge, specifically the architectural diagram and code elements, results showed the highest N/A responses. These higher levels of N/As were associated with lower levels of expertise (eg, coding experience) since only 2 of the 11 evaluators had computer science backgrounds. However, the overall score excluding the N/As was positive for the technical component.

Discussion

Results Summary

This study leveraged ChatGPT-4 to recreate content features and software development of PAMS. ChatGPT served as a usable facilitator for researchers engaging in the software development life cycle, from product conceptualization to feature identification, and user story development to code generation. GenAI technologies facilitated effective communication and understanding within our multidisciplinary team by providing well-described features and supporting the role of a software engineer. Our findings indicate that the ChatGPT-generated output is comprehensive, albeit with occasional ambiguities that required clarification or adjustment by the research team. The ChatGPT-generated output exhibited a high level of accuracy in capturing the intended requirements. We found that ChatGPT supported a highly efficient development process, producing over 5 days what initially required more than 200 human hours from content and technical experts. The results suggested that by efficiently prompting ChatGPT and leveraging the expertise of our team, we could have significantly reduced the time we invested in initial system modeling and conceptualization phases as well as technical phases of software development (coding). Overall, GenAI technologies like ChatGPT offer a promising approach to efficient software development.

While promising, some significant limitations to ChatGPT's outputs should be noted. In the design phase, while ChatGPT was able to provide general guidance in tool design (eg, app vs web-based vs EHR solution) it was unable to provide evidence to support its rationale for these choices. This lack of reference support has been well-documented and has a material impact on researchers looking to build upon an evidence base for their health technology interventions. Similarly, when asked to provide theoretical frameworks to support behavior change, it offered only a partial list, initially excluding the COM-B (capability, opportunity, motivation, behavior) model upon which the original PAMS intervention was based, and needed prompting from our behavior change expert to provide more specific guidance. In the context of code generation, we focused on testing a specific function, namely the Calculating Patient Engagement feature, which is the core functionality of our software. Initially, we tasked ChatGPT with generating a function to compute a 3-week patient engagement trend. However, the initially generated code deviated from the intended objective and instead calculated a weekly engagement score. Through subsequent iterations, we were able to obtain the desired code. However, the initial attempts exhibited nonidiomatic constructs and contained bugs (no efficient loops and wrong logic). Finally, we observed that ChatGPT overlooked certain suggested features during the design phase, resulting in the generated code occasionally demonstrating unnecessary complexity and disregarding some of the best practices and features of the target programming language. We believe that further iterations would have improved the code quality, encompassing better adherence to coding standards and the inclusion of desired business features, such as handling edge cases and capturing more nuanced engagement trends. Nevertheless, we reached a point of diminishing returns with ChatGPT where we determined that engaging an experienced developer would have expedited the code generation process and ensured a more robust implementation.

These limitations highlight the ongoing importance of human expertise in the development process, especially in scenarios where theoretical expertise, intricate coding practices, and business-specific requirements are involved. The lack of rationale to support the generated results shows the value of having human experts on the team who can interpret the results. ChatGPT needs to be used as a support tool but not the source of truth; thus, we always trusted and relied on human experts to validate the ChatGPT-generated results before moving to the next phase. Overall, it is important to have human experts in the system development process to guide the outputs in terms of reprompting the system (support the decision-making on acceptable output) and ensuring their accuracy. Moreover, results are highly dependent on the quality of the prompts which emphasizes the role of prompt engineering. The results show that well-structured prompts (role + problem description + ask) that infuse human expertise into every iteration are key to obtaining good results (Figure 1). As part of our prompt framework described in the methodology section, results showed that detailed problem explanations, clear asks, and roleplaying are an excellent combination to guide accurate results. We suggest asking ChatGPT questions using different roles, asking for clarification if needed, and in cases of wrong outputs, redirecting the prompts.

Related Work

There is near-universal interest in understanding the impacts of GenAI and large language models (LLMs) on human social structures, including the experience of work and the production of work-related outputs in health care and more broadly [15,16]. In health care, LLMs are poised to impact everything from care delivery experience, diagnostic reasoning and cognitive skills, training and education, and the overall composition of the workforce [17]. These theoretical disruptions are tempered, however, by acknowledging that in its current state, GenAI tools remain suboptimal, with ongoing issues in accuracy, reliability, usability, cost, equity, and ethics.

In commercial spaces, ChatGPT-enabled products designed to assist with coding and software development are already being developed (eg, OpenAI Codex [OpenAI] and CodeGPT [CodeGPT]). These tools can help generate novel code, debug and analyze code issues, assist in code refactoring, and provide code documentation. As yet, however, their usefulness in terms of quality has not been extensively evaluated, and costs and other considerations may make them inaccessible to health care researchers. ChatGPT-enabled tools for front-end design (eg, integrating ChatGPT with Figma [Figma, Inc]), user testing (including synthetic user testing), and prototyping have also been created, all allowing health technology research teams with limited design resources to take advantage of tools from product and experience design to create their interventions. Overall, commercial LLMs have been demonstrated to improve worker efficiency and productivity, through "co-pilot" support services that automate low-skills tasks, organize and present information, and surface insights [18]. Brynjolfsson et al [18] found that a ChatGPT-supported tool providing conversational guidance for customer support agents increased worker productivity by almost 14%. The authors further found that these productivity benefits accrued disproportionately to less-experienced and lower-skill workers, allowing less-skilled or newer workers to experience more rapid gains; the authors posit that high-skill workers may have less to gain from artificial intelligence assistance due to tacit knowledge reinforcement rather than new knowledge or skill development. Our work suggests that both less-experienced, lower-skill workers and high-skill workers can benefit, with novices benefitting more from new knowledge (if accurate) and skill development and experts benefiting from knowledge validation and offloading of high-effort low-value tasks.

In the academic computer science literature, ChatGPT has been evaluated as a tool for collaborative software design [4], including to improve code quality refactoring, requirements elicitation, and general design solutions [5], and fix programming bugs [19]. Similar findings are reflected in our work, including the caveats of requiring human oversight. Other authors have identified important ethical issues in using GenAI solutions for software engineering, which were not considered in this study [20].

Within health care, a growing body of research has explored the feasibility of GenAI tools (mostly ChatGPT) in a variety of use cases, including answering patient questions [3,21], creating suggestions to optimize clinical decision support [22], generating a history of present illness summaries [23], and overall examination performance [24]. In general, these papers find promising signals for the accurate and acceptable use of GenAI tools, but with many current-state caveats for their optimal, safe, and scaled use. Key areas of concern include reliability (particularly around hallucinations and citation fabrication), reproducibility, and recency of data inputs. While research in this area will continue to grow, as more test cases comparing GenAI performance to that of clinical staff will be undertaken, further work is needed to create validated and generalizable outcome measures. Future work must also ensure that the variety of GenAI tools (including general commercial LLMs, health care–specific LLMs, and internally developed tools) are equally evaluated.

Limitations

There are several limitations to this study. First, no research team members have expertise in prompt generation for GenAI tools; as a result, our prompting reflects the a priori perspectives, biases, and knowledge gaps of our team, and are therefore particularly subject to issues of framing, recall, and confirmation bias that may influence the interpretation of the results. Second, our research team members, who acted as prompt engineers in this study, were highly familiar with the project and participated in the human-based design process; thus, they were aware of what deviations from human-based design to address by reprompting the system. As a result, we have introduced bias in the prompting process and results reflect higher accuracy. Third, the absence of robust tools to objectively measure the "quality" of current ChatGPT outputs poses challenges to accurately and objectively assess its performance. Furthermore, in this case, the output reviewers were not blinded to the human vs ChatGPT outputs, given the complexity of this study and the difficulty in providing enough research context to support independent blind review. Finally, broader limitations of the technology, such as potential hallucinations and concerns about behavioral changes of responses over time, deserve acknowledgment, as they could have implications for the practical applications and long-term viability of GenAI in health care research contexts. Future research efforts should address these limitations to enhance and replicate our findings.

Implications and Future Directions for Exploration

We are considering several future directions for the use of ChatGPT in our digital health intervention development. We envision increasing our expertise in prompt engineering (add expert prompt engineers to the team) to actively use ChatGPT to further develop PAMS features, particularly for additional messaging content. We anticipate this will save our research team considerable time and effort. We may also use ChatGPT to facilitate more time-consuming aspects of our research documentation, including both coding documentation and larger research archival work (eg, meeting minutes and recording intervention decision-making). Overall, we feel ChatGPT and related tools can be effectively leveraged within health care technology research teams with a spectrum of technical expertise, serving to both augment existing skills and supplement skill gaps. For those with expertise in computer science or programming, we imagine ChatGPT can assist by automating high-effort, low-impact tasks or repetitive work that is considered important but often deprioritized as more urgent tasks arise (eg, code documentation). For those without preexisting programming skills, we imagine ChatGPT can offer technical support, including educational tools and skill-building opportunities. Overall, this process will both validate existing knowledge and create new knowledge for teams, as well as potentially improve interteam communication and collaboration.

Conclusions

In this study, we explored the use of the GenAI tool ChatGPT to recreate a novel digital behavior change intervention which our research team had previously developed to support patient engagement and adherence to a commercial dDPP. Specifically, we reviewed and evaluated the capacity and limitations of ChatGPT to support digital health research intervention ideation, design, and software development, finding it a feasible and potential time- and resource-saving tool to support research teams in developing novel digital health products and technologies. At the same time, we identified gaps in ChatGPT outputs that may limit its effective use for both novel and advanced technology developers, particularly around the completeness of outputs. Future directions will include the development of more targeted artificial intelligence–based tools to support health care researchers with all levels of software or engineering skills, as well as the development of improved tools to objectively evaluate GenAI outputs.

Acknowledgments

The National Institute of Diabetes and Digestive and Kidney Diseases (1R18DK118545-01A1, principal investigator: DM) funded this research.

Conflicts of Interest

None declared.

References

1. Risling TL, Risling DE. Advancing nursing participation in user-centred design. J Res Nurs 2020;25(3):226-238. [doi: 10.1177/1744987120913590] [Medline: 34394630]
2. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. Front Artif Intell 2023;6:1169595 [FREE Full text] [doi: 10.3389/frai.2023.1169595] [Medline: 37215063]
3. Liu S, Wright AP, Patterson BL, Wanderer JP, Turer RW, Nelson SD, et al. Using AI-generated suggestions from ChatGPT to optimize clinical decision support. J Am Med Inform Assoc 2023;30(7):1237-1245 [FREE Full text] [doi: 10.1093/jamia/ocad072] [Medline: 37087108]
4. Ahmad A, Waseem M, Liang P, Fahmideh M, Aktar MS, Mikkonen T. Towards human-bot collaborative software architecting with ChatGPT. 2023 Presented at: EASE '23: Proceedings of the 27th International Conference on Evaluation and Assessment in Software Engineering; June 14-16, 2023; Oulu, Finland p. 279-285. [doi: 10.1145/3593434.3593468]
5. White J, Hays S, Fu Q, Spencer-Smith J, Schmidt DC. ChatGPT prompt patterns for improving code quality, refactoring, requirements elicitation, and software design. ArXiv Preprint posted online on March 11 2023 [FREE Full text] [doi: 10.48550/arXiv.2303.07839]
6. Rodriguez DV, Lawrence K, Luu S, Yu JL, Feldthouse DM, Gonzalez J, et al. Development of a computer-aided text message platform for user engagement with a digital diabetes prevention program: a case study. J Am Med Inform Assoc 2021;29(1):155-162 [FREE Full text] [doi: 10.1093/jamia/ocab206] [Medline: 34664647]
7. Lawrence K, Rodriguez DV, Feldthouse DM, Shelley D, Yu JL, Belli HM, et al. Effectiveness of an integrated engagement support system to facilitate patient use of digital diabetes prevention programs: protocol for a randomized controlled trial. JMIR Res Protoc 2021;10(2):e26750 [FREE Full text] [doi: 10.2196/26750] [Medline: 33560240]
8. Liu V, Chilton LB. Design guidelines for prompt engineering text-to-image generative models. 2022 Presented at: CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems; April 29-May 5, 2022; New Orleans, LA p. 1-23. [doi: 10.1145/3491102.3501825]
9. Zhou Y, Muresanu AI, Han Z, Paster K, Pitis S, Chan H, et al. Large language models are human-level prompt engineers. ArXiv Preprint posted online on November 03 2022 [FREE Full text]
10. White J, Fu Q, Hays S, Sandborn M, Olea C, Gilbert H, et al. A prompt pattern catalog to enhance prompt engineering with chatGPT. ArXiv Preprint posted online on February 21 2023 [FREE Full text] [doi: 10.48550/arXiv.2302.11382]
11. Wang J, Liu Z, Zhao L, Wu Z, Ma C, Yu S, et al. Review of large vision models and visual prompt engineering. Meta-Radiology 2023;1(3):100047 [FREE Full text] [doi: 10.1016/j.metrad.2023.100047]
12. Nadareishvili I, Mitra R, McLarty M, Amundsen M. Microservice Architecture: Aligning Principles, Practices, and Culture. Sebastopol, CA: O'Reilly Media, Inc; 2016.
13. Alshuqayran N, Ali N, Evans R. A systematic mapping study in microservice architecture. 2016 Presented at: 2016 IEEE 9th International Conference on Service-Oriented Computing and Applications (SOCA); November 4-6, 2016; Macau, China p. 44-51. [doi: 10.1109/soca.2016.15]
14. Mathew S, Varia J. Overview of Amazon Web Services. Amazon Whitepapers. 2014. URL: https://d1.awsstatic.com/whitepapers/aws-overview.pdf [accessed 2024-02-01]
15. Eloundou T, Manning S, Mishkin P, Rock D. GPTs are GPTs: an early look at the labor market impact potential of large language models. ArXiv Preprint posted online on March 17 2023 [FREE Full text] [doi: 10.48550/arXiv.2303.10130]
16. The impact of artificial intelligence on the future of workforces in the European Union and the United States of America. The White House. 2022. URL: https://www.whitehouse.gov/wp-content/uploads/2022/12/TTC-EC-CEA-AI-Report-12052022-1.pdf [accessed 2024-01-18]
17. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med 2023;29(8):1930-1940. [doi: 10.1038/s41591-023-02448-8] [Medline: 37460753]
18. Brynjolfsson E, Li D, Raymond LR. Generative AI at work. National Bureau of Economic Research. 2023. URL: https://www.nber.org/papers/w31161 [accessed 2024-01-18]
19. Surameery NMS, Shakor MY. Use Chat GPT to solve programming bugs. Int J Inf Technol Comput Eng 2023;3(1):17-22 [FREE Full text] [doi: 10.55529/ijitc.31.17.22]
20. Akbar MA, Khan AA, Liang P. Ethical aspects of ChatGPT in software engineering research. IEEE Trans Artif Intell 2023:1-14. [doi: 10.1109/tai.2023.3318183]

21. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence Chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 2023;183(6):589-596. [doi: 10.1001/jamainternmed.2023.1838] [Medline: 37115527]
22. Biswas S, Logan NS, Davies LN, Sheppard AL, Wolffsohn JS. Assessing the utility of ChatGPT as an artificial intelligence-based large language model for information to answer questions on myopia. Ophthalmic Physiol Opt 2023;43(6):1562-1570 [FREE Full text] [doi: 10.1111/opo.13207] [Medline: 37476960]
23. Nayak A, Alkaitis MS, Nayak K, Nikolov M, Weinfurt KP, Schulman K. Comparison of history of present illness summaries generated by a Chatbot and senior internal medicine residents. JAMA Intern Med 2023;183(9):1026-1027. [doi: 10.1001/jamainternmed.2023.2561] [Medline: 37459091]
24. Strong E, DiGiammarino A, Weng Y, Kumar A, Hosamani P, Hom J, et al. Chatbot vs medical student performance on free-response clinical reasoning examinations. JAMA Intern Med 2023;183(9):1028-1030. [doi: 10.1001/jamainternmed.2023.2909] [Medline: 37459090]

Abbreviations

AWS: Amazon Web Services
COM-B: capability, opportunity, motivation, behavior
dDPP: digital diabetes prevention program
GenAI: generative artificial intelligence
LLM: large language model
N/A: not applicable
PAMS: personalized automatic messaging system
REDCap: Research Electronic Data Capture

Edited by A Kushniruk; submitted 18.09.23; peer-reviewed by S Mulvaney, I Oakley-Girvan; comments to author 20.11.23; revised version received 27.11.23; accepted 15.12.23; published 06.03.24.

Please cite as:
Rodriguez DV, Lawrence K, Gonzalez J, Brandfield-Harvey B, Xu L, Tasneem S, Levine DL, Mann D
Leveraging Generative AI Tools to Support the Development of Digital Solutions in Health Care Research: Case Study
JMIR Hum Factors 2024;11:e52885
URL: https://humanfactors.jmir.org/2024/1/e52885
doi:10.2196/52885
PMID:38446539

Original Paper

Testing Two Online Symptom Checkers With Vulnerable Groups: Usability Study to Improve Cognitive Accessibility of eHealth Services

Kaisa Savolainen[1], DA (Art and Des), MSc (tech); Sari Kujala[1], PhD

[1]Department of Computer Science, Aalto University School of Science, Espoo, Finland

Corresponding Author:
Kaisa Savolainen, DA (Art and Des), MSc (tech)
Department of Computer Science
Aalto University School of Science
Konemiehentie 2
Espoo, 02150
Finland
Phone: 358 503101457
Email: kaisa.savolainen@aalto.fi

Abstract

Background: The popularity of eHealth services has surged significantly, underscoring the importance of ensuring their usability and accessibility for users with diverse needs, characteristics, and capabilities. These services can pose cognitive demands, especially for individuals who are unwell, fatigued, or experiencing distress. Additionally, numerous potentially vulnerable groups, including older adults, are susceptible to digital exclusion and may encounter cognitive limitations related to perception, attention, memory, and language comprehension. Regrettably, many studies overlook the preferences and needs of user groups likely to encounter challenges associated with these cognitive aspects.

Objective: This study primarily aims to gain a deeper understanding of cognitive accessibility in the practical context of eHealth services. Additionally, we aimed to identify the specific challenges that vulnerable groups encounter when using eHealth services and determine key considerations for testing these services with such groups.

Methods: As a case study of eHealth services, we conducted qualitative usability testing on 2 online symptom checkers used in Finnish public primary care. A total of 13 participants from 3 distinct groups participated in the study: older adults, individuals with mild intellectual disabilities, and nonnative Finnish speakers. The primary research methods used were the thinking-aloud method, questionnaires, and semistructured interviews.

Results: We found that potentially vulnerable groups encountered numerous issues with the tested services, with similar problems observed across all 3 groups. Specifically, clarity and the use of terminology posed significant challenges. The services overwhelmed users with excessive information and choices, while the terminology consisted of numerous complex medical terms that were difficult to understand. When conducting tests with vulnerable groups, it is crucial to carefully plan the sessions to avoid being overly lengthy, as these users often require more time to complete tasks. Additionally, testing with vulnerable groups proved to be quite efficient, with results likely to benefit a wider audience as well.

Conclusions: Based on the findings of this study, it is evident that older adults, individuals with mild intellectual disability, and nonnative speakers may encounter cognitive challenges when using eHealth services, which can impede or slow down their use and make the services more difficult to navigate. In the worst-case scenario, these challenges may lead to errors in using the services. We recommend expanding the scope of testing to include a broader range of eHealth services with vulnerable groups, incorporating users with diverse characteristics and capabilities who are likely to encounter difficulties in cognitive accessibility.

(JMIR Hum Factors 2024;11:e45275) doi:10.2196/45275

KEYWORDS

eHealth; online symptom checkers; usability; cognitive accessibility; web accessibility; qualitative research

Introduction

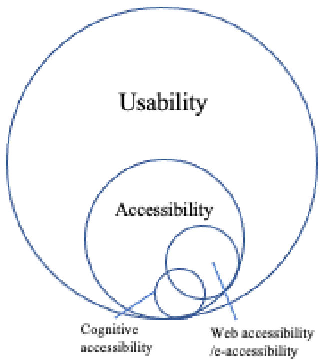
Background

Given the widespread use and popularity of eHealth services, there is a growing need for more accessible services to all potential user groups [1]. In recent years, more emphasis has been placed on accessibility and inclusion; for example, the European Union Accessibility Act has been incorporated into and enforced as national law since June 2022 [2]. As health care services are often public services, it is important that they serve a broad range of users. Furthermore, usability has been recognized as a key component of eHealth applications, and users may face problems with using the applications due to their health conditions [3]. In addition, patients with chronic illness have been reported to encounter more cognitive challenges [4]. Thus, extra attention should be paid to the usability of eHealth applications.

Universal design and design for all address these requirements by aiming at designing services that are usable by and accessible to all user groups regardless of their age, abilities, or possible disabilities [5]. Usability is a high-level term that indicates how a system can be used by specified users in a certain context of use to achieve specific goals with regard to effectiveness, efficiency, and satisfaction [6]. Accessibility, which is a part of usability, describes how a system can be used by people with the widest range of needs, characteristics, and capabilities [6,7]. Thus, accessibility covers all sorts of users with different limitations. A concept that has been addressed by several research papers [8,9] is web accessibility (or e-accessibility), which refers to the accessibility of web services.

In this paper, we address cognitive accessibility, which refers to accessibility beyond physical and sensory capabilities, and thus takes into account varied human characteristics such as intellectual disabilities, attention difficulties, reading problems, autism spectrum disorders, and low language skills [10]. Cognitive accessibility is an important aspect of web accessibility as it involves a large number of users and has a high impact on usability [10]. A summary of the relationship between these concepts is presented in Figure 1.

Figure 1. The relation of cognitive accessibility to usability and accessibility. Note that the sizes and positions of the circles are indicative.



This research focuses on cognitive accessibility within the context of the 2 most frequently used online symptom checkers in Finnish public primary care across numerous municipalities in Finland. Online symptom checkers are used by people seeking health-related guidance, and these services typically provide an urgent assessment and suggest guidance based on the symptoms reported by the user [11]. Patients can use the 2 examined symptom checkers to book appointment times for doctors and laboratory tests or obtain medical help for the most common health issues. First, patients report their symptoms and submit them to the health care center through the symptom checker. Health care professionals receive patient inquiries with an urgency rating, decide on actions to be taken, and inform patients.

Patients are generally highly satisfied with symptom checkers, but younger and more highly educated people have been more likely to use them [11]. For example, symptom checkers enable patients to access health care anytime and anywhere. Therefore, it is essential to ensure that all user groups, including individuals in vulnerable situations, can use these services effectively. Symptom checkers can also empower users as a means of facilitating their health care [12]. However, the accuracy of the symptom checkers depends on how well patients are able to communicate their symptoms when using the tools [13]. As these services spread and are used by a wider range of individuals, it is crucial to also evaluate their usability and accessibility with a more diverse set of users.

Prior Work

Vulnerable Groups

Many public eHealth services and their poor usability and accessibility can cause challenges for certain user groups [14]. These user groups are, thus, in a potentially vulnerable situation in using the service and at risk of digital exclusion [1]. This is especially problematic because research has shown that digital exclusion can cause social exclusion [15]. Public health services must, thus, address the needs of potentially vulnerable groups, including people who are disadvantaged by health, economic, cultural, or social conditions [16], such as older adults, migrants, mental health service users, and the unemployed [16,17].

Older adults are the largest group to face challenges in using digital health services [18,19]. As people age, their cognitive abilities may weaken, with cognitive load being identified as the most significant accessibility barrier for older adults [20]. Memory changes can also affect learning, information processing, and language comprehension [21,22]. Additionally, older adults often struggle with focusing their attention, particularly when multitasking [21,22]. Moreover, older age groups tend to use eHealth services less frequently than younger demographics. A Finnish study examining an online symptom checker (referred to as service A in this study) observed that individuals aged 20-39 years used the service more actively compared with older age groups, relative to their representation in the population [23]. This suggests that enhancing service usage entails prioritizing usability and accessibility from the perspective of older users as well.

Migrants represent a growing demographic that often faces challenges when accessing health services in their new country of residence [1]. Language barriers and a lack of digital skills are common issues encountered by this group [1]. Additionally, individuals with intellectual disability are another vulnerable population impacted by the digitalization of health services [24]. They have been noted to experience more difficulties in finding information on the internet and understanding online information compared with the general population [25].

Previous research suggests that vulnerable groups, such as older adults and individuals with mild intellectual disabilities, encounter cognitive challenges when using technology [26,27]. Therefore, the development of more accessible eHealth services would enable these groups to access health information more easily [25,28], thereby enhancing their sense of empowerment concerning their health issues.

The preferences or needs of older adults or individuals with mild intellectual disabilities are often overlooked in the majority of eHealth studies [29,30]. It is imperative to better consider these user groups during the design of eHealth services [17,28]. Many eHealth applications could greatly benefit from the application of universal design principles [29], which facilitate understanding the needs of potentially vulnerable groups and inform the design of more inclusive and usable services [31,32]. Consequently, this enables vulnerable groups to derive as much benefit from eHealth systems as the rest of the population [33,34]. Indeed, universal access approaches can offer benefits to anyone [35]. Therefore, to gain a better understanding of the challenges faced by vulnerable groups when using services, it is essential to conduct testing with a diverse group of users.

Usability Testing of Symptom Checkers

The usability of symptom checkers has been examined in prior research; however, there has been limited emphasis on potentially vulnerable user groups, such as older adults, migrants, and those with intellectual disability [36-38]. Moreover, research on usability in the eHealth domain frequently concentrates on quantitative aspects (eg, the number of errors, task completion times, and usability questionnaires) and typically involves a large number of users [12,36,38]. However, the qualitative aspect of usability studies is also crucial for gaining a deeper understanding of the thoughts and reasons behind errors, as well as capturing the patient's perspective at a broader level [3,39]. Additionally, while a System Usability Scale (SUS) questionnaire provides a numeric score for experienced usability, it alone is not adequate for evaluating usability. Instead, it should be complemented with other measures, such as task completion rates or more qualitative approaches, to ascertain which aspects of a service require improvement and how best to address them [39].

Marco-Ruiz et al [13] conducted research on symptom checkers and emphasized the significance of testing with real users to comprehend the cognitive processes involved when using a new system to record health data. Furthermore, they noted that the user base accessing symptom checkers is highly diverse, with some individuals possessing higher health literacy and experience in recording online information, while others may have very limited or no experience [13].

Goal of the Study

The goal of our study is to gain a deeper understanding of cognitive accessibility in the context of eHealth services. Therefore, our paper focuses on addressing the following research questions:

- What kind of challenges do vulnerable groups face in using eHealth services?
- What needs to be considered when testing with vulnerable groups?

The structure of this paper is as follows: In the next section, we describe the methods used in this study, followed by the presentation of results. Subsequently, we discuss the findings and overarching contributions of this study, concluding with our final remarks.

Methods

Approach and Researcher Background

Our qualitative study adopts a case study approach, wherein the cognitive accessibility of eHealth services was assessed through usability testing of 2 online symptom checkers. The research team comprised 3 researchers: The first researcher, a human-computer interaction student, conducted the initial 8 tests as part of their master's thesis work. Subsequently, a second researcher, a senior researcher with expertise in human-computer interaction (who served as the thesis advisor), conducted the remaining 5 tests. Additionally, a third senior researcher with backgrounds in human-computer interaction and eHealth oversaw the entire study.

Context and Study Setting

We conducted a usability test of 2 Finnish online symptom checkers in 2 phases in Finland during the Spring and Fall of 2021. The tested services were Omaolo (DigiFinland Oy) [40] and Klinik Access (Klinik Healthcare Solutions Oy) [41], which are the 2 most-used symptom checkers in Finnish public primary care. Omaolo has been actively used since 2019, while Klinik Access, which is also used internationally, has been in use since 2015. Both services are designed to assist patients in obtaining appropriate care. Users answer a set of questions regarding their symptoms, following which the symptom checkers use artificial intelligence to assess the urgency of care. If necessary, the services guide patients to contact emergency care services.

The Omaolo symptom checker comprises 15 specialized symptom checkers tailored for different types of symptoms, along with a generic symptom checker. Each symptom checker prompts the user with a specific set of questions and subsequently recommends the next steps they should take. Additionally, if the user provides their home municipality, the service displays recommended actions specific to the area, offers contact details, and may even facilitate direct contact with health care professionals if deemed necessary. The Omaolo symptom checker served as the primary COVID-19 symptom checker in Finland, enabling users to schedule appointments for COVID-19 tests. Consequently, its user base experienced a significant surge [23].

The Klinik Access symptom checker enables users to initially select the part of the body where their main symptoms are located. Subsequently, it prompts for more specific symptoms. The responses can then be forwarded to the medical staff responsible for the patient's care before their appointment, ensuring the patient is directed to the appropriate type of health care professional. The primary distinction between these services lies in their user interface (UI): Klinik Access features a more visual UI with a list of clickable symptoms, whereas Omaolo presents users with multiple-choice questions describing the symptoms. Henceforth, the Omaolo service will be denoted as service A, and Klinik Access will be referred to as service B. It is important to note that both services are classified as medical devices and must adhere to specific safety requirements, such as repetitive questions, which may impact usability.

Sampling Strategy

Purposive sampling [42] was used to recruit participants, who were sourced through personal contacts and various associations representing the targeted user groups. These associations included initiatives such as the Selkeästi meille, which focuses on enhancing cognitive accessibility, and Väylä ry, which is dedicated to improving the employment opportunities of individuals with intellectual disability. It is important to note that the test facilitator did not have a close personal relationship with the participants, such as being a friend or family member, during any of the test sessions.

A total of 13 participants were recruited to partake in the study. Notably, an evaluation of sample sizes within the field of human-computer interaction has indicated that 12 is the most common sample size for usability studies [43].

Ethical Approval

The study received approval from the ethical review board of Aalto University (D/902/03.04/2021). Each participant provided informed consent by signing a consent form after confirming their understanding of the study's purpose and how their information would be handled. Reporting has been conducted in such a manner that individual participants cannot be identified.

Data Collection Methods

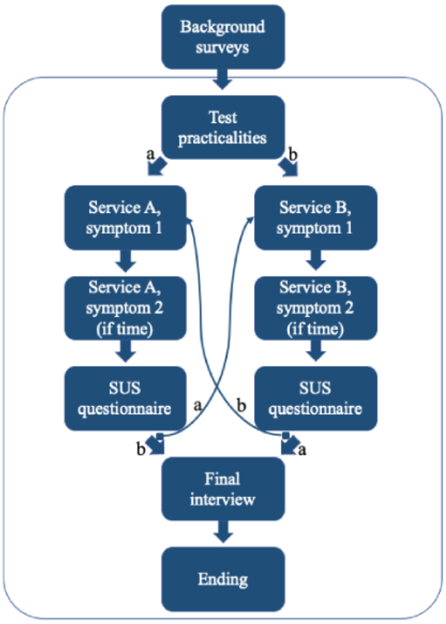
Overview

The main methods used in this study were thinking aloud, observations, questionnaires, and semistructured interviews. Before the actual tests, a pilot test was conducted to identify any potential inconsistencies and to ensure that the questions and instructions were comprehensible. Minor adjustments to the test setup were made based on the findings from the pilot test.

Test Procedure

An overview of the test sessions is presented in Figure 2.

Figure 2. An overview of the usability test sessions with older adults, mildly intellectually disabled individuals, and nonnative speakers (N=13). Half of the participants started with Service A and the other half with Service B.



Each participant tested both services, and the order of service usage was counterbalanced. During the testing phase, participants were presented with 2 symptom vignettes, each providing a brief description of the symptoms they were instructed to imagine having. These vignettes were used 1 at a time. Participants were then asked to open the service and imagine they had the symptoms described in the first vignette, aiming to determine how they should proceed. The vignettes and mode of distribution between the participants are presented in Multimedia Appendix 1.

After using the first service, participants were instructed to take the second vignette and attempt to use the service again. However, if the first part of the test had exceeded 40 minutes, the second vignette was omitted for the first service to prevent the overall test time from exceeding 90 minutes. Following their interaction with each service, participants were asked to evaluate the respective service.

After testing the first service, participants were instructed to open the second service and follow the same procedure. Upon completion of both testing phases, participants were asked to compare the 2 systems and select the one they preferred.

Data Collection Instruments

Test Sessions

The test sessions were conducted via the Microsoft Teams videoconferencing platform, which facilitates screen sharing, screen recording, and voice recording functionalities. The decision to conduct remote testing was primarily influenced by the COVID-19 pandemic situation, but it also aligned well with the nature of the tests, as the services being evaluated were online. Participants used their personal computers to access the services during the testing sessions.

Symptom Vignettes

To streamline the usability test and eliminate the necessity for participants to input their personal medical information into the services, each participant was provided with 2 standardized clinical vignettes featuring predefined symptoms. These vignettes were selected from a list compiled by Semigran et al [44], encompassing a total of 6 conditions with varying severity levels. The selection included conditions with different severity levels to account for the fact that individuals may use symptom checkers in both urgent and nonurgent situations [45].

In line with the recommendations provided by Semigran et al [44], the selected vignettes encompassed 3 categories of triage urgency: conditions necessitating emergency care, conditions warranting nonemergency care, and conditions deemed unnecessary for medical visits, thus manageable with self-care. Moreover, we opted for conditions commonly observed within the age group under study to ensure relevance. These conditions encompass ailments such as acute bronchitis, back pain, and meningitis. To ensure clarity and relevance to the participants, the selected conditions were translated from English to Finnish and simplified. The English versions of the vignettes used can be found in Multimedia Appendix 2.

Background Questionnaires

Before the actual test session, participants were requested to complete a brief background survey and the health literacy survey HLS-EU-Q16 [46]. The background information collected were the participant's gender; age; the frequency of doctor visits in the preceding 2 years; the number of doctor-diagnosed medical conditions; their previous usage frequency of digital health care services; and their frequency of digital device usage, such as smartphones or computers. These questions aimed to ascertain whether participants met the study's target demographic criteria in terms of age and their ability to independently use electronic devices such as computers. The health literacy survey provided insights into participants' understanding of health-related topics.

Interview and Questionnaire

After interacting with each service, participants were asked to evaluate the tested services. This involved administering an SUS questionnaire [47] to gauge the perceived usability of the system, as well as posing 4 interview questions:

- Would you use the service again in the future?
- Were the summary and the instructions about what to do next clear enough?
- Would you actually follow the instructions given?
- Given the option, would you use the service using your phone?

Data Processing and Analysis

The test sessions were recorded using Microsoft Teams. The voice recordings of the initial 8 tests were transcribed in full, while for the remaining 5 tests, notes were taken from the recordings, and user comments were documented to streamline the process. An experienced researcher could identify the issues encountered by users as well as their comments without requiring a complete transcription. The notes and transcriptions underwent anonymization. Qualitative content analysis was used in this study. Using the notes and recordings, all usability issues were identified and compiled. This encompassed problems mentioned by participants as well as those observed during testing or evident from the recordings. The identified usability problems were coded and categorized based on their similarities. When new problems were identified, they were compared with existing ones, and if deemed similar, they were grouped under the same code. Eventually, these groups were consolidated under higher-level descriptive categories. Furthermore, user comments were collected to bolster the analysis and reporting process.

The background questionnaires were analyzed by aggregating the responses to obtain an overview of participant characteristics. Additionally, the health literacy surveys were analyzed according to the guidelines [46] to determine the groups to which participants belonged. The SUS questionnaires were analyzed by computing the SUS scores as per the guidelines [47], resulting in scores of up to 100 points, which were then compared with the general score.

To ensure the quality and trustworthiness of the study, a senior researcher (the second author) supervised the entire research process and provided support for the analysis work. Two other researchers (the first author and the master's thesis worker) conducted the actual tests and analyzed the data. Therefore, a total of 3 researchers participated in the process, ensuring that data gathering and processing proceeded appropriately.

Results

Overview

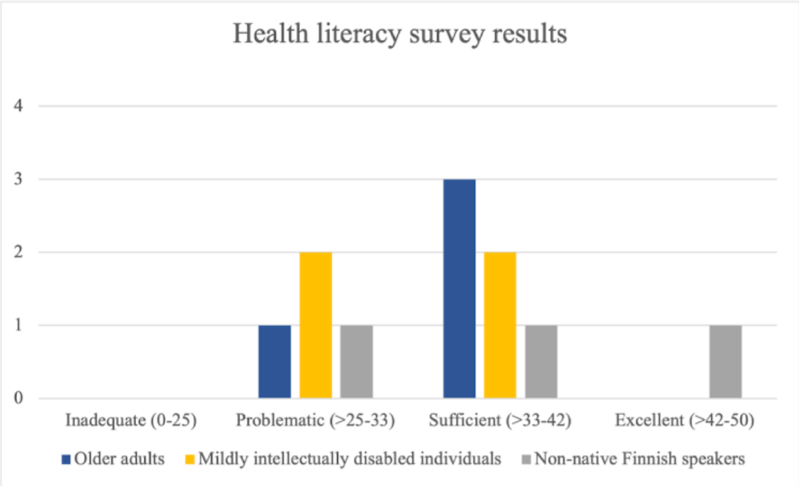
The subsequent sections present the principal findings of the study. We commence with an overview of the participants' characteristics, followed by an examination of the identified usability issues. Finally, we present additional findings that emphasize the characteristics of these user groups.

Test Participants

A total of 13 individuals participated in the study in Finland. Among them, 4 were individuals with mild intellectual disability, 4 were older adults (aged 75-79 years), and 5 were nonnative Finnish speakers. Therefore, all test users potentially encountered cognitive accessibility challenges with the services. The background characteristics of the participants are detailed in Multimedia Appendix 3.

The HLS-EU-Q16 questionnaire results were calculated in accordance with the guidelines [46], with each participant receiving a score corresponding to a quartile representing their health literacy level. The results were computed only for participants who responded to at least 80% of the questions, as recommended in the guidelines [46]. The questionnaire includes an "I don't know" answer option, which was interpreted as the question not being answered. Consequently, the results of 2 of the nonnative participants were excluded, as they chose this answering option too frequently. Figure 3 depicts the distribution of health literacy among the 3 groups.

Figure 3. The results of the HLS-EU-Q16 health literacy assessment of older people, mildly intellectually disabled individuals, and non-native speakers divided into four categories.



Usability Problems

Cognitive Accessibility Issues

The study identified a total of 65 usability problems with the 2 systems. Specifically, 36 usability problems were discovered with service A, while 29 problems were identified with service B. These issues occurred across 99 and 91 individual user instances, respectively. The problems were classified into 14 usability problem categories. A comprehensive list of the usability problem categories is provided in Multimedia Appendix 4. For the purpose of this discussion, we will focus on issues related to cognitive accessibility, primarily concerning terminology, text volume, and UI clarity.

Terminology-Related Issues

The most prevalent issues were associated with terminology and answering options. eHealth services frequently incorporate specialized language and specific terminology, posing challenges for users with cognitive limitations. Nearly all users encountered confusion with certain terms or inadvertently mixed them up with similar ones. Furthermore, lengthy words and extensive blocks of text, such as lengthy paragraphs, presented challenges, a sentiment that was also echoed during the interviews. Users with cognitive restrictions often encounter challenges when confronted with long words and extensive passages of text.

As one user commented,

> *It takes time to go through all the texts.* [ID10, nonnative]

Related issues were reported and commented on by users across all user groups. In addition to contributing to usability problems, these issues slowed down the usage of the services and occasionally led users to select incorrect symptoms.

Issues Related to the Clarity of the UI

Another area where users encountered difficulties was with the visibility of information and the lack of clarity in the UI. It is crucial for the most important information and elements of the UI to be clearly visible, facilitating easy comprehension for users. Additionally, problems arose when users' attention was diverted to unimportant features. These issues are especially pronounced among user groups with cognitive difficulties, as they require additional attention to comprehend the content and must focus more intently. Furthermore, some users found the input methods challenging; initially, they struggled to discern the type of information required for input in a field and how the inputting should be performed.

The most prevalent individual usability problems we identified regarding the logic and functionality of the UI, observed across all 3 user groups, are detailed in Textbox 1.

Textbox 1. Individual usability problems identified.

1. **Users making an incorrect selection due to an item being highlighted in the user interface:**

   Service A was, at the time of the study, the prevalent symptom checker for COVID-19 in Finland; COVID-19 was highlighted at the top of the home page of service A and was thus the first item to attract the users' attention and be selected.

2. **Difficulties in making the correct selection from a long list of items:**

   Service A had a list of 15 symptom checkers from which the user had to choose, making it difficult for the users to select the correct symptom checker to continue with.

3. **Not remembering what questions needed to be answered after the questions disappeared:**

   Service B presented questions as placeholders to describe symptoms in open answers, and these questions disappeared when the user started typing in the field; as a result, the user might not fully describe their symptoms.

4. **Being confused by long lists of apparently uncategorized symptoms:**

   Service B had long lists of symptoms as selectable buttons that seemed to be unorganized and caused anxiety and confusion.

5. **The logic and functionality of submenus were not understood by the users**:

   Service B had additional submenus and dialog boxes that were not fully understood by the users. There was a small arrow that opened the submenu and the logic of how the items were selected or the submenus opened was unclear.

6. **The users did not understand the logic of the input fields that combined several user interface items:**

   The way in which service B required the duration of symptoms to be input meant that the user needed to enter the number in one field and then select the unit from different options. However, the unit selection was not clearly related to the textbox where the user inputs the number.

These individual usability problems highlight issues with how information is presented to users, with clarity being particularly emphasized among this user group. In some instances, the selection or input options were unclear, and the services featured lengthy lists of symptoms.

Clarity was a recurring theme in several test sessions. As one user commented:

> *...if you think about this in real life, if you have a fever and you're doing this and you start to scroll all these selection choices and you're evaluating which one would fit best, the options are quite broad, so it might be quite difficult to do in practice...* [ID7, user with mild intellectual disability]

Similarly, one user suggested:

> *I don't know you could kind of put those in order like one row and another row, these are quite...your eye kind of jumps, but otherwise those are clear.* [ID2, older adult]

One user preferred the structure of service A and, again, referred to the clarity with which the information is presented:

> *Well, [I prefer Service A] because it was maybe better organized, there was one thing and one question and then one answer. After this, the next question and so on. In the other one [Service B], you had to read all the small boxes and look for your symptom. [...]* [ID8, user with mild intellectual disability]
>
> *Well, maybe what is the most [difficult], this one had so many small boxes that at least for me, it was difficult to find my own symptom, the one I needed to select from there. So, if I wanted to know what fit me, I had to read through them all and then, since they are not in any order, they just are there, I had to read them all, to see if I could find the one I have at the moment.* [ID8, user with mild intellectual disability]

It is worth noting that the symptoms were arranged in alphabetical order; however, the layout was such that users did not realize this ordering method had been used.

Differences Between the User Groups

Some differences between the user groups were evident, although the majority of the usability problems were consistent across all user groups. Nonnative Finnish speakers found the service to be particularly slow to use, often taking an extended period to read the texts. One user commented regarding service B that:

> *Reading and writing text is not easy for an immigrant. When you can click on an item it is easy, you don't have to write.* [ID13, nonnative]

The older adults did not encounter as many issues with longer texts. Instead, they faced more challenges in understanding the logic of the services and remembering to scroll down to view all the provided information. However, this scrolling also frustrated some nonnative users; as one user commented:

> *And again, we're scrolling, this is terrible!* [ID12, nonnative]

The task completion times were also measured and presented for the initial tasks of both services. As depicted in Table 1, aside from the older adults, there were no significant differences in the completion times between the services. However, for the older adults, service B, which featured more clickable elements to choose from, appeared to be quicker to use. Nonnative speakers took the longest time to complete the tasks, primarily because they often needed to translate some of the terms used in the services. Three of the users used an online translator (eg, Google Translator), and at times, users asked the facilitator about specific terms. Overall, the task completion times were quite lengthy, suggesting that these user groups require ample time to use these services effectively.

Table 1. Average task completion times (first task) for both services. For older adults there was a clear difference in favor of service B; for the 2 other groups service A got a slightly better time.

Task completion times	Service A, hh:mm:ss[a]	Service B, hh:mm:ss
Older adults	0:14:30	0:08:00
User with mild intellectual disability	0:14:54	0:16:00
Nonnatives	0:15:46	0:18:16

[a]hh:mm:ss: hours:minutes:seconds.

For the few users who had the opportunity to test the services twice, the second time was generally much faster than the first, indicating good learnability. As one user mentioned:

> *Now I know that I need to select this and not the other, which I didn't know previously.* [ID8, user with mild intellectual disability]

The SUS scores are provided in Multimedia Appendix 5, illustrating how participants evaluated the usability of the services. The SUS score ranges from 0 to 100 points. It has been assessed for numerous services, and according to Bangor et al [48], a satisfactory SUS score is above 70, with superior products typically scoring 80 or higher. However, it is important to note that the interpretation of SUS scores can vary depending on the type of product and its development phase. When evaluating the SUS scores of the tested services, which are predominantly below 75, it is evident that the perceived usability was not considered very good, except for nonnative Finnish speakers, as their scores hovered around 80.

From the interviews, we found that older adults tended to prefer computers over mobile devices when using the symptom checkers, whereas nonnative speakers mostly preferred mobile devices. The preference among users with mild intellectual disability was evenly divided. Nonetheless, the advantage of this type of online symptom checker was evident, as all participants expressed willingness to use the services again. The nonnative participants particularly valued a service that enabled them to input information at their own pace, as opposed to speaking on the phone. However, their preference for the service they would use was fairly evenly split, with no clear consensus: 7 participants favored service A, while 6 participants favored service B.

Discussion

Principal Findings

Testing for cognitive accessibility with 2 symptom checkers revealed that older adults, individuals with mild intellectual disability, and nonnative speakers may encounter numerous challenges when using the services. Primarily, problems arise concerning the terminology used. This highlights the need for greater emphasis on ensuring that the vocabulary used in the health sector, while specialized, remains understandable to a broad audience when services are intended for universal use. Furthermore, complications arose from the intricate structure and layout of the services. The significance of simplifying services, minimizing lengthy lists, and using more understandable terminology was highlighted in nearly all the test sessions. Implementing these improvements to the services would likely benefit a broader range of users [5].

There were distinct differences observed among the 3 user groups. Primarily, nonnative speakers assigned notably higher usability ratings to the services compared with the other 2 groups. One possible reason for this could be their overall satisfaction with the existence of such services, which enable them to seek help for their health issues without having to converse over the phone in a language that is not their native tongue.

One notable distinction between the user groups pertained to their preference for using either a computer or a mobile device. It was evident that older adults favored using computers, likely because of their larger screens and the familiarity that older adults have with them. Conversely, most nonnative Finnish speakers showed a preference for mobile devices, with some noting that they solely rely on their mobile devices and do not even own a computer. This preference may be influenced in part by financial constraints, which limit the number of devices a person can afford. Additionally, in our sample, older adults encountered fewer difficulties with processing long pieces of text compared with the other groups.

The promotion of online symptom checkers as a means to decrease unnecessary clinic visits [13] underscores the importance of ensuring they do not inadvertently increase contact with health care staff. Therefore, greater attention should be directed toward enhancing the cognitive accessibility of these tools, thereby enabling a wider range of users to use them effectively. In this study, users' incomplete understanding of the questions or answer options led them to select additional symptoms, resulting in more serious care recommendations and advising users to seek emergency health care.

In ensuring the cognitive accessibility of eHealth services, it is imperative to involve vulnerable groups in testing. Testing with vulnerable groups provides valuable insights. First, it emphasizes the need for well-planned test sessions with a manageable number of tasks. This approach ensures that participants can fully engage and provide meaningful feedback without being overwhelmed. All of these groups required considerable time to complete the test tasks, with most participants unable to finish both planned tasks with either service. Moreover, they necessitated more detailed instructions and support during the test sessions, as many participants within these groups were not at ease with using eHealth services.

Based on the findings of this study and as supported by the broader universal design literature [5], several design guidelines can be outlined. Foremost among these is the emphasis on clarity. (1) The options provided to the user should be clear and understandable. The user should understand what the differences between different options are and what actions are available for them. (2) It should be made clear to the user where they should be focusing on. This is particularly important in services that contain a lot of information and options. (3) Long or uncommon words and difficult compound words should be avoided. This is especially relevant in health-related terminology, as the user might not understand the special terms and might confuse different terms. (4) Navigating the services should be easy and effortless. The user should be presented with as few options as possible, and excessive scrolling should be minimized. This is because the user may inadvertently overlook relevant information.

Limitations

There are, naturally, some limitations to this study. First, the sample size of 13 participants was rather small, albeit quite typical for this type of qualitative study [43]. However, given the diverse nature of the user group and potential challenges related to cognitive accessibility, a more diverse participant pool could have been beneficial. Specifically, a wider age range of older adults could have been tested, considering their versatility as a group. Additionally, nonnative Finnish speakers could have been recruited from a more geographically diverse range of countries of origin. Moreover, testing should involve other diverse human characteristics, such as neurodiversity (including conditions such as attention-deficit/hyperactivity disorder, attention-deficit disorder, and various forms of autism). Given society's rapid transition toward digitalized services, it is crucial to broaden the scope to include other groups at risk of digital exclusion.

Another limitation of this study is its focus on only 2 online symptom checkers. While the range of available online symptom checkers is already extensive, it is important to include testing of other eHealth services designed for use by all citizens. Additionally, this study only examines a limited list of symptoms and assesses usage on a 1-time or 2-time basis.

In conclusion, we recommend conducting testing with a more diverse user group, with a specific focus on accessibility and cognitive accessibility. Additionally, adopting a broader test setup that encompasses a wider range of symptoms and includes other eHealth services intended for broad usage would be beneficial.

Comparison With Prior Work

Usability issues were efficiently identified during testing with special user groups. In a study by Liu et al [36], which involved 350 participants, similar problems were discovered with service A as found in our study. The authors observed comparable challenges related to understanding questions and terminology, along with a need to enhance the visual layout and instructions for users. However, a notable disparity was observed in completion times: their participants completed the symptom checkers in an average of 4 minutes and 9 seconds, whereas users in our study required, on average, 3 times longer. In addition to uncovering issues that notably impact cognitive accessibility, our study identified similar usability problems as other assessments. Furthermore, as highlighted by Jormanainen et al [23], the same service was used over 1.5 million times for COVID-19 evaluation, suggesting its successful use by a vast number of users. Moreover, challenges with terminology have been recognized in other services [20].

This study has concentrated on cognitive accessibility with 3 distinct user groups. Comparable user groups have been used in other studies that center on eHealth services [1,29,30]. Upon comparing our findings with these studies, we observe that the necessity for clearer language and terminology, along with the clarity of the service, has previously been recognized through interviews and focus groups [1,29]. Our study provides more nuanced insights into how these issues manifest in practical usage.

Conclusions

In this study, we conducted a qualitative usability evaluation of 2 online symptom checkers, with a particular emphasis on the cognitive accessibility of the services. The evaluation targeted potentially vulnerable groups at risk of digital exclusion. Three distinct user groups participated in the tests: older adults, individuals with mild intellectual disabilities, and nonnative Finnish speakers. Our findings revealed that these groups encountered numerous difficulties with the tested services, particularly concerning their clarity and the language/terminology used. Furthermore, when testing with these groups, several key points must be considered: test sessions should be meticulously planned, instructions need to be clear, sessions should not be overly prolonged, and sufficient time must be allocated for each task.

In general, we found that testing with vulnerable groups was both useful and efficient. The rate of usability problems identified was notably high compared with the number of participants, and these issues were readily uncovered. These user groups encountered similar challenges related to information processing. It is imperative to provide them with better support through services that are clear, presenting less information and fewer options at once, and incorporating fewer long and complex words and selection lists. Additionally, following the principles of universal design, the proposed improvements are such that they will also benefit a more general user group. Therefore, we highly recommend testing with potentially vulnerable groups and, furthermore, expanding the user groups to include a representation of a broader variety of cognitive characteristics and challenges.

Acknowledgments

This research was conducted as part of the “Toward a Socially Inclusive Digital Society: Transforming Service Culture” project (DigiIN) and was funded by the Strategic Research Council at the Academy of Finland (grant numbers 352501and 352503) and NordForsk (project 100477). We thank the service providers DigiFinland Oy and Klinik Healthcare Solutions for allowing us to test the services, and Selkeästi meille initiative (a joint project between The Finnish Association on Intellectual and Developmental Disabilities [FAIDD] and Kehitysvammatuki 57 ry) and Väylä ry for helping us to contact the test users. Finally, we thank Felipe González Carceller for his excellent contribution to the research.

Data Availability

The data sets generated during or analyzed during this study are not publicly available due to the sensitive nature of the data but the numeric data are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The distribution of the symptoms and services.
[DOCX File , 17 KB - humanfactors_v11i1e45275_app1.docx]

Multimedia Appendix 2
Symptom vignettes used.
[DOCX File , 16 KB - humanfactors_v11i1e45275_app2.docx]

Multimedia Appendix 3
The background information of the participants. All the participants used digital services multiple times a day. MID: mildly intellectually disabled.
[DOCX File , 15 KB - humanfactors_v11i1e45275_app3.docx]

Multimedia Appendix 4
Usability problem categories.
[DOCX File , 15 KB - humanfactors_v11i1e45275_app4.docx]

Multimedia Appendix 5
Average SUS scores for both services. For older adults the two services got the same results, for the two other groups Service A got a slightly better score.
[DOCX File , 13 KB - humanfactors_v11i1e45275_app5.docx]

References

1. Kaihlanen A, Virtanen L, Buchert U, Safarov N, Valkonen P, Hietapakka L, et al. Towards digital health equity - a qualitative study of the challenges experienced by vulnerable groups in using digital health services in the COVID-19 era. BMC Health Serv Res 2022 Feb 12;22(1):188 [FREE Full text] [doi: 10.1186/s12913-022-07584-4] [Medline: 35151302]
2. European Commission. European accessibility act. European Commission. URL: https://ec.europa.eu/social/main.jsp?catId=1202 [accessed 2022-10-28]
3. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: a scoping review. Int J Med Inform 2019 Jun;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
4. Archer N, Keshavjee K, Demers C, Lee R. Online self-management interventions for chronically ill patients: cognitive impairment and technology issues. Int J Med Inform 2014 Apr;83(4):264-272 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.01.005] [Medline: 24507762]
5. Story MF, Mueller JL, Mace RL. Designing for People of All Ages and Abilities. Raleigh, NC: The Center for Universal Design, NC State University; 1998.
6. International Organization for Standardization (ISO). Ergonomics of human-system interaction: Part 11: Usability: Definitions and concepts (ISO 9241-11:2018). International Organization for Standardization (ISO). 2018. URL: https://www.iso.org/standard/63500.html#:~:text=ISO%209241%2D11%3A2018%20provides,services%20(including%20technical%20and%20personal [accessed 2024-02-24]

7. International Organization for Standardization (ISO). ISO/IEC 25010:2019: Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- System and software quality models. International Organization for Standardization (ISO). URL: https://iso25000.com/index.php/en/iso-25000-standards/iso-25010 [accessed 2024-02-24]
8. Bernard R, Sabariego C, Cieza A. Barriers and facilitation measures related to people with mental disorders when using the web: a systematic review. J Med Internet Res 2016 Jun 09;18(6):e157 [FREE Full text] [doi: 10.2196/jmir.5442] [Medline: 27282115]
9. Muñoz-Hernández S, Benac-Earle C, Herranz Nieva A, Gonzalez-McGuinness M. Sustainable methodology for operational and formal digital skills acquisition: a case study of e-health inclusion. Sustainability 2021 Aug 30;13(17):9698. [doi: 10.3390/su13179698]
10. Miesenberger K, Edler C, Heumader P, Petz A. In: Yesilada Y, Harper S, editors. Tools and Applications for Cognitive Accessibility. London, UK: Springer; 2019:523-546.
11. Chambers D, Cantrell AJ, Johnson M, Preston L, Baxter SK, Booth A, et al. Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review. BMJ Open 2019 Aug 01;9(8):e027743 [FREE Full text] [doi: 10.1136/bmjopen-2018-027743] [Medline: 31375610]
12. Meyer AND, Giardina TD, Spitzmueller C, Shahid U, Scott TMT, Singh H. Patient perspectives on the usefulness of an artificial intelligence-assisted symptom checker: cross-sectional survey study. J Med Internet Res 2020 Jan 30;22(1):e14679 [FREE Full text] [doi: 10.2196/14679] [Medline: 32012052]
13. Marco-Ruiz L, Bønes E, de la Asunción E, Gabarron E, Aviles-Solis JC, Lee E, et al. Combining multivariate statistics and the think-aloud protocol to assess Human-Computer Interaction barriers in symptom checkers. J Biomed Inform 2017 Oct;74:104-122 [FREE Full text] [doi: 10.1016/j.jbi.2017.09.002] [Medline: 28893671]
14. Alajarmeh N. Evaluating the accessibility of public health websites: an exploratory cross-country study. Univers Access Inf Soc 2022 Jan 27;21(3):771-789 [FREE Full text] [doi: 10.1007/s10209-020-00788-7] [Medline: 33526996]
15. Helsper EJ. A corresponding fields model for the links between social and digital exclusion. Commun Theor 2012 Oct 15;22(4):403-426. [doi: 10.1111/j.1468-2885.2012.01416.x]
16. Mulvale G, Moll S, Miatello A, Robert G, Larkin M, Palmer VJ, et al. Codesigning health and other public services with vulnerable and disadvantaged populations: insights from an international collaboration. Health Expect 2019 Jun;22(3):284-297 [FREE Full text] [doi: 10.1111/hex.12864] [Medline: 30604580]
17. Arsenijevic J, Tummers L, Bosma N. Adherence to electronic health tools among vulnerable groups: systematic literature review and meta-analysis. J Med Internet Res 2020 Feb 06;22(2):e11613 [FREE Full text] [doi: 10.2196/11613] [Medline: 32027311]
18. Fang ML, Siden E, Korol A, Demestihas M, Sixsmith J, Sixsmith A. A scoping review exploration of the intended and unintended consequences of eHealth on older people: a health equity impact assessment. Human Technology 2018 Nov 30:297-323. [doi: 10.17011/ht/urn.201811224835]
19. Kampmeijer R, Pavlova M, Tambor M, Golinowska S, Groot W. The use of e-health and m-health tools in health promotion and primary prevention among older adults: a systematic literature review. BMC Health Serv Res 2016 Sep 05;16 Suppl 5(Suppl 5):290 [FREE Full text] [doi: 10.1186/s12913-016-1522-3] [Medline: 27608677]
20. Sayago S, Blat J. Telling the story of older people e-mailing: an ethnographical study. International Journal of Human-Computer Studies 2010 Jan;68(1-2):105-120 [FREE Full text] [doi: 10.1016/j.ijhcs.2009.10.004]
21. Nunes F, Silva PA, Abrantes F. Human-Computer Interaction and the older adult: an example using user research and personas. In: Proceedings of the 3rd International Conference on Pervasive Technologies Related to Assistive Environments, PETRA 2010, Samos, Greece, June 23-25, 2010. 2010 Presented at: The 3rd International Conference on Pervasive Technologies Related to Assistive Environments (PETRA 2010); June 23-25, 2010; Samos, Greece p. 1-8 URL: https://www.researchgate.net/publication/221410323_Human-Computer_Interaction_and_the_older_adult_An_example_using_user_research_and_personas [doi: 10.1145/1839294.1839353]
22. Satre DD, Knight BG, David S. Cognitive-behavioral interventions with older adults: integrating clinical and gerontological research. Professional Psychology: Research and Practice 2006 Oct;37(5):489-498. [doi: 10.1037/0735-7028.37.5.489]
23. Jormanainen V, Soininen L. Use and users of the web-based Omaolo COVID-19 symptom self-assesment tool in Finland since March 16, 2020. Stud Health Technol Inform 2021 May 27;281:739-743. [doi: 10.3233/SHTI210270] [Medline: 34042674]
24. Lussier-Desrochers D, Normand CL, Romero-Torres A, Lachapelle Y, Godin-Tremblay V, Dupont M, et al. Bridging the digital divide for people with intellectual disability. CP 2017 May 31;11(1):Article 1. [doi: 10.5817/cp2017-1-1]
25. Alfredsson Ågren K, Kjellberg A, Hemmingsson H. Digital participation? Internet use among adolescents with and without intellectual disabilities: a comparative study. New Media & Society 2019 Nov 18;22(12):2128-2145. [doi: 10.1177/1461444819888398]
26. Marston HR, Genoe R, Freeman S, Kulczycki C, Musselwhite C. Older adults' perceptions of ICT: main findings from the Technology In Later Life (TILL) study. Healthcare (Basel) 2019 Jul 04;7(3):7030086 [FREE Full text] [doi: 10.3390/healthcare7030086] [Medline: 31277387]

27. Setchell J, Barlott T, Torres M. A socio-emotional analysis of technology use by people with intellectual disabilities. J Intellect Disabil Res 2021 Feb 22;65(2):149-161. [doi: 10.1111/jir.12796] [Medline: 33225540]
28. Fernández Da Silva Á, Buceta BB, Mahou-Lago XM. eHealth policy in Spain: a comparative study between general population and groups at risk of social exclusion in Spain. Digit Health 2022;8:20552076221120724 [FREE Full text] [doi: 10.1177/20552076221120724] [Medline: 35990110]
29. Oudshoorn CEM, Frielink N, Nijs SLP, Embregts PJCM. eHealth in the support of people with mild intellectual disability in daily life: a systematic review. J Appl Res Intellect Disabil 2020 Nov 23;33(6):1166-1187 [FREE Full text] [doi: 10.1111/jar.12758] [Medline: 32578361]
30. Ware P, Bartlett SJ, Paré G, Symeonidis I, Tannenbaum C, Bartlett G, et al. Using eHealth technologies: interests, preferences, and concerns of older adults. Interact J Med Res 2017 Mar 23;6(1):e3 [FREE Full text] [doi: 10.2196/ijmr.4447] [Medline: 28336506]
31. Kadir SA, Jamaludin M. Universal design as a significant component for sustainable life and social development. Procedia - Social and Behavioral Sciences 2013 Sep;85:179-190 [FREE Full text] [doi: 10.1016/j.sbspro.2013.08.349]
32. Ruzic L, Sanfod J. Universal design mobile interface guidelines (UDMIG) for an aging population. In: Marston HR, Freeman S, Musselwhite C, editors. Mobile e-Health. Cham, Switzerland: Springer International Publishing; 2017:17-37.
33. Raspa M, Fitzgerald T, Furberg R, Wylie A, Moultrie R, DeRamus M, et al. Mobile technology use and skills among individuals with fragile X syndrome: implications for healthcare decision making. J Intellect Disabil Res 2018 Oct;62(10):821-832 [FREE Full text] [doi: 10.1111/jir.12537] [Medline: 30105880]
34. Watfern C, Heck C, Rule C, Baldwin P, Boydell KM. Feasibility and acceptability of a mental health website for adults with an intellectual disability: qualitative evaluation. JMIR Ment Health 2019 Mar 28;6(3):e12958 [FREE Full text] [doi: 10.2196/12958] [Medline: 30920378]
35. Persson H, Åhman H, Yngling AA, Gulliksen J. Universal design, inclusive design, accessible design, design for all: different concepts—one goal? On the concept of accessibility—historical, methodological and philosophical aspects. Univ Access Inf Soc 2014 May 7;14(4):505-526. [doi: 10.1007/s10209-014-0358-z]
36. Liu VD, Sellgren L, Kaila M, Koskela T. Sähköisten oirearvioiden käytettävyys. FinJeHeW 2021 Jun 18;13(2):97020. [doi: 10.23996/fjhw.97020]
37. Kim H. A study on usability improvement of mobile health care services. Int J Adv Smart Converg 2017;6(2):72-81. [doi: 10.7236/IJASC.2017.6.2.72]
38. Knitza J, Mohn J, Bergmann C, Kampylafka E, Hagen M, Bohr D, et al. Accuracy, patient-perceived usability, and acceptance of two symptom checkers (Ada and Rheport) in rheumatology: interim results from a randomized controlled crossover trial. Arthritis Res Ther 2021 Apr 13;23(1):112 [FREE Full text] [doi: 10.1186/s13075-021-02498-8] [Medline: 33849654]
39. Broekhuis M, van Velsen L, Peute L, Halim M, Hermens H. Conceptualizing usability for the eHealth context: content analysis of usability problems of ehealth applications. JMIR Form Res 2021 Jul 27;5(7):e18198 [FREE Full text] [doi: 10.2196/18198] [Medline: 34313594]
40. Omaolo. DigiFInland Oy. URL: https://www.omaolo.fi [accessed 2022-07-10]
41. Klinik Access. Klinik Healthcare Solutions Oy. URL: http://www.klinik.fi [accessed 2022-07-10]
42. Etikan I. Comparison of convenience sampling and purposive sampling. AJTAS 2016;5(1):1. [doi: 10.11648/j.ajtas.20160501.11]
43. Caine K. Local standards for sample size at CHI. In: CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 2016 May Presented at: CHI '16: The 2016 CHI Conference on Human Factors in Computing Systems; May 7-12, 2016; San Jose, CA p. 981-992. [doi: 10.1145/2858036.2858498]
44. Semigran H, Linder J, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ 2015 Jul 08;351:h3480. [doi: 10.1136/bmj.h3480] [Medline: 26157077]
45. North F, Ward WJ, Varkey P, Tulledge-Scheitel SM. Should you search the internet for information about your acute symptom? Telemed J E Health 2012 Apr;18(3):213-218. [doi: 10.1089/tmj.2011.0127] [Medline: 22364307]
46. Eronen J, Paakkari L, Portegijs E, Saajanaho M, Rantanen T. Assessment of health literacy among older Finns. Aging Clin Exp Res 2019 Apr 21;31(4):549-556 [FREE Full text] [doi: 10.1007/s40520-018-1104-9] [Medline: 30578457]
47. Brooke J. SUS: a 'quick and dirty' usability scale. In: Usability Evaluation In Industry (1st Edition). London, UK: CRC Press; 1996:1-6.
48. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. International Journal of Human-Computer Interaction 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]

Abbreviations

SUS: System Usability Scale
UI: user interface

Edited by Y Gong; submitted 22.12.22; peer-reviewed by C Preiser, E Bønes; comments to author 16.02.23; revised version received 08.03.23; accepted 03.02.24; published 08.03.24.

Please cite as:
Savolainen K, Kujala S
Testing Two Online Symptom Checkers With Vulnerable Groups: Usability Study to Improve Cognitive Accessibility of eHealth Services
JMIR Hum Factors 2024;11:e45275
URL: https://humanfactors.jmir.org/2024/1/e45275
doi:10.2196/45275
PMID:38457214

Original Paper

Head Protection Device for Individuals at Risk for Head Injury due to Ground-Level Falls: Single Trauma Center User Experience Investigation

Susan Haag[1*], PhD; John Kepros[1*], MD

Scottsdale Osborn Medical Center, Scottsdale, AZ, United States

[*]all authors contributed equally

Corresponding Author:
Susan Haag, PhD
Scottsdale Osborn Medical Center
7400 E Osborn Rd
Scottsdale, AZ, 85251
United States
Phone: 1 480 323 4018
Email: susan.haag@ymail.com

Abstract

Background: Falls represent a large percentage of hospitalized patients with trauma as they may result in head injuries. Brain injury from ground-level falls (GLFs) in patients is common and has substantial mortality. As fall prevention initiatives have been inconclusive, we changed our strategy to injury prevention. We identified a head protection device (HPD) with impact-resistant technology, which meets head impact criteria sustained in a GLF. HPDs such as helmets are ubiquitous in preventing head injuries in sports and industrial activities; yet, they have not been studied for daily activities.

Objective: We investigated the usability of a novel HPD on patients with head injury in acute care and home contexts to predict future compliance.

Methods: A total of 26 individuals who sustained head injuries, wore an HPD in the hospital, while ambulatory and were evaluated at baseline and 2 months post discharge. Clinical and demographic data were collected; a usability survey captured HPD domains. This user experience design revealed patient perceptions, satisfaction, and compliance. Nonparametric tests were used for intragroup comparisons (Wilcoxon signed rank test). Differences between categorical variables including sex, race, and age (age group 1: 55-77 years; age group 2: 78+ years) and compliance were tested using the chi-square test.

Results: Of the 26 patients enrolled, 12 (46%) were female, 18 (69%) were on anticoagulants, and 25 (96%) were admitted with a head injury due to a GLF. The median age was 77 (IQR 55-92) years. After 2 months, 22 (85%) wore the device with 0 falls and no GLF hospital readmissions. Usability assessment with 26 patients revealed positive scores for the HPD post discharge regarding satisfaction (mean 4.8, SD 0.89), usability (mean 4.23, SD 0.86), effectiveness (mean 4.69, SD 0.54), and relevance (mean 4.12, SD 1.10). Nonparametric tests showed positive results with no significant differences between 2 observations. One issue emerged in the domain of aesthetics; post discharge, 8 (30%) patients had a concern about device weight. Analysis showed differences in patient compliance regarding age (χ_1^2=4.27; *P*=.04) but not sex (χ_1^2=1.58; *P*=.23) or race (χ_1^2=0.75; *P*=.60). Age group 1 was more likely to wear the device for normal daily activities. Patients most often wore the device ambulating, and protection was identified as the primary benefit.

Conclusions: The HPD intervention is likely to have reasonably high compliance in a population at risk for GLFs as it was considered usable, protective, and relevant. The feasibility and wearability of the device in patients who are at risk for GLFs will inform future directions, which includes a multicenter study to evaluate device compliance and effectiveness. Our work will guide other institutions in pursuing technologies and interventions that are effective in mitigating injury in the event of a fall in this high-risk population.

(JMIR Hum Factors 2024;11:e54854) doi:10.2196/54854

KEYWORDS

health care interventions and technologies; user experience research; usability; brain injury; ground-level fall (GLF); head protection device (HPD); fall risk; patient compliance

Introduction

Frailty in aging is represented by a decline in functioning, with a risk of poor outcomes, including falls, which have implications for clinical practice and public health [1]. Falls are the primary cause of injury-related death in aging adults as 33% of adults 65 years and older fall each year [2,3].

Falls also represent a large percentage of hospitalized older patients as they may result in multiple injuries, including head trauma [4-6]. A head injury can be a common cause of disability and mortality and may be as mild as a bump, bruise (contusion), or cut and can be moderate to severe due to a concussion. Head injury may lead to premature nursing home admissions and increased hospital length of stay (LOS) with undesirable results for patients and hospitals [7,8]. Due to the aging population worldwide, the incidence of falls will continue to rise [9,10].

Studies have shown a clear pattern of increased health care costs associated with falls and frail individuals and various fall prevention initiatives have been promoted. Of the fall prevention interventions studied, some results have been favorable, such as those with well-developed educational programs [11]. However, others have been inconclusive [12-14], prompting our center to include head injury prevention, and therefore, we investigated a head protection device (HPD), similar to a helmet. In many fields, such as construction and sports, helmets have shown efficacy in preventing head injury risks, especially moderate to severe head injury [15-17]. The human head is vulnerable to even moderate impact as it can cause injury or death. A greater emphasis has been placed on job safety in industries like construction particularly to protect the head from injury, and hard hats and helmets have been required [18]. However, historically, helmets have not been used for normal daily living.

Health care systems are increasingly looking for contexts that provide accessible and efficient care and for medical devices and interventions to improve the patient experience and health outcomes [19,20]. Human factors, a scientific discipline, is important in clinical practice as it reveals how humans interact with interventions, such as devices, regarding expectations and limitations. User experience (UX) focuses on having a deep understanding of users and what they need and value [21-23]. UX research has been used to ascertain user domains such as adherence, usability, and perceived impact and has assisted with intervention development and refinement [24]. Adopting a UX research design will help ensure that new devices are easy to use and meet the needs of most patients.

Clinical practices should target effective strategies that improve individuals' quality of life and independence including screenings and interventions to manage injuries associated with falls [25,26]. Screenings that measure activities of daily living (ADLs) are essential, as the ability to perform daily tasks safely without exhaustion is a critical component of healthy aging, thus allowing older individuals to maintain their independence and quality of life [27]. Measurement of daily activities is important as these may be predictors of early admission to assisted care facilities or the need for alternative living arrangements [28,29].

Recent literature advocates change toward tailored interventions that preserve an individual's independence by promoting furthering advancements in evidence-based treatment options and identifying cost-effective strategies [2,3]. Due to an increasing incidence of head injuries after ground-level falls (GLFs) in our trauma center, we designed a study that examined the effects of a low-cost HPD that has the potential to prevent head injury due to a fall.

The purpose of this UX research was to assess compliance by investigating the usability of an HPD from a patient's perspective in both acute care (hospital) and home contexts. We hypothesized that consented patients would follow the research protocol as recommended and wear the device in the hospital and at the 2 months post discharge. The primary limitation in an aging population is compliance, which we approached first. This in-hospital and home-based UX investigation concerning a low-cost treatment option may serve clinicians to better manage frailty and mitigate injury due to falls in their clinical practice.

Methods

Study Design

We considered the UX of frail individuals at this developmental, exploratory stage of a device to examine patient adherence and use. The UX assessment instrument adopted UX domains with a 5-point scale showing a more positive rating (rating of 5) and a lower rating (rating of 1). UX domains included device credibility, satisfaction, usability, adherence, effectiveness, relevance, and aesthetics. The primary outcome variable is patient compliance regarding wearing the device for 2 months. Additional data collected included the frequency of wearing the device during normal daily activities. Consistent with the literature, ADLs (such as ambulating and preparing meals) are critical for independence in aging populations [29].

Recruitment

Participants were recruited from among patients who were treated at our level 1 trauma center and subsequently admitted to the hospital for observation due to head injury. Protocol inclusion criteria included the following: patients admitted to the hospital with a fall sustaining a head injury, patients with fall risk (eg, patients who fell within the prior year or other physical conditions aligned with fall risk), and patients who were ambulatory and 55 years or older. Head injuries included in the study were patients with a concussion, contusion, lacerations, or loss of consciousness. The individuals recruited did not experience trauma that required surgical intervention. After signing the consent in the hospital, individuals were given an HPD at no cost to wear while ambulatory. After consenting and wearing the HPD for in-hospital observation (and just before discharge), the hospital team asked whether the patients would wear the HPD at home. If the patient agreed, we indicated that the research team would follow up post discharge for additional observations using the UX survey.

Ethical Considerations

In total, 26 patients, who experienced a fall and sustained a head injury, wore an HPD in hospital, while ambulatory and were evaluated at baseline (before discharge) and at 2 months post discharge. The study protocol was approved by the institutional review board for research ethics and subsequently approved (IRB 1804935). Informed consent was obtained from the 26 patients who met the inclusion criteria and were willing to participate. Confidentiality of information was maintained. The data are anonymized and patients are deidentified. Each patient was assigned a discrete number in the study and data are secured by the research scientist. There was no compensation for patient participation in the study.

HPD

The HPD includes an impact-resistant technological insert for additional head protection. It helps protect against bumps, scrapes, bruises, and other head injuries. The HPD is designed with ventilation to provide airflow for breathability without compromised protection. The HPD size can be adjusted with a hook and loop strap to give a quick, secure fit. Figure 1 displays the HPD, which looks like a typical baseball cap.

Figure 1. Head protection device.

Usability Survey

A multidisciplinary health care team comprised of physicians, a research scientist, and physical therapists collaborated on the study design, developing a usability survey for patients who are at risk of fall, which led to a tangible and targeted intervention strategy. UX (usability) domain definitions were identified in the literature. Existing domain definitions were examined such as credibility, usability, and satisfaction [24], and additional domains were defined such as effectiveness, relevance, and aesthetics. The domains were refined, used on the usability survey instrument, and functioned as outcome measures. Textbox 1 shows the domains and UX definitions. UX domain data were collected on the instrument using a 5-point scale (5=strongly agree, 4=agree, 3=neutral, 2=disagree, and 1=strongly disagree). Patients were asked if they would recommend the HPD. The survey was intended to evaluate the HPD's usability and was administered after patients concluded their interaction with the HDP in the hospital. Those who agreed to wear the HPD at home were provided a device and were reevaluated post discharge.

Textbox 1. Domain and user experience definitions.

- Credibility: whether the user perceives the device to be trustworthy (eg, accuracy and quality of information presented in the patient consent)
- Satisfaction: the user's overall experience and interaction with the device
- Usability: the user's perceived ease of use of the device based on technical factors
- Adherence: whether the patient followed the device research protocol and continued to use the device as recommended (compliance) completing outcome measures
- Effectiveness: the extent the user perceives the overall value of the device, including safety and whether they would recommend it to another fall risk individual
- Relevance: the extent to which the device is appropriate for their situation and whether they perceive it meets their needs (provides protection to their head and helps them maintain a sense of independence)
- Aesthetics: factors such as color, pattern, size, shape, and weight

Data Collection

Quantitative data included demographics (age, sex, and race) and clinical data such as hospital LOS, number of GLFs, readmission to the hospital due to a GLF, and Glasgow Coma Scale. Data were also captured on the usability survey including domains such as device satisfaction, effectiveness, relevance, and aesthetics. Qualitative data were also collected on the usability survey, and patient comments were recorded regarding HPD benefits and opportunities for improvement.

Statistical Analysis

This UX research methodology included multiple patient observations and differences between observations were examined. Nonparametric tests, used to analyze ordinal and categorical data, were used for intragroup comparisons (Wilcoxon signed rank test). We used descriptive statistics, such that patterns might emerge from the data. Frequencies and percentages are reported for categorical variables. Medians and means with SDs are reported for continuous variables as appropriate. All computations included 26 patients. Group comparisons were made using chi-square tests or Fisher exact tests, where numbers were small and were reported as numbers (%). All variables were assessed for normality. Analyses of categorical variables (age) and patient adherence were tested using the chi-square statistic. Statistical tests are 2-tailed, with a significance level of an α of .05. All statistical analyses were performed using SPSS Statistics for Windows (version 28.0; IBM Corp).

Open-ended patient comments (qualitative data) were analyzed using a 3-step process: data reduction, data display, and conclusion drawing and verification. Data reduction helped sort and compile data excerpts (to organize the data) and assist in developing assertions regarding patient perceptions surrounding wearability (eg, comfort and weight) and modifications of HPD, if necessary. Excerpts were annotated with topics such as the benefits of HPD: positive feedback (aspects recorded as positive by the patient participants regarding HPD experience and interaction) and negative feedback (points considered negative by the patients pertaining to interaction with the device). As a next step, we analyzed the themes that emerged and categorized them based on whether they were related to the usability of the HPD or the health support the device offered.

Results

Study Population

Among the 26 participants, 12 (46%) were female and 5 (19%) were non-White, with a median age of 77 (IQR 55-92) years. The average hospital LOS was 3.8 (SD 3.65) days. The majority (n=25, 96%) of patients who experienced head trauma were admitted to the hospital with a head injury due to a GLF (n=1, 4% were other types of falls); 22 (85%) had prior falls in the last 12 months and 16 (62%) had a hospital visit due to a head injury related to a fall within the year; 18 (69%) were on anticoagulants. The mean Glasgow Coma Score was 14.2 (SD 0.44). The age category was divided into 2 groups for analysis: age group 1 comprised of those who were 55 to 77 years and age group 2 comprised of patients 78 years and older.

Usability Survey Domain Results

In the hospital, all 26 consented patients wore the device with 0 falls recorded. After 2 months, 22 (85%) were wearing the HPD, had 0 falls, and had no hospital readmissions due to GLFs. At 6 months, 16 (62%) patients were compliant with wearing the device, with 0 falls and no hospital readmissions due to a GLF. The results showed positive scores, with no significant differences between ratings in hospital and post discharge regarding device credibility (0.42), satisfaction (0.60), usability (0.80), adherence (0.06), effectiveness (0.53), and relevance (0.09). A difference emerged for the domain of aesthetics. After the discharge, 8 (30%) patients had concerns regarding the device's weight, saying it was slightly heavier than a typical cap. Overall, users had a positive experience with the HPD and scores revealed that patients felt it was effective and relevant. Thus, post discharge, users would recommend the HPD to others at risk for falls (mean 4.52, SD 0.51). Users were compliant by wearing the device in hospital and at 2 months post discharge, supporting the research hypothesis. Table 1 displays the UX domain means (SDs) for 2 observations.

Differences between categorical variables (age group 1: 55-77 years, group 2: 78 years and older, sex, and race) and protocol adherence were analyzed. Chi-square analysis showed differences in compliance regarding age (χ_1^2=4.27; *P*=.04) but not sex (χ_1^2=1.58; *P*=.23) or race (χ_1^2=0.75; *P*=.60). Age group 1 was more likely to wear the device for normal daily activities.

Table 1. In-hospital and postdischarge intragroup domain differences.

User experience domains	Hospital, mean (SD)	Postdischarge, mean (SD)	*P* value
Credibility	3.91 (0.80)	4.01 (0.84)	.42
Satisfaction	4.15 (0.88)	4.80 (0.89)	.60
Usability	4.27 (0.66)	4.23 (0.86)	.80
Adherence	4.50 (0.86)	4.30 (1.06)	.06
Effectiveness (value)	4.62 (0.49)	4.69 (0.54)	.53
Relevance	4.42 (0.75)	4.12 (1.10)	.09
Aesthetics	3.38 (1.30)	2.96 (1.83)	.003

Patient Device Use in Daily Activities

The usability survey data captured patient device use during typical ADLs at 2 weeks and at 2 months post discharge. Users were provided a list of daily activities and were asked to rate the frequency of wearing the device. Consistent with the literature, ADLs, such as ambulating and preparing meals, are critical for independence in an aging population [29]. The highest score on the usability instrument was a “5” which indicated that the patient would wear the HPD “most often.” In-home contexts, patients indicated they most often wore the device ambulating and when driving (to meals and doctor appointments) and less often for personal hygiene. Table 2 shows within-group differences in device use in daily activities.

Table 2. Within-group differences in device use in daily activities.

Daily activities	Two weeks, mean (SD)	Two months, mean (SD)	*P* value
Ambulating	4.31 (0.92)	4.15 (1.12)	.47
Driving (or being driven)	4.04 (0.77)	4.12 (0.76)	.16
Grocery shopping or shopping	3.69 (1.28)	3.58 (1.23)	.54
Relaxing (TV)	4.00 (1.06)	3.31(1.10)	.20
Housekeeping	3.35 (1.09)	3.27 (1.00)	.67
Preparing meals	2.77 (0.99)	2.50 (1.06)	.07
Personal hygiene	2.42 (0.94)	2.27 (0.96)	.49

Positive Patient Feedback

Open-ended questions on the usability instrument elicited patient qualitative comments regarding HPD benefits and opportunities for improvement. As a result, 2 dominant themes emerged, namely HPD usability and HPD as health support (protection). Usability was associated with the use of the device and functionality in terms of wearability. Health support included themes that were aligned with head protection for a patient.

Usability and relevance from the patients’ perspective translated into wearability, and the majority of patients wore the device after 2 months post discharge. Participants felt that the HPD was comfortable and easy to wear. However, 8 (30%) patients mentioned that the HPD was not as light as a typical cap due to the protective “technology insert” and suggested the HPD could be lighter in weight. One male participant stated,

> *The cap is heavier than a usual baseball cap and it took me longer to get used to it. I would like it a bit lighter in weight if possible and more air vents to let in air.*

Health support from the participant’s perspective sufficed as the primary benefit, as 18 (69%) commented that the device protected their head in the event of a fall. Patients called the device a “cap” as it resembles a baseball cap. One patient stated, “Protection for my head is important. I will wear it going out to eat and to doctor appointments.” Another female participant indicated, “I wear it eight hours a day to protect my head.” Two patients (male and female) indicated post discharge, they hit their heads on cabinets, as 1 commented:

> *I already bent over and hit my head on a cabinet; it protected me from another head injury. Since wearing the cap, I have not had a fall, only a bump and I had on my cap.*

A 74-year-old female participant stated, “I fell last year and I will wear this walking whenever possible. It protects my head.” A male participant noted, “The device is protective and comfortable; I forgot I had it on.” From patient comments, the HPD is cognate with head protection.

Discussion

Principal Findings

Using a UX design, we investigated the usability of a novel HPD on patients with head injury in acute care and home contexts to predict future compliance. All 26 patients provided positive scores for the HPD post discharge regarding satisfaction, usability, effectiveness, and relevance. Nonparametric tests showed positive results, with no significant differences between 2 observations at 2 months. Chi-square analysis showed a significant difference in HPD compliance regarding age but not sex or race as age group 1 was more likely to wear the device for normal daily activities. Patients most often wore the device ambulating and head protection was identified as the primary benefit. Thus, patients were most likely to recommend the HPD to others at risk of GLFs.

Due to the consistently high rate of head injuries after GLFs in our center, the targeted team strategy for an HPD and UX research design was developed. We realized that patient compliance in the geriatric population has been a limiting factor and approached that aspect first. Patients adhered to the research protocol by wearing the device in the hospital and post discharge, in the home, supporting the research hypothesis. At 2 months, 22 (85%) patients wore the device with 0 falls recorded and no readmissions due to falls.

Our multidisciplinary team, a diverse group of medical professionals, consisting of physicians, research scientists, and physical therapists, studied a device to be worn during daily activities in home environments. Recent literature has advocated for home care strategies [30] and interventions to be used in home contexts where falls most often occur [31]. Managing falls in this high-risk population is complex, requiring a systemic and collaborative approach directed by a multidisciplinary team focused on improving patient outcomes [3].

Limitations

Accuracy is critical regarding the collection of patient data, and the in-hospital data collection was conducted under medical supervision. However, the limitations of the UX research included the nature of self-reporting by participants post discharge at 2 and 6 months. One measure to counter this bias was to include a family member during the evaluation to corroborate the patient's self-reported data and responses. Another issue and limitation, we noted, was the difficulty of trying to reconnect or contact this population at follow-up due to cognitive decline, the extent and severity of head trauma, and other injuries associated with a GLF.

Conclusions

The results show our proposed HPD intervention will have a high compliance rate in those at risk for GLFs as it was considered usable, protective, and relevant. Managing individuals with fall risk may include future investigations of specific interventions and low-cost devices that preserve a patient's independence and physical function, and research that contributes to further advancements in evidence-based treatment options. The feasibility and wearability of the device in patients with GLF with head injuries will inform future directions, which includes a multicenter study to evaluate compliance and device effectiveness. Our work will guide other health care institutions in pursuing cost-effective treatments and technological interventions that are usable and effective in improving outcomes for this fall risk population.

Conflicts of Interest

None declared.

References

1. Dent E, Martin FC, Bergman H, Woo J, Romero-Ortuno R, Walston JD. Management of frailty: opportunities, challenges, and future directions. Lancet 2019;394(10206):1376-1386. [doi: 10.1016/S0140-6736(19)31785-4] [Medline: 31609229]
2. Hsieh KL, Fanning JT, Rogers WA, Wood TA, Sosnoff JJ. A fall risk mHealth app for older adults: development and usability study. JMIR Aging 2018 Jan 12;1(2):1-12 [FREE Full text] [doi: 10.2196/11569] [Medline: 31518234]
3. Jacobson CL, Foster LC, Arul H, Rees A, Stafford RS. A digital health fall prevention program for older adults: feasibility study. JMIR Form Res 2021 Dec 23;5(12):1-11 [FREE Full text] [doi: 10.2196/30558] [Medline: 34837492]
4. Dunnell L, Barnard AN, Chu K, Barling A, Birns J, Walker G. In situ simulation training for frailty. Geriatrics (Basel) 2023 Feb 17;8(1):1-9 [FREE Full text] [doi: 10.3390/geriatrics8010026] [Medline: 36826368]
5. Khurrum M, Chehab M, Ditillo M, Richards J, Douglas M, Bible L, et al. Trends in geriatric ground-level falls: report from the National Trauma Data Bank. J Surg Res 2021 May 24;266:261-268. [doi: 10.1016/j.jss.2021.02.047] [Medline: 34034061]
6. Tricco AC, Thomas SM, Veroniki AA, Hamid JS, Cogo E, Strifler L, et al. Comparisons of interventions for preventing falls in older adults: a systematic review and meta-analysis. JAMA 2017 Nov 07;318(17):1687-1699 [FREE Full text] [doi: 10.1001/jama.2017.15006] [Medline: 29114830]
7. Jones DM, Connelly DW, Burgess JR. Ground level falls in patients over 60, do race and socioeconomic status affect outcome? Am Surg 2022 Apr 23;88(9):2127-2131. [doi: 10.1177/00031348221091961] [Medline: 35465738]
8. Cameron ID, Dyer SM, Panagoda CE, Murray GR, Hill KD, Cumming RG, et al. Interventions for preventing falls in older people in care facilities and hospitals. Cochrane Database Syst Rev 2018 Sep 07;9(9):1-255 [FREE Full text] [doi: 10.1002/14651858.CD005465.pub4] [Medline: 30191554]
9. Grabel ZJ, Lunati MP, Segal DN, Kukowski NR, Yoon ST, Jain A. Thoracolumbar spinal fractures associated with ground level falls in the elderly: an analysis of 254,486 emergency department visits. J Clin Orthop Trauma 2020 Apr 10;11(5):916-920 [FREE Full text] [doi: 10.1016/j.jcot.2020.04.009] [Medline: 32879581]
10. Yee G, Jain A. Geriatric head injury. In: StatPearls. Treasure Island (FL): StatPearls Publishing; Jan 07, 2021:1-10.
11. Heng H, Jazayeri D, Shaw L, Kiegaldie D, Hill AM, Morris ME. Hospital falls prevention with patient education: a scoping review. BMC Geriatr 2020 Apr 15;20(1):140-147 [FREE Full text] [doi: 10.1186/s12877-020-01515-w] [Medline: 32293298]
12. Lamb SE, Bruce J, Hossain A, Ji C, Longo R, Lall R, et al. Screening and intervention to prevent falls and fractures in older people. N Engl J Med 2020 Nov 05;383(19):1848-1859 [FREE Full text] [doi: 10.1056/NEJMoa2001500] [Medline: 33211928]
13. Rezaei-Shahsavarloo Z, Atashzadeh-Shoorideh F, Gobbens RJJ, Ebadi A, Harouni GG. The impact of interventions on management of frailty in hospitalized frail older adults: a systematic review and meta-analysis. BMC Geriatr 2020 Dec 03;20(1):526-529 [FREE Full text] [doi: 10.1186/s12877-020-01935-8] [Medline: 33272208]
14. Williams CT, Whyman J, Loewenthal J, Chahal K. Managing geriatric patients with falls and fractures. Orthop Clin North Am 2023 May 31;54(3S):e1-e12. [doi: 10.1016/j.ocl.2023.04.001] [Medline: 37349065]
15. McIntosh AS, Andersen TE, Bahr R, Greenwald R, Kleiven S, Turner M, et al. Sports helmets now and in the future. Br J Sports Med 2011 Dec 04;45(16):1258-1265. [doi: 10.1136/bjsports-2011-090509] [Medline: 22117017]
16. Olivier J, Creighton P. Bicycle injuries and helmet use: a systematic review and meta-analysis. Int J Epidemiol 2017 Feb 01;46(1):278-292 [FREE Full text] [doi: 10.1093/ije/dyw153] [Medline: 27450862]

17. Long J, Yang J, Lei Z, Liang D. Simulation-based assessment for construction helmets. Comput Methods Biomech Biomed Engin 2015 Mar 18;18(1):24-37. [doi: 10.1080/10255842.2013.774382] [Medline: 23495784]
18. Wu JZ, Pan CS, Wimer BM, Rosen CL. Finite element simulations of the head-brain responses to the top impacts of a construction helmet: effects of the neck and body mass. Proc Inst Mech Eng H 2017 Dec 21;231(1):58-68 [FREE Full text] [doi: 10.1177/0954411916678017] [Medline: 28097935]
19. Bitkina OV, Jeong H, Lee BC, Park J, Park J, Kim HK. Perceived trust in artificial intelligence technologies: a preliminary study. Hum Factors Ergon Manuf Serv Ind 2020;30(4):282-290. [doi: 10.1002/hfm.20839]
20. Williams PAH, Lovelock B, Cabarrus JA. A sense of coherence approach to improving patient experience using information infrastructure modeling: design science research. JMIR Form Res 2022 Apr 12;6(4):e35418 [FREE Full text] [doi: 10.2196/35418] [Medline: 35307641]
21. Sauer J, Sonderegger A, Schmutz S. Usability, user experience and accessibility: towards an integrative model. Ergonomics 2020 Jun 12;63(10):1207-1220. [doi: 10.1080/00140139.2020.1774080] [Medline: 32450782]
22. Allen J, Hutchinson AM, Brown R, Livingston PM. User experience and care integration in transitional care for older people from hospital to home: a meta-synthesis. Qual Health Res 2017 Jul 28;27(1):24-36. [doi: 10.1177/1049732316658267] [Medline: 27469975]
23. Chagas BA, Pagano AS, Prates RO, Praes EC, Ferreguetti K, Vaz H, et al. Evaluating user experience with a chatbot designed as a public health response to the COVID-19 pandemic in Brazil: mixed methods study. JMIR Hum Factors 2023 Apr 04;10:e43135 [FREE Full text] [doi: 10.2196/43135] [Medline: 36634267]
24. Newton AS, March S, Gehring ND, Rowe AK, Radomski AD. Establishing a working definition of user experience for eHealth interventions of self-reported user experience measures with eHealth researchers and adolescents: scoping review. J Med Internet Res 2021 Oct 26;23(12):e25012 [FREE Full text] [doi: 10.2196/25012] [Medline: 34860671]
25. Abrahamson K, Hass Z, Arling G. Shall I stay or shall I go? The choice to remain in the nursing home among residents with high potential for discharge. J Appl Gerontol 2020 Oct 26;39(8):863-870. [doi: 10.1177/0733464818807818] [Medline: 30366510]
26. Cordes T, Bischoff LL, Schoene D, Schott N, Voelcker-Rehage C, Meixner C, et al. A multicomponent exercise intervention to improve physical functioning, cognition and psychosocial well-being in elderly nursing home residents: a study protocol of a randomized controlled trial in the PROCARE (prevention and occupational health in long-term care) project. BMC Geriatr 2019 Dec 23;19(1):369 [FREE Full text] [doi: 10.1186/s12877-019-1386-6] [Medline: 31870314]
27. Poncumhak P, Amput P, Sangkarit N, Promsrisuk T, Srithawong A. Predictive ability of the 2-minute step test for functional fitness in older individuals with hypertension. Ann Geriatr Med Res 2023 Jul 28;27(3):228-234 [FREE Full text] [doi: 10.4235/agmr.23.0070] [Medline: 37500190]
28. Zhang C, Dong X, Ding M, Chen X, Shan X, Ouyang H, et al. Executive control, alerting, updating, and falls in cognitively healthy older adults. Gerontology 2020 Aug 25;66(5):494-505. [doi: 10.1159/000509288] [Medline: 32841943]
29. Costenoble A, Knoop V, Vermeiren S, Vella RA, Debain A, Rossi G, et al. A comprehensive overview of activities of daily living in existing frailty instruments: a systematic literature search. Gerontologist 2021 Apr 03;61(3):e12-e22 [FREE Full text] [doi: 10.1093/geront/gnz147] [Medline: 31872238]
30. Liu-Ambrose T, Davis JC, Best JR, Dian L, Madden K, Cook W, et al. Effect of a home-based exercise program on subsequent falls among community-dwelling high-risk older adults after a fall: a randomized clinical trial. JAMA 2019 Jun 04;321(21):2092-2100 [FREE Full text] [doi: 10.1001/jama.2019.5795] [Medline: 31162569]
31. Gazibara T, Kurtagic I, Kisic-Tepavcevic D, Nurkovic S, Kovacevic N, Gazibara T, et al. Falls, risk factors and fear of falling among persons older than 65 years of age. Psychogeriatrics 2017 Jan 27;17(4):215-223. [doi: 10.1111/psyg.12217] [Medline: 28130862]

Abbreviations

ADL: activities of daily living
GLF: ground-level fall
HPD: head protection device
LOS: length of stay
UX: user experience

Edited by A Kushniruk; submitted 26.11.23; peer-reviewed by DD Larson MD, S Li, D Low; comments to author 31.01.24; revised version received 08.02.24; accepted 18.02.24; published 19.03.24.

Please cite as:
Haag S, Kepros J
Head Protection Device for Individuals at Risk for Head Injury due to Ground-Level Falls: Single Trauma Center User Experience Investigation
JMIR Hum Factors 2024;11:e54854
URL: https://humanfactors.jmir.org/2024/1/e54854
doi:10.2196/54854
PMID:38502170

Original Paper

A Digital Health Intervention Platform (Active and Independent Management System) to Enhance the Rehabilitation Experience for Orthopedic Joint Replacement Patients: Usability Evaluation Study

Petros Papadopoulos[1*], BSc, PhD; Mario Soflano[2*], BSc, PhD; Thomas Connolly[3*], BSc, PhD

[1]University of Strathclyde, Glasgow, United Kingdom
[2]Glasgow Caledonian University, Glasgow, United Kingdom
[3]DS Partnership, Glasgow, United Kingdom
[*]all authors contributed equally

Corresponding Author:
Petros Papadopoulos, BSc, PhD
University of Strathclyde
26 Richmond Street
Glasgow, G1 1XH
United Kingdom
Phone: 44 7920756076
Email: petros.papadopoulos@gmail.com

Abstract

Background: Optimal rehabilitation programs for orthopedic joint replacement patients ensure faster return to function, earlier discharge from hospital, and improved patient satisfaction. Digital health interventions show promise as a supporting tool for re-enablement.

Objective: The main goal of this mixed methods study was to examine the usability of the AIMS platform from the perspectives of both patients and clinicians. The aim of this study was to evaluate a re-enablement platform that we have developed that uses a holistic systems approach to address the *de-enablement* that occurs in hospitalized inpatients, with the older adult population most at risk. The Active and Independent Management System (AIMS) platform is anticipated to deliver improved patient participation in recovery and self-management through education and the ability to track rehabilitation progression in hospital and after patient discharge.

Methods: Two well-known instruments were used to measure usability: the System Usability Scale (SUS) with 10 items and, for finer granularity, the User Experience Questionnaire (UEQ) with 26 items. In all, 26 physiotherapists and health care professionals evaluated the AIMS clinical portal; and 44 patients in hospital for total knee replacement, total hip replacement, or dynamic hip screw implant evaluated the AIMS app.

Results: For the AIMS clinical portal, the mean SUS score obtained was 82.88 (SD 13.07, median 86.25), which would be considered *good/excellent* according to a validated adjective rating scale. For the UEQ, the means of the normalized scores (range −3 to +3) were as follows: attractiveness=2.683 (SD 0.100), perspicuity=2.775 (SD 0.150), efficiency=2.775 (SD 0.130), dependability=2.300 (SD 0.080), stimulation=1.950 (SD 0.120), and novelty=1.625 (SD 0.090). All dimensions were thus classed as *excellent* against the benchmarks, confirming the results from the SUS questionnaire. For the AIMS app, the mean SUS score obtained was 74.41 (SD 10.26), with a median of 77.50, which would be considered *good* according to the aforementioned adjective rating scale. For the UEQ, the means of the normalized scores were as follows: attractiveness=2.733 (SD 0.070), perspicuity=2.900 (SD 0.060), efficiency=2.800 (SD 0.090), dependability=2.425 (SD 0.060), stimulation=2.200 (SD 0.010), and novelty=1.450 (0.260). All dimensions were thus classed as *excellent* against the benchmarks (with the exception of novelty, which was classed as *good*), providing slightly better results than the SUS questionnaire.

Conclusions: The study has shown that both the AIMS clinical portal and the AIMS app have *good* to *excellent* usability scores, and the platform provides a solid foundation for the next phase of research, which will involve evaluating the effectiveness of the platform in improving patient outcomes after total knee replacement, total hip replacement, or dynamic hip screw.

(JMIR Hum Factors 2024;11:e50430) doi:10.2196/50430

KEYWORDS

mobile health; mHealth; digital health intervention; total knee replacement; TKR; total hip replacement; THR; dynamic hip screw; DHS; rehabilitation; usability; mobile phone

Introduction

Background

According to the World Health Organization's 2019 Global Burden of Disease study, approximately 1.71 billion people globally experience musculoskeletal conditions. Low back pain is the most common condition, affecting an estimated 568 million people [1]. In the United Kingdom, it has been estimated that musculoskeletal conditions affect >20 million people, approximately a third of the population [2]. Musculoskeletal conditions are the second greatest contributor to disability worldwide and is a significant burden to the individual and society [3]. It is expected that the impact of musculoskeletal conditions on the health service and on society will continue to rise as life expectancy increases [4]. Many different approaches have been explored to reduce this burden, including medical interventions, work-related approaches (reducing stress at work as well as improving health and safety regulations), social education (improving awareness of exercise and healthy eating), and the use of technology.

Musculoskeletal conditions comprise >150 different disorders, diseases, and syndromes that affect bones, joints, muscles, the spine, and soft tissues [3]. While some conditions are short lived, such as sprains and fractures, others can be lifelong conditions requiring ongoing treatment. Pain is a common symptom of musculoskeletal conditions. Back and neck pain, osteoarthritis, rheumatoid arthritis, and fractures are among the most disabling conditions and can be a significant barrier to healthy aging [5]. Musculoskeletal conditions can be classified by the body part affected (eg, knee pain and shoulder pain), whether the condition is noninflammatory (such as osteoarthritis) or inflammatory (such as rheumatoid arthritis), and whether the condition is restricted to the musculoskeletal system or more widespread (such as systemic lupus erythematosus) [4]. To compound matters, musculoskeletal issues tend to be associated with other diseases, such as heart or respiratory disease and stroke, and lead to an increase in disabilities and deaths [6-8]. It has been estimated that musculoskeletal conditions account for up to 21% of annual general practitioner consultations across England [9], and health service costs from inability to work and sickness absence in the United Kingdom are approximately £100 billion (US $125 billion) annually [10]. It is important to find solutions that will help reduce the significant burdens on the individual, society, the economy, and the health service. While many solutions will be of a medical nature, technology has a significant part to play in easing the burdens. In the next subsection, we discuss some digital health interventions (DHIs) for musculoskeletal conditions.

A number of different terminologies exist in the health domain for software solutions generally. The terms eHealth and mobile health (mHealth) have been used for a number of years. More recently, the more encompassing term *digital health* has been introduced. This is defined as "encompassing eHealth [which includes mHealth] as well as developing areas such as the use of advanced computing sciences (in the fields of 'big data,' genomics and artificial intelligence, for example)" [11]. Examples of digital health solutions include primary and secondary care IT systems; patient portals that provide secure web-based access to a range of health services, such as My Diabetes My Way and PatientView [12]; personal health data stores such as Mydex [13]; telehealth systems such as Attend Anywhere and Near Me [14]; and health-related mobile apps. It is believed that these systems can benefit health care delivery by improving different outcomes, such as effectiveness, efficiency, accessibility, safety, and personalization [15]. There has been a growing public interest in DHIs because they can allow individuals to monitor, manage, and improve their health and quality of life in a more personalized way, potentially more cost-effectively, and at a time that suits them [16-18].

Optimal rehabilitation programs for orthopedic joint replacement patients ensure faster return to function, earlier discharge from hospital, and improved patient satisfaction [4,19-21] as well as prevent further deconditioning [22]. The aim of this study was to evaluate the usability of a re-enablement platform called Active and Independent Management System (AIMS) that was developed to address the *de-enablement* that occurs in hospitalized inpatients for one of the groups considered to be most at risk, that is, older adults. The platform is capable of delivering digital rehabilitation plans and tracking the progression of the plans in real time; in addition, it can be used both in hospital and at home after a patient is discharged. The rationale for using such a system is to help reduce the time spent in hospital and improve patient satisfaction through self-management.

Re-Enablement DHIs

This subsection examines some recent literature related to the use of DHIs for total knee replacement (TKR) or total knee arthroplasty (TKA) and total hip replacement (THR) or total hip arthroplasty (THA). Hussain et al [23] developed a TKR platform comprising a mobile phone app, a wrist-worn activity tracker, and a clinical web portal. The purpose-built iOS and Android apps included weekly psychoeducation sessions and tasks that were delivered by a program guide via text and voice recordings. By obtaining the data from the tracker and the app, the clinician could monitor patient progress and the configured physiotherapy programs, while the patient care team could review the progress and the designated programs using the web portal. Physiotherapy programs were mostly from a library of videos created for TKR rehabilitation, which were made available in the app once set by the clinician. The authors planned to conduct a 13-month multisite unblinded randomized controlled trial in which participants were assigned to 1 of 2 study groups [23]. The participants for the experiment were patients who underwent TKR, and the study included an active intervention period from the time the patients were scheduled for surgery (approximately 4 weeks before surgery) to 12 weeks after the surgery, followed by a 40-week free-living period until 1 year after surgery.

Timmers et al [24] investigated the effect of a mobile app for day-to-day postoperative care education on TKR patients regarding the level of pain compared to those who only received standard information about their recovery through the app. The study involved 114 patients in the intervention group and 99 patients in the control group. In the intervention group, 93 patients downloaded and used the app. The results showed that, in comparison with standard patient education, the active education and coaching of patients on a day-to-day basis via the app in the 4 weeks after TKR resulted in a significant decrease, among other things, in the patients' levels of pain and a significant improvement in patients' physical functioning and quality of life, as well as their ability to perform physiotherapy exercises and activities of daily self-care.

Van Dijk-Huisman et al [25] developed a mobile app to prevent the negative effects of inactivity in hospital. The app supported objective activity monitoring, gave patients a view of their recovery progress, and offered a customized exercise program. The aim of the study was to investigate the potential of the app to enhance physical activity levels and functional recovery after orthopedic surgery discharge. In all, 97 patients undergoing TKA and THA were recruited for the evaluation. The control group (n=64) received standard physiotherapy, while the intervention group (n=33) used the mobile app in addition to physiotherapy. The time spent in active and functional recovery on postoperative day 1 (POD1) was measured. The app use, corrected for age, resulted in patients standing and walking on POD1 for an average increase of 28.43 (95% CI 5.55-51.32) minutes. The odds of achieving functional recovery on POD1 were 3.08 times higher (95% CI 1.14-8.31) with the use of the mobile app. The authors concluded that a mobile app combined with an accelerometer demonstrated the potential to enhance patients' activity levels and functional recovery during their hospital stay [25].

Wijnen et al [26] investigated the effectiveness of a home-based rehabilitation program using a tablet app and remote coaching for patients after THA. Existing data from 2 studies were combined: patients from a single-arm intervention study were matched with the historical controls from an observational study. Patients aged 18 to 65 years who had undergone THA were included. The intervention group had a 12-week home-based rehabilitation program with instructional videos on a tablet device and remote coaching. Patients were asked to perform strengthening and walking exercises at least 5 days a week. The intervention group was compared with a control group that included patients who received usual care. Effectiveness was measured at 4 points (preoperatively and 4 weeks, 12 weeks, and 6 months postoperatively) by means of functional tests and self-reported questionnaires. The intervention group performed functional tests significantly faster at 12 weeks and 6 months postoperatively and also scored significantly higher on the subscales *function in sport and recreational activities* and *hip-related quality of life* of the Hip Disability and Osteoarthritis Outcome Questionnaire, as well as on the subscale *physical role limitations* of the Short Form Health Survey-36 at 12 weeks and 6 months postoperatively. Large effect sizes were found on functional tests at 12 weeks and 6 months, endorsed by effect sizes on the self-reported outcomes. The authors concluded that the results demonstrated larger effects in the intervention group than in the historical controls, indicating that a home-based rehabilitation program using a mobile app after THA can be more effective than usual care [26].

Bell et al [27] ran a controlled pilot study for TKR patients, investigating the feasibility and effectiveness of interACTION, a remote (wearable) rehabilitation monitoring platform developed for use by patients after TKR. The InterACTION platform has portable motion sensors placed on either side of a joint to collect joint orientation data using a custom mobile app and then send the data to the clinician's web-based portal. The mobile app also contains 30 knee-specific home exercises for TKR rehabilitation that the physical therapist can personalize remotely through a web-based clinical portal. The study compared 2 groups: 19 patients who used the interACTION platform and a control group with 19 patients who used standard postoperative outpatient rehabilitation with a physical therapist (2-3 sessions per week over a maximum of 10 weeks), supplemented with a home exercise program. The primary outcome measured was value, operationally defined as the change in the activities of daily living scale of the Knee Outcome Survey at 10 weeks divided by the total cost of rehabilitation (determined from the total number of physical therapy sessions and the billable charges for each session during the 10 weeks the patients were enrolled in the study). In terms of this measure, no statistical differences were found between the groups. The study showed relatively low and not significant differences between the groups in terms of attrition rates, indicating that both interventions were acceptable. There was a small decrease in clinic visits by patients in the interACTION group, and all patients and physical therapists in the group indicated that they would use the system again.

Bäcker et al [28] developed a mobile app with a GenuSport sensor that allows isokinetic exercises to improve postoperative quadriceps weakness and knee motion. The sensor was placed underneath the patient's knee, and gamified exercise routines were presented through the app consisting of two exercises: (1) *high striker game*, where the patient has to push the knee onto the sensor for 5 seconds; and (2) *flight simulator*, where the player is supposed to keep the knee in the air for 100 seconds. The authors carried out a randomized controlled trial with a 2-year follow-up to evaluate the effectiveness of the app-based rehabilitation for patients after TKA [28]. In all, 35 patients completed the study and were randomly assigned to 2 groups: 20 patients received the app-based exercise program, and 15 patients were included in the control group. Patients in the app group used an external device to measure knee range of motion starting on the day of surgery, whereas patients in the control group underwent regular physiotherapy. Functional outcome scores using the Knee Injury and Osteoarthritis Outcome Score, the Knee Society Scoring System, and a visual analog scale for pain were analyzed. The results showed that, in the short term, the app group performed significantly better than the control group when taking a 10-minute walk, with less pain. In the longer term, the app group also performed significantly better, with higher Knee Society Scoring System scores as well as requiring fewer painkillers. In addition, the app group participants were more likely to participate in sports.

Colomina et al [29] developed an mHealth system for older patients with complex chronic conditions undergoing elective THA or TKA. The mHealth system formed part of the Personalized Connected Care for Complex Chronic Patients platform, which contained a web-based smart adaptive case management system for health care professionals that seamlessly integrated with a patient self-management mHealth system that supported communication between health care professionals and patients. The authors assessed the effectiveness and cost-effectiveness of implementing an mHealth-enabled integrated care (IC) model for patients with complex chronic conditions undergoing TKA or THA versus usual care [29]. A prospective pragmatic 2-arm parallel implementation trial was conducted in the rural region of Lleida in Catalonia, Spain, for 3 months. A total of 29 patients with complex chronic conditions undergoing TKA or THA and their caregivers received the IC program, while 30 patients with statistically comparable baseline characteristics, such as age, sex, and type of arthroplasty, were recruited for the usual care group. The results suggested that both treatment models significantly improved the physical and mental health status of the patients; however, IC significantly reduced the number of unplanned visits related to the surgery procedure and consequently significantly lowered the patients' expenses.

Rian et al [30] presented a web tool called Eir for symptom registration at home after knee arthroplasty. Given that the system was previously used in cancer care, a separate patient module was designed for patient-reported postoperative symptom assessment and medication registration after fast-track TKA that consisted of measurements of pain and side effects, as well as detailed registration of the use of analgesic drugs. The authors conducted a usability and feasibility study using a randomized controlled trial involving 134 participants [30]. The tool's usability was assessed with the use of the System Usability Scale (SUS) by 119 of the 134 participants, while the feasibility data were collected qualitatively. The results showed that 70% of the participants managed to use the tool at home without any technical support, although they indicated technical challenges related to the log-in procedure or internet access. The usability was rated high, with a mean SUS score of 89.6 (median 92.5; range 22.5-100).

Two literature reviews assessing the use of app-based rehabilitation for TKA or THA were conducted recently [31,32]. Bäcker et al [31] examined the functional outcomes of app-based rehabilitation of patients after TKA or THA. The review identified 420 entries from MEDLINE or PubMed and Google databases, but only 9 publications met the inclusion criteria, covering 518 patients in the intervention groups and 549 patients in the control groups. Five studies used app-based exercise instructions delivered via a mobile device, and 4 studies used a sensor or motion tracker. The average follow-up was 9.5 (SD 8.1; range 3-23.4) months. Overall, significantly lower activity visual analog scale values were observed for the interventional groups in the short term (P=.002). There were no other significant differences observed between the 2 groups. The study found that there were significant short-term improvements in the mobile app group. The authors concluded that mobile apps provide an alternative to in-person sessions that may improve access to physical activity for patients after TKA or THA, and, in combination with a Bluetooth-enabled sensor for isometric exercises, patients can additionally receive real-time feedback after TKA or THA [31].

Constantinescu et al [32] conducted a systematic literature review on the use of commercially available smartphone apps and wearable devices to assist rehabilitation interventions after TKA from the PubMed, Cochrane Library, MEDLINE, and Web of Science databases. Of the 60 full-text studies identified (published between January 2020 and September 2021), a total of 15 met the inclusion criteria, of which 4 studies used smartphone apps, 7 used wearable devices, and 4 used both to monitor physical activity and patient status after TKA. In terms of primary outcomes, 3 studies examined device accuracy, 3 recovery prediction, 2 functional recovery, 2 physical activity promotion, 2 patient compliance, 2 pain control, and 1 study examined health care use. The authors concluded that commercially available apps and wearable devices can capably monitor physical activity and improve patient engagement after TKA, making them approaches that support or replace traditional rehabilitation programs [32]. Using different strategies in interventions, such as setting step goals, using app-based patient engagement platforms, and establishing patient-specific benchmarks for recovery, can enhance the effectiveness of the treatment.

The AIMS Platform

It is well established that musculoskeletal conditions contribute to a large number of disabilities worldwide, and the projections show that this number will continue to rise. Initiatives are ongoing to combat this problem proactively (eg, reducing stress, improving health and safety regulations, exercising, and healthy eating). Furthermore, reactive approaches of optimal prehabilitation and rehabilitation programs are also undergoing development to optimize operations and ensure the best use of available resources while improving patient satisfaction. The aim of this study was to evaluate a rehabilitation platform in an effort to combat the lack of enablement in hospitalized older adults considered more vulnerable. The platform is capable of delivering digital rehabilitation plans and tracking the progression of these plans in real time; in addition, it can be used both in hospital and at home after patient discharge. The rationale for using such a system is to help reduce time spent in hospital and improve patient satisfaction through self-management.

The AIMS platform helps manage patients' rehabilitation programs. Each patient is registered by a clinician at the beginning of their patient pathway, and the system collects certain relevant information about the patient as they move through their journey. Rehabilitation clinicians use this platform to create, monitor, and adjust a patient's rehabilitation package as and when required. A team consisting of stakeholders is assigned to each patient and is responsible for the delivery of the program. A library of physiotherapy exercises and educational videos (eg, the use of a walking aid or how to apply a patient's splint) have been recorded and uploaded into the system (Figure 1). Each staff member can attach a series of video exercises specific to the patient's needs that can help the re-enablement process.

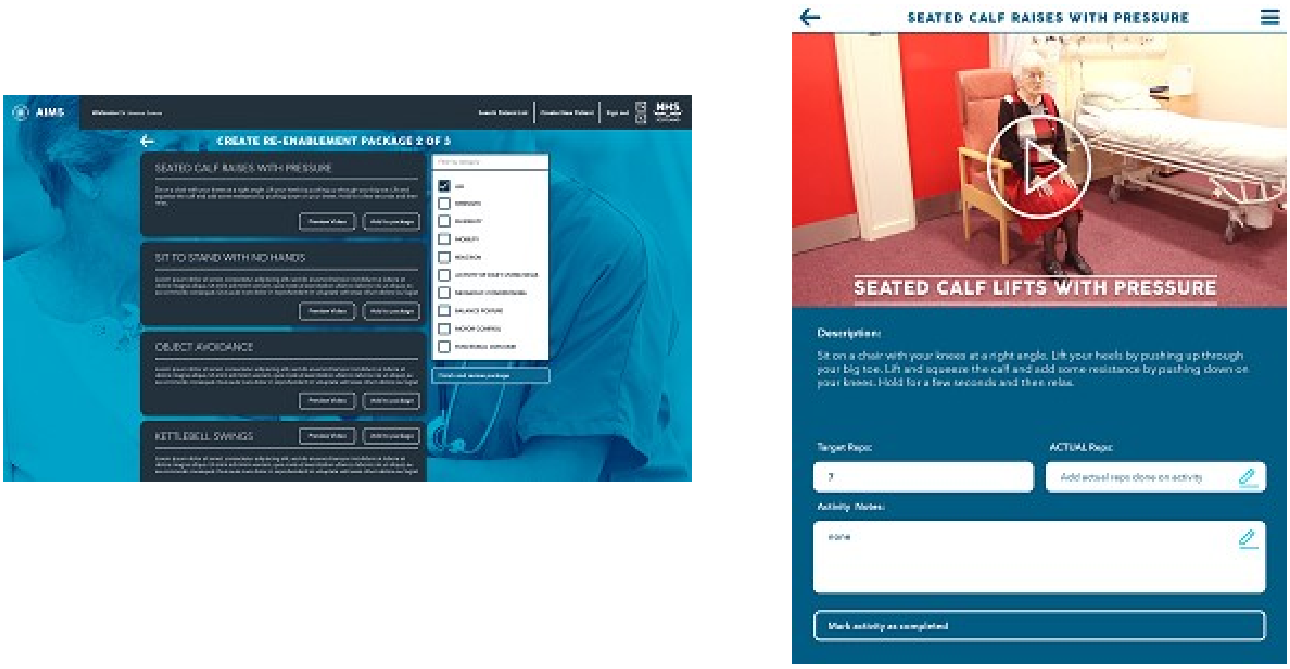
The platform consists of 2 components: a web content management system used by clinicians to create rehabilitation plans for postoperative patients and a mobile app designed to deliver these rehabilitation plans to the patients with a series of exercises to be completed by them.

The clinician starts by creating a user account for a patient (all patient information is anonymized with a random unique ID number that is later used to gain access to the rehabilitation plan through the app; Figure 1). The clinician then sets up a specific rehabilitation plan for this patient according to their needs. The clinician can search from all available exercises by using general search terms to filter what is available and also preview the associated video to make sure the appropriate plan is created (Figure 1A). The clinician must then determine the number of repetitions for each exercise and the frequency at which they need to be performed each day, usually 4 sessions a day. After the plan is complete, it becomes active on the patient's device, and the clinician then demonstrates to the patient how to use the app. This enables the clinician to monitor a patient and their progress each time they complete an exercise. There is also capability for the patient to comment on any particular issues with any of the exercises, and the clinician will be able to view the comments using the content management system and modify the plan accordingly. A typical example of a rehabilitation plan for a patient consists of some simple but very effective exercises (eg, heel slides, knee extensions, and knee flexions). The patient would be asked to perform 10 repetitions of all exercises in 4 daily sessions. The rehabilitation process starts after the operation for as long as the patient remains in hospital, and there are physiotherapists available to offer assistance during the patient rehabilitation process; the app does not prompt patients to complete their daily rehabilitation plan because these sessions are already scheduled in the hospital ward. It is up to the patient to continue using the app for rehabilitation after hospital discharge (the app is available for free download from app stores).

The patient uses a tablet device provided by the hospital to gain access to the AIMS mobile app and work on their rehabilitation plan. Each user is given a random ID number generated by the clinician that is required to log in to the app; no password is required because all information is anonymized. After this, the user can use the app and work on their specific rehabilitation plan and set of exercises and also view their daily plan progress. The patient is provided a textual description of the exercise and a video with audio explaining how it should be performed and how many repetitions should be performed (Figure 1B). At the bottom of the page featuring each exercise, feedback can be provided on how many repetitions were achieved as well as any comments if there were any issues when performing the exercise.

Typically, the app would be used by a member of the staff or a member of the family during visiting hours to help the patient with their exercises by encouraging them or participating with them and achieving successful completion of the rehabilitation plan.

Figure 1. (A) The Create Rehabilitation Re-Enablement Package Screen, and (B) the patient exercise screen.



A

B

Aims of the Study

The aim of this study was to investigate the usability of the AIMS platform from the perspectives of both clinicians and patients. Two well-known instruments were used to measure usability: the SUS [33] with 10 items and, for finer granularity, the User Experience Questionnaire (UEQ) [34] with 26 items.

The evaluation aims to answer the following 2 research questions (RQs):

- RQ1: does the AIMS clinical portal provide a solution that could be usable by clinicians?
- RQ2: does the AIMS app provide a solution that could be usable by patients?

Methods

Overview

The World Health Organization defines evaluation as "the systematic and objective assessment of an ongoing or completed project [with the aim of determining] the relevance and fulfilment of objectives, development efficiency, effectiveness, impact and sustainability" [35]; and the guide for monitoring and evaluating DHIs outlines 7 stages of DHI maturity, ranging from preprototype to full deployment. This project is considered to be at the prototype stage of maturity, which would include usability testing. Ways to improve the system would also be investigated.

Usability is recognized as a significant quality indicator that determines the success of software applications [36-39]. Johnson et al [40] defines three main approaches to evaluate usability: (1) user based (a sample of prospective users use the system), (2) expert based (≥1 usability or human-computer interaction experts evaluate the system), and (3) model based (formal methods are used to predict user performance). Our health board members were keen on using the user-based approach to evaluate the DHI; hence, this approach was chosen.

Many validated usability instruments have been proposed in the literature with varying numbers of questions. In this study, we used 2 well-known validated instruments: the SUS and the UEQ. The SUS [33] consists of 10 statements (5 positive and 5 negative) that the users rate on a scale ranging from 1=*strongly disagree* to 5=*strongly agree*. The questionnaire alternates between positive and negative statements to avoid random answers. The aggregated score out of 100 can be compared with the average SUS benchmark score of 68.0. To represent SUS scores, Bangor et al [41] defined a 7-point adjective rating scale: *best imaginable*, *excellent*, *good*, *OK*, *poor*, *awful*, and *worst imaginable*.

The UEQ assesses the extent to which (1) the product meets expectations and (2) a product can be compared with other systems using a published benchmark. Schrepp et al [42] developed an adjective rating scale for benchmarking, and a mean score of >1.75 would be considered in the 10% best results. While the UEQ provides finer detail than the SUS, it was felt that asking busy clinicians to rate 26 statements may result in a smaller number of responses compared to asking them to rate 10 SUS statements; therefore, it was decided to use the SUS with all participants and the UEQ with a small number of participants.

Some qualitative information was also gathered using open-ended questions to gain a deeper understanding of participants' views of the AIMS platform.

Ethical Considerations

Ethics approval for this study was obtained from Hairmyres University Hospital, Lanarkshire. One of the conditions of approval was that all personal information from the study should be removed and that patient information should be kept private and safe (Data Protection Impact Assessment Questionnaire for Active Independent Mobility System [AIMS] Pilot Study Hairmyres University Hospital, Lanarkshire; May 29, 2019). All participants provided consent before participating in the study.

Participants

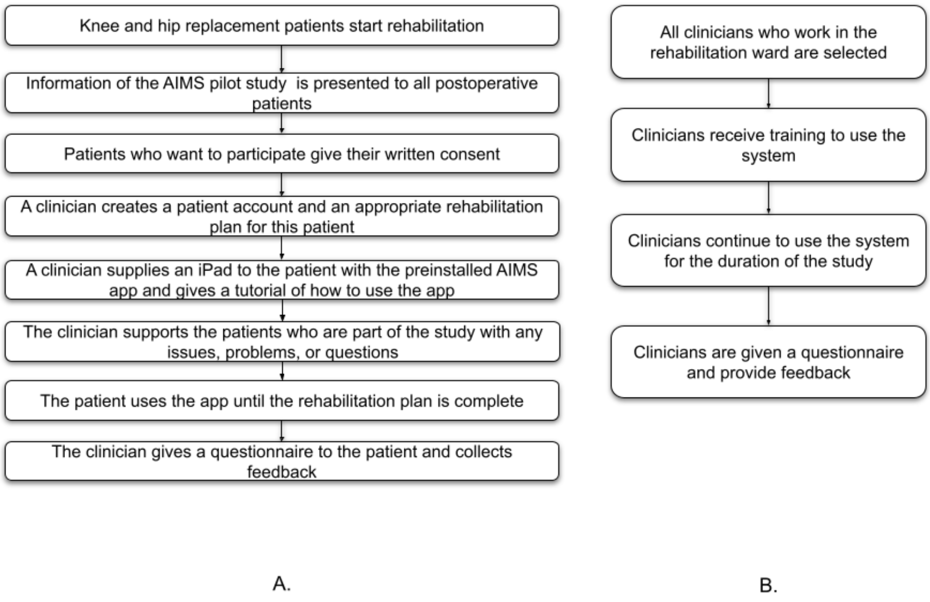
In all, 26 physiotherapists and health care professionals volunteered to evaluate the AIMS clinical portal; and 44 patients in hospital for TKR, THR, or dynamic hip screw (DHS) agreed to participate in evaluating the AIMS app. The study was carried out on May 5, 2019, or November 14, 2019.

Test Protocol

The 6-month-long study was undertaken in the rehabilitation ward in an Hairmyres University Hospital, Lanarkshire, that specializes in TKR or THR surgery. Only clinicians had access to the patients during their stay in the hospital, and a member of the team of clinicians (lead study clinician) had oversight over recruiting and running the experiment. Technical support was provided by the study team to all clinicians during the rehabilitation sessions in the hospital, but this was not in the rehabilitation ward. Questionnaires were given to the participants in a paper-based form, which they were asked to complete and hand back to the lead study clinician toward the end of their rehabilitation stay. The study was conducted using 10 hospital-supplied second-generation iPad Air 2 devices running iOS 10.3 with a 9.7-inch display in portrait orientation (refer to the patient test protocol presented in Figure 2A).

Clinicians of the rehabilitation team in the hospital were all given training on how to use the portal to create rehabilitation packages for patients and how they would look in the app. They were also involved in the development and design process with focus groups and early prototyping, which enabled most of them to develop a good understanding of the AIMS platform. As the study was taking place alongside patients who were not part of the study, everyone had to be able to help the patients, which is why they were all trained to use the system. Clinicians were given a questionnaire to complete after they had used the platform a few times (refer to the clinician test protocol presented in Figure 2B).

Figure 2. (A) Patient test protocol. (B) Clinician test protocol. AIMS: Active and Independent Management System.



Results

For the SUS, the analysis was carried out using Excel (Microsoft Corp); and for the UEQ, the analysis was carried out using the standard UEQ spreadsheet.

AIMS Clinical Portal: SUS Results

All 26 participants completed the SUS questionnaire (100% response rate). The mean SUS score obtained was 82.88 (SD 13.07), with a median of 86.25. This score would be considered *good/excellent* according to the adjective rating scale developed by Bangor et al [41]. A breakdown of the participants' answers to the SUS questions regarding the AIMS clinical portal is provided in Table 1.

Table 1. Participants' answers to the System Usability Scale (SUS) questions for the Active and Independent Management System clinical portal (n=26).

Statements	Participants agreeing, n (%)	Participants disagreeing, n (%)	SUS scores, mean (SD)
Positive statements			
I think that I would like to use this system frequently	23 (88)	2 (8)	4.00 (0.76)
I thought the system was easy to use	24 (92)	1 (4)	4.50 (0.77)
I found the various functions in this system were well integrated	22 (85)	2 (8)	4.50 (0.96)
I would imagine that most people would learn to use this system very quickly	23 (88)	1 (4)	4.19 (0.65)
I felt very confident using this system	24 (92)	1 (4)	4.38 (1.42)
Negative statements			
I found the system unnecessarily complex	2 (8)	23 (88)	2.04 (0.73)
I think that I would need the support of a technical person to be able to use this system	1 (4)	24 (92)	1.50 (0.77)
I thought there was too much inconsistency in this system	0 (0)	23 (88)	1.62 (0.86)
I found the system very cumbersome to use	3 (88)	21 (81)	1.73 (1.05)
I needed to learn a lot of things before I could get going with this system	0 (0)	24 (92)	1.42 (0.65)

AIMS Clinical Portal: UEQ Results

Invitations were sent to 12 (46%) of the 26 participants. Of these 12 participants, 10 (83%) completed the UEQ questionnaire. The means of the normalized scores (range −3 to +3) for the AIMS clinical portal were as follows: attractiveness=2.683 (SD 0.100), perspicuity=2.775 (SD 0.150), efficiency=2.775 (SD 0.130), dependability=2.300 (SD 0.080), stimulation=1.950 (SD 0.120), and novelty=1.625 (SD 0.090). Figure 3 shows the bar chart of the results for the AIMS clinical portal against the benchmarks, showing all dimensions classed as *excellent* and confirming the results from the SUS questionnaire. Table 2 provides the mean (SD) and variance of the normalized values for the items in the UEQ questionnaire for the AIMS clinical portal. In most cases, the values are very encouraging, with the exception of *conservative* and *innovative*, although this is still rated *good*. Figure 4 shows the bar chart of the data grouped into the 6 UEQ dimensions.

Figure 3. Bar chart of the Active and Independent Management System clinical portal User Experience Questionnaire results against the benchmarks.

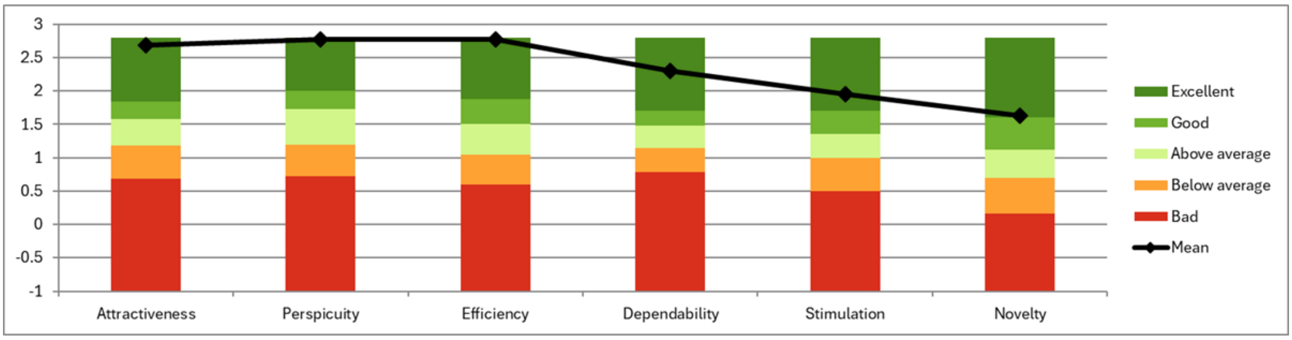
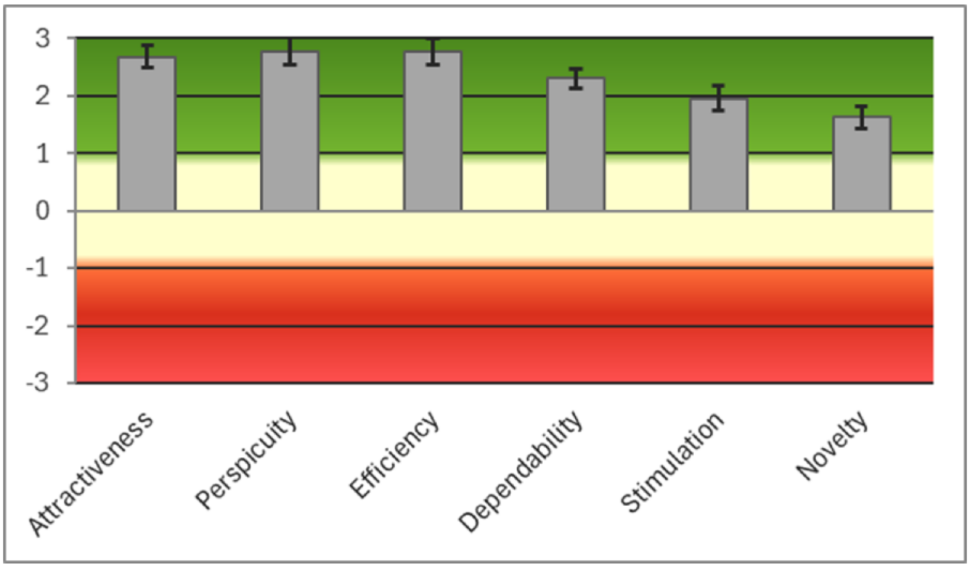


Table 2. Mean (SD) and variance of the normalized values for the items in the User Experience Questionnaire (UEQ) for the Active and Independent Management System clinical portal (n=10).

Scale	Left anchor of the scale	Right anchor of the scale	UEQ scores, mean (SD)	Variance
Attractiveness	Annoying	Enjoyable	1.9 (0.3)	0.1
Perspicuity	Not understandable	Understandable	2.6 (0.7)	0.5
Novelty	Creative	Dull	1.9 (0.3)	0.1
Perspicuity	Easy to learn	Difficult to learn	2.9 (0.3)	0.1
Stimulation	Valuable	Inferior	2.2 (0.6)	0.4
Stimulation	Boring	Exciting	1.5 (0.5)	0.3
Stimulation	Not interesting	Interesting	2.1 (0.6)	0.3
Dependability	Unpredictable	Predictable	1.8 (0.4)	0.2
Efficiency	Fast	Slow	2.8 (0.4)	0.2
Novelty	Inventive	Conventional	1.9 (0.3)	0.1
Dependability	Obstructive	Supportive	2.2 (0.6)	0.4
Attractiveness	Good	Bad	2.8 (0.4)	0.2
Perspicuity	Complicated	Easy	2.7 (0.5)	0.2
Attractiveness	Unlikable	Pleasing	2.7 (0.5)	0.2
Novelty	Usual	Leading edge	1.2 (0.4)	0.2
Attractiveness	Unpleasant	Pleasant	2.9 (0.3)	0.1
Dependability	Secure	Not secure	2.5 (0.5)	0.3
Stimulation	Motivating	Demotivating	2.0 (0.5)	0.2
Dependability	Meets expectations	Does not meet expectations	2.7 (0.5)	0.2
Efficiency	Inefficient	Efficient	2.8 (0.4)	0.2
Perspicuity	Clear	Confusing	2.9 (0.3)	0.1
Efficiency	Impractical	Practical	2.7 (0.5)	0.2
Efficiency	Organized	Cluttered	2.8 (0.4)	0.2
Attractiveness	Attractive	Unattractive	2.9 (0.3)	0.1
Attractiveness	Friendly	Unfriendly	2.9 (0.3)	0.1
Novelty	Conservative	Innovative	1.5 (0.5)	0.3

Figure 4. Bar chart of the Active and Independent Management System clinical portal data grouped into the 6 User Experience Questionnaire dimensions.



AIMS Clinical Portal: Qualitative Feedback

To gain further insight into how users perceived the AIMS clinical portal, 3 additional questions were asked (refer to the following subsections).

Q1: What Do You Think Are the Advantages of This Portal?

Of the 26 participants, 20 (77%) answered this question. All clinicians (20/20, 100%) who answered the question thought that providing customized exercise videos after an operation was very useful, particularly for patients being able to use the system at home; 15 (75%) of the 20 clinicians also suggested that receiving immediate feedback on how patients were coping with the exercise regime was very helpful and meant that the regime could be easily customized for each patient based on how they were coping, which was a key advantage. In addition, 60% (12/20) of the clinicians considered ease of use an advantage. Example comments were as follows:

> *Really liked the exercise videos for the patients; they were professionally produced and highly relevant for rehabilitation.* [Physiotherapist A]

> *I'm pleased to see that patients automatically receive some feedback on how they are progressing with the rehabilitation exercises.* [Physiotherapist B]

Q2: What Do You Think Are the Disadvantages of This Portal?

Of the 26 participants, 12 (46%) answered this question. Of these 12 clinicians, 7 (58%) thought that the integration of the portal with the current IT systems may be a challenge, 3 (25%) thought that getting the staff to agree to use the portal may be a possible issue, and 2 (17%) thought that some staff members would need training on how to use it. Example comments were as follows:

> *One big issue that will have to be addressed at some point is integrating the software with hospital systems, as we ultimately need to have the patient progress data in their EHR [electronic health record].* [Physiotherapist B]

> *While the current system was intuitive and easy to use, I wonder whether some training will need to be provided when the features to add further videos and provide more customised feedback are added.* [Physiotherapist C, an academic]

Q3: Would You Change Anything?

Of the 26 participants, 17 (65%) answered this question. Of these 17 clinicians, 12 (71%) suggested that the ability to create more self-help advice for patients would be useful, and 5 (29%) suggested that having a larger data bank of exercise regimes would be helpful. Example comments were as follows:

> *It would be very helpful if more self-help could be added to the app to reduce the dependency on the volume of information sheets we provide to patients.* [Physiotherapist D]

> *The current set of videos are very relevant and of a high quality; however, it would be beneficial to be able to have a wider selection of videos to be able [to] select from.* [Physiotherapist E]

AIMS App: SUS Results

The participants were selected during their first postoperative rehabilitation session (opportunistic recruitment). The recruitment of patients was carried out by a physiotherapist who would ask patients during their first session whether they were willing to participate in the study. The physiotherapist provided an information leaflet that explained what the study was about and how it could be used. All postoperative patients automatically qualified for the study; no one was excluded based on age, sex, or technical competency. The study did not collect any age- or sex-related information (a condition of the ethics approval for the study); therefore, it was not possible to provide information about patient demographics.

Of the 44 patients, 38 (86%) completed the SUS questionnaire. The mean SUS score obtained was 74.41 (SD 10.26), with a median of 77.50. This score would be considered *good* according to the adjective rating scale developed by Bangor et al [41]. A breakdown of the participants' answers to the SUS questions for the AIMS app is provided in Table 3.

Table 3. Participants' answers to the System Usability Scale (SUS) questions for the Active and Independent Management System app (n=38).

Statements	Participants agreeing, n (%)	Participants disagreeing, n (%)	SUS scores, mean (SD)
Positive statements			
I think that I would like to use this system frequently	34 (89)	1 (3)	4.16 (0.68)
I thought the system was easy to use	31 (82)	1 (3)	4.13 (0.62)
I found the various functions in this system were well integrated	34 (89)	1 (3)	4.16 (0.68)
I would imagine that most people would learn to use this system very quickly	32 (84)	2 (5)	3.92 (0.67)
I felt very confident using this system	33 (87)	2 (5)	3.97 (0.68)
Negative statements			
I found the system unnecessarily complex	1 (3)	32 (84)	2.00 (0.66)
I think that I would need the support of a technical person to be able to use this system	1 (3)	32 (84)	1.92 (0.71)
I thought there was too much inconsistency in this system	1 (3)	32 (84)	2.11 (0.56)
I found the system very cumbersome to use	1 (3)	28 (74)	2.24 (0.59)
I needed to learn a lot of things before I could get going with this system	2 (5)	26 (68)	2.32 (0.66)

AIMS App: UEQ Results

Invitations were sent to 12 (27%) of the 44 participants. Of these 12 patients, 10 (83%) completed the UEQ questionnaire. The means of the normalized scores (range −3 to +3) for the AIMS app were as follows: attractiveness=2.733 (SD 0.070), perspicuity=2.900 (SD 0.060), efficiency=2.800 (SD 0.090), dependability=2.425 (SD 0.060), stimulation=2.200 (SD 0.010), and novelty=1.450 (0.260). Figure 5 shows the bar chart of the results for the AIMS app against the benchmarks, with all dimensions classed as *excellent* (with the exception of *novelty*, which was classed as *good*), providing slightly better results than the SUS questionnaire. Table 4 gives the mean (SD) and variance of the normalized values for the items in the UEQ questionnaire for the AIMS app. In this case, all values are very encouraging. Figure 6 shows the bar chart for the data grouped into the 6 UEQ dimensions.

Figure 5. Bar chart of the Active and Independent Management System app User Experience Questionnaire results against the benchmarks.

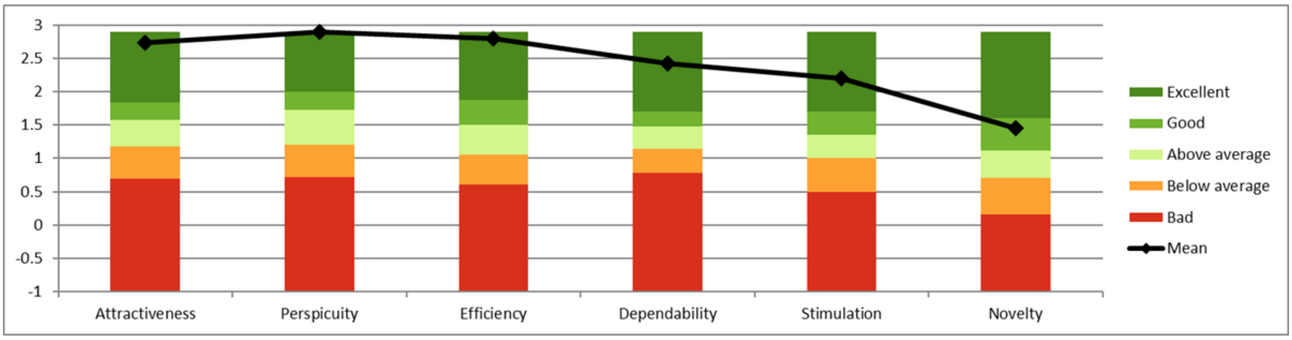
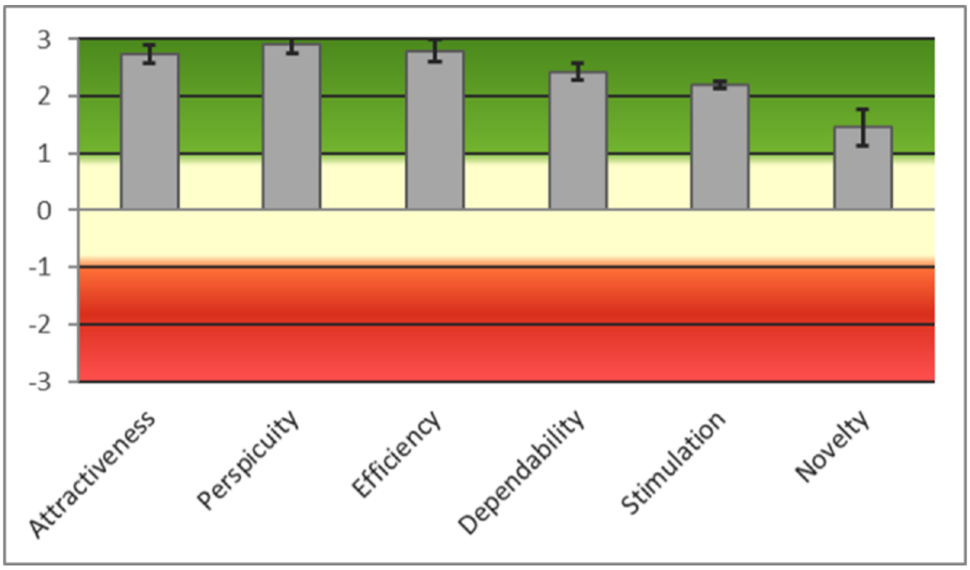


Table 4. Mean (SD) and variance of the normalized values for the items in the User Experience Questionnaire (UEQ) for the Active and Independent Management System app (n=10).

Scale	Left anchor of the scale	Right anchor of the scale	UEQ scores, mean (SD)	Variance
Attractiveness	Annoying	Enjoyable	1.9 (0.3)	0.1
Perspicuity	Not understandable	Understandable	2.6 (0.7)	0.5
Novelty	Creative	Dull	1.9 (0.3)	0.1
Perspicuity	Easy to learn	Difficult to learn	2.9 (0.3)	0.1
Stimulation	Valuable	Inferior	2.2 (0.6)	0.4
Stimulation	Boring	Exciting	1.5 (0.5)	0.3
Stimulation	Not interesting	Interesting	2.1 (0.6)	0.3
Dependability	Unpredictable	Predictable	1.8 (0.4)	0.2
Efficiency	Fast	Slow	2.8 (0.4)	0.2
Novelty	Inventive	Conventional	1.9 (0.3)	0.1
Dependability	Obstructive	Supportive	2.2 (0.6)	0.4
Attractiveness	Good	Bad	2.8 (0.4)	0.2
Perspicuity	Complicated	Easy	2.7 (0.5)	0.2
Attractiveness	Unlikable	Pleasing	2.7 (0.5)	0.2
Novelty	Usual	Leading edge	1.2 (0.4)	0.2
Attractiveness	Unpleasant	Pleasant	2.9 (0.3)	0.1
Dependability	Secure	Not secure	2.5 (0.5)	0.3
Stimulation	Motivating	Demotivating	2.0 (0.5)	0.2
Dependability	Meets expectations	Does not meet expectations	2.7 (0.5)	0.2
Efficiency	Inefficient	Efficient	2.8 (0.4)	0.2
Perspicuity	Clear	Confusing	2.9 (0.3)	0.1
Efficiency	Impractical	Practical	2.7 (0.5)	0.2
Efficiency	Organized	Cluttered	2.8 (0.4)	0.2
Attractiveness	Attractive	Unattractive	2.9 (0.3)	0.1
Attractiveness	Friendly	Unfriendly	2.9 (0.3)	0.1
Novelty	Conservative	Innovative	1.5 (0.5)	0.3

Figure 6. Bar chart of the Active and Independent Management System app data grouped into the 6 User Experience Questionnaire dimensions.



AIMS App: Qualitative Feedback

To gain further insight into how users perceived the AIMS app, 3 additional questions were asked (refer to the following subsections).

Q1. What Do You Think Are the Advantages of This App?

Most of the participants (33/44, 75%) answered this question. Of the 33 participants, 26 (79%) thought that the exercise videos provided after an operation were very useful, 30 (91%) considered clinicians having immediate access to patient progress an advantage, and 24 (73%) considered ease of use an advantage. Example comments were as follows:

> *Having exercise videos that I can use both in the hospital and at home is a great help. While there is help on hand in the hospital if needed, being able to view the videos while at home is great.* [Patient A]
>
> *Loved the being able to access the videos on the tablet, was very helpful and the app was so easy to use.* [Patient B]

Q2. What Do You Think Are the Disadvantages of This App?

Only 8 (18%) of the 44 participants answered this question, and very few disadvantages were listed: 1 (13%) participant thought that the app could include some embedded videos for generic stretching exercises; 1 (13%) thought that the app might be too simple, and more functionality was required; and 6 (75%) thought that a self-help section would be beneficial. An example comment was as follows:

> *While the hospital provide[s] a number of leaflets on what to expect after the knee replacement, it would be handier of [sic] these were part of the app.* [Patient C]

Q3. Would You Change Anything?

Of the 44 participants, 17 (39%) answered this question. Of these 17 participants, 6 (35%) suggested more self-help, and 5 (29%) suggested having the ability to keep a daily or weekly diary of symptoms or pain. An example comment was as follows:

> *Would it be possible to have a section in the app to record how I am getting on with the videos and make notes on any symptoms I'm getting after the operation, particularly once I'm home?* [Patient D]

Discussion

Principal Findings

The main goal of this mixed methods study was to examine the usability of the AIMS platform from the perspectives of both patients and clinicians. Two well-known validated instruments were used to measure usability: the SUS and the UEQ. In all, 26 physiotherapists and health care professionals evaluated the AIMS clinical portal; and 44 patients in hospital for TKR, THR, or DHS evaluated the AIMS app. In terms of the RQs, the study has shown that both the AIMS clinical portal (RQ1) and the AIMS app (RQ2) have *good* to *excellent* usability scores, and this platform provides a solid foundation for the next phase of research, which will involve evaluating its effectiveness in improving patient outcomes after TKR, THR, or DHS. In addition, useful qualitative information was obtained from participants through a set of open-ended questions.

On the basis of the literature reviewed in the Re-Enablement DHIs subsection, it seems that smartphones and the web are the 2 main platforms used to provide re-enablement DHIs after TKA or THA. The platforms have been identified to be used by patients who will receive the instructions in the form of video, text, and interactive game as well as by clinicians who can create custom treatment plans for patients. The AIMS platform provides similar functionality to the systems found in the literature, with a web-based clinical portal and a mobile app for patients. The AIMS platform mainly presents content in video and text, which is similar to the majority of the systems discussed in the Re-Enablement DHIs subsection. Text and video are considered to be effective in presenting rehabilitation content to patients because they allow a wider level of proficiency in information and communications technology. Compared to static images, we considered videos to be more engaging, although further research should be conducted to investigate this. Some studies, such as those by Hussain et al [23], van Dijk-Huisman et al [25], and Bell et al [27], used sensors from wearable devices and mobile phones, while Bäcker et al [28] developed their own custom sensor. Personalization features that allow the system to customize activities for patients were only evident in the studies by Hussain et al [23], van Dijk-Huisman et al [25], and Bell et al [27]. Currently, the AIMS platform does not use sensors or have any personalization features, but these will be considered for the next phase of the research. A summary comparing the literature reviewed with our study can be found in Multimedia Appendix 1 [23-30].

In terms of limitations, to overcome major privacy concerns, a condition of the ethics approval for the study was that all data had to be anonymized; therefore, neither could we perform a demographic analysis nor conduct follow-up monitoring of the progress of a more informed patient after they left the hospital. As this study's main focus was on usability, the recruitment of participating patients was carried out during rehabilitation sessions. This method did not allow us to conduct a randomized study, and there were no control and experimental groups. Furthermore, due to ethics approval restrictions, we were not able to directly observe the experiment and had to use questionnaires and interviews conducted by the clinicians during the rehabilitation sessions. The experiment did not use any additional sensing technologies to monitor user progress and relied on the patient's input and feedback. Future studies will aim to overcome these limitations.

Since this study was carried out, the platform has been improved to include additional support videos for patients, ideas for which emerged from the qualitative feedback, and a second usability study is underway to ensure that results are consistent with this initial study (Multimedia Appendix 1).

Conclusions

This study aimed to assess the usability of a re-enablement platform called AIMS, designed to address the *de-enablement* often experienced by hospitalized older adults most at risk. Usability was measured using 2 common validated instruments: the 10-item SUS and, for more detailed analysis, the 26-item UEQ. The AIMS clinical portal was evaluated by 26 physiotherapists and health care professionals; and 44 patients undergoing TKR, THR, or DHS assessed the AIMS app. Overall, both the AIMS clinical portal and the AIMS app received *good* to *excellent* usability scores, providing a solid foundation for future research on their effectiveness in improving patient outcomes after joint replacements. Optimal rehabilitation programs for orthopedic joint replacement patients can lead to a quicker return to normal function, faster hospital discharge, and higher patient satisfaction.

Acknowledgments

The authors acknowledge the resources allocated to this study by Hairmyres University Hospital, Lanarkshire.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Comparison with the literature reviewed.

[XLSX File (Microsoft Excel File), 6 KB - humanfactors_v11i1e50430_app1.xlsx]

References

1. Mackey S, Gilam G, Darnall B, Goldin P, Kong JT, Law C, et al. Mindfulness-based stress reduction, cognitive behavioral therapy, and acupuncture in chronic low back pain: protocol for two linked randomized controlled trials. JMIR Res Protoc 2022 Sep 27;11(9):e37823 [FREE Full text] [doi: 10.2196/37823] [Medline: 36166279]
2. The state of musculoskeletal health. Versus Arthritis. URL: https://www.versusarthritis.org/about-arthritis/data-and-statistics/the-state-of-musculoskeletal-health/ [accessed 2022-11-03]
3. Hewitt S, Sephton R, Yeowell G. The effectiveness of digital health interventions in the management of musculoskeletal conditions: systematic literature review. J Med Internet Res 2020 Jun 05;22(6):e15617 [FREE Full text] [doi: 10.2196/15617] [Medline: 32501277]
4. Nagel J, Wegener F, Grim C, Hoppe MW. Effects of digital physical health exercises on musculoskeletal diseases: systematic review with best-evidence synthesis. JMIR Mhealth Uhealth 2024 Jan 23;12:e50616 [FREE Full text] [doi: 10.2196/50616] [Medline: 38261356]
5. Briggs AM, Woolf AD, Dreinhöfer K, Homb N, Hoy DG, Kopansky-Giles D, et al. Reducing the global burden of musculoskeletal conditions. Bull World Health Organ 2018 Apr 12;96(5):366-368. [doi: 10.2471/blt.17.204891]
6. Duffield SJ, Ellis BM, Goodson N, Walker-Bone K, Conaghan PG, Margham T, et al. The contribution of musculoskeletal disorders in multimorbidity: implications for practice and policy. Best Pract Res Clin Rheumatol 2017 Apr;31(2):129-144 [FREE Full text] [doi: 10.1016/j.berh.2017.09.004] [Medline: 29224692]
7. Ross EL, Jamison RN, Nicholls L, Perry BM, Nolen KD. Clinical integration of a smartphone app for patients with chronic pain: retrospective analysis of predictors of benefits and patient engagement between clinic visits. J Med Internet Res 2020 Apr 16;22(4):e16939 [FREE Full text] [doi: 10.2196/16939] [Medline: 32297871]
8. Williams A, Kamper SJ, Wiggers JH, O'Brien KM, Lee H, Wolfenden L, et al. Musculoskeletal conditions may increase the risk of chronic disease: a systematic review and meta-analysis of cohort studies. BMC Med 2018 Sep 25;16(1):167 [FREE Full text] [doi: 10.1186/s12916-018-1151-2] [Medline: 30249247]
9. Keavy R. The prevalence of musculoskeletal presentations in general practice: an epidemiological study. Br J Gen Pract 2020 Jun 18;70(suppl 1):bjgp20X711497. [doi: 10.3399/bjgp20x711497]
10. Health matters: health and work. United Kingdom Government. URL: https://www.gov.uk/government/publications/health-matters-health-and-work/health-matters-health-and-work [accessed 2022-11-02]
11. mHealth: use of appropriate digital technologies for public health: report by the Director-General. World Health Organization. 2017. URL: https://iris.who.int/handle/10665/274134 [accessed 2024-04-16]
12. Cunningham SG, Wake DJ, Waller A, Morris AD, Walker J. My diabetes my way: an electronic personal health record for diabetes. Diabetes Vasc Dis 2013 Jun 26;13(3):143-149. [doi: 10.1177/1474651413493336]
13. Janssen H, Cobbe J, Norval C, Singh J. Decentralized data processing: personal data stores and the GDPR. Int Data Privacy Law 2020 Nov;10(4):356-384 [FREE Full text] [doi: 10.1093/idpl/ipaa016]
14. Wherton J, Greenhalgh T, Shaw SE. Expanding video consultation services at pace and scale in Scotland during the COVID-19 pandemic: national mixed methods case study. J Med Internet Res 2021 Oct 07;23(10):e31374 [FREE Full text] [doi: 10.2196/31374] [Medline: 34516389]

15. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
16. Garell C, Svedberg P, Nygren JM. A legal framework to support development and assessment of digital health services. JMIR Med Inform 2016 May 25;4(2):e17 [FREE Full text] [doi: 10.2196/medinform.5401] [Medline: 27226391]
17. Classification of digital health interventions v1.0: a shared language to describe the uses of digital technology for health. World Health Organization. URL: https://apps.who.int/iris/bitstream/handle/10665/260480/WHO-RHR-18.06-eng.pdf [accessed 2022-12-03]
18. Saleem M, Kühne L, De Santis KK, Christianson L, Brand T, Busse H. Understanding engagement strategies in digital interventions for mental health promotion: scoping review. JMIR Ment Health 2021 Dec 20;8(12):e30000 [FREE Full text] [doi: 10.2196/30000] [Medline: 34931995]
19. Hardy A, Courgeon M, Pellei K, Desmeules F, Loubert C, Vendittoli PA. Improved clinical outcomes of outpatient enhanced recovery hip and knee replacements in comparison to standard inpatient procedures: a study of patients who experienced both. Orthop Traumatol Surg Res 2022 Oct;108(6):103236 [FREE Full text] [doi: 10.1016/j.otsr.2022.103236] [Medline: 35150928]
20. Zhong M, Liu D, Tang H, Zheng Y, Bai Y, Liang Q, et al. Impacts of the perioperative fast track surgery concept on the physical and psychological rehabilitation of total hip arthroplasty: a prospective cohort study of 348 patients. Medicine (Baltimore) 2021 Aug 13;100(32):e26869 [FREE Full text] [doi: 10.1097/MD.0000000000026869] [Medline: 34397901]
21. Galbraith AS, McGloughlin E, Cashman J. Enhanced recovery protocols in total joint arthroplasty: a review of the literature and their implementation. Ir J Med Sci 2018 Feb 16;187(1):97-109. [doi: 10.1007/s11845-017-1641-9] [Medline: 28623570]
22. Swinnerton E, Price A. Recognising, reducing and preventing deconditioning in hospitalised older people. Nurs Older People 2023 Apr 03;35(2):34-41. [doi: 10.7748/nop.2023.e1396] [Medline: 36751897]
23. Hussain MS, Li J, Brindal E, van Kasteren Y, Varnfield M, Reeson A, et al. Supporting the delivery of total knee replacements care for both patients and their clinicians with a mobile app and web-based tool: randomized controlled trial protocol. JMIR Res Protoc 2017 Mar 01;6(3):e32 [FREE Full text] [doi: 10.2196/resprot.6498] [Medline: 28249832]
24. Timmers T, Janssen L, van der Weegen WJ, Das D, Marijnissen W, Hannink G, et al. The effect of an app for day-to-day postoperative care education on patients with total knee replacement: randomized controlled trial. JMIR Mhealth Uhealth 2019 Oct 21;7(10):e15323 [FREE Full text] [doi: 10.2196/15323] [Medline: 31638594]
25. van Dijk-Huisman HC, Weemaes AT, Boymans TA, Lenssen AF, de Bie RA. Smartphone app with an accelerometer enhances patients' physical activity following elective orthopedic surgery: a pilot study. Sensors (Basel) 2020 Aug 02;20(15):4317 [FREE Full text] [doi: 10.3390/s20154317] [Medline: 32748876]
26. Wijnen A, Hoogland J, Munsterman T, Gerritsma CL, Dijkstra B, Zijlstra WP, et al. Effectiveness of a home-based rehabilitation program after total hip arthroplasty driven by a tablet app and remote coaching: nonrandomized controlled trial combining a single-arm intervention cohort with historical controls. JMIR Rehabil Assist Technol 2020 Apr 27;7(1):e14139 [FREE Full text] [doi: 10.2196/14139] [Medline: 32338621]
27. Bell KM, Onyeukwu C, Smith CN, Oh A, Devito Dabbs A, Piva SR, et al. A portable system for remote rehabilitation following a total knee replacement: a pilot randomized controlled clinical study. Sensors (Basel) 2020 Oct 27;20(21):6118 [FREE Full text] [doi: 10.3390/s20216118] [Medline: 33121204]
28. Bäcker HC, Wu CH, Schulz MR, Weber-Spickschen TS, Perka C, Hardt S. App-based rehabilitation program after total knee arthroplasty: a randomized controlled trial. Arch Orthop Trauma Surg 2021 Sep 06;141(9):1575-1582 [FREE Full text] [doi: 10.1007/s00402-021-03789-0] [Medline: 33547927]
29. Colomina J, Drudis R, Torra M, Pallisó F, Massip M, Vargiu E, et al. Implementing mHealth-enabled integrated care for complex chronic patients with osteoarthritis undergoing primary hip or knee arthroplasty: prospective, two-arm, parallel trial. J Med Internet Res 2021 Sep 02;23(9):e28320 [FREE Full text] [doi: 10.2196/28320] [Medline: 34473068]
30. Rian T, Sand K, Skogvoll E, Klepstad P, Wik TS. A web-based communication tool for postoperative follow-up and pain assessment at home after primary knee arthroplasty: feasibility and usability study. JMIR Form Res 2022 Apr 28;6(4):e34543 [FREE Full text] [doi: 10.2196/34543] [Medline: 35482392]
31. Bäcker HC, Wu CH, Pförringer D, Petersen W, Stöckle U, Braun KF. A review of functional outcomes after the app-based rehabilitation of patients with TKA and THA. J Pers Med 2022 Aug 21;12(8):1342 [FREE Full text] [doi: 10.3390/jpm12081342] [Medline: 36013291]
32. Constantinescu D, Pavlis W, Rizzo M, Vanden Berge D, Barnhill S, Hernandez VH. The role of commercially available smartphone apps and wearable devices in monitoring patients after total knee arthroplasty: a systematic review. EFORT Open Rev 2022 Jul 05;7(7):481-490 [FREE Full text] [doi: 10.1530/EOR-21-0115] [Medline: 35900191]
33. Brooke J. SUS: a 'quick and dirty' usability scale. In: Usability Evaluation In Industry. Boca Raton, FL: CRC Press; 1996.
34. Laugwitz B, Held T, Schrepp M. Construction and evaluation of a user experience questionnaire. In: Proceedings of the 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society. 2008 Presented at: USAB 2008; November 20-21, 2008; Graz, Austria. [doi: 10.1007/978-3-540-89350-9_6]
35. WHO evaluation practice handbook. World Health Organization. 2013. URL: http://apps.who.int/iris/bitstream/10665/96311/1/9789241548687_eng.pdf [accessed 2023-01-04]

36. Seffah A, Donyaee M, Kline RB, Padda HK. Usability measurement and metrics: a consolidated model. Softw Qual J 2006 Jun;14(2):159-178. [doi: 10.1007/s11219-006-7600-8]
37. BS EN ISO 9241-11:1998 ergonomic requirements for office work with visual display terminals (VDTs) - Guidance on usability. British Standards Institution. 1998. URL: https://knowledge.bsigroup.com/products/ergonomic-requirements-for-office-work-with-visual-display-terminals-vdts-guidance-on-usability?version=standard [accessed 2024-04-16]
38. ISO/IEC 9126 in software engineering. GeeksforGeeks. URL: https://www.geeksforgeeks.org/iso-iec-9126-in-software-engineering/ [accessed 2024-04-16]
39. Hornbæk K. Current practice in measuring usability: challenges to usability studies and research. Int J Hum Comput Stud 2006 Feb;64(2):79-102. [doi: 10.1016/j.ijhcs.2005.06.002]
40. Johnson SG, Potrebny T, Larun L, Ciliska D, Olsen NR. Usability methods and attributes reported in usability studies of mobile apps for health care education: scoping review. JMIR Med Educ 2022 Jun 29;8(2):e38259 [FREE Full text] [doi: 10.2196/38259] [Medline: 35767323]
41. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
42. Schrepp M, Hinderks A, Thomaschewski J. Applying the user experience questionnaire (UEQ) in different evaluation scenarios. In: Proceedings of the Design, User Experience, and Usability: Theories, Methods, and Tools for Designing the User Experience. 2014 Presented at: DUXU 2014; June 22-27, 2014; Heraklion, Greece. [doi: 10.1007/978-3-319-07668-3_37]

Abbreviations

AIMS: Active and Independent Management System
DHI: digital health intervention
DHS: dynamic hip screw
IC: integrated care
mHealth: mobile health
POD1: postoperative day 1
RQ: research question
SUS: System Usability Scale
THA: total hip arthroplasty
THR: total hip replacement
TKA: total knee arthroplasty
TKR: total knee replacement
UEQ: User Experience Questionnaire

Edited by A Parush; submitted 30.06.23; peer-reviewed by C Baxter, F Hussain; comments to author 14.01.24; revised version received 08.03.24; accepted 18.03.24; published 14.05.24.

Please cite as:
Papadopoulos P, Soflano M, Connolly T
A Digital Health Intervention Platform (Active and Independent Management System) to Enhance the Rehabilitation Experience for Orthopedic Joint Replacement Patients: Usability Evaluation Study
JMIR Hum Factors 2024;11:e50430
URL: https://humanfactors.jmir.org/2024/1/e50430
doi:10.2196/50430
PMID:38743479

Original Paper

Evaluation of the Parkinson's Remote Interactive Monitoring System in a Clinical Setting: Usability Study

Bronwyn Bridges[1], BSc; Jake Taylor[2], BSc; John Thomas Weber[1], BS, MS, PhD

[1]School of Pharmacy, Memorial University, St. John's, NL, Canada

[2]School of Exercise Science, Physical & Health Education, University of Victoria, Victoria, BC, Canada

Corresponding Author:
John Thomas Weber, BS, MS, PhD
School of Pharmacy
Memorial University
Health Sciences Centre
300 Prince Philip Drive
St. John's, NL, A1B 3V6
Canada
Phone: 1 709 864 7261
Fax: 1 709 864 6941
Email: jweber@mun.ca

Abstract

Background: The fastest-growing neurological disorder is Parkinson disease (PD), a progressive neurodegenerative disease that affects 10 million people worldwide. PD is typically treated with levodopa, an oral pill taken to increase dopamine levels, and other dopaminergic agonists. As the disease advances, the efficacy of the drug diminishes, necessitating adjustments in treatment dosage according to the patient's symptoms and disease progression. Therefore, remote monitoring systems that can provide more detailed and accurate information on a patient's condition regularly are a valuable tool for clinicians and patients to manage their medication. The Parkinson's Remote Interactive Monitoring System (PRIMS), developed by PragmaClin Research Inc, was designed on the premise that it will be an easy-to-use digital system that can accurately capture motor and nonmotor symptoms of PD remotely.

Objective: We performed a usability evaluation in a simulated clinical environment to assess the ease of use of the PRIMS and determine whether the product offers suitable functionality for users in a clinical setting.

Methods: Participants were recruited from a user sign-up web-based database owned by PragmaClin Research Inc. A total of 11 participants were included in the study based on the following criteria: (1) being diagnosed with PD and (2) not being diagnosed with dementia or any other comorbidities that would make it difficult to complete the PRIMS assessment safely and independently. Patient users completed a questionnaire that is based on the Movement Disorder Society–sponsored revision of the Unified Parkinson's Disease Rating Scale. Interviews and field notes were analyzed for underlying themes and topics.

Results: In total, 11 people with PD participated in the study (female individuals: n=5, 45%; male individuals: n=6, 55%; age: mean 66.7, SD 7.77 years). Thematic analysis of the observer's notes revealed 6 central usability issues associated with the PRIMS. These were the following: (1) the automated voice prompts are confusing, (2) the small camera is problematic, (3) the motor test exhibits excessive sensitivity to the participant's orientation and position in relation to the cameras, (4) the system poses mobility challenges, (5) navigating the system is difficult, and (6) the motor test exhibits inconsistencies and technical issues. Thematic analysis of qualitative interview responses revealed four central themes associated with participants' perspectives and opinions on the PRIMS, which were (1) admiration of purpose, (2) excessive system sensitivity, (3) video instructions preferred, and (4) written instructions disliked. The average system usability score was calculated to be 69.2 (SD 4.92), which failed to meet the acceptable system usability score of 70.

Conclusions: Although multiple areas of improvement were identified, most of the participants showed an affinity for the overarching objective of the PRIMS. This feedback is being used to upgrade the current PRIMS so that it aligns more with patients' needs.

(JMIR Hum Factors 2024;11:e54145) doi:10.2196/54145

KEYWORDS

Parkinson disease; usability; remote monitoring; motor examination; movement disorders; thematic analysis; System Usability Scale; mobile phone

Introduction

Background

Overview

The fastest-growing neurological disorder is Parkinson disease (PD) [1]. PD is a progressive neurodegenerative disease that affects 10 million people worldwide [2]. The incidence and prevalence of PD is rising sharply in countries with aging populations, and in the last 2 decades, the burden of PD has more than doubled, with estimations predicting 1,238,000 cases in North America by 2023 [3,4]. The disease affects the basal nuclei in the central nervous system causing the progressive deterioration of dopaminergic neurons. The loss of these neurons causes motor and nonmotor dysfunctions [5]. Motor system deficits result in symptoms such as tremor, rigidity, bradykinesia, and postural instability [6]. Other symptoms include cognitive problems, gastrointestinal upset, and urinary control issues [7]. Due to these mal effects, PD is linked to morbidity, high economic burden, and decreased quality of life for patients and caregivers. The annual estimated direct and indirect costs of the condition in the United States alone are close to US $52 billion [8]. Neurologists are struggling to manage the increasing prevalence of PD, leading to clinician burnout [9] and lengthy appointment wait times for patients [10]. However, studies show that the management of these symptoms in the early stages of the disease can achieve positive results. In contrast, the consequences of late or faulty diagnoses negatively impact patients and the health care system [11-13].

Medication Management

PD is typically treated with levodopa, an oral pill taken to increase dopamine levels, and other dopaminergic agonists. However, as the disease progresses, the effects of the drugs wane. This requires medication dosage adjustments to properly manage symptoms throughout the day [14]. This can be a difficult task for physicians as symptoms are constantly fluctuating and may appear and disappear throughout the day with a hard-to-establish pattern. Some physicians ask their patients to keep diaries where they note the time of day and a description of their symptoms. However, adherence to this method is typically poor and does not provide meaningful information [15]. Therefore, remote monitoring systems that provide more detailed and accurate information on a patient's condition regularly are a valuable tool for clinicians and patients to manage their medication.

Evaluation of PD

The evaluation of PD is commonly performed using clinical rating scales that are essential to the quantification of neurological disorders [16]. These rating scales enable clinicians and researchers to evaluate PD symptoms, progression, treatment efficacy, and disease severity [16,17]. One of the most widely used clinical scales for PD assessment is the Movement Disorder Society–sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS) [18].

The MDS-UPDRS is a revised form of the original Unified Parkinson's Disease Rating Scale [17] and incorporates both motor and nonmotor aspects into the assessment. It consists of 65 elements and, on average, requires approximately 30 minutes of administration time. There are four parts to the questionnaire: (1) nonmotor experiences of daily living (13 elements), (2) motor experiences of daily living (13 elements), (3) motor examination (33 elements), and (4) motor complications (6 elements) [18]. Elements are scored from 0 to 4, where 0=normal, 1=slight, 2=mild, 3=moderate, and 4=severe. There are some elements that patients could possibly administer themselves as they are multiple-choice questions asking about personal symptom experience, whereas others are rated by an examiner (typically a neurologist or other clinician) based on observation and physical examination.

While the MDS-UPDRS represents the international gold standard in PD rating scales and has undergone strict validation through clinical studies, it still remains a clinician-based scale; this means that a clinician assigns a score based on their own personal qualitative observations of a patient. Therefore, the assessments are often subjective and biased to the examiner's skill and knowledge. The assessments will also vary from one examiner to the other in this way [19-21]. Studies have shown that there is variability between assessments conducted by nurses and neurologists [22,23]. In these situations, it is difficult to compare and interpret the scores, as they may differ based on a patient's condition or simply due to the clinician performing the assessment. The MDS-UPDRS is also time consuming for clinicians. It requires approximately 30 minutes of an examiner's time, which makes it impractical for routine practice [18]. Examiners must also be highly trained to improve the validity of the scores. Many of the elements in the MDS-UPDRS must be completed by a patient themselves, which adds to the time burden of the questionnaire when performed in a clinician's office. The typical assessment performed in a clinical setting rarely assesses a patient's day-to-day symptoms, which usually vary over time, and only captures a snapshot of an individual's condition at the moment of their appointment [24]. Patients also typically have long wait times in between their appointments, which makes it difficult to remember their symptoms since their last visit [10]. This way, medical decisions are now influenced by recall bias and patient attitudes instead of by reliable patient data. In addition, it is an inconvenience for patients to travel to clinics due to transportation, long commutes, and their conditions, especially if they are in the advanced stages of PD. Therefore, there is a need for objective, accurate, and reliable assessment tools that can help increase the chance of effective treatment. These could aid patients with their disease management, thus cutting down on health care costs [25].

Digital Health Technologies and the Parkinson's Remote Interactive Monitoring System

An emerging solution to some access to health care issues are video-based visits. These bring care directly into a patient's home, which improves access in a patient-centered manner and minimizes the burden on people with PD and their caregivers [26]. In addition, due to the largely visual nature of a PD examination, it tends to work well in a video-based visit. Studies have shown that web-based appointments with neurologists are feasible and valuable [27,28]. It has also been shown that a modified version of the MDS-UPDRS motor examination (excluding the test of rigidity and postural stability) can be successfully administered remotely [29]. However, as virtual visits still require a clinician's time, they still only provide a brief snapshot of a patient's condition. There need to be other methods of assessing PD without occupying already overburdened clinicians.

Digital health technologies that alleviate the need for medical professionals to assess disease progression have been on the rise. These technologies offer possibilities for self-assessment and improved health care [30]. Some of the technologies developed for PD include wearable sensors and mobile apps. These devices have been used extensively to monitor motor symptoms and complications of people with PD in their home environments [31]. These wearable sensors and mobile apps can accurately track the progression of PD [32-34] and other neurological conditions [35]. Examples of these devices on the market are the Global Kinetics Corporation's Personal KinetiGraph Watch [36,37] and Rune Labs' StrivePD mobile app [38]. APDM Wearable Technologies has also developed multiple sensors that can accurately monitor tremor and dyskinesia symptoms of PD that have been used in many clinical studies [32,34,39,40]. Other wearables that collect contextual data include DynaPort MiniMod Hybrid (a sensor worn on the lower back), Shimmer (records gait), SENSE-PARK (records walking, hypokinesia, dyskinesia, and sleep), activPAL, and StepWatch (gait and basic movement parameters) [20,41-45]. The problem is that these technologies only generate a small amount of patient data (mainly tremors and other motor symptoms, moods, and sleep characteristics). Therefore, although these devices provide an objective means of tracking PD characteristics, they do not provide a complete assessment of the condition. Wearable sensors also have inherent risks [46] and do not follow the gold standard clinical scales such as the MDS-UPDRS. These risks encompass potential interference with the daily activities of patients with PD, impacting their natural movements and behaviors. In addition, behavioral modifications stemming from the feedback provided by sensors can yield both positive and negative outcomes. On the positive side, such modifications may encourage beneficial lifestyle changes and provide meaningful data. However, as a downside, they may also contribute to increased anxiety and foster a dependency on the wearable device [46].

To address the need for reliable tools to objectively assess PD symptoms that do not require a clinician's involvement, the Parkinson's Remote Interactive Monitoring System (PRIMS) was developed. The PRIMS is a digitized version of the MDS-UPDRS in the form of a desktop application that people with PD can complete themselves. This way, the PRIMS provides a complete picture of PD assessment via its capacity to comprehensively measure both motor and nonmotor symptoms without the need for wearable sensors and its potential to serve as a valuable tool in a clinical or home setting. If validated through further investigation, the PRIMS has the potential of delivering a standard in PD assessment. The PRIMS also has the potential to be valuable in a home setting, offering a user-operated system capable of capturing a significant portion of the MDS-UPDRS (considered the gold standard). The system provides patients with a means of tracking their condition remotely and offers clinicians reliable data for better medication management. This comprehensive approach enhances understanding and facilitates more effective monitoring of the progression and individual symptoms of a patient.

Usability Testing

The development of any system that is used by patients and clinicians for the management of biomedical data should always involve usability evaluations, which aim to understand whether such a product is easy to use and has the appropriate functionality for the users. *Usability* is a term used to define how easily people can use a tool or object to accomplish a specific task [47,48]. In this way, when developing interfaces, it is imperative that they can be learned quickly and are easy to navigate. The system's layout should avoid and manage operational errors efficiently and provide users with appropriate feedback [47]. Usability must also address user satisfaction and provide solutions to the problem that the system was designed to solve [49]. A common method of assessing usability is the System Usability Scale (SUS). The SUS has been used in multiple studies, such as the evaluation of a mobile app for people with PD [50]. Structured interviews are common practice for these types of studies [51]. Field notes can also be a valuable tool for qualitative researchers to collect and analyze [52]. Observational notes can capture information such as the nonverbal reactions of users while they interact with the system. This study used multiple methods to assess the usability of the PRIMS.

Study Objectives

This study aimed to assess the functionality, usability, and user experience aspects of the most recent version of the PRIMS in a clinical setting from the perspectives of people with PD. Use issues identified in this study will guide designers in creating a more effective commercial product. Using multiple methods, including interviews and field notes along with SUS surveys, we evaluated the user experience of the PRIMS.

Methods

Participants

Participants were recruited from a user sign-up web-based database owned by PragmaClin Research Inc. The study was also advertised by the Parkinson Society Newfoundland and Labrador on their weekly newsletter. Interested participants who contacted us were given a questionnaire that determined their eligibility for the study. The inclusion criteria for study participation were the following: (1) being diagnosed with PD and (2) not being diagnosed with dementia or any other comorbidities that would make it difficult to complete the PRIMS assessment safely and independently. Participants were recruited on a first come, first served basis. Informed consent was obtained from all participants via a web-based consent form emailed to them before study completion. Paper copies were also available to participants at the time of their scheduled session.

Ethical Considerations

This study received ethics approval from the National Research Council of Canada Institutional Review Board (protocol 2021-137). Informed consent was obtained, and the possible consequences of the study were explained. All data were deidentified. No compensation was provided to participants.

Description of the PRIMS

The PRIMS was developed by PragmaClin Research Inc and was designed on the premise that it will be an easy-to-use digital system that can accurately quantify motor and nonmotor symptoms of PD remotely. The PRIMS has the capability to interact with patients in real time, delivering results promptly through a dedicated patient dashboard. Patients can access their dashboard by logging into the web-based platform to see a history of their assessments. Patient users complete a questionnaire that is based on the MDS-UPDRS. The questionnaire comprises 4 sections shown in Figure 1. Of these sections, 3 are multiple-choice questions based on daily living experiences; an example is shown in Figure 2, and there is also a motor examination where users perform tasks similar to those outlined in the MDS-UPDRS. Data are captured via 2 depth cameras (Intel models D435 and D455) that track a patient's movement in 3D. Before completing the motor examination, there is a series of ability questions that determine whether the user can safely perform all motor tasks; an example is shown in Figure 3. Motor tasks are explained in written form on the screen along with a demonstration video that presents users with a visual walk-through of the movement; an example is shown in Figure 4. The intelligent software scores each motor task based on the same parameters as the MDS-UPDRS. However, it is important to note that the system's scoring has not yet been validated. After users complete the 4 sections, a participant's responses are analyzed to put an individual on a PD rating scale from 0 to 4 (0=normal, 1=slight, 2=mild, 3=moderate, and 4=severe). A summary of a user's scores is presented on the home page, which can be seen in Figure 5. The survey was intentionally crafted and edited from the original MDS-UPDRS to use layperson language for easy comprehension. Although some technical terms appeared in titles or examples, they were not essential for answering questions or comprehending instructions.

Figure 1. The 4 sections of the Parkinson's Remote Interactive Monitoring System (PRIMS) questionnaire based on the Movement Disorder Society–sponsored revision of the Unified Parkinson's Disease Rating Scale.

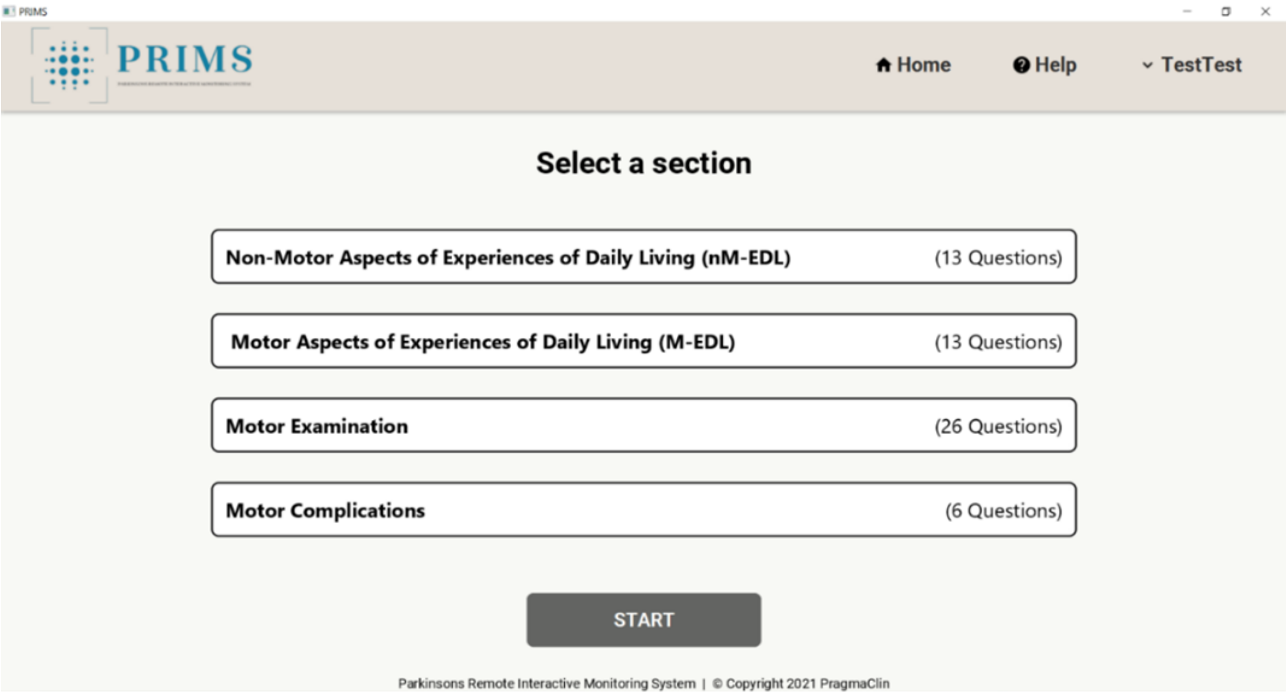


Figure 2. Example multiple-choice question.

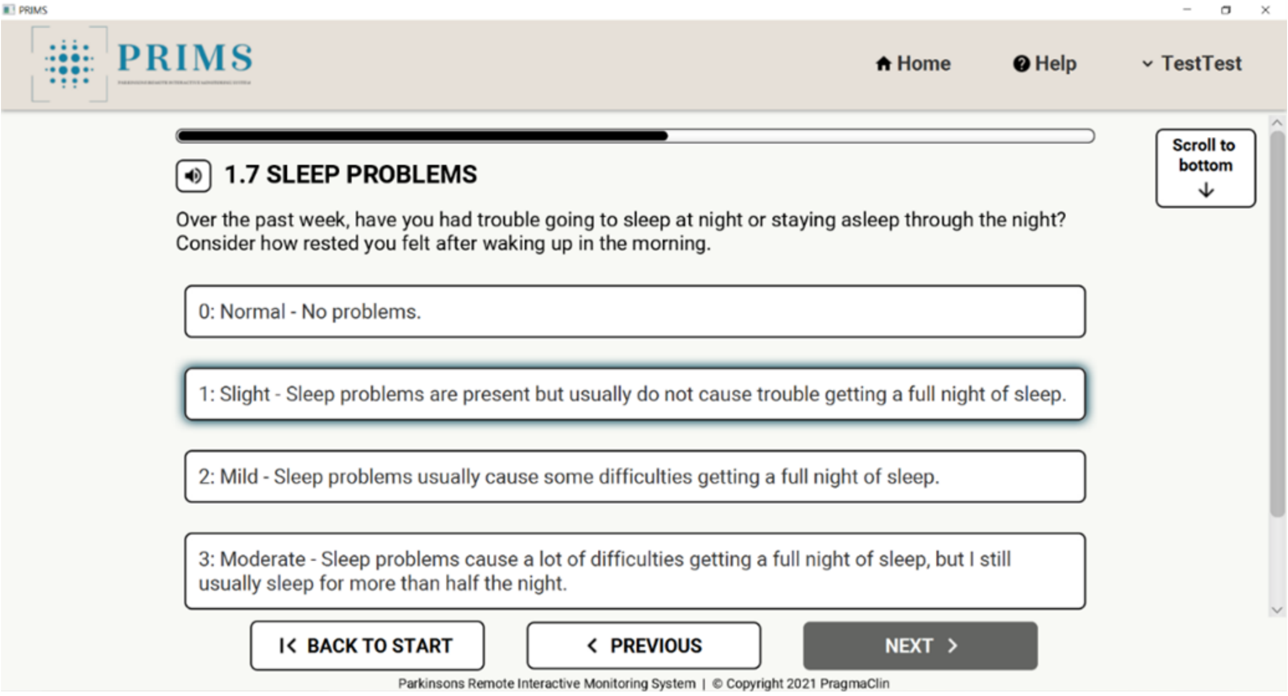


Figure 3. Multiple-choice question assessing an individual’s ability to stand.

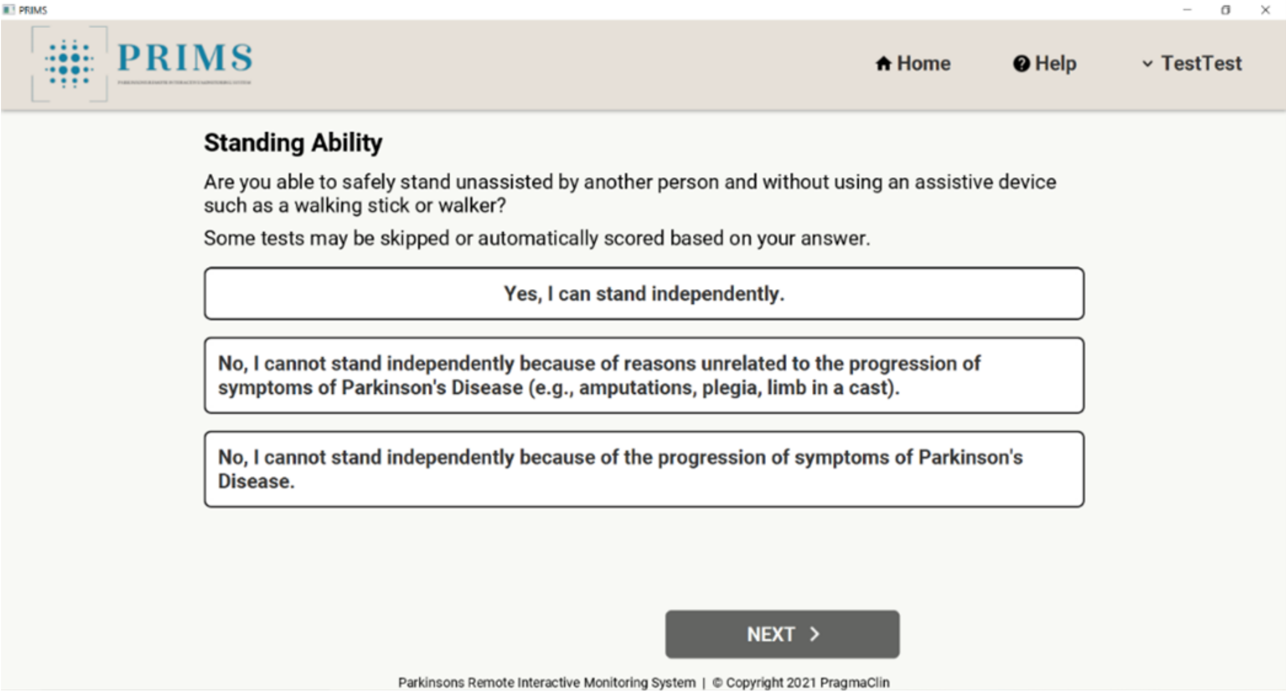


Figure 4. Example of a motor task page.

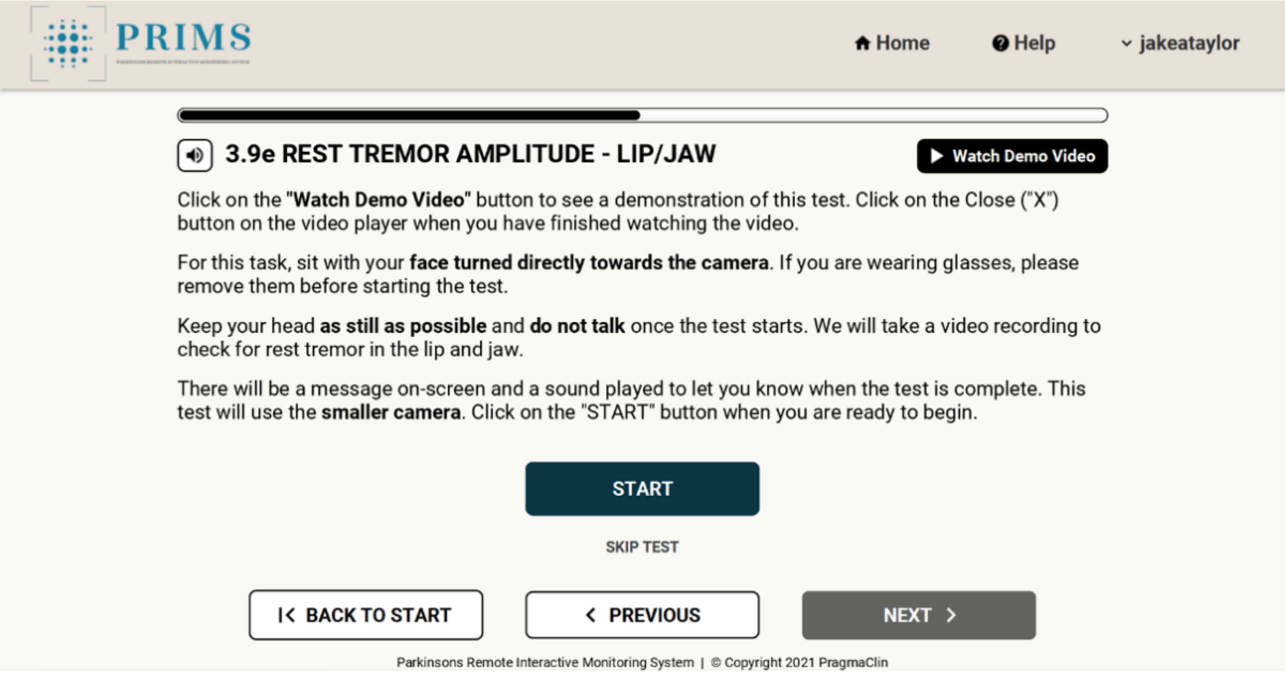
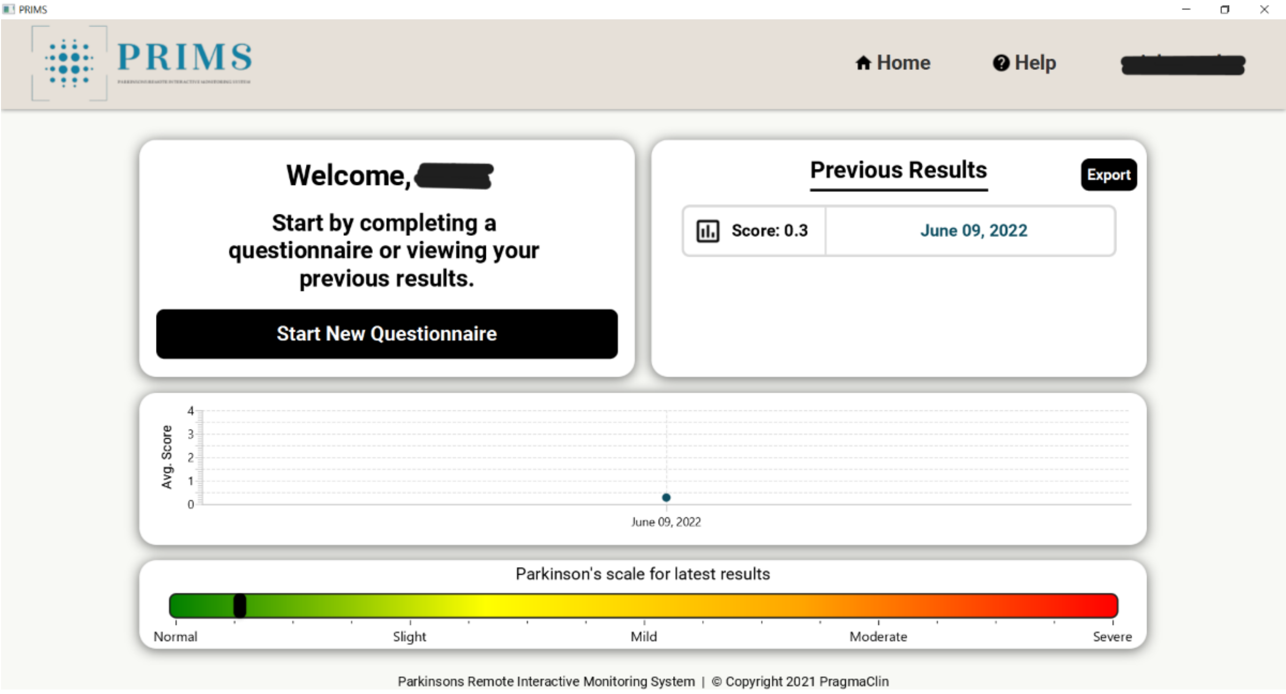


Figure 5. Screenshot of the Parkinson's Remote Interactive Monitoring System (PRIMS) home page.



Equipment

The PRIMS was run on a Dell G15 laptop computer, and Intel RealSense D435 (small—stand-alone mini tripod beside the laptop) and D455 (large—mounted on the computer) depth cameras were used. Participants sat on a contemporary midback task office chair with wheels for the entire questionnaire and had the option of using a Kensington Pro Fit wireless computer mouse. Interviews were recorded using a HyperX SoloCast stand-alone microphone. Audacity (Muse Group) was used as an audio recording and processing software. The computer-assisted qualitative coding software Delve (Twenty to Nine) was used for thematic analysis. All audio files were transcribed into Word (Microsoft Corp) before being uploaded to Delve. Qualtrics XM (Qualtrics International Inc) was used to administer the SUS survey and the virtual consent form.

Usability Testing Protocol and Procedure

Usability testing occurred at PragmaClin Research Inc's office site and was carried out by a trained research assistant (RA). Before the start of testing, the RA explained the study objective and research protocol to the participants. The RA also provided detailed information about the test procedures and described the purpose of the PRIMS. The example script is provided in Multimedia Appendix 1. As the research team was interested in how participants interact with the system when there is no one present to assist them, the RA was not allowed to help unless deemed necessary. The necessity to intervene in the form of helpful hints or prompts or manipulating the system was operationally defined as any circumstance in which the participant was unable to progress through the system without aid. An observer was present during the entire session. All participants used the PRIMS only once and reported user experience from this single use.

After the participants gave informed consent, they were instructed to start using the PRIMS. The initial PRIMS developed by PragmaClin Research Inc was used during the usability testing. While participants worked their way through the assessment, they were encouraged to vocalize any confusion or ask any questions.

Data were recorded in the form of field notes, a short qualitative interview, and an SUS survey administered to participants after they completed the PRIMS questionnaire. Recorded data were used to identify a set of usability issues.

Field Notes

Structured observation was used to analyze the users' interactions with the system. During the session, the RA was instructed to observe and note any issues that arose along with any user critiques, comments, questions, difficulties, or observations about their interaction with the system. Thematic analysis was performed on these notes in Delve. Usability issues were identified via this method of analysis of the written notes.

Qualitative Interviews

After participants were finished with the PRIMS, they completed a short qualitative interview. The interview consisted of six questions: (1) What things did you like most about the PRIMS? (2) What things did you like least about the PRIMS? (3) Were there things about the PRIMS that you found confusing or frustrating? (4) What would you like to change about the PRIMS? (5) Are there any features that you would like to see added to the PRIMS? (6) Do you have any overall comments on the PRIMS?

Audio was transcribed and analyzed in the qualitative coding software Delve. Thematic analysis was performed following the framework by Braun and Clarke [53]. The themes were discussed, reviewed, and interpreted by the research team.

SUS Survey

After the interview was finished, participants completed a short-answer quantitative questionnaire following the standard SUS approach devised by Lewis and Sauro [54] in 2018. A copy of the survey is provided in Multimedia Appendix 2.

SUS scores were output in Qualtrics XM and then analyzed in Microsoft Excel. To calculate the SUS score, first the score contributions from each item (question) were summed. Each item's score contribution ranged from 0 to 4. For items 1, 3, 5, 7, and 9, the score contribution was the scale position minus 1. For items 2, 4, 6, 8, and 10, the contribution was 5 minus the scale position. We multiplied the sum of the scores by 2.5 to obtain the overall value of the SUS score. SUS scores have a range of 0 to 100. SUS scores of >70 points are considered acceptable usability (according to various other usability studies), and scores of >85 are regarded as excellent usability [55]. The curved grading scale by Lewis and Sauro [54] that was used to interpret the scores from the SUS is provided in Multimedia Appendix 3.

Results

Participants

A total of 11 people with PD participated in our study (female individuals: n=5, 45%; male individuals: n=6, 55%; age: mean 66.7, SD 7.77 years). Data from 91% (10/11) of the participants were fully analyzed as 1 user dropped out during the testing session. The 10 participants took, on average, 67.7 (SD 16.4) minutes to complete the motor examination and 84.2 (SD 23.3) minutes to complete the entire PRIMS questionnaire. Participants skipped 2.9 (SD 1.97) motor tests on average (total of 29 skipped tests). Table 1 shows which tests were skipped the most and how many times they were skipped.

Table 1. Number of times each motor test was skipped (tests ranked from the highest to lowest number of skips).

Motor test	Skipped tests, n
3.15	12
3.14	6
3.13	6
3.3	2
3.8	1
3.9e	1
3.10	1
3.1	0
3.2	0
3.4 and 3.5	0
3.6	0
3.7	0
3.9a-d	0
3.11	0
3.12	0

Field Notes

Thematic analysis of the observer's notes revealed 6 central usability problems associated with the PRIMS. These were the following: (1) automated voice prompts are confusing, (2) the small camera is problematic, (3) the motor test exhibits excessive sensitivity to the participant's orientation and position in relation to the cameras, (4) the system poses mobility challenges, (5) navigating the system is difficult, and (6) the motor test exhibits inconsistencies and technical issues.

Automated Voice Prompts Are Confusing

The RA noted on multiple occasions that the participants found the automated voice prompts to be confusing. Frequently, participants would begin the test and align themselves in a good position; however, when presented with audio prompts such as "make sure hand is not tilted" or "adjust hand position or angle," users would move in all directions:

> *The automated voice prompts were confusing and making it difficult. Even though P8 was turned in the right direction, the system still prompted him that he was turned the wrong way.*

> *A lot of automated verbal instructions were getting fired at P8, this made the task confusing and frustrating since they will be in the correct position and the software tells them wrong direction, turn right, etc.*

The prompt "Adjust hand position or angle" was especially frustrating and confusing to participants. Many verbalized their confusion with the statement. On multiple occasions, the RA noted that participants were becoming frustrated with the motor tests when the audio prompts began giving them instructions:

> *The prompt—no hand detected—is vague and confusing. P6 was in a good position, with their hand fully in view...Plus, P6 did not move and then immediately after there is a prompt saying—no hand detected...the audio prompts confused and frustrated P6.*

The RA also noted that some participants complained that the voice commands were authoritative and unfriendly:

> *...get in position and stand still—is a little authoritative! P8 mentioned this, participant did not like the automated voice instructions, said it needs to be more comforting / friendly. The automated voice prompts were authoritative P3 mentioned.*

Comments also often noted the repetitiveness of the automated voice prompts. They were repeated too often and in bizarre patterns, which caused confusion and frustration among users:

> *"don't move body in good position" was constantly repeating...was repeated ~5 times over, and the test wouldn't start.*

> *"Make sure hand is fully in view" is repeated even when the hand is fully in view, The system was prompting repeatedly "make sure hand is not tilted," motor test was very particular on positioning here. This made hand and body measurements very difficult for P2. They were moving their hand in all directions trying to figure out what tilted meant.*

Overall, participants found the automated voice prompts to be vague and irritating, as they rarely provided useful corrective feedback.

The Small Camera Is Problematic

The RA frequently noted problems associated with the tests that used the small camera or issues directly related to the small camera itself. The narrower field of view was an issue for multiple tests; participants frequently moved out of the camera view midway through a task:

> *Finger to nose movements were difficult for P7, they had a hard time staying in the cameraview. They would be prompted that they are in a good position, then move out of the cameraview when performing their finger to nose movements. Hand moves out of frame...when the participant does hand rotations...small camera issue. P4 had great difficulty staying in the cameraview for these hand rotation tasks.*

The system would frequently tell users that they were in a good position, but when they started performing the task, they would move out of the camera view. The setup of the small camera also created problems for users. Users had to manipulate and adjust the small camera on the tripod to align themselves in a proper position. Frequently, this would lead to participants becoming uncomfortable due to the poor ergonomics of the system:

> *Face measurements using the small camera were not comfortable for P4. Small camera moved P4 into an uncomfortable position.*
>
> *The fact that P3 had to look at the camera for the test but look at the screen to get into position was causing difficulties.*
>
> *Getting into position was difficult for P9. Again, ergonomics is poor here for performing the test on two different sides when the camera is on one side.*

The tests that asked users to adjust the small camera were especially problematic. The RA frequently noted that people who pulled the cords out when laying the camera down for the final 2 tests did the following:

> *P2 moved the tripod/small camera, it was sloppy and difficult to work with. P2 pulled the cord out while adjusting the small camera.*
>
> *P11 unplugged the camera when they moved it for 3.15...issue with adjusting camera.*

Overall, the RA noted far more issues with the tests that used the small camera compared to those that used the larger camera.

The Motor Test Is Excessively Sensitive

It was clear from the RA's notes that participants had difficulty getting into what the system would consider to be valid positions to score during the motor examination. At times, the system would prompt users to stay still:

> *P9 had difficulty with hand measurements, they were unable to hold their hand still (issue since the software is for PwP). System is far too particular. The hand and body measurements require users to stay still to capture the measurement; P11 had great difficulty with this. System is far too particular on positioning here, which is just not feasible for those with PD.*

The system in its current state is very sensitive, which posed challenges for users:

> *The software is too picky on the positioning, participant's dyskinesia made it very difficult to stay in position, when the software told P7 that they were in a good position, they only had to move very slightly for the software to tell them that they needed to "adjust hand position or angle."*
>
> *A slight tilt is all it took for P8 to move out of position. The system was very particular on the positioning of the limb.*

The RA noted on multiple occasions that the system would repeat certain prompts when users were close to getting into the correct position:

> *"Don't move, hand is in good position" repeats a lot when P6 was "on the edge" of a good position. And when you start rotating your hand "no hand detected."*
>
> *"Don't move, hand is in good position" repeated a lot when P5 was "on the edge" of being in a good position.*
>
> *The system is far too sensitive, "Don't move face is in good position" kept repeating even though P3's face was in a good position. The motor test is too picky, its needs to be able to get the measurements from a broader range of places.*

Overall, it was clear from the recorded field notes that the motor examination was difficult for users:

> *3.14, and 3.15 P9 had a lot of difficulty keeping their hand in the correct position, P8 had difficulties getting their hand perfectly parallel to the camera face, The fact that P3 had to look at the camera for the test but look at the screen to get into position was causing difficulties.*

The System Poses Mobility Challenges

The setup of the PRIMS posed various mobility challenges for users. The RA noted on multiple occasions that users felt that the chair and constant movement of users were both issues. In its current state, the PRIMS requires participants to move back and forth from the computer to go from one task to another. The RA noted on several occasions that this posed a challenge for users:

> *Moving back and forth from the computer was difficult...the system requires too much movement of the chair, and to and from the computer, P7 vocalized this.*

This, coupled with the fact that users are constantly moving the chair in and out of the camera view, made the motor examination tiring for participants:

> *There is constant movement to and from the system that was tiring P8.*

The chair was difficult to work around due to the nature of the test; the RA noted repeatedly that participants failed to complete tests due to the chair obstructing the camera view:

> *...chair was in cameraview during these tests, posed an issue, was also difficult for participants to work around it.*

Some participants were also obstructed by the legs of the chair:

> *P5 needed to work around the chair legs for the arising from chair task, they vocalized this. The legs of the chair were in the way.*

Multiple participants also pointed out that they thought the chair was a safety issue:

> *The chair with wheels concerned P2. They thought it was a safety issue.*
>
> *P3 mentioned a few times that people with PD are told NOT to use chairs with wheels.*
>
> *For safety reasons, the chair with wheels is a huge problem, P7 vocalized this.*

Overall, the maneuvering required to complete the motor examination was an issue for users.

Navigating the System Is Difficult

There were frequent comments made by the RA on navigational issues users had while working through the system. Many participants had issues with the required amount of scrolling:

> *Scrolling is too difficult. a bigger screen would allow the entire survey to fit on one screen. P5 found the scrolling to be a challenge right away.*
>
> *P9 again demonstrated issues with scrolling and navigation, they had trouble scrolling to the bottom of the screen to select next on multiple tasks.*

Some users even found that they made mistakes due to the need to scroll to the bottom of the page:

> *P7 found that the scrolling led to mistakes. Choices they didn't mean to select.*

The RA also noted that the amount of clicking was an issue, specifically accurate clicking:

> *Skip test button is too small. Navigation issue, P9 had trouble clicking it due to dyskinesia, since it was so small.*
>
> *P7 had trouble closing the demo videos. Too much accurate clicking / total clicking required.*

Users also had issues with the computer mouse and mentioned that a touch screen interface would be preferred:

> *P9 had significant difficulties with the mouse. They said that they would prefer a touch screen.*
>
> *P2 didn't like using the mouse...Stated right away that they wanted to use a touch screen.*

The test window and demonstration video window also caused issues for users. Demonstration videos would frequently open in inconsistent sizes and locations of the screen, making them difficult to close:

> *Demo videos were opening in small windows at the top of the screen. Made them difficult to close and to watch.*

The test window did not show a married image of the person, which made it confusing to get into position:

> *Screen being non-mirrored is an issue. P5 had trouble moving into position because of this.*

The RA also noted that, as the software was not entirely full screen, users would frequently open other programs by accidentally clicking on the bottom task bar:

> *Full screen should eliminate the lower task bar (desktop), P4 ended up clicking things below or bringing up the news.*

Overall, users had difficulty navigating the system to progress.

The Motor Test Exhibits Technical Issues

There were frequent notes made on the system not operating correctly. Users would perform tests correctly or be told that they were in a good position yet would still be asked to try again as the system did not capture enough valid measurements to score:

> *Hand movements test stated that there were not enough valid measurements to score, after the participant did everything correctly.*

There were also occurrences in which users would perform tasks incorrectly and the test would still function:

> *Postural stability test worked even though the participant was not in the correct position at all.*
>
> *The foot tapping test ran even though P3 tapped the wrong foot, the system still gave him a score. They performed the measurement incorrectly, yet the system still considered it to be valid.*

This would lead to confusion among users as they would go through tests being told that they were in a good position without any other corrective feedback only to be asked to try again:

> *P6 performed the test correctly without any prompts to change position yet the system still prompted them to try again. The test will prompt people to start walking, and will run through without any corrective feedback, but may still state that there were not enough valid measurements to score.*

Some tests also tended to shut off very early and inconsistently. Other tests would often produce nonsense automated voice prompts:

> *...while performing the finger to nose movements: the audio prompt "multiple hands detected" was repeated even though there was only one hand in the camera view.*

Overall, the motor test presented frequent glitches causing usability trouble for participants.

Qualitative Interviews

Thematic analysis of qualitative interview responses revealed 4 central themes associated with participants' opinions on the PRIMS. These were (1) admiration of purpose, (2) excessive system sensitivity, (3) video instructions preferred, and (4) written instructions disliked.

Admiration of Purpose

Most of the participants showed an affinity for the overarching objective of the PRIMS:

> *I know what the main objective is, and I applaud that, that is a good objective.* [Participant 8]

They were excited about the system being available to people with PD:

> *I think it is awesome that people will have access to this.* [Participant 11]

> *I just like the fact that this is available for people.* [Participant 11]

> *I am sure a lot of people would be thrilled to have this at their doctor's office.* [Participant 11]

The intended purpose of the PRIMS was also well received and understood. Participants liked the idea that this will give their physicians a better view of their condition and support their ability to do their job:

> *Doctors often don't have a lot of time to do the examination in depth. My in-person examinations with my Neurologist are very fleeting, and scratching the surface in my view, but if that is the norm, and my Neurologist got a good reputation, then something like this would be very very helpful.* [Participant 8]

> *I like how then your doctor would have a better idea of what you are doing really, rather than based on that little scope of time kind of thing. Yea...that would be good.* [Participant 6]

> *I like that it can be used for long distance. And in our new post covid medical system, we need to free up time for our doctors.* [Participant 3]

Overall, participants admired the system and what it is trying to achieve and were excited to see the finished product in the future:

> *I think it would be worthwhile, if it was something that was worked out, if all the bugs and stuff were worked out it could be used as a tool...* [Participant 4]

> *Overall, I think it's a pretty good system. I think it will help patients or people.* [Participant 1]

Excessive System Sensitivity

Most participants found that the motor examination was very sensitive, which made it difficult to get into the proper position for the tests:

> *Yea like I say it's too sensitive, cause we have Parkinson's, and most people, you know you can be [shaky] and there's no way you are going to be able to stop it [tremors].* [Participant 2]

Users also found it frustrating and time-consuming:

> *Well I didn't like how it kept telling me that my hand is in a good position and then it's not or they don't detect it or those kinds of things, it can get a bit frustrating.* [Participant 1]

> *The only thing is the actual working of it in those couple of times where no matter what I did or didn't do, everything was as still as I could make it and it says you are fine then it says nope you have to start again...oh my gentle god...maybe it's just too sensitive or something.* [Participant 2]

> *It is time consuming too.* [Participant 11]

> *It is long and...you know...as Parkinson's patients you get tired easily.* [Participant 6]

The frustration seemed to stem from the fact that the system was very particular on how it wanted users to be positioned. Participants had the greatest trouble with the hand movements and tasks that required users to stay still:

> *The ones we have the most difficulty with are the hand.* [Participant 8]

> *Because holding still is a challenge for some people with Parkinson's...some people have tremor, and some don't. For those that do, holding still is a real challenge.* [Participant 11]

Overall, users found that the system was difficult to use in its current state due to its sensitivity:

> *I think that that [PRIMS] would be difficult for some people...unless you had extensive training.* [Participant 4]

Video Instructions Preferred

Users found that the demonstration videos were far more helpful, and much less confusing, than the written instructions:

> *The video was a good tool because we have a lot of brain fog, and reading can be confusing and looking at the video makes things much easier.* [Participant 3]

> *...you are able to see a video of the man actually doing what you are supposed to do you know is quite helpful too I thought.* [Participant 8]

> *There are a lot of words there, in the instructions, again I think if you had it in bullet form maybe. It's a lot easier to watch the video.* [Participant 1]

> *The video was good to show how to do the testing.* [Participant 1]

Users suggested more video instructions and less written instructions and even suggested an introduction video outlining what the system entails:

> *Instead of just jumping right in there, if you had, well I guess it would be a video, but if you had a synopsis of what the testing involved. Maybe if we had a 10-minute video overlooking the whole test at first.* [Participant 1]

Overall, the videos were one of the most liked aspects of the entire system:

> *The things that I thought worked best were the videos.* [Participant 5]

> *I like how there is a video...watching what they do is much more clear.* [Participant 11]

Written Instructions Disliked

Many participants found that the written instructions were vague and confusing:

> *Some of the tests I found confusing, but again that was the written instructions that were somewhat confusing.* [Participant 5]
>
> *The instructions were really kind of vague.* [Participant 1]

Multiple users stated that these instructions were annoying and far too wordy:

> *Too much instruction, yea, but I know you have to have the instruction down, but it was a lot of reading.* [Participant 1]
>
> *I would say all the text that open on the screen, yea it's like going to a presentation...mostly just the instructions, I mean you're asking me, in my gut, kind of what I found to be annoying about it...and...the text was annoying.* [Participant 7]

Participants suggested that more concise bullet-point instructions would be preferred over written paragraphs:

> *I think the instructions was too many...If it was concise and shorter instructions, I think it would make it a little better.* [Participant 1]

SUS Survey

The average SUS score was calculated to be 69.2, which corresponds to a C on our curved grading scale [54]. The PRIMS failed to meet the acceptable SUS score of 70.

Discussion

Principal Findings

We conducted a multiple methods study to assess the usability, functionality, and user experience of the PRIMS. Thematic analysis of interview transcripts and field notes revealed multiple themes and usability issues, respectively, that describe the tested product. An SUS survey also gave us a key objective insight into the system and its user experience.

One of the key findings of this study was that video instructions were preferred over written instructions. Thematic analysis of interview transcripts revealed these 2 themes (*written instructions disliked* and *video instructions preferred*). Multiple participants stated that the video instructions were much less confusing and much more informative than the on-screen text. The written instructions were designed to give all the necessary information to complete the task. This may have resulted in users feeling unmotivated to read the entire set of instructions as there was an intimidating amount of text present on-screen. Other investigations comparing video to written instructions have found similar results. Cosford et al [56] evaluated the effectiveness of video and handout instructions during a veterinary student examination. Their findings revealed that students using video instructions achieved notably higher scores, suggesting a better understanding of the tasks compared to those using handouts [56]. Shah and Gupta [57] found that video instructions were significantly more effective than written instructions in teaching inhaler use technique. Video instructions provide both a visual and audio description of each task, which can make the instructions both clearer and less time-consuming.

Another principal finding was that the PRIMS motor examination was too sensitive and particular on users' body positions during the tests. Thematic analysis of field notes and interview transcripts unveiled 2 areas of issues, namely, system sensitivity and the motor test's positioning specificity, which exhibited alignment in their respective scopes. To quantify each motor task performed during the PRIMS questionnaire, the depth cameras would require participants to be oriented in a "good position." From the RA's observational notes and the interview transcripts, it was clear that the system asked too much of users, which led to frustration and difficulties. Systems designed for those with movement disorders must be accommodating to their needs. The PRIMS, in its current state, asks users to stay still in certain situations and adopt specific and uncomfortable positions to score their movements. Future versions of the PRIMS will need to address this in their design and implementation.

Another theme revealed from analysis of field notes was that the automated voice prompts that are used during each motor test are confusing to participants (*automated voice prompts are confusing*). The prompts would also do more harm than good when it came to helping participants align themselves in the correct position for each motor test. It was noted that users tended to move in all directions in response to the automated voice. This could be due to the vague nature of the instructions provided by the prompts. They also led to frustration and confusion, making them an ineffective tool to guide users through the tests. Some users even stated that they found the automated voice to be authoritative and unfriendly, which only increased their frustration with the system. The consensus of this key finding was that these automated prompts did not provide any useful corrective feedback and only led to confusion and frustration among participants. Mays et al [58] delved into how people in the United States perceive automated communication, such as interactive voice response systems. They found that older respondents especially did not enjoy the automated voice system and exhibited greater levels of frustration toward it [58]. Most people with PD are older individuals; therefore, it would be best practice to tailor the system's instructions and prompts to their typical preferences.

Our next key finding is that the small camera tended to cause more problems for users than the large camera. Thematic analysis performed on field notes revealed this theme (*the small camera is problematic*). The smaller-depth camera (Intel D435) has a narrower field of view and was primarily used for the hand movement tests during the motor examination. A common problem that users faced was staying within the camera view for these tests. As the RA noted on several occasions, it was common for participants to have difficulties with the narrow view. The position of the small camera also caused trouble for users. Unlike the larger camera, the smaller camera is placed on a tripod on either side of the computer (Figure 6). It was noted frequently in the field notes that users had to manipulate and adjust this camera to align themselves in the proper position. The placement of the camera also negatively impacted the system's ergonomics. In addition, for tests 3.14 and 3.15, users were prompted to flip the tripod down so that the camera was facing the ceiling. Frequently, users accidentally disconnected the cords from the camera when moving it. In general, a setup in which users do not have to adjust any equipment would be preferential. To our knowledge, there is no direct study to compare this finding to.

Figure 6. Diagram of hardware setup showing the position of the laptop and 2 depth cameras.

Another principal finding of our thematic analysis was that the PRIMS has associated mobility challenges for users. A big issue that the design of the system has is the constant movement to and from the computer to go from one task to another. At times when the users' feet or whole body had to be visible to the camera, participants would have to move backward until this was the case. To move onto the next task, users would have to return to the laptop to select it. This way, users are constantly moving to and from the computer and frequently being obstructed by the chair. Several users even pointed out that they felt as though others would have problems with the back-and-forth nature of the system. When designing systems for those with movement disorders, it is important to consider the user experience as a whole.

Thematic analysis of field notes also revealed that users had difficulties navigating the system (*navigating the system is difficult*). In its tested state, the PRIMS was operated using a standard computer mouse or laptop touchpad (depending on user preference). Navigating the system using either of these tools caused difficulty among participants. Scrolling or clicking to move in between sections of the questionnaire was frequently noted as a challenge for users and even led to mistakes in some circumstances. The accuracy of the clicking to move through the system was also a big issue and would often lead to opening other applications or closing the PRIMS software. It is important to remember that, when designing systems for those with movement disorders, there must be special considerations taken. A viable option could be a touch screen device featuring prominently sized buttons, eliminating the need for scrolling. However, it is worth noting that touch screens can lead to increased postural discomfort during use [59]. Thus, offering a variety of system operation methods might be the most effective way to cater to diverse user needs and preferences. Enhancing usability is paramount not only for optimizing human-computer interactions but also for ensuring the system's social and practical acceptance [48]. A system's usability should be of a standard that facilitates effortless task execution by the user. Given that the PRIMS posed challenges for users in performing certain tasks, there is a clear need for usability enhancements.

Another principal finding that came from analyzing the field notes is that the motor examination did not function perfectly or as intended. Similar to any new software system, the PRIMS had its share of technical issues. One of the most frequent issues noted by the RA was that users would perform tests correctly yet the system would fail to score their movements. This is a problem especially when the system does not prompt any corrective feedback yet still informs users that there were not enough valid measurements to score. The software issues caused frustration among the participants and led to multiple usability issues. Usability is tied to functionality, although they are not exactly the same. When a product is not functioning correctly, it ultimately impacts its usability. Thus, ensuring that the system works as it is intended must be a priority for future developers to improve its usability.

Through the analysis of interview transcripts, a prominent theme of appreciation emerged in relation to the PRIMS. Designed with the primary objective of enhancing care for individuals living with PD, the PRIMS was met with significant enthusiasm. Users recognized the immense potential of such a remote monitoring solution, expressing eagerness about its availability. This positive reception underscores the importance of aligning the product's design with the needs of its intended users. The participants' commendation of both the system and its mission suggests that the overarching principle of the product is robust. Previous studies have emphasized the pivotal role of consumer perspectives in determining the success of a product [60]. Given this context, such a positive reception indicates a promising trajectory for the future deployment of the PRIMS.

Our last principal finding from this usability study was that our recorded SUS score was 69.7, which failed to meet our acceptable usability score of 70. There are several reasons that could explain why participants felt that the system was not as user-friendly as it should be. Binyamin et al [61] used the SUS to evaluate a learning management system in an educational setting. Their system also failed to meet an acceptable usability score of 70. A distinct aspect of their study was that participants engaged with the system repeatedly throughout a summer term. They observed a direct relationship between the frequency of system use and the SUS score, suggesting that increased familiarity led to improved usability ratings [61]. Drawing from this, it is conceivable that, if participants in our study had interacted with the system over an extended duration, as intended for the PRIMS, the SUS score might have been more favorable due to enhanced user familiarity by the study's conclusion.

Severity of Usability Problems

Within the spectrum of presented usability problems, a hierarchical assessment of severity becomes imperative considering factors ranging from potential risks to participant safety to issues causing minor hindrances in task completion. We ranked the issues posing threats to safety and mobility as top priorities to be addressed, followed by those that led to difficulty and frustration, with minor issues that may have slowed participants down being of the least concern.

Foremost among the identified challenges were those associated with mobility constraints, standing out as the most severe due to their inhibiting impact on participants with mobility challenges. Beyond impeding the use of the PRIMS, these challenges pose a risk to participant safety by potentially placing individuals in vulnerable positions. Notably, the constant need for movement to and from the computer for task transitions emerges as a top priority for resolution.

A usability problem of a lesser degree of severity pertains to the system's sensitivity, specifically in quantifying motor tasks during the PRIMS questionnaire. The requirement for participants to be consistently in a "good position" proved overly demanding, leading to frustration and difficulties, as observed in RA notes and interview transcripts. This underscores the importance of designing systems for individuals with movement disorders to be accommodating to their unique needs. The current state of the PRIMS, requiring users to stay still in certain situations and adopt uncomfortable positions, resulted in skipped tests, increasing the priority of addressing this issue.

Issues that made completion difficult included challenges in navigating using the computer mouse and occasional malfunctions in the motor examination. These concerns, while not as severe as mobility-related issues or the system's sensitivity, warrant attention as they contribute to user frustration and impact task completion.

Finally, minor difficulties associated with operating the small camera and managing automated voice prompts and written instructions require fine refinements rather than constituting significant hurdles. While not impeding overall system navigation, these issues contributed to slower completion and user frustration. In the hierarchy of severity, they represent areas for enhancement rather than critical concerns requiring immediate attention.

Implications

This study conducted a comprehensive evaluation of a system specifically tailored for individuals with motor and cognitive conditions, shedding light on critical considerations for the development of technology for this population. First, we advise against the use of desktop applications requiring a computer mouse, scrolling, and intricate clicking, recognizing the potential challenges faced by users. Furthermore, our findings emphasize the superiority of visual instructions over written ones, also suggesting that automated voice prompts should be used judiciously and presented in a friendly manner and offer clear instructions, especially during confirmation processes. To address the mobility challenges commonly faced by this population, systems necessitating movements in front of a camera should minimize the need for multiple adjustments as these can introduce errors. In addition, our study underscores the importance of system flexibility, allowing for a significant margin of error in data capture without imposing the requirement for participants to remain perfectly still during calibration—an often-unattainable feat for those with motor conditions. As we navigate future developments of the PRIMS, these insights will serve as a guide in creating a product that effectively addresses the outlined issues, emphasizing a visually guided interface requiring minimal effort for seamless operation, aligning with the unique needs of our target user base.

Limitations and Future Investigation

There were several limitations to this study. Our small sample size may not have revealed all the usability issues [62,63] as testing with a small number of participants tends to only reveal the major flaws or glitches in the system. However, our main objective was to uncover the biggest areas of concern rather than identifying every problem associated with the system.

Another limitation to this study was our methodology. Other common qualitative data recording techniques for usability studies include the think-aloud technique [64] and focus groups. The think-aloud technique is the process in which users are encouraged to verbalize their perceptions as they interact with the system [64], which can provide insight into the user experience. Focus groups with study participants following the interviews could have produced richer information. This would have given users the opportunity to compare their ideas and thoughts on the PRIMS. However, while conducting this study, we made a deliberate decision not to use the think-aloud technique. This choice was grounded in our consideration for the unique challenges faced by individuals with PD, particularly those with motor and speech difficulties. The think-aloud technique traditionally involves participants verbalizing their thoughts as they navigate through a system. However, given the potential speech impediments, tremors, and other motor-related challenges associated with PD, we anticipated that asking participants to vocalize their thoughts could introduce unnecessary stress and frustration, and we did not want to pile on any extra cognitive load. To ensure a more comfortable and authentic testing environment for individuals with PD, we opted for direct observation followed by an interview after they were finished using the system, allowing us to carefully note any challenges users encountered as they interacted with the software.

We also acknowledge that other methods, for example, mixed methods research [65], are applicable for usability and user experience research. Mixed methods may allow for a more comprehensive understanding of a user's experience, which can enable researchers to identify specific usability issues [66]. There are also other scales that we could have used to quantify user satisfaction, for example, the Post-Study System Usability Questionnaire [67]. As our product is in the early stages of development, we opted for a simpler multiple methods study to uncover the major flaws in our system. Future usability studies on the PRIMS can use a mixed methods design to gain a deeper understanding of usability issues.

Considering the intricacies involved in designing systems for users with movement disorders, a promising direction for future research could entail conducting user-centered design studies to tackle the identified usability challenges. This approach, which is advocated by other authors as an effective methodology for achieving a usable product, aims to design products that consider the needs and interests of end users [68-71]. Salinas et al [72] have reviewed the techniques and tools used in the successful redesigns of graphical user interfaces of software products following the user-centered design approach. While some of these techniques align with those used in our evaluation, the key lies in using these techniques throughout the design process instead of solely during product testing. Commonly reported methods of user testing include prototyping, pre- and postdesign interviews, heuristic evaluation, and surveys or questionnaires [72]. Therefore, upcoming research endeavors concerning the redesigned PRIMS should embrace a user-centered design methodology to guarantee the satisfaction of end users' needs and explore the integration of the aforementioned effective techniques. Moreover, future investigations could concentrate on crafting customizable interface options enabling users to tailor their interaction experience according to their individual capabilities and preferences. This might entail the incorporation of adjustable settings for font size, button layout, and navigation pathways such as voice activation or remote controllers to cater to a diverse range of users.

The current iteration of the PRIMS faces practicality challenges for home use. A more feasible adaptation would necessitate enhanced usability, reduced equipment costs, and minimal space requirements. Substantial updates are imperative to transform the PRIMS into a valuable home-based tool. This entails enhancing user-friendliness, optimizing the product to function seamlessly on common smartphone or tablet cameras, and refining the interface for a user-friendly experience.

It is important to emphasize that a direct score comparison between the PRIMS and a clinician was not conducted in this study. The PRIMS did assign scores to each movement in the motor test using an algorithm developed by PragmaClin Research Inc. However, it is essential to clarify that this study exclusively focused on usability and did not assess the validity of the scoring process. The validation of scoring algorithms remains a subject for future investigations.

Conclusions

In conclusion, the PRIMS currently exhibits several usability challenges that hinder its efficient use by individuals with PD. For the system to achieve successful implementation and gain broad acceptance, it is imperative to address these identified issues. Feedback from this study is being used to upgrade the PRIMS so that it better aligns with patients' needs. This study contributes significantly to the growing literature on usability testing, particularly emphasizing design nuances for systems tailored to those with movement disorders. Moving forward, it would be beneficial for future research to explore diverse interaction methods with digital devices, aiming to pinpoint optimal usability practices.

Acknowledgments

The authors would like to thank the Parkinson Society Newfoundland and Labrador for assistance with recruitment and the Mitacs Accelerate program for funding this project. They also extend their heartfelt gratitude to the participants, whose invaluable contributions have been instrumental in shaping this research endeavor. This study is one of several manuscripts that will collectively contribute to BB's PhD thesis.

Data Availability

The data sets generated during and analyzed during this study are available in the Memorial University Dataverse-Borealis repository [73].

Conflicts of Interest

BB and JT are employees of PragmaClin Research Inc, the company that developed the Parkinson's Remote Interactive Monitoring System. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Usability study script.
[PDF File (Adobe PDF File), 142 KB - humanfactors_v11i1e54145_app1.pdf]

Multimedia Appendix 2
System usability survey.
[PDF File (Adobe PDF File), 46 KB - humanfactors_v11i1e54145_app2.pdf]

Multimedia Appendix 3
Curved grading scale by Lewis and Sauro [55].
[PDF File (Adobe PDF File), 96 KB - humanfactors_v11i1e54145_app3.pdf]

References

1. Dorsey ER, Sherer T, Okun MS, Bloem BR. The emerging evidence of the Parkinson pandemic. J Parkinsons Dis 2018 Dec 18;8(s1):S3-S8. [doi: 10.3233/jpd-181474]
2. Ball N, Teo WP, Chandra S, Chapman J. Parkinson's disease and the environment. Front Neurol 2019 Mar 19;10:218 [FREE Full text] [doi: 10.3389/fneur.2019.00218] [Medline: 30941085]
3. Marras C, Beck JC, Bower JH, Roberts E, Ritz B, Ross GW, et al. Prevalence of Parkinson's disease across North America. NPJ Parkinsons Dis 2018 Jul 10;4:21 [FREE Full text] [doi: 10.1038/s41531-018-0058-0] [Medline: 30003140]
4. Wijeratne T, Grisold W, Trenkwalder C, Carroll W. World Brain Day 2020: move together to end Parkinson's disease. Lancet Neurol 2020 Aug;19(8):643. [doi: 10.1016/S1474-4422(20)30215-5] [Medline: 32702326]
5. Davie CA. A review of Parkinson's disease. Br Med Bull 2008;86:109-127. [doi: 10.1093/bmb/ldn013] [Medline: 18398010]
6. Cacabelos R. Parkinson's disease: from pathogenesis to pharmacogenomics. Int J Mol Sci 2017 Mar 04;18(3):551 [FREE Full text] [doi: 10.3390/ijms18030551] [Medline: 28273839]
7. Antonini A, Barone P, Marconi R, Morgante L, Zappulla S, Pontieri FE, et al. The progression of non-motor symptoms in Parkinson's disease and their contribution to motor disability and quality of life. J Neurol 2012 Dec;259(12):2621-2631. [doi: 10.1007/s00415-012-6557-8] [Medline: 22711157]
8. Statistics. Parkinson's Foundation. URL: https://www.parkinson.org/Understanding-Parkinsons/Statistics [accessed 2024-05-08]
9. Busis NA, Shanafelt TD, Keran CM, Levin KH, Schwarz HB, Molano JR, et al. Author response: burnout, career satisfaction, and well-being among US neurologists in 2016. Neurology 2017 Oct 10;89(15):1650-1651. [doi: 10.1212/WNL.0000000000004484] [Medline: 28993531]
10. Dall TM, Storm MV, Chakrabarti R, Drogan O, Keran CM, Donofrio PD, et al. Supply and demand analysis of the current and future US neurology workforce. Neurology 2013 Jul 30;81(5):470-478 [FREE Full text] [doi: 10.1212/WNL.0b013e318294b1cf] [Medline: 23596071]
11. Schapira AH, Obeso J. Timing of treatment initiation in Parkinson's disease: a need for reappraisal? Ann Neurol 2006 Mar;59(3):559-562. [doi: 10.1002/ana.20789] [Medline: 16489611]
12. Caslake R, Taylor KS, Counsell CE. Parkinson's disease misdiagnosed as stroke. BMJ Case Rep 2009;2009:bcr07.2008.0558 [FREE Full text] [doi: 10.1136/bcr.07.2008.0558] [Medline: 21686747]
13. DeMaagd G, Philip A. Parkinson's disease and its management: part 1: disease entity, risk factors, pathophysiology, clinical presentation, and diagnosis. P T 2015 Aug;40(8):504-532 [FREE Full text] [Medline: 26236139]
14. Morgan J, Sethi KD. Levodopa and the progression of Parkinson's disease. Curr Neurol Neurosci Rep 2005 Jul;5(4):261-262. [doi: 10.1007/s11910-005-0069-1] [Medline: 15987608]
15. Papapetropoulos SS. Patient diaries as a clinical endpoint in Parkinson's disease clinical trials. CNS Neurosci Ther 2012 May;18(5):380-387 [FREE Full text] [doi: 10.1111/j.1755-5949.2011.00253.x] [Medline: 22070400]
16. Martinez-Martin P, Rodriguez-Blazquez C, Forjaz MJ, Chaudhuri KR. Motor impairment and disability scales. In: Guide to Assessment Scales in Parkinson’s Disease. Tarporley: Springer Healthcare; Oct 17, 2014.
17. Ramaker C, Marinus J, Stiggelbout AM, Van Hilten BJ. Systematic evaluation of rating scales for impairment and disability in Parkinson's disease. Mov Disord 2002 Sep;17(5):867-876. [doi: 10.1002/mds.10248] [Medline: 12360535]
18. Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, Martinez-Martin P, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. Mov Disord 2008 Nov 15;23(15):2129-2170. [doi: 10.1002/mds.22340] [Medline: 19025984]
19. Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. Sci Data 2016 Mar 03;3:160011 [FREE Full text] [doi: 10.1038/sdata.2016.11] [Medline: 26938265]
20. Ossig C, Antonini A, Buhmann C, Classen J, Csoti I, Falkenburger B, et al. Wearable sensor-based objective assessment of motor symptoms in Parkinson's disease. J Neural Transm (Vienna) 2016 Jan;123(1):57-64. [doi: 10.1007/s00702-015-1439-8] [Medline: 26253901]
21. Silva de Lima AL, Hahn T, de Vries NM, Cohen E, Bataille L, Little MA, et al. Large-scale wearable sensor deployment in Parkinson's patients: the Parkinson@Home study protocol. JMIR Res Protoc 2016 Aug 26;5(3):e172 [FREE Full text] [doi: 10.2196/resprot.5990] [Medline: 27565186]

22. Post B, Merkus MP, de Bie RM, de Haan RJ, Speelman JD. Unified Parkinson's disease rating scale motor examination: are ratings of nurses, residents in neurology, and movement disorders specialists interchangeable? Mov Disord 2005 Dec;20(12):1577-1584. [doi: 10.1002/mds.20640] [Medline: 16116612]
23. Palmer JL, Coats MA, Roe CM, Hanko SM, Xiong C, Morris JC. Unified Parkinson's disease rating scale-motor exam: inter-rater reliability of advanced practice nurse and neurologist assessments. J Adv Nurs 2010 Jun;66(6):1382-1387 [FREE Full text] [doi: 10.1111/j.1365-2648.2010.05313.x] [Medline: 20546368]
24. Bhidayasiri R, Martinez-Martin P. Clinical assessments in Parkinson's disease: scales and monitoring. Int Rev Neurobiol 2017;132:129-182. [doi: 10.1016/bs.irn.2017.01.001] [Medline: 28554406]
25. Weintraub D, Comella CL, Horn S. Parkinson's disease--part 1: pathophysiology, symptoms, burden, diagnosis, and assessment. Am J Manag Care 2008 Mar;14(2 Suppl):S40-S48 [FREE Full text] [Medline: 18402507]
26. Larson DN, Schneider RB, Simuni T. A new era: the growth of video-based visits for remote management of persons with Parkinson's disease. J Parkinsons Dis 2021;11(s1):S27-S34 [FREE Full text] [doi: 10.3233/JPD-202381] [Medline: 33492246]
27. Dorsey ER, Achey MA, Beck CA, Beran DB, Biglan KM, Boyd CM, et al. National randomized controlled trial of virtual house calls for people with Parkinson's disease: interest and barriers. Telemed J E Health 2016 Jul;22(7):590-598 [FREE Full text] [doi: 10.1089/tmj.2015.0191] [Medline: 26886406]
28. Adams JL, Myers TL, Waddell EM, Spear KL, Schneider RB. Telemedicine: a valuable tool in neurodegenerative diseases. Curr Geriatr Rep 2020 Jun;9(2):72-81 [FREE Full text] [doi: 10.1007/s13670-020-00311-z] [Medline: 32509504]
29. Dorsey ER, Wagner JD, Bull MT, Rizzieri A, Grischkan J, Achey MA, et al. Feasibility of virtual research visits in fox trial finder. J Parkinsons Dis 2015;5(3):505-515 [FREE Full text] [doi: 10.3233/JPD-150549] [Medline: 26406130]
30. Espay AJ, Bonato P, Nahab FB, Maetzler W, Dean JM, Klucken J, et al. Technology in Parkinson's disease: challenges and opportunities. Mov Disord 2016 Sep;31(9):1272-1282 [FREE Full text] [doi: 10.1002/mds.26642] [Medline: 27125836]
31. Rovini E, Maremmani C, Cavallo F. How wearable sensors can support Parkinson's disease diagnosis and treatment: a systematic review. Front Neurosci 2017 Oct 06;11:555 [FREE Full text] [doi: 10.3389/fnins.2017.00555] [Medline: 29056899]
32. Dewey DC, Miocinovic S, Bernstein I, Khemani P, Dewey RB3, Querry R, et al. Automated gait and balance parameters diagnose and correlate with severity in Parkinson disease. J Neurol Sci 2014 Oct 15;345(1-2):131-138 [FREE Full text] [doi: 10.1016/j.jns.2014.07.026] [Medline: 25082782]
33. Wang J, Gong D, Luo H, Zhang W, Zhang L, Zhang H, et al. Measurement of step angle for quantifying the gait impairment of Parkinson's disease by wearable sensors: controlled study. JMIR Mhealth Uhealth 2020 Mar 20;8(3):e16650 [FREE Full text] [doi: 10.2196/16650] [Medline: 32196458]
34. Anand V, Bilal E, Ramos V, Naylor M, Demanuele C, Zhang H, et al. F62. Automatic detection of ON/OFF states in Parkinson disease patients using wearable inertial sensors. Clin Neurophysiol 2018 May;129(Supplement 1):e90. [doi: 10.1016/j.clinph.2018.04.225]
35. Lipsmeier F, Simillion C, Bamdadian A, Tortelli R, Byrne LM, Zhang YP, et al. A remote digital monitoring platform to assess cognitive and motor symptoms in Huntington disease: cross-sectional validation study. J Med Internet Res 2022 Jun 28;24(6):e32997 [FREE Full text] [doi: 10.2196/32997] [Medline: 35763342]
36. Joshi R, Bronstein JM, Keener A, Alcazar J, Yang DD, Joshi M, et al. PKG movement recording system use shows promise in routine clinical care of patients with Parkinson's disease. Front Neurol 2019 Oct 1;10:1027 [FREE Full text] [doi: 10.3389/fneur.2019.01027] [Medline: 31632333]
37. Chaudhuri KR, Hand A, Obam F, Belsey J. Cost-effectiveness analysis of the Parkinson's KinetiGraph and clinical assessment in the management of Parkinson's disease. J Med Econ 2022;25(1):774-782 [FREE Full text] [doi: 10.1080/13696998.2022.2080437] [Medline: 35593687]
38. StrivePD homepage. StrivePD. URL: https://www.strive.group/ [accessed 2024-05-08]
39. Wearable sensors - APDM wearable technologies. APDM Wearable Technologies. URL: https://apdm.com/wearable-sensors/ [accessed 2024-05-02]
40. Ramos V, Naylor M, Demanuele C, Zhang H, Amato S, Wacnik P, et al. T153. Wearable inertial sensors produce reliable endpoints in healthy volunteers and detect levodopa -induced changes in Parkinson’s disease patients. Clin Neurophysiol 2018 May;129(Supplement 1):e61. [doi: 10.1016/j.clinph.2018.04.154]
41. Piau A, Wild K, Mattek N, Kaye J. Current state of digital biomarker technologies for real-life, home-based monitoring of cognitive function for mild cognitive impairment to mild Alzheimer disease and implications for clinical care: systematic review. J Med Internet Res 2019 Aug 30;21(8):e12785 [FREE Full text] [doi: 10.2196/12785] [Medline: 31471958]
42. Pasluosta CF, Gassner H, Winkler J, Klucken J, Eskofier BM. An emerging era in the management of Parkinson's disease: wearable technologies and the internet of things. IEEE J Biomed Health Inform 2015 Nov;19(6):1873-1881. [doi: 10.1109/JBHI.2015.2461555] [Medline: 26241979]
43. Hergenroeder AL, Barone Gibbs B, Kotlarczyk MP, Kowalsky RJ, Perera S, Brach JS. Accuracy of objective physical activity monitors in measuring steps in older adults. Gerontol Geriatr Med 2018 Jun 17;4:2333721418781126 [FREE Full text] [doi: 10.1177/2333721418781126] [Medline: 29977979]
44. Heldman DA, Giuffrida JP, Cubo E. Wearable sensors for advanced therapy referral in Parkinson's disease. J Parkinsons Dis 2016 Jul 02;6(3):631-638. [doi: 10.3233/JPD-160830] [Medline: 27392872]

45. Ramdhani RA, Khojandi A, Shylo O, Kopell BH. Optimizing clinical assessments in Parkinson's disease through the use of wearable sensors and data driven modeling. Front Comput Neurosci 2018 Sep 11;12:72 [FREE Full text] [doi: 10.3389/fncom.2018.00072] [Medline: 30254580]
46. Schukat M, McCaldin D, Wang K, Schreier G, Lovell NH, Marschollek M, et al. Unintended consequences of wearable sensor use in healthcare. Contribution of the IMIA wearable sensors in healthcare WG. Yearb Med Inform 2016 Nov 10(1):73-86 [FREE Full text] [doi: 10.15265/IY-2016-025] [Medline: 27830234]
47. Folador JP, Vieira MF, Pereira AA, Andrade AO. Open-source data management system for Parkinson's disease follow-up. PeerJ Comput Sci 2021 Feb 17;7:e396 [FREE Full text] [doi: 10.7717/peerj-cs.396] [Medline: 33817042]
48. Hartson R, Pyla PS. The UX Book: Process and Guidelines for Ensuring a Quality User Experience. Amsterdam, The Netherlands: Elsevier Science; 2012.
49. Taktak A, Ganney P, Long D, Axell R. Clinical Engineering: A Handbook for Clinical and Biomedical Engineers. Amsterdam, The Netherlands: Elsevier Science; 2019.
50. Garzo A, Silva PA, Garay-Vitoria N, Hernandez E, Cullen S, Cochen De Cock V, et al. Design and development of a gait training system for Parkinson's disease. PLoS One 2018 Nov 12;13(11):e0207136 [FREE Full text] [doi: 10.1371/journal.pone.0207136] [Medline: 30418993]
51. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: a scoping review. Int J Med Inform 2019 Jun;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
52. Phillippi J, Lauderdale J. A guide to field notes for qualitative research: context and conversation. Qual Health Res 2018 Feb;28(3):381-388. [doi: 10.1177/1049732317697102] [Medline: 29298584]
53. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
54. Lewis JR, Sauro J. Item benchmarks for the system usability scale. J Usability Stud 2018 May;13(3):158-167 [FREE Full text]
55. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput Interact 2008;24(6):574-594. [doi: 10.1080/10447310802205776]
56. Cosford K, Briere J, Ambros B, Beazley S, Cartwright C. Effect of instructional format on veterinary students' task performance and emotional state during a simulation-based canine endotracheal intubation laboratory: handout versus video. J Vet Med Educ 2020 Apr;47(2):239-247. [doi: 10.3138/jvme.0618-077r1] [Medline: 31194627]
57. Shah RF, Gupta RM. Video instruction is more effective than written instruction in improving inhaler technique. Pulm Pharmacol Ther 2017 Oct;46:16-19. [doi: 10.1016/j.pupt.2017.08.005] [Medline: 28797611]
58. Mays KK, Katz JE, Groshek J. Mediated communication and customer service experiences. Psychological and demographic predictors of user evaluations in the United States. Period Polytech Soc Manag Sci 2022 Jan 03;30(1):1-11. [doi: 10.3311/PPso.16882]
59. Kargar N, Choobineh AR, Razeghi M, Keshavarzi S, Meftahi N. Posture and discomfort assessment in computer users while using touch screen device as compared with mouse-keyboard and touch pad-keyboard. Work 2018;59(3):341-349. [doi: 10.3233/WOR-182685] [Medline: 29630576]
60. Lackermair G, Kailer D, Kanmaz K. Importance of online product reviews from a consumer's perspective. Adv Econ Bus 2013;1(1):1-5. [doi: 10.13189/aeb.2013.010101]
61. Binyamin S, Rutter M, Smith S. The utilization of system usability scale in learning management systems: a case study of Jeddah community college. In: Proceedings of the 9th annual International Conference of Education, Research and Innovation. 2016 Presented at: ICERI2016; November 14-16, 2016; Seville, Spain. [doi: 10.21125/iceri.2016.2290]
62. Nielsen J. Why you only need to test with 5 users. Nielsen Norman Group. 2000 Mar 18. URL: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/ [accessed 2022-08-22]
63. Virzi RA. Refining the test phase of usability evaluation: how many subjects is enough? Hum Factors 1992;34(4):457-468. [doi: 10.1177/001872089203400407]
64. Moradian S, Krzyzanowska MK, Maguire R, Morita PP, Kukreti V, Avery J, et al. Usability evaluation of a mobile phone-based system for remote monitoring and management of chemotherapy-related side effects in cancer patients: mixed-methods study. JMIR Cancer 2018 Dec 21;4(2):e10932 [FREE Full text] [doi: 10.2196/10932] [Medline: 30578238]
65. Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs-principles and practices. Health Serv Res 2013 Dec;48(6 Pt 2):2134-2156 [FREE Full text] [doi: 10.1111/1475-6773.12117] [Medline: 24279835]
66. Green CA, Duan N, Gibbons RD, Hoagwood KE, Palinkas LA, Wisdom JP. Approaches to mixed methods dissemination and implementation research: methods, strengths, caveats, and opportunities. Adm Policy Ment Health 2015 Sep;42(5):508-523 [FREE Full text] [doi: 10.1007/s10488-014-0552-6] [Medline: 24722814]
67. Al-Tahat KS. Arabic translation, cultural adaptation and psychometric validation of the post-study system usability questionnaire (PSSUQ). Int J Hum Comput Interact 2021 May 11;37(19):1815-1822. [doi: 10.1080/10447318.2021.1913844]
68. Cornet VP, Toscos T, Bolchini D, Rohani Ghahari R, Ahmed R, Daley C, et al. Untold stories in user-centered design of mobile health: practical challenges and strategies learned from the design and evaluation of an app for older adults with heart failure. JMIR Mhealth Uhealth 2020 Jul 21;8(7):e17703 [FREE Full text] [doi: 10.2196/17703] [Medline: 32706745]

69. Graham AK, Munson SA, Reddy M, Neubert SW, Green EA, Chang A, et al. Integrating user-centered design and behavioral science to design a mobile intervention for obesity and binge eating: mixed methods analysis. JMIR Form Res 2021 May 10;5(5):e23809 [FREE Full text] [doi: 10.2196/23809] [Medline: 33970114]
70. Moldestad M, Petrova VV, Tirtanadi K, Mishra SR, Rajan S, Sayre G, et al. Improving the usability of written exposure therapy for therapists in the department of veterans affairs telemental health: formative study using qualitative and user-centered design methods. JMIR Form Res 2023 Nov 06;7:e47189 [FREE Full text] [doi: 10.2196/47189] [Medline: 37930747]
71. Sonney J, Cho EE, Zheng Q, Kientz JA. Refinement of a parent-child shared asthma management mobile health app: human-centered design study. JMIR Pediatr Parent 2022 Feb 17;5(1):e34117 [FREE Full text] [doi: 10.2196/34117] [Medline: 35175214]
72. Salinas E, Cueva R, Paz F. A systematic review of user-centered design techniques. In: Proceedings of the Design, User Experience, and Usability. Interaction Design. 2020 Presented at: DUXU 2020; July 19-24, 2020; Copenhagen, Denmark. [doi: 10.1007/978-3-030-49713-2_18]
73. John W, Bridges B, Taylor J. Evaluation of the Parkinson's remote interactive monitoring system (PRIMS): a usability study. JMIR Preprint posted online October 31, 2023 2023 [FREE Full text] [doi: 10.2196/54145]

Abbreviations

MDS-UPDRS: Movement Disorder Society–sponsored revision of the Unified Parkinson's Disease Rating Scale
PD: Parkinson disease
PRIMS: Parkinson's Remote Interactive Monitoring System
RA: research assistant
SUS: System Usability Scale

Edited by P Santana-Mancilla; submitted 31.10.23; peer-reviewed by I Straand, J Edwards; comments to author 22.01.24; revised version received 15.03.24; accepted 14.04.24; published 24.05.24.

Please cite as:
Bridges B, Taylor J, Weber JT
Evaluation of the Parkinson's Remote Interactive Monitoring System in a Clinical Setting: Usability Study
JMIR Hum Factors 2024;11:e54145
URL: https://humanfactors.jmir.org/2024/1/e54145
doi:10.2196/54145
PMID:38787603

Original Paper

Assessing the Relationship Between Digital Trail Making Test Performance and IT Task Performance: Empirical Study

Tanguy Depauw[1], MSc; Jared Boasen[1,2], PhD; Pierre-Majorique Léger[1], PhD; Sylvain Sénécal[1], PhD

[1]Tech3lab, HEC Montréal, Montréal, QC, Canada

[2]Faculty of Health Sciences, Hokkaido University, Sapporo, Hokkaido, Japan

Corresponding Author:
Jared Boasen, PhD
Tech3lab
HEC Montréal
5540 Av Louis-Colin
Montréal, QC, H3T 1T7
Canada
Phone: 1 514 340 6000
Email: jared.boasen@hec.ca

Abstract

Background: Cognitive functional ability affects the accessibility of IT and is thus something that should be controlled for in user experience (UX) research. However, many cognitive function assessment batteries are long and complex, making them impractical for use in conventional experimental time frames. Therefore, there is a need for a short and reliable cognitive assessment that has discriminant validity for cognitive functions needed for general IT tasks. One potential candidate is the Trail Making Test (TMT).

Objective: This study investigated the usefulness of a digital TMT as a cognitive profiling tool in IT-related UX research by assessing its predictive validity on general IT task performance and exploring its discriminant validity according to discrete cognitive functions required to perform the IT task.

Methods: A digital TMT (parts A and B) named Axon was administered to 27 healthy participants, followed by administration of 5 IT tasks in the form of CAPTCHAs (Completely Automated Public Turing tests to Tell Computers and Humans Apart). The discrete cognitive functions required to perform each CAPTCHA were rated by trained evaluators. To further explain and cross-validate our results, the original TMT and 2 psychological assessments of visuomotor and short-term memory function were administered.

Results: Axon A and B were administrable in less than 5 minutes, and overall performance was significantly predictive of general IT task performance ($F_{5,19}$=6.352; P=.001; Λ=0.374). This result was driven by performance on Axon B ($F_{5,19}$=3.382; P=.02; Λ=0.529), particularly for IT tasks involving the combination of executive processing with visual object and pattern recognition. Furthermore, Axon was cross-validated with the original TMT (P_{corr}=.001 and P_{corr}=.017 for A and B, respectively) and visuomotor and short-term memory tasks.

Conclusions: The results demonstrate that variance in IT task performance among an age-homogenous neurotypical population can be related to intersubject variance in cognitive function as assessed by Axon. Although Axon's predictive validity seemed stronger for tasks involving the combination of executive function with visual object and pattern recognition, these cognitive functions are arguably relevant to the majority of IT interfaces. Considering its short administration time and remote implementability, the Axon digital TMT demonstrates the potential to be a useful cognitive profiling tool for IT-based UX research.

(JMIR Hum Factors 2024;11:e49992) doi:10.2196/49992

KEYWORDS

Trail Making Test; user experience; cognitive profile; information technology; task performance; cognitive assessment; human factors; cognitive function; CAPTCHA

Introduction

Cognitive functional ability is a fundamental factor widely recognized to influence IT usability [1-3]. The classical approach to control for cognitive functional ability is to target participants according to general demographics based on age, education, or other factors [4,5]. However, this approach intrinsically precludes the ability to control for or assess how cognitive functional ability impacts IT usability in individual users, thereby limiting the extent which insight can be gained within a demographic or for an individual. Moreover, this approach is incongruent with the rapid advancement of IT toward products that adapt to individual user characteristics, thus necessitating a more granular understanding of individual cognitive abilities [6-8].

To obtain a granular characterization of individual cognitive function, hitherto, research has typically used cognitive assessment batteries [9-11]. Dumont et al [12] used the National Institutes of Health Toolbox, which is a battery of cognitive tests that can be completed in 40 minutes [13] to develop a cognitive analysis grid to be able to draw statistical parallels between the cognitive demands of an information systems interface and the performance of a user. Other batteries of tests were also used, such as the Kit of Factor-Referenced Cognitive Tests [10], which was used by Wagner et al [1] to study the impact of age on website usability and by Allen [14] in his research to study the combination of users' cognitive abilities and specific information system functionalities that can be implemented to create system usability. This battery is typically administered in 144 minutes [15]. Another approach for assessing individual cognitive ability is to use clinically administered tests such as the Montreal Cognitive Assessment (MoCA) or the Mini-Mental State Examination (MMSE). Although typically used in medical settings to evaluate cognitive impairment in patients with neurological disorders [9,11], MoCA and MMSE have been reportedly used to measure the cognitive abilities of participants in human-computer interaction experiments [3,16-18]. However, while detailed and accurate, these cognitive assessment batteries are too lengthy to practically administer during typical user experience (UX) testing time frames [19,20]. Furthermore, while clinically administered tests such as MoCA and MMSE are comparatively shorter than other assessment batteries, they require a trained administrator to administer and score the test [3]. This level of expertise may not always be available, particularly in UX research settings where mostly nonclinically trained research personnel are conducting the experiments.

Correspondingly, there have been calls from across health, UX, and IT domains for a more practical yet accurate means of assessing cognitive function [12,21,22]. One solution would be to identify a short test with reduced scope but which nevertheless targets cognitive functions important for using IT. Based on research conducted to understand the impact of cognitive functions on the use of technology by older people [23,24], and on existing models of cognitive architecture in human-computer interaction [25], we identified 5 key cognitive functions important for IT use: visual perception, motor function, executive function, inhibitory control, and working memory. Visual perception is important for finding relevant information cues on a web page [23]. Motor functions are involved in tasks such as data entry using the keyboard, navigation using the mouse, or other tool to perform a digital task [26]. Executive functions come into play in order to make decisions and prioritize action [23]. Inhibitory control, also called "response inhibition" [27], is the functional ability to inhibit or override motor commands or other executive processing, such as when an external stimulus interferes with goal-driven behavior as in a task-switching situation [28,29]. Finally, short-term or working memory capacity may be important in IT task performance, for example, for remembering options or system output at a later stage [23].

One potential preexisting cognitive assessment candidate that targets these cognitive functions related to IT use is the Trail Making Test (TMT). First developed for the Army Individual Test Battery [30], the TMT is one of the most widely used instruments in neuropsychological assessment as an indicator of cognitive processing speed and executive functioning [31-35]. Many studies have been conducted to determine which cognitive abilities are engaged during the completion of this 2-part test (TMT-A and TMT-B). After a comprehensive review of the literature on the topic, Sánchez-Cubillo et al [36] explored the contributions of certain cognitive functions and found that part A of the TMT (TMT-A) mainly requires visual-perceptual abilities, and that part B (TMT-B) reflects primarily working memory, executive function, and task-switching ability. Finally, although its contribution in the TMT has been questioned by the study of Sánchez-Cubillo et al [36], it is interesting to note that psychomotor ability has been mentioned numerous times as one of the abilities required for both parts of the TMT (Groff and Hubble [37] in both parts, Schear and Sato [38], Gaudino et al [39], and Crowe [40] in part B). The primary objective of this study was to test the validity of using the TMT as a cognitive profiling tool to predict or explain the variance in IT task performance. With an interest in a practical tool for cognitive profile assessments in UX testing of digital artifacts, we chose to use a digital version of the TMT. To further support and explain our results, we additionally cross-validated the digital TMT with the original TMT, a visual search task assessing visuomotor processing [41,42], and a hidden path learning task assessing visuomotor-processing speed, spatial working memory, and error-monitoring ability [43]. We had two hypotheses: (1) TMT times would be predictive of general IT task performance and (2) that the predictive power of the TMT would be stronger for tasks requiring the use of cognitive functions that are congruent with those assessed by the TMT.

Methods

Sample

To test our hypothesis, we conducted a laboratory experiment with 27 healthy participants (12 men and 15 women), between 18 and 36 (mean 24, SD 4.22) years of age, who were mostly university students (n=22, 85%).

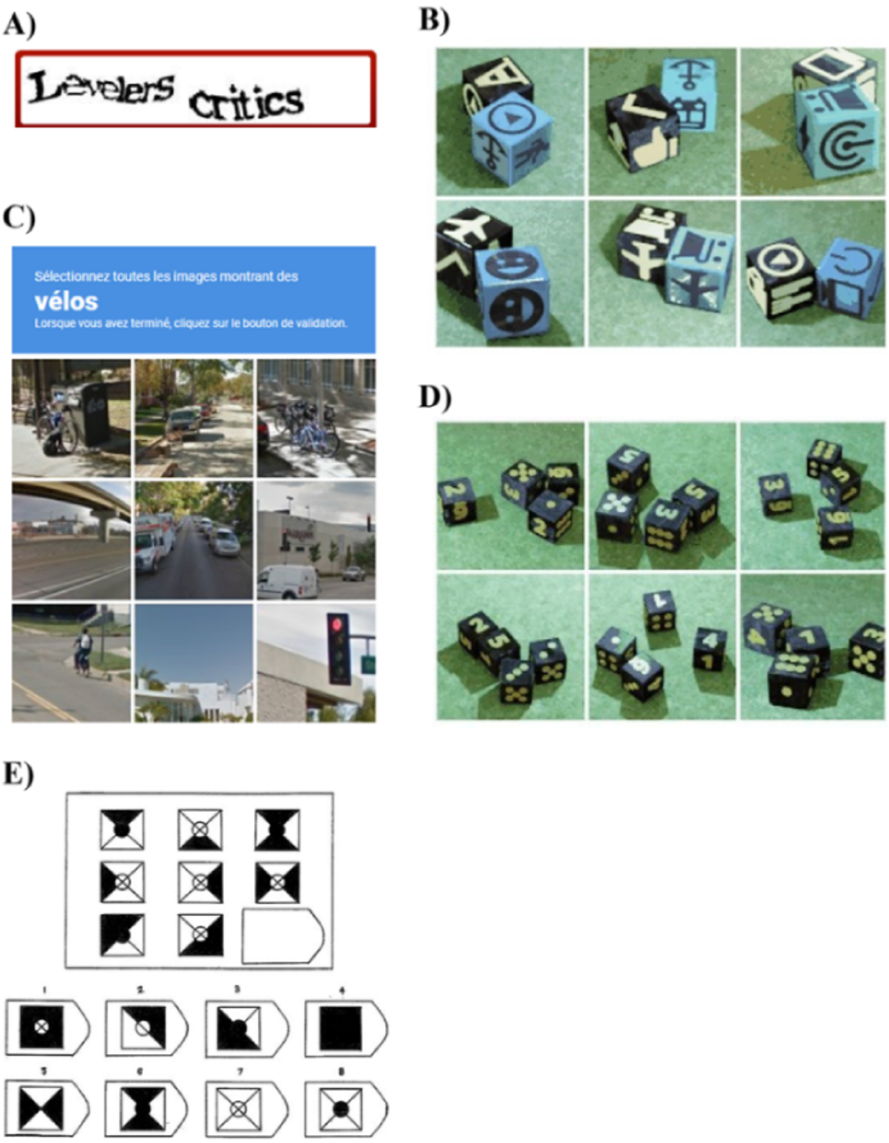
Ethical Considerations

Written informed consent was obtained from all subjects via a signed form at the beginning of the experiment. This project was approved by our institution's research ethics committee (#2021-4108). A monetary compensation of CAD $25 (US $18.35) was provided to each subject upon completion of the experiment. Data from 1 subject were lost due to technical issues, thus leaving data from 26 participants available for analysis. All data were anonymized prior to analysis and stored in encrypted servers only accessible by authorized researchers.

IT Tasks

Two types of general IT tasks were used in the experiment. One type of IT task was based on CAPTCHA (Completely Automated Public Turing Tests to Tell Computers and Humans Apart). This type of Turing test is widely used in IT to ensure the cybersecurity of many internet services, as they prevent a number of attacks from automated programs (often referred to as bots), by distinguishing legitimate users from computer bots while requiring minimal effort by the human user [44]. Four CAPTCHAs were based on typical existent CAPTCHAs and included Google reCAPTCHA (Google), pictogram recognition (PicRec), numerical recognition (NumRec), and text recognition (Text). A Fifth task was taken from Raven's Progressive Matrices (RPM) and presented in a CAPTCHA format. RPM are a collection of widely used standardized intelligence tests consisting of analogy problems in which a matrix of geometric figures is presented with 1 entry missing, and the correct missing entry must be selected from a set of answer choices [45]. A 3×3 RPM was selected as it was considered that it offered the best trade-off between cognitive effort and the time required to complete it. The final 5 IT tasks, shown in Figure 1, were embedded on a Qualtrics questionnaire. For this study, we targeted IT task completion time, measured as the time from the display of each task to when subjects responded and pressed the "next" button, based on 30 fps screen recordings.

Figure 1. The 5 information technology tasks. (A) Text-based Completely Automated Public Turing tests to tell Computers and Humans Apart (CAPTCHA): subjects had to type the 2 words in an input field below the text image. (B) Pictogram recognition CAPTCHA: subjects had to recognize and click on the image showing the 2 dice with the same pictogram on the top face. (C) Google reCAPTCHA: subjects had to recognize and click on the images showing the bicycles. (D) Number recognition CAPTCHA: subjects had to recognize and click on the image showing dice summing to 14 on the top faces (numerals and dots combined). (E) Raven's Progressive Matrix: subjects had to click from among the 8 proposed images the one which most appropriately fit in the missing corner of the basic matrix.



The other type of IT task was a website design evaluation to assess perceived usability using Aladwani and Palvia's [46] user-perceived web quality measurement scale. Screenshots of the home pages of the following 5 websites were used: Vignerons d'Exception [47], Renaud-Bray [48], LesPAC [49], [50], and [51]. One website was presented subsequent to each CAPTCHA. Participants were told that the website evaluation was the primary task of the experiment and that the CAPTCHAs were present as a security measure to access our database housing the website screenshots. However, the website evaluations were actually dummy tasks, and participant responses were not analyzed. The IT tasks really targeted and analyzed in this study were the CAPTCHAs.

Cognitive Function Characterization of CAPTCHAs

The principal reason CAPTCHAs were chosen as our general IT tasks is because they are ubiquitous in IT and because they are often distinguishable from one another according to task-specific demands such as math, 3D orientation, text recognition, and visual search, suggesting that different underlying cognitive processing required them. However, there is a paucity of studies regarding the examination of the specific cognitive functions of CAPTCHAs. Therefore, we formed a panel of 11 trained, nonexpert evaluators to rank the selected CAPTCHAs on a 5-point agreement scale according to the 5 cognitive functions mentioned in the Introduction section, which have been deemed relevant to IT tasks and the TMT: visuospatial perception, motor function, executive function, inhibitory control, and working memory. The evaluation scores permitted each CAPTCHA to be assigned a rank according to the extent the cognitive functions required to perform it overlapped with those of the TMT. In order of highest to lowest alignment, the rankings were as follows: (1) RPM, (2) NumRec, (3) PicRec, (4) Google, and (5) Text, as shown in Table 1. For details of how this evaluation was conducted and how the process was validated, see Multimedia Appendix 1.

Table 1. Convergence ranks of IT tasks with the TMT[a].

IT task	RPM[b] (E)	NumRec[c] (D)	PicRec[d] (B)	Google (C)	Text (A)
Executive function, mean (SD)	5.00 (0.00)	4.91 (0.30)	4.45 (1.04)	4.45 (1.04)	3.82 (1.4)
Visual object recognition, mean (SD)	4.09 (0.70)	4.27 (0.65)	4.64 (0.67)	4.82 (0.60)	4.18 (1.17)
Visual pattern recognition, mean (SD)	4.91 (0.30)	4.45 (0.93)	4.64 (0.67)	3.82 (0.98)	4.64 (0.67)
Working memory, mean (SD)	4.18 (0.60)	3.91 (1.38)	2.91 (1.51)	2.45 (1.21)	2.73 (1.27)
Evaluation score for reliable convergent dimensions, mean (SD)	4.55 (0.48)	4.39 (0.42)	4.16 (0.84)	3.89 (1.04)	3.84 (0.81)
Convergence rank with TMT following the evaluation[e]	1	2	3	4	5

[a]TMT: Trail Making Test.

[b]RPM: Raven's Progressive Matrices.

[c]NumRec: numerical recognition.

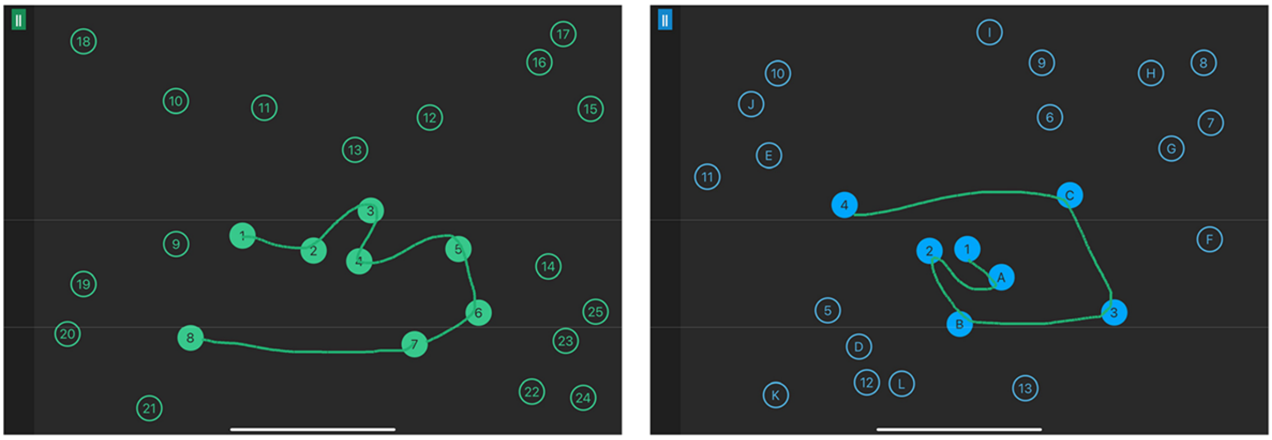
[d]PicRec: pictogram recognition.

[e]Based on the average evaluation scores of IT tasks on the reliable cognitive dimensions considered convergent with the TMT. A, B, C, D, and E refer to the labels of the IT tasks presented in Figure 1.

Digital TMT

Because we are interested in cognitive assessment for UX testing of IT and because it was convenient to present all the tasks on the same device, we chose to use a digital version of the TMT called "Axon" (Language Research Development Group). This version emulates the original TMT as an iPad app, allowing the user to draw the trail on the touch screen with 1 finger. The 2 parts (A and B) of the TMT were completed, each with 25 circles to connect. Axon TMT was designed with a canvas generation algorithm, meaning that the test canvas for each subject for each TMT A and B was different. As shown in Figure 2, both tests were presented in full screen on the iPad with 25 circles of 1-cm diameter placed randomly on the digital canvas in a homogeneous way. The rules of Axon were identical to those of the original TMT, as outlined by Bowie and Harvey [52]. Participants had to connect the circles in ascending order: from 1 to 25 for part A and from 1 to 13 for part B, alternating numbers and letters in ascending order (ie, 1, A, 2, B, 3, C, etc). Errors such as lifting the finger off the screen, crossing trails, or connecting a wrong circle resulted in the line for the latest segment to be automatically erased and subjects had to return to the last successfully reached circle in order to continue. The measures chosen for this study were the completion time for each of the 2 parts of the test, from the moment the layout was displayed until the last circle was reached. These measures were exported from the app after the completion of the study and used in our statistical analyses.

Figure 2. Screenshots of Axon A and Axon B. Subjects had to draw to connect the circles in ascending order (from 1 to 25 for part A and from 1 to 13 and A to L for part B, alternating numbers and letters) on a single line, without crossing paths or lifting their finger from the screen. In case of errors in drawing, the app automatically guided subjects back to the last correct circle from which they continued the test.



Cross-Validation of the Digital TMT

Overview

To better support and explain our results, we cross-validated Axon with the original TMT and a working memory and a visual search task.

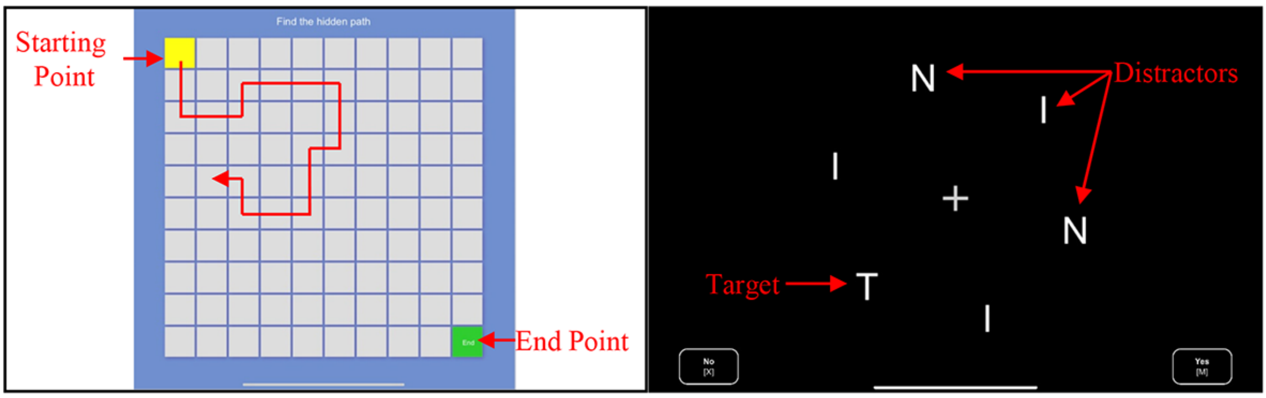
Original TMT

The original TMT was administered as outlined by Bowie and Harvey [52] at the end of the study. The practice step was skipped in the interest of time and with the knowledge that the subject had already performed the digital TMT earlier in the study.

Hidden Path Learning Task

To cross-validate Axon's ability to measure working memory and spatial ability, we administered a hidden maze learning task, based on the Groton Maze Learning Test developed by Pietrzak et al [43]. Our task was called the "hidden path learning" task and was based on a 10 × 10 grid. Five trials were administered on the iPad via the Cognition Lab platform (BeriSoft, Inc), following similar guidelines as the Groton Maze Learning Test [43]. The hidden path learning task is particularly targeted at working memory, as the user has to call on it to navigate between tiles and remember any errors they may have made before [53,54]. Correspondingly, working memory ability is associated with the extent to which completion time decreases over trials, revealing a learning curve. Thus, the metrics used for these analyses were the difference between the completion times of each consecutive trial on the task. A depiction of the hidden path learning task is shown in Figure 3 (left). Measures were automatically collected on the Cognition Lab server.

Figure 3. Cross-validation tasks. In the hidden path learning 10×10 matrix (left), subjects had to go from the yellow starting point to the green end point 1 tile at a time. In the visual search task (right), there were 6 items, with 1 target and 5 distractors. In the I+N sequence (shown), participants had to touch "Yes" at the bottom-right if they saw the target, "No" at the bottom-left otherwise.



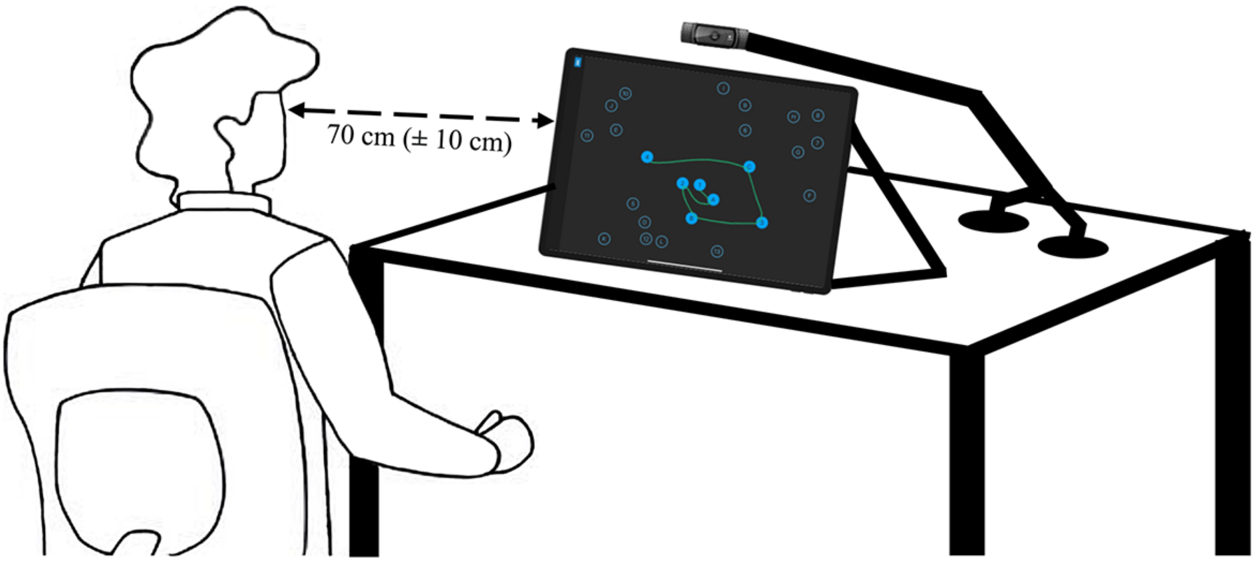
Visual Search Task

To cross-validate Axon's ability to measure visuomotor function, we administered a visual search task on the Cognition Lab platform (BeriSoft, Inc). This task was based on the work by Treisman and Gelade [42] and involved finding a target among distractors. Participants had to touch the right side of the screen when they saw the target, the left side otherwise, therefore involving visual and psychomotor response ability. Three stimuli configurations were used, with 3 distractor sets. Configurations were displayed with 24 trials for each stimulus, leading to a total of 72 trials. For each trial, 3, 6, or 9 symbols were displayed (letters or shapes), with even and randomized distribution among each stimulus sequence. A depiction of this task is shown in Figure 3 (right). Again, measures were automatically collected on the Cognition Lab server. Reaction times were used for the present analyses.

Procedures

Upon arrival and after signing the informed consent form, subjects were asked to sit on a chair facing the iPad Air (fourth generation) running on iPadOS 15.3 (Apple Inc) placed on a desk and were asked to adjust the chair's height so that they were comfortable using the iPad, and they were within the camera recording frame. The experimental setup is presented in Figure 4. They were asked to move the chair closer or further away to maintain an approximate distance of 70 (±10) cm between their eyes and the iPad screen to give enough space for hand movement during the tasks. The camera was fixed independently from the iPad to avoid unwanted movements on the video when the participant presses the screen while doing the tasks. After a presentation of the study and the tools used, the participants were asked to complete the 2 parts of the TMT (A and then B) on the Axon app. Task instructions were given in a protocol format to ensure that all participants received the same instructions and that the data would be comparable. Participants were verbally and visually guided through the rules of the TMT using a tutorial embedded in the app.

Figure 4. Experimental setup diagram. The subject was seated at a chair in front of a desk where the iPad Air 4 was placed. A Logitech C920 camera was independently fixed to the desk via a camera stand and duct tape.



After completing parts A and B on the Axon app, participants were administered the hidden path learning and the visual search tasks. Then, participants commenced the IT task portion of the experiment. As previously mentioned, participants were told that the primary objective was to evaluate 5 interfaces of more or less popular websites, each interface being on a secure server accessible only after the completion of a CAPTCHA. Thus, subjects completed a CAPTCHA, observed a web interface for a few minutes, and then completed the user-perceived web quality measurement scale [46]. This sequence was repeated 5 times, with the tasks presented in random order, each preceded by a distinct CAPTCHA. At the end of the study, for ethical reasons, subjects were told orally that they were in fact being evaluated on their performance on the CAPTCHAs.

Statistical Analyses

To test the ability of the Axon TMT to predict performance on the 5 CAPTCHA IT tasks, a repeated-measures multivariate analysis of covariance (RM MANCOVA) was performed with Axon A completion time and Axon B completion time as independent predictors and the completion time for each of the 5 IT tasks as the dependent covariates.

To further interpret our results, we tested the relationship between Axon TMT completion times and visuomotor function by performing an RM MANCOVA with Axon A and Axon B times as independent predictors and the mean reaction time of each of the 3 visual search tasks (the shape of an arrow as a target among the triangle shapes as distractors, the letter T as a target among the letters I and N as distractors, and the letter T as a target among the letters I and Z as distractors) as the dependent covariates. In addition, we tested the relationship between Axon TMT completion times and working memory function by performing an RM MANCOVA with Axon A and Axon B times as independent predictors and the difference between the completion time of each consecutive trial on the hidden path learning task as the dependent covariates. Finally, we cross-validated the relationship between the Axon TMT and the original TMT using 2 Pearson correlation tests, 1 each for tests A and B.

For all RM MANCOVAs performed in the analysis, omnibus results and multivariate results for each independent predictor are reported. In the case of significant multivariate results, simple main effects based on parameter estimates are reported for dependent covariates, which were significantly predicted by Axon.

All statistical analyses were conducted using the IBM SPSS Statistics software (version 28.0.1.1; IBM Corp) with a threshold for statistical significance set at $P \leq .05$, using the Bonferroni correction to adjust for multiple comparisons.

Results

Axon TMT Cross-Validation

Axon Versus Original TMT

The mean scores of Axon A and B were 48.04 (SD 25.80) and 56.88 (SD 25.53) seconds, respectively. The mean scores on the original TMT A and B were 29.22 (SD 12.26) and 51.62 (SD 19.07) seconds, respectively. Pearson correlation tests revealed that Axon is highly correlated with TMT results, with a significant positive correlation between Axon A and TMT A (r=0.688; P_{corr}=.001) and a significant positive correlation between Axon B and TMT B (r=0.505; P_{corr}=.017).

Axon TMT Versus Hidden Path Learning

The difference in consecutive trial times was (2–1) –29.87 (17.70), (3–2) –5.48 (6.01) seconds, (43) –4.30 (4.80) seconds, and (5–4) –1.50 (4.25) seconds. The omnibus test of the RM MANCOVA revealed that Axon A and Axon B combined are significant to explain the variance in the decrease in completion times across consecutive trials ($F_{4,20}$=4.119; P=.01; Λ=0.548). However, multivariate results revealed that the decrease in completion times across trials was not predicted by Axon A ($F_{4,20}$=1.923; P=.15; Λ=0.722) or Axon B ($F_{4,20}$=1.106; P=.38; Λ=0.819) alone. Thus, a predictive relationship appears to exist between Axon and working memory in the hidden path learning task as a function of Axon A and B combined.

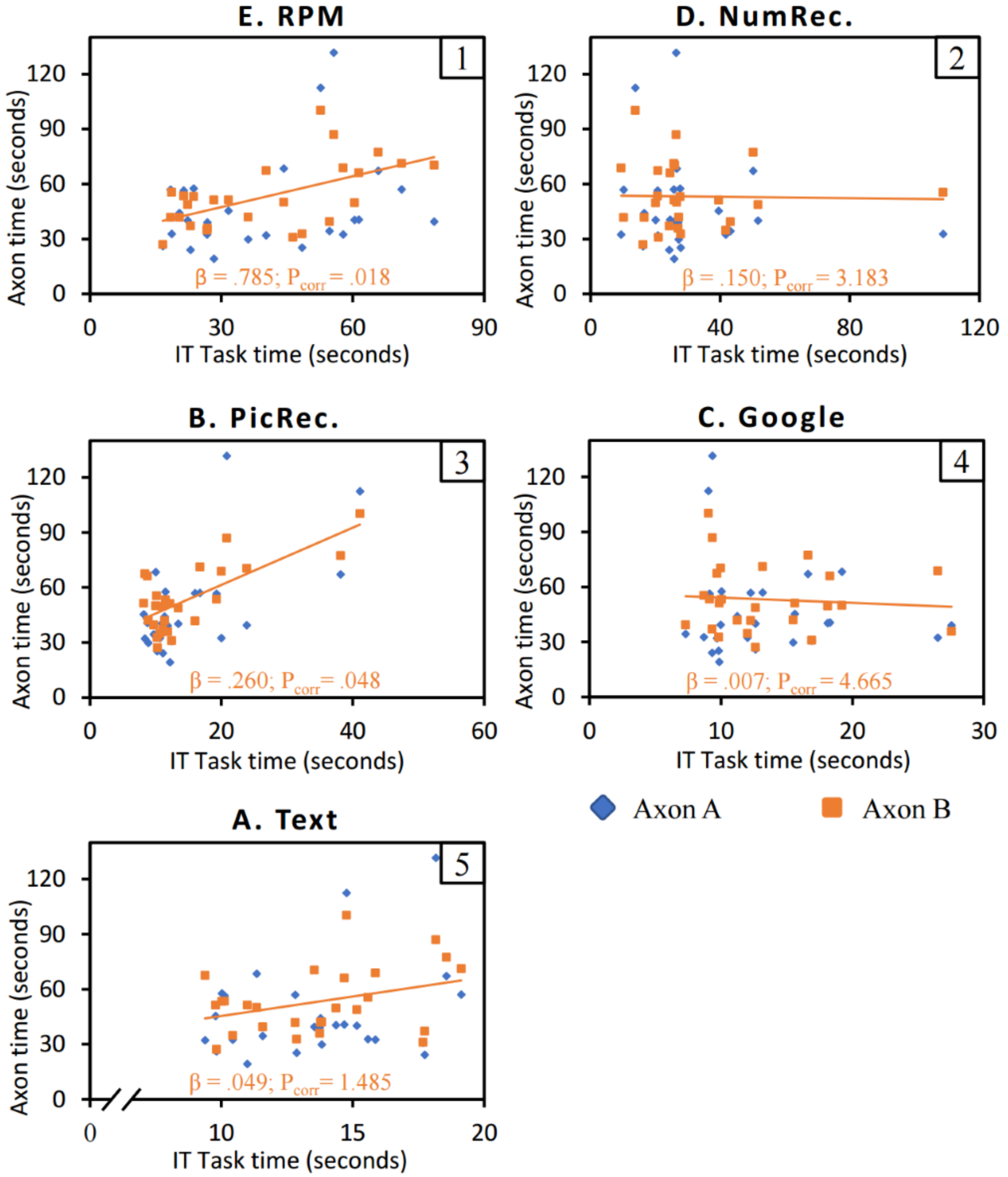
Axon TMT Versus Visual Search

Reaction times for the T among letters I and N, T among letters I and Z, and arrow among triangles were 0.80 (0.14) milliseconds, 0.78 (0.15) milliseconds, and 0.68 (0.14) milliseconds, respectively. The omnibus test of the RM MANCOVA revealed that Axon A and Axon B combined significantly explained the variance in visuomotor function assessed with reaction time to the 3 stimuli in the visual search task ($F_{3,21}$=3.125; P=.048; Λ=0.691). Multivariate results revealed that this result was driven mainly by Axon A ($F_{3,21}$=3.220; P=.043; Λ=0.685) rather than Axon B ($F_{3,21}$=0.502; P=.69; Λ=0.933). Parameter estimates revealed that Axon A was marginally significantly predictive of reaction times to the letter T among letters I and N stimulus (β=3.573; t_{21}=2.767; P_{corr}=.055) and significant for letter T among letters I and Z (β=4.353; t_{21}=3.156; P_{corr}=.02) and arrow among triangles (β=3.725; t_{21}=3.158; P_{corr}=.02) stimuli.

Axon TMT Predicts Overall IT Performance

The primary hypothesis assumed that there was a positive predictive relationship between TMT performance and IT task performance. The omnibus test of the RM MANCOVA revealed that Axon A and Axon B combined significantly explain the variance in IT tasks performance ($F_{5,19}$=6.352; P=.001; Λ=0.374), thereby supporting the primary hypothesis. Multivariate results revealed that this effect was driven by performance on Axon B ($F_{5,19}$=3.382; P=.03; Λ=0.53). Figure 5 shows the distribution of Axon completion times in relation to IT task completion times.

Figure 5. Distribution of Axon A and B completion times in relation with the completion times of the 5 IT tasks (N=26). Axon B trendlines and parameter estimates (β and P) show the relationship between Axon B IT task performance. Number in upper right corner of plot area is hypothesized convergence rank (Table 1). IT: information technology; NumRec: numerical recognition; PicRec: pictogram recognition; RPM: Raven's Progressive Matrices. Letters A through E refer to the labels used for each task in Figure 1.



Axon TMT Better Predicts Performance on Convergent IT Tasks

The second hypothesis assumed that the predictive relationship between TMT performance and IT task performance would be stronger if the cognitive abilities involved in the performance were congruent. To test our hypothesis, we analyzed the parameter estimates for the multivariate results of Axon B. These revealed that Axon B was significantly predictive of IT task C (RPM task; β=.785; t_{19}=3.240; P_{corr}=.018) and IT task B (PicRec task; β=.260; t_{19}=2.824; P_{corr}=.048). However, IT task D (number recognition task), which was rated the second most congruent task with Axon, was not significantly predicted by Axon B (β=.150; t_{19}=0.479; P_{corr}=3.183). Our secondary hypothesis is therefore partially supported. These results are shown in Figure 5, where the effects of individual factors of Axon B on performance on IT tasks are represented (β and *P* values).

Discussion

Principal Findings

Cognitive functional ability may well affect task performance in UX and other research experimentation, leading to variance in performance measures among the target population and confounding the effects of experimental factors. Although detailed cognitive assessment batteries exist and can be used to control intersubject differences in cognitive abilities [12], they are not time efficient and thus impractical to implement within typical experimental time frames. Here, this study tested the validity of using the Axon TMT, which takes only a few minutes to administer, to predict or explain the variance in IT task performance in an age-homogenous subject population.

The mean age of the subject population of this sample was 24 (SD 4.22) years. This is typical of many research studies, UX related or otherwise, relying on student recruitment through the parent institution [55-57]. Despite the relatively low SD of age, the SD in Axon TMT scores was broad, at 25.80 (mean 48.04) and 25.53 (mean 56.9) seconds, respectively, for Axon A and B, suggesting a large distribution of cognitive functional abilities among this age-homogeneous neurotypical population. Notably, the means and SDs for the Axon TMT, particularly for Axon A, were higher than what is typically reported in the literature for neurotypical subjects in this age bracket [58-60]. This may be due to the fact that, unlike in the implementation of the paper-based TMT, subjects did not practice a mini version of the test before performing Axon A or B. Thus, some portion of the time taken to complete the test must be attributable to familiarization with task demands. This would also explain why the mean scores for Axon B, whose task demands are similar to Axon A in many respects, are closer to typically reported TMT B means. Nevertheless, for the purposes of this study, it is not absolute Axon TMT scores that are important. Rather, it is the relative distribution of the variance in Axon scores and their correlation to other metrics that is essential. To that end, both Axon A and B significantly correlated to their respective paper-based TMT counterparts showed a combined predictive validity toward working memory via the hidden path learning task. Furthermore, it was Axon A, not B, which was the predominant driver of the significant correlation with visual search performance. This is logical, as the visual search task does not involve working memory–related processing [42,61]. Instead, it requires an emphasis on target identification, cognitive control, and motor output, precisely the dominant cognitive functions involved in TMT A [36,39,40]. Thus, far from being problematic, implementing Axon A and B without a preliminary minitest for practice was time-efficient and yielded a reliable distribution of scores, which could be cross-correlated with expected cognitive functions.

This cross-validation lends credibility to our observation that Axon A and B combined were significantly predictive of IT task performance, supporting our primary hypothesis. Interestingly, for the IT tasks chosen, it was Axon B that appeared as the stronger driver of predictive validity, suggesting that it may be more powerful in capturing the executive decision-making involved in an ecologically valid IT task. Moreover, simple main effects tests revealed that Axon B significantly correlated with 2 out of the top 3 tasks ranked as requiring congruent cognitive functions as the TMT, thereby partially supporting our secondary hypothesis. Contrary to our expectations, the NumRec task, which had the second-highest congruence rank, was not significantly correlated with Axon B. We speculate that the confound here relates to the underlying mathematical operations involved in solving that CAPTCHA. Although raters classified this as executive decision-making, it certainly can be said that neither TMT A nor TMT B requires arithmetic. Therefore, there must be cognitive processes involved that are simply not recruited during the performance of the TMT, which our ranking system was not granularized enough to capture, hence explaining the lack of correlation between the NumRec task and Axon B. Meanwhile, Axon B was most strongly correlated with the RPM and PicRec task, suggesting that it is well suited for tasks involving visual pattern and object recognition in combination with higher-order executive processing to orient this visual information. These kinds of processing are arguably crucial for interface navigation, virtual reality, gaming, or using simulators, which are extremely common IT tasks investigated in UX research [62-64]. Thus, while Axon does appear to be better aligned with IT tasks involving convergent cognitive processing, such tasks may well comprise a major proportion of those studied in UX research.

Finally, there are a few points worth emphasizing. First, the complete administration of Axon took less than 5 minutes, far shorter than the strategy used by Dumont et al [12] or any other cognitive assessment that we are aware of. Second, considering Axon's ability to differentiate from among an age-homogeneous neurotypical population, it would likely perform even better among populations where a larger variance in cognitive function would be expected, such as in older adults, children, stroke survivors, or other individuals with atypical cognitive function. This is important because understanding how to design appropriate and accessible IT for these populations has become a topic of increasing concern in UX research [65-67]. Moreover, Axon is suitable for remotely moderated experimentation, a popular strategy since the COVID-19 pandemic [68] and one that mitigates subject recruitment challenges for all population types. Finally, the current advancement in technology, particularly in the field of artificial intelligence, is trending toward a more personalized and user-centric approach, adapting technology to individual user characteristics such as preferences and interests [8,69,70]. Part of this personalization could be to tailor technology according to the cognitive abilities of users. Axon could potentially facilitate this advancement, serving as a quick and reliable metric to train the artificial intelligence technology adaptation algorithm.

Limitations

There are some limitations that should be acknowledged with this study. First, because the Axon app is designed to produce TMT canvases according to an algorithm with every test instance, the Axon A and B canvas layouts were not constant across subjects. This means that some of the variance in Axon A and B times is intrinsically attributable to factors such as differences in the straight-line drawing path length of the test or the extent of visual interference between each drawing segment. On the other hand, the fact that Axon A and B were significantly cross-validated with the original TMT and the visual search and hidden path learning tasks in spite of canvas layout differences between participants suggests that the variance these differences cause is small and does not detract from the use of Axon as a cognitive profiling tool in UX testing. Second, this study tested the predictive validity of Axon on simple and discrete IT tasks. This was necessary as a proof of concept for our hypotheses. However, readers should use caution when generalizing the present results. Further research is needed to investigate the extent to which Axon retains predictive validity for more complex IT tasks in different contexts and across various user demographics, including neuroatypical and cognitively impaired users.

Conclusions

This study tested the ability of the Axon digital TMT to predict performance on discrete IT tasks. The results indicate that variance in IT task performance among an age-homogenous neurotypical population can be related to intersubject variance in cognitive function as assessed by Axon. Although the findings suggest that Axon's predictive validity may be strongest for IT tasks involving the combination of decision-making with visual object and pattern recognition, these types of cognitive processing would arguably be relevant to the majority of IT interfaces. Considering its short administration time and remote implementability, the Axon digital TMT has the potential to be a useful cognitive profiling tool for IT-based UX research.

Acknowledgments

The authors would like to thank David Brieugne, Salima Tazi, Xavier Côté, and all the operational staff at the Tech3Lab for their operational assistance in the execution of this study. This study was made possible thanks to funding from NSERC (ALLRP 571020-21), Prompt (176_Sénécal-BNC, CAE, CN, SRC_2021.02), and Mitacs (IT18280).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Evaluation of CAPTCHAs–congruent cognitive functions in the information technology tasks.

[DOCX File , 17 KB - humanfactors_v11i1e49992_app1.docx]

References

1. Wagner N, Hassanein K, Head M. The impact of age on website usability. Computers in Human Behavior 2014 Aug;37:270-282. [doi: 10.1016/j.chb.2014.05.003]
2. Karahoca D, Karahoca A, Güngör A. Assessing effectiveness of the cognitive abilities and individual differences on e-learning portal usability evaluation. New York, NY, United States: Association for Computing Machinery; 2008 Presented at: Proceedings of the 9th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing; June 12-13, 2008; Gabrovo, Bulgaria. [doi: 10.1145/1500879.1500942]
3. Martins AI, Silva AG, Pais J, Cruz VT, Rocha NP. The impact of users' cognitive function on evaluator perceptions of usability. Sci Rep 2022 Aug 12;12(1):13753 [FREE Full text] [doi: 10.1038/s41598-022-17441-3] [Medline: 35962003]
4. Aykin NM, Aykin T. Individual differences in human-computer interaction. Comput Ind Eng 1991;20(3):373-379. [doi: 10.1016/0360-8352(91)90009-U]
5. Lewis JR. Usability testing. In: Salvendy G, editor. Handbook of Human Factors and Ergonomics, Fourth Edition. Hoboken, NJ: John Wiley & Sons; 2012:1267-1312.
6. Ji H, Yun Y, Lee S, Kim K, Lim H. An adaptable UI/UX considering user's cognitive and behavior information in distributed environment. Cluster Comput 2017;21(1):1045-1058. [doi: 10.1007/s10586-017-0999-9]
7. Andreessen LM, Gerjets P, Meurers D, Zander TO. Toward neuroadaptive support technologies for improving digital reading: a passive BCI-based assessment of mental workload imposed by text difficulty and presentation speed during reading. User Model User-Adap Inter 2020;31(1):75-104 [FREE Full text] [doi: 10.1007/s11257-020-09273-5]
8. Gajos KZ, Weld DS, Wobbrock JO. Automatically generating personalized user interfaces with Supple. AI 2010;174(12-13):910-950 [FREE Full text] [doi: 10.1016/j.artint.2010.05.005]
9. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
10. Ekstrom RB, French JW, Harman HH, Dermen D. Manual for Kit of Factor-Referenced Cognitive Tests, 1976. Princeton, NJ: Education Testing Service; 1976.
11. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
12. Dumont L, Chénier-Leduc G, de Guise É, de Guinea AO, Sénécal S, Léger PM. Using a cognitive analysis grid to inform information systems design. In: Davis FD, Riedl R, vom Brocke J, Léger PM, Randolph AB, editors. Information Systems and Neuroscience: Gmunden Retreat on NeuroIS 2015. Cham: Springer International Publishing; 2015:193-199.
13. Weintraub S, Dikmen SS, Heaton RK, Tulsky DS, Zelazo PD, Bauer PJ, et al. Cognition assessment using the NIH Toolbox. Neurology 2013 Mar 12;80(11 Suppl 3):S54-S64 [FREE Full text] [doi: 10.1212/WNL.0b013e3182872ded] [Medline: 23479546]

14. Allen B. Cognitive abilities and information system usability. Inf Process Manage 1994;30(2):177-191. [doi: 10.1016/0306-4573(94)90063-9]
15. Wothke W, Curran LT, Augustin JW, Guerrero C, Bock RD, Fairbank BA, et al. Factor Analytic Examination of the Armed Services Vocational Aptitude Battery (ASVAB) and the KIT of Factor-Referenced Tests. San Antonio, TX: Operational Technologies Corp; 1991. URL: http://www.iapsych.com/wmfhcaarchive/LinkedDocuments/WOTH01.pdf [accessed 2024-04-26]
16. Oliveira J, Gamito P, Alghazzawi DM, Fardoun HM, Rosa PJ, Sousa T, et al. Performance on naturalistic virtual reality tasks depends on global cognitive functioning as assessed via traditional neurocognitive tests. Appl Neuropsychol Adult 2018;25(6):555-561. [doi: 10.1080/23279095.2017.1349661] [Medline: 28805447]
17. Czaja SJ, Charness N, Fisk AD, Hertzog C, Nair SN, Rogers WA, et al. Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). Psychol Aging 2006;21(2):333-352 [FREE Full text] [doi: 10.1037/0882-7974.21.2.333] [Medline: 16768579]
18. Khalili-Mahani N, Assadi A, Li K, Mirgholami M, Rivard ME, Benali H, et al. Reflective and reflexive stress responses of older adults to three gaming experiences in relation to their cognitive abilities: mixed methods crossover study. JMIR Ment Health 2020;7(3):e12388 [FREE Full text] [doi: 10.2196/12388] [Medline: 32213474]
19. Schatz R, Egger-Lampl S, Masuch K. The impact of test duration on user fatigue and reliability of subjective quality ratings. J Audio Eng Soc 2012;60(1):63-73.
20. Lallemand C, Gronier G. UX design methods: 30 fundamental methods to design for optimal experiences. Paris, France: Eyrolles; 2018.
21. Isaacs E, Oates J, ILSI Europe a.i.s.b.l. Nutrition and cognition: assessing cognitive abilities in children and young people. Eur J Nutr 2008;47(Suppl 3):4-24. [doi: 10.1007/s00394-008-3002-y] [Medline: 18683026]
22. Robbins TW, Sahakian BJ. Computer methods of assessment of cognitive function. In: Copeland JRM, Abou-Saleh MT, Blazer DG, editors. Principles and Practice of Geriatric Psychiatry. Chichester, West Sussex, England: Wiley; 2002:147-151.
23. Slegers K, Van Boxtel MPJ, Jolles J. The efficiency of using everyday technological devices by older adults: the role of cognitive functions. Ageing Soc 2009;29(2):309-325. [doi: 10.1017/s0144686x08007629]
24. Chan MY, Haber S, Drew LM, Park DC. Training older adults to use tablet computers: does it enhance cognitive function? Gerontologist 2016;56(3):475-484 [FREE Full text] [doi: 10.1093/geront/gnu057] [Medline: 24928557]
25. Kieras DE, Meyer DE. An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. Hum–Comput Int 1997;12(4):391-438. [doi: 10.1207/s15327051hci1204_4]
26. Czaja SJ, Sharit J, Nair S, Rubert M. Understanding sources of user variability in computer-based data entry performance. Behav Inf Technol 1998;17(5):282-293. [doi: 10.1080/014492998119355]
27. Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A, Wager TD. The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. Cogn Psychol 2000;41(1):49-100. [doi: 10.1006/cogp.1999.0734] [Medline: 10945922]
28. Kiesel A, Steinhauser M, Wendt M, Falkenstein M, Jost K, Philipp AM, et al. Control and interference in task switching—a review. Psychol Bull 2010;136(5):849-874. [doi: 10.1037/a0019842] [Medline: 20804238]
29. Gade M, Schuch S, Druey M, Koch I. Inhibitory control in task switching. In: Grange J, Houghton G, editors. Task Switching and Cognitive Control. New York, NY: Oxford University Press; 2014:137-159.
30. US Army. Army individual test battery. In: Manual of Directions and Scoring. Washington, DC: War Department, Adjunct General's Office; 1944.
31. Lezak MD. Executive functions and motor performance. In: Neuropsychological Assessment. New York: Oxford University Press; 1995:650-685.
32. Mitrushina M, Boone KB, Razani J, D'Elia L. Handbook of Normative Data for Neuropsychological Assessment, Second Edition. New York, NY: Oxford University Press; 2005.
33. Reitan RM. Validity of the trail making test as an indicator of organic brain damage. Percept Mot Skills 2016;8(3):271-276. [doi: 10.2466/pms.1958.8.3.271]
34. Reitan RM, Wolfson D. Conventional intelligence measurements and neuropsychological concepts of adaptive abilities. J Clin Psychol 1992;48(4):521-529. [doi: 10.1002/1097-4679(199207)48:4<521::aid-jclp2270480414>3.0.co;2-c] [Medline: 1517447]
35. Strauss E, Sherman EMS, Spreen O. A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary. Oxford: Oxford University Press; 2006.
36. Sánchez-Cubillo I, Periáñez JA, Adrover-Roig D, Rodríguez-Sánchez JM, Ríos-Lago M, Tirapu J, et al. Construct validity of the trail making test: role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. J Int Neuropsychol Soc 2009;15(3):438-450. [doi: 10.1017/S1355617709090626] [Medline: 19402930]
37. Groff MG, Hubble LM. A factor analytic investigation of the Trail Making Test. Clin Neuropsychol 1981;3(4):11-13. [doi: 10.1007/978-0-387-79948-3_217]
38. Schear JM, Sato SD. Effects of visual acuity and visual motor speed and dexterity on cognitive test performance. Arch Clin Neuropsychol 1989;4(1):25-32. [doi: 10.1016/0887-6177(89)90004-8]

39. Gaudino EA, Geisler MW, Squires NK. Construct validity in the trail making test: what makes Part B harder? J Clin Exp Neuropsychol 1995;17(4):529-535. [doi: 10.1080/01688639508405143] [Medline: 7593473]
40. Crowe SF. The differential contribution of mental tracking, cognitive flexibility, visual search, and motor speed to performance on parts A and B of the trail making test. J Clin Psychol 1998;54(5):585-591. [doi: 10.1002/(sici)1097-4679(199808)54:5<585::aid-jclp4>3.0.co;2-k]
41. Meegan DV, Tipper SP. Visual search and target-directed action. J Exp Psychol Hum Percept Perform 1999;25(5):1347-1362. [doi: 10.1037//0096-1523.25.5.1347]
42. Treisman AM, Gelade G. A feature-integration theory of attention. Cogn Psychol 1980;12(1):97-136. [doi: 10.1016/0010-0285(80)90005-5] [Medline: 7351125]
43. Pietrzak RH, Maruff P, Mayes LC, Roman SA, Sosa JA, Snyder PJ. An examination of the construct validity and factor structure of the Groton Maze Learning Test, a new measure of spatial working memory, learning efficiency, and error monitoring. Arch Clin Neuropsychol 2008;23(4):433-445 [FREE Full text] [doi: 10.1016/j.acn.2008.03.002] [Medline: 18448309]
44. Gossweiler R, Kamvar M, Baluja S. What's up CAPTCHA? A CAPTCHA based on image orientation. New York, NY: Association for Computing Machinery; 2009 Presented at: Proceedings of the 18th international conference on World wide web; April 20-24, 2009; Madrid, Spain p. 841-850. [doi: 10.1145/1526709.1526822]
45. Kunda M, McGreggor K, Goel AK. A computational model for solving problems from the Raven's progressive matrices intelligence test using iconic visual representations. Cogn Syst Res 2013;22-23:47-66. [doi: 10.1016/j.cogsys.2012.08.001]
46. Aladwani AM, Palvia PC. Developing and validating an instrument for measuring user-perceived web quality. Inf Manag 2002;39(6):467-476. [doi: 10.1016/s0378-7206(01)00113-6]
47. Vignerons d'Exception. URL: https://www.vigneronsdexception.com/en/ [accessed 2024-06-03]
48. Renaud-Bray. URL: https://www.renaud-bray.com/accueil.aspx [accessed 2024-06-03]
49. LesPAC. URL: https://www.lespac.com/en/ [accessed 2024-06-05]
50. Lufa Farms. URL: https://montreal.lufa.com/en/#/ [accessed 2024-06-03]
51. Craigslist. URL: https://montreal.craigslist.org/ [accessed 2024-06-03]
52. Bowie CR, Harvey PD. Administration and interpretation of the Trail Making Test. Nat Protoc 2006;1(5):2277-2281. [doi: 10.1038/nprot.2006.390] [Medline: 17406468]
53. Thomas E, Snyder PJ, Pietrzak RH, Maruff P. Behavior at the choice point: decision making in hidden pathway maze learning. Neuropsychol Rev 2014;24(4):514-536. [doi: 10.1007/s11065-014-9272-7] [Medline: 25348261]
54. Gould MC, Perrin FAC. A comparison of the factors involved in the maze learning of human adults and children. J Exp Psychol 1916;1(2):122-154. [doi: 10.1037/h0072916]
55. Cuvillier M, Léger P, Sénécal S. Studies on the reliability, validity and sensitivity of single-item measurement scales in user experience. HEC Montréal 2021:57-93. [doi: 10.2307/j.ctv10qqx59.12]
56. Passalacqua M, Léger PM, Nacke LE, Fredette M, Labonté-Lemoyne E, Lin X, et al. Playing in the backstore: interface gamification increases warehousing workforce engagement. Ind Manag Data Syst 2020;120(7):1309-1330. [doi: 10.1108/imds-08-2019-0458]
57. Bracken MR, Mazur-Mosiewicz A, Glazek K. Trail making test: comparison of paper-and-pencil and electronic versions. Appl Neuropsychol Adult 2019;26(6):522-532. [doi: 10.1080/23279095.2018.1460371] [Medline: 30265569]
58. Tombaugh TN. Trail making test a and B: normative data stratified by age and education. Arch Clin Neuropsychol 2004;19(2):203-214 [FREE Full text] [doi: 10.1016/S0887-6177(03)00039-8] [Medline: 15010086]
59. Giovagnoli AR, Del Pesce M, Mascheroni S, Simoncelli M, Laiacona M, Capitani E. Trail making test: normative values from 287 normal adult controls. Ital J Neurol Sci 1996;17(4):305-309. [doi: 10.1007/BF01997792] [Medline: 8915764]
60. Arango-Lasprilla JC, Rivera D, Aguayo A, Rodríguez W, Garza MT, Saracho CP, et al. Trail making test: normative data for the Latin American Spanish speaking adult population. NeuroRehabilitation 2015;37(4):639-661 [FREE Full text] [doi: 10.3233/NRE-151284] [Medline: 26639932]
61. Shiffrin RM, Schneider W. Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. Psychol Rev 1977;84(2):127-190. [doi: 10.1037//0033-295x.84.2.127]
62. Karran AJ, Demazure T, Leger P, Labonte-LeMoyne E, Senecal S, Fredette M, et al. Toward a hybrid passive BCI for the modulation of sustained attention using EEG and fNIRS. Front Hum Neurosci 2019;13:393 [FREE Full text] [doi: 10.3389/fnhum.2019.00393] [Medline: 31780914]
63. Guertin-Lahoud S, Coursaris CK, Sénécal S, Léger PM. User experience evaluation in shared interactive virtual reality. Cyberpsychol Behav Soc Netw 2023;26(4):263-272 [FREE Full text] [doi: 10.1089/cyber.2022.0261] [Medline: 37071640]
64. Courtemanche F, Labonté-LeMoyne E, Léger PM, Fredette M, Senecal S, Cameron A, et al. Texting while walking: an expensive switch cost. Accid Anal Prev 2019;127:1-8. [doi: 10.1016/j.aap.2019.02.022] [Medline: 30826692]
65. Giroux F, Couture L, Lasbareille C, Boasen J, Stagg CJ, Fleming MK, et al. Usability evaluation of assistive technology for ICT accessibility: lessons learned with stroke patients and able-bodied participants experiencing a motor dysfunction simulation. In: Davis FD, Riedl R, vom Brocke J, Léger PM, Randolph AB, Müller-Putz GR, editors. Information Systems and Neuroscience: NeuroIS Retreat 2022. Cham: Springer International Publishing; 2022:349-359.

66. Zolyomi A, Snyder J. Social-emotional-sensory design map for affective computing informed by neurodivergent experiences. Proc ACM Hum-Comput Interact 2021;5(CSCW1):1-37 [FREE Full text] [doi: 10.1145/3449151]
67. Demmans Epp C, McEwen R, Campigotto R, Moffatt K. Information practices and user interfaces: student use of an iOS application in special education. Educ Inf Technol 2015;21(5):1433-1456. [doi: 10.1007/s10639-015-9392-6]
68. Vasseur A, Léger PM, Courtemanche F, Labonte-Lemoyne E, Georges V, Valiquette A, et al. Distributed remote psychophysiological data collection for UX evaluation: a pilot project. In: Human-Computer Interaction. Theory, Methods and Tools: Thematic Area, HCI 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings, Part I. Cham, Switzerland: Springer Nature; 2021.
69. Amershi S, Weld D, Vorvoreanu M, Fourney A, Nushi B, Collisson P, et al. Guidelines for human-AI interaction. New York, NY, United States: Association for Computing Machinery; 2019 Presented at: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland, UK p. 1-13. [doi: 10.1145/3290605.3300233]
70. Hussain J, Ul Hassan A, Muhammad Bilal HS, Ali R, Afzal M, Hussain S, et al. Model-based adaptive user interface based on context and user experience evaluation. J Multimodal User Interfaces 2018;12(1):1-16. [doi: 10.1007/s12193-018-0258-2]

Abbreviations

CAPTCHA: Completely Automated Public Turing tests to tell Computers and Humans Apart
MMSE: Mini-Mental State Examination
MoCA: Montreal Cognitive Assessment
NumRec: numerical recognition
PicRec: pictogram recognition
RM MANCOVA: repeated-measures multivariate analysis of covariance
RPM: Raven’s Progressive Matrices
TMT: Trail Making Test
UX: user experience

Edited by A Kushniruk; submitted 22.06.23; peer-reviewed by L Loeb, C Baxter; comments to author 20.01.24; revised version received 13.02.24; accepted 20.02.24; published 14.06.24.

Please cite as:
Depauw T, Boasen J, Léger PM, Sénécal S
Assessing the Relationship Between Digital Trail Making Test Performance and IT Task Performance: Empirical Study
JMIR Hum Factors 2024;11:e49992
URL: https://humanfactors.jmir.org/2024/1/e49992
doi:10.2196/49992
PMID:38875007

Original Paper

Capturing Usability Problems for People Living With Dementia by Applying the DEMIGNED Principles in Usability Evaluation Methods: Mixed Methods Study

Thomas Engelsma[1,2,3], MSc; Simone Heijmink[1], BSc; Heleen M A Hendriksen[4,5], MSc; Leonie N C Visser[4,5,6,7,8], PhD; Afina W Lemstra[4,5], PhD; Monique W M Jaspers[1,2], PhD; Linda W P Peute[1,2], PhD

[1]eHealth Living & Learning Lab Amsterdam, Department of Medical Informatics, Amsterdam UMC, location University of Amsterdam, Amsterdam, Netherlands

[2]Amsterdam Public Health, Digital Health, Amsterdam, Netherlands

[3]Amsterdam Public Health, Aging & Later Life, Amsterdam, Netherlands

[4]Alzheimer Center Amsterdam, Neurology, Vrije Universiteit Amsterdam, Amsterdam UMC, Location VUmc, Amsterdam, Netherlands

[5]Amsterdam Neuroscience, Neurodegeneration, Amsterdam, Netherlands

[6]Amsterdam Public Health, Quality of Care, Amsterdam, Netherlands

[7]Division of Clinical Geriatrics, Center for Alzheimer Research, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden

[8]Department of Medical Psychology, Amsterdam UMC location University of Amsterdam, Amsterdam, Netherlands

Corresponding Author:
Thomas Engelsma, MSc
eHealth Living & Learning Lab Amsterdam
Department of Medical Informatics
Amsterdam UMC, location University of Amsterdam
Meibergdreef 9
Amsterdam, 1105AZ
Netherlands
Phone: 31 657570009
Email: t.engelsma@amsterdamumc.nl

Abstract

Background: Dementia-related impairments can cause complex barriers to access, use, and adopt digital health technologies (DHTs). These barriers can contribute to digital health inequities. Therefore, literature-based design principles called DEMIGNED have been developed to support the design and evaluation of DHTs for this rapidly increasing population.

Objective: This study aims to apply the DEMIGNED principles in usability evaluation methods to (1) capture usability problems on a mobile website providing information resources for people visiting a memory clinic, including those living with subjective cognitive decline (SCD), mild cognitive impairment (MCI), or dementia, and (2) investigate the realness of usability problems captured by the DEMIGNED principles in expert testing, specifically for mobile websites that act as a means of providing DHTs.

Methods: First, a heuristic evaluation was conducted, with the DEMIGNED principles serving as domain-specific guidelines, with 3 double experts (experienced in both usability and dementia) and 2 usability engineering experts. Second, think-aloud sessions were conducted with patients visiting a memory clinic who were living with SCD, MCI, or dementia.

Results: The heuristic evaluation resulted in 36 unique usability problems. A representative sample of 7 people visiting a memory clinic participated in a think-aloud session, including 4 (57%) with SCD, 1 (14%) with MCI, and 2 (29%) with dementia. The analysis of the think-aloud sessions revealed 181 encounters with usability problems. Of these encounters, 144 (79.6%) could be mapped to 18 usability problems identified in the heuristic evaluation. The remaining 37 (20.4%) encounters from the user testing revealed another 10 unique usability problems. Usability problems frequently described in the think-aloud sessions encompassed difficulties with using the search function, discrepancies between the user's expectations and the content organization, the need for scrolling, information overload, and unclear system feedback.

Conclusions: By applying the DEMIGNED principles in expert testing, evaluators were able to capture 79.6% (144/181) of all usability problem encounters in the user testing of a mobile website for people visiting a memory clinic, including people living with dementia. Regarding unique usability problems, 50% (18/36) of the unique usability problems identified during the heuristic evaluation were captured by the user-testing sessions. Future research should look into the applicability of the DEMIGNED principles to other digital health functionalities to increase the accessibility of digital health and decrease digital health inequity for this complex and rapidly increasing population.

(JMIR Hum Factors 2024;11:e54032) doi:10.2196/54032

KEYWORDS

dementia; design principles; digital health; memory clinic; usability evaluation; mobile phone

Introduction

Background

Digital health technologies (DHTs) have the potential to improve health outcomes, information access, patient monitoring and self-management, treatment adherence, and disease diagnostics and prevention [1-4]. However, susceptible patient groups, such as older adults or people with disabilities, can experience challenges with accessing DHTs, creating digital health inequities [5,6]. Besides well-known accessibility issues such as financial or geographical barriers, human factors such as insufficient digital skills, poor motivation, low health literacy, poor health conditions, low socioeconomic status, and declining cognitive and physical abilities can also decrease the accessibility of DHTs [5,7-9]. These factors may also result in specific challenges to using DHTs that should be accounted for during the development of DHTs to increase their usability. Usability is defined as the extent to which DHTs can be used to achieve specified goals with effectiveness, efficiency, and satisfaction [10]. Problems related to the usability of DHTs arise when their design is not tailored to the needs of specific end users, thereby hampering their acceptability, user engagement, adoption, and successful use [2,11-13].

A vulnerable patient group that can experience challenges with accessing and using DHTs is people living with dementia. The prevalence of dementia is increasing globally as the population ages, making it a significant public health concern. It is expected that >150 million people in 2050 will be living with dementia, with 10 million new cases each year [14]. There are many underlying causes of dementia. Therefore, different types of dementia can be distinguished, such as dementia due to Alzheimer disease, vascular dementia, Lewy body dementia, and frontotemporal dementia [15]. People living with dementia face specific barriers when trying to use DHTs, which can worsen disparities in digital health care access [9,16]. These barriers relate to their cognitive, perceptive, and physical decline; changing frame of mind; and decreasing speech and language skills [9]. Even though these barriers to using DHTs for people living with dementia exist, the availability of DHTs increases, as they are suggested to improve the quality of life for this population by providing, among others, assistance with activities of daily living; improve their social engagement; and monitor their cognitive functions [17-19].

Objectives

To improve the usability and accessibility of DHTs for people living with dementia, their design can benefit from context-specific design guidance. While accessibility standards presented in, for example, International Organization for Standardization Technical Specification 82304-2 [20], the Web Content Accessibility Guidelines [21], and xCertia Guidelines [22] can serve as valuable foundations for inclusive design, "understanding user diversity and applying this in the development process" [23], their use during the development of DHTs and evaluation is currently limited [24]. In addition, these standards and guidelines may not comprehensively address the complex, unique, and multifaceted needs of vulnerable populations, such as people living with dementia or other cognitive and physical impairments [24-26]. Ensuring the accessibility and usability of DHTs for such vulnerable user groups necessitates the implementation of additional context-related and detailed design specifications, something often missing in traditional guidelines and principles [24]. Therefore, in previous research, we developed literature-based design principles, or DEMIGNED principles, to be considered when developing DHTs for this population [27,28]. However, empirical evidence needs to be collected to further investigate the applicability of the DEMIGNED principles and their potential refinements. To collect this empirical evidence, this study aimed to apply the DEMIGNED principles in usability evaluation methods to (1) capture usability problems on a mobile website providing information resources for people visiting a memory clinic, including those living with subjective cognitive complaints, mild cognitive impairment (MCI), or dementia, and (2) investigate the realness of usability problems captured by the DEMIGNED principles in expert testing, specifically for mobile websites that act as a means of providing DHTs.

Methods

Study Design

Investigating the realness of usability problems captured by the DEMIGNED principles can encompass usability evaluation research [29]. According to Hartson et al [30], a usability problem is real if it is "a predictor of a problem that users will encounter in real work-context usage and that will have an impact on usability (user performance, productivity, and/or satisfaction)" [30]. Therefore, in the first part of this study, usability issues on a mobile website for patients visiting a memory clinic, including people with dementia, were captured by applying the DEMIGNED principles. This was conducted through a heuristic evaluation approach. A heuristic evaluation is a low-cost method where usability or domain experts evaluate a system's navigation structure, screen layout, and interaction structure, typically on a set of predefined generic heuristics or guidelines [31]. This resulted in a list of unique usability problems that violate a specific heuristic or guideline. To investigate the realness of these usability problems, these findings were used to map the results from inclusive user-based testing with patients visiting a memory clinic with subjective cognitive decline (SCD), MCI, or dementia. As there are challenges to inclusive-based user testing with people living with dementia, we applied considerations proposed by human factors experts to build trust and decrease potential stress [32]. To conduct the 2 usability evaluations, the mobile website of Alzheimer Center Amsterdam, a diagnostic and treatment center for (early-onset) dementia, was examined [33]. Their website contains information and other resources for people who are diagnosed with (or concerned about having) dementia and their relatives, researchers, funders, and others interested in learning about dementia (research). On the basis of the findings, recommendations for redesign are presented.

Heuristic Evaluation Approach

By applying the heuristic evaluation method, the overall user interface and structure of the mobile website of Alzheimer Center Amsterdam were assessed. In this study, potential usability problems were captured by assessing violations of the DEMIGNED principles (Table 1). These principles were developed by TE, LWPP, and MWMJ [27,28]. The design principles have been mapped to the categories of the mobile health for older adults with dementia–usability framework (MOLDEM-US), which captures barriers to using DHTs, specifically mobile health technologies, for people living with dementia [9]. These categories of barriers relate to cognition, perception, frame of mind, and speech and language.

Table 1. Overview of the DEMIGNED design principles for digital health technologies for people living with dementia.

Category and design principle	Abbreviation[a]
Cognition	
Support the monitoring of action progress	C-monitoring
Provide tutorials with short instructions to guide the digital tool	C-tutorials
Provide functionalities and actions adjustable to the user’s cognitive abilities	C-abilities
Allow easy navigation to functions and content in a digital tool	C-navigation
Implement representative and understandable icons	C-icon use
Perception	
Provide visually compartmentalized user interfaces	P-compartmentalize
Provide appropriate system feedback	P-system feedback
Implement distinguishable colors	P-color use
Allow distinguishable clickable and nonclickable areas	P-click ability
Allow easily processable elements	P-elements
Frame of mind	
Provide continuous support	F-support
Ensure no time pressure	F-time
Provide positive feedback for correct action completion	F-positive feedback
Implement app settings adjustable to personal preferences	F-preferences
Provide attractive and respectful content	F-content
Speech and language	
Ensure the use of understandable words and sentences that feel comfortable	S-understandability
Allow user input through both speech and text	S-user input

[a]Each principle has an abbreviation presented with the first letter of the category, followed by a keyword describing the principle.

Participants and Procedures

Previous heuristic evaluations have shown that 3 to 5 experts can identify 74% to 87% of usability problems [34]. This evaluation was performed by 5 evaluators: 3 double experts, TE, SH, and LWPP (knowledgeable in both usability engineering and dementia), and 2 usability engineering experts (Table 2). After confirming participation, TE familiarized the evaluators with the DEMIGNED principles in an introductory session where the evaluators were able to discuss and ask for clarifications on the principles.

A total of 3 use cases were determined based on the information presented on the website: patient care, scientific research, and information about dementia. The evaluators assessed these use cases in an overall manner to gain a sense of the website’s navigation and structure before a more thorough assessment of the user interface components. Because of the high number of pages on the website and the uncertainty of which pages participants will encounter in the think-aloud session, example tasks were composed for each use case to guide the evaluation (Table 3).

The DEMIGNED principles were used as a set of heuristics by the evaluators and applied to check the design of the website's user interface and structure. If a principle was violated, it was given a severity rating based on the following scales, defined by Nielsen [35]: "(0) I do not agree that this is a usability problem at all (1) Cosmetic problem only: need not to be fixed unless extra time is available on project, (2) Minor usability problem: fixing this should be given low priority, (3) Major usability problem: important to fix, so should be given high priority or (4) Usability catastrophe: imperative to fix this before the product can be released." Moreover, the evaluators reported for each identified usability issue the location on the website and the violated DEMIGNED principle. Usability issues can be related to multiple DEMIGNED principles, allowing the experts to report >1 principle for a unique usability issue.

Table 2. Overview of experts participating in the heuristic evaluation.

Degree	Sex	Research expertise	Experience (years), n	Occupation
PhD	Female	Human factors engineering in health care	20	Senior researcher
MSc	Female	User experience	8	UX[a] designer and university lecturer
MSc	Male	Design for people living with dementia	6	Assistant professor
MSc	Female	Design for vulnerable populations	1	PhD student
BSc	Female	Working with people living with dementia and experience with HE[b]	3	Medical informatics student

[a]UX: user experience.

[b]HE: heuristic evaluation.

Table 3. Tasks conducted in the heuristic evaluation.

Use case	Task
Patient care	1. Find information about the *Screeningsdag*
	2. Find the information video on the lumbar puncture
	3. Go to the log-in page of *MijnDossier*
	4. Find information on how to participate in scientific research
	5. Find the clinician who will see you during the *Screeningsdag*
	6. Find the physical address of the center
	7. Find the phone number of the center
	8. Find information about getting a second opinion
Scientific research	9. Find information about ongoing research projects from the center
	10. Find information about completed research projects from the center
	11. Find information about hersenonderzoek.nl
Information about dementia	12. Find information about whether you have dementia or not
	13. Find information about the treatment of dementia
	14. Find tips about how to live with dementia
	15. Find personal stories from patients
	16. Find the frequent asked questions
	17. Find information about getting dementia at a young age
	18. Find information about "corticobasal degeneration"

Data Analysis

The identified usability problems were coded by performing a deductive thematic analysis. The DEMIGNED principles were used to predefine usability themes in which issues could arise. All usability issue encounters were combined in 1 master list by SH, after which duplicates were summed independently by TE and SH. The issues were summed overall rather than per task, as the aim was to obtain an overall report of the usability issues the DEMIGNED principles can capture on a mobile website. This led to a final set of unique usability issues on which consensus was reached with LWPP. The severity ranking for each unique issue was calculated by taking the average severity given by the evaluators who identified the usability issue. It was decided to report an average severity score rather than a consensus score to moderate potential extreme views in assessing usability problems, given the novelty of the DEMIGNED principles. This offers a balanced representation of the severity score, eliminating the need for evaluators to reach a consensus.

Think-Aloud Method

Participants

For the think-aloud sessions, participants were recruited at Alzheimer Center Amsterdam. The aim of the recruitment was to include a representative sample of the (heterogenous) population that is most likely to use the website of Alzheimer Center Amsterdam: those who visit the memory clinic to seek support for their cognitive complaints. Alzheimer Center Amsterdam has a focus on patients living with dementia at a younger age (<65 years). Within this group, the most frequent diagnosis is SCD, followed by dementia and MCI [36]. Therefore, for the sample to be representative, more people with SCD were recruited, followed by people with dementia and MCI. All participants were presented as patients at the memory clinic of Alzheimer Center Amsterdam, where they received a standardized dementia diagnostic workup. Clinical diagnosis was made in a multidisciplinary meeting and discussed with patients during a second appointment. Subsequently, patients were offered annual or biannual follow-up. People scheduled for such a follow-up appointment were called to participate when they had previously given permission to be approached for research; had an appointment on June 5, June 8, or July 6, 2023; were not already participating in other research (that day); and were Dutch speaking. Patients who were unable to give informed consent were excluded from participation. Those who showed interest in participating received an information letter. In addition, a follow-up telephone call was scheduled to answer any questions and confirm participation after participants had the opportunity to read the information letter in depth.

Study Procedures

After confirming participation, a think-aloud session was scheduled either before or after the participant's appointment at Alzheimer Center Amsterdam. We aimed to adopt an empathetic approach to ensure there was trust between the participant and the evaluators (LWPP, SH, and TE) and to make the participant feel comfortable [32]. After a short informal chat, participants provided informed consent and completed a paper-based questionnaire about their age, sex, and technology use. Thereafter, the think-aloud session started. The participant was then asked to sit behind the smartphone. The researcher sitting next to the participant first explained the goal of the session and showed how to verbalize thoughts through an example task on the smartphone provided by the researcher (eg, "Find what the weather will be this Saturday."). Participants were then asked to complete 7 tasks on the mobile website while verbalizing their thoughts (Textbox 1). For each use case, these tasks were derived from the example tasks for the heuristic evaluation but were specified to specific end points deemed relevant to the end users, such as information resources. The completion of the tasks was monitored by a second researcher, sitting behind the computer that recorded the session and across from the participant. If a partner or relative was present, they were instructed to only provide motivational reactions if the participant became silent. Therefore, only usability problems detected by the participants were gathered and analyzed. Even though this may introduce bias, this decision was made to make the participant feel more comfortable and simulate a more real-life setting for the participant [32].

The facilities at eHealth Living & Learning Lab Amsterdam were used to both audio record and video record the think-aloud session. This allowed researchers to capture rich data from both the user's interactions with the mobile device and the verbalizations from the think-aloud session. The facilities used to capture these data include (1) a video camera to record the participant's hand interaction with the smartphone, (2) a voice recorder to record the verbalized thoughts of the participant, (3) a smartphone device for the participant to access the mobile website, and (4) two laptops. One laptop was used to capture the screen from the smartphone through screen casting. The other laptop was used for Viso software (Noldus) to bring together the video recording, the screen capture recording, and the audio recording [37].

Textbox 1. The 7 tasks conducted in the think-aloud sessions.

Tasks

1. Find the date of the next event of Alzheimer Center Amsterdam
2. Find the 4 health care professionals scheduled for the screening day
3. Find information about the nurse consultants
4. Find the video about lumbar puncture
5. Find the conversation guide for the screening day
6. Find information about hersenonderzoek.nl
7. Find details about a fitness to drive statement

Data Analysis

First, all audio recordings were transcribed. Segments that did not convey verbal thoughts, such as instances when the participants read the website, were excluded. Second, the video recordings were analyzed and used to amplify the transcribed audio recordings. Third, the resulting transcripts were openly coded first and axially encoded afterward. Finally, the frequency of problems per category was registered. The resulting usability problems were mapped onto the findings from the heuristic evaluation by TE, SH, and LWPP. Each usability problem encounter was mapped onto a theme from the DEMIGNED principles (cognition, perception, frame of mind, or speech and language). Afterward, if applicable, a subtheme from the heuristic evaluation findings was linked to the usability problem encounter, followed by a specific issue. Findings from the think-aloud sessions that could not be verified with the findings from the heuristic evaluation will be presented separately.

Ethical Considerations

The study was approved by the Amsterdam University Medical Center medical ethical review committee with the number 2023.0240. All participants received detailed information about the study and provided written informed consent before participation in the think-aloud study. All data used in analysis have been anonymized. Each participant in the think aloud sessions received a €10 (US $10.90) gift card.

Results

Heuristic Evaluation Approach

A total of 4 evaluators completed the heuristic evaluation. One evaluator conducted only the example tasks related to patient care and was further involved in the development of the master list. The heuristic evaluation resulted in a final set of 36 unique usability issues, as shown in Multimedia Appendix 1. The usability issues identified in the cognition theme can impact people living with dementia when interacting with the website. The malfunctioning search function, nonintuitive navigation processes, misaligned information headings, inconsistent structures, nonlinear pathways, duplicated content, and scrolling difficulties can lead to confusion and fatigue. In addition, the lack of logical menu structures, faulty filter functions, and challenges in finding essential features such as the information videos and the patient portal can exacerbate cognitive impairments associated with dementia, hindering effective navigation and information retrieval. Furthermore, external navigation links may disrupt the user's cognitive flow and understanding.

Think-Aloud Method

A total of 7 participants were included in the think-aloud sessions, composing a representative sample of people visiting the memory clinic at Alzheimer Center Amsterdam (Table 4). Each session lasted approximately 30 to 50 minutes. A total of 5 audio recordings and 7 video recordings were used in the analysis.

The think-aloud sessions revealed 181 usability problem encounters, of which 144 (79.6%) were mapped to 18 usability problems identified during expert testing (Multimedia Appendix 1). Most frequent usability problem encounters that verified the findings from the heuristic evaluation relate to the user expectations (48/181, 26.5%), information overload (20/181, 11%), system feedback (19/181, 10.5%), search function results (12/181, 6.6%), and scrolling (9/181, 5%). Examples of these issues are provided by means of quotes in Table 5.

Table 4. Characteristics of patients visiting a memory clinic who participated in the think-aloud sessions (n=7).

Characteristics	Values, n (%)
Diagnosis	
Subjective cognitive decline	4 (57)
Mild cognitive impairment	1 (14)
Dementia	2 (29)
Sex	
Male	3 (43)
Female	4 (57)
Age (years)	
50-59	2 (29)
60-69	2 (29)
70-79	2 (29)
80-89	1 (14)
Smartphone use	
Yes	6 (86)
No	1 (14)
Smartphone use (min/day)	
0-30	2 (29)
30-60	2 (29)
60-120	1 (14)
>120	1 (14)
Tablet use	
Yes	5 (71)
No	2 (29)
Tablet use (min/day)	
0-30	3 (43)
30-60	2 (29)
First time viewing the website	
Yes	3 (43)
No	4 (57)

Table 5. Example quotes from transcript analysis for the most frequently encountered usability problems during the think-aloud sessions.

Usability problem	Example quote from transcript analysis
User expectations	• "I wouldn't know how to search for this and under which heading it falls. [Read aloud: about dementia, patient care, professionals, scientific research]. I don't see a section called 'what if you've received a diagnosis.' I can't find it; I wouldn't know" (Participant 3). • "My Record [Dutch: 'MijnDossier,' a patient portal]. When I click on that, there's nothing about me personally. My record means my record, but there are general things there. And I would like to access my record, and it exists because it's listed here below, if I see it correctly" (Participant 1).
Information overload	• "Well, I just saw it, and I'm just trying to remember where. They're, of course, trying to provide a lot of information" (Participant 2). • "Forms of dementia. [Looks at the menu screen for a while]. Yes, there's so much on it; it's making me a bit fidgety, and then I think, what was it I wanted to look up again? I'm getting completely distracted by all those things" (Participant 5).
System feedback	• "What I notice here is that there are long lists, and you're working with people who have dementia or the caregivers, of course. With those long lists, you touch them, and nothing happens, but then the information is listed below" (Participant 4). • "Yes, I was just there earlier. Then we go back to the preparation. [Clicks three times on 'preparation' in the menu but doesn't see anything change]" (Participant 2).
Search function	• [Searches for "lumbar puncture" in the search bar; clicks on a result.] "I have the protein research here, so that's a lumbar puncture. That's what you're looking for, right? Or are you looking for what it looks like?" (Participant 3) • "Let me see. I've entered your question, and I actually expect an answer, but it's not doing that. Yes, because when I do this on Google, I get 20 answers, and then I can choose. But here, I don't get there" (Participant 1).
Scrolling	• "They naturally have a lot of information, and then you click on this, and it's listed below [under the menu], and personally, you know, I just want to see it at the top because, yeah. I've had times when I thought 'I'm not there or something,' but it's actually there." (Participant 2). • "Yes, you see, I have to go down (scroll) every time. Lumbar puncture, well, look" (Participant 1).

Additional Usability Issues Identified Through the Think-Aloud Analysis

A total of 37 usability problem encounters, solely identified in the think-aloud analysis, were thematically categorized into 10 unique usability problems and further classified into 7 themes (Table 6).

The findings revealed additional issues with identifying and using the search function, such as (recovering from) typing errors; the use of too many search terms leading to an overload of search results; and confusion caused by the visibility of the search history from the mobile device, rather than solely the queries on the website's search engine. Some participants limited their mental model for information finding solely to the search engine, leading to insecurities when asked to apply other navigation structures, such as clicking through the menu structure. Moreover, irrelevant search results or typing errors caused difficulties with remembering the task they were asked to perform. Finally, 3 additional usability issues related to the participants' frame of mind while interacting with the website were identified: insecurity, stress, and habit. Participants verbally conveyed questions about their own ability to interact with the website and stressful responses while conducting tasks, and these were coded as shown in Table 6.

Table 6. Usability issues identified only through the think-aloud analysis.

Theme and usability problem	Impact	Severity score[a], average	Frequency of the problem, n
Touch screen sensitivity			
The touch screen of the smartphone is too sensitive for participants who experience slower performance speed.	It causes unwanted user actions, such as accidentally pressing external links or pressing too hard instead of swiping or scrolling.	2	2
Search function issues			
Typing errors in a search query when using the search function leads to no or unwanted search results.	When typing error goes unnoticed, the user does not understand why information cannot be found, leading to frustration and inability to complete tasks.	4	6
The search function cannot be accessed via the home page and can be accessed only after clicking on the menu icon.	If a user limits their mental model for information finding solely to the search function and this cannot be found, it prevents them from completing tasks.	4	3
Too many search terms in a search query lead to an overload of results.	When the user types "driving and dementia," many results are shown, including all pages that include "and," leading to frustration and inability to complete tasks.	4	3
Search history is shown from when typing a search query.	When a user wants to type a search query in the search bar but the search history appears below the bar, this can be seen as a dropdown menu with an option to select from, causing confusion.	2	1
Return button			
The return button is placed under the menu structure but above the content of the page, making the user miss information.	When a user sees the return button under the menu structure, they expect there is no more content on the page when they do not scroll down. Therefore, they might miss valuable information, preventing them from completing an action.	3	3
Insecurity			
Users verbally question their own interactions with the website while performing tasks.	Insecurity can lead to decreased motivation and willingness to use the website.	2	5
Stress			
Users verbally convey stress while interacting with the website.	Stress can lead to decreased motivation and willingness to use the website.	2	10
Habit			
Users verbalize having their own habits while using websites, which are not aligned with the website under evaluation, such as changing a search query, only using Google search and button recognition.	When a user is too stuck in such habits, it makes it difficult to adjust to the website's design, causing difficulties in completing a task.	2	3
Clicks			
The user needs too many clicks to find the desired information, while the user expects to find information faster.	The user cannot reach the page to complete a task.	3	1

[a]Average severity score based on Nielsen's severity ranking [35].

Discussion

This study aimed to apply the DEMIGNED principles in usability evaluation methods to (1) capture usability problems on a mobile website providing information resources for people visiting a memory clinic, including those living with subjective cognitive complaints, MCI, or dementia, and (2) investigate the realness of usability problems captured by the DEMIGNED principles in expert testing, specifically for mobile websites that act as a means of providing DHTs. In addition, this study provided insights for the future refinements of these principles.

Principal Findings

The mobile website of Alzheimer Center Amsterdam underwent both heuristic evaluation and evaluation through the think-aloud method to investigate the realness of usability problems detected by applying the DEMIGNED principles in expert testing. To this end, usability problems derived from user testing were mapped to findings from expert testing. The heuristic evaluation, covering most of the website's pages, revealed 36 distinct usability issues. Subsequently, during the think-aloud sessions with people living with SCD, MCI, or dementia, 18 unique usability problems were mapped to the findings from the heuristic evaluation. In addition, the think-aloud sessions revealed 10 new usability issues. Despite potential differences in the number of pages viewed between the heuristic evaluation and the think-aloud session, given the uncertainty surrounding the information-seeking strategies of participants in the think-aloud session, these results suggest a 50% (18/36) validity score for the DEMIGNED principles in an expert test based on the usability evaluation method validity score measurement [30]. However, this first investigation aimed to explore the realness of usability problems that can be captured by applying the novel DEMIGNED principles, gaining insights into the use of the DEMIGNED principles. Follow-up research should determine the usability evaluation method effectiveness of applying the DEMIGNED principles to detecting usability issues through heuristic testing. Nevertheless, a validity score of 50% (18/36) in this study suggests the potential of the DEMIGNED principles for future heuristic evaluation approaches. This contrasts with other research showing a validity score of 40% when using commonly used heuristics [38]. Moreover, looking at the number of times a usability problem was encountered, 79.6% (144/181) of the usability problem encounters were captured by experts using the DEMIGNED principles.

The findings indicate the potential for the DEMIGNED principles to be further used as a set of guidelines for a heuristic evaluation of DHTs for people living with SCD, MCI, or dementia. This can be valuable, as current sets of heuristics may not sufficiently capture usability issues experienced by people living with dementia. Over time, numerous sets of heuristics have been devised to assess the usability of mobile devices, focusing primarily on their physical attributes and phone interfaces [39,40]. Widely acknowledged heuristics of Nielsen [34] are valuable for evaluating general usability in website interfaces. Nevertheless, these guidelines do not fully encompass the unique challenges associated with designing DHTs for people living with dementia. Heuristics to capture usability problems in the design of DHTs related to cognitive decline, such as memory loss, reduced attention span, and decision-making difficulties, are lacking. Similarly, heuristics proposed by Neto and da Graça Pimentel [41], inspired by Nielsen [34], focus on assessing the usability of mobile apps but lack specific heuristics to capture usability problems experienced by people living with dementia. These heuristics are tailored for general mobile interface design and do not comprehensively address specific barriers to using DHTs for this population, such as attention deficits and challenges in learning and adapting to new interfaces. For medical devices, heuristics proposed by Zhang et al [42] are commonly used. However, these might not be entirely suitable for evaluating DHTs for people living with dementia. Heuristics proposed by Zhang et al [42] prioritize the detection of usability issues for medical devices, such as issues with accuracy, reliability, and safety. However, these are not attuned to capture usability problems caused by, for example, emotional fluctuations or cognitive decline.

Refinement of the DEMIGNED Principles

The think-aloud sessions revealed 10 usability problems that were not found through expert testing that may require further refinement of the DEMIGNED principles. First, 3 (30%) of these problems relate to the frame of mind of the user. It was observed that interacting with the website can cause verbal reactions of stress and insecurities. This was also the case when a participant had a certain habit of finding web-based information that was not in line with how the evaluated website can be searched. These barriers may lead to unsuccessful task completion and should, therefore, be accounted for and included in the DEMIGNED principles. However, the 3 usability issues related to the frame of mind seemed to be related to navigational structure, suggesting that a more linear navigational structure might make information easier to find and, therefore, reduce stress and insecurity. Second, the design of the search function was revealed to be a critical usability problem that prevented successful task completion. Even though 2 potential usability problems were identified by one of the evaluators during the heuristic evaluation, 3 more issues were revealed in the think-aloud analysis when people with SCD, MCI, or dementia used the engine. Using the search function revealed issues with typing errors, too many search terms leading to an overload of results, and a misunderstanding of the search history. Therefore, the DEMIGNED principle of navigation should provide design guidelines specific to search functions. Finally, 3 usability problems were encountered by a few participants. First, the amount of clicking was found to be too much to find the information they were looking for, which was also found in other studies where participants showed problems finding content because of the navigation structure [43,44]. Limiting the amount of clickable content on a single screen may support directing the users to relevant information [43]. Second, the sensitivity of the touch screen led to unwanted actions, which were caused by the device itself. Currently, the DEMIGNED principles only focus on user interface and structures rather than the hardware; therefore, this problem may not have been found by the evaluators in the heuristic evaluation. However, touch screen sensitivity could be tackled by providing visible or audible system feedback when a user clicks too fast or a pop-up to confirm an action [28]. Third, it was observed that the position of the "return" button above the content caused challenges, as one of the participants mistakenly perceived that no additional information was available below the button. This provides insights into further refining the DEMIGNED principles, such as icon use and system feedback, with guidelines regarding button placement as part of the content organization.

Recommendations for Redesign

Overview

The website of Alzheimer Center Amsterdam not only is informative for people living with SCD, MCI, or dementia and others having an appointment at the memory clinic but also contains information for researchers, funders, family members, etc. However, these recommendations aim to enhance the website's accessibility and user experience for people living with dementia. Moreover, they can also benefit those interested in redesigning for improved digital health accessibility in this context. Overall, it is suggested to redesign the website with general principles that can improve accessibility, such as by adding an option to read text out loud to the user, increasing the font size without decreasing readability, and magnifying options for images and other graphics. Furthermore, 5 specific areas of redesign were identified based on the findings from this study, namely navigation, information overload, scrolling, content and user experience, and system feedback.

Navigation

Navigation often requires sufficient decision-making skills, the awareness of location, and a sequence of clicks. These skills may decline due to memory impairment, decreased learnability, and difficulties with processing information. Therefore, it is recommended to implement linear navigation. This means the user can move only forward or backward through the website and its content. Starting with the home page, it is recommended to redesign this to full-screen navigation, showing clickable main options to navigate to “patient care,” “scientific research,” and “living with dementia.” After selecting one of these main options, a full-screen menu with suboptions to choose from should open. This may tackle the usability problem of users not noticing that the “main option” is also a page with unique information. Additional research prioritizing information needs and formulating clear suboptions may support this redesign, given the perceived discrepancies between the current information headings and accompanying content. In addition, this requires removing potentially irrelevant, outdated, or duplicate information and empty pages to ease the process of scanning and finding information.

Furthermore, participants experienced challenges with using the search function. In general, people often rely on searching as their main strategy [45]. In this study, those who relied solely on the search function experienced difficulties when navigating the website. Familiarity with search engines such as Google may make them more inclined to use the search feature as their default interaction method [43]. Overall, the availability of a function that eases information finding on a website is important for people living with dementia [43,44]. However, some participants experienced difficulties finding the search function, as it was accessible only after clicking on the menu icon, or using the function, as it did not work as expected. This led to confusion, frustration, poor user experience, and hindered task completion during think-aloud sessions. This reduced mental model of finding information may be caused by barriers related to the user’s frame of mind, such as a lack of trust in their abilities or perceived complexity [9]. In addition, a reason for using the search function as the default interaction method may be due to affected language and communication skills. Therefore, it is suggested to redesign the search function to prevent typing errors, too many and irrelevant results, and the presentation of results in an illogical order. Implementing a Google-like search function with positive system feedback can be considered, where suggestions for typing errors are made, an overview of the most relevant results are presented, and selecting for filters that are of relevance to the user is possible. Such a search function could also be relevant for the “team” page to ease the user’s process of finding information about their physician. Moreover, the search function can be placed on the home page to increase its accessibility.

Information Overload

An excessive amount of information on a single page and in the menu (eg, the number of subpages to choose from) can not only overwhelm users living with cognitive problems, making it challenging to process and remember the content, but also impair decision-making skills by hindering the ability to effectively filter out the relevant information. This creates a poor user experience and can cause frustration and disorientation. In addition, it can compromise the usefulness and accessibility of such web pages because users may struggle to find the specific information they need within a cluttered interface, hindering their ability to use the website effectively. To increase usefulness and decrease information overload for people living with dementia, it is suggested to provide content that is comprehensive, practical, and reassuring [43]. Moreover, categories should be less broad and listed alphabetically [44].

Scrolling

The need for scrolling caused participants to miss useful information and prevented successful task completion during the think-aloud sessions. A user with cognitive impairments can experience challenges in remembering the need to scroll every time a new web page is loaded. In addition, excessive scrolling can lead to increased cognitive load, as it requires processing information sequentially, potentially overwhelming the working memory. It has been suggested from observational data that reducing the cognitive load of a web page (eg, less information on a page) can contribute to tackling scrolling issues, as scrolling appears to have a high cognitive load [46]. Moreover, the physical need for scrolling may be challenging and cause fatigue [46].

Content and User Expectations

It was found that 48 usability problems related to user expectations were primarily caused by the discrepancy between user expectations and actual content organization. This led to frustration and confusion among participants, making it difficult to locate the information they sought. The findings emphasize the importance of accurately representing content and containing words that users feel comfortable with, especially in dementia-friendly DHTs. Future iterations of redesigning the website should focus on aligning user expectations with content presentation through co-design, which involves people living with dementia from the start of the process. This has been shown to be beneficial for both the person living with dementia and the design process itself [47-50]. However, a recent systematic review found that only 23% of studies involved people living with dementia in their design approach [51]. Additional recommendations to better align the content with the user expectations include considering a potential artificial intelligence–generated language check to ensure that all information relevant for patients is provided at the B1 level using: (1) a clear title and subheadings, (2) an active writing style with examples, (3) simple words that everyone knows, and (4) short and clear sentences. Moreover, terminology should be consistent, familiar, and comfortable and in Dutch (avoid

English terms; for example “second opinion” was found to be confusing).

System Feedback

Incomprehensible or seemingly invisible system feedback can lead to confusion and difficulties with navigating a website. For example, this occurred when clicking on a main item on the home page or the menu that refreshes the current page without any visible changes. It hinders the ability to understand the system’s responses, making it challenging to navigate the website effectively. This lack of clarity also reduces users’ confidence in their interactions, potentially causing them to make errors. Providing visually clear, short, and positive instructions may improve the understanding of users living with MCI or dementia about the system feedback on their interactions with the website [9,43]. In essence, clear and meaningful system feedback is essential for providing guidance and minimizing user errors. To tackle these issues, it is recommended to clearly distinguish the clickable and nonclickable areas to prevent confusion or endless clicks. After pressing a clickable area, the consequences of this click should be clearly visible or audible. This distinction can be made with colors, icons, buttons, vibrations, or audible beeps.

Limitations

This study had several limitations. First, the goal of the study was to conduct a heuristic evaluation with the DEMIGNED principles to investigate the reality of these principles in capturing usability problems that people living with dementia may experience. The realness was determined by mapping the findings from the think-aloud sessions to the findings from the heuristic evaluation. These insights suggest that 50% (18/36) of the findings by the experts were real usability problems experienced by potential end users. Even though this shows the potential of the DEMIGNED principles in heuristic evaluations, it may be possible that some usability issues identified from the heuristic evaluation were too subjective and hence not encountered in the user-testing method. This can be caused by methodological differences between the subjective nature of heuristic evaluations (as they predominantly rely on expert judgment founded on the DEMIGNED principles) and the objective user feedback derived from think-aloud tests involving direct user interactions. Second, an explanation as to why some usability problems from the heuristic evaluation were not encountered in the think-aloud sessions may be the fact that 18 tasks to guide the evaluators through most of the website were performed during the expert testing, while only 7 tasks were performed during user testing. Conducting 18 tasks was deemed to be too challenging in terms of cognitive overload, concentration, and motivation of the participants. The think-aloud tasks were composed with the aim of catching most of the pages that were evaluated during the heuristic evaluation. However, such a variety in the number of tasks could introduce bias in agreement scores calculations, such as Cohen κ, and was, therefore, left out of the analysis. Third, the software used for capturing the think-aloud sessions introduced some limitations. Due to the setup, participants were to use the smartphones provided by the researchers. This allowed the researchers to set up the screen-sharing function of the smartphone rather than going through the participants’ smartphone, which might cause stress or anxiety related to their privacy before starting the think-aloud session. This may introduce bias to the results, as the smartphone could be new for the participant. However, the study was conducted within an internet browser consistent across various smartphones. In addition, due to unannounced software updates, the sixth and seventh think-aloud session recordings suffered from audio issues. Therefore, only the observations from the video recordings were used in the analysis. However, saturation was reached, as no new usability issues were encountered during these sessions, which is important in think-aloud research [52]. Nevertheless, dementia does present a wide range of symptoms, so it might be possible that additional usability problems would arise for those with symptoms varying from the symptoms the participants experienced during the think-aloud sessions. Fourth, 1 user profile for the think-aloud session was created that captured a representative sample of people most likely to use the website of Alzheimer Center Amsterdam: those who visit the memory clinic to seek support for their cognitive complaints. The most frequent diagnosis for people aged <65 years, which is the focus of Alzheimer Center Amsterdam, is SCD, followed by the diagnoses of dementia and MCI [36]. This was reflected in the included sample for user testing. Therefore, we assume that with this group of participants, most usability issues were detected. However, these varying diagnoses may also influence the types and severities of usability problems. Nevertheless, in this study, the reality of the principles has been examined for this group as a whole, given the novelty of the DEMIGNED principles and their scientific basis of study, including people living with varying cognitive complaints, MCI, or dementia. Nevertheless, an approach to further investigate the reality of the DEMIGNED principles could be to compare the detected usability problems between people with varying levels of cognitive abilities and those without such cognitive issues to rule out the possibility that usability problems exist independently of cognitive abilities. Fifth, 1 evaluator only conducted approximately 50% of the tasks in the heuristic evaluation due to time constraints. This may have influenced the number of times a usability problem was encountered. However, the researcher was involved in the further analysis of the heuristic evaluation data. Finally, the sessions were conducted at the memory clinic. Even though we emphasized to the participants to imagine that they were in a home setting and allowed them to ask their partner or loved one for support, results may be different when this study is conducted in a home setting.

Future Research

Future research should focus on further evaluating the use of the DEMIGNED principles in expert reviews to assess their effectiveness in identifying real usability issues experienced by people living with varying types and stages of dementia. For example, digital health tools for people living with dementia that offer different functionalities from those offered by the mobile information resource evaluated in this study should be evaluated using the DEMIGNED principles, such as health monitoring, medication management, or participating in leisure or social activities. This may support categorizing the applicability of DEMIGNED principles per digital health functionality, as not all principles apply to each digital health tool. For example, in this study, the principle "ensure no time pressure" was not violated, as it was not applicable for the website. In addition, it is important to further evaluate the relevance of DEMIGNED principles for other types of user interactions, such as input controls (text fields, dropdown lists, toggles, etc) and serious games. This hopefully leads to more accessible digital health tools and the validation and refinement of the DEMIGNED principles. In the short term, the recommendations for redesign should be implemented into a new prototype for the website. The prototype should again be evaluated to ensure its usability. To enrich this evaluation, eye tracking and emotion readers could be used in think-aloud sessions to obtain more insights into the user experience because verbalizing what they are doing can be challenging for people living with dementia [53]. The emotion readers could also support in further investigating potential stress reactions of these people while conducting tasks on the website. Moreover, more extensive user testing could be conducted with the redesigned website, including metrics such as task completion rates and times, to gain more insights into the usability of the website. These metrics have been suggested to produce the most reliable results when conducting user testing with people living with dementia [53]. Finally, the literature suggests that an early decline in cognitive function may be detected from the input that people deliver when using a search engine [54]. The findings from this study related to using the search functionality initiate future research opportunities to optimize the search engine's functionality and explore opportunities to use this in research to detect cognitive decline.

Conclusions

This study showed that applying the DEMIGNED principles in expert testing can capture usability problems that people living with SCD, MCI, or dementia can experience when using a mobile website. The think-aloud analysis revealed 10 additional usability issues that were not captured in the heuristic evaluation approach. This shows the importance of involving end users in usability evaluations. Moreover, these findings provided insights into refining the DEMIGNED principles, mainly related to the use of a search function and barriers caused by the user's frame of mind. Future research should look into the applicability of the DEMIGNED principles for other functionalities that DHTs provide to increase the accessibility of digital health and decrease digital health inequity for this complex and rapidly increasing population.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Overview of unique usability issues captured from the heuristic evaluation.

[DOCX File , 32 KB - humanfactors_v11i1e54032_app1.docx]

References

1. Brands MR, Gouw SC, Beestrum M, Cronin RM, Fijnvandraat K, Badawy SM. Patient-centered digital health records and their effects on health outcomes: systematic review. J Med Internet Res 2022 Dec 22;24(12):e43086 [FREE Full text] [doi: 10.2196/43086] [Medline: 36548034]
2. Aapro M, Bossi P, Dasari A, Fallowfield L, Gascón P, Geller M, et al. Digital health for optimal supportive care in oncology: benefits, limits, and future perspectives. Support Care Cancer 2020 Oct 12;28(10):4589-4612 [FREE Full text] [doi: 10.1007/s00520-020-05539-1] [Medline: 32533435]
3. Phan P, Mitragotri S, Zhao Z. Digital therapeutics in the clinic. Bioeng Transl Med 2023 Jul 03;8(4):e10536 [FREE Full text] [doi: 10.1002/btm2.10536] [Medline: 37476062]
4. Gentili A, Failla G, Melnyk A, Puleo V, Tanna GL, Ricciardi W, et al. The cost-effectiveness of digital health interventions: a systematic review of the literature. Front Public Health 2022;10:787135 [FREE Full text] [doi: 10.3389/fpubh.2022.787135] [Medline: 36033812]
5. Kaihlanen AM, Virtanen L, Buchert U, Safarov N, Valkonen P, Hietapakka L, et al. Towards digital health equity - a qualitative study of the challenges experienced by vulnerable groups in using digital health services in the COVID-19 era. BMC Health Serv Res 2022 Feb 12;22(1):188 [FREE Full text] [doi: 10.1186/s12913-022-07584-4] [Medline: 35151302]
6. Pettersson L, Johansson S, Demmelmaier I, Gustavsson C. Disability digital divide: survey of accessibility of eHealth services as perceived by people with and without impairment. BMC Public Health 2023 Jan 27;23(1):181 [FREE Full text] [doi: 10.1186/s12889-023-15094-z] [Medline: 36707791]
7. Yao R, Zhang W, Evans R, Cao G, Rui T, Shen L. Inequities in health care services caused by the adoption of digital health technologies: scoping review. J Med Internet Res 2022 Mar 21;24(3):e34144 [FREE Full text] [doi: 10.2196/34144] [Medline: 35311682]
8. Banerjee A. Digital health interventions and inequalities: the case for a new paradigm. BMJ Evid Based Med 2021 Jun 21;26(3):77-78. [doi: 10.1136/bmjebm-2019-111282] [Medline: 31753847]

9. Engelsma T, Jaspers MW, Peute LW. Considerate mHealth design for older adults with Alzheimer's disease and related dementias (ADRD): a scoping review on usability barriers and design suggestions. Int J Med Inform 2021 Aug;152:104494 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104494] [Medline: 34015657]
10. ISO 9241-11:2018: ergonomics of human-system interaction: part 11: usability: definitions and concepts. International Organization for Standardization. 2018. URL: https://www.iso.org/standard/63500.html [accessed 2023-09-21]
11. Giebel GD, Speckemeier C, Abels C, Plescher F, Börchers K, Wasem J, et al. Problems and barriers related to the use of digital health applications: scoping review. J Med Internet Res 2023 May 12;25:e43808 [FREE Full text] [doi: 10.2196/43808] [Medline: 37171838]
12. Cahn A, Akirov A, Raz I. Digital health technology and diabetes management. J Diabetes 2018 Jan;10(1):10-17. [doi: 10.1111/1753-0407.12606] [Medline: 28872765]
13. Wei Y, Zheng P, Deng H, Wang X, Li X, Fu H. Design features for improving mobile health intervention user engagement: systematic review and thematic analysis. J Med Internet Res 2020 Dec 09;22(12):e21687 [FREE Full text] [doi: 10.2196/21687] [Medline: 33295292]
14. Fact sheets of dementia. World Health Organization. 2023. URL: https://www.who.int/news-room/fact-sheets/detail/dementia [accessed 2023-08-16]
15. Montine TJ, Koroshetz WJ, Babcock D, Dickson DW, Galpern WR, Glymour MM, ADRD 2013 Conference Organizing Committee. Recommendations of the Alzheimer's disease-related dementias conference. Neurology 2014 Aug 26;83(9):851-860 [FREE Full text] [doi: 10.1212/WNL.0000000000000733] [Medline: 25080517]
16. Conway A, Ryan A, Harkin D, Mc Cauley C, Goode D. A review of the factors influencing adoption of digital health applications for people living with dementia. Digit Health 2023 Mar 15;9:20552076231162985 [FREE Full text] [doi: 10.1177/20552076231162985] [Medline: 36937696]
17. Sohn M, Yang J, Sohn J, Lee JH. Digital healthcare for dementia and cognitive impairment: a scoping review. Int J Nurs Stud 2023 Apr;140:104413. [doi: 10.1016/j.ijnurstu.2022.104413] [Medline: 36821951]
18. Di Lorito C, Bosco A, Rai H, Craven M, McNally D, Todd C, et al. A systematic literature review and meta-analysis on digital health interventions for people living with dementia and Mild Cognitive Impairment. Int J Geriatr Psychiatry 2022 May 04;37(6). [doi: 10.1002/gps.5730]
19. Tsolaki M, Makri M, Tsatali M, Teichmann B. European projects for patients with dementia and their caregivers. Adv Exp Med Biol 2023;1425:609-618. [doi: 10.1007/978-3-031-31986-0_59] [Medline: 37581834]
20. ISO/TS 82304-2:2021: health software: part 2: health and wellness apps — quality and reliability. International Organization for Standardization. 2021. URL: https://www.iso.org/standard/78182.html [accessed 2023-09-21]
21. Web content accessibility guidelines (WCAG) 2.0. World Wide Web Consortium. URL: https://www.w3.org/TR/WCAG20 [accessed 2023-09-21]
22. Abts NA, McNicol SA, Branaghan RJ. Guidelines for Mobile Health Applications. Proc Int Symp Hum Factors Ergon Healthc 2019 Sep 15;8(1):197-200. [doi: 10.1177/2327857919081050]
23. Waller S, Bradley M, Hosking I, Clarkson PJ. Making the case for inclusive design. Appl Ergon 2015 Jan;46 Pt B:297-303. [doi: 10.1016/j.apergo.2013.03.012] [Medline: 23538129]
24. Jonsson M, Johansson S, Hussain D, Gulliksen J, Gustavsson C. Development and evaluation of eHealth services regarding accessibility: scoping literature review. J Med Internet Res 2023 Aug 17;25:e45118 [FREE Full text] [doi: 10.2196/45118] [Medline: 37590050]
25. Neal D, Engelsma T, Tan J, Craven MP, Marcilly R, Peute L, et al. Limitations of the new ISO standard for health and wellness apps. Lancet Digit Health 2022 Feb;4(2):e80-e82 [FREE Full text] [doi: 10.1016/S2589-7500(21)00273-9] [Medline: 35090677]
26. James A, Draffan EA, Wald M. Designing web-apps for all: how do we include those with cognitive disabilities? Stud Health Technol Inform 2017;242:665-668. [Medline: 28873869]
27. Engelsma T, Jaspers M, Peute L. Design principles and guidelines for inclusive mHealth design for people living with dementia. In: Niedderer K, Ludden G, Dening T, Holthoff-Detto V, editors. Design for Dementia, Mental Health and Wellbeing: Co-Design, Interventions and Policy. Oxfordshire, UK: Routledge; 2024:304-317.
28. Engelsma T, Jaspers M, Peute L. Enhancing inclusive mHealth design for people living with dementia: examples from literature. Stud Health Technol Inform 2023 May 18;302:646-650. [doi: 10.3233/SHTI230228] [Medline: 37203770]
29. Alcaraz-Martínez R, Ribera M, Adeva FA, Pascual AA. Validation of a heuristic set to evaluate the accessibility of statistical charts. In: Proceedings of the XXII International Conference on Human Computer Interaction. 2022 Presented at: Interacción '22; September 7-9, 2022; Teruel, Spain p. 1-9 URL: https://dl.acm.org/doi/10.1145/3549865.3549893 [doi: 10.1145/3549865.3549893]
30. Hartson HR, Andre TS, Williges RC. Criteria for evaluating usability evaluation methods. Int J Hum Comput Interact 2003 Feb;15(1):145-181. [doi: 10.1207/s15327590ijhc1501_13]
31. Jaspers MW. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. Int J Med Inform 2009 May;78(5):340-353. [doi: 10.1016/j.ijmedinf.2008.10.002] [Medline: 19046928]
32. Peute LW, Wildenbos G, Engelsma T, Lesselroth BJ, Lichtner V, Monkman H, et al. Overcoming challenges to inclusive user-based testing of health information technology with vulnerable older adults: recommendations from a human factors

engineering expert inquiry. Yearb Med Inform 2022 Aug;31(1):74-81 [FREE Full text] [doi: 10.1055/s-0042-1742499] [Medline: 35654432]
33. Home page. Alzheimercentrum Amsterdam. URL: http://www.alzheimercentrum.nl/ [accessed 2023-08-30]
34. Nielsen J. Finding usability problems through heuristic evaluation. In: Proceedings of the 1992 SIGCHI Conference on Human Factors in Computing Systems. 1992 Presented at: CHI '92; May 3-7, 1992; Monterey, CA p. 373-380 URL: https://dl.acm.org/doi/10.1145/142750.142834 [doi: 10.1145/142750.142834]
35. Nielsen J. Reliability of severity estimates for usability problems found by heuristic evaluation. In: Posters and Short Talks of the 1992 SIGCHI Conference on Human Factors in Computing Systems. 1992 Presented at: CHI '92; May 3-7, 1992; Monterey, CA p. 129-130 URL: https://dl.acm.org/doi/10.1145/1125021.1125117 [doi: 10.1145/1125021.1125117]
36. van der Flier WM, Scheltens P. Amsterdam dementia cohort: performing research to optimize care. J. Alzheimer's Dis 2018 Mar 13;62(3):1091-1111. [doi: 10.3233/jad-170850]
37. Viso. Noldus. URL: https://www.noldus.com/viso [accessed 2023-08-23]
38. Hvannberg ET, Law EL, Lárusdóttir MK. Heuristic evaluation: comparing ways of finding and reporting usability problems. Interact Comput 2007 Mar;19(2):225-240. [doi: 10.1016/j.intcom.2006.10.001]
39. Bertini E, Catarci T, Dix A, Gabrielli S, Kimani S, Santucci G. Appropriating heuristic evaluation for mobile computing. In: Lumsden J, editor. Human-Computer Interaction and Innovation in Handheld, Mobile and Wearable Technologies. Philadelphia, PA: IGI Global; 2011:20-41.
40. Inostroza R, Rusu C, Roncagliolo S, Jimenez C, Rusu V. Usability heuristics validation through empirical evidences: a touchscreen-based mobile devices proposal. In: Proceedings of the 31st International Conference of the Chilean Computer Science Society. 2012 Presented at: SCCC '12; November 12-16, 2012; Valparaiso, Chile p. 60-68 URL: https://ieeexplore.ieee.org/document/6694074 [doi: 10.1109/sccc.2012.15]
41. Neto OM, da Graça Pimentel M. Heuristics for the assessment of interfaces of mobile devices. In: Proceedings of the 19th Brazilian Symposium on Multimedia and the Web. 2013 Presented at: WebMedia '13; November 5-8, 2013; Salvador, Brazil p. 93-96 URL: https://dl.acm.org/doi/10.1145/2526188.2526237 [doi: 10.1145/2526188.2526237]
42. Zhang J, Johnson TR, Patel VL, Paige DL, Kubose T. Using usability heuristics to evaluate patient safety of medical devices. J Biomed Inform 2003 Feb;36(1-2):23-30. [doi: 10.1016/S1532-0464(03)00060-1]
43. Zheng J, Gresham M, Phillipson L, Hall D, Jeon YH, Brodaty H, et al. Exploring the usability, user experience and usefulness of a supportive website for people with dementia and carers. Disabil Rehabil Assist Technol 2024 May 22;19(4):1369-1381. [doi: 10.1080/17483107.2023.2180546] [Medline: 37086036]
44. Thomas M, Henderson D, Trudel C, Thomas N. Usability of a community-based dementia resource website: mixed methods study. JMIR Aging 2023 Apr 20;6:e40762 [FREE Full text] [doi: 10.2196/40762] [Medline: 37079355]
45. Is navigation useful? Nielsen Norman Group. URL: https://www.nngroup.com/articles/is-navigation-useful/ [accessed 2023-08-30]
46. Freeman ED, Clare L, Savitch N, Royan L, Litherland R, Lindsay M. Improving website accessibility for people with early-stage dementia: a preliminary investigation. Aging Ment Health 2005 Sep;9(5):442-448. [doi: 10.1080/13607860500142838] [Medline: 16024403]
47. Wang G, Marradi C, Albayrak A, van der Cammen TJ. Co-designing with people with dementia: a scoping review of involving people with dementia in design research. Maturitas 2019 Sep;127:55-63 [FREE Full text] [doi: 10.1016/j.maturitas.2019.06.003] [Medline: 31351521]
48. Chopra S, Pradhan A, Dixon E, Radnofsky ML, Ganesh K, Lazar A. Designing for and with people with dementia using a human rights-based approach. Ext Abstr Hum Factors Computing Syst 2021;44:1-8 [FREE Full text] [Medline: 34423336]
49. Morbey H, Harding AJ, Swarbrick C, Ahmed F, Elvish R, Keady J, et al. Involving people living with dementia in research: an accessible modified Delphi survey for core outcome set development. Trials 2019 Jan 06;20(1):12 [FREE Full text] [doi: 10.1186/s13063-018-3069-6] [Medline: 30612587]
50. Suijkerbuijk S, Nap HH, Cornelisse L, IJsselsteijn WA, de Kort YA, Minkman MM. Active involvement of people with dementia: a systematic review of studies developing supportive technologies. J Alzheimers Dis 2019 Jun 18;69(4):1041-1065 [FREE Full text] [doi: 10.3233/JAD-190050] [Medline: 31156158]
51. de Oliveira Cruz D, Chechetti C, Brucki SM, Takada LT, Nunes FL. A comprehensive systematic review on mobile applications to support dementia patients. Pervasive Mob Comput 2023 Mar;90:101757. [doi: 10.1016/j.pmcj.2023.101757]
52. How many participants for a UX interview? Nielsen Norman Group. 2021 Oct. URL: https://www.nngroup.com/articles/interview-sample-size/ [accessed 2023-08-30]
53. Gibson A, McCauley C, Mulvenna M, Ryan A, Laird L, Curran K, et al. Assessing usability testing for people living with dementia. In: Proceedings of the 4th Workshop on ICTs for Improving Patients Rehabilitation Research Techniques. 2016 Presented at: REHAB '16; October 13-14, 2016; Lisbon, Portugal p. 25-31 URL: https://dl.acm.org/doi/10.1145/3051488.3051492 [doi: 10.1145/3051488.3051492]
54. Austin J, Hollingshead K, Kaye J. Internet searches and their relationship to cognitive function in older adults: cross-sectional analysis. J Med Internet Res 2017 Sep 06;19(9):e307 [FREE Full text] [doi: 10.2196/jmir.7671] [Medline: 28877864]

Abbreviations

DHT: digital health technology
MCI: mild cognitive impairment
MOLDEM-US: mobile health for older adults with dementia–usability framework
SCD: subjective cognitive decline

Edited by A Kushniruk; submitted 03.11.23; peer-reviewed by N Blase, R Marcilly; comments to author 22.12.23; revised version received 15.02.24; accepted 06.03.24; published 31.07.24.

Please cite as:
Engelsma T, Heijmink S, Hendriksen HMA, Visser LNC, Lemstra AW, Jaspers MWM, Peute LWP
Capturing Usability Problems for People Living With Dementia by Applying the DEMIGNED Principles in Usability Evaluation Methods: Mixed Methods Study
JMIR Hum Factors 2024;11:e54032
URL: https://humanfactors.jmir.org/2024/1/e54032
doi:10.2196/54032
PMID:

Original Paper

Patients' Attitudes Toward the Use of Artificial Intelligence as a Diagnostic Tool in Radiology in Saudi Arabia: Cross-Sectional Study

Leena R Baghdadi[1], MBBS, Master CliEpi, PhD; Arwa A Mobeirek[2]; Dania R Alhudaithi[2]; Fatimah A Albenmousa[2]; Leen S Alhadlaq[2]; Maisa S Alaql[2]; Sarah A Alhamlan[2]

[1]Department of Family and Community Medicine, College of Medicine, King Saud University, Riyadh, Saudi Arabia

[2]College of Medicine, King Saud University, Riyadh, Saudi Arabia

Corresponding Author:
Leena R Baghdadi, MBBS, Master CliEpi, PhD
Department of Family and Community Medicine
College of Medicine
King Saud University
3332 King Khalid Road
Riyadh, 12372
Saudi Arabia
Phone: 966 114670836
Fax: 966 114670836
Email: lbaghdadi@ksu.edu.sa

Abstract

Background: Artificial intelligence (AI) is widely used in various medical fields, including diagnostic radiology as a tool for greater efficiency, precision, and accuracy. The integration of AI as a radiological diagnostic tool has the potential to mitigate delays in diagnosis, which could, in turn, impact patients' prognosis and treatment outcomes. The literature shows conflicting results regarding patients' attitudes to AI as a diagnostic tool. To the best of our knowledge, no similar study has been conducted in Saudi Arabia.

Objective: The objectives of this study are to examine patients' attitudes toward the use of AI as a tool in diagnostic radiology at King Khalid University Hospital, Saudi Arabia. Additionally, we sought to explore potential associations between patients' attitudes and various sociodemographic factors.

Methods: This descriptive-analytical cross-sectional study was conducted in a tertiary care hospital. Data were collected from patients scheduled for radiological imaging through a validated self-administered questionnaire. The main outcome was to measure patients' attitudes to the use of AI in radiology by calculating mean scores of 5 factors, distrust and accountability (factor 1), procedural knowledge (factor 2), personal interaction and communication (factor 3), efficiency (factor 4), and methods of providing information to patients (factor 5). Data were analyzed using the student *t* test, one-way analysis of variance followed by post hoc and multivariable analysis.

Results: A total of 382 participants (n=273, 71.5% women and n=109, 28.5% men) completed the surveys and were included in the analysis. The mean age of the respondents was 39.51 (SD 13.26) years. Participants favored physicians over AI for procedural knowledge, personal interaction, and being informed. However, the participants demonstrated a neutral attitude for distrust and accountability and for efficiency. Marital status was found to be associated with distrust and accountability, procedural knowledge, and personal interaction. Associations were also found between self-reported health status and being informed and between the field of specialization and distrust and accountability.

Conclusions: Patients were keen to understand the work of AI in radiology but favored personal interaction with a radiologist. Patients were impartial toward AI replacing radiologists and the efficiency of AI, which should be a consideration in future policy development and integration. Future research involving multicenter studies in different regions of Saudi Arabia is required.

(JMIR Hum Factors 2024;11:e53108) doi:10.2196/53108

KEYWORDS

artificial intelligence; diagnostic radiology; patients; attitudes; questionnaire; patient; attitude; diagnostic tool; diagnostic tools; AI; artificial intelligence; radiologists; prognosis; treatment; Saudi Arabia; sociodemographic factors; sociodemographic factor; sociodemographic; cross-sectional study; participant; men; women; analysis; distrust; trust

Introduction

Introduction to Artificial Intelligence

Artificial intelligence (AI) is a branch of computer science focused on creating systems that mimic human intelligence, enabling machines to learn, solve problems, and understand language and visuals. The aim is to develop machines capable of performing cognitive tasks traditionally associated with human intelligence. It is anticipated that this revolution in data science and technology will provide practical advantages in many areas [1]. AI has many subsets, including algorithmic machine learning and autonomous decision-making. AI is used in diverse specialties, including health care [2]. In health care, AI decreases the workload of health care providers [3] and improves disease prevention, diagnosis, management, and treatment; thereby, improving patient outcomes and decreasing the economic burden [1,2].

Numerous studies have explored the implementation of AI in diverse aspects of medicine [4-6]. AI-based technologies have been used in surgery, observation of pulmonary metastasis on computed tomography scans, diagnosis and evaluation of diabetic retinopathy, Alzheimer disease, heart arrhythmias, onychomycosis, vertebral compressions, cerebral aneurysms, brain neoplasms, assessment of psoriasis, detection of subarachnoid hemorrhage, and for glioblastoma prognosis after bevacizumab treatment [5-8]. AI is also expected to show advancements in differentiating lung diseases and improvements in breast and skin cancer detection and screening [4,7,9]. However, to integrate AI appropriately, it is crucial to grasp the viewpoints of various stakeholders, including patients and health care professionals like students, physicians, and caregivers, regarding the use of AI in clinical practices.

Implementation of AI in Radiology

Radiology is an important digital data generator. Due to advancements in technology, increasing rates of diagnostic procedures among older patients in an aging population, and improvements in screening programs [1,2], there is a notable rise in the number of scans to be evaluated and diagnosed. However, this demand coincides with a lack of trained radiologists in certain specialties [10]. As a result, radiologists face an increased workload, leading to marked delays in diagnosis and reduced interpretation power. Such delays could affect patients' prognosis and treatment outcomes [10-12].

Saudi Arabia has prioritized changing this practice to be more time efficient and improve the health care system, especially for clinical cases where time is crucial for quick diagnosis and immediate management (stroke, trauma, and cancer). In radiology, a surge of evidence has demonstrated the potential of AI to outperform physicians in some clinical aspects. The use of AI would decrease the time between patient diagnosis and administration of treatment. Fewer delays in diagnosis and treatment can help mitigate legal implications, such as malpractice cases, where timely and accurate diagnoses supported by AI can minimize the risk of misdiagnoses or delayed treatments that may result in patient harm [10]. Consequently, the implementation of AI in radiology became a heated debate and a subject of discussion in around 800 related published papers in 2017 [13]. AI developers have increased efforts to provide trustworthy technologies that aid image recognition and diagnostic tasks minimizing radiologists' workloads [12]. AI can achieve faster-customized diagnoses and recommendations equal to or superior to highly qualified radiologists for performance [4,7,11,14,15], precision, and accuracy [8], thus increasing diagnostic reliability [3].

The use of AI also involves ethical, legal, and societal concerns, which are essential and must be addressed. These involve protecting autonomy, ensuring transparency and accountability, and promoting human well-being, safety, and equity. Legal concerns include adhering to protection laws, establishing principles for the use of AI in health, and complying with bioethics regulations [16]. Recognizing these concerns, it is crucial for developers to engage all stakeholders, especially the patients that AI intends to serve [11,14,15]. The World Health Organization [16] states the public should be informed about the development of AI in health care as it will facilitate their understanding of data use and allocation and help them voice their concerns and anticipation of the use of AI technologies. The public should be encouraged to have further knowledge of AI technologies to determine those acceptable for use. Therefore, there is an increasing need to explore patient attitudes toward using AI in radiology and address social and ethical considerations.

Patients are important stakeholders in the decision-making process and understanding their perspective is necessary to ensure widespread implementation of these technological advances. By studying patients' attitudes, we can uncover potential barriers and facilitators to the integration of AI, ultimately informing strategies to enhance patient engagement, trust, and satisfaction. Moreover, exploring patients' viewpoints adds a crucial dimension to the discourse on AI in health care, fostering patient-centered approaches and ensuring that AI implementations align with patients' needs and preferences [4,7]. Positive patient attitudes to AI as a diagnostic tool will indicate readiness and support, whereas a lack thereof will indicate patient education is necessary. To the best of our knowledge, no studies have examined patients' attitudes toward the use of AI in radiology or health care in Saudi Arabia.

This study aimed to investigate the attitudes of patients toward the use of AI in diagnostic radiology at King Khalid University Hospital (KKUH) in Riyadh, Saudi Arabia, and to evaluate potential associations between patients' attitudes and specific sociodemographic factors. Our study could influence future policy making on the integration of AI into health care by defining critical points of supervision, it will guide program development and primary education on a national level, ensuring AI implementation caters to societal needs.

Methods

Recruitment

This was an analytical cross-sectional study conducted between July and December 2022 at the tertiary care hospital KKUH, Riyadh, Saudi Arabia. Eligible participants were outpatients scheduled for any type of radiological imaging involving any part of the body, aged ≥18 years old, literate, and able to speak Arabic or English. Mental health patients were excluded. We used convenience sampling techniques for ease of recruitment and cost-effectiveness. Participants were approached for enrollment in the radiology waiting rooms by 1 of 6 investigators. Participants who voluntarily agreed to participate and provided their explicit consent were provided a brief description of the role of AI in diagnostic radiology and asked to complete a self-administered electronic questionnaire via Google form in Arabic or English based on their preferences. Questionnaires were completed by participants from August 9, 2022, to September 1, 2022.

Study Variables

The main outcome of the study was patients' attitudes to using AI in radiology at KKUH. Demographic and socioeconomic data (age, sex, nationality, marital status, education level, place of residence, family income, employment status, specialty, self-reported health status, and self-reported prior knowledge about AI, and experience of diagnostic errors) were collected.

Survey Tools

The questionnaire had 2 sections, sociodemographic data and attitude toward AI in radiology. The first section had participant characteristics, including age, sex, nationality, marital status, education level, place of residence, family income, employment status, and field of specialization, and additional information such as self-reported health status, self-reported prior knowledge about AI, and experience of diagnostic errors.

The second section measured attitudes toward AI in radiology using a validated questionnaire (Multimedia Appendix 1) developed by Ongena et al [17]. It is a 39-item tool that calculates the average score of 5 dimensions, which are distrust and accountability (15 items), procedural knowledge (8 items), personal interaction (7 items) and communication, efficiency (5 items), and ways patients are informed of their imaging results and prognosis (4 items). Respondents assessed each item on a 5-point Likert scale ranging from "Strongly Disagree" to "Strongly Agree." The mean of statements within each factor was then calculated. Subsequently, the scores for the 5 factors were categorized into levels of attitude: "Strongly Negative Attitude" (1.00-1.80), "Moderately Negative Attitude" (1.81-2.60), "Neutral Attitude" (2.61-3.40), "Moderately Positive Attitude" (3.41-4.20), and "Strongly Positive Attitude" (4.21-5.00), using criteria corresponding to the Likert scale values. The higher scores indicate being more negative toward the use of AI in radiology.

The tool was originally developed in English and conceptually translated into Arabic by a certified translator whose first language is Arabic and who is fluent in English. A panel of 12 experts (radiologists, public health consultants, epidemiologists, and family and community medicine consultants) reviewed the Arabic version of the questionnaire and compared it with the original English version. The panel provided several suggestions for improving the translation. A pilot study was conducted and responses from the pilot study were excluded from the analysis.

Statistical Analysis

Based on previous reports of a 1.10 difference between participants' attitudes toward AI [17], the minimum sample size consisted of 384 participants in total and had more than 80% power (α=.05) to detect small differences between participants. Ten percent was added to account for nonresponses, yielding a total sample size of 423 participants.

Data were analyzed using the SPSS (version 27; IBM Corp). Descriptive statistics (frequencies, percentages, means, and SDs) were used to describe the categorical and quantitative variables. The hypothesis testing was 1-tailed. Associations were determined between the patients' attitudes to using AI in radiology and demographic and socioeconomic data (such as age, sex, nationality, marital status, and education level) through a bivariate analysis using the student *t* test for independent samples or a one-way analysis of variance followed by post hoc analysis to test for quantitative outcome variables to compare the mean values in relation to the categorical study variables, which have 2 and >2 options, respectively. Common confounders, such as sociodemographic factors, were analyzed either via student *t* test or ANOVA as appropriate. The Pearson correlation coefficient was used to determine the association between continuous quantitative study variables (age) and the outcome variables.

Multivariable Analysis

Multivariable regression models examined whether any associations between the sociodemographic variables and common confounders and quantitative outcome variables were retained after controlling for selected covariates. Sociodemographic variables that were found statistically significant (or approaching significance level) in bivariate analysis were entered as predictors in the multivariable regression model. Four regression models were built (factors 1, 2, and 5, overall score). Factors 3 and 4 only had one significant (or approaching significance) predictor, therefore multivariable analysis was not warranted. Categorical predictors were included with the most suitable category being chosen as the reference category. For each regression model, we reported the overall model fit, significance of each predictor (omnibus ANOVA), regression coefficients (including 95% CI and standardized estimate), and checked for multicollinearity assumption (variance inflation factor values). Variance inflation factor values <5 indicate no presence of multicollinearity.

A level of significance of 0.05 was used for all inferential analyses. The 95% CI was reported where applicable. *P* values were reported for inferential tests, with *P*<.05 interpreted as statistically significant. Missing data such as occupation and social status were collected from patients' files through the electronic system for integrated health information (e-SIHI). Invalid entries were excluded from the analysis, such as patients who had nondiagnostic radiological imaging or those who were not waiting for a radiology appointment regarding their own health.

Ethical Considerations

The study was approved by the Institutional Review Board of the College of Medicine at King Saud University on July 24, 2022 (research project number E-22-6966). All participants were verbally informed and were given sufficient time to read the consent form and give written consent before enrollment (Multimedia Appendix 1). Participation was voluntary, and subjects retained the autonomy to withdraw from the study at any point without incurring any adverse consequences.

Results

Descriptive Statistics

There were 382 completed surveys with a nonresponse rate of ≈11%. Participants were 18-86 years old (mean 39.51, SD 13.26 years), 273 (71.5%) women and 109 (28.5%) men, 281 (73.6%) were married, and 293 (76.7%) were college-educated individuals. There were 180 (47.1%) employed participants and 131 (34.3%) of participants had a monthly household income of US $ 2666-5331. The participants' detailed demographic information is in Table 1.

The participants were scheduled for ultrasound imaging (209/382, 54.7%), magnetic resonance imaging (46/382, 12%), x-rays (43/382, 11.3%), computed tomography scans (40/382, 10.5%), mammograms (20/382, 5.2%), echocardiography (18/382, 4.7%), and angiography (6/382, 1.6%). Some participants (8.6%) had previously experienced diagnostic medical errors (Table S1, Multimedia Appendix 2). Most respondents evaluated their prior knowledge of AI as average (36.1%) or very good (27.7%; Table S2, Multimedia Appendix 2). The top 3 sources of their AI information were internet sources (80.4%), social media (66.2%), and friends and peers (42.7%; Table S3, Multimedia Appendix 2).

Table 2 presents the average scores for patients' attitudes toward AI in radiology per statement and factor. Participants were neutral in their trust of AI taking over radiologists' diagnostic interpretation tasks for accuracy, communication, and confidentiality, 3.16 was the average score for factor 1 (distrust and accountability). The average score for factor 2 (procedural knowledge) was 4.08, signifying that patients are interested in understanding how radiological images are obtained, interpreted, and disseminated. The average score for factor 3 (personal interaction) was 4.06, indicating that patients favored personal interaction with a radiologist over AI-based communication. The average score for factor 4 (efficiency) was 2.89, suggesting patients were ambiguous about AI improving the diagnostic procedure. Factor 5 (being informed) had an average score of 3.65, showing patients favor obtaining full disclosure of their medical findings and predictions of any future diseases they might develop, in addition to full-body scans performed by AI rather than scans of selected parts of the body.

Table 1. Sociodemographic and health characteristics of the participants (N=382).

Characteristics	Values, n (%)
Age group (years)	
18-19	8 (2.1)
20-29	78 (20.4)
30-39	143 (37.4)
40-49	68 (17.8)
≥50	85 (22.3)
Sex	
Male	109 (28.5)
Female	273 (71.5)
Nationality	
Saudi	355 (92.9)
Non-Saudi	27 (7.1)
Residence	
Riyadh region	342 (89.5)
Outside Riyadh region	40 (10.5)
Education	
No formal education	3 (0.8)
Literacy school	3 (0.8)
Elementary education	10 (2.6)
Intermediate education	13 (3.4)
Secondary education	60 (15.7)
College (Diploma, Bachelor, Master, or PhD)	293 (76.7)
Field of specialization (n=376)	
Health sciences	49 (12.8)
Scientific field	41 (10.7)
Humanities	139 (36.4)
Technology and computer science	26 (6.8)
Administrative field	73 (19.1)
Other	3 (0.8)
None	45 (11.8)
Employment status	
Student	41 (10.7)
Employed	180 (47.1)
Unemployed	39 (10.2)
Retired	41 (10.7)
Housewife	81 (21.2)
Marital status	
Married	281 (73.6)
Unmarried	73 (19.1)
Divorced or widowed	28 (7.3)
Monthly household income (US $)	
≤1333	84 (22)

Characteristics	Values, n (%)
1333-2665	121 (31.7)
2666-5331	131 (34.3)
5332-7997	26 (6.8)
>7997	20 (5.2)
Self-reported health status	
Excellent	122 (31.9)
Very good	159 (41.6)
Average	89 (23.3)
Fair	10 (2.6)
Poor	2 (0.5)

Table 2. Results of the 39 questionnaire statements that provide scores for 5 factors.

Items	Score, mean[a] (SD)
Factor 1 (distrust and accountability)	**3.16 (0.59)**
1. A computer can never compete against the experience of a specialized doctor (radiologist)	3.19 (1.04)
2. Through human experience, a radiologist can detect more than the computer	3.30 (0.99)
3. Humans have a better overview than computers on what happens in my body	3.52 (0.97)
4. It worries me when computers analyze scans without interference of humans	3.31 (1.12)
5. I wonder how it is possible that a computer can give me the results of a scan	2.88 (1.11)
6. Artificial intelligence makes doctors lazy	2.83 (1.18)
7. Humans and artificial intelligence can complement each other	2.70 (1.04)
8. I think replacement of doctors by artificial intelligence will happen in the far future	2.98 (1.12)
9. I would never blindly trust a computer	3.30 (1.08)
10. Artificial intelligence can only be implemented to check human judgment	3.56 (1.10)
11. I find it worrisome that a computer does not take feelings into account	3.50 (1.13)
12. It is unclear to me how computers will be used in evaluating scans	3.19 (1.02)
13. Even if computers are better in evaluating scans, I still prefer a doctor	3.31 (1.15)
14. When artificial intelligence is used, my personal data may fall into the wrong hands	3.01 (1.16)
15. Artificial intelligence may prevent errors[b]	2.77 (0.98)
Factor 2 (procedural knowledge)	**4.08 (0.71)**
1. I find it important to have a good understanding of the results of a scan	4.16 (0.83)
2. I find it important to be able to ask questions personally about the results of a scan	4.17 (0.84)
3. I find it important to talk with someone about the results of a scan	4.19 (0.87)
4. I find it important that a scan provides as much information about my body as possible	4.18 (0.88)
5. I find it important to get the results of a scan as fast as possible	4.13 (0.85)
6. I find it important to ask questions on the reliability of the results	4.11 (0.92)
7. I find it important to be well informed about how a scan is made	4.02 (0.94)
8. I find it important to read how radiologists work before I get a scan	3.68 (0.99)
Factor 3 (personal interaction)	**4.06 (0.71)**
1. When discussing the results of a scan, humans are indispensable	4.16 (0.91)
2. Getting the results involves personal contact	3.96 (0.93)
3. As a patient, I want to be treated as a person, not as a number	4.14 (0.87)
4. When a computer gives the result, I would miss the explanation	3.69 (1.12)
5. I find it important to ask questions when getting the result	4.18 (0.85)
6. Even when computers are used to evaluate scans, humans always remain responsible	4.13 (0.92)
7. Humans and artificial intelligence can complement each other	4.17 (0.94)
Factor 4 (efficiency)	**2.89 (0.46)**
1. As far as I am concerned, artificial intelligence can replace doctors in evaluating scans[b]	3.08 (1.04)
2. The sooner I get the results, even when this is from a computer, the more I am at ease	3.50 (1.01)
3. Because of the use of artificial intelligence, fewer doctors and radiologists are required[b]	2.61 (1.07)
4. Evaluating scans with artificial intelligence will reduce health care waiting times[b]	2.30 (0.95)
5. In my opinion, humans make more errors than computers[b]	2.95 (0.93)
Factor 5 (being informed)	**3.65 (0.62)**
1. If it does not matter in costs, a computer should always make a full-body scan instead of looking at specific body parts	3.56 (1.07)

Items	Score, mean[a] (SD)
2. If a computer would give the results, I would not feel emotional support	3.53 (0.98)
3. A computer should only look at body parts that were selected by my doctor	3.48 (1.05)
4. When a computer can predict that I will get a disease in the future, I want to know that no matter what	4.02 (0.97)

[a]The mean score of statements measured on a 5-point scale (strongly disagree-strongly agree). For all factors, higher scores indicate being more negative toward the use of artificial intelligence in radiology.

[b]Items marked are recoded to measure in the same direction.

Associations

The associations of factors with participant characteristics are presented in Table 3. Factor 1 was significantly associated with the participants' study specialization, the level of distrust in AI was lower among individuals in the administrative field (2.88, SD 0.62) compared with those specializing in humanities (3.20, SD 0.58; P=.003), health sciences (3.28, SD 0.63; P=.005), and those with no specialty (3.36, SD 0.61; P<.001). Factor 1 was significantly related to employment status ($F_{4,377}$=2.74, P=.03) and self-reported health status ($F_{3,378}$=2.88, P=.04). However, post hoc analysis did not reveal a significant difference between employment status and the self-reported health status subgroups.

With regard to self-reported health status, a statistically significant difference was noted in the mean scores of factor 5. Participants who reported an excellent health status (3.50, SD 0.62) had significantly lower scores for factor 5 than those who evaluated their health status as average (3.74, SD 0.70, P=.03) or fair/poor (4.17, SD 0.59; P=.002).

Factors 1, 2, and 3 showed a statistically significant association with marital status on univariate analysis; divorced and widowed individuals showed a higher level of distrust of AI in radiology (factor 1, 3.46, SD 0.72), a higher need for active engagement (factor 2, 4.42, SD 0.50), and a higher appreciation for personal interaction (factor 3, 4.40, SD 0.55) when compared with married individuals (factor 1, 3.15, SD 0.57, P=.02; factor 2, 4.04, SD 0.75, P=.02; factor 3, 4.02, SD 0.74, P=.02). Divorced or widowed individuals had a higher level of distrust of AI in radiology than unmarried participants (factor 1, 3.08, SD 0.59, P=.01).

On univariate analysis, factor 1 was weakly positively associated with age (r=0.124, P=.02; Figure 1). When participants were categorized into different age groups, no significant differences were observed. None of the factors showed statistically significant associations with sex, nationality, residence, education level, income, or self-reported knowledge about AI.

Table 3. Comparison of mean scores of the 5 factors in relation to sociodemographic and health characteristics of the participants.

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Age group (years), mean (SD)					
18-19	3.09 (0.56)	4.05 (0.45)	4.29 (0.58)	2.83 (0.55)	3.72 (0.45)
20-29	2.98 (0.39)	4.14 (0.74)	4.11 (0.62)	2.82 (0.45)	3.66 (0.64)
30-39	3.16 (0.56)	4.11 (0.71)	4.06 (0.72)	2.90 (0.49)	3.60 (0.61)
40-49	3.14 (0.65)	4.01 (0.66)	4.00 (0.77)	2.91 (0.46)	3.66 (0.62)
≥50	3.25 (0.65)	4.04 (0.77)	4.05 (0.73)	2.90 (0.40)	3.71 (0.66)
F test (age group; 4, 377)	0.966	0.485	0.402	0.515	0.508
P value (age group)	.43	.75	.81	.73	.73
Marital status, mean (SD)					
Unmarried	3.08 (0.59)[b]	4.13 (0.60)	4.07 (0.58)	2.88 (0.46)	3.72 (0.56)
Married	3.15 (0.57)[b]	4.04 (0.75)[b]	4.02 (0.74)[b]	2.89 (0.46)	3.61 (0.63)
Divorced or widowed	3.46 (0.72)[b]	4.42 (0.50)[b]	4.40 (0.55)[b]	2.88 (0.46)	3.86 (0.71)
F test (marital status; 2, 379)	4.449	3.918	3.592	0.004	2.514
P value (marital status)	.01[a]	.02[a]	.03[a]	.99	.08
Education, mean (SD)					
No formal education	3.24 (0.71)	4.29 (0.40)	4.71 (0.25)	3.00 (0.40)	4.25 (0.66)
Literacy school	3.87 (0.71)	4.54 (0.29)	4.33 (0.50)	2.80 (0.20)	4.25 (0.90)
Elementary education	3.45 (0.79)	3.88 (1.28)	3.89 (1.14)	2.92 (0.51)	3.83 (0.69)
Intermediate education	3.24 (0.77)	3.85 (0.83)	3.91 (0.92)	2.85 (0.52)	3.71 (0.92)
Secondary education	3.16 (0.62)	3.96 (0.82)	4.00 (0.85)	3.03 (0.49)	3.55 (0.72)
College	3.14 (0.57)	4.12 (0.66)	4.08 (0.65)	2.86 (0.45)	3.65 (0.58)
F test (education; 5, 376)	1.482	1.270	0.973	1.516	1.643
P value (education)	.20	.28	.43	.18	.15
Field of specialization, mean (SD)					
Humanities	3.20 (0.58)[b]	4.10 (0.74)	4.07 (0.69)	2.88 (0.46)	3.60 (0.62)
Health sciences	3.28 (0.63)[b]	4.19 (0.65)	4.12 (0.59)	2.86 (0.51)	3.68 (0.57)
Scientific field	3.06 (0.47)	4.18 (0.57)	4.19 (0.62)	2.86 (0.42)	3.79 (0.57)
Technology and computer science	3.21 (0.44)	4.14 (0.58)	4.14 (0.70)	2.66 (0.41)	3.78 (0.59)
Administrative	2.88 (0.62)[b]	4.03 (0.71)	3.96 (0.75)	2.99 (0.44)	3.57 (0.59)
Other	3.51 (0.17)	4.33 (0.58)	4.52 (0.50)	2.80 (0.35)	4.00 (0.50)
None	3.36 (0.61)[c]	3.92 (0.89)	3.94 (0.87)	2.92 (0.47)	3.70 (0.81)
F test (field of specialization; 6, 369)	4.563	0.865	1.006	1.730	1.077
P value (field of specialization)	<.001[a]	.52	.42	.11	.38
Employment status, mean (SD)					
Student	3.01 (0.51)	4.01 (0.71)	3.97 (0.64)	2.96 (0.51)	3.65 (0.68)
Employed	3.12 (0.60)	4.12 (0.68)	4.06 (0.73)	2.88 (0.50)	3.63 (0.63)
Unemployed	3.07 (0.61)	4.26 (0.60)	4.18 (0.59)	2.86 (0.41)	3.72 (0.49)
Retired	3.31 (0.53)	4.02 (0.68)	4.05 (0.64)	2.92 (0.34)	3.73 (0.53)
Housewife	3.29 (0.62)	3.97 (0.85)	4.06 (0.77)	2.87 (0.43)	3.61 (0.68)
F test (employment status; 4, 377)	2.740	1.463	0.432	0.365	0.408

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
P value (employment status)	.03[a]	.21	.79	.83	.80
Self-reported health status, mean (SD)					
Excellent	3.05 (0.52)	4.03 (0.82)	3.96 (0.84)	2.96 (0.49)	3.50 (0.62)[b]
Very good	3.20 (0.57)	4.11 (0.61)	4.10 (0.61)	2.84 (0.43)	3.67 (0.55)
Average	3.20 (0.65)	4.06 (0.74)	4.09 (0.67)	2.87 (0.48)	3.74 (0.70)[b]
Fair/poor[d]	3.46 (0.90)	4.36 (0.68)	4.32 (0.73)	2.88 (0.34)	4.17 (0.59)[b]
F test (self-reported health status; 3, 378)	2.878	0.948	1.520	1.517	6.042
P value (self-reported health status)	.04[a]	.42	.21	.21	<.001[a]
Self-reported AI knowledge, mean (SD)					
Excellent	3.13 (0.66)	4.16 (0.82)	4.10 (0.82)	2.90 (0.55)	3.66 (0.74)
Very good	3.09 (0.52)	4.07 (0.64)	4.06 (0.66)	2.85 (0.48)	3.60 (0.58)
Average	3.18 (0.60)	4.10 (0.68)	4.09 (0.69)	2.89 (0.47)	3.64 (0.63)
Fair	3.20 (0.56)	3.99 (0.79)	3.95 (0.72)	2.92 (0.35)	3.64 (0.61)
Poor	3.24 (0.73)	4.08 (0.81)	4.07 (0.75)	2.89 (0.36)	3.83 (0.56)
F test (self-reported AI knowledge; 4, 377)	0.714	0.422	0.456	0.240	0.868
P value (eslf-reported AI knowledge)	.58	.79	.77	.92	.48
Monthly household income US$, mean (SD)					
≤1,333	3.21 (0.59)	4.03 (0.80)	3.97 (0.78)	2.93 (0.45)	3.64 (0.76)
1,333-2,665	3.12 (0.60)	4.02 (0.72)	4.02 (0.68)	2.93 (0.46)	3.60 (0.58)
2,666-5,331	3.16 (0.61)	4.11 (0.69)	4.10 (0.73)	2.84 (0.43)	3.66 (0.61)
5,332-7,997	3.12 (0.59)	4.23 (0.63)	4.25 (0.59)	2.74 (0.55)	3.71 (0.51)
>7,997	3.19 (0.51)	4.24 (0.48)	4.17 (0.56)	2.90 (0.57)	3.80 (0.48)
F test (monthly household income; 6, 369)	0.301	0.883	1.119	1.500	0.571
P value (monthly household income)	.88	.47	.35	.20	.68
Sex, mean (SD)					
Male	3.09 (0.59)	4.00 (0.76)	3.99 (0.74)	2.90 (0.48)	3.65 (0.63)
Female	3.18 (0.60)	4.11 (0.70)	4.09 (0.69)	2.88 (0.45)	3.65 (0.62)
t value (sex; 380)	–1.310	1.463	0.432	0.365	0.408
P value (sex)	.19	.21	.79	.83	.80
95% CI (sex)	–0.220 to 0.044	–0.271 to 0.047	–0.261 to 0.054	–0.089 to 0.116	–0.142 to 0.136
Nationality, mean (SD)					
Saudi	3.14 (0.59)	4.07 (0.72)	4.06 (0.72)	2.89 (0.47)	3.64 (0.63)
Non-Saudi	3.33 (0.68)	4.23 (0.62)	4.11 (0.54)	2.86 (0.38)	3.71 (0.57)
t value (nationality; 380)	–1.528	–1.099	–0.341	0.312	–0.551
P value (nationality)	.13	.27	.73	.76	.58
95% CI (nationality)	–0.414 to 0.052	–0.437 to 0.124	–0.326 to 0.230	–0.152 to 0.209	–0.314 to 0.176
Residence, mean (SD)					
Riyadh region	3.16 (0.57)	4.10 (0.68)	4.06 (0.70)	2.89 (0.45)	3.65 (0.61)
Outside Riyadh region	3.15 (0.79)	3.89 (0.93)	4.11 (0.79)	2.83 (0.54)	3.62 (0.75)
t value (residence; 380)	0.096	1.816	–0.436	0.812	0.326
P value (residence)	.92	.07	.66	.42	.74

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
95% CI (residence)	–0.183 to 0.208	–0.018 to 0.450	–0.284 to 0.181	–0.089 to 0.213	–0.171 to 0.239

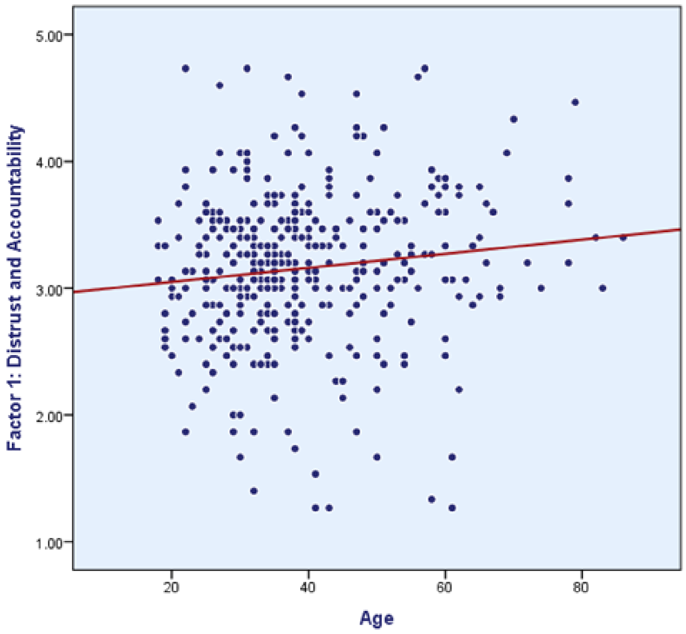
[a]A significant difference between means within a variable.

[b]Groups with a significant mean difference at the .05 level.

[c]Groups with a significant mean difference at the 0.001 level.

[d]"fair" and "poor" categories were combined for analysis purposes.

Figure 1. Pearson correlation between the ages of participants and factor 1 (distrust and accountability).



Regression Models

Overall, model 1 (factor 1, distrust and accountability) was statistically significant ($F_{15,356}$=3.48, P<.001) with a coefficient of determination (R^2) of 0.128. The model explains a 12.8% variability in factor 1 scores. Out of 5 predictors (age, marital status, field of specialization, employment, and self-reported health status), only marital status and field of specialization were statistically significant (P=.04 and P<.001, respectively). Controlling for other factors in the model, participants who specialized in scientific and administrative fields still showed lower levels of distrust compared with those with no specialty; the scores were 0.27 lower (P=.047) and 0.44 lower (P<.001) than that of no-specialty participants, respectively. Divorced or widowed participants also showed higher scores when compared with married participants by 0.29 (P=.01), controlling for other factors in the model (Table 4).

Overall, model 2 (factor 2, procedural knowledge) was statistically significant ($F_{3,378}$=3.41, P=.02) with a coefficient of determination (R^2) of 0.026. The model explains only 2.6% variability in factor 2 scores. Out of 2 predictors (marital status and living region), only marital status was statistically significant (P=.03). Divorced or widowed participants showed factor 2 scores 0.36 higher compared with married participants (P=.01), controlling for living region (Table 4).

Overall, model 3 (factor 5, being informed) was statistically significant ($F_{6,374}$=3.99, P<.001) with a coefficient of determination (R^2) of 0.060. The model explains only 6.0% variability in factor 5 scores. Out of 3 predictors (age, marital status, and self-reported health status), only self-reported health status was statistically significant (P=.004).

Compared with excellent health status, participants with very good, average, fair/poor health status reported significantly higher factor 5 scores by 0.17 (P=.02), 0.23 (P=.01), and 0.58 (P=.002) respectively, controlling for age and marital status (Table 4).

Overall, model 4 (overall score) was statistically significant ($F_{10,362}$=3.22, P<.001) with a coefficient of determination (R^2) of 0.082. The model explains 8.2% variability in the overall scores. Out of 3 predictors (marital status, field of specialization, and self-reported health status, only marital status and health status remained statistically significant (P=.004 and P=.02 respectively). Divorced or widowed participants reported overall scores 0.28 higher compared with married participants (P<.001), controlling for field of specialization and self-reported health status. Compared with excellent health status, participants with very good, average, fair/poor health status reported significantly higher overall scores by 0.13 (P=.01), 0.12 (P=.04), and 0.29 (P=.02) respectively, controlling for field of specialization and marital status (Table 4).

Table 4. Multivariable models of mean scores of factors 1, 2, and 5 and overall score.

	Model 1 (DV[a]: Factor 1 score [Distrust and accountability])	Model 2 (DV: Factor 2 score [Procedural knowledge])	Model 3 (DV: Factor 5 score [Being informed])	Model 4 (DV: Overall score)
Model fit				
F test (*df*)	3.48 (15, 356)	3.41 (3, 378)	3.99 (6, 374)	3.22 (10, 362)
P value	<.001	.02	<.001	<.001
R^2	0.128	—[b]	0.060	0.082
Age (years)				
RC[c] (95% CI)	0.001 (–0.01 to 0.01)	—	0.003 (–0.002 to 0.01)	—
P value	.87	—	.27	—
R^2	—	0.026	—	—
Marital status (ref[d]: married)				
RC (95% CI)				
Unmarried	0.05 (–0.13 to 0.23)	0.08 (–0.10 to 0.27)	0.16 (–0.02 to 0.34)	0.02 (–0.09 to 0.13)
Divorced/widowed	0.29[e] (0.06 to 0.53)	0.36[e] (0.08 to 0.64)	0.19 (–0.05 to 0.43)	0.28[f] (0.11 to 0.45)
P value	.04	.03	.08	.004
Specialty (ref: none)				
RC (95% CI)				
Humanities	–0.12 (–0.33 to 0.09)	—	—	0.02 (–0.12 to 0.16)
Health sciences	0.03 (–0.23 to 0.30)	—	—	0.10 (–0.07 to 0.27)
Scientific field	–0.27[e] (–0.52 to –0.004)	—	—	0.02 (–0.16 to 0.19)
Technology and Computer science	–0.03 (–0.33 to 0.26)	—	—	0.06 (–0.14 to 0.26)
Administrative field	–0.44[f] (–0.68 to –0.20)	—	—	–0.12 (–0.28 to 0.03)
P value	<.001	—	—	—
Employment (ref: unemployed)				
RC (95% CI)				
Student	–0.07 (–0.35 to 0.20)	—	—	—
Employed	0.11 (–0.11 to 0.32)	—	—	—
Retired	0.28 (–0.05 to 0.61)	—	—	—
Housewife	0.16 (–0.08 to 0.40)	—	—	—
P value	.28	—	—	—
Region (ref: outside Riyadh)				
RC (95% CI)				
Riyadh region	—	0.18 (–0.05 to 0.42)	—	—
P value	—	.13	—	—
Health status (ref: excellent)				
RC (95% CI)				
Very good	0.14 (–0.0004 to 0.28)	—	0.17[e] (0.02 to 0.32)	0.13[e] (0.03 to 0.23)

	Model 1 (DV[a]: Factor 1 score [Distrust and accountability])	Model 2 (DV: Factor 2 score [Procedural knowledge])	Model 3 (DV: Factor 5 score [Being informed])	Model 4 (DV: Overall score)
Average	0.13 (–0.04 to 0.29)	—	0.23[e] (0.05 to 0.40)	0.12[e] (0.01 to 0.24)
Fair/poor	0.28 (–0.07 to 0.64)	—	0.58[g] (0.21 to 0.96)	0.29[e] (0.04 to 0.53)
P value	.16	—	.004	.02

[a]DV: dependent variable.

[b]Not applicable.

[c]RC: regression coefficient.

[d]ref: reference category.

[e]*P*<.05.

[f]*P*<.001.

[g]*P*<.01.

Discussion

Principal Findings

To the best of our knowledge, this is the first study in Saudi Arabia to examine the attitudes to AI as a diagnostic tool from the patient's perspective. In this cross-sectional study with 382 participants, patients in the radiology waiting rooms at KKUH had a moderately positive attitude toward the use of AI as a diagnostic tool in radiology. This was similar to our hypothesis with reference to Jutzi et al [8] and Young et al [18] and contrary to Ongena et al [17] and Lennartz et al [4].

Regarding factor 1 (distrust and accountability), even though patients were neutral in their trust of AI taking over radiologists' diagnostic interpretation tasks; they believe that AI might enhance the accuracy of radiological diagnosis. This finding is similar to the conclusion by Jutzi et al [8], Young et al [18], and Lennartz et al [4] that AI is perceived to enhance the accuracy of radiological diagnosis and uses the latest in diagnostic procedures. As we had hypothesized that patients would show a positive attitude toward perceiving the knowledge behind radiological diagnosis and personal interaction with average scores of around 4.5 and 4.4, respectively; and in line with previous studies [4,17,18], we found that patients were interested in understanding how radiological images are obtained, interpreted, and disseminated, they favored personal interaction with a radiologist over AI-based communication, were ambiguous about AI improving the diagnostic procedure, and favored obtaining full disclosure of their medical findings and predictions of future diseases they might develop, in addition to full-body scans performed by AI rather than scans of selected body parts.

With regards to the associations between age and patients' attitudes toward AI as a diagnostic tool, our study showed a weak positive association between age and factor 1 (distrust and accountability) on univariate analysis although multivariable analysis showed no statistical significance. Ongena et al [17] indicated that age was weakly positively associated with factor 2 (procedural knowledge) and weakly negatively associated with factor 4 (efficiency). Young et al [18] reported that younger university students had more positive attitudes toward AI as a diagnostic tool and aimed to use AI as a diagnostic tool in radiology. Contrary to the findings of Young et al [18], who reported that males were more accepting of AI as a radiological diagnostic tool and Jutzi et al [8] who reported female participants had a more positive attitude toward the use of AI in radiology than males, our study and that of Ongena et al [17] showed no significant associations between sex and the factors determining attitudes to AI as a diagnostic tool in radiology.

While Ongena et al [17] observed an increase in trust in AI-based technologies with higher levels of education, our study found no significant association between education levels and patients' attitudes toward AI use in radiology. Nevertheless, our results align with the findings by Jutzi et al [8]; the similarity in educational backgrounds, with a notable percentage of participants in both studies holding undergraduate or postgraduate degrees (121 participants, 40.6% in Jutzi et al [8] and 293 participants, 76% in our study), coupled with a middle-aged demographic, suggests a potential wariness towards AI. This wariness may stem from concerns about privacy invasion, or limited digital skills and financial resources required for technology use [19].

The field of specialization showed a statistically significant association (*P*<.001) with patients' trust in the use of AI in radiology. Participants in scientific and administrative fields reported lower levels of distrust in AI compared with participants with no specialty (*P*<.001). People who are well educated in a certain subject tend to build more trust in it. Thus, AI experts showed greater confidence and positivity in their views about AI implementation in medicine and health care when compared with the general public in the United States [20]. These results might be due to their previous exposure to AI for work or study, which facilitated the intention to use it in health care.

Participants who reported excellent health status have also expressed a higher need to obtain full disclosure and be informed by the AI diagnostic tool about their overall health status when compared with participants who reported their health status as average or fair/poor (*P*<.05). This was consistent with Lennartz et al [4], who reported that patients with severe disease had a negative attitude toward the use of AI in diagnosis. However, Jutzi et al [8] observed that patients with melanomas were more likely to accept the AI diagnosis than healthy people.

Although the relationship between prior knowledge of AI and patients' attitudes toward the use of AI in radiology is the most influential determinant of its acceptance in radiological diagnosis [4], such a relationship was not statistically significant (P>.05) in our study. This could be a result of our participants' age and level of education, which made them less interested in AI and learning about it [19].

Divorced or widowed participants had a higher level of distrust of AI use in radiology and expressed a greater need for active engagement when compared with married participants. This could be due to the minimal health assistance they receive from family members in their household. Without the support of a partner, they carry a heavier burden of disease and could be unaware of the minor health-related details that are usually picked up by a partner [21]. Divorced patients are more prone to illness anxiety disorders, including hypochondriasis [22,23].

According to the World Health Organization, efficient implementation of AI requires a good interpretation of the patient's attitudes toward the use of AI in medicine in order to build their trust. One of the objectives of this study is to understand patients' perspectives on these technologies to ensure their widespread implementation. This research may aid in adding insight into future integration policies and ensuring the suitability of AI programs to meet societal needs. The strength of this study is that we recruited participants from a large specialized referral hospital in Riyadh. Thus, the population of this study had different health statuses and needed different scans. We also considered sociodemographic differences and social determinants of health.

Limitations

This study is subject to certain limitations, notably selection bias introduced by the convenience sampling method and potential variability in participants' comprehension of the topic. The selection of waiting rooms for our convenience sampling approach might be influenced by several factors, including the frequency of x-ray and magnetic resonance imaging examinations performed in each facility, weighting time as well as the availability of waiting areas conducive to survey administration. While x-ray examinations are indeed more commonly performed procedures, our higher percentage of participants waiting for magnetic resonance imaging scans (or longer waiting time) may be attributed to the specific scheduling patterns and patient volumes observed during the data collection period. Additionally, variations in appointment scheduling and patient flow within different departments or clinics may have influenced the distribution of participants across waiting areas. Future research in this field should include a multicenter study population and studies that examine the predictors of distrust among patients in different hospitals, as well as identify useful methods for addressing the lack of knowledge and misconceptions that few patients hold with regard to AI.

Conclusions

In conclusion, patients were keen to understand the work of AI in radiology but favored personal interaction with a radiologist. If an AI system was implemented, patients would prefer full-body scans and full disclosure of medical findings. Patients were impartial toward AI replacing radiologists and the efficiency of AI. This preference expressed by patients for AI could have implications for clinicians and policymakers. Clinicians may consider incorporating these preferences into the design and implementation of AI systems in radiology, ensuring that the technology aligns with patient preferences for imaging procedures and information disclosure. Policymakers, on the other hand, may use this feedback to inform regulations and guidelines surrounding the use of AI in health care, emphasizing patient-centric approaches and ethical considerations in the integration of AI technologies. Therefore, our findings provide insight for future integration policies and help adapt AI to societal needs.

Acknowledgments

The authors convey special thanks for the support of the College of Medicine Research Center, Deanship of Scientific Research, King Saud University Riyadh, Saudi Arabia. The authors express our gratitude to Yfke P Ongena, Marieke Haan, Derya Yakar, and Thomas C Kwee for their contributions in providing the standardized questionnaire used in this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Questionnaire and consent - English.

[PDF File (Adobe PDF File), 90 KB - humanfactors_v11i1e53108_app1.pdf]

Multimedia Appendix 2

Supplementary results' tables.

[PDF File (Adobe PDF File), 90 KB - humanfactors_v11i1e53108_app2.pdf]

References

1. Panch T, Szolovits P, Atun R. Artificial intelligence, machine learning and health systems. J Glob Health 2018;8(2):020303 [FREE Full text] [doi: 10.7189/jogh.08.020303] [Medline: 30405904]
2. López-Robles JR, Otegi-Olaso JR, Porto Gómez I, Cobo MJ. 30 years of intelligence models in management and business: a bibliometric review. Int J Inf Manage 2019;48(C):22-38 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2019.01.013]
3. Alelyani M, Alamri S, Alqahtani MS, Musa A, Almater H, Alqahtani N, et al. Radiology community attitude in Saudi Arabia about the applications of artificial intelligence in radiology. Healthcare (Basel) 2021;9(7):834 [FREE Full text] [doi: 10.3390/healthcare9070834] [Medline: 34356212]
4. Lennartz S, Dratsch T, Zopfs D, Persigehl T, Maintz D, Große Hokamp N, et al. Use and control of artificial intelligence in patients across the medical workflow: single-center questionnaire study of patient perspectives. J Med Internet Res 2021;23(2):e24221 [FREE Full text] [doi: 10.2196/24221] [Medline: 33595451]
5. Palmisciano P, Jamjoom AB, Taylor D, Stoyanov D, Marcus HJ. Attitudes of patients and their relatives toward artificial intelligence in neurosurgery. World Neurosurg 2020;138:e627-e633. [doi: 10.1016/j.wneu.2020.03.029] [Medline: 32179185]
6. Martini K, Blüthgen C, Eberhard M, Schönenberger ALN, De Martini I, Huber FA, et al. Impact of vessel suppressed-CT on diagnostic accuracy in detection of pulmonary metastasis and reading time. Acad Radiol 2021;28(7):988-994. [doi: 10.1016/j.acra.2020.01.014] [Medline: 32037256]
7. Esmaeilzadeh P, Mirzaei T, Dharanikota S. Patients' perceptions toward human-artificial intelligence interaction in health care: experimental study. J Med Internet Res 2021;23(11):e25856 [FREE Full text] [doi: 10.2196/25856] [Medline: 34842535]
8. Jutzi TB, Krieghoff-Henning EI, Holland-Letz T, Utikal JS, Hauschild A, Schadendorf D, et al. Artificial intelligence in skin cancer diagnostics: the patients' perspective. Front Med (Lausanne) 2020;7:233 [FREE Full text] [doi: 10.3389/fmed.2020.00233] [Medline: 32671078]
9. Alsharif W, Qurashi A. Effectiveness of COVID-19 diagnosis and management tools: a review. Radiography (Lond) 2021;27(2):682-687 [FREE Full text] [doi: 10.1016/j.radi.2020.09.010] [Medline: 33008761]
10. Antes AL, Burrous S, Sisk BA, Schuelke MJ, Keune JD, DuBois JM. Exploring perceptions of healthcare technologies enabled by artificial intelligence: an online, scenario-based survey. BMC Med Inform Decis Mak 2021;21(1):221 [FREE Full text] [doi: 10.1186/s12911-021-01586-8] [Medline: 34284756]
11. Ongena YP, Yakar D, Haan M, Kwee TC. Artificial intelligence in screening mammography: a population survey of women's preferences. J Am Coll Radiol 2021;18(1 Pt A):79-86 [FREE Full text] [doi: 10.1016/j.jacr.2020.09.042] [Medline: 33058789]
12. Lennox-Chhugani N, Chen Y, Pearson V, Trzcinski B, James J. Women's attitudes to the use of AI image readers: a case study from a national breast screening programme. BMJ Health Care Inform 2021;28(1):e100293 [FREE Full text] [doi: 10.1136/bmjhci-2020-100293] [Medline: 33795236]
13. Pesapane F, Codari M, Sardanelli F. Artificial intelligence in medical imaging: threat or opportunity? radiologists again at the forefront of innovation in medicine. Eur Radiol Exp 2018;2(1):35 [FREE Full text] [doi: 10.1186/s41747-018-0061-6] [Medline: 30353365]
14. Yakar D, Ongena YP, Kwee TC, Haan M. Do people favor artificial intelligence over physicians? a survey among the general population and their view on artificial intelligence in medicine. Value Health 2022;25(3):374-381 [FREE Full text] [doi: 10.1016/j.jval.2021.09.004] [Medline: 35227448]
15. Richardson JP, Smith C, Curtis S, Watson S, Zhu X, Barry B, et al. Patient apprehensions about the use of artificial intelligence in healthcare. NPJ Digit Med 2021;4(1):140 [FREE Full text] [doi: 10.1038/s41746-021-00509-1] [Medline: 34548621]
16. Ethics and governance of artificial intelligence for health: WHO guidance. World Health Organization. Geneva: World Health Organization; 2021. URL: https://www.who.int/publications/i/item/9789240029200 [accessed 2021-06-28]
17. Ongena YP, Haan M, Yakar D, Kwee TC. Patients' views on the implementation of artificial intelligence in radiology: development and validation of a standardized questionnaire. Eur Radiol 2020;30(2):1033-1040 [FREE Full text] [doi: 10.1007/s00330-019-06486-0] [Medline: 31705254]
18. Young AT, Amara D, Bhattacharya A, Wei ML. Patient and general public attitudes towards clinical artificial intelligence: a mixed methods systematic review. Lancet Digit Health 2021;3(9):e599-e611 [FREE Full text] [doi: 10.1016/S2589-7500(21)00132-1] [Medline: 34446266]
19. Kebede AS, Ozolins LL, Holst H, Galvin K. Digital engagement of older adults: scoping review. J Med Internet Res 2022;24(12):e40192 [FREE Full text] [doi: 10.2196/40192] [Medline: 36477006]
20. O'Shaughnessy MR, Schiff DS, Varshney LR, Rozell CJ, Davenport MA. What governs attitudes toward artificial intelligence adoption and governance? Sci Public Policy 2022;50(2):161-176 [FREE Full text] [doi: 10.1093/scipol/scac056]
21. Cajita MI, Hodgson NA, Lam KW, Yoo S, Han HR. Facilitators of and barriers to mHealth adoption in older adults with heart failure. Comput Inform Nurs 2018;36(8):376-382 [FREE Full text] [doi: 10.1097/CIN.0000000000000442] [Medline: 29742549]
22. Williams K. The transition to widowhood and the social regulation of health: consequences for health and health risk behavior. J Gerontol B Psychol Sci Soc Sci 2004;59(6):S343-S349. [doi: 10.1093/geronb/59.6.s343] [Medline: 15576865]

23. Goel N, Terman M, Terman JS. Depressive symptomatology differentiates subgroups of patients with seasonal affective disorder. Depress Anxiety 2002;15(1):34-41. [doi: 10.1002/da.1083] [Medline: 11816051]

Abbreviations

AI: artificial intelligence
e-SIHI: electronic system for integrated health information
KKUH: King Khalid University Hospital

Edited by A Kushniruk; submitted 26.09.23; peer-reviewed by R Toomey, S Papikyan, A Torchyan; comments to author 23.02.24; revised version received 15.03.24; accepted 22.06.24; published 07.08.24.

Please cite as:
Baghdadi LR, Mobeirek AA, Alhudaithi DR, Albenmousa FA, Alhadlaq LS, Alaql MS, Alhamlan SA
Patients' Attitudes Toward the Use of Artificial Intelligence as a Diagnostic Tool in Radiology in Saudi Arabia: Cross-Sectional Study
JMIR Hum Factors 2024;11:e53108
URL: https://humanfactors.jmir.org/2024/1/e53108
doi:10.2196/53108
PMID:

Original Paper

Mobile App for Patients With Chronic Obstructive Pulmonary Diseases During Home-Based Exercise Care: Usability Study

Shih-Ying Chien[1,2,3], PhD

[1]Department of Industrial Design, Chang Gung University, Taoyuan, Taiwan

[2]Physical Medicine & Rehabilitation, Chang Gung Memorial Hospital, Taoyuan, Taiwan

[3]Department of Public Health & Medical Humanities, School of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan

Corresponding Author:
Shih-Ying Chien, PhD
Department of Industrial Design
Chang Gung University
No.259, Wen-Hwa 1st Rd, Kweishan, Taoyuan, Taiwan, R.O.C.
Taoyuan, 333
Taiwan
Phone: 886 3 2118800 ext 5669
Email: shihying@mail.cgu.edu.tw

Abstract

Background: Digital health tools have demonstrated promise in the treatment and self-management of chronic diseases while also serving as an important means for reducing the workload of health care professionals (HCPs) and enhancing the quality of care. However, these tools often merely undergo large-scale testing or enter the market without undergoing rigorous user experience analysis in the early stages of their development, leading to frequent instances of low use or failure.

Objective: This study aims to assess the usability of and satisfaction with a mobile app designed for the clinical monitoring of patients with chronic obstructive pulmonary disease undergoing pulmonary rehabilitation at home.

Methods: This study used a mixed methods approach involving two key stakeholders—patients with chronic obstructive pulmonary disease and HCPs—across three phases: (1) mobile app mock-up design, (2) usability testing, and (3) satisfaction evaluation. Using convenience sampling, participants were grouped as HCPs (n=12) and patients (n=18). Each received a tablet with mock-ups for usability testing through interviews, with audio recordings transcribed and analyzed anonymously in NVivo12.0, focusing on mock-up features and usability insights. Task difficulty was rated from 1 (very easy) to 5 (very difficult), with noncompletion deemed a critical error. Usability satisfaction was measured on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree).

Results: The research indicated a notable difference in app usability perceptions: 66% (8/12) of HCPs found tasks “very easy,” compared to only 22% (4/18) of patients. Despite this, no participant made critical errors or withdrew, and satisfaction was high. HCPs completed tasks in about 20 minutes, while patients took 30. Older adults faced challenges with touch screens and scroll menus, suggesting the need for intuitive design aids like auditory support and visual health progress indicators, such as graphs. HCPs noted potential data delays affecting service, while non–native-speaking caregivers faced interpretation challenges. A secure pairing system for privacy in teleconsultations proved difficult for older users; a simpler icon-based system is recommended. This study highlights the need to consider stakeholder abilities in medical app design to enhance function implementation.

Conclusions: Most HCPs (11/12, 91%) found the app intuitive, though they recommended adding icons to show patient progress to support clinical decisions. In contrast, 62% (11/18) of patients struggled with tablet navigation, especially with connectivity features. To ensure equitable access, the design should accommodate older users with diverse abilities. Despite challenges, both groups reported high satisfaction, with patients expressing a willingness to learn and recommending the app. These positive usability evaluations suggest that, with design improvements, such apps could see increased use in home-based care.

(JMIR Hum Factors 2024;11:e60049) doi:10.2196/60049

KEYWORDS

digital health; chronic obstructive pulmonary disease; COPD; usability; telerehabilitation; mobile health app

Introduction

Background

Chronic pulmonary diseases severely impair the respiratory system and impact the daily activities of those affected [1-4]. With approximately 3 million deaths annually, chronic pulmonary diseases rank among the top 3 global causes of death [5,6]. Chronic obstructive pulmonary disease (COPD) is a major subset of chronic pulmonary diseases, with nearly 36% of patients developing comorbidities, such as hypertension and cardiovascular diseases [7]. Globally, 1 person succumbs to COPD every 10 seconds; in Taiwan, over 5000 deaths are attributed to COPD annually [8]. Severe cases of COPD may lead to systemic manifestations, emphasizing the urgency of interventions [9,10]. Patients, particularly older adults, often struggle with dyspnea and thus experience a diminished quality of life [11]. Depressive symptoms also commonly arise along with the aforementioned problems [12,13]. The significant impact of chronic pulmonary diseases on respiratory function and overall well-being highlights the urgency of addressing this issue. COPD presents with 3 major symptoms: cough, sputum production, and wheezing and is therefore, often mistaken as the common cold [14,15]. Coupled with low awareness of obstructive lung diseases among the public, individuals frequently underestimate their condition, leading to a delayed diagnosis and treatment as well as a significantly elevated risk of health deterioration [16,17].

Exercise is recognized as one of the most effective therapies for chronic pulmonary diseases [18]. It serves to alleviate respiratory symptoms, enhance cardiorespiratory function, improve quality of life, and consequently increase the overall well-being of patients [19,20]. Statistics indicate that individuals with COPD face a reduced life expectancy of 6 to 10 years [21]. In Taiwan, the 1-year mortality rate after the first hospitalization for obstructive lung diseases is as high as 20% [22]. To prevent recurrent hospitalizations due to acute deterioration in pulmonary function, postdischarge priorities for patients with chronic pulmonary diseases include maintaining regular exercise habits and receiving precise exercise prescriptions tailored to their conditions [23,24]. This approach aims to train and enhance pulmonary capacity, thereby extending life expectancy [20,25]. To ensure that home-based pulmonary rehabilitation exercises meet clinical requirements, digital health (eHealth) tools are considered potential aids in the treatment of chronic diseases because they assist patients in self-managing their conditions [26,27]. Simultaneously, these tools can provide real-time assistance to patients, caregivers, and health care professionals (HCPs) in achieving the vision of individual disease management and monitoring [28,29]. Moreover, eHealth tools serve as the optimal instruments for HCPs to provide real-time guidance to patients at home in developing health care skills and managing chronic diseases, especially during circumstances when in-person treatments are challenging (such as during the COVID-19 pandemic) [30]. Research confirms that the application of eHealth tools not only effectively reduces incidence rates, disease exacerbation, and recurrent hospitalizations but has also proven to be an efficient means of alleviating the clinical workload of HCPs [31,32].

Chronic diseases are a primary driver of the global increase in health care and caregiving expenditures [32,33]. In this context, eHealth tools could benefit various stakeholders (including HCPs, patients, and caregivers) by offering a health service [34,35]. However, these tools often face challenges in achieving seamless operations among the stakeholders involved, leading to outcomes below expectations [36]. Consequently, patients after their discharge exhibit low adherence to using these tools, hindering the comprehensive realization of their intended benefits [37,38]. Therefore, the design of eHealth tools should prioritize a methodology centered on user experience [32,39,40].

Objectives

For the aforementioned reasons, this study conducted a user experience–based evaluation of the usability of a home-based pulmonary rehabilitation app. The assessment focused on operational proficiency, information comprehension, interface design, and system acceptance among discharged patients and HCPs. User feedback and results were collected, and based on the survey findings, posttrial design modifications were implemented to enhance the effectiveness and user adherence to the eHealth tool. The objective was to ascertain the usability of and satisfaction with the mobile app in tracking home-based pulmonary rehabilitation exercise therapy for users.

Methods

Overview

This study was a nonpharmacological clinical trial that used a questionnaire-based interview approach to evaluate the operational usability of a mobile app designed for remote health care delivery. The evaluation involved 2 key stakeholder groups: the HCPs and the patients. The insights gained from this assessment will be used as a reference for refining the design of eHealth tools, ensuring their applicability for both clinical and home-based remote health care delivery.

Ethical Considerations

This study received approval from the Institutional Review Board of the Chang Gung Medical Foundation (registration number: 202200070B0) prior to its implementation. Before the commencement of the experiment, informed consent was obtained from all participants, who signed the necessary consent forms. To protect participants' privacy, all personal information was de-identified and stored with encoded data for analysis. Throughout the data collection and processing, strict adherence to confidentiality and privacy protection principles was maintained. Furthermore, all participants voluntarily joined the study, and no monetary compensation or gifts were offered to ensure the objectivity and impartiality of the experiment.

App Development and Usability Testing

Before initiating this study, the researchers conducted a survey on the construction requirements of a home-based pulmonary rehabilitation system. The aim was to establish the fundamental elements of system design based on user experience and needs. Building upon the outcomes of the preliminary research, 2 apps were developed: one for HCPs to clinically monitor exercise activities and another for patients or caregivers to execute the prescribed exercises at home. This study concentrated on assessing the usability of these apps for the 2 participant groups. The survey results of this study will contribute to refining the design of the home-based pulmonary rehabilitation app. The research investigation comprised the following two stages: (1) observing and recording user interactions to understand the ease of use regarding system functionalities and (2) conducting usability tests through app operations to elucidate user satisfaction with the app design. This study was conducted at the Chang Gung Hospital in Taiwan with convenient sampling. The HCP group consisted of respiratory care and pulmonary care professionals, including physicians, therapists, and nurses (n=12), who (1) were aged ≥25 years with at least 1 year of experience in respiratory or pulmonary disease care, (2) had provided care for patients with chronic pulmonary diseases in the past 3 years, and (3) had experience in tracking patient rehabilitation. The patient group comprised patients with chronic pulmonary diseases (n=18) who (1) were aged ≥58 years, (2) had been diagnosed with chronic pulmonary diseases for ≥3 years with a history of hospitalization and had experienced pulmonary rehabilitation, (3) willingly participated and signed an informed consent form, and (4) possessed the ability to express themselves independently.

Considering the participants were older adults, the questionnaires were administered and filled out by the trial executor. Figure 1 illustrates the main functionalities of the app and the information distribution on each screen. Usability tests were conducted on the 10 key functional pages of the app, representing the most crucial functions of the system. The details of these tasks are outlined in Textbox 1. Throughout the testing process, when participants encountered notable operational challenges, the researcher proactively inquired about their difficulties or potential consideration for discontinuation. If participants successfully accomplished a task, it was documented as “operation success” on the questionnaire. Conversely, if participants failed to complete a task in 3 consecutive attempts without adhering to instructions, the researcher categorized it as “operation failure.” Irrespective of the participants’ success, posttask feedback was collected, and a Likert scale was used to evaluate task difficulty, using the following ratings: 1=very easy, 2=easy, 3=moderately easy, 4=difficult, and 5=very difficult. To evaluate app usability satisfaction, a survey was developed by adapting and modifying validated questionnaires from previous studies to align with the objectives of this research. Regarding the satisfaction of app usability, the survey comprised 10 questions each for the HCPs and the patients (Textbox 2). Responses were assessed on a 4-point Likert scale, ranging from 1 (totally disagree) to 4 (totally agree).

Figure 1. Usability test mock-ups. (A) App’s functional interface for health care professionals. Professionals can prescribe exercise regimens based on the patient’s condition, adjusting them dynamically by referencing clinical data and psychophysical states. This aids in facilitating the self-management of postdischarge patients at home. (B) App’s functional interface for patients. The interface for the home pulmonary rehabilitation app guides patients through lung exercises at home using visualizations and encompassing features, such as receiving exercise prescriptions, activity tracking, and progress monitoring.

Textbox 1. Usability test items.

Usability test items for assessing app functionality for health care professionals

1. Create an account
2. Log in
3. Patient information management including entering or adding the patient’s name, age, sex, contact number, and a brief description of the patient’s physical condition
4. Prescription management: set patient prescriptions and transmit them to the patient’s end
5. Record review and analysis: view the patient’s exercise history records, including visualizing graphs showing trends in clinical signs
6. Clinical monitoring: monitor patient cycle variations, including whether the heart rate falls within the “safe heart rate” range
7. Remote monitoring: establish a connection and pair it with the patient’s end for we-based monitoring of the patient’s exercise status
8. Consultation and reporting: view exercise outcome reports and conduct web-based consultations
9. Prescription adjustment: online adjustment of exercise prescriptions, including intensity and difficulty.
10. Record management: view, edit, or delete member records

Usability test items for assessing app functionality for patients

1. Log in
2. Select personal information to view individual details
3. Generate a pairing code to connect with the hospital’s end
4. Retrieve exercise prescription
5. Review exercise prescription details
6. View personal exercise history records (including visualizing graphs showing trends in clinical signs)
7. Pause the exercise (can be paused at any time if discomfort is experienced)
8. Examine exercise cycle variations, including checking safe heart rate and heart rate variability
9. Receive remote connection requests for web-based consultations
10. Complete web-based questionnaires (eg, Borg Rating of Perceived Exertion scale)

Textbox 2. Usability satisfaction assessment of the app.

Response items for health care professionals

1. Overall, this app is user-friendly.
2. The app interface is well-designed and aligns with clinical information needs.
3. On the basis of patient-generated data, the graphs are easy to interpret.
4. The app substantially assists clinical care professionals.
5. The app facilitates monitoring the home-based exercise rehabilitation of postdischarge patients.
6. Would you recommend colleagues to use a similar app?
7. The operations of the app are straightforward and easy to remember.
8. The actions performed, whether on the web or offline, are straightforward to me.
9. The app design is comprehensive with no missing or incorrect information.
10. I believe that it is safe for patients to use at home.

Questions for patients

1. Overall, this app is user-friendly.
2. The app has a simple and easy-to-understand interface.
3. The app facilitates easy recall of exercise prescription information at home.
4. The app comprehensively records the entire exercise process, ensuring no loss of vital information.
5. When encountering issues, I can easily resolve them.
6. Virtual reality contributes to my increased focus and enjoyment during rehabilitation.
7. I intend to continue using this app.
8. Even without assistance, I can operate the app on my own.
9. I feel safe using this app at home.
10. I find some features a bit complex.

In addition, to ensure the accuracy of the participants' feedback, the entire experimental process was recorded with meticulous time control to ensure that each participant completed the interview within the designated time frame (50-60 minutes). Participants had the opportunity to express their perspectives, opinions, and experiences regarding the tasks during the experiment (such as the difficulties or simplicity in the task). Subsequently, this information underwent verbatim transcription analysis with relevant usability keywords marked. Detailed information on participants' sociodemographic and clinical characteristics can be found in subsequent sections. Furthermore, the quantitative data of this study were analyzed using the statistical software SPSS (version 22.0; IBM Corp).

Results

Overview

This study collected data from two groups: HCPs (n=12) and patients (n=18). The majority (8/12, 66%) of HCPs were respiratory therapists with an average age of 46 (SD 5) years. They had >1 year of experience in respiratory care within the past 3 years. The majority (16/18, 89%) of patients with chronic lung disease were male, with an average age of 66 (SD 5) years. They had a history of chronic lung disease for >3 years and had undergone pulmonary rehabilitation for ≥3 years. Detailed information regarding the sociodemographic and clinical backgrounds of these two participant groups can be found in Table 1.

Table 1. Sociodemographic characteristics of participants.

Participants and characteristics	Values
HCPs[a] (n=12)	
Age (y), mean (SD)	46 (5)
Sex, n (%)	
Male	3 (25)
Female	9 (75)
Job title, n (%)	
Respiratory therapist	8 (67)
Thoracic surgeon	1 (8)
Physiotherapist	1 (8)
Pulmonary rehabilitation specialist	2 (17)
Experience in caring for chronic respiratory diseases (y), n (%)	
1-3	3 (25)
>3	9 (75)
Patients (n=18)	
Age (y), mean (SD)	66 (5)
Gender, n (%)	
Male	16 (89)
Female	2 (11)
Education level, n (%)	
Elementary	10 (56)
High school	3 (17)
Bachelor	3 (17)
Bachelor's degree or higher	2 (11)
Duration of illness (y), n (%)	
1-3	5 (28)
>3	13 (72)

[a]HCP: health care professional.

User Operations and Usability Perception Interview Survey

Following user interaction with the system, we conducted one-on-one semistructured interviews to gather insights into operational experiences. The participants described any difficulties encountered as well as their feelings while executing the tasks. The entire interview process was recorded, transcribed verbatim, and then subjected to qualitative analysis using NVivo (version 12.0; Lumivero) for content analysis and synthesis.

Among the HCPs, 66% (8/12) acknowledged the need for a brief transitional period to familiarize themselves with the app interface and functionalities. During the transformation of data into graphical representations, more diverse visualizations were preferred. Specifically, 58% (7/12) expressed a preference for observing changes in heart rhythm and having graphical representations illustrating cyclic patterns to aid in explaining the progression of pulmonary function in patients.

Moreover, 84% (10/12) of the HCPs emphasized from the hospital's standpoint, the concern regarding inadequate self-health management in patients after discharge, especially considering that these patients belonged to a high-risk demographic and reported that prioritizing safety should be a top consideration. Therefore, it is recommended that the fundamental design of the app incorporate an emergency cessation mechanism or a real-time notification feature that activates upon detecting physiological abnormalities, aligning with the fundamental requisites for medical applications.

Within the patient participant group, 83% (15/18) of respondents reported a sense of complexity during their initial exposure to the app. Among them, 62% (11/18) indicated an inability to operate the app independently; this challenge was attributed to the age of the majority of participants (mean 66, SD 5 years) and their inherent skepticism regarding their operational capabilities. Despite the user interface of the patient-side app being menu-based and devoid of text input requirements, using tablets and apps as interactive media was still perceived as challenging. Moreover, 45% (8/18) expressed difficulty in discerning or comprehending the information presented on the screen, contributing to feelings of unease and anxiety; 62% (11/18) believed that understanding the operational procedures of the app without guidance was challenging. Notably, the majority (12/18, 67%) of patients specifically highlighted the difficulties encountered during the initial step of entering email and password information to log in to the system. Lastly, 39% (7/18) conveyed an inability to comprehend the numerical representations in the postexercise feedback reports, expressing curiosity or confusion regarding the meaning of the graphs. Table 2 presents detailed feedback from participants.

Table 2. Operational challenges: findings from the semistructured interview.

Participant and participant's feedback	Pain points
HCPs[a]	
• "Um... Honestly, at first, without any instructions, I was a bit unsure where to start, but luckily, I quickly found the "+" sign to create an account." [HCP 6]	• The "+" symbol for account creation might not be very clear, but fortunately, it was quickly resolved.
• "I found it easy to fill in basic patient information, but prescribing medication is more challenging for me since I'm not a doctor." [HCP 1]	• Different backgrounds may entail different responsibilities.
• "I'm not sure if it's just me being unfamiliar with the system, but currently, while the screen displays patient heart rate and related data well, for clinical staff, besides linear charts illustrating historical backgrounds, it would be helpful to have icons indicating categories or different charts displaying changes in various physiological values for clarity." [HCP 11]	• Clinical staff require more varied graphical representations or symbols to express the significance of diverse clinical data.
• "I'm not sure if it's a network issue or a system problem, but I feel the screen updates a bit slowly." [HCP 4] • "Perhaps there's some network delay; when I modify prescriptions, the patient's screen doesn't always match what I intend to prescribe." [HCP 12]	• It is essential to ensure data conversion speed and the ability to promptly provide accurate information.
• "Hmm... Typically, it takes some time to accumulate clinical data to see the effectiveness. At this stage, although the feature exists, there might not be enough data to discern changes." [HCP 1]	• A vast amount of data are required for clinical efficacy to be evident in the system.
• "I'm afraid I might accidentally delete a patient's record." [HCP 5]	• There is a lack of mechanism for recovery or error compensation.
Patients	
• "I can't read, and I can't see what's written on the screen." [Ps 8 and Ps 17]	• Patients who are older adults have lower levels of education or have poor vision and may not be comfortable with operating the app alone.
• "I can use a tablet, but I don't know how to type." [Ps 17]	• In addition to typing, other input functions and voice commands need to be added.
• "I don't really understand what these numbers mean...And what is pairing connection...Do I just press it, or do I need to input something?" [Ps 5, Ps 9, and Ps 12]	• The connection mechanism poses a challenge for both the older adults and caregivers.
• "I often accidentally press the pause button... I don't know how to get back to the exercise screen...It makes me very anxious." [Ps 13]	• There are too many function keys on the screen, making it difficult for users to navigate.
• "I see my heartbeat and heart rate...I don't quite understand them, so it would be better if there were colors or lines to remind me when to slow down." [Ps 10]	• The interface should be more user-friendly to avoid excessive use of numbers and scientific charts and should make good use of patterns, icons, or voice commands.
• "I received a message asking to connect and to input numbers. I find this very difficult." [Ps 8]	• Text may be difficult to understand; replacing it with diagrams or call-in features might be easier to comprehend.
• "This thing is too advanced; I can't figure it out." [Ps 7]	• The digital divide is a significant barrier for many rural and older adult populations.

[a]HCP: health care professional.

Evaluation of User Operational Difficulty

During the assessment of the usability of the app's function, HCPs encountered minimal challenges in tasks, such as setting up user and patient accounts as well as modifying patient information using tablets. Specifically, 8 (67%) of the 12 participants rated these tasks as "very easy." More than half (7/12, 58%) of the HCPs found it easy to set and transmit prescriptions to the patient's end. In addition, the majority (9/12, 75%) found it relatively straightforward to access patient exercise histories and visualize health status charts, although some recommended potential enhancements.

Concerning the web-based monitoring of changes in patient physiological readings, the majority (9/12, 75%) of HCPs perceived the interface to be clear and easy to understand, with operations being very straightforward. However, when establishing connections, most (6/12, 50%) participants initially found it somewhat difficult. Nevertheless, after becoming familiar with the process, they regarded it as relatively simple (8/12, 66%).

While tasks like adjusting web-based prescriptions and conducting remote consultations were found easy by most (9/12, 75%) HCPs, 7 (58%) of the 12 HCPs expressed that accessing patient information and engaging in web-based consultations were comparatively complex, requiring more time for comprehension.

Managing patient information was deemed straightforward by 10 (83%) of the 12 HCPs. They noted that the clear interface facilitated easy access to and deletion of information. Screen delays were reported by only 42% (5/12) of participants, classifying this issue as "moderate" (Table 3). Importantly, no task was rated as "difficult" or "too difficult."

Table 3. Perceived difficulty of participant operations.

Participant and task	Very easy, n (%)	Easy, n (%)	Moderately difficult, n (%)	Difficult, n (%)	Very difficult, n (%)
HCPs[a] (N=12)					
T1: Create an account	8 (67)	3 (25)	1 (8)	0 (0)	0 (0)
T2: Log in	8 (67)	3 (25)	1 (8)	0 (0)	0 (0)
T3: Patient information management	8 (67)	4 (33)	0 (0)	0 (0)	0 (0)
T4: Prescription management	4 (33)	7 (58)	1 (8)	0 (0)	0 (0)
T5: Record review and analysis	3 (25)	9 (75)	0 (0)	0 (0)	0 (0)
T6: Clinical monitoring	2 (17)	9 (75)	1 (8)	0 (0)	0 (0)
T7: Remote monitoring	2 (17)	8 (67)	2 (17)	0 (0)	0 (0)
T8: Consultation and reporting	3 (25)	9 (75)	0 (0)	0 (0)	0 (0)
T9: Prescription adjustment	3 (25)	4 (33)	5 (42)	0 (0)	0 (0)
T10: Record management	2 (17)	10 (83)	0 (0)	0 (0)	0 (0)
Patients (N=18)					
T1: Log in	4 (22)	8 (44)	1 (6)	5 (28)	0 (0)
T2: Viewing personal information	5 (28)	6 (33)	3 (17)	4 (22)	0 (0)
T3: Pairing for connection	4 (22)	3 (17)	5 (28)	6 (33)	0 (0)
T4: Retrieving exercise prescription	6 (33)	3 (17)	4 (22)	5 (28)	0 (0)
T5: Reviewing prescription details	6 (33)	11 (61)	1 (6)	0 (0)	0 (0)
T6: Viewing past exercise history	6 (33)	11 (61)	1 (6)	0 (0)	0 (0)
T7: Pausing exercise	12 (67)	6 (33)	0 (0)	0 (0)	0 (0)
T8: Examining cycle variations	5 (28)	7 (39)	4 (22)	2 (11)	0 (0)
T9: Requesting web-based consultations	1 (6)	2 (11)	7 (39)	8 (44)	0 (0)
T10: Completing questionnaires on the web	3 (17)	15 (83)	0 (0)	0 (0)	0 (0)

In contrast, the patient group experienced substantial challenges during system log-in. Most (6/18,34 %) patients initially struggled to understand how to use the tablet. After reminders and demonstrations, 45% (8/18) of the patients eventually considered the log-in process as "easy," while 28% (5/18) found it challenging. When viewing personal information, 34% (6/18) individuals expressed that it was relatively simple, despite requiring some time for searching and consideration.

Regarding receiving pairing codes and connecting to the hospital end, most (7/18, 39%) patients could input the pairing code to establish a connection; however, 34% (6/18) still encountered difficulties due to a lack of familiarity and the absence of assistance from caregivers. Despite some (9/18,50 %) patients expressing confusion regarding the functionality on the screen to receive prescriptions from HCPs, 33% (6/18) and 17% (3/18) patients found it “very easy” and “easy,” respectively.

In addition, concerning viewing exercise prescriptions and accessing personal exercise history records (including visualized charts displaying clinical symptom trends), the majority (11/18, 61%) of patients considered these tasks relatively simple. When experiencing discomfort, 67% (12/18) knew which button to press to pause and considered this step as “very easy.”

Regarding viewing personal physiological information, 7 (39%) of the 18 patients found it relatively easy; however, 8 (44%) patients expressed uncertainty about how to initiate web-based consultations. Nonetheless, although participants within the patient group encountered operational difficulties, none described the tasks as “too difficult” in the questionnaire interviews (Table 3).

Usability Testing of System Task Accomplishment

In usability testing with HCPs, system tasks, such as setup, log-in, exercise prescription, and remote connectivity were assessed. None of the tasks were completely successful on the first attempt. However, log-in and clinical monitoring tasks showed a higher success rate, reaching 83% (10/12; Tables 4 and 5). In the testing process, each participant was given 5 opportunities for operation. The majority (8/12, 66%) of HCPs required some time to adapt to and familiarize themselves with the app features. About 42% (5/12) committed 1 to 2 errors during operations, while 25% (3/12) made 3 or more errors. Nevertheless, 58% (7/12) ultimately succeeded in completing the tasks within the specified time limit. The HCPs attributed these errors primarily to the unfamiliarity with the interface and the occasional accidental presses. However, they noted that the design of the system was not overly complex, and the inclusion of a “back” mechanism allowed for quick error correction.

Table 4. Success rates of participant actions (N=122).

Task	Success, n (%)	Failure, n (%)
T1: Create an account	9 (75)	3 (25)
T2: Log in	10 (83)	2 (17)
T3: Patient information management	8 (66)	4 (33)
T4: Prescription management	8 (66)	4 (33)
T5: Record review and analysis	9 (75)	3 (25)
T6: Clinical monitoring	10 (83)	2 (17)
T7: Remote monitoring	8 (66)	4 (33)
T8: Consultation and reporting	7 (58)	5 (42)
T9: Prescription adjustment	8 (66)	4 (33)
T10: Record management	9 (75)	3 (25)

Table 5. Success rates of participant (patient) actions (n=18).

Task	Success, n (%)	Failure, n (%)
T1: Logging in	8 (44)	10 (56)
T2: Viewing personal information	9 (50)	9 (50)
T3: Pairing for connection	6 (33)	12 (67)
T4: Retrieving exercise prescription	9 (50)	9 (50)
T5: Reviewing prescription details	11 (61)	7 (39)
T6: Viewing past exercise history	13 (72)	5 (28)
T7: Pausing exercise	16 (89)	2 (11)
T8: Examining cycle variations	10 (56)	8 (44)
T9: Requesting web-based consultations	4 (22)	14 (78)
T10: Completing questionnaires on the web	18 (100)	0 (0)

Conversely, the majority (12/18, 67%) of the patients were able to complete tasks with assistance. However, tasks such as “Pairing for connection” and “Request web-based consultations” remained challenging for many, with failure percentages of 67% (12/18) and 78% (14/18), respectively. Among these patients, 44% (8/18) made 1 to 2 mistakes, while 39% (7/18) made 3 or more errors on the test (Table 5).

During the study, the main reason for errors among the patients was the fact that they were older adults, which may have affected their ability to operate the tablet as we would have expected. Notably, the patients achieved a 100% success rate in Task 10, “Completing questionnaires on the web,” primarily because this functionality was operated by the HCPs at this stage, thereby encountering fewer issues in operation.

Moreover, the experiment was conducted in a medical setting; as a result, none of the participants in either group made critical mistakes.

Mobile App Satisfaction Survey

Regarding the satisfaction survey, over half (11/12, 91%) of the HCPs expressed satisfaction with the functionality design of the app, finding it relatively satisfactory with minimal operational issues. A substantial proportion, constituting 75% (9/12), perceived the app as highly user-friendly, and 66% (8/12) expressed considerable satisfaction with the design of the interface. Moreover, the HCPs could promptly access patients' physiological information in real time through the app, with 58% (7/12) indicating such capability. However, satisfaction levels slightly declined when monitoring patients' physiological conditions through web-based connectivity. Only 33% (4/12) perceived this function as comprehensive, while 17% (2/12) expressed concerns regarding the inability of the current connection quality to facilitate real-time monitoring. Furthermore, 25% (3/12) suggested that a more stable network connection would enhance safety. Finally, the extraction of information and the generation of graphical representations were considered crucial by most (11/12,91%) HCPs. In this study, 42% (5/12) and 33% (4/12) HCPs expressed extreme satisfaction and satisfaction, respectively, with the design of this app. Respondents anticipated that these functionalities would aid HCPs in assessing patients' physical conditions and prescribing medical treatments.

In the patient group, satisfaction with operational aspects notably lagged behind that of the HCPs across functions, such as operation, information retrieval, connectivity, and message access. Only 22% (4/18) of the patients perceived the interface design of the app as user-friendly, with nearly 33% (6/18) unable to provide a proper evaluation. This is attributed to the significant operational challenges faced by the patients, with 50% (9/18) of the users unable to comprehend each interface function (neutral 5/18, 28%, disagree 2/18, 11%, and strongly disagree 2/18, 11%). Moreover, a high percentage (13/18, 72%) were unaware of how to connect remotely, and 39% (7/18; neutral 2/18, 11%, disagree 4/18, 22%, and strongly disagree 1/18, 6%) were unsure of how to access historical rehabilitation exercise records. In addition, 61% (11/18) were uncertain about accessing personal messages, and over a quarter (5/18, 28%) expressed confusion about the significance of real-time values. Nonetheless, 61% (11/18; strongly agree 6/18, 34% and agree 5/18, 28%) expressed willingness to recommend and continue using the app (Table 6).

Table 6. App satisfaction survey.

Participant and task	Strongly agree, n (%)	Agree, n (%)	Neutral, n (%)	Disagree, n (%)	Strongly disagree, n (%)
HCPs[a] (n=12)					
Easy to use	9 (75)	2 (17)	1 (8)	0 (0)	0 (0)
User-friendly interface	8 (67)	3 (25)	1 (8)	0 (0)	0 (0)
Well-designed patient information management	6 (50)	3 (25)	1 (8)	2 (17)	0 (0)
Effective real-time patient data access and analysis	7 (58)	4 (33)	1 (8)	0 (0)	0 (0)
Effective web-based health monitoring functionality	4 (33)	4 (33)	2 (17)	2 (17)	0 (0)
Clear medical information and graphics	5 (42)	4 (33)	1 (8)	2 (17)	0 (0)
Well-implemented connectivity features	4 (33)	4 (33)	1 (8)	3 (25)	0 (0)
Data aiding decision on rehabilitation prescription	3 (25)	4 (33)	3 (25)	2 (17)	0 (0)
Easy web-based prescription setup and adjustment	4 (33)	5 (42)	1 (8)	2 (17)	0 (0)
Convenient record retrieval	4 (33)	5 (42)	1 (8)	2 (17)	0 (0)
Patients (n=18)					
User-friendly interface	4 (22)	3 (17)	6 (33)	4 (22)	1 (6)
Easy access to personal information	5 (28)	3 (17)	4 (22)	4 (22)	2 (11)
Understanding of each feature on the interface	4 (22)	5 (28)	5 (28)	2 (11)	2 (11)
Easy-to-use connectivity features	2 (11)	3 (17)	5 (28)	4 (22)	4 (22)
Convenient access to historical records	7 (39)	4 (22)	2 (11)	4 (22)	1 (6)
Well-designed personal prescription collection and execution	5 (28)	4 (22)	4 (22)	4 (22)	1 (6)
Implementation of security mechanisms	6 (33)	5 (28)	2 (11)	4 (22)	1 (6)
Personal exercise variations incorporated	5 (28)	3 (17)	5 (28)	3 (17)	2 (11)
Complete personal message records	3 (17)	4 (22)	6 (33)	4 (22)	1 (6)
Willingness to recommend this app	6 (33)	5 (28)	4 (22)	3 (17)	0 (0)

[a]HCP: health care professional.

Discussion

Principal Findings

This study has revealed a dichotomy in app operation and satisfaction levels between the participating HCPs and patients. Both groups exhibited a demand for adapting to the app, with the majority (11/12, 91%) of the HCPs potentially acquiring app operation skills and understanding its functions through learning and adaptation. Conversely, some (9/18, 50%) participants in the patient group found it challenging to use tablets to assist them in rehabilitation programs. Overall, there exists a correlation between operational capability and satisfaction levels. Although both groups of participants provided suggestions for the app, they all recognized that the use of this app would contribute to personal health management and remote home health monitoring. In addition, the results of the usability tests and the suggestions provided will aid us in devising improvements to the design of the mobile app.

Considering Human Factors for Enhancing Health Care App Usability Design

Technological advancements offer various methods to enhance health care service quality. Despite the longstanding application of digital technology in medical facilities and home-based care, the practical implementation and user acceptance of eHealth tools remain limited [41]. A key reason is the absence of user-centric interface design [42]. A 2024 study highlights the importance of considering all stakeholders' perspectives and needs in product, system, and service design, identifying the significant challenge posed by the absence of such a focus on digital health care technology [32]. The study reveals that software usability directly impacts the smooth delivery of health care products and services and determines users' operational capability, acceptance, and satisfaction [43].

Our research indicated that the principle of user-centric design can be further refined to tailor designs to different stakeholders. Regarding the app interface proposed in this study, some HCPs required a short period to learn and adapt to the app to assess whether its functionalities met their needs or required improvement. Similarly, many patients appeared to lack the skills and abilities to sufficiently manage eHealth technology;

therefore, they were possibly unable to provide effective evaluations on functionality evaluations or suggestions for improvement or even determine the potential of the app for future home use. Previous research has emphasized that the challenges arising from implementing home-based care often stem from overly complex, expensive, or bulky equipment [44]. This study builds upon current literature and suggests that the technology used at home could pose challenges for patients who are older adults [44]. One challenge is attributed to issues with interface design between software and hardware, which negatively impact performance [45]. Thus, home-based care devices and equipment could be perceived as unhelpful and introducing such technology into patient rehabilitation at home could be rendered futile. Consequently, interface design should prioritize not only the capabilities and preferred modes of communication of app users but also how the information is transmitted and visually displayed. For instance, considering the educational levels of older adults and their unfamiliarity with consumer electronics, graphical representations should be favored over text and clickable options preferred over dropdown menus [46].

Feasibility of Digital Technology in Home Health Care

The benefits of digital technology alleviate the workload of clinical personnel in addition to providing more precise references for clinical decision-making [47]. An immediate assessment of patients' current physiological status and a prediction of their needs can be derived from the data generated by user interaction behavior [48]. However, effective information communication and exchange rely on the capabilities of caregiver and patients. Therefore, a well-designed user interface becomes a crucial element in message transmission and is also a reason for users to accept, trust, and rely on technology [49].

Or et al [50] explored technology acceptance among patients with early-stage disease, with a primary focus on factors relevant to older adult populations. The authors suggested that patients are more likely to adopt technologies if they perceive them as useful or are satisfied with the recommendations provided by HCPs through technological assistance. Conversely, users may reject or discontinue communication with HCPs when they fail to recognize the benefits of technology, struggle to understand its significance due to technological barriers, or experience negative emotions, such as increased anxiety, uncertainty, or fear.

The feasibility and acceptance of technology for managing chronic diseases among older adults entail key considerations, primarily due to potential limitations in perceptual abilities [51]. Factors such as diminished cognitive function, tactile sensitivity, and visual acuity may hinder the use of technology, consequently impacting the efficacy and acceptance of using mobile apps. Research findings indicate significant challenges among most older individuals in navigating touch-screen controls alongside difficulties in comprehending textual information displayed on screens. Hence, presenting relevant information through visual representations or incorporating voice prompts better caters to the needs of the older adults.

On the contrary, both patients and HCPs with higher education levels and tablet familiarity preferred using visual aids instead of text to illustrate patient recovery progress. In addition, they advocated for longer data cycles and considered them instrumental in enhancing the accuracy of medical decisions and influencing patient recovery results and timelines. As irreversible conditions, chronic lung diseases necessitate long-term monitoring and health management, making home rehabilitation and health management an inevitable trend among patients. The use of digital technology not only benefits HCPs but also closely relates to the health of patients with chronic disease. However, the benefits of digital technology can only be realized by leveraging its functional advantages, which allows all parties involved to benefit from and consequently accept eHealth tools. The usability tests conducted in this study revealed that despite numerous suggestions and challenges encountered by both groups of participants in using the app, there was still a high level of satisfaction with the intervention method. Importantly, most (11/18, 61%) participants expressed willingness to adopt this approach as a future method of health management.

Recommendations for App Design Based on the Results of Usability Testing

The usability study revealed challenges for older adults in using tablets or other touch-screen interfaces [52]. In addition to lower-than-expected tactile responses, their limited visual acuity and cognitive abilities may hinder their understanding of on-screen information [53]. Some studies have suggested that presenting information in visual formats is an effective method to alleviate communication barriers and enhance comprehension, particularly benefiting older adults [54]. Those with limited tactile perception may require additional aids such as auditory cues or styluses.

Both participant groups in this study recommended that designs be easily understandable and advocated for presenting clinically relevant graphical data in a simple and intuitive manner. Furthermore, during the trial period, some patients struggled with input commands (eg, drop-down menus) due to unfamiliarity with tablet touch-screen interfaces, leading to feelings of helplessness or confusion. Consequently, such input interfaces may not be suitable for older adults, and alternatives are required to meet their needs, for example, click-based interactions coupled with imagery, which are widely used pain assessment tools.

In terms of operational capability, most (11/12, 91%) HCPs encountered a few issues in tasks such as creating account, logging in, accessing patient information, managing or adjusting web-based prescriptions, and monitoring patient dynamics. However, some HCPs expressed concerns about the impact of connection quality on delivering medical services. Unstable connections could lead to message asynchrony as well as screen delays and lagging. While some HCPs considered this a normal occurrence due to data processing and screen updates, others believed that it might affect the effectiveness of medical services.

In contrast, patients could face more challenges in operation. Research observations have indicated that the primary reason that most patients can operate technology smoothly is due to assistance from caregivers. However, as the interface of this app is primarily displayed in Chinese, caregivers whose native language is not Chinese likely require additional explanations and perhaps training to understand the app. To address this issue, future improvements could consider using icons to replace textual instructions for broader user adaptation.

To enhance patient privacy and security, the app implemented a bidirectional web-based consultation feature. However, this feature may not be user-friendly for patients, as many of them are unfamiliar with how to input pairing connection values via touch screens on tablets. Consequently, even if the hospital sends a request message, patients may struggle to comprehend or operate this feature. To address this issue, it is suggested that a simpler approach be adopted, such as using a basic incoming call icon as a prompt for physician calls or as a signal for connection.

Our findings demonstrate that placing human-centered design at the forefront is paramount. Considering the abilities, cognition, and perceptions of different stakeholders will aid in improving the design, usability, and acceptance of eHealth apps. Throughout the study period, no critical mistakes, such as incomplete tasks, were reported by the patients and the HCPs. Moreover, a high degree of satisfaction was expressed. Despite significant operational challenges faced by the patients, they indicated a willingness to learn to use the app and recommend it to others. This suggests that such a model has the potential to be used in home-based care settings; however, further considerations in design details are necessary.

This study involved 2 participant groups and aimed to assess the usability, cognition, and acceptance of eHealth interfaces from their perspectives. The study included a total of 30 participants, comprising 12 (40%) HCPs and 18 (60%) patients; this met the required sample size for such studies. Nielsen [55] suggested that 5 participants are typically adequate to identify most issues; however, the actual sample size may vary depending on different categories, which has also been noted by numerous related studies. The findings of this study are consistent with those of other investigations evaluating the usability of eHealth technology tools. Nonetheless, there is a need for further refinement, particularly in simplifying the communication hardware and app interfaces used by patients. Tablets may not be the most suitable tools for older adults, and other alternative solutions such as voice calls instead of text input should be considered.

Conclusions

The survey results of this study indicated that most (11/12, 91%) HCPs found the app intuitive and easy to use, while most (5/18, 28%) older patients found it challenging to operate independently. Nevertheless, both groups of participants exhibited a high level of satisfaction in the usability satisfaction survey. It is imperative for us to focus more on discussing and addressing the difficulties and dissatisfactions encountered by participants during app use to enhance the potential for future home-based apps. By examining errors during actual operations, conducting usability satisfaction surveys, and analyzing qualitative data, we can gain a better understanding of the use of mobile health apps during testing. The results from user testing aided in comprehending the actual functioning of mobile apps and served as a basis for design modifications. Moreover, our usability tests underscore the importance of tailoring app designs to accommodate contextual factors and user characteristics, such as age, education, and functional conditions, especially in populations with chronic diseases like COPD. Addressing these variables is pivotal for ensuring the effective adoption and acceptance of digital technologies in health care settings.

Acknowledgments

This study received support from the National Science and Technology Council through grant 111-2222-E-182-02-MY2. The authors sincerely thank all the participants for their valuable contributions to this study. Special appreciation goes to the Respiratory Therapy Department at Chang Gung Memorial Hospital for their invaluable clinical support.

Data Availability

The datasets used in this study are not publicly accessible due to the need to safeguard participants' personal data. However, interested parties may obtain access to the datasets by contacting the corresponding author and making a reasonable request.

Authors' Contributions

SYC contributed to the study design, data analysis, and drafting of the manuscript.

Conflicts of Interest

None declared.

References

1. Seeger W, Adir Y, Barberà JA, Champion H, Coghlan JG, Cottin V, et al. Pulmonary hypertension in chronic lung diseases. J Am Coll Cardiol 2013 Dec 24;62(25 Suppl):D109-D116 [FREE Full text] [doi: 10.1016/j.jacc.2013.10.036] [Medline: 24355635]

2. Seemungal TA, Donaldson GC, Paul EA, Bestall JC, Jeffries DJ, Wedzicha JA. Effect of exacerbation on quality of life in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med 1998 May;157(5 Pt 1):1418-1422. [doi: 10.1164/ajrccm.157.5.9709032] [Medline: 9603117]
3. Siafakas N, Vermeire P, Pride N, Paoletti P, Gibson J, Howard P, et al. Optimal assessment and management of chronic obstructive pulmonary disease (COPD). The European Respiratory Society task force. Eur Respir J 1995 Aug 01;8(8):1398-1420 [FREE Full text] [doi: 10.1183/09031936.95.08081398] [Medline: 7489808]
4. Donaldson G, Seemungal TA, Bhowmik A, Wedzicha JA. Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. Thorax 2002 Oct;57(10):847-852 [FREE Full text] [doi: 10.1136/thorax.57.10.847] [Medline: 12324669]
5. Viegi G, Maio S, Fasola S, Baldacci S. Global burden of chronic respiratory diseases. J Aerosol Med Pulm Drug Deliv 2020 Aug;33(4):171-177. [doi: 10.1089/jamp.2019.1576] [Medline: 32423274]
6. GBD Chronic Respiratory Disease Collaborators. Prevalence and attributable health burden of chronic respiratory diseases, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Respir Med 2020 Jun;8(6):585-596 [FREE Full text] [doi: 10.1016/S2213-2600(20)30105-3] [Medline: 32526187]
7. Chatila WM, Thomashow BM, Minai OA, Criner GJ, Make BJ. Comorbidities in chronic obstructive pulmonary disease. Proc Am Thorac Soc 2008 May 01;5(4):549-555. [doi: 10.1513/pats.200709-148ET] [Medline: 18453370]
8. Guo C, Zhang Z, Lau AK, Lin CQ, Chuang YC, Chan J, et al. Effect of long-term exposure to fine particulate matter on lung function decline and risk of chronic obstructive pulmonary disease in Taiwan: a longitudinal, cohort study. Lancet Planet Health 2018 Mar;2(3):e114-e125 [FREE Full text] [doi: 10.1016/S2542-5196(18)30028-7] [Medline: 29615226]
9. Vestbo J, Hurd SS, Agustí AG, Jones PW, Vogelmeier C, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med 2013 Feb 15;187(4):347-365. [doi: 10.1164/rccm.201204-0596PP] [Medline: 22878278]
10. Rabe KF, Hurd S, Anzueto A, Barnes PJ, Buist SA, Calverley P, Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med 2007 Sep 15;176(6):532-555. [doi: 10.1164/rccm.200703-456SO] [Medline: 17507545]
11. Jones P, Brusselle G, Dal Negro R, Ferrer M, Kardos P, Levy M, et al. Health-related quality of life in patients by COPD severity within primary care in Europe. Respir Med 2011 Jan;105(1):57-66 [FREE Full text] [doi: 10.1016/j.rmed.2010.09.004] [Medline: 20932736]
12. Mikkelsen RL, Middelboe T, Pisinger C, Stage KB. Anxiety and depression in patients with chronic obstructive pulmonary disease (COPD). A review. Nord J Psychiatry 2004 Jul 12;58(1):65-70 [FREE Full text] [doi: 10.1080/08039480310000824] [Medline: 14985157]
13. van Manen JG, Bindels PJ, Dekker FW, IJzermans CJ, van der Zee JS, Schadé E. Risk of depression in patients with chronic obstructive pulmonary disease and its determinants. Thorax 2002 May;57(5):412-416 [FREE Full text] [doi: 10.1136/thorax.57.5.412] [Medline: 11978917]
14. Pauwels RA, Rabe KF. Burden and clinical features of chronic obstructive pulmonary disease (COPD). Lancet 2004 Aug;364(9434):613-620 [FREE Full text] [doi: 10.1016/S0140-6736(04)16855-4] [Medline: 15313363]
15. Kızılırmak D, Yorgancıoğlu A. Acute and chronic bronchitis: diagnosis and treatment. In: Cingi C, Yorgancıoğlu A, Muluk NB, Cruz AA, editors. Airway Diseases. Cham, Switzerland: Springer; 2023:1-8.
16. Safiri S, Carson-Chahhoud K, Noori M, Nejadghaderi SA, Sullman MJ, Ahmadian Heris J, et al. Burden of chronic obstructive pulmonary disease and its attributable risk factors in 204 countries and territories, 1990-2019: results from the Global Burden of Disease study 2019. BMJ 2022 Jul 27;378:e069679 [FREE Full text] [doi: 10.1136/bmj-2021-069679] [Medline: 35896191]
17. Inoue D, Watanabe R, Okazaki R. COPD and osteoporosis: links, risks, and treatment challenges. Int J Chron Obstruct Pulmon Dis 2016;11:637-648. [doi: 10.2147/COPD.S79638] [Medline: 27099481]
18. Pedersen BK, Saltin B. Exercise as medicine - evidence for prescribing exercise as therapy in 26 different chronic diseases. Scand J Med Sci Sports 2015 Dec 25;25 Suppl 3(S3):1-72 [FREE Full text] [doi: 10.1111/sms.12581] [Medline: 26606383]
19. Gillison FB, Skevington SM, Sato A, Standage M, Evangelidou S. The effects of exercise interventions on quality of life in clinical and healthy populations; a meta-analysis. Soc Sci Med 2009 May;68(9):1700-1710 [FREE Full text] [doi: 10.1016/j.socscimed.2009.02.028] [Medline: 19297065]
20. Spruit MA, Singh SJ, Garvey C, ZuWallack R, Nici L, Rochester C, ATS/ERS Task Force on Pulmonary Rehabilitation. An official American thoracic society/European respiratory society statement: key concepts and advances in pulmonary rehabilitation. Am J Respir Crit Care Med 2013 Oct 15;188(8):e13-e64. [doi: 10.1164/rccm.201309-1634ST] [Medline: 24127811]
21. GBD 2015 Chronic Respiratory Disease Collaborators. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet Respir Med 2017 Sep;5(9):691-706 [FREE Full text] [doi: 10.1016/S2213-2600(17)30293-X] [Medline: 28822787]

22. Ho T, Tsai Y, Ruan S, Huang C, Lai F, Yu C, HINT Study Group. In-hospital and one-year mortality and their predictors in patients hospitalized for first-ever chronic obstructive pulmonary disease exacerbations: a nationwide population-based study. PLoS One 2014 Dec 9;9(12):e114866 [FREE Full text] [doi: 10.1371/journal.pone.0114866] [Medline: 25490399]
23. Horowitz MB, Littenberg B, Mahler DA. Dyspnea ratings for prescribing exercise intensity in patients with COPD. Chest 1996 May;109(5):1169-1175 [FREE Full text] [doi: 10.1378/chest.109.5.1169] [Medline: 8625662]
24. Sasso JP, Eves ND, Christensen JF, Koelwyn GJ, Scott J, Jones LW. A framework for prescription in exercise-oncology research. J Cachexia Sarcopenia Muscle 2015 Jun;6(2):115-124 [FREE Full text] [doi: 10.1002/jcsm.12042] [Medline: 26136187]
25. Mereles D, Ehlken N, Kreuscher S, Ghofrani S, Hoeper MM, Halank M, et al. Exercise and respiratory training improve exercise capacity and quality of life in patients with severe chronic pulmonary hypertension. Circulation 2006 Oct 03;114(14):1482-1489. [doi: 10.1161/CIRCULATIONAHA.106.618397] [Medline: 16982941]
26. Marquis N, Larivée P, Dubois M, Tousignant M. Are improvements maintained after in-home pulmonary telerehabilitation for patients with chronic obstructive pulmonary disease? Int J Telerehabil 2014 Jan 29;6(2):21-30 [FREE Full text] [doi: 10.5195/ijt.2014.6156] [Medline: 25945226]
27. Nazir A, Moeliono MA. Monitored home-based exercise rehabilitation for patient with chronic respiratory and cardiovascular diseases during COVID-19 pandemic: a case report. Indo J Phys Med Rehabil 2021 Jul 30;9(02):113-123 [FREE Full text] [doi: 10.36803/ijpmr.v9i2.300]
28. Velardo C, Shah SA, Gibson O, Clifford G, Heneghan C, Rutter H, EDGE COPD Team. Digital health system for personalised COPD long-term management. BMC Med Inform Decis Mak 2017 Feb 20;17(1):19 [FREE Full text] [doi: 10.1186/s12911-017-0414-8] [Medline: 28219430]
29. Vitacca M, Montini A, Comini L. How will telemedicine change clinical practice in chronic obstructive pulmonary disease? Ther Adv Respir Dis 2018 Feb 07;12:1753465818754778 [FREE Full text] [doi: 10.1177/1753465818754778] [Medline: 29411700]
30. Alsharif AH. Cross sectional e-health evaluation study for telemedicine and m-Health approaches in monitoring COVID-19 patients with chronic obstructive pulmonary disease (COPD). Int J Environ Res Public Health 2021 Aug 12;18(16):8513 [FREE Full text] [doi: 10.3390/ijerph18168513] [Medline: 34444261]
31. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of eHealth interventions: systematic review of the literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
32. Chien SY, Wong AM, Tseng W, Hu H, Cho H. Feasibility and design factors for home-based pulmonary rehabilitation of patients with chronic obstructive pulmonary disease and chronic lung diseases based on a people-object-environment framework: qualitative interview study. JMIR Hum Factors 2024 Mar 07;11:e51150 [FREE Full text] [doi: 10.2196/51150] [Medline: 38452366]
33. Epping-Jordan JE, Pruitt SD, Bengoa R, Wagner EH. Improving the quality of health care for chronic conditions. Qual Saf Health Care 2004 Aug;13(4):299-305 [FREE Full text] [doi: 10.1136/qhc.13.4.299] [Medline: 15289634]
34. Faessen JL, van Vugt R, Veldhuizen R, Stoot JH. Using an e-Health application for post-operative monitoring after inguinal hernia repair: a feasibility study. World J Surg 2023 Jan 23;47(1):182-189 [FREE Full text] [doi: 10.1007/s00268-022-06590-1] [Medline: 35604449]
35. Alshahrani A, Stewart D, MacLure K. A systematic review of the adoption and acceptance of eHealth in Saudi Arabia: views of multiple stakeholders. Int J Med Inform 2019 Aug;128:7-17 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.05.007] [Medline: 31160014]
36. Farahani B, Firouzi F, Chang V, Badaroglu M, Constant N, Mankodiya K. Towards fog-driven IoT eHealth: promises and challenges of IoT in medicine and healthcare. Future Gener Comput Syst 2018 Jan;78:659-676 [FREE Full text] [doi: 10.1016/j.future.2017.04.036]
37. Neubeck L, Coorey G, Peiris D, Mulley J, Heeley E, Hersch F, et al. Development of an integrated e-health tool for people with, or at high risk of, cardiovascular disease: the consumer navigation of electronic cardiovascular tools (CONNECT) web application. Int J Med Inform 2016 Dec;96:24-37 [FREE Full text] [doi: 10.1016/j.ijmedinf.2016.01.009] [Medline: 26847070]
38. Garfan S, Alamoodi A, Zaidan B, Al-Zobbi M, Hamid RA, Alwan JK, et al. Telehealth utilization during the COVID-19 pandemic: a systematic review. Comput Biol Med 2021 Nov;138:104878 [FREE Full text] [doi: 10.1016/j.compbiomed.2021.104878] [Medline: 34592585]
39. De Rouck S, Jacobs A, Leys M. A methodology for shifting the focus of e-health support design onto user needs: a case in the homecare field. Int J Med Inform 2008 Sep;77(9):589-601 [FREE Full text] [doi: 10.1016/j.ijmedinf.2007.11.004] [Medline: 18248846]
40. Harte R, Glynn L, Rodríguez-Molinero A, Baker PM, Scharf T, Quinlan LR, et al. A human-centered design methodology to enhance the usability, human factors, and user experience of connected health systems: a three-phase methodology. JMIR Hum Factors 2017 Mar 16;4(1):e8 [FREE Full text] [doi: 10.2196/humanfactors.5443] [Medline: 28302594]
41. Ossebaard HC, Van Gemert-Pijnen L. eHealth and quality in health care: implementation time. Int J Qual Health Care 2016 Jun 29;28(3):415-419 [FREE Full text] [doi: 10.1093/intqhc/mzw032] [Medline: 27029590]

42. Arsand E, Demiris G. User-centered methods for designing patient-centric self-help tools. Inform Health Soc Care 2008 Sep 12;33(3):158-169 [FREE Full text] [doi: 10.1080/17538150802457562] [Medline: 18850399]
43. Negash YT, Calahorrano Sarmiento LS. Smart product-service systems in the healthcare industry: intelligent connected products and stakeholder communication drive digital health service adoption. Heliyon 2023 Feb;9(2):e13137 [FREE Full text] [doi: 10.1016/j.heliyon.2023.e13137] [Medline: 36820023]
44. Chan M, Estève D, Escriba C, Campo E. A review of smart homes- present state and future challenges. Comput Methods Programs Biomed 2008 Jul;91(1):55-81 [FREE Full text] [doi: 10.1016/j.cmpb.2008.02.001] [Medline: 18367286]
45. Chen JY, Haas EC, Barnes MJ. Human performance issues and user interface design for teleoperated robots. IEEE Trans Syst Man Cybern 2007 Nov;37(6):1231-1245. [doi: 10.1109/tsmcc.2007.905819]
46. Harle CA, Golembiewski EH, Rahmanian KP, Krieger JL, Hagmajer D, Mainous AG, et al. Patient preferences toward an interactive e-consent application for research using electronic health records. J Am Med Inform Assoc 2018 Mar 01;25(3):360-368 [FREE Full text] [doi: 10.1093/jamia/ocx145] [Medline: 29272408]
47. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020;3:17 [FREE Full text] [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
48. Li Q, Luximon Y. Older adults and digital technology: a study of user perception and usage behavior. In: Proceedings of the AHFE 2016 International Conference on Physical Ergonomics and Human Factors. 2016 Presented at: APEHF '16; July 27-31, 2016; Walt Disney World, FL p. 27-31 URL: https://doi.org/10.1007/978-3-319-41694-6_16 [doi: 10.1007/978-3-319-41694-6_17]
49. Johnson J. Designing with the Mind in Mind: Simple Guide to Understanding User Interface Design Guidelines. Boston, MA: Morgan Kaufmann Publishers; 2020.
50. Or CK, Karsh B, Severtson DJ, Burke LJ, Brown RL, Brennan PF. Factors affecting home care patients' acceptance of a web-based interactive self-management technology. J Am Med Inform Assoc 2011;18(1):51-59 [FREE Full text] [doi: 10.1136/jamia.2010.007336] [Medline: 21131605]
51. Holthe T, Halvorsrud L, Karterud D, Hoel KA, Lund A. Usability and acceptability of technology for community-dwelling older adults with mild cognitive impairment and dementia: a systematic literature review. Clin Interv Aging 2018;13:863-886 [FREE Full text] [doi: 10.2147/CIA.S154717] [Medline: 29765211]
52. Page T. Touchscreen mobile devices and older adults: a usability study. Int J Hum Factors Ergon 2014;3(1):65 [FREE Full text] [doi: 10.1504/ijhfe.2014.062550]
53. Ocak N, Cagiltay K. Comparison of cognitive modeling and user performance analysis for touch screen mobile interface design. Int J Hum Comput Interact 2016 Dec 23;33(8):633-641 [FREE Full text] [doi: 10.1080/10447318.2016.1274160]
54. Houts PS, Doak CC, Doak LG, Loscalzo MJ. The role of pictures in improving health communication: a review of research on attention, comprehension, recall, and adherence. Patient Educ Couns 2006 May;61(2):173-190 [FREE Full text] [doi: 10.1016/j.pec.2005.05.004] [Medline: 16122896]
55. Nielsen K, Randall R, Holten A, González ER. Conducting organizational-level occupational health interventions: what works? Work Stress 2010 Jul;24(3):234-259 [FREE Full text] [doi: 10.1080/02678373.2010.515393]

Abbreviations

COPD: chronic obstructive pulmonary disease
HCP: health care professional

Edited by A Kushniruk, E Borycki; submitted 29.04.24; peer-reviewed by K Heslop-Marshall, YW Shau; comments to author 12.08.24; revised version received 15.08.24; accepted 13.09.24; published 15.11.24.

Please cite as:
Chien SY
Mobile App for Patients With Chronic Obstructive Pulmonary Diseases During Home-Based Exercise Care: Usability Study
JMIR Hum Factors 2024;11:e60049
URL: https://humanfactors.jmir.org/2024/1/e60049
doi:10.2196/60049
PMID:

Social Media Recruitment as a Potential Trigger for Vulnerability: Multistakeholder Interview Study

Nina Matthes[1]; Theresa Willem[1,2,3], MA; Alena Buyx[1], Prof Dr Med; Bettina M Zimmermann[1,2,4], PhD

[1]Institute of History and Ethics in Medicine, TUM School of Medicine and Health, Technical University of Munich, Ismaningerstr. 22, Munich, Germany
[2]Institute of Molecular Immunology, Klinikum Rechts der Isar, TUM School of Medicine and Health, Technical University of Munich, Munich, Germany
[3]Department of Science, Technology and Society (STS), School of Social Sciences and Technology, Technical University of Munich, Munich, Germany
[4]Institute of Philosophy, Multidisciplinary Center for Infectious Diseases, University of Bern, Bern, Switzerland

Corresponding Author:
Nina Matthes
Institute of History and Ethics in Medicine, TUM School of Medicine and Health, Technical University of Munich, Ismaningerstr. 22, Munich, Germany

Abstract

Background: More clinical studies use social media to increase recruitment accrual. However, empirical analyses focusing on the ethical aspects pertinent when targeting patients with vulnerable characteristics are lacking.

Objective: This study aims to explore expert and patient perspectives on vulnerability in the context of social media recruitment and seeks to explore how social media can reduce or amplify vulnerabilities.

Methods: As part of an international consortium that tests a therapeutic vaccine against hepatitis B (TherVacB), we conducted 30 qualitative interviews with multidisciplinary experts in social media recruitment (from the fields of clinical research, public relations, psychology, ethics, philosophy, law, and social sciences) about the ethical, legal, and social challenges of social media recruitment. We triangulated the expert assessments with the perceptions of 6 patients with hepatitis B regarding social media usage and attitudes relative to their diagnosis.

Results: Experts perceived social media recruitment as beneficial for reaching hard-to-reach populations and preserving patient privacy. Features that may aggravate existing vulnerabilities are the acontextual point of contact, potential breaches of user privacy, biased algorithms disproportionately affecting disadvantaged groups, and technological barriers such as insufficient digital literacy skills and restricted access to relevant technology. We also report several practical recommendations from experts to navigate these triggering effects of social media recruitment, including transparent communication, addressing algorithm bias, privacy education, and multichannel recruitment.

Conclusions: Using social media for clinical study recruitment can mitigate and aggravate potential study participants' vulnerabilities. Researchers should anticipate and address the outlined triggering effects within this study's design and proactively define strategies to overcome them. We suggest practical recommendations to achieve this.

(JMIR Hum Factors 2024;11:e52448) doi:10.2196/52448

KEYWORDS

vulnerability; social media; clinical study enrollment; clinical study recruitment; clinical trials; stigma; discrimination; injustice; recruitment; clinical study; hepatitis B; TherVacB; clinical research; attitudes; patient privacy; utilization

Introduction

Vulnerability, defined as "an increased likelihood of being wronged or of incurring additional harm," [1] has been a topic of significant interest in research ethics. Types of vulnerability include cognitive, juridic, deferential, medical, allocational, and infrastructural vulnerability [2]. The Declaration of Helsinki underscores that vulnerable populations require heightened protection due to their increased risk of harm, stating that research involving them is only justified when it directly addresses their health needs and cannot be conducted with nonvulnerable groups [1]. Yet, labeling whole population groups as vulnerable has been criticized as overly simplistic [3]. Accordingly, the revised Council for International Organizations of Medical Sciences guidelines from 2016 emphasize the importance of avoiding the exclusion of entire groups considered vulnerable from research participation under the guise of protecting their well-being, which has led to limited knowledge about diseases affecting these groups:

> *In the past, groups considered vulnerable were excluded from participation in research because it was considered the most expedient way of protecting those groups (for example, children, women of reproductive age, and pregnant women). As a*

> *consequence of such exclusions, information about the diagnosis, prevention, and treatment of diseases that afflict such groups is limited. This has resulted in a serious injustice.… The need to redress these injustices by encouraging the participation of previously excluded groups in basic and applied biomedical research is widely recognized.* [4]

To mitigate such effects, research ethicists suggested a more dynamic and context-dependent understanding of vulnerability. Luna [3] recognized that individuals may experience varying degrees of vulnerability and proposed that vulnerability is constituted by layers. Victor et al [5] emphasized the importance of identifying and addressing triggers that may exacerbate certain layers of vulnerability. Aggregating such layers of vulnerability might lead to cascading effects, a chain reaction of consequences that occur because of the activation of one vulnerability layer [5].

As such, these authors understand vulnerability as the sum of such layers (ie, personal characteristics or situational circumstances) that put a person at risk of being harmed or disadvantaged in specific contexts. An individual's vulnerability is shaped by a combination of factors (eg, social, economic, and health-related) that can influence one another in ways that heighten or reduce overall vulnerability.

One potential factor that might aggravate—or mitigate—vulnerability in the research context is using social media as a recruitment channel for research studies. On the one hand, social media recruitment allows for effective reaching of populations that are challenging to reach through other recruitment channels since social media have targeted advertising features that enable researchers to tailor recruitment efforts to specific characteristics, interests, and behaviors. The accessibility of social media allows individuals to participate in research studies from anywhere with an internet connection, removing barriers related to geographic location, mobility limitations, or time constraints [6]. Thus, in specific contexts, social media recruitment may achieve higher enrollment rates and be more cost-effective than other recruitment methods [7]. On the other hand, however, social media can also involve considerable risks for specific target groups, including potential privacy violations, the risk of stigmatization, and challenges in ensuring proper informed consent [8]. Thus, to ensure ethical recruitment strategies, it is essential to identify how social media recruitment might trigger or mitigate vulnerabilities, especially in the context of clinical studies, and examine ways to address these triggering effects.

Previous studies investigating the ethical benefits and challenges regarding social media recruitment found that ethical benefits primarily include reaching a broader and more diverse pool of participants, and cost-effectiveness and direct engagement with potential participants. However, challenges such as ensuring data privacy, navigating regulatory compliance, and managing the quality of information dissemination require careful attention to maintain ethical standards [7-10]. Key recommendations from these studies stress the importance of ensuring recruitment strategies adhere to relevant ethical norms and comply with federal and state laws. Transparency is also crucial for building trust and upholding ethical standards throughout the recruitment process [10].

Despite these contributions, to our knowledge, no empirical studies specifically address vulnerability in the context of social media recruitment. For understanding the recruitment for clinical studies via social media as a "stimulus condition" [11] that potentially triggers or aggravates existing vulnerabilities, this paper aims to explore experts' perceptions of vulnerability and those of patients with hepatitis B in the context of social media recruitment. We asked multidisciplinary experts and patients diagnosed with hepatitis B about their experiences and perceptions of recruiting vulnerable people through social media. We addressed the following research questions: (1) How can social media recruitment mitigate existing vulnerabilities in the context of clinical study recruitment? (2) What social media features can trigger and exacerbate vulnerabilities and why? (3) How do interviewees suggest navigating these triggering effects?

Chronic hepatitis B infection serves as a use case for vulnerability, underscoring the interplay between medical and psychosocial vulnerabilities experienced by affected individuals. Hepatitis B, a viral liver infection caused by the hepatitis B virus, leads to chronic liver disease with severe complications such as cirrhosis and hepatocellular carcinoma, presenting a significant public health challenge worldwide [12]. Individuals with hepatitis B encounter medical vulnerability due to their medically exigent state, compounded by the absence of a curative treatment. This renders them susceptible to exploitation, as they may enroll in clinical trials with inflated hopes of accessing potentially effective treatment. Their vulnerability is underscored by a persistent but often misguided hope for a cure, leading them to enter studies with unrealistic expectations of success [13]. Beyond medical vulnerabilities, individuals living with hepatitis B also face profound psychosocial challenges. Stigma and discrimination persist in many societies, driven by misconceptions about transmission routes and fear of contagion. This stigma manifests in various parts of life, including employment, health care settings, and social interactions, resulting in feelings of shame, isolation, and psychological distress [14].

Methods

Study Design

As part of the international research consortium "TherVacB - A Therapeutic Vaccine to Cure Hepatitis B," we conducted semistructured interviews with multidisciplinary experts and patients with hepatitis B to explore the ethical, social, and regulatory issues of social media recruitment for clinical studies. In this paper, the analysis focused especially on the aspects of vulnerability. We followed the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist to report on the research team, study design, and data analysis [15].

Ethical Considerations

Human Subject Ethics Review Approval

This study received approval from the Ethics Committee of the Technical University of Munich (431/20 S-KH).

Informed Consent

All participants were provided with an information sheet detailing the study's purpose and scope. Written informed consent was obtained from each participant before the interviews.

Privacy and Confidentiality

Interview transcripts were pseudonymized by replacing identifiable information with placeholders.

Compensation Details

Participants did not receive any financial compensation for their participation in the study.

Recruitment

Stakeholders with practical or theoretical experience in social media–based recruitment were eligible to participate. They were recruited via snowballing, convenient sampling through the TherVacB network, and screening the corresponding authors of relevant scientific publications in the field. The stepwise recruitment process was guided by considerations of theoretical saturation [16] and was stopped when additional interviews were not expected to reveal any relevant new findings based on the ongoing data analysis.

Moreover, we included patients with hepatitis B as an additional stakeholder group. Patients with hepatitis B aged 18 years and above with at least one social media account and English or German language skills were qualified to participate. They were recruited during regular hepatitis B-related check-ups at a German University Hospital. Due to the COVID-19 pandemic, patient recruitment was challenging. Therefore, no more than 6 patients with hepatitis B participated in this study. Consequently, theoretical saturation is limited for the patient population alone. We considered the insights from these interviews as an additional triangulation point to further substantiate our qualitative analysis.

Data Collection

TW and BMZ held the semistructured, qualitative interviews in German or English via phone or videoconferencing. We used separate interview guidelines for experts and patients with hepatitis B. The expert interview guide focused on the ethical, legal, social, and practical risks and benefits of social media recruitment. The patient interview guide included questions about their experiences with social media and their disease, drawing particular attention to the potential stigma and privacy issues as essential aspects of vulnerability in the context of hepatitis B and social media. Experts were also asked "The patients eligible for the TherVacB clinical trial often have particularly vulnerable characteristics (eg economic, social). What do you think is noteworthy to consider in this context?"

The interviews were conducted between August 2020 and September 2021. Each interview lasted between 25 and 60 minutes. Interviews were recorded and transcribed verbatim.

Data Analysis

The research team (BMZ; TW; Nina Goldman, PhD; and NM) transcribed audio recordings verbatim. Interview transcripts were coded based on inductive thematic analysis [17,18], using Atlas.ti (version 9; Scientific Software Development GmbH) software. Interview transcripts were analyzed based on reflexive thematic analysis (see Multimedia Appendix 1 for the final list of codes). Based on the first 6 expert interviews, BMZ and TW developed a preliminary coding scheme, which they applied to these 6 interviews. Then, they refined and reviewed the coding scheme in subsequent interviews until they identified no additional relevant codes. Next, the research team summarized the contents of each code across all interviews in analytical memos and related them to each other, with existing literature and theory. Thereby, we found vulnerability to be a central aspect of the interviews that warrants specific empirical and conceptual attention. In multiple rounds of discussion, NM, BMZ, and TW discussed aspects pertinent to the data inductively and refined them iteratively. Illustrative quotes in German were translated to English by NM and double-checked by BMZ.

Results

Overview

We conducted qualitative interviews with 6 patients with hepatitis B and 30 multidisciplinary experts from clinical research, public relations, psychology, ethics, philosophy, law, and social sciences. The patients resided in Germany; the experts were from Australia, Canada, Germany, Spain, Switzerland, and the United States. The findings from the qualitative analysis of these interviews are structured as follows: first, we outline how experts perceived social media recruitment to help mitigate pertinent vulnerabilities (eg, being affected by a stigmatized condition or belonging to historically disadvantaged populations) in the context of clinical study recruitment. Second, we present four features of social media that may trigger vulnerabilities. Third, we present experts' recommendations from the interviews on mitigating some of these triggering effects.

How Social Media Recruitment Can Help Mitigate Vulnerabilities

Reaching Hard-to-Reach Populations

Many experts have emphasized the potential of social media recruitment to reach traditionally disadvantaged hard-to-reach populations for research effectively. The paternalistic approach, often excluding these groups under the guise of protection, has led to a lack of research findings tailored to their specific needs and circumstances. An ethicist cautioned against such paternalistic views of vulnerability, underscoring the importance of respecting individuals' autonomy:

> *I believe that…it is important that the truly crucial concept of vulnerability is not used in an overly paternalistic way. To [avoid] thinking that anyone vulnerable is generally not autonomous.* [Ethicist 1, Switzerland]

Achieving inclusivity in research necessitates shifting from paternalistic protectionism toward empowerment and collaboration with vulnerable populations. Meaningful inclusion entails valuing the autonomy of vulnerable individuals and actively involving them as partners in the research process while safeguarding their rights. Social media recruitment emerges as a valuable tool in this context, enabling researchers to reach these hard-to-reach populations more effectively, thereby promoting inclusivity and equity in research participation. Another ethicist further stressed the tension between protecting and recruiting patients with vulnerable characteristics:

> *And maybe at some point, you can say, you know, you shouldn't even be using social media, at least maybe in certain ways to recruit those really kinds of stigmatized groups. Again, I think we have to keep the benefits in mind here due to our earlier discussion because some of these folks are going to be/ Maybe this is the only way you're going to reach them, right? Well, it's sort of/It's going to be the best way for you to reach them, so I guess I'm committed to kind of finding ways to do it sort of safely.* [Ethicist 3, United States]

Consequently, several expert participants, especially ethicists, perceived social media as a helpful tool to reach underserved and hard-to-reach populations for research studies [19].

Preserving Patient Privacy

Despite some large social media platforms' reputation for having limited privacy standards [20], social media recruitment may, conversely, also preserve patient privacy. By using social media platforms, researchers can reach individuals experiencing stigmatized conditions or traits, such as hepatitis B, outside a personal, clinical setting. Social media users with stigmatized conditions or traits who view social media recruitment advertisements may, for instance, be more likely to participate in survey research that does not require direct contact with the research team. They may prefer online recruitment settings that offer increased discretion and anonymity.

> *In our case, and I'm thinking about the HIV study I'm involved in, we are trying very hard to reach people who may not be in healthcare. When we talk about putting flyers up around the hospital or an outpatient clinic we are already excluding people that don't have access to regular care. We may not want to show up at a venue known to be frequented for privacy purposes, so if someone is going through [inaudible] or dating website [inaudible] in a way that can protect their privacy while reaching them whenever they get a chance to see an ad.* [Ethicist 2, United States]

As we will show below, however, the predominant view among interviewees was that social media recruitment infringed on user privacy. This indicates the importance of assessing recruitment strategies in a context-sensitive manner, particularly regarding data privacy. The context relevance of who exactly the target population is may determine the risks and benefits related to this privacy aspect.

Social Media Recruitment as a Trigger for Vulnerabilities

Overview

This section is structured along the features of social media that can trigger and exacerbate (existing) vulnerabilities in the context of social media recruitment. The relative importance of vulnerabilities depends on this study's design and the target group of the social media recruitment strategy. Interviewees gave a range of examples of characteristics they considered making people vulnerable, including having sexually transmitted or infectious diseases (eg, hepatitis B), being affected by severe or untreatable conditions (eg, cancer), holding stigmatized traits (eg, sexual orientation or psychological conditions), or identifying with historically disadvantaged populations (eg, immigrants or people of color). If targeted for study recruitment via social media, these individuals may face various harms, including social exclusion, discrimination, and limited access to opportunities. For instance, those with sexually transmitted or infectious diseases may experience social exclusion and stigma. At the same time, individuals with severe illnesses may encounter financial burdens and emotional distress.

Similarly, individuals with stigmatized traits may confront prejudice, discrimination, and barriers to employment. Moreover, historically disadvantaged groups may encounter systemic inequalities, social injustices, and unequal access to employment and health care resources. While these risks of harm also exist beyond the context of clinical study participation, social media recruitment might trigger or exacerbate them.

Acontextual Point of Contact

When promoting studies via social media, individuals may encounter recruitment advertisements at unexpected moments. In the context of medical research, interviewees emphasized that learning about a clinical study outside a clinical setting might cause distress or exacerbate existing emotional challenges in participants' lives, triggering or aggravating existing vulnerabilities:

> *You don't know the timing of your reach out. You know, it could be something very disturbing. It could be not just inappropriate timing,…, but just inappropriate in the context of a person's life.… So, there is this contextual aspect of…recruiting people on social media, which can be hugely problematic.* [Ethicist 1, United States]

Relatedly, a communication specialist with practical expertise in social media recruitment deemed social media inappropriate for studies investigating severe, uncurable diseases when social media content was usually about fun and happiness because this could be off-putting for those affected. He underscored the potential mismatch between the typically light-hearted nature of social media content and the serious nature of such health conditions:

> *I believe that certain studies should simply not be promoted on social media.... We have recently received a request regarding a blood cancer study it is about the fact that the patient cannot be helped.... I consider it very unethical to advertise such things on social media platforms like Facebook, Instagram, and TikTok, at least not to offer such advertising.* [Communication specialist 3, Germany]

Moreover, it was mentioned that the acontextual and impersonal communication on social media increased the risk of therapeutic misconceptions and false hopes regarding receiving an effective or curative treatment through the clinical study. One interviewed patient with hepatitis B supported this by stating:

> *We are always told that there is never a chance for us to cure this disease,...and sometimes this thought overwhelms me. I want to search on social media to see what can be done in this regard.* [Patient 2 with hepatitis B, Germany]

Thus, acontextual communication through social media might trigger both medical vulnerabilities that come with severe, untreatable, or otherwise burdensome conditions and cognitive vulnerabilities, such as insufficient health literacy.

Public Space

A second feature of social media is that at least some communication happens in public, where others can see and interact with the content. This feature can breach peoples' need for privacy and unwantedly expose the medical information of social media users, constituting a risk of discrimination and people's right to medical data privacy. A Canadian ethicist explained that patients might find it intrusive and irritating if they received targeted messages about their condition without actively seeking that information. When people publicly announce their illness, they may not expect that researchers looking for study participants with specific conditions could encounter their shared information. Thus, sensitivity is required in the initial engagement with potential study participants via social media, as receiving a targeted advertisement or private message based on this public disclosure may cause discomfort and feel like an intrusion into their privacy:

> *And so, I think the form of initial engagement has to be sensitive to the fact that people might not expect that they've made that information public to this particular audience when they made a public announcement. And so, if you're targeting a particular condition to sort of barge in and say, hey, so-and-so with condition X, that might be chilling to the person who's receiving that message. He might feel like, how do you know this? And that could be off-putting and change their experience.* [Ethicist, Canada]

Several patients with hepatitis B confirmed the importance of privacy. They expressed fear of being exposed to their diagnosis on social media.

> *If I were to post something about my Hepatitis B to someone somewhere [on social media].... They would essentially have something in writing from me. And they could forward it or repost it at any time.* [Patient 1 with hepatitis B, Germany]

Regarding stigmatized medical conditions, a communication expert raised concerns about hate speech and the potential for discriminatory comments on social media platforms, which can be off-putting for people, meaning that it could discourage them from participating in this study. The challenge is to create recruitment posts that attract potential participants while avoiding harmful reactions in the comments section.

> *But how do you implement [social media recruitment ads] in a way that firstly appeals to the patients, and how do I also manage to avoid the hatred from healthy participants who have no understanding of what it means to have Parkinson's? Because if a person with Parkinson's reads something like this and is exposed to such advertisements, if they read those comments, what do you think will happen? They will not sign up for this study, and that's problematic.* [Communication specialist 3, Germany]

Thus, the feature of many social media platforms to operate (partly) in a public communication sphere triggers vulnerability for people with high privacy needs, such as people with stigmatized traits who would not want to be exposed. Such privacy infringements can lead to cascading effects, making people with stigmatized traits or severe diseases more vulnerable due to unsettling comments from other users or risks of discrimination.

Biased Social Media Algorithms

A third feature of social media that might trigger vulnerabilities in the context of historical discrimination is biased algorithms on social media. Several participants mentioned the issue of social media algorithms being potentially biased and discriminatory toward disadvantaged social groups. Consequently, using such algorithms for clinical study recruitment could trigger vulnerabilities for these groups. For example, one US-based ethicist working for an institutional review board stated:

> *…we see a ton of social media, almost for every major clinical trial, from big industry sponsors like Novartis, and they almost always have some kind of social media outreach. I think there has been a growing desire to address potential inequities, so I know that Facebook has been criticized for, you know, when you run a clinical trial ad, who exactly is it going to reach? Is it going to reach people who are underserved and historically disadvantaged? There has been criticism of the algorithms they use to target people for clinical trial recruitment. And I think there is some desire amongst Facebook even to try and do better on that point. But I don't know/ I guess I haven't seen it slowing down.* [Ethicist 3, United States]

While there is no right to participate in a clinical trial, the systematic neglect of research participants with certain traits might lead to lower health care standards. This occurs because clinical trials that lack diversity may produce results that are not generalizable to the entire population. For instance, if clinical trials primarily include participants from specific demographics, the findings may not accurately reflect how treatments affect other groups, such as those with different ages, genders, ethnic backgrounds, or health conditions. This can result in less effective treatments or unforeseen side effects in underrepresented groups, thereby lowering the overall quality and equity of health care. This issue is particularly relevant because social media is often perceived as a means to reach populations that are otherwise hard to engage in research, such as minority communities, rural populations, or those with limited access to traditional health care settings. However, this perception does not always hold in practice, as algorithmic biases may limit the effectiveness of social media recruitment [21]. As a result, specific populations may remain underrepresented in clinical research, perpetuating the cycle of inequitable health care standards.

Technological Tool

Finally, social media, being technological tools, cause effects that might trigger existing vulnerabilities. Several interviewed ethicists pointed out that users needed "a certain technical expertise and familiarity in dealing with such media" (ethicist 2, Switzerland). Thus, some patients may be interested in clinical studies but lack the digital skills to access this information via social media. As one expert pointed out, this can be a matter of justice if individuals with insufficient digital literacy skills are excluded from clinical studies solely advertised on social media:

> *Especially with vulnerable groups, it is certainly an issue: Who doesn't even have the opportunity or the chance to either see or respond appropriately? In that sense, it is indeed a matter of justice, but it's probably not immediately apparent under the label of justice. It might be framed differently.* [Clinical researcher 3, Germany]

In addition to insufficient literacy, restricted access to relevant technology can render those lacking this access as more vulnerable as they do not have the same access to information as others:

> *You need the hardware for it, internet access or a stable network, and so on. Many things are required, and by that, I may inadvertently exclude certain people who don't have access, who can't afford it, or who may be hesitant to engage or participate in something like this. And just like that, we have a selection bias again. I think we need to be careful here. It can even take on a discriminatory character if I only focus on such individuals.* [Ethicist 2, Switzerland]

Because social media are technological tools that require skills and hardware, their use can systematically exclude specific populations from clinical studies. This may lead to lower health care quality and aggravate medical vulnerabilities because drugs might work less efficiently. In the following section, we will present experts' suggestions for mitigating these triggering effects of social media recruitment.

Mitigating Triggering Effects of Social Media Recruitment

Overview

The interviewed experts pointed to several practical recommendations that help mitigate the aforementioned triggering effects of social media recruitment.

Communicating Transparently

First, several experts pointed to the importance of transparency when communicating with potential research participants on social media. One ethicist referred to people with vulnerable characteristics as requiring a "higher standard of transparency and consent" (ethicist 1, United States). Investigators should, therefore, always be transparent about who they are and what they are contacting the person for, as a clinical researcher from the United States pointed out. To further support patient autonomy, researchers should make sure that patients can ask questions before consenting to this study:

> *…at least granting the opportunity to ask questions. What risks it would entail for me, for example? Even if it is formulated in writing, there might be instances where someone doesn't understand it or so.… So, providing information that is as comprehensible as possible, I believe that's what it is about.* [Social scientist 1, Germany]

While transparency in research communication is a fundamental principle across all recruitment methods, it is particularly important in the context of social media recruitment, as users on social media platforms may receive numerous messages and requests. Transparency about the nature of the contact can help recipients differentiate legitimate research inquiries from spam or phishing attempts. This transparency helps establish trust and credibility with potential participants. By contrast, recruitment for clinical studies featured in newspapers or on the radio typically undergoes critical observation by the editorial teams beforehand. A clinical researcher from Germany also emphasized the need for personal interaction "to make sure that the patients understand that the information they are about to give you might contain personal health issues" (Clinical researcher 2, Germany).

Analyzing Algorithm Bias

To address the issue of biased algorithms, experts suggested researchers collaborate with ethicists and social media experts to review and assess the algorithms used for recruitment, identifying any potential biases and working toward mitigating them. Researchers should actively engage with Facebook and other social media platforms to understand how their algorithms work and to be aware of any changes or updates that may impact the fairness and inclusivity of their recruitment offers. Depending on the social media platform, this could be laborious, as this expert acknowledged:

> *But Facebook changes its rules every second day, and you have to have someone whose job it is to monitor those.… It's really, really hard from a privacy and confidentiality point of view because you're not in control of what Facebook does.* [Social scientist 2, Australia]

Protecting Privacy

One ethicist from the United States emphasized the importance of raising awareness about the potential misuse of data on social media. He suggested that instead of solely focusing on finding ways to protect privacy, it would be beneficial to educate people about the challenges of maintaining privacy online:

> *I feel as though we need to be much more proactive about creating tools to protect all sorts of people in social media space from harm that may result from the misuse of their data. Rather than trying to figure out ways/ I think it would be healthier if we educated people that they've given up a ton of privacy, and maintaining privacy by keeping your information inaccessible is hard in these spaces and that we need to create an environment that creates consequences for the misuse of people's information. So we shifted towards that sort of dynamic, maybe [inaudible] have a better understanding that but limited when you're participating in a lot of these spaces.* [Ethicist 4, United States]

A clinical researcher with practical experience in social media recruitment explained that his research team did not allow comments on their Facebook page to prevent patients from sharing information "about their diagnosis status or ask[Ing] questions about medication management" (clinical researcher 3, United States). To prevent patients from inadvertently disclosing personal health information, it is advisable not to specifically target individuals with a particular diagnosis in recruitment posts on social media. A US-based ethicist suggested making a disclaimer saying that users should not directly reply to study-related social media advertisements. Still, experts acknowledge the challenge of protecting patient privacy on social media because platforms may change their privacy policy anytime.

Using Multichannel Recruitment Approaches

Because not everybody has access to social media and might not see information regarding relevant clinical studies, one ethicist from the United States suggested that it may be necessary to use alternative offline strategies to reach populations with limited social media access:

> *Social media...may not be available to everyone.... So we have to sort of keep that in mind. There is a place where you might want to drive up and hand out flyers outside of a soup kitchen, depending on who you're trying to talk to.* [Ethicist 2, United States]

Therefore, researchers should consider the possible drawbacks of relying solely on social media as a recruitment method and instead devise an approach that integrates various strategies (both online and offline) to promote inclusivity in clinical studies and increase the likelihood of reaching a more diverse and representative sample of participants. In addition, one expert suggested learning the target population's behavior to "track down what data is available in terms of what websites they are on" (ethicist, Canada). While the other presented strategies involve specific actions to mitigate the vulnerability-triggering effects of social media recruitment, multichannel recruitment consists of integrating online and offline recruitment methods to enhance inclusivity.

Discussion

Principal Findings

In many clinical studies, social media offer one of several ways to recruit potential study participants. Most often, participants are recruited within the clinical setting, through snowballing techniques, or via more traditional forms of advertisement such as billboards, newspaper advertisements, and community networks. Gelinas et al [10] argued for a nonexceptional approach to evaluating the ethical implications of social media recruitment because the same research ethics principles (respect for persons, justice, and beneficence) apply. Yet, as we delve into this discourse, we recognize the need for a deeper examination. In this context, we propose to extend Gelinas and colleagues' [10] recommendations. We offer the first empirical evidence of a more nuanced assessment of social media recruitment regarding participant vulnerability by showing features specific to social media that might mitigate or trigger vulnerabilities of potential study participants in the context of clinical study recruitment.

We will now discuss how the identified vulnerability triggers of social media recruitment are exclusive to social media. The vulnerabilities associated with social media recruitment are not mere extensions of issues seen in more traditional recruitment approaches. Instead, social media themselves shape vulnerabilities due to their unique characteristics. We argue that social media uniquely combines ethically relevant features of other recruitment channels and, thus, warrants special attention and ethical scrutiny. First, social media are unique because they operate in a semipublic sphere [22]: within this sphere, content can straddle the line between public and private, creating a complex web of accessibility. For the average user, determining who can access which specific activities and digital traces on social media is not trivial [23], particularly when considering the constantly changing terms of use imposed by social media platforms. Second, and this distinguishes social media recruitment from other online recruitment activities, is the deeply personal nature of these platforms. Social media platforms serve as more than just spaces for sharing information; they also collect vast amounts of data about individuals, including their preferences, behaviors, and interests.

Users recognize that the content they encounter on these platforms is tailor-made for them based on their previous interactions and the data collected about them. This personalization sets social media recruitment apart from traditional methods such as billboard advertisements, which are generally not personalized. This individual targeting might be off-putting for users because they may feel that their privacy is being invaded, which is not a concern with more traditional recruitment approaches. Finally, social media recruitment poses unique challenges related to privacy and algorithmic discrimination [20]. The power wielded by social media platforms in the digital realm is unprecedented, and their ability to shape the recruitment process is substantial. These platforms' algorithmic decisions can inadvertently contribute to inequalities in health care and research participation [21]. Thus, when researchers use nondisclosed social media algorithms for recruitment, they may unintentionally exacerbate vulnerabilities for marginalized communities by targeting specific demographic groups and inadvertently excluding others, potentially exacerbating disparities in health care access and research participation. The use of such algorithms for clinical study recruitment has the potential to trigger vulnerabilities for marginalized groups. It is essential to consider the far-reaching impact of social media platforms and their algorithms on fairness in the recruitment process. As social media are used to recruit potential study participants, these risks of triggering or exacerbating vulnerabilities exist independent of actual research participation. Consequently, potential research participants could be harmed even without consenting to participate in a study, making it even more important to consider vulnerabilities in the context of social media recruitment carefully.

Particularly in Europe, privacy concerns and related legal uncertainty often cause researchers to refrain from using social media as a recruitment tool for clinical studies [9]. Yet, to not aggravate existing injustice, researchers and research ethics committees should not completely refrain from social media recruitment out of concern for vulnerability [24]. This should be especially considered in the context of historically underserved, hard-to-reach populations and people with stigmatized conditions, who may benefit from the platform's accessibility and reach. Consequently, we conclude from our findings neither promoting nor declining positions on the use of social media for clinical study recruitment. Instead, we highlight the importance of a context-sensitive assessment of study designs regarding the role of social media in triggering vulnerabilities and report on concrete ways researchers could develop appropriate safeguards tailored to address the specific types of vulnerability involved [2]. In developing such appropriate safeguards, assessing the risks and benefits of using social media recruitment regarding vulnerability is essential.

Limitations

The results of this study should be considered against the background of some methodological limitations. Due to limited access to patients during the COVID-19 pandemic, only 6 patients were recruited. While we found during the analysis that they represented a high variability in terms of age, digital literacy, and attitudes toward technology and social media, additional interviews might have revealed more nuanced patient views. Additionally, this study's focus on patients with hepatitis B introduces another limitation. The vulnerability and ethical considerations related to social media recruitment can differ substantially across various diseases and patient populations, thus the findings may not be directly applicable to other medical conditions. Yet, this study provides relevant insights for future studies investigating vulnerability in the context of social media recruitment in the context of other disorders.

Another potential limitation is that some of the interviews were held via the phone. The absence of visual cues might have hurt the richness and quality of the empirical data compared to face-to-face interviews, particularly for patient interviews. Due to the COVID-19 pandemic, it was not possible to conduct these interviews face-to-face.

Finally, throughout the interview process, we did not provide a specific definition of social media to the participants. Instead, we asked them about their understanding of social media. As a result, the practical recommendations presented in this paper are expected to have a more universal scope rather than being tailored to specific social media platforms. Thus, the provided recommendations may depend on the particular social media platform being considered, the target population, the kind of study being conducted, and the envisaged study design [8].

Conclusions

The use of social media for clinical study recruitment can mitigate but also trigger or aggravate existing vulnerabilities. To avoid the systematic neglect of certain groups in research studies, vulnerability should be anticipated in the study design, and ways to mitigate them should be defined upfront. To facilitate this, we have reported a range of practical recommendations to address vulnerability and presented a practical case to do this. As such, social media recruitment should be designed and reviewed in a way to mitigate effects that render people more vulnerable. Our expert participants proposed that studies targeting people with stigmatized conditions or historically disadvantaged populations should make sure that the recruitment design allows for transparent communication and protection of privacy. Expertise in analyzing potential algorithm bias and using multichannel recruitment strategies are other practical recommendations for certain target populations.

Acknowledgments

We thank Nina Goldman, PhD, for data management and contributing to the initial analysis. The research for this paper received funding from the European Union's Horizon 2020 research and innovation program under grant agreement number 848223 (TherVacB). This publication reflects only the authors' views, and the European Commission is not liable for any use that may be made of the information contained therein.

Data Availability

The datasets generated and analyzed during this study are not publicly available because of the German Data Protection Regulation. Anonymized excerpts are available from the corresponding author upon reasonable request.

Authors' Contributions

NM conceptualized this paper, analyzed and interpreted the data, wrote the initial draft, and implemented revisions. BMZ conducted interviews, conceptualized this paper, analyzed and interpreted the data, wrote the initial draft, and implemented revisions. TW conducted interviews, analyzed and interpreted the data, and critically revised this paper. AB acquired funding, conceptualized this study, discussed findings, and critically revised this paper. All authors read and approved the final paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Final list of codes.
[DOCX File, 22 KB - humanfactors_v11i1e52448_app1.docx]

References

1. WMA Declaration of Helsinki—ethical principles for medical research involving human subjects. World Medical Association. URL: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/ [accessed 2023-08-18]
2. Gordon BG. Vulnerability in research: basic ethical concepts and general approach to review. Ochsner J 2020;20(1):34-38. [doi: 10.31486/toj.19.0079] [Medline: 32284680]
3. Luna F. Elucidating the concept of vulnerability: layers not labels. IJFAB: Int J Feminist Approaches Bioethics 2009 Mar;2(1):121-139. [doi: 10.3138/ijfab.2.1.121]
4. International Ethical Guidelines for Health-Related Research Involving Humans: Council for International Organizations of Medical Sciences (CIOMS); 2016. [doi: 10.56759/rgxl7405]
5. Victor E, Luna F, Guidry-Grimes L, Reiheld A. Vulnerability in practice: peeling back the layers, avoiding triggers, and preventing cascading effects. Bioethics 2022 Jun;36(5):587-596. [doi: 10.1111/bioe.13023] [Medline: 35481605]
6. Russomanno J, Patterson JG, Jabson Tree JM. Social media recruitment of marginalized, hard-to-reach populations: development of recruitment and monitoring guidelines. JMIR Public Health Surveill 2019 Dec 2;5(4):e14886. [doi: 10.2196/14886] [Medline: 31789598]
7. Benedict C, Hahn AL, Diefenbach MA, Ford JS. Recruitment via social media: advantages and potential biases. Dig Health 2019;5:2055207619867223. [doi: 10.1177/2055207619867223] [Medline: 31431827]
8. Zimmermann BM, Willem T, Bredthauer CJ, Buyx A. Ethical issues in social media recruitment for clinical studies: ethical analysis and framework. J Med Internet Res 2022 May 3;24(5):e31231. [doi: 10.2196/31231] [Medline: 35503247]
9. Goldman N, Willem T, Buyx A, Zimmermann BM. Practical benefits, challenges, and recommendations on social media recruitment: multi-stakeholder interview study. J Med Internet Res 2023 May 22;25:e44587. [doi: 10.2196/44587] [Medline: 37213177]
10. Gelinas L, Pierce R, Winkler S, Cohen IG, Lynch HF, Bierer BE. Using social media as a research recruitment tool: ethical issues and recommendations. Am J Bioeth 2017 Mar;17(3):3-14. [doi: 10.1080/15265161.2016.1276644] [Medline: 28207365]
11. Luna F. Identifying and evaluating layers of vulnerability—a way forward. Dev World Bioeth 2019 Jun;19(2):86-95. [doi: 10.1111/dewb.12206] [Medline: 30058768]
12. Burns GS, Thompson AJ. Viral hepatitis B: clinical and epidemiological characteristics. Cold Spring Harb Perspect Med 2014 Oct 30;4(12):a024935. [doi: 10.1101/cshperspect.a024935] [Medline: 25359547]
13. Sheppard MK. Vulnerability, therapeutic misconception and informed consent: is there a need for special treatment of pregnant women in fetus-regarding clinical trials? J Med Ethics 2016 Feb;42(2):127-131. [doi: 10.1136/medethics-2014-102595] [Medline: 26585882]
14. Enescu A, Mitrut P, Balasoiu M, Turculeanu A, Enescu AS. Psychosocial issues in patients with chronic hepatitis B and C. Curr Health Sci J 2014;40(2):93-96. [doi: 10.12865/CHSJ.40.02.02] [Medline: 25729588]
15. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
16. Low J. A pragmatic definition of the concept of theoretical saturation. Sociol Focus 2019 Apr 3;52(2):131-139. [doi: 10.1080/00380237.2018.1544514]
17. Braun V, Clarke V. Toward good practice in thematic analysis: avoiding common problems and be(com)ing a knowing researcher. Int J Transgend Health 2023;24(1):1-6. [doi: 10.1080/26895269.2022.2129597] [Medline: 36713144]
18. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exerc Health 2019 Aug 8;11(4):589-597. [doi: 10.1080/2159676X.2019.1628806]
19. Matsuda Y, Brooks JL, Beeber LS. Guidelines for research recruitment of underserved populations (EERC). Appl Nurs Res 2016 Nov;32:164-170. [doi: 10.1016/j.apnr.2016.07.009] [Medline: 27969022]

20. Mühlhoff R, Willem T. Social media advertising for clinical studies: ethical and data protection implications of online targeting. Big Data Soc 2023 Jan;10(1):20539517231156127. [doi: 10.1177/20539517231156127]
21. Tilmes N. Disability, fairness, and algorithmic bias in AI recruitment. Ethics Inf Technol 2022 Jun;24(2):21. [doi: 10.1007/s10676-022-09633-2]
22. Fuchs C. Social media and the public sphere. TripleC 2014;12(1):57-101. [doi: 10.31269/vol12iss1pp57-101]
23. Hinds J, Williams EJ, Joinson AN. “It wouldn’t happen to me”: privacy concerns and perspectives following the Cambridge Analytica scandal. Int J Hum Comput Stud 2020 Nov;143:102498. [doi: 10.1016/j.ijhcs.2020.102498]
24. Racine E, Bracken-Roche D. Enriching the concept of vulnerability in research ethics: an integrative and functional account. Bioethics 2019 Jan;33(1):19-34. [doi: 10.1111/bioe.12471] [Medline: 30136737]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research

Edited by A Kushniruk; submitted 04.10.23; peer-reviewed by A Hassan, E Vashishtha, E Malbois; revised version received 22.08.24; accepted 05.09.24; published 30.12.24.

Please cite as:
Matthes N, Willem T, Buyx A, Zimmermann BM
Social Media Recruitment as a Potential Trigger for Vulnerability: Multistakeholder Interview Study
JMIR Hum Factors 2024;11:e52448
URL: https://humanfactors.jmir.org/2024/1/e52448
doi:10.2196/52448

The Promise of AI for Image-Driven Medicine: Qualitative Interview Study of Radiologists' and Pathologists' Perspectives

Jojanneke Drogt[1], MA, MSc; Megan Milota[1], PhD; Wouter Veldhuis[1], MD, PhD; Shoko Vos[2], MD, PhD; Karin Jongsma[1], PhD

[1]University Medical Center Utrecht (UMC Utrecht), Heidelberglaan 100, Utrecht, Netherlands

[2]Radboud University Medical Center, Nijmegen, Netherlands

Corresponding Author:
Jojanneke Drogt, MA, MSc
University Medical Center Utrecht (UMC Utrecht), , Heidelberglaan 100, Utrecht, , Netherlands

Abstract

Background: Image-driven specialisms such as radiology and pathology are at the forefront of medical artificial intelligence (AI) innovation. Many believe that AI will lead to significant shifts in professional roles, so it is vital to investigate how professionals view the pending changes that AI innovation will initiate and incorporate their views in ongoing AI developments.

Objective: Our study aimed to gain insights into the perspectives and wishes of radiologists and pathologists regarding the promise of AI.

Methods: We have conducted the first qualitative interview study investigating the perspectives of both radiologists and pathologists regarding the integration of AI in their fields. The study design is in accordance with the consolidated criteria for reporting qualitative research (COREQ).

Results: In total, 21 participants were interviewed for this study (7 pathologists, 10 radiologists, and 4 computer scientists). The interviews revealed a diverse range of perspectives on the impact of AI. Respondents discussed various task-specific benefits of AI; yet, both pathologists and radiologists agreed that AI had yet to live up to its hype. Overall, our study shows that AI could facilitate welcome changes in the workflows of image-driven professionals and eventually lead to better quality of care. At the same time, these professionals also admitted that many hopes and expectations for AI were unlikely to become a reality in the next decade.

Conclusions: This study points to the importance of maintaining a "healthy skepticism" on the promise of AI in imaging specialisms and argues for more structural and inclusive discussions about whether AI is the right technology to solve current problems encountered in daily clinical practice.

(JMIR Hum Factors 2024;11:e52514) doi:10.2196/52514

KEYWORDS

digital medicine; computer vision; medical AI; image-driven specialisms; qualitative interview study; digital health ethics; artificial intelligence; AI; imaging; imaging informatics; radiology; pathology

Introduction

Image-driven specialisms such as radiology and pathology are at the forefront of technological innovation in medicine, and many believe that artificial intelligence (AI) is the next innovation to reshape these fields [1-3]. AI refers to a broad range of machine-based systems designed to influence the environment by producing an output (predictions, recommendations, or decisions) for a given set of objectives [4]. AI is considered promising for image-driven medical fields because the work involves pattern recognition and is often digitalized, meaning rich datasets are available for AI training. Some have already argued that the professional roles of radiologists and pathologists will drastically change due to AI; they will become "information specialists" [5] or "imaging consultants" [6] who seamlessly use AI to help interpret patient data. Jha and Topol [5] even speculate that the fields will most likely merge, leading to "a natural fusion of human talent and artificial intelligence. United, radiologists and pathologists can thrive with the rise of artificial intelligence."

Despite the great promises for image-driven diagnostics and Dr Geoffrey Hinton's prediction that radiology as a specialization would now be extinct, the implementation of AI in routine patient care is often lagging [7,8]. One cause is the lingering uncertainty among professionals about the added value for clinical practice. Another contributing factor is the large variance in acceptance and trust of direct and indirect adopters [9]. While fears about an upcoming "AI winter" [10] are likely unfounded, expectations must be tempered to prevent disillusionment. It is therefore relevant to consider "how to actually deploy AI in clinical practice" and investigate whether the high expectations of AI in radiology and pathology require substantial changes in these fields—and in the current implementation approaches used by AI vendors [7].

Empirical studies have investigated how image-driven professionals view AI innovations. For example, professionals in radiology [9,11-14] and pathology [15-17] have a wide range of predominantly positive expectations for AI, yet they remain divided on the roles AI should have in their daily workflows. Studies have also called for a more thorough incorporation of medical professionals' views in AI design and implementation [12,18]. This paper aims to add to the understanding of image-driven professionals' views on the future of AI in radiology and pathology by highlighting how their views relate to current discussions on AI. As far as we are aware, this is the first qualitative interview study to combine views from both fields. By doing so, we hope to provide a more comprehensive perspective on AI's influence on medical imaging. These insights are also intended to help inform the responsible integration of AI in image-driven medicine.

Methods

Overview

This study is part of a broader research project focusing on the ethical integration of AI in image-driven medicine, and the main research question is, "how should AI be responsibly integrated and used in image-driven medicine?" In order to answer this question, we use empirical research methods such as qualitative interviews and participant observations [19] to ground and inform our ethical analysis. The interview study design is in accordance with the consolidated criteria for reporting qualitative research (COREQ) [20]. In another publication, we reported the perspectives of pathologists, laboratory technicians, and computer scientists from 2 Dutch hospitals regarding the development and implementation of AI in pathology [15]. For the previous paper, we focused on the perceived roles and responsibilities of AI according to professionals working in pathology. In this paper, we compare and contrast the perspectives of professionals working in 2 departments—radiology and pathology—within 1 Dutch hospital and focus on the perceived promise of AI for image-driven medicine.

Research Design

To gain insight into the promise of AI for image-driven medicine, we conducted an inductive qualitative analysis of recorded conversations with radiologists, pathologists, and computer scientists [21-24]. The interviews with computer scientists have been used to contextualize and steer the interpretation of our findings.

Sampling in a High-Resource Context

Radiologists and pathologists working at 1 academic hospital in the Netherlands, the University Medical Center Utrecht (UMCU), were invited to participate in this study via a department-wide call. Potential participants were also directly approached by the research team or a contact person at the department to reach a representative group of professionals. We personally approached radiologists and pathologists who were less involved in AI integration in these fields because we found it important to include their perspectives in the study. Computer scientists working with these departments were also asked to participate to provide additional context. For several reasons, we chose to focus on professionals from 1 innovation-driven medical center. First, these departments are relatively far along in their AI implementation processes compared with other Dutch hospitals. We hypothesized that this would correlate with a greater familiarity with AI, meaning respondents would be more likely to relate their opinions and expectations to practical encounters with AI. Second, we also recognize that context matters for AI integration and that it can be challenging to compare different medical contexts [25]. Focusing on radiologists and pathologists from 1 medical center enabled the comparison of perspectives on AI innovations between the departments, as, in general, the 2 departments function in the same context (eg, same region, managerial structures, and access to high-quality data), and both have access to internal computer science teams to support AI development and use. Nevertheless, we also recognized that conducting the study at 1 medical center would present a practical challenge. As there were a fixed number of radiologists and pathologists working at these sites, and we were dependent on their willingness to participate, the number of respondents for this study was finite. As our primary aim was to conduct a comprehensive exploration of perspectives on the promise of AI for image-driven medicine, we have focused on including a range of perspectives present in the departments to ensure broad representation instead of purely focusing on the sample size. This means we have taken meaning saturation into account in the analysis of the data to ensure that the quality of data is high and that the elicited views are representative of the perspectives present in the departments (except perhaps for those respondents who remained unwilling to discuss the potential of AI; see the Discussion section), but we mainly reflected on the information power in our in-depth interviews [26,27].

Data Collection and Analysis

Interviews were conducted between June 2020 and December 2021. Because of the pandemic, many of the conversations took place via the telephone; JD and MM conducted interviews individually and as a team. A semistructured topic list was used to guide the conversations (see Table 1 for sample questions). The recorded interviews were transcribed verbatim by a professional transcription service and checked for reliability by JD. The transcripts were then coded for confidentiality, and identifying information was removed. The interviews were conducted in Dutch and translated to English by JD and MM.

Table . Sample interview questions of semistructured topic list.'

Designators	Questions
A. General questions about respondent's background	
A1	How long have you worked as a radiologist/pathologist?
A2	Why did you choose to specialize in the field?
A3	What technology developments did you encounter during the time you have worked as a radiologist/pathologist?
B. Question(s) on conceptualization of AI [a]	
B1	In your view, how would you define AI?
C. Questions about respondent's thoughts and opinions about AI: ***What is the perceived effect of the introduction of AI in relation to ideas about professional identity and expertise?***	
C1	In general, what do you think AI could mean for radiology/pathology?
C2	To which extent are you involved in AI integration in your field?
C3	What kinds of AI applications would be most helpful or useful to you?
C4	In what ways do you think AI might impact your decision-making process?
C5	What (new) skill(s) or knowledge do you foresee yourself needing if AI becomes more prevalent?
C6	In the next 10 years, how do you expect AI to impact radiology/pathology?
D. Questions about desirable ethical guidance for AI in image-driven medicine	
D1	What ethical issues do you foresee with the increased use of AI in your work?
D2	Do you think special guidelines should be established for using AI? If so, what kinds of issues should be addressed?
E. Exit questions	
E1	Do you have any other thoughts or opinions about the use of AI in your department that you'd like to share with me?
E2	Is there anything you think we missed? Is there an important question I forgot to ask?

[a]AI: artificial intelligence.

The data selection and analysis occurred inductively and iteratively [28] using constant comparison [29]. The software program NVivo (version 12; Lumivero) supported the data analysis. JD and MM read individual interview transcripts and independently identified conversation fragments or units of meaning [21-24] they considered relevant to the research question; they met regularly to compare their observations. They used the code tree from the analysis for the earlier publication [15] as a baseline, adapted the code tree to fit the new dataset, and supplemented it with new descriptive categories. JD and MM then sampled and independently coded 4 interviews, compared the results, and refined the code tree. JD then coded the remaining transcripts, adjusting the code tree when necessary. Finally, MM and JD performed an intercoder reliability check by recoding 2 interviews (1 pathologist and 1 radiologist) and comparing their results. Meaning saturation and information power were taken into account throughout this process [27,30]. During the analysis, JD, MM, and KJ kept track of new AI developments in radiology and pathology; in consultation with WV and SV, we evaluated the relevance of the data to current situations on the work floor and included current literature in the discussion.

Data Statement

The data have been presented by means of in-text illustrative quotes, carefully selected to represent the arguments presented in the interviews and do justice to the variety of perspectives captured in the interviews. We have considered whether the quotes could be understood without the context in which they were originally uttered. The complete datasets are not publicly available because privacy of individual participants could be compromised. The individual privacy of the participants was particularly important as their statements included political opinions and philosophical beliefs regarding the ways in which AI should be adopted. These are deemed sensitive and, therefore, fall under the protection of the General Data Protection Regulation (GDPR: article 9).

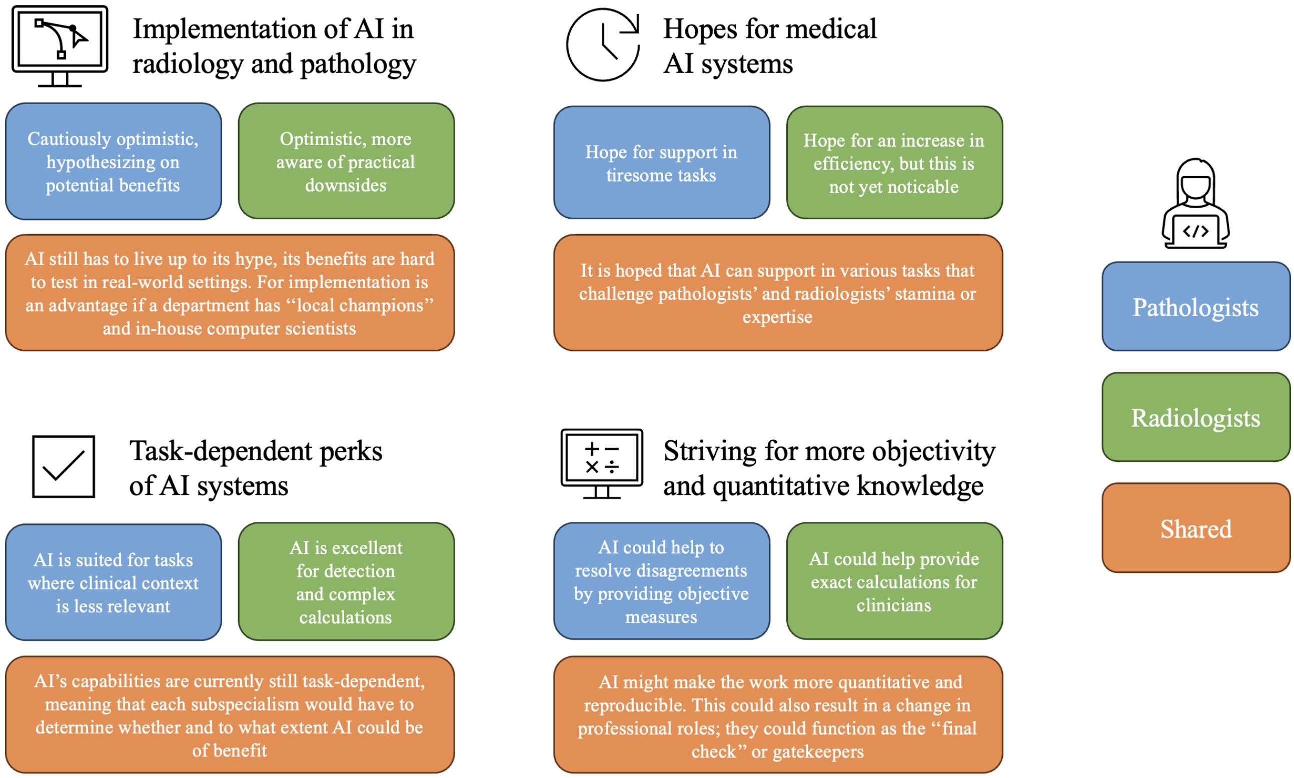
Ethical Considerations

This study constitutes part of the Responsible Artificial Intelligence in clinical DecisIOn making (RAIDIO) study. Ethical approval for the RAIDIO study was obtained from the Medical Research Ethics Committee of the UMCU (WAG/mb/20/014090). The Medical Research Ethics Committee determined that this study was exempt from the Medical Research Involving Human Subjects Act. Written or oral informed consent was obtained from all participating respondents. Data were deidentified through pseudonymization and stored in a protected digital environment of the UMCU. Participants of this study did not receive financial compensation.

Results

In total, 21 participants (7 pathologists, 10 radiologists, and 4 computer scientists) agreed to be interviewed, provided written or oral informed consent, and were included in this study. The following sections present how participants perceived current AI developments and AI's promise for image-driven medicine; we pay special attention to similarities and differences between radiologists and pathologists (for an overview of respondents' perspectives, see Figure 1).

Figure 1. General overview of respondents' perspectives. AI: artificial intelligence.



Implementing AI in Radiology and Pathology

Respondents from both fields considered AI a novel technology. The extent to which participants could elaborate on technical or development issues of AI in their respective fields depended on their familiarity and previous experiences with these technologies. Nevertheless, it was striking that *all* respondents could refer to a landmark AI system as a point of reference; for the pathologists, this was a mitosis counting algorithm, and for the radiologists, it was a pulmonary embolism detection algorithm. To some extent, these 2 systems shaped the ways respondents envisioned future AI integrations in their departments. Because of the initial success of the mitosis counting algorithm, pathologists were cautiously optimistic about AI in their field. Many expected that other "simple" tasks could be supported or performed by an AI system, and some considered it a matter of time before applications for more complex tasks would be developed. Most radiologists were also optimistic about AI systems, but multiple respondents referred to minor flaws in the pulmonary embolism detection algorithm when considering possible future AI applications. For example, although the accuracy of the detection tool was very high, they still had to actively verify the algorithm's outcomes. Some also considered the notifications of possible pulmonary embolisms a disturbance to their workflow.

Both pathologists and radiologists, independent of their respective familiarity with AI, admitted that medical AI in practice had yet to live up to its hype. Besides the fact that it is technically complicated to integrate multiple AI systems into the workflow of professionals, another roadblock respondents mentioned was the difficulty in getting an AI system tested in real-world settings. As a pathologist (P7) described:

> *testing [the AI system] in the real world is something that is not standard practice, and you encounter all kinds of problems when you try to do it (…) [AI] functions suboptimally when employed outside the research setting.*

Respondents mentioned that they were lucky to have computer scientists working in or with their departments who were available to help problem solve when integrating new AI systems into their daily practice. Respondents were aware that the close collaboration with data scientists made their departments unique. They considered their departments a “frontrunner” (P7) and example for other hospitals in regard to AI implementation. Both departments also had “local champions,” pathologists, or radiologists who were exceptionally knowledgeable about AI and could “speak the language” of medical specialists and computer scientists. These “local champions” helped accelerate the adoption of AI. As a radiologist (R8) described, AI implementation is site-specific:

> *In our hospital, some people are very invested in this topic. We probably encounter more changes than other hospitals. It might go relatively quickly at our department.*

For these reasons, the pathology and radiology departments of the UMCU may have a head start in working toward (more elaborate) AI implementation and are at a point in which they must decide how to proceed. In the following sections, we describe how pathologists and radiologists articulated the changes they were experiencing and how they envisioned AI’s impact on the future of their specialties.

Hopes for Medical AI

Most radiologists and pathologists in this study argued that AI could help improve medical care by supporting tasks that challenged their stamina or expertise. This is illustrated by an overview of AI systems that were present in the departments and which were referred to in the interviews (Table 2). These systems fall into one of three categories: (1) fixes for tiresome or time-consuming tasks, (2) support in cases where context and classification are challenging according to current standards, and (3) prognosis and therapeutic response (prediction) generators.

Table . Overview of AI[a] systems at the radiology and pathology departments[b].

Field	Specific medical task	Potential role of AI algorithms	Relevant to which (sub)specialisms	Benefit mentioned by pathologists and radiologists	Level of risk involved	Stage of development
Pathology	Determining the aggressiveness of a tumor	Counting the number of mitoses on a digital slide Calculating the percentage of Ki67 positive tumor cells (proliferation index)	Pathology, medical oncology, pulmonary medicine, endocrinology, etc	Tiresome task, less subjectivity	Low—can be checked manually	Implemented
Pathology	Grading cancer; for instance, grading of breast and prostate cancer	For example, Bloom and Richardson grading score for breast cancer	Pathology, medical oncology, urology, etc	Tiresome task, less subjectivity	Low—can be checked manually	In development
Pathology	Analyzing the inflammatory response	Identifying as well as quantitative measurement of number and distribution of immune cells, for example, in/around tumors	Pathology, medical oncology, internal medicine, etc	Tiresome task, less subjectivity	Low—can be checked manually	In development
Pathology	Deciding whether to proceed to surgical (vs endoscopic) resection in case of early colon carcinoma	Analysis of tumor characteristics related to the patient's chance of developing (future) metastases of colon carcinoma	Pathology, gastroenterology, and surgery	Less invasive treatment for the patient, personalized medicine	High—not easy to check	Research phase
Pathology	Determining prognosis and treatment options for patients with cancer	Analysis of prognostic and (treatment) predictive tumor characteristics	Pathology, medical oncology, pulmonary medicine, etc	Less subjectivity, personalized medicine	High—not easy to check	Research phase
Pathology	Analyzing naevi and other melanocytic lesions on signs of malignancy	Analysis of characteristics associated with malignancy, providing reasons for why the sample is malignant or not, or calculate the risk of malignancy	Pathology, dermatology, and medical oncology	Less uncertainty about the diagnosis, learning from the algorithm	High—not easy to check	Research phase
Pathology	Generating the pathology report	Generating an initial pathology report for a pathologist to check	Pathology	Tiresome task, less variation in reporting style between pathologists	Low—can be checked and changed manually	Research phase
Pathology	Checking images on possible metastases in lymph nodes	Initial screening of lymph nodes on possible metastases	Pathology	Tiresome task	Low—can be checked	Research phase
Radiology	Confirm/rule out pulmonary embolism	Detect/rule out suspected pulmonary embolisms on dedicated CTPA[c] scans	Radiology, internal medicine, and cardiology	Faster diagnostic process in case of confirmed high accuracy	Low in terms of patient risk: dedicated CTPA are always checked for the primary rule out PE[d] question. But high level of trust required for benefit to be realized	Implemented

Field	Specific medical task	Potential role of AI algorithms	Relevant to which (sub)specialisms	Benefit mentioned by pathologists and radiologists	Level of risk involved	Stage of development
Radiology	Detect incidental pulmonary embolism	Detect incidental pulmonary embolism on CT[e] scans made for other indications	Radiology, oncology, trauma, internal medicine, cardiology, etc	Earlier detection of unsuspected PE in nonprioritized scans + increased detection rate of unsuspected incidental PE in general	Low—is primarily added value of current standard of care	Implemented
Radiology	Measure prostate in 3D and manually calculate both corresponding prostate volume estimate and its ratio with the plasma PSA[f] value to determine PSA density correlated with risk of prostate cancer being present	Measure actual prostate volume in 3D (+ provide PSA density)	Radiology, urology, oncology, and radiotherapy	Tedious and repetitive task	Low—volume calculation performed by AI and the corresponding segmentation on which the calculation depends is easily visually checked by the radiologist	Implemented
Radiology	Determine age of pediatric patient on the basis of hand x-ray	Independently perform the bone-age determination	Radiologists and pediatricians	Fully automated procedure	High—bone-age independently determined by AI—with only a visual check of the correctness of joint segmentation by the radiologist	Implemented
Radiology	Detecting and measuring lung nodules	Detecting lung nodules including quantitative 3D volumetry	Radiologists, pulmonologists, oncologists, etc	Tedious, repetitive task, and possible reduction in number of missed nodules	Intermediate—aids in detection, volume calculation more quantitative than radiologist, but correlation with prior scans (crucial for determining growths over time) still lacking in reliability and intuitiveness	Implemented
Radiology	Detecting cervical spine fractures	Detecting fractures in cervical vertebral bodies on CT scans that include the neck	Radiologists, trauma surgeons, orthopedic surgeons	Quicker diagnostic process—theoretically reduced number of missed fractures	Low—always checked	Implemented
Radiology	Quantify cerebral white matter disease	Quantifying the volume of white matter lesions on MRI[g] of the brain	Radiologists, neurologists	More quantitative and more reproducible measurements, including individualized comparison to reference standard	Low—correctness is easily and reliably visually verified	Research phase
Radiology	Working toward body composition–derived prognostication and personalized treatment	Quantifying the volume of multiple different muscle groups and of subcutaneous and visceral fat	Radiologists, any clinical profession ordering CT scans containing the abdomen	Impossible to perform by radiologists (far too time-consuming, would be hours of work per scan)	Low—with respect to the correctness of segmentations (important to understand that the prognostic application is not part of the AI output)	Actively used in research setting in clinical trials

Field	Specific medical task	Potential role of AI algorithms	Relevant to which (sub)specialisms	Benefit mentioned by pathologists and radiologists	Level of risk involved	Stage of development
Radiology	Working toward body composition for personalized drug dosing; from contrast agents to chemotherapeutics	Quantifying the volume of multiple different muscle groups and of subcutaneous and visceral fat	Radiologists, any clinical profession ordering CT scans containing the abdomen	Impossible to perform by radiologists (far too time-consuming, would be hours of work per scan)	High—while the segmentation is reliably verified, subsequent drug dose calculations require extensive validation	In development
Radiology	Segmenting the liver: both the organ and its internal liver segments for subsequent clinical and treatment calculation that depend on liver/segment volumetry	Segmentation of liver and liver segments	Radiologists, interventional radiologists, nuclear medicine, HPB[h] surgery, oncology	Time-consuming, tedious, task-automated, and made more reproducible	Intermediate—easily, visually checked, still requires some manual corrections	In development
Radiology	Working toward body composition for creatinine clearance calculations that are both more personalized and do not require 24-hour urine samples	Quantifying the volume of multiple different muscle groups and of subcutaneous and visceral fat	Radiologists, any clinical profession ordering CT scans containing the abdomen	Impossible to perform by radiologists (far too time-consuming, would be hours of work per scan)	High—while the segmentation is reliably verified, subsequent drug dose calculations require extensive validation	In development
Radiology	Deciding whether a patient undergoing a breast MRI for detection of breast cancer needs additional imaging or can exit the scanner	Triage of patients with and without possible breast cancer based upon the initial phases of enhancement directly after contrast injection	Radiologist, MRI technologist	Reducing examination time from ±25 minutes to ±5 minutes	High—not easy to check	In development

[a]AI: artificial intelligence.

[b]Check for completeness by radiologists and pathologists working at the department.

[c]CTPA: computed tomography pulmonary angiogram.

[d]PE: pulmonary embolism.

[e]CT: computed tomographic.

[f]PSA: prostate-specific antigen density.

[g]MRI: magnetic resonance imaging.

[h]HPB: hepato-pancreato-biliary surgery.

The hope that AI can support or take over time-consuming tasks was especially prominent in the interviews. Many respondents were concerned about the increased work pressure, as clinicians often depend on radiologists' and pathologists' knowledge to diagnose and treat patients. One respondent (R7) described their relationship with clinicians as follows:

> *There is almost no patient who (…) is treated without scans. We are constantly discussing patients with [other] medical disciplines. These meetings cost a terrible amount of time. Everyone wants you to be at their beck and call.*

Respondents appreciated that their fields were seen as essential to the medical system and that their perspectives were valued. Still, many worried about the workload and the limited time to assess cases, write reports, and prepare for multidisciplinary meetings. Pathologists and radiologists were optimistic about AI's future role in time-consuming activities, such as tissue or tumor segmentations, calculation of abnormalities such as deviations in heart function, and detection of the evolution of brain metastases or the presence of tumor cells in lymph nodes. In other words, by supporting these kinds of tasks, AI could help them refocus on the more "enjoyable" aspects of the job, such as diagnosing complex disease patterns. As a radiologist (R9) mentioned:

> *It would be fantastic if part of our routine work (…) could be taken over. I hope that this will be possible in the future, so we radiologists can again focus on the fun things.*

Nevertheless, some also worried that AI applications would not increase efficiency and might even cost them extra time; as one respondent stated:

> *In the meantime, the amount of scans increases, and I'll also have to manage the AI. That's something to think about. Eventually, we'll just be doing our jobs. But hopefully, the quality will become a little bit better. [R10]*

Respondents thus speculated about the impact of AI on the work pressure they experienced. This was particularly apparent for radiologists who noticed that the pulmonary embolism algorithm resulted in a quicker diagnostic process. Using the algorithm also meant that they had to recheck a patient's images when notified of a possible embolism.

Task-Dependent Perks

Notably, radiologists and pathologists put their hopes of AI in perspective by remarking that it is task-dependent, meaning that each subspecialism would have to determine whether and to what extent AI could benefit their work. Besides naming technical hurdles, many respondents noted that the amount of input needed to make a diagnosis or prognosis would likely determine whether AI would be suitable for their diagnostic process. For instance, respondents did not expect AI to be able to make complex integrations between different sources of knowledge or to prioritize information. As R7 argued, work in the field requires:

> *Integrating everything you've learned in your medical education and training as a radiologist, (...) I sometimes wonder how AI could help me with this. I think it will be useless on this front. AI (...) doesn't know how to search through old reports and gather the relevant information for my scans. I think this will remain—as I currently see it—a skill particular to medical experts.*

As the quote illustrates, respondents questioned whether AI would be useful when a radiologist or a pathologist had to determine which information was relevant for interpreting a medical image, a common practice in all but the most straightforward cases. Although some respondents mentioned that AI could provide a differential diagnosis based on context-related information (eg, age, gender, or laboratory results), many doubted whether AI could prioritize or "weigh" this information in the same way they did. As one pathologist (P1) commented:

> *I think that context certainly matters. AI could go wrong because it insufficiently weighs the context. (...) Plus I also think a lot of histological images look similar. But the clinical context of one patient may be very different from another patient and will result in another diagnosis, even when the images look completely the same. (...) One histological image could indicate 20 different clinical diagnoses, especially if you're looking at images of inflammatory disease. This won't be easy for AI.*

Radiologists and pathologists often viewed "real interpretation" (R3) as something exceeding the capabilities of AI. They mentioned that AI could be good at detecting certain things (such as lung nodules or other conspicuous manifestations of cancer) and might even gain an "associative capacity" (R1) similar to their own, but that it would increase their workload if adopted in areas where they did not need it. Some respondents had not ruled out the possibility of AI becoming better at specific, well-defined "expert tasks" and found it an exciting thought that AI could become more competent than humans in the interpretation process. Nevertheless, respondents also stated that it would be hard for an algorithm to learn to independently evaluate pathological processes with respect to the clinical context, making many of the potential uses for AI more speculatory than an inevitability.

Striving for More Objectivity and Quantitative Knowledge

Although there were varying views on the tasks best suited for AI, many radiologists and pathologists stated that AI could improve the quality of their work. This was often mentioned with the expectation that AI might make the work less "subjective" and more quantitative and reproducible. For instance, pathologists, in particular, talked about the possibility of AI (sometimes referred to as "the computer") helping resolve disagreements in their fields by offering an additional, objective interpretation of medical images. As one pathologist (P5) stated:

> *If you have a tumor cell with a nucleus that is a little bit enlarged, one pathologist could say something like "alright, it's probably reactive," and the other pathologist says "oh no, it is malignant." But a computer could precisely measure the nucleus and determine "Okay, there is a lot of chromatin, it is irregular, this is the intensity of the chromatin." These are all objective measures by which you could say whether it is benign, malignant or reactive.*

Multiple respondents also discussed a potential beneficial characteristic of AI, namely, that it could keep track of minute details in medical images. It might, therefore, become better than humans at recalling and comparing image characteristics.

AI was also described as a tool to help radiologists and pathologists better understand the data by quantitively measuring multiple aspects of medical conditions. AI systems have already been designed to compute a patient's fat and muscle mass, the amount of white matter in the brain, the volumes of various parts of the brain, and the histological parameters of a tumor. Some participants mentioned that these AI applications mirror a broader trend in radiology and pathology to approach medical findings in a more quantitative manner. A radiologist (R8) described this as an ongoing shift in the way medical images are used in the field, adding: "It's not just about the interpretation of images, but also the generation of scores and the production of numbers." Some radiologists also referred to clinicians' wishes that they provide exact calculations. One radiologist (R5) even called this "the ultimate goal" of their specialism: to precisely identify a patient's condition for the clinician. Many respondents mentioned that such quantitative measures might also lead to more reliable and precise prognoses by giving the clinician more relevant information to determine a patient's treatment.

Both pathologists and radiologists also reflected broadly on how AI might change how they form medical judgments. While some radiologists imagined that they would become “data specialists” (R8) or “translators” (R4) who would mainly check the algorithms’ reports, most radiologists and pathologists were inclined to describe themselves as the “final check” or gatekeeper. In other words, they were comfortable letting AI do some of the primary work but wanted the medical specialist to make the final judgment and bear the responsibility for the diagnosis. We observed some slight differences between radiologists and pathologists regarding the role of AI in their specialism. Radiologists seemed more inclined to describe specialism-wide changes initiated by AI and viewed AI as a more significant force that could become an integral part of their specialization. Pathologists primarily focused on AI as an innovation from which they could learn. Respondents from both fields indicated that they were unsure of the ultimate impact AI would have on their specialisms, and when asked for their expectations for the coming 10 years, most replied that they did not expect any fundamental changes to their professional roles or responsibilities.

Discussion

Principal Findings

This qualitative interview study investigated how professionals from the 2 most image-driven medical specialisms perceive the promise of AI for their respective fields. Overall, our analysis shows that pathologists and radiologists have comparable views on AI’s possible benefits and drawbacks. Differences between radiologists’ and pathologists’ perspectives were mostly a level of degree; for instance, the use of AI for quantification purposes seems to be somewhat more pronounced in radiology. One reason for this discrepancy might be that the radiology department in our study currently has more experience with implementing AI systems in practice.

The radiologists and pathologists in our study echoed some of the findings of earlier empirical studies concerning the potential of AI in these fields. Respondents in this study also argued that AI could provide them with quantitative data [11] and were interested in AI systems that could perform simple yet time-consuming and repetitive tasks [12]. However, they also worried that AI could result in more work [11] and hypothesized that AI would be less suitable for complex, variable, or intellectually challenging tasks [31]. Respondents from both disciplines (irrespective of their experience with AI) also cautioned about overstating the benefits of AI and tried to shift the focus to task-specific advantages. This resembles the results of the study by Hendrix et al [14], where respondents emphasized that AI-based decision support is contingent on its specific features and functionality. In our study, most respondents had a positive yet realistic view of AI, keeping in mind the current limits of AI and roadblocks for successful implementation.

As our findings reflect the unique combination of pathologists’ and radiologists’ perspectives from a technologically innovative academic medical center, the interview data can indicate how to proceed with the implementation of AI. In the following sections, we discuss the implications of our findings in relation to broader questions about AI integration within image-driven medicine.

Will AI Reduce or Increase the Workload of Image-Driven Professionals?

Radiologists and pathologists in this study often mentioned that their workload had expanded over the last decades and that they increasingly participated in multidisciplinary meetings. Therefore, many respondents expressed the hope that AI would help them tackle their demanding workloads. This is consistent with other studies and literature, which point to the possibility of designing AI for the most tiresome and repetitive tasks in radiology and pathology [32-35]. Both radiologists and pathologists in this study mentioned that AI had already been developed for several time-consuming tasks in their departments, and some also hoped that AI would someday help them write their reports.

At the same time, many respondents questioned AI’s ability to contribute to increased efficiency. Many studies confirm that AI should not be considered an augmentation or support tool, not a direct replacement for pathologists or radiologists [36,37]. AI involvement would also result in new tasks for professionals, such as validating AI systems and checking outcomes. Professionals would also have to become more skilled in dealing with AI in their daily work. The amount of extra effort it costs to work with AI highly depends on the specific task and the trust radiologists and pathologists have in the algorithm’s functioning. This was illustrated by respondents’ emphasis that they wanted to remain involved in the final medical conclusion. A similar argument can be found, for example, in the study by Ranjan et al [37]. Literature on the successful adoption of AI in clinical workflows often stresses that physicians should have epistemic trust in AI functioning, adding that many open questions still exist on the level of control physicians should have over AI [38] and which kinds of outcomes physicians should trust [39,40].

The results of this study have highlighted the dichotomous role AI could play in the high workload of professionals working in image-driven fields; they also point to the importance of contemplating the amount of work AI could and should impose. Although AI could create welcome changes in the workflows of these professionals, it also has the potential to become another technology for them to manage and may not always be a legitimate aid to their already busy schedules.

What Will AI Mean for the Future of Radiology and Pathology?

Many authors contend that AI could lead to significant changes in the professional roles of radiologists and pathologists [6,41-43], and some have even argued that the fields will eventually merge to become the “information specialists” of the medical system [5]. In this study, participants shared the belief that AI could greatly impact how they perform their work and could change their professional roles. At the same time, they emphasized that many of these changes were speculative and unlikely to occur soon.

Because of the speculative nature of the grander promises of AI, Saboury et al [44] argue that “it is critical to improve our understanding of the pitfalls of deep learning and maintain a healthy and constructive skepticism as we explore the tremendous potential of the technology." Karhade and Schwab [45] also state that this kind of "healthy skepticism"—along with engagement and collaboration with technical experts—can support "the development of AI systems that complement and expand our abilities to diagnose, predict and operate," help sustain informed dialogue, and ask the right questions concerning the use of AI in clinical practice. Therefore, it may be essential to focus on the actual impact AI can have on radiology and pathology and maintain a skeptical attitude in order to ultimately maximize the advantages of AI. For now, this could also mean focusing on AI's task- and specialism-dependent benefits rather than its broader potential for integrating multiple medical specialisms—even though bridging disciplinary boundaries between radiology, pathology, and other medical fields may eventually benefit the quality of care [46,47].

How Can We Incorporate Critical Voices in AI Innovation?

There is currently a push toward AI in image-driven diagnostics, illustrated by assertions such as "radiologists who use AI will replace radiologists who don't." [48,49] Yet, the question is, who are the radiologists (and pathologists) who do not want to use AI? Who is going to be replaced? The positive voices about AI still outweigh the more critical voices in existing qualitative interview studies [50,51], and it is hard to find medical professionals who contest the possible advantages of using AI in image-driven medicine. Although those who refuse to work with AI altogether may be a relatively small group, we noticed in our recruitment process that professionals who were less convinced of the benefits of AI or were working in subspecialisms less suited to AI were more reluctant to participate in our study than individuals who were already involved in the validation and implementation of AI [9]. We successfully recruited some individuals with skeptical views but were unable to include those few radiologists and pathologists who remained unwilling to consider the potential of AI for radiology and pathology.

While not everyone has to participate in the debate about medical AI, it is important to be aware of the possibility of perpetuating existing bias in empirical studies about AI. Concerns about the issues that could arise by using AI (such as deskilling and the effects of changing practice patterns on AI) [52] persist; we, therefore, urge radiology and pathology departments to create ways to include critical voices in the development of AI in their fields. Accelerating AI integration could force some professionals to use it even when they believe their field "is not ready for AI" [53]. Ideally, consideration should be paid to how *all* users respond to and can accept the involvement of AI in their workflows. As Krupinski [54] formulates:

> *Technology development and deployment are critical to improve patient care, health outcomes, and the efficacy and efficiency with which our health care systems achieve these goals, but it cannot take place without considering how it will be accepted and integrated in routine daily use by all stakeholders.*

Although there may be practical roadblocks to ensuring all voices are represented, inclusive communication will help ensure that more specialists are familiar with specific AI systems; this will also ease the transition to using AI in their workflows. A broad representation of perspectives could also benefit developers by supporting them in detecting blind spots in the design and implementation of medical AI and might facilitate trust in the development process.

What Could Future Research on AI in Image-Driven Specialisms Focus on?

Besides the importance of maintaining healthy skepticism and focusing on the inclusion of critical voices, this study offers additional recommendations for future research. Future research could, for instance, repeat this study when (both) departments are further along in integrating AI into their workflows. Our study was limited in the sense that, although we selected departments that were relatively far along in implementing AI, the integration of AI in health care is generally still in its early phases. This meant that some questions were answered hypothetically. We expect that perspectives will become more concrete when AI becomes more thoroughly implemented into these specialisms. Another consideration for future research is that it is unclear whether the perspectives mentioned here would also apply to the implementation of AI in low-resource settings. We consciously focused on the integration of AI in 2 high-resource departments, which made the perspectives on AI between these departments more comparable and likely also with other high-resource settings. Yet, it is essential to state that the results of this study should not be taken at face value for low-resource deployment environments [25]. Further research is necessary to determine the extent to which the perspectives presented here are also mirrored in low-resource sites.

Acknowledgments

This study was funded by the Dutch Research Council (Nederlandse Organisatie voor Wetenschappelijk Onderzoek, NWO; project 406.DI.19.089). We extend our gratitude to our fellow RAIDIO project members, Sally Wyatt and Flora Lysen, for the valuable collaboration and insights that significantly supported writing this article.

Data Availability

The datasets used and analyzed during this study are unavailable, as they are protected under the General Data Protection Regulation (GDPR: article 9).

Conflicts of Interest

None declared.

References

1. Chauhan C, Gullapalli RR. Ethics of AI in pathology: current paradigms and emerging issues. Am J Pathol 2021 Oct;191(10):1673-1683. [doi: 10.1016/j.ajpath.2021.06.011] [Medline: 34252382]
2. Nakata N. Recent technical development of artificial intelligence for diagnostic medical imaging. Jpn J Radiol 2019 Feb;37(2):103-108. [doi: 10.1007/s11604-018-0804-6] [Medline: 30706381]
3. Mandal S, Greenblatt AB, An J. Imaging intelligence: AI is transforming medical imaging across the imaging spectrum. IEEE Pulse 2018;9(5):16-24. [doi: 10.1109/MPUL.2018.2857226] [Medline: 30273136]
4. OECD. OECD AI principles overview. 2023. URL: https://oecd.ai/en/ai-principles [accessed 2023-05-30]
5. Jha S, Topol EJ. Adapting to artificial intelligence: radiologists and pathologists as information specialists. JAMA 2016 Dec 13;316(22):2353-2354. [doi: 10.1001/jama.2016.17438] [Medline: 27898975]
6. NVVR. De rol van de radioloog in 2020. Strategische visie nederlanse vereniging voor radiologie [The role of the radiologist in 2020. Strategic vision Dutch Society of Radiology]. 2016 URL: de_rol_van_de_radioloog_in_2020.pdf (radiologen.nl)
7. Leiner T, Bennink E, Mol CP, Kuijf HJ, Veldhuis WB. Bringing AI to the clinic: blueprint for a vendor-neutral AI deployment infrastructure. Insights Imaging 2021 Feb 2;12(1):11. [doi: 10.1186/s13244-020-00931-1] [Medline: 33528677]
8. Jha S, Topol EJ. Upending the model of AI adoption. Lancet 2023 Jun;401(10392):1920. [doi: 10.1016/S0140-6736(23)01136-4]
9. Strohm L, Hehakaya C, Ranschaert ER, Boon WPC, Moors EHM. Implementation of artificial intelligence (AI) applications in radiology: hindering and facilitating factors. Eur Radiol 2020 Oct;30(10):5525-5532. [doi: 10.1007/s00330-020-06946-y] [Medline: 32458173]
10. van de Sande D, Van Genderen ME, Smit JM, et al. Developing, implementing and governing artificial intelligence in medicine: a step-by-step approach to prevent an artificial intelligence winter. BMJ Health Care Inform 2022 Feb;29(1):e100495. [doi: 10.1136/bmjhci-2021-100495] [Medline: 35185012]
11. Coppola F, Faggioni L, Regge D, et al. Artificial intelligence: radiologists' expectations and opinions gleaned from a nationwide online survey. Radiol Med 2021 Jan;126(1):63-71. [doi: 10.1007/s11547-020-01205-y] [Medline: 32350797]
12. Chen Y, Stavropoulou C, Narasinkan R, Baker A, Scarbrough H. Professionals' responses to the introduction of AI innovations in radiology and their implications for future adoption: a qualitative study. BMC Health Serv Res 2021 Aug 14;21(1):813. [doi: 10.1186/s12913-021-06861-y] [Medline: 34389014]
13. Eltorai AEM, Bratt AK, Guo HH. Thoracic radiologists' versus computer scientists' perspectives on the future of artificial intelligence in radiology. J Thorac Imaging 2020 Jul;35(4):255-259. [doi: 10.1097/RTI.0000000000000453] [Medline: 31609778]
14. Hendrix N, Lowry KP, Elmore JG, et al. Radiologist preferences for artificial intelligence-based decision support during screening mammography interpretation. J Am Coll Radiol 2022 Oct;19(10):1098-1110. [doi: 10.1016/j.jacr.2022.06.019] [Medline: 35970474]
15. Drogt J, Milota M, Vos S, Bredenoord A, Jongsma K. Integrating artificial intelligence in pathology: a qualitative interview study of users' experiences and expectations. Mod Pathol 2022 Nov;35(11):1540-1550. [doi: 10.1038/s41379-022-01123-6] [Medline: 35927490]
16. Sarwar S, Dent A, Faust K, et al. Physician perspectives on integration of artificial intelligence into diagnostic pathology. NPJ Digit Med 2019;2(1):28. [doi: 10.1038/s41746-019-0106-0] [Medline: 31304375]
17. King H, Williams B, Treanor D, Randell R. How, for whom, and in what contexts will artificial intelligence be adopted in pathology? A realist interview study. J Am Med Inform Assoc 2023 Feb 16;30(3):529-538. [doi: 10.1093/jamia/ocac254] [Medline: 36565465]
18. Huo W, Yuan X, Li X, Luo W, Xie J, Shi B. Increasing acceptance of medical AI: the role of medical staff participation in AI development. Int J Med Inform 2023 Jul;175:105073. [doi: 10.1016/j.ijmedinf.2023.105073] [Medline: 37119693]
19. Milota MM, Drogt J, Jongsma KR. Making AI's impact on pathology visible: using ethnographic methods for ethical and epistemological insights. Digital Soc 2023 Dec;2(3):47. [doi: 10.1007/s44206-023-00072-0]
20. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
21. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
22. Graneheim UH, Lindgren BM, Lundman B. Methodological challenges in qualitative content analysis: a discussion paper. Nurse Educ Today 2017 Sep;56:29-34. [doi: 10.1016/j.nedt.2017.06.002] [Medline: 28651100]
23. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]

24. Kiger ME, Varpio L. Thematic analysis of qualitative data: AMEE Guide No. 131. Med Teach 2020 Aug;42(8):846-854. [doi: 10.1080/0142159X.2020.1755030] [Medline: 32356468]
25. Price W, Nicholson I. Medical AI and contextual bias. Harv J Law Technol 2019;33:66 [FREE Full text]
26. Braun V, Clarke V. To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. Qual Res Sport Exerc Health 2021 Mar 4;13(2):201-216. [doi: 10.1080/2159676X.2019.1704846]
27. Asselbergs FW, Fraser AG. Artificial intelligence in cardiology: the debate continues. Eur Heart J Digit Health 2021 Dec;2(4):721-726. [doi: 10.1093/ehjdh/ztab090] [Medline: 36713089]
28. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. Nurs Health Sci 2013 Sep;15(3):398-405. [doi: 10.1111/nhs.12048]
29. Dierckx de Casterlé B, Gastmans C, Bryon E, Denier Y. QUAGOL: a guide for qualitative data analysis. Int J Nurs Stud 2012 Mar;49(3):360-371. [doi: 10.1016/j.ijnurstu.2011.09.012] [Medline: 21996649]
30. Walter MM. Social Research Methods: An Australian Perspective: Oxford University Press; 2013:416.
31. Aquino YSJ, Rogers WA, Braunack-Mayer A, et al. Utopia versus dystopia: professional perspectives on the impact of healthcare artificial intelligence on clinical roles and skills. Int J Med Inform 2023 Jan;169:104903. [doi: 10.1016/j.ijmedinf.2022.104903] [Medline: 36343512]
32. Mintz Y, Brodie R. Introduction to artificial intelligence in medicine. Minim Invasive Ther Allied Technol 2019 Mar 4;28(2):73-81. [doi: 10.1080/13645706.2019.1575882]
33. van der Kamp A, Waterlander TJ, de Bel T, et al. Artificial intelligence in pediatric pathology: the extinction of a medical profession or the key to a bright future? Pediatr Dev Pathol 2022;25(4):380-387. [doi: 10.1177/10935266211059809] [Medline: 35238696]
34. Froelich JW, Salavati A. Artificial intelligence in PET/CT is about to make whole-body tumor burden measurements a clinical reality. Radiology 2020 Feb;294(2):453-454. [doi: 10.1148/radiol.2019192425] [Medline: 31825291]
35. Singh S, Karimi S, Ho-Shon K, Hamey L. Show, tell and summarise: learning to generate and summarise radiology findings from medical images. Neural Comput Appl 2021 Jul;33(13):7441-7465. [doi: 10.1007/s00521-021-05943-6]
36. Hainc N, Federau C, Stieltjes B, Blatow M, Bink A, Stippich C. The bright, artificial intelligence-augmented future of neuroimaging reading. Front Neurol 2017;8:489. [doi: 10.3389/fneur.2017.00489] [Medline: 28983278]
37. Ranjan A, Parpaleix A, Cardoso J, Adeleke S. AI vs FRCR: what it means for the future. Eur J Radiol 2023 Aug;165:110918. [doi: 10.1016/j.ejrad.2023.110918] [Medline: 37311341]
38. Verma H, Mlynar J, Schaer R, et al. Rethinking the role of AI with physicians in oncology: revealing perspectives from clinical and research workflows. 2023 Apr 19 Presented at: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems; Apr 19, 2023; Hamburg Germany. [doi: 10.1145/3544548.3581506]
39. Gille F, Jobin A, Ienca M. What we talk about when we talk about trust: theory of trust for AI in healthcare. Intell-Based Med 2020 Nov;1-2:100001. [doi: 10.1016/j.ibmed.2020.100001]
40. Hatherley JJ. Limits of trust in medical AI. J Med Ethics 2020 Jul;46(7):478-481. [doi: 10.1136/medethics-2019-105935]
41. Scheek D, Rezazade Mehrizi MH, Ranschaert E. Radiologists in the loop: the roles of radiologists in the development of AI applications. Eur Radiol 2021 Oct;31(10):7960-7968. [doi: 10.1007/s00330-021-07879-w] [Medline: 33860828]
42. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer 2018 Aug;18(8):500-510. [doi: 10.1038/s41568-018-0016-5] [Medline: 29777175]
43. Cui M, Zhang DY. Artificial intelligence and computational pathology. Lab Invest 2021 Apr;101(4):412-422. [doi: 10.1038/s41374-020-00514-0] [Medline: 33454724]
44. Saboury B, Morris M, Siegel E. Future directions in artificial intelligence. Radiol Clin North Am 2021 Nov;59(6):1085-1095. [doi: 10.1016/j.rcl.2021.07.008] [Medline: 34689876]
45. Karhade AV, Schwab JH. Introduction to The Spine Journal special issue on artificial intelligence and machine learning. Spine J 2021 Oct;21(10):1601-1603. [doi: 10.1016/j.spinee.2021.03.028] [Medline: 33785475]
46. Lundström CF, Gilmore HL, Ros PR. Integrated diagnostics: the computational revolution catalyzing cross-disciplinary practices in radiology, pathology, and genomics. Radiology 2017 Oct;285(1):12-15. [doi: 10.1148/radiol.2017170062] [Medline: 28926318]
47. Sorace J, Aberle DR, Elimam D, Lawvere S, Tawfik O, Wallace WD. Integrating pathology and radiology disciplines: an emerging opportunity? BMC Med 2012 Sep 5;10(1):100. [doi: 10.1186/1741-7015-10-100] [Medline: 22950414]
48. Langlotz CP. Will Artificial Intelligence Replace Radiologists?: Radiological Society of North America; 2019. [doi: 10.1148/ryai.2019190058]
49. Hustinx R. Physician centred imaging interpretation is dying out—why should I be a nuclear medicine physician? Eur J Nucl Med Mol Imaging 2019 Dec;46(13):2708-2714. [doi: 10.1007/s00259-019-04371-y] [Medline: 31175395]
50. Laï MC, Brian M, Mamzer MF. Perceptions of artificial intelligence in healthcare: findings from a qualitative survey study among actors in France. J Transl Med 2020 Jan 9;18(1):14. [doi: 10.1186/s12967-019-02204-y] [Medline: 31918710]
51. Abdullah R, Fakieh B. Health care employees' perceptions of the use of artificial intelligence applications: survey study. J Med Internet Res 2020 May 14;22(5):e17620. [doi: 10.2196/17620] [Medline: 32406857]
52. Nakagawa K, Moukheiber L, Celi LA, et al. AI in pathology: what could possibly go wrong? Semin Diagn Pathol 2023 Mar;40(2):100-108. [doi: 10.1053/j.semdp.2023.02.006] [Medline: 36882343]

53. Kim B, Koopmanschap I, Mehrizi MHR, Huysman M, Ranschaert E. How does the radiology community discuss the benefits and limitations of artificial intelligence for their work? A systematic discourse analysis. Eur J Radiol 2021 Mar;136:109566. [doi: 10.1016/j.ejrad.2021.109566] [Medline: 33556686]
54. Krupinski EA. Translating computational innovations into reality: focus on the users!. Presented at: Digital and Computational Pathology; 2023; San Diego, United States. [doi: 10.1117/12.2663067]

Abbreviations

AI: artificial intelligence
COREQ: consolidated criteria for reporting qualitative research
RAIDIO: Responsible Artificial Intelligence in Clinical DecisIOn making
UMCU: University Medical Center Utrecht

Edited by A Kushniruk, E Borycki; submitted 06.09.23; peer-reviewed by D Liu, H Veldandi, H Pokkalla; revised version received 31.03.24; accepted 13.09.24; published 21.11.24.

Please cite as:
Drogt J, Milota M, Veldhuis W, Vos S, Jongsma K
The Promise of AI for Image-Driven Medicine: Qualitative Interview Study of Radiologists' and Pathologists' Perspectives
JMIR Hum Factors 2024;11:e52514
URL: https://humanfactors.jmir.org/2024/1/e52514
doi:10.2196/52514

Supporting Patients' Use of Digital Services in Primary Health Care in England: Synthesis of Evidence From a Mixed Methods Study of "Digital Facilitation"

Jon Sussex[1], MA, MSc; Helen Atherton[2], MPH, PhD; Gary Abel[3], PhD; Christopher Clark[3], PhD; Emma Cockcroft[3], PhD; Brandi Leach[1], PhD; Christine Marriott[4]; Jennifer Newbould[1], PhD; Emma Pitchforth[3], PhD; Rachel Winder[3], BSc; John Campbell[3], MD

[1]RAND Europe, Eastbrook House, Shaftesbury Road, Cambridge, United Kingdom

[2]University of Southampton, Southampton, United Kingdom

[3]University of Exeter Medical School, Exeter, United Kingdom

[4]National Institute of Health and Care Research Collaboration South West Peninsula Patient Engagement Group, University of Exeter Medical School, Exeter, United Kingdom

Corresponding Author:
Jon Sussex, MA, MSc
RAND Europe, , Eastbrook House, Shaftesbury Road, Cambridge, , United Kingdom

Abstract

Background: General medical practitioners and other staff at primary care medical practices have an important role in facilitating patient access to online services in the National Health Service in England. These services range from online ordering of repeat prescriptions to conducting online consultations with health care professionals. We have defined "digital facilitation" as that range of processes, procedures, and personnel that seeks to support patients in their uptake and use of online services.

Objective: We report how we have synthesized the evidence from a mixed methods study of digital facilitation in primary care in England. The study's objectives were to identify, characterize, and explore the benefits and challenges of different models of digital facilitation in general medical practices in England and to design a framework for evaluation of the effectiveness and costs of digital facilitation interventions.

Methods: Our study comprised scoping review of literature, survey of staff in general practices, survey of patients, and ethnography at case study practices plus stakeholder interviews. We compiled a triangulation matrix of the findings from individual work packages through an iterative process whereby each work package's results were first analyzed separately and were then cumulatively combined across work packages in 3 successive workshops. From the resulting matrix, we developed a program theory and an implementation theory and constructed a framework for evaluations of digital facilitation in primary care. The final step of the synthesis process was to discuss the results with national and regional National Health Service stakeholders.

Results: Triangulation yielded a combined set of findings summarized within 11 thematic groupings: 3 setting the scene within which digital facilitation takes place, and 8 related to different types of digital facilitation, their implementation, and effectiveness. Some thematic groupings were evident in the findings of all 4 of the research work packages; others were not addressed in all the work packages but were evident from those where they were addressed. Throughout the synthesis, there were no instances where findings from one work package contradicted the findings of another. Findings either reinforced each other or offered complementary or additional insights. The discussion at the stakeholder meeting held at the end of the study resulted in the research team clarifying some findings but not changing any of them.

Conclusions: Digital facilitation can take many forms, though much of what is currently done in primary care practices in England is reactive and passive. Clear lines of responsibility, digital tools and platforms that work well for patients and practice staff, and investment in staff time and training are all needed if digital facilitation is to deliver on its promise. We propose a framework for future evaluations of the effectiveness and costs of digital facilitation interventions.

(JMIR Hum Factors 2024;11:e52516) doi:10.2196/52516

KEYWORDS

web-based health services; primary care; digital facilitation; evidence synthesis; medical practitioners; digital services; digital intervention; mixed methods study; scoping review; ethnography

Introduction

The National Health Service (NHS) offers all UK residents primary care physician services delivered by general medical practitioners (GPs). GPs work in local practices in multidisciplinary teams with nurses and other health care professionals such as health care assistants, pharmacists, and paramedics, all supported by staff and administrative and reception staff. Each individual patient registers with a GP practice of their choice, usually near to where they live, and the practice is funded via a capitation fee and other payments from the NHS; GP services are free of charge to patients. That practice is responsible for their registered patients' care. GPs and other primary care staff consequently have an important role in facilitating access to the growing range of digital NHS services. It is crucial to note that colloquially within the health service in England, patient-facing services that use the web are known as "online services" [1] but that at policy level and in the academic literature, these services are among what are referred to as "digital" services [2], with digital used as an umbrella term. As a consequence, we use the 2 terms within this paper applying the term "online" when referring to the use of services within general practice in the NHS in England, and "digital" to refer to the wider context.

This paper reports how we synthesized the evidence from multiple parts of a mixed methods study of such "digital facilitation" in NHS primary care in 4 regions of England [3] and the overall findings of that synthesis.

Health care policy in England, in common with many other countries, highlights the role and potential of digital services for patients. There have been increasing contractual requirements for GPs to offer and promote a range of online services following the publication of the NHS Long Term Plan [4] and associated 5-year framework for GP contract reform [5] in 2019. Online services in primary care range from ordering repeat prescriptions, viewing test results, and booking appointments with primary care professionals, to online consultations between patients and doctors. There is considerable diversity in the online services offered across more than 6000 general practices in England. The move to increase reliance on online means of accessing services stems from assumptions that they improve choice, convenience, and ease of access for users; improve triage systems; and streamline service delivery [6-8]. The exigencies of the COVID-19 pandemic from March 2020 onward reinforced and accelerated this initiative [9]. Increased emphasis on many types of digital services has occurred in may jurisdictions [10], associated with the development of policy responses to support the continuing use and growth of such services. In New Zealand, for example, there has been recognition that a longer-term strategy is needed to provide comprehensive and less fragmented digital services in primary care [11].

Offering digital services is not enough to ensure that people use them. Patients may need encouragement and help engage with and use digital services. Our research addresses the experience and future potential of digital facilitation by GPs and other practice staff (ie, providers of primary care). We have defined "digital facilitation" as that range of processes, procedures, and personnel, which seeks to support patients in their uptake and use of online services. This is a pragmatic definition intended to include actions that specifically support patients to take up and use online services already available within the NHS. Although our study focuses on primary care in England, it has relevance for any health care system offering digital services to patients. A scoping literature review undertaken as the first stage of our research found international evidence that digital facilitation can be effective [12].

Our research had 2 principal aims overall: first, to identify, characterize, and explore the potential benefits and challenges of different models of digital facilitation in use in NHS primary care practices in England; and second, to design a framework for future evaluations of the effectiveness and costs of such interventions.

Methods

Overview

We conducted four main, interrelated research work packages over the period 2020 - 2022: (1) a scoping review of literature [12] to determine the types of digital facilitation relevant to primary care, their effectiveness, and cost-effectiveness; (2) a survey of staff in 156 general practices in 4 regions of England: East of England and North London, North-West, South-West, and West Midlands; (3) a survey of 3051 patients from a sample of general practices in East of England and North London, South-West, and West Midlands; and (4) qualitative exploration, based on rapid ethnography undertaken at 8 case study practices and 19 additional interviews with senior stakeholders, to understand in-depth and from the perspective of staff, patients, and other stakeholders the benefits and challenges of different models of digital facilitation. Full details of the methods and findings of each work package are presented elsewhere [3]. This paper focuses on how we have synthesized the evidence from the 4 work packages and the implications of that synthesis.

We have used as our theoretical framework Weiss's approach to theory-based evaluation [13]. The essence of our approach was to compile a triangulation matrix [14] of the findings from individual work packages. The matrix was developed in an iterative process that exploited the staggered completion of the different work packages. First, the researchers for each work package analyzed their findings. The work package findings were then brought together and analyzed cumulatively in a series of 3 half-day workshop meetings of researchers from the relevant work packages and members of the public. Two of the researchers who participated in all 3 workshops were part-time staff in general practices, 1 as a GP (JC) and 1 as an administrator (RW). Among the other participants were members of the patient advisory group established for the research project. Briefing material was circulated to all participants in advance and plenty of time was allowed at the start of each workshop for discussion and clarification of that material. In each workshop the participants identified themes arising from the developing data set and discussed how different sources of evidence reinforced, added to, or differed from one another in their findings.

The first synthesis workshop took place online in November 2021. Nine researchers and 3 patient advisory group members analyzed the findings from the practice survey and the scoping literature review. At the second synthesis workshop in July 2022, also online, 12 researchers and 3 patient advisory group members added the emerging findings from the qualitative research. The final synthesis workshop was held in-person in September 2022 and involved 15 researchers and 3 patient advisory group members, bringing together the entire evidence set from all work packages. The output from the third workshop was an agreed set of 11 thematic groupings identified from the evidence.

Following Weiss's approach [13], we used this knowledge to develop a "program theory," which specifies the mechanism of change, and an "implementation theory" of how digital facilitation is operationalized and what facilitates or hinders its implementation. From the 11 thematic groupings, the program theory and implementation theory, and with reference to UK Medical Research Council guidance on evaluating complex interventions [15], we then constructed a framework for future evaluations of digital facilitation approaches in primary care.

The final step of the synthesis process was to discuss the themes, program and implementation theories, and evaluation framework at an online meeting in December 2022 with representatives of NHS stakeholders at national and regional levels. The participants were from the national body responsible for the operation of the NHS in England (NHS England); the national body specifically responsible for implementation of digital health care services (a part of NHS England); and one of the 42 integrated care boards, which are NHS bodies responsible for commissioning and coordinating health services at a regional level. We shared findings and invited the challenge and review of participants, considering whether change in findings, interpretation, or emphases was warranted.

Ethical Considerations

Ethics approval was obtained for the patient survey and ethnographic case studies within the research project from the North East—Newcastle and North Tyneside 2 Research Ethics Committee on April 27, 2021. Health Research Authority approval was obtained in July 2021 (IRAS number 289425, protocol number L01886). All methods were carried out within the ethical and data governance guidance overseeing this project. This research was carried out in compliance with the World Medical Association Declaration of Helsinki. Ethics approval was not required for the practice survey element (as advised by the Health Research Authority) because the survey did not intend to change practice or patient care. Patients were deemed to have consented to participate in the patient survey if they returned a questionnaire either by post or online (implied consent). The research team did not ask for any personal data from survey participants, although participants could provide their contact details (which were kept separate from other survey data) if they wished to take part in the prize draw. Information on processing of personal data on the participant information sheet provided an explanation of our approach to handling personal data. Practices responding to the practice survey were entered into a prize draw for 1 of 10 £250 (US $316) vouchers. A voluntary prize draw for 1 of 10 £25 (US $32) vouchers was offered as an incentive for patients participating in the patients survey. Potential patient survey respondents were informed that consent would be assumed upon return of a questionnaire either by post or online. Analysis of General Practice Patient Survey data was deemed service evaluation not requiring ethics approval.

Results

Themes

Overview

The detailed results from the literature review, practice survey, patient survey, and qualitative research are reported elsewhere [3,12]. Triangulation of the results from these sources yielded a combined set of findings summarized within 11 thematic groupings: 3 of them setting the scene within which digital facilitation takes place and 8 related to different types of digital facilitation, their implementation, and their effectiveness. Some of the themes (we use this shorter notation from here on) were evident in the findings of all 4 of the research work packages. Other themes were not addressed in all the work packages but were evident from 1, 2, or 3 of them. Throughout the synthesis work, there were no instances where findings from one work package contradicted the findings of another. Findings either reinforced each other or offered complementary or additional insights. The discussion at the stakeholder meeting at the end of the study resulted in the research team clarifying some of its findings but not changing any of them.

The 11 themes are described in turn in the following paragraphs. We first present the 3 scene-setting themes. These were identified from the qualitative exploration. Thus, they emerged from the discussion in the second synthesis workshop—when the emerging results of the qualitative exploration were added into the evidence synthesis process. The third scene-setting scene, related to COVID-19, had also been implicit in some of the practice survey data that reflected the major changes that primary care underwent in response to the pandemic, including reductions in face-to-face consultations and increased reliance on remote methods of ordering prescriptions, for example. The sources that led to the generation of the other themes are explained for each theme in turn.

Theme 1 (Scene Setting)

The *value and purpose of digital services* determines the usefulness of facilitating use of those services. Our qualitative research found that the value of some digital services, and what they are expected to achieve, is not always clear to primary care staff or patients and views and understanding about the value and purpose of such services may differ.

Theme 2 (Scene Setting)

"Digital" conflates online with other routes to access primary care. The qualitative research showed that while patients are greatly interested in navigating the system to gain access to health care, the distinction between "online services" and other access routes may not be important to them. Indeed, any way

of accessing primary care remotely is sometimes seen by patients as "digital," including telephone consultations.

Theme 3 (Scene Setting)

Our survey of GP practices confirmed that the *COVID-19* pandemic and the measures taken in response to it had led to major changes in primary care, including much-increased reliance on remote ordering of prescriptions as well as remote consultations. The qualitative research at case study sites found that at the same time as the need for digital facilitation increased, the pandemic led to the cessation of some digital facilitation initiatives (such as using tablet computers to sign patients up to online services while they sat in the practice waiting room).

Theme 4

Defining and identifying digital facilitation: at the beginning of the study, the research team had defined "digital facilitation" as stated earlier and based the literature review on this, but practice staff may have different definitions. The practice survey, considered with the literature review findings at the first synthesis workshop, revealed that in some practices ad hoc facilitation efforts, such as receptionists answering queries from patients about using online services, would not be seen as digital facilitation but in other practices they would be. The patient survey and the qualitative research then added to this theme by showing that patients may not be aware of digital facilitation per se.

Theme 5

Types of digital facilitation can be active or passive, reactive, or proactive. This was already clear after the literature review and hence the range of types of digital facilitation was already an emergent theme at the first of the 3 synthesis workshops. We revisited the typology at each subsequent synthesis workshop. An example of active digital facilitation is a member of a practice's staff recommending to a patient that they use an online service and showing them how to do that. A poster in a waiting room, or a message played to callers to the practice's telephone line, is passive facilitation. Helping patients only when they ask for assistance is reactive, whereas offering help before it is sought is proactive. Our literature review found that most published studies concern active digital facilitation, while respondents to our survey of GP practices more commonly reported passive and reactive approaches rather than active and proactive, and this was borne out by the qualitative study. The survey of patients demonstrated low awareness of any type of digital facilitation taking place, alongside an often-unfulfilled wish for proactive support to use online services.

Theme 6

Supporting *initial sign-up to online services versus sustained use* thereafter: the literature we reviewed implies that most digital facilitation is focused on achieving initial patient sign-up, rather than supporting sustained use thereafter. The questions we asked in the practice survey were about digital facilitation in general and did not ask separately about support for ongoing use of online services as compared with supporting initial registration to use those services. Our qualitative research findings, added at the second synthesis workshop and confirmed at the third, are consistent with this: digital facilitation by primary care practices more often concerns getting people signed up rather than providing ongoing support to patients subsequently. The patient survey results, discussed at the third synthesis workshop, showed that initial registration for an online service in general practice can be a hurdle for patients.

Theme 7

Who delivers the digital facilitation was not a focus of the literature found in our review. However, the examples in the literature mostly refer to primary care physicians and nurses helping with digital facilitation. Our survey of practices found that GPs and practice staff view the responsibility for digital facilitation as shared between practices and other parts of the NHS regionally and nationally, and that within practices it is the administrative staff who provide most support. Our qualitative research, discussed at the second and third synthesis workshops, found a "bystander effect": different staff groups, patients, and other stakeholders identify other groups as being responsible for digital facilitation. NHS stakeholders at national and regional levels placed responsibility with clinicians and other practice staff. Primary care clinicians seem to place responsibility on practice reception staff (and on patients to sort themselves out); patients viewed it as the responsibility of the NHS more widely.

Theme 8

Enablers of digital facilitation were referred to in many papers in the literature we reviewed. Enablers include staff and patients' perceptions of the usefulness of the online services, along with the user-friendliness of the digital platforms that patients are to use. The practice survey results, discussed at the first synthesis, found that the majority of practices had positive perceptions of online services (that they lead to operational efficiencies). The qualitative exploration found other enablers of digital facilitation, all also mentioned in the literature, including funding or paid time for staff to deliver the facilitation, expectation that an online service will be useful to patients or bring operational efficiencies for the practice or both, the existence of guidelines, 1 or more of the practice staff having specific responsibility for digital facilitation, and patients' trust in practice staff. The patient survey found that only 13% (392/2935) of patients reported having been given help to use online services but also yielded suggestions for enabling digital facilitation.

Theme 9

Barriers to digital facilitation can include staff attitudes toward online services and stereotyped assumptions about the capabilities of some patients, such as the elderly. The multiplicity of different platforms for online services makes digital facilitation more difficult. The literature review highlighted the need for staff time and capacity to deliver digital facilitation, and the practice survey found that most practices consider that they lack the staff time and ability to sufficiently support patients in using online services. Our qualitative research confirmed the existence of all these barriers and also found that digital facilitation can be a low priority for practice staff. The patient survey found that some patients are unaware of online services and others are unhappy about them and resist using the services regardless of facilitation efforts.

Theme 10

Inequalities in digital access and digital facilitation between subgroups in the population: we found little information in the literature about how well digital facilitation works for different population subgroups, although some literature pointed out the risk of practice staff concentrating on the most digitally literate patients and neglecting others who might be in more need of support. In our practice survey, we asked whether digital facilitation was particularly focused on any subgroups of the population. Practices responding frequently reported targeting digital facilitation at older adults. The qualitative research reinforced that a patient's age may be assumed (not necessarily correctly) by staff to be an indicator of digital competence. But many practices also reported "targeting" all the other patient subgroups suggested in the questionnaire, which calls into question what such "targeting" amounts to. The patient survey found that older patients were less likely to be aware of or use digital facilitation.

Theme 11

Effectiveness of digital facilitation: the literature review revealed examples of where digital facilitation had successfully encouraged initial registration with online services, including when GPs and other practice staff personally recommend online services and when practices run introductory sessions for patients. Ongoing guidance and support that is delivered within primary care consultations may be effective not only in increasing initial registration with online services but also in sustaining their subsequent use [12]. We did not address this theme in our survey of practices, which was more concerned to identify the extent and type of digital facilitation activities being undertaken, and enablers and barriers for such activities. The qualitative exploration, discussed at the second synthesis workshop, found that some practices highlight how many patients sign up to online services as evidence of effectiveness—but without knowing the contribution to that of digital facilitation efforts. Our patient survey findings, discussed at the third (final) synthesis workshop, suggest that practices that report using displays, social media, workshops, or events for digital facilitation were more likely to have patients who are aware of and use digital facilitation. We found no such difference, however, for practices that use leaflets, text messages (or emails), or online approaches to digital facilitation for use of online services. Practices that use a "practice champion" for online services have more patients who report being told about such services than other practices.

Program Theory

From the findings of our research as captured in the 11 themes, we developed the program theory using a realist approach to describe provision of digital facilitation in terms of the context in which it takes place, the flow of activities comprising the intervention, and the theory and assumptions underlying the intervention, including the intended outcomes.

Digital facilitation in primary care in England takes place in a context of NHS policy to encourage greater use of digital services, which are seen as a way to improve patients' access to care, improve triage systems, and streamline service delivery. But our research found that practice staff and patients do not always share this view of the value of online services, which makes it challenging to facilitate their use. The context of digital facilitation is also characterized by a multiplicity of platforms for delivering online services. Individual GP practices choose which to use, and the result is a patchwork of different systems and approaches. The COVID-19 pandemic and the measures taken in response greatly accelerated the introduction and use of online services. Our qualitative research found that at the same time some practices stopped providing digital facilitation activities during the pandemic and did not resume them.

The activities that could comprise digital facilitation are varied. Active forms of facilitation include practice champions, training, and workshops. Passive forms include informational leaflets, text messages, and recorded messages on practices' telephone lines. We found that most facilitation is reactive, relying on the digital skills of practice staff—especially reception and administrative staff—to respond to a patient's immediate needs, rather than forming part of a wider effort to enable participation in a health service that increasingly relies on online services. There is evidently some confusion over who has responsibility for supporting patients to use online services. Although practice staff are largely undertaking the digital facilitation that is taking place, they and patients saw a role also for other parts of the NHS at regional and national levels, for example, to tackle the need for wider efforts to educate patients about the benefits of booking appointments online.

Digital facilitation can help patients both directly—to access services such as ordering repeat prescriptions online —and indirectly—by making them more confident users of online services. Our research has shown that the pathways linking digital facilitation to expected beneficial outcomes can be complex. Patients want help to access care and not necessarily help with online services. It is, therefore, unsurprising that we found that digital facilitation is frequently just responding to immediate issues of patient access rather than being aimed at building patients' capacity to access and continue to use online services generally. The path from digital facilitation to its hoped-for benefits requires not only that patients sign up to online services but also that they continue to use them thereafter. Digital facilitation needs to be focused on the latter rather more than is currently the case. Digital facilitation could also be more responsive to inequalities in accessing NHS online services. Our survey showed that practices are aware of the need particularly to support older age groups, non-White ethnicities, lower socioeconomic groups, those in poorer health, and those in rural settings, who may struggle to access digital services. But it was unclear in what ways digital facilitation was being tailored for such patients. Our survey of patients found that ethnic minorities and those for whom English is not a first language are more likely to be aware of and use digital facilitation. However, we also found that older patients are less likely than others to be aware of, or make use of, digital facilitation and (with assumptions made by staff about the impact of older age) are less likely to be told about digital services or helped to use them.

Implementation Theory

Implementation of digital facilitation varies across general practices, both in the capacity to provide it and in the types of facilitation used. Practice populations also differ—for example, in health needs, age structure, and ethnicity—and hence have varying needs for digital facilitation. Within themes 8 and 9 we describe the range of enablers of, and barriers to, implementing digital facilitation. The staff time and resources available to a practice clearly are a major constraint. Most practices we surveyed felt that they had insufficient capacity to provide digital facilitation to the extent they would like. Reactive and passive approaches are more commonly used and they require less staff time (at least in the short term) than more active approaches.

The quality and usefulness of online services as perceived by practice staff and patients affect how, and how far, digital facilitation is implemented. Online services that are easy to use and with clearly apparent benefits require less facilitation effort. For example, repeat prescription ordering online was found in our patient survey and qualitative case studies to be relatively well used. Our qualitative research indicated that issues with more difficult to use services are seen by staff as not their responsibility to resolve. The diverse and changing mix of online services not only presents challenges to patients in understanding what is available and with what support but can also create issues for staff in maintaining knowledge of the online services and the requisite skills to provide facilitation.

Unclear lines of responsibility can also hinder implementation. Practice staff, patients, and other stakeholders (suppliers of digital technology; other parts of the NHS) may each assume that some responsibility for digital facilitation lies with other parties. When no one considers it their responsibility to support patients with broader issues of digital access such as digital literacy and confidence, then these aspects are likely to be neglected. In general, implementation would be aided by all members of staff in a general practice having a clear, shared understanding of what digital facilitation its patients need and how to deliver that.

Evaluation Framework

One aim of our research was to design a framework for future evaluations of the effectiveness and costs of digital facilitation interventions. Based on our synthesis of the findings from all 4 work packages, we propose an evaluation framework consisting of 4 aspects. The high-level nature of the proposed framework reflects what we have learned about the awareness and extent of digital facilitation in the NHS in England. The implementation of future evaluations will require engagement with all stakeholders and with policy makers. Within the bounds of timescales and funding available, an iterative approach to evaluation, with progressive and cumulative learning, is, as ever, desirable.

1. Digital facilitation as an intervention: digital facilitation may be defined as support to enable patients to achieve access to care services digitally. Many types of digital facilitation interventions are possible, are not mutually exclusive, and are often complex [15].
2. Responsibility for digital facilitation: clarity about who is responsible for (which) digital facilitation is key to it happening: which practice staff are responsible and how this fits with their role; what is expected of patients; are any third parties (eg, technology suppliers and charities working with patients) involved in delivering or supporting digital facilitation and how do they interface with practice staff and patients; and how does digital facilitation fit into the wider health care community beyond general practice?
3. Patient groups and potential for inequalities: for which groups of patients is digital facilitation most needed; do such groups differ in the extent to which they would benefit from, or be burdened by, online services and will they have different views of the importance of online services; and which types of facilitation are most effective for which groups?
4. Outcome and cost measures: Potential measures of outcomes (intended and unintended) include the following and should be collected in a way that permits determination of inequalities between subgroups in the population: patient awareness of, registration with, and sustained use over time of digital services; measures of access to GP services and rates of digital access within that; patient-reported experiences of engaging with facilitation and of the online services used; practice staff–reported experiences with facilitation (ease of delivery, impact on workload, and views on whether it is working); costs to general practices of digital facilitation (training, staff time, materials, and equipment); costs or savings to practices from changed use of online services by patients; costs or savings to the rest of the health care system; and costs or savings (money and time) to patients from using online services. Within the bounds of feasibility and funding, the longer the period over which outcome and cost data can be collected the more complete will be the understanding of longer-term impacts.

Discussion

Principal Findings

To the research team's knowledge, our research is the first to explore the range and extent of digital facilitation in primary care in a health care system. We have found that the digital facilitation being provided by general practices varies from place to place but is often passive and reactive, rather than active or proactive, and is focused on immediate difficulties rather than on broader and longer-term support for patients generally to use online primary care services. By identifying themes emerging from our findings when taken together, we have developed a program theory and an implementation theory of digital facilitation in primary care, which in turn have enabled us to construct a framework for future evaluations. These lead collectively to the following implications for policy, practice, and research.

Implications for Policy

A range of policy implications is evident and may well apply in many other countries than just England. Digital facilitation has a role in achieving the move to greater use of digital services in health care. But there is a disconnect between policy makers' expectations about what use of online services might achieve and the limited efforts at digital facilitation that are occurring at general practice level. A first step for policy makers is to recognize the existence of this disconnect. Rectifying it implies a need to better articulate to service users and practice staff the hoped-for benefits of the online services. To deliver digital facilitation requires investment in staff time and training. It also requires clarity about how the responsibility for supporting patients to use online services is distributed across different parties. Although some responsibility for digital facilitation falls on practice staff, there is a role for complementary support from health service organizations nationally or regionally [16], and this needs to be recognized and acted upon. An additional option is for the providers of online service platforms to be encouraged, or mandated, to offer support to patients either directly or by helping general practice staff to deliver digital facilitation (see, for example, eConsult [17]).

Implications for Practice

There is clear potential for digital facilitation to help patients and practices use online services. Realizing this potential requires investment in, as well as by, general practices. It also requires clear, shared understanding of which staff are responsible for doing what [18]. Practice leaders and managers need to take responsibility for their practice's digital facilitation strategy and associated training and resources being provided for those staff tasked to deliver facilitation activities [19]. There exists a wide range of facilitation types and it is likely that a combination of approaches would be appropriate. Attention needs to be paid to the likely differing digital facilitation needs of subgroups of the patient population, rather than assuming that a generic approach will suffice. Getting patients signed up with online services is a necessary but insufficient condition for realizing the benefits of the services. Monitoring and support for continued use after initial sign-up are also needed but appear to receive insufficient attention currently. Finally, it is inevitable that some patients will never be able or willing to use some (or any) online services; hence adequate, equitable, and nondigital access to care will need to remain an option however good digital facilitation becomes.

Implications for Research

There is great scope for useful research around digital facilitation. There is a need to explore the association between patients' awareness of online services and their use of them, and how different digital facilitation approaches affect that. Measuring the effectiveness of digital facilitation efforts requires a holistic approach in line with guidance on complex interventions, which suggests that evaluation goes beyond whether an intervention works [15]. Our research suggests several areas for future evaluation of digital facilitation, where current evidence is lacking, including:

- the extent of patient demand for online services to access general practice, and how this varies between population subgroups,
- the level and mix of digital facilitation interventions needed,
- the role of general practice administrative staff in supporting online services,
- the relative effectiveness and costs of different approaches to digital facilitation,
- how different approaches to digital facilitation work for different population subgroups, and
- focusing on supporting sustained patient use of online services, not just initial registration.

Strengths and Limitations

A strength of the research project was the collection of evidence from 4 main sources and the triangulation between them that enabled. The 4 work packages were deliberately staggered and yielded results at different times. We exploited that through a cumulative design of the evidence synthesis, comprising 3 workshops (as successive work packages yielded findings) and a final stakeholder challenge meeting, which took place at intervals over 13 months. This design was chosen to ensure a thorough, effective, and efficient process with time for detailed challenge and discussion. Members of the public were actively involved throughout, including at the 3 synthesis workshops, and they were supported throughout by a researcher in our team specializing in that role. The public participants made pertinent challenges and shared experiences in each workshop leading to clearer thinking about the context and meaning of the research findings.

The principal limitation of the synthesis reported here is that it is based on evidence gathered during a period of great change in primary care in the United Kingdom (as elsewhere): 2020 - 2022. The COVID-19 pandemic prompted great and sudden changes in the practice of primary care, with much greater emphasis on providing services remotely rather than in person. The full implications of those changes are not yet wholly apparent. We have noted their importance and taken them into account to the extent possible so far.

Conclusions

Digital facilitation is important in the context of increasing opportunities for online access to services in primary care. It can take many forms, though much of what is currently done in GP practices in England is reactive and passive. There is scope to develop an approach to facilitation that more actively engages patients. There seems to be a disconnect between stakeholders' expectations and perceptions of how digital facilitation could help and the reality seen in everyday practice. Digital facilitation requires staff time and resources, along with clarity over responsibilities. The establishment of clear lines of responsibility, the development of digital tools and platforms that work well for patients and practice staff, and investment in staff time and training will all be needed if digital facilitation is to deliver on its promise. Based on synthesis of the findings from our research, we propose a framework for future evaluations of the effectiveness and costs of digital facilitation interventions.

Acknowledgments

This study was funded by the National Institute for Health and Care Research (NIHR) Health and Social Care Delivery Research Programme 128268. The views expressed in this publication are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. The authors gratefully acknowledge the following people for their contributions to the research project for which the evidence synthesis is reported here. Research team colleagues: Nurunnahar Akter, Abodunrin Aminu, Wiktoria Bak, Carol Bryce, Hamish Evans, Evangelos Gkousis, Georgia Jenkins, Caroline Jenkinson, Nada Khan, Sarah Parkinson, Joanne Parsons, Laura Sheard, Stephanie Stockwell, Chloe Thomas, and Bethan Treadgold. Patient advisory group: Laiya Bailey, Julie Harvey, Diana Lewis-Frost, Ben Marshall, Andrea Shelly, Lynn Tatnell, Penny Thompson, and Malcolm Turner. Steering group committee: Jeremy Horwood, David Parkes, and Amrit Takhar. Sponsors: Pam Baxter and Vivienne Shaw. Sponsor team: Sarah Hider and Alexander Philips. Administrator: Ellie Kingsland. Survey support: Jodie Button, Mayam Gomez-Cano, Sine MacDonald, Isabelle Mayne, Charlotte Reburn, and Claire Wright. Project development: Richard Moore. Clinical Research Networks: East of England, North Thames, North West Coast, South West Peninsula, and West Midlands. All patients, general practices, and general practice staff who supported and participated in the study.

Authors' Contributions

JS conceived the original research idea, contributed to the overall study design, led the synthesis of evidence, and led the drafting of the journal article. JC led the overall study, contributed to the overall study design, contributed to the synthesis of evidence, and contributed to the drafting of the journal article. HA and GA contributed to the overall study design, contributed to the synthesis of evidence, and contributed to drafting the journal article. CEC, EC, BL, CM, JN, EP, and RW contributed to the synthesis of evidence and to drafting the journal article.

Conflicts of Interest

None declared.

References

1. NHS England. GP online services. 2024. URL: https://www.england.nhs.uk/gp-online-services/ [accessed 2024-03-18]
2. NHS England Transformation Directorate. Our strategy to digitise, connect and transform. 2024. URL: https://transform.england.nhs.uk/digitise-connect-transform/our-strategy-to-digitise-connect-and-transform/ [accessed 2024-03-18]
3. Abel G, Atherton H, Sussex J, et al. Current experience and future potential of facilitating access to digital NHS primary care services in England: the Di-Facto mixed-methods study. Health Soc Care Deliv Res 2024;12(32):1-197. [doi: 10.3310/JKYT5803]
4. NHS England. The NHS long term plan. 2019. URL: https://www.longtermplan.nhs.uk/wp-content/uploads/2019/08/nhs-long-term-plan-version-1.2.pdf [accessed 2023-07-25]
5. NHS England. A five-year framework for GP contract reform to implement The NHS Long Term Plan. 2019. URL: https://www.england.nhs.uk/publication/gp-contract-five-year-framework/ [accessed 2023-07-25]
6. Department of Health and Social Care. A plan for digital health and social care Appendix A: our action plan for delivering a digital future. 2022. URL: https://www.gov.uk/government/publications/a-plan-for-digital-health-and-social-care/a-plan-for-digital-health-and-social-care [accessed 2023-06-29]
7. Clarke G, Dias A, Wolters A. Access to and delivery of general practice services: a study of patients at practices using digital and online tools.: The Health Foundation; 2022. URL: https://www.health.org.uk/publications/access-to-and-delivery-of-general-practice-services [accessed 2023-06-29]
8. NHS Digital. Why digital inclusion matters to health and social care. 2022. URL: https://digital.nhs.uk/about-nhs-digital/corporate-information-and-documents/digital-inclusion/digital-inclusion-in-health-and-social-care [accessed 2023-06-29]
9. Hutchings R. The impact of Covid-19 on the use of digital technology in the NHS.: Nuffield Trust; 2020. URL: https://www.nuffieldtrust.org.uk/files/2020-08/the-impact-of-covid-19-on-the-use-of-digital-technology-in-the-nhs-web-2.pdf [accessed 2023-06-29]
10. Jonnagaddala J, Godinho MA, Liaw ST. From telehealth to virtual primary care in Australia? A rapid scoping review. Int J Med Inform 2021 Jul;151:104470. [doi: 10.1016/j.ijmedinf.2021.104470] [Medline: 34000481]
11. Ang A. New Zealand’s national health plan seeks “greater use” of digital tools. : Heatlhcare IT News; 2022. URL: https://www.healthcareitnews.com/news/anz/new-zealands-national-health-plan-seeks-greater-use-digital-tools [accessed 2023-07-25]
12. Leach B, Parkinson S, Gkousis E, et al. Digital facilitation to support patient access to web-based primary care services: scoping literature review. J Med Internet Res 2022 Jul 14;24(7):e33911. [doi: 10.2196/33911] [Medline: 35834301]
13. Weiss CH. Theory‐based evaluation: past, present, and future. N D E 1997 Dec;1997(76):41-55. [doi: 10.1002/ev.1086]
14. O’Cathain A, Murphy E, Nicholl J. Three techniques for integrating data in mixed methods studies. BMJ 2010 Sep 17;341:c4587. [doi: 10.1136/bmj.c4587] [Medline: 20851841]
15. Skivington K, Matthews L, Simpson SA, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021 Sep 30;374:n2061. [doi: 10.1136/bmj.n2061] [Medline: 34593508]

16. NHS England. Next steps for integrating primary care: fuller stocktake report. 2022. URL: https://www.england.nhs.uk/wp-content/uploads/2022/05/next-steps-for-integrating-primary-care-fuller-stocktake-report.pdf [accessed 2023-01-18]
17. eConsult. How to use eConsult. 2023. URL: https://econsult.net/nhs-patients/how-to-use-econsult [accessed 2023-01-11]
18. Hammerton M, Benson T, Sibley A. Readiness for five digital technologies in general practice: perceptions of staff in one part of southern England. BMJ Open Qual 2022 Jun;11(2):e001865. [doi: 10.1136/bmjoq-2022-001865] [Medline: 35768171]
19. Crotty BH, Somai M. Bugs in the virtual clinic: confronting telemedicine's challenges through empathy and support. J Particip Med 2022 Apr 22;14(1):e25688. [doi: 10.2196/25688] [Medline: 35452399]

Abbreviations

GP: general medical practitioner (primary care physician)
NHS: National Health Service (UK)

Edited by A Kushniruk, E Borycki; submitted 06.09.23; peer-reviewed by F Savira, M Lariviere; revised version received 18.03.24; accepted 13.09.24; published 04.12.24.

Please cite as:
Sussex J, Atherton H, Abel G, Clark C, Cockcroft E, Leach B, Marriott C, Newbould J, Pitchforth E, Winder R, Campbell J
Supporting Patients' Use of Digital Services in Primary Health Care in England: Synthesis of Evidence From a Mixed Methods Study of "Digital Facilitation"
JMIR Hum Factors 2024;11:e52516
URL: https://humanfactors.jmir.org/2024/1/e52516
doi:10.2196/52516

Factors Impacting the Adoption and Potential Reimbursement of a Virtual Reality Tool for Pain Management in Switzerland: Qualitative Case Study

Josefine Lurtz[1,2], MSc; Thomas C Sauter[3*], Prof Dr Med, MME; Christine Jacob[1*], PhD

[1]FHNW - University of Applied Sciences Northwestern Switzerland, Riggenbachstrasse 16, Olten, Switzerland

[2]University of Oxford, Oxford, United Kingdom

[3]Department of Emergency Medicine, Inselspital, Bern University Hospital, Bern, Switzerland

[*]these authors contributed equally

Corresponding Author:
Christine Jacob, PhD
FHNW - University of Applied Sciences Northwestern Switzerland, , Riggenbachstrasse 16, Olten, , Switzerland

Abstract

Background: Pain and its adequate treatment are an issue in hospitals and emergency departments (EDs). A virtual reality (VR) tool to manage pain could act as a valuable complement to common pharmaceutical analgesics. While efficacy could be shown in previous studies, this does not assure clinical adoption in EDs.

Objective: The main aim of this study was to investigate which factors affect the adoption and potential reimbursement of a VR tool for pain management in the ED of a Swiss university hospital.

Methods: Key informant interviews were conducted using in-depth semistructured interviews with 11 participants reflecting the perspectives of all the relevant stakeholder groups, including physicians, nurses, patients, health technology providers, and health insurance and reimbursement experts. The interviews were recorded and transcribed, and the extracted data were systematically analyzed using a thematic analysis and narrative synthesis of emergent themes. A consolidated framework for eHealth adoption was used to enable a systematic investigation of the topic and help determine which adoption factors are considered as facilitators or barriers or as not particularly relevant for the tool subject of this study.

Results: According to the participants, the three key facilitators are (1) organizational environment; (2) tension for change, ease of use, and demonstrability; and (3) employee engagement. Further, the three key barriers to adoption are (1) workload, (2) changes in clinical workflow and habit, and (3) reimbursement.

Conclusions: This study concludes that the adoption of a VR tool for pain management in the ED of the hospital subject of this study, although benefiting from a high tension for change in pain and workload management, is highly dependent on the respective organizational environment, engagement of the clinical staff, and reimbursement considerations. While tailored incentive structures and ambassador roles could benefit initial adoption, a change in the reimbursement landscape and further investigation of the positive effects on workflow effectiveness are required to drive long-term adoption.

(JMIR Hum Factors 2024;11:e59073) doi:10.2196/59073

KEYWORDS

eHealth; mobile health; mHealth; digital health; reimbursement; emergency; technology assessment; technology adoption; implementation; VR; virtual reality; pain; experience; attitude; opinion; perception; acceptance; adoption; qualitative; interview; hospital

Introduction

Background

Pain management in emergency departments (EDs) has been shown to be a challenge, with oligoanalgesia, the undertreatment of pain, being a major issue [1], leading to inadequate treatment of pain [2]. According to the European Society for Emergency Medicine, effective pain relief, whether through nondrug methods or medication, involves several key steps: consistently assessing pain levels, using age-appropriate pain management techniques, selecting suitable pain relief medications, understanding potential side effects, and regularly reevaluating both the patient's pain and their pain management plan to adjust as needed [3]. This implies that addressing acute pain in emergency situations necessitates a personalized approach, taking into account the specific characteristics of each pain management method [4].

Within EDs, pain treatment is impaired by the fact that pain is highly individual, knowledge and education are often scarce, the individual expression of pain is culturally dependent and the workload in EDs is high [1]. While this issue may be addressed through specific education and guidelines on pain treatment [2], tools such as virtual reality (VR) could support clinicians as a complementary pain management option to common pharmaceutical analgesics [5,6]. VR is generally defined as “the use of computer modeling and simulation that enables a person to interact with an artificial 3D visual or other sensory environment. VR applications immerse the user in a computer-generated environment that simulates reality through the use of interactive devices, which send and receive information and are worn as goggles, headsets, gloves, or bodysuits” [7].

VR interventions have emerged as a promising nonpharmacological treatment option for managing pain in patients. Multiple benefits have been demonstrated including decreasing pain intensity perceived, anxiety, unpleasantness, and the time spent thinking about pain during medical procedures [8,9]. Hoffman et al [10] showed that the use of VR tools can result in pain alleviation comparable to a moderate opioid dose, and that it can be an effective complement to common analgesics. As a complementary pain management practice, VR has shown potential for both acute and chronic pain conditions [11]. However, as this study focused on an ED setting, it is centered around acute pain interventions.

The use of VR in pain management is based on the concept of providing immersive, multimodal stimuli to engage the patient, thus distracting patients from their pain [12]. These stimuli may include, but are not limited to, a device, projecting visuals (eg, in the form of a head-mounted display), earphones (potentially noise canceling), and handheld steering devices [12]. The resulting multimodal sensory inputs are an important differentiating factor to passive diversion such as watching television, as the stimuli provide the patient the illusion of being in a different environment [12].

Researchers studying VR as a tool to attenuate pain explain that the underlying methodology of immersive distraction has been identified as a main factor contributing to a reduction in the experienced pain of patients [11]. Hoffman et al [13] argued that VR is redirecting the limited attentional capacity of humans. Thereby, VR treatments would leave less attention available to process and direct input from pain receptors. Strong et al [14] found a similar neuronal explanation, reasoning that neural pathways are not available for transmitting pain signals while occupied with VR input. Gupta et al [11] argue that while there is evidence on VR affecting pain beyond distraction, the respective research is focused on chronic rather than acute pain.

Objectives

This study investigates the adoption of a VR simulation for pain and anxiety at a Swiss university hospital ED. The VR tool, called Healthymind (HEALTHY MIND), is a patented, CE (conformité européenne) marked, International Organization for Standardization 13485 certified, and registered Class I medical device that is commercially available [15]. The intervention consisted of the application of the Healthymind VR simulation, using a Pico G2 4 K VR headset (Pico Interactive Inc) with a resolution of 1920×2160 pixels and a diagonal field of view of 101 degrees and Bose Quiet Comfort 35 II noise-canceling headphones (Bose Corporation) as an adjunct to usual care in the ED.

The ED team in the hospital subject of this study has conducted an initial research about Virtual Reality for Pain Relief in the Emergency Room to investigate the feasibility of deployment of a VR simulation in the busy setting of the ED for an adult population presenting with traumatic or nontraumatic pain, and the effectiveness of the VR simulation in pain and anxiety control. This work builds on and complements this initial study that was published earlier [5], and TCS shares coauthorship of both papers.

While the tool was introduced with intense training and the highly engaged VR tool advocates pushing adoption, its use has not yet set into clinical routine. Therefore, the key objectives of this work were to assess what factors impact the adoption of the VR tool in question, and whether they act as barriers or facilitators. Furthermore, potential solutions for the identified barriers were also investigated. The authors were guided in their thinking by the consolidated framework for eHealth adoption by Jacob et al [16]. The consolidated framework, informed by the sociotechnical theory [17], categorizes adoption factors in three key clusters: (1) organizational and policy, (2) technical and material, (3) social and personal factors.

Methods

Study Design

The qualitative paradigm was chosen due to its emphasis on prioritizing “the voices of participants” and the rich insights it offers [18]. Qualitative methods allow for a deeper understanding of the participants’ individual perceptions in ways that cannot always be attained through quantitative approaches [19].

Data Collection

Data were gathered through in-depth, semistructured interviews, all held digitally. Additionally, physical artifacts such as screenshots of the VR tool, compatible devices, and examples of user feedback were collected to provide a comprehensive assessment of the tool under study (Healthymind) [20]. The data collection period spanned from March to June 2023, during which a total of 11 interviews were conducted (out of 62 individuals contacted). The interviews were conducted and recorded in German by the first author (JL), with durations ranging from 15 to 60 minutes depending on the participant’s depth of perspective. The interview topic guide is available in Multimedia Appendix 1. Research themes and questions were developed in accordance with Jacob et al’s [16] consolidated framework of the factors impacting eHealth adoption. As per this framework, themes in the interview guide were categorized into three groups: social, organizational, and technical. Data collection continued until a satisfactory level of saturation was achieved, indicating that new data no longer yielded novel insights [19].

Sampling Techniques and Participant Profiles

Considering the specific expertise required for this research, purposive sampling was used, in which potential participants were selected based on their ability to provide rich and in-depth information about the topic. We used a purposive sampling method, selecting participants based on their capacity to provide detailed and comprehensive information about the app and its use [18,21]. Initially, key informants within the Swiss university hospital ED were approached, and subsequently, snowball sampling was used to identify suitable participants. The primary selection criteria included participants being key stakeholders in the Swiss health care system and in the adoption discussion around VR tools. A “stakeholder” is defined by Effective Health Care Program as a person or group with a vested interest in a particular clinical decision and the evidence that supports that decision [22]; in the context of this study, this includes physicians, nurses, patients, health insurance professionals, policy experts, and employees of the hospital provider of the VR tool and experts of the Swiss Federal Office of Public Health as well as of the cantonal health department. Thereby, this research is incorporating insights from various points of views in the Swiss health care context to portray an inclusive picture on the current situation. Participant recruitment was conducted through tailored LinkedIn and email outreaches, including a detailed study information sheet as shown in Multimedia Appendix 2. Table 1 lists the participant profiles showing their diverse angles that represent the different stakeholders in the Swiss health care ecosystem.

Table . Participant profiles.

Participant ID	Participant perspective	Organization
P1	Policy expert	Hospital subject of this study
P2	Physician	Hospital subject of this study
P3	Physician	Hospital subject of this study
P4	Nurse	Hospital subject of this study
P5	Technology provider	Provider of the virtual reality tool subject of this study
P6	National policy expert	Member of the Eidgenössischen Kommission für Leistungen und Grundsatzfragen, which translates to Federal Commission for Benefits and Policy Issues
P7	Reimbursement expert	Mandatory insurance provider
P8	National policy expert	From the Federal Office of Public Health
P9	Cantonal policy expert	From the canton where the hospital subject of this study operates
P10	Nurse	Hospital subject of this study
P11	Patient	Hospital subject of this study

Data Analysis

Thematic analysis was used to identify and extract relevant themes, as well as to interpret their potential meanings and interrelationships [21,23]. For data coding, computer-assisted qualitative data analysis software, specifically Atlas.ti (Lumivero and ATLAS.ti Scientific Software Development GmbH), was used. Excerpts were selected to construct a narrative for each theme, aiding in enhancing comprehension of the analysis. The primary author (JL) conducted the interviews and the analysis and coding. CJ reviewed the coding, and any instances of disagreement were resolved through discussion with the third author (TCS).

The deductive themes were predefined according to the consolidated framework for eHealth adoption by Jacob et al [16]. Figure 1 shows the deductive themes that served as the starting point for the thematic analysis; the inductive themes emerged from the data as some factors were marked as facilitators, barriers, or not particularly relevant for the case subject of this study.

Figure 1. Deductive themes according the consolidated framework for eHealth adoption by Jacob et al [16].

Ethical Considerations

All participants signed an informed consent form prior to the qualitative interview. This research was part of an overarching study approved by the local ethics institutional review board Kantonale Ethikkommission Bern (Req-2020‐01,266). All participants were briefed about the research background and signed a consent form agreeing to participate. Participants did not receive any payment for their participation.

Results

Overview

Based on the interview analyses, this section synthesizes which factors are currently supporting or inhibiting the adoption of the VR tool for pain management in the ED setting of the Swiss university hospital subject of this study. Multimedia Appendix 3 includes direct quotes from the participants supporting the narrative synthesis of the different themes reported below.

Organizational and Policy-Related Adoption Factors

Organizational Factors

The interview analysis confirms that the overall organizational environment is a critical facilitating factor for adoption. A policy expert within the hospital pointed that the organizational circumstances are indeed beneficial for the adoption of such a technology, as the innovational spirit in a university hospital is high. The physician's perspective underscores the advantage of university settings in fostering early adoption of new technologies or approaches. Since these environments are often research-focused, financial concerns can take a back seat during pilot phases. This allows for a more exploratory and experimental approach without immediate financial pressures, thus facilitating innovation and the testing of new methods or technologies.

A factor that was unanimously mentioned as a strong facilitator is the inner setting and strong focus around training and education within the hospital, particularly on innovative treatment methods. Physicians and nurses appreciate the promotion of tools such as VR within the hospital and the general focus on new technologies. Specifically, the teaching assignment of the university hospital promotes the adoption of new tools by not merely offering the option to use the device, but by carefully introducing it to the relevant personnel. Therefore, the importance of proper training and the effort required for it were emphasized.

Leadership commitment and management support regarding this and similar projects that go in the direction of digital solutions were explicitly mentioned as a facilitator. Furthermore, participants noted that the VR tool might act as a unique selling point of the organization and its ED in clinician recruitment efforts, which could act as a component reinforcing implementation, especially in the context of a hospital engaging in education.

Lastly, the organizational factor of "tension for change" and the current trend toward such interventions can create traction, as innovations in this direction are anticipated. Additionally, pain management in the ED was confirmed as a pressing concern; having an additional option to alleviate patient pain will facilitate adoption.

Workflow-Related Factors

The analysis of workflow-related factors revealed that changes to clinical workflow, workflow fit, resources, and workload are highly interrelated. The required changes in clinical work to implement VR were brought up as a barrier. Workload was mentioned frequently as a main barrier to adoption into the clinical workflow and routine, as time is scarce in the ED.

While VR adoption requires habituation and adaptation, and pharmacological options might initially be faster, a clinician believed that VR may lead to time savings in the long run, benefiting the organization. Although nurses did not explicitly confirm this efficiency gain, the tool could allow them to allocate their time more effectively by distracting patients in pain before further attention is needed. Additionally, the tool's adoption may support the transformation away from purely pharmacological pain management, which was mentioned as a positive change of clinical care.

Regarding compatibility and adaptability, the provider of the tool emphasized that trials on further adaptation of the software to specific environments or medical uses did not yield better results. That is why they provide environments which are applicable to as many use cases as possible. However, within the possible simulation environments that the provider supplies, customization is possible in a sense of selection of the preferred environment through the patient, which is a facilitating factor.

Although the potential threat of new technologies for clinicians was mentioned during interviews, this does not seem applicable to the VR case at the hospital studied. On the contrary, especially for nurses, this tool is seen as an opportunity to expand their capabilities and skill sets. Additionally, it is beneficial that nurses can decide to use the VR tool without needing to escalate up the hospital hierarchy or require a physician's prescription.

Policy and Regulations

Since the tool is already approved and certified, research participants did not perceive regulatory approval as a barrier, as is often the case in similar studies. During discussions on "policy and regulations," participants did not raise this point; instead, the focus was on reimbursement aspects, particularly in the context of Swiss health care. From a health insurance perspective, participants emphasized that the Swiss insurance system operates on a fee-for-service payment model, where reimbursement is based on the volume of care services provided. This model incentivizes providers to perform more services, such as filling hospital beds and conducting procedures. Therefore, a solution adds the most value if it increases efficiency, allowing more patients to be seen and generating more billing revenue. This could act as a barrier to adoption for digital health interventions such as the VR tool studied, which focus more on quality and patient experience improvements rather than increasing service volume.

The overarching concern was how the VR application could be funded for patients in pain in the ED. Reimbursement may happen via the mandatory insurance or through a supplementary route. One suggestion for reimbursement through the mandatory health insurance was via the so-called MiGeL (German: *mittel und gegenständeiste* which translates to *means and object list*; Nova Cantica). This exhaustive list for reimbursable tools for the use outside the hospital, was proposed by the federal ministry of health. However, as the VR tool subject of this study is applied inside the hospital, the MiGeL-related reimbursement is not applicable. Instead, reimbursement would happen via TARMED, the current Swiss tariff system for outpatients in hospitals, as the VR tool is to be categorized as a mandatory service provided in the hospital. Such a categorization is not substantiated on an exhaustive list, but by a so-called principle of trust, an assumption that medical insurers reimburse all performed examinations and treatments if they are appropriate according to the WZW criteria (in German: *wirksamkeit, zweckmässigkeit, und wirtschaftlichkeit*; translates to *effectiveness, expediency, and efficiency*). As the categorization as a mandatory service is not currently doubted or questioned by insurers or other parties, reimbursement of mandatory health insurers would be expected. However, the categorization as a mandatory service cannot definitely be confirmed according to the policy experts participating in this study. This categorization would rule out options of reimbursement through complementary insurances, which was the initial reimbursement method suggested by clinicians and payers. Further it rules out self-payment of patients, as mandatory services fall under a so-called tariff protection.

For the categorization as a mandatory service the WZW criteria must be fulfilled. The service provider, in this case the hospital subject of this study and its physicians, must determine whether the WZW criteria are fulfilled by the tool, by demonstrating effectiveness, appropriateness, and economic efficiency. If this can be confirmed, reimbursement within the mandatory health insurance could be given. However, the categorization as a mandatory service through the fulfillment of the criteria above, does neither automatically lead to a cost-covering reimbursement amount, nor does it create tariff positions in the TARMED system for "standard" reimbursement, as TARMED is not kept up to date anymore because it will be replaced in the upcoming years.

Reimbursement could also be pursued through individual case invoices, where a physician would need to justify the necessity and appropriateness of the tool to the insurer. However, this approach is resource-intensive and not advisable in this case. Instead, other short-term solutions should be explored, such as assessing whether the tool's adoption offers economic advantages, making it attractive even without immediate reimbursement. In the long-term, participating policy experts strongly recommend that the hospital engage in advocacy and help shape the new outpatient tariff system. This would ensure that VR for pain management and other digital tools have an appropriate category within the new system, allowing for seamless reimbursement if categorized as a mandatory service.

Patient-Related Factors

Patient access to care was not a concern for the participants, due to the fact that the Swiss EDs are open to all citizens, thanks to compulsory health insurance, anyone living in Switzerland has access to medical care [24]. Similarly, patient safety was not a major concern as the VR tool subject of this study does not prompt any known safety risks according to this study's participants. However, patient condition and engagement were discussed.

Patient condition was mentioned in the context that the severity of the patients' pain may impact the tools' effectiveness, as the necessary immersion may be harder to achieve. While this can have a negative impact on adoption, the general difficulty to appropriately treat the condition "pain" in the ED, including the reluctance for the use of opioids, can be seen as a facilitator, as another treatment option expands the clinicians' toolkit according to this study's participants.

The nurses in this study highlighted that patient education requires significant time and effort, which might not always be available. The participating patient confirmed the time taken to inform them and noted their high involvement in deciding whether to use the VR tool. This process substantially increased patient engagement and their willingness to use the tool.

User Engagement

Participating physicians emphasized that nurses may perceive the tool as an expansion of their capabilities, which may have a positive effect on their feelings about their job. Confirming this perspective, the interviewed nurses expressed excitement about the application of the new tool. Furthermore, the clinicians explained that fostered engagement may positively affect adoption through the enhancement of employee engagement and other factors such as habit. However, the nurse view also revealed that different nurse team members and physicians individually vary in their perception, which affects teamwork and adoption.

Technical and Material Adoption Factors

Usefulness

The tool's usefulness was of high relevance for all participants. User perception was generally positive, and the participating clinicians highlighted the great potential of VR for pain management. As for efficiency of care, the picture is more complex. This is as installing and applying the tool takes about 20 minutes, which the care professionals must integrate into their already busy schedule. However, it may be argued that while the patient is occupied with the tool for 20 minutes, the care team may take advantage of this time more efficiently. Therefore, the factor efficiency can be seen as a barrier at a first glance, but it can also be perceived as a facilitator.

Communication was also discussed, with the care team noting that explaining the function and usefulness of the device could take significant time, depending on the patient's prior knowledge. However, this effort can enhance the patient experience. Both patients and clinicians confirmed that positive stories about the tool increase patients' willingness to try it. Therefore, clear guidelines on how to effectively present the tool to patients may facilitate its adoption.

IT Capability and Capacity

Although IT capability and capacity can affect adoption for certain tools, it was not perceived as of relevance for the VR tool subject of this study as it does not require any additional infrastructure and does not necessitate any integration with other IT systems in the hospital. Even though the hospital subject of this study was going through a major IT transformation at the time of this research, this did not seem to impact use or adoption, which may be due to the tool lacking the need for system integration.

Data-Related Factors

Data-related factors were also mentioned as a general challenge for eHealth adoption but were not particularly relevant for the tool subject of the case study, as it does not generate health data. While data quality, exchange, and storage were general considerations shared by policy experts, these were mostly not applicable to the VR tool in question.

User Experience

Regarding the design content and quality for patients, the participating nurses highlighted recent software updates of the visual simulation, which made the VR environments more realistic and thereby more immersive and enjoyable. The option to select specific scenarios based on personal preferences was especially appreciated by patients. User experience was generally positive, which suggests good content design and quality. This was further emphasized by the tool's developer, who shared insights into their design testing and optimization for ideal immersion. One negative aspect which the participating patient brought up, is that certain medical conditions impacting the mobility of the neck, may not be suitable for the application of VR in a lying position, as putting on the headset was experienced as uncomfortable.

Ease of Use

High design quality, especially regarding user guidance was indicated by the developer, as they specifically referred to user-centered design and how users are guided through the application step by step on the tablet steering the VR simulation. It was specified that even without training, a person less familiar with technology can handle the device. However, this is partially contradicted by the nurse perspective, which revealed that although the tool was perceived as intuitively usable by some, others had issues using the tool, especially in connection with their personal innovativeness. Specifically, the lack of technical affinity is mentioned to affect ease-of-use on a personal level.

The physician perspective on habit was discussed in relation to using VR tools with patients. The lack of a routine in using VR tools was attributed to workload and time constraints, as conventional medication procedures are habitual and potentially quicker. This issue could be generalized for other innovative treatments. However, the intuitive nature of the VR tool's handling suggests that it may not pose a significant barrier. From the clinician's viewpoint, the importance of external encouragement in breaking previous habits and adopting the tool was highlighted, though it was noted that this transition might require ongoing effort.

Monetary Factors

While reimbursement potentially finances the use of the device through clinicians and possibly amortizes the equipment costs over time, the initial cost and affordability of the device are not considered, neither are the maintenance and software updates costs. With the limited information at hand, the analysis remains inconclusive whether this factor may act as a barrier, especially as it is highly interrelated with reimbursement considerations.

Social and Personal Adoption Factors

Personal Characteristics

Personal characteristics were indicated to play a relevant role in successful adoption. Participating clinicians explained that patients' awareness and personal attitudes could be affected through persistent and educative communication. Comfort and acceptance were reported to be positively related to technical skills and experience, and are highly personal. Overall, personal characteristics can therefore be a barrier or a facilitator, depending on the user in question.

Social and Cultural Factors

Clinicians' endorsement is key for the acceptance of the tool and may positively influence individual decisions. The power of endorsement was highlighted by the technology provider and clinicians. It was emphasized that even only one promoting individual physician serving as a VR tool advocate can socially influence others to adopt VR as a tool for patients in pain, especially connected with a push toward a change of habits. Further, the importance of the team was highlighted regarding social influence. When no other team member uses the device, that may therefore inhibit adoption. Conversely, the participating patient voiced that they did not feel impacted by culture or their environment in their decision to use VR for pain management.

Moderating Factors

The moderating effects of gender and age were brought up by several participants, implying that younger age was frequently associated with higher comfort and acceptability as well as skills and attitude toward novel technologies. The participating nurses also confirmed that older patients generally require more on-boarding time. Additionally, male gender was associated with a higher openness to try VR for pain management, potentially due to personal experience of gaming and leisure time with VR applications.

Discussion

Most Prominent Barriers and Facilitators

Overview

The analysis showed that some factors had a more prominent facilitating or inhibiting impact on user adoption than others that had limited relevance for the VR tool subject of this case study. Therefore, the focus of the discussion will be the factors that had a clear impact in this specific case and their respective potential implications for practice.

Facilitator 1—Organizational Environment

The organizational environment in which the tool is implemented is considered a critical facilitator to eHealth adoption according to several published studies [25-28], a finding that was confirmed by the outcomes of this research. The participants acknowledged that the environment in the ED of the hospital under study was conducive to adoption. The overall culture of openness to innovation within the hospital was seen as a facilitator, particularly in terms of leadership commitment and the emphasis on training and educating staff in the use of new devices, consistent with findings from similar research [29,30].

Facilitator 2—Tension for Change, Ease of Use, and Demonstrability

This study's participants emphasized that a high tension for change toward more efficient ways of working and pain management is perceived as a facilitator, especially when the adoption of the VR tool allows nurses to focus on other tasks while the patient is immersed in VR and, at the same time, shows clinical effectiveness to reduce pain. Similar studies correspondingly suggested that this demonstrability of the tool's usefulness, especially in combination with ease of use may indeed positively impact adoption [25,26,31]. Technical affinity has been discussed in relation to the tool's ease of use; in alignment with similar studies, the easier a technology is perceived in use, the less technical affinity is required [32,33]. In this context, the ease of use of the VR tool under study has been perceived as a facilitator. Furthermore, immediate positive effects on patients after the application regarding anxiety and pain levels could be observed by clinical personnel and could also be proven by the hospital subject of this study [5]. This high level of demonstrability may be used to convince clinicians and patients alike that the VR tool is a valuable, easy-to-use option, especially as the tension for change in pain management within EDs is rising.

Facilitator 3—Employee Engagement

The clinicians participating in this study emphasized that employee engagement, particularly because the VR tool is primarily administered by the care team, is a crucial facilitator. Analysis of the interviews suggested that the new tool could have a positive impact on nurses' attitudes toward their profession by enriching their jobs. This finding aligns with other studies indicating that eHealth tools may empower and expand the roles of health care professionals in certain cases [34,35]. However, further insights revealed that as initial excitement for a technology subsides, this engagement may decrease if not actively managed.

Barrier 1—Workload

The first important adoption barrier according to this study's participants is the experienced workload, especially among nurses that administer the VR tool. As suggested by the literature, workload is a general concern in the health care sector [36,37], and at the same time, health care professionals are expected to increase their work quality and efficiency. This was confirmed by the interviewees to be an inhibiting factor of the implementation of a new tool such as VR for patients in pain. Especially since the tool is currently perceived as an additional task that consumes more time rather than saves time through the set-up of the patient, but especially given the patient education required to ensure willingness to use and overall satisfaction by the patient.

Barrier 2—Changes in Clinical Workflow and Habit

A further important barrier that could be observed is the required change in clinical work and therefore in the work habits of nurses and physicians, in alignment with previous studies that emphasized the workflow changes associated with eHealth use as a potential adoption challenge [38,39]. Changing the clinical workflow from the current standard of care's medication-only approach toward the usage of an additional, complementary tool, requires a substantial adjustment of the standard workflow and work habits of clinicians.

Barrier 3—Reimbursement

Finally, the current reimbursement situation in Switzerland is a very important potential barrier for the VR tool's adoption. The analysis of Swiss reimbursement structures in the outpatient setting of the ED showed that even if the tool fulfills the relevant WZW criteria for reimbursement and is therefore considered a mandatory service under the mandatory health insurance, it could currently not be efficiently reimbursed, given the lack of an appropriate position in TARMED [40]. This is primarily due to the current billing system not being properly maintained as it is to be discontinued [41]. Reimbursing through supplementary insurance is an option, but it would demand significant resources to establish agreements with various commercial insurance companies and would only reach a limited number of potential patients. Moreover, categorizing the tool as a noncompulsory treatment would impede reimbursement through mandatory health insurance in the future, as transitioning from noncompulsory to mandatory service is challenging.

Practical Implications

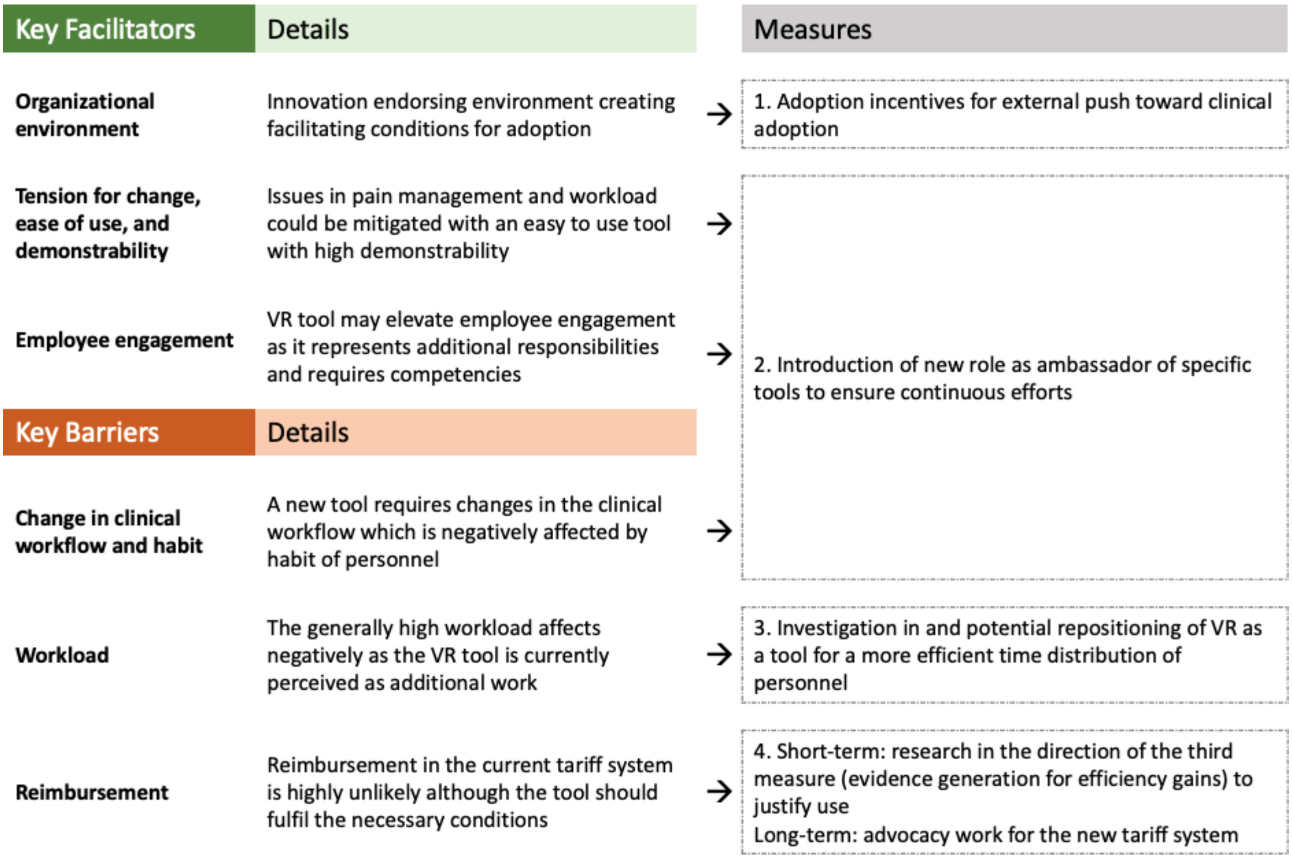
The analysis revealed a nuanced perspective for some factors, as they may act both as facilitators as well as barriers, depending on the specific circumstances such as personal characteristics of the user (patient and nurse) and future reimbursement landscape. This study's participants helped us identify three key facilitators and three key barriers. To enhance the impact of the key facilitators and mitigate the effects of the main barriers observed, four measures are proposed as visualized in Figure 2 and explained in the subsequent paragraphs. The numbers on the figure represent the four measures proposed to tackle the corresponding barriers and facilitators.

To foster the beneficial organizational impact, an incentive structure within the ED to drive the adoption of new tools, such as VR for pain management, allows for an external push toward adoption. The introduction of a specific ambassador role is proposed, as a further extrinsic push factor. The ambassador could capitalize on the high tension for change, the ease-of-use of the tool, and the demonstrability as arguments to remind the clinical staff of the valuable option of VR as a complementary pain management option. Moreover, the ambassador could demonstrate and accompany its use to increase employee engagement. Ambassadors could thereby also mitigate the barrier combination of workflow and habit, as their continuous efforts could promote clinical adoption.

To tackle the key barrier of workload, further investigation is proposed in the direction of the potential of the VR tool to support clinical staff in allocating their time as effectively as possible, thereby potentially changing the perception of the tool as additional workload. As patients are immersed in VR, this could potentially lead to less simultaneous attention requirements of the nurse, who is then able to focus on the remaining patients and tasks. Simultaneously, the application could lead to higher patient satisfaction.

Since it is unlikely that reimbursement will improve soon, it is recommended to explore how hospitals can financially benefit from adopting these tools. This could involve assessing their impact on workflow efficiency to see if they reduce workload barriers. This argument suggests that the hospital currently cannot bill the VR service to the patient’s health insurances to achieve reimbursement, but by increasing the effectiveness of the workflows in the ED, the hospital may eventually yield a positive return on investment. However, in the long term, investment in advocacy work by the hospital is suggested to shape the upcoming revision of the Swiss tariff system for outpatient treatment to enable adequate reimbursement of such mHealth tools in the future and have clear positions in the tariff system for VR tools in pain management.

Figure 2. Key facilitators, barriers, and proposed measures. VR: virtual reality.



Limitations and Recommendations for Future Research

This study has some limitations worth noting. We focused on a specific VR tool in one Swiss hospital during a set period of time, so it is hard to generalize our findings to other contexts that might have different characteristics, such as a different regulatory landscape. Although the target was to cover all relevant stakeholder perspectives on the matter, the recruitment of insurance professionals and patients for their perspectives was especially challenging, leading to only one interview for each of those stakeholder types. To minimize the selection bias that can be introduced by purposive sampling, we used snowballing; this involved asking participants to suggest other colleagues who were willing to participate. Lastly, while this research used the comprehensive list of factors in the consolidated framework for eHealth adoption by Jacob et al [16] as a basis for the interview guide and subsequent analysis, it observed connections between factors and disclaimed them when applicable. More research on their interrelatability may be useful when developing measures to address barriers and facilitators. Future studies could explore other tools in different settings to overcome some of these limitations.

Conclusions

Key facilitators and barriers identified in this study and the respective suggested measures may help improve adoption of the VR tool in the hospital subject of this study. Key facilitators include the supportive atmosphere in settings such as the hospital subject of this case study, which encourages VR adoption and is backed by leadership commitment and training; high demand for change in pain management and VR’s effectiveness and usability, which promote its adoption; and continuous support, which is crucial to sustain user engagement over time.

Key barriers include nurses' perception that the VR tool is adding to their workload, particularly in patient education and setup; incorporating the VR tool requires the staff to adapt their standard workflows, posing challenges; and the current reimbursement systems lack appropriate codes for VR services, hindering financial incentives.

To address these factors, some measures are recommended: establishing proper incentive structures can encourage VR adoption, ambassador roles can offer ongoing support and advocacy, further research into VR's impact on workflow efficiency is necessary, and advocacy efforts are needed to influence reimbursement system revisions. By leveraging facilitators and mitigating barriers, hospitals can optimize VR's benefits in pain management and enhance patient care.

Acknowledgments

The authors would like to express their gratitude to Mayella Favre (MAS in health economics and management) and Ulla Strobel (CAS in digital health) for generously sharing their in-depth expertise of the Swiss reimbursement system with the research team.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview topic guide.
[PDF File, 181 KB - humanfactors_v11i1e59073_app1.pdf]

Multimedia Appendix 2
Study information sheet.
[PDF File, 75 KB - humanfactors_v11i1e59073_app2.pdf]

Multimedia Appendix 3
Direct quotes from the participants supporting the narrative synthesis of the different themes reported in this study (translated from German).
[PDF File, 153 KB - humanfactors_v11i1e59073_app3.pdf]

References

1. Van Woerden G, Van Den Brand CL, Den Hartog CF, Idenburg FJ, Grootendorst DC, Van Der Linden MC. Increased analgesia administration in emergency medicine after implementation of revised guidelines. Int J Emerg Med 2016 Dec;9(1):4. [doi: 10.1186/s12245-016-0102-y] [Medline: 26860533]
2. Guéant S, Taleb A, Borel-Kühner J, et al. Quality of pain management in the emergency department: results of a multicentre prospective study. Eur J Anaesthesiol 2011 Feb;28(2):97-105. [doi: 10.1097/EJA.0b013e3283418fb0] [Medline: 21119516]
3. Hachimi-Idrissi S, Coffey F, Hautz WE, et al. Approaching acute pain in emergency settings: European Society for Emergency Medicine (EUSEM) guidelines-part 1: assessment. Intern Emerg Med 2020 Oct;15(7):1125-1139. [doi: 10.1007/s11739-020-02477-y] [Medline: 32930965]
4. Hachimi-Idrissi S, Dobias V, Hautz WE, et al. Approaching acute pain in emergency settings; European Society for Emergency Medicine (EUSEM) guidelines-part 2: management and recommendations. Intern Emerg Med 2020 Oct;15(7):1141-1155. [doi: 10.1007/s11739-020-02411-2] [Medline: 32930964]
5. Birrenbach T, Bühlmann F, Exadaktylos AK, Hautz WE, Müller M, Sauter TC. Virtual Reality for Pain Relief in the Emergency Room (VIPER) - a prospective, interventional feasibility study. BMC Emerg Med 2022 Jun 21;22(1):113. [doi: 10.1186/s12873-022-00671-z] [Medline: 35729502]
6. Abdolrazaghnejad A, Banaie M, Tavakoli N, Safdari M, Rajabpour-Sanati A. Pain management in the emergency department: a review article on options and methods. Adv J Emerg Med 2018;2(4):e45. [doi: 10.22114/AJEM.v0i0.93] [Medline: 31172108]
7. Lowood HE. Virtual reality. Britannica. 2024. URL: https://www.britannica.com/technology/virtual-reality [accessed 2024-11-07]
8. Zolfaghari E, Ridout B, Medlow S, et al. Exploring the use of virtual reality to manage distress in adolescent patients in emergency departments: a feasibility study. Emerg Med Australas 2022 Oct;34(5):687-693. [doi: 10.1111/1742-6723.13945] [Medline: 35238143]
9. Sikka N, Shu L, Ritchie B, Amdur RL, Pourmand A. Virtual reality-assisted pain, anxiety, and anger management in the emergency department. Telemed J E Health 2019 Dec;25(12):1207-1215. [doi: 10.1089/tmj.2018.0273] [Medline: 30785860]

10. Hoffman HG, Richards TL, Van Oostrom T, et al. The analgesic effects of opioids and immersive virtual reality distraction: evidence from subjective and functional brain imaging assessments. Anesth Analg 2007 Dec;105(6):1776-1783. [doi: 10.1213/01.ane.0000270205.45146.db] [Medline: 18042882]
11. Gupta A, Scott K, Dukewich M. Innovative technology using virtual reality in the treatment of pain: does it reduce pain via distraction, or is there more to it? Pain Med 2018 Jan 1;19(1):151-159. [doi: 10.1093/pm/pnx109] [Medline: 29025113]
12. Li A, Montaño Z, Chen VJ, Gold JI. Virtual reality and pain management: current trends and future directions. Pain Manag 2011 Mar;1(2):147-157. [doi: 10.2217/pmt.10.15] [Medline: 21779307]
13. Hoffman HG, Chambers GT, Meyer WJ 3rd, et al. Virtual reality as an adjunctive non-pharmacologic analgesic for acute burn pain during medical procedures. Ann Behav Med 2011 Apr;41(2):183-191. [doi: 10.1007/s12160-010-9248-7] [Medline: 21264690]
14. Strong J, Van Griensven H, Unruh A. Pain: A Textbook for Health Professionals: Churchill Livingstone; 2013.
15. Therapeutic virtual reality to reduce pain and anxiety. Healthy Mind. 2024. URL: https://healthymind.fr/en/accueil-en [accessed 2024-11-07]
16. Jacob C, Sanchez-Vazquez A, Ivory C. Understanding clinicians’ adoption of mobile health tools: a qualitative review of the most used frameworks. JMIR mHealth uHealth 2020 Jul 6;8(7):e18072. [doi: 10.2196/18072] [Medline: 32442132]
17. Leonardi PM. Methodological guidelines for the study of materiality and affordances. Routledge Comp to Qual Res Org Stud 2017:279-290. [doi: 10.4324/9781315686103-18]
18. Creswell J, Poth C. Qualitative Inquiry and Research Design: Choosing Among Five Approaches: SAGE Publications, Inc; 2017.
19. Braun V, Clarke V. What can “thematic analysis” offer health and wellbeing researchers? Int J Qual Stud Health Well-being 2014;9:26152. [doi: 10.3402/qhw.v9.26152] [Medline: 25326092]
20. Yin R. Case Study Research and Applications: Design and Methods, 6th edition: SAGE Publications, Inc; 2017.
21. Braun V, Clarke V. Successful Qualitative Research: A Practical Guide For Beginners: SAGE Publications Ltd; 2013. [doi: 10.1002/jmr.2361]
22. The effective health care program stakeholder guide. : Agency for Healthcare Research and Quality; 2014 14-EHC010-EF.
23. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
24. Health insurance. FOPH - Federal Office of Public Health Health Insurance in Switzerland. 2024. URL: https://www.bag.admin.ch/bag/en/home/versicherungen/krankenversicherung.html [accessed 2024-11-07]
25. Jacob C, Sanchez-Vazquez A, Ivory C. Clinicians’ role in the adoption of an oncology decision support app in Europe and its implications for organizational practices: qualitative case study. JMIR mHealth uHealth 2019 May 3;7(5):e13555. [doi: 10.2196/13555] [Medline: 31066710]
26. Jacob C, Sanchez-Vazquez A, Ivory C. Factors impacting clinicians’ adoption of a clinical photo documentation app and its implications for clinical workflows and quality of care: qualitative case study. JMIR mHealth uHealth 2020 Sep 23;8(9):e20203. [doi: 10.2196/20203] [Medline: 32965232]
27. Gagnon MP, Orruño E, Asua J, Abdeljelil AB, Emparanza J. Using a modified technology acceptance model to evaluate healthcare professionals’ adoption of a new telemonitoring system. Telemed J E Health 2012;18(1):54-59. [doi: 10.1089/tmj.2011.0066] [Medline: 22082108]
28. Ehrler F, Ducloux P, Wu DTY, Lovis C, Blondon K. Acceptance of a mobile application supporting nurses workflow at patient bedside: results from a pilot study. Stud Health Technol Inform 2018;247:506-510. [Medline: 29678012]
29. Varsi C, Ekstedt M, Gammon D, Ruland CM. Using the consolidated framework for implementation research to identify barriers and facilitators for the implementation of an internet-based patient-provider communication service in five settings: a qualitative study. J Med Internet Res 2015 Nov 18;17(11):e262. [doi: 10.2196/jmir.5091] [Medline: 26582138]
30. Bagot KL, Cadilhac DA, Bladin CF, et al. Integrating acute stroke telemedicine consultations into specialists’ usual practice: a qualitative analysis comparing the experience of Australia and the United Kingdom. BMC Health Serv Res 2017 Nov 21;17(1):751. [doi: 10.1186/s12913-017-2694-1] [Medline: 29157233]
31. Adenuga KI, Iahad NA, Miskon S. Towards reinforcing telemedicine adoption amongst clinicians in Nigeria. Int J Med Inform 2017 Aug;104:84-96. [doi: 10.1016/j.ijmedinf.2017.05.008] [Medline: 28599820]
32. Jacob C, Sanchez-Vazquez A, Ivory C. Social, organizational, and technological factors impacting clinicians’ adoption of mobile health tools: systematic literature review. JMIR mHealth uHealth 2020 Feb 20;8(2):e15935. [doi: 10.2196/15935] [Medline: 32130167]
33. Jacob C, Sezgin E, Sanchez-Vazquez A, Ivory C. Sociotechnical factors affecting patients’ adoption of mobile health tools: systematic literature review and narrative synthesis. JMIR mHealth uHealth 2022 May 5;10(5):e36284. [doi: 10.2196/36284] [Medline: 35318189]
34. O’Connor S, Andrews T. Smartphones and mobile applications (apps) in clinical nursing education: a student perspective. Nurse Educ Today 2018 Oct;69:172-178. [doi: 10.1016/j.nedt.2018.07.013] [Medline: 30096510]
35. Li L, Cotton A. A systematic review of nurses’ perspectives toward the telemedicine intensive care unit: a basis for supporting its future implementation in China? Telemed e-Health 2019 May;25(5):343-350. [doi: 10.1089/tmj.2018.0006]

36. Schmalz O, Jacob C, Ammann J, et al. Digital monitoring and management of patients with advanced or metastatic non-small cell lung cancer treated with cancer immunotherapy and its impact on quality of clinical care: interview and survey study among health care professionals and patients. J Med Internet Res 2020 Dec 21;22(12):e18655. [doi: 10.2196/18655] [Medline: 33346738]
37. Öberg U, Orre CJ, Isaksson U, Schimmer R, Larsson H, Hörnsten Å. Swedish primary healthcare nurses' perceptions of using digital health services in support of patient self‐management. Scand J Caring Sci 2018 Jun;32(2):961-970. [doi: 10.1111/scs.12534] [Medline: 28960451]
38. Giraldo L, Schachner B, Luna D, Benítez S. Exploring nurses' perceptions and expectations toward a BMCA implementation using a mobile app and workstations as a change management strategy. Stud Health Technol Inform 2018;250:134-138. [Medline: 29857405]
39. Penny RA, Bradford NK, Langbecker D. Registered nurse and midwife experiences of using videoconferencing in practice: a systematic review of qualitative studies. J Clin Nurs 2018 Mar;27(5-6):e739-e752. [doi: 10.1111/jocn.14175] [Medline: 29149507]
40. Health insurance: benefits and tariffs. Federal Office of Public Health FOPH. 2023. URL: https://www.bag.admin.ch/bag/en/home/versicherungen/krankenversicherung/krankenversicherung-leistungen-tarife.html [accessed 2024-11-07]
41. Tarif 10.07.2022 Tarmed 01.09_br (KVG) online browser - v.01.09. FMH the Swiss Medical Association Änderungen der Tarifstruktur folgen den Vorgaben der Verordnung des Bunderates über die Anpassung von Tarifstrukturen in der Krankenversicherung. 2023. URL: https://browser.tartools.ch/de/tarmed_kvg [accessed 2024-11-07]

Abbreviations

CE: conformité européenne
ED: emergency department
MiGeL: mittel und gegenständeiste
VR: virtual reality
WZW: wirksamkeit, zweckmässigkeit, und wirtschaftlichkeit

Edited by A Kushniruk, E Borycki; submitted 01.04.24; peer-reviewed by T Dege, YT Kuan; revised version received 12.06.24; accepted 13.09.24; published 04.12.24.

Please cite as:
Lurtz J, C Sauter T, Jacob C
Factors Impacting the Adoption and Potential Reimbursement of a Virtual Reality Tool for Pain Management in Switzerland: Qualitative Case Study
JMIR Hum Factors 2024;11:e59073
URL: https://humanfactors.jmir.org/2024/1/e59073
doi:10.2196/59073

Optimizing a Novel Smartphone App to Prevent Postpartum Depression Adapted From an Evidence-Based Cognitive Behavioral Therapy Program: Qualitative Study

Adam K Lewkowitz[1,2], MPHS, MD; Melissa Guillen[1], BA; Katrina Ursino[1], BA; Rackeem Baker[1], BA; Liana Lum[1], BA; Cynthia L Battle[3], PhD; Crystal Ware[1], BSN; Nina K Ayala[1], MD, SCM; Melissa Clark[1,4], PhD; Megan L Ranney[5], MPH, MD; Emily S Miller[1,3,4], MPH, MD; Kate M Guthrie[3], PhD

[1]Department of Obstetrics and Gynecology, Warren Alpert Medical School at Brown University, 222 Richmond Street, Providence, RI, United States

[2]Center for Digital Health, Brown School of Public Health, Providence, RI, United States

[3]Department of Psychiatry and Human Behavior, Warren Alpert Medical School at Brown University, Providence, RI, United States

[4]Department of Health Services Policy & Practice, Brown School of Public Health, Providence, RI, United States

[5]Yale School of Public Health, New Haven, CT, United States

Corresponding Author:
Adam K Lewkowitz, MPHS, MD
Department of Obstetrics and Gynecology, Warren Alpert Medical School at Brown University, , 222 Richmond Street, Providence, RI, , United States

Abstract

Background: Low-income pregnant patients are at high risk of postpartum depression (PPD). Mothers and Babies (MB) is a cognitive behavioral therapy–based program that prevents up to 50% of de novo PPD when provided in person to low-income Spanish- and English-speaking people who are pregnant without depression. MB is limited by the need for trained personnel to support it. Transforming MB into a smartphone app may mitigate this key barrier.

Objective: We aimed to use qualitative data from target end users to create and optimize MBapp, a novel app centered on the MB program.

Methods: Draft wireframes of MBapp were created in English and Spanish with cognitive behavioral therapy–based modules adapted from MB. These wireframes included several features shown previously to sustain app engagement: (1) push notifications delivered at participant-preferred times; (2) text-, graphic-, and video-based content; and (3) gamification with digital rewards for app engagement. English- or Spanish-speaking individuals with public health insurance who were between 32 weeks gestation and 6 months post partum and owned smartphones were eligible to consent for individual in-depth interviews. Individuals with prior or current depression were excluded. Interviews were recorded, transcribed, and analyzed using deductive and inductive codes to characterize opinions about MBapp and perceptions of challenges and facilitators of use of MBapp or other perinatal or mental health apps. End user feedback led to major modifications to the wireframes. Each of these changes was categorized according to the FRAME (Framework for Modification and Adaptation), an established method of systematically reporting adaptations and modifications to evidence-based interventions via end user feedback. Recruitment ceased with content saturation, defined as 3 successive participants providing only positive feedback on MBapp's wireframe, without further suggestions for improvement.

Results: A total of 25 interviews were completed. Participants were racially and ethnically diverse, generally representing our target end user population, and 48% (n=12) of interviews were conducted in Spanish. Participants' suggestions to improve MBapp were categorized within the FRAME as adaptations that improved either content or context to optimize reach, retention, engagement, and fit for end users. Specifically, the following features were added to MBapp secondary to end user feedback: (1) audio narration; (2) "ask a clinician" nonurgent questions; (3) on-demand module summaries accessible upon module completion; and (4) choice to defer assessments and start the next module. Participants also provided insights into features of perinatal or mental health apps they found appealing or unappealing to understand preferences, challenges, and negotiables or nonnegotiables for MBapp.

Conclusions: Adapting MBapp to incorporate end users' perspectives optimized our digital PPD prevention intervention, ideally increasing its appeal to future users. Our team's next steps will confirm that MBapp is a feasible, acceptable intervention among English- and Spanish-speaking perinatal people at risk of PPD.

(JMIR Hum Factors 2024;11:e63143) doi:10.2196/63143

KEYWORDS

cognitive behavioral therapy; mothers and babies program; digital health; postpartum depression; smartphone application; FRAME for intervention adaptation; Framework for Modification and Adaptation; behavioral therapy; mental health apps

Introduction

Approximately 15% of all pregnant women in the United States develop postpartum depression (PPD) [1-6], and more than 95% of people with PPD have persistent symptoms [7]. The reasons for such a high burden of persistent symptomatic burden from PPD are complex but include the following barriers: (1) obstetric providers who are not trained to care for psychiatric conditions and who do not routinely screen and treat for mental health conditions such as PPD; (2) a shortage of trained mental health providers to care for those who screen positive; and (3) stigma of receiving in-person mental health care, which prevents those with PPD from receiving effective treatment [8-14].

Thus, preventing PPD from occurring is crucial, particularly among individuals with unmet social needs such as financial stress or food insecurity that place them at higher risk for this condition than those without unmet social needs [1-6]. The US Preventive Services Task Force recommends that women at risk for PPD participate in 1 of 2 evidence-based psychotherapy programs shown to effectively prevent PPD from developing [15,16]. One of these programs is an in-person cognitive behavioral therapy (CBT)–based program, Mothers and Babies (MB) [17-19]. MB was created to prevent PPD among low-income English- or Spanish-speaking pregnant women and has been shown to be effective in diverse communities of low-income perinatal patients [17-19]. Though MB has been adapted to an asynchronous online intervention [18,20], access to the program generally remains limited to people who receive prenatal care in clinical settings affiliated with MB-trained providers, reducing MB's dissemination potential.

App-provided mental health care has dramatically expanded access to mental health services through their ability to be scaled without reducing fidelity and their ability to reduce barriers and potentially stigma related to receiving in-person mental health care [8,21]. In addition, app-provided mental health care has been shown to effectively treat depressive symptoms and is commonly used in pregnancy (the majority of people who are pregnant download apps regardless of income) [22-24]. From a perinatal patient perspective, engaging with an app to improve somatic and mental health symptoms has become commonplace [25,26]. However, most mental health apps do not contain evidence-based psychotherapy [27,28], and, of the apps that do contain CBT elements, most do not contain content that has been specifically created for peripartum people [29]. For these reasons, we aimed to adapt the evidence-based MB course, ultimately creating a 9-module, app-based intervention, entitled MBapp, from the original 6-session in-person group intervention. In this paper, we present findings from individual in-depth interviews with pregnant or postpartum individuals regarding optimal perinatal or mental health apps. We also use an established method of systematic reporting of adaptations and modifications to evidence-based interventions via end user feedback to describe how incorporating end users' suggestions optimized MBapp for future pregnant or postpartum users. To our knowledge, this study represents the first to systematically report the process of adapting an existing in-person curriculum to a perinatal mental health app before the intervention was deployed.

Methods

Creating MBapp Draft Wireframe

MB materials are publicly available [30]. Once downloaded, MB's curriculum for course leaders and the participant workbook served as the starting point for MBapp. The draft wireframe was created by this study's principal investigator (AKL), who was supported by his coauthors—a multidisciplinary team with expertise in perinatal mental health, digital health interventions, and qualitative research methods. To create the draft wireframe, MB's curriculum was first adapted to be delivered asynchronously on a smartphone app. Prior digital adaptations of MB reported that end users requested additional alterations to the digital draft content by further simplifying language and adding more visuals [20,31,32]. Thus, in MBapp's initial draft wireframe, paragraphs of text or worksheets in MB were re-restructured to include short, direct phrases, colorful graphics (some from the MB curriculum), and pictures of pregnant and postpartum people and young infants. Some of MB's videos were included within MBapp to allow users to learn the curriculum via a combination of text, graphics, and videos. The draft wireframe also contained ecological momentary assessments and gamification, features that have been shown in other perinatal mental health apps to increase the likelihood of sustained user engagement [33-35]. Specifically, MBapp included daily push notifications at users' preferred times that prompt them to complete a mood scale. Their scores on this scale send them to pertinent content within the app. In terms of gamification, responding to at least 3 of the 7 daily ecological momentary assessment requests to complete the mood scale generated a digital reward, which can be visualized in a digital trophy case. Original MB creators confirmed the draft wireframe was adherent to CBT principles and retained fidelity to MB. Once the draft wireframe was completed, qualitative feedback was obtained from target end users to further adapt the intervention.

Study Population

Participants were recruited from among those receiving perinatal care at a clinic that exclusively serves patients with government pregnancy-related health insurance (Medicaid), as having perinatal Medicaid insurance is only available for low-income patients who are pregnant in our state (Rhode Island, United States) [36]. To be eligible for this study, participants were English- or Spanish-speaking, aged at least 18 years, owned a smartphone, and were either in the third trimester of pregnancy or had an infant that was born within the prior 6 months. As MB prevents—but does not treat—PPD, volunteers were excluded if they reported prepregnancy or current anxiety or depression, an active (or within last 5 years) prescription for psychotropic medications, active engagement in psychotherapy, untreated substance use disorder, or if they were unable to consent. Potential participants were approached in person during routine antenatal or postpartum care visits. Those interested, who remained eligible after screening, provided written informed consent in person before the interview was scheduled. Participants who did not complete their interview after 3 no-shows at interview appointments were not rescheduled.

Data Collection

Individual in-depth interviews were all conducted via video teleconferencing at the participant's preferred schedule and in their preferred language (English or Spanish). Each interview was conducted by the same investigator (AKL), who followed a semistructured interview agenda that divided the interview into two components. First, the interview discussed the participant's personal experiences, preferences, and challenges with apps, particularly those that focus on perinatal or mental health. Then, the investigator shared his computer screen to review distinct components of MBapp's wireframe with the participant and obtain page-by-page feedback. All interviews were audio-recorded, transcribed by a professional transcription service, and reviewed for accuracy with identifiers redacted. Participants were asked their race and ethnicity.

Using an Evidence-Based Framework to Determine Sample Size

Similar to other web-based digital adaptations of MB [18,20], a human-centered design approach was used to optimize MBapp based on end user feedback about the intervention itself and other, optimal perinatal or mental health apps. We recorded MBapp's evolution according to the FRAME (Framework for Modification and Adaptation) [37], an established method of systematic reporting adaptations and modifications to evidence-based interventions via end user feedback. The FRAME provides insight about when MBapp's wireframe was modified during the development process and the reason for the modification. Feedback on MBapp's themes, content, and function were extracted from the data and, after 5 to 7 interviews, the draft wireframes were revised to incorporate this feedback. The new iteration of the wireframe was then presented to subsequent participants to allow revisions suggested by earlier participants to be seen and commented on by subsequent participants. We anticipated recruiting up to 36 participants and planned to cease recruitment with content saturation, defined as 3 successive participants providing only positive feedback on MBapp's wireframe, without further suggestions for improvement.

Data Analysis

Authors AKL, MG, and KMG generated a preliminary codebook for the interviews based on the semistructured interview agenda. This initial version of the codebook was limited to deductive codes. Authors AKL, MG, KU, RB, and LL then collaborated via consensus coding in stages. First, two transcripts were coded collaboratively, and additional codes were added inductively to the codebook. Some of these codes were added de novo whereas other inductive codes were based on deductive codes that were reorganized to better capture data. Second, each of these authors independently coded a third transcript, and the authors met as a group to reconcile coding discrepancies. In this manner, consensus coding was achieved and a codebook comprising deductive and inductive codes was finalized. Then, all transcripts—including those that had already been reviewed—were coded independently by two authors (AKL and either KU, RB, or LL). A third coauthor reviewed transcripts and codes when coding disagreements could not be resolved (MG). All coding was then inputted into NVivo (Lumivero), and coded excerpts were organized and synthesized into thematic memos and discussed by the coding team. This study was reviewed and approved by Women & Infant's Hospital of Rhode Island's institutional review board prior to the consenting of the first participant.

Ethical Considerations

This study was approved by the institutional review board of Women & Infants Hospital of Rhode Island (submission WIH 21-0021; approved on September 7, 2021). All participants were assigned a study identification number, which was stored on a password protected file and was only utilized to coordinate the $50 electronic gift card, which was provided to each participant after the study interview was completed. Otherwise, all data generated during the interview was deidentified and stored in a password protected computer network.

Results

Characteristics of Study Participants

From August 2022 to August 2023, a total of 33 patients were consented, and 25 individual in-depth interviews were conducted. There were no differences in language of scheduled interview, patient-reported race, or parity between those who consented and did versus did not complete the interview. The sociodemographic and obstetric characteristics of participants who completed their interviews are described in Table 1. In brief, 24% of participants were aged 35 years or older, 24% self-reported as White, and nearly half (n=12, 48%) of interviews were conducted in Spanish. Approximately the same proportion of participants were pregnant versus post partum, and over half of participants were first-time parents (ie, either about to deliver or had recently birthed their first child).

Table . Sociodemographic and obstetric characteristics of study participants (N=25).

Characteristic	Value
Age (years)	
Median (IQR)	29 (24-35)
Minimum, Maximum	20, 43
Advanced maternal age (35 years or older at time of delivery), n (%)	6 (24)
Race, n (%)	
American Indian or Alaska Native	3 (12)
Asian	2 (8)
Black	5 (20)
White	6 (24)
Other[a]	9 (36)
Ethnicity, n (%)	
Hispanic	13 (52)
Language of interview[b], n (%)	
English	13 (52)
Spanish	12 (48)
Primiparous (pregnant with first child or recently delivered first child), n (%)	13 (52)
Pregnancy status, n (%)	
Pregnant	13 (52)
Postpartum	12 (48)

[a]Other: multiracial (White and Black): 4; multiracial (White and American Indian or Alaska Native): 3; multiracial (Black and American Indian or Alaska Native): 2.

[b]Four participants spoke a combination of English and Spanish. Their interview language was categorized as the language that most of the interview was conducted in.

Perspectives on Optimal Features Within Perinatal or Mental Health Apps

Twenty-one of the 25 participants reported using at least one pregnancy-related app, and 2 of the 25 participants accessed a mental health related app during the perinatal period. The general consensus among perinatal app users was they were more likely to engage with an app if it provided pertinent, updated, and easy-to-understand text and video education either via a push notification or an email and contained digital rewards or games. Many perinatal app users noted that they turned to apps to receive perinatal education between prenatal care appointments or to supplement perinatal education received during these appointments:

> *The doctor focuses on talking about the pregnancy, focuses on the baby or they just look at your vagina and then [participant waves arms to mimic dismissing someone, as if doctor had instructed her to leave the room]. It was not much about me…[My app] helped me learn about what I wanted to know* [Postpartum, English-speaking, first-time mother]

Another consistent theme that emerged was that many participants engaged with perinatal apps to learn from people similar to them (those who were pregnant or had recently given birth):

> *[After my baby was born] they would call me, maybe it was a doctor or nurse, and they would ask me questions, but it's not the same. I can't explain it. It's not the same as seeing a mom that is the same as me [who] is saying, "Look, do this. No, do this. Look," and they show me with their own baby.* [Pregnant, Spanish-speaking mother with prior children]

Though apps were consistently described as useful throughout the perinatal period, the majority of users who had delivered at least one child at the time of the interview noted that apps were particularly helpful after delivery. The need for perinatal apps in the postpartum period was described in language suggesting that apps help fill the gap between in-person medical care from birth until 6 weeks after delivery. The long interval between medical appointments left participants feeling unsupported by their prenatal care team, and apps helped to provide support:

> *[When I left the hospital] they did give me a pamplet that had some…breastfeeding classes and stuff like that [but] I lost the paper…Now I have questions [on breastfeeding] and don't have anywhere to go, but wait for the next doctor appointment… [My app] gives*

> *an idea of how much milk I should give him. Super helpful.* [Postpartum, English-speaking first-time mother]

While prenatal apps were often used for additional perinatal education, there was a recurrent theme that these apps—such as prenatal care—focus more on the baby than on the mother:

> *Apps for mother and baby focus more on the baby than the mother…[just like in real life]. An example, when the mother gives birth to the baby, everyone is concerned about the baby. Hardly anybody worries about the mother, but the mothers go through a difficult process.* [Pregnant, Spanish-speaking mother with prior children]

Participants in this sample were much less likely to use mental health apps than perinatal apps. Mostly, participants attributed this difference to a lack of knowledge that such resources were available, not due to lack of interest in perinatal mental health apps. For example, one participant shared that she had depressive symptoms after delivery and informed her midwife of these symptoms at her postpartum visit:

> *She told me it was normal. She [gave] me a sheet of paper to…see if I had suicidal thoughts, if the depression was severe…and things like that. [Then] she told me to find a counselor. That I needed one. I started to look for a counselor…I wish she had told me about this [mental health apps]. I would have used one and found it useful.* [Postpartum, Spanish-speaking first-time mother]

Once the interview shifted toward talking about how MBapp aimed to provide app-based mental health support for perinatal people, the majority of participants responded very favorably, even before viewing any of MBapp's wireframe:

> *I feel like an app that talks about mental health and they make it fun with the game and the video and the colors and everything, I think it could be useful. I would've liked something like that—or even now, I would like it.* [Pregnant, English-speaking first-time mother]

Adapting the MBapp Wireframe

The MBapp wireframe was adapted in multiple ways through end user feedback. Participants were asked about specific features in perinatal or mental health apps, and whether each feature was appealing (or not) and useful (or not). In addition, each participant viewed at least two distinct sections of MBapp's wireframe and provided direct feedback on the appearance, appeal, content, and usability of each page. This feedback led to minor changes in MBapp's font size or color scheme. Each participant also was asked general questions on how MBapp should be modified to encourage initial and sustained user engagement from the third trimester until 6 months post partum. All feedback from participants was considered by the study team and adaptations were incorporated into MBapp's wireframe. Table 2 highlights the major adaptations made to MBapp's wireframe during qualitative interviews and categorizes these changes according to the FRAME [37]. Of note, participant feedback resulted in both significant and nuanced changes to MBapp but did not alter MBapp's core CBT content; thus, these adaptations did not affect MBapp's fidelity to the evidence-based MB curriculum.

MBapp's wireframe was originally created to mirror MB's programmatic structure, in which a module on parenting education and infant care occurred midway through the curriculum and participants were instructed to complete daily mood scales between each in-person session, regardless of content. Based on end user feedback, MBapp was restructured such that the parenting module was first. We also edited the wireframe to provide users with the option to proceed with the daily mood scales for one week before accessing the next module or to defer these assessments in lieu of moving on immediately to the next module.

Participants recommended adding multiple new features to increase the likelihood of sustained engagement. These included modest financial reimbursement tied to responding to the daily mood scale and the option for audio narration, as well as an "ask a clinician" feature in which users can ask a nonurgent question to a member of the MBapp clinical team, who would respond through the platform. End users also desired short module summaries to be added after each module so they could review high-yield content without repeating the entire module. Module summaries now appear automatically on a digital bookshelf upon completion of each module. Figure 1 provides an example of how MBapp's wireframe was adapted from the initial to final versions.

Table . Major end user–requested adaptations to MBapp's wireframe, categorized according to the FRAME[a] [37].

MB[b] feature	MBapp adaptation	Classification[c]	Reasons[c]	Nature of the content modification[c]	Timing of change implementation	Participant recommendation addressed
Not in MB	"Ask a clinician" nonurgent questions	Content: adding elements	Recipient: access to resources	Improve fit with recipients	After interview #16	Resource for perinatal or mental health advice, support, or education
Not in MB	Audio narration, to be turned on or off on each page	Context: adding elements	Recipient: first or spoken languages	Improve fit with recipients	After interview #11	Increase engagement with MBapp via additional method to deliver content
Mood assessments given between in-person classes	Choice to defer week of mood assessments and pass immediately to next module	Context: format	Recipient: cognitive capacity	Increase retention	After interview #11	Flexible schedule
Parenting education module in middle of curriculum	MBapp module order restructuring, with parenting module in the beginning	Content: reordering of intervention modules or segments	Recipient: cognitive capacity	Improve fit with recipients	After interview #6	Optimize MBapp flow by removing parenting education from middle of curriculum
Not in MB	Module summaries to access high-yield content after module is completed	Content: adding elements	Recipient: cognitive capacity	Increase retention	After interview #6	Practical, high-yield content for fast review
Not in MB	Small financial reimbursement based on number of responses to ecological momentary assessment prompts	Content: tailoring, tweaking, and refining	Recipient: motivation and readiness	Increase reach or engagement	After interview #21	Increase meaningful engagement with MBapp via gamification

[a]FRAME: Framework for Reporting Adaptations and Modifications.
[b]MB: Mothers and Babies.
[c]As defined by FRAME [37].

Figure 1. Visual representation of major adaptations to MBapp's draft wireframe (left) and final wireframe (right), using the digital trophy case as an example. In MBapp's draft wireframe (left), the graphic was smaller; the digital rewards were trophies; and the icons in the header (which appear on every page) were limited to the table of contents, the digital reward icon, and the bookshelf (from left to right). Also, modules 4 and 5 were mislabeled. The final wireframe (right) represents adaptations to MBapp due to end user feedback, which included larger graphic size, medals as digital rewards, more text description for each module, less text description on the page itself (the text is spread over to two pages), and additional icons in the header (audio feature is in the middle, and the question mark icon is how users ask a provider a nonurgent question). The updated header appears on every page.

Discussion

Principal Findings

To our knowledge, this study represents the first to systematically report the process of adapting an in-person curriculum to a perinatal mental health app before the intervention was deployed. Indeed, this paper presents qualitative data from 25 diverse pregnant or postpartum perinatal patients with Medicaid insurance, describing their perspectives on perinatal or mental health apps and design recommendations for MBapp, a novel app based on an in-person CBT program shown to effectively prevent PPD among low-income patients [15,16]. The majority of participants used perinatal apps to receive on-demand pregnancy or postpartum education, and many preferred learning from others with similar experiences. Comparatively, mental health app use was less common, but this lack of uptake was attributed to lack of awareness, not disinterest.

Participants' perspectives and recommendations were incorporated into MBapp's wireframe, which led to minor adjustments in MBapp's aesthetics and major adaptations to MBapp's structure and features. All major adaptations were categorized according to the FRAME [37]. As each modification was derived directly from qualitative feedback from individual end users, incorporating minor and major adaptations to MBapp's appearance, structure, and features should increase its appeal to future users.

Comparison With Prior Work

MB was originally created as 6 in-person 2-hour group sessions [17] and has been implemented successfully in nontrial, community settings [38]. Since its inception, MB has been adapted into multiple formats. For example, 1 in-person MB curriculum comprises 10 individual sessions occurring during prenatal visits or at home [19] and can include supplemental text messages to participants [32]. There are also two asynchronous, online adaptations of MB, one for perinatal women and one for pregnant adolescents at risk for PPD [20]. Each digital adaptation of MB (text-based or online) included mixed methods research to optimize the intervention according to end user perspectives, and this feedback led to simplified language and more visuals in the final digital MB intervention [20,31,32]. Based on prior investigators' experience with adapting MB to a digital intervention, we assumed that adapting

MBapp would also need to include simplified and direct language and visual engagement via graphics and pictures of perinatal individuals and infants. The initial wireframe was developed with these features. Thus, all participants could provide granular feedback on features that were included or should be added to MBapp to encourage engagement, instead of focusing on giving the investigative team feedback on reducing the amount of text or adding visual appeal to the wireframe. It is important to note that only the online MB program for high-risk perinatal adolescents used formative qualitative interviews to adapt MB via a structured framework such as the FRAME, as is described in this paper. Incorporating this evidence-based approach ensured that all major MBapp modifications were well-justified and provided transparency to inform how future data on MBapp's effectiveness may be interpreted.

Strengths and Limitations

Our study had many strengths. Each aspect of intervention creation and adaptation was grounded in evidence-based practices: (1) MBapp's core content was derived from an in-person CBT curriculum known to be highly effective for PPD prevention [17-19]; (2) the individual in-depth interviews were performed according to established and recommended qualitative research practices as described by coauthors of our paper [39-41]; and (3) MBapp was adapted using a standard, commonly used framework for modifying established interventions [37]. In addition, our participants were diverse in terms of race, ethnicity, language, and parity, helping to ensure that different perspectives and opinions were included in our digital perinatal mental health intervention. Lastly, MBapp was created with a multidisciplinary team with expertise in perinatal mental health, digital health, and qualitative research, ensuring the digital health intervention was created via best practices.

Nevertheless, our study is not without limitations. First, our study population was comprised of 25 participants. Though a standard sample for a qualitative research study, this study population cannot possibly represent all opinions and perspectives of perinatal patients. Second, recruitment was limited to one obstetric clinic, which may reduce transferability of findings to individuals within different types of clinics (eg, academic vs community-based obstetric clinics). Third, MBapp will be a fully functional native app available on iTunes (Apple Inc) and Google Play (Google LLC). Compared to a web-based intervention that can be easily and cheaply modified after creation, native apps such as MBapp incur significant software developer costs for any modifications postcoding. Though we cannot alter MBapp postproduction as has been carried out in web-based perinatal mental health interventions [42], we are optimistic that using the FRAME and end user feedback has optimized MBapp's wireframe to the point that additional edits are unnecessary. Lastly, though MBapp was deemed to have preserved fidelity to MB's curriculum by MB's original creators, it remains an empirical question whether an asynchronous digital education delivered via smartphone app is comparable to a 12-hour in-person group education led by a trained facilitator. Thus, the effect of MBapp on PPD will need to be explicitly and intentionally tested in future randomized trials.

Conclusions

In conclusion, the majority of our participants—pregnant or postpartum individuals with public health insurance—used apps to obtain on-demand perinatal education, and this expertise was leveraged to update and enhance a novel smartphone adaptation of MB. MBapp's wireframe was tailored to the unique needs of app-based interventions and to feedback from target end users, which may result in sustained user engagement once the app is deployed. Our team's next steps will be to confirm that MBapp is a feasible, acceptable intervention among English- or Spanish-speaking pregnant and postpartum people at risk of PPD through a pilot randomized trial. Ultimately, we aim to demonstrate that MBapp is an effective, scalable intervention that can prevent PPD on a population level.

Acknowledgments

The authors would like to acknowledge the Qualitative Science and Methods Training Program (QSMTP), Department of Psychiatry and Human Behavior, Alpert Medical School of Brown University, through which multiple authors (AKL, MG, and MLR) received training in qualitative research. Pertinent to this publication, AKL was trained via the QSMTP in the conceptualization and design of this study and the preparation, implementation, and analyses of the data presented here. AKL is supported by the Eunice Kennedy Shriver National Institute for Child Health and Human Development (NICHD; K23HD103961). CLB is supported by the NICHD (R01HD81868; R21HD110912), the National Institute on Drug Abuse (NIDA; R34DA055317), and the American Psychological Association (6NU87PS004366-03-02). MLR is supported by the NICHD (R01HD104187 and R01HD093655), Centers for Disease Control and Prevention (R01CE003267), and NIDA (501DA054698). ESM is supported by the NICHD (R01HD105499), National Institute of Nursing Research (NINR; R01NR021126), and National Institute of Mental Health (NIMH; R34 MH130969). These funding sources had no role in study design, collection, analysis, or interpretation of the data, or in the decision to submit this paper for publication. Portions of this paper were presented as an oral presentation at the Biennial Conference of the International Marcé Society in Barcelona, Spain from September 4-7, 2024.

Conflicts of Interest

None declared.

References

1. Ko JY, Rockhill KM, Tong VT, Morrow B, Farr SL. Trends in postpartum depressive symptoms — 27 states, 2004, 2008, and 2012. MMWR 2017 Feb 17;66(6):153-158. [doi: 10.15585/mmwr.mm6606a1] [Medline: 28207685]
2. The American College of Obstetricians and Gynecologists. ACOG committee opinion no. 757: screening for perinatal depression. Obstet Gynecol 2018;132(5):e208-e212. [doi: 10.1097/AOG.0000000000002927]
3. Aris-Meijer J, Bockting C, Stolk R, et al. What if pregnancy is not seventh heaven? the influence of specific life events during pregnancy and delivery on the transition of antenatal into postpartum anxiety and depression. Int J Environ Res Public Health 2019 Aug 9;16(16):2851. [doi: 10.3390/ijerph16162851] [Medline: 31405014]
4. Robertson E, Grace S, Wallington T, Stewart DE. Antenatal risk factors for postpartum depression: a synthesis of recent literature. Gen Hosp Psychiatry 2004;26(4):289-295. [doi: 10.1016/j.genhosppsych.2004.02.006] [Medline: 15234824]
5. Bauman BL, Ko JY, Cox S, et al. Vital signs: postpartum depressive symptoms and provider discussions about perinatal depression - United States, 2018. MMWR Morb Mortal Wkly Rep 2020 May 15;69(19):575-581. [doi: 10.15585/mmwr.mm6919a2] [Medline: 32407302]
6. O'hara MW, Swain AM. Rates and risk of postpartum depression—a meta-analysis. Int Rev Psychiatry 1996 Jan;8(1):37-54. [doi: 10.3109/09540269609037816]
7. Cox EQ, Sowa NA, Meltzer-Brody SE, Gaynes BN. The perinatal depression treatment cascade: baby steps toward improving outcomes. J Clin Psychiatry 2016 Sep;77(9):1189-1200. [doi: 10.4088/JCP.15r10174] [Medline: 27780317]
8. Thiel de Bocanegra H, Braughton M, Bradsberry M, Howell M, Logan J, Schwarz EB. Racial and ethnic disparities in postpartum care and contraception in California's Medicaid program. Am J Obstet Gynecol 2017 Jul;217(1):47. [doi: 10.1016/j.ajog.2017.02.040] [Medline: 28263752]
9. Bennett WL, Chang HY, Levine DM, et al. Utilization of primary and obstetric care after medically complicated pregnancies: an analysis of medical claims data. J Gen Intern Med 2014 Apr;29(4):636-645. [doi: 10.1007/s11606-013-2744-2] [Medline: 24474651]
10. Baldwin MK, Hart KD, Rodriguez MI. Predictors for follow-up among postpartum patients enrolled in a clinical trial. Contraception 2018 Sep;98(3):228-231. [doi: 10.1016/j.contraception.2018.04.016] [Medline: 29750924]
11. Dennis CL, Chung-Lee L. Postpartum depression help-seeking barriers and maternal treatment preferences: a qualitative systematic review. Birth 2006 Dec;33(4):323-331. [doi: 10.1111/j.1523-536X.2006.00130.x] [Medline: 17150072]
12. Hadfield H, Wittkowski A. Women's experiences of seeking and receiving psychological and psychosocial interventions for postpartum depression: a systematic review and thematic synthesis of the qualitative literature. J Midwif Womens Health 2017 Nov;62(6):723-736. [doi: 10.1111/jmwh.12669] [Medline: 29210501]
13. Bodnar-Deren S, Benn EKT, Balbierz A, Howell EA. Stigma and postpartum depression treatment acceptability among black and white women in the first six-months postpartum. Matern Child Health J 2017 Jul;21(7):1457-1468. [doi: 10.1007/s10995-017-2263-6] [Medline: 28102504]
14. Callister LC, Beckstrand RL, Corbett C. Postpartum depression and help-seeking behaviors in immigrant Hispanic women. J Obstet Gynecol Neonatal Nurs 2011;40(4):440-449. [doi: 10.1111/j.1552-6909.2011.01254.x] [Medline: 21639863]
15. US Preventive Services Task Force, Curry SJ, Krist AH, et al. Interventions to prevent perinatal depression: US Preventive Services Task Force recommendation statement. JAMA 2019;321:580-587. [doi: 10.1001/jama.2019.0007]
16. O'Connor E, Senger CA, Henninger ML, Coppola E, Gaynes BN. Interventions to prevent perinatal depression: evidence report and systematic review for the US Preventive Services Task Force. JAMA 2019 Feb 12;321(6):588-601. [doi: 10.1001/jama.2018.20865] [Medline: 30747970]
17. Muñoz RF, Le HN, Ippen CG, et al. Prevention of postpartum depression in low-income women: development of the mamás y bebés/mothers and babies course. Cogn Behav Pract 2007 Feb;14(1):70-83. [doi: 10.1016/j.cbpra.2006.04.021]
18. Barrera AZ, Wickham RE, Muñoz RF. Online prevention of postpartum depression for Spanish- and English-speaking pregnant women: a pilot randomized controlled trial. Internet Interv 2015 Sep 1;2(3):257-265. [doi: 10.1016/j.invent.2015.06.002] [Medline: 26273567]
19. Tandon SD, Perry DF, Mendelson T, Kemp K, Leis JA. Preventing perinatal depression in low-income home visiting clients: a randomized controlled trial. J Consult Clin Psychol 2011 Oct;79(5):707-712. [doi: 10.1037/a0024895] [Medline: 21806298]
20. Ronen K, Gewali A, Dachelet K, et al. Acceptability and utility of a digital group intervention to prevent perinatal depression in youths via Interactive Maternal Group for Information and Emotional Support (IMAGINE): pilot cohort study. JMIR Form Res 2024 Feb 2;8:e51066. [doi: 10.2196/51066] [Medline: 38306159]
21. Werner EF, Has P, Kanno L, Sullivan A, Clark MA. Barriers to postpartum glucose testing in women with gestational diabetes mellitus. Am J Perinatol 2019 Jan;36(2):212-218. [doi: 10.1055/s-0038-1667290] [Medline: 30060291]
22. Lewkowitz AK, Whelan AR, Ayala NK, et al. The effect of digital health interventions on postpartum depression or anxiety: a systematic review and meta-analysis of randomized controlled trials. Am J Obstet Gynecol 2024 Jan;230(1):12-43. [doi: 10.1016/j.ajog.2023.06.028] [Medline: 37330123]
23. Bae H, Shin H, Ji HG, Kwon JS, Kim H, Hur JW. App-based interventions for moderate to severe depression: a systematic review and meta-analysis. JAMA Netw Open 2023 Nov 1;6(11):e2344120. [doi: 10.1001/jamanetworkopen.2023.44120] [Medline: 37983028]

24. Frid G, Bogaert K, Chen KT. Mobile health apps for pregnant women: systematic search, evaluation, and analysis of features. J Med Internet Res 2021 Oct 18;23(10):e25667. [doi: 10.2196/25667] [Medline: 34524100]
25. Osma J, Barrera AZ, Ramphos E. Are pregnant and postpartum women interested in health-related apps? Implications for the prevention of perinatal depression. Cyberpsychol Behav Soc Netw 2016 Jun;19(6):412-415. [doi: 10.1089/cyber.2015.0549] [Medline: 27327069]
26. Zingg A, Singh T, Franklin A, et al. Digital health technologies for peripartum depression management among low-socioeconomic populations: perspectives from patients, providers, and social media channels. BMC Pregnancy Childbirth 2023 Jun 3;23(1):411. [doi: 10.1186/s12884-023-05729-9] [Medline: 37270494]
27. Camacho E, Cohen A, Torous J. Assessment of mental health services available through smartphone apps. JAMA Netw Open 2022 Dec 1;5(12):e2248784. [doi: 10.1001/jamanetworkopen.2022.48784] [Medline: 36576737]
28. Tsai Z, Kiss A, Nadeem S, et al. Evaluating the effectiveness and quality of mobile applications for perinatal depression and anxiety: a systematic review and meta-analysis. J Affect Disord 2022 Jan 1;296:443-453. [doi: 10.1016/j.jad.2021.09.106] [Medline: 34624434]
29. Lau N, O'Daffer A, Colt S, et al. Android and iPhone mobile apps for psychosocial wellness and stress management: systematic search in app stores and literature review. JMIR Mhealth Uhealth 2020 May 22;8(5):e17798. [doi: 10.2196/17798] [Medline: 32357125]
30. Tools to support pregnant and parenting people in managing their stress. Mothers & Babies. 2021. URL: https://www.mothersandbabiesprogram.org [accessed 2024-11-07]
31. Ramphos ES, Kelman AR, Stanley ML, Barrera AZ. Responding to women's needs and preferences in an online program to prevent postpartum depression. Internet Interv 2019 Dec;18:100275. [doi: 10.1016/j.invent.2019.100275]
32. Barrera AZ, Hamil J, Tandon D. Integrating SMS text messages into a preventive intervention for postpartum depression delivered via in-home visitation programs: feasibility and acceptability study. JMIR Form Res 2021 Nov 18;5(11):e30995. [doi: 10.2196/30995] [Medline: 34792478]
33. van Genugten CR, Schuurmans J, Lamers F, et al. Experienced burden of and adherence to smartphone-based ecological momentary assessment in persons with affective disorders. J Clin Med 2020 Jan 23;9(2):322. [doi: 10.3390/jcm9020322] [Medline: 31979340]
34. Colombo D, Fernández-Álvarez J, Patané A, et al. Current state and future directions of technology-based ecological momentary assessment and intervention for major depressive disorder: a systematic review. J Clin Med 2019 Apr 5;8(4):465. [doi: 10.3390/jcm8040465] [Medline: 30959828]
35. Ebner-Priemer UW, Trull TJ. Ecological momentary assessment of mood disorders and mood dysregulation. Psychol Assess 2009 Dec;21(4):463-475. [doi: 10.1037/a0017075] [Medline: 19947781]
36. RIte care. The Executive Office of Health and Human Services State of Rhode Island. 2024. URL: https://eohhs.ri.gov/Consumer/FamilieswithChildren/RIteCare.aspx#:~:text=for%20RIte%20Care%20is%20based,253%20percent%20of%20the%20FPL [accessed 2024-11-07]
37. Wiltsey Stirman S, Baumann AA, Miller CJ. The FRAME: an expanded framework for reporting adaptations and modifications to evidence-based interventions. Implement Sci 2019 Jun 6;14(1):58. [doi: 10.1186/s13012-019-0898-y] [Medline: 31171014]
38. Le HN, Perry DF, Mendelson T, Tandon SD, Muñoz RF. Preventing perinatal depression in high risk women: moving the mothers and babies course from clinical trials to community implementation. Matern Child Health J 2015 Oct;19(10):2102-2110. [doi: 10.1007/s10995-015-1729-7] [Medline: 25673369]
39. Ranney ML, Meisel ZF, Choo EK, Garro AC, Sasson C, Morrow Guthrie K. Interview-based qualitative research in emergency care part II: data collection, analysis and results reporting. Acad Emerg Med 2015 Sep;22(9):1103-1112. [doi: 10.1111/acem.12735] [Medline: 26284572]
40. Choo EK, Garro AC, Ranney ML, Meisel ZF, Morrow Guthrie K. Qualitative research in emergency care part I: research principles and common applications. Acad Emerg Med 2015 Sep;22(9):1096-1102. [doi: 10.1111/acem.12736] [Medline: 26284696]
41. Guthrie KM, Rosen RK, Vargas SE, et al. User input in iterative design for prevention product development: leveraging interdisciplinary methods to optimize effectiveness. Drug Deliv Transl Res 2017 Oct;7(5):761-770. [doi: 10.1007/s13346-017-0397-0] [Medline: 28653286]
42. Milgrom J, Danaher BG, Gemmill AW, et al. Internet cognitive behavioral therapy for women with postnatal depression: a randomized controlled trial of MumMoodBooster. J Med Internet Res 2016 Mar 7;18(3):e54. [doi: 10.2196/jmir.4993] [Medline: 26952645]

Abbreviations

CBT: cognitive behavioral therapy
FRAME: Framework for Modification and Adaptation
MB: Mothers and Babies
PPD: postpartum depression

Edited by A Kushniruk, J Waterson; submitted 11.06.24; peer-reviewed by A Finlay, J Griffiths; revised version received 25.09.24; accepted 14.10.24; published 09.12.24.

Please cite as:
Lewkowitz AK, Guillen M, Ursino K, Baker R, Lum L, Battle CL, Ware C, Ayala NK, Clark M, Ranney ML, Miller ES, Guthrie KM
Optimizing a Novel Smartphone App to Prevent Postpartum Depression Adapted From an Evidence-Based Cognitive Behavioral Therapy Program: Qualitative Study
JMIR Hum Factors 2024;11:e63143
URL: https://humanfactors.jmir.org/2024/1/e63143
doi:10.2196/63143

Work Systems Analysis of Emergency Nurse Patient Flow Management Using the Systems Engineering Initiative for Patient Safety Model: Applying Findings From a Grounded Theory Study

Ellen Benjamin[1,2], RN, PhD; Karen K Giuliano[3], RN, MBA, PhD

[1]Elaine Marieb College of Nursing, University of Massachusetts Amherst, Amherst, MA, United States

[2]Donna M and Robert J Manning College of Nursing and Health Sciences, University of Massachusetts Boston, 100 Morrissey Blvd, Boston, MA, United States

[3]Elaine Marieb Center for Nursing and Engineering Innovation, Institute for Applied Life Sciences, University of Massachusetts Amherst, Amherst, MA, United States

Corresponding Author:
Ellen Benjamin, RN, PhD
Elaine Marieb College of Nursing, University of Massachusetts Amherst, , Amherst, MA, , United States

Abstract

Background: Emergency nurses actively manage the flow of patients through emergency departments. Patient flow management is complex, cognitively demanding work that shapes the timeliness, efficiency, and safety of patient care. Research exploring nursing patient flow management is limited. A comprehensive analysis of emergency nursing work systems is needed to improve patient flow work processes.

Objective: The aim of this paper is to describe the work system factors that impact emergency nurse patient flow management using the System Engineering Initiative for Patient Safety model.

Methods: This study used grounded theory methodologies. Data were collected through multiple rounds of focus groups and interviews with 27 emergency nurse participants and 64 hours of participant observation across 4 emergency departments between August 2022 and February 2023. Data were analyzed using coding, constant comparative analysis, and memo-writing. Emergent themes were organized according to the first component of the System Engineering Initiative for Patient Safety model, the work system.

Results: Patient flow management is impacted by diverse factors, including personal nursing characteristics; tools and technology; external factors; and the emergency department's physical and socio-organizational environment. Participants raised concerns about the available technology's functionality, usability, and accessibility; departmental capacity and layout; resource levels across the health care system; and interdepartmental teamwork. Other noteworthy findings include obscurity and variability across departments' staff roles titles, functions, and norms; the degree of provider involvement in patient flow management decisions; and management's enforcement of timing metrics.

Conclusions: There are significant barriers to the work of emergency patient flow management. More research is needed to measure the impact of these human factors on patient flow outcomes. Collaboration between health care administrators, human factors engineers, and nurses is needed to improve emergency nurse work systems.

(JMIR Hum Factors 2024;11:e60176) doi:10.2196/60176

KEYWORDS

patient flow; throughput; emergency department; nursing; emergency nursing; organizing work; cognitive work; human factors; ergonomics; SEIPS model

Introduction

Background

Emergency department crowding poses a grave threat to global health [1]. Hospitals face high patient volumes and acuity, limited bed capacity, staffing shortages, and financial constraints [1,2]. These issues necessitate an urgent optimization of patient flow to provide timely, efficient, high-quality care [2,3]. Emergency nurses play a central role in effective patient flow management, but their work is poorly understood [4-6]. This paper describes the human factors that impact the work of emergency nurse patient flow management.

Patient Flow Management

Patient flow analysis is the study of progressive patient movement through a unit, hospital, and wider health care system [6,7]. Research analyzing patient flow has grown over the last

2 decades to address patient crowding and limited health care resources [1,2]. The study of patient flow is especially important within emergency departments, where overcrowding and access block pose great dangers to patient safety [1]. Significant contributions from patient flow research include the identification of strategies to reduce demand for emergency services, improve hospitalwide management of system capacity, and expedite emergency throughput through process improvements such as patient streaming, point of care laboratory tests, fast-track treatment zones, and short stay observation units [1,2,6].

Throughout this research, emergency patient flow has been widely conceptualized using linear models that illustrate patient transitions into, through, and out of the emergency department [1,8]. However, studies that represent patient flow as a simplified sequential process may fail to account for the impact of the human agents who actively manage patient flow [9-12]. Research that does not adequately consider human factors has been criticized for poorly capturing the complexity of real-world systems, lacking generalizability across emergency departments, and failing to provide a deeper understanding of how and why patient flow interventions succeed or fail [4,6,13,14]. Emergency nurses are autonomous decision makers who exert active agency over care processes and have a demonstrated impact on patient flow outcomes [5,10,15]. Nevertheless, few scholars have investigated this aspect of nursing work [5,15].

To respond to this gap in knowledge, this paper is the second in a series that describes emergency nurse patient flow management. For the purposes of this research, "emergency nurses" describes registered nurses employed in hospital-based or freestanding emergency departments. The first paper described how emergency nurses conceptualize patient flow management as the balancing practice of optimizing patient care without exhausting department resources [16]. This balance is guided by the primary goal of patient safety [12,16]. Patient flow management encompasses five critical tasks: (1) information gathering, (2) continuous triage, (3) resource management, (4) throughput management, and (5) care oversight [16]. By engaging in these tasks, emergency nurses monitor fluctuating resources and patient care needs and aim to ethically allocate limited resources to the right patient. Emergency nurses also manage a tension between the desire to expedite patient throughput and the desire to provide care that comprehensively meets patient needs [16].

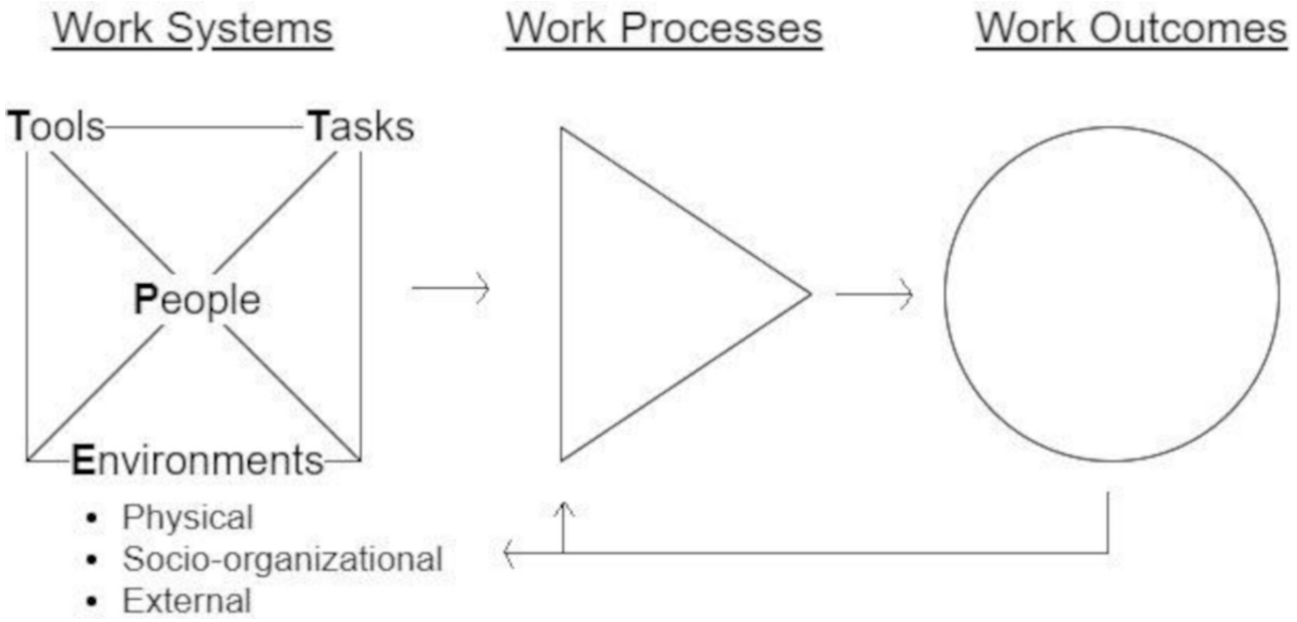
More specifically, nurses manage the flow of patients by making real-time, frontline decisions to shape patient care [12,16]. Nurses are often responsible for triaging patients and determining the order of bedding assignments, placing patients in appropriate treatment spaces, choosing which patients receive limited department resources such as cardiac monitors and specialized treatment rooms, and managing the size and acuity of nursing assignments [16-18]. Nurses also influence the timeliness and order of patients to receive treatment, diagnostic testing, inpatient bed assignment, and discharge or transfer [16-18]. Patient flow management is understood to be complex, dynamic, and cognitively demanding [6,12,15]. This paper adds to previous descriptions of emergency patient flow management by analyzing the factors that shape nursing work systems.

Human Factors and the Systems Engineering Initiative for Patient Safety Model

A purposeful investigation of interactions between humans, processes, and their environment is needed to effectively design systems, promote worker well-being, and understand and improve worker performance [19]. However, human factors methods have been underutilized in emergency medicine, and the organizational, cultural, and interpersonal factors that impact emergency patient care have yet to be fully explored [20-22]. Barriers to applying human factors approaches within emergency medicine include a lack of resources within hospitals to conduct human factors research, policies that impede human factors experts from accessing clinical settings, and limited dissemination of published examples [20].

The Systems Engineering Initiative for Patient Safety (SEIPS) model provides a human factors–based framework to investigate health care work [23]. The SEIPS model was first proposed in 2006 and has been refined several times, including the recently simplified SEIPS 101 model (Figure 1) [23-26]. SEIPS builds upon Donabedian's classic System-Process-Outcomes model [27] to further explicate the elements of work systems, work processes, and work outcomes [23]. Within emergency health care research, this model has been successfully used to investigate physician documentation practices, physician disposition decision-making, and the transitions of older adult patients back home [28-30]. Within nursing, the SEIPS model has been used to better understand the work systems factors that impact nursing medication errors in critical care units, cardiac nursing workflow, and endotracheal cuff pressure management [31-33].

Figure 1. SEIPS 101 model by Holden and Carayon [26], published under Creative Commons Attribution-Noncommercial 4.0 International License [34]. SEIPS: Systems Engineering Initiative for Patient Safety.



This manuscript focuses on the work system, which is the first element of the SEIPS model [26] (Figure 1). The work system comprises 6 interconnected components: people, tasks, tools, physical environment, socio-organizational environment, and the external environment. These elements interact to form the work processes that directly impact outcomes such as health care quality, safety, and worker well-being [23].

Methods

Study Design

This study employed constructivist and situational analysis grounded theory methodologies, as articulated by Kathy Charmaz and Adele Clarke. Each methodology involves inductive, qualitative analysis [35,36]. Constructivist grounded theory focuses on exploring social processes [36], while situational analysis supports analysis of the disparate, nonhuman elements that shape situations [35]. To minimize bias and preconceptions, grounded theorists may postpone formal literature reviews until after data collection and analysis [36]. Therefore, this study started with the broad question, “How do emergency nurses perform patient flow management?”

Data Collection

Data collection strategies included individual interviews, focus groups, and observations across 4 emergency departments. Interview and focus group participants were recruited using purposeful and snowball sampling through email and social media platforms. Participation inclusion criteria were the following: English-speaking, over the age of 18 years, and registered nurses with at least 90 days of experience working in an emergency department. Initial interviews and focus groups lasted approximately 60 minutes and were guided by broad interview guides. Meetings were held remotely over Zoom and were audio/video recorded and transcribed with manual verification.

After initial rounds of focus groups and interviews, theoretical sampling was used to recruit existing participants for additional rounds of follow-up and think-aloud scenario interviews. Think-aloud scenario interviews [37] prompted participants to verbalize their patient flow management considerations when presented with a mock, simulated emergency department tracking board. These subsequent interviews were conducted individually, lasted between 30 - 60 minutes, and were used to support and further develop coding categories. In all, a total of 5 focus groups and 24 interviews were conducted across 27 participants.

Participant observations were concurrently conducted at 4 emergency departments in the northeastern United States. Departments had varying community settings, sizes, trauma designations, and annual patient visit volumes. The researcher recorded handwritten narrative field notes to describe the departments and staff behavior. Field notes also described brief interactions with staff that were used to clarify their actions and decision-making strategies. Observations were conducted in 4-hour blocks at variable times throughout a 24-hour period, for a total of 64 hours of observation.

Data Analysis

Data analysis occurred simultaneously with data collection and continued until saturation was reached. Data analysis relied on coding in NVivo 12 Plus (Lumivero) and constant comparative analysis. Line-by-line in vivo and gerund coding were used for initial coding [36]. As data analysis progressed, incident-by-incident coding was also used to code larger segments of data [36]. Constant comparative analysis is a method of qualitative analysis where coding of incidents and categories are compared to previous analysis to clarify emerging themes, inform additional data collection, and generate a theory [38]. Memo-writing was used to investigate coding categories, clarify theory development, and prompt reflexivity. Data collection and analysis were performed by a researcher experienced in emergency nursing; thus, reflexivity was an active process of investigating assumptions and personal experiences, as well as reflecting upon potential biases [36].

After emergent themes were developed, a review of the literature was performed, and study findings were found to closely align with the SEIPS model (Table 1). Emergent themes were then recategorized according to the SEIPS framework. As Carayon et al state, the SEIPS model can be applied during data analysis even if it was not used to guide data collection [24].

Table . Emergent theme and corresponding SEIPS component.[a]

Grounded theory emergent theme	Corresponding SEIPS component
Tasks of patient flow management	Tasks
Factors that impact patient flow management	
Individual nursing factors	Person factors
Departmental factors (structural: resources, technology, physical layout; interpersonal: communication, staff roles and norms, department culture)	Technology and tools, and physical and socio-organizational environment factors
Interdepartmental factors	External environment factors

[a]SEIPS: Systems Engineering Initiative for Patient Safety.

Trustworthiness

Strategies to increase study trustworthiness included prolonged engagement with data, negative case analysis, triangulation of participant samples and data collection strategies, and use of an audit trail. Three formal member checking interviews were performed and affirmed study findings. Study design, data analysis, and theory generation were informed by consultation with methodology and subject matter experts.

Ethical Considerations

This study was approved by the institutional review boards of the participating health care system and the University of Massachusetts Amherst (number 017066 - 00002). A Certificate of Confidentiality from the National Institutes of Health was obtained to increase participant confidentiality. Informed consent was obtained from focus group and interview participants using Qualtrics surveys. Observed participants were informed of the study using an information sheet and interactions were guided by an oral script with verbal confirmation of participation. Focus group and interview participants were compensated at a rate of US $35/hour using Amazon gift cards.

Results

Overview

Study findings are categorized according to elements of the SEIPS work system: person, technology and tools, socio-organizational environment, physical environment, and external environment (Figure 2). Participant demographics, emergency department setting characteristics, and patient flow management tasks are described in detail elsewhere [16]. Table 2 summarizes supporting evidence.

Figure 2. A SEIPS model of emergency nurse patient flow management. SEIPS: Systems Engineering Initiative for Patient Safety.

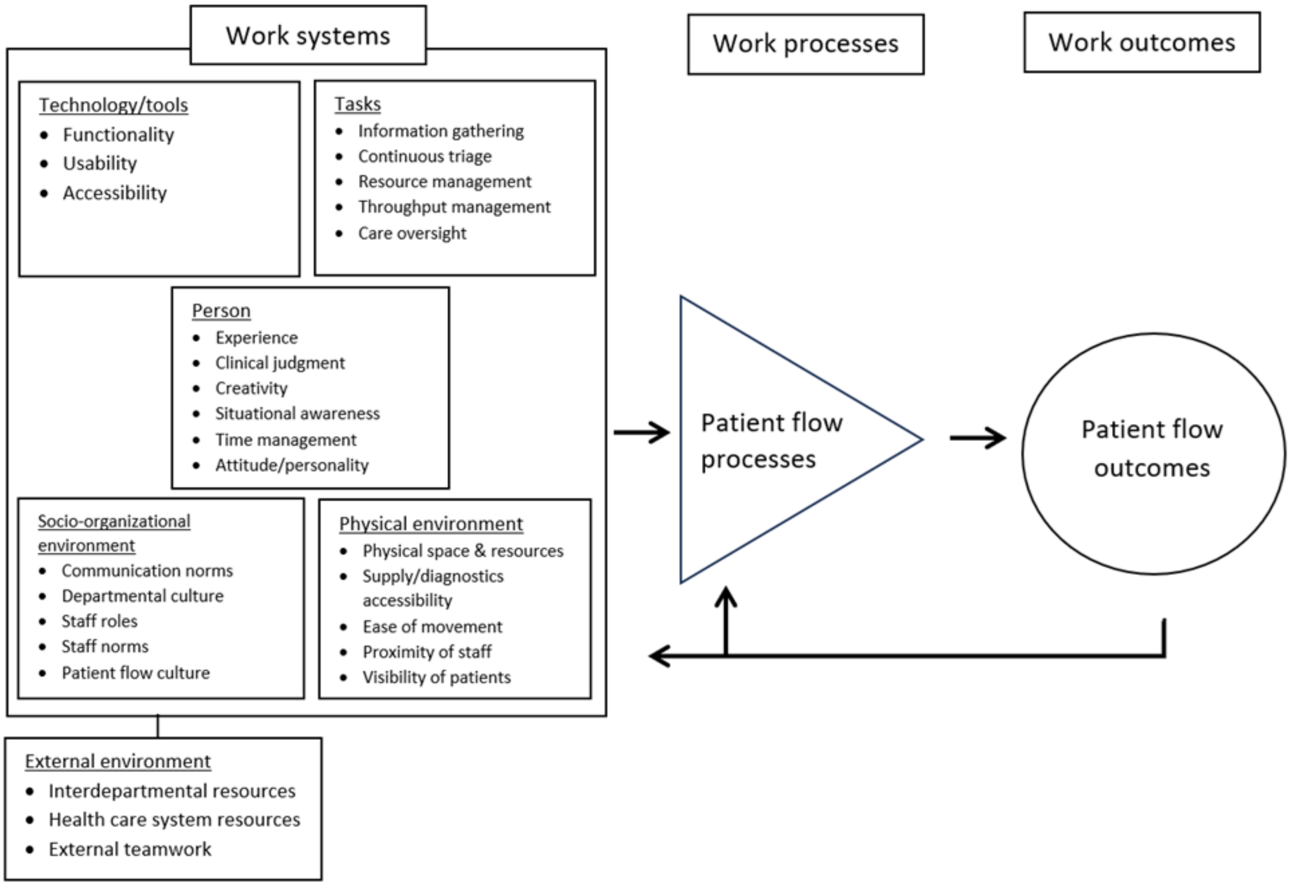


Table . Summary of findings with supporting evidence.

Work system element, major theme, and minor theme			Supporting quotes
Person			
	Experience		
		Emergency department	It's taking him a long time to adjust to this environment...in the ICU, "patients are there for so long, their plan of care is established." (RN2) – Field Note 14
		Overcrowding experience	Community hospitals have a harder time dealing with stress influxes on patient resource management, while larger hospitals are used to being consistently stressed… and handle that patient flow management under stress a little bit better. (Int5)
		Experience in specialized flow roles	When you're able to take those experiences, it opens your eyes up to the bigger picture…of patient flow. (Int11)
	Clinical judgment		It has to be somebody with...that has a good core knowledge as a base. (Int6)
	Creativity		Part of the charge nurse's job is to be really creative… you have to be innovative. (FG5)
	Time management		
		Correct prioritization of tasks	So I guess prioritizing…having a great awareness of prioritizing what is important at the time. (Int7)
		Ability to multitask	I think that's where a lot of nurses really get caught up, is they don't have the ability to effectively multitask and remember the 7 different directions that they need to go at once. (Int10)
	Situational awareness		Our charge nurses, I think of them as up in the balcony, watching the orchestra play. They're imperative to the functioning of the team, but they've gotta be up a level so that they can actually see what's going on, and so they've got that complete view that isn't skewed because they're, you know, sinking in the quicksand with everybody else. (Int1)
	Attitude		
		Sense of accountability to department	The nurses are wrapped up in their own little section, and they don't want more because they're overwhelmed. (FG3)
		Degree of motivation	And then you have bedside nurses who will sit on patients, sit on orders, to hold on to as many patients as they can, so that they don't have openings and flow. (FG5)
		Level of burnout	I've never seen nurses this exhausted, physically and emotionally. (Int6)

Work system element, major theme, and minor theme			Supporting quotes
	Personality		
		Compassion for patients and colleagues	I think they're good people… it's definitely a personality thing. They're friendlier, more socially aware of other people's struggles, and they care about it. I think other nurses don't give a shit. (Int9)
		Adaptability	There's some people that they're like 'Chicken Little the sky is falling' all the time, and it's like, 'Take a deep breath. It's gonna be okay.' (Int2)
Technology and tools			
	Functionality		We all have the computers on wheels and they're just all over the place…and you know, of course, half of them don't work. (Int8)
	Usability		I was trying to login so I could get report, and it pops up, 'Your patient is overdue for a bath.' It's like, 'Why, why?!'...and you can't get out of the screen unless you chart something, and if you chart "not responsible," it tries to prompt you as to why?... and then it's, you're clicking out of the boxes... it's alarm fatigue, but for the charting side. It's charting fatigue. (Int15)
	Information accessibility		So when an ambulance comes in, I don't like to be like, 'Oh, hold on let me change my screen.' I just wanna know, visually, like what I have available, and who has what, and that kind of thing, at all times. (Int7)
Physical environment			
	Physical capacity and resources		There's just no space, no room. Even if we try to bring them back, there's just nowhere to put them. (FG12)
	Access to supplies and diagnostics		Our isolation room has a little anteroom…we were kind of using it for storage, and we would just shove stuff in there, all kinds of stuff. And a lot of equipment has to be plugged in all the time, so their solution was to mount an extension cord strip type thing to the wall… so all that equipment is just right there, to get to sterile gloves, I have to move things out of the way. (Int8)
	Ease of movement		There wasn't even a hallway space to put the patient in. It was like, 'Okay, I have 14 inches here in a hallway.' Everybody just turned sideways to go around this trauma patient. (FG11)
	Proximity of staff		Just having everybody kind of close to the nurses' station just really helps communication. (Int7)

Work system element, major theme, and minor theme			Supporting quotes
	Visibility of patients		The worst setup that we did was fast-track. They were made all private rooms, and you couldn't see the patients 'cause they were behind a closed door. (Int6)
Socio-organizational environment			
	Communication norms		If you have [staff that] don't openly communicate, you are going to have significant delays in your patient flow management. (Int5)
	Staff roles		There was no flow coordinator. I had never heard of that until my friend was talking about their flow coordinator, I'm like, 'Who is that?' (Int15)
	Staff norms		
		Role flexibility	If I'm lead that day, like, I have my plan… I have my system. And if people do try to help …I think it adds to the chaos. (Int7)
		Role preference	That's a bad job by the charge person, because you've got all your buddies doing all the fun jobs for the day, and then everyone else has to do the other ones. And if you don't think they notice that, you're out of your mind. So that's a cultural problem that will slow things down. (Int9)
		Role hierarchy	Good patient flow has to be somebody who's willing to not sit in the White Tower… or 'I'm gonna sit in the pod chair and get 'charge butt' and never move.' You have to move. (Int3)
	Departmental culture		
		Teamwork and camaraderie	They would step over your quivering body on the floor to get to where they're going, instead of picking you up or helping you. (Int9)
		Respect between providers and nurses	If they walk up to a doctor, whether it's a third or fourth year resident or an attending, and they say 'You need to come right now and see this patient.' That doctor will stop what they're doing and they will go see their patient. (Int6)
		Relationship between staff and administration	It's so frustrating, 'cause this is the 6th hospital I've worked in, and they're all the same. They're greedy, greedy corporations and they don't care about patients. (RN6)
		Capacity for change	RN2 tells me some of the older nurses are "stuck in the old ways" when they only saw 5 patients in their whole shift and they've had trouble adjusting. — Field Note 11
	Patient flow culture		

Work system element, major theme, and minor theme			Supporting quotes
		Perception of provider incentives	I'll have a patient come in from the waiting room and they're already admitted. And it's for bullshit reasons…They need to look at what they're admitting and send more people home. (RN7)
		Provider-driven versus nurse-driven flow	So now, it's interesting for me to come back and see this clash between what the doctors are used to. They're used to nurses running things, and the nurses are used to doctors running things. (Int12)
		Role of patient flow metrics	Like door-to-doc time less than 20 minutes, discharge time less than 7 minutes...it was engraved in everybody's head. As soon as that manager left—we just got a new manager—and there's no real guidance for this new manager. (Int7)
External environment			
	Interdepartmental resources		That's our biggest pushback... we've had a CT scanner down and they're understaffed right now, and we had a 4-hour delay in CTs yesterday. (Int2)
	Health care system resources		There's this backlog… when patients haven't been efficiently discharged from upstairs, and there's a delay to the folks who have been admitted. And then, obviously, it's probably the same thing that's happening in the bigger hospitals, so there's no transfers available. (FG5)
	External teamwork		But the house supervisor will come and help—sometimes they'll send…a nurse over here, who comes in kicking and screaming. (Int8)

Person

Emergency patient flow management relies on nurses' individual characteristics, including their level of experience, cognitive skills, and personal attributes. In addition to overall years of nursing experience, participants emphasized a need for experience working specifically in emergency settings. Emergency departments have their own unique rhythm and pace of care. In observations, this was exemplified by a senior intensive care unit (ICU) nurse who struggled with the speed and turnover of emergency patients.

> *RN2 has significant ICU experience, including 8 years working in a MICU, SICU, NICU [medical, surgical, neonatal ICU]. "It's a totally different environment and flow than I'm used to."* [Field Note 14]

Emergency department experience is also needed to gain knowledge of emergency medical treatment and care processes. Strong patient assessment skills, critical thinking, and knowledge of emergency patient care are perceived as the "core knowledge" (Int6) of emergency patient flow management. Notably, patient flow management is viewed as work that nurses learn after they master these fundamentals of emergency nursing. As one participant stated, "a new nurse is just trying not to kill a patient. And not miss anything" (RN3).

Grounded in clinical judgment, patient flow management also relies on nurses' creativity, time management, and situational awareness. To successfully manage flow, nurses must think on their feet and creatively problem-solve in real time, while simultaneously managing multiple, urgent patient care priorities. This creativity is fostered by experience in overcrowded settings, where nurses learn to "think outside the box" (Int1) and develop a repertoire of innovative strategies.

Above all, participants emphasized the need for situational awareness. Situational awareness was described as the ability to maintain a holistic perspective of fluctuating patients, care needs, staff, and resources. Nurses with strong situational awareness are knowledgeable about the entire emergency department within its wider contexts of the hospital, emergency medical system, and surrounding health care system.

> *So you need to be able to look at the big picture, not just your one team of patients, or even your one ER.*

> *You have to be able to look and see globally what's going on.* [Int4]

The cognitive skills required to manage patient flow are strengthened through experience in specialized flow roles. Nurses in lead, float, flow coordinator, journey navigator, resource, or charge roles are immersed in the work of managing patient flow across an entire pod, zone, or department. As one participant described, these roles "open your eyes up to the bigger picture of patient flow" (Int11) and provide nurses with a deeper knowledge of the policies, organizational procedures, and internal politics of their institution.

Patient flow management is also impacted by nurses' attitudes and personalities. Nurses engage more in patient flow management when they have a sense of accountability to the wider department. Nurses who focus more narrowly on their own patients often lack an awareness of the department's patient volume, the number of waiting patients, or of coworkers struggling with heavy assignments.

> *They're just focused on their assignment…they work on their own pace and are in their own bubble. They're not hustling as much as the charge nurse would be, or somebody who's trying to make room for other folks.* [FG5]

Nurses' engagement in patient flow management also varies according to their level of motivation. Many nurses demonstrate a sense of urgency to expedite patient throughput. Other nurses are described as "unmotivated, complacent, or lazy." Burnout was seen as an especially common barrier to patient flow management engagement. Participants reported widespread physical, mental, and emotional exhaustion among emergency staff. For disengaged nurses, expediting patient flow is undesirable because it leads to new patients and additional work. As one participant summarized, "the more you kick in quicksand, the faster you sink" (FG3).

> *Well, the faster I get this out, the faster you're just gonna give me something new, and this never gonna end, so why should I hurry?* [FG4]

Regarding personality, patient flow management is impacted by nurses' level of compassion and adaptability. Nurses are perceived as more highly engaged when they hold deep compassion for their patients and colleagues that drives a desire to help and provide better care. Effective patient flow management also relies on the ability to remain calm and level-headed under pressure.

> *None of my nurses really spin, you know, they don't get caught up and start freaking out about things. My team is so tight…we just don't get that excited over stuff.* [Int10]

To summarize the personal factors, patient flow management appears to be facilitated by highly engaged nurses who are experienced in emergency department care, have deep clinical and institutional knowledge, and have the cognitive capacity to perform creative, complex, dynamic work.

Technology and Tools

Technology was found to significantly shape emergency nurse patient flow management. Access to Pyxis medication dispensers, portable computers, pneumatic tube systems, portable communication devices, and online applications that provide pharmacy or education support were observed to impact nurse workflows. Health information systems, including emergency tracking boards, electronic medical records, bed tracking systems, and ambulance tracking systems, appear to be especially consequential to patient flow management because they shape emergency nurses' ability to gather information about department resources and patient care.

When describing technology, participants reported concerns with functionality, usability, and information accessibility. Inconsistent access to functioning technology is a common and time-consuming barrier to patient flow management. Participants frequently dealt with failures and breakdowns such as broken medication scanners, computer downtimes, stalled pneumatic tube systems, and challenges with portable communication devices.

> *Nothing is working...yawn...the computers aren't working. (RN3)* [Field Note 11]

Participants also struggled with technology usability, often lacking a full understanding of how to operate and troubleshoot devices or make sense of presented information. Emergency nurses heavily rely on emergency tracking boards, which present a departmentwide overview of patient assignments and care progress. Multiple participants acknowledged uncertainty in the meaning of the icons, symbols, and colors displayed on the emergency tracking board.

> *… the stars indicate if an order is late by 15 minutes or 30 minutes. They both laugh, "See? I didn't even know that." (RN3)* [Field Note 8]

Although some participants reported satisfaction with their health information systems, many described information platforms that were overwhelming, contained burdensome alert functions, and had poorly designed interfaces.

> *I ask about their experience using [product]. "It's overwhelming," "Way too much on the eyes. I don't know how pod leads can look at this board." (RN10)* [Field Note 8]

Last, information accessibility appears to be impacted by the number of different health information systems required to access patient information and the transparency between different systems. Participants reported outdated programs, the need to navigate several different interfaces, and time-consuming workarounds.

> *To look up the actual lab numbers you have to go to [a] different screen, to look up radiology results you have to go to [a] different program, and so there's a lot of minimizing, you know, getting to this screen, or that screen, or this screen, or that screen, to figure out what needs to be done.* [Int8]

Socio-Organizational Environment

Patient flow management is impacted by the social and organizational characteristics that influence how nurses communicate and work together. First, participants emphasized the importance of clear, closed-loop, and frequent communication to keep one another apprised of department resources and patient care. Observations revealed that, in practice, nurses often rely on indirect strategies, such as writing updates on the emergency tracking board, to communicate changes in patient locations. Communication failures were especially common during handover with prehospital personnel, during bedding assignments, and when calling patient report to inpatient units.

> *"Do you have that guy?" (RN5) "Oh, so I probably have that guy. I don't know, I just came out and my name was on them. Alright, I guess I'll go see him." (RN4) The patient they're talking about has been in the department for 2 hours.* [Field Note 3]

Patient flow management is also shaped by the culture of emergency staff, providers, and administration. Departments with high levels of teamwork and camaraderie are able to work together toward common patient flow goals. Participants described respect between providers and nurses as the foundation for patient safety and optimized flow. Without a sense of mutual respect, nurses report a reluctance to ask questions, raise concerns, and share their opinions. Whereas nurses praised departmental cultures where staff and providers were a unified team, participants often expressed a disdain or distrust of hospital administrators. Nurses' suspicion of administrators' intentions can exacerbate a reluctance to embrace change or engage in patient flow improvement initiatives.

> *With this model and this process that they want to try—that was directed by our medical director, who is not a nurse, and he doesn't push for us to get these patients out like he should, or stand up to the people with the dollars and the pulling of strings.* [Int14]

Staff roles were found to vary widely between departments, encompassing a broad array of role titles and job functions. Responsibility for patient flow management decisions may be held by staff nurses, charge nurses, or by a variety of specialized flow roles such as flow coordinators, triage nurses, navigator nurses, pivot nurses, pod leads, float nurses, streamer nurses, or bed czars. Individual roles fluctuate according to staffing levels, such that one staff member may have to take on the duties of several roles during periods of poor staffing. Patient flow management is also shaped by the presence of other staff and provider roles, such as technicians, medical assistants, paramedics, orderlies, transporters, and medical residents.

> *And then the roles within the department, your typical staff nurse—clinical nurse, but we also have a resource nurse, which is just a charge nurse. We also have a flow nurse and then triage nurses as well.* [FG17]

In addition to differing staff role titles and functions, departments vary in their role norms. Some departments demonstrate high levels of flexibility in responsibility for patient flow management decisions, while others have more rigid role expectations. For example, in some departments, the responsibilities of answering the ambulance radio, assigning patients to emergency rooms, or shuffling patient locations are strictly held by a charge or flow coordinator nurse. In other departments, nurses in bedside roles readily take on these tasks when they see an opportunity to help. Departmental norms further impact the desirability of certain roles and role hierarchy. High degrees of hierarchy that create distance between the charge nurse or flow coordinators and the patients' bedside may impede that nurse's familiarity with direct patient care and create animosity between staff.

> *I feel like when some people level-up they forget how it is to be in an assignment. And then they're in an assignment and they're drowning and they're asking you for your help.* [RN4]

Role norms also impact whether patient flow management is perceived as primarily provider- or nurse-driven. Departments perceived to be more "doctor-driven" were described as those where the providers more closely monitor the waiting room patients and voice their opinions about who should be assigned emergency beds, and where providers delegate frequently and engage less in collaborative decision-making.

> *RN2 said that post-COVID, with all the newer nurses, that patient care is more "micromanaged" by the doctors. "It never used to be like that." (RN2)* [Field Note 8]

Finally, patient flow management is influenced by the incentives provided to physicians and nurses. Participants criticized the perception that providers may be incentivized to empty the waiting room, perform excessive diagnostic imaging, and admit high proportions of emergency patients as especially impactful on patient flow management.

> *Our doctors… hate waiting room times. So at (hospital), what I appreciated was it was okay to have a waiting room. That's what the waiting room is for.* [FG17]

Emergency departments vary in the extent to which management has established clear incentives to meet patient flow metric goals. These metric goals might include specifications for the length of time that patients should wait to be assessed or transported to an assigned bed. Nurses differ in their opinions of these patient flow standards. Although some participants felt that strict timing expectations were helpful to increase staff engagement, others felt that they compromised nursing judgment and patient care.

> *It made it harder to be really thorough. Like you'd think 'Hmm, I can either do a really thorough assessment, or I can meet my time.'* [RN2]

Physical Environment

Participants perceive access to resources to be the single most important factor influencing patient flow management. Patient flow management is impeded by insufficient beds, rooms, equipment and supplies, and inadequate staffing.

> *The halls are narrow, especially when we have the hallway beds in there. So, you know, when you're pushing a gurney through there, you're kind of having to wiggle a little bit, make sure you don't hit somebody.* [Int8]

In addition to its physical capacity, the layout of a department impacts nurses' access to supplies and diagnostic testing, physical movement, the proximity of staff, and visibility of patients. These characteristics impact the efficiency and ease of emergency nurses' work. In observations, departments were found to be cluttered, cramped, disorganized, and lacking sufficient capacity for patients, medical equipment, and supplies.

> *So layout is important…because there are rooms where you're kind of isolated, and it's hard because they don't hear the doctors talking, or updates on patients, so that has an impact on patient flow.* [Int7]

External Environment

Last, emergency patient flow management is shaped by factors external to the work system of the department. Interdepartmental resources—including staffing levels of diagnostic departments, environmental services, transport services, and inpatient units—impact the efficiency of patient care and movement out of the department. Participants also acutely felt the consequences of working within an overburdened health care system that limits their ability to successfully transfer and discharge patients. Nurses emphasized a frustration with the lack of adequate ambulance transportation.

> *Our EMS service's so short-staffed, they can hardly handle the 9-1-1 calls, let alone handling the transports out. Like, how do we get these people out?* [FG4]

Teamwork between the emergency department and other hospital departments impacts the ability to coordinate and advance patient care processes. Tension and pushback from inpatient floors during reports from the emergency department to floors is a common experience. Participants reported that inpatient nurses are often unavailable to take report, reluctant to accept patients, and engage in delaying tactics. This pushback from the floor results in challenges transporting emergency patients out of the department.

> *It's been like hand-to-hand combat with the floors, trying to get patients upstairs.* [Int10]

Discussion

Principal Findings

This is the first paper, to our knowledge, to comprehensively describe the work system of emergency nurse patient flow management. Using this approach, we have framed patient flow processes around the work processes of nurses rather than sequential patient transitions. The articulation of 5 discrete patient flow management tasks is described more thoroughly in the first paper in this series [16]. Conceptualizing patient flow management as a balance of tasks offers a new framework for improving patient flow by supporting the nursing work of information gathering, continuous triage, resource management, throughput management, and care oversight.

The SEIPS model places a "person" at the heart of the work system. This human-centric approach theoretically links nurses' individual characteristics to patient flow outcomes. Study results propose that patient flow outcomes are influenced by emergency nurses' attitude, personality, experience level, and cognitive abilities. Collectively, these findings emphasize the importance of strengthening nurse training, retention, and support. Developing nurse patient flow management training has been proposed as a potentially cost-effective approach to improve patient flow [15]. The current lack of training may be due, in part, to a limited understanding of needed skills and traits [39]. Nursing education also predominately focuses on individual, patient-centered care rather than collective decision-making across multiple patients [16,40]. This study has clarified that clinical judgment, time management, and situational awareness are essential for patient flow management and should be a priority for training efforts. In addition to training support, findings suggest that hospitals may consider investing in nurse retention and well-being as a patient flow intervention.

Patient flow management depends on several structural factors, including department technology, physical capacity, layout, and internal and external resources. These elements impact the speed and ease of nursing work, including the ability to gather information, access needed supplies, collaborate with peers, and visualize patients. Broadly, participants described emergency departments as ill-suited to support the work of patient flow management. Nurses criticized the accessibility, functionality, and usability of health information systems, and physical work environments were found to be cramped, cluttered, and lacking adequate supply and staffing resources. Although the physical capabilities and resource levels of emergency departments are often challenging to change, there is potential to better support patient flow managers by redesigning health information systems [41]. Notably, a recent systematic review found that, despite their importance to patient flow, how and why health information systems impact patient flow processes remain poorly understood [42], re-emphasizing the need to better understand emergency nursing work processes.

Finally, patient flow management is shaped by social and organizational factors. Although the importance of departmental communication and culture to promote patient flow are apparent, the variability of staff roles, staff norms, and patient flow incentives are noteworthy findings. Patient flow management roles, norms, and incentives were found to be highly inconsistent, context dependent, and largely informal. Staff roles and norms lack uniformity between departments and fluctuate as staffing levels, individual staff members' preferences, and levels of expertise change. Departments also vary in the degree to which patient flow management is perceived to be doctor- or nurse-driven, and these expectations are unwritten and obscure. Unclear responsibility for patient flow management decision-making is further complicated in departments with high role flexibility, where flow decisions are made on an ad hoc basis by available or nearby nurses. Other scholars have noted this ambiguity and blurring of emergency department flow management roles and have called for greater clarity in staff responsibilities [43-45].

Further, findings suggest a great disparity between departments' enforcement of patient flow metric expectations. Some nurse participants described working in environments with highly stringent timing guidelines, while other nurses were unsure if their hospitals had any timing expectations at all. The impact of these patient flow incentives is unclear. Scholars have recognized an inherent tension between quality and speed of care in the work of patient flow management [9,16,45], but research is needed to understand how emergency nurses manage this balance. Overall, strategies to clarify and promote consistency between patient flow management roles, norms, and incentives may benefit emergency nursing work.

Implications and Applications

This paper has addressed a gap in research describing emergency nurse patient flow management. We have identified numerous work system elements that shape emergency nursing work. These work system elements are theoretically linked to patient flow outcomes, but more research is needed to verify and measure their impact. The presented SEIPS model should therefore serve as a guiding framework for future studies that investigate the facilitators, barriers, and motivators to emergency nursing patient flow management.

This SEIPS work system analysis further demonstrates the complexity and difficulty of patient flow management. Patient flow researchers and health care administrators should embrace the study of human factors to support health care delivery. Increasing scholarly attention to the work of emergency nurses may offer new strategies to improve patient flow.

Impact Statement

The work of emergency nurse patient flow management has been poorly described. This nursing research study has used a human factors model to analyze the work system of emergency patient flow management. Study findings offer a theoretical framework to further investigate the impact of emergency nursing work on patient flow outcomes and identify novel patient flow solutions.

Conclusion

The SEIPS model has been applied in many health care projects and sectors to investigate care delivery [24]. Findings from this inductive, qualitative study provide further empirical support for the validity and usefulness of the SEIPS model. The unique contribution of this paper is the integration of the SEIPS model with the expertise of emergency nurses to describe their work system. There are many opportunities to better support emergency nurses in the complex, dynamic work of patient flow management.

Acknowledgments

The authors wish to thank Drs Cynthia Jacelon, Hari Balasubramanian, and Shannon Roberts for their guidance and support. This research was supported by a Beta Zeta at-Large grant from Sigma Theta Tau International Honor Society of Nursing.

Conflicts of Interest

None declared.

References

1. Javidan A, Hansen K, Higginson I, et al. White paper from the emergency department crowding and access block task force. : International Federation for Emergency Medicine; 2020. URL: https://assets.nationbuilder.com/ifem/pages/270/attachments/original/1650595379/ED-Crowding-and-Access-Block-Report-Final-June-30-2020.pdf?1650595379 [accessed 2024-11-26]
2. Rutherford PA, Anderson A, Kotagal UR, et al. Achieving hospital-wide patient flow. : Institute for Healthcare Improvement; 2020 URL: https://www.ihi.org/sites/default/files/IHIAchievingHospitalWidePatientFlowWhitePaper.pdf [accessed 2024-11-29]
3. Benjamin E, Jacelon C. An analysis of the concept of patient flow management. Nurs Forum 2022 May;57(3):429-436. [doi: 10.1111/nuf.12681] [Medline: 34862615]
4. Saghafian S, Austin G, Traub SJ. Operations research/management contributions to emergency department patient flow optimization: review and research prospects. IIE Trans Healthc Syst Eng 2015 Apr 3;5(2):101-123. [doi: 10.1080/19488300.2015.1017676]
5. Hybinette K, Praetorius G, Ekstedt M, Pukk Härenstam K. Exploring patient flow management through a lens of cognitive systems engineering. Ergonomics 2023 Dec;66(12):2106-2120. [doi: 10.1080/00140139.2023.2186321] [Medline: 36872878]
6. De Freitas L, Goodacre S, O'Hara R, Thokala P, Hariharan S. Interventions to improve patient flow in emergency departments: an umbrella review. Emerg Med J 2018 Oct;35(10):626-637. [doi: 10.1136/emermed-2017-207263] [Medline: 30093379]
7. Hall RW, editor. Patient Flow: Reducing Delay in Healthcare Delivery, 2nd edition: Springer; 2013.
8. Asplin BR, Magid DJ, Rhodes KV, Solberg LI, Lurie N, Camargo CA Jr. A conceptual model of emergency department crowding. Ann Emerg Med 2003 Aug;42(2):173-180. [doi: 10.1067/mem.2003.302]
9. Nugus P, Braithwaite J. The dynamic interaction of quality and efficiency in the emergency department: squaring the circle? Soc Sci Med 2010 Feb;70(4):511-517. [doi: 10.1016/j.socscimed.2009.11.001] [Medline: 19942332]

10. Nugus P, Forero R, McCarthy S, et al. The emergency department “carousel”: an ethnographically-derived model of the dynamics of patient flow. Int Emerg Nurs 2014 Jan;22(1):3-9. [doi: 10.1016/j.ienj.2013.01.001] [Medline: 23669027]
11. Nugus P, Holdgate A, Fry M, Forero R, McCarthy S, Braithwaite J. Work pressure and patient flow management in the emergency department: findings from an ethnographic study. Acad Emerg Med 2011 Oct;18(10):1045-1052. [doi: 10.1111/j.1553-2712.2011.01171.x] [Medline: 21996069]
12. Benjamin E, Wolf LA. “ Nurses are every bit of the flow:” emergency department nurses’ conceptualization of patient flow management. Nurs Forum 2022 Nov;57(6):1407-1414. [doi: 10.1111/nuf.12834] [Medline: 36398596]
13. Bergs J, Vandijck D, Hoogmartens O, et al. Emergency department crowding: time to shift the paradigm from predicting and controlling to analysing and managing. Int Emerg Nurs 2016 Jan;24:74-77. [doi: 10.1016/j.ienj.2015.05.004] [Medline: 27170954]
14. Mohiuddin S, Busby J, Savović J, et al. Patient flow within UK emergency departments: a systematic review of the use of computer simulation modelling methods. BMJ Open 2017 May 9;7(5):e015007. [doi: 10.1136/bmjopen-2016-015007] [Medline: 28487459]
15. Sharma S, Rafferty AM, Boiko O. The role and contribution of nurses to patient flow management in acute hospitals: a systematic review of mixed methods studies. Int J Nurs Stud 2020 Oct;110:103709. [doi: 10.1016/j.ijnurstu.2020.103709] [Medline: 32745787]
16. Benjamin E. The work of patient flow management: a grounded theory study of emergency nurses. Int Emerg Nurs 2024 Jun;74:101457. [doi: 10.1016/j.ienj.2024.101457] [Medline: 38744106]
17. Wolf L, Delao A, Simon C, Clark P, Burchill CN. Ensuring throughput: development and validation of charge nurse competencies for United States emergency care settings. J Emerg Nurs 2024 Jul;50(4):523-536. [doi: 10.1016/j.jen.2024.02.009] [Medline: 38573297]
18. Reay G, Rankin JA, Then KL. Momentary fitting in a fluid environment: a grounded theory of triage nurse decision making. Int Emerg Nurs 2016 May;26:8-13. [doi: 10.1016/j.ienj.2015.09.006] [Medline: 26522731]
19. Carayon P, editor. Handbook of Human Factors and Ergonomics in Health Care and Patient Safety, 2nd edition: CRC Press; 2012.
20. Hayden EM, Wong AH, Ackerman J, et al. Human factors and simulation in emergency medicine. Acad Emerg Med 2018 Feb;25(2):221-229. [doi: 10.1111/acem.13315] [Medline: 28925571]
21. Gifford R, van der Vaart T, Molleman E, van der Linden MC. Working together in emergency care? How professional boundaries influence integration efforts and operational performance. IJOPM 2022 Dec 19;42(13):54-78. [doi: 10.1108/IJOPM-10-2021-0644]
22. Chang AM, Cohen DJ, Lin A, et al. Hospital strategies for reducing emergency department crowding: a mixed-methods study. Ann Emerg Med 2018 Apr;71(4):497-505. [doi: 10.1016/j.annemergmed.2017.07.022] [Medline: 28844764]
23. Carayon P, Schoofs Hundt A, Karsh BT, et al. Work system design for patient safety: the SEIPS model. Qual Saf Health Care 2006 Dec;15 Suppl 1(Suppl 1):i50-i58. [doi: 10.1136/qshc.2005.015842] [Medline: 17142610]
24. Carayon P, Wetterneck TB, Rivera-Rodriguez AJ, et al. Human factors systems approach to healthcare quality and patient safety. Appl Ergon 2014 Jan;45(1):14-25. [doi: 10.1016/j.apergo.2013.04.023] [Medline: 23845724]
25. Carayon P, Wooldridge A, Hoonakker P, Hundt AS, Kelly MM. SEIPS 3.0: Human-centered design of the patient journey for patient safety. Appl Ergon 2020 Apr;84:103033. [doi: 10.1016/j.apergo.2019.103033] [Medline: 31987516]
26. Holden RJ, Carayon P. SEIPS 101 and seven simple SEIPS tools. BMJ Qual Saf 2021 Nov;30(11):901-910. [doi: 10.1136/bmjqs-2020-012538] [Medline: 34039748]
27. Donabedian A. The quality of care. How can it be assessed? JAMA 1988;260(12):1743-1748. [doi: 10.1001/jama.260.12.1743] [Medline: 3045356]
28. Fouquet SD, Fitzmaurice L, Chan YR, Palmer EM. Doctors documenting: an ethnographic and informatics approach to understanding attending physician documentation in the pediatric emergency department. J Am Med Inform Assoc 2021 Feb 15;28(2):239-248. [doi: 10.1093/jamia/ocaa252] [Medline: 33175154]
29. Rutkowski RA, Scheer E, Carlson C, et al. A scoping review of work system elements that influence emergency department disposition decision-making. Hum Fact Healthc 2023 Dec;4:100059. [doi: 10.1016/j.hfh.2023.100059]
30. Werner NE, Rutkowski R, Graske A, et al. Exploring SEIPS 2.0 as a model for analyzing care transitions across work systems. Appl Ergon 2020 Oct;88:103141. [doi: 10.1016/j.apergo.2020.103141] [Medline: 32421635]
31. Kebapci A, Ozkaynak M. Endotracheal tube cuff pressure management: an observational study guided by the SEIPS model. Dimens Crit Care Nurs 2022;41(2):64-75. [doi: 10.1097/DCC.0000000000000512] [Medline: 35099152]
32. Danesh MK, Garosi E, Mazloumi A, Najafi S. Identifying factors influencing cardiac care nurses’ work ability within the framework of the SEIPS model. Work 2020;66(3):569-577. [doi: 10.3233/WOR-203199] [Medline: 32623418]
33. Frith KH. Medication errors in the intensive care unit: literature review using the SEIPS model. AACN Adv Crit Care 2013;24(4):389-404. [doi: 10.1097/NCI.0b013e3182a8b516] [Medline: 24153217]
34. Attribution-noncommercial 4.0 international. Creative Commons. URL: https://creativecommons.org/licenses/by-nc/4.0/ [accessed 2024-12-09]
35. Clarke AE. Situational Analysis: Grounded Theory After the Postmodern Turn: Sage Publications; 2005.
36. Charmaz K. Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis: Sage Publications; 2006.

37. Fonteyn M, Fisher A. Use of think aloud method to study nurses' reasoning and decision making in clinical practice settings. J Neurosci Nurs 1995 Apr;27(2):124-128. [doi: 10.1097/01376517-199504000-00012] [Medline: 7622950]
38. Glaser BG, Anselm LS. The Discovery of Grounded Theory: Strategies for Qualitative Research: AdlineTransaction; 1967.
39. Young C, Patey C, Norman P, et al. Identifying relevant topics and training methods for emergency department flow training. CJEM 2022 Dec;24(8):837-843. [doi: 10.1007/s43678-022-00390-1] [Medline: 36242732]
40. Hossain F, Clatty A. Self-care strategies in response to nurses' moral injury during COVID-19 pandemic. Nurs Ethics 2021 Feb;28(1):23-32. [doi: 10.1177/0969733020961825] [Medline: 33124492]
41. Almasi S, Rabiei R, Moghaddasi H, Vahidi-Asl M. Emergency department quality dashboard; a systematic review of performance indicators, functionalities, and challenges. Arch Acad Emerg Med 2021;9(1):e47. [doi: 10.22037/aaem.v9i1.1230] [Medline: 34405145]
42. Nguyen Q, Wybrow M, Burstein F, Taylor D, Enticott J. Understanding the impacts of health information systems on patient flow management: a systematic review across several decades of research. PLoS ONE 2022;17(9):e0274493. [doi: 10.1371/journal.pone.0274493] [Medline: 36094946]
43. Pryce A, Unwin M, Kinsman L, McCann D. Delayed flow is a risk to patient safety: a mixed method analysis of emergency department patient flow. Int Emerg Nurs 2021 Jan;54:100956. [doi: 10.1016/j.ienj.2020.100956] [Medline: 33360361]
44. Wise S, Duffield C, Fry M, Roche M. Nurses' role in accomplishing interprofessional coordination: lessons in "almost managing" an emergency department team. J Nurs Manag 2022 Jan;30(1):198-204. [doi: 10.1111/jonm.13464] [Medline: 34436800]
45. Boiko O, Edwards M, Zschaler S, Miles S, Rafferty AM. Interprofessional barriers in patient flow management: an interview study of the views of emergency department staff involved in patient admissions. J Interprof Care 2021;35(3):334-342. [doi: 10.1080/13561820.2020.1760223] [Medline: 32506989]

Abbreviations

ICU: intensive care unit
SEIPS: Systems Engineering Initiative for Patient Safety

Edited by A Kushniruk, E Borycki; submitted 03.05.24; peer-reviewed by AS Kallberg, E Bai, PV Asaro; revised version received 17.09.24; accepted 23.09.24; published 10.12.24.

Please cite as:
Benjamin E, Giuliano KK
Work Systems Analysis of Emergency Nurse Patient Flow Management Using the Systems Engineering Initiative for Patient Safety Model: Applying Findings From a Grounded Theory Study
JMIR Hum Factors 2024;11:e60176
URL: https://humanfactors.jmir.org/2024/1/e60176
doi:10.2196/60176

Stunting Super App as an Effort Toward Stunting Management in Indonesia: Delphi and Pilot Study

Kadek Ayu Erika[1]; Nur Fadilah[1]; Aulia Insani Latif[1]; Nurhikmawaty Hasbiah[1]; Aidah Juliaty[2]; Harun Achmad[3]; Anugrayani Bustamin[4]

[1]Faculty of Nursing, Hasanuddin University, Jl. Perintis Kemerdekaan Kampus Tamalanrea Km.10, Makassar, Indonesia
[2]Faculty of Medicine, Hasanuddin University, Makassar, Indonesia
[3]Faculty of Dentistry, Hasanuddin University, Makassar, Indonesia
[4]Faculty of Engineering, Hasanuddin University, Makassar, Indonesia

Corresponding Author:
Kadek Ayu Erika
Faculty of Nursing, Hasanuddin University, Jl. Perintis Kemerdekaan Kampus Tamalanrea Km.10, Makassar, Indonesia

Abstract

Background: Currently, 30 million children are experiencing acute malnutrition, and 8 million children are severely underweight.

Objective: This study aimed to develop a stunting super app, a one-stop app designed to prevent and manage stunting in Indonesia.

Methods: This study consisted of three stages. Stage 1 used a 3-round Delphi study involving 12 experts. In stage 2, 4 experts and a parent of children with stunted growth created an Android app containing stunting educational materials. In stage 3, a pilot study involving a control group was conducted to evaluate parents' knowledge about stunting prevention through the app and standard interventions.

Results: In the Delphi study, 11 consensus statements were extracted; arranged in three major themes, including maternal health education, child health education, and environmental education; and applied in the form of the Sistem Evaluasi Kesehatan Anak Tumbuh Ideal (SEHATI) app. This app was assessed using a content validity index, with a cumulative agreement of ≥80% among the 5 individuals. The pilot study showed an increase in the knowledge of mothers of toddlers with stunted growth before and after the educational intervention (*P*=.001).

Conclusions: The SEHATI app provides educational content on stunting prevention that can increase the knowledge of mothers of toddlers with stunted growth.

(JMIR Hum Factors 2024;11:e54862) doi:10.2196/54862

KEYWORDS

stunting; stunting prevention; mobile app

Introduction

Stunting is an effect of malnutrition characterized by suboptimal physical growth caused by malnutrition in the first 1000 days of life, which can affect children's development and cognitive abilities [1]. Currently, >30 million children in 15 countries (ie, Afghanistan, Burkina Faso, Chad, Democratic Republic of the Congo, Ethiopia, Haiti, Kenya, Madagascar, Mali, Niger, Nigeria, Somalia, South Sudan, Sudan, and Yemen) are experiencing acute malnutrition, with 8 million of them being severely underweight [2]. In Indonesia, the Basic Health Research data show that the prevalence of stunting in toddlers in 2018 reached 30.8% [3]. Further, based on the Indonesian Toddler Nutrition Status Survey, the prevalence of stunting in 2021 was 24.4% (5.33 million children) [4]. This proves that the prevalence of stunting in Indonesia is still a public health issue because the figure exceeds the limit set by the World Health Organization, which is <20% [4]. The prevalence of stunting is increasing, owing to the lack of knowledge on the importance of meeting a balanced diet for toddlers during their golden age; the lack of awareness of primary caregivers, particularly mothers, in monitoring their food; and poor public knowledge on the early prevention of stunting [5]. Early stunting prevention can be started by increasing the understanding and role of the community in stunting prevention and increasing early detection programs in toddlers, which are expected to directly reduce this issue [6,7]. If inappropriately managed, stunting can lead to short- and long-term risks, such as decreased cognitive, motor, and language development, which will affect school performance, learning capacity, and other child potentials [8]. This can also incur high health expenditures and more frequent requirements for childcare, which will ultimately increase morbidity and mortality [9,10].

Currently, governments have implemented various national programs to ensure the health of mothers and children, including (1) specific intervention activities for giving iron tablets as a supplement to adolescent girls, brides-to-be, and pregnant women; (2) the promotion of exclusive breastfeeding and breastfeeding assistance; (3) classes for pregnant women; (4) information on a balanced diet; and (5) the Healthy Living Community Movement, which consists of increasing physical activity, vegetable and fruit consumption, and early disease detection [11]. At present, the prevalence of stunting in Indonesia is still very high and thus must be addressed through more innovative and creative interventions [12]. Currently, the most widely offered intervention is the use of IT, which offers easier access to health information anytime and anywhere.

The use of digital technology for health has become a popular practice when compared with other ways [13]. This is supported by the increasing use of mobile phones, as evidenced by global mobile user data reaching more than two-thirds (67.1%) of the world's population [14]. A form of technology usage in the health sector is the use of a smartphone app, which is commonly presented as a mobile health or eHealth app and can be further developed into a super app. A super app is a holistic app concept that has recently gained popularity, particularly in transport and other services offered by start-up companies such as Gojek, Grab, and Shopee. However, its use in the health sector remains unexplored. eHealth, or a super app, uses the interactions among medical informatics that are expected to reduce the prevalence of stunting in Indonesia, by referring to health services and information delivered or improved through technology [15]. In addition, assessment through this app is expected to be able to detect stunting early, monitor the condition of toddlers with stunted growth easily, and bridge stunting education to parents.

Preventing stunting is important to ensure the growth and physical development of children [16]. Having an app that aims to prevent stunting can improve the parents' and the community's knowledge about nutrition [17]. It can also provide easily accessible and relevant information and easily monitor the growth and development of children.

Several studies have evaluated the role of eHealth in stunting prevention. For example, Permana et al [18] took advantage of the use of "Nutrimo" or nutrition monitoring. Rianti et al [19] developed an app to prevent babies from being born underweight, which focuses on the health of pregnant women during pregnancy, particularly to comply with taking iron tablets. Kasjono and Suryani [20] developed Gerakan Anti Stunting (GASING, antistunting movement), an app for the delivery of information on a balanced diet and the promotion of clean and healthy living behaviors. Utario and Sutriyanti [21] developed an offline app for stunting management for Integrated Service Post task forces. However, current apps are neither comprehensive nor integrated into one platform and do not focus on stunting prevention. Therefore, this study is interested in developing a stunting super app, a one-stop app that can help prevent and handle malnutrition and thereby facilitate the eradication of stunting in Indonesia.

Methods

Study Design

Overview

This study consisted of three stages. In the first stage, content development related to stunting prevention education was performed using the Delphi study. The Delphi study is a method used to create context and collect opinions from experts on a particular topic [22]. In the second stage, the app was developed, which was later named Sistem Evaluasi Kesehatan Anak Tumbuh Ideal (SEHATI). In the final stage, a knowledge assessment was performed before and after the health education using this app with standard educational media.

Stage 1: Content Development by Experts

In stage 1, a classic Delphi study and web-based Delphi study were performed in 3 rounds for content development. This study involved 12 experts who met the inclusion criteria: expert educational backgrounds in medicine, nursing, nutrition, or physiotherapy who had experience with stunting for >2 years in Indonesia or parents who had children with stunting. In the first round of the Delphi study (first questionnaire), a classical Delphi study was performed by distributing a questionnaire with open-ended questions. Participants were asked, "What do you think can be done to prevent stunting?" The questionnaire was sent as a Google Form via WhatsApp version 2.24.23.78 (Meta Platforms). The answers obtained on the first questionnaire were collected and analyzed.

In the second Delphi round (second questionnaire), the analyzed answers were then sent back to the experts and scored using a 4-point Likert scale: highly relevant, relevant, irrelevant, and highly irrelevant. The answers were analyzed with a cumulative agreement indicator of ≥80%. In the third Delphi round (third questionnaire), answers that obtained a cumulative approval of ≥80% were returned for final validation, and expert approval was achieved by assessing the appropriateness of each answer component based on the Likert scale. These stages generated the final consensus.

Stage 2: App Development by IT Experts

IT experts developed the app based on the consensus obtained in the first stage. The content validity index (CVI) test involved 5 experts, who were not the experts employed in the previous stage, with the CVI standard of ≥80%. These participants were experts in nursing, medicine, IT, or language and a parent who had toddlers with stunted growth.

The app was developed beginning with understanding the needs of the target audience through stakeholder meetings with health care professionals and potential users. We defined the app's core features, such as growth monitoring, educational content, and consultation tools. We then chose Flutter 3.13.0 (Google) for its responsive and visually appealing user interface, and then we moved on to designing and prototyping using human-computer interaction principles like accessibility, usability, and user experience. Wireframes were created using Figma 24.33.0 (Figma) to outline the app's layout, and detailed designs were developed focusing on user experience and navigation. An interactive prototype to simulate how the app will work and gather feedback from stakeholders to make improvements was then built. The app was then developed using Flutter and Dart 3.3.0 (Google). Flutter allowed for a responsive and attractive user interface, while Dart was the programming language used for app development. Firebase was set up for back-end services, including data management and real-time updates, and Firebase Authentication was implemented to manage user log-ins and security. We continuously tested the app to find and fix any issues. This included unit tests for individual parts and integration tests for how different features worked together. The app was prepared for launch by submitting it to the Google Play Store, ensuring it met all guidelines. After the launch, we continued to support the app by fixing bugs, adding new features, and improving performance based on user feedback.

Stage 3: Stunting Super App Evaluation

A pilot study was conducted to evaluate changes in parents' knowledge about stunting prevention after intervention using SEHATI. Knowledge assessment took place using the contents developed in stage 1. In this stage, the pilot study involved 30 participants selected based on the inclusion and exclusion criteria. Inclusion criteria were participants who could use a mobile phone and mothers who have children younger than 5 years. Meanwhile, the exclusion criteria were participants who use iOS software on their mobile phones. In the pilot study, the recommended sample size was at least 12 participants [23]. Changes in knowledge of stunting prevention were assessed, and changes attributed to the use of SEHATI and other educational media were compared. Parents (1) with toddlers aged <2 years (2) who possessed and could operate a smartphone (3) and were willing to participate in the study were enrolled.

The app was given once to respondents, starting with an explanation of the manual book by the research team. After the app was installed and respondents could use the SEHATI app, respondents were given educational interventions to prevent stunting through the app. Respondents could access the app anytime and anywhere because it could be used offline, so it was also suitable for use in rural areas that have limited internet access.

Data Analysis and Statistical Test

Descriptive statistical tests were used to assess the characteristics of the respondents. The results are displayed as mean, SD, and percentage. The results obtained from the consensus of the experts were integrated into an app, and the CVI test was taken with expert judgments. The CVI indicator was considered valid if the value was ≥0.78. Furthermore, bivariate analysis was performed to assess the increase in knowledge in the group using the nonparametric Wilcoxon test and Mann-Whitney test. All data were analyzed using IBM SPSS Statistics for Microsoft Windows version 21 (IBM Corp).

Ethical Considerations

This study received Health Research Ethics Commission approval (#126/UN4.6.4.5.31/PP36/2023). Before the study, all experts received a consent form, which explained the purpose, procedures, benefits, confidentiality, risks of the research, and the right to refuse or freely withdraw participation in the study. All expert data were anonymized. Participants were given compensation in the form of internet data vouchers (US $18.84).

Results

Delphi Study

The educational theme for mothers consisted of 4 contents with a cumulative agreement presentation of ≥80%, including preparing prepregnancy nutrition, meeting nutritional needs during pregnancy, routine pregnancy checks, and regular intake of vitamins during pregnancy. Meanwhile, the educational theme for children that met the cumulative agreement presentation of ≥80% had 4 contents, namely, exclusive breastfeeding, provision of complementary feeding, routine monitoring of children's growth and development, and complete immunizations. Meanwhile, the theme for environmental health education included 3 contents, with a cumulative agreement presentation of ≥80%, including environmental cleanliness, parenting, and socioeconomic support.

In stage 1, the Delphi study was conducted to develop a stunting prevention educational app involving 12 experts, and the 11 contents identified were grouped into three major themes, namely, educational content for mothers, children, and the environment (Table 1).

Table . Table 1. Result of stage I of the Delphi study: educational contents for stunting prevention.

Topic	Prevention components	Likert scale responses (n=12), n (%)				Cumulative agreement, n (%)
		Highly irrelevant	Irrelevant	Relevant	Highly relevant	
Maternal health educational contents						
Preparing prepregnancy nutrition	Brides prepare nutrition before pregnancy (prepregnancy) by consuming a balanced diet (carbohydrates, vegetables, animal proteins, vitamins, and minerals) and iron tablets at the age of 12 - 18 years with a dose of one tablet per week. Mothers also need to take height and weight measurements to monitor BMI.	0 (0)	0 (0)	3 (25)	9 (75)	12 (100)
Meeting nutritional needs during pregnancy	Pregnant women must meet their nutritional needs by consuming macronutrients and micronutrients. a. Macronutrients • Carbohydrates such as rice, cassava, corn, sweet potatoes, and others • Proteins from plants such as beans, tofu, tempe, and others • Proteins from animals such as meat, chicken, eggs, and milk, and others b. Micronutrients • Multivitamins contained in vegetables and fruits • Minerals such as calcium, folic acid, iron, and iodine	0 (0)	0 (0)	3 (25)	9 (75)	12 (100)

Topic	Prevention components	Likert scale responses (n=12), n (%)				Cumulative agreement, n (%)
		Highly irrelevant	Irrelevant	Relevant	Highly relevant	
Routine pregnancy check	Mothers routinely undergo pregnancy checks at least six times during pregnancy, as follows: • First trimester (gestational age 0‐12 weeks): 1 time • Second trimester (gestational age>12‐24 weeks): 2 times • Third trimester (gestational age>24‐40 weeks): 3 times	0 (0)	0 (0)	2 (17)	10 (83)	12 (100)
Routine vitamin consumption	Mothers routinely: • Take folic acid tablets once a day after meals during pregnancy • Take one iron tablet every day during pregnancy, for at least 90 tablets, to prevent blood deficiency (anemia) and start as early as possible	0 (0)	0 (0)	3 (25)	9 (75)	12 (100)
Children’s health educational content						
Exclusive breastfeeding	Mothers give breast milk exclusively to newborns up to 6 months of age.	0 (0)	0 (0)	3 (25)	9 (75)	12 (100)
Complementary feeding	Mothers feed complementary foods that are adjusted to the age (month) of the child	0 (0)	1 (8)	1 (8)	10 (84)	11 (93)

Topic	Prevention components	Likert scale responses (n=12), n (%)				Cumulative agreement, n (%)
		Highly irrelevant	Irrelevant	Relevant	Highly relevant	
Routinely monitor children’s growth and development	Mothers routinely monitor the growth and development of children by weighing and measuring the length or height at each visit to the Integrated Service Post with the following frequency: • Monthly visit for toddlers aged 0 - 12 months • Every 3 months for toddlers aged 12 - 24 months • Every 6 months for toddlers aged 24 - 72 months • Every 3 months for head circumference measurement at the age of 0 - 12 months • Every 6 months for head circumference measurement at the age of 18 - 72 months	0 (0)	1 (8)	1 (8)	10 (84)	11 (93)
Complete immunization, vitamin A, and deworming		0 (0)	1 (7)	3 (22)	8 (71)	11 (93)

Topic	Prevention components	Likert scale responses (n=12), n (%)				Cumulative agreement, n (%)
		Highly irrelevant	Irrelevant	Relevant	Highly relevant	
	Mothers routinely take their children for complete immunization, vitamin A supplementation, and deworming with the following schedule: • From birth to 24 hours old: hepatitis B, BCG[a], and polio drops 1 • 1 month old: BCG and polio drops 1 • 2 months old: DPT-HB-Hib[b] 1 and polio drops 2 • 3 months old: DPT-HB-Hib 2 and polio drops 3 • 4 months old: DPT-HB-Hib 3, polio drops 4, and injectable polio vaccine • 9 months old: measles-rubella • 18 months old: continued measles-rubella and DPT-HB-Hib • Administration of blue vitamin A capsule once at the age of 6 - 11 months • Administration of red vitamin A capsule once every 6 months at the age of 12 - 59 months • Deworming from the age of 12 months, taken twice a year or every 6 months					
Environmental educational content						

Topic	Prevention components	Likert scale responses (n=12), n (%)				Cumulative agreement, n (%)
		Highly irrelevant	Irrelevant	Relevant	Highly relevant	
Keeping the environment clean	Sanitation, clean latrines, stay away from cigarette smoke and wear proper clothing.	0 (0)	0 (0)	3 (25)	9 (75)	12 (100)

Topic	Prevention components	Likert scale responses (n=12), n (%)				Cumulative agreement, n (%)
		Highly irrelevant	Irrelevant	Relevant	Highly relevant	
Parenting	Parenting plays an important role in preventing stunting. Parenting is divided into *asah*, *asih*, and *asuh*. • *Asah* refers to the stimulation of children's intelligence, such as the provision of educational tools that help them become smarter • *Asih* refers to efforts to develop a child's affection, spirituality, independence, and need for a sense of security and comfort • *Asuh* encompasses physical and biological needs, including nutritional requirements, immunization, personal and environmental cleanliness, medical care, and physical activity and play	0 (0)	0 (0)	3 (25)	9 (75)	12 (100)
Socioeconomic support	Parents must be aware that economic support plays an important role in helping children obtain their nutritional needs. One of the things that can prevent malnutrition is to save money during pregnancy, locally known as *Tabungan Ibu Bersalin* (*Tabulin*). It can be started at the beginning of pregnancy to meet maternity needs and the initial needs of the baby at birth.	0 (0)	1 (7)	3 (29)	8 (64)	11 (93)

[a]BCG: Bacillus Calmette-Guérin.

[b]DPT-HB-Hib: diptheria, pertussis, tetanus, hepatitis B, *Haemophilus influenzae* type b pediatric unspecified.

Stunting Super App Development: Content Validity

In the second stage, the validity of the super app contents was evaluated. The contents were assessed by 4 experts from various disciplines (nursing, medicine, IT, and language) and a parent of children with stunted growth (Table 2).

Table . Quantitative evaluation of SEHATI[a] app contents.

Assessment items	Likert scale responses (n=5), n (%)				Cumulative agreement, n (%)
	Highly irrelevant	Irrelevant	Relevant	Highly relevant	
Simple design for easy use	0 (0)	0 (0)	1 (20)	4 (80)	5 (100)
Uses language and sentences that are easy to understand	0 (0)	0 (0)	2 (40)	3 (60)	5 (100)
Provides features based on user categories	0 (0)	0 (0)	1 (20)	4 (80)	5 (100)
Combines images with text	0 (0)	0 (0)	0 (0)	5 (100)	5 (100)
Avoids distracting elements (such as excessive accessories or styles)	0 (0)	0 (0)	2 (40)	3 (60)	5 (100)
Users cannot edit the app's contents	0 (0)	0 (0)	4 (80)	1 (20)	5 (100)
Provides structured information for easy comprehension	0 (0)	0 (0)	3 (60)	2 (40)	5 (100)
Users can adjust notification sounds (alarm notifications)	0 (0)	0 (0)	1 (20)	4 (80)	5 (100)
Users can set alarms	0 (0)	0 (0)	3 (60)	2 (40)	5 (100)
Uses multiple-choice mode	0 (0)	0 (0)	1 (20)	4 (80)	5 (100)
Users can download the app through WhatsApp, Shareit, Bluetooth, or email	0 (0)	0 (0)	2 (40)	3 (60)	5 (100)
Users can take pictures	0 (0)	0 (0)	3 (60)	2 (40)	5 (100)
The app does not exceed 30 MB in size	0 (0)	1 (20)	2 (40)	2 (40)	4 (80)

[a]SEHATI: Sistem Evaluasi Kesehatan Anak Tumbuh Ideal.

There were 13 scoring items using 4 Likert scales (highly irrelevant, irrelevant, relevant, and highly relevant). The validity limit of the app contents was set to a cumulative agreement of ≥80%, and if the evaluation value was <80%, then adjustments were made to the app content. All items in the CVI test met the cumulative agreement presentation (≥80%). In addition, the results of the qualitative evaluation of SEHATI are shown in Table 3.

Table . Qualitative evaluation of SEHATI[a] content.

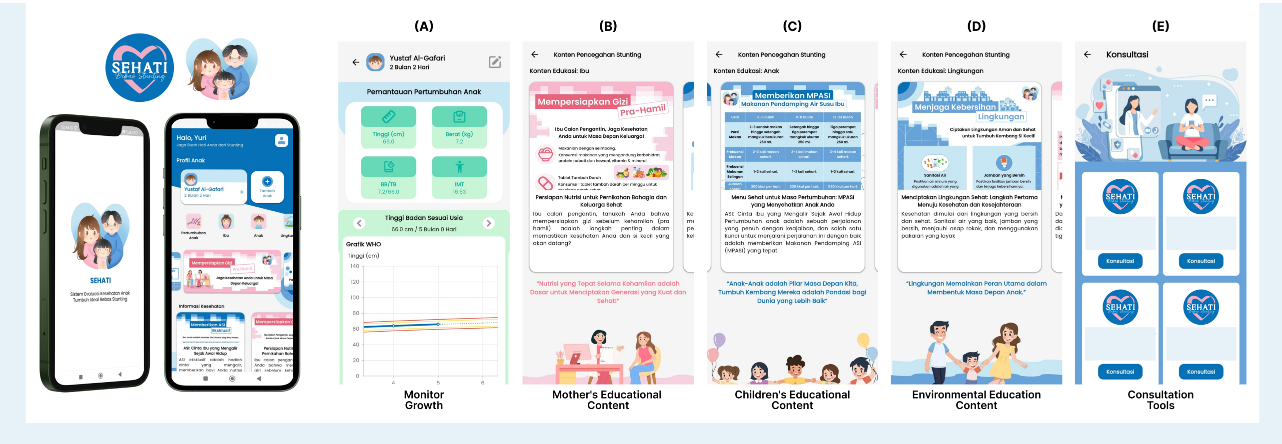
Reviewer field	Comments of the reviewers
Nursing	*This app has provided educational needs from every critical stage for stunting prevention starting from before pregnancy in mothers until 1000 days after giving birth. In the future, there needs to be regular content updates and addition of information needed by users according to the latest developments in child health issues.* [Reviewer Ar]
Medicine	*The app is very good; however, it is necessary to add information for adolescent girls and pregnant women and monitor growth and development based on BW/BH, BMI/A, MUAC, and HC.* [Reviewer Aj]
IT	*The interface of the system is already decent and has met the needs of the users. It needs to consider space and memory upgrade to anticipate increasing users.* [Reviewer U]
Indonesian linguist	*This app benefits parents to avoid stunting. Its clear and communicative language makes this app easy to use. The design and important information regarding the health of the mother, child, and the environment are displayed very well. This app also has a consultation feature, which makes it easier for users to consult their health if they experience symptoms. Hopefully, it can be useful for the public given its simple operation and economical use.* [Reviewer M]
Mother of toddlers with stunted growth	*This app is good enough because it covers all stages of development starting from the womb, after birth, and subsequent development. Even so, several components must be added such as myths and facts related to pregnancy and children, foods to be avoided, dos and don'ts, and traditions that may ruin child development.* [Reviewer R]

[a]SEHATI: Sistem Evaluasi Kesehatan Anak Tumbuh Ideal.

The results showed that content must be updated regularly according to the development of the latest child health issues (nursing). Also, additional information is necessary for adolescent girls and pregnant women and for growth and development monitoring (medicine), and space and storage upgrades are required if the number of users increases (IT). The app language was clear, communicative, and economical in its use (linguist); provided consultation features; and covered all stages of development. However, some components must be added, including myths and facts about pregnancy, childhood, and inappropriate traditions (mother; Table 3).

This app was named SEHATI (Sistem Evaluasi Kesehatan Anak Tumbuh Ideal, ie, Ideal Growth Child Health Evaluation System), which means a comprehensive, one-stop health intervention app for stunting prevention and management. The features of SEHATI are summarized in Figure 1.

Figure 1. Sistem Evaluasi Kesehatan Anak Tumbuh Ideal (SEHATI) app features. SEHATI is an app designed to simplify child growth monitoring and includes stunting information. Parents can track their children's growth and ensure their well-being with an easy-to-use interface. (A) Track and display a child's growth through charts: weight-for-age, length/height-for-age, weight-for-length/height, and BMI-for-age. (B) Informational resources that focus on maternal health to guide mothers in making the right health decisions. (C) Learning materials to support a child's development and improve health awareness. (D) Information on environmental factors that have an impact on children's growth. (E) Users can consult with health experts for personalized advice regarding a child's growth.



Based on the results of the observations and interviews with researchers, the experience of users of this app was good. Respondents were happy because the app contained health information on stunting prevention that could be accessed easily at any time. This app was helpful for mothers in monitoring child growth, and because the app was equipped with growth charts, data could be input by mothers themselves every month. In addition, there was a consultation feature that was helpful for mothers in dealing with their children's health.

Pilot Study: Increasing the Knowledge of Stunting Through SEHATI

This stage was conducted to track changes in the knowledge of mothers of toddlers with stunted growth regarding stunting prevention. A total of 30 mothers who had babies or toddlers and could operate smartphones and the SEHATI app received educational intervention. Their knowledge was evaluated twice: once before education (pretest) and once after education (posttest). The average age of mothers was 29 years and the average age of the children was 6 months. Most participants had a bachelor's degree (n=20, 66.7%), with the dominant occupation being a housewife (n=20, 66.7%; Table 4).

Table . Characteristics of the respondents by age, education, and occupation.

Characteristics of respondents	Values
Age of mothers (years; n=30)	
Mean (SD)	29 (5.240)
Min-max	17 - 39
Age of children (months; n=30)	
Mean (SD)	6 (7.617)
Min-max	1 - 38
Education (n=30), n (%)	
Primary school	1 (3)
Junior high school	2 (7)
Senior high school	7 (23)
Undergraduate degree	20 (67)
Occupation (n=30), n (%)	
Housewife	20 (67)
Government employee	9 (30)
Self-employed	1 (3)

The statistical test used to determine the difference between pretests and posttests was a 2-tailed *t* test because the data was normally distributed. The results showed a significant difference after participants received stunting prevention education using the SEHATI app (*P*=.001), which means there was an increase in maternal knowledge after participating in education with the app (Table 5).

Table . Differences in knowledge before and after stunting prevention education.

	Mean test score (SD)[a]	SE	*P* value[b]
Scores (n=30)			.001
Pretest	14.63 (1.326)	0.242	
Posttest	16.20 (1.495)	0.273	

[a]The number of pretest and posttest questions was 20.

[b]A paired sample, 2-tailed *t* test was done to determine significance.

Discussion

Preparation of Stunting Prevention Education Contents

Initially, the Delphi technique was used in a series of studies by the RAND Corporation in the 1950s [24]. It is one of the methods that the group used to survey and gather the opinions of experts on a particular topic [25]. In this study, a combination of classic and web-based Delphi techniques were used, which began with open questions to experts, and the study was conducted in three rounds. The classic Delphi used an open-ended questionnaire administered in three rounds, whereas the web-based Delphi used Google Forms distributed through WhatsApp [26]. The Delphi study was conducted in 3 rounds as adapted from Abrar et al [27].

The preparation of the app's contents involved 12 experts, who were selected based on education level, work experience, and stunting training certificates, and they extracted three major themes, namely, maternal, child, and environmental health educational content. In addition [28], the authors involved parents who had experience in caring for toddlers with stunting, provided information about caring for toddlers with stunted growth, and gained knowledge on preventing children from experiencing the adverse effects of stunting again [29].

A total of 11 themes on stunting prevention were identified in the Delphi study. This is in line with the component of needs regarding stunting prevention in terms of maternal health in the study by Ekayanthi and Suryani [30]. In their study, the nutrition and health of pregnant women were improved by meeting the requirement for macronutrients and micronutrients, giving nutritional supplements (iron tablets), and regular monitoring of the health of pregnant women.

Stunting prevention based on children's health is carried out by good parenting, exclusive breastfeeding, intake of appropriate complementary foods, immunization, provision of a psychosocial stimulus to children, and basic health care [31]. Environmental education also contributes to stunting prevention by practicing clean and healthy living behavior, increasing access to clean water and sanitation facilities, and maintaining environmental cleanliness [32].

Content Validity Evaluation of SEHATI

In the second stage, the validity of the app was tested by expert judgment. The authors chose Android as the educational medium because of its effectiveness as an information tool and its increasing use for educational media [24]. According to Putra et al [33], the use of Android-based educational media is an effective medium to increase public knowledge and attitudes toward stunting. In line with this finding, Fitriami and Galaresa [34] found that Android is a comprehensive means that helps improve maternal nutritional behavior.

In this study, all items in the CVI assessment component on the SEHATI app met the cumulative agreement percentage (≥80%), and they recommended updating the contents regularly so that mothers have easy access to the latest information regarding child development issues and stunting prevention [33]. Further, young women need additional information to prepare them for the next stage of life as mothers so that stunting can be prevented [35]. In addition, more content about pregnancy myths and facts must be considered to prevent fetal stunting in mothers believing such myths [36].

In addition, the SEHATI app's sustainability plan is that the development team will collaborate with health stakeholders for sustainable maintenance. Funding and maintenance of the SEHATI app will be charged to stakeholder institutions. SEHATI is designed to complement existing stunting prevention programs in Indonesia through comprehensive growth monitoring features and easy access to educational materials. SEHATI also strengthens the government's efforts to monitor and improve the nutritional status of children. The app supports existing systems by providing structured, real-time growth data that can be accessed by relevant parties for further analysis under certain conditions. SEHATI also allows users to consult directly with health experts, facilitating coordination and early intervention in high-risk cases.

Change in the Level of Knowledge

The results showed that the participants demonstrated an increase in knowledge after using the SEHATI app. The urgency of disseminating knowledge has been stated in the Regulation of the Minister of Health of the Republic of Indonesia No. 67 of 2016, which focuses on increasing true and comprehensive knowledge regarding prevention, transmission, treatment, and a clean and healthy lifestyle [37]. The app is considered effective because it is associated with efficient use of time and energy in providing education to parents who have malnourished toddlers [29]. In addition, Android apps can be accessed by phone and by everyone, allow for multitasking, and easily notify parents or the community [34]. Regarding app use by parents, a reliable, up-to-date, and evidence-based app is crucial, particularly when using apps in the context of routine care for children [38].

Limitation

The limitation of this study was related to the sampling approach, in which experts were chosen based on the researchers' judgment. Thus, to prevent this bias, a purposeful sampling technique was used in the selection of expert panelists, who were clinically competent and participated voluntarily.

Conclusions

The Delphi study resulted in 11 consensus themes for stunting prevention. The results were integrated into the SEHATI app, which was assessed using the CVI and the cumulative percentage of agreement from 4 experts in nursing, medicine, IT, language, and a parent of toddlers with stunted growth was ≥80%. SEHATI as a stunting prevention educational app contributes significantly to the increase in stunting knowledge for mothers who have toddlers with stunted growth.

Acknowledgments

The authors thank all who participated in this study. The authors also wish to thank their colleagues at the Faculty of Nursing who have helped in this study and the Hasanuddin University Research and Community Service Institute.

Authors' Contributions

KAE, NF, AIL, NH, AJ, HA, and AB all equally participated in the preparation of proposals, research processes, and preparation of the final report.

Conflicts of Interest

None declared.

References

1. Akombi BJ, Agho KE, Hall JJ, Wali N, Renzaho AMN, Merom D. Stunting, wasting and underweight in sub-Saharan Africa: a systematic review. Int J Environ Res Public Health 2017 Aug 1;14(8):863. [doi: 10.3390/ijerph14080863] [Medline: 28788108]
2. Urgent action needed as acute malnutrition threatens the lives of millions of vulnerable children. World Health Organization. 2023 Jan 12. URL: https://www.who.int/news/item/12-01-2023-urgent-action-needed-as-acute-malnutrition-threatens-the-lives-of-millions-of-vulnerable-children [accessed 2024-11-19]
3. Badan Penelitian dan Pengembangan Kesehatan. Hasil riset kesehatan dasar tahun 2018. In: Laporan Nasional Riskesdas 2018: Lembaga Penerbit Badan Penelitian dan Pengembangan Kesehatan; 2018. URL: https://repository.badankebijakan.kemkes.go.id/id/eprint/3514/ [accessed 2024-11-25]
4. Hasil Survei Status Gizi Indonesia (SSGI) 2022. Repositori Badan Kebijakan Pembangunan Kesehatan. 2022. URL: https://repository.badankebijakan.kemkes.go.id/id/eprint/4855/3/Buku%20Saku%20SSGI%202022%20rev%20270123%20OK.pdf
5. Isni K, Dinni SM. Pelatihan pengukuran status gizi balita sebagai upaya pencegahan stunting sejak dini pada ibu di Dusun Randugunting, Sleman, DIY. Panrita Abdi J Pengabd pada Masy 2020;4(1):60. [doi: 10.20956/pa.v4i1.7299]
6. Puspitasari A, Putra WD, Amir H. Pencegahan stunting pada anak di desa Tamangapa kec. Ma'rang kab. Pangkep. J Pengabdi Masy 2021;1(1):05-08. [doi: 10.53690/ipm.v1i1.3]
7. Erika KA, Mulhaeriah M, Amir NA. The relations between birth weight and the growth of children under five in Puskesmas Goarie subdistrict Marioriwawo regency Soppeng. Indones Contemp Nurs J 2019;3(2):15. [doi: 10.20956/icon.v3i2.4049]
8. Rafika M. Dampak Stunting Pada Kondisi Psikologis Anak. Bul Jagaddhita 2019;1(1):1-4 [FREE Full text]
9. Hendrawati S, Adistie F, Maryam NNA. Effectiveness of developmental care on physiological functions' low birth weight babies: a literature review. Indones Contemp Nurs J 2020;4(2):52-63 [FREE Full text]
10. Pratiwi R. Dampak status gizi pendek (stunting) terhadap prestasi belajar. J Nurs Updat Ed Khusus 2021;12(2):11-23 [FREE Full text] [doi: 10.36089/nu.v12i2.317]
11. Pus Data dan Informasi, Kementerian Kesehatan RI. Cegah stunting, itu penting. Ministry of Health of Republic Indonesia. 2018. URL: https://ayosehat.kemkes.go.id/cegah-stunting-itu-penting [accessed 2024-11-25]
12. Huey SL, Mehta S. Stunting: the need for application of advances in technology to understand a complex health problem. EBioMedicine 2016 Apr;6:26-27. [doi: 10.1016/j.ebiom.2016.03.013] [Medline: 27211543]
13. WHO guideline: recommendations on digital interventions for health system strengthening. World Health Organization. 2019. URL: https://www.who.int/publications/i/item/9789241550505 [accessed 2024-11-19]
14. Persentase penduduk yang memiliki/menguasai telepon seluler menurut provinsi dan klasifikasi daerah 2018-2020. Badan Pusat Statistik. 2022. URL: https://www.bps.go.id/id/statistics-table/2/Mzk1IzI=/persentase-penduduk-yang-memiliki-menguasai-telepon-seluler-menurut-provinsi-dan-klasifikasi-daerah.html [accessed 2024-11-25]
15. Syaroni W, Munir Z. Pemanfaatan aplikasi android dalam mendiagnosa dan memonitoring kasus stunting lebih dini. J Tek Elektro dan Komput 2020;9(3):189-196. [doi: 10.35793/jtek.v9i3.30377]
16. Saleh A, Syahrul S, Hadju V, Andriani I, Restika I. Role of maternal in preventing stunting: a systematic review. Gac Sanit 2021;35 Suppl 2:S576-S582. [doi: 10.1016/j.gaceta.2021.10.087] [Medline: 34929905]
17. Ponum M, Khan S, Hasan O, et al. Stunting diagnostic and awareness: impact assessment study of sociodemographic factors of stunting among school-going children of Pakistan. BMC Pediatr 2020 May 19;20(1):232. [doi: 10.1186/s12887-020-02139-0] [Medline: 32429876]
18. Permana AA, Perdana AT, Handayani N, Destriana R. A stunting prevention application "Nutrimo" (nutrition monitoring). J Phys Conf Ser 2021 Mar 1;1844(1):012023. [doi: 10.1088/1742-6596/1844/1/012023]
19. Rianti E, Triwinarto A, Lukman E. Aplikasi cegah anak lahir stunting berbasis android. Pros Forum Ilm Tah IAKMI (Ikatan Ahli Kesehat Masy Indones) 2020:1-6 [FREE Full text]
20. Kasjono HS, Suryani E. Pengaruh aplikasi pencegahan stunting "gasing" terhadap perilaku pencegahan stunting pada siswi SMA di wilayah kecamatan Kalibawang Kulon Progo. J Nutr 2020;22(1):16-22. [doi: 10.29238/jnutri.v22i1.200]
21. Utario Y, Sutriyanti Y. Aplikasi offline stunting untuk meningkatkan pengetahuan Kader Posyandu Di Puskesmas Perumnas kabupaten Rejang Lebong. J Abdimas Kesehat Perintis 2020;2(1):25-30 [FREE Full text]
22. Barrett D, Heale R. What are Delphi studies? Evid Based Nurs 2020 Jul;23(3):68-69. [doi: 10.1136/ebnurs-2020-103303] [Medline: 32430290]
23. Moore CG, Carter RE, Nietert PJ, Stewart PW. Recommendations for planning pilot studies. Clin Transl Res 2011;4(5). [doi: 10.1111/j.1752-8062.2011.00347.x]
24. Lyu F, Zheng C, Wang H, Nie C, Ma X, Xia X, et al. Establishment of a clinician-led guideline on the diagnosis and treatment of hirayama disease using a modified Delphi technique. Clin Neurophysiol 2020;131(6):1311-1319. [doi: 10.1016/j.clinph]
25. Drumm S, Bradley C, Moriarty F. 'More of an art than a science'? The development, design, and mechanics of the Delphi technique. Res Soc Admin Pharm 2022;8(1):2230-2236. [doi: 10.1016/j.sapharm.2021.06.027]

26. Keeney S, Hasson F, McKenna H. The Delphi Technique in Nursing and Health Research: Wiley-Blackwell; 2011. [doi: 10.1002/9781444392029]
27. Abrar EA, Yusuf S, Sjattar EL, Rachmawaty R. Development and evaluation educational videos of diabetic foot care in traditional languages to enhance knowledge of patients diagnosed with diabetes and risk for diabetic foot ulcers. Prim Care Diabetes 2020 Apr;14(2):104-110. [doi: 10.1016/j.pcd.2019.06.005] [Medline: 31311727]
28. Latif AI, Sjattar EL, Erika KA. Models and benefits of mobile health application to support patient with tuberculosis: a literature review. Enferm Clín 2020 Mar;30:163-167. [doi: 10.1016/j.enfcli.2019.07.069]
29. Rahmawati E, Faozi A, Rahmat DY, Sopiah P. Pengalaman ibu dalam merawat balita stunting usia (3-5) tahun. J Kesehat Tambusai 2023;4(2 SE-Articles):1079-1085. [doi: 10.31004/jkt.v4i2.15034]
30. Ekayanthi NWD, Suryani P. Edukasi gizi pada ibu hamil mencegah stunting pada kelas ibu hamil. Jurnal Kesehatan 2019;10(3):312-319. [doi: 10.26630/jk.v10i3.1389]
31. Aninora NR, Satria E. Pelaksanaan edukasi pencegahan stunting pada ibu hamil di wilayah Kerja Puskesmas Lubuk Buaya Kota Padang. EBIMA J Edukasi Bidan Di Masy 2021;2(2):10-14. [doi: 10.36929/ebima.v2i2.425]
32. Landi S, Regaletha TAL, Muntasir M. Pencegahan penyakit infeksi menular melalui upaya sanitasi dan higiene personal untuk pencegahan stunting di desa Tablolong Kecamatan Kupang Barat Kabupaten Kupang NTT. Jurnal Pengabdian Pada Masyarakat Kepulauan Lahan Kering 2023;4(2):71-83. [doi: 10.51556/jpkmkelaker.v4i2.243]
33. Putra ME, Femelia W, Gizi J, Masyarakat FK, Andalas U, Studi P, et al. Efektifitas aplikasi edukasi gizi remaja berbasis Android untuk pencegahan stunting. J Endur 2022;6(2):443-451 [FREE Full text]
34. Fitriami E, Galaresa AV. Edukasi pencegahan stunting berbasis aplikasi Android dalam meningkatkan pengetahuan dan sikap ibu. Citra Delima Sci J Citra Int Inst 2021;5(2):78-85. [doi: 10.33862/citradelima.v5i2.258]
35. Muchtar F, Rejeki S, Elvira I, Hastian H. Edukasi pengenalan stunting pada remaja putri. Lamahu J Pengabdi Masy Terintegrasi 2023;2(2):138-144. [doi: 10.34312/ljpmt.v2i2.21400]
36. Marpaung NL, Hutabarat S, Amri R, Siahaan CN, Octa M, Gultom V. Training of stunting prevention movement in Tasik Serai Barat village. . 2023(4) p. 182-190. [doi: 10.31258/cers.3.4.182-190]
37. Peraturan menteri kesehatan RI no67 tahun 2016. Database Peraturan. 2016. URL: https://peraturan.bpk.go.id/Details/114486/permenkes-no-67-tahun-2016 [accessed 2024-11-25]
38. Cheng H, Tutt A, Llewellyn C, et al. Content and quality of infant feeding smartphone apps: five-year update on a systematic search and evaluation. JMIR Mhealth Uhealth 2020 May 27;8(5):e17300. [doi: 10.2196/17300] [Medline: 32459187]

Abbreviations

CVI: content validity index
GASING: Gerakan Anti Stunting
SEHATI: Sistem Evaluasi Kesehatan Anak Tumbuh Ideal

Edited by A Kushniruk; submitted 10.06.24; peer-reviewed by M Munikumar, MN Widyawati; revised version received 05.09.24; accepted 14.10.24; published 17.12.24.

Please cite as:
Erika KA, Fadilah N, Latif AI, Hasbiah N, Juliaty A, Achmad H, Bustamin A
Stunting Super App as an Effort Toward Stunting Management in Indonesia: Delphi and Pilot Study
JMIR Hum Factors 2024;11:e54862
URL: https://humanfactors.jmir.org/2024/1/e54862
doi:10.2196/54862

Integrating Health and Disability Data Into Academic Information Systems: Workflow Optimization Study

Abdulrahman Jabour, PhD

Department of Health Informatics, Faculty of Public Health and Tropical Medicine, Jazan University, Jazan, Saudi Arabia

Corresponding Author:
Abdulrahman Jabour, PhD
Department of Health Informatics, Faculty of Public Health and Tropical Medicine, Jazan University, , Jazan, , Saudi Arabia

Abstract

Background: Integrating health information into university information systems holds significant potential for enhancing student support and well-being. Despite the growing body of research highlighting issues faced by university students, including stress, depression, and disability, little has been done in the informatics field to incorporate health technologies at the institutional level.

Objective: This study aims to investigate the current state of health information integration within university systems and provide design recommendations to address existing gaps and opportunities.

Methods: We used a user-centered approach to conduct interviews and focus group sessions with stakeholders to gather comprehensive insights and requirements for the system. The methodology involved data collection, analysis, and the development of a suggested workflow.

Results: The findings of this study revealed the shortcomings in the current process of handling health and disability data within university information systems. In our results, we discuss some requirements identified for integrating health-related information into student information systems, such as privacy and confidentiality, timely communication, task automation, and disability resources. We propose a workflow that separates the process into 2 distinct components: a health and disability system and measures of quality of life and wellness. The proposed workflow highlights the vital role of academic advisors in facilitating support and enhancing coordination among stakeholders.

Conclusions: To streamline the workflow, it is vital to have effective coordination among stakeholders and redesign the university information system. However, implementing the new system will require significant capital and resources. We strongly emphasize the importance of increased standardization and regulation to support the information system requirements for health and disability. Through the adoption of standardized practices and regulations, we can ensure the smooth and effective implementation of the required support system.

(JMIR Hum Factors 2024;11:e54859) doi:10.2196/54859

KEYWORDS

disability; health data; student health; health measures; disability data; university setting; university system; student system; academic system; health informatics; health-related; health information; support; well-being; user-centered; data collection; analysis; development; privacy; confidentiality; timely communication; task automation; resources; quality of life; wellness; advisor; advisors; support system; interview; interviews; administrative staff; admin staff; physician; physicians; faculty; student; students; thematic analysis; focus group

Introduction

In recent years, there has been a growing body of research highlighting the profound impact of health and well-being on students' academic achievements [1]. Numerous studies have documented the significance of maintaining good physical and mental health in order to support students' overall well-being and optimize their educational outcomes [1-3]

One particularly alarming trend is the increasing prevalence of stress, depression, and anxiety among students. These mental health challenges have been shown to have detrimental effects on students' physical health and their academic performance [1]. Consequently, it has become essential to address these issues and implement effective strategies to support students' well-being throughout their educational journey [4].

Furthermore, students with disabilities and special needs form an integral part of university life. Inclusive education is not only a moral imperative but is also mandated by standards and regulatory requirements. However, students with disabilities often face unique challenges that can hinder their full participation in academic and social activities on campus. Navigating the complex landscape of campus resources, overcoming the stigma associated with invisible and physical disabilities, and combating feelings of isolation are just a few of the obstacles that these students encounter [5-7].

In light of these challenges, it is crucial to explore and address the role of health information technologies in supporting the health and well-being of university students. Despite recent advancements in health information technologies, there is a noticeable gap in research focusing on health information and university information systems. These systems play a central role in managing various aspects of university life, including academic support, administrative processes, and campus resources. However, the incorporation of health information into these systems remains relatively unexplored.

Recent studies highlight the vital role of information and communication technologies (ICTs) in supporting students with disabilities in higher education. Fichten et al [8] discuss how, despite generally adequate support on campuses, significant gaps exist in home and e-learning environments. Sharby and Roush [9] propose a decision-making model tailored to the needs of allied health students, emphasizing tailored accommodations and maintaining academic integrity. Stein et al [10] show how a knowledge-based system can effectively support disability accommodations under the Americans with Disabilities Act (ADA), enhancing institutional support and compliance. These insights reveal an urgent need for universities to improve ICT integration and support to foster inclusivity and aid students with disabilities. Furthermore, these gaps underline the necessity for better adaptive technology access off campus and increased faculty training in disability accommodation practices.

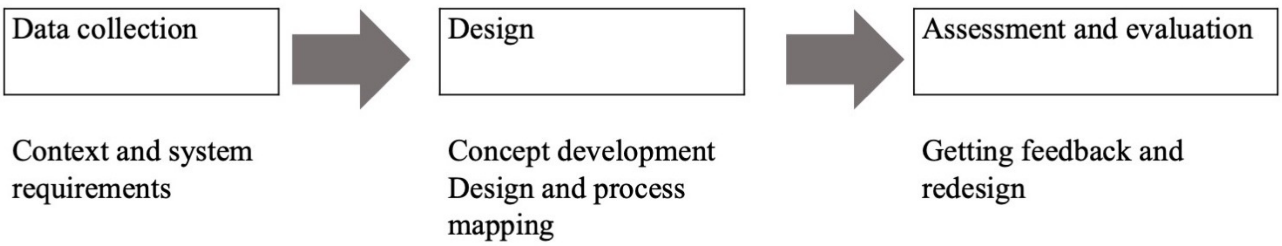
Understanding the potential benefits and challenges associated with integrating health information into university information systems is essential for enhancing the overall support and well-being of students. By leveraging the capabilities of modern information systems, universities can better address the unique health needs of their student population, promote inclusivity, and empower students to achieve their full academic potential. The analytical decision-making model proposed by Sharby and Roush [9] offers a comprehensive theoretical framework for integrating health information into university information systems. This model's systematic approach to accommodating diverse needs can be adapted to evaluate how health data are integrated into university information systems, ensuring these systems meet the varied requirements of students, faculty, and staff effectively. By applying this model, universities can better identify gaps and implement tailored solutions that enhance accessibility and functionality within their health information ecosystems.

Therefore, the aim of this study is to investigate the current state of health information integration within university information systems and provide design recommendations. By examining the existing system and stakeholders needs, this study seeks to shed light on gaps and opportunities for integrating health information technologies into university systems. The findings of this research will contribute to the growing body of knowledge in health information technologies and inform the development of strategies and interventions aimed at improving student well-being in higher education settings.

Methods

In this study, we used a user-centered approach to integrate health data into the university information system [11,12]. The methodology consisted of data collection, design, and assessment and evaluation [13] (Figure 1). The following subsections describe each phase in detail.

Figure 1. Summary of the study design steps describing each phase.



Data Collection

To gather comprehensive insights and users' requirements for the system, a series of interviews and focus group sessions were conducted with stakeholders. The interview phase involved engaging with a diverse group of participants, including 3 academic advisors, 3 administrative staff working on student affairs related to health and academic performance, 2 physicians/health counselors, 3 health professionals, and 4 teaching faculty members. The participants' experience ranged from 6 to 13 years. These stakeholders were chosen based on their expertise and previous involvement in activities and cases related to students' health information. In the methodological design of this study, we considered content saturation to be achieved within the 15-participant sample, as no additional unique insights were observed beyond this point. Consistent with findings from the usability testing literature, studies such as that by Faulkner [14] have shown that increasing the sample size from 5 to 10 participants can capture up to 95% of usability issues, thus providing a robust basis for our sample size decision and ensuring comprehensive data collection for system evaluation.

Additionally, 2 focus groups were conducted with 10 students. All students were undergraduates, and their age ranged from 19 to 24 years. The aim of these focus groups was to understand students' perspectives and needs.

Before the interviews, participants were given introductory information about the research project. Participation was voluntary, and they were allowed to skip questions, withdraw, or end the interview at any time without any conditions. No compensation was offered for participation. Prior to initiating the interview, participants were briefed on the study's scope and provided their informed consent. The interviews were unstructured, and probing questions were introduced as needed (Multimedia Appendix 1). The main areas covered during the interviews are described in the following sections.

Background and Context

We sought to understand the stakeholders' roles and responsibilities, as well as their relationship to the student health information system being developed.

Goals and Objectives

Our aim was to identify the stakeholders' primary goals, objectives, and desired outcomes related to the system. This included understanding what they hoped to achieve or improve through the use of the system.

System Requirements

We gathered requirements from stakeholders regarding the desired functionalities, features, and capabilities of the health information system. This included understanding the specific data elements, information flows, and interactions that are important to stakeholders.

Use Cases and Workflows

We examined the typical use cases and workflows of stakeholders within the context of the health information system. This involved understanding how stakeholders interact with the system, the sequence of steps they take, and any dependencies or constraints they encounter.

Data Privacy and Confidentiality

This involved discussing stakeholders' concerns and requirements related to data privacy, confidentiality, and compliance with relevant regulations. This included understanding their expectations for data protection, access control, and confidentiality.

Initial Phase and Workflow Analysis

During the initial phase of data collection, the focus was on gathering system requirements, as well as understanding the use cases, flow of information, restrictions, and other pertinent factors. The interviews and focus groups were guided by open-ended questions and prompts to encourage participants to provide detailed insights and suggestions.

The data collection process also included an analysis of the existing workflow and an artifact analysis. The artifacts included the existing forms, screenshots, templates, instructions, and guidelines used for data gathering and communication.

Design

Following the data collection phase, thematic analysis was conducted to identify patterns, common themes, and user requirements [15]. This analysis informed the design phase, where a conceptual design of the workflow was developed. The design process involved translating the gathered requirements into system functionalities and a workflow. In addition, an artifact analysis was conducted to examine existing forms, templates, and guidelines related to the workflow. This analysis involved systematically reviewing and coding these documents to identify recurring themes, inefficiencies, and opportunities for improvement. The artifacts were evaluated based on criteria such as clarity, efficiency, and alignment with the workflow.

To establish a holistic approach, the design phase was informed by the systems model of Bowman and Marzouk [16] for implementing the ADA in higher education. Their model, based on general systems theory, encompasses a comprehensive structure with input, throughput, and output subsystems. By adopting this systems model, the workflow design addressed core components such as accessibility, learning resources, and quality control.

To enhance the design methodology, insights were incorporated from previous studies to provide a broader understanding of user requirements and system functionalities. Fichten et al [8] contributed valuable information on variables related to how well the ICT needs of students with different disabilities are being met at institutions of higher education, at home, and in e-learning contexts. Their work also explored the disciplines and programs pursued by students with varying disabilities and the specialized ICTs used. Stein et al [10] provided a knowledge-based system design to assist university administrators in meeting the requirements of disability-related legislation. This ensured that the proposed system was comprehensive and aligned with compliance standards. Finally, Sharby and Roush [9] delivered an analytical decision-making model for addressing the specific needs of allied health students with disabilities, offering accommodations that are both practical and inclusive.

Assessment and Evaluation

Once the conceptual design was drafted, participants were presented with the proposed design and concept. Their feedback and suggestions were collected to ensure their perspectives were considered in the development process. An iterative and agile approach was adopted to incorporate the feedback received from participants.

Ethical Considerations

The study was approved by the ethical review committee at Jazan University (REC-45/11/1100). Participation for both staff and students was entirely voluntary, and no compensation was provided to participants. Participants were allowed to withdraw at any time without any conditions. Informed consent was obtained and all data collected were anonymized to ensure the privacy and confidentiality of participant information.

Results

Current System

Based on our analysis, we have identified several systemic challenges and shortcomings in the existing system, which are summarized in the following sections.

Variation and Inaccuracy of Student Information

We found that the information entered by students varies in nature and is often incomplete or inaccurate. This creates additional work for university staff and committees, who must gather more information and assess individual cases. Students tend to understate or overstate their conditions, further complicating the process.

Absence of University Health Care Center Representation

Currently, the university health care center is not involved in the process. When students require medical reports for absences or other requests, they upload the report from their accounts. This lack of integration creates inefficiencies and delays in the system.

Exclusion of Academic Advising Unit

The academic advising unit is not part of the process, which means that advisors cannot proactively identify or address students' needs unless a request is specifically referred to them by a staff member. This disconnect hinders effective support and guidance for students.

Neglect of Disability-Related Accommodations and Student Needs During Registration

The current system does not adequately consider disability-related accommodations and students' individual needs during the registration and enrollment process. There is no distinction made between different classrooms or labs, nor is there information on the availability of relevant resources. This oversight can lead to accessibility issues and hinder students' academic experience.

Design Considerations

Our results identified several requirements that need to be taken into consideration for integrating health-related information into the student information system. We categorized these needs into the following 4 categories: privacy and confidentiality, timely communication, workload and automation, and disability resources.

Privacy and Confidentiality

Our results revealed that students' health information requires a particular degree of privacy. Disseminating such information indiscriminately among all stakeholders encompassed within the system or involved in the process may not align with best practices. To overcome this challenge, we proposed a system that empowers the student advising unit. This system allows for the evaluation of received information, assessment of specific requirements, and addition of labels, tags, or notes detailing students' needs. This will enable the effective safeguarding of the privacy of students. Additionally, in scenarios where medical reports are needed by the university administration or external committees for request assessments or legal matters, exceptions can be formulated to ensure appropriate sharing.

Timely Communication

Our interviews with stakeholders highlighted recurrent instances where students had a low attendance rate due to health conditions that required regular hospital visits, such as sickle cell anemia or cancer. We found that these students were sometimes reluctant to address their needs openly. Many were only discovered during the consultation session with the student's advisor after being called for a low attendance rate. Providing detailed health information about students early on in their academic journey can help student advisors support them better.

Workload and Automation

In our study, we discovered that a significant portion of the tasks related to managing health information were being carried out manually. This manual approach posed several challenges, including inconsistencies and variations in practice, as well as an increased workload for staff members. To address these issues and streamline the process, the proposed workflow emphasizes the importance of automating certain activities. For example, by implementing automation, we can effectively match students' needs to appropriate classrooms; ensure the availability of necessary equipment and infrastructure; send timely notifications for attention; calculate students' health measures, such as stress, depression, and anxiety; and provide tailored advice and recommendations to support students in overcoming difficulties. By leveraging automation, we can alleviate the burden on staff members and enhance the overall efficiency and effectiveness of managing health information for a large number of students.

Disability Resources

While some university systems contain sections for disabilities, information regarding specific health conditions or academic needs are not always incorporated into academic information systems [17]. Moreover, the inclusion of disability information is not fully streamlined or integrated in the workflow, such as by automated flagging and registering of students with a disability that need specialized resources and accommodations in unequipped rooms or laboratories. These needs span from relatively minor adjustments such as left-handed tables to conditions that may need wheelchair accessibility or technological aids for visual and auditory impairments. The integration of health conditions within the academic information system is recommended to empower administrators and registrars in the seamless allocation of classrooms and laboratories suitably equipped to address these distinctive needs.

Data and Information

Based on the interviews, it is evident that the design of student information systems needs to consider various aspects of students' health. In this study, we categorized the types of information required into 3 categories, which we describe in the following sections.

Information Related to Disabilities and Special Needs

This category pertains to students with disabilities; examples include physical and mental disabilities, as well as vision and hearing impairments. It is crucial to consider the specific needs of these students to ensure an inclusive educational environment. Understanding their disabilities and special needs allows educational institutions to provide appropriate accommodations and support services, promoting equal opportunities for learning and participation.

Information Related to Health and Chronic Conditions

This category encompasses both preexisting chronic health conditions that may impact a student's educational journey and health-related events that might affect their attendance. Our results suggest that students with chronic health conditions face unique challenges in maintaining academic progress, managing their health, and navigating the educational system. Incorporating information about these conditions into students' information systems enables educational institutions to provide interventions, accommodations, and support tailored to individual needs.

Information About Continuous Health Measures

While the previous 2 categories focus on students with specific health issues, we propose that continuous health measures should be included for all students. These measures may involve assessing stress, anxiety, and depression levels. This can be very valuable for monitoring and addressing mental health concerns in educational settings to enhance overall well-being, academic performance, and retention rates. Collecting data on continuous health measures can facilitate early intervention and the implementation of appropriate support systems.

Workflow Description

To design the workflow, we separated the process into two different processes: (1) a health and disability system and (2) quality of life (QoL) and wellness measures (Table 1).

Table . Description of the data needed and frequency of collection.

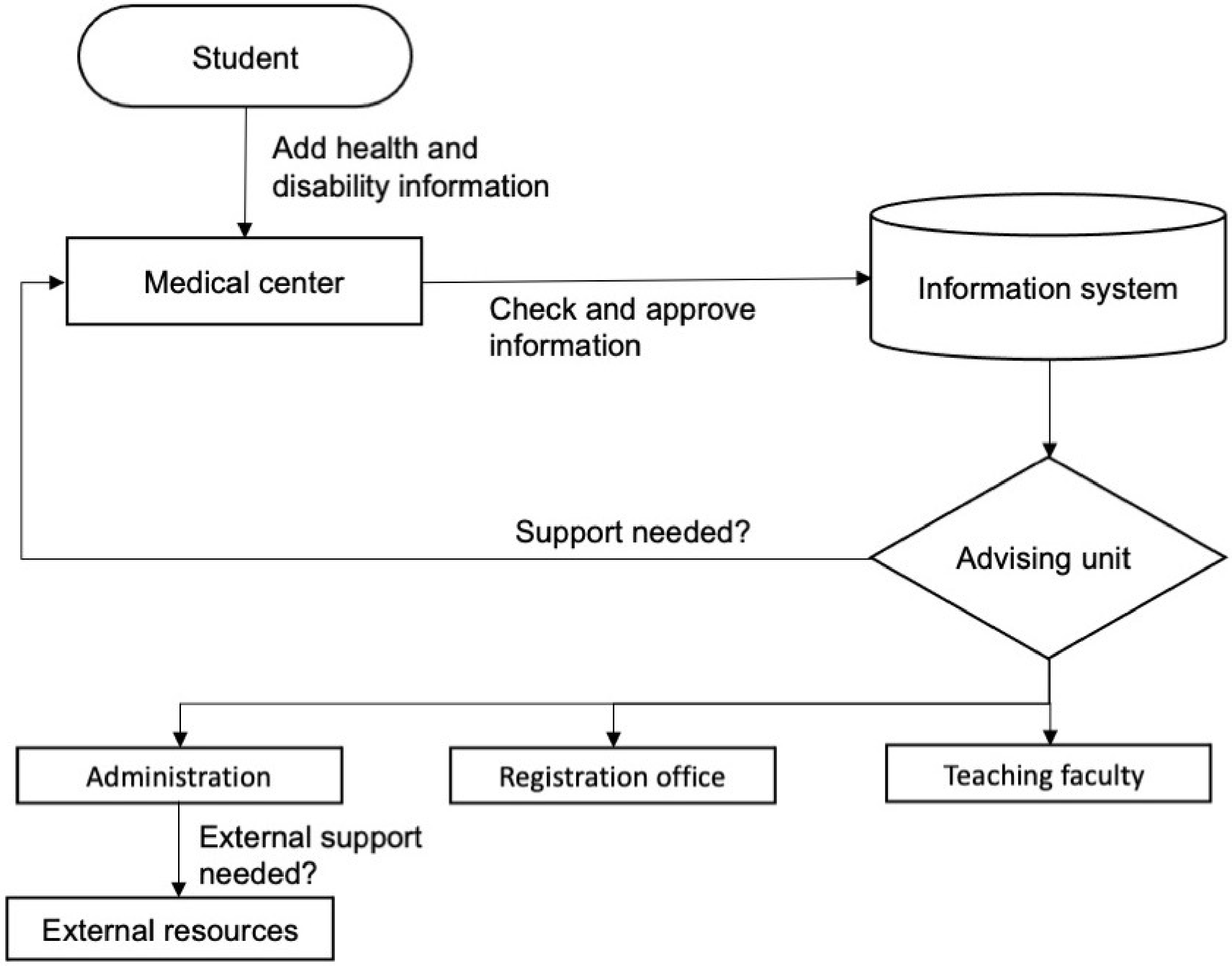
Process and data		Collection frequency
Health and disability system		
	Information related to disability and special needs	Entered once at the beginning of enrollment and updated as needed
	Information related to health and chronic conditions (preexisting chronic health conditions)	Entered once at the beginning of enrollment and updated as needed
	Information related to health and nonchronic conditions (health-related events)	Entered as needed
QoL[a] and wellness system		
	Information about continuous QoL and wellness measures including stress, depression, and anxiety	Systemic collection, once every semester

[a]QoL: quality of life.

Health and Disability System Workflow

Information related to disability and special needs are entered by students for health center review and approval (Figure 2). The university health center approves and enters the students' health-related information. Based on the type of disability selected, the system suggests a classroom or laboratory based on the availability of necessary accommodations that the student may need, including academic, technological, or administrative needs. The advising unit receives notification of registrations that have a mismatch between disabilities and resources for further attention. If needed, the student advising unit identifies the necessary accommodation that a student may need and sends messages to communicate with the registration office, administration, and teaching faculties.

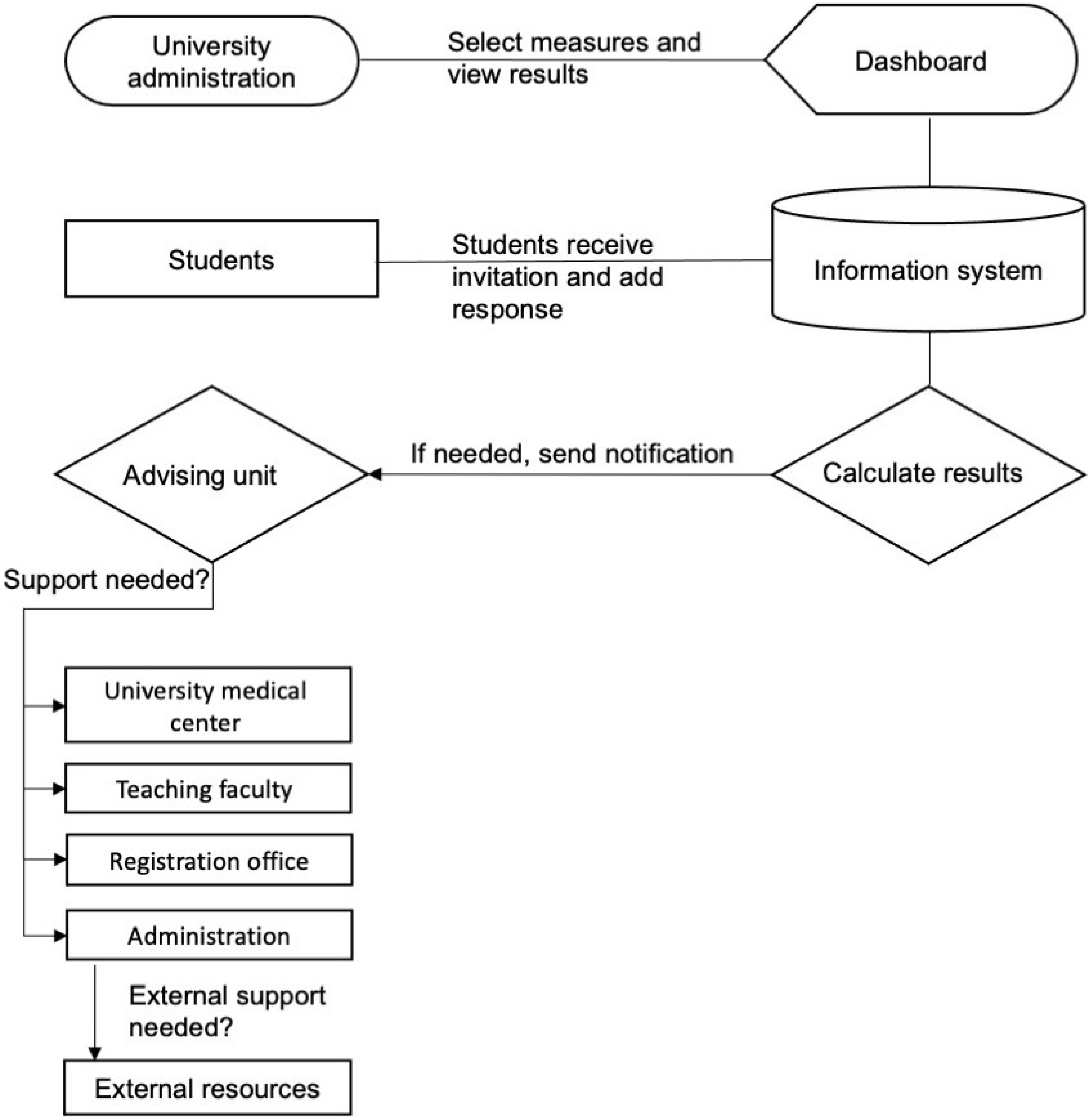
Figure 2. Health and disability system workflow.



QoL and Wellness System Workflow

The university administration selects the health measures they are interested in and specifies the frequency of measurement (Figure 3). Students receive the invitation and complete their responses. The university administration uses the dashboard to view the results, including summary information and a data overview. Responses with low scores that needs attention are identified by the system, triggering a notification that is sent to the advising unit for attention. The advising unit reviews the triggered student response and takes the necessary action.

Figure 3. Quality of life and wellness system workflow.



Discussion

Principal Findings

In this study, we investigated the current status of integrating health and disability information within university systems, identified the existing gaps, and provided design considerations. Using a user-centered design approach, we assessed the impact of these design considerations on students' success and inclusion. Our findings revealed several challenges. These challenges include privacy concerns, underidentification of stress, anxiety, and depression, and reluctance among students with hidden disabilities to disclose their conditions due to fear of stigma and discrimination. Our user-centered design approach addressed these issues by emphasizing the role of advising units in maintaining privacy and facilitating support. Additionally, the design focused on resource allocation to meet the diverse needs of students with various disabilities, ensuring accessibility and adaptive learning support. By incorporating health and disability information into university systems, the proposed design aims to enhance inclusivity, improve resource allocation, and promote better academic outcomes for all students. In the subsequent sections we discuss these findings and provide recommendations for future work.

Incorporating health and disability information is crucial for promoting students' academic success. However, effectively integrating this information into existing systems remains challenging. To address this challenge, there is a need to incorporate the use of health and disability information into the academic process and daily activities. The appropriate inclusion and use of health and disability information within university information systems offers a promising approach to integrate these data into daily tasks and processes, leading to enhanced effectiveness and improved outcomes. In our study, we adopted a user-centered approach to design the integration of health and disability information into university information systems. This approach prioritizes the needs and perspectives of users, ensuring that the system is tailored to their requirements and preferences. By considering the user experience and incorporating user feedback throughout the design process, we designed a system process that effectively incorporates health and disability information, ultimately enhancing its usability and impact on student success.

Our study uncovered many challenges in the existing system. Among these challenges was the privacy and confidentiality of students' health information. While obtaining knowledge about students' health events is essential for providing adequate support and accommodation, some students were reluctant to share their health information due to concerns about privacy breaches, stigma, or fear of discrimination.

Another significant challenge was the underidentification and underdisclosure of psychological conditions such as stress, anxiety, and depression among students. Many students perceive stress and anxiety as a normal part of the academic journey and may not seek professional help to differentiate between healthy stress and unhealthy stress. The systematic collection of stress, anxiety, and depression data via validated measures offers great potential for timely interventions and support. Furthermore, students with hidden disabilities often encounter barriers that hinder their university experiences [18]. Due to a desire to avoid the label of "disability," many students with invisible disabilities choose not to access the supports they are entitled to, which can hinder their academic success [18,19].

To address these challenges, the proposed design emphasizes the role of student advising units in facilitating support while maintaining students' privacy. Advising units can play a crucial role in creating a safe and supportive environment where students feel comfortable sharing their health information and accessing the necessary support services [20,21]. By establishing trust and providing guidance, advising units can help students navigate the system while respecting their privacy needs.

One of the key use cases found during interviews was the allocation of resources to meet the diverse requirements of students with disabilities. It was evident that different types of disabilities present unique needs and challenges. For instance, students with hearing or vision impairments may require special technological resources to facilitate their learning experience [22,23]. On the other hand, students with physical disabilities require accessibility measures, such as wheelchair access, that should be available in various locations across the campus.

Furthermore, the interviewees emphasized the importance of addressing the adaptive learning needs of students. This includes considering students with conditions such as cognitive disabilities and other health conditions that may impact their attendance [21]. While the specific needs may vary in each case, being aware of these needs is crucial for effective preparation and support.

By incorporating health and disability information into the university information system, administrators and faculty members can gain a comprehensive understanding of the diverse requirements of students. This knowledge enables them to allocate resources more effectively and provide appropriate support, ultimately promoting inclusive education and enhancing student success [20,21].

In order to ensure comprehensive inclusion and accessibility, it is crucial for accreditation bodies and policy frameworks to address the needs of individuals with disabilities. While physical accommodations such as wheelchair access, left-handed writing tables, and disability-friendly bathrooms are typically addressed in these standards, there appears to be a significant gap in incorporating disability-related information into IT standards and guidelines. Despite the growing reliance on IT in various domains, there is a lack of explicit guidance on how to handle disability-related information and promote its integration into IT systems. This oversight is concerning given the potential of IT to enhance accessibility and empower individuals with disabilities. By failing to address the inclusion of disability-related information, accreditation bodies and policy frameworks miss an opportunity to ensure equal access and participation for all.

While the current academic accreditation includes requirements related to students' disabilities in the physical environment, such as wheelchair access, it does not address requirements related to student information systems [24]. To address this gap, it is essential for accreditation bodies and policy guidelines to recognize the importance of incorporating disability-related information into IT standards. This could involve establishing guidelines for designing accessible user interfaces, ensuring compatibility with assistive technologies, and promoting the provision of alternative formats for individuals with sensory impairments. Additionally, the safe handling and protection of disability-related data should be emphasized to safeguard the privacy and confidentiality of individuals with disabilities.

By actively addressing the inclusion of disability-related information within IT standards, accreditation bodies and policy frameworks can contribute to a more inclusive and accessible society. This proactive approach will enhance the usability and effectiveness of IT systems, promoting equal opportunities and empowering individuals with disabilities to fully participate in various aspects of life.

Limitations and Future Work

The proposed design offers potential solutions to address the issue of late discovery of health-related cases through systematic data collection and communication with student health clinics. However, challenges persist due to students' reluctance to disclose their health information out of concerns related to privacy breaches, stigma, and fear of discrimination.

The potential side effects of health information disclosure are a crucial area that requires careful attention. The threat of privacy breaches, stigma, and discrimination needs to be carefully assessed, and risk management strategies should be established to ensure the protection of students' privacy and prevent any negative consequences.

Although the design process described here incorporated user input, our study was limited to a relatively small sample size. For future work, we recommend prototype usability testing and evaluation of the redesigned model with a larger and more diverse group of users. This would provide a broader range of perspectives, facilitating further evaluation and improvements to the design. In addition, interpretation of our findings should be approached with caution, as the homogeneity of the participant pool, all drawn from a single institution, may limit the generalizability of our results

Future studies should also explore strategies to address privacy concerns, enhance security measures, and develop clear communication strategies to alleviate students' fears regarding the disclosure of their health information. Additionally, assessing the impact of the proposed design on student outcomes, such as academic performance and well-being, would provide valuable insights into its effectiveness and usefulness.

Conclusion

In conclusion, our study has identified several shortcomings in the current process of handling health and disability data. Based on our findings, we recommend implementing a new, redesigned process that addresses these issues. Effective coordination among stakeholders is crucial for the success of this process, and redesigning the university information system is essential for organizing and streamlining the workflow.

While implementing these changes will require significant resources and commitment, we strongly advocate for increased standardization and regulation to support information system requirements related to health and disability. By adopting standardized practices and regulations, we can ensure that the necessary information system is in place to effectively support these needs.

Acknowledgments

The author would like to thank all participants who participated in this research. The author also extends their appreciation to the King Salman Center for Disability Research for funding this work through research group KSRG-2023-237. The funding agency had no influence on the design or outcome of this work.

Data Availability

The data sets supporting the conclusions of this study will be available upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview script.
[DOCX File, 17 KB - humanfactors_v11i1e54859_app1.docx]

References

1. Shankar NL, Park CL. Effects of stress on students' physical and mental health and academic success. Int J Sch Educ Psychol 2016 Jan 2;4(1):5-9. [doi: 10.1080/21683603.2016.1130532]
2. Eisenberg D, Golberstein E, Hunt JB. Mental health and academic success in college. BE J Econ Anal Policy 2009;9(1):40. [doi: 10.2202/1935-1682.2191]
3. Hysenbegasi A, Hass SL, Rowland CR. The impact of depression on the academic productivity of university students. J Ment Health Policy Econ 2005 Sep;8(3):145-151. [Medline: 16278502]
4. Ressa T. Dreaming college: transition experiences of undergraduate students with disabilities. Psychol Sch 2022 Jun;59(6):1175-1191. [doi: 10.1002/pits.22675]
5. Bessaha M, Reed R, Donlon AJ, Mathews W, Bell AC, Merolla D. Creating a more inclusive environment for students with disabilities: findings from participatory action research. Disabil Stud Q 2020;40(3). [doi: 10.18061/dsq.v40i3.7094]
6. Bradley A, Healey M, Fuller M. Incorporating disabled students within an inclusive higher education environment. Disabil Soc 2004 Sep;19(5):455-468. [doi: 10.1080/0968759042000235307]
7. Holloway S. The experience of higher education from the perspective of disabled students. Disabil Soc 2001 Jun;16(4):597-615. [doi: 10.1080/09687590120059568]
8. Fichten CS, Asuncion JV, Wolforth J, et al. Information and communication technology related needs of college and university students with disabilities. Res Learn Technol 2012;20(4). [doi: 10.3402/rlt.v20i0.18646]
9. Sharby N, Roush SE. Analytical decision-making model for addressing the needs of allied health students with disabilities. J Allied Health 2009;38(1):54-62. [Medline: 19361024]
10. Stein EW, Manco MP, Manco SA. A knowledge-based system to assist university administrators in meeting disability act requirements. Expert Syst Appl 2001 Aug;21(2):65-74. [doi: 10.1016/S0957-4174(01)00027-6]

11. Gulliksen J, Göransson B, Boivie I, Blomkvist S, Persson J, Cajander Å. Key principles for user-centred systems design. Behav Inform Technol 2003 Nov;22(6):397-409. [doi: 10.1080/01449290310001624329]
12. Abras C, Maloney-Krichmar D, Preece J. User-centered design. In: Bainbridge W, editor. Encyclopedia of Human-Computer Interaction: Sage Publications; 2004, Vol. 37:445-456.
13. Reijers H, Limanmansar S. Best practices in business process redesign: an overview and qualitative evaluation of successful redesign heuristics. Omega (Westport) 2005 Aug;33(4):283-306. [doi: 10.1016/j.omega.2004.04.012]
14. Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. Behav Res Methods Instr Comput 2003 Aug;35(3):379-383. [doi: 10.3758/BF03195514]
15. Braun V, Clarke V. Thematic analysis. In: Cooper H, Camic PM, Long DL, Panter AT, Rindskopf D, Sher KJ, editors. APA Handbook of Research Methods in Psychology, Vol 2: Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological: American Psychological Association; 2012:57-71. [doi: 10.1037/13620-004]
16. Bowman OJ, Marzouk DK. Implementing the Americans with Disabilities Act of 1990 in higher education. Am J Occup Ther 1992 Jun;46(6):521-533. [doi: 10.5014/ajot.46.6.521] [Medline: 1534969]
17. Istenic Starcic A, Bagon S. ICT‐supported learning for inclusion of people with special needs: review of seven educational technology journals, 1970–2011. Brit J Educ Tech 2014 Mar;45(2):202-230. [doi: 10.1111/bjet.12086]
18. Wolf LE. College students with ADHD and other hidden disabilities. Ann N Y Acad Sci 2001 Jun;931(1):385-395. [doi: 10.1111/j.1749-6632.2001.tb05792.x]
19. Lovett BJ, Nelson JM, Lindstrom W. Documenting hidden disabilities in higher education: analysis of recent guidance from the Association on Higher Education and Disability (AHEAD). J Disabil Policy Stud 2015;26(1):44-53. [doi: 10.1177/1044207314533383]
20. Garrison-Wade DF. Listening to their voices: factors that inhibit or enhance postsecondary outcomes for students with disabilities. Int J Spec Educ 2012;27(2):113-125 [FREE Full text]
21. Francis GL, Duke JM, Fujita M, Sutton JC. “It’s a constant fight:” experiences of college students with disabilities. J Postsecond Educ Disabil 2019;32(3):247-262 [FREE Full text]
22. Hasselbring TS, Glaser CH. Use of computer technology to help students with special needs. Fut Child 2000;10(2):102-122. [doi: 10.2307/1602691] [Medline: 11255702]
23. Mulloy AM, Gevarter C, Hopkins M, Sutherland KS, Ramdoss ST. Assistive technology for students with visual impairments and blindness. In: Lancioni G, Singh N, editors. Assistive Technologies for People with Diverse Abilities. Autism and Child Psychopathology Series: Springer; 2014:113-156. [doi: 10.1007/978-1-4899-8029-8_5]
24. National Center for Academic Accreditation and Evaluation. Standards for institutional accreditation. 2022. URL: https://www.etec.gov.sa/en/service/Institutional-Accreditation/servicegoal [accessed 2024-08-26]

Abbreviations

ADA: Americans with Disabilities Act
ICT: information and communication technology
QoL: quality of life

Edited by A Kushniruk; submitted 25.11.23; peer-reviewed by J Chang, K Ayers; revised version received 30.05.24; accepted 11.07.24; published 04.09.24.

Please cite as:
Jabour A
Integrating Health and Disability Data Into Academic Information Systems: Workflow Optimization Study
JMIR Hum Factors 2024;11:e54859
URL: https://humanfactors.jmir.org/2024/1/e54859
doi:10.2196/54859

Original Paper

Physical Therapists' Acceptance of a Wearable, Fabric-Based Sensor System (Motion Tape) for Use in Clinical Practice: Qualitative Focus Group Study

Audrey Lee[1], MS; Patricia Dionicio[2], MPH; Emilia Farcas[3], PhD; Job Godino[3], PhD; Kevin Patrick[3,4], MD, MS; Elijah Wyckoff[5], MS; Kenneth J Loh[5], PhD; Sara Gombatto[6], PT, PhD

[1]Department of Bioengineering, San Diego State University, San Diego, CA, United States

[2]Joint Doctoral Program in Public Health, San Diego State University and University of California San Diego, San Diego, CA, United States

[3]Qualcomm Institute, University of California San Diego, La Jolla, CA, United States

[4]School of Public Health, University of California San Diego, La Jolla, CA, United States

[5]Active, Responsive, Multifunctional, and Ordered-materials Research (ARMOR) Laboratory, Department of Structural Engineering, University of California San Diego, La Jolla, CA, United States

[6]School of Exercise & Nutritional Sciences, College of Health & Human Services, San Diego State University, San Diego, CA, United States

Corresponding Author:
Sara Gombatto, PT, PhD
School of Exercise & Nutritional Sciences
College of Health & Human Services
San Diego State University
5500 Campanile Drive
San Diego, CA, 92182-7251
United States
Phone: 1 619 594 3288
Email: sgombatto@sdsu.edu

Abstract

Background: Low back pain (LBP) is a costly global health condition that affects individuals of all ages and genders. Physical therapy (PT) is a commonly used and effective intervention for the management of LBP and incorporates movement assessment and therapeutic exercise. A newly developed wearable, fabric-based sensor system, Motion Tape, uses novel sensing and data modeling to measure lumbar spine movements unobtrusively and thus offers potential benefits when used in conjunction with PT. However, physical therapists' acceptance of Motion Tape remains unexplored.

Objective: The primary aim of this research study was to evaluate physical therapists' acceptance of Motion Tape to be used for the management of LBP. The secondary aim was to explore physical therapists' recommendations for future device development.

Methods: Licensed physical therapists from the American Physical Therapy Association Academy of Leadership Technology Special Interest Group participated in this study. Overall, 2 focus groups (FGs; N=8) were conducted, in which participants were presented with Motion Tape samples and examples of app data output on a poster. Informed by the Technology Acceptance Model, we conducted semistructured FGs and explored the wearability, usefulness, and ease of use of and suggestions for improvements in Motion Tape for PT management of LBP. FG data were transcribed and analyzed using rapid qualitative analysis.

Results: Regarding wearability, participants perceived that Motion Tape would be able to adhere for several days, with some variability owing to external factors. Feedback was positive for the low-profile and universal fit, but discomfort owing to wires and potential friction with clothing was of concern. Other concerns included difficulty with self-application and potential skin sensitivity. Regarding usefulness, participants expressed that Motion Tape would enhance the efficiency and specificity of assessments and treatment. Regarding ease of use, participants stated that the app would be easy, but data management and challenges with interpretation were of concern. Physical therapists provided several recommendations for future design improvements including having a wireless system or removable wires, customizable sizes for the tape, and output including range of motion data and summary graphs and adding app features that consider patient input and context.

Conclusions: Several themes related to Motion Tape's wearability, usefulness, and ease of use were identified. Overall, physical therapists expressed acceptance of Motion Tape's potential for assessing and monitoring low back posture and movement, both within and outside clinical settings. Participants expressed that Motion Tape would be a valuable tool for the personalized treatment of LBP but highlighted several future improvements needed for Motion Tape to be used in practice.

(JMIR Hum Factors 2024;11:e55246) doi:10.2196/55246

KEYWORDS

low back pain; physical therapy; physical therapist; wearable sensor; technology acceptance model; motion tape; kinesiology tape

Introduction

Prevalence and Impact of Low Back Pain

Low back pain (LBP) is one of the world's leading causes of disability [1-3]. In 2019, there were approximately 568.4 million prevalent cases, 223.5 million incident cases, and 63.7 million cases of years lived with disability owing to LBP reported globally [4]. LBP affects all ages and genders, but its prevalence increases with age, peaking at the age of approximately 45 to 54 years [4]. Approximately 70% to 85% of adults are expected to experience at least 1 episode of LBP in their lifetime [5]. Once predisposed to LBP, individuals are twice as likely to experience recurrent episodes of LBP [6]. Annually, LBP in the United States results in 149 million missed work days [7]. The total costs of LBP worldwide amount to approximately US $100 billion a year, with two-thirds of this amount owing to lost wages and decreased work productivity [8].

Treatment of LBP With Physical Therapy

Physical therapy (PT) is a common, effective, evidence-based treatment for LBP [9,10]. Specifically, active interventions including exercises prescribed by a physical therapist are effective for prevention and treatment of LBP [11,12]. During an initial examination, a physical therapist can identify musculoskeletal and neuromuscular impairments associated with the LBP problem by conducting assessment of the patient's posture and movement. Then, the physical therapist and patient can work together to promote strength, stability, and mobility with in-clinic sessions and an assigned home exercise program with the goal of decreasing pain and disability [10,13]. Monitoring the patient's posture and movement can provide a basis for determining individualized factors associated with the LBP problem, which can then be addressed through targeted interventions.

Incorporation of Technology in PT

Whether at home or at work, specific movement patterns that are performed repeatedly have been identified as a significant risk factor for the development and persistence of LBP [2,14,15]. These movement patterns of the low back region can be characterized by evaluating the angle, velocity, and acceleration [16] and can assist in LBP diagnosis, treatment, and prevention. There are several approaches to monitoring spine posture and movement. Generally, when conducting a PT examination, clinicians visually monitor posture and movement or use tools that measure the range of movement such as goniometers or inclinometers [17], but an alternative approach is to use technology to help better quantify the objective measures of spine posture and movement and offer potential benefits such as remote monitoring [16,18,19] while the patient is away from the clinic.

Technologies for Monitoring LBP

To date, existing technologies used to measure spine posture and movement in research and practice include optical motion capture, inertial measurement units (IMUs), and other wearable sensors [20-22]. Despite the variety of systems available, they generally present ≥1 limitation. Optical motion capture systems offer great precision and accuracy in monitoring human movement. However, their applications are limited owing to space needs, cost, and level of expertise needed. IMUs are portable devices that measure metrics such as acceleration and orientation [23] and include a variety of wearable sensors such as accelerometers, gyroscopes, and magnetometers, making them ideal for collecting data in a free-living environment. However, when used for monitoring human movement, IMUs have several limitations including decreased accuracy and precision for measuring slow movements [24,25], difficulty with measuring the axial plane movement accurately, inability to account for the multisegmented nature of the spine [26], and the need for multiple IMUs to triangulate posture and movement of a segment that can be cumbersome to the wearer [27].

Motion Tape

Owing to the limitations of existing sensor systems for measuring spine posture and movement, there is a need to explore new sensor innovations to address this issue. Ideally, such an approach would be wearable, unobtrusive, and usable in a clinical environment during PT sessions and in a person's natural environment to support home-based care. Another desired requirement would be high accuracy while collecting posture and movement data for a prolonged period.

Motion Tape, developed by Loh and Lin [28], is a disposable, self-adhesive skin-strain sensor system made using graphene nanosheets coated onto commercially available kinesiology tape (also known as K-Tape) [29-33]. Motion Tape has piezoresistive properties based on the deformation of the integrated graphene nanosheets in the tape that makes it sensitive to strain [33]. In previous studies, Motion Tape has demonstrated stable performance under cyclic strains [33,34]. In addition, the Motion Tape sensor system has been tested on human participants [33,34], displaying accuracy in measuring skin strains and angles across biceps, knees, shoulders, wrists, and various other body regions when compared with IMUs and skin strains estimated using optical motion capture systems [35]. Overall, Motion Tape offers noninvasive, comfortable, and practical skin-strain measurements and can comprehensively capture complex movements and muscle engagement, especially when applied as a network of sensors [35].

Motion Tape for a Low Back Use Case

When used for a low back use case, Motion Tape provides a means to capture the lumbar spine's multisegmental nature and multiplanar movements [36]. Motion Tape's low-profile and stretchable nature allows it to be worn throughout the day for all human shapes and sizes, and it could be suitable for use in an individual's natural environment with minimal interference to their daily activities. Motion Tape provides unique sensing streams that can be used in machine learning and artificial intelligence models to optimize inferences related to the management of LBP. Specifically, Motion Tape for a low back use case can address several key issues in a physical therapist's management of LBP, including the following: expanding on the level of detail available during the clinical assessment of posture and movement, assessing spinal posture and movement in a free-living environment, use for the promotion of engagement and adherence with and precise performance of a prescribed home exercise program, and using the patient's response to treatment to make informed decisions for future treatment or other patients [37]. Although there are several potential benefits that Motion Tape may add to personalized health care for LBP, the acceptability of Motion Tape among physical therapists has yet to be assessed.

Physical Therapists' Acceptance of Motion Tape

The success of this device is dependent on user acceptance or one's belief that the device will help them perform their work better (ie, perceived usefulness) and that the device's performance benefits outweigh the effort of using the device (ie, perceived ease of use) [38]. Thus, it is vital to understand physical therapists' perspectives about Motion Tape and their willingness to use it in their practice, to inform future developments and improvement of the technology.

Problem Statement

The primary aim of this research study was to evaluate physical therapists' acceptance of Motion Tape for the management of LBP. The secondary purpose was to explore physical therapists' current needs and recommendations regarding future development of Motion Tape.

Methods

Device Description and Stage of Development

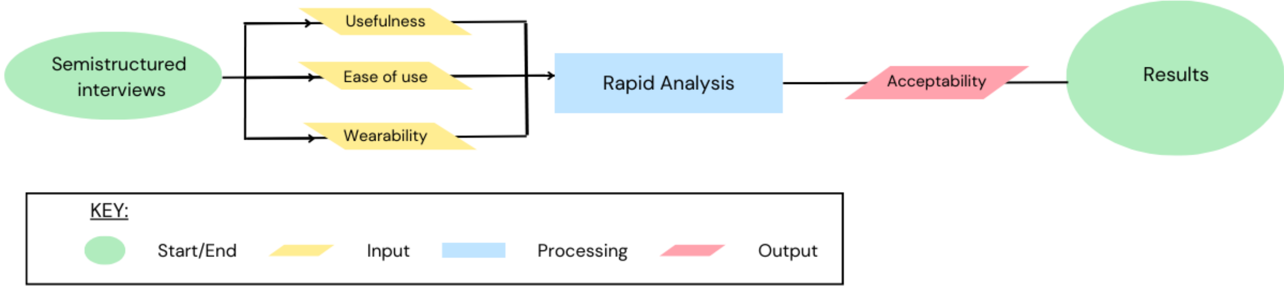
In this study, licensed physical therapists evaluated a prototype of Motion Tape and examples of data streams from the app for a low back use case. The Motion Tape samples evaluated in this study included the Motion Tape sensor system with conductive wire leads connected to both sides of the sample (Figure 1).

Figure 1. Motion Tape sample with conductive wire leads given to the physical therapists for evaluation.

Study Design

This exploratory, qualitative study was designed to explore physical therapists' acceptance of Motion Tape to provide a basis for future device development (Figure 2). The study was conducted from a constructivist point of view, with the goal of gaining insightful accounts and narrations of clinicians' lived experiences with technology and patients, rather than identifying an absolute truth [39]. We used semistructured focus groups (FGs) that incorporated human factor considerations to uncover real-world needs and obstacles and to ensure that the development of the sensor system can be informed by real-world PT clinical needs.

Figure 2. Study design overview—the evaluation of Motion Tape's acceptability.



Theoretical Framework and Constructs

The Technology Acceptance Model (TAM) framework was used in this study to assess two determinants of user acceptance of or willingness to use a technology: (1) perceived usefulness and (2) perceived ease of use [38,40]. An additional factor of wearability was also assessed to examine physical therapists' perceptions about patient-centered issues that would affect whether the device would be worn [41]. Recommendations for future improvements were also investigated to collect insight into data, device, and app developments that clinicians would like to see for Motion Tape.

> *Perceived usefulness was defined as the degree to which the use of Motion Tape would enhance the physical therapists' management of LBP [39-42], and this was assessed using the following constructs: (1) productivity, (2) effectiveness, (3) ability to make their job easy, and (4) benefits to PT treatment and recovery. Perceived ease of use was defined as the degree to which the use of Motion Tape would be effortless when used for managing LBP [39-42], and this was assessed using the following constructs: (1) how easy it would be for physical therapists to learn how to use it, (2) what level of instruction would physical therapists need to use it, and (3) how clear and understandable Motion Tape was in its current state. Wearability was defined as the degree to which Motion Tape would fit well and be comfortable for patients to wear on their back [42], and this was assessed based on (1) adhesion, (2) fit, (3) feel, and (4) how comfortable physical therapists would feel about applying and prescribing Motion Tape.*

Participants and Setting

This study was conducted at the American Physical Therapy Association's (APTA's) Combined Sections Meeting (San Diego, California) on February 24, 2023. Participants were recruited by sending study information via email to physical therapists who were members of the APTA Academy of Leadership Technology Special Interest Group. Members were also offered an opportunity to participate when they attended the Technology Special Interest Group in-person meeting at the APTA Combined Sections Meeting. Individuals were included in this study if they were a licensed physical therapist and were excluded from participating if they were unable to respond to questions in English. In total, 8 physical therapists were eligible and agreed to participate in 2 FGs of 4 clinicians each. A sample size of 8 people, in 2 FGs, was considered sufficient for this qualitative study to provide adequate variability and data saturation [43] and to provide a basis for device improvement. In addition, after data from the 2 FGs were collected and analyzed, the data were deemed saturated (ie, no new themes or codes were generated) and no further FGs were needed.

Ethical Considerations

The study protocol was considered to be exempt from ethics approval by the San Diego State University institutional review board. Each participant provided written consent before participating.

FG Methods

An FG guide (Multimedia Appendix 1) was used to lead the group's discussion. The FG guide was developed by investigators (AL, PD, and SG) to be semistructured with open-ended questions to explore the participants' perspectives about the usefulness, usability, and wearability of Motion Tape and to collect insight into future improvements for the sensors and data visualization (Textbox 1). A template of the FG guide was piloted with a Doctor of Physical Therapy student and a physical therapist at San Diego State University to ensure credibility [44]. General domains for each construct were prespecified to correspond with each interview question. Domains were defined based on the TAM framework and included perceived usefulness and perceived ease of use. An additional domain of wearability also was assessed.

Textbox 1. Guiding questions from the focus group guide.

Perceived wearability (W)

- How secure do you think the Motion Tape adhesive will be? (W-adhesion)
- To what degree do you think these sensors would fit your patients' anatomy (ie, their low back)? (W-fit)
- To what degree do you think your patients would feel the sensors on their back? (W-feel)
- How do you predict the Motion Tape Sensors would feel when being removed? (W-feel)

Perceived usefulness (U)

- To what degree would the usage of Motion Tape sensors affect how quickly you can assess your patient's posture, movement, or exercise performance? (U-efficiency)
- How effective do you think the Motion Tape sensors will be to capture valid data on your patients in the clinic? (U-effectiveness)
- How effective do you think the Motion Tape will be to capture valid data on your patients in their daily routine and normal environment? (U-effectiveness)
- To what degree would the usage of Motion Tape sensors affect the level of difficulty of your job as a clinician/physical therapist? (U–make job easier)
- What features, if any, would make the Motion Tape more useful to you? (U-useful)

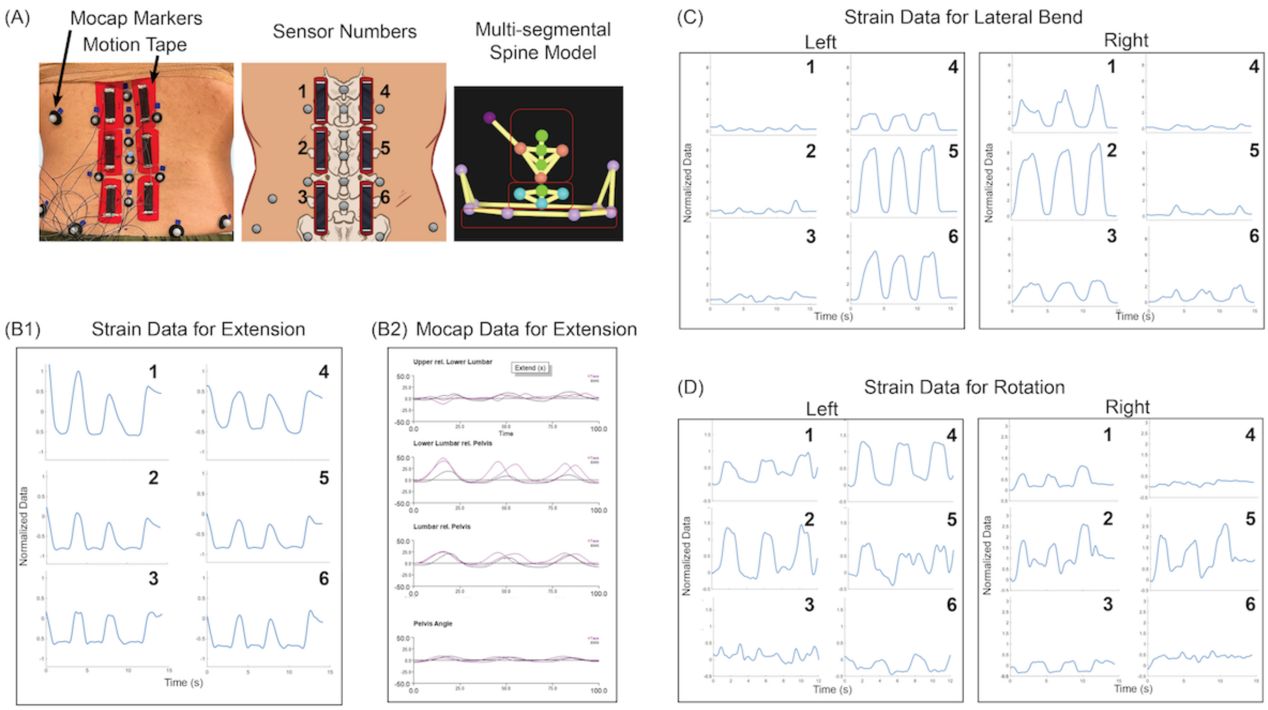
Perceived ease of use (EU)

- How easy do you think it would be to learn how to use Motion Tape? (EU-easy to learn)
- How comfortable would you feel prescribing Motion Tape to a patient to monitor their movements at home? (EU-comfort in usage)
- What level of knowledge do you think a clinician/PT would need to use the Motion Tape? (EU-clear and understandable)
- How easy/difficult do you think it would be for a clinician/PT to apply the Motion Tape to the patient's back? (EU-easy to use)
- What features, if any, would make the Motion Tape easier for you to use? (EU-easy to use)

FGs were conducted by AL (a female Master of Science student investigator) and PD (a female PhD student investigator). Reflexivity was maintained by the research team by discussing assumptions and biases that may influence how the clinicians responded to the FG moderators, who were not licensed physical therapists. As SG is a licensed physical therapist and member of APTA, she was able to provide valuable insight during the development of the interview guide, analysis, and interpretation to ensure credibility of the findings [44].

FGs were anonymized, and each participant was assigned a color as a name to ensure confidentiality. Each FG lasted approximately 1 hour and was recorded using digital voice recorders (Olympus Voice Recorder; WS-853). Before asking the participants questions, the investigators gave each participant a sample of Motion Tape. Participants were then oriented to a poster that displayed the Motion Tape placement and app data output streams (Figure 3).

Figure 3. Poster of Motion Tape placement and app data output for a low back use case. (A) The laboratory setup with 6 pieces of Motion Tape and several optical motion capture markers on anatomical landmarks of the lumbar spine. (B) The graphs display the following: (1) blue—the normalized strain data for extension, captured by the 6 Motion Tapes, and (2) purple—the kinematics for extension in degrees, captured by the optical motion capture system (reference standard). (C) The normalized strain data for right and left lateral bending obtained from the 6 Motion Tapes. (D) The normalized strain data for right and left rotation obtained from the 6 Motion Tapes.



Data Processing and Analysis

All FG audio data were downloaded to a HIPAA (Health Insurance Portability and Accountability Act)-compliant laboratory server, accessible only to the research staff, and removed from the digital voice recorder. The recordings were then transcribed, first using computer-based transcription (Word; Microsoft Corp). An investigator then checked and verified each transcription by listening to the original audio and reviewing and correcting the computer-based transcription.

Considering the need for timely feedback in the sensor development process, we adopted a rapid qualitative analysis (RQA) approach to explore themes regarding the acceptability and wearability of Motion Tape [45]. RQA was conducted by 3 investigators to assess the FG responses effectively and efficiently and to identify major themes. Codes and themes for RQA were deductively developed based on the TAM framework and the study objective [41]. We then used an inductive approach to generate RQA codes and themes, allowing for quick sorting of FG dialogue.

To ensure rigor and consistency of the method, a constant comparative approach with investigator triangulation was used at each stage [46]. First, the 3 investigators independently completed a summary report for each FG, with quotes and relevant topics under the respective themes and codes. Once the individual coding and summary reports for both FGs were completed, the investigators consolidated them into a combined rapid analysis summary report for each FG, unifying themes and reconciling discrepancies by consensus through discussion. The summary reports for each FG were then transferred into a matrix in which each row was a participant quote and each column was a domain. From this matrix, investigators identified the underlying themes and subthemes between the 2 FGs.

Results

Overview

In total, 8 physical therapists (n=5, 63% men and n=3, 38% women), with a mean age of 47.5 (SD 5.6) years participated in this study. Participants reported obtaining PT degrees ranging from a bachelor's degree to a Doctorate in Physical Therapy and had, on average, 20 (SD 8.5) years of clinical practice experience, and most reported practicing in an outpatient orthopedic setting. Of the 8 participants, 5 (63%) reported having advanced doctoral degrees (3/5, 60% PhD; 2/5, 40% EdD).

The qualitative results from the FGs were organized using the TAM for the acceptance of Motion Tape [38,40-42]. Data were organized based on the 3 main domains relevant to user acceptance (perceived wearability, perceived usefulness, and perceived ease of use) and 21 subthemes (Textbox 2). Subthemes were further designated using *positive*, *negative*, and *neutral* valences. Positive valence indicates that the FG participants perceived the Motion Tape attribute as positive. Negative valence indicates that the FG participants perceived the attribute as negative. Neutral valence indicates that the FG participants perceived the attribute as neither positive nor negative.

Textbox 2. Themes (n=3), subthemes (n=21), and valences of user acceptance of Motion Tape.

Theme 1: perceived wearability

- Positive
  - Motion Tape has a small, universal fit.
  - The feeling of Motion Tape on the skin would decrease over time.
- Negative
  - Patients may feel Motion Tape's wires snagging or sensors rubbing on clothes.
  - Motion Tape does not consider people with skin sensitivities.
- Neutral
  - Motion Tape adheres for 3-4 days but may adhere less owing to external factors.
  - The feeling of Motion Tape being removed depends on the physical therapist.

Theme 2: perceived usefulness

- Positive
  - Motion Tape could increase specificity of physical therapy management of low back pain (LBP).
  - Motion Tape could be effective for the diagnosis, management, and monitoring of low back pain (LBP).
  - The feeling of Motion Tape and the awareness of Motion Tape monitoring would increase adherence to a home exercise program.
  - Motion Tape would be beneficial in telerehabilitation and hybrid sessions.
  - Motion Tape could increase the physical therapist's awareness of the pain source.
- Negative
  - Motion Tape brings legal concerns with data responsibility.
  - Motion Tape's reliability could be affected by external factors.
- Neutral
  - Motion Tape could increase the efficiency of assessments, but set up could take more time.

Theme 3: perceived ease of use

- Positive
  - Motion Tape would be easy for a physical therapist to apply.
- Negative
  - Motion Tape has a lot of data to sift through.
  - Motion Tape data are hard to interpret in their current state.
  - The self-application of Motion Tape would be difficult.
  - Motion Tape is designed for single use.
- Neutral
  - The prescription of Motion Tape is subjective to many factors.
  - The user interface would dictate how much knowledge would be needed to use Motion Tape.

Domain 1: Perceived Wearability

Regarding perceived *wearability*, all physical therapists were familiar with commercially available kinesiology tape. Thus, their thoughts about perceived wearability reflected their experience with kinesiology tape. For example, the physical therapists expected Motion Tape to last about 3 to 4 days. A physical therapist mentioned the following:

> *Oh, I've used the K-Tape for four days before it started peeling off. Sometimes it lasts more than five days actually. Three to four days I think is average.* [FG1]

However, some physical therapists clarified that the *longevity* of Motion Tape's adhesion depends on several factors. For example, 2 of the physical therapists expressed the following:

> *How secure it is depends on a lot of factors, like moisture on the skin. It depends on not just moisture, but how clean your skin is and how much hair is on the skin.* [FG2]

> *Some of them, specifically on the low back, tend to have more oily skin, and that depreciates the life of the tape.* [FG1]

Regarding the *fit* aspect of wearability, physical therapists also believed that Motion Tape's size was sufficiently small to be universal to the wearer and the placement location. They expressed the following:

> *In my experience with tapes like this, it fits most of the clientele that I've worked with, both inpatient and long-term post-acute.* [FG1]

> *If it was that little strip, I think it would be great to use anywhere.* [FG2]

Regarding the *feel* aspect of wearability, generally, physical therapists felt that patients would feel Motion Tape at first when applied but would become less aware over time until the tape starts to peel off:

> *They'd know that they're there, and they'd probably become less aware of over time.* [FG1]

However, physical therapists generally felt that with Motion Tape's current design, patients would *feel* the wires snagging or the sensors rubbing on clothing. A physical therapist explained it as follows:

> *So contraptions with wires will always have that uncomfortable feeling. Always. But if you go the wireless route, then probably after two days, the patient will be more comfortable until the tape starts peeling off. However, what I'm wearing right now, something that goes above my PSIS, if I go to the bathroom or do something, I'm going to, it's gonna move around, it might get pulled on it by my clothes.* [FG1]

When removing Motion Tape, physical therapists said that patient feelings about the removal process would be quite variable. Some physical therapists felt that it was subjective to how the therapist removed Motion Tape and how much hair or oil the individual has on their skin. A physical therapist explained it as follows:

> *I'm just thinking of whoever is taking it off. You know, like, it depends on you, like, some people just rip. And some people are just gentle. So subjective. So it depends on the training of the therapist and concern if they're empathetic to our patients.* [FG1]

The physical therapists mentioned some wearability concerns during the FGs. A concern was about how patients with skin sensitivities would be able to use Motion Tape. A physical therapist asked the following:

> *For those with skin allergy. Can you put an under wrap under this?* [FG1]

Domain 2: Perceived Usefulness

Physical therapists expressed mixed feelings about whether Motion Tape would increase their efficiency with assessments of lumbar spine posture and movement. Some expressed that if all they had to do was apply the tape, then there would be increased efficiency:

> *If it's easy to objectively document, by understanding the graph, I think it's a night and day difference versus getting into the goniometers and doing manual assessment. Instead, you put on the tape, ask the patient to rotate their trunk, lean forward, reach forward, extend their back. And then if I have it digitally by email or direct messaging, it would save a bunch of time.* [FG1]

However, others felt that it would reduce efficiency. A physical therapist explained the following:

> *Regarding the speed of assessment, I would be a little doubtful. I think by the time that you took this and you put it on the patient, you hooked up all the wires to it, you did the calibration, if you need to do a calibration, it might take just as long as doing an assessment. I would have concerns around the accuracy of this, to give you a number, an accurate range of motion, particularly for things like rotation. But if the data was convincing that everyone, if it was validated for everyone that gave you an accurate number, I think it could improve the quality of assessment.* [FG1]

A physical therapist felt that for the in-clinic assessments, Motion Tape would improve specificity:

> *I don't feel like it [Motion Tape] would improve speed, it would improve specificity.* [FG2]

Physical therapists also mentioned that they could envision Motion Tape as a useful tool for self-management and remote monitoring when used in combination with in-clinic PT. A physical therapist mentioned that the ability to monitor patients outside the clinic would be very meaningful:

> *That's the best place to actually observe them, their normal environment. If they're in therapy, they're being observed, coached, cued by a skilled clinician. Their performance is definitely going to be different. So if they're at home, and we're able to monitor them at home, I think the treatment will be more, and your adjustment and progression will be more meaningful.* [FG1]

Some physical therapists suggested that having patients wear these sensors would increase their awareness of being monitored and thus increase engagement with and adherence to the home exercise program:

> *I think that what it has to offer is improving...adherence with our programs. I think that's your potential.* [FG1]

> *When you tell someone, I'm looking at your posture right now, you change [gesturing to posture]. If they think you are watching, they'll do better.* [FG1]

Physical therapists expected patients to have a phenomenal experience with Motion Tape when used in a hybrid setting:

> *I think to his point that if it's applied properly in the clinics, it's hybridized, and you can take a call, and there's no technical involvement on the patient side, and all they do is open up the app, they'd have this really phenomenal experience.* [FG2]

Specifically, several physical therapists expressed that Motion Tape would help with the identification of postures and movements in free-living environments that provoke pain, allowing for more meaningful interventions:

> *I think for it to be very useful. It would have to compare with the app where you've got user input as to what's going on...where he's got these flags and the data that was pain here, pain here, pain here, and you can look, you know, to the periods of time before that.* [FG2]

Some physical therapists did have some concerns about the usefulness of Motion Tape. A physical therapist expressed legal concerns regarding data responsibility:

> *As long as you collect data, someone's then responsible for it. So who's going to look at it? What's the liability then that person takes on by having that information?...if something goes wrong, and the therapist hasn't looked at the data, I'd like to know, are they liable?* [FG1]

Another concern was knowing what external factors affect Motion Tape's signal and data reliability, mentioning that the use of Motion Tape in practice was "gonna depend on the reliability of the data" (FG2).

Several physical therapists felt that there were a variety of variables that might affect the reliability of the signal or data. They expressed the following:

> *And what other factors affect them, the sensors, as far as humidity, water, other environmental factors that might affect it? You know, what if they have a compression garment around the trunk, for example, does that affect the sensors?* [FG2]

> *Whether, getting it wet and getting so some things on it changes the conductivity, and therefore the calibration over time.* [FG2]

> *You get variability in the readings based on amount of tension that people put on it when they applied it.* [FG2]

> *Whether that's different from person to person because of different makeups morphology.* [FG1]

Domain 3: Perceived Ease of Use

Physical therapists felt that it would be easy to apply Motion Tape, given their background knowledge in human anatomy. A physical therapist stated the following:

> *You would need to know basic clinical knowledge of the application for where to look for the muscles, you know, right. So, they need to be clinician to have knowledge of the body.* [FG2]

When asked whether they would feel comfortable using Motion Tape with their patients, there were mixed responses among physical therapists. Some mentioned that it would depend on "cost and buy-in" (FG2) or how it was going to be "incorporated into the plan of care" (FG1). A physical therapist even explained the variability as follows:

> *Depends on the situation, honestly. I mean, I have some families that I'll show them how to do the application. And I'll see them three weeks later, and they've reapplied four times and done it great. And then I've seen others that I'm like, "Oh, no! This is nope."* [FG2]

There were also several concerns about the ease of use. Some physical therapists felt that they would have challenges with ease of use, specifically regarding interpretation of the data:

> *I think in its current form, easy to apply. Hard to interpret.* [FG1]

> *It depends on the interface and how much it interprets the points. The tape will be easy, but it's all the other pieces.* [FG2]

Additional concerns about the ease of use included that the amount of data presented was excessive and the type of data displayed was difficult to interpret. The physical therapists expressed a desire to see the range of motion displayed in degrees rather than resistance in ohms:

> *I think I'm probably realistically just correlated with what they report has been painful. Because I don't know that I've ever been so interested in all of that. Like, it might be too much data. For a patient, like I don't necessarily need to know their range of motion during every single activity, I need to know when it is relevant to them. And when it is impacting whatever condition they're here for.* [FG1]

> *And again, I think for a clinician, it's going to have to be meaningful data. It's gonna have be Range of motion data not ohms.* [FG1]

> *So then, conceivably, would it be helpful instead of giving you normalized strain,...if they could interpret it, would convert this over to degrees of rotation and flexibility?* [FG2]

> *If you could get range of motion kind of information, I think that would be great.* [FG2]

Another concern was about how challenging the self-application would be for patients:

> *How are people actually going to apply this on their own, someone that doesn't know how?* [FG2]

Finally, another concern was that Motion Tape is a single-use product. A physical therapist explained the challenge of a single-use product as follows:

> *Okay, now how about waste? So it's like a single use thing? Now I'm gonna throw in a whole planet into this is single usage. Or can you reapply?* [FG2]

Future Recommendations

Future recommendations from the physical therapists were organized into 3 categories (Textbox 3): data, physical features, and app features.

Textbox 3. Themes (n=3) and subthemes of future recommendations for Motion Tape.

Theme 1: Data recommendations

- Motion Tape data should be easy to read at a glance.
- Motion tape data should account for differing patient morphology.
- Physical therapists should be aware of factors that affect Motion Tape data.

Theme 2: Physical feature recommendations

- Motion Tape should be made wireless or with removable wires.
- Motion Tape should be reusable.
- Motion Tape should be customizable in length.

Theme 3: App feature recommendations

- Motion Tape app should include BMI input.
- Motion Tape app should include input for a patient's change in activity.
- Motion Tape app should allow flagging events.
- Motion Tape app should include comparative data.

Regarding the data *recommendations*, physical therapists expressed that data should be summarized in the form of an at-a-glance graph with 1 overall meaningful number, reflecting the range of motion. They would also like to know how the data change from person to person owing to morphology and how external factors (water, application stretch variability, and skin movement) affect the data. Additional data that would be useful for their job included comparative data, graphs with a color scale, and information about muscle activation. Participants in an FG expressed the following:

> *Take a baseline and have them rotate from that position and determined by the volume strain, whether they are tension either degrees, or even if it's yellow, green, yellow, red, like if they're moving within if they can't pinpoint it specifically, but you know, within a range, would that be helpful?* [FG2]

> *I think even just having comparative data would be helpful, right? Because, you know, I keep telling my students, "Don't tell me, 'I want to increase range 10 degrees.'" Because that doesn't tell me, "Can they walk?" Right? But, "Are they doing better now than they were doing when we started?" That's useful. So even if we get baseline data that could be translated into amount of motion and then follow up data that says, "hey, it's more, it's more fluid, it's better, it's whatever." I think that can be really useful. Now I know the payers are gonna want, how much rotation did you get? How much lateral flexion did you get?* [FG2]

> *And I think beyond the range of motion, I work in neuro. I think just like muscle activation would be interesting, you know, like, how much activation did you get today, for example, versus six weeks ago, post stroke or, you know, spinal cord or something? I think that would be really interesting just to see the firing muscle activation. And on the flip side, and I don't know if that's possible, but looking at specificity. Could that be something to monitor changes in specimen specificity? Post- X Y & Z intervention, right? That could also be interesting. So it's not really about range of motion, we're also activity known as firing or not?* [FG2]

Regarding *physical feature* recommendations, physical therapists wanted a way to mitigate the wires, either by moving to a wireless system or making them removable. Physical therapists were also concerned about the limited stretchability of the short pieces of tape, as it would not be long enough in length for typical kinesiology tape use, and recommended making the length customizable to the physical therapists' needs. Physical therapists were also concerned about Motion Tape's single-use design and were curious about whether it could be reusable to reduce waste:

> *Again, I'm thinking like, in the future, no wires, you've got a strip of graphene that you could customize length to, with those couple millimeters around the edge. And if we wanted a whole length, we cut whole lengths. And if we want segments, we can cut segments. And it feeds the data to the app somehow tailor it to someone's body.* [FG2]

> *So you can imagine that maybe something like this could be a roll of tape. Yeah, the width of duct tape. And there's actually two pieces on this roll. There's one section, that's the conductive piece, that you can cut it to length, and then next to it there are maybe*

> *there's a wire section, that's conductive tape that you can pull off and put on the ends of whatever you choose. So you get one roll of tape. And then one of them is the is the graphene is the other piece that you tear off to the appropriate length is the conductive tape that connects it to the box. And then it's a solution, you can customize length and you have your conductive piece and then your graphene.* [FG2]

Regarding future *app feature* recommendations, physical therapists expressed a need for the capability to input factors such as BMI, activity changes, "flags" for events, and changes in pain to help label, compare, or contextualize the data.

Discussion

Overview

There is a gap in the research between rehabilitation device development and evaluation of clinicians' acceptance of such devices. Most existing studies have considered patient or user satisfaction [47,48], whereas others that consider the clinician's perspective have not specifically evaluated sensors for measuring spine posture and movement [49,50]. In this study, several themes relating to physical therapists' perspectives about Motion Tape's wearability, usefulness, and ease of use for a low back use case were identified.

Domain 1: Perceived Wearability

One of the most common challenges for wearable sensors is ensuring that they are unobtrusive to the wearer's natural movement and environment [39]. The small form and fit of Motion Tape was considered by physical therapists to be ideal for a wearable sensor. However, similar to previous studies, the wires in the current design were considered to be not ideal [37]. Studies have shown that wireless technologies tend to be more widely used in many fields, especially in the field of wearable devices for health care [51]. Thus, a future iteration of Motion Tape without wires would be considered optimal. On the basis of feedback obtained from physical therapists, wearability for people with skin sensitivities also should be considered. Previous studies have shown that skin irritation is the most common concern when using kinesiology tape for extended periods of time [52,53]. Thus, future studies should explore whether a medium or substrate can be used under Motion Tape to mitigate skin irritation, possibly as an extension of recent research that integrated Motion Tape with elastic fabric for respiration monitoring [54].

Domain 2: Perceived Usefulness

There were mixed feelings among physical therapist participants about how efficient Motion Tape would be in the clinic. Overall, most physical therapists felt that Motion Tape would increase the specificity of their assessments, a characteristic that has been shown to be beneficial for LBP diagnosis and treatment [55]. Furthermore, Motion Tape's ability to monitor the patient's movements remotely was considered beneficial, as this feature may increase adherence to home exercise programs, which is an important component of effective treatment for LBP [56,57].

Domain 3: Perceived Ease of Use

On the basis of physical therapists' perspectives, Motion Tape would be easy to apply, but data would be difficult to interpret. Creating a device that is easy to use and understand is crucial because it predicts consumer use behavior [38,41]. Recommendations included presenting the data in units that physical therapists are more familiar with (ie, degrees of range of motion) and creating an app that requires minimal time for the physical therapists to use. These changes may promote increased device use and acceptance in PT.

Future Recommendations

On the basis of clinician feedback, Motion Tape appears to be a promising new technology that could be used for monitoring lumbar spine posture and movement in the management of patients with LBP. Future device development will be needed to address clinician recommendations obtained from this study in the domains of wearability and ease of use. In addition, future studies will be needed to validate Motion Tape in laboratory, clinical, and free-living environments and to investigate patient acceptance of Motion Tape.

Limitations

A limitation of this study is that participants were physical therapists who were part of a Technology Special Interest Group and are likely to be more receptive to using technology in practice. Thus, this study's results regarding Motion Tape's acceptability may be biased in favor of Motion Tape's ease of use, usefulness, and wearability. Future studies should also assess the acceptability of Motion Tape for clinicians who do not regularly use technology in their practice. Another limitation is that the physical therapists were not presented with active samples of Motion Tape with live data streams in the app. Instead, participants were given inactive samples of Motion Tape and presented with a poster with examples of app data streams. Future studies should provide an opportunity for physical therapists to apply Motion Tape to a person and use it with the app interface. Finally, there was a potential for investigator bias in the interpretation of the results, as several investigators of this study are actively working on the development of this device. However, 2 of the 3 investigators who conducted data analysis were outside the primary research team.

Conclusions

Physical therapists expressed overall acceptance of Motion Tape for its potential to monitor and assess low back posture and movement, both within and outside clinical settings. Physical therapist participants expressed that Motion Tape would be a valuable tool for personalized treatment of LBP but highlighted several future improvements needed for Motion Tape to be used in practice.

Acknowledgments

The authors would like to thank Nicolette Jaghab for her assistance with qualitative data analysis. This study is part of the Multi-Sensor Adaptive Data Analytics for Physical Therapy project and was supported by US National Science Foundation (award IIS-2205093).

Conflicts of Interest

KJL is a cofounder of JAK Labs Inc, a company that may potentially benefit from the study results. JAK Labs intends to commercialize Motion Tape for the physical therapy and rehabilitation market, among other markets. The terms of this arrangement have been reviewed and approved by the University of California San Diego in accordance with its conflicts of interest policies.

Multimedia Appendix 1

Focus group guide.

[DOCX File , 21 KB - humanfactors_v11i1e55246_app1.docx]

References

1. Hartvigsen J, Hancock MJ, Kongsted A, Louw Q, Ferreira ML, Genevay S, et al. What low back pain is and why we need to pay attention. Lancet 2018 Jul 09;391(10137):2356-2367 [FREE Full text] [doi: 10.1016/S0140-6736(18)30480-X] [Medline: 29573870]
2. Silva FG, da Silva TM, Maher C, L Ferreira M, Costa LD, Costa LO. What triggers an episode of acute low back pain? A protocol of a replication case-crossover study. BMJ Open 2021 Mar 05;11(2):e040784 [FREE Full text] [doi: 10.1136/bmjopen-2020-040784] [Medline: 33550235]
3. Wu A, March L, Zheng X, Huang J, Wang X, Zhao J, et al. Global low back pain prevalence and years lived with disability from 1990 to 2017: estimates from the Global Burden of Disease Study 2017. Ann Transl Med 2020 Mar;8(6):299 [FREE Full text] [doi: 10.21037/atm.2020.02.175] [Medline: 32355743]
4. Chen S, Chen M, Wu X, Lin S, Tao C, Cao H, et al. Global, regional and national burden of low back pain 1990-2019: a systematic analysis of the Global Burden of Disease study 2019. J Orthop Translat 2022 Jan;32:49-58 [FREE Full text] [doi: 10.1016/j.jot.2021.07.005] [Medline: 34934626]
5. Henchoz Y, Kai-Lik So A. Exercise and nonspecific low back pain: a literature review. Joint Bone Spine 2008 Oct;75(5):533-539. [doi: 10.1016/j.jbspin.2008.03.003] [Medline: 18801686]
6. Hestbaek L, Leboeuf-Yde C, Manniche C. Low back pain: what is the long-term course? A review of studies of general patient populations. Eur Spine J 2003 May;12(2):149-165 [FREE Full text] [doi: 10.1007/s00586-002-0508-5] [Medline: 12709853]
7. Kahere M, Hlongwa M, Ginindza TG. A scoping review on the epidemiology of chronic low back pain among adults in sub-Saharan Africa. Int J Environ Res Public Health 2022 Mar 03;19(5):2964 [FREE Full text] [doi: 10.3390/ijerph19052964] [Medline: 35270657]
8. Katz JN. Lumbar disc disorders and low-back pain: socioeconomic factors and consequences. J Bone Joint Surg Am 2006 May;88 Suppl 2:21-24. [doi: 10.2106/JBJS.E.01273] [Medline: 16595438]
9. Anar S. The effectiveness of home-based exercise programs for low back pain patients. J Phys Ther Sci 2016 Oct;28(10):2727-2730 [FREE Full text] [doi: 10.1589/jpts.28.2727] [Medline: 27821923]
10. Marich AV, Lanier VM, Salsich GB, Lang CE, Van Dillen LR. Immediate effects of a single session of motor skill training on the lumbar movement pattern during a functional activity in people with low back pain: a repeated-measures study. Phys Ther 2018 Jul 01;98(7):605-615 [FREE Full text] [doi: 10.1093/ptj/pzy044] [Medline: 29660077]
11. Saragiotto BT, Maher CG, Yamato TP, Costa LO, Menezes Costa LC, Ostelo RW, et al. Motor control exercise for chronic non-specific low-back pain. Cochrane Database Syst Rev 2016 Jan 08;2016(1):CD012004 [FREE Full text] [doi: 10.1002/14651858.CD012004] [Medline: 26742533]
12. Steffens D, Maher CG, Pereira LS, Stevens ML, Oliveira VC, Chapple M, et al. Prevention of low back pain: a systematic review and meta-analysis. JAMA Intern Med 2016 Mar;176(2):199-208. [doi: 10.1001/jamainternmed.2015.7431] [Medline: 26752509]
13. Nelson BW, O'Reilly E, Miller M, Hogan M, Wegner JA, Kelly C. The clinical effects of intensive, specific exercise on chronic low back pain: a controlled study of 895 consecutive patients with 1-year follow up. Orthopedics 1995 Oct;18(10):971-981. [doi: 10.3928/0147-7447-19951001-05] [Medline: 8584467]
14. Steffens D, Ferreira ML, Latimer J, Ferreira PH, Koes BW, Blyth F, et al. What triggers an episode of acute low back pain? A case-crossover study. Arthritis Care Res (Hoboken) 2015 Mar;67(3):403-410 [FREE Full text] [doi: 10.1002/acr.22533] [Medline: 25665074]
15. Karayannis NV, Jull GA, Hodges PW. Physiotherapy movement based classification approaches to low back pain: comparison of subgroups through review and developer/expert survey. BMC Musculoskelet Disord 2012 Mar 20;13:24 [FREE Full text] [doi: 10.1186/1471-2474-13-24] [Medline: 22348236]

16. Laird RA, Gilbert J, Kent P, Keating JL. Comparing lumbo-pelvic kinematics in people with and without back pain: a systematic review and meta-analysis. BMC Musculoskelet Disord 2014 Jul 10;15(1):229 [FREE Full text] [doi: 10.1186/1471-2474-15-229] [Medline: 25012528]
17. Zheng Z, Wang Y, Wang T, Wu Y, Li Y. A systematic review and meta-analysis on comparative kinematics in the lumbopelvic region in the patients suffering from spinal pain. J Healthc Eng 2022;2022:7369242 [FREE Full text] [doi: 10.1155/2022/7369242] [Medline: 35345658]
18. Alschuler KN, Hoodin F, Murphy SL, Geisser ME. Ambulatory monitoring as a measure of disability in chronic low back pain populations. Clin J Pain 2011 Oct;27(8):707-715. [doi: 10.1097/AJP.0b013e318217b7d0] [Medline: 21505322]
19. Schaller A, Rudolf K, Dejonghe L, Grieben C, Froboese I. Influencing factors on the overestimation of self-reported physical activity: a cross-sectional analysis of low back pain patients and healthy controls. Biomed Res Int 2016;2016:1497213 [FREE Full text] [doi: 10.1155/2016/1497213] [Medline: 27298820]
20. Shum GL, Crosbie J, Lee RY. Symptomatic and asymptomatic movement coordination of the lumbar spine and hip during an everyday activity. Spine (Phila Pa 1976) 2005 Dec 01;30(23):E697-E702. [doi: 10.1097/01.brs.0000188255.10759.7a] [Medline: 16319739]
21. Shum GL, Crosbie J, Lee RY. Effect of low back pain on the kinematics and joint coordination of the lumbar spine and hip during sit-to-stand and stand-to-sit. Spine (Phila Pa 1976) 2005 Oct 01;30(17):1998-2004. [doi: 10.1097/01.brs.0000176195.16128.27] [Medline: 16135992]
22. Bauer CM, Rast FM, Ernst MJ, Kool J, Luomajoki H, Suni J, et al. Validity and reliability of inertial measurement units when measuring lumbar range of motion, movement control, repetetive movement and reposition error. Physiotherapy 2015 May;101:e914-e915. [doi: 10.1016/j.physio.2015.03.1754]
23. Victorino MN, Jiang X, Menon C. Chapter 7 - wearable technologies and force myography for healthcare. In: Tong RK, editor. Wearable Technology in Medicine and Health Care. Cambridge, MA: Academic Press; 2018.
24. Ricci L, Taffoni F, Formica D. On the orientation error of IMU: investigating static and dynamic accuracy targeting human motion. PLoS One 2016;11(9):e0161940 [FREE Full text] [doi: 10.1371/journal.pone.0161940] [Medline: 27612100]
25. Zhang Y, Haghighi PD, Burstein F, Yap LW, Cheng W, Yao L, et al. Electronic skin wearable sensors for detecting lumbar-pelvic movements. Sensors (Basel) 2020 Mar 09;20(5):1510 [FREE Full text] [doi: 10.3390/s20051510] [Medline: 32182928]
26. Papi E, Koh WS, McGregor AH. Wearable technology for spine movement assessment: a systematic review. J Biomech 2017 Dec 07;64:186-197 [FREE Full text] [doi: 10.1016/j.jbiomech.2017.09.037] [Medline: 29102267]
27. Yun WS, Kim H, Ahn JH, Park YB, Park YJ. Individual characteristics of reliable lumbar coupling motions. Eur Spine J 2015 Oct;24(9):1917-1925. [doi: 10.1007/s00586-015-4081-0] [Medline: 26100279]
28. Loh KJ, Lin YA. Smart elastic fabric tape for distributed skin strain, movement, and muscle engagement monitoring. U.S. Utility Patent Application. 2021. URL: https://patentimages.storage.googleapis.com/a8/81/5b/4117f6d1403cc2/US20220087565A1.pdf [accessed 2024-02-05]
29. Loh KJ, Kim J, Lynch JP, Kam NW, Kotov NA. Multifunctional layer-by-layer carbon nanotube–polyelectrolyte thin films for strain and corrosion sensing. Smart Mater Struct 2007 Feb 09;16(2):429-438. [doi: 10.1088/0964-1726/16/2/022]
30. Loh KJ, Lynch JP, Shim BS, Kotov NA. Tailoring piezoresistive sensitivity of multilayer carbon nanotube composite strain sensors. J Intell Mater Syst Struct 2007 Oct 17;19(7):747-764. [doi: 10.1177/1045389x07079872]
31. Loyola BR, La Saponara V, Loh KJ. In situ strain monitoring of fiber-reinforced polymers using embedded piezoresistive nanocomposites. J Mater Sci 2010 Jul 31;45(24):6786-6798. [doi: 10.1007/s10853-010-4775-y]
32. Wang L, Loh KJ. Wearable carbon nanotube-based fabric sensors for monitoring human physiological performance. Smart Mater Struct 2017 Apr 13;26(5):055018. [doi: 10.1088/1361-665x/aa6849]
33. Lin YA, Zhao Y, Wang L, Park Y, Yeh YJ, Chiang WH, et al. Graphene K‐Tape meshes for densely distributed human motion monitoring. Adv Mater Technol 2021 Jan;6(1):2000861. [doi: 10.1002/admt.202000861]
34. Lin YA, schraefel MC, Chiang WH, Loh KJ. Wearable nanocomposite kinesiology tape for distributed muscle engagement monitoring. MRS Advances 2021 Jan 28;6(1):6-13. [doi: 10.1557/s43580-021-00005-4]
35. Lin YA, Mhaskar Y, Silder A, Sessoms PH, Fraser JJ, Loh KJ. Muscle engagement monitoring using self-adhesive elastic nanocomposite fabrics. Sensors (Basel) 2022 Oct 07;22(18):6768 [FREE Full text] [doi: 10.3390/s22186768] [Medline: 36146120]
36. Gombatto SP, Brock T, DeLork A, Jones G, Madden E, Rinere C. Lumbar spine kinematics during walking in people with and people without low back pain. Gait Posture 2015 Oct;42(4):539-544. [doi: 10.1016/j.gaitpost.2015.08.010] [Medline: 26380913]
37. Hodges PW, van den Hoorn W. A vision for the future of wearable sensors in spine care and its challenges: narrative review. J Spine Surg 2022 Mar;8(1):103-116 [FREE Full text] [doi: 10.21037/jss-21-112] [Medline: 35441093]
38. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
39. Bunniss S, Kelly DR. Research paradigms in medical education research. Med Educ 2010 May;44(4):358-366. [doi: 10.1111/j.1365-2923.2009.03611.x] [Medline: 20444071]

40. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Oct;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
41. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Mar;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
42. Lange J, Schneider A, Jordi C, Lau M, Disher T. Formative study on the wearability and usability of a large-volume patch injector. Med Devices Evid Res 2021 Nov;Volume 14:363-377. [doi: 10.2147/mder.s337670]
43. Guest G, Namey E, McKenna K. How many focus groups are enough? Building an evidence base for nonprobability sample sizes. Field Methods 2016 Jul 24;29(1):3-22 [FREE Full text] [doi: 10.1177/1525822x16639015]
44. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis: striving to meet the trustworthiness criteria. Int J Qual Methods 2017 Oct 02;16(1) [FREE Full text] [doi: 10.1177/1609406917733847]
45. Hamilton AB, Finley EP. Qualitative methods in implementation research: an introduction. Psychiatry Res 2019 Oct;280:112516 [FREE Full text] [doi: 10.1016/j.psychres.2019.112516] [Medline: 31437661]
46. Carter N, Bryant-Lukosius D, DiCenso A, Blythe J, Neville AJ. The use of triangulation in qualitative research. Oncol Nurs Forum 2014 Oct;41(5):545-547. [doi: 10.1188/14.ONF.545-547] [Medline: 25158659]
47. Su WC, Yeh SC, Lee SH, Huang HC. A virtual reality lower-back pain rehabilitation approach: system design and user acceptance analysis. In: Proceedings of the Universal Access in Human-Computer Interaction. Access to Learning, Health and Well-Being. 2015 Presented at: Universal Access in Human-Computer Interaction. Access to Learning, Health and Well-Being; August 2-7, 2015; Los Angeles, CA. [doi: 10.1007/978-3-319-20684-4_37]
48. Lin WY, Chou WC, Tsai TH, Lin CC, Lee MY. Development of a wearable instrumented vest for posture monitoring and system usability verification based on the technology acceptance model. Sensors (Basel) 2016 Dec 17;16(12):2172 [FREE Full text] [doi: 10.3390/s16122172] [Medline: 27999324]
49. Blanchard A, Nguyen SM, Devanne M, Simonnet M, Le Goff-Pronost M, Rémy-Néris O. Technical feasibility of supervision of stretching exercises by a humanoid robot coach for chronic low back pain: the R-COOL randomized trial. Biomed Res Int 2022 Mar 9;2022:5667223-5667210 [FREE Full text] [doi: 10.1155/2022/5667223] [Medline: 35309176]
50. Luciani B, Braghin F, Pedrocchi AL, Gandolla M. Technology acceptance model for exoskeletons for rehabilitation of the upper limbs from therapists' perspectives. Sensors (Basel) 2023 Mar 03;23(3):1721 [FREE Full text] [doi: 10.3390/s23031721] [Medline: 36772758]
51. Sun X, Zhao C, Li H, Yu H, Zhang J, Qiu H, et al. Wearable near-field communication sensors for healthcare: materials, fabrication and application. Micromachines (Basel) 2022 May 17;13(5):784 [FREE Full text] [doi: 10.3390/mi13050784] [Medline: 35630251]
52. Andrýsková A, Lee JH. The guidelines for application of kinesiology tape for prevention and treatment of sports injuries. Healthcare (Basel) 2020 May 26;8(2):144 [FREE Full text] [doi: 10.3390/healthcare8020144] [Medline: 32466467]
53. Cheatham SW, Baker RT, Abdenour TE. Kinesiology tape: a descriptive survey of healthcare professionals in the United States. Int J Sports Phys Ther 2021 Jul 01;16(3):778-796 [FREE Full text] [doi: 10.26603/001c.22136] [Medline: 34123530]
54. Yun-An, L, Noble E, Chin-Hsiung L, Loh KJ. Respiration monitoring using a motion tape chest band and portable wireless sensing node. J Commerce Biotechnol 2022;27(1). [doi: 10.5912/jcb1026]
55. George SZ, Lentz TA, Beneciuk JM, Bhavsar NA, Mundt JM, Boissoneault J. Framework for improving outcome prediction for acute to chronic low back pain transitions. Pain Rep 2020;5(2):e809 [FREE Full text] [doi: 10.1097/PR9.0000000000000809] [Medline: 32440606]
56. Collado-Mateo D, Lavín-Pérez AM, Peñacoba C, Del Coso J, Leyton-Román M, Luque-Casado A, et al. Key factors associated with adherence to physical exercise in patients with chronic diseases and older adults: an umbrella review. Int J Environ Res Public Health 2021 Mar 19;18(4):2023 [FREE Full text] [doi: 10.3390/ijerph18042023] [Medline: 33669679]
57. Palazzo C, Klinger E, Dorner V, Kadri A, Thierry O, Boumenir Y, et al. Barriers to home-based exercise program adherence with chronic low back pain: patient expectations regarding new technologies. Ann Phys Rehabil Med 2016 May;59(2):107-113 [FREE Full text] [doi: 10.1016/j.rehab.2016.01.009] [Medline: 27050664]

Abbreviations

APTA: American Physical Therapy Association
FG: focus group
HIPAA: Health Insurance Portability and Accountability Act
IMU: inertial measurement unit
LBP: low back pain
PT: physical therapy
RQA: rapid qualitative analysis
TAM: Technology Acceptance Model

Edited by A Kushniruk; submitted 08.12.23; peer-reviewed by O Ang, S Oh; comments to author 06.01.24; revised version received 13.01.24; accepted 17.01.24; published 29.02.24.

Please cite as:
Lee A, Dionicio P, Farcas E, Godino J, Patrick K, Wyckoff E, Loh KJ, Gombatto S
Physical Therapists' Acceptance of a Wearable, Fabric-Based Sensor System (Motion Tape) for Use in Clinical Practice: Qualitative Focus Group Study
JMIR Hum Factors 2024;11:e55246
URL: https://humanfactors.jmir.org/2024/1/e55246
doi:10.2196/55246
PMID:38421708

Original Paper

Physicians' and Patients' Expectations From Digital Agents for Consultations: Interview Study Among Physicians and Patients

Andri Färber[1,2], MSc; Christiane Schwabe[2], PhD; Philipp H Stalder[1], MD; Mateusz Dolata[2], PhD; Gerhard Schwabe[2], Prof Dr

[1]ZHAW School of Management and Law, Zurich University of Applied Sciences, Winterthur, Switzerland
[2]Department of Informatics, University of Zurich, Zurich, Switzerland

Corresponding Author:
Andri Färber, MSc
ZHAW School of Management and Law
Zurich University of Applied Sciences
Theaterstr. 17
Winterthur, 8401
Switzerland
Phone: 41 79 642 08 85
Email: xfbe@zhaw.ch

Abstract

Background: Physicians are currently overwhelmed by administrative tasks and spend very little time in consultations with patients, which hampers health literacy, shared decision-making, and treatment adherence.

Objective: This study aims to examine whether digital agents constructed using fast-evolving generative artificial intelligence, such as ChatGPT, have the potential to improve consultations, adherence to treatment, and health literacy. We interviewed patients and physicians to obtain their opinions about 3 digital agents—a silent digital expert, a communicative digital expert, and a digital companion (DC).

Methods: We conducted in-depth interviews with 25 patients and 22 physicians from a purposeful sample, with the patients having a wide age range and coming from different educational backgrounds and the physicians having different medical specialties. Transcripts of the interviews were deductively coded using MAXQDA (VERBI Software GmbH) and then summarized according to code and interview before being clustered for interpretation.

Results: Statements from patients and physicians were categorized according to three consultation phases: (1) silent and communicative digital experts that are part of the consultation, (2) digital experts that hand over to a DC, and (3) DCs that support patients in the period between consultations. Overall, patients and physicians were open to these forms of digital support but had reservations about all 3 agents.

Conclusions: Ultimately, we derived 9 requirements for designing digital agents to support consultations, treatment adherence, and health literacy based on the literature and our qualitative findings.

(JMIR Hum Factors 2024;11:e49647) doi:10.2196/49647

KEYWORDS

adherence to treatment; digital agents; eHealth; electronic medical records; health literacy; mobile health; mHealth; mobile phone

Introduction

Motivation

Consultations are less productive than what physicians and patients would wish [1,2], which hampers health literacy, shared decision-making, and treatment adherence. The recent rise of generative artificial intelligence (AI), such as ChatGPT, has sparked the interest of digital health developers, as they explore how this technology can improve shared decision-making, physician-patient communication, adherence to treatment, and health literacy. In this study, we sought to discover what physicians and patients expect from digital agents (functional requirements) and how this functionality should be provided (nonfunctional requirements). A user-centric perspective is essential for guiding the development of digital agents because it prepares physicians for changes in their consultation methods and allows patients to understand what the new technology can offer.

Through in-depth interviews (refer to the Methods section), we described 3 digital agents to physicians and patients, analyzed their impressions and expectations (refer to the Results section), and deduced a set of design requirements (refer to the Discussion section). An introduction to the related work and concepts for the 3 different digital agents is provided in the following sections.

Related Work and Concepts

Relevant Medical Concepts

Overall, four medical concepts are essential when supporting medical consultations with digital agents: (1) shared decision-making, (2) physician-patient communication, (3) adherence to treatment, and (4) health literacy.

Consultations involve a participatory process between patients and physicians to reach an agreement regarding treatment goals and their implementation [3,4]. "Shared decision-making" has emerged as the gold standard for this participatory process [5-10] as it strives to reach a mutual agreement about therapy [6,7]. However, a systematic review of shared decision-making regarding clinical decisions found that the humanistic aspects of physician-patient communication were rarely assessed [11]. Good "physician-patient communication" is not only about technique or process but also involves understanding the whole person, finding common ground, and enhancing the patient-physician relationship [4]. In this way, physician-patient communication can have a therapeutic effect and influence health benefits [12].

The therapeutic process continues after the patient has left the consultation [3]. Once at home, it is up to the patient to implement the therapy plan, and the extent to which this occurs is referred to as "adherence to treatment" [13]. Adherence focuses on patients taking responsibility for their treatment and physicians collaborating more with their patients [14,15]. However, despite some progress, adherence to treatment remains insufficient [13,16-18]. First, there is a lack of "health literacy" when following the given instructions. Physicians may explain medical issues and treatment options during consultations, but their time is limited, and they must convey as much information as possible. Second, patients are in a stressful situation, which restricts their ability to absorb and hinders their recall [19-24]. Third, physicians may use medical terminology [25] with the following consequences: patients either do not understand or quickly forget what was discussed [26,27]. Brochures and leaflets are typically used to support health literacy, and modern approaches include video, multimedia, computer-assisted learning, mobile apps, and other web-based aids [28-32].

Digital Agents

Digital agents are computers that undertake tasks previously performed by humans. As such, they function autonomously, react to environmental situations, initiate actions, communicate with humans or machines, and behave intelligently [33]. An increasing volume of digitized data, improved algorithms, and better hardware has vastly enhanced the range of tasks that digital agents can perform. The most noticeable aspect is the recent success of generative AI. Nevertheless, the expanding capabilities of digital agents also raise concerns about AI in general and digital agents in particular [34]. Examples include their potential misuse, how they can be controlled, and whether they exhibit bias [35]. Besides these general concerns, researchers are interested in understanding exactly how digital agents interact with humans. Although humanlike behavior may be helpful in some situations, task performance may be impeded by excessive humanness [36,37] such as in situations where humans prefer a digital agent with a background function. This issue is critical in institutional settings [38], where professionalism is vital.

Discussion about the capabilities of digital agents and their suitability has also reached the medical domain [33,39,40]. Conceptually, the dyadic physician-patient consultation becomes triadic [41-44] if a digital agent is included. The presence of digital agents changes the consultation dynamics [45,46] and alters how patients and physicians behave [41]. Despite such insights, the discussion lacks a clear conceptualization of the digital agent's role in the professional context of physician-patient consultation. Consequently, discussing what physicians and patients expect from digital agents during and between consultations has not been possible.

Current Digital Support for Consultation, Adherence to Treatment, and Health Literacy

Physicians use electronic medical records (EMRs) and encounter patient decision aids (PDAs) during consultations, which provides patients access to their data through patient portals. Patients may also store data in their personal health records (PHRs) and take advantage of mobile health (mHealth) apps between sessions.

EMRs support physicians in documenting medical history, including physical examinations and laboratory results. They are intended to reduce costs, improve patient safety, increase efficiency [47], and safeguard data [48,49]. As EMRs are designed primarily for documentation purposes [50], it is the physician's responsibility to determine how to use them in patient interactions. Proper use of EMRs by trained health care professionals can improve health literacy and adherence to treatment compared with paper-based records [51], for example, if physicians share their EMR screens with patients during consultations [52,53]. However, when used ineptly, physicians lose control of the consultation owing to increased gaze shifts and multitasking, which hinders their medical reasoning [47,54]. In the presence of a computer, preexisting positive and negative communication skills are amplified [55,56].

Encounter PDAs support physician-patient consultations by providing decision-related information and choices [57-61]. Although they tend to be simple in design [61], physicians complain that lack of training and experience and insufficient content and format impede meaningful use of encounter PDAs [57,58]. Another challenge is keeping encounter PDAs updated with the latest information [60].

Patient portals provide patients with access to their data stored in EMRs [62]. In such tethered patient portals, the responsibility for maintaining the data lies with the physician. To be understood by patients, information from EMRs must be translated [62], and this applies to language, graphs, and other multimedia material.

Unlike patient portals, in electronic PHRs, patients themselves enter and maintain their health data [63]. Although PHRs can accumulate more information than patient portals, quality control and manageability are challenging. There is a consensus that more needs to be done (eg, patients also need to understand what they get from the PHR and need to act on what they understand) to enhance health outcomes or treatment adherence than just providing patients with access to their data [64,65]. Better-informed patients are not necessarily healthier patients [64], but there is (1) value and (2) potential in patient portals and PHRs. First, patients want access to their data to review it again at home, discuss it with their families, and use it as a starting point for further online research [62,64]. Second, there is evidence suggesting that patient portals and PHRs are more effective when they are interactive, when they are combined with other services such as reminders or interactive decision support, and when physicians actively promote their use [62,64].

Digital interventions based on mHealth apps promise to support patients' health literacy and adherence to treatment. In 2017, >300,000 health apps were available in online app stores [66]. Not all are considered effective, convenient, or of high quality [67-69], and many have low success rates and high dropout rates [70-72]. Nevertheless, despite their limitations, mHealth apps appear to support patients effectively in treatment adherence [67,73,74]. If they pass the medical quality requirements, they can even be prescribed in the same manner as medicine [75,76]. Physicians are best placed to assist with their use, but this requires their integration into workflows and EMRs [74,77,78], and the security of patient data must be guaranteed [79].

Digital Agents to Support Consultation, Adherence to Treatment, and Health Literacy

Overview

We conceptualized 3 general roles for digital health agents, which tie together the modern medical concepts and previous studies of digital agents with current digital support for consultation, adherence to treatment, and health literacy. These served as a basis for our empirical study, when introducing our selected physicians and patients to digital agents.

A digital agent can be a "digital expert" that provides the right aids at the right time or offers a second opinion about diagnosis and treatment. It can stay in the background of the consultation as a "silent digital expert" or actively participate in the consultation as a "communicative digital expert." Alternatively, it can be a "digital companion" (DC), which supports the patient between consultations. DCs provide patients with comprehensible information about diagnosis and ongoing treatment.

Silent Digital Expert

This is an extension of EMRs, providing the physician with contextual and real-time advice and additional information. The silent digital expert is designed to free the physician from searching vast information sources and allows more time for face-to-face consultation, thereby improving physician-patient communication [4,12]. For example, the silent digital agent can alert physicians to different diagnoses and drug interactions or offer prompts for further questions. The silent digital agent also supports diagnosis and suggests appropriate treatment in a shared decision-making process [5-10]. It acts as an aid to the physician and is visible and accessible only to the physician, and with patient consent, it can record, transcribe, analyze, and summarize the consultation.

Communicative Digital Expert

As the third party in a triadic consultation, the communicative digital expert offers the same functionality as the silent digital expert. However, it actively participates in the consultation by extending the functionality of EMRs and encounter PDAs through an agency. It may be physically represented as a humanlike robot, smart speaker, or device of any shape. As the third party, the communicative digital expert can be invited to comment about the decision-making process of physicians or patients [5-10] and become active in explaining medical topics, thereby improving health literacy [80-83]. As such, it can be considered as a physician's assistant or patient's advocate, thus improving physician-patient communication [4,12]. For example, it might interrupt the dialogue if a physician is very brief or dominant, thereby providing both parties with further information, diagnosis considerations, and treatment recommendations. It acts in an empathetic, patient-centered manner and is capable of identifying and taking patient preferences into consideration.

Digital Companion

This agent is intended to support patients *between* consultations by extending patient portals and PHRs and combining them with an mHealth app. It relies on data from EMRs and supports patient treatment behavior. Its primary goals are to improve the recall of recommendations and information, promote health literacy [80-83], and support treatment adherence [12-18,84]. DC captures the critical points of the physician-patient consultation, translates them into everyday language, enriches them with multimedia elements (audio, picture, diagram, and video), and makes them conveniently accessible to patients or their families at any time. It also provides the patient with curated additional information and interactively supports their health care education based on individual preferences. Using sensor data from various devices (eg, smartphones, smartwatches, pedometers, and blood glucose monitors) and patient's interaction with DC, adherence to the treatment plan is measured, analyzed, and fed back to the patient (and with the patient's consent, to the physician). DC provides context-specific, adaptive interventions [85-88] based on adherence measurement, individual treatment agreement, and patient preferences. For example, adherence support might include diet recipes, exercise instructions, morale-boosting talks, and so on.

Methods

Research Approach

This study aims to understand what physicians and patients require from digital agents. These requirements should be grounded not only on technical vision but also on current consultation practices, with a focus on problem-solving.

Our research approach was inspired by the practice-oriented approach popular in computer-supported cooperative work (CSCW). CSCW is an interdisciplinary field of research involving, among others, computer science, psychology, and sociology, to analyze the potential and the shortcomings of digital assistance in consultations [89-91]. CSCW mainly uses qualitative methods and focuses on how human collaboration can be supported by technical means [89,92]. As these means must be applied within a professional context, this also involves studying work practices from the perspective of those involved [93,94].

Our study embraced this tradition by following an exploratory paradigm, striving for deep, contextualized insights [95,96]. We conducted an interview-based qualitative study with 47 participants—22 (47%) physicians and 25 (53%) patients. Our analysis combined bottom-up thematic analysis and interpretive research, allowing for both broad coverage and deep insight.

Overall, the chosen methodological approach respected the need to understand patients' and physicians' perspectives regarding their work practices and the potential use of technologies. We addressed variation and triangulation, whereby multiple researchers conducted the interviews with different patients and physicians. We ensured audit throughout the process by mutual control among researchers and by assigning a quality manager role to one of the authors. The first author was directly engaged in data collection during a preliminary study [97] and guided data collection during this study to ensure adequate engagement in data collection activities. In summary, the study used various strategies to ensure the reliability and validity of the presented results [98] and followed the COREQ (Consolidated Criteria for Reporting Qualitative Research) guidelines for reporting qualitative research [99].

Ethical Considerations

The Ethics Committee of the Zurich canton confirmed that this study was not subject to the Swiss Human Research Act (Business Administration System for Ethics Committees [BASEC]–Nr Req-2018-00847). Nevertheless, written informed consent was obtained from all participants before their interviews according to the World Medical Association Declaration of Helsinki [100].

Sampling and Recruitment

Exploratory studies require a variety of opinions, but they do not seek to be representative. To ensure variety, we interviewed both physicians and patients. We also relied on purposive sampling using a maximum variation strategy [101], which allowed us to search for a broad range of physicians and patients. Given that 5 interviewers acquired the patients and physicians independently, we can assume the coverage to be better than that of strategies involving sampling through a single researcher. Table 1 shows the demographic characteristics of the study participants.

Table 1. Demographic data of the interviewed physicians and patients.

Characteristics	Physicians (n=22)	Patients (n=25)
Sex, n (%)		
Male	12 (55)	14 (56)
Female	10 (45)	11 (44)
Age (y), mean (SD; range)	50 (14; 25-66)	46 (19; 20-86)

Of the 22 physicians, 13 (59%) are active in primary care, and the others work in hospitals; 11 (50%) are general practitioners or specialize in internal medicine. Other specializations include pediatrics, gynecology, radio-oncology, and dentistry. The educational background of the 25 patients ranged from unskilled workers to professionals and academics. The patients presented a broad spectrum of conditions, including diabetes, multiple sclerosis, heart conditions, tick-borne encephalitis, and epilepsy.

We conducted 46 in-depth interviews that resulted in audio recordings with 32 hours of interview time, amounting to an average length of 42 minutes and 46 seconds (SD 13 min and 47 s). Of the 46 interviews, 45 (98%) were conducted with 1 interviewee per session, and 1 (2%) involved 2 respondents. The sample size assured data saturation—the topics emerging in the interviews began to overlap after about 18 to 20 interviews for each group [102]. Consistent with the practice for purposive sampling and maximum variation [101], we used various channels to establish the initial contact with the interviewees (email, face-to-face, and telephone). After confirming the time and date for a potential interview and giving their consent, no one dropped out of the study.

Data Collection

In total, 5 researchers conducted in-depth interviews based on the respective interview guides—separate guides for patients and physicians [96]. The interview guides were developed based on the literature about physician-patient communication; adherence to treatment; existing solutions in the field of medical informatics; and the authors' own experiences in the medical domain, including their research background. The overall structure of the interviews was informed by CSCW practice-oriented studies [93,94]. The interview guides were pretested in a preliminary study (with 11 health care professionals and 7 patients) published elsewhere [97]. Interviews for this study were conducted between January 2019 and May 2019, with patient interviews being conducted mostly in their homes and health care professional interviews in their professional setting. Before the interviews, all researchers underwent interview training sessions to ensure that they had the same understanding of the questions and knew how to conduct the interview. The interviews were structured around 3 areas: current situation or practice (format of and preparation for a consultation), future developments (expectations from and attitudes toward digital health care), and closure (other points that were not already covered).

When discussing about digital developments, we suggested potential ideas because users often lack the necessary imagination when asked about future products or services [103]. Nevertheless, when prompted, many users can express helpful, subjective opinions about specific ideas [103]. Therefore, in the spirit of design thinking [104], we exposed the users to key design ideas by describing the digital experts and DC and asking for their perceptions, expectations, and preferences regarding digital agent support. As is typical in design thinking, the discussion focused on the desirability of critical capabilities but did not include a detailed discussion about feasibility.

Data Analysis

All the interviews were audio recorded and transcribed. The analysis combined deductive thematic research and interpretive research, allowing for broad coverage and deep insight simultaneously. During the top-down analysis, the transcripts were coded according to a codebook derived inductively from a small preliminary study [97]. A professor of nursing science cross-checked the codebook. Again, all researchers attended a training session to ensure that they had the same understanding of the codebook. All interviews were then deductively coded using MAXQDA (VERBI Software GmbH) [105]. The designated quality manager conducted quality assurance activities by controlling all code assignments and correcting them to ensure a consistent basis for analysis. We achieved thematic saturation—all themes from the specified coding schema appeared in the data with high frequency (the most frequent code was assigned 274 times and the least frequent was assigned 25 times; overall, we had 1954 assignments across all codes) [102]. Finally, all interviews were summarized by code; for each theme, we obtained a summary of participant opinions related to the code. These summaries formed the basis for further analysis, and the results were then used for interpretation.

To interpret the data, we organized 2 interpretation workshops involving the authors. The workshops aimed to establish a shared and consistent understanding of the most essential insights between the authors. The interpretive process involved iterative restructuring of the summaries along various dimensions, with 2 dimensions emerging as crucial for forming a consistent data view. First, we differentiated the problems, current practices that emerged to mitigate those problems, and potential technological solutions to address the problems that occurred during the interviews. Second, we observed that the issues aligned with the phases of a patient's journey: (1) consultation, (2) "transition" between consultations and period between consultations, and (3) actual period between consultations. These differentiations provided the framework for reporting our results, and the proposed structure covered all the challenges and problems identified during coding.

In our presentation of the results, we refer to the frequency of specific challenges because, after identifying the framework and distributing the significant challenges for each element in the framework, we returned to the coded data to classify the coded passages. In the following section, we have presented the quantified data about the frequency of passages pertaining to the challenges. However, it is important to clarify that we do not assert the representativeness of these figures, as the analyzed population was not chosen to be representative of the broader population. Instead, the numbers ensured the thematic saturation mentioned previously.

Results

Through analysis, we categorized the results into 3 steps in the patient journey: first, the consultation; then, incorporating information from the consultation into their lives; and finally, the time between consultations.

Problems and Agent-Based Solutions During a Consultation

During consultation, the main challenge, according to physicians and patients, is conveying complex information in minimal time to laypeople with various backgrounds, expectations, and abilities while building or maintaining a relationship of trust. Table 2 summarizes the problems voiced by physicians and patients, current practices (as presented by the interview partners), and envisioned solutions offered by the 2 different versions of digital experts.

Table 2. Problems and solutions suggested during a consultation, along with the number of mentions in interviews.

Problems during the consultation	Current practices	Solutions offered by the silent digital experts	Solutions offered by communicative digital experts
Time pressures (physicians: 5/22, 23%; patients: 3/25, 12%)	—[a]	• Physicians can concentrate on a thorough and engaging consultation using digital situational information	• Physicians can concentrate on a thorough and engaging consultation using digital situational information

Medical information is complex (physicians: 9/22, 41%; patients: 18/25, 72%)	• Physicians use graphics, visualizations, videos, and 3D models from brochures, books, and online sources (physicians: 14/22, 64%; patients: 7/25, 28%)
	• Physicians draw illustrations themselves (physicians: 13/22, 59%; patients: 4/25, 16%)
		• The digital expert provides physicians with the following:
		  • The right visual aid at the right time
		  • Graphic templates or blank drawing areas that they can use for their own drawings
			• The digital expert suggests text, images, audios, and videos tailored to individual patient needs

Not all patients respond to medical advice and information in the same manner (physicians: 12/22, 55%; patients: no matching question)	• Most physicians try to approach patients individually by adapting their language to a patient's educational background or medical knowledge (physicians: 4/22, 18%; patients: no matching question)	• The digital expert provides visual aids tailored to a patient's educational background, numerical ability, or language skills	• The digital expert intervenes if it determines (eg, through sentiment analysis [106]) that a patient does not understand the physician
Patients expect more transparency and control over the treatment process (physicians: 2/22, 9%; patients: 21/25, 84%)	• Many patients engage in conversations with physicians and take responsibility for their treatment (patient: 3/25, 12%), and physicians try to support this (physicians: 4/22, 18%; patients: 5/25, 20%)	—	• The digital expert intervenes when physicians do not give their patients enough time to talk, and it can empower patients to take more control
Some patients do not agree with the proposed treatment plan (physicians: 4/22, 18%; patients: 6/25, 24%)	• Physicians respond with more intensive explanations (physicians: 8/22, 36%) • Physicians protect themselves by documenting the conversation • Physicians do not enforce treatment	• The digital expert offers arguments, statistics, and figures to support the physician's point of view	• The digital expert advocates for the patient (by putting the physician's thoughts or guidelines into perspective) or for the physician (by supporting the physician's thoughts or guidelines)
The computer distracts the physician and interrupts communication, and use of computer amplifies inferior communication skills	—	• The user interface of the digital expert is designed to be self-explanatory and user-friendly • Instead of the physician, the digital expert searches for information and offers context-related content	• The digital expert supports the physician and the patient, for example, through active listening • It will only interfere by assisting an already impaired conversation

[a]Nothing mentioned in the interviews.

Regarding current practices, patients and physicians report that there is very little time for a thorough and engaging conversation:

> *I just felt like I was being processed. Quick assessment with the question: What's the problem? And I felt that I couldn't even say what I had because it was already clear to the physician. After a quarter of an hour, I was out of there again, and I was no wiser.* [Male patient; aged 60 years; D07]

> *I frequently make lifestyle recommendations. Costs time too, by the way, cannot be done in a 20-minute consultation that's just long enough for issuing a prescription.* [Male general practitioner; aged 64 years; hospital; ST09]

Most physicians in this sample practice shared decision-making. Some use the explicit term during the interview, whereas others simply implement shared decision-making without labeling it as such:

> *Then I say, we could try pharmacy, we could try herbs, we could try acupuncture or this or that. I'll let the patient have a say. Because then the patient's adherence is also much better.* [Female general practitioner; aged 65 years; medical office; MA10]

All interviewed patients favored a silent digital expert as an aid to the physician; they did not object to physicians using online sources to obtain additional information during a consultation:

> *I don't like having a doctor who introduces him- or herself as "I am the all-knowing one." For me, that*

> *tends to inspire confidence when a physician says: I don't know, I have to work with the exclusion procedure.* [Male patient; aged 74 years; F01]

However, patients expect uninterrupted attention, which requires a sufficiently high level of expertise by the physician in using the computer:

> *He kept asking and reading to me while he was writing and asking me if that was correct. This was great for me because then I knew what he was writing.* [Female patient; aged 52 years; S10]

Most physicians in this study would welcome a silent digital expert to facilitate multitasking, and some already use drug interaction assistants, risk or score calculators:

> *You can't read through the books in the evening. That would mean an insane amount of time or such a head. That's why these are important tools, I think for rare conditions it's certainly a good idea.* [Female gynecologist; aged 35 years; hospital; MA02]

However, the benefits of a digital expert are assessed differently by those in different medical disciplines. A physician was concerned about the transfer of responsibility to the digital expert, whereas another physician worried about a decline in interprofessional communication. A young physician was concerned that this would cause them to acquire very little experience and self-confidence:

> *You rely too firmly on that afterward. Then you believe too firmly in that. Then it takes over your task, so to speak.* [Female dentist; aged 29 years; dental surgery; MA03]

Most patients in this sample view communicative digital experts positively. Those against them are concerned that they might be disruptive or could be manipulated by the physician:

> *I do not know what the physician can enter there, and then it is clear that the computer represents the opinion of the physician.* [Female patient; aged 51 years; S07]

The opinions of those in favor of it differ. Some consider a communicative digital expert as helping less skillful physicians and others consider it as helping competent physicians. Some would like a digital expert to be a physician's assistant, whereas others consider it as a patient's advocate:

> *As a patient, you are always subordinate to the physician, in that sense. I don't think it's a bad thing when someone else is on my side.* [Female patient; aged 28 years; S06]

Approximately two-thirds of the interviewed physicians reject the communicative digital expert. For them, credibility, decision-making authority, and their patients' trust are at stake. Some consider empathy between the physician and patient as essential for patient adherence to treatment and, therefore, do not believe that a digital expert can help. A physician found communicative digital experts annoying but assumed that physicians and patients would get used to them over time:

> *In principle, I say, there is still an interpersonal level that artificial intelligence cannot comprehend.* [Female general practitioner; aged 48 years; medical office; MA08]

Problems and Agent-Based Solutions for Transitioning From Consultations to the Period Between Consultations

Problems during the consultation may also hinder treatment because poor consultations can impair health literacy and adherence to treatment. Table 3 provides an overview of the voiced consultation issues that affect the time between consultations and the envisioned solutions offered through an interaction of the digital expert and DC.

Table 3. Problems and envisioned solutions for transitioning from consultations to the period between consultations, along with the number of mentions in the interviews.

Problems resulting from the consultation	Current practices	Solutions offered by the digital experts connecting to the digital companion
Patients cannot remember everything that the physician says (physicians: 0/22, 0%; patients: 10/25, 40%)	• Patients do the following: ◦ Bring companions to the consultation ◦ Consult brochures or online sources (patients: 2/25, 8%) ◦ Use reminders on smartphones (patients: 2/25, 8%) ◦ Take notes (patients: 6/25, 24%) • Physicians do the following: ◦ Repeat (physicians: 2/22, 9%) ◦ Use active listening techniques	The digital expert records, transcribes, and summarizes the conversation for the patient (quality assurance)
Identifying and introducing clinically relevant mHealth[a] apps is time consuming and difficult	• Patients search for apps themselves, but use dropout rates are high	The digital expert suggests quality-assured mHealth apps or equivalent features of the digital companion

[a]mHealth: mobile health.

Most physicians in this sample see potential in automated recording and transcription. A physician hoped that digital experts would give them more time to communicate with patients. However, physicians doubt whether a computer can separate relevant statements from irrelevant ones and produce relevant summaries. Some physicians stress that the notes they make for themselves about the case cannot be directly shared with the patient but need to be translated. Others insist on control over the information that is shared with patients:

> *Therefore, the software must either be able to guarantee this or otherwise it is legally difficult to prove that the patient has been informed correctly.* [Male radio-oncologist; aged 35 years; hospital; MA01]

Besides technical difficulties, the interviewed physicians see another reason to avoid automatic summaries—subjective perceptions are often only discussed verbally or communicated via telephone owing to fear of litigation:

> *Certain things, incidents and so on, or special experiences or special stories that are told that could have legal relevance. I don't list them in the computer.* [Male general practitioner; aged 62 years; medical office; ST02]

Another physician takes precisely the opposite position. They would appreciate transcripts of complex consultations in which, for example, discussions about child protection or off-label prescriptions of medication are involved. A physician did not believe that a consultation's significant first and last seconds would be transcribed with the necessary weighting.

Patients also have different opinions about digital experts. Only a few patients in this study raised data protection concerns regarding the consultation transcripts and other information recorded during the consultation. Some patients indicated that they would benefit from this evidence of what was said in the event of disagreement or malpractice. A patient was worried about a decline in care because physicians were afraid of malpractice lawsuits:

> *I tend to think I get worse treatment because most physicians have way too much fear of someone coming in afterward and saying, "I'm going to sue you – you told me something wrong."* [Male patient; aged 61 years; S02]

Problems and Agent-Based Solutions for the Period Between Consultations

The consultation cannot cover all the questions and issues arising between consultation appointments, and patients must rely on their own judgment or a tool that assists them during this period. Table 4 presents the problems that arise between consultations that lead to poor adherence and the solutions offered by DC.

Table 4. Problems and envisioned solutions for the period between consultations, along with the number of mentions in the interviews.

Problems arising between consultations	Current practices	Solutions offered by DC[a]
• Patients lack information because of the following: • Insufficient time for explanations during the consultation (physicians: 5/22, 23%; patients: 3/25, 12%) • Poor recall of the consultation (physicians: 3/22, 14%; patients: 11/25, 44%) • More questions arising later (physicians: 0/22, 0%; patients: 11/25, 44%)	• Patients do the following: • Use online sources, but they are skeptical, and some distrust online forums in particular (physicians: 6/22, 27%; patients: 6/25, 24%) • Read brochures (patients: 3/25, 12%), attend public lectures, or even attend anatomy courses • Physicians provide brochures to guide patients away from online self-diagnosis (physicians: 2/22, 9%).	DC provides curated content and web links tailored to the patient's diagnosis. This reduces misinformation and false self-diagnosis. In addition, it fosters more trust in health care information.
• Patients lack clear instructions and specific information but instead experience information overload (physicians: 0/22, 0%; patients: 8/25, 32%)	• Physicians provide paper-based instructions regarding medication, exercises, and lifestyle changes (physicians: 3/22, 14%; patients: 8/25, 32%)	DC tailors content to patient preferences, contexts, and specific circumstances. This includes content presentation in different formats (simple or sophisticated text, images, audios, and videos).
• Patients are on their own between consultations (physicians: 3/22, 14%; patients: 1/25, 4%)	• Patients report little interaction with their physicians between consultations • Some use email but only sparingly (physicians: 2/22, 9%; patients: 7/25, 28%)	DC provides low-barrier access to the physician between consultations. A chatbot covers part of the conversation to protect physicians from huge workload.
• Patients are overchallenged when taking their medication (physicians: 2/22, 9%; patients: 8/25, 32%)	• Medication apps can support complicated medication regimes (physicians: 1/22, 5%; patients: 2/25, 8%)	DC supports adherence by providing the patient with individualized interventions that consider patient preferences, contexts, and specific circumstances.
• Treatment success or failure goes unnoticed (physicians: 4/22, 18%; patients: 0/25, 0%)	• Physicians ask patients to maintain diaries or journals, mostly paper based (physicians: 8/22, 36%; patients: 3/25, 12%)	DC offers easy-to-maintain diaries and journals, including data captured from digital devices (eg, wearables). The collected data can be shared with physicians (with the patient's consent).
• Measuring adherence is difficult	• Adherence is rarely measured, and often, it is only based on the purchase of medicines (physicians: 1/22, 5%; patients: no corresponding question)	DC offers adherence measurements in an easy-to-understand format.

[a]DC: digital companion.

Most patients in this study would welcome a DC; however, a few are skeptical or undecided. Patients are open to using electronic tools and online services regarding current practices. However, this is not always helpful to physicians:

> *People practically come with a diagnosis, and after that, we first have to come back to the symptoms. And I have to say, "hey, we have to start all over again."* [Male general practitioner; aged 66 years; medical office; ST01]

Many physicians who were interviewed could see the potential of a DC. Some hoped this would improve adherence to medical advice, whereas a physician saw a significant benefit in making the DC genuinely personalized and tailored to an individual patient's needs. Regarding monitoring patient behavior between consultations, less than one-third of the physicians reported adherence measurement (which is usually based on the purchase of medications):

> *That's why I'm very happy when the patients order medication from us because then I can see on the computer when they have picked up their medication. I don't see that when they buy medicines from the pharmacy.* [Female general practitioner; aged 48 years; medical office; MA08]

Most physicians in this sample are open to receiving and interpreting monitoring data from patients and their mobile devices. However, they have the following reservations. First, there is an unmanageable number of mobile apps. Second, they fear data overload and being forced to respond to monitoring results, which requires additional time that physicians do not have. Third, physicians see a risk that such monitoring will negatively influence patient behavior. A physician raised the possibility that neurosis could result from constant introspection. Another concern was that patients would abdicate responsibility for their condition by transmitting data and threshold violations. Despite these concerns, confronting patients regarding their threshold violations encourages them to reflect on their condition and possible lifestyle changes. Therefore, patients can become "experts" on their condition:

Because that is certainly one aspect when patients think about it: Why did my sugar do that now? That's the most instructive. And the goal is that they become the "expert" and I coach them. [Female general practitioner; aged 39 years; hospital; ST08]

Discussion

Overview

Problems in physician-patient interaction that ultimately hamper treatment adherence can be classified into 3 categories: problems regarding the consultation itself, problems from the consultation but appearing between consultations, and new problems arising between consultations. These problems overlap and, therefore, need to be addressed using integrated support systems. On the basis of the scenario, a support system consisting of digital agents assisting in the consultation and a companion for the periods between consultations is proposed. To qualify for the task, these agents need to meet the expectations of physicians and patients and improve health outcomes. In the following sections, we discuss design recommendations for the 3 digital agents that are active in the consultation and act as the patient's companion between consultations.

Requirements for Digital Experts During the Consultation

Digital experts reveal their capabilities during the consultation by integrating and extending the functionalities of EMRs and encounter PDAs with the characteristics of digital agents [33]. These include autonomous and intelligent behavior, reactions to environmental situations, and communication with humans or machines.

The Digital Agent Should Make Its Role in the Triadic Consultation Transparent

Our interviews asked for opinions about including medically skilled digital agents as part of a physician's EMR [45,46]. These can facilitate conversations between physicians and patients or offer second opinions regarding diagnosis and treatment. In such cases, the digital agent functions as an additional physician. Although most patients would welcome this triadic consultation, some fear that physicians could manipulate their DCs. These reservations arise from an understanding that digital agents could adopt the role of a second physician and a trusted family member, spouse, or friend [41,42]. Such roles include informational or emotional support (eg, taking notes, ensuring understanding, and reassuring patients) [42]. Accordingly, the role of a digital agent in consultation must be clearly defined and transparent to patients. Further studies might explore what patients require to trust and benefit most from these digital agents in the role of a second physician, family member, spouse, or friend.

The Digital Agent Should Encourage Trust and Support the Physician-Patient Relationship While Safeguarding the Physician's Credibility

The literature and interviews with physicians and patients agree on the importance of trust and good relationships between physicians and patients in a medical setting [4,12]. Although traditional health IT (eg, EMRs and encounter PDAs) does not seem to interfere with patient-physician relationships [53], the situation changes when digital agents act as medical experts or DCs during a consultation. Most interviewed patients like the idea of a digital agent and do not think it will harm the physician-patient relationship. At the same time, many physicians have an opposing view, fearing loss of credibility and decision-making authority. Therefore, a challenge for DC is to foster trust and support, rather than undermine, the relationship between physicians and patients. Such digital agents must support patients but not unduly contradict physicians or disrupt the natural flow of conversation. This means that digital agents must recognize whether a piece of medical advice will strengthen or damage the relationship.

The Digital Agent Should Help Physicians to Focus on the Patient During the Consultation

The interviewed patients expect their physicians' full attention even when interacting with a computer. In a traditional practice setting, computer screens create a barrier between patients and physicians and can be a serious distraction [47,54]. However, digital agents act independently or are triggered by voice control to provide information or document the conversation, requiring less attention from the physician. The form of digital agents integrated into the conversation can range from shared screens or smart speakers to humanlike robots. Technological advances have brought such user interfaces and digital agents more close to reality. Further studies should indicate what patients and physicians are most likely to accept.

The Digital Agent Should Support Physicians by Taking Over Administrative Duties

Administrative duties prevents physicians from doing what they were trained to do (at considerable expense) and reduces their job satisfaction. The time pressure resulting from these administrative duties is a well-known problem that affects patient health outcomes [1,2,12]. This issue surfaced in the interviews with physicians and patients who were dissatisfied with their treatment. Therefore, a significant role for digital experts is to relieve physicians from as many administrative duties as possible. However, it is essential for physicians that their medical reasoning is considered as something more than mere administration. Recording, transcribing, and summarizing the conversation is necessary, but it is not the whole story. Digital experts should support medical reasoning of physicians and ask for it if not already done, rather than impeding it.

Requirements for Handover From Digital Experts to DCs

To ensure a seamless patient experience, information collected and discussed during the consultation must be passed from the digital experts supporting the consultation to a patient's DC.

The Digital Agent Should Tailor Information and Patient Education to Individual Patient Needs and Preferences

In supporting consultation, digital experts could, for example, provide appropriate information at the appropriate time. After consultation, DCs could continue patient education between consultations, which is tailored to their information needs and preferences. This can give physicians extra time during consultations [1,2] and assist patients in recalling recommendations and information [19,23,24]. In contrast to reading widely circulated brochures, leaflets, and generalized online sources [28,29,31,32,107], patients receive personalized information matching their specific circumstances and treatment plans. This saves time by reducing the need to guide patients away from potentially incorrect self-diagnosis [30].

Our interviews indicated that physicians effectively tailor information to their patients' needs and backgrounds. Therefore, digital agents in the form of digital experts and companions must keep up with or even outperform physicians to add value. To achieve this, digital experts should either be able to draw on predefined patient profiles or interpret and assess patient preferences and backgrounds correctly. Physicians understandably insist on maintaining overall control as they are liable for the information they give their patients. A suboptimal solution would require physicians to verify the information they provide patients via the DC. In contrast, a better solution would ensure (in a trusted manner) that the information offered was consistent with the physician's directions.

Requirements for the DC in the Period Between Consultations

DCs support patients as digital agents between consultations by integrating and extending the functionalities of patient portals, PHRs, and mHealth apps.

The Digital Agent Should Offer Adaptive Interventions for Behavior Change

In conventional lifestyle change treatment, adaptive interventions are standard, and physicians and patients adapt and agree about the treatment every few weeks or months, ideally in a shared decision-making process [3,4,6,7,9]. However, adjustment cycles are dependent on consultation cycles, and in the meantime, patients may treat themselves incorrectly or discontinue a treatment owing to a lack of corrective measures. Here, digital agents in the form of DCs can shorten the cycle considerably. Depending on a patient's mood, context, experience, and feedback, the DC can adjust the treatment within days, hours, minutes, or even seconds [85,86]. In our interviews, patients welcomed the idea of such functional flexibility. However, the challenge for the digital agent is to offer adaptive interventions that align with the respective physician's recommendations, comply with medical device regulations, and fulfill safety and performance requirements. Further studies must demonstrate that this type of adaptive intervention will improve treatment adherence.

The Digital Agent Should Measure and Monitor Patients' Adherence to Treatment and Provide Physicians With Easy-to-Read and Easy-to-Interpret Summaries

Measuring patients' adherence to treatment is a prerequisite for adaptive interventions [13]. Our interviews indicate scope for improvement regarding the measurement of treatment adherence—particularly for exercise and lifestyle changes. DCs are well suited to measure adherence based on objective data from sensors and subjective data such as chatbot conversations with patients. The interviewed physicians indicated that they would accept patient behavior monitoring if DCs aggregated the monitoring results and communicated them directly to EMRs. The literature also calls for this type of workflow integration [62,74,77,78]. However, the DC must be able to recognize red-flag situations and respond appropriately because the responsibility and workload of constantly monitoring the results cannot solely rely on physicians.

Further studies are needed to determine how patients respond to behavioral monitoring. The interviewed physicians anticipate positive effects, such as patients becoming "experts" on their condition, and adverse effects, such as patients relinquishing responsibility for their actions. Therefore, digital agents must monitor patients in a supportive manner and report the results in a form that assists rather than overloads the physician.

The Digital Agent Should React to Feedback and Questions From Patients in the Period Between Consultations

The more sophisticated the DC's communication and interaction skills are, the greater the expectation patients have for them to react appropriately. It is insufficient to simply give patients access to information through patient portals or PHRs [62,63] or have chatbots handling patient questions and feedback. In certain circumstances, patients still wish to talk to their human physician. In such cases, a triage mechanism might involve physicians only when necessary. However, the associated liability issues affecting the physicians (eg, in the case of suicidal intent) must be resolved.

Requirements for the Integration of Digital Experts and DCs

Only the integration of digital experts and DCs can unlock the full potential of these agents to support the entire consultation process for the mutual benefit of patients and physicians.

The Digital Agent Should Integrate Consultation Support (Digital Experts) and Patient Apps (DCs)

Integrating digital experts and DCs closes the loop from one consultation to the next and synergistically increases the benefits of both agents [108]. From a digital expert to a DC, personalized information about the diagnosis and treatment is transmitted immediately at the end of the consultation. This avoids media discontinuity, overcomes the problem of poor recall of recommendations or information, and allows patients to implement correct therapy immediately. Some of this functionality is already part of patient portals or PHRs [62,63]. However, making this information available in an mHealth app supported by digital agents allows for better interactivity, adherence support, and measurement. As access to information alone has not proven to be effective [64,65], the mHealth approach promises greater effectiveness. Adherence measurements are fed from the DC to the digital expert based on sensor data and patient-reported outcome measures (eg, diary entries and chatbot threads). This allows physicians to prepare for the next consultation and saves time because patients do not have to report verbally what they have already entered into the app. The interviewed physicians and patients welcomed this focus and time-saving measure, and the literature also calls for workflow integration along these lines [77,109-111].

Limitations

We derived the requirements for the design of digital agents to support consultation, adherence to treatment, and health literacy solely based on the statements obtained from our in-depth interviews with patients and physicians. Therefore, the 9 resulting requirements cannot be described as exhaustive. In particular, many necessary nonfunctional requirements are still lacking.

Furthermore, this study was conducted in Switzerland, which has one of the most expensive health care systems in the world. According to participating physicians, the standard consultation time is 20 minutes, which is significantly longer than that in many other countries. The responses from patients and physicians in other places and cultures might differ considerably. Further limitations may have arisen from the nature of a qualitative study based on a purposive sample. Although such a study results in a broad picture and deep insights, it may not be representative, not even for Switzerland. In addition, it is impossible to quantify the importance of the issues, suggested solutions, participant feedback, or the derived design requirements. For such purposes, surveys based on the insights obtained from this study are better suited. In addition, we cannot draw any conclusions related to specific user groups or medical disciplines. The fact that interview partners from very diverse backgrounds made similar observations and judgments indicates that our findings could be applied to various disciplines and user groups.

Conclusions and Future Studies

With the introduction of generative AI such as ChatGPT, the time for digital agents to support consultation, adherence to treatment, and health literacy may have arrived. There is enormous potential for patients and physicians to benefit from this new technology. Through in-depth interviews, both parties revealed their opinions about a silent and a communicative digital expert to support consultation and a DC to accompany patients between consultations. Their responses are synthesized into the following 9 requirements for the design of digital agents to support consultations.

The digital agent should do the following:

1. Make its role in the triadic consultation transparent
2. Encourage trust and support the physician-patient relationship while safeguarding physician credibility
3. Help physicians to focus on the patient during the consultation
4. Support physicians by taking over administrative duties
5. Tailor information and patient education to individual patient needs and preferences
6. Offer adaptive interventions for behavior change
7. Measure and monitor patient adherence to treatment and provide physicians with easy-to-read and easy-to-interpret summaries
8. React to feedback and questions from patients in the period between consultations
9. Integrate consultation support (digital experts) and patient apps (DCs).

Some recommendations for future studies were also offered in Requirements for Digital Experts During the Consultation section and Requirements for the DC Between Consultations section in the Discussion section. In addition, we suggest the following:

1. Obtain a complete set of requirements for the design of digital agents for consultation; a full requirement engineering approach would need to be followed and explored in the field. This would include an analysis of the technical feasibility and economic viability [104] of the system, with the results of this study serving as a starting point.
2. Depending on where the digital agents are to be deployed, this study could be replicated with local patients and physicians.

Acknowledgments

The authors thank Andrea Koppitz Prof Dr, Professor of Nursing, for reviewing the codebook and assisting them with their profound health care knowledge. Furthermore, the authors would like to thank Dominik Spescha and Daniel Degenhardt for their support in conducting, transcribing, and analyzing the interviews.

Authors' Contributions

AF was involved in conceptualization, methodology, validation, investigation, data curation, writing the original draft, review and editing, visualization, and project administration. CS was involved in writing the original draft and review and editing. PHS contributed to the investigation. MD was involved in conceptualization and review and editing. GS contributed to conceptualization; reviewing, editing, and rewriting some sections; and supervision.

Conflicts of Interest

None declared.

References

1. Costello KL. Impact of patient-provider communication on online health information behaviors in chronic illness. Proc Assoc Inf Sci Technol 2016 Dec 27;53(1):1-10. [doi: 10.1002/pra2.2016.14505301060]

2. Golder L, Jans C, Tschöpe S, Venetz A, Herzog N. Verändertes arbeitsumfeld und einstellung zu neuen finanzierungsmodellen: auswirkungen leistungsorientierung im gesundheitswesen erkennbar. gfs.bern. 2018 Jan. URL: https://www.fmh.ch/files/pdf20/2018_02_15_Begleitforschung_FMH_Langversion_SB_final.pdf [accessed 2024-02-27]
3. Kissling B, Ryser P. Die Ärztliche Konsultation Systemisch-Lösungsorientiert. Göttingen, Germany: Vandenhoeck & Ruprecht GmbH & Co KG; 2019.
4. Stewart M, Brown JB, Weston W, McWhinney IR, McWilliam CL, Freeman T. Patient-Centered Medicine: Transforming the Clinical Method, 3rd Edition. Boca Raton, FL: CRC Press; 2013.
5. Gafni A, Charles C, Whelan T. The physician-patient encounter: the physician as a perfect agent for the patient versus the informed treatment decision-making model. Soc Sci Med 1998 Aug;47(3):347-354. [doi: 10.1016/s0277-9536(98)00091-4] [Medline: 9681904]
6. Elwyn G, Frosch D, Thomson R, Joseph-Williams N, Lloyd A, Kinnersley P, et al. Shared decision making: a model for clinical practice. J Gen Intern Med 2012 Oct;27(10):1361-1367 [FREE Full text] [doi: 10.1007/s11606-012-2077-6] [Medline: 22618581]
7. Stiggelbout AM, Pieterse AH, De Haes JC. Shared decision making: concepts, evidence, and practice. Patient Educ Couns 2015 Oct;98(10):1172-1179. [doi: 10.1016/j.pec.2015.06.022] [Medline: 26215573]
8. Shay LA, Lafata JE. Where is the evidence? A systematic review of shared decision making and patient outcomes. Med Decis Making 2015 Jan;35(1):114-131 [FREE Full text] [doi: 10.1177/0272989X14551638] [Medline: 25351843]
9. Bomhof-Roordink H, Gärtner FR, Stiggelbout AM, Pieterse AH. Key components of shared decision making models: a systematic review. BMJ Open 2019 Dec 17;9(12):e031763 [FREE Full text] [doi: 10.1136/bmjopen-2019-031763] [Medline: 31852700]
10. Makoul G, Clayman ML. An integrative model of shared decision making in medical encounters. Patient Educ Couns 2006 Mar;60(3):301-312. [doi: 10.1016/j.pec.2005.06.010] [Medline: 16051459]
11. Kunneman M, Gionfriddo MR, Toloza FJ, Gärtner FR, Spencer-Bonilla G, Hargraves IG, et al. Humanistic communication in the evaluation of shared decision making: a systematic review. Patient Educ Couns 2019 Mar;102(3):452-466. [doi: 10.1016/j.pec.2018.11.003] [Medline: 30458971]
12. Street RLJ, Makoul G, Arora NK, Epstein RM. How does communication heal? Pathways linking clinician-patient communication to health outcomes. Patient Educ Couns 2009 Mar;74(3):295-301. [doi: 10.1016/j.pec.2008.11.015] [Medline: 19150199]
13. Sabaté E, World Health Organization. Adherence to Long-term Therapies: Evidence for Action. Geneva, Switzerland: World Health Organization; 2003.
14. Castro EM, Van Regenmortel T, Vanhaecht K, Sermeus W, Van Hecke A. Patient empowerment, patient participation and patient-centeredness in hospital care: a concept analysis based on a literature review. Patient Educ Couns 2016 Dec;99(12):1923-1939. [doi: 10.1016/j.pec.2016.07.026] [Medline: 27450481]
15. Lutfey KE, Wishner WJ. Beyond "compliance" is "adherence". Improving the prospect of diabetes care. Diabetes Care 1999 Apr;22(4):635-639. [doi: 10.2337/diacare.22.4.635] [Medline: 10189544]
16. Krass I, Schieback P, Dhippayom T. Adherence to diabetes medication: a systematic review. Diabet Med 2015 Jun;32(6):725-737. [doi: 10.1111/dme.12651] [Medline: 25440507]
17. Nieuwlaat R, Wilczynski N, Navarro T, Hobson N, Jeffery R, Keepanasseril A, et al. Interventions for enhancing medication adherence. Cochrane Database Syst Rev 2014 Nov 20;2014(11):CD000011 [FREE Full text] [doi: 10.1002/14651858.CD000011.pub4] [Medline: 25412402]
18. Rosli MR, Neoh CF, Wu DB, Hassan NW, Mahmud M, Rahimi A, et al. Evaluation of home medication review for patients with type 2 diabetes mellitus by community pharmacists: a randomised controlled trial. Pharm Pract (Granada) 2021;19(3):2397 [FREE Full text] [doi: 10.18549/PharmPract.2021.3.2397] [Medline: 34621450]
19. Kandula NR, Malli T, Zei CP, Larsen E, Baker DW. Literacy and retention of information after a multimedia diabetes education program and teach-back. J Health Commun 2011;16 Suppl 3:89-102. [doi: 10.1080/10810730.2011.604382] [Medline: 21951245]
20. Kravitz RL, Hays RD, Sherbourne CD. Recall of recommendations and adherence to advice among patients with chronic medical conditions. Arch Intern Med 1993 Aug 23;153(16):1869-1878. [doi: 10.1001/archinte.1993.00410160029002]
21. van Weert JC, Jansen J, Spreeuwenberg PM, van Dulmen S, Bensing JM. Effects of communication skills training and a question prompt sheet to improve communication with older cancer patients: a randomized controlled trial. Crit Rev Oncol Hematol 2011 Oct;80(1):145-159. [doi: 10.1016/j.critrevonc.2010.10.010] [Medline: 21075644]
22. Fallows RR, Hilsabeck RC. Comparing two methods of delivering neuropsychological feedback. Arch Clin Neuropsychol 2013 Mar;28(2):180-188. [doi: 10.1093/arclin/acs142] [Medline: 23315402]
23. Meth M, Calamia M, Tranel D. Does a simple intervention enhance memory and adherence for neuropsychological recommendations? Appl Neuropsychol Adult 2016;23(1):21-28. [doi: 10.1080/23279095.2014.996881] [Medline: 26111163]
24. Sherlock A, Brownie S. Patients' recollection and understanding of informed consent: a literature review. ANZ J Surg 2014 Apr;84(4):207-210. [doi: 10.1111/ans.12555] [Medline: 24812707]
25. Castro CM, Wilson C, Wang F, Schillinger D. Babel babble: physicians' use of unclarified medical jargon with patients. Am J Health Behav 2007;31 Suppl 1:S85-S95. [doi: 10.5555/ajhb.2007.31.supp.S85] [Medline: 17931142]

26. Schillinger D, Piette J, Grumbach K, Wang F, Wilson C, Daher C, et al. Closing the loop: physician communication with diabetic patients who have low health literacy. Arch Intern Med 2003 Jan 13;163(1):83-90. [doi: 10.1001/archinte.163.1.83] [Medline: 12523921]
27. McCarthy DM, Waite KR, Curtis LM, Engel KG, Baker DW, Wolf MS. What did the doctor say? Health literacy and recall of medical instructions. Med Care 2012 Apr;50(4):277-282 [FREE Full text] [doi: 10.1097/MLR.0b013e318241e8e1] [Medline: 22411440]
28. Tan SS, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
29. Ciciriello S, Johnston RV, Osborne RH, Wicks I, deKroo T, Clerehan R, et al. Multimedia educational interventions for consumers about prescribed and over-the-counter medications. Cochrane Database Syst Rev 2013 Apr 30(4):CD008416. [doi: 10.1002/14651858.CD008416.pub2] [Medline: 23633355]
30. Sim V, Galbraith K. Effectiveness of multimedia interventions in the provision of patient education on anticoagulation therapy: a review. Patient Educ Couns 2020 Oct;103(10):2009-2017. [doi: 10.1016/j.pec.2020.05.003] [Medline: 32532633]
31. Ab Malik N, Zhang J, Lam OL, Jin L, McGrath C. Effectiveness of computer-aided learning in oral health among patients and caregivers: a systematic review. J Am Med Inform Assoc 2017 Jan;24(1):209-217 [FREE Full text] [doi: 10.1093/jamia/ocw045] [Medline: 27274013]
32. Kuwabara A, Su S, Krauss J. Utilizing digital health technologies for patient education in lifestyle medicine. Am J Lifestyle Med 2019 Dec 13;14(2):137-142 [FREE Full text] [doi: 10.1177/1559827619892547] [Medline: 32231478]
33. Nealon J, Moreno A. Agent-based applications in health care. In: Moreno A, Nealon JL, editors. Applications of Software Agent Technology in the Health Care Domain. Basel, Switzerland: Birkhäuser; 2003.
34. Niederman F, Baker EW. Ethics and AI issues: old container with new wine? Inf Syst Front 2022 Jun 30;25:9-28. [doi: 10.1007/s10796-022-10305-1]
35. Feuerriegel S, Dolata M, Schwabe G. Fair AI. Bus Inf Syst Eng 2020 May 11;62:379-384. [doi: 10.1007/s12599-020-00650-3]
36. McMahan RP, Lai C, Pal SK. Interaction fidelity: the uncanny valley of virtual reality interactions. In: Proceedings of the 8th International Conference, VAMR 2016, Held as Part of HCI International 2016. 2016 Presented at: 8th International Conference, VAMR 2016, Held as Part of HCI International 2016; July 17-22, 2016; Toronto, ON. [doi: 10.1007/978-3-319-39907-2_6]
37. Mori M, MacDorman KF, Kageki N. The uncanny valley [from the field]. IEEE Robot Autom Mag 2012 Jun;19(2):98-100. [doi: 10.1109/MRA.2012.2192811]
38. Dolata M, Kilic M, Schwabe G. When a computer speaks institutional talk: exploring challenges and potentials of virtual assistants in face-to-face advisory services. In: Proceedings of the 52nd Hawaii International Conference on System Sciences (HICSS 2019). 2019 Presented at: HICSS 2019; January 8, 2019; Maui, Hawaii URL: https://www.zora.uzh.ch/id/eprint/173360/1/hicss_agent_v14_camera_zora.pdf [doi: 10.24251/hicss.2019.014]
39. Mazzi C, Ganguly P, Kidd M. Healthcare applications based on software agents. In: Proceedings of the 10th World Congress on Medical Informatics. 2001 Presented at: WCIM 2001; September 2-5, 2001; London, UK.
40. Linders GM, Vaitonytė J, Alimardani M, Mitev KO, Louwerse MM. A realistic, multimodal virtual agent for the healthcare domain. In: Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents. 2022 Presented at: IVA '22; September 6-9, 2022; Faro, Portugal. [doi: 10.1145/3514197.3551250]
41. Troy E, Doltani D, Harmon D. The role of a companion attending consultations with the patient. A systematic review. Ir J Med Sci 2019 Aug;188(3):743-750. [doi: 10.1007/s11845-018-1920-0] [Medline: 30374801]
42. Laidsaar-Powell RC, Butow PN, Bu S, Charles C, Gafni A, Lam WW, et al. Physician-patient-companion communication and decision-making: a systematic review of triadic medical consultations. Patient Educ Couns 2013 Apr;91(1):3-13. [doi: 10.1016/j.pec.2012.11.007] [Medline: 23332193]
43. Wolff JL, Roter DL. Family presence in routine medical visits: a meta-analytical review. Soc Sci Med 2011 Mar;72(6):823-831 [FREE Full text] [doi: 10.1016/j.socscimed.2011.01.015] [Medline: 21353358]
44. Bracher M, Stewart S, Reidy C, Allen C, Townsend K, Brindle L. Partner involvement in treatment-related decision making in triadic clinical consultations - a systematic review of qualitative and quantitative studies. Patient Educ Couns 2020 Feb;103(2):245-253. [doi: 10.1016/j.pec.2019.08.031] [Medline: 31477515]
45. Pearce C, Dwan K, Arnold M, Phillips C, Trumble S. Doctor, patient and computer--a framework for the new consultation. Int J Med Inform 2009 Jan;78(1):32-38. [doi: 10.1016/j.ijmedinf.2008.07.002] [Medline: 18752989]
46. Pearce C, Arnold M, Phillips C, Trumble S, Dwan K. The patient and the computer in the primary care consultation. J Am Med Inform Assoc 2011;18(2):138-142 [FREE Full text] [doi: 10.1136/jamia.2010.006486] [Medline: 21262923]
47. Mørck P, Langhoff TO, Christophersen M, Møller AK, Bjørn P. Variations in oncology consultations: how dictation allows variations to be documented in standardized ways. Comput Supported Coop Work 2018 May 18;27:539-568. [doi: 10.1007/s10606-018-9332-2]
48. Barrows RC, Clayton PD. Privacy, confidentiality, and electronic medical records. J Am Med Inform Assoc 1996;3(2):139-148 [FREE Full text] [doi: 10.1136/jamia.1996.96236282] [Medline: 8653450]
49. Keshta I, Odeh A. Security and privacy of electronic health records: concerns and challenges. Egypt Inform J 2021 Jul;22(2):177-183. [doi: 10.1016/j.eij.2020.07.003]

50. Pearce C, Sandoval M. Consulting with a computer: new frontiers. Aust J Gen Pract 2020 Sep;49(9):612-614 [FREE Full text] [doi: 10.31128/AJGP-05-20-5400] [Medline: 32864673]
51. Rose D, Richter LT, Kapustin J. Patient experiences with electronic medical records: lessons learned. J Am Assoc Nurse Pract 2014 Dec;26(12):674-680 [FREE Full text] [doi: 10.1002/2327-6924.12170] [Medline: 25234112]
52. Asan O, Young HN, Chewning B, Montague E. How physician electronic health record screen sharing affects patient and doctor non-verbal communication in primary care. Patient Educ Couns 2015 Mar;98(3):310-316 [FREE Full text] [doi: 10.1016/j.pec.2014.11.024] [Medline: 25534022]
53. Alkureishi MA, Lee WW, Lyons M, Press VG, Imam S, Nkansah-Amankra A, et al. Impact of electronic medical record use on the patient-doctor relationship and communication: a systematic review. J Gen Intern Med 2016 May;31(5):548-560 [FREE Full text] [doi: 10.1007/s11606-015-3582-1] [Medline: 26786877]
54. Gawande A. Why doctors hate their computers. The New Yorker. 2018 Nov 5. URL: https://www.newyorker.com/magazine/2018/11/12/why-doctors-hate-their-computers [accessed 2019-02-18]
55. Frankel R, Altschuler A, George S, Kinsman J, Jimison H, Robertson NR, et al. Effects of exam-room computing on clinician-patient communication: a longitudinal qualitative study. J Gen Intern Med 2005 Aug;20(8):677-682 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.0163.x] [Medline: 16050873]
56. Kazmi Z. Effects of exam room EHR use on doctor-patient communication: a systematic literature review. Inform Prim Care 2013;21(1):30-39 [FREE Full text] [doi: 10.14236/jhi.v21i1.37] [Medline: 24629654]
57. Scalia P, Durand MA, Berkowitz JL, Ramesh NP, Faber MJ, Kremer JA, et al. The impact and utility of encounter patient decision aids: systematic review, meta-analysis and narrative synthesis. Patient Educ Couns 2019 May;102(5):817-841. [doi: 10.1016/j.pec.2018.12.020] [Medline: 30612829]
58. Stacey D, Légaré F, Lewis K, Barry MJ, Bennett CL, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2017 Apr 12;4(4):CD001431 [FREE Full text] [doi: 10.1002/14651858.CD001431.pub5] [Medline: 28402085]
59. Scalia P, Tighe KM, Elwyn G, Bagley PJ, Blunt HB, Boh B, et al. A systematic review of decision aids for gender affirming therapy. Transl Androl Urol 2021 Jun;10(6):2574-2582 [FREE Full text] [doi: 10.21037/tau-20-1000] [Medline: 34295744]
60. Agoritsas T, Heen AF, Brandt L, Alonso-Coello P, Kristiansen A, Akl EA, et al. Decision aids that really promote shared decision making: the pace quickens. BMJ 2015 Feb 10;350:g7624 [FREE Full text] [doi: 10.1136/bmj.g7624] [Medline: 25670178]
61. Elwyn G, Lloyd A, Joseph-Williams N, Cording E, Thomson R, Durand MA, et al. Option Grids: shared decision making made easier. Patient Educ Couns 2013 Feb;90(2):207-212. [doi: 10.1016/j.pec.2012.06.036] [Medline: 22854227]
62. Irizarry T, DeVito Dabbs A, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
63. Dameff C, Clay B, Longhurst CA. Personal health records: more promising in the smartphone era? JAMA 2019 Jan 29;321(4):339-340. [doi: 10.1001/jama.2018.20434] [Medline: 30633300]
64. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res 2012 Nov 26;14(6):e162 [FREE Full text] [doi: 10.2196/jmir.2238] [Medline: 23183044]
65. Goldzweig CL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Miake-Lye I, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med 2013 Nov 19;159(10):677-687. [doi: 10.7326/0003-4819-159-10-201311190-00006] [Medline: 24247673]
66. mHealth economics 2017 – current status and future trends in mobile health. Research2Guidance. URL: https://research2guidance.com/product/mhealth-economics-2017-current-status-and-future-trends-in-mobile-health/ [accessed 2021-04-12]
67. Garabedian LF, Ross-Degnan D, Wharam JF. Mobile phone and smartphone technologies for diabetes care and self-management. Curr Diab Rep 2015 Dec;15(12):109 [FREE Full text] [doi: 10.1007/s11892-015-0680-8] [Medline: 26458380]
68. Holmen H, Wahl AK, Cvancarova Småstuen M, Ribu L. Tailored communication within mobile apps for diabetes self-management: a systematic review. J Med Internet Res 2017 Jun 23;19(6):e227 [FREE Full text] [doi: 10.2196/jmir.7045] [Medline: 28645890]
69. McMillan KA, Kirk A, Hewitt A, MacRury S. A systematic and integrated review of mobile-based technology to promote active lifestyles in people with type 2 diabetes. J Diabetes Sci Technol 2017 Mar;11(2):299-307 [FREE Full text] [doi: 10.1177/1932296816656018] [Medline: 27334301]
70. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
71. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]

72. Fukuoka Y, Gay C, Haskell W, Arai S, Vittinghoff E. Identifying factors associated with dropout during prerandomization run-in period from an mHealth physical activity education study: the mPED trial. JMIR Mhealth Uhealth 2015 Apr 13;3(2):e34 [FREE Full text] [doi: 10.2196/mhealth.3928] [Medline: 25872754]
73. Lobelo F, Kelli HM, Tejedor SC, Pratt M, McConnell MV, Martin SS, et al. The wild wild west: a framework to integrate mHealth software applications and wearables to support physical activity assessment, counseling and interventions for cardiovascular disease risk reduction. Prog Cardiovasc Dis 2016;58(6):584-594 [FREE Full text] [doi: 10.1016/j.pcad.2016.02.007] [Medline: 26923067]
74. Zhang Y, Li X, Luo S, Liu C, Liu F, Zhou Z. Exploration of users' perspectives and needs and design of a type 1 diabetes management mobile app: mixed-methods study. JMIR Mhealth Uhealth 2018 Sep 21;6(9):e11400 [FREE Full text] [doi: 10.2196/11400] [Medline: 30249580]
75. Gordon WJ, Landman A, Zhang H, Bates DW. Beyond validation: getting health apps into clinical practice. NPJ Digit Med 2020;3:14 [FREE Full text] [doi: 10.1038/s41746-019-0212-z] [Medline: 32047860]
76. Heidel A, Hagist C. Potential benefits and risks resulting from the introduction of health apps and wearables into the German statutory health care system: scoping review. JMIR Mhealth Uhealth 2020 Sep 23;8(9):e16444 [FREE Full text] [doi: 10.2196/16444] [Medline: 32965231]
77. Street RLJ, Liu L, Farber NJ, Chen Y, Calvitti A, Zuest D, et al. Provider interaction with the electronic health record: the effects on patient-centered communication in medical encounters. Patient Educ Couns 2014 Sep;96(3):315-319 [FREE Full text] [doi: 10.1016/j.pec.2014.05.004] [Medline: 24882086]
78. Cahn A, Akirov A, Raz I. Digital health technology and diabetes management. J Diabetes 2018 Jan;10(1):10-17. [doi: 10.1111/1753-0407.12606] [Medline: 28872765]
79. Nurgalieva L, O'Callaghan D, Doherty G. Security and privacy of mHealth applications: a scoping review. IEEE Access 2020;8:104247-104268. [doi: 10.1109/access.2020.2999934]
80. Baker DW, Wolf MS, Feinglass J, Thompson JA, Gazmararian JA, Huang J. Health literacy and mortality among elderly persons. Arch Intern Med 2007 Jul 23;167(14):1503-1509. [doi: 10.1001/archinte.167.14.1503] [Medline: 17646604]
81. Magnani JW, Mujahid MS, Aronow HD, Cené CW, Dickson VV, Havranek E, et al. Health literacy and cardiovascular disease: fundamental relevance to primary and secondary prevention: a scientific statement from the American Heart Association. Circulation 2018 Jul 10;138(2):e48-e74 [FREE Full text] [doi: 10.1161/CIR.0000000000000579] [Medline: 29866648]
82. Michou M, Panagiotakos DB, Costarelli V. Low health literacy and excess body weight: a systematic review. Cent Eur J Public Health 2018 Sep;26(3):234-241 [FREE Full text] [doi: 10.21101/cejph.a5172] [Medline: 30419628]
83. N'Goran AA, Pasquier J, Deruaz-Luyet A, Burnand B, Haller DM, Neuner-Jehle S, et al. Factors associated with health literacy in multimorbid patients in primary care: a cross-sectional study in Switzerland. BMJ Open 2018 Feb 13;8(2):e018281 [FREE Full text] [doi: 10.1136/bmjopen-2017-018281] [Medline: 29440210]
84. Cramer JA. Consequences of intermittent treatment for hypertension: the case for medication compliance and persistence. Am J Manag Care 1998 Nov;4(11):1563-1568 [FREE Full text] [Medline: 10338902]
85. Collins LM. Optimization of Behavioral, Biobehavioral, and Biomedical Interventions: The Multiphase Optimization Strategy (MOST). Cham, Switzerland: Springer International Publishing; 2018.
86. Hardeman W, Houghton J, Lane K, Jones A, Naughton F. A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. Int J Behav Nutr Phys Act 2019 Apr 03;16(1):31 [FREE Full text] [doi: 10.1186/s12966-019-0792-7] [Medline: 30943983]
87. Dahodwala M, Geransar R, Babion J, de Grood J, Sargious P. The impact of the use of video-based educational interventions on patient outcomes in hospital settings: a scoping review. Patient Educ Couns 2018 Dec;101(12):2116-2124. [doi: 10.1016/j.pec.2018.06.018] [Medline: 30087021]
88. Miller CK. Adaptive intervention designs to promote behavioral change in adults: what is the evidence? Curr Diab Rep 2019 Jan 25;19(2):7. [doi: 10.1007/s11892-019-1127-4] [Medline: 30684109]
89. Gross T, Koch M. Computer-supported cooperative work - concepts and trends. In: Proceedings of the 11th International Conference of the Association Information and Management (AIM). 2006 Presented at: AIM 2006; June 8-9, 2006; Luxembourg.
90. Wilson P. Computer supported cooperative work: an overview. Intell Tutor Media 1990 Jan;1(3):103-116. [doi: 10.1080/14626269009408268]
91. Grudin J. Computer-supported cooperative work: history and focus. Computer 1994 May;27(5):19-26. [doi: 10.1109/2.291294]
92. Carroll JM, Rosson MB, Convertino G, Ganoe CH. Awareness and teamwork in computer-supported collaborations. Interact Comput 2006 Jan;18(1):21-46. [doi: 10.1016/j.intcom.2005.05.005]
93. Dolata M, Schwabe G. Paper practices in institutional talk: how financial advisors impress their clients. Comput Support Coop Work 2017 Jun 16;26(4-6):769-805. [doi: 10.1007/s10606-017-9279-8]
94. Willis M. I'm trying to find my way of staying organized: the socio-technical assemblages of personal health information management. Comput Support Coop Work 2019 May 15;28(6):1073-1102. [doi: 10.1007/s10606-019-09346-z]
95. Stebbins RA. Exploratory Research in the Social Sciences. Thousand Oaks, CA: SAGE Publications; 2001.

96. Nicholls CM, Lewis J, Ritchie J, Ormston R. Qualitative Research Practice: A Guide for Social Science Students and Researchers. Thousand Oaks, CA: SAGE Publications; 2003.
97. Färber A, Zigan N, Dolata M, Stalder P, Koppitz A, Schwabe G. The digital transformation of physician–patient consultations: identifying problems and approaches to improve adherence. In: Proceedings of the 52nd Hawaii International Conference on System Sciences. 2019 Presented at: HICSS 2019; January 8-11, 2019; Maui, Hawaii. [doi: 10.24251/hicss.2019.500]
98. Merriam SB, Grenier RS. Qualitative Research in Practice: Examples for Discussion and Analysis. Hoboken, NJ: John Wiley & Sons; 2019.
99. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
100. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
101. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm Policy Ment Health 2015 Sep;42(5):533-544 [FREE Full text] [doi: 10.1007/s10488-013-0528-y] [Medline: 24193818]
102. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
103. Gordon AV, Rohrbeck R, Schwarz JO. Escaping the ‘faster horses’ trap: bridging strategic foresight and design-based innovation. Technol Innov Manag Rev 2019 Sep;9(8):30-42 [FREE Full text] [doi: 10.22215/timreview/1259]
104. Dolata M, Schwabe G. Design thinking in IS Research Projects. In: Brenner W, Uebernickel F, editors. Design Thinking for Innovation. Cham, Switzerland: Springer; 2016.
105. Mayring P. Qualitative Inhaltsanalyse. Grundlagen und Techniken. Weinheim, Germany: Deutscher Studien Verlag; 2000.
106. Birjali M, Kasri M, Beni-Hssane A. A comprehensive survey on sentiment analysis: approaches, challenges and trends. Knowl Based Syst 2021 Aug;226:107134. [doi: 10.1016/j.knosys.2021.107134]
107. Schnitman G, Wang T, Kundu S, Turkdogan S, Gotlieb R, How J, et al. The role of digital patient education in maternal health: a systematic review. Patient Educ Couns 2022 Mar;105(3):586-593. [doi: 10.1016/j.pec.2021.06.019] [Medline: 34183217]
108. Färber A, de Spindler A, Moser A, Schwabe G. Closing the loop for patients with chronic diseases - from problems to a solution architecture. In: Proceedings of the IEEE 11th International Conference on Healthcare Informatics (ICHI). 2023 Presented at: ICHI 2023; June 26-29, 2023; Houston, TX, USA URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10337012 [doi: 10.1109/ICHI57859.2023.00055]
109. Hämäläinen A, Hirvonen H. Electronic health records reshaping the socio-technical practices in long-term care of older persons. Technol Soc 2020 Aug;62:101316. [doi: 10.1016/j.techsoc.2020.101316]
110. Janett RS, Yeracaris PP. Electronic medical records in the American health system: challenges and lessons learned. Cien Saude Colet 2020 Mar;25(4):1293-1304 [FREE Full text] [doi: 10.1590/1413-81232020254.28922019] [Medline: 32267432]
111. Wisner K, Lyndon A, Chesla CA. The electronic health record's impact on nurses' cognitive work: an integrative review. Int J Nurs Stud 2019 Jun;94:74-84. [doi: 10.1016/j.ijnurstu.2019.03.003] [Medline: 30939418]

Abbreviations

AI: artificial intelligence
COREQ: Consolidated Criteria for Reporting Qualitative Research
CSCW: computer-supported cooperative work
DC: digital companion
EMR: electronic medical record
mHealth: mobile health
PDA: patient decision aid
PHR: personal health record

Edited by B Price; submitted 04.06.23; peer-reviewed by L Weinert, Y Zhang, C Yu, C Baur; comments to author 04.08.23; revised version received 09.12.23; accepted 15.01.24; published 18.03.24.

Please cite as:
Färber A, Schwabe C, Stalder PH, Dolata M, Schwabe G
Physicians' and Patients' Expectations From Digital Agents for Consultations: Interview Study Among Physicians and Patients
JMIR Hum Factors 2024;11:e49647
URL: https://humanfactors.jmir.org/2024/1/e49647
doi:10.2196/49647
PMID:38498022

Original Paper

Preferences for mHealth Intervention to Address Mental Health Challenges Among Men Who Have Sex With Men in Nepal: Qualitative Study

Kamal Gautam[1*], MPH; Camille Aguilar[2*]; Kiran Paudel[1,3], BPH; Manisha Dhakal[4], MBS; Jeffrey A Wickersham[5], PhD; Bibhav Acharya[6,7], MD, PhD; Sabitri Sapkota[7], MPH, PhD; Keshab Deuba[8], PhD; Roman Shrestha[1,5], MPH, PhD

[1]Department of Allied Health Sciences, University of Connecticut, Storrs, CT, United States

[2]School of Nursing and Health Studies, University of Miami, Miami, FL, United States

[3]Nepal Health Frontiers, Tokha-5, Kathmandu, Nepal

[4]Blue Diamond Society, Dhumbarahi Height, Kathmandu, Nepal

[5]Department of Internal Medicine, Section of Infectious Diseases, Yale School of Medicine, New Haven, CT, United States

[6]Department of Psychiatry and Behavioral Sciences, School of Medicine, UCSF Weill Institute for Neurosciences, San Francisco, CA, United States

[7]Possible, Kathmandu, Nepal

[8]Department of Global Public Health, Karolinska Institutet, Stockholm, Sweden

[*]these authors contributed equally

Corresponding Author:
Roman Shrestha, MPH, PhD
Department of Allied Health Sciences
University of Connecticut
358 Mansfield Road
Storrs, CT, 06269
United States
Phone: 1 860 486 2834
Email: roman.shrestha@uconn.edu

Abstract

Background: Men who have sex with men (MSM) are disproportionately burdened by poor mental health. Despite the increasing burden, evidence-based interventions for MSM are largely nonexistent in Nepal.

Objective: This study explored mental health concerns, contributing factors, barriers to mental health care and support, and preferred interventions to improve access to and use of mental health support services among MSM in Nepal.

Methods: We conducted focus groups with MSM in Kathmandu, Nepal, in January 2023. In total, 28 participants took part in 5 focus group sessions. Participants discussed several topics related to the mental health issues they experienced, factors contributing to these issues, and their suggestions for potential interventions to address existing barriers. The discussions were recorded, transcribed, and analyzed using Dedoose (version 9.0.54; SocioCultural Research Consultants, LLC) software for thematic analysis.

Results: Participants reported substantial mental health problems, including anxiety, depression, suicidal ideation, and behaviors. Contributing factors included family rejection, isolation, bullying, stigma, discrimination, and fear of HIV and other sexually transmitted infections. Barriers to accessing services included cost, lack of lesbian, gay, bisexual, transgender, intersex, queer, and asexual (LGBTIQA+)–friendly providers, and the stigma associated with mental health and sexuality. Participants suggested a smartphone app with features such as a mental health screening tool, digital consultation, helpline number, directory of LGBTIQA+-friendly providers, mental health resources, and a discussion forum for peer support as potential solutions. Participants emphasized the importance of privacy and confidentiality to ensure mobile apps are safe and accessible.

Conclusions: The findings of this study have potential transferability to other low-resource settings facing similar challenges. Intervention developers can use these findings to design tailored mobile apps to facilitate mental health care delivery and support for MSM and other marginalized groups.

(JMIR Hum Factors 2024;11:e56002) doi:10.2196/56002

KEYWORDS

mental health; MSM; mHealth; smartphone apps; digital health; Nepal; gay; homosexual; homosexuality; men who have sex with men; focus group; focus groups; qualitative; barrier; barriers; thematic; mHealth; mobile health; app; apps; applications; applications

Introduction

Gay, bisexual, and other men who have sex with men (MSM) have poorer mental health and experience more mental distress than their cisgender heterosexual counterparts [1-3]. Studies have shown a high proportion of MSM's experiences such as mood swings, disordered eating behavior, anxiety disorder, depression, suicidal ideation and behaviors, substance abuse, and body image disorders [4-7]. A recent systematic review and meta-analysis found that the prevalence of depression among MSM in Asia was 37% [6]. These mental health issues experienced by MSM are often linked to stressors triggered by a homophobic environment, particularly due to their sexual orientation [8].

In the context of Nepal, homosexuality is not criminalized, and the rights of MSM are guaranteed by the constitution [9,10]. Despite these legal safeguards, the prevailing cultural norms and societal attitudes pose significant challenges. Traditional and cultural values emphasize heterosexual marriages and family structures and traditional expectations of relationships, and a lack of family support often marginalizes individuals with diverse sexual orientations [11]. These social and cultural characteristics create a heteronormative and stigmatizing environment for MSM, which is detrimental to their mental health. Past studies have found that a very high number of MSM in Nepal had clinically significant depression (54%) and lifetime prevalence of suicidal thoughts (26%) [12,13]. Despite these dire mental health statistics, MSM encounter barriers in accessing health care, particularly mental health services, due to social stigma, discrimination, financial constraints, and insensitivity among health care providers [11,12,14-17]. These barriers to seeking mental health and psychosocial support among MSM, who not only have the highest needs but also the highest unmet needs, give rise to health disparities in this population. In order to reduce these disparities, improving access is crucial for advancing their overall health and well-being.

Mobile health (mHealth), especially mobile apps, offers a promising solution to bridge this gap. It can offer tailored and cost-effective interventions without the need for in-person contact and can provide convenience, improve mental health literacy and easy accessibility, eliminate travel hassles, and encourage help-seeking behavior [18,19]. With Nepal experiencing significant growth in mobile phone ownership of 96% and over 70% using the internet through smartphones, mobile app–based interventions tailored to the needs of MSM in Nepal are potentially feasible [20]. Recognizing the potential of mHealth, we conducted this study to (1) identify the mental health challenges and barriers to accessing mental health and psychosocial support services among MSM and (2) understand their preferences for smartphone apps (eg, functionality, format, design, and attributes) that could enable their access to mental health and psychosocial support services access.

Methods

Study Setting and Recruitment

This qualitative study is part of a larger HIV biobehavioral survey that was conducted among 250 MSM participants in Kathmandu, Nepal [13]. Five focus group (FG) sessions were conducted with MSM participants in January 2023. Four of these sessions included 6 MSM participants in each, while the remaining session had 4 participants (N=28). FG sessions were conducted until a point of theoretical saturation was achieved. Eligibility criteria for participation included: (1) 18 years or older, (2) self-identified as cisgender MSM, and (3) proficiency in Nepali or English.

Participants were recruited using respondent-driven sampling, a network-based sampling method often used for hard-to-reach populations. The recruitment chain was initiated with 5 MSM "seeds," purposively selected based on recommendations from a community-based organization providing services to MSM. Each seed who completed the interviewer-administered questionnaire was given 5 recruitment coupons to recruit potential peers. Subsequent participants were, in turn, given 5 coupons to recruit additional peers. In total, 28 (~11%) of the survey participants were randomly selected for the FG sessions.

Study Procedure

FG sessions were conducted inside the community-based organization's office and lasted about 90 minutes. A semistructured FG topic guide with appropriate probes was developed that guided the discussion. A trained facilitator led the FG sessions, and a cofacilitator took the notes. Both the facilitator and cofacilitator identified themselves as MSM.

Before the discussion, participants completed an interviewer-administered Qualtrics survey that included sociodemographic, sexual health, alcohol, smoking, violence, and mental health–related questions. The participants' exposure to violence was assessed using the 4-item Hurt, Insult, Threat, and Scream screening tool, using a 5-point frequency format (scores 4-20). Final scores were classified as normal (0-10) or violence (11-25) [21]. Depressive symptoms were evaluated with the Patient Health Questionnaire instrument, scoring each of the 9 *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition, criteria (0-3). A composite score of 0-27 was computed, with a score exceeding 10 indicating moderate to severe depressive symptoms [22].

The FGs involved questions and discussions about traumatic life events. Participants were made aware that they did not have to answer any questions that they felt were distressing and could leave the FG session at any time if they felt uncomfortable. A study team member was also present at all 5 FG sessions to refer to a counselor or provide any additional support needed in the case of a distressing situation. While conducting the FG sessions, a trained facilitator approached participants sensitively, respecting moments of silence and their willingness to continue discussions—statements like "I am fine" or "we can continue" followed silence. Despite the sensitive topics discussed, none of the participants requested support, including speaking with counselors. At the end of all FG sessions, participants also disclosed that they were glad to have had the opportunity to share their experiences.

Data Analysis

SPSS (version 29.0.0 software; IBM Corp) was used to calculate descriptive statistics (frequencies and percentages) for the variables collected via a Qualtrics survey. FG transcripts were transcribed and checked for accuracy before coding. The 2 coders (KG and CA) read and reread transcriptions to identify key ideas and recurring themes. A codebook was developed with mutually agreed-upon codes derived from the FG transcripts, and coding was completed independently by 2 researchers (KG and CA). To ensure reliability, codes were constantly compared for agreement and discussed between the coders, and the senior author (RS) cross-checked all codes. Dedoose (version 9.0.54) was used for data management and analysis. The themes were gathered as child codes and then placed into a broad category as root codes. Each theme with its qualitative quotes to best illustrate the findings are presented in the results section.

Ethical Considerations

The study protocol was approved by the institutional review boards at the University of Connecticut (H22-0039) and the Nepal Health Research Council (2391-2022 P). All the participants provided verbal informed consent before their participation. Participants were explained the importance of maintaining the confidentiality of FGs and requested not to discuss the experiences and comments shared during the FGs with others. All the sessions were conducted in Nepali and were audio recorded, transcribed, and translated. Participants were compensated NRs 1000 (~US $8) for their time and participation. FG transcripts were deidentified before the analysis, and the survey data were anonymous.

Results

Participant Characteristics

Table 1 provides information on participants' characteristics. The mean age of study participants was 25.3 (SD 6.1) years. Most of the 28 participants were Hindu (n=22, 79%), had a high school or higher degree (n=21, 75%), and identified as gay (n=22, 79%). A total of 21% (n=6) of participants had depressive symptoms, and 14% (n=4) had experienced violence in their life. A little over half (n=15, 54%) of participants had used health-related mobile apps, and almost 90% (n=25) used digital devices to search for health-related information.

Table 1. Participant characteristics (N=28).

Sociodemographic factors	Values
Age (years), mean (SD)	25.3 (6.1)
Religion, n (%)	
Hindu	22 (79)
Buddhist	5 (18)
Others	1 (4)
Level of education, n (%)	
Up to grade 10	7 (25)
High school and above	21 (75)
Employment, n (%)	
No	15 (54)
Yes	13 (46)
Income level, n (%)	
Less than NRs 20,000 (~US $150)	12 (43)
NRs 20,000 (~US $150) and above	16 (57)
Sexual orientation, n (%)	
Gay	22 (77)
Bisexual	6 (21)
Relationship status, n (%)	
Single	19 (68)
With partner	9 (32)
Depressive symptoms, n (%)	
No	22 (79)
Yes	6 (21)
Ever experienced violence, n (%)	
No	24 (86)
Yes	4 (14)
Daily smoker, n (%)	
No	5 (18)
Yes	23 (82)
Alcohol use (past 12 months), n (%)	
No	6 (21)
Yes	12 (79)
HIV status, n (%)	
Positive	1 (4)
Negative	27 (96)
Syphilis status, n (%)	
Positive	8 (29)
Negative	20 (71)
Engaged in anal sex (past 6 months), n (%)	
No	6 (21)
Yes	22 (79)
Condomless sex (past 6 months), n (%)	

Sociodemographic factors	Values
No	11 (39)
Yes	17 (61)
Sexual partners in (past 6 months), n (%)	
Single	17 (61)
Multiple	11 (39)
Engagement in group sex (past 6 months), n (%)	
No	26 (93)
Yes	2 (7)
Engagement in sex work (past 6 months), n (%)	
No	26 (93)
Yes	2 (7)
Has any health insurance, n (%)	
No	24 (86)
Yes	4 (14)
Use of health-related apps in mobile, n (%)	
No	13 (46)
Yes	15 (54)
Use of mobile or technological devices to search for health-related information, n (%)	
No	3 (11)
Yes	25 (89)

FG Results

Overview

Throughout the data analysis, 3 overarching themes emerged in the codebook with their own subthemes (Multimedia Appendix 1): (1) mental health challenges, (2) barriers to accessing mental health services, and (3) preference for mental health mobile apps with desired features and attributes.

Mental Health Challenges

Mental health challenges faced by the participants involve a multifaceted interaction of factors, including sexual orientation, emotional distress, stigma, discrimination and victimization, and social exclusion. Moreover, they frequently encounter barriers to accessing support services that could enhance their mental well-being. Participants not only vividly described their day-to-day challenges but also shared insights into the collective experiences of the lesbian, gay, bisexual, transgender, intersex, queer, and asexual (LGBTIQA+) community. Their comprehensive perspective underscored the profound impact of prevailing societal biases on their mental well-being (Textbox 1).

A constant fear of societal judgment and family pressure to conform to traditional gender norms has intensified issues like anxiety, depression, and suicidal thoughts. Participants in all FGs highlighted the pressure to enter heterosexual marriages, causing emotional turmoil as they navigate their identities and societal expectations (Textbox 2).

Participants disclosed coping mechanisms, such as drug use, drinking alcohol, smoking, engaging in sexual risk behaviors (eg, multiple sex partners), and self-harm. These strategies were described as providing temporary relief from the immense emotional turmoil they experience.

> *I have personally known someone who started risky sexual behavior from a young age because that was how they felt validated. They wanted others to make them feel better. So, they would often engage in multiple sexual encounters, thinking it would help them cope with their struggles... I also know people who turned to drugs and alcohol to cope with themselves.* [18-year-old participant from FG1]

> *…due to tension and mental pressure, it was tough for me to control myself, so I started to cut my hands with a razor; I did it many times. I was also thinking of taking tablets for suicide.* [30-year-old participant from FG3]

Several participants talked about and shared their experiences of intense anxiety and fear surrounding the possibility of contracting HIV and other sexually transmitted infections following sexual encounters with their partners.

> *I have extreme fear about whether I contracted it [HIV] or not… even the close friends I know have contracted HIV, and because of that, I also have a fear and anxiety of whether I contracted HIV or not after the sex is done.* [21-year-oldparticipant from FG1]

Textbox 1. Social acceptance and lack of family support heighten mental health challenges.

- "…there is no one, and when we open up there is no family support. Family supporting queer people, it is like gold which is rare. We only open out in this [LGBTIQA+] community; You can imagine how bad is our mental status and the situation." [26-year-old participant from FG2]
- "…because of my sexuality, sometimes I suffer from social anxiety, 'are they judging me because of my looks, voice or the way I dress." [24-years-old participant from FG5]
- "I study in 12th grade, and most of the time, I am bullied by my male classmates… even the teachers ask, "Why do you act like a girl?" And most of them do know I use TikTok, and everyone knows about me, so I think bullying is also another part, and I think mental health or stress is a common occurrence for everyone in LGBTIQA+ people." [22-year-old participant from FG1]

Textbox 2. Mental health challenges from societal pressure and identity concealment.

- "…if I have to be 'me' or do something feminine, then there is a fear of being judged by other people, so I have to pretend as a closeted man, I have to pretend masculine, have a masculine voice, and all the stereotypes and stigmas the people in the community have kept, which fuels the anxiety." [22-year-old participant from FG1]
- "I faced immense pressure from my family about marriage, being the only son. I kept my sexual orientation hidden, making it harder in our community. The talks of marriage became unbearable… I felt so distressed that I left home and was even suicidal. My relatives found out, but the misunderstanding about my identity remained... these struggles took a toll on my mental health, forcing me to search for ways to cope and maintain my well-being." [28-year-old participant from FG3]
- "I was so low that all I had in mind was suicide." [29-year-old participant from FG2]

Participants also shared that they fear the potential disclosure of their HIV status because they anticipate that others may treat them differently after learning their status.

> *When it comes to HIV, if a person status has HIV, he is afraid that if his status leaks, then people will look differently.* [23-year-old participant from FG4]

One participant recounted how their colleague, upon learning their HIV-positive status, tragically died by suicide, underscoring the emotional toll and mental health challenges.

> *There was one colleague of mine who died by suicide as soon as the HIV test result came back positive.* [25-year-old participant from FG1]

Barriers to Accessing Mental Health Services

Participants shared that many gay, bisexual, and MSM do not seek mental health services because they perceive themselves as mentally healthy and believe their lives are going well, leading them to overlook the need for such support.

> *The reason I believe that our community members do not seek mental health services is because they think they are alright, that their life is just going on, they think they are fine and healthy and feel they don't need such services.* [25-year-old participant from FG3]

Participants also shared that individuals tend to become more open and willing to seek help if they are aware of mental health services like counseling and therapy.

> *…if people are aware of counseling and therapy, people will be more willing to go there.* [24-year-old participant from FG2]

Participants discussed that individuals still closeted about their identity find it challenging to trust others, creating a communication barrier. Their hesitancy to trust stems from a history of hiding aspects of themselves, hindering open communication and sharing true feelings and experiences.

> *It is hard for people to trust. There is also a communication barrier because they are still closeted and grew up hiding things from the beginning. If the person themselves is not trusting them, then how can they trust the person in front of them.* [35-year-old participant from FG2]

Stigma and discrimination associated with mental health and sexuality were major concerns for participants. Many participants brought up fears of being labeled "*pagal*" (a pejorative that is closest to "crazy" in English) as a barrier to accessing mental health services.

> *…there is a stigma against mental health, that is the reason we do not seek mental health services. If we visit a health care center, then people will talk about it, and the peer groups and society will think of us as pagal (crazy); they will say that this person is taking medicines, so that is another reason we do not visit mental health care centers.* [26-year-old participant from FG3]

Others discussed the impact that homophobia can have on MSM seeking mental health services. Homophobia and heterosexism still exist in Nepal's society and can have significant impacts on MSM decisions.

> *A stereotypical saying "how can men like men?" is still prevalent in society, so, to not get judged by others, people don't attend these [mental health] sessions.* [29-year-old participant from FG2]

Many participants expressed their frustration with medical professionals who, instead of addressing their health concerns seriously, tend to label them, dismiss their issues, and attribute symptoms to perceived psychological factors such as overthinking, thereby hindering their access to necessary services.

> *Often, the doctor calls us with names, gives us a tag, they do not give us a priority, they only say "there's*

> *nothing wrong with you, it's only because you overthink" which has an impact on seeking services.* [26-year-old participant from FG3]

Many participants expressed the financial strain posed by mental health services for MSM in Nepal. The consensus was that the cost associated with psychiatrist visits, along with their limited financial resources, significantly affects MSMs' ability to access the necessary mental health services.

> *The main reason is finances because it is still very expensive, like we have to pay NRs 800 to 1000 (approximately US $8 to 10) per visit. It is expensive, even more so in private clinics.* [23-year-old participant from FG5]

Several participants across all FG sessions expressed concerns about time limitations and transportation challenges when it came to accessing mental health services. In an FG, a majority of participants agreed on the considerable difficulty that MSM faces in securing transportation to be able to go to a physical mental health appointment. In another session, everyone unanimously agreed and nodded in agreement with the following statement:

> *I think so too, because not everywhere has access to transportation, and for some places, we might even have to walk a lot to reach there.* [27-year-old participant from FG2]

When it comes to time constraints, participants talked about the difficulty of scheduling mental health appointments within the confines of work or school hours. They highlighted the difficulty of taking leave from work or school to attend counseling sessions during times of need.

> *I can go frequently, but the counseling appointment has to be time-friendly. Some of us are employed from 10 am to 5 pm or even 6 pm. If the counseling session is around that time, then I might come for a few days, taking a leave from work, but if my office does not allow me, then even if I had a mental health support need, I would not be able to attend.* [26-year-old participant from FG3]

Solution: Mental Health Smartphone Apps With Desired Features and Attributes

Overview

Participants expressed a preference for a smartphone app with a variety of features and attributes compared to traditional clinical settings. They foresaw that such an app could enhance understanding of mental health, offer convenience, improve accessibility, reduce the necessity for travel and associated expenses, and deliver services in a confidential and nonjudgmental setting.

> *During our young age, we didn't have any type of apps to help with our issues or any sort of networking apps like Grindr, but now people are more open to using apps, so creating an app to help solve the mental health issue and counsel can be a great idea.* [35-year-old participant from FG2]

Desired Features of the Mobile App

Participants recommended using creative approaches, such as fun activities to assess individuals' mental health for early detection, moving away from more direct approaches.

> *Something creative, not a direct approach, but through games or other ways we could assess the mental health status of the people for early detection.* [25-year-old participant from FG3]

Participants emphasized the importance of using the app to schedule regular counseling appointments with mental health professionals for those requiring assistance. There was a strong preference for using Zoom over platforms like Viber and WhatsApp for digital counseling, citing its widespread use during the COVID-19 pandemic.

> *…those who are in need of mental health services should get counseling appointments from a professional by selecting them once or twice a week in the app.* [29-year-old participants from FG4]
>
> *Rather than Viber and WhatsApp, Zoom is good for e-counseling, as in COVID many people are using it.* [18-year-old participants from FG4]

All participants underscored the importance of mental health and psychosocial service providers being qualified, friendly, and supportive of the LGBTIQA+ community. They stressed the need for an environment where individuals feel safe and comfortable to share their concerns.

> *First of all, they should be very friendly towards LGBTIQA+, no matter whether they are a community member or not, and we have to feel safe and able to share everything. Is qualified and has studied the related field.* [23-year-old participant from FG5]

Some participants had suggestions that would help make MSM more comfortable in participating in digital counseling, such as making cameras not compulsory.

> *We can do it through audio calls. Zoom counseling sessions are fine, but opening cameras should not be necessary or compulsory.* [21-year-old participant from FG5]

An additional recommendation included providing convenient hours, allowing users of the app to secure digital counseling appointments relatively quickly. This would accommodate individuals who work or go to school, ensuring continued accessibility to the services.

> *People will schedule according to their needs and how big their problem is, if you are having a problem now and get an appointment for a session after a month, it is not possible.* [21-year-old participant from FG2]

Participants suggested incorporating a toll-free helpline number within the smartphone app. They shared their experiences with toll-free helplines that did not function as intended in the past. Additionally, they provided suggestions for improving the toll-free helplines within the mobile app.

> *We can use a Toll-free helpline number, but even I tried to use toll-free service every time it was busy. So, the missed call system [call back system] is good.* [21-year-old participant from FG5]

Several participants suggested including a feature to message counselors in addition to the toll-free helpline that could help those who do not want to or cannot talk over the phone.

> *Some might not want to speak; they could talk through chat.* [30-year-old participant FG3]

Several participants shared their difficulties in finding friendly mental health and psychosocial service providers. To address this issue, participants suggested having a directory of LGBTIQA+-friendly providers on a mobile app that would help show MSM where to go when they require help.

> *I searched, and I came to know. It took me a lot of effort, and it was hard to find psychosocial counselors.* [29-year-old participant from FG2]

Participants also suggested to include mental health educational resources, especially in the form of videos.

> *Many are hidden, they do not even want to come out of the house, because of the fear of society. But they use mobile apps, they could connect to the app, and even with information and educational videos, we could reach them.* [21-year-old participant from FG5]

Many participants suggested a feature to connect with peers and other members of the MSM community through a communication channel within the app. They highlighted the importance of such a platform for sharing experiences and emphasized the value of peer support.

> *I think a discussion forum would be a good addition. The forum can help you share and make you feel like you are not the only one who is going through the same trauma and hardships, and we will be sharing with each other.* [21-year-old participant from FG1]

Attributes of the App

Many participants suggested placing special emphasis on the privacy and confidentiality of data collected by the app. They recommended that app developers and health care providers should commit to privacy and confidentiality clauses in their contracts, with strict consequences for any breaches of information.

> *The staff, app developers, and providers should sign on privacy and confidentiality in their contract. If leakage of information is found, they need to know that strong steps will be taken.* [26-year-old participant from FG3]

When discussing the user interface and colors of the app, several participants suggested that the mobile app should not overtly appear targeted exclusively at the LGBTIQA+ community. The participant expresses a desire for the app to have a discreet appearance, in contrast to the distinctiveness of dating apps targeted toward LGBTIQA+.

> *Through application maybe, the application should not look like for only LGBTIQ. It must look normal, not like Grindr.* [23-year-old participant from FG5]

Participants showed a strong interest in an engaging activity for user engagement and retention, particularly one that incorporates entertainment. One participant mentioned:

> *There has to be an environment in the mobile app so that I feel like going and using it again.* [26-year-old participant from FG3]

Discussion

Principal Findings

This study revealed a complex interplay between mental health challenges, including depression, anxiety, and suicidal behavior, among MSM in Nepal. The findings further highlight the barriers to accessing mental health care and support services among Nepali MSM due to factors such as insufficient mental health literacy, privacy concerns, financial strain, stigma, and discrimination. This underscores the urgent need for tailored and accessible mental health interventions. Participants overwhelmingly preferred smartphone app interventions to address the identified barriers and challenges, emphasizing their preference for accessible and confidential mental health support through digital platforms.

The major concern among MSM, where individuals perceive themselves as "all right" without the need for mental health services and less help-seeking attitudes, likely indicates a lack of mental health literacy, which is similar to the findings from studies among men and other minority populations [23-27]. Participants in this study expressed a preference for mental health resources and screening tools integrated into the app. Few studies have demonstrated that a smartphone app with an easily accessible and comprehensive mental health education module, resources, and engaging screening tools has the potential to combat this issue by fostering a proactive attitude toward mental well-being, the importance of seeking support, and the early detection of mental health problems [28-30].

The stigma and discrimination faced by MSM, both within society and health care settings, contribute to hesitancy in seeking mental health support. This fear of stigma and reluctance aligns with the findings from studies of various marginalized populations [26,31-33]. In response to this, participants expressed a preference for features within the mobile app that could link participants with LGBTIQA+-friendly mental health professionals through video sessions, automated text messages, or phone calls, emphasizing the crucial role of trust and understanding in the provider-patient relationship. Few interventions have integrated such features into digital interventions [34,35]. This feature could help to overcome this barrier by connecting individuals with LGBTIQA+-friendly and supportive mental health professionals and fostering a more inclusive, judgment-free, and accessible mental health support system.

In line with a substantial body of research, the findings emphasize that various stressors, particularly those related to societal biases, discrimination, fear of HIV, and other sexually transmitted infection results, contribute to psychological distress, and these influence maladaptive coping behaviors among MSM [7,8,36-38]. By incorporating features such as mental health resources, coping strategies, and peer support discussion forums, the app can have the potential to empower MSM to navigate these challenges more effectively.

The privacy and confidentiality concerns expressed by MSM underscore the need for a sensitive approach to mental health support. This apprehension aligns with findings from studies of various minority populations [39-42]. Participants in this study articulated the desire for a mobile app that explicitly addresses these concerns through robust consent forms, privacy features, and secure messaging platforms. The app could have features that aim to ensure privacy and confidentiality, potentially fostering MSM trust and addressing barriers related to sharing personal mental health information. Integrating these features into app design could significantly contribute to alleviating privacy concerns and establishing a secure environment that encourages seeking mental health support.

The cost of accessing mental health services was a major concern for participants in the study, which aligns with previous research on the cost of mental health in Nepal [43]. It is important to address financial strain in any intervention that is created to help MSM in Nepal with mental health [44-46]. Studies have found that, by reducing travel expenses, mHealth interventions help allow access for sexual minority individuals to mental health care [47,48]. This not only addresses the financial challenges faced by Nepali MSM but also alleviates the transportation struggles [35].

Strengths and Limitations

This study is of particular value due to the lack of participant involvement in the development of mental health interventions, with LGBTIQA+ consultation being notably rare when it comes to the creation of health interventions, policies, or guidelines [49,50]. Using FGs, the participants' perspectives can be used to create a more tailored and effective digital health intervention. However, this study has its own limitations. One of these limitations is the presence of social desirability bias, which is a common occurrence in FG discussions. This bias can influence participants to express socially acceptable opinions rather than their true thoughts and feelings. Additionally, it is worth noting that the study was done in Kathmandu, Nepal, which can differ in culture and access to mental health services than other areas of Nepal, limiting the transferability of the study findings mainly on the challenges and barriers. Finally, it is important to consider that the desire to participate in a given intervention does not automatically guarantee its real-world adoption. Evaluating the actual usage and effectiveness of the intervention in real-life scenarios is crucial to fully understand its impact and potential benefits. Therefore, it is necessary to evaluate real-world usage.

Conclusions

The study highlights the mental health challenges encountered by MSM in Nepal and the barriers they face in accessing mental health support services. The participants' direct quote, "invisible in the corner of the room," captures the hidden nature of their struggles intimately tied to the intersectional stigma surrounding mental health and sexuality. Emphasizing the potential of mobile apps, our findings suggest that incorporating user-friendly features like accessible resources, mental health screening tools, and digital counseling with LGBTIQA+-friendly providers can bring visibility to the mental health challenges of MSM. The mobile app has the ability to establish an open and supportive space, breaking down barriers and offering a pathway for MSM in Nepal to identify and address their mental health concerns with ease and confidence.

Acknowledgments

RS received funding from the National Institute on Drug Abuse (award number K01DA051346).

Conflicts of Interest

RS is an editorial board member of JMIR mHealth and uHealth.

Multimedia Appendix 1
Parent and child codes with description.
[DOCX File , 17 KB - humanfactors_v11i1e56002_app1.docx]

References

1. Sivasubramanian M, Mimiaga MJ, Mayer KH, Anand VR, Johnson CV, Prabhugate P, et al. Suicidality, clinical depression, and anxiety disorders are highly prevalent in men who have sex with men in Mumbai, India: findings from a community-recruited sample. Psychol Health Med 2011 Aug;16(4):450-462 [FREE Full text] [doi: 10.1080/13548506.2011.554645] [Medline: 21749242]
2. Sandfort TGM, Bakker F, Schellevis FG, Vanwesenbeeck I. Sexual orientation and mental and physical health status: findings from a Dutch population survey. Am J Public Health 2006 Jun;96(6):1119-1125. [doi: 10.2105/AJPH.2004.058891] [Medline: 16670235]
3. Salway T, Ross LE, Fehr CP, Burley J, Asadi S, Hawkins B, et al. A systematic review and meta-analysis of disparities in the prevalence of suicide ideation and attempt among bisexual populations. Arch Sex Behav 2019 Jan;48(1):89-111. [doi: 10.1007/s10508-018-1150-6] [Medline: 29492768]

4. Bostwick WB, Boyd CJ, Hughes TL, McCabe SE. Dimensions of sexual orientation and the prevalence of mood and anxiety disorders in the United States. Am J Public Health 2010 Mar;100(3):468-475 [FREE Full text] [doi: 10.2105/AJPH.2008.152942] [Medline: 19696380]
5. Alvy LM, McKirnan DJ, Mansergh G, Koblin B, Colfax GN, Flores SA, Project MIX Study Group. Depression is associated with sexual risk among men who have sex with men, but is mediated by cognitive escape and self-efficacy. AIDS Behav 2011 Aug;15(6):1171-1179. [doi: 10.1007/s10461-010-9678-z] [Medline: 20217471]
6. Nouri E, Moradi Y, Moradi G. What is the global prevalence of depression among men who have sex with men? A systematic review and meta-analysis. Ann Gen Psychiatry 2022 Sep 12;21(1):38 [FREE Full text] [doi: 10.1186/s12991-022-00414-1] [Medline: 36096814]
7. King M, Semlyen J, Tai SS, Killaspy H, Osborn D, Popelyuk D, et al. A systematic review of mental disorder, suicide, and deliberate self harm in lesbian, gay and bisexual people. BMC Psychiatry 2008 Aug 18;8:70 [FREE Full text] [doi: 10.1186/1471-244X-8-70] [Medline: 18706118]
8. Meyer IH. Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: conceptual issues and research evidence. Psychol Bull 2003 Sep;129(5):674-697 [FREE Full text] [doi: 10.1037/0033-2909.129.5.674] [Medline: 12956539]
9. Byrne J, Russel C, Quan A. Legal gender recognition: a multi-country legal and policy review in Asia. Asia Pacific Transgender Network and United Nations Development Programme. 2017. URL: https://weareaptn.org/wp-content/uploads/2019/05/UPDATED-ACKNOWLEDGMENTS-2019_rbap-hhd-2017-legal-gender-recognition-1.pdf [accessed 2024-03-05]
10. The constitution of Nepal. National Law Commission: Nepal. 2015. URL: https://ag.gov.np/files/Constitution-of-Nepal_2072_Eng_www.moljpa.gov_.npDate-72_11_16.pdf [accessed 2024-03-05]
11. UNDP, USAID. Being LGBT in Asia: Nepal country report. United Nations Development Programme and United States Agency for International Development. 2014. URL: https://www.undp.org/asia-pacific/publications/being-lgbt-asia-nepal-country-report [accessed 2024-03-05]
12. Storm S, Deuba K, Shrestha R, Pandey LR, Dahal D, Shrestha MK, et al. Social and structural factors associated with depression and suicidality among men who have sex with men and transgender women in Nepal. BMC Psychiatry 2021 Sep 29;21(1):476 [FREE Full text] [doi: 10.1186/s12888-021-03477-8] [Medline: 34587934]
13. Paudel K, Gautam K, Bhandari P, Wickersham JA, Dhakal M, Sharma S, et al. Suicidal ideation, plan, and attempt among men who have sex with men in Nepal: Findings from a cross-sectional study. PLOS Glob Public Health 2023;3(11):e0002348 [FREE Full text] [doi: 10.1371/journal.pgph.0002348] [Medline: 37992034]
14. Giri PD, Adhikari A, Pradhan M, Yogi I, Khanal S. Barriers in access to health care services among lesbian, gay, bisexual, transgender (LGBT). Int J New Technol Res 2019;5(3):15-20. [doi: 10.31871/IJNTR.5.3.24]
15. Oli N, Onta SR. Self-perception of stigma and discrimination among men having sex with men. J Nepal Health Res Counc 2012 Sep;10(22):197-200. [Medline: 23281450]
16. Singh S, Pant SB, Dhakal S, Pokhrel S, Mullany LC. Human rights violations among sexual and gender minorities in Kathmandu, Nepal: a qualitative investigation. BMC Int Health Hum Rights 2012 May 16;12:7 [FREE Full text] [doi: 10.1186/1472-698X-12-7] [Medline: 22591775]
17. Subedi P, Jha A. Status of healthcare in LGBTQI+ community in Nepal: challenges and possibilities. JNMA J Nepal Med Assoc 2023 Jan 01;61(257):95-97 [FREE Full text] [doi: 10.31729/jnma.7948] [Medline: 37203914]
18. Rauschenberg C, Schick A, Hirjak D, Seidler A, Paetzold I, Apfelbacher C, et al. Evidence synthesis of digital interventions to mitigate the negative impact of the COVID-19 pandemic on public mental health: rapid meta-review. J Med Internet Res 2021 Mar 10;23(3):e23365 [FREE Full text] [doi: 10.2196/23365] [Medline: 33606657]
19. Gilbey D, Morgan H, Lin A, Perry Y. Effectiveness, acceptability, and feasibility of digital health interventions for LGBTIQ+ young people: systematic review. J Med Internet Res 2020 Dec 03;22(12):e20158 [FREE Full text] [doi: 10.2196/20158] [Medline: 33270039]
20. Ministry of Health and Population, New ERA, The DHS Program ICF. Nepal demographic and health survey 2022: key indicators report. Ministry of Health and Population. Kathmandu, Nepal; 2022. URL: https://dhsprogram.com/pubs/pdf/PR142/PR142.pdf [accessed 2024-03-05]
21. Sherin KM, Sinacore JM, Li XQ, Zitter RE, Shakil A. HITS: a short domestic violence screening tool for use in a family practice setting. Fam Med 1998;30(7):508-512. [Medline: 9669164]
22. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
23. Kim JE, Saw A, Zane N. The influence of psychological symptoms on mental health literacy of college students. Am J Orthopsychiatry 2015 Nov;85(6):620-630 [FREE Full text] [doi: 10.1037/ort0000074] [Medline: 26052815]
24. Klineberg E, Biddle L, Donovan J, Gunnell D. Symptom recognition and help seeking for depression in young adults: a vignette study. Soc Psychiatry Psychiatr Epidemiol 2011 Jun;46(6):495-505. [doi: 10.1007/s00127-010-0214-2] [Medline: 20358174]
25. Swami V. Mental health literacy of depression: gender differences and attitudinal antecedents in a representative British sample. PLoS One 2012;7(11):e49779 [FREE Full text] [doi: 10.1371/journal.pone.0049779] [Medline: 23166769]

26. Garcia C, Grant E, Treharne GJ, Arahanga-Doyle H, Lucassen MFG, Scarf D, et al. 'Is it worth potentially dealing with someone who won't get it?': LGBTQA+ university students' perspectives on mental health care. J R Soc New Zealand 2023 Jul 16:1-15 [FREE Full text] [doi: 10.1080/03036758.2023.2235297]
27. Narendorf SC, Munson MR, Ben-David S, Cole AR, Scott LD. Race and gender differences in attitudes toward help seeking among marginalized young adults with mood disorders: a mixed-methods investigation. Psychiatr Rehabil J 2018 Dec;41(4):277-289. [doi: 10.1037/prj0000312] [Medline: 30507242]
28. Shrestha R, Altice FL, Khati A, Azwa I, Gautam K, Gupta S, et al. Clinic-integrated smartphone app (JomPrEP) to improve uptake of HIV testing and pre-exposure prophylaxis among men who have sex with men in Malaysia: mixed methods evaluation of usability and acceptability. JMIR Mhealth Uhealth 2023 Feb 16;11:e44468 [FREE Full text] [doi: 10.2196/44468] [Medline: 36795465]
29. Drydakis N. mHealth apps and physical and mental health outcomes of sexual minorities. J Homosex 2023 Dec 06;70(14):3421-3448. [doi: 10.1080/00918369.2022.2095240] [Medline: 35904851]
30. Dubov A, Fraenkel L, Goldstein Z, Arroyo H, McKellar D, Shoptaw S. Development of a smartphone app to predict and improve the rates of suicidal ideation among transgender persons (TransLife): qualitative study. J Med Internet Res 2021 Mar 24;23(3):e24023 [FREE Full text] [doi: 10.2196/24023] [Medline: 33596181]
31. Hsieh N, Shuster SM. Health and health care of sexual and gender minorities. J Health Soc Behav 2021 Sep;62(3):318-333 [FREE Full text] [doi: 10.1177/00221465211016436] [Medline: 34528481]
32. Cronin TJ, Pepping CA, Halford WK, Lyons A. Mental health help-seeking and barriers to service access among lesbian, gay, and bisexual Australians. Australian Psychologist 2021 Mar 17;56(1):46-60. [doi: 10.1080/00050067.2021.1890981]
33. Dunbar MS, Sontag-Padilla L, Ramchand R, Seelam R, Stein BD. Mental health service utilization among lesbian, gay, bisexual, and questioning or queer college students. J Adolesc Health 2017 Sep;61(3):294-301. [doi: 10.1016/j.jadohealth.2017.03.008] [Medline: 28549595]
34. Samarasekera U. The rise of racial minority health apps. Lancet Digit Health 2022 Apr;4(4):e218-e219 [FREE Full text] [doi: 10.1016/S2589-7500(22)00043-7] [Medline: 35337641]
35. Beattie A, Shaw A, Kaur S, Kessler D. Primary-care patients' expectations and experiences of online cognitive behavioural therapy for depression: a qualitative study. Health Expect 2009 Mar;12(1):45-59 [FREE Full text] [doi: 10.1111/j.1369-7625.2008.00531.x] [Medline: 19250152]
36. Hatzenbuehler ML, Pachankis JE. Stigma and minority stress as social determinants of health among lesbian, gay, bisexual, and transgender youth: research evidence and clinical implications. Pediatr Clin North Am 2016 Dec;63(6):985-997. [doi: 10.1016/j.pcl.2016.07.003] [Medline: 27865340]
37. Chard AN, Metheny NS, Sullivan PS, Stephenson R. Social stressors and intoxicated sex among an online sample of men who have sex with men (MSM) drawn from seven countries. Subst Use Misuse 2018 Jan 02;53(1):42-50. [doi: 10.1080/10826084.2017.1322985] [Medline: 28792280]
38. Dowshen N, Binns HJ, Garofalo R. Experiences of HIV-related stigma among young men who have sex with men. AIDS Patient Care STDS 2009 May;23(5):371-376 [FREE Full text] [doi: 10.1089/apc.2008.0256] [Medline: 19320600]
39. Khati A, Wickersham JA, Rosen AO, Luces JRB, Copenhaver N, Jeri-Wahrhaftig A, et al. Ethical issues in the use of smartphone apps for HIV prevention in Malaysia: focus group study with men who have sex with men. JMIR Form Res 2022;6(12):e42939 [FREE Full text] [doi: 10.2196/42939] [Medline: 36563046]
40. Shrestha R, Fisher C, Wickersham JA, Khati A, Kim R, Azwa I, et al. Privacy and confidentiality concerns related to the use of mHealth apps for HIV prevention efforts among Malaysian men who have sex with men: cross-sectional survey study. JMIR Form Res 2021 Dec 16;5(12):e28311 [FREE Full text] [doi: 10.2196/28311] [Medline: 34924355]
41. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Privacy and security in mobile health apps: a review and recommendations. J Med Syst 2015 Jan;39(1):181. [doi: 10.1007/s10916-014-0181-3] [Medline: 25486895]
42. Morera EP, de la Torre Díez I, Garcia-Zapirain B, López-Coronado M, Arambarri J. Security recommendations for mHealth apps: elaboration of a developer's guide. J Med Syst 2016 Jun;40(6):152. [doi: 10.1007/s10916-016-0513-6] [Medline: 27147515]
43. Rajan S, Rathod SD, Luitel NP, Murphy A, Roberts T, Jordans MJD. Healthcare utilization and out-of-pocket expenditures associated with depression in adults: a cross-sectional analysis in Nepal. BMC Health Serv Res 2020 Mar 25;20(1):250 [FREE Full text] [doi: 10.1186/s12913-020-05094-9] [Medline: 32213188]
44. Regmi SK, Pokharel A, Ojha SP, Pradhan SN, Chapagain G. Nepal mental health country profile. Int Rev Psychiatry 2004;16(1-2):142-149. [doi: 10.1080/09540260310001635186] [Medline: 15276946]
45. Knapp M, Funk M, Curran C, Prince M, Grigg M, McDaid D. Economic barriers to better mental health practice and policy. Health Policy Plan 2006 May;21(3):157-170 [FREE Full text] [doi: 10.1093/heapol/czl003] [Medline: 16522714]
46. Charlton BM, Gordon AR, Reisner SL, Sarda V, Samnaliev M, Austin SB. Sexual orientation-related disparities in employment, health insurance, healthcare access and health-related quality of life: a cohort study of US male and female adolescents and young adults. BMJ Open 2018 Jul 26;8(6):e020418 [FREE Full text] [doi: 10.1136/bmjopen-2017-020418] [Medline: 30049672]

47. Meiksin R, Melendez-Torres GJ, Miners A, Falconer J, Witzel TC, Weatherburn P, et al. E-Health interventions targeting STIs, sexual risk, substance use and mental health among men who have sex with men: four systematic reviews. Public Health Res 2022;10(4):1-322 [FREE Full text] [doi: 10.3310/BRWR6308] [Medline: 35344312]
48. Lucassen MFG, Hatcher S, Fleming TM, Stasiak K, Shepherd MJ, Merry SN. A qualitative study of sexual minority young people's experiences of computerised therapy for depression. Australas Psychiatry 2015 Jun;23(3):268-273. [doi: 10.1177/1039856215579542] [Medline: 25881962]
49. Anthony P, Crawford P. Service user involvement in care planning: the mental health nurse's perspective. J Psychiatr Ment Health Nurs 2000 Oct;7(5):425-434. [doi: 10.1046/j.1365-2850.2000.00332.x] [Medline: 11933469]
50. McNair RP, Hegarty K. Guidelines for the primary care of lesbian, gay, and bisexual people: a systematic review. Ann Fam Med 2010;8(6):533-541 [FREE Full text] [doi: 10.1370/afm.1173] [Medline: 21060124]

Abbreviations

FG: focus group
mHealth: mobile health
MSM: men who have sex with men
LGBTIQA+: lesbian, gay, bisexual, transgender, intersex, queer, and asexual

Edited by G Eysenbach, A Kushniruk; submitted 02.01.24; peer-reviewed by A AL-Asadi, A Nyamathi; comments to author 31.01.24; revised version received 01.02.24; accepted 18.02.24; published 29.03.24.

Please cite as:
Gautam K, Aguilar C, Paudel K, Dhakal M, Wickersham JA, Acharya B, Sapkota S, Deuba K, Shrestha R
Preferences for mHealth Intervention to Address Mental Health Challenges Among Men Who Have Sex With Men in Nepal: Qualitative Study
JMIR Hum Factors 2024;11:e56002
URL: https://humanfactors.jmir.org/2024/1/e56002
doi:10.2196/56002
PMID:38551632

Original Paper

Patient Perspectives on Communication Pathways After Orthopedic Surgery and Discharge and Evaluation of Team-Based Digital Communication: Qualitative Exploratory Study

Lili Worre Høpfner Jensen[1], BSc, MSc; Ole Rahbek[1], MD, PhD; Rikke Emilie Kildahl Lauritsen[1], BSc, MSc; Søren Kold[1], MD, PhD; Birthe Dinesen[2], MSc, PhD

[1]Interdisciplinary Orthopaedics, Orthopaedic Surgery Department, Aalborg University Hospital, Aalborg, Denmark

[2]Laboratory for Welfare Technologies – Digital Health & Rehabilitation, Sport Sciences – Performance & Technology, Department of Health Science and Technology, Aalborg University, Aalborg East, Denmark

Corresponding Author:
Lili Worre Høpfner Jensen, BSc, MSc
Interdisciplinary Orthopaedics
Orthopaedic Surgery Department
Aalborg University Hospital
Hobrovej 18-22
Aalborg, 9000
Denmark
Phone: 45 60229406
Email: lili.jensen@rn.dk

Abstract

Background: The transition from hospital to home after orthopedic surgery requires smooth communication and coordination between patients and their team of care to avoid fragmented care pathways. Digital communication is increasingly being used to facilitate easy and accessible asynchronous communication between patients and health care professionals across settings. A team-based approach to digital communication may provide optimized quality of care in the postoperative period following orthopedic surgery and hospital discharge.

Objective: This study was divided into two phases that aimed to (1) explore the perspectives of patients undergoing orthopedic surgery on current communication pathways at a tertiary hospital in Denmark and (2) test and explore patients’ experiences and use of team-based digital communication following hospital discharge (eDialogue).

Methods: A triangulation of qualitative data collection techniques was applied: document analysis, participant observations (n=16 hours), semistructured interviews with patients before (n=31) and after (n=24) their access to eDialogue, and exploration of use data.

Results: Findings show that patients experience difficult communication pathways after hospital discharge and a lack of information due to inadequate coordination of care. eDialogue was used by 84% (26/31) of the patients, and they suggested that it provided a sense of security, coherence, and proximity in the aftercare rearranging communication pathways for the better. Specific drivers and barriers to use were identified, and these call for further exploration of eDialogue.

Conclusions: In conclusion, patients evaluated eDialogue positively and suggested that it could support them after returning home following orthopedic surgery.

(JMIR Hum Factors 2024;11:e49696) doi:10.2196/49696

KEYWORDS

digital communication; patient-provider communication; continuity of care; interdisciplinary communication; hospital discharge; orthopedic surgery; postoperative care; text messaging; mobile phone

Introduction

Across the health care system, digital communication is being implemented as an addition to traditional communication pathways [1,2]. Digital communication is a form of eHealth [3] that facilitates asynchronous 2-way text messaging between patients and health care professionals (HCPs). Digital communication is typically facilitated through email [4,5]; secure text messaging in patient portals [2,6]; or as a feature in mobile health apps developed for specific purposes, for example, postoperative monitoring [7,8] and neonatal tele-homecare [9,10]. Establishing the effects of using digital communication is still challenging [11,12]; however, an increasing number of studies suggest that it can support patients in taking care of their own health [12] and address unmet communication needs after hospital discharge [13,14]. When digital communication is used with the purpose of facilitating team-based communication across settings, studies indicate that it may contribute to improving continuity of care (COC) in transitions from hospital to home [14-16]. COC is essential for patients undergoing complex and long-term procedures [17]. Patients who receive care across time and settings are susceptible to fragmented care, and the absence of consistent professional support and communication may lead to neglect that ultimately affects patient safety [18-21]. Because of the growing population in need of orthopedic surgery, workforce shortage [22], and optimized surgery techniques, patients undergoing orthopedic surgery are discharged earlier [23]. Day surgery is increasingly used, and even patients undergoing complex treatments are hospitalized for a shorter time. Common to patients undergoing orthopedic surgery is a need for continuing rehabilitation across settings, supported by adequate communication and home symptom monitoring between follow-up visits [24,25]. Even so, only a few studies have addressed the use of team-based digital communication involving patients and HCPs across settings, and primarily in other patient populations, such as patients with cancer [14,15,26] and children with cerebral palsy [27]. To our knowledge, no studies have investigated the use of team-based digital communication after hospital discharge in orthopedic surgery, although these patients often have long periods of rehabilitation, where cross-disciplinary and cross-sectoral communication is pivotal [28].

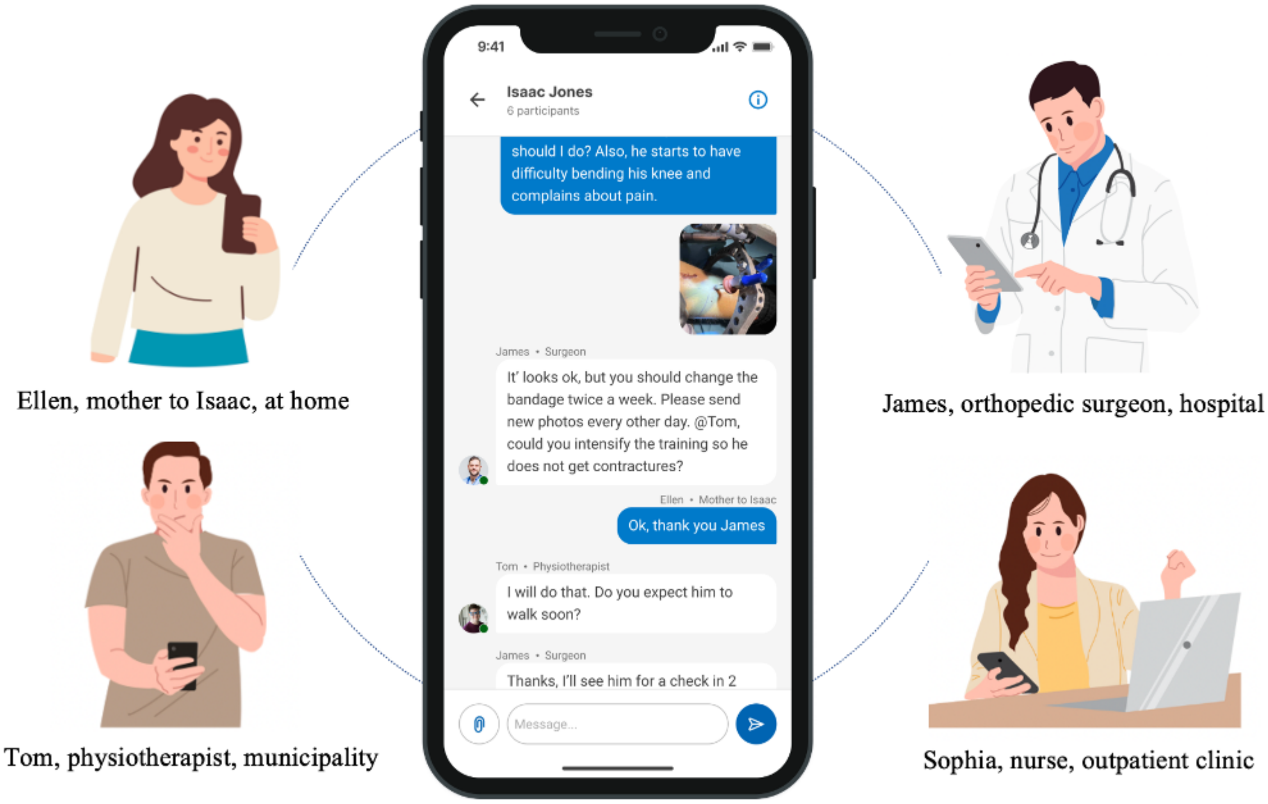
Therefore, the aim of this study was to explore the perspectives of patients undergoing orthopedic surgery on current communication pathways (phase 1) and to subsequently test and explore their experiences and use of a team-based digital communication solution (eDialogue) to evaluate whether the solution can support their needs after hospital discharge (phase 2).

Methods

The eDialogue Intervention

The technical solution used in this study was a simple General Data Protection Regulation–compliant solution, developed for team-based communication, that lets users chat directly with each other with texts and photos (“LetDialog” by Visma) [29] (Figure 1).

Figure 1. Illustration of the team-based digital communication (eDialogue) used in this study, where patients and health care professionals across settings could text and send photos to communicate about postdischarge issues.



The solution was accessed through an app for smartphones or through a website. Users could choose how they accessed it individually. To ensure compliance with the current legislation, user profiles were created with a digital signature (NemID), and the digital dialogues were stored in a secure cloud-based solution. A data processor agreement was made among the North Denmark Region, Aalborg University Hospital, and Visma before this study.

The features were basic asynchronous text messaging and exchange of photos. Photos could be taken directly or uploaded and sent through the solution for review by the health care team. Team-based digital communication was organized in teams, defined by the individual patient, in a shared chat. Notifications were sent to all the participants when there were new posts. Key HCPs from the orthopedic surgery department at the hospital were identified and recruited for participation before the study (surgeons, nurses, and physiotherapists). Other HCPs from municipal or private settings were recruited ad hoc and based on patients' wishes (eg, physiotherapists from the municipality).

Study Design

The study was exploratory, using a triangulation of qualitative data collection techniques, including document analysis, participant observations [30], semistructured interviews [31], and use data, with the purpose of obtaining in-depth knowledge of patients' perspectives and the context.

Theoretical Framework

The theoretical framework for this study was inspired by the concepts of COC [17,32], which is used as a measure of quality of care in health care transitions. COC includes *informational continuity*, described as the use of medical or personal information to provide appropriate care over time; *management continuity*, which refers to the provision of timely, coordinated, and complementary services that are responsive to patients' needs to connect care over time; and *relational continuity*, which involves the consistency and quality of relationships between patients and providers as a means of connecting care over time [32]. All 3 dimensions should be integrated to achieve COC, and thus, COC is maximized when planning for patient-provider continuity, information exchange, and seamless coordination of services in the period of transition from hospital to home [32-34]. For this study, COC has inspired the data collection and analysis of interviews and observations as well as the use of team-based digital communication to prevent fragmented care experiences after hospital discharge.

Participants and Setting

The study was conducted at the Orthopedic Surgery Department of Aalborg University Hospital, Denmark. The recruitment of participants began in May 2021 and ended in November 2021. The final follow-up interviews were conducted 2 months later in January 2022.

In phase 1, participants were recruited consecutively based on predefined inclusion criteria: (1) patients, or their parents if the patient was aged <15 years, undergoing deformity correction (DC) surgery or anterior cruciate ligament (ACL) reconstruction; (2) those who were able to read and write Danish; (3) those who were discharged to their own home and had planned follow-up in the outpatient clinic; and (4) those who owned a smartphone and had access to a secure digital signature. The exclusion criteria were (1) those who were not able to understand Danish and (2) those who were not cognitively able to participate in interviews.

The 2 patient groups, DC and ACL, were selected because they represent 2 different orthopedic surgical care pathways. Involving both patient groups allowed us to gain an insight into the different needs of patients undergoing orthopedic surgery. ACL is performed as a day surgery (ie, discharge on the same day), whereas patients in the DC group most often have longer hospitalizations and prolonged treatments.

The same recruitment procedure was used for patients undergoing DC or ACL. The patients were approached by secretaries at the hospital with an invitation to participate. If the patients agreed to be called by phone with information about participation in the study, the first author (LWHJ) would call them to provide oral participant information. Written participant information was then sent by email, and the patients were given time to consider participation. One patient did not want to participate after receiving oral information due to a lack of mental capacity to participate in the interviews. Another patient could not be contacted by telephone after he had initially registered his telephone number. Both patients were from the ACL group.

In phase 2, patients and parents (if the patient was a minor) were onboarded to eDialogue on the day of discharge. The orthopedic surgeon, who had performed the surgery, was invited to join the patients' dialogue, as were nurses from the outpatient clinic and physiotherapists across sectors who were involved in the patient's care and rehabilitation after discharge. Thus, the patients were connected with known HCPs and were able to use eDialogue as needed from the day of discharge until 2 months after discharge. The patients could send texts and photos whenever it suited them, but they were told that a 24-hour response time on weekdays (Monday to Friday) would be aimed for. As such, messages sent during weekends and holidays would be responded to on the next weekday. It was pointed out, both verbally and in the participant information letter, that in case of emergency, patients should not use the solution but instead call, as they usually would have done before access to eDialogue. Thus, eDialogue was an addition to traditional communication channels (eg, telephone calls and email) and an extra opportunity for communication after discharge.

Data Collection

A triangulation of data collection techniques was performed to achieve exhaustive knowledge of current communication pathways, patients' perspectives, and their experiences with eDialogue.

Phase 1

First, document analysis was performed on documents and guidelines for postdischarge communication between patients and HCPs followed by participant observations of workflows (n=16 hours). The aim of the document analysis was to obtain knowledge of the policies and context of the study. The aim of observations was to document the current communication pathways for patients following hospital discharge. Participant observations were performed by LWHJ and followed a predefined observation guide [30]. Observations were carried out at the orthopedic surgery ward and the outpatient clinic at the hospital and documented in Word files (Microsoft Corp). This involved, for example, secretaries' handling of incoming phone calls from patients, registration of patient inquiries, procedures for passing on messages to nurses and orthopedic surgeons, and HCPs' calls with patients. In addition, existing systems for communication with discharged patients were reviewed, including written communication to patients via "E-box," (a secure digital mail system for communication from Danish authorities) correspondence between HCPs across hospitals and municipalities in the local electronic health record, and interprofessional communication related to patients' phone calls.

Second, semistructured interviews were conducted at the point of inclusion for each participant (N=31). The aim was to explore patients' and parents' perspectives on current communication pathways. Interviews were performed using video 5 to 7 days before surgery for patients from the ACL group (n=14) and physically at the ward for patients and parents from the DC group (n=17) because they were all hospitalized in connection with their operation. All interviews were conducted by LWHJ based on a predefined semistructured interview guide (Multimedia Appendix 1). The guide was developed based on the theoretical framework for this study and combined with exploratory questions. It was pilot-tested in 2 patients similar to the study participants and revised accordingly. The interviews were carried out until data saturation had been reached, defined by the point where no new insights into participants' responses occurred, indicating the achievement of a comprehensive understanding of the participants' perspectives [31]. The interviews were audio recorded using a digital voice recorder (DM-450; Olympus) and lasted for 40 to 60 minutes. They were continuously transcribed and documented in Word files. During and at the end of each interview, key points were summarized to ensure the credibility of the meanings expressed.

Phase 2

Semistructured follow-up interviews were performed with the same patients and parents 2 months after hospital discharge (24/31, 77%). The aim was to explore their experiences of using eDialogue for team-based communication in the postdischarge period. The interviews were performed by LWHJ, audio recorded, and followed a predefined interview guide that was pilot-tested (Multimedia Appendix 1). The interviews were conducted until data saturation was reached for each patient group [31]. They lasted between 30 and 60 minutes. Both users and those who did not use eDialogue after getting access were interviewed. A total of 6 patients (DC: n=3; ACL: n=3) were reached by phone, their experiences were discussed, and a short report was written. Nothing new emerged from these conversations. One parent of a child from the DC group was lost to follow-up as she did not return our calls. Interviews were performed face-to-face at the ward or digitally based on the preferences of the participants. Participants were most likely to choose web-based interviews due to convenience and distance to the hospital, and data collection was conducted at the same time as the COVID-19 pandemic.

Use data of eDialogue was collected through registration of events and manual counts of messages exchanged in all digital dialogues. Data included the total number of messages exchanged in eDialogue during the 2-month study period, the number of text messages and photos sent by patients or parents, and the number of text messages that actually needed a reply from HCPs. In addition, the distribution of text messages per week per patient group was collected and displayed to show the differences between groups. Content analysis [31] of the messages sent by the patients and the parents was performed to provide insight into question categories as well as how they were distributed between the patient groups.

Data Analysis

Data analysis was carried out in NVivo (version 20.6.2; Lumivero), inspired by Brinkmann and Kvale [31], with the aim of achieving an in-depth understanding and connection of the participants' expressed perspectives on current communication pathways (phase 1) and experiences using eDialogue (phase 2).

Separate data analyses were carried out for phase 1 and phase 2 and for each patient group (DC and ACL), all involving 3 steps: meaning coding, meaning condensation, and meaning interpretation (Textbox 1).

In phase 1, observational data were integrated into the data set to enhance the understanding of existing communication pathways for patients in need of postdischarge contact.

Use data from eDialogue were analyzed and presented using simple descriptive statistics and basic content analysis to present the overall question categories.

The reporting of this study followed the Consolidated Criteria for Reporting Qualitative Research checklist [35].

Textbox 1. The process of the thematic analysis.

Meaning coding

- Coding of the transcribed interviews was initially performed individually by 2 of the authors (LWHJ and REKL) by randomly selecting 4 interviews from each patient group and from each phase. Coding was conducted with an inductive approach with the aim of reflecting the meanings expressed by the participants.
- To achieve intersubjectivity before and during the analysis, interviews were individually read and reread by LWHJ and REKL, notes were made for initial ideas for codes, and these were then compared and discussed until agreement. In phase 1, this resulted in 21 codes in the deformity correction (DC) group and 18 codes in the anterior cruciate ligament (ACL) group. In phase 2, we identified 18 codes in the DC group and 15 codes in the ACL group. LWHJ continued the data analysis of the remaining interviews by applying the same codes to the entire data.

Meaning condensation

- The codes were then reread and condensed in discussion with REKL and the last author (BD) through several iterations, and this process resulted in 12 codes for the DC group and 8 codes for the ACL group in phase 1 and 11 codes for the DC group and 7 codes for the ACL group in phase 2. We merged and we left behind codes that did not directly address the research questions or were only described vaguely by 1 participant.
- The remaining codes were then discussed with all authors to achieve further condensation and to define and name subthemes and themes that would capture the essence of the data. It was clear to the authors that the 2 patient groups had expressed similar perspectives on the phenomena of interest, and therefore, codes could be merged between the 2 groups in this step. Theme generation was based on a systematic identification and organization of recurring patterns, topics, or concepts within the data set. This process resulted in 3 overall themes and 6 subthemes for phase 1 and 3 overall themes and 3 subthemes for phase 2.

Meaning interpretation

- Themes were defined and described narratively, and data extracts were chosen for presentation in the manuscript before writing the findings.

Ethical Considerations

Before the study started, the Ethics Committee of Northern Jutland was approached, and it was found that the study did not require approval, as eDialogue was an extra opportunity for patients to communicate directly with their team of HCPs across sectors. This was confirmed by email on March 18, 2021 (2021-000438). The study was registered with the Regional Committee on Health Research and approved (ID number 2021-057). All participants received thorough oral and written information and guidance in the use of eDialogue before discharge. The study followed the Helsinki Declaration, and the participants signed an informed consent form and were able to leave the study without explanation or effects on usual care. All patients or parents had access to eDialogue for 2 months after hospital discharge. If they wanted, patients were allowed to keep the possibility of eDialogue with their team of HCPs after 2 months and until their follow-up in the outpatient clinic was completed. An administrator from the project group was passively present in all dialogues to continuously observe whether the patients used the solution for emergencies against the given advice.

Results

Participants' Characteristics

Table 1 provides the baseline description of the 31 patients included in this study. The patients were recruited from 2 different subgroups of orthopedic surgery: DC (17/31, 55%) and ACL (14/31, 45%).

The patients in the DC group were, for example, patients with malalignment or limb length discrepancy, and they were all hospitalized for >1 day. The patients in the ACL group were all treated with ACL reconstruction, and they had the procedure performed as day surgery. Of the 14 patients with ACL injuries, 7 (50%) had a concurrent meniscal injury.

Across the groups, most patients were male (22/31, 71%), and patients ranged in age from 1 to 59 years. Patients from the DC group were discharged from the hospital after an average of 6.1 (range 1-9) days, and patients from the ACL group were all discharged on the same day of surgery (<9 hours of admission). All the included patients were discharged to their own home. Of the 31 patients included, 14 (45%) had previously undergone orthopedic surgery at Aalborg University Hospital, and thus, they were able to reflect on previous experiences with postdischarge communication during the initial interviews in phase 1. In the DC group, 5 patients lived outside the North Jutland Region.

A total of 42% (13/31) of the patients were children aged <15 years, and thus, their parents were the primary users of eDialogue. Therefore, the baseline characteristics of all users of eDialogue are presented in Table 2.

The table shows 33 users in total because 2 patients aged 16 and 17 years had a parent joining the dialogue with them. Of the 13 parents who were users of eDialogue with or on behalf of their child, 77% (10/13) were female (mothers). The mean age of the parents was 43 (range 37-48) years on the day of discharge. All users of eDialogue used a smartphone on a daily basis.

Table 1. Characteristics of all patients across groups (DC[a], n=17; ACL[b], n=14; N=31).

Characteristics	Values
Sex (DC/ACL), n (%)	
Female	5 (29)/4 (29)
Male	12 (71)/10 (71)
Age at discharge (years), mean (range)	
DC	19.2 (1-59)
ACL	29.1 (17-46)
Length of hospital stay, mean (range)	
DC	6.1 (1-9) days
ACL	1 (7-9) hours
Previously had orthopedic surgery (yes/no), n (%)	
DC	12 (71)/5 (29)
ACL	2 (14)/12 (86)
Highest education level (DC/ACL), n (%)	
Primary or high school	12 (71)/5 (36)
Vocational education (skilled worker)	2 (12)/2 (14)
Short education, 2-3 years	1 (6)/2 (14)
Bachelor's degree, 3-5 years	2 (12)/4 (29)
Academic education, 5-8 years	0 (0)/1 (7)
Work status (DC/ACL), n (%)	
Student	13 (76)/7 (50)
Unemployed	1 (6)/2 (14)
Employed	3 (18)/5 (36)
Civil status (DC/ACL), n (%)	
Living alone	3 (18)/4 (29)
Cohabiting	14 (82)/10 (71)

[a]DC: deformity correction.
[b]ACL: anterior cruciate ligament.

Table 2. Baseline characteristics of all users of eDialogue (DC[a], n=18; ACL[b], n=15; patients and parents; N=33).

Characteristics	Values
Distribution of users (DC/ACL), n (%)	
Patients	6 (33)/14 (93)
Parents	12 (67)/1 (7)
Sex (DC/ACL), n (%)	
Female	12 (67)/5 (33)
Male	6 (33)/10 (67)
Age at discharge (years), mean (range)	
DC	39.8 (16-59)
ACL	28.8 (17-46)
Highest education level (DC/ACL), n (%)	
Primary or high school	2 (11)/5 (33)
Vocational education (skilled worker)	3 (17)/2 (13)
Short education, 2-3 years	3 (17)/2 (13)
Bachelor's degree, 3-5 years	8 (44)/5 (33)
Master's degree, 5-8 years	2(11)/1 (7)
Work status (DC/ACL), n (%)	
Student	2 (11)/7 (47)
Unemployed	1 (6)/2 (13)
Employed	14 (78)/6 (40)
Disability pensioner	1 (6)/0 (0)
Civil status (DC/ACL), n (%)	
Living alone	4 (22)/3 (20)
Cohabiting	14 (78)/12 (80)

[a]DC: deformity correction.

[b]ACL: anterior cruciate ligament.

Phase 1: Perspectives on Current Communication Pathways

Themes and Subthemes

Through the initial interviews, 3 themes and associated subthemes were revealed across the groups. Overall, patients and parents from the DC and ACL groups had similar experiences of, and perspectives on, current communication pathways. However, some subthemes were more prominent in one group than the other. This is illustrated by showing how many patients and parents from each group expressed experiences related to the specific subtheme (Table 3).

Table 3. Themes and subthemes of patients' and parents' perspectives on current communication pathways with HCPs[a] after hospital discharge (N=31).

Themes and subthemes	DC[b] (n=17), n (%)	ACL[c] (n=14), n (%)
Difficult communication pathways		
Doubts about who to contact and when	8 (47)	7 (50)
Withhold questions or forget to ask	7 (41)	9 (64)
Lack of information due to inadequate coordination of care		
Knowledge is not shared sufficiently	8 (47)	6 (43)
Hard to be "the messenger" between HCPs	9 (53)	5 (36)
Relations and communication provide "peace of mind"		
Relational continuity matters	15 (88)	4 (29)
Contacts provides a sense of being cared for	10 (59)	2 (14)

[a]HCP: health care professional.

[b]DC: deformity correction.

[c]ACL: anterior cruciate ligament.

Difficult Communication Pathways

Most patients and parents expressed frustrations related to difficult communication pathways when they needed contact with HCPs. They were in doubt about who to contact regarding specific issues both before and after surgery and discharge:

> *It was like a week after discharge, and I didn't know who to ask. Should I contact the department, the outpatient clinic or my own physician? I didn't know that. They kept telling me to call a new location.* [Mother of patient 2, DC]

The patients also described how they would often forget to ask questions at the outpatient clinic or they would withhold questions because they found it difficult to assess whether their issues were "severe enough" to take up HCPs time. A patient explains how it had previously led to concerns and worsening of symptoms:

> *I couldn't lift up my leg like I had been able to before...The next morning, the knee was barely visible due to swelling. Well, I should probably have done something the day before, but I didn't. You just know that when you call the hospital, you must go through several people, and I don't want to be a nuisance either.* [Patient 4, DC]

Lack of Information Due to Inadequate Coordination of Care

Patients in the ACL group highlighted a lack of information before surgery. Similarly, they described missing information in the first weeks after discharge, before their postoperative follow-up visit, and before starting rehabilitation with a physiotherapist:

> *Actually, I didn't know what I was supposed to do. Maybe I didn't ask enough questions before discharge. The first week (after discharge) I didn't do anything. I was wearing this DonJoy bandage and I didn't put stress on my leg or anything. And it turns out that I really should have done that.* [Patient 1, ACL]

They had questions about rehabilitation and restrictions associated with the operation, and this led to Google searches, which usually left them more confused:

> *I felt like I was in a no man's land and didn't really know what to do.* [Patient 3, ACL]

In the DC group, the patients and the parents described how knowledge is not shared across sectors in a sufficient and timely fashion. The fact that HCPs in the municipality did not have specialty-specific knowledge, as did those from the hospital, was perceived as unsafe and uncertain. They described situations in which home care nurses or physiotherapists had little or no experience with their treatment and care. That placed a massive burden on the patients or the parents to be in "control" of everything. Lack of information and coordination across sectors also led to confusion regarding the rehabilitation, for example, when the physiotherapist understood the rehabilitation plan differently than the patient remembered it. The patients and the parents from the DC group pointed out how they become the "messengers" and thus responsible for passing on information between the hospital and municipal providers. They viewed this as burdensome, expressing insecurity about accurately conveying all crucial information:

> *It's the fact that it is our interpretation of what is heard. You know, it is not necessarily medical language that we pass on to the next professional.* [Mother of patient 13, DC]
>
> *The physiotherapists often ask questions like "what did the surgeon say?" But when you have no professional knowledge, and you are busier with being there for your child, then there might be things I do not remember or consider as being important.* [Mother of patient 12, DC]

Relations and Communication Provide "Peace of Mind"

Patients and parents from both groups highlighted the importance of the relationship and communication with HCPs.

However, they had different perceptions of their actual needs. For the patients in the ACL group, the most important thing was that the HCPs were "competent." This was also valid in the DC group, but they unanimously expressed that the relationship and contact with known HCPs were just as important to them. The mother of a boy, who had been through several operations throughout his childhood, described what the relationship between her son and the HCPs at the hospital meant:

> *It gives, well, it gives you peace. It gives peace of mind even before you have to leave home (to attend surgery or follow-up visit). He can say: "Well, now we're going home to Aalborg again soon," and people will say "You don't live in Aalborg, do you?." And then he would respond: "Well, a lot of my time, I do."* [Mother of patient 7, DC]

The same perspective was elaborated by the mother of another boy:

> *I think it's about safety, trust, and recognizability, and we don't refer to it as the "doctor," we say we're going to see him (the surgeon) or her (the nurse).* [Mother of patient 15, DC]

During the initial interviews, it became clear that some patients undergoing long-term treatments in the DC group already used email or SMS text messaging for communication with the orthopedic surgeon or the physiotherapist. This was described as a workaround because traditional communication pathways did not meet their needs, such as calling the secretary, who would leave a note for the nurse or the surgeon to call the patient. The patients and the parents expressed that it made them feel supported, and thus, they largely understood the intention of eDialogue. When asked about their expectations of eDialogue, most patients and parents who had previous experiences with orthopedic surgery expressed that they wished they had had the opportunity of team-based digital communication the first time. Thus, they expected that their previous experiences of "being a patient" would minimize their need for eDialogue at this time.

Phase 2: Experiences With, and Use of, eDialogue After Discharge

Themes and Subthemes

All 31 patients or their parents included in this study were given access to eDialogue for 2 months after discharge with their team of HCPs across sectors. Interviews with 77% (24/31) of the patients and parents led to 3 overall themes and associated subthemes identified across the groups. As in the initial interviews, some subthemes were more prominent in one group than the other and thus highlighted in the table (Table 4).

Table 4. Themes and subthemes of patients and parents' experiences of using eDialogue with HCPs[a] after discharge (n=24).

Themes and subthemes	DC[b] (n=13), n (%)	ACL[c] (n=11), n (%)
Digitally enhanced coherence and proximity		
A sense of security at home	13 (100)	7 (64)
Sharing knowledge between patients and HCPs	9 (69)	5 (45)
Drivers and barriers to use		
Recognizable, informal tool and easy to use	11 (85)	8 (73)
To "be invited" to dialogue by HCPs allows use	6 (46)	4 (36)
Worry about overburdening HCPs	10 (77)	2 (18)
eDialogue rearranges communication pathways		
Reduces the need for phone calls	12 (92)	6 (55)
Text messages and photos are adequate	9 (69)	7 (64)

[a]HCP: health care professional.

[b]DC: deformity correction.

[c]ACL: anterior cruciate ligament.

Digitally Enhanced Coherence and Proximity

Across groups, patients and parents unanimously reported that the possibility of easy and direct communication with HCPs after discharge provided them with a sense of security at home. Although eDialogue was used sparingly by some patients, the possibility made them feel at ease during the rehabilitation period. For the patients who used eDialogue more, it was expressed that it helped them get through the first period after discharge because they felt "closer" to the HCPs and as if they had a constant "back up":

> *For me, it is very much about security, I almost feel that I have the surgeon by my side all the time. The first time (of surgery and discharge), I felt that he was far away.* [Patient 4, DC]

The patients in the ACL group appreciated the opportunity to ask questions, but the need for communication was most evident in the first weeks after discharge and before the first clinical follow-up and exercise sessions with physiotherapists:

> *Before my first checkup, I encountered some problems that I really wanted answered, so that I didn't have to go and wait and worry if there was something*

> *wrong. It was solved immediately in eDialogue.* [Patient 8, ACL]

For the patients and the parents in the DC group, eDialogue specifically helped HCPs share important information across sectors. They described how no longer being responsible for passing on information between the surgeon and the physiotherapist at the municipality brought relief and was highly appreciated:

> *Then we could see that they had the dialogue and then we knew that when we showed up for training next time, the physiotherapist knew it, so we didn't have to explain, which we found difficult anyway.* [Father of patient 10, DC]

In other cases, the patients described how municipal HCPs would use eDialogue indirectly to keep updated with the patient's progress just by reading the messages exchanged between the HCPs from the hospital and the patient. This provided a basis for a common point of view at the patient's next training session.

The parents of minor children described how they used eDialogue to calm their child or explain the treatment plan to them by reading them messages from HCPs.

Drivers and Barriers to Use

In both groups, the patients and the parents agreed that eDialogue presented as a recognizable and informal tool that was easy to use and that this promoted their use. The short response time was also highlighted as a main reason to use eDialogue:

> *I don't remember a day has passed, more like minutes or hours. So, it's been cool. It would never have been the case if I had to call.* [Patient 1, ACL]

Few patients experienced a late or no response. If it happened with their first question, they explained that it made them lose courage to use eDialogue another time. In general, the patients and the parents felt that the use of eDialogue was less intrusive than calling, but they also expressed worry about overburdening the HCPs. By contrast, they expected HCPs to manage their working hours themselves and assess when they had the time to respond:

> *To begin with, I thought that I would not burden the system unnecessarily...but it probably became a little more urgent and I worried about the way he was feeling, so I texted them and got a reply shortly after.* [Mother of patient 12, DC]

No patients expected answers out of hours, but some sent messages at these times to be relieved. However, they all emphasized that they could have waited for a response until the next weekday. A patient from the ACL group described her reflections about sending a message on a Friday night:

> *And of course, I thought, Oh no, now I hope he doesn't feel obliged to answer, but I also thought that they must be professional and decide for themselves.* [Patient 11, ACL]

Some patients and parents described how, before discharge, some HCPs would urge them to use eDialogue if needed and that the feeling of being invited made them more inclined to use it after coming home. The patients from the ACL group also described how eDialogue opened up the possibility to ask about "minor issues," which they might not have called about.

Among nonusers or those who used eDialogue sparingly, it was expressed that they simply did not have the need, as everything went as planned. Nonuse was also attributed to having frequent follow-ups at the outpatient clinic or attending physiotherapy several times a week.

eDialogue Rearranges Communication Pathways

The patients and the parents highlighted how the use of eDialogue had prevented phone calls or additional physical attendance after discharge; this was particularly prominent for the patients in the DC group:

> *Well, to start with we used eDialogue quite a bit I would say. As soon as we had any questions, we texted them and did not need any other forms of communication.* [Mother of patient 8, DC]

In a few cases, messages in eDialogue developed into a need for phone calls or an extra checkup in the outpatient clinic. The time of the phone call or attendance was then arranged through eDialogue. However, digital communication was perceived as adequate in most cases. There were instances where follow-up questions from HCPs were necessary, yet patients quickly felt understood and equally comprehended the answers they received:

> *Although we have not spoken on the phone, I have received sufficient information and I also feel that I have managed to communicate well.* [Mother of patient 1, DC]

A patient from the ACL group described how eDialogue was used as an extra contact for a him to "fully guard" himself. He was in doubt if the photo sent in eDialogue could show his concerns regarding the surgical site clearly enough, and therefore, he contacted his general practitioner and texted the team in eDialogue at the same time:

> *There was a situation where I had sent a message in the morning, and so, I thought I might as well, while there was still phone time at the GP, call to see if he had an available appointment. Then I came to my GP, and actually got exactly the same answer as I received on the phone (eDialogue) an hour later. So, it wasn't something that was needed as such, but now that I had the opportunity, I thought I might as well do it.* [Patient 8, ACL]

No patients expressed feelings of being misunderstood in their communication with HCPs in eDialogue. They experienced digital communication as being sufficient for their needs; however, they reflected on the risk of misunderstandings when communicating via texts:

> *I think it's a much more optimized way of doing it, because I don't need a physical conversation by phone. I'm fine with texting, but obviously there can*

be some misunderstandings or something that can go wrong and then you have to call. [Mother of patient 15, DC]

The use of photos was mentioned as being very important to support texts. A few patients explained that they lacked the possibility of sending and receiving videos; however, they emphasized that it was not a necessity for their use:

> *If I hadn't been able to send photos, then maybe I would have had to explain something visual by phone, and then I would have had to come in for a checkup, and then I would have wasted a whole day.* [Patient 1, ACL]

> *Video could be nice, but then again, the photos could effectively illustrate how the position of her leg is and show how much she has actually been able to stretch, in what positions it hurts, and so on.* [Mother of patient 17, DC]

The mother of a minor patient explained how she used eDialogue as a photo diary to keep the HCPs across sectors updated on the progress of her son's surgical wound:

> *So, when she (the home care nurse) came and changed the dressings, we took some photos before she put on new ones, and then we kind of had it (photos) from time to time and could follow how it progressed...It was smart as hell, and when it wasn't the same home care nurse coming by, we showed them the photos and at the same time kept the surgeon at the hospital up to date.* [Mother of patient 15, DC]

Use of eDialogue 2 Months After Discharge

The need for support and communication for both patient groups after discharge was expressed through the actual use of eDialogue (Table 5).

Table 5. Patients' and parents' use of eDialogue 2 months after hospital discharge.

	Total number of messages, n	Average number of messages per patient, n	Maximum number of messages per patient, n
DC[a] (n=17)			
All text messages exchanged[b]	338	19.9	54
Text messages sent by patients	189	11.2	34
Actual questions that needed a reply[c]	128	7.5	20
Photos sent by patients[d]	127	7.5	53
ACL[e] (n=14)			
All text messages exchanged	126	9.0	36
Text messages sent by patients	68	4.9	19
Actual questions that needed a reply[f]	55	3.9	14
Photos sent by patients[c]	13	0.9	6

[a]DC: deformity correction.

[b]The total number of text messages exchanged between patients and health care professionals (HCPs) 2 months after discharge.

[c]Text messages sent from the patient or their parents to the HCPs in eDialogue. The minimum number of messages or photos sent per patient was 0, as some patients did not use eDialogue at all.

[d]Actual questions that needed a reply from the HCPs are the number of individual text messages from patients or parents that were formulated as a question; thus, this does not include the back-and-forth 2-way communication that 1 question could lead to (eg, saying thank you).

[e]ACL: anterior cruciate ligament.

[f]Photos refer to the number of photos taken by the patients or parents and sent for review by the HCPs.

Of the patients or their parents, 88% (15/17) in the DC group and 79% (11/14) in the ACL group used eDialogue to ask questions to HCPs after discharge. In the DC group, 13 (87%) of the 15 active users used photos, and in the ACL group, 5 (45%) of the 11 active users sent photos to support communication. Upon inclusion in the study, the patients and the parents were informed that they could expect a response time of 24 hours during the weekdays. This was complied with in 96.2% (176/183) of the cases where a message that required a response from HCPs was sent, and the distribution was equal across groups.

Among users of eDialogue in the DC group, the minimum number of per-patient questions that needed a reply from HCPs was 2, and the maximum was 20. For the ACL group, there was a minimum of 1 and a maximum of 14 questions that needed a reply in 1 dialogue. Thus, there was a marked difference in the individual's use of eDialogue during the study period in both groups.

Most of the communication took place from Monday to Friday; thus, 84.7% (155/183) of the questions that needed a reply from the HCPs were sent and replied to during the weekdays.

The patients and the parents in the DC group used eDialogue throughout the 2 months (Figures 2 and 3), and 15 (88%) of the

17 patients requested to keep on using it after the data collection stopped at 2 months. The patients in the ACL group primarily used eDialogue for the first 2 to 3 weeks after discharge (Figures 2 and 3), and use then faded. Only 2 (6%) of the 31 patients or parents expressed a need to continue with eDialogue after 2 months.

Content analysis of the messages in eDialogue revealed 9 overall categories, including treatment-related issues, rehabilitation and restrictions, concerns about symptoms and complications, medication, psychological support, interdisciplinary and cross-sectoral dialogue, coordination and practical needs, updates and gratitude, HCP ask for feedback. The categories were identified across groups; however, some categories were more prominent in one group than the other (Multimedia Appendix 2).

Figure 2. The number of individual text messages sent from patients or parents to the health care professionals in eDialogue per week 8 weeks after discharge. ACL: anterior cruciate ligament; DC: deformity correction.

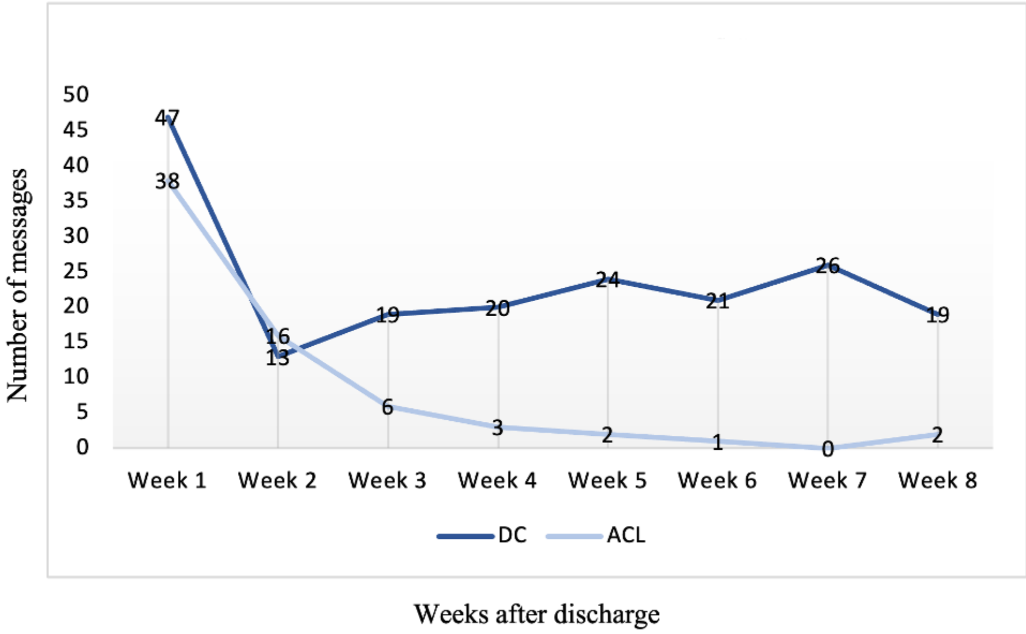
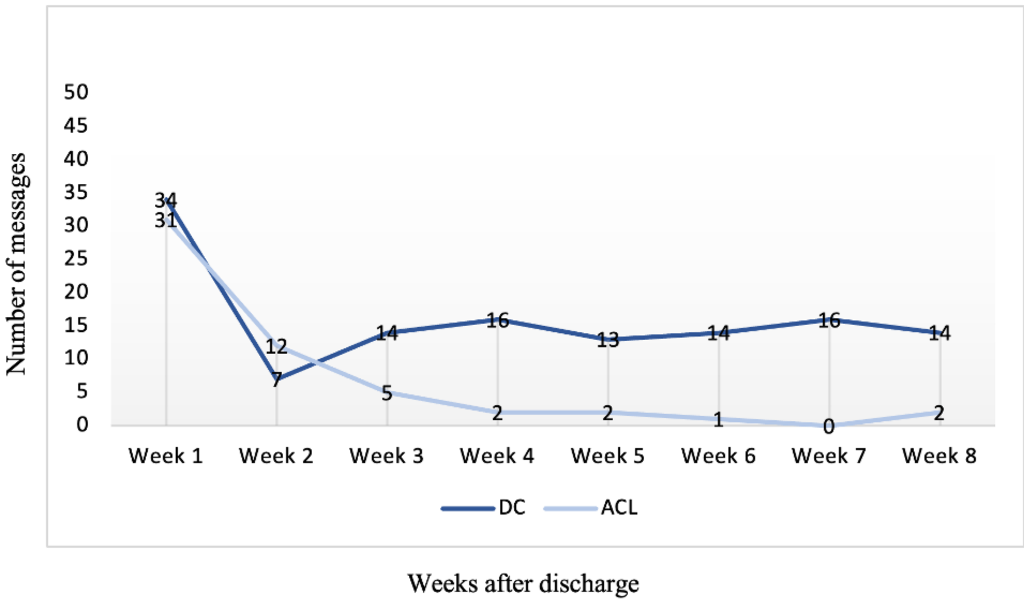


Figure 3. The number of messages sent by patients or parents that required a response from health care professionals, that is, messages phrased as a question, per week 8 weeks after discharge. ACL: anterior cruciate ligament; DC: deformity correction.



Discussion

Principal Findings

Overview

This study explored the perspectives of patients undergoing orthopedic surgery on current communication pathways (phase 1), and their subsequent experiences of using eDialogue after discharge, as well as the actual use of the solution (phase 2).

In phase 1, we identified unmet needs among patients regarding communication with HCPs after discharge. The themes involved perspectives of difficult communication pathways, lack of information due to inadequate coordination of care, and that relation and communication provide "peace of mind." In phase 2, the participants were set up to use eDialogue for 2 months after surgery and discharge, providing them access to direct digital communication with their individual health care team across settings. Through follow-up interviews, they articulated the following themes: digitally enhanced coherence and proximity, drivers and barriers to use, and that eDialogue rearranges communication pathways. Use data of eDialogue supported the experiences expressed in the interviews and provided an overview of the actual use. These findings will be discussed with the theoretical framework of COC and previous research.

Signs of Improved COC With eDialogue

Through initial interviews, the patients and the parents expressed a need for more clear communication pathways after discharge. A patient expressed that it felt like being in a "no man's land." As such, they lacked communicative support at home as well as optimized sharing of knowledge between the HCPs involved in their treatment and care across settings, indicating that informational and management COC is under pressure [32]. Similar findings are described in other studies on patients' experiences of the transition from hospital to home following surgery [24,28], and this emphasizes the need to address communicative challenges around hospital discharge.

The patients and the parents in complex and long-term orthopedic treatments (DC) experienced a greater need for continuous contact with their known health care team than those undergoing day surgery (ACL). Thus, the relationship, trust, and mutual understanding with the HCPs were described as being of great importance for their experience of security. For these patients, access to eDialogue was particularly useful, suggesting that eDialogue may play a role in facilitating relational COC. The patients in the ACL group, despite still having an unmet need for information, expressed that "less" would have been suitable for them. As digital communication becomes more prevalent in health care [1,2,4,5,7,9], comprehensive evaluations are crucial, including efficiency and optimal resource use considerations. Some patients may find less resource-intensive options, such as automated text message interventions, sufficient [36].

Through follow-up interviews, the patients and the parents across groups highlighted that eDialogue provided easy access to relevant HCPs and facilitated coherence and proximity after returning home, leading to "a sense of security." These findings corroborate previous studies [14,37] and support our assumption that team-based digital communication may contribute to improving patients' experiences of COC in transitions from hospital to home [32]. Other studies have also highlighted that COC is one of the factors that can be positively influenced by the use of team-based digital communication [15,16]. Voruganti et al [15] evaluated the feasibility of integrating a web-based communication tool for collaborative care in a pilot randomized controlled trial and found evidence indicating an increase in COC scores in the intervention group; however, the study was unpowered to show the effect statistically. Another study by Lindkvist et al [16] described how access to and use of an eHealth device for text-based communication, image exchange, and data reports between HCPs and parents of preterm infants or pediatric surgery was experienced positively in the transfer period from hospital to home. Moreover, they reported that parents felt it gave a sense of "shared responsibility," which was also expressed by the patients and parents in this study. Thus, they highlighted that eDialogue facilitated the sharing of information, so they no longer had to be the ones passing on information and knowledge between HCPs. This was a role that they often disliked or mistrusted that they could fulfill adequately. The findings from this study indicate, in line with other studies [14-16], that digital team-based communication has the potential to set the framework for interdisciplinary and cross-sector collaboration that supports COC following hospital discharge. Whether team-based digital communication can actually enhance levels of COC to an extent where it can be measured remains to be investigated.

Patients Want to Communicate Digitally

As seen in other studies on digital asynchronous communication [15,16,38], use data demonstrated that most patients and parents across groups used eDialogue (26/31, 84%). The drivers to use eDialogue involved that the tool was recognizable and easy to use. Employing a messenger-like tool, made available to patients on their own smartphone, was a strength, as we did not encounter technical challenges as described in other studies, where devices were newly developed and delivered to participants [16]. The simple solution only allowed for communication in text and photos, and it may lack other options for patients who cannot use the text-based medium. Although previous studies involving text-based digital communication for health care purposes show that patients largely adopt this form of communication across settings and needs [4,10,37,39], digital inclusion in eHealth interventions is important to acknowledge both in regard to the hardware as well as patients' ability to use the solutions [40]. As such, if the patients cannot use the tool, no value has been added. Other studies have integrated several means of communication into their solutions, including text, video, photos, and voice recordings, and found that video communication was especially useful [16,41,42]. This is in contrast to our findings, where patients expressed that the text-based medium was sufficient for them in the postoperative period. However, we acknowledge that eDialogue, as used in this study, may not be sufficient for all patients. When designing and implementing digital communication solutions, considering patients' literacy and eHealth literacy becomes crucial to ensure equal access to health care [40,43]. Integrating multiple communication modalities within a single solution could serve as a means to achieve this goal.

A driver mentioned in this study was the informality of the solution, and that it felt less interrupting than calling by phone. Similar results have been found in other studies of digital text-based communication [16,37], and this indicates a high degree of acceptance and usability of the solution from the patients' perspectives. With an increasing level of smartphone use in the general population, digital communication becomes a more natural choice when patients need to contact providers. Thus, statistics show that the use of smartphones worldwide is increasing significantly, and in Denmark, it is estimated that 90% of all households own a smartphone [44]. As a barrier to use, the patients expressed concerns about wasting the HCPs' time. This is important to consider when implementing solutions for digital communication. Our findings indicate that this may be offset by a more inviting approach from the HCPs, as some patients and parents expressed this as a facilitator to their use. Previous studies have pointed out the importance of clearly communicating response times when using digital communication [16,37]. Similar findings were highlighted in follow-up interviews of this study, where patients and parents described quick response times, or alternatively late responses, as a driver and a barrier, respectively.

Across groups, the patients and the parents expressed that eDialogue, despite only being an addition to existing communication channels, had rearranged the communication pathways significantly. This became obvious as the patients and parents described a reduced need to call the hospital, as they found eDialogue adequate and exhaustive for their needs. These findings corroborate previous studies showing a potential decrease in phone calls to the hospital after discharge when digital communication is being used [16,45]. By contrast, another study reported, in line with our study, that some questions asked by patients in a digital communication tool were not something they would have called about and thereby indicate that access to digital communication may contribute to an increased consumption of health care resources [16]. To evaluate the effect on resource use, a randomized controlled trial should be performed. Future studies designed to demonstrate the effects on health resource use are desired to shed light on whether digital communication actually reduces patients' use of other forms of communication channels or adds on. In addition, it should be considered whether digital communication provides better quality, for example, defined as COC, patient satisfaction, and security for patients.

This study adds to the knowledge of patients' perspectives on current communication pathways and the sparse evidence of their experiences and use of digital team-based communication, specifically in an orthopedic surgery setting. This may inform future interventions of team-based digital communication, from its application in clinical practice to organizational and management levels.

Limitations

The study has limitations that may affect the interpretation of our results. First, inclusion criteria were participants who owned and used a smartphone and could speak and write Danish well enough to send text messages. Second, we explored the perspectives of 2 selected groups of patients undergoing orthopedic surgery. Therefore, the external validity of the results is unknown for other groups of patients undergoing orthopedic surgery, than the ones we explored.

In planning the study, we decided that initial interviews with patients and parents in phase 1, who were subsequently recruited to use eDialogue after discharge, were appropriate to identify patients' perspectives on current communication pathways. However, some patients found it difficult to express themselves about this, as they had no or little previous experience of an orthopedic surgery context. In addition, there was a risk that the use of initial interviews combined with follow-up interviews within a short timespan (2 months) may have influenced the patients' expressed attitudes in favor of the intervention in the follow-up interviews. Reflecting on this, it might have been better to perform initial interviews with a group of patients who were not given access to eDialogue afterwards.

In this study, we did not use log files to summarize the use data, as other studies have done [16,26], and this may be perceived as a limitation. However, we argue that log files, which report the number of log-in attempts, database entries, messages sent in total and the like, would not show the actual use as it presented to the participants in clinical practice. Therefore, manual counts were used to remove messages saying “thank you” or similar, as these are not considered relevant to the use of eDialogue in a health care setting.

Overall, the 24-hour weekday response time was met in this study and some patients reported extremely fast responses from HCPs. This finding must be interpreted with caution, as we cannot rule out that it is due to the Hawthorne effect, which suggests that people behave better when they are observed [46]. Conversely, it can also be an expression of the flexibility that lies in the digital asynchronous form of communication, giving HCPs the possibility to answer when they have the time for it, or it may simply reflect that the HCPs replied instantly (when able to) not to forget it. Nevertheless, an exclusively positive interpretation of compliance with the response time in this study may result in blindness toward the possible pitfalls that can occur in the real world if eDialogue is implemented. Insights from the perspective of HCPs can reveal this.

Conclusions

The findings from this study indicate that the patients and the parents experienced an unmet need related to communication and collaboration following hospital discharge. eDialogue was overall evaluated positively, and the patients and parents perceived team-based digital communication as correspondent to their needs and suggested that it provided a sense of security after returning home. COC may be enhanced by assembling the team of HCPs in a simple digital communication solution with patients. However, eDialogue should be further evaluated and tested. Future research has to explore HCPs' perspectives on the solution as well as establish the effects and organizational and economic incentives to use team-based digital communication in the context of orthopedic surgery care pathways.

Acknowledgments

The authors would like to thank all patients, parents, health care professionals, and management from departments across hospitals and municipalities in Northern Jutland, Denmark, who agreed to participate in this study. The study was partly funded by the Research Fund of Danish Health Region North (2021-0019). Visma, the company who owns the technical solution used in this study ("LetDialog"), had no role in the design of the study; collection, analyses, or interpretation of data; or in the writing of the manuscript or the decision to publish the results.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guides.
[DOCX File , 16 KB - humanfactors_v11i1e49696_app1.docx]

Multimedia Appendix 2
The number of patients from the deformity correction (DC) and anterior cruciate ligament (ACL) group who sent messages within the respective categories. HCP: health care professional.
[PNG File , 88 KB - humanfactors_v11i1e49696_app2.png]

References

1. Hamza M, Khan HS, Sattar ZA, Hanif M. Doctor-patient communication in surgical practice during the coronavirus (COVID-19) pandemic. Br J Surg 2020 Jun;107(7):e193 [FREE Full text] [doi: 10.1002/bjs.11661] [Medline: 32364265]
2. Cronin RM, Davis SE, Shenson JA, Chen Q, Rosenbloom ST, Jackson GP. Growth of secure messaging through a patient portal as a form of outpatient interaction across clinical specialties. Appl Clin Inform 2015;6(2):288-304 [FREE Full text] [doi: 10.4338/ACI-2014-12-RA-0117] [Medline: 26171076]
3. Using e-health and information technology to improve health. World Health Organization. URL: https://www.who.int/westernpacific/activities/using-e-health-and-information-technology-to-improve-health [accessed 2024-03-21]
4. Newhouse N, Lupiáñez-Villanueva F, Codagnone C, Atherton H. Patient use of email for health care communication purposes across 14 European countries: an analysis of users according to demographic and health-related factors. J Med Internet Res 2015 Mar 06;17(3):e58 [FREE Full text] [doi: 10.2196/jmir.3700] [Medline: 25798912]
5. Baer D. Patient-physician e-mail communication: the kaiser permanente experience. J Oncol Pract 2011 Jul;7(4):230-233 [FREE Full text] [doi: 10.1200/JOP.2011.000323] [Medline: 22043186]
6. McAlearney AS, Sieck CJ, Gaughan A, Fareed N, Volney J, Huerta TR. Patients' perceptions of portal use across care settings: qualitative study. J Med Internet Res 2019 Jun 06;21(6):e13126 [FREE Full text] [doi: 10.2196/13126] [Medline: 31172960]
7. Semple JL, Armstrong KA. Mobile applications for postoperative monitoring after discharge. CMAJ 2017 Jan 09;189(1):E22-E24 [FREE Full text] [doi: 10.1503/cmaj.160195] [Medline: 27920015]
8. Mehta SJ, Hume E, Troxel AB, Reitz C, Norton L, Lacko H, et al. Effect of remote monitoring on discharge to home, return to activity, and rehospitalization after hip and knee arthroplasty: a randomized clinical trial. JAMA Netw Open 2020 Dec 01;3(12):e2028328 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.28328] [Medline: 33346847]
9. Danbjørg DB, Wagner L, Kristensen BR, Clemensen J. Intervention among new parents followed up by an interview study exploring their experiences of telemedicine after early postnatal discharge. Midwifery 2015 Jun;31(6):574-581. [doi: 10.1016/j.midw.2015.02.007] [Medline: 25765743]
10. Garne Holm K, Brødsgaard A, Zachariassen G, Smith AC, Clemensen J. Parent perspectives of neonatal tele-homecare: a qualitative study. J Telemed Telecare 2019 May;25(4):221-229. [doi: 10.1177/1357633X18765059] [Medline: 29792079]
11. Atherton H, Sawmynaden P, Sheikh A, Majeed A, Car J. Email for clinical communication between patients/caregivers and healthcare professionals. Cochrane Database Syst Rev 2012;11:CD007978. [doi: 10.1002/14651858.CD007978.pub2] [Medline: 23152249]
12. de JC, Ros WJ, Schrijvers G. The effects on health behavior and health outcomes of internet-based asynchronous communication between health providers and patients with a chronic condition: a systematic review. J Med Internet Res 2014;16(1):e19 [FREE Full text] [doi: 10.2196/jmir.3000] [Medline: 24434570]
13. Wibe T, Hellesø R, Varsi C, Ruland C, Ekstedt M. How does an online patient-nurse communication service meet the information needs of men with recently diagnosed testicular cancer? ISRN Nurs 2012;2012:260975 [FREE Full text] [doi: 10.5402/2012/260975] [Medline: 23251816]

14. Ekstedt M, Børøsund E, Svenningsen IK, Ruland CM. Reducing errors through a web-based self-management support system. Stud Health Technol Inform 2014;201:328-334. [Medline: 24943563]
15. Voruganti T, Grunfeld E, Jamieson T, Kurahashi AM, Lokuge B, Krzyzanowska MK, et al. My team of care study: a pilot randomized controlled trial of a web-based communication tool for collaborative care in patients with advanced cancer. J Med Internet Res 2017 Jul 18;19(7):e219 [FREE Full text] [doi: 10.2196/jmir.7421] [Medline: 28720558]
16. Lindkvist RM, Sjöström-Strand A, Landgren K, Johnsson BA, Stenström P, Hallström IK. "In a way we took the hospital home"-a descriptive mixed-methods study of parents' usage and experiences of eHealth for self-management after hospital discharge due to pediatric surgery or preterm birth. Int J Environ Res Public Health 2021 Jun 15;18(12):6480 [FREE Full text] [doi: 10.3390/ijerph18126480] [Medline: 34203985]
17. Continuity and coordination of care: a practice brief to support implementation of the WHO framework on integrated people-centred health services. World Health Organization. 2018. URL: https://www.who.int/publications/i/item/9789241514033 [accessed 2024-03-21]
18. Naylor MD, Aiken LH, Kurtzman ET, Olds DM, Hirschman KB. The care span: the importance of transitional care in achieving health reform. Health Aff (Millwood) 2011 Apr;30(4):746-754 [FREE Full text] [doi: 10.1377/hlthaff.2011.0041] [Medline: 21471497]
19. Forster AJ, Murff HJ, Peterson JF, Gandhi TK, Bates DW. The incidence and severity of adverse events affecting patients after discharge from the hospital. Ann Intern Med 2003 Feb 4;138(3):161-167. [Medline: 12558354]
20. Coleman EA, Parry C, Chalmers S, Min S. The care transitions intervention: results of a randomized controlled trial. Arch Intern Med 2006 Sep 25;166(17):1822-1828. [doi: 10.1001/archinte.166.17.1822] [Medline: 17000937]
21. Goodwin JS. Continuity of care matters in all health care settings. JAMA Netw Open 2021 Mar 01;4(3):e213842 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.3842] [Medline: 33760086]
22. Oslock WM, Satiani B, Way DP, Tamer RM, Maurer J, Hawley JD, et al. A contemporary reassessment of the US surgical workforce through 2050 predicts continued shortages and increased productivity demands. Am J Surg 2022 Jan;223(1):28-35. [doi: 10.1016/j.amjsurg.2021.07.033] [Medline: 34376275]
23. Wainwright T, Middleton R. An orthopaedic enhanced recovery pathway. Curr Anaesth Crit Care 2010 Jun;21(3):114-120. [doi: 10.1016/j.cacc.2010.01.003]
24. Halding AG, Hovland G, Johansen VA, Hege Eide P, Åslaug Kringeland T. A planned but vulnerable pathway towards self-management following day surgery. Orthopaedic patients' experiences. Scand J Caring Sci 2021 Dec;35(4):1114-1122. [doi: 10.1111/scs.12926] [Medline: 33128304]
25. Jensen LW, Madsen NA, Dinesen B. Developing a digital platform for telerehabilitation of patients treated with external fixation device after complex tibia fractures. In: Proceedings of the 17th Scandinavian Conference on Health Informatics. 2019 Presented at: SHI '19; November 12-13, 2019; Oslo, Norway p. 154-159 URL: https://vbn.aau.dk/en/publications/developing-a-digital-platform-for-telerehabilitation-of-patients-
26. Husain A, Cohen E, Dubrowski R, Jamieson T, Kurahashi AM, Lokuge B, et al. A clinical communication tool (Loop) for team-based care in pediatric and adult care settings: hybrid mixed methods implementation study. J Med Internet Res 2021 Mar 03;23(3):e25505 [FREE Full text] [doi: 10.2196/25505] [Medline: 33656445]
27. Gulmans J, Vollenbroek-Hutten M, van Gemert-Pijnen L, van Harten W. A web-based communication system for integrated care in cerebral palsy: experienced contribution to parent-professional communication. Int J Integr Care 2012;12:e9 [FREE Full text] [doi: 10.5334/ijic.672] [Medline: 22977435]
28. Balakrishnan AS, Palmer NR, Odisho AY. "Can you hear me now?": postoperative patient-initiated communication with providers. Am J Manag Care 2020 Oct 01;26(10):e333-e341 [FREE Full text] [doi: 10.37765/ajmc.2020.88507] [Medline: 33094946]
29. Kom tættere på dine borgere med sikker SMS: first agenda. LetDialog. URL: https://letdialog.dk/ [accessed 2023-02-02]
30. DeWalt KM, DeWalt BR. Participant Observation: A Guide for Fieldworkers. Walnut Creek, CA: AltaMira Press; 2011.
31. Brinkmann S, Kvale S. InterViews: Learning the Craft of Qualitative Research Interviewing. København, Denmark: SAGE Publications; 2015.
32. Haggerty JL, Reid RJ, Freeman GK, Starfield BH, Adair CE, McKendry R. Continuity of care: a multidisciplinary review. BMJ 2003 Nov 22;327(7425):1219-1221 [FREE Full text] [doi: 10.1136/bmj.327.7425.1219] [Medline: 14630762]
33. Bahr SJ, Weiss ME. Clarifying model for continuity of care: a concept analysis. Int J Nurs Pract 2019 Apr;25(2):e12704. [doi: 10.1111/ijn.12704] [Medline: 30393894]
34. Reid R, Haggerty J, McKendry R. Defusing the confusion: concepts and measures of continuity of healthcare. Canadian Health Services Research Foundation. 2002. URL: https://www.researchgate.net/publication/245856177_Defusing_the_Confusion_Concepts_and_Measures_of_Continuity_of_Health_Care#fullTextFileContent [accessed 2024-03-23]
35. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]

36. Gordon CR, Rezzadeh KS, Li A, Vardanian A, Zelken J, Shores JT, et al. Digital mobile technology facilitates HIPAA-sensitive perioperative messaging, improves physician-patient communication, and streamlines patient care. Patient Saf Surg 2015;9:21 [FREE Full text] [doi: 10.1186/s13037-015-0070-9] [Medline: 26136830]
37. Alexander KE, Ogle T, Hoberg H, Linley L, Bradford N. Patient preferences for using technology in communication about symptoms post hospital discharge. BMC Health Serv Res 2021 Feb 15;21(1):141 [FREE Full text] [doi: 10.1186/s12913-021-06119-7] [Medline: 33588829]
38. Lau C, Churchill RS, Kim J, Matsen FA, Kim Y. Asynchronous web-based patient-centered home telemedicine system. IEEE Trans Biomed Eng 2002 Dec;49(12):1452-1462. [doi: 10.1109/TBME.2002.805456] [Medline: 12542241]
39. Hefner JL, MacEwan SR, Biltz A, Sieck CJ. Patient portal messaging for care coordination: a qualitative study of perspectives of experienced users with chronic conditions. BMC Fam Pract 2019 May 03;20(1):57 [FREE Full text] [doi: 10.1186/s12875-019-0948-1] [Medline: 31053063]
40. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
41. DeAntonio JH, Kang HS, Cockrell HC, Rothstein W, Oiticica C, Lanning DA. Utilization of a handheld telemedicine device in postoperative pediatric surgical care. J Pediatr Surg 2019 May;54(5):1005-1008. [doi: 10.1016/j.jpedsurg.2019.01.032] [Medline: 30782441]
42. Gund A, Sjöqvist BA, Wigert H, Hentz E, Lindecrantz K, Bry K. A randomized controlled study about the use of eHealth in the home health care of premature infants. BMC Med Inform Decis Mak 2013 Feb 09;13:22 [FREE Full text] [doi: 10.1186/1472-6947-13-22] [Medline: 23394465]
43. WHO guideline: recommendations on digital interventions for health system strengthening. Licence: CC BY-NC-SA 3.0 IGO. World Health Organization. 2019. URL: https://www.who.int/publications/i/item/9789241550505 [accessed 2024-03-21]
44. Tassy A, Berg CE. It-anvendelse i befolkningen 2021. Danmarks Statistik: Forskning, Teknologi og Kultur. 2021. URL: https://www.dst.dk/Site/Dst/Udgivelser/GetPubFile.aspx?id=39431&sid=itbef2021 [accessed 2024-03-21]
45. Armstrong KA, Coyte PC, Brown M, Beber B, Semple JL. Effect of home monitoring via mobile app on the number of in-person visits following ambulatory surgery: a randomized clinical trial. JAMA Surg 2017 Jul 01;152(7):622-627 [FREE Full text] [doi: 10.1001/jamasurg.2017.0111] [Medline: 28329223]
46. Silverman WA. Hawthorne effects. In: Silverman WA, editor. Hawthorne Effects. In Where's The Evidence? Debates in Modern Medicine. New York, NY: Oxford University Press; 1999:58-51.

Abbreviations

ACL: anterior cruciate ligament
COC: continuity of care
DC: deformity correction
HCP: health care professional

Edited by A Kushniruk; submitted 06.06.23; peer-reviewed by N Finn, L Shang; comments to author 24.01.24; revised version received 29.01.24; accepted 20.02.24; published 29.03.24.

Please cite as:
Jensen LWH, Rahbek O, Lauritsen REK, Kold S, Dinesen B
Patient Perspectives on Communication Pathways After Orthopedic Surgery and Discharge and Evaluation of Team-Based Digital Communication: Qualitative Exploratory Study
JMIR Hum Factors 2024;11:e49696
URL: https://humanfactors.jmir.org/2024/1/e49696
doi:10.2196/49696
PMID:38551641

Original Paper

The Asthma App as a New Way to Promote Responsible Short-Acting Beta2-Agonist Use in People With Asthma: Results of a Mixed Methods Pilot Study

Liselot N van den Berg[1,2], MSc; Cynthia Hallensleben[1,2]; Lisa AE Vlug[1,2], MSc; Niels H Chavannes[1,2], MD, PhD; Anke Versluis[1,2], PhD

[1]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, Netherlands

[2]National eHealth Living Lab, Leiden, Netherlands

Corresponding Author:
Liselot N van den Berg, MSc
Department of Public Health and Primary Care
Leiden University Medical Center
Hippocratespad 21
Leiden, 2300 RC
Netherlands
Phone: 31 71 526 57 61
Email: L.N.van_den_Berg@lumc.nl

Abstract

Background: Approximately 262 million people worldwide are affected by asthma, and the overuse of reliever medication—specifically, short-acting beta2-agonist (SABA) overuse—is common. This can lead to adverse health effects. A smartphone app, the Asthma app, was developed via a participatory design to help patients gain more insight into their SABA use through monitoring and psychoeducation.

Objective: This pilot study aims to evaluate the feasibility and usability of the app. The preliminary effects of using the app after 3 months on decreasing asthma symptoms and improving quality of life were examined.

Methods: A mixed methods study design was used. Quantitative data were collected using the app. Asthma symptoms (measured using the Control of Allergic Rhinitis and Asthma Test) and the triggers of these symptoms were collected weekly. Quality of life (36-Item Short-Form Health Survey) was assessed at baseline and after 3, 6, and 12 months. User experience (System Usability Scale) was measured at all time points, except for baseline. Furthermore, objective user data were collected, and qualitative interviews, focusing on feasibility and usability, were organized. The interview protocol was based on the Unified Theory of Acceptance and Use of Technology framework. Qualitative data were analyzed using the Framework Method.

Results: The baseline questionnaire was completed by 373 participants. The majority were female (309/373, 82.8%), with a mean age of 46 (SD 15) years, and used, on average, 10 SABA inhalations per week. App usability was rated as good: 82.3 (SD 13.2; N=44) at 3 months. The Control of Allergic Rhinitis and Asthma Test score significantly improved at 3 months (18.5) compared with baseline (14.8; β=.189; SE 0.048; P<.001); however, the obtained score still indicated uncontrolled asthma. At 3 months, there was no significant difference in the quality of life. Owing to the high dropout rate, insufficient data were collected at 6 and 12 months and were, therefore, not further examined. User data showed that 335 users opened the app (250/335, 74.6%, were returning visitors), with an average session time of 1 minute, and SABA registration was most often used (7506/13,081, 57.38%). Qualitative data (from a total of 4 participants; n=2, 50% female) showed that the participants found the app acceptable and clear. Three participants stated that gaining insight into asthma and its triggers was helpful. Two participants no longer used the app because they perceived their asthma as controlled and, therefore, did not use SABA often or only used it regularly based on the advice of the pulmonologist.

Conclusions: The initial findings regarding the app's feasibility and usability are encouraging. However, the notable dropout rate underscores the need for a cautious interpretation of the results. Subsequent studies, particularly those focusing on implementation, should explore the potential integration of the app into standard treatment practices.

(JMIR Hum Factors 2024;11:e54386) doi:10.2196/54386

KEYWORDS

asthma; short-acting beta2-agonist; SABA overuse; app; eHealth; feasibility; usability; mobile phone

Introduction

Asthma is a common chronic inflammatory disease, which is estimated to affect 262 million people worldwide [1]. Step 1 of medical treatment involves the prescription of short-acting beta2-agonist (SABA) as a reliever medication. In contrast to inhaled corticosteroids (ICS), SABA does not have an anti-inflammatory effect on the respiratory tract [2,3]. In 2019, step 1 was modified in the Global Initiative for Asthma guidelines [2]. Specifically, the option of a low dose of ICS-formoterol, as needed, was added because asthma control is often suboptimal [3-5]. According to guidelines, using SABA more than twice a week indicates suboptimal, uncontrolled asthma [3]. Approximately half of the patients with asthma have uncontrolled asthma [5-7]. The overuse of SABA is linked to an increased risk of asthma exacerbations, which are associated with damage to the respiratory tract, asthma-related hospitalization, and visits to the emergency department [8-12].

The overuse of SABA is common for different reasons. First, individuals often overuse their SABA instead of taking ICS to achieve a rapid relief from an asthma attack [13-15]. Second, individuals may lack knowledge about the medication and insight into the actual frequency of medication use [13,16]. For example, the REcognise Asthma and LInk to Symptoms and Experience study [17] found that 80% of the participants thought they had controlled asthma, although 40% had used their SABA ≥3 times during the past week. A post hoc analysis of the Dutch participants from this study showed that 60% of the patients with asthma overused their SABA in the previous week [18].

Previous studies have shown that self-management apps can help reduce the frequency of SABA use, increase SABA-free days, and improve overall asthma control [19,20]. These apps can also boost individuals' confidence in managing asthma and improve their quality of life (QoL) [21-23]. Often, these self-management tools include education, self-monitoring, and feedback to support the end users in managing their disease daily [21,22,24,25]. Most apps are developed using state-based models, such as the Waterfall Model, and agile methods [26]. These traditional methods do not engage end users in the development process, which may result in lower usability and adherence of end users [27]. Therefore, an app was developed in collaboration with end users and other relevant stakeholders (eg, health care professionals) using a participatory design. This design can be used to engage relevant stakeholders during the development process, which may improve the usability and adherence to an app. The objective of the app is to help patients gain more insight into their SABA use while also promoting responsible SABA use. This may eventually decrease SABA overuse. In a previous study, we described the development process of the Asthma app [28].

This pilot study, using a mixed methods design, aims to examine (1) the feasibility and usability of the app in people with asthma and (2) the preliminary effects of using the app after 3 months on decreasing asthma symptoms and improving QoL.

Methods

Design and Population

The pilot study had a mixed methods design. Initially, the study was purely quantitative, with data collected through questionnaires administered in the app to examine the usability and preliminary effects of the app. Individuals were eligible to participate if they (1) were aged ≥18 years and (2) had asthma. Individuals who did not meet these inclusion criteria were excluded from the study; however, they could still use the app. The study period for the participants was 12 months. The study was conducted from January 15, 2021, to December 6, 2022; however, user data were collected until December 31, 2021. User data collection was stopped earlier because the costs for collecting these data increased after 2021, and this could no longer be funded.

During the study, we noticed that most participants used the app only in the first week after downloading. Owing to the high dropout rate, an additional qualitative study was conducted to examine the feasibility and usability of the app in more detail. Individuals who use, had used, or had downloaded the app once were included in the semistructured interviews. Individuals who participated in the qualitative interviews did not necessarily participate in the quantitative study. Qualitative interviews were held until data saturation was reached; data saturation was expected after 6 to 12 interviews [29,30]. Data were collected between November 7, 2022, and December 13, 2022.

Ethical Considerations

According to the Medical Ethics Committee of the Leiden University Medical Center, this study did not fall within the scope of the Dutch Medical Research Involving Human Subjects Act (N20.103). Subsequently, a declaration of no objection was obtained from the Medical Ethics Committee. Participants provided informed consent and were able to opt out (see the *Procedure* section). The quantitative data were collected anonymously, and the qualitative data were collected pseudonymously.

Asthma App

The Asthma app (a Dutch app developed by the Leiden University Medical Center and Innovattic; Figure 1) allows end users to register their SABA use. Moreover, users can register asthma symptoms weekly (they receive a notification to do so), and they can register the triggers of these symptoms at any time. A graph shows how SABA use, asthma symptoms, and their triggers are related. The amount of SABA used was compared with the existing guidelines [2,3] or, when applicable, with health care professional's advice. Psychoeducation is also included, covering topics such as *what is asthma* and *types of medication and their function* [28]. The app was available free of charge in the App Store and Google Play Store.

Figure 1. Visuals of the final version of the Asthma app: (A) landing page where users can register their short-acting beta2-agonist (SABA); (B) the graphical overview or statistics (in Dutch statistieken) where users can get insight into their SABA use, asthma symptoms, and asthma triggers; and (C) psychoeducation or information (in Dutch informatie) where users can learn more about their asthma and the app. On the landing page, users can receive three different messages based on the number of registered SABA compared with the prescription: (1) "You can still use your SABA # times this week," (2) "You are at the maximum recommended dose of SABA for this week," or (3) "You needed more SABA this week than advised." After downloading the app, users receive an explanation on how to interpret the graphical overview, and this explanation can also be found in the informational part of the app.

Procedure

Quantitative Data

Different channels were used to announce the app's go-live and to recruit participants. Relevant organizations (eg, Lung Foundation Netherlands and National eHealth Living Lab) posted the information on their website and social media, or only on their website or social media, and the closed Facebook group *Asthma and Peers* in the Netherlands published the information as well. The information was further communicated through publications (ie, via the COPD Asthma General Practitioners Advice Group in a magazine for pharmacists assistants and in a national newspaper in the Netherlands). Moreover, flyers were distributed via general practices.

After downloading and installing the app, individuals were asked 2 questions to determine their eligibility for the study (ie, whether they were aged ≥18 years and had asthma). Eligible individuals were given information about the study and could decide whether they wanted to participate by signing an informed consent form in the app. In the app, participants could view the informed consent form and withdraw from the study at any moment if they wanted to. If individuals chose to withdraw their consent, they could continue using the app.

Next, participants were asked to complete the demographic and clinical characteristics questionnaire and the baseline questionnaire about QoL (ie, 36-Item Short-Form Health Survey [SF-36] [31]) and intentions to change behavior (ie, a short version of the Theoretical Domains Framework [32]). Asthma symptoms were measured weekly using the Control of Allergic Rhinitis and Asthma Test (CARAT) [33,34]. The triggers of the asthma symptoms, such as dust mites and hay fever, were asked at the end of the CARAT, and the user could also enter additional triggers throughout the week. At 3, 6, and 12 months, user experience (ie, System Usability Scale [SUS]; [35]) and QoL were assessed. No compensation was provided for completing the questionnaire.

Qualitative Data

To gain more insight into the usability and experiences with the app, the following recruitment text was used: "NeLL is looking for (former) users of the Asthma app to get more insight into the usability and experiences with the app, during a one-time interview." We recruited participants for the semistructured interviews via relevant organizations (eg, Asthma Association of the Netherlands and Davos and National eHealth Living Lab) that posted the information on their website and social media, or only on their website or social media; the information was also posted in the closed Facebook group *Asthma and Peers* in the Netherlands. To increase the interview response rate, participants were recruited via the personal channels of the researchers. When a participant was recruited via personal channels, the researcher did not conduct the interview.

Interested individuals could contact the researchers via email. Subsequently, 1 of the researchers (LNvdB and AEV) would contact them to determine whether they were eligible to participate (ie, aged ≥18 years, having asthma, and [at least] having downloaded the app). Eligible individuals interested in participating received the informed consent form via email. The participants could sign the informed consent form digitally via Castor (ie, a digital, secure research environment) [36]. After signing the informed consent form, the participants received an email invitation to schedule the semistructured interview. We aimed to enroll individuals who use the app and former users (ie, those who had at least downloaded the app).

An interview protocol was developed (Multimedia Appendix 1) based on the Unified Theory of Acceptance and Use of Technology (UTAUT) framework [37]. These interviews were conducted to better understand the perceived usability and feasibility. Interviews were conducted web-based via Microsoft Teams and lasted between 30 and 45 minutes. The participants received a gift card of 30 euros (US $31.2).

Outcome Measures

Demographic and Clinical Characteristics

General information about the participants and their asthma was obtained, including gender, age (birth year), level of education, type of asthma, degree of asthma control, and type of medication. Multiple answers could be selected when answering the question about the type of asthma (ie, allergic asthma, nonallergic asthma, exercise asthma, severe asthma, and do not know) and medication (ie, SABA, ICS, long-acting beta2-agonist (LABA), ICS+LABA, do not know, and no medication use). Furthermore, the participants were asked whether they had received specific advice from their general practitioner on how much SABA they could use per week. When the participant had not received specific advice or did not know whether they had received specific advice, the existing guideline of a maximum of 2 SABA intakes per week was used. When the participants received specific advice from their general practitioner on their SABA use, they could indicate how much SABA they could use per week.

The question “How much SABA did you use last week?” was used as a baseline measure of SABA use. To examine whether an individual’s asthma was stable or unstable during the last week and differed from their average SABA use, an additional question was asked: “How much SABA do you use on average per week?”. In the app, individuals could register their weekly SABA use by clicking on the plus sign shown on the home screen.

The Intention to Change Behavior

The intention to change behavior was assessed using 3 items of the subscale “Intentions” of the Theoretical Domains Framework questionnaire [32]. The original subscale consisted of 4 items, but 1 of the items did not apply to this study and was, therefore, omitted. Items were answered on a 7-point Likert scale ranging from strongly disagree (1) to strongly agree (7). An example of an item is “In the next three months, I intend to use my SABA as prescribed.” A higher score (with a maximum of 21) signified more intent to use their SABA as prescribed in the next 3 months.

Feasibility and Usability

Different types of user data in the app were collected via an analytics platform (ie, PIWIK), namely, (1) which pages are visited in the app (ie, home screen, psychoeducation, user settings page, questionnaires, and the graph) and (2) events (ie, when the app is opened; SABA registrations; number of user clicks on notifications; and, when applicable, made changes in the maximum intake of SABA as advised by the health care professional).

The usability of the app was measured quantitatively using the 10-item SUS [35]. The items were rated on a 5-point Likert scale ranging from strongly disagree (0) to strongly agree (4). The scores were multiplied by 2.5 to obtain the total score ranging from 0 to 100. A higher score indicated that the app was more user-friendly.

A qualitative assessment of the feasibility and usability was conducted through interviews. The interview protocol was based on the UTAUT framework [37], which identified four main factors that influence the intention and use of technology (in this case, an app): (1) performance expectancy, (2) effort expectancy, (3) facilitating conditions, and (4) social influence. Textbox 1 presents an explanation of these factors. Moreover, the UTAUT framework includes four moderating factors: (1) gender, (2) age, (3) experience, and (4) voluntariness of use [37]. These factors and moderating factors were discussed during the interviews.

Textbox 1. Explanation of the factors within the Unified Theory of Acceptance and Use of Technology framework.

Description

- Performance expectancy: the general benefits associated with app use and feasibility of the app
- Effort expectancy: ease of use and usability of the app
- Facilitating conditions: having sufficient resources and knowledge to use the app
- Social influence: the influence of other people (eg, family, friends, and acquaintances) to start and keep using the app and whether they would recommend the app to others

Preliminary Effects

Asthma symptoms and triggers of these symptoms were measured using the 10-item CARAT [33,34]. An example item is "During the last week, because of your asthma/rhinitis/allergy, how many times, on average, did you experience sneezing?." Items were rated on a 4-point scale ranging from 0 (never) to 3 (almost every day). All items were reverse scored, and the total score ranged from 0 (minimal control) to 30 (maximum control). A score of 24 or higher indicated controlled asthma. In addition to the total score, 2 subscales were calculated: a score of the upper airway and a score of the lower airway. The upper airway score ranged from 0 (minimal control) to 12 (maximum control), and the lower airway score ranged from 0 (minimal control) to 18 (maximum control). An additional question was added to identify the symptom triggers: dust mites, animals, smoke, weather, hay fever or pollen, air pollution, smells, and exertion or exercise. Participants were able to select multiple triggers.

Participants' health and health-related QoL were measured using the SF-36 [31]. The SF-36 consists of 2 main categories: physical and mental health [38]. Physical health entailed the physical components and consisted of the following subscales: physical functioning (10 items), role limitations due to physical problems (4 items), bodily pain (2 items), and general health perceptions (5 items). Mental health entailed the mental components and consisted of the following subscales: social functioning (2 items), general mental health (5 items), role limitations due to emotional problems (3 items), and vitality (4 items) [31,39]. All items were recoded into scores ranging from 0 (the poorest level of physical or mental health) to 100 (the best level of physical or mental health) [40], with higher scores indicating better health and higher QoL.

Statistical Analysis

All quantitative data were analyzed using SPSS (version 25.0; IBM Corp) [41]. Descriptive analyses (eg, means, SDs, and percentages) were used to describe the demographic and clinical characteristics of the participants, intention to change behavior, user experience, QoL, and user data (eg, frequency of weekly SABA use). A mixed model was used to determine the change in asthma symptoms over time from the first week of using the app to 3 months after baseline. QoL at 3 months was compared with baseline data using Wilcoxon signed rank tests. The effects at 6 and 12 months were not examined because of the high dropout rate during the study period (88.2% at 3 months and more dropouts beyond that).

Interviews were audiotaped for subsequent analyses, and all audio records were transcribed intelligent verbatim by 1 researcher (AEV). Qualitative data analyses were performed by 2 researchers (LNvdB and AEV) according to the principles of the Framework Method [42] using Atlas.ti (version 22.0) [43]. The Framework Method is a systematic and flexible approach often used for the thematic analysis of semistructured interview data. Following transcription, the 2 researchers immersed themselves in the interviews to gain a comprehensive understanding. Subsequently, a deductive approach was adopted to code the interviews based on a predefined concept codebook developed beforehand based on the UTAUT framework [37]. The coding process was conducted independently by the 2 researchers, followed by a comparison of the codes. Additional codes were incorporated into the codebook, where applicable. A framework matrix was used to organize the data comprehensively, featuring relevant quotes from the participants. Finally, the characteristics and distinctions within the data set were identified. Throughout the process, the steps and data were discussed with the researchers CH and AV.

Results

Demographic and Clinical Characteristics

In the quantitative study, 485 individuals participated at baseline. Of these 485 individuals, 373 (76.9%) reported that they used SABA. Only these individuals were included in the analysis. Most of the participants were female (309/373, 82.8%) with a mean age of 46 (SD 15) years, had a secondary vocational education or higher (316/373, 84.7%), and had allergic asthma (187/373, 50.1%). At baseline, participants stated that they used, on average, 10 SABA per week and 10 SABA in the week before using the app. Moreover, the mean intention to change behavior was 17.1. This indicates that the participants wanted to use their SABA as prescribed for the next 3 months. Table 1 shows an overview of the demographic and clinical characteristics and the intention to change behavior.

Table 1. Baseline demographic and clinical characteristics of the participants and the intention to change behavior in the quantitative study.

Characteristic	Values
Gender (n=373), n (%)	
Male	63 (16.9)
Female	309 (82.8)
Rather not say	1 (0.3)
Age[a] (y; n=371), mean (SD; range)	46.1 (15; 18-81)
Educational level (n=373), n (%)	
Primary school	7 (1.9)
Secondary education	50 (13.4)
Secondary vocational education	136 (36.5)
Higher professional education	121 (32.4)
University education	59 (15.8)
Type of asthma[b] (n=373), n (%)	
Allergic asthma	187 (50.1)
Nonallergic asthma	126 (33.8)
Exercise asthma	164 (44)
Severe asthma	104 (27.9)
Do not know	31 (8.3)
Self-reported asthma control (n=373), n (%)	
Good control	134 (35.9)
Insufficient control	151 (40.5)
Do not know	86 (23.1)
Medication type used[b] (n=373), n (%)	
SABA[c,d]	373 (100)
ICS[e]	198 (53.1)
LABA[f]	150 (40.2)
ICS+LABA	127 (34)
Do not know	0 (0)
No medication use	0 (0)
Average SABA use in the last week: self-reported (n=373), mean (SD; range)	10.5 (12.6; 0-60)
Average SABA use per week: self-reported (n=373), mean (SD; range)	9.7 (11.6; 0-60)
Had medication advice from the health care professional (n=373), n (%)	
Yes	246 (66)
No	110 (29.5)
Do not know	17 (4.6)
Average maximum prescribed SABA[g] (n=246), mean (SD; range)	22.2 (16.9; 0-60)
Intention to change behavior (n=373), mean (SD; range)	17.1 (4.5, 3-21)

[a]The birth year of 2 participants was missing. These participants were excluded from the calculation of the mean age.

[b]Participants were able to select multiple answers.

[c]SABA: short-acting beta2-agonist.

[d]51 participants only used SABA and no other inhalers.

[e]ICS: inhaled corticosteroids.

[f]LABA: long-acting beta2-agonist.

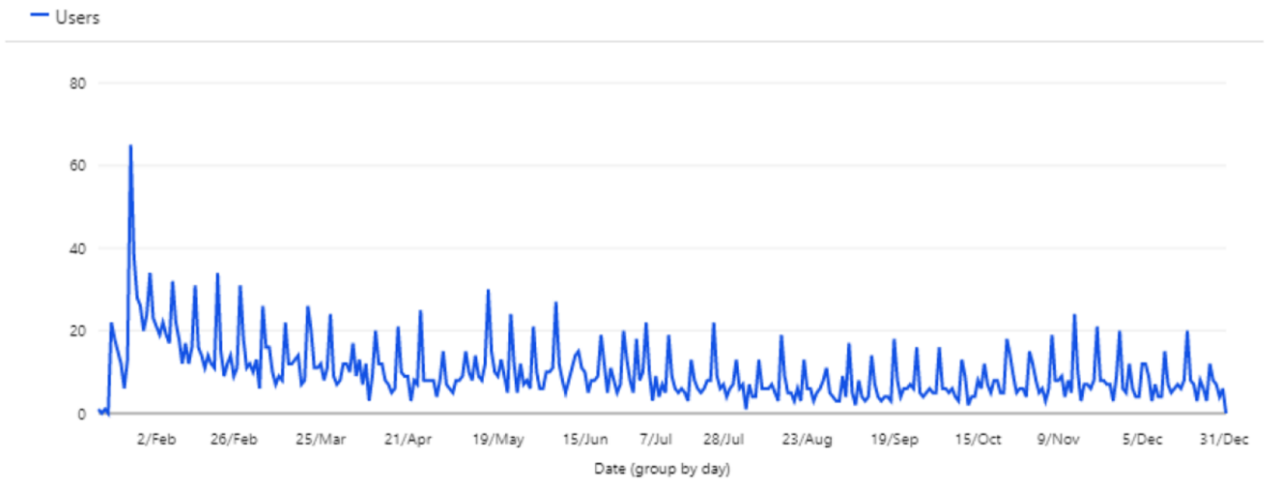
[g]The maximum number of SABA inhalations per week, as prescribed by the participant's health care professional.

In the qualitative part of the study, among the 6 to 12 participants that we planned to recruit, only 4 participants could be included and interviewed. Half of the interviewed participants were female (2/4, 50%), with a mean age of 55 (range 21-78) years. One participant completed senior general secondary education, 1 completed secondary vocational education, and 2 had higher professional education. Two participants stated that they still used the app: they had both been using it for 1 year and 5 months. Regarding *social influence* from the UTAUT framework [37], the app was recommended by the hospital to one participant, and the other participant found it via the asthma association. Two participants stated that they no longer used the app but had used it for approximately 1 or 2 weeks. They both started using the app after the recommendation from a family member.

Feasibility and Usability

User data showed that 335 unique users opened the app, of which 250 (74.6%) were returning visitors, with an average session time of 1 minute. An overview of the number of users during the study period is shown in Figure 2.

Figure 2. Number of users during the study period (January 15, 2021, to December 31, 2021).



Most users opened the app via their smartphone (303/335, 90.4%), followed by a tablet (27/335, 8.1%) and a phablet (4/335, 1.2%). On average, the users had 5 events (ie, starting the app, adding SABA, removing SABA, changing the maximum amount of SABA, and clicking on 1 of the notifications) per session. Registration of SABA (ie, add-function) was most often used (7506/13,081 times, 57.38%). An overview of the events used per week is shown in Figures 3 and 4. At 3, 6, and 12 months, users registered an average of 5 SABA intakes per week (Table 2).

Figure 3. Unique events used per week: "add" means registering short-acting beta2-agonist (SABA), "remove" means removing a SABA registration, "set-max" means changing the maximum amount of SABA, and "start" means starting the app after giving informed consent and filling in the first questionnaires.

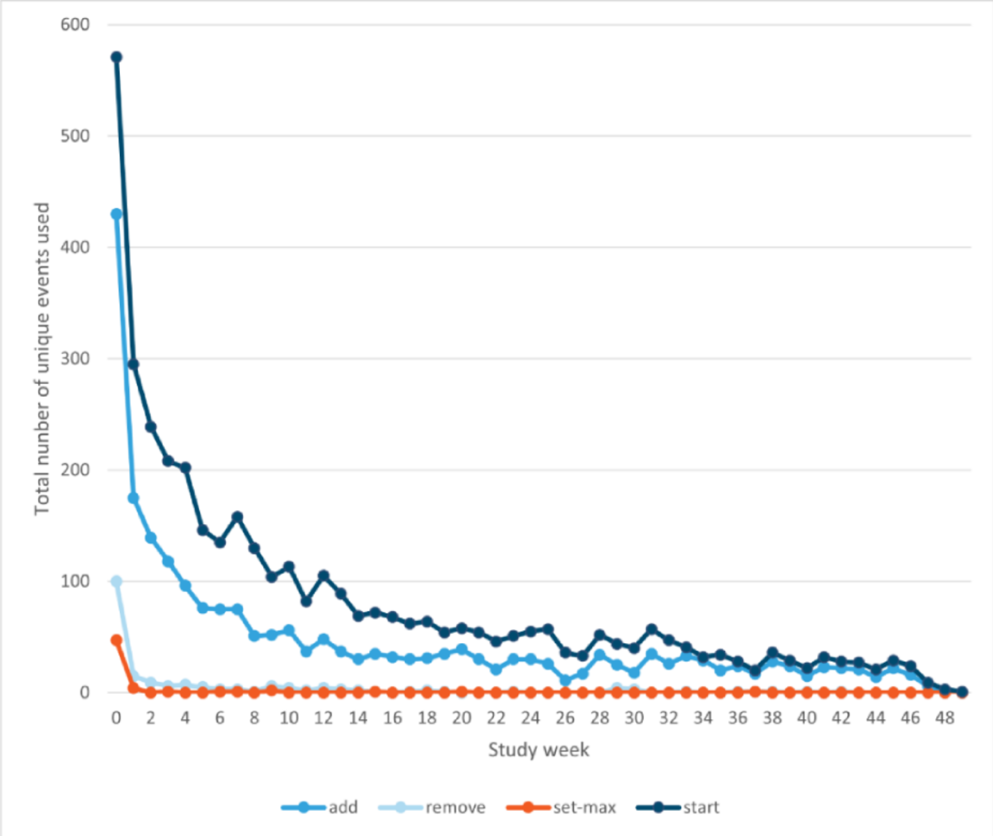


Figure 4. Events used per week, whether participants opened the app via 1 of the notifications: "open-intake-registration-reminder" is the notification users received when they did not register any short-acting beta2-agonist before the end of the week (Sunday); "open-review-questionnaire-reminder" is the notification for the questionnaires used at 3, 6 and 12 months; and "open-weekly questionnaire-reminder" is the notification for the weekly Control of Allergic Rhinitis and Asthma Test.

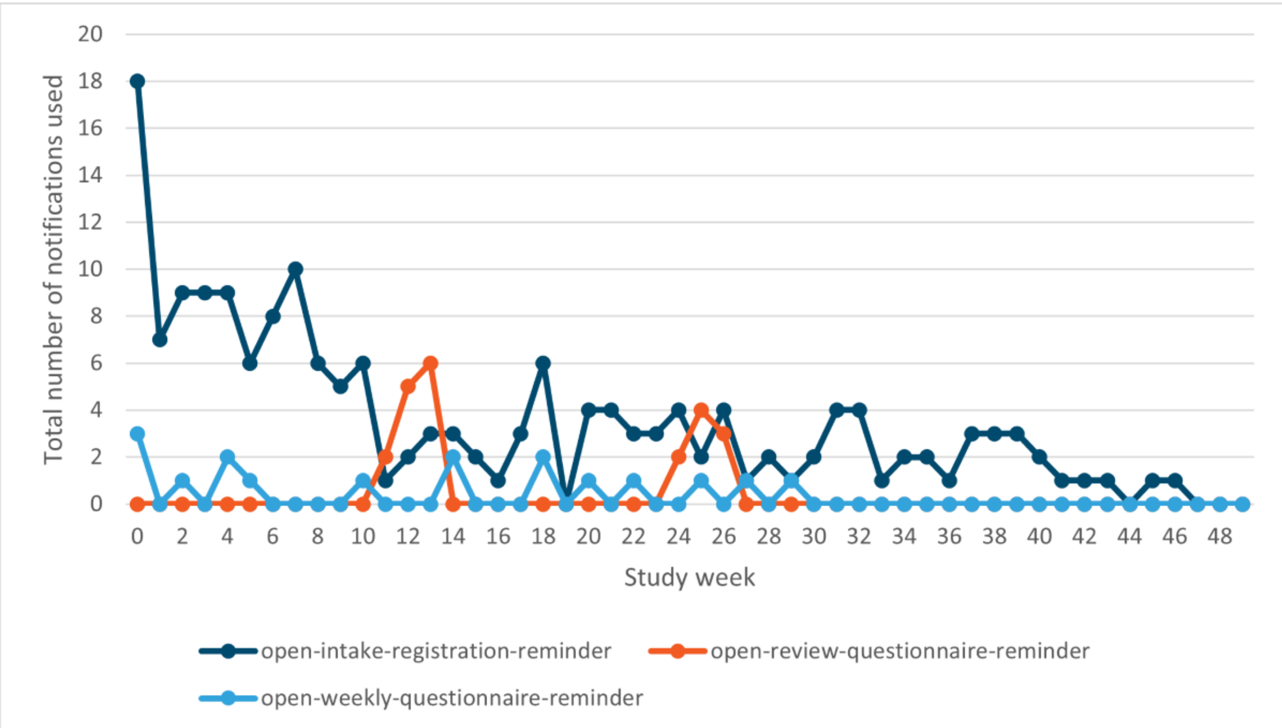


Table 2. Registered short-acting beta2-agonist use per week.

Time point	Values, mean (SD; 95% CI)
Baseline (n=373)	
Baseline questionnaire	10.5 (0.5; N/A[a])
3 months (n=19)	
12 weeks after baseline	4.67 (2.5; 1.64-13.29)
6 months (n=11)	
25 weeks after baseline	4.82 (2.6; 1.68-13.81)
Latest time point[b] (n=2)	
48 weeks after baseline	5.24 (3; 1.71-16.09)

[a]N/A: not applicable.

[b]Measures ended within 1 year (mid-January 2021 until the end of December 2021).

Usability of the app, as assessed with the SUS, was good over the entire study period: 82.3 (SD 13.2; n=44) at 3 months, 84 (SD 13.6; n=26) at 6 months, and 82.3 (SD 13.4; n=11) at 12 months (Table 3).

Table 3. Questionnaire results regarding usability and quality of life.

Questionnaire	Baseline	3 months	6 months	12 months
Usability				
Values, n (%)	N/A[a]	44 (100)	26 (100)	11 (100)
Values, mean (SD)	N/A	82.3 (13.2)	84 (13.6)	82.3 (13.4)
Quality of life: physical health				
Values, n (%)	373 (100)	44 (100)	26 (100)	11 (100)
Values, mean (SD)	53.6 (22.4)	56.1 (23.9)	57.6 (21.8)	54.9 (24.7)
Quality of life: mental health				
Values, n (%)	373 (100)	44 (100)	26 (100)	11 (100)
Values, mean (SD)	57.4 (21.2)	62.9 (22.3)	59.4 (20.1)	67.8 (18.1)

[a]N/A: not applicable.

Qualitative data showed that 3 (75%) of the 4 participants had experience using other health apps. The users mentioned that they wanted to use health apps that were fun and useful:

> *[...] I only want apps which I like or which are useful.* [Male user, 78 years old]

The participants found the app acceptable and clear in terms of *performance expectancy*. Three participants stated that gaining insight into asthma and its triggers was helpful. Another participant explained that it was not helpful at the moment because he considered his asthma to be controlled. This was for both participants who no longer used the app. One participant did not use SABA often, and the other participant only used it regularly based on the advice of his pulmonologist:

> *First impression was, well, I think, it looks clear. It was pretty clear to me on my own what I could do with it. After using it, yeah, I think it just looks like a nice app, not too old-fashioned. But just fairly new, as you expect from an app in this day and age. And it was also very quickly clear to me exactly what I could do with it.* [Male user, 21 years old]

> *As very useful; you open the app and click on the plus icon how many times if you use it at that time. And also very nice that you get a notification every now and then like, "hey, it is the end of the week; make sure you fill in the amount." Especially if you forget to fill it in. That is nice.* [Male user, 21 years old]

Regarding *effort expectancy* and *facilitating conditions*, 3 participants stated that the app is easy to use and straightforward and does not require much effort to register SABA use. One of these participants also stated that the app was well written and easy to read. The fourth participant did not say anything about ease of use. However, 1 participant experienced difficulties in interpreting the questions and answering the possibilities of the CARAT:

> *So with a few questions, I got, well you already noticed that I have some difficulties with choosing the right one.* [Male user, 78 years old]

Of the participants who continue to use the app (2/4, 50%), they use it multiple times per week, with a minimum frequency of once per week and often 2 or 3 times per week. Opening the app was, for 1 participant, mostly completed after receiving a notification. A former user mentioned that he would use the app once a week to fill in all the SABA intakes for that week.

Multiple possibilities for improvement were mentioned during the interviews. One participant wanted to be able to fill in triggers that were not listed in the app and also wanted to have the possibility to add more types of medication. Another participant missed contact with other patients with asthma in the app to discuss, for example, medication use. Someone else would change the CARAT based on their experienced difficulties. The last participant missed more background information about SABA use and why SABA should not be used more than twice a week:

> *There is one question that I do not understand. I filled it in good conscience in, and it immediately gave a number that should be decisive, but that I think "yes, but this does not apply to me." In the app, it asked "how many times a night do you wake up?" [...] I do wake up but with a different cause [...] I personally think, but that is my opinion, there should stand "Do you wake up at night, because of your asthma?* [Male user, 78 years old]

Further exploring *social influence* showed that all the participants would recommend the app to others because they experienced that it provided more insight into their medication use and they received more information about the complete picture of asthma. One participant had already recommended the app to an acquaintance, who also started using the app. The 2 former users would specifically recommend it to certain patients: people with severe asthma or uncontrolled asthma or people who do not take their reliever medication as intended. Furthermore, 3 participants also found it useful to show the app to their health care professional during a consultation:

> *I would recommend it, especially to people who do not really have a case like mine. I would also not recommend it to people who, like me, only use salbutamol for sports. Yes, I do not know those people who just do it for sports just like me. Then it does not make much sense to keep track of how often you use it. You just know how often you exercise, and if you first use salbutamol then you know "hey, I use it so often." But for people who use it often, it seems to me that is a very handy app, especially if you can see in that graph how often you have used it per week and in which week more and in which week less.* [Male user, 21 years old]

Preliminary Effects

At week 1, the mean CARAT score was 14.8. This indicated that the participants' asthma was uncontrolled. Their CARAT score improved significantly to a mean score of 18.5 after 12 weeks (ie, 3 months; β=.189; SE 0.048; P<.001); however, this mean score still indicated that their asthma was uncontrolled. This was also the case for both the upper airway score, which significantly improved from a mean score of 6.8 to 7.7 after 12 weeks (β=.073; SE 0.027; *P*=.009), and the lower airway score, which significantly improved from a mean score of 8 to 10.8 after 12 weeks (β=.121; SE 0.037; *P*=.002).

The top three asthma triggers reported in week 1 were (1) weather (321/435, 73.8%), (2) exertion or exercise (305/435, 70.1%), and (3) smoke (197/435, 45.3%). After 12 weeks (ie, 3 months), the top three triggers were (1) weather (25/37, 68%), (2) exertion or exercise (19/37, 51%), and (3) hay fever or pollen (17/37, 46%).

At 3 months, there was no significant difference compared with baseline regarding the mean physical and mental health scores (Z=−0.074; *P*=.94 and Z=−0.117; *P*=.91, respectively; Table 3).

Discussion

Principal Findings

This study aimed to determine the feasibility and usability of a newly developed app, the Asthma app. Furthermore, the preliminary effects of using the app after 3 months on decreasing asthma symptoms and improving QoL were examined. The quantitative data showed that the usability was good. This was also found in the qualitative data: the app was considered easy to use, and it did not take much effort to register SABA. Furthermore, most participants stated that the app was useful for gaining insight into asthma, triggers, and medication use, and therefore, the app was considered feasible and usable. Multiple improvement possibilities were mentioned during the interviews, such as adding additional personal triggers next to the existing standard list of triggers and the availability of a social support network to contact others with asthma easily. In addition, former users, who no longer used the app, stated that they would recommend the app to people with severe asthma or uncontrolled asthma or people who do not take their reliever medication as intended.

As for the preliminary effects, an improvement in asthma symptoms was found after 3 months; however, the mean asthma symptom score still indicated that the asthma was uncontrolled. Improvement in asthma symptoms was also found in other eHealth studies [19,20]. The mean asthma symptom score in our study, indicating uncontrolled asthma, could be explained by the low intensity and noninvasive nature of the intervention (eg, users could use the app whenever and how often they wanted). A systematic review [44] also found that asthma control did not significantly improve in other studies. They proposed additional well-designed studies to gather more robust findings on what is necessary to achieve optimal asthma control [44]. In terms of QoL, no significant improvement was observed after 3 months. No effect was observed because poor asthma control was associated with worsened QoL [45,46]. The average uncontrolled asthma scores at week 1 and 3 months after baseline can be related to the low QoL scores at the same time points. Moreover, a systematic review [47] demonstrated that eHealth interventions have an inconsistent impact on QoL in people with asthma. The systematic enhancement of clinical outcomes such as QoL was mostly observed within the whole-systems approach, taking into account patient, professional, and organizational elements.

The data from this study should be interpreted with caution because of the high dropout rate, which resulted in insufficient data for conducting analyses at 6 and 12 months. Although a high dropout rate is frequently seen in studies investigating digital applications, we envisioned that the dropout rate would be lower in this study, considering the participatory design process [28]. The dropout may be explained by the higher probability of dropout in people with chronic diseases when they are impacted physically and mentally by the condition [48]. Most of the participants in this study had uncontrolled asthma and, therefore, more symptoms throughout the day and night. This could have resulted in lower or no app use, which was directly linked to the withdrawal from the study. Another explanation for the high dropout rate could be, as described in our previous study [28], that only a minimal viable product was evaluated. Not all features recommended by the patients, such as registering additional controller medication, were implemented. Therefore, the app might not fit the needs of all the users and cause them to stop using the app.

Using the UTAUT framework [37], performance expectancy was positively associated with the use of the app for the current users. The app will help them gain more insight into asthma, triggers, and medication use. Performance expectancy was lower for former users who stated that their asthma was controlled; therefore, the aim of the app did not align with their needs. Effort expectancy was positively associated with both the intention to use and the actual use of the app, largely because of its user-friendly interface, minimal effort required for SABA registration, and language simplicity. The only aspect that was negatively related to the effort expectancy factor was difficulty with one of the questionnaires by a former user. Facilitating conditions were positively associated with the use of the app. The participants had the appropriate knowledge and resources to use the app. Technical support was not discussed during the interviews; however, clarity regarding the appropriate contact for technical issues could enhance user experience. Finally, social influence played an essential role in intention and use; all interviewees initiated app use through social media discovery or recommendations from health care professionals or family members. They would also recommend the app to others, and 1 participant had already recommended the app to an acquaintance. However, in future studies, this could be further explored in relation to voluntariness of use, which was not thoroughly explored in this study. This is also the case for other moderating factors such as *gender* and *age*. The sample size was too small to explore the associations between the moderating factors, factors, and intention and use of the app. Notably, prior experience with health apps positively influenced the intention and use of the app in this study, and current experience was positively influenced by effort expectancy, facilitating conditions, and social influence for current users.

Strengths and Limitations

This study has several strengths and limitations. A notable strength was the use of a real-life setting for evaluating the Asthma app, allowing a comprehensive understanding of its feasibility and usability. In addition, interviews with both current and former users provided a nuanced perspective on user satisfaction and the factors influencing app use.

In addition to the previously mentioned high dropout rate, another limitation was that the questionnaires were exclusively offered in the app environment. Therefore, former users were no longer able to complete the study questionnaires, thus limiting the availability of their data at later time points (ie, after baseline). To obtain the perspectives of former users on feasibility and usability, they were included in the qualitative interviews. Nevertheless, the recruitment of this group was difficult, and only 2 former users could be included.

Finally, the intended target of 6 to 12 interviews to achieve data saturation [29,30] was not attained. This was partially attributed to the difficulty in reaching former users who may have lost interest in the app or study. Despite the small number of interviews conducted, similar findings were found during data collection between the 2 users and the 2 former users.

Implications for Future Research and Practice

A minimal viable product was examined in this study. During the next development round, feedback gathered during the cocreation of the app could be re-evaluated [28], or new cocreation sessions could be organized to further enhance the app. In future studies, with a newer version of the app, the outcomes of this study could be further examined with more data at more time points, and clinical outcomes, such as the impact of the app on medication adherence, could be explored. A smart asthma inhaler [49,50] could also be linked to the app to gather real-time objective data instead of self-reported registration, which is more sensitive to biases.

This study has a high dropout rate. Renzi et al [51] stated in their review that reminders are often used to improve medication adherence in eHealth interventions but that this improvement is reduced over time. Typically, after 6 months, users tend to revert to their previous behaviors as the novelty of the eHealth intervention wanes [51]. This could also be the case in this study, especially because of the anonymous nature and the use of in-app questionnaires. In future studies, it would be advisable to collect data pseudonymously and send questionnaires via email to achieve a higher response rate. In this way, participants will also be less likely to withdraw from the study and stay involved for longer.

In the new version of the app, additional information about the treatment guidelines should be implemented, such as the fact that users should follow the advice from their health care professional if they receive any. It should be clarified that the app is specifically for people with asthma who only use SABA (and not ICS), which has been the first step of treatment for decades. Potential users could be reached via general practitioners, specialized practice nurses, or pharmacists when they prescribe or distribute SABA. Currently, the Asthma app is a stand-alone app, which means that it is used by patients without the involvement of health care professionals. However, involving health care professionals via “blended care” could improve the quality of care [52]. Moreover, health care professionals can offer additional education and guidance based on the data from the app [53]. To incorporate the app into standard treatment, it is necessary to develop a plan together with asthma associations and health care professionals. A designated implementation team can improve the success rate of the implementation [54], and it is important to explore context-specific strategies that align with the implementation process phase [55]. Certain barriers (eg, technical issues, time and attention requirements for use, low engagement from health care professionals, and shortage of funding) and facilitators (eg, stakeholder engagement and enthusiasm, minimizing workflow interruptions, and access to information about the app) should be taken into account when implementing the app in standard care [27,56-58]. In addition, more education about SABA overuse could make health care professionals more aware of the risks, which could prioritize the use of the app.

Conclusions

This study evaluated the feasibility and usability of a new app for people with asthma. The initial results regarding usability were positive. Nevertheless, it is essential to exercise caution when interpreting these results because of the high dropout rate in this study. Two former users would recommend the app to people with severe asthma or uncontrolled asthma or people who do not use their reliever medication as intended. Future (implementation) studies could evaluate the potential of incorporating the app into standard treatment practices. Moreover, the actual impact of the app on clinical outcomes, such as medication adherence, should be further examined.

Acknowledgments

This study was funded by Astra Zeneca (no grant number). The authors would like to thank Andrea Sellink and Eva van den Berg who contributed to the review of the manuscript.

Authors' Contributions

LNvdB, CH, NHC, and AV contributed to the conceptualization and reviewed and edited the manuscript. LNvdB and AEV prepared the original draft and conducted the quantitative and qualitative analyses. LNvdB provided the visualization (preparation, creation, and presentation of the published work, specifically visualization or data presentation). LNvdB, CH, and AV contributed to the methodology, validation, investigation, resources, and data curation. CH, NHC, and AV supervised the project and acquired funding for the project. AV was responsible for the project administration. All authors have read and agreed to the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Semistructured interview protocol about the Asthma app.
[DOCX File , 15 KB - humanfactors_v11i1e54386_app1.docx]

References

1. The Global Asthma Report 2022. Int J Tuberc Lung Dis 2022 Nov 25;26(Supp 1):1-104. [doi: 10.5588/ijtld.22.1010] [Medline: 36303302]
2. Global strategy for asthma management and prevention. Global Initiative for Asthma. URL: https://ginasthma.org/reports/ [accessed 2020-09-08]
3. Bottema JW, Bouma M, Broekhuizen L, Chavannes NH, Frankemölle LA, Hallensleben C, et al. Astma bij volwassenen (derde herziening). Nederlands Huisartsen Genootschap. 2020 Jul. URL: https://www.nhg.org/standaarden/volledig/nhg-standaard-astma-bij-volwassenen [accessed 2020-08-12]
4. Belhassen M, Nibber A, Van Ganse E, Ryan D, Langlois C, Appiagyei F, et al. Inappropriate asthma therapy-a tale of two countries: a parallel population-based cohort study. NPJ Prim Care Respir Med 2016 Oct 13;26(1):16076 [FREE Full text] [doi: 10.1038/npjpcrm.2016.76] [Medline: 27735927]
5. Demoly P, Annunziata K, Gubba E, Adamek L. Repeated cross-sectional survey of patient-reported asthma control in Europe in the past 5 years. Eur Respir Rev 2012 Mar 01;21(123):66-74 [FREE Full text] [doi: 10.1183/09059180.00008111] [Medline: 22379176]
6. Kosse RC, Murray E, Bouvy ML, de Vries TW, Stevenson F, Koster ES. Potential normalization of an asthma mHealth intervention in community pharmacies: applying a theory-based framework. Res Social Adm Pharm 2020 Feb;16(2):195-201 [FREE Full text] [doi: 10.1016/j.sapharm.2019.05.004] [Medline: 31109820]
7. Kritikos V, Price D, Papi A, Infantino A, Ställberg B, Ryan D, et al. A multinational observational study identifying primary care patients at risk of overestimation of asthma control. NPJ Prim Care Respir Med 2019 Dec 05;29(1):43 [FREE Full text] [doi: 10.1038/s41533-019-0156-4] [Medline: 31804501]
8. Stanford RH, Shah MB, D'Souza AO, Dhamane AD, Schatz M. Short-acting beta-agonist use and its ability to predict future asthma-related outcomes. Ann Allergy Asthma Immunol 2012 Dec;109(6):403-407. [doi: 10.1016/j.anai.2012.08.014] [Medline: 23176877]

9. Nwaru BI, Ekström M, Hasvold P, Wiklund F, Telg G, Janson C. Overuse of short-acting β2-agonists in asthma is associated with increased risk of exacerbation and mortality: a nationwide cohort study of the global SABINA programme. Eur Respir J 2020 Apr 16;55(4):1901872 [FREE Full text] [doi: 10.1183/13993003.01872-2019] [Medline: 31949111]
10. Rennard SI, Farmer SG. Exacerbations and progression of disease in asthma and chronic obstructive pulmonary disease. Proc Am Thorac Soc 2004;1(2):88-92. [doi: 10.1513/pats.2306026] [Medline: 16113418]
11. FitzGerald JM, Tavakoli H, Lynd LD, Al Efraij K, Sadatsafavi M. The impact of inappropriate use of short acting beta agonists in asthma. Respir Med 2017 Oct;131:135-140 [FREE Full text] [doi: 10.1016/j.rmed.2017.08.014] [Medline: 28947020]
12. Anis AH, Lynd LD, Wang XH, King G, Spinelli JJ, Fitzgerald M, et al. Double trouble: impact of inappropriate use of asthma medication on the use of health care resources. CMAJ 2001 Mar 06;164(5):625-631 [FREE Full text] [Medline: 11258208]
13. O'Byrne PM, Jenkins C, Bateman ED. The paradoxes of asthma management: time for a new approach? Eur Respir J 2017 Sep 09;50(3):1701103 [FREE Full text] [doi: 10.1183/13993003.01103-2017] [Medline: 28889114]
14. Partridge MR, van der Molen T, Myrseth SE, Busse WW. Attitudes and actions of asthma patients on regular maintenance therapy: the INSPIRE study. BMC Pulm Med 2006 Jun 13;6(1):13 [FREE Full text] [doi: 10.1186/1471-2466-6-13] [Medline: 16772035]
15. Janson C, Menzies-Gow A, Nan C, Nuevo J, Papi A, Quint JK, et al. SABINA: an overview of short-acting β2-agonist use in asthma in European countries. Adv Ther 2020 Mar 24;37(3):1124-1135 [FREE Full text] [doi: 10.1007/s12325-020-01233-0] [Medline: 31981105]
16. Amin S, Soliman M, McIvor A, Cave A, Cabrera C. Understanding patient perspectives on medication adherence in asthma: a targeted review of qualitative studies. Patient Prefer Adherence 2020 Mar;14:541-551 [FREE Full text] [doi: 10.2147/PPA.S234651] [Medline: 32210541]
17. Price D, Fletcher M, van der Molen T. Asthma control and management in 8,000 European patients: the REcognise Asthma and LInk to Symptoms and Experience (REALISE) survey. NPJ Prim Care Respir Med 2014 Jun 12;24(1):14009 [FREE Full text] [doi: 10.1038/npjpcrm.2014.9] [Medline: 24921985]
18. Baron AJ, Flokstra-de Blok BM, Kerstjens HA, Koopmans-Klein G, Price DB, Sellink AA, et al. High use of SABAs is associated with higher exacerbation rate in Dutch patients with asthma. J Asthma Allergy 2021 Jul;14:851-861 [FREE Full text] [doi: 10.2147/JAA.S292943] [Medline: 34285512]
19. Barrett MA, Humblet O, Marcus JE, Henderson K, Smith T, Eid N, et al. Effect of a mobile health, sensor-driven asthma management platform on asthma control. Ann Allergy Asthma Immunol 2017 Nov;119(5):415-21.e1. [doi: 10.1016/j.anai.2017.08.002] [Medline: 29150069]
20. Merchant RK, Inamdar R, Quade RC. Effectiveness of population health management using the propeller health asthma platform: a randomized clinical trial. J Allergy Clin Immunol Pract 2016;4(3):455-463. [doi: 10.1016/j.jaip.2015.11.022] [Medline: 26778246]
21. Gibson PG, Powell H, Coughlan J, Wilson AJ, Abramson M, Haywood P, et al. Self-management education and regular practitioner review for adults with asthma. Cochrane Database Syst Rev 2003(1):CD001117. [doi: 10.1002/14651858.CD001117] [Medline: 12535399]
22. Pinnock H, Parke HL, Panagioti M, Daines L, Pearce G, Epiphaniou E, et al. Systematic meta-review of supported self-management for asthma: a healthcare perspective. BMC Med 2017 Mar 17;15(1):64 [FREE Full text] [doi: 10.1186/s12916-017-0823-7] [Medline: 28302126]
23. Morrison D, Wyke S, Agur K, Cameron EJ, Docking RI, Mackenzie AM, et al. Digital asthma self-management interventions: a systematic review. J Med Internet Res 2014 Feb 18;16(2):e51 [FREE Full text] [doi: 10.2196/jmir.2814] [Medline: 24550161]
24. Pinnock H. Supported self-management for asthma. Breathe (Sheff) 2015 Jun 03;11(2):98-109 [FREE Full text] [doi: 10.1183/20734735.015614] [Medline: 26306110]
25. James DR, Lyttle MD. British guideline on the management of asthma: SIGN Clinical Guideline 141, 2014. Arch Dis Child Educ Pract Ed 2016 Dec 20;101(6):319-322. [doi: 10.1136/archdischild-2015-310145] [Medline: 27206454]
26. Jabangwe R, Edison H, Duc AN. Software engineering process models for mobile app development: a systematic literature review. J Syst Softw 2018 Nov;145:98-111. [doi: 10.1016/j.jss.2018.08.028]
27. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 05;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
28. van den Berg LN, Hallensleben C, Chavannes NH, Versluis A. Developing a smartphone application that promotes responsible short-acting beta2-agonist use in people with asthma: a participatory design. Int J Environ Res Public Health 2022 Jul 12;19(14):8496 [FREE Full text] [doi: 10.3390/ijerph19148496] [Medline: 35886348]
29. Guest G, Bunce A, Johnson L. How many interviews are enough? An experiment with data saturation and variability. Field Methods 2016 Jul 21;18(1):59-82. [doi: 10.1177/1525822X05279903]
30. Guest G, Namey E, Chen M. A simple method to assess and report thematic saturation in qualitative research. PLoS One 2020 May 5;15(5):e0232076 [FREE Full text] [doi: 10.1371/journal.pone.0232076] [Medline: 32369511]

31. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992 Jun;30(6):473-483. [Medline: 1593914]
32. Huijg JM, Gebhardt WA, Crone MR, Dusseldorp E, Presseau J. Discriminant content validity of a theoretical domains framework questionnaire for use in implementation research. Implement Sci 2014 Jan 15;9(1):11 [FREE Full text] [doi: 10.1186/1748-5908-9-11] [Medline: 24423394]
33. Azevedo P, Correia de Sousa J, Bousquet J, Bugalho-Almeida A, Del Giacco SR, Demoly P, et al. Control of Allergic Rhinitis and Asthma Test (CARAT): dissemination and applications in primary care. Prim Care Respir J 2013 Mar 14;22(1):112-116 [FREE Full text] [doi: 10.4104/pcrj.2013.00012] [Medline: 23412110]
34. The control of allergic rhinitis and asthma test. The Control of Allergic Rhinitis and Asthma Test Network. URL: https://www.new.caratnetwork.org/fastcarat/index.html [accessed 2021-05-24]
35. Brooke J. SUS: a 'quick and dirty' usability scale. In: Usability Evaluation In Industry. Boca Raton, FL: CRC Press; 1996.
36. Castor EDC. URL: https://www.castoredc.com/ [accessed 2020-02-17]
37. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003 Sep;27(3):425-478. [doi: 10.2307/30036540]
38. McHorney CA, Ware JE, Raczek AE. The MOS 36-item short-form health survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care 1993 Mar;31(3):247-263. [doi: 10.1097/00005650-199303000-00006] [Medline: 8450681]
39. Ware JE, Kosinski M, Keller SD. SF-36 Physical and Mental Health Summary Scales: A User's Manual. Boston, MA: Health Assessment Lab; 1994.
40. Grassi M, Nucera A, European Community Respiratory Health Study Quality of Life Working Group. Dimensionality and summary measures of the SF-36 v1.6: comparison of scale- and item-based approach across ECRHS II adults population. Value Health 2010;13(4):469-478 [FREE Full text] [doi: 10.1111/j.1524-4733.2009.00684.x] [Medline: 20088893]
41. IBM SPSS Statistics 25.0. IBM. URL: https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-25 [accessed 2020-03-24]
42. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13:117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
43. ATLAS.ti homepage. ATLAS.ti Scientific Software Development GmbH. URL: https://atlasti.com [accessed 2022-01-13]
44. van Buul AR, Kasteleyn MJ, Arends JM, Shi T, Kelly DP, Chavannes NH, et al. eHealth only interventions and blended interventions to support self-management in adolescents with asthma: a systematic review. Clinical eHealth 2020;3:49-62. [doi: 10.1016/j.ceh.2020.06.001]
45. Gonzalez-Barcala FJ, de la Fuente-Cid R, Tafalla M, Nuevo J, Caamaño-Isorna F. Factors associated with health-related quality of life in adults with asthma. A cross-sectional study. Multidiscip Respir Med 2012 Oct 02;7(1):32 [FREE Full text] [doi: 10.1186/2049-6958-7-32] [Medline: 23031194]
46. Chen H, Gould MK, Blanc PD, Miller DP, Kamath TV, Lee JH, et al. Asthma control, severity, and quality of life: quantifying the effect of uncontrolled disease. J Allergy Clin Immunol 2007 Aug;120(2):396-402. [doi: 10.1016/j.jaci.2007.04.040] [Medline: 17561244]
47. Hui CY, Walton R, McKinstry B, Jackson T, Parker R, Pinnock H. The use of mobile applications to support self-management for people with asthma: a systematic review of controlled studies to identify features associated with clinical effectiveness and adherence. J Am Med Inform Assoc 2017 May 01;24(3):619-632 [FREE Full text] [doi: 10.1093/jamia/ocw143] [Medline: 27694279]
48. Beller J, Geyer S, Epping J. Health and study dropout: health aspects differentially predict attrition. BMC Med Res Methodol 2022 Jan 30;22(1):31 [FREE Full text] [doi: 10.1186/s12874-022-01508-w] [Medline: 35094681]
49. Sulaiman I, Greene G, MacHale E, Seheult J, Mokoka M, D'Arcy S, et al. A randomised clinical trial of feedback on inhaler adherence and technique in patients with severe uncontrolled asthma. Eur Respir J 2018 Jan 04;51(1):1701126 [FREE Full text] [doi: 10.1183/13993003.01126-2017] [Medline: 29301919]
50. van de Hei SJ, Poot CC, van den Berg LN, Meijer E, van Boven JF, Flokstra-de Blok BM, et al. Effectiveness, usability and acceptability of a smart inhaler programme in patients with asthma: protocol of the multicentre, pragmatic, open-label, cluster randomised controlled ACCEPTANCE trial. BMJ Open Respir Res 2022 Dec;9(1):e001400 [FREE Full text] [doi: 10.1136/bmjresp-2022-001400] [Medline: 36522130]
51. Renzi E, Baccolini V, Migliara G, De Vito C, Gasperini G, Cianciulli A, et al. The impact of eHealth interventions on the improvement of self-care in chronic patients: an overview of systematic reviews. Life (Basel) 2022 Aug 17;12(8):1253 [FREE Full text] [doi: 10.3390/life12081253] [Medline: 36013432]
52. van der Kleij RM, Kasteleyn MJ, Meijer E, Bonten TN, Houwink EJ, Teichert M, et al. SERIES: eHealth in primary care. Part 1: concepts, conditions and challenges. Eur J Gen Pract 2019 Oct;25(4):179-189 [FREE Full text] [doi: 10.1080/13814788.2019.1658190] [Medline: 31597502]
53. Song X, Hallensleben C, Zhang W, Jiang Z, Shen H, Gobbens RJ, et al. Blended self-management interventions to reduce disease burden in patients with chronic obstructive pulmonary disease and asthma: systematic review and meta-analysis. J Med Internet Res 2021 Mar 31;23(3):e24602 [FREE Full text] [doi: 10.2196/24602] [Medline: 33788700]

54. Fixsen DL, Blasé KA, Timbers GD, Wolf MM. In search of program implementation: 792 replications of the teaching-family model. Behav Anal Today 2007;8(1):96-110. [doi: 10.1037/h0100104]
55. Versluis A, van Luenen S, Meijer E, Honkoop PJ, Pinnock H, Mohr DC, et al. SERIES: eHealth in primary care. Part 4: addressing the challenges of implementation. Eur J Gen Pract 2020 Dec;26(1):140-145 [FREE Full text] [doi: 10.1080/13814788.2020.1826431] [Medline: 33025820]
56. Khalid A, Dong Q, Chuluunbaatar E, Haldane V, Durrani H, Wei X. Implementation science perspectives on implementing telemedicine interventions for hypertension or diabetes management: scoping review. J Med Internet Res 2023 Mar 14;25:e42134 [FREE Full text] [doi: 10.2196/42134] [Medline: 36917174]
57. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
58. Tossaint-Schoenmakers R, Versluis A, Chavannes N, Talboom-Kamp E, Kasteleyn M. The challenge of integrating eHealth into health care: systematic literature review of the Donabedian model of structure, process, and outcome. J Med Internet Res 2021 May 10;23(5):e27180 [FREE Full text] [doi: 10.2196/27180] [Medline: 33970123]

Abbreviations

CARAT: Control of Allergic Rhinitis and Asthma Test
ICS: inhaled corticosteroids
QoL: quality of life
SABA: short-acting beta2-agonist
SF-36: 36-Item Short-Form Health Survey
SUS: System Usability Scale
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by G Eysenbach, A Kushniruk; submitted 08.11.23; peer-reviewed by UK Chalwadi, D Drummond; comments to author 06.01.24; revised version received 26.01.24; accepted 18.02.24; published 04.04.24.

Please cite as:
van den Berg LN, Hallensleben C, Vlug LAE, Chavannes NH, Versluis A
The Asthma App as a New Way to Promote Responsible Short-Acting Beta2-Agonist Use in People With Asthma: Results of a Mixed Methods Pilot Study
JMIR Hum Factors 2024;11:e54386
URL: https://humanfactors.jmir.org/2024/1/e54386
doi:10.2196/54386
PMID:38574348

Original Paper

A Smartphone App to Support Self-Management for People Living With Sjögren's Syndrome: Qualitative Co-Design Workshops

Claire McCallum[1*], PhD; Miglena Campbell[2*], PhD; John Vines[3], PhD; Tim Rapley[4], PhD; Jason Ellis[5], PhD; Vincent Deary[5], PhD; Katie Hackett[4], PhD

[1]Faculty of Engineering, University of Bristol, Bristol, United Kingdom

[2]Institute for Collective Place Leadership, Teesside University, Middlesbrough, United Kingdom

[3]School of Informatics, University of Edinburgh, Edinburgh, United Kingdom

[4]Department of Social Work, Education and Community Wellbeing, Northumbria University, Newcastle upon Tyne, United Kingdom

[5]Department of Psychology, Northumbria University, Newcastle upon Tyne, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Katie Hackett, PhD
Department of Social Work, Education and Community Wellbeing
Northumbria University
Coach Lane Campus West
Newcastle upon Tyne, NE7 7XA
United Kingdom
Phone: 44 191 215 6098
Email: kate.hackett@northumbria.ac.uk

Abstract

Background: Sjögren's syndrome (SS) is the second most common autoimmune rheumatic disease, and the range of symptoms includes fatigue, dryness, sleep disturbances, and pain. Smartphone apps may help deliver a variety of cognitive and behavioral techniques to support self-management in SS. However, app-based interventions must be carefully designed to promote engagement and motivate behavior change.

Objective: We aimed to explore self-management approaches and challenges experienced by people living with SS and produce a corresponding set of design recommendations that inform the design of an engaging, motivating, and evidence-based self-management app for those living with SS.

Methods: We conducted a series of 8 co-design workshops and an additional 3 interviews with participants who were unable to attend a workshop. These were audio recorded, transcribed, and initially thematically analyzed using an inductive approach. Then, the themes were mapped to the Self-Determination Theory domains of competency, autonomy, and relatedness.

Results: Participants experienced a considerable demand in the daily work required in self-managing their SS. The condition demanded unrelenting, fluctuating, and unpredictable mental, physical, and social efforts. Participants used a wide variety of techniques to self-manage their symptoms; however, their sense of competency was undermined by the complexity and interconnected nature of their symptoms and affected by interactions with others. The daily contexts in which this labor was occurring revealed ample opportunities to use digital health aids. The lived experience of participants showed that the constructs of competency, autonomy, and relatedness existed in a complex equilibrium with each other. Sometimes, they were disrupted by tensions, whereas on other occasions, they worked together harmoniously.

Conclusions: An SS self-management app needs to recognize the complexity and overlap of symptoms and the complexities of managing the condition in daily life. Identifying techniques that target several symptoms simultaneously may prevent users from becoming overwhelmed. Including techniques that support assertiveness and communication with others about the condition, its symptoms, and users' limitations may support users in their interactions with others and improve engagement in symptom management strategies. For digital health aids (such as self-management apps) to provide meaningful support, they should be designed according to human needs such as competence, autonomy, and relatedness. However, the complexities among the 3 Self-Determination Theory constructs should be carefully considered, as they present both design difficulties and opportunities.

(JMIR Hum Factors 2024;11:e54172) doi:10.2196/54172

KEYWORDS

self-management; mobile health; mHealth; eHealth; Sjögren's syndrome; patient participation; patient involvement; fatigue; chronic disease; focus groups; complex intervention development; mobile phone

Introduction

Background

The need to improve the accessibility and quality of care for those with long-term conditions (LTCs) is an international priority [1]. In England alone, LTCs affect 15 million people [2] and account for 70% of health care spending [3]. Rheumatic diseases are LTCs with a particularly high prevalence in the United Kingdom and worldwide, having been estimated to affect up to one-fourth of Europeans [4,5] and a similar proportion of the population in the global south [6]. Sjögren's syndrome (SS) is thought to be the second most common autoimmune rheumatic disease [7] and is associated with poor quality of life [8] and high disease burden [9].

SS is a heterogeneous LTC, with a constellation of unpredictable and diverse symptoms [10,11]. A key characteristic of SS is mucosal dryness due to the destruction of exocrine (moisture-producing) glands by the body’s immune system, which particularly affects the eyes, mouth, and vagina [12]. In addition to dryness, common extraglandular features include persistent fatigue [13], chronic pain [14], sleep disturbances [15], and anxiety and depression [16]. People with SS report experiencing these symptoms as being interconnected, with the exacerbation of one symptom impacting others [17-19].

Similar to many other autoimmune diseases, SS does not have a cure [20]. Therefore, intervention efforts have focused on reducing the severity of symptoms; for instance, topical treatments are used for managing dryness [21]. Drug treatments for the systemic management of SS, such as hydroxychloroquine and rituximab, have had disappointing results in clinical trials [22,23]. Behavioral interventions that aim to improve the quality of life are a promising alternative; however, few interventions have been developed, and evaluations of their impact have been of low quality [21,24]. A recent stakeholder engagement study found that support for self-managing symptoms was a key priority for people with SS [25]. The term self-management has been defined as “the individual’s ability to manage the symptoms, treatment, physical and psychosocial consequences and lifestyle changes inherent in living with a chronic condition” [26,27]. To support the knowledge, behaviors, and attitudes required, self-management interventions should deliver a range of educational, behavioral, and cognitive techniques [28]. In SS, a targeted “complex” intervention is required, which delivers multiple techniques and targets multiple SS symptoms [29]. Our previous body of work with patients with SS found that they require different levels of support. Some require more complex individual support, but most people require lower levels of support with access to written information and digital self-management tools [29], which could be provided in the form of a website or smartphone app.

Apps as a Support for Self-Management

SS shares multiple symptom and self-management similarities with other LTCs [30], including but not limited to neurological and autoimmune conditions such as rheumatoid arthritis, myalgic encephalomyelitis, and multiple sclerosis. Smartphone apps are a promising approach to support self-management of these LTCs [31,32] and other conditions such as type 2 diabetes [33], asthma [34], and hypertension [35]. Their increasing availability and functionalities enable complex intervention techniques to be delivered in the context of users’ daily lives when they are designed with consideration of users’ routines and choices [36]. User-centered design studies of LTCs have produced various app features and content [37] to support, for example, user education and cognitive strategies. However, app effectiveness can be limited by very low levels of user engagement [38,39]. Therefore, intervention developers must design apps that are more engaging and carefully consider how such engagement will ultimately lead to long-term behavior change [40]. For example, beyond simply providing information about how to perform techniques, apps can be designed to promote a sense of autonomy and motivation to engage in self-management behaviors over time [41,42].

To increase user engagement, apps should be user centered and person centered [43], that is, designed to fit within individuals’ current lives and daily activities [44]. People are more likely to use a new intervention if it can be incorporated into their existing habits, routines, and contexts [45,46]. Therefore, self-management interventions should account for and actively support how people manage their conditions currently [31,47]. Thus, to develop a useful, effective, and engaging app-based intervention that supports those with SS, there is a need to first understand their current self-management opportunities and challenges. To date, limited studies have only been conducted to understand the lived experience of symptoms [17,19,48] and have not explored the self-management of multiple diverse symptoms.

To gain an understanding of individuals’ self-management contexts, co-design and user-centered methods are useful [49]. These can involve practical design activities that elicit conversations regarding a topic of interest (such as self-management) to inform the development of a design, product, or intervention and have been used to develop digital health interventions [50,51]. Then, to understand how users in these contexts might best be supported in changing their self-management behavior, co-design findings can be interpreted using theories of motivation and behavior change [41].

Self-Determination Theory (SDT) [52] is one motivational theory widely used in interventions promoting health behavior change [53,54], including those for self-managing chronic illnesses [55,56]. SDT proposes that the constructs of competence, autonomy, and relatedness are required for individuals to be internally motivated to perform behaviors and sustain these changes over time. Situating qualitative findings within theoretical constructs facilitates the development of apps that are based on theory [42,57]. While intervention developers use SDT to inform their interventions, many do not explicitly link the theoretical constructs directly to their individual components, and we aimed to bridge this gap. To the best of our knowledge, there are no evidence-based, theory-driven, self-management apps for SS.

Study Aims

We aimed to use an SDT framework to explore self-management challenges and approaches used by people with SS and to produce a set of design and therapeutic recommendations for a supportive and engaging app to aid self-management.

Methods

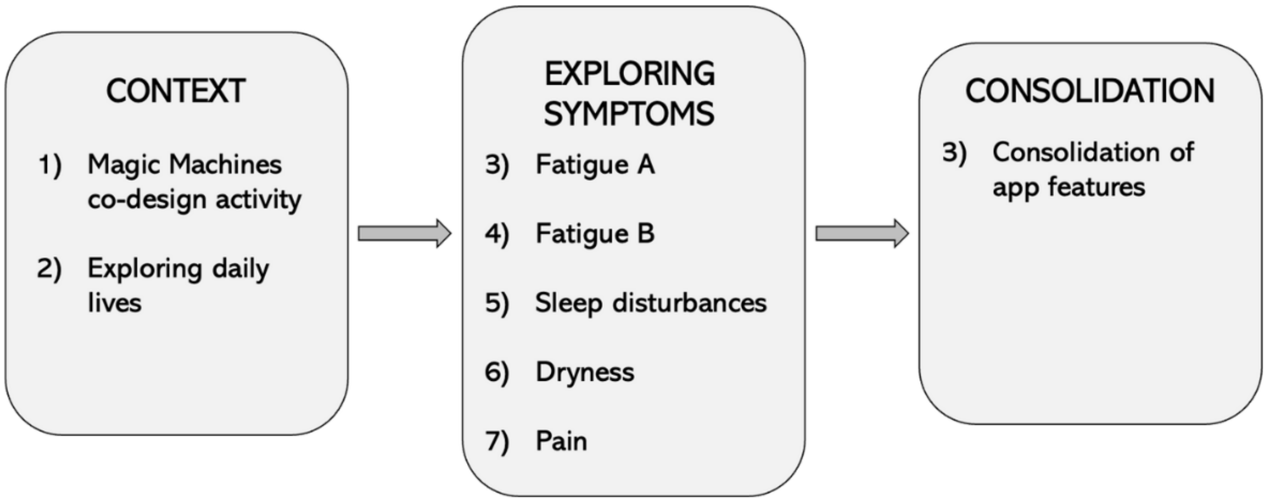
The methods and subsequent results have been reported according to the COREQ (Consolidated Criteria for Reporting Qualitative Research) guidelines (Multimedia Appendix 1 [58]).

Study Design

A consecutive series of 8 workshops with people living with SS was conducted over 7 months, each involving design activities and focused discussions (Figure 1). The first 2 workshops were open ended to broadly understand participants' contexts (ie, key self-management challenges and overall self-management routines) and enable participants to become familiar with each other and to feel comfortable while discussing potentially sensitive and personal topics. We decided in advance to include a series of workshops, with each workshop dedicated to in-depth understanding of the self-management activities, challenges, and opportunities for each symptom. However, the order of symptom workshops and their exact discussion topics and activities were not predetermined; their sequential nature enabled us to iteratively design topics based on the findings from the previous session. For example, a clear theme emerged regarding symptom interrelatedness, so subsequent workshops included discussions about how participants managed interrelations among their symptoms. Furthermore, fatigue was a priority for all workshop participants (14/14, 100%), and therefore, 2 workshops were dedicated to this symptom.

Participants were given the option to attend ≥1 workshops. Several workshops were repeated to suit participants' availability. To enable those who could not access any workshops due to other commitments and to include the experiences of younger people living with SS, 3 one-off semistructured interviews were conducted. These focused on the key self-management practices and challenges experienced by the respective participant.

Figure 1. The procedural flow and topics of the 8 design workshops.



Ethical Considerations

This qualitative study received ethics approval from Northumbria University ethics committee (reference 11130). Informed consent was obtained before data collection, and travel costs were reimbursed.

Recruitment

Workshop participants were purposively recruited from a regional UK SS support group (Northeast Sjögren's Syndrome Association). Advertisements were distributed via their member mailing list and Facebook page, and the research team presented the project at a support group meeting. The invitation was open to those diagnosed with SS by a physician, and potential participants were invited to attend as many workshops as they liked. Interested participants who were unable to attend due to their location or life commitments were invited to attend a video web-based interview. Additional participants were recruited via social media (a single tweet on Twitter [subsequently rebranded as X]) and invited to participate in the interviews only.

Data Collection Activities

Overview

Workshops were conducted at Northumbria University, lasted approximately 90 minutes, and included a 10-minute comfort break. The interviews lasted 30 to 60 minutes and were conducted via telephone or videoconferencing software. Workshops were facilitated by 3 authors (CM, MC, and KH)—all were female postdoctoral (PhD) researchers trained in qualitative research methods and experienced in conducting qualitative research interviews and focus groups; one of them was also an occupational therapist (KH) with experience in SS symptom management. Several workshop participants (4/14, 29%) had attended clinics (conducted by KH), and 14% (2/14) had participated in previous studies (conducted by KH and VD). All participants (17/17, 100%) were briefed about the aims of the study. All workshops and interviews were audio recorded, and facilitators took field notes. In the following sections, we have outlined the focus of each workshop. The individual workshop topic guides are presented in Multimedia Appendix 2 [59-62].

Workshop 1: Magic Machines Co-Design Activity

This workshop introduced the series of workshops, included discussion about some key self-management issues experienced by participants, and involved a Magic Machines [54,59] craft activity where participants created some imaginative design solutions for another workshop participant. The Magic Machines activity aimed to elicit a broad range of knowledge about participants' personal and technological needs through discussions about everyday problems related to their condition and potential solutions. Participants were asked to create an object, which addressed their partner's daily challenge, using household objects and craft items. Data capture was focused on the conversations between participants about their "problem" while making their objects (a potential "solution") and when describing their object to the main group at the end of the session.

Workshop 2: Exploring Daily Lives

The second workshop explored individuals' "daily lives" and the self-management of symptoms. The discussion about daily lives invited participants to discuss their "typical day" in managing SS (ie, their habits and routines), how SS self-management was incorporated into their routines, and any related challenges that they experienced.

Workshops 3 to 7: Exploring Symptoms

These workshops explored the self-management of specific symptoms and their interrelationships through group discussions and invited participants to engage in basic sketching to articulate their self-management experiences and challenges. We preselected the symptoms for discussion based on our previous study where patients identified them as being important and impacting their daily activities [25].

Workshop 8: Consolidation

Sketching was used to explore how an app might be structured to support symptom interconnectedness and complexity. This design activity also elicited discussion about user experience and usability issues. All participants (5/5, 100%) attending this workshop engaged in sketching, but if time was insufficient, they were encouraged to further develop their ideas by articulating them verbally.

Interviews

Following the workshops, 3 semistructured, web-based interviews were conducted by CM. The interviews followed a schedule of open-ended questions to allow for flexibility (Multimedia Appendix 3).

Data Analysis

Audio data were transcribed verbatim, pseudonymized, and combined into a corpus for analysis using NVivo (version 12; QSR International). Analysis was conducted in 2 phases using a hybrid approach, incorporating both inductive and deductive methodologies, to harness the advantages of both methods [63]. First, an inductive thematic analysis approach [64] was used, where 2 researchers (CM and MC) independently coded the data, generating an initial set of codes related to participants' self-management perceptions and experiences. Then, these codes were applied and refined through the arrival of each new transcript, and independent coding was subsequently conducted by CM. Discussions during regular research team meetings (with CM, MC, and KH) related to the codes and their connections, importance, and relevance were conducted to group codes into themes.

Then, these inductive themes were mapped to the 3 SDT [65] constructs of competency (the sense of capability to perform activities and tasks), autonomy (experience of having control and choice over one's actions and decisions), and relatedness (feeling of connection and belonging and meaningful interaction with others) by CM. SDT was chosen over other motivational theories because it emphasizes social context as a key factor in helping or hindering motivation, which matched a prominent theme in our inductive thematic analysis of social relations, along with other major themes we found related to empowerment, autonomy, and capability (or "competency" in SDT). The theory is also highly translational, enabling findings to inform intervention design [66]. Regular research team meetings were conducted to review and reach consensus regarding the categorization of themes based on the SDT constructs. Opportunities to support participants' challenges associated with these themes through an app were also identified through discussions. Methodological rigor and credibility of findings were pursued through development of a codebook, maintenance of ongoing reflexivity, peer debriefings, and data triangulation (from interviews, focus groups, and observations during workshop activities).

Results

Participants

In total, 17 people with SS participated in the workshops and interviews: 14 (82%) of the 17 participated in the workshops (13/14, 93% women and 1/14, 7% men) and 3 (18%) of the 17 participated in the web-based interviews (3/3, 100% women). Participants' ages ranged from 33 to 76 (mean 56.5, SD 13.95) years, and 82% (14/17) of them had a diagnosis of primary SS. The remaining 18% (3/17) of the participants had a diagnosis of secondary SS. The mean number of years since diagnosis was 7.5 (SD 7.88) years. Regarding employment status, of the 17 participants, 8 (47%) were retired, 6 (35%) were working full time, 1 (6%) was in part-time employment, and 2 (12%) were not working currently. All workshop participants (14/14, 100%) had links to a local SS support group in the north of England. Of the 3 interviewees, 1 (33%) was part of the same support group, and the remaining 2 (67%) were from Spain and Canada, respectively (both were aged <35 years). Workshop group numbers ranged from 2 to 7 participants and 2 to 3 facilitators in each session. Some participants attended some sessions and not others, whereas a “core” group of 5 participants attended most sessions. One participant (1/14, 7%) attended only the first workshop; others attended at least 2 sessions.

Overview

Participants engaged in a wide range of self-management behaviors, including using prescribed and over-the-counter medications and treatments (ie, applying eye drops and gels; bathing and massaging the eyes; using humidifiers, skin creams, and vaginal lubricants; following mouth care routines; using pain medication; and using hot and cold compresses). They also used cognitive and behavioral techniques including activity pacing, goal setting, general exercise, relaxation, mindfulness, distraction, napping, sleep management and wind-down routines, and social support. Participants used various tools to support and facilitate the learning, use, and practice of these techniques, including books (eg, about managing fatigue), diaries (paper based and digital), websites and forums (eg, National Health Service or SS associations as both knowledge resources and social support), apps on smartphones and tablets (such as for yoga, breathing exercises, and mindfulness), wearables (to track physical activity), and other devices (eg, for relaxing music, “mindless” television, or distracting podcasts and comedy). Not all participants owned or used smartphones. Tools were used in addition to visiting friends and holistic wellness centers (eg, spas and mindfulness classes) and learning self-care techniques directly from health care professionals (eg, when to apply eye drops and more complex techniques such as activity pacing and graded exercise).

In the following sections, we have described the challenges that participants faced in managing their condition and their psychosocial needs in terms of competency, autonomy, and relatedness.

Competency

Participants varied in the extent to which they felt competent and successful in self-managing their SS, and this was related to how well they had established a self-care routine. One participant had a very “strict regime,” which they felt was required to maintain their level of functioning. While hearing about such self-management strategies from others, Jim reflected about his competencies:

> *I’ve still got quite a lot to learn...although it has been a few years now, I think I still haven’t got a good routine...I listen to your explanation [of another participant’s routine] and I think, why can’t I get myself like that? I’m supposed to be Mr Organised. I am known as that in my life. My working life and my own home life. Yet with this, I have not gotten organised yet.* [Jim]

Regardless of whether participants had routines or described habitual self-management behaviors, their sense of competency in self-managing SS was still impeded by the complex nature of their symptoms. Isolating and targeting individual symptoms was not only perceived to be difficult to perform (“You can’t separate the different symptoms” [Jim]) but were also sometimes unhelpful, as it did not account for their accumulative negative impact:

> *It is the overall effect to me. That three [symptoms] I can cope with and then the next day one raises its head and floors me...That straw that broke the camel’s back effect, you know. [Patricia]*

Several participants believed that they could better manage their symptoms through self-management techniques capable of improving multiple symptoms simultaneously. Some had discovered these types of techniques accidentally. For example, participants recounted noticing, with surprise, that eye drops had helped not only their dryness but also mental and physical fatigue. Other participants purposefully sought and regularly used techniques that targeted multiple symptoms simultaneously. Mindfulness and relaxation techniques provided a sense of control and the ability to “keep a cap on” multiple symptoms before they became very severe. Others agreed that seeking these techniques was worthwhile if they resulted in minor improvements across multiple symptoms. Despite valuing self-management techniques that targeted multiple symptoms, most did not feel confident or knowledgeable about which techniques were beneficial.

Another challenge to participants’ sense of competency is how SS symptoms are not static but change over time. Participants described instances where individual symptoms would rapidly fluctuate in severity:

> *[They] come and go...one day you might have a headache, the next day you don’t. [Jim]*

Participants also explained experiencing longer periods where multiple symptoms were severe (described using phrases such as a “flare,” “phase,” or “wave”) or individual symptoms persisted (such as “a dry patch”). While, sometimes, the onset of symptoms appeared “gradually,” at other times, they changed rapidly, leaving participants feeling unprepared (“a phase hits you”). Fatigue and pain were felt to be particularly volatile and could become severe with no warning and “like somebody just switched a switch” (Penny).

Participants varied in how they managed such changing symptoms. Many attended to symptoms as they arose or increased in severity on a moment-to-moment basis (ie, an adaptive or reactive approach). However, this often meant devising complex and intricate strategies and sequences to manage the new combination of symptoms experienced in that moment. For example, sleep disturbances that might be attributed to pain, dryness, or anxiety required participants to change their approach to getting back to sleep accordingly (“depending on how I am” [Penny] or “what problem I am having” [Jim]). Other participants seemed to disregard the changing combination of symptoms and addressed symptoms “one at a time” based on whether they felt successfully managed. For example, Julie noted the following:

> *I tend to find like I feel like my feet are sorted, so I am now sorting my eyes, so I’m kind of going through this list.*

Addressing symptoms required constant adjustments for participants. Their variable nature meant that just at the point that the individual starts to feel in control of one symptom, a flare of another may occur.

A final layer of complexity impacting participant competency was how symptoms often change due to environmental factors. For example, dryness was exacerbated by air conditioning, bright lighting, and other people's aftershaves and perfumes, whereas navigating new and busy places could exacerbate mental fatigue. The unpredictable nature of environments outside the home made self-managing symptoms more challenging. While home was characterized as "familiar" and "unchanging," participants felt that they needed to continuously estimate the potential impact of environments on their symptoms and plan accordingly:

> *You have to be very wary of where you're going...you've got to be careful. I will not walk through [the shopping mall] in the perfumery because there is always somebody going to...pick up a bottle of perfume and [spray]...I go, oh my eyes! [Geraldine]*

This planning itself was exhausting to several participants, and it also meant that they lacked spontaneity in their lives. Participants also felt that symptoms were easier to manage at home because they could easily perform physical relaxation and self-care techniques when required, particularly during a flare. During such times Sarah remarked as follows:

> *I just don't want to leave the house, I don't want to do anything. I just want to go and have like 2-3 baths per day.*

In contrast, when symptoms left them debilitated outside the house, participants had to adopt different self-management techniques such as soothing self-talk or be "rescued" by a taxi or friend. Overall, being outside the home meant that participants were less in control of their environments; had to continuously plan and predict how the environment might impact their symptoms, which was mentally and physically tiring; and had to use different techniques to suit different environments.

Autonomy

Our analysis identified many examples of participants feeling that they had autonomy in the self-management of their condition; however, sometimes, the same factors that promoted autonomy also reduced confidence and competence. Participants believed that the availability of various techniques meant that they had options in their self-management; there were multiple "different ways" they could try to improve symptoms. The plurality of techniques appeared to provide reassurance that at least one would be likely to be effective:

> *I have six choices...I don't beat myself up when it doesn't work because I've already got something else in mind.* [Patricia]

This plurality and optimism could provide a strong drive to continue in their self-management activities.

Participants varied in how they kept track of different available techniques. One participant had self-help books at various locations in their home. Another participant explained that they had collated several techniques to create their own book:

> *I wrote myself a little book...[of] top tips...I just wrote maybe two dozen messages across the book at random, things that might give me a clue. [Debra]*

Other participants used an experimental approach:

> *It's about learning...through trial and error...you'll notice a pattern...you don't know until you've done it for a few months.* [Michelle]

These were similarly characterized by the desire to try different techniques and to keep track of their effectiveness:

> *With time and experience you begin to realise what works and what doesn't.* [Penny]

It was acknowledged that this required continuous effort and perseverance.

Having personal choice to decide which techniques to try, as opposed to being directed by a health care professional, provided some participants with a sense of control. Debra likened creating her book of techniques to developing a tailored smartphone app:

> *It is basically my own app that I've written for myself...I didn't feel like being ordered around by anybody else...I don't necessarily follow it. If it's inside my book, I think, well alright, maybe I'll try something else...I've still got some kind of control over things. [Debra]*

Therefore, developing this herself meant that she did not feel obligated to try any 1 technique. Although participants appreciated having the autonomy to choose techniques in a personalized manner, the credibility of these techniques was also very important to them. Perceived credibility seemed to give them confidence to go ahead and try them. Some participants indicated that they understood the distinction between evidence-based information and hearsay:

> *I am pretty much someone who will try anything once if there's some evidence to support its effectiveness...Some people suggest real outlandish things, like you hear it and you're like, "okay!" I mean, I'm glad that it works for you, but I'm not really sold on trying that just yet.* [Ellie]

Participants felt that information about their condition or how to manage symptoms should be credible. For example, Jim explained that simply being presented with multiple self-management techniques and options, without a rationale for why they might be helpful, would not suit him. Others stated that knowing information sources was "useful...[for] controlling symptoms and trying to minimise [them]" (Edith). Information from websites such as the UK National Health Service or regional and national SS organizations was deemed trustworthy.

Although participants respected expert advice and implemented it in their self-management, expert authority was often only 1 element in an autonomous process of symptom management decision-making. For instance, when faced with a conflict between their preferred routines and expert advice, participants trusted their own expertise and experience. Jim outlined how he fell asleep with the help of music or old comedy shows and that he would simply "ignore" any potential prompts about adjusting his bedtime routine if it meant removing his music from the bed (as may be advised as part of a sleep intervention).

For some participants, smartphones and associated apps appeared to contribute to feelings of autonomy regarding their self-management of SS. Those who used a smartphone reported using basic note apps to track symptoms or calendars built into the operating system to track feelings of fatigue. Experienced smartphone users described how their ubiquitous nature enabled quick access to information and could give them access to techniques whenever and wherever needed, regardless of their location. In sessions where feedback was given about potential app designs, participants expressed the value they would see in new apps that brought various techniques together, provided reminders to apply eye drops, and helped track symptoms in a simpler manner. For instance, Julie suggested that “a tracker or a journal...or something like that on the app would be helpful” as this could help her manage her forgetfulness, which she referred to as “brain fog.”

However, while smartphones could enable autonomy, they also posed challenges that could impact the users’ SS symptoms. It was also noted that looking at the screen of a computer or smartphone for a very long time could exacerbate eye dryness:

> *It’s okay [when it’s] short, but you can’t spend a long time looking at the screen, because your eyes are just too sore.* [Mel]

To overcome this, participants used their smartphones differently. Some described deliberately limiting the amount of time they used them in 1 session, and others described changing their device settings to increase the font size or darken the screens. In addition, participants mentioned improving on-screen accessibility to reduce their eye strain and listening to audio instead of reading text. Patricia, noted that when “I am having my brain fog” the complexity of most apps “would blow my mind.” Among participants, there was a sense that smartphone use was closely related to experiences of mental fatigue from their SS.

Overall, participants valued the diversity of SS self-management techniques that are available and experienced this as enhancing autonomy. Smartphones and both generic and SS-specific apps were viewed as an important part of this diversity and could provide in situ tailored support. However, the apparent abundance of techniques and availability of smartphones also posed a challenge to autonomy. Patricia recounted that soon after being diagnosed with SS, she was overwhelmed by the need to learn about multiple symptoms and techniques from many sources. However, for her, this felt similar to being “shot at” from multiple angles. Sometimes, the factors that enable autonomy can also constrain it.

Relatedness

Relatedness refers to the manner in which participants operated in their social worlds and how their practices of managing SS were related to it. Participants explained that SS profoundly impacted their familial interactions, friendships, and other forms of social contact. Participants enjoyed social activities and cherished positive relationships as a source of social support. Socializing and participating in activities with others provided a positive “distraction” from their symptoms. However, self-management tasks could impact their ability to socialize and interrupt the flow of conversations:

> *When in company if you are out and about and talking to people...You have to keep popping off to go and put eyedrops in, in the loo.* [Edith]

Furthermore, engaging in certain social activities, such as going to the cinema with friends, required participants to perform additional self-care, for example, applying eye drops more frequently, which could irritate the skin around the eyes. Geraldine explained that although she enjoyed going to live theatre performances, she was now reluctant to go based on previously being “crucified” by a smoke machine.

Pacing was a helpful technique to manage fatigue, but it was not always received well by others in social situations and workplaces. Patricia recounted that she had been regarded as “selfish” by family and friends for cancelling plans while trying to manage her energy and fatigue levels. She also recognized that having to “book” people into her diary well in advance to support her planning and pacing efforts “frightens some people off.” Edith recalled that the need to take more breaks meant that she had to decide to leave her walking group as she was no longer able to keep up with her friends. In turn, this negatively impacted her feeling of belonging.

Communication was key while managing illness demands and relationships. Some participants created their own SS information sheets to give to friends and health care professionals. Creating opportunities to explain difficulties was conducive to receiving valuable social support. Penny’s husband had delegated several household tasks to her, which were conflicting with her pacing technique. Penny explained that after discussing the issue with him, he subsequently understood the need to balance activities and that they were able to do this together. Ellie noted the following:

> *I do think that it is helpful to have people that you can talk to about Sjögren’s. I mean I have a very close relationship with my family, and I have close friends who I do feel like I can confide in, and that is really helpful for me.* [Ellie]

The freedom and ability to be open and honest about their SS symptoms with trustworthy family members and friends were central to well-being and helped with symptom self-management. However, despite all efforts to communicate effectively, many participants believed that, often, family, friends, and even health care professionals did not fully understand SS. They felt frustrated that symptoms were dismissed, normalized, or incorrectly attributed to other issues such as “getting old” or menopause. Dealing with invisible, ever-changing symptoms was difficult. Multiple general practitioner visits with complaints about seemingly benign symptoms such as fatigue and thirst were sometimes received with skepticism, and the transient nature of these symptoms made the situation worse:

> *Then you’re fine and you think, “they’ll think I am putting it on.”* [Geraldine]

Any respite from symptoms made some participants worry that those around them would not believe them the rest of the time. Carol knew that relative to other conditions that may have 1 visible "major" symptom, her multiple symptoms were unlikely to garner support and understanding because of "Sjögren's [and] all the little things that it has" (Carol). Some participants had stopped attempting to explain their symptoms to family and friends, saying that some symptoms were "very difficult for you to articulate...to somebody who doesn't feel it" (Joan). This was particularly detrimental to relationships with health care professionals. When health care professionals seemed uncompassionate about their symptoms, some participants talked about "shutting down" and making a choice to no longer discuss their SS in consultations. This had negative consequences on participants who ended up feeling rejected and disengaged, and there was a perception that, sometimes, health care practitioners were not even aware of this relational and motivational shift.

When participants felt disbelieved, it led to experiences of self-doubt. Ellie said she was "bounced around like 4-5 practitioners" to the point where she questioned her illness "almost as if it is in your head." Carol resorted to maintaining an activity diary, in part to monitor her fatigue and to preserve her sanity. For her, the diary data provided a sense of external objectivity and an opportunity to feel validated when being questioned by other people:

> *By doing the [diary] you think, yes...I've got a problem and that graph tells me...it is a physical thing, it's not in my mind.* [Carol]

Being diagnosed with SS was a lonely experience for some participants due to the challenges of family, friends, colleagues, and health professionals not relating to participants' symptoms or condition. Social isolation was particularly pronounced for a younger person with SS:

> *I don't know anyone else who has it. So, it is kind of isolating...I also had a hard time finding people who are...my age. So, I mean, I would definitely be interested in meeting younger women who are working, who are finding strategies.* [Ellie]

Overall, connecting with others with SS was important, and participants sought opportunities to meet others with SS, learn, and find the validation and understanding they did not receive from others without SS. Some joined support groups and attended scientific conferences to expand their social circle with other people with SS.

However, not all social contact with others with SS was deemed helpful:

> *Some of the interactions I had honestly more scared me than helped me because it was people who were really in the throes of severe illness and some who weren't coping well, and it was sort of anxiety-provoking.* [Ellie]

Therefore, support from others with SS was generally more welcome when it was helpful and positive, as interactions with those who were struggling to cope could have a negative impact on participants.

Within the construct of relatedness, even positive self-management was found to impact social activity, but having highly supportive friends and family could mitigate this to some extent. Describing and explaining the various, ever-changing symptoms to colleagues, friends, and health professionals who did not fully understand the condition or symptoms could be particularly challenging, but external resources such as using diary data could be a helpful tool to aid communication.

Discussion

Principal Findings

We sought to understand the current self-management approaches used by people with SS to inform the therapeutic ingredients and design recommendations for a self-management smartphone intervention. To date, most studies of lived experience with SS have focused on how specific symptoms are experienced [17,18,48]. To the best of our knowledge, no studies have explored how people with SS perform the day-to-day work of managing their condition and navigating challenges as they do so. This is an important consideration when designing interventions, as those that draw upon users' expertise are more likely to be used [38]. Therefore, we analyzed qualitative data collated through a series of workshops and interviews with people with SS inductively before mapping the themes to the 3 constructs of SDT (competency, autonomy, and relatedness) [52]. This theory was used because it can help identify the psychosocial and practical requirements to support autonomous motivation to adopt and sustain healthy behaviors and to improve well-being in a population [52,67]. Our findings were consistent with what Cartner [68] first described in her qualitative study with participants with SS: the *labor* of living with SS. For her and our participants, competency was an ongoing effort, never a completed achievement. The complex, multisymptomatic, volatile, and unpredictable nature of the condition meant that their hard-earned expertise was being constantly challenged. Having to adapt to an ever-changing and unpredictable challenge evokes the concept of stress, but more specifically, it is captured by the notion of *allostasis:* the work that needs to be performed to find stability within a situation that is constantly changing. When allostasis is frequent or continuous, more work needs to be performed, and our emotional, cognitive, and biological resources can become dysregulated. This is known as *allostatic load*—the psychophysiological wear and tear that occurs to a system that is constantly having to adapt—which has clear links to anxiety, depression, morbidity, and mortality [69].

The labor of the participants and its costs were also evident in the SDT domain of autonomy. Often, there was a degree of forced autonomy, with participants having to perform the epistemic labor of determining how to manage their condition for themselves. This involved ongoing research and even compiling their own resources. Discernment and discrimination were required to determine what advice to trust and follow and how to balance that advice against their own experience. Although this process was enabling, it was also potentially disabling as the process of gathering and compiling information worsened some SS symptoms.

Finally, in the realm of relationships, managing SS requires significant social labor. Often, participants were required to manage the expectations, lack of understanding, skepticism, and disbelief of others, including health professionals, and these efforts were often only partially successful, leading to self-doubt, isolation, and lack of adequate care and support for their illness. This is not dissimilar to the experiences others face with other fatiguing LTCs such as stroke, fibromyalgia, multiple sclerosis, and ankylosing spondylitis [70].

Design Recommendations

Multimedia Appendix 4 [52] summarizes our key findings, which have been mapped to the 3 constructs of SDT, with identified therapeutic approaches and design solutions for each. The findings within these SDT domains were identified as targets for intervention by the participants. In the following sections, we have reviewed these domains and suggested what interventions might help and how the interventions could be incorporated into an app to support self-management.

A key finding within the competency domain was that SS was multisymptomatic, volatile, and unpredictable. Participants were keen for interventions that would impact >1 symptom at a time. A previous study that investigated patient strategies for self-management of inflammatory bowel disease had similar findings [71]. Several treatment approaches and their components discussed during the workshops could potentially address several symptoms simultaneously. For example, activity and sleep management strategies such as pacing and reflective activity diaries have been used to support self-management of pain, sleep disturbances, and fatigue [72-76], and previous studies that evaluated interventions targeting several symptoms have shown promising results. Therefore, we suggest that when designing complex interventions for LTCs, intervention developers should map the potential, identified intervention content to behaviors and symptoms and select techniques that target >1 symptom where possible, thereby placing a smaller demand on the user. While this may not always be possible, streamlining the intervention content where practicable is likely to decrease the possibility of becoming overwhelmed and thereby supports user competency.

The key challenge in the autonomy domain was the amount of work required by participants to determine how to manage their condition on their own. As with many other LTCs, a large part of the "burden of treatment" is shouldered by the person with the condition [77]. Our findings broadly indicate that technology-enabled symptom management could help with this work of illness management. Participants liked the idea of a smartphone app to support self-management. However, merely operationalizing technology is not sufficient to promote and support self-management. Güldenpfennig et al [78] found that poor design and well-meaning paternalism, for example, through automated support that takes active choice away from the user, may compromise autonomy and proactive self-management. Furthermore, intervention designers should aim to strike a more careful balance between the input of experts by experience and those of professional experts [25,78]. In our study, we found that people with SS managed their symptoms using different approaches but that all of them had arrived at their own set of strategies and management regimes through experience, research, and trial and error. Acknowledging the individuality of self-management and the necessity to experiment with different approaches would be a key part of any intervention. Having a repository of strategies in 1 centralized app, which would also allow them to add their own strategies, would seem to be a potentially useful resource. This aligns with previous studies of apps that provide resources while allowing customization and thus may support a user's sense of autonomy [79-82] and move away from a top-down paternalistic or prescriptive approach to LTC management [83,84]. An app for SS would need to combine recognition of the labor of self-management while helping to support it in a manner that honors the user's autonomy and existing wisdom, providing the ability to choose from a range of therapeutic content and to determine the order in which they interact with it.

The most difficult and often fruitless area of labor was observed within the relatedness domain. Participants were required to manage others' expectations, lack of understanding, skepticism, and disbelief, often leading to a smaller social world, isolation, and difficulties in accessing help from health professionals. Again, any intervention needs to begin by acknowledging this labor and the emotional and social costs of having a poorly understood and invisible illness. Our findings also showed that there was often a tension between illness management and maintaining relationships. For example, it could become difficult to implement strategies such as pacing when others were involved, particularly when the person with SS had not fully disclosed their symptoms or condition to the people whom their self-management strategy may affect. Therefore, saying "no" could also be hard for participants, particularly when it was perceived that others would not understand. Other participants had found a solution by working on their means of communicating their difficulties with those around them. Winger et al [85] have found that greater practice of assertiveness and communication skills was associated with reduced pain interference and psychological distress in people with lung cancer, and assertiveness and communication is also a key component of an effective fatigue management intervention for people with rheumatoid arthritis [60]. Therefore, we recommend including assertiveness and communication strategies within a therapeutic self-management app for SS. When considering the design of the app, we recommend including some text to help the user provide a brief explanation about their condition, its symptoms, and their impact to share with health professionals, colleagues, or people in social settings, as needed. We also recommend designing opportunities to practice assertiveness and communication skills within the app for those who may find it helpful.

In summary, our findings suggest that some of the key areas of concern for participants were potentially addressable through an intervention. A common starting point for any approach should be an acknowledgment of the real costs and the daily hard work of having an unpredictable, volatile, and multisymptomatic LTC. Any therapeutic approach needs to be designed to help with this labor; to acknowledge the social, emotional, and physical costs of having and managing SS; and to appreciate the wisdom that the "end user" of the app or intervention will have already accumulated. Strategies obtained from Acceptance and Commitment Therapy [86] and Compassion Focused Therapy [87] could be useful as they have been used to target the psychosocial impact of other related health complaints such as chronic pain [88]. Next, specific strategies (eg, pacing and sleep management) that could help target multiple symptoms or single symptoms in sequence would be useful. Finally, support to perform some of the social labor involved in living with SS should be a key component. In Multimedia Appendix 4, we have further specified the areas of intervention and suggested the broad therapeutic approaches that might be useful.

Regarding our use of the SDT framework, while it was useful to structure our thinking about intervention development, we also noted that the constructs of SDT often existed in a state of tension with each other, where successfully fulfilling the requirements of one construct leads to reduced functioning in another. As noted previously, this tension occurred between competence and relatedness, where symptom management conflicted with maintaining social bonds. Similar tension existed between autonomy and competence, where participants struggled to feel competent if presented with several self-management options. The SDT states that all 3 fundamental needs have to be met for internally motivated, self-determined behavior to occur [52], but we tentatively suggest that the theory needs to consider moments when some needs stand in opposition to each other. Making the nature of these tensions explicit to the users of an intervention or app would be a key part of its opening narrative.

Limitations

The extent of transferability of our findings to other LTCs is not yet known. However, studies of other autoimmune conditions have demonstrated the same need to self-manage complexity—people with inflammatory bowel disease reported that symptoms (pain, fatigue, and diarrhea) changed over time and could be interconnected at different times, and they required a highly individualized management strategy to "balance" the illness and attend to dynamic fluctuations in symptoms [71]. Overall, our findings may provide insight into how several other autoimmune conditions are self-managed or could be self-managed with the use of an app. However, owing to the nature of the complexity we captured in this study, transferability of our findings to other contexts may be limited. The 17 participants in this study included only 1 (6%) man, which may mean that any unique difficulties experienced by men with SS have been missed in our study. However, SS has a female-to-male incidence rate of 16:1 [89], and the gender makeup of our participants is representative of the wider SS population. Another limitation was that we did not formally collate information about smartphone ownership from participants. Such data should be collated in future similar studies. A final limitation is that most of this study was conducted within the United Kingdom, with only 67% (2/3) of the interviewees living outside the United Kingdom. Therefore, we cannot assume that similar findings would be replicated in other geographical contexts.

Future Studies

Future studies should operationalize the findings of this study to construct an intervention protocol that could be implemented via a smartphone app for the management of SS and empirically optimize its content through pilot and feasibility testing. Furthermore, future studies may explore the transferability of our findings to the self-management contexts of other autoimmune and fluctuating conditions. Our target users were those with primary or secondary SS; future studies should consider how user age influences the design requirements in this patient group.

Conclusions

In conclusion, therapeutic and design approaches for SS should be constructed in both bottom-up (ie, based on the self-management challenges that prospective users already experience) and top-down (according to the most effective treatments documented for SS) formats. For people with SS, choosing to involve an app in their self-management has the possibility of being counterproductive—by adding to their experience of fatigue and becoming overwhelmed. Therefore, the design of a self-management app for SS should support the user in performing the physical, cognitive, emotional, and social work of self-management and should be careful not to add to their already high self-management costs.

Acknowledgments

This study was funded by a grant from Versus Arthritis (grant 22026).

We would like to thank the participants for their time and contribution. In addition we would like to express our gratitude to Michelle Harrison and the North East Sjögren's Syndrome Association, Dr Elizabeth Price, Dr Helen Cartner and Julie Norris for their expert opinion and suggestions regarding the study and the app content.

Conflicts of Interest

None declared.

Multimedia Appendix 1
COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.
[PDF File (Adobe PDF File), 482 KB - humanfactors_v11i1e54172_app1.pdf]

Multimedia Appendix 2
Workshop topic guide.
[DOCX File , 24 KB - humanfactors_v11i1e54172_app2.docx]

Multimedia Appendix 3
Interview schedule.
[DOCX File , 14 KB - humanfactors_v11i1e54172_app3.docx]

Multimedia Appendix 4
Key findings with potential therapeutic and design recommendations mapped to the Self-Determination Theory domains.
[DOCX File , 23 KB - humanfactors_v11i1e54172_app4.docx]

References

1. Burt J, Roland M, Paddison C, Reeves D, Campbell J, Abel G, et al. Prevalence and benefits of care plans and care planning for people with long-term conditions in England. J Health Serv Res Policy 2012 Jan;17 Suppl 1:64-71. [doi: 10.1258/jhsrp.2011.010172] [Medline: 22315479]
2. Long term conditions compendium of information. Department of Health, United Kingdom Government. 2012 May 30. URL: https://assets.publishing.service.gov.uk/media/5a7c638340f0b62aff6c154e/dh_134486.pdf [accessed 2024-03-28]
3. Care and support for long term conditions. Nuffield Trust. URL: https://tinyurl.com/4h9s46dw [accessed 2023-03-27]
4. Anagnostopoulos I, Zinzaras E, Alexiou I, Papathanasiou AA, Davas E, Koutroumpas A, et al. The prevalence of rheumatic diseases in central Greece: a population survey. BMC Musculoskelet Disord 2010 May 26;11:98 [FREE Full text] [doi: 10.1186/1471-2474-11-98] [Medline: 20504294]
5. Monjardino T, Lucas R, Barros H. Frequency of rheumatic diseases in Portugal: a systematic review. Acta Reumatol Port 2011;36(4):336-363 [FREE Full text] [Medline: 22472925]
6. Chopra A, Abdel-Nasser A. Epidemiology of rheumatic musculoskeletal disorders in the developing world. Best Pract Res Clin Rheumatol 2008 Aug;22(4):583-604. [doi: 10.1016/j.berh.2008.07.001] [Medline: 18783739]
7. Vivino FB. Sjogren's syndrome: clinical aspects. Clin Immunol 2017 Sep;182:48-54. [doi: 10.1016/j.clim.2017.04.005] [Medline: 28428095]
8. Tarn J, Lendrem D, McMeekin P, Lendrem C, Hargreaves B, Ng WF. Primary Sjögren's syndrome: longitudinal real-world, observational data on health-related quality of life. J Intern Med 2022 Jun;291(6):849-855 [FREE Full text] [doi: 10.1111/joim.13451] [Medline: 35018685]
9. Kalouche-Khalil L, Gandhi R, Hao L, Smolkina E, Schumacher F. POS0319 disease burden of primary Sjogren syndrome: a retrospective United States claims database study. Ann Rheum Dis 2022 May 23;81(Suppl 1):410.1-410.1. [doi: 10.1136/annrheumdis-2022-eular.2548]
10. Gairy K, Ruark K, Sinclair SM, Brandwood H, Nelsen L. An innovative online qualitative study to explore the symptom experience of patients with primary Sjögren's syndrome. Rheumatol Ther 2020 Sep;7(3):601-615 [FREE Full text] [doi: 10.1007/s40744-020-00220-9] [Medline: 32725407]
11. Noaiseh G, Baer AN. Toward better outcomes in Sjögren's syndrome: the promise of a stratified medicine approach. Best Pract Res Clin Rheumatol 2020 Feb;34(1):101475. [doi: 10.1016/j.berh.2019.101475] [Medline: 32005417]
12. Brito-Zerón P, Baldini C, Bootsma H, Bowman SJ, Jonsson R, Mariette X, et al. Sjögren syndrome. Nat Rev Dis Primers 2016 Jul 07;2:16047. [doi: 10.1038/nrdp.2016.47] [Medline: 27383445]
13. Miyamoto ST, Lendrem DW, Ng WF, Hackett KL, Valim V. Managing fatigue in patients with primary Sjögren's syndrome: challenges and solutions. Open Access Rheumatol 2019;11:77-88 [FREE Full text] [doi: 10.2147/OARRR.S167990] [Medline: 31118841]
14. Giacomelli C, Talarico R, Baldini C, Bazzichi L. Pain in Sjögren's syndrome. Reumatismo 2014 Jun 06;66(1):39-43 [FREE Full text] [doi: 10.4081/reumatismo.2014.763] [Medline: 24938195]
15. Hackett KL, Gotts ZM, Ellis J, Deary V, Rapley T, Ng WF, et al. An investigation into the prevalence of sleep disturbances in primary Sjögren's syndrome: a systematic review of the literature. Rheumatology (Oxford) 2017 Apr 01;56(4):570-580 [FREE Full text] [doi: 10.1093/rheumatology/kew443] [Medline: 28013207]
16. Cui Y, Xia L, Li L, Zhao Q, Chen S, Gu Z. Anxiety and depression in primary Sjögren's syndrome: a cross-sectional study. BMC Psychiatry 2018 May 16;18(1):131 [FREE Full text] [doi: 10.1186/s12888-018-1715-x] [Medline: 29769121]
17. Hackett KL, Deary V, Deane KH, Newton JL, Ng WF, Rapley T. Experience of sleep disruption in primary Sjögren's syndrome: a focus group study. Br J Occup Ther 2018 Apr;81(4):218-226 [FREE Full text] [doi: 10.1177/0308022617745006] [Medline: 29657352]
18. Ngo DY, Thomson WM, Nolan A, Ferguson S. The lived experience of Sjögren's Syndrome. BMC Oral Health 2016 Feb 01;16:7 [FREE Full text] [doi: 10.1186/s12903-016-0165-4] [Medline: 26831141]

19. Stack RJ, Southworth S, Fisher BA, Barone F, Buckley CD, Rauz S, et al. A qualitative exploration of physical, mental and ocular fatigue in patients with primary Sjögren's syndrome. PLoS One 2017;12(10):e0187272 [FREE Full text] [doi: 10.1371/journal.pone.0187272] [Medline: 29088284]
20. Chen X, Wu H, Wei W. Advances in the diagnosis and treatment of Sjogren's syndrome. Clin Rheumatol 2018 Jul;37(7):1743-1749. [doi: 10.1007/s10067-018-4153-8] [Medline: 29804149]
21. Price EJ, Rauz S, Tappuni AR, Sutcliffe N, Hackett KL, Barone F, et al. The British Society for Rheumatology guideline for the management of adults with primary Sjögren's syndrome. Rheumatology (Oxford) 2017 Oct 01;56(10):1828. [doi: 10.1093/rheumatology/kex375] [Medline: 28957572]
22. Bowman SJ, Everett CC, O'Dwyer JL, Emery P, Pitzalis C, Ng WF, et al. Randomized controlled trial of rituximab and cost-effectiveness analysis in treating fatigue and oral dryness in primary Sjögren's syndrome. Arthritis Rheumatol 2017 Jul;69(7):1440-1450 [FREE Full text] [doi: 10.1002/art.40093] [Medline: 28296257]
23. Gottenberg JE, Ravaud P, Puéchal X, Le Guern V, Sibilia J, Goeb V, et al. Effects of hydroxychloroquine on symptomatic improvement in primary Sjögren syndrome: the JOQUER randomized clinical trial. JAMA 2014 Jul 16;312(3):249-258. [doi: 10.1001/jama.2014.7682] [Medline: 25027140]
24. Hackett KL, Deane KH, Strassheim V, Deary V, Rapley T, Newton JL, et al. A systematic review of non-pharmacological interventions for primary Sjögren's syndrome. Rheumatology (Oxford) 2015 Nov;54(11):2025-2032 [FREE Full text] [doi: 10.1093/rheumatology/kev227] [Medline: 26135587]
25. Hackett KL, Deane KH, Newton JL, Deary V, Bowman SJ, Rapley T, et al. Mixed-methods study identifying key intervention targets to improve participation in daily living activities in primary Sjögren's syndrome patients. Arthritis Care Res (Hoboken) 2018 Jul;70(7):1064-1073 [FREE Full text] [doi: 10.1002/acr.23536] [Medline: 29409110]
26. Barlow J. How to use education as an intervention in osteoarthritis. Best Pract Res Clin Rheumatol 2001 Oct;15(4):545-558. [doi: 10.1053/berh.2001.0172] [Medline: 11567538]
27. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48(2):177-187. [doi: 10.1016/s0738-3991(02)00032-0] [Medline: 12401421]
28. Lorig KR, Holman HR. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [doi: 10.1207/s15324796abm2601_01]
29. Hackett KL. Developing a non-pharmacological intervention model to improve function and participation in people with primary Sjo¨gren's syndrome. Newcastle University. 2017. URL: https://theses.ncl.ac.uk/jspui/handle/10443/3802 [accessed 2024-03-28]
30. McQueenie R, Jani BD, Siebert S, McLoone P, McCowan C, Macdonald S, et al. Prevalence of chronic pain in LTCs and multimorbidity: a cross-sectional study using UK Biobank. J Multimorb Comorb 2021;11:26335565211005870 [FREE Full text] [doi: 10.1177/26335565211005870] [Medline: 35004337]
31. Morton K, Dennison L, May C, Murray E, Little P, McManus RJ, et al. Using digital interventions for self-management of chronic physical health conditions: a meta-ethnography review of published studies. Patient Educ Couns 2017 Apr;100(4):616-635 [FREE Full text] [doi: 10.1016/j.pec.2016.10.019] [Medline: 28029572]
32. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016 May 16;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
33. Wu X, Guo X, Zhang Z. The efficacy of mobile phone apps for lifestyle modification in diabetes: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Jan 15;7(1):e12297 [FREE Full text] [doi: 10.2196/12297] [Medline: 30664494]
34. Morrison D, Mair FS, Yardley L, Kirby S, Thomas M. Living with asthma and chronic obstructive airways disease: using technology to support self-management - an overview. Chron Respir Dis 2017 Nov;14(4):407-419 [FREE Full text] [doi: 10.1177/1479972316660977] [Medline: 27512084]
35. Duan H, Wang Z, Ji Y, Ma L, Liu F, Chi M, et al. Using goal-directed design to create a mobile health app to improve patient compliance with hypertension self-management: development and deployment. JMIR Mhealth Uhealth 2020 Feb 25;8(2):e14466 [FREE Full text] [doi: 10.2196/14466] [Medline: 32130161]
36. Wulfovich S, Fiordelli M, Rivas H, Concepcion W, Wac K. "I must try harder": design implications for mobile apps and wearables contributing to self-efficacy of patients with chronic conditions. Front Psychol 2019 Oct 23;10:2388 [FREE Full text] [doi: 10.3389/fpsyg.2019.02588] [Medline: 31749733]
37. Osborne CL, Juengst SB, Smith EE. Identifying user-centered content, design, and features for mobile health apps to support long-term assessment, behavioral intervention, and transitions of care in neurological rehabilitation: An exploratory study. Brit J Occup Ther 2020 Oct 07;84(2):101-110. [doi: 10.1177/0308022620954115]
38. Druce KL, Dixon WG, McBeth J. Maximizing engagement in mobile health studies: lessons learned and future directions. Rheum Dis Clin North Am 2019 May;45(2):159-172 [FREE Full text] [doi: 10.1016/j.rdc.2019.01.004] [Medline: 30952390]
39. Maharaj A, Lim D, Murphy R, Serlachius A. Comparing two commercially available diabetes apps to explore challenges in user engagement: randomized controlled feasibility study. JMIR Form Res 2021 Jun 16;5(6):e25151 [FREE Full text] [doi: 10.2196/25151] [Medline: 34132640]

40. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
41. Morrison LG. Theory-based strategies for enhancing the impact and usage of digital health behaviour change interventions: a review. Digit Health 2015;1:2055207615595335 [FREE Full text] [doi: 10.1177/2055207615595335] [Medline: 29942544]
42. Peters D, Davis S, Calvo RA, Sawyer SM, Smith L, Foster JM. Young people's preferences for an asthma self-management app highlight psychological needs: a participatory study. J Med Internet Res 2017 Apr 11;19(4):e113 [FREE Full text] [doi: 10.2196/jmir.6994] [Medline: 28400353]
43. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
44. Rogers Y, Marshall P. Research in the wild. In: Synthesis Lectures on Human-Centered Informatics. San Rafael, CA: Morgan & Claypool Publishers; Apr 04, 2017:i-97.
45. Morris ME. Preventing falls in older people. BMJ 2012 Aug 07;345:e4919. [doi: 10.1136/bmj.e4919] [Medline: 22872696]
46. van Middelaar T, Beishuizen CR, Guillemont J, Barbera M, Richard E, Moll van Charante EP. Engaging older people in an internet platform for cardiovascular risk self-management: a qualitative study among Dutch HATICE participants. BMJ Open 2018 Jan 21;8(1):e019683 [FREE Full text] [doi: 10.1136/bmjopen-2017-019683] [Medline: 29358447]
47. Schermer M. Telecare and self-management: opportunity to change the paradigm? J Med Ethics 2009 Nov;35(11):688-691. [doi: 10.1136/jme.2009.030973] [Medline: 19880706]
48. Lackner A, Ficjan A, Stradner MH, Hermann J, Unger J, Stamm T, et al. It's more than dryness and fatigue: the patient perspective on health-related quality of life in primary Sjögren's Syndrome - a qualitative study. PLoS One 2017;12(2):e0172056 [FREE Full text] [doi: 10.1371/journal.pone.0172056] [Medline: 28182787]
49. Sanz MF, Acha BV, García MF. Co-design for people-centred care digital solutions: a literature review. Int J Integr Care 2021 Apr 30;21(2):16 [FREE Full text] [doi: 10.5334/ijic.5573] [Medline: 33981193]
50. Grenier C, Ibrahim R, Paixão-Barradas S. Designing and innovating in digital healthcare: co-design for taking patients' needs into account. In: Béranger J, Rizoulières R, editors. The Digital Revolution in Health. Hoboken, NJ: John Wiley & Sons; 2021.
51. Harrington CN, Wilcox L, Connelly K, Rogers W, Sanford J. Designing health and fitness apps with older adults: examining the value of experience-based co-design. In: Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2018 Presented at: PervasiveHealth '18; May 21-24, 2018; New York, NY. [doi: 10.1145/3240925.3240929]
52. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
53. Gillison FB, Rouse P, Standage M, Sebire SJ, Ryan RM. A meta-analysis of techniques to promote motivation for health behaviour change from a self-determination theory perspective. Health Psychol Rev 2019 Mar;13(1):110-130. [doi: 10.1080/17437199.2018.1534071] [Medline: 30295176]
54. Peters D, Calvo RA, Ryan RM. Designing for motivation, engagement and wellbeing in digital experience. Front Psychol 2018;9:797 [FREE Full text] [doi: 10.3389/fpsyg.2018.00797] [Medline: 29892246]
55. Eassey D, Reddel HK, Ryan K, Smith L. 'It is like learning how to live all over again' a systematic review of people's experiences of living with a chronic illness from a self-determination theory perspective. Health Psychol Behav Med 2020 Jul 23;8(1):270-291 [FREE Full text] [doi: 10.1080/21642850.2020.1794879] [Medline: 34040872]
56. Stenberg N, Gillison F, Rodham K. How do peer support interventions for the self-management of chronic pain, support basic psychological needs? a systematic review and framework synthesis using self-determination theory. Patient Educ Couns 2022 Nov;105(11):3225-3234 [FREE Full text] [doi: 10.1016/j.pec.2022.07.017] [Medline: 35985906]
57. Hekler EB, Klasnja P, Riley WT, Buman MP, Huberty J, Rivera DE, et al. Agile science: creating useful products for behavior change in the real world. Transl Behav Med 2016 Jun;6(2):317-328 [FREE Full text] [doi: 10.1007/s13142-016-0395-7] [Medline: 27357001]
58. Tong A, Sainsbury P, Craig J. Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
59. Andersen K, Wakkary R. The magic machine workshops: making personal design knowledge. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19; May 4-9, 2019; Glasgow, UK. [doi: 10.1145/3290605.3300342]
60. Hewlett S, Almeida C, Ambler N, Blair PS, Choy EH, Dures E, RAFT Study Group. Reducing arthritis fatigue impact: two-year randomised controlled trial of cognitive behavioural approaches by rheumatology teams (RAFT). Ann Rheum Dis 2019 Apr;78(4):465-472 [FREE Full text] [doi: 10.1136/annrheumdis-2018-214469] [Medline: 30793700]
61. Hackett KL, Lambson RL, Strassheim V, Gotts Z, Deary V, Newton JL. A concept mapping study evaluating the UK's first NHS generic fatigue clinic. Health Expect 2016 Oct;19(5):1138-1149 [FREE Full text] [doi: 10.1111/hex.12405] [Medline: 26332418]

62. Ericsson KA, Simon HA. Verbal reports as data. Psychol Rev 1980 May;87(3):215-251. [doi: 10.1037/0033-295X.87.3.215]
63. Proudfoot K. Inductive/deductive hybrid thematic analysis in mixed methods research. J Mixed Method Res 2022 Sep 20;17(3):308-326. [doi: 10.1177/15586898221126816]
64. Braun V, Clarke V. Successful Qualitative Research: A Practical Guide for Beginners. Thousand Oaks, CA: SAGE Publications; Mar 22, 2013.
65. Deci EL, Ryan RM. Self-determination theory: a macrotheory of human motivation, development, and health. Can Psychol 2008 Aug;49(3):182-185. [doi: 10.1037/a0012801]
66. Ryan RM, Deci EL. Chapter four - brick by brick: the origins, development, and future of self-determination theory. In: Elliot AJ, editor. Advances in Motivation Science. Cambridge, MA: Academic Press; 2019.
67. Ntoumanis N, Ng JY, Prestwich A, Quested E, Hancox JE, Thøgersen-Ntoumani C, et al. A meta-analysis of self-determination theory-informed intervention studies in the health domain: effects on motivation, health behavior, physical, and psychological health. Health Psychol Rev 2021 Jun;15(2):214-244 [FREE Full text] [doi: 10.1080/17437199.2020.1718529] [Medline: 31983293]
68. Cartner H. 'A Taste of my Life'. Altered Eating and Sjogren's Syndrome: A Participant-led Exploration of Eating Experience Using Application of a Novel Methodology, 'Food Play', in Department of Psychology 2024;Northumbria University:1-371 (forthcoming).
69. Parker HW, Abreu AM, Sullivan MC, Vadiveloo MK. Allostatic load and mortality: a systematic review and meta-analysis. Am J Prev Med 2022 Jul;63(1):131-140. [doi: 10.1016/j.amepre.2022.02.003] [Medline: 35393143]
70. Eilertsen G, Ormstad H, Kirkevold M, Mengshoel AM, Söderberg S, Olsson M. Similarities and differences in the experience of fatigue among people living with fibromyalgia, multiple sclerosis, ankylosing spondylitis and stroke. J Clin Nurs 2015 Jul;24(13-14):2023-2034. [doi: 10.1111/jocn.12774] [Medline: 25661994]
71. Dibley L, Khoshaba B, Artom M, Van Loo V, Sweeney L, Syred J, et al. Patient strategies for managing the vicious cycle of fatigue, pain and urgency in inflammatory bowel disease: impact, planning and support. Dig Dis Sci 2021 Oct;66(10):3330-3342. [doi: 10.1007/s10620-020-06698-1] [Medline: 33164146]
72. Andrews NE, Strong J, Meredith PJ, Branjerdporn GS. Approach to activity engagement and differences in activity participation in chronic pain: a five-day observational study. Aust Occup Ther J 2018 Dec;65(6):575-585. [doi: 10.1111/1440-1630.12516] [Medline: 30178871]
73. Carney CE, Buysse DJ, Ancoli-Israel S, Edinger JD, Krystal AD, Lichstein KL, et al. The consensus sleep diary: standardizing prospective sleep self-monitoring. Sleep 2012 Feb 01;35(2):287-302 [FREE Full text] [doi: 10.5665/sleep.1642] [Medline: 22294820]
74. Thomas S, Thomas PW, Nock A, Slingsby V, Galvin K, Baker R, et al. Development and preliminary evaluation of a cognitive behavioural approach to fatigue management in people with multiple sclerosis. Patient Educ Couns 2010 Feb;78(2):240-249. [doi: 10.1016/j.pec.2009.07.001] [Medline: 19665337]
75. Antcliff D, Keeley P, Campbell M, Oldham J, Woby S. The development of an activity pacing questionnaire for chronic pain and/or fatigue: a Delphi technique. Physiotherapy 2013 Sep;99(3):241-246. [doi: 10.1016/j.physio.2012.12.003] [Medline: 23395263]
76. McCrae CS, Williams J, Roditi D, Anderson R, Mundt JM, Miller MB, et al. Cognitive behavioral treatments for insomnia and pain in adults with comorbid chronic insomnia and fibromyalgia: clinical outcomes from the SPIN randomized controlled trial. Sleep 2019 Mar 01;42(3):1-15 [FREE Full text] [doi: 10.1093/sleep/zsy234] [Medline: 30496533]
77. Mair FS, May CR. Thinking about the burden of treatment. BMJ 2014 Nov 10;349:g6680. [doi: 10.1136/bmj.g6680] [Medline: 25385748]
78. Güldenpfennig F, Mayer P, Panek P, Fitzpatrick G. An autonomy-perspective on the design of assistive technology experiences of people with multiple sclerosis. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19; May 4-9, 2019; Glasgow, UK. [doi: 10.1145/3290605.3300357]
79. Seah ML, Koh KT. The efficacy of using mobile applications in changing adolescent girls' physical activity behaviour during weekends. Eur Phys Educ Rev 2020 Jun 17;27(1):113-131. [doi: 10.1177/1356336x20930741]
80. Marathe S, Shyam Sundar S. What drives customization?: control or identity? In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011 Presented at: CHI '11; May 7-12, 2011; Vancouver, BC. [doi: 10.1145/1978942.1979056]
81. Harrison D, Marshall P, Bianchi-Berthouze N, Bird J. Activity tracking: barriers, workarounds and customisation. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Presented at: UbiComp '15; September 7-11, 2015; Osaka, Japan. [doi: 10.1145/2750858.2805832]
82. Zhang R, Ringland KE, Paan M, Mohr DC, Reddy M. Designing for emotional well-being: integrating persuasion and customization into mental health technologies. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021 Presented at: CHI '21; May 8-13, 2021; Yokohama, Japan. [doi: 10.1145/3411764.3445771]
83. Carswell C, Brown JV, Lister J, Ajjan RA, Alderson SL, Balogun-Katung A, et al. The lived experience of severe mental illness and long-term conditions: a qualitative exploration of service user, carer, and healthcare professional perspectives on self-managing co-existing mental and physical conditions. BMC Psychiatry 2022 Jul 19;22(1):479 [FREE Full text] [doi: 10.1186/s12888-022-04117-5] [Medline: 35850709]

84. Howell D, Mayer DK, Fielding R, Eicher M, Verdonck-de Leeuw IM, Johansen C, et al. Management of cancer and health after the clinic visit: a call to action for self-management in cancer care. J Natl Cancer Inst 2021 May 04;113(5):523-531 [FREE Full text] [doi: 10.1093/jnci/djaa083] [Medline: 32525530]
85. Winger JG, Rand KL, Hanna N, Jalal SI, Einhorn LH, Birdas TJ, et al. Coping skills practice and symptom change: a secondary analysis of a pilot telephone symptom management intervention for lung cancer patients and their family caregivers. J Pain Sympt Manag 2018 May;55(5):1341-9.e4 [FREE Full text] [doi: 10.1016/j.jpainsymman.2018.01.005] [Medline: 29366911]
86. Hayes SC, Lillis J. Acceptance and Commitment Therapy. Washington, DC: American Psychological Association; 2012.
87. Gilbert P. Introducing compassion-focused therapy. Adv Psychiatr Treat 2018 Jan 02;15(3):199-208. [doi: 10.1192/apt.bp.107.005264]
88. McCracken LM, Yu L, Vowles KE. New generation psychological treatments in chronic pain. BMJ 2022 Feb 28;376:e057212. [doi: 10.1136/bmj-2021-057212] [Medline: 35228207]
89. Brandt JE, Priori R, Valesini G, Fairweather D. Sex differences in Sjögren's syndrome: a comprehensive review of immune mechanisms. Biol Sex Differ 2015;6:19 [FREE Full text] [doi: 10.1186/s13293-015-0037-7] [Medline: 26535108]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
LTC: long-term condition
SDT: Self-Determination Theory
SS: Sjögren's syndrome

Edited by C Jacob; submitted 01.11.23; peer-reviewed by S Stones, A Chen, R Grainger; comments to author 26.12.23; revised version received 26.01.24; accepted 31.01.24; published 17.04.24.

Please cite as:
McCallum C, Campbell M, Vines J, Rapley T, Ellis J, Deary V, Hackett K
A Smartphone App to Support Self-Management for People Living With Sjögren's Syndrome: Qualitative Co-Design Workshops
JMIR Hum Factors 2024;11:e54172
URL: https://humanfactors.jmir.org/2024/1/e54172
doi:10.2196/54172
PMID:38630530

Original Paper

Effective Communication Supported by an App for Pregnant Women: Quantitative Longitudinal Study

Lukas Kötting[1], MSc; Vinayak Anand-Kumar[1], MSc; Franziska Maria Keller[2], MSc, PhD; Nils Tobias Henschel[1], MSc; Sonia Lippke[1], Dr

[1]Psychology and Methods, School of Business, Social & Decision Sciences, Constructor University Bremen gGmbH, Bremen, Germany
[2]Klinikum Bremerhaven Reinkenheide gGmbH, Bremerhaven, Germany

Corresponding Author:
Sonia Lippke, Dr
Psychology and Methods
School of Business, Social & Decision Sciences
Constructor University Bremen gGmbH
Campus Ring 1
Bremen, 28759
Germany
Phone: 49 421 200 4730
Email: s.lippke@jacobs-university.de

Abstract

Background: In the medical field of obstetrics, communication plays a crucial role, and pregnant women, in particular, can benefit from interventions improving their self-reported communication behavior. Effective communication behavior can be understood as the correct transmission of information without misunderstanding, confusion, or losses. Although effective communication can be trained by patient education, there is limited research testing this systematically with an app-based digital intervention. Thus, little is known about the success of such a digital intervention in the form of a web-app, potential behavioral barriers for engagement, as well as the processes by which such a web-app might improve self-reported communication behavior.

Objective: This study fills this research gap by applying a web-app aiming at improving pregnant women's communication behavior in clinical care. The goals of this study were to (1) uncover the potential risk factors for early dropout from the web-app and (2) investigate the social-cognitive factors that predict self-reported communication behavior after having used the web-app.

Methods: In this study, 1187 pregnant women were recruited. They all started to use a theory-based web-app focusing on intention, planning, self-efficacy, and outcome expectancy to improve communication behavior. Mechanisms of behavior change as a result of exposure to the web-app were explored using stepwise regression and path analysis. Moreover, determinants of dropout were tested using logistic regression.

Results: We found that dropout was associated with younger age (P=.014). Mechanisms of behavior change were consistent with the predictions of the health action process approach. The stepwise regression analysis revealed that action planning was the best predictor for successful behavioral change over the course of the app-based digital intervention (β=.331; P<.001). The path analyses proved that self-efficacy beliefs affected the intention to communicate effectively, which in turn, elicited action planning and thereby improved communication behavior (β=.017; comparative fit index=0.994; Tucker–Lewis index=0.971; root mean square error of approximation=0.055).

Conclusions: Our findings can guide the development and improvement of apps addressing communication behavior in the following ways in obstetric care. First, such tools would enable action planning to improve communication behavior, as action planning is the key predictor of behavior change. Second, younger women need more attention to keep them from dropping out. However, future research should build upon the gained insights by conducting similar internet interventions in related fields of clinical care. The focus should be on processes of behavior change and strategies to minimize dropout rates, as well as replicating the findings with patient safety measures.

Trial Registration: ClinicalTrials.gov identifier: NCT03855735; https://classic.clinicaltrials.gov/ct2/show/NCT03855735

(JMIR Hum Factors 2024;11:e48218) doi:10.2196/48218

KEYWORDS

clinical care; health action process approach; HAPA; intention; communication behavior; patient safety; patient education; internet intervention; dropout; digital health; behavior change; prediction; obstetric; pregnant women; pregnancy; safe communication; health behaviors; obstetric care

Introduction

Background

In the dynamic landscape of medical internet research, the pursuit of effective interventions and preventive health programs demands a comprehensive understanding of diverse populations, including pregnant women and their unique needs. This paper unveils the outcomes of formative research and preliminary results within the realm of medical and preventive health, exploring an innovation and technology in terms of a digital intervention, that is, apps aiming at improving communication behavior. Formative research, characterized by its emphasis on gathering insights from the intended beneficiaries, emerges as a fundamental tool for tailoring interventions to meet the unique requirements of diverse communities [1]. By demonstrating the integral role of formative research in the early stages of program development, we aim to provide a compelling case for its incorporation in the toolkit of researchers and experts working in the field of medical internet research. In this paper, we outline the potential of using digital tools like apps for improving communication behavior for patient safety and the risks involved with regard to dropouts in app-based interventions. Lastly, we outline a behavior change theory to model communication behavior, which may help to map out the health-related behavior more systematically and which was used for our research investigating communication behavior and app usage.

Patient Safety in Health Care

Patient safety is defined as the absence of harm that could have been prevented in patients. For achieving patient safety, health care should be delivered in an optimal manner, trust should be built among all involved individuals, and misunderstanding, information loss, or error occurrence should be prevented [2]. Thus, patient safety requires effective communication behaviors among health care professionals, patients, their partners, or accompanying persons [2]. In particular, in obstetric care, this holds true [2,3] because women in labor have to express their needs and wishes even in the face of stress and barriers to ensure their active role in the obstetric process. Communication behavior can be measured and taught [3-5] and is a reliable approach for improving patient safety [6,7]. Communication behavior involves multiple individuals, including patients, health care workers, and partners [8-13]. This encompasses not only the importance of perceiving a supportive environment that guarantees an open exchange of concerns and potential solutions but also the individual's competency to communicate safely. Such competency consists of the self-reported communication skills that are based on Rider and Keefer's [14] competencies and are impacted by determinants of the communication behavior, that is, self-efficacy, intention formation, and planning [15-19].

Communication Behavior

Communication is defined as a process involving the exchange of cognitions and emotions through verbal and nonverbal actions [20,21]. In this work, we define patients' communication similarly to communication in health care workers to keep definitions for both groups aligned. Previous work [8,10] performed over the scope of this project has defined communication in line with Rider and Keefer's definition [14]. They describe a set of skills including the creation and sustainability of a therapeutic relationship, use of effective listening, prompt and effective responding, and effective communication [14]. Effective communication is the correct transmission of information without misunderstanding, confusion, or losses.

Although effective communication has been shown to be of importance in preventing errors in medical care as well as in patient-provider relationships [6,22-24], only few studies have investigated effective communication behavior among those receiving obstetric care. Moreover, there is limited evidence for innovative tools aiming at increasing effective communication among pregnant women and their support networks [8,11,25]. Previous research has mainly investigated face-to-face interventions in clinical care or hospital settings [26,27]. Although traditional face-to-face interventions demonstrate efficacy, they tend to show several disadvantages concerning feasibility, such as higher financial constraints, limited utilization due to mobility constraints, or scheduling and time issues [28-31]. These constraints of traditional face-to-face interventions also call for cost-effective, convenient, instantly available, and scalable alternative solutions. One of these alternatives, successfully implemented across multiple therapeutic areas, including the promotion of health behavior change, is support via the internet, digital interventions, and apps in the medical field [32-35].

Digital Interventions and Apps

Digital interventions and apps (also called as medical internet support or web-based communication training) have shown several advantages over traditional face-to-face interventions, such as increased ease of accessibility and personalized interactions with real-time feedback. Furthermore, they offer the opportunity for scalability to larger populations, including individuals who live in remote areas. Moreover, such digital interventions can be relatively cost-effective compared to traditional formats [36,37]. Although there is clear general evidence regarding digital interventions, there is scarcity of research on those targeting to foster effective communication.

The same holds true regarding the applicability and integrability of traditional health behavior change theories such as the health action process approach (HAPA) to explain health behavior changes in digital interventions (ie, smartphone apps). Indeed, literature shows how interventions supporting motivational and volitional processes prove effective [8,9,38]. However, the

HAPA model has been rarely applied to interventions targeting effective communication [8,9], and it is hardly ever used to explain communication behavior in the context of digital interventions or their dropout of pregnant women. Therefore, we review dropout in more detail.

Factors Associated With App Usage and Dropout From Digital Interventions

Early dropout from digital interventions is a key problem [39], as the intervention use is discontinued. This needs more attention because if users drop out, which might occur as often as 1 in 2 cases [40], efficacy is limited, and the reach and generalizability of the obtained results are diminished [39]. However, little is known about the factors associated with dropout [39]. Accordingly, more research investigating and identifying such factors is needed, especially in the context of communication behavior and giving birth.

Looking at the general literature on the potential risk factors for early dropout in digital interventions, the following sociodemographic and behavior change factors were identified: age, education, and social support [41,42]. Although there is no previous study on dropout from digital interventions addressing effective communication of pregnant women, evidence from other areas with digital interventions exist. For example, Wu and colleagues [43] investigated dropout in a blended care cognitive behavioral intervention. They highlighted that a higher dropout rate was associated specifically with female gender, poorer financial status, and the absence of a college degree. Additionally, Gao et al [44] found that younger patients and those who were less educated were more likely to drop out from digital intervention studies. Other factors associated with early dropout were marital status (higher probability of divorced individuals to drop out) and ethnicity [45]. Besides the sociodemographic factors, according to Davis and Addis [46], psychological determinants should also be considered while examining dropouts from digital health interventions. According to [47,48], users with low intention to change their behavior have been found to drop out more often from digital interventions. A study by Schroé and colleagues [49] further investigated why users discontinued the use of digital health interventions. Their results highlighted that whereas sociodemographic factors were predictive of early dropout, psychological determinants such as action planning and self-monitoring were associated with completion of digital interventions [49]. This is in line with other research highlighting that self-monitoring [50] and higher intrinsic motivation were associated with lower attrition rates [51]. A theory that could bring the different factors together to enable systematic research is the aforementioned HAPA model, which is described in more detail below.

HAPA Model to Understand and Improve Behavior Change

Self-reported communication behavior constitutes a preventive health behavior [25] and may be fostered by the same factors and processes that health psychology literature has repeatedly showcased [52-54]. HAPA proved to be a useful theory [11] essentially since it considers the interplay of resources, barriers, as well as the well-known behavior intention gap [54,55]. The HAPA model is divided into 2 distinct phases: (1) the motivational phase in which individuals consider their competencies' determinants such as self-efficacy, expectations about behavioral outcomes (outcome expectancies) and formulate a behavioral intention (eg, to communicate in the birthing context), and (2) the volitional phase, wherein pregnant women develop and enact behavioral plans in order to bring the intentions to behavioral actions. This whole process is shaped by social-cognitive barriers and facilitators that may originate externally or stem from women's personal belief, which is also called self-efficacy [54,56]. According to the HAPA model, individuals need to first form an intention, which is based on outcome expectancies and self-efficacy, before acting accordingly. Hence, the pathway of intention on the actual behavior is mediated by action planning [8,11,57]—with action planning being more proximal to behavior, and intention, outcome expectancies, and self-efficacy being more distal to behavior.

In order to improve communication behavior, interventions must be tailored to social-cognitive barriers and facilitators of the target population. Previous evidence has demonstrated that classical face-to-face interventions based on motivational and volitional theories such as HAPA are effective in improving self-reported communication [8,9,38]. It should be noted that most of these findings stem from interventions that were solely offered to health care workers [25], but more attention needs to be paid to patient education. This is the basis of our study with pregnant women randomized into an intervention group or a waitlist control group.

Goal of This Study

As previously outlined, there is a need for further studies to investigate effective communication behaviors of pregnant women within the context of a digital intervention. The goals of our study were 2-fold. First, we aimed to uncover the potential risk factors for early dropout from a digital intervention. Second, we aimed to investigate the social-cognitive factors that would predict the self-reported communication behavior after having used the digital intervention. Thus, the hypotheses are as follows:

1. Hypothesis 1: Sociodemographic factors play a larger role in predicting dropout during a digital intervention relative to behavior change variables.
2. Hypothesis 2: The social-cognitive factors outlined by HAPA (self-efficacy, outcome expectancy, and action planning) predict self-reported communication behavior in pregnant women over the course of the app-based intervention.
3. Hypothesis 3: More distal HAPA variables (intention, outcome expectancies, and self-efficacy) indirectly relate with self-reported communication behavior mediated by action planning.

Methods

TeamBaby Project

This study stems from a larger project named TeamBaby, which was tasked with developing interventions to improve communication behavior between those who receive and provide obstetric care. One of the interventions was a digital intervention, that is, an app (actually, a web-app). Data collected from the TeamBaby web-app were used to investigate our hypotheses. The TeamBaby project was funded by the German Innovation Fund (project 01VSF18023) of the Gemeinsamer Bundesausschuss and preregistered (ClinicalTrials.gov identifier: NCT03855735) on February 27, 2019.

Recruitment and Procedures

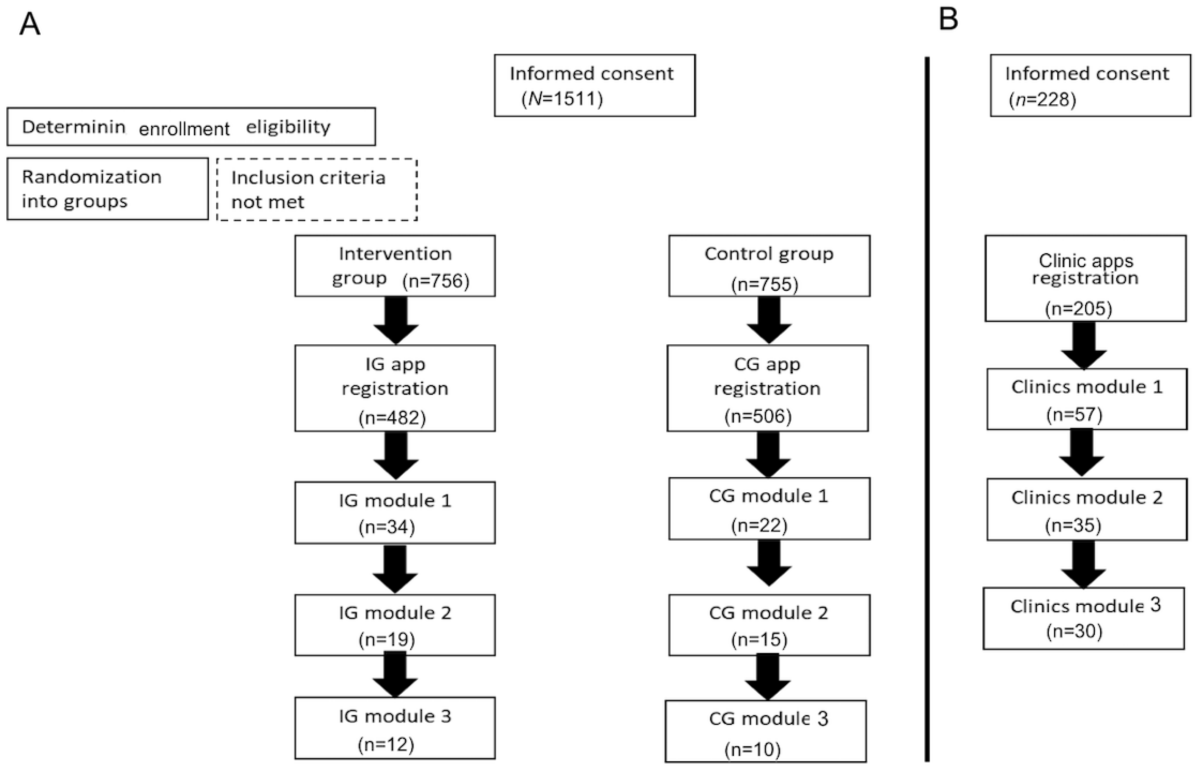
Ethics Approval

Ethics approval for data collection in the maternity clinics was granted by the ethics committee for human research of the University Hospital Ulm (114/19) and the ethics committee for medical research of the University Hospital Frankfurt am Main (19-292). Informed consent was provided in the registration process, and all data were anonymized by providing users with a random ID that could not be linked to user emails or personal IDs. No compensation was provided for participation in this study.

Participants

Participants were recruited into this study, as outlined in Figure 1. Participants represented a pragmatic sample. Sample calculations were performed prior to data collection for an assumed dropout of 20%. We estimated that 176 or more individuals would be needed to recruit [9]. All recruited women were able to register to use the TeamBaby web-app if they were residing in Germany, either during the time of our study (if randomized into the intervention group) or 2 weeks later (if randomized into the waitlist-control arm), that is, pregnant women and their support persons.

Figure 1. Flowchart of the study participants. (A) Study participants using the app as intervention group versus being randomized to the control group and using the app only 2 weeks later. (B) Study flow for the clinic's intervention group. CG: control group; IG: intervention group.



Pregnant women were recruited through 1 of the 2 recruitment channels (Figure 1): (1) those who sought treatment during pregnancy in project-affiliated clinics or (2) those who were currently pregnant and based anywhere in Germany. With respect to the former group, project members worked together with obstetricians to recruit patients. The Germany-wide recruitment utilized social media and targeted advertising to promote this study. In addition, flyers were placed in health care clinics and pharmacies across the country. Women were eligible to participate if they were currently pregnant, had sufficient knowledge of German, and were at least 18 years old.

Participants were recruited into this study upon completing a web-based questionnaire. Women recruited in the clinics were invited to register with the web-app. Women from the Germany-wide recruitment were randomized into either a treatment group or waitlist-control group. The treatment group was presented with a link to use the web-app directly. The waitlist-control arm was provided with a link to the web-app 2 weeks later. All users of the web-app were presented with a series of questions at regular timepoints to determine whether there were changes in HAPA variables and communication behavior as users progressed through the 3 modules of the web-app. Participants who completed less than 2 modules were considered as early dropouts.

Intervention Content

The TeamBaby web-app provided guidance on how to work effectively with health care workers. The web-app consisted of 10 lessons, wrapped in 3 modules, which were developed and structured based on the processes of behavior change as set out by the HAPA model. The first set of lessons was designed to increase the outcome expectations of effective communication behavior and create an intention to adopt self-reported communication practices. The subsequent lessons were designed to increase the belief or trust in oneself to employ effective communication and enable users to make tangible plans for implementing self-reported communication behavior. More detailed information about the modules and the content can be found in Multimedia Appendix 1.

Measures

Participants were asked to complete questions relating to communication behavior and HAPA variables at 4 timepoints: before starting the first module and after completing each subsequent module. The assessment of self-reported communication behavior was based on Rider and Keefer's competencies [14] and adapted to address pregnant women's behavior in previous publications [12]. As the aim of the research was to understand the underlying social-cognitive processes of communication, the items were developed to capture self-reported communication behavior. Table 1 presents the items [14,58,59] used to evaluate each variable.

Table 1. Measured health action process approach and self-reported communication variables.

Variables	Item example	Range[a]
Communication behavior (7 items) [14]	During pregnancy, I always have communicated my needs clearly.	1-6
Outcome expectancy (single item) [58]	If I communicate well with doctors and midwives, my preferences can be considered during childbirth.	1-6
Intention (single item) [59]	I intend to always make sure that I communicate effectively with the doctors and midwives.	1-6
Self-efficacy (single item) [59]	I am sure I can communicate well even when I am tired or exhausted.	1-6
Action planning (single item) [58]	I have planned precisely how to communicate well while giving birth.	1-6

[a]Range of 1-6 spans from "does not apply at all" to "does apply fully."

Sociodemographic Data

In addition to behavior change measures, demographic information was collected. Participants reported their age, marital status, highest level of education, and nationality.

Aggregated Variables

Participant communication scores were combined into a single item for each individual by taking the average across the different communication behaviors. This was expressed as overall communication. The implicit assumption here is that more effective communication within the described obstetric setting should facilitate a safer birth.

Statistical Analysis

All analyses were conducted using the R [60] and RStudio [61] software. Significance was determined at the 5% level. The aim was to determine what variables would predict early dropout from the web-app. Early dropout was expressed as a binary variable: participants were marked as dropping out early if less than 2 modules were completed. For example, participants who completed 2 or 3 modules were marked as 0 (ie, not dropped out early), while participants who completed only 1 module or none were marked as 1 (ie, dropped out early).

To investigate hypothesis 1, a general logistic regression model using the glm function was built to identify whether HAPA variables, age, marital status, education, and recruitment channel predicted dropout. Recruitment channel was a categorical variable that reflected entry into the app. Participants entered the app either directly through clinical recruitment or in the Germany-wide recruitment after randomization into the intervention group or after the waiting time when being randomized into the waitlist control arm. In the logistic regression model, the clinical recruitment group was the reference group to which the other recruitment channels were compared to.

A hierarchical regression was performed to investigate hypothesis 2, using the "lm" function available in Base R; HAPA variables were sequentially added to build a final model that predicts post web-app communication. Table 2 outlines how the predictor and outcome variables were operationalized in the model. Following the construction of a model to explain changes in post web-app communication, possible processes for behavior change were proposed. To investigate hypothesis 3, a structural equation model using the lavaan package [62] was built to identify how HAPA variables related to one another and in turn contributed to changes in post web-app communication.

Table 2. Overview of the operationalization of social-cognitive predictors and communication behavior.

Type	Variable	Operationalization	Example
Outcome	• Post web-app communication, expressed as C_t	Respondent's most recent overall communication score after having completed at least 1 module of the web-app. Only individuals who completed at least 1 module were included in the analyses to ensure the data captured those that used the web-app.	If an individual had communication scores after modules 1, 2, and 3, only the response after module 3 was used.
Predictor	• Outcome expectancy at the preceding timepoint, expressed as OE_{t-1} • Intention at the preceding timepoint, expressed as $Intention_{t-1}$ • Self-efficacy at the preceding timepoint, expressed as SE_{t-1} • Action planning at the preceding timepoint, expressed as AP_{t-1}	A respondent's HAPA[a] variable score at the timepoint preceding the last available communication score	If an individual had a communication score after module 3, their HAPA variable scores after module 2 were used for the predictor variables.

[a]HAPA: health action process approach.

Multimedia Appendix 2 outlines the intercorrelation between all used social-cognitive HAPA determinants as well as self-reported communication behavior over the course of the app-based intervention. Since the abovementioned analysis includes as many timepoints as possible, all variables at the different timepoints were included.

Results

Study Participants

Overall, 1187 women were recruited into this study, of which 988 were from the Germany-wide recruitment (Figure 1A). Of those in the Germany-wide sample, 506 were randomized into the waitlist control arm (control group app registration), and 482 were randomized into the intervention arm (intervention group app registration). In the clinics, 199 pregnant women were recruited (after the registration of 205 women with the app). The majority of the participants were aged 30-39 years (n=881), had a higher education status (n=763), and were married (n=759).

Descriptive Statistics

Sociodemographic data are depicted in detail in Table 3 below.

Table 3. Sociodemographic characteristics of expectant mothers (N=1187).

Characteristics	Values, n (%)
Age[a] (years)	
18-29	159 (14.23)
30-39	881 (78.87)
40-49	77 (6.89)
Marital status[b]	
Single	31 (2.67)
In a committed relationship	366 (31.5)
Married/registered partnership	759 (65.32)
Divorced/separated/widowed	6 (0.51)
Highest educational level[c]	
No school-leaving qualification	21 (1.81)
Secondary or elementary school leaving	78 (6.74)
Secondary school diploma	137 (11.83)
A-levels	763 (65.89)
Completed vocational training	25 (2.16)
University degree[d]	34 (2.94)
University degree[e]	100 (8.64)

[a]76 missing values for age.
[b]31 missing values for marital status.
[c]35 missing values for highest educational level.
[d]Special German university degree (Hochschule).
[e]University degree.

Predicting Dropout

Of the 1187 pregnant women who were recruited and started using the web-app, 1124 dropped out of the intervention, as indicated by completion of less than 2 modules. A general logistic model was estimated to investigate hypothesis 1 and to determine whether social-cognitive HAPA variables and communication behavior as well as sociodemographic characteristics might predict early dropout (completing less than 2 modules). Thereby, the predictive capacity of 4 HAPA variables and behavior along with age, education, and marital status was tested: intention, outcome expectancy, self-efficacy, and action planning, as well as sex, education, and marital status. As Table 4 highlights, only age was a significant predictor of early dropout. In other words, younger pregnant women were more likely to drop out from the digital intervention at an earlier stage. Accordingly, hypothesis 1 can be empirically supported.

Table 4. Parameter table of the generalized linear model predicting early dropout from health action process approach variables and sociodemographic characteristics.

Variable	Estimate (SE)	*t* test *(df)*[a]	*P* value
Outcome expectancy$_{tt1}$[b]	.005 (.164)	0.034 (418)	.97
Intention$_{tt1}$	–.254 (.153)	–1.658 (418)	.09
Self-efficacy$_{tt1}$	.184 (.121)	1.516 (418)	.13
Action planning$_{t1}$[c]	.204 (.112)	1.816 (418)	.07
Age	–.094 (.038)	–2.469 (418)	.014[d]
Education	–.178 (.161)	–1.106 (418)	.27
Marital status	–.363 (.302)	–1.200 (418)	.23
Recruitment channel=Germany-wide recruitment, randomized into the intervention group (compared to clinical recruitment)	.880 (.523)	1.683 (418)	.09
Recruitment channel=Germany-wide recruitment, randomized into the waitlist control arm (compared to clinical recruitment)	.374 (.540)	0.692 (418)	.49

[a]2-sided *t* test.

[b]tt1: measurement after completing module 1.

[c]t1: module 1 (lessons 1-3).

[d]β is significant at *P*=.05; R^2=0.13.

Predictors of Self-Reported Communication Behavior

Hypothesis 2 tests whether socio-demographic variables and social-cognitive HAPA variables (self-efficacy, outcome expectancy, and action planning) would predict self-reported communication behavior in pregnant women over the course of the app-based intervention (see Table 5 for details). In the first series of models, each sociodemographic variable was added in a stepwise fashion to predict communication behavior. Adding age ($F_{1,93}$=1.16; *P*=.28), education ($F_{1,92}$=0.12; *P*=.66), and family status ($F_{1,91}$=0.39; *P*=.54) did not significantly improve the prediction of communication behavior scores. In the subsequent models, HAPA variables were added (Table 5), and intention was added. Upon inclusion, most HAPA variables improved the model fit. For the motivational phase of the HAPA model, outcome expectancy ($F_{1,90}$=4.88; *P*=.03) and intention ($F_{1,98}$=8.65; *P*=.004) improved the model fit, while task self-efficacy ($F_{1,88}$=2.11; *P*=.15) did not improve the model fit. For the volitional phase, action planning ($F_{1,87}$=17.74; *P*<.001) improved the prediction of communication scores.

After including all the variables of HAPA along with sociodemographic variables into the model (for model comparisons, see Table 5), only action planning (β=.331; *P*<.001) significantly predicted communication behavior (Table 6 for further parameter estimates). Accordingly, hypothesis 2 could be partially empirically supported.

Table 5. Hierarchical regression model comparison of sociodemographic and social-cognitive health action process approach variables predicting communication behavior.

Model name	Comparison model	Predictors	*F* test *(df)*	*P* value
Age	Null	Age	2.165 (1, 92)	.15
Education status	Age	Age + education	0.045 (1, 91)	.83
Marital status	Education status	Age + education + marital status	0.365 (1, 90)	.55
Outcome expectancy	N/A[a]	Outcome expectancy	4.430 (1, 89)	.04[b]
Intention	Outcome expectancy	Outcome expectancy + intention	8.457 (1, 88)	.005[c]
Self-efficacy	Intention	N/A	1.955 (1, 87)	.17
Action planning	Self-efficacy	N/A	17.68 (1, 86)	<.001[d]

[a]N/A: not applicable.

[b]β is significant at *P*=.05.

[c]β is significant at *P*=.01.

[d]β is significant at *P*=.001.

Table 6. Parameter table of hierarchical regression model predicting communication behavior.

Variable	Estimate (95% CI)	SE	*P* value
Age	–. 016 (–.051 to .018)	.02	.34
Education status	–. 047 (–.154 to .060)	.053	.39
Marital status	–. 100 (–.442 to .243)	.172	.57
Outcome expectancy	.14 (–.047 to .336)	.096	.14
Intention	.046 (–.143 to .236)	.095	.63
Self-efficacy	.007 (–.136 to .149)	.072	.93
Action planning	.305[a] (–.161 to .449)	.072	<.001

[a]β is significant at *P*=.001.

Mediation Model

To test hypothesis 3 and thus whether distal HAPA variables (intention, outcome expectancies, self-efficacy) are mediated through planning, a path model was facilitated (Figure 2 and Table 7). Indeed, a sequential mediation emerged from self-efficacy to intention to action planning to self-reported communication behavior. This likewise entailed the indirect mediation from intention via action planning to self-reported communication behavior. Conversely, no serial mediation was found from outcome expectancies to intention via action planning to self-reported communication behavior. Accordingly, hypothesis 3 could only be empirically supported regarding 2 of the 3 distal HAPA variables, namely, self-efficacy and intention.

Figure 2. Regression model of social-cognitive health action process approach variables and safe communication behavior across all groups. Taken together, communication behavior is significantly predicted by action planning, and action planning mediates the impact of self-efficacy and intention on self-reported communication behavior with β=.017 (comparative fit index=0.994; Tucker–Lewis index=0.971; root mean square error of approximation=0.055). AP: action planning; COM: safe communication behavior; INT: intention; SE: self-efficacy; t: reflects a relative timepoint.

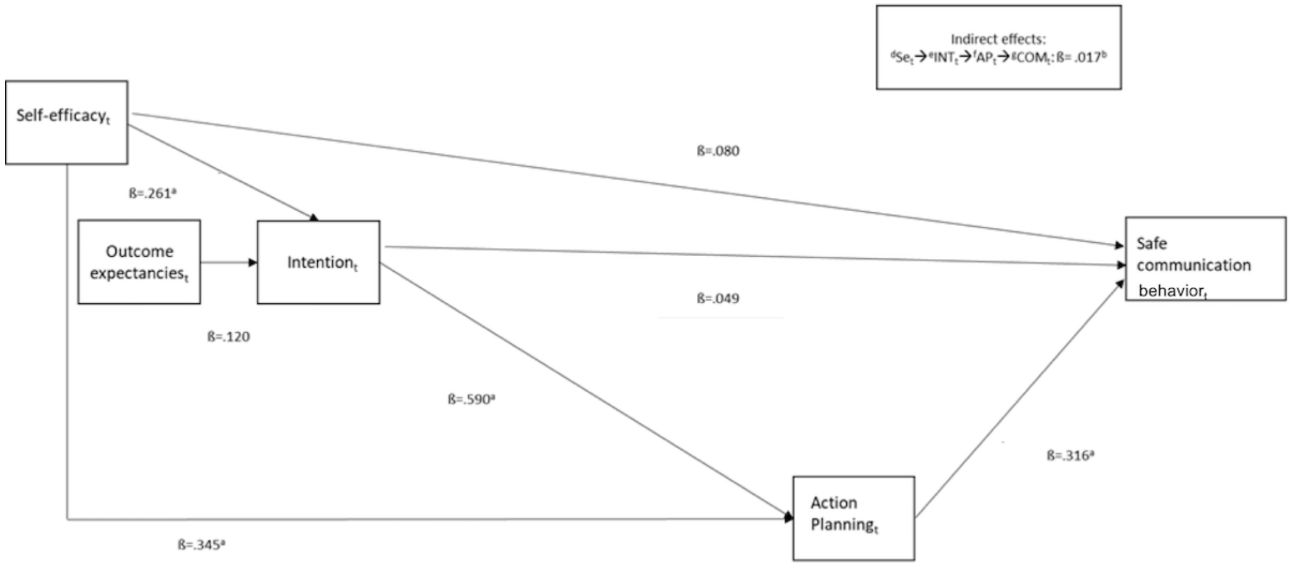


Table 7. Results of the path model depicted in Figure 2.

Predictor variable	Outcome variable	Estimate (95% CI)	SE	*P* value
Self-efficacy	Intention	.353[a] (.152 to .555)	.103	.001
Outcome expectancy	Intention	.105 (–.108 to .318)	.109	.33
Intention	Action planning	.408[a] (.242 to .574)	.085	<.001
Self-efficacy	Planning	.322[a] (.148 to .497)	.089	<.001
Planning	Communication behavior	.471[a] (.311 to .631)	.082	<.001
Intention	Communication behavior	.053 (–.163 to .270)	.111	.63
Self-efficacy	Communication behavior	.108 (–.069 to .285)	.090	.23

[a]β is significant at *P*=.001.

Discussion

Principal Findings

In this study, we investigated the determinants of dropout from an app-based intervention for pregnant women and the mechanisms of adopting self-reported communication behavior. Regarding both aspects, variables from HAPA were measured and evaluated with respect to their predictive capability. Consistent with hypothesis 1, dropout analyses found that only age was predictive, and none of the HAPA variables played a role. Communication behavior was only predicted significantly by one of the HAPA variables, namely, action planning. A serial mediation emerged from intention to self-reported communication behavior via action planning. In detail, communication behavior was significantly predicted by action planning, and action planning mediated the impact of self-efficacy and intention on self-reported communication behavior. These findings match previous findings that identified age as a predictor of dropout from digital interventions [41,42]. However, in contrast to the aforementioned studies, no other sociodemographic predictors emerged. Moreover, this study deviates from previous dropout investigations, as no HAPA variables predicted early dropout. This might be a result of differences in the target group or context of the digital tool between studies; predictors of dropout from digital interventions might depend on aspects of specific intervention types and could vary based on the timepoint of dropout, as revealed in a previous research [63]. Regarding the latter, it should be noted that our study only investigated early dropout.

With respect to mechanisms of adopting self-reported communication behavior, the results were in line with hypotheses 2 and 3: the data demonstrated that HAPA variables predicted the self-reported communication behavior. Whether this is a result of the behavior change context (ie, the app) still needs to be determined. However, the findings are in line with previous evidence that HAPA assumptions match the data and accordingly are able to explain the changes in self-reported communication behavior [8,10].

Our findings of early dropout from the app-based intervention are partially in line with those of other studies, in which younger participants were more likely to drop out from digital interventions [42,64-66]. The relationship between age and dropout could be a result of higher perceived need among older women, that is, older women, through more life experiences and previous pregnancies, may realize a greater need for communication interventions and in turn adhere to the app.

Behavior change variables did not predict dropout from the app-based intervention. This is in contrast to the results of previous studies that have shown that behavior change variables are associated with dropout [63]. Among others, this study shows outcome expectancies as a crucial predictor of retention in digital interventions and likewise concludes that the perception of unmet needs and expectations might be a determinant of dropout [63]. In the context of self-reported communication behavior, expectant mothers' outcome expectancies focused on safe child delivery instead of self-reported communication behavior and the accommodation of individual preferences as the item wording suggests. An alternative explanation would be that the relationship between behavior change variables and dropout is context-specific, as a previous study implicates [63]. In any case, our study shows the determinants of dropout from an app-based intervention. This is an important insight for patient safety interventions in obstetrics because it can be used by future tools to prevent early dropout and maximize the amount of support that pregnant women receive. However, future research should further investigate the contextual variability of predictors of early dropout in digital intervention studies. This is particularly important because app-based interventions generally show high dropout rates [67-70].

Previous studies [8,11] have shown that social-cognitive variables are associated with pregnant women's safe communication behavior in general. Our study demonstrates that some of these associations also drive change in self-reported communication behavior during a digital intervention. This is of importance for theoretical understanding and practitioners; it shows how apps might elicit and affect changes in self-reported communication behavior, highlighting pathways which future interventions can focus on and improve its effectiveness. This might be useful for designing future apps in the specific field of pregnancy and giving birth.

It is striking that not all associations in HAPA emerged as theorized. First, it became apparent that action planning was the single best predictor of change in self-reported communication behavior. Although the predictive capacity of action planning is expected from its association with behavior within the HAPA model, it was not hypothesized that action planning would emerge as the sole predictor of behavior change. The reason could be pregnant women participating in the app were taught to think of how and when they might communicate effectively, that is, making concrete action plans, which worked well in the app, while other variables were not addressed as effectively. In addition, action planning was targeted in the last lesson of the app, which might have resulted in stronger effects due to a shorter time lag and recency effect.

Relating to the process by which pregnant women improve their communication behavior in clinical care, it is likewise striking why only 2 of the indirect effects specified in the HAPA framework emerged. First, there was an indirect effect of self-efficacy on self-reported communication behavior via intention and subsequently via action planning. Second, self-efficacy also directly impacted action planning and thereby indirectly impacted self-reported communication behavior. Notably, the same predictive capacity in explaining self-reported communication behavior in this study replicated findings from a cross-sectional research [8,11]. Indeed, the HAPA model seems to be applicable for predicting several kinds of behavior change in digitally supported interventions like the app used in this study and has shown similar findings overall [71-73].

In a previous randomized trial in patients with insomnia, both action planning and coping planning in the HAPA model were shown to be effective mediators in improving sleep hygiene [71]. In another randomized trial testing a digital tool to promote active lifestyles in patients with type 2 diabetes, the intervention group showed a significant intervention effect for action planning, whereas the control group exhibited a significant effect for coping planning and self-efficacy [72]. Lastly, in an earlier study primarily focusing on reducing salt intake to prevent high blood pressure, both intention and outcome expectancies as well as risk perception were found to be improved by the digital intervention [73].

In the future, digital and nondigital face-to-face interventions should be compared, especially when aiming to improve self-reported communication behavior in obstetrics and preventing dropout [10]. Different app modes showed various degrees of effectiveness in a study on depression [74]. A meta-analytic review [74] showed that apps in combination with personal contact with a therapist are more effective than self-help apps. However, no differences were found between smartphone-based apps and computer- and internet-based interventions. Similarly, there seems to be no difference between human-guided digital interventions and face-to-face psychotherapy [74]. In other areas of research, gamifications in apps have proven to be beneficial [68]. Among other findings, feedback, leaderboards (participants can compare their own progress with that of others), and storytelling (context within the app to create an alternate reality and guide the user) have been shown to be advantageous for digital interventions [68-70]. Those findings provide some guidance for future teams aiming at developing apps and internet interventions in this field.

Findings from this body of research set the stage for iterating on existing apps in clinical care and for developing new apps. However, it seems questionable, what kind of intervention might be sufficient or helpful for those participants at risk for dropping out. On the one hand, flexible digital tools, which allow an automatic dynamic change of modules and learning intensity, might be helpful. On the other hand, an overload of information or special attention to these participants might make them even more prone to drop out. To conclude, future research should further uncover the reasons for dropout of such participants, so that optimal strategies for prevention can be devised.

Given that younger patients are at increased risk of discontinuing the digital intervention at an early stage, it is crucial to make the underlying behavior visible and targetable. We made the first attempt to explain this phenomenon. In the literature, the possibility of using behavior change theories to predict dropout behavior has been demonstrated [63]. Future research should be conducted using different intervention modes as well as different digital incentives (eg, optimal level of gamification, possibility to exchange with other users vs personal contact with midwives, doctors, or other birthing professionals, or more intensive self-help vs person-guided self-help).

Limitations and Suggestions for Future Research

This study, as a formative research and with preliminary results, has certain limitations with regard to the conclusions drawn from the results. First, there is a possibility that confounding external factors could have been at play during the course of the internet intervention, such as physiological or mental health risk factors. Second, bias and self-selection might have confounded the web-app data in the sense that only certain women volunteered to participate and continue the app-based intervention. Future internet intervention studies should try to recruit a more diverse sample and find concrete reasons for them completing the app or dropping out. It seems possible that both flexible feasibility questions in the context of the app as well as effectiveness ratings and satisfaction ratings could provide more information about usage behavior.

For the time frame between intervention start and the last timepoint, we have conducted a test of factors predicting dropout, from which we concluded that only younger age at intervention start predicted early dropout. However, factors associated with the selection to participate in the intervention could not be uncovered in our presented design. Future research should target a wider age range of pregnant women to gain further insight between age groups and dropout via subgroup analyses. With a larger sample, it also seems possible to examine the age categories and dropout behavior in more detail.

As mentioned previously, our data do not allow conclusions regarding the motivations of dropout and study retainment—a topic that future studies should investigate further. Relatedly, it would have been important to have more finely spaced time intervals for measurement points, which could add valuable information on the interplay of the processes underlying behavioral change and the topics of the particular lessons that were covered. Additionally, it should be acknowledged that the scales used to measure the social-cognitive variables of HAPA were not previous validated in German language based on evaluating the communication behavior. Hence, the measurement qualities of the scales in the German population might be limited.

The main contribution of this study can be seen in that it is the first attempt of employing a digitally enhanced internet intervention aiming at fostering self-reported communication behavior within a clinical sample in the context of obstetric care. This has innovation potential, as it shows that technology in terms of a digital intervention, that is, apps aiming at improving communication behavior can make a difference. Therefore, this study sheds light on the mechanisms underlying self-reported communication behavior and its improvement while also investigating the potential predictors of dropout in an app-based intervention. This contribution yields both practical and theoretical implications. On a theoretical note, our study contributes to a deeper understanding of the genesis of self-reported communication behavior, thereby highlighting various points of the psychological processes that future interventions could address, such as action planning and self-efficacy. Likewise, practical implications arise as our study presents an initial framework for improving effective communication via the app in a clinical sample and explores how to maximize its effectiveness by retaining participants at risk of early dropout. The variables of the HAPA model can function as a toolkit, with a particular focus on action planning, self-efficacy beliefs, intention to communicate effectively, and app users’ age.

Acknowledgments

We thank Katharina Schiffer for proofreading the manuscript regarding referencing style and language. We are thankful and appreciative of the help from Dr Marina Schmiedhofer, Christina Derksen (PhD), Prof Dr Frank Louwen, Prof Dr Frank Reister, Johanna Dietl, Freya Häußler, and Sandra Riedmüller in the data collection. This study was funded by the German Innovation Fund (project 01VSF18023) of the Federal Joint Committee (Gemeinsamer Bundesausschuss) as part of a third-party project funding.

Data Availability

The data that support the findings of this study are available on request from the corresponding author SL. The data are not allowed to be made publicly available due to privacy and data security reasons of the research participants.

Authors' Contributions

LK, VA-K, and SL were involved in data collection and monitoring as well as in the conceptual aspects of this study. LK analyzed and described the data statistically and wrote all parts of this paper. SL advised on the methodology and structure. FMK, VA-K, and NTH gave advice on the rationale and structure of this paper. All coauthors approved the final version and contributed to the preparation as well as revision of the final paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Specifications of the TeamBaby web-app.

[DOCX File , 16 KB - humanfactors_v11i1e48218_app1.docx]

Multimedia Appendix 2

Intercorrelation between all used social-cognitive determinants as well as self-reported communication behavior over the course of the app-based intervention.

[DOCX File , 34 KB - humanfactors_v11i1e48218_app2.docx]

References

1. McLaughlin T, Yan Z. Diverse delivery methods and strong psychological benefits: A review of online formative assessment. Computer Assisted Learning 2017 May 25;33(6):562-574. [doi: 10.1111/jcal.12200]
2. Runciman W, Hibbert P, Thomson R, Van Der Schaaf T, Sherman H, Lewalle P. Towards an international classification for patient safety: key concepts and terms. Int J Qual Health Care 2009 Feb;21(1):18-26 [FREE Full text] [doi: 10.1093/intqhc/mzn057] [Medline: 19147597]
3. Eulmesekian PG, Alvarez JP, Ceriani Cernadas JM, Pérez A, Berberis S, Kondratiuk Y. The occurrence of adverse events is associated with increased morbidity and mortality in children admitted to a single pediatric intensive care unit. Eur J Pediatr 2019 Dec 08;179(3):473-482. [doi: 10.1007/s00431-019-03528-z]
4. Schrappe M, Spahn J, Berwick D, Durkin P. APS-Weißbuch. Sicherheit in der Gesundheitsversorgung: Medizinisch Wissenschaftliche Verlagsgesellschaft; 2018.

5. Shojania KG, Marang-van de Mheen PJ. Identifying adverse events: reflections on an imperfect gold standard after 20 years of patient safety research. BMJ Qual Saf 2020 Apr;29(4):265-270. [doi: 10.1136/bmjqs-2019-009731] [Medline: 32066571]
6. Antony J, Zarin W, Pham B, Nincic V, Cardoso R, Ivory JD, et al. Patient safety initiatives in obstetrics: a rapid review. BMJ Open 2018 Jul 06;8(7):e020170 [FREE Full text] [doi: 10.1136/bmjopen-2017-020170] [Medline: 29982200]
7. Gittell JH. Relationships between service providers and their impact on customers. Journal of Service Research 2016 Jun 29;4(4):299-311. [doi: 10.1177/1094670502004004007]
8. Derksen C, Kötting L, Keller FM, Schmiedhofer M, Lippke S. Psychological intervention to improve communication and patient safety in obstetrics: examination of the health action process approach. Front Psychol 2022;13:771626 [FREE Full text] [doi: 10.3389/fpsyg.2022.771626] [Medline: 35250715]
9. Lippke S, Wienert J, Keller FM, Derksen C, Welp A, Kötting L, et al. Communication and patient safety in gynecology and obstetrics - study protocol of an intervention study. BMC Health Serv Res 2019 Nov 28;19(1):908 [FREE Full text] [doi: 10.1186/s12913-019-4579-y] [Medline: 31779620]
10. Derksen C, Dietl JE, Haeussler FE, Steinherr Zazo M, Schmiedhofer M, Lippke S. Behavior change training for pregnant women's communication during birth: A randomized controlled trial. Appl Psychol Health Well Being 2023 Aug;15(3):865-883. [doi: 10.1111/aphw.12413] [Medline: 36380576]
11. Kötting L, Henschel N, Keller FM, Derksen C, Lippke S. Social-cognitive correlates of expectant mothers' safe communication behaviour: Applying an adapted HAPA model. Cogent Psychology 2023 Feb 19;10(1):2173996-2173996. [doi: 10.1080/23311908.2023.2173996]
12. Bertakis KD, Franks P, Azari R. Effects of physician gender on patient satisfaction. J Am Med Womens Assoc (1972) 2003;58(2):69-75. [Medline: 12744418]
13. Dolan JG. Shared decision-making--transferring research into practice: the Analytic Hierarchy Process (AHP). Patient Educ Couns 2008 Dec;73(3):418-425 [FREE Full text] [doi: 10.1016/j.pec.2008.07.032] [Medline: 18760559]
14. Rider EA, Keefer CH. Communication skills competencies: definitions and a teaching toolbox. Med Educ 2006 Jul;40(7):624-629. [doi: 10.1111/j.1365-2929.2006.02500.x] [Medline: 16836534]
15. Attanasio LB, McPherson ME, Kozhimannil KB. Positive childbirth experiences in U.S. hospitals: a mixed methods analysis. Matern Child Health J 2014 Jul;18(5):1280-1290 [FREE Full text] [doi: 10.1007/s10995-013-1363-1] [Medline: 24072597]
16. Delaney AL, Singleton G. Information and relationship functions of communication between pregnant women and their health care providers. Communication Studies 2020 Aug 20;71(5):800-822. [doi: 10.1080/10510974.2020.1807376]
17. Kim SC, Boren D, Solem SL. The Kim Alliance Scale: development and preliminary testing. Clin Nurs Res 2001 Aug;10(3):314-331. [doi: 10.1177/c10n3r7] [Medline: 11881945]
18. Nicoloro-SantaBarbara J, Rosenthal L, Auerbach MV, Kocis C, Busso C, Lobel M. Patient-provider communication, maternal anxiety, and self-care in pregnancy. Soc Sci Med 2017 Oct;190:133-140. [doi: 10.1016/j.socscimed.2017.08.011] [Medline: 28863336]
19. Garnweidner-Holme LM, Lukasse M, Solheim M, Henriksen L. Talking about intimate partner violence in multi-cultural antenatal care: a qualitative study of pregnant women's advice for better communication in South-East Norway. BMC Pregnancy Childbirth 2017 Apr 19;17(1):123 [FREE Full text] [doi: 10.1186/s12884-017-1308-6] [Medline: 28420328]
20. Berry D. Risk, Communication & Health Psychology. UK: McGraw-Hill Education; 2004.
21. Rundell S. A study of nurse-patient interaction in a high dependency unit. Intensive Care Nurs 1991 Sep;7(3):171-178. [doi: 10.1016/0266-612x(91)90006-d] [Medline: 1918904]
22. Yang Y, Zhang X, Lee PK. Improving the effectiveness of online healthcare platforms: An empirical study with multi-period patient-doctor consultation data. International Journal of Production Economics 2019 Jan;207:70-80. [doi: 10.1016/j.ijpe.2018.11.009]
23. Simpson M, Catling C. Understanding psychological traumatic birth experiences: A literature review. Women Birth 2016 Jun;29(3):203-207. [doi: 10.1016/j.wombi.2015.10.009] [Medline: 26563636]
24. Hollander MH, van Hastenberg E, van Dillen J, van Pampus MG, de Miranda E, Stramrood CAI. Preventing traumatic childbirth experiences: 2192 women's perceptions and views. Arch Womens Ment Health 2017 Aug;20(4):515-523. [doi: 10.1007/s00737-017-0729-6] [Medline: 28553692]
25. Lippke S, Derksen C, Keller FM, Kötting L, Schmiedhofer M, Welp A. Effectiveness of communication interventions in obstetrics—a systematic review. IJERPH 2021 Mar 05;18(5):2616. [doi: 10.3390/ijerph18052616]
26. Smith KA, Bishop FL, Dambha-Miller H, Ratnapalan M, Lyness E, Vennik J, et al. Improving empathy in healthcare consultations-a secondary analysis of interventions. J Gen Intern Med 2020 Oct;35(10):3007-3014 [FREE Full text] [doi: 10.1007/s11606-020-05994-w] [Medline: 32666488]
27. D'Agostino TA, Atkinson TM, Latella LE, Rogers M, Morrissey D, DeRosa AP, et al. Promoting patient participation in healthcare interactions through communication skills training: A systematic review. Patient Educ Couns 2017 Jul;100(7):1247-1257 [FREE Full text] [doi: 10.1016/j.pec.2017.02.016] [Medline: 28238421]
28. West R, May S, West M, Croghan E, McEwen A. Performance of English stop smoking services in first 10 years: analysis of service monitoring data. BMJ 2013 Aug 19;347:f4921 [FREE Full text] [doi: 10.1136/bmj.f4921] [Medline: 23963106]

29. Schauer GL, Malarcher AM, Zhang L, Engstrom MC, Zhu S. Prevalence and correlates of quitline awareness and utilization in the United States: an update from the 2009-2010 National Adult Tobacco Survey. Nicotine Tob Res 2014 May;16(5):544-553. [doi: 10.1093/ntr/ntt181] [Medline: 24253378]
30. Moshe I, Terhorst Y, Philippi P, Domhardt M, Cuijpers P, Cristea I, et al. Digital interventions for the treatment of depression: A meta-analytic review. Psychol Bull 2021 Aug;147(8):749-786. [doi: 10.1037/bul0000334] [Medline: 34898233]
31. Dores AR, Geraldo A, Carvalho IP, Barbosa F. The use of new digital information and communication technologies in psychological counseling during the COVID-19 pandemic. Int J Environ Res Public Health 2020 Oct 21;17(20):7663 [FREE Full text] [doi: 10.3390/ijerph17207663] [Medline: 33096650]
32. Rose T, Barker M, Maria Jacob C, Morrison L, Lawrence W, Strömmer S, et al. A systematic review of digital interventions for improving the diet and physical activity behaviors of adolescents. J Adolesc Health 2017 Dec;61(6):669-677 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.05.024] [Medline: 28822682]
33. Kaner EF, Beyer FR, Garnett C, Crane D, Brown J, Muirhead C, et al. Personalised digital interventions for reducing hazardous and harmful alcohol consumption in community-dwelling populations. Cochrane Database Syst Rev 2017 Sep 25;9(9):CD011479 [FREE Full text] [doi: 10.1002/14651858.CD011479.pub2] [Medline: 28944453]
34. Vilardaga R, Casellas-Pujol E, McClernon JF, Garrison KA. Mobile applications for the treatment of tobacco use and dependence. Curr Addict Rep 2019 Jun;6(2):86-97 [FREE Full text] [doi: 10.1007/s40429-019-00248-0] [Medline: 32010548]
35. Marler JD, Fujii CA, Utley DS, Tesfamariam LJ, Galanko JA, Patrick H. Initial assessment of a comprehensive digital smoking cessation program that incorporates a mobile app, breath sensor, and coaching: cohort study. JMIR Mhealth Uhealth 2019 Feb 04;7(2):e12609 [FREE Full text] [doi: 10.2196/12609] [Medline: 30670372]
36. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y, Dobson R. Mobile phone text messaging and app-based interventions for smoking cessation. Cochrane Database Syst Rev. (CD006611). PMC680 2019;10:4292. [doi: 10.1002/14651858.cd006611.pub5]
37. Marcu G, Ondersma SJ, Spiller AN, Broderick BM, Kadri R, Buis LR. The perceived benefits of digital interventions for behavioral health: qualitative interview study. J Med Internet Res 2022 Mar 30;24(3):e34300 [FREE Full text] [doi: 10.2196/34300] [Medline: 35353045]
38. Kripalani S, LeFevre F, Phillips CO, Williams MV, Basaviah P, Baker DW. Deficits in communication and information transfer between hospital-based and primary care physicians: implications for patient safety and continuity of care. JAMA 2007 Feb 28;297(8):831-841. [doi: 10.1001/jama.297.8.831] [Medline: 17327525]
39. Karin E, Dear BF, Heller GZ, Crane MF, Titov N. "Wish You Were Here": examining characteristics, outcomes, and statistical solutions for missing cases in web-based psychotherapeutic trials. JMIR Ment Health 2018 Apr 19;5(2):e22 [FREE Full text] [doi: 10.2196/mental.8363] [Medline: 29674311]
40. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: A systematic review and meta-analysis. J Affect Disord 2020 Feb 15;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
41. Karyotaki E, Kleiboer A, Smit F, Turner DT, Pastor AM, Andersson G, et al. Predictors of treatment dropout in self-guided web-based interventions for depression: an 'individual patient data' meta-analysis. Psychol Med 2015 Oct;45(13):2717-2726 [FREE Full text] [doi: 10.1017/S0033291715000665] [Medline: 25881626]
42. Moshe I, Terhorst Y, Paganini S, Schlicker S, Pulkki-Råback L, Baumeister H, et al. Predictors of dropout in a digital intervention for the prevention and treatment of depression in patients with chronic back pain: secondary analysis of two randomized controlled trials. J Med Internet Res 2022 Aug 30;24(8):e38261 [FREE Full text] [doi: 10.2196/38261] [Medline: 36040780]
43. Wu MS, Chen S, Wickham RE, Leykin Y, Varra A, Chen C, et al. Predicting non-initiation of care and dropout in a blended care CBT intervention: Impact of early digital engagement, sociodemographic, and clinical factors. Digit Health 2022;8:20552076221133760 [FREE Full text] [doi: 10.1177/20552076221133760] [Medline: 36312847]
44. Gao L, Keller FM, Becker P, Dahmen A, Lippke S. Predictors of dropout among psychosomatic rehabilitation patients during the COVID-19 pandemic: secondary analysis of a longitudinal study of digital training. J Med Internet Res 2023 Nov 27;25:e43584 [FREE Full text] [doi: 10.2196/43584] [Medline: 37903289]
45. Cerdà-Navarro A, Quintana-Murci E, Salvà-Mut F. Reasons for dropping out of intermediate vocational education and training in Spain: the influence of sociodemographic characteristics and academic background. Journal of Vocational Education & Training 2022 Mar 06:1-25. [doi: 10.1080/13636820.2022.2049625]
46. Davis MJ, Addis ME. Predictors of attrition from behavioral medicine treatments. Ann Behav Med 1999;21(4):339-349. [doi: 10.1007/BF02895967] [Medline: 10721442]
47. Vöhringer M, Knaevelsrud C, Wagner B, Slotta M, Schmidt A, Stammel N, et al. Should I stay or must I go? Predictors of dropout in an internet-based psychotherapy programme for posttraumatic stress disorder in Arabic. Eur J Psychotraumatol 2020;11(1):1706297 [FREE Full text] [doi: 10.1080/20008198.2019.1706297] [Medline: 32082510]
48. Murray E, White IR, Varagunam M, Godfrey C, Khadjesari Z, McCambridge J. Attrition revisited: adherence and retention in a web-based alcohol trial. J Med Internet Res 2013 Aug 30;15(8):e162 [FREE Full text] [doi: 10.2196/jmir.2336] [Medline: 23996958]

49. Schroé H, Crombez G, De Bourdeaudhuij I, Van Dyck D. Investigating when, which, and why users stop using a digital health intervention to promote an active lifestyle: secondary analysis with a focus on health action process approach-based psychological determinants. JMIR Mhealth Uhealth 2022 Jan 31;10(1):e30583 [FREE Full text] [doi: 10.2196/30583] [Medline: 35099400]
50. Maenhout L, Peuters C, Cardon G, Crombez G, DeSmet A, Compernolle S. Nonusage attrition of adolescents in an mHealth promotion intervention and the role of socioeconomic status: secondary analysis of a 2-arm cluster-controlled trial. JMIR Mhealth Uhealth 2022 May 10;10(5):e36404 [FREE Full text] [doi: 10.2196/36404] [Medline: 35536640]
51. Donkin L, Glozier N. Motivators and motivations to persist with online psychological interventions: a qualitative study of treatment completers. J Med Internet Res 2012 Jun 22;14(3):e91 [FREE Full text] [doi: 10.2196/jmir.2100] [Medline: 22743581]
52. Davis R, Campbell R, Hildon Z, Hobbs L, Michie S. Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. Health Psychol Rev 2015;9(3):323-344 [FREE Full text] [doi: 10.1080/17437199.2014.941722] [Medline: 25104107]
53. Dillon K, Rollo S, Prapavessis H. A combined health action process approach and mHealth intervention to reduce sedentary behaviour in university students - a randomized controlled trial. Psychol Health 2022 Jun;37(6):692-711. [doi: 10.1080/08870446.2021.1900574] [Medline: 33780297]
54. Schwarzer R. Modeling health behavior change: how to predict and modify the adoption and maintenance of health behaviors. Applied Psychology 2008 Jan 30;57(1):1-29. [doi: 10.1111/j.1464-0597.2007.00325.x]
55. Every newborn: an action plan to end preventable deaths. WHO. 2014. URL: https://www.who.int/initiatives/every-newborn-action-plan [accessed 2023-01-03]
56. Schwarzer R, Luszczynska A. Self-efficacy. Division of Cancer Control and Population Sciences (DCCPS). URL: https://cancercontrol.cancer.gov/brp/research/constructs/self-efficacy [accessed 2023-02-16]
57. Schwarzer R, Hamilton K. Changing behavior using the Health Action Process Approach. In: The Handbook of Behavior Change. Cambridge, England: Cambridge University Press; 2020:89-103.
58. Keller FM, Derksen C, Kötting L, Schmiedhofer M, Lippke S. Development of the perceptions of preventable adverse events assessment tool (PPAEAT): measurement properties and patients' mental health status. Int J Qual Health Care 2021 Apr 16;33(2):mzab063. [doi: 10.1093/intqhc/mzab063] [Medline: 33822086]
59. Gholami M, Schwarzer R, Reyes-Fernandez B, Rica C. Brief scales for the multilingual assessment of HAPA variables. Health Psychology. URL: https://www.psyc.de/hapascales.pdf [accessed 2023-01-10]
60. Accessed April 3, 2023. R: The R Project for Statistical Computing. URL: https://www.r-project.org/ [accessed 2023-01-05]
61. RStudio T. RStudio: integrated development for R. RStudio. URL: http://www.rstudio.com/ [accessed 2023-01-02]
62. Rosseel Y. lavaan: An R package for structural equation modeling. Journal of Statistical Software 2012;48:1-36. [doi: 10.18637/jss.v048.i02]
63. Lippke S, Gao L, Keller FM, Becker P, Dahmen A. Adherence with online therapy vs face-to-face therapy and with online therapy vs care as usual: secondary analysis of two randomized controlled trials. J Med Internet Res 2021 Nov 03;23(11):e31274 [FREE Full text] [doi: 10.2196/31274] [Medline: 34730541]
64. Linnet J, Hertz SPT, Jensen ES, Runge E, Tarp KHH, Holmberg TT, et al. Days between sessions predict attrition in text-based internet intervention of binge eating disorder. Internet Interv 2023 Mar;31:100607 [FREE Full text] [doi: 10.1016/j.invent.2023.100607] [Medline: 36819741]
65. Jiandani D, Wharton S, Rotondi MA, Ardern CI, Kuk JL. Predictors of early attrition and successful weight loss in patients attending an obesity management program. BMC Obes 2016;3:14 [FREE Full text] [doi: 10.1186/s40608-016-0098-0] [Medline: 26966544]
66. Van der Mispel C, Poppe L, Crombez G, Verloigne M, De Bourdeaudhuij I. A self-regulation-based eHealth intervention to promote a healthy lifestyle: investigating user and website characteristics related to attrition. J Med Internet Res 2017 Jul 11;19(7):e241 [FREE Full text] [doi: 10.2196/jmir.7277] [Medline: 28698168]
67. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
68. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
69. Hernández-Rodríguez JC, García-Muñoz C, Ortiz-Álvarez J, Saigí-Rubió F, Conejo-Mir J, Pereyra-Rodriguez J. Dropout rate in digital health interventions for the prevention of skin cancer: systematic review, meta-analysis, and metaregression. J Med Internet Res 2022 Dec 09;24(12):e42397 [FREE Full text] [doi: 10.2196/42397] [Medline: 36485027]
70. Peels DA, Bolman C, Golsteijn RHJ, De Vries H, Mudde AN, van Stralen MM, et al. Differences in reach and attrition between Web-based and print-delivered tailored interventions among adults over 50 years of age: clustered randomized trial. J Med Internet Res 2012 Dec 17;14(6):e179 [FREE Full text] [doi: 10.2196/jmir.2229] [Medline: 23246790]
71. Rajabi Majd N, Broström A, Ulander M, Lin C, Griffiths MD, Imani V, et al. Efficacy of a theory-based cognitive behavioral technique app-based intervention for patients with insomnia: randomized controlled trial. J Med Internet Res 2020 Apr 01;22(4):e15841 [FREE Full text] [doi: 10.2196/15841] [Medline: 32234700]

72. Poppe L, De Bourdeaudhuij I, Verloigne M, Shadid S, Van Cauwenberg J, Compernolle S, et al. Efficacy of a self-regulation-based electronic and mobile health intervention targeting an active lifestyle in adults having type 2 diabetes and in adults aged 50 years or older: two randomized controlled trials. J Med Internet Res 2019 Aug 02;21(8):e13363 [FREE Full text] [doi: 10.2196/13363] [Medline: 31376274]
73. Fuchs K, Meusburger D, Haldimann M, Ilic A. NutritionAvatar: designing a future-self avatar for promotion of balanced, low-sodium diet intention: framework design and user study. 2019 Presented at: Proceedings of the 13th Biannual Conference of the Italian SIGCHI; September; Italy p. 1-10. [doi: 10.1145/3351995.3352044]
74. Moshe I, Terhorst Y, Philippi P, Domhardt M, Cuijpers P, Cristea I, et al. Digital interventions for the treatment of depression: A meta-analytic review. Psychol Bull 2021 Aug;147(8):749-786. [doi: 10.1037/bul0000334] [Medline: 34898233]

Abbreviations

HAPA: health action process approach

Edited by A Kushniruk; submitted 16.04.23; peer-reviewed by A Hannawa, W Zhang, M Gasmi ; comments to author 10.10.23; revised version received 31.01.24; accepted 20.02.24; published 26.04.24.

Please cite as:
Kötting L, Anand-Kumar V, Keller FM, Henschel NT, Lippke S
Effective Communication Supported by an App for Pregnant Women: Quantitative Longitudinal Study
JMIR Hum Factors 2024;11:e48218
URL: https://humanfactors.jmir.org/2024/1/e48218
doi:10.2196/48218
PMID:38669073

Original Paper

Comparing Attitudes Toward Different Consent Mediums: Semistructured Qualitative Study

Xengie Doan[1], MS; Arianna Rossi[2], PhD; Marietjie Botes[1,3], LLD; Annika Selzer[4,5], PhD

[1]SnT, University of Luxembourg, Esch-sur-Alzette, Luxembourg

[2]LIDER Lab, DIRPOLIS, Scuola Superiore Sant'Anna, Pisa, Italy

[3]Department of Medicine, Stellenbosch University, Stellenbosch, South Africa

[4]Fraunhofer Institute for Secure Information Technology, Darmstadt, Germany

[5]ATHENE, National Research Center for Applied Cybersecurity, Darmstadt, Germany

Corresponding Author:
Xengie Doan, MS
SnT
University of Luxembourg
2 place de l'Université
Esch-sur-Alzette, 4365
Luxembourg
Phone: 352 46 66 44 1
Email: xengie.doan@uni.lu

Abstract

Background: As consent for data sharing evolves with the digital age, plain-text consent is not the only format in which information can be presented. However, designing a good consent form is highly challenging. The addition of graphics, video, and other mediums to use can vary widely in effectiveness; and improper use can be detrimental to users.

Objective: This study aims to explore the expectations and experiences of adults toward consent given in infographic, video, text, newsletter, and comic forms in a health data sharing scenario to better understand the appropriateness of different mediums and identify elements of each medium that most affect engagement with the content.

Methods: We designed mock consent forms in infographic, video, text, newsletter, and comic versions. Semistructured interviews were conducted with adults who were interviewed about their expectations for consent and were then shown each consent medium and asked about engaging elements across mediums, preferences for consent mediums, and the value of document quality criteria. We transcribed and qualitatively co-coded to identify themes and perform analyses.

Results: We interviewed 24 users and identified different thematic archetypes based on participant goals, such as the Trust Seeker, who considered their own understanding and trust in organizations when making decisions. The infographic was ranked first for enhancing understanding, prioritizing information, and maintaining the proper audience fit for serious consent in health data sharing scenarios. In addition, specific elements such as structure, step-by-step organization, and readability were preferred engaging elements.

Conclusions: We identified archetypes to better understand user needs and elements that can be targeted to enhance user engagement with consent forms; this can help inform the design of more effective consent in the future. Overall, preferences for mediums are highly contextual, and more research should be done.

(JMIR Hum Factors 2024;11:e53113) doi:10.2196/53113

KEYWORDS

consent; transparency; data governance; visualization; health data sharing

Introduction

Overview

Consent is a cornerstone of ethical research, allowing people to be informed about the risks and benefits of research and demonstrate their autonomy. Consent has been discussed since the Nuremberg trials and takes on a pivotal role throughout European Union (EU) regulations for data protection, such as the General Data Protection Regulation (GDPR), but there are still challenges as bioethical consent and data protection consent collide. Digital decision-making about one's own data can be influenced or misled through interface design choices (ie, through so-called dark patterns [1-8]), while the consent experience of most European users corresponds to nagging cookie consent requests with profiling and advertisements that induce consent fatigue while trying to access a needed service [8]. Decades of research in the biomedical domain show that study participants' consent can rarely be deemed actually informed [9], often due to the complexity of language [10] and lack of health literacy [11], as well as the lack of data literacy of the individuals [12].

Engaging individuals in a user-friendly consent experience is thus fundamental to enabling them to meaningfully and freely make decisions with a sense of satisfaction [10] and agency. Improving the readability and comprehensibility of consent notices is one aspect of this, but research is also being done to explore visual communication techniques. Current research often focuses on the effect of multimedia on understanding [11,12], which can have a varied effect based on different studies. Multimedia also spans many formats, and most studies reviewed for their effect on understanding compared 2-3 different formats [12]. The Article 29 Working Party also refers to visual design means, such as "cartoons, infographics, flowcharts," to enhance the comprehensibility of information, and specifically to "comics/cartoons, pictograms, animations" [13]. However, they do not offer further guidance about what mediums to use and for what purpose (eg, how one might prioritize skimming, while another might be better for complex information). Therefore, we experimented with 5 different mediums of consent in this study, building on studies researching the use of a comic [14,15], video [16], infographic, and illustrated text [17], with plain text as a control [18].

In this paper, we substantially built on our previous work [19] by analyzing more mediums beyond the comic and infographic and specific engaging elements. The study presented a fictional scenario with a data trustee who would assist organizations (eg, research institutions, hospitals, etc) in finding suitable participants for clinical trials in a privacy-friendly manner. Participants were given a scenario where they were individuals who may benefit from a clinical trial organized by a hospital, so the data intermediary requested their consent to share their contact information with the hospital.

The objective of this study was to better profile user expectations and their attitudes toward different consent mediums, which included infographic, video, text, newsletter, and comic versions. We specifically analyzed how different elements of consent mediums (eg, narrative, color, and audio) affected participant engagement to survey the different affordances of each medium. Each medium has its own strengths and weaknesses in representing various kinds of information and can achieve various informational goals (eg, the video is low effort but can be skimmed, while the text can be skimmed but boring) [20]. As we intended to understand whether there are benefits to using one medium over another and why participants would prefer different mediums, we compared multiple mediums in this study based on semistructured interviews and dived into participant motivations, expectations, and experiences.

The results hinted at diverse goals among participants. We also identified the elements of document design that make the information concise, structured, and appropriate for the audience. We also found a large influence of context (eg, cookie consent or consent with different trusted institutions) on participant perceptions and expectations. Thus, we offer recommendations on how to better design consent documents to address different general participant profiles using layering and to engage the audience more effectively with a suitable medium. This has a pivotal role in the digital health data sharing space to give more effective transparency to participants who are deciding whether to share sensitive data. Our results can be leveraged by designers of digital consent experiences for more efficient multimedia use.

Background

Consent and Transparency

The European data strategy [21,22] aims to create a single market for data to allow for the free flow of data to benefit businesses, research, and public administrations within the EU. It is built on the GDPR, which aims to give users more control over their personal data.

Informed consent (IC) is a legal requirement specified in the GDPR as "freely given, specific, informed and unambiguous" (article 4(11)); easily withdrawn (article 7(3)); presented in an intelligible and easily accessible form using clear and plain language (article 7(2)); explicitly given for biomedical and genome data categorized as sensitive data (article 9); transparent in terms of completeness, comprehensibility, and accessibility of the information disclosures (articles 12, 13, and 14); and compliant with the principles of data protection by design and by default (article 25) [23]. IC requires user-centric design elements in consent to help achieve the general principle of transparency, which encompasses the "quality, accessibility, and comprehensibility of the information" [14]. The GDPR also contains obligations for "transparency by design" wherein privacy and consent notices should be purposefully designed to adequately inform the intended audience [24]. In addition, the GDPR also refers to other visual design methods like comics, videos, and infographics.

However, most existing informed decision-making solutions fail to reconcile theoretical demands with actual transparency. Conventional data privacy communication is characterized by lengthy, off-putting walls of complex jargon that impact the readability, comprehensibility, navigability, and memorability of information [20]. In addition, it is often standard, vague, or boilerplate instead of customized to the different needs and abilities of the intended audiences [25] and the type of data and processing activity. Reaching beyond plain language, in the last few years, there has been a renewed attention (and quite some experimentation) toward legal document design criteria [26] that more holistically relate to the language, writer-reader relationship, information design, and content.

Profiling User Needs Using Archetypes

Human-computer interaction research has used the persona technique (wherein imaginary users are assigned different profiles or personas with different goals and personalities based on demographic data) to better understand different users and needs and design suitable solutions [27]. However, it is a lengthy process that is often used for designing IT systems, not the consent processes. User need assessments have been conducted in relation to different demographics in health studies, but rather than focusing on the IC aspect, they focus on the health symptoms and how to address specific health-related needs [18,28,29]. Beyond health-related needs, we are interested more broadly in how the general adult population would interact with consent process to share information for downstream health reasons and what elements would be engaging when making informed decisions. This aspect has not been studied, to the best of our knowledge, but would be important for understanding how to strategically create effective information disclosures. Thus, we wanted to create archetypes, which capture general profiles, instead of personas, which are representations of imaginary individuals with specific population characteristics.

Multimedia Tools and Engagement With Digital Consent

The digitalization of data collection and use authorization allows for multimedia tools to be used during the consent process, which can have a positive outcome for participants. Overall, a systematic review of multimedia consent with videos, interactive programs, so on for surgical procedures found increased patient satisfaction for usability and informational availability [30]. However, for clinical trial consent, videos did not improve understanding [31]. Diving into the reasons that multimedia consent may be preferred to conventional text, one study compared animated videos, slideshows with voice-overs, comics, and text consent for medical practices and found that a dual-channel approach combining audio with visuals helped participant understanding [31]. This study supported older research showing that repetition of information using different multimedia means increases retention [32]. However, the specific elements of videos, comics, and text that contributed to effective communication in more general health consent were not studied—a gap that we intend to bridge with our work.

Even in other domains, studies strive to understand how to achieve effective communication of complex information by analyzing participant engagement, understanding, and recall of the information [33]. In the study by Wang et al [33], engagement refers to the time spent and fun experienced reading a form; and infographics, illustrated text, and data comics of complex economic data were tested. They found that students from different countries (aged from 18 to 35 years) preferred data comics, as they enable the greatest understanding, engagement, and enjoyment of all mediums, while the infographic performed best in esthetics and exploration, and the illustrated text performed the worst. As similar studies had not been performed on consent forms in a health scenario, we sought to study engagement as a factor of effective communication, as it might help understand what gains and retains attention within a complex digital attention economy.

Traditionally, engagement studies in biomedical consent refer to patient engagement with the research or biomedical process. Such engagement refers to participants interacting with the results of a study, updating information, or changing consent [34-36]. However, we are interested in participant motivations to consume the information in a consent form and give their initial and continued attention to a conventionally tedious process while competing in an attention economy [37,38]. Can consent forms be interesting and attention-grabbing?

Research Questions

The previous section has gathered evidence about the interplay between GDPR transparency requirements in data protection, the use of archetypes, and multimedia tools to enhance the experience. However, we lack an understanding of user needs, the impacts of different mediums on the user experience, and user engagement. The research questions that this study sought to answer are shown in Textbox 1.

Textbox 1. Research questions.

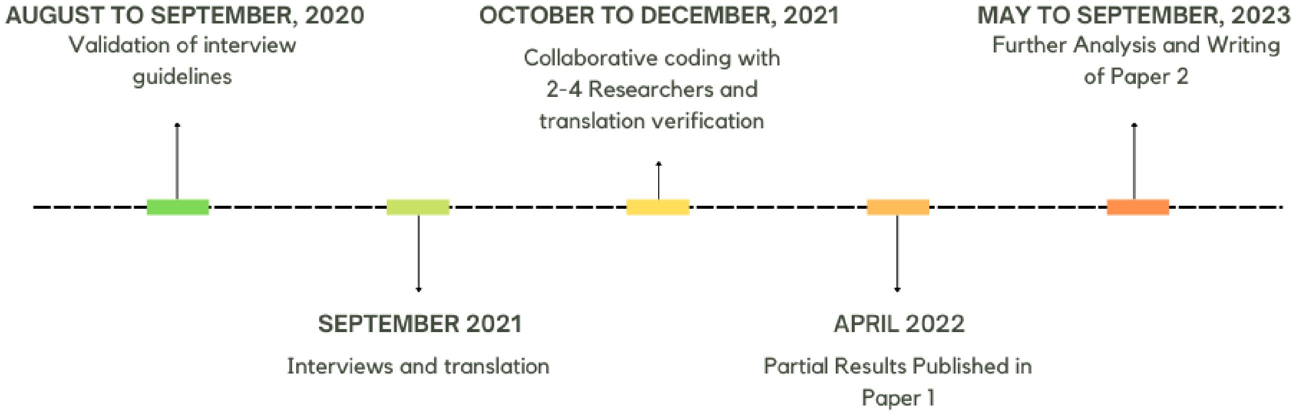
1. What kind of goal-oriented archetypes can be created to better understand participant needs for consent?
2. Across the 5 analyzed mediums (ie, infographic, video, text, newsletter, and comic forms), what were the participant rankings of different engaging elements?
3. After exposure to the consent mediums, we asked the following questions:
    a. What were the participants’ rankings of consent mediums?
    b. What elements reportedly influence their preference for mediums?
    c. What document quality criteria concerning language, design, content, and relationship with the reader did participants value?

Methods

Overview

AS carried out 24 semistructured interviews in September 2021 in Germany (Figure 1). We created an interview guideline (Multimedia Appendix 1), which was validated with 3 potential participants to ensure clarity, comprehensibility, and precision of the questions.

Figure 1. A time line of key activities and elaboration of results.



Recruitment

We searched for 24 participants by word of mouth. The demographic included adults from a cross-section of the German adult population by age, sex, and education level (Table 1).

Table 1. Participant demographics (N=24).

Characteristics	Participants, n (%)
Age range (y)	
18-30	8 (33)
31-55	8 (33)
56-90	8 (33)
Sex	
Male	12 (50)
Female	12 (50)
Highest degree	
School leaving or apprenticeship	12 (50)
College or university	12 (50)

The sample size and participant characteristics were based on a systemic review of unbiased citizens' juries for health policies [39]. Within the age ranges, there was an equal distribution of men and women with the highest degree obtained. All participants were native German speakers and lived in Germany, and the interviews took an average of 60 to 75 minutes.

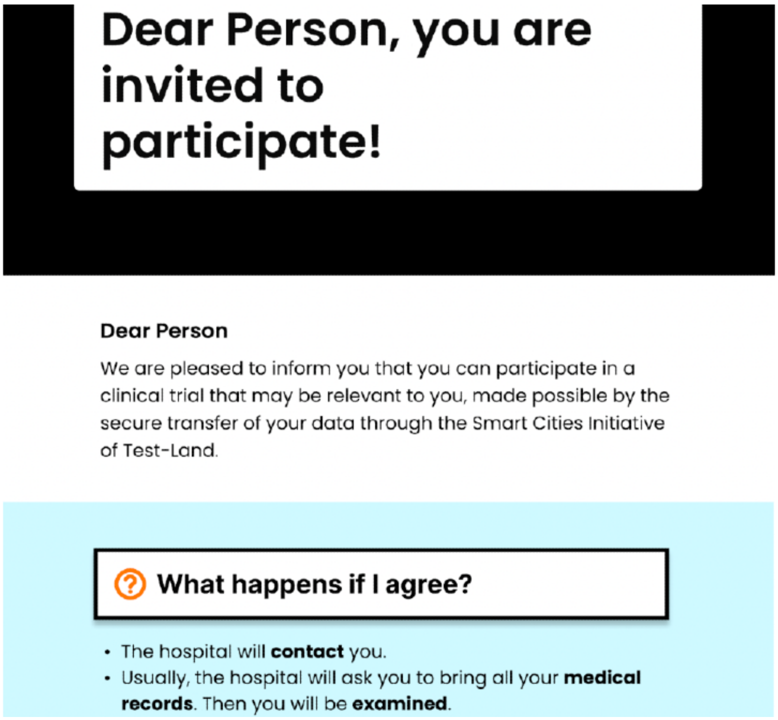
Study Material

AS created an example plain text document that asked for consent for the transfer of personal data from an intermediation service to another organization, a hospital. On the basis of the plain text, XD designed 4 additional variations in different mediums: an infographic, a comic, a newsletter, and a video. These 4 variations only included the subsection "What happens if you agree?" of the consent form. All 5 consent forms (ie, plain text, infographic, comic, newsletter, and video forms) differed in design, but the core consent text was the same across all mediums. XD followed best practices for information transparency, designed documents for each medium with different subsets of engaging elements, and adapted them for the mediums (ie, additional ellipses between comic text) for the purposes of the study, consulting coauthors during design. (a) The video (Multimedia Appendix 2) was created to test the use of color, audio, and animations and illustrate the text using free resources on Biteable website (eg, "a doctor will call you" conveyed as an animation of a waving physician). The audio was provided by AS (a native German speaker) out of convenience. (b) The infographic (Figure 2) was designed with a step-by-step format and color from a health template on Canva (Canva Pty Ltd), with icons describing the text (eg, scheduling an appointment had a calendar icon; Multimedia Appendix 3). (c) The comic (Figure 3) used a story element and color and was designed in Figma with input from all authors, and it used simple figures to expedite the creation of the comics. The drawings sought to describe the text as literally as possible (eg, "you will be contacted" depicts a ringing phone; Multimedia Appendix 4). (d) The newsletter (Figure 4) used open format and color and was created in Figma (Figma, Inc; a popular website for user interface or user experience design) based on an existing newsletter template's structure. The newsletter was thought to be a more familiar medium with more graphics than text (eg, newsletters sent via email; Multimedia Appendix 5).

Figure 2. A translated section of the infographic study material designed with a step-by-step format, color, and structured sections.

Figure 3. A translated section of the comic study material designed with a story and color.

Figure 4. A translated section of the newsletter study material designed with an open format, color, and structured sections.



Study Design

All interviews took place via a web conferencing system. No recordings were taken, and a summary transcription was written after each question and finalized after each interview. This method was chosen to respect participant anonymity and COVID-19 protocols.

The interviewer invited participants to imagine that they were contacted by a data intermediary to obtain their consent to share their name, email, and allergy information with a hospital that wanted to carry out a clinical trial for lactose allergies. A verbal explanation was given along with the full plain text version of the consent form, and participants could ask questions at any time.

To answer RQ1, participants were asked about their previous experience with consent forms and desires regarding consent.

To address RQ2, after participants were shown all mediums of consent, they were asked to rank 8 design elements of a consent form: the use of colors, audio, animated elements, readability of text (eg, if it is not too technical or complicated), story element (eg, using examples and people in the forms), structured sections, step-by-step elements (eg, having an order to the information with text or visuals), and an open format (eg, being able to skip around to sections) from the most to the least engaging with an option for “other.”

To answer RQ3, we showed them a subsection of the full consent form, “What happens if I agree?” in different mediums (ie, comic, infographic, plain text, newsletter, and video versions) in a random order per participant. Participants were asked to rank the different forms according to their preferences and clarify why.

Data Analysis

The interviews were documented in German, and anonymized answers were translated into English via DeepL (DeepL SE) and proofread by AS to ensure the translations’ adherence to the original meaning and to collaboratively analyze them with XD, AR, and MB (all non-German). Translation verification continued throughout the qualitative coding process in various sessions from November 2021 to April 2022 with the multidisciplinary team. To code the interview, the software MAXQDA (VERBI GmbH) was used. The expertise of the coding team spanned data protection law, usable privacy, bioethics, bioinformatics, and legal design.

To code the interviews, we inductively and iteratively established a codebook over three 2-hour sessions (Multimedia Appendix 6). The codebook combines a bottom-up approach through analysis of the data (eg, the concept of trust stemming from participant answers) with a top-down approach derived from the criteria for good documents given by Waller [26] (Table 2) to answer RQ3 (c).

Table 2. Document quality criteria elaborated by Waller [26].

Criteria	Description
Language	
Directness	Using direct language to make it clear who is acting
Plain words	The extent to which the vocabulary is easily understood
Grammar	Conformity with good standard English practices
Readability	Ease with which the reader can follow arguments
Design	
Legibility	Use of legible fonts and text layout
Graphic elements	Use of tables, bullet lists, graphs, charts, icons, etc
Structure	Quality of document organization for function
Impression	Attractiveness and approachability, overall appearance
Relationship	
Who from	Is it clear who is communicating?
Contact	Whether there are clear contacts or means of contact
Audience fit	Appropriateness to the knowledge and skills of users
Tone	Matching the style and language of the context
Content	
Relevance	How relevant is the content to the recipient?
Subject	If it is clear what the communication is about
Action	Clarity about what action is required of the user
Alignment	Compliance with the organization's intended aims and values

Participant consent expectations have been organized into archetypes depending on the salience of reported goals and relevant features. A matrix was created with the participant number, expected features, expected goals, and expected behaviors to help group similar profiles.

Ethical Considerations

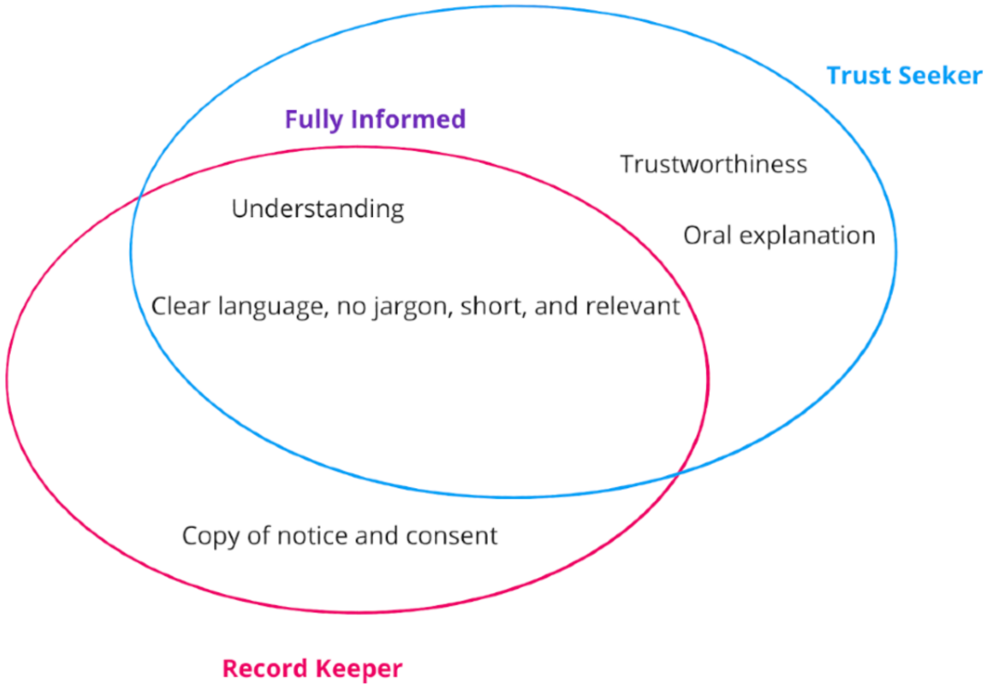
The study design has been authorized by the Research Ethics Committee of the University of Luxembourg (ERP 21-038 LeAds), and best practices were followed. We chose a summary transcription to enable easier anonymization of the interview. Once manually anonymized, transcripts were securely shared with the authors from the other organization. The interviewees were compensated €30 for their time.

Results

User Desires via Archetypes

The interview findings from the questions, which explore participant expectations, desires, and needs, have been organized into 3 goal-oriented archetypes: the Fully Informed, the Record Keeper, and the Trust Seeker. Not all participants reported specific goals, while some participants reported multiple goals. Thus, the archetypes are based on grouping similar features (Figure 5).

Figure 5. Venn diagram describing the core goals of different consent archetypes.



The Fully Informed archetype wanted relevant and fitting information to understand what they were consenting to. This aligns with the most common goal explicitly reported by participants (14/24, 58%):

> *[A]s an affected person, I would like to see a few examples to get a better understanding of what may be done with my data.* [P1]

The information must also be appropriate for them as an audience:

> *A simple explanation that everyone understands would be my preference.* [P12]

The Record Keeper sought understanding while specifically wanting to remember what they had agreed to (3/24, 13%) or to have a copy for their records (4/24, 17%). For example, participant 13 had a clear idea of the elements they wanted to understand and retain a clear memory of:

> *It needs to be clear to me what the consent is for, who it is from, and exactly what data is being processed for what purpose.* [P13]

In addition, participant 4 stated the following:

> *It doesn't matter to me if it is paper or digital. The main thing is that I receive a copy of the text to which I have consented.* [P4]

The Trust Seeker also sought understanding but was cautious toward the system or desired a trustworthy system (6/24, 25%):

> *I must have the impression that the data trustee is a reliable company or that there is an expertise that proves that I can trust this data trustee.* [P3]

> *[I would rather avoid] to invest time and read through stuff...[and be able to trust] since I've already given my data...that my data will just be handled well.* [P7]

When considered together, the archetypes lie on a spectrum where the Fully Informed archetype relies more on individual responsibility and capacity to make informed decisions, while the Trust Seeker also considers the context of organizational reputation and trust in making their decisions. In addition, the Record Keeper could be seen as an individual who wants to manage their consent decision over time, while those who do not want to review or revise their consent accept a one-time decision without records.

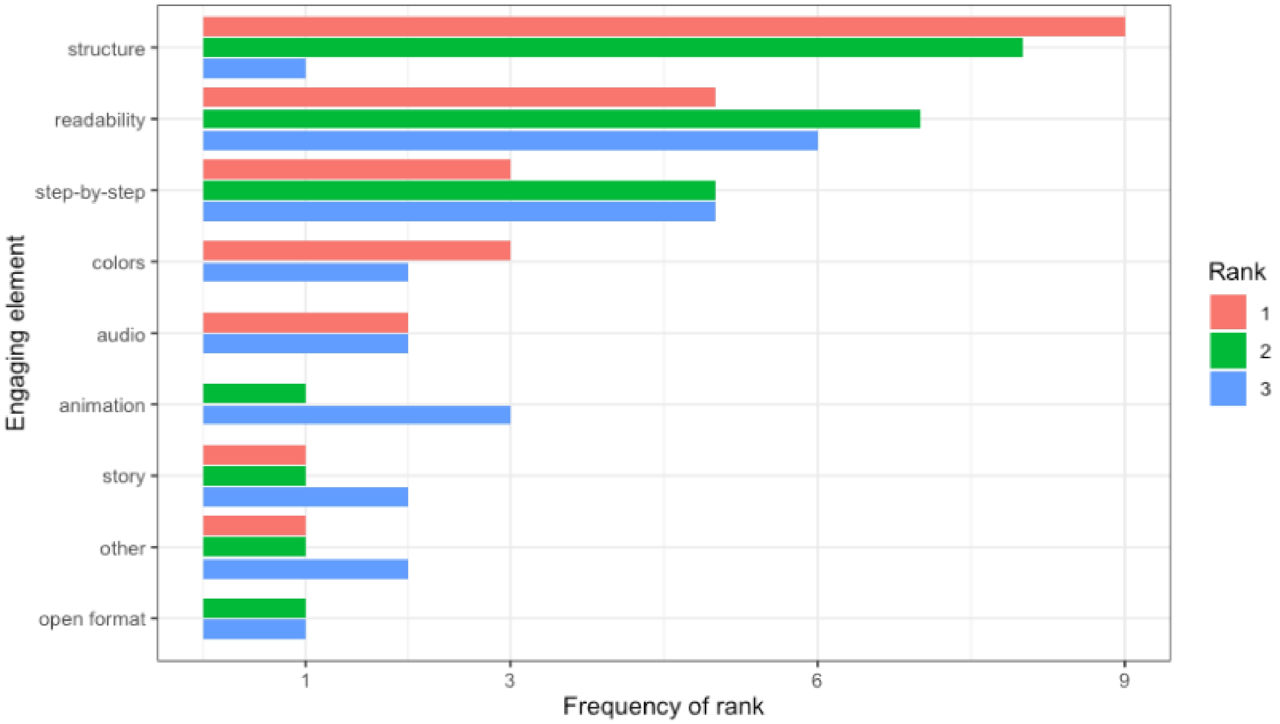
In addition to finding patterns based on common goals, some individuals stood out for their unique consent desires, including using more technical jargon (2/24, 8%). The use of jargon seemed to enable the process more time saving for some participants:

> *If I had to choose between short technical language and simple but longer language that is easy for everyone to understand, I would choose the short technical language.* [P15]

Top Engaging Elements

To better understand RQ2, about how different elements across mediums were perceived by participants, they were asked to rank the listed elements after experiencing all mediums. The most frequent element ranked first was structure, followed by readability, colors, step-by-step elements (tied with "colors"), audio, story, and others (also tied with "story"; Figure 6). The top element at rank 2 was also structure, and the top element at rank 3 was readability. When the option "others" was chosen, not all participants elaborated on what "other" element they referred to, but when they did, personal engagement (4/24, 17%) was most commonly cited. In ranks 2 and 3, structure, readability, and step-by-step were also frequently cited engaging elements.

Figure 6. The frequency of each engaging element ranked from 1 to 3 by participants.

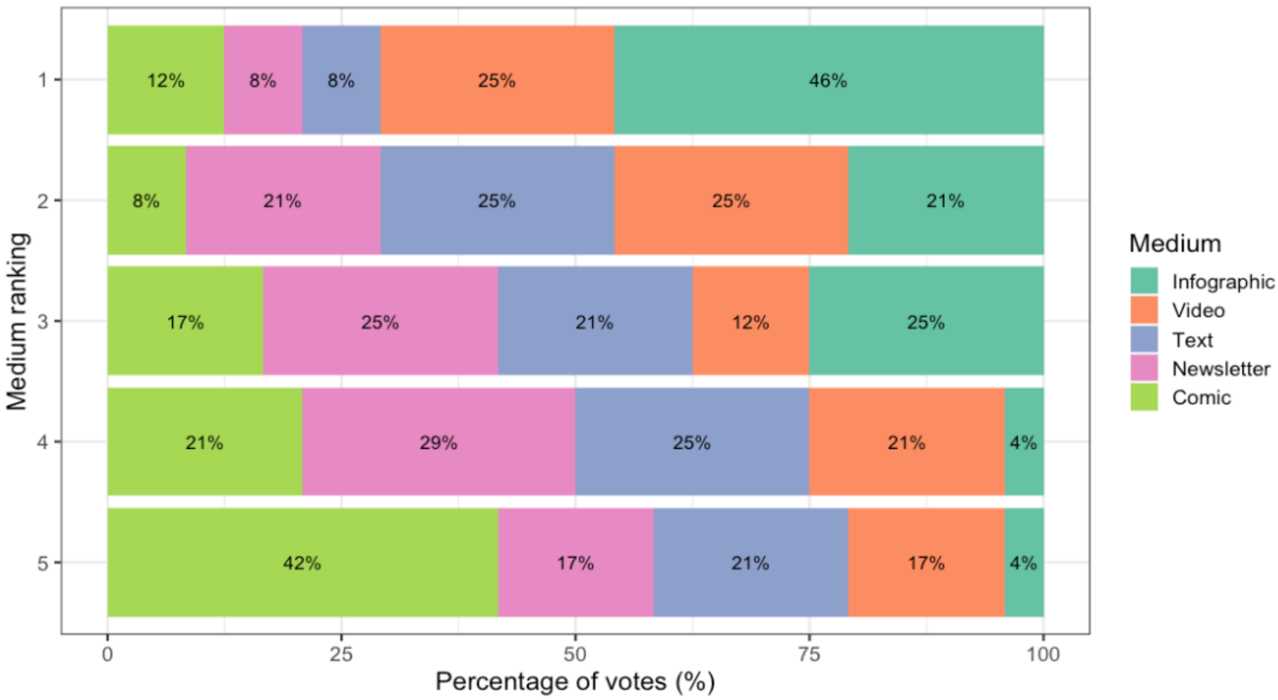


Preferred Mediums and Document Criteria

Overview

To answer RQ3, we report the results about participants' ranking for their preferred consent form after being shown each medium in Figure 7. The infographic was the overall winner and the comic the overall loser, while the video, text, and newsletter had varying trends (eg, the text was uniformly distributed across ranks 2 to 5, while the video was most often ranked 1, 2, or 4). Interestingly, no medium had consensus across the 24 participants.

Figure 7. Participant ranking of mediums (where 1 corresponds to the first choice and 5 to the last choice) by percentage of votes.



In the following sections, each medium is discussed based on (1) the top 3 factors that influenced the ranking and (2) the top 3 positive or negative document criteria adapted from good document criteria by Waller [26] (Table 1). A participant could share multiple influencing factors or document criteria. We instead looked at the number of unique coded segments within their answer. Participants could share as few (though they were prompted to try to give at least 1) or as many factors as they desired. The coded segments for influencing factors, such as the element of time, could be positive (time-saving) or negative (time-wasting). This was to help us identify the categories that were most important to participants. Then, we contextualize the data and report whether important factors were positive or negative and their respective coded segments volume in the detailed section. Finally, Table 3 offers a summary of rank, top influencing factors, and document criteria per medium.

Table 3. Overview of the top 3 influencing factors and document criteria per medium with overall participant ranking.

	Medium				
	Infographic (rank 1)	Video (rank 2)	Text (rank 3)	Newsletter (rank 4)	Comic (rank 5)
Influencing factors					
Positive element	Understanding, time, and interest	Understanding and effort	Understanding and time-saving	Prioritization and understanding	Understanding and interest
Negative element	N/A[a]	N/A	Boring	Association with advertisements	Inappropriate fit for the context
Mixed	N/A	Time-saving and time-wasting	N/A	N/A	N/A
Document criteria					
Positive element	Numbered lists, icons, bold headings, and graphic elements	Audio, step-by-step elements, and interplay of text and graphics	Structured layout	Bolded key text, sections, and open format for skimming	Text and graphics
Negative element	Extraneous or leading icons	N/A	Lacked highlighting	Advertisement impression	Tone and audience fit

[a]N/A: not applicable.

Infographic Medium

The infographic was strongly preferred, with one-third of the participants (15/49, 31%) citing understanding as a positive influencing factor, while time and interest were in a close second. Elements such as numbered bullet points, bold headings, and icons were referenced:

> *With the bullets, you know right away what each is about in the text written underneath. In general, this is easy to grasp...* [P3]

The top 3 influencing factors were overall positive, in contrast to the other mediums.

Most positive document criteria concern design criteria such as step-by-step elements, icons, bold headings, bullet points, and color. There were much fewer negatively received elements, which were also related to the design criteria: the overuse of color and icons and the large size of the infographic. Participants had specific reactions to different icons, such as the hospital or medical professionals at the top and bottom that did not support any text or specific icons that might seem manipulative:

> *[W]ith the consent form, the “thumbs up” graphic makes it look like I’m being preempted from making a decision.* [P14]

Video Medium

Approximately one-third of the participants (14/44, 32%) reported that it influenced their understanding, followed by time and effort. Understanding was largely positive, partially due to the format that they’re “forced to watch it from beginning to end, so that you perceive the whole content” (P15).

On the other hand, time was slightly more positive than negative because while most participants felt that compared to reading, the video saved time, some felt it was inefficient compared to their reading speed, or they wanted to review material but felt rewinding would be time-wasting. Saving effort was wholly positive, with participants saying that it was more accessible, entertaining, or less attention draining while still being understanding. One minor interesting influencing factor unique to the video was a perceived feeling of trust from the audio, with 2 participants mentioning that a human voice engendered confidence in the process.

More than half of the positive feedback about the video mentioned the audio element, followed by the sequential nature and use of animation and images. Less than one-fourth of the participants (12/54, 22%) liked the content, which included the interplay between text and graphics and the story element:

> *What I like about the video is that...you see movements that show what you hear at the same time via audio.* [P3]

There were about half as many negative elements as positive ones, and most were due to the video pacing. Some wanted it faster, while others wanted it slower. Interestingly, 1 participant noted the following:

> *I have the feeling that with a video like this, people are rather uncritical of the content of the consent form. One is rather tempted to agree to something. If, for example, a button appeared after the video that allowed me to consent, I would probably consent.* [P17]

Text Medium

Approximately one-third of the participants (12/42, 29%) indicated that interest and understanding were most influenced by the text. Interest was a complex influencing factor that was slightly more negative. Those stating that it negatively influenced attention felt that the text medium was boring or lacked interest compared to other mediums. The participants who viewed it positively said that the text had a simple, clean layout allowing for quick skimming, and those who felt it was neutral felt like participant 21:

> *This is the format that I know and have simply accepted by now.* [P21]

The understanding was generally a positive influencing factor, with many saying that it was clear, concise, and short; however, some felt that it was difficult to skim or that the text was confusing or dry. Some participants also felt that it saved time by being short and concise.

The most cited positive document elements of the text were the use of clear sections, headlines, and bullet points. The positive elements were twice as common as the negative elements; the majority of both positive and negative elements also stemmed from design. Participants wanted more highlighting of key facts via colored, bold, italicized, or underlined words. Less than one-third of the participants (5/18, 28%) also cited the negative impression the document gave them:

> *[I]t is still a bit boring and trivial, so you might not read it properly if you get it as a letter home, for example.* [P5]

Newsletter Medium

More than a quarter of participants (6/22, 27%) mentioned prioritization as an influencing factor, less than a quarter (5/22, 23%) mentioned understanding, and 18% (4/22) mentioned interest. Prioritization and understanding were positive influencing factors, with participants saying that the bold words and ability to skip sections allowed them to roughly understand the contents because the bold text highlighted the important information in sentences. However, the interest factor was equally mixed, with the positive influence surrounding the bolded text and headers, while the negative influence was mainly attributed to the association with advertising spam. More than half of the participants (13/24, 54%) agreed with participant 1, who stated the following:

> *It looks like advertising, by the structure and the "headline," which is repetitive.* [P1]

Although it had positive influencing factors, the negative interest likely had a large impact on the lower ranking of this form, as participant 5 said the following:

> *[I]t looks like advertising and I don't like that.... I am rather annoyed by it. The bold as highlighting and the textual design I find good.* [P5]

The newsletter's positive elements were largely regarding the design and use of structure, headings, bold text, sectioning, and the open format for skimming. The negative elements also similarly mentioned the design criteria because it looked like advertising based on prior experiences. The use of color was also disliked because the black header was too strong and off-putting.

Comic Medium

The main influencing factor was understanding, with one-third of the participants (7/24, 29%) mentioning it both positively and negatively. A slight majority (4/7, 57%) cited a positive influence on understandability. Interest was generally a positive influencing factor because the medium was novel. Less than one-fourth of coded segments showed that the comic had an overall negative influence on skimming, as the narrative-driven step-by-step format made it difficult to prioritize, reread for specific elements, or gain a quick overview. In addition, many participants disliked the comic medium as a whole, even if they could find some helpful design elements:

> *I found the comic a bit inappropriate for the topic...the message is better visualized by the little pictures, which may be better remembered but I don't like it.* [P20]

Almost half of the positive feedback for the comic stemmed from the support of text with graphics, narrative elements, and illustrations. A third felt that the tone and audience fit suited them. However, negative impressions were almost double the positive ones because audience fit and tone were unsatisfactory for more than half of the participants (14/24, 58%):

> *I'm out of the age where I still like comics.... I don't feel like I'm being taken seriously as a customer with a consent form like this.* [P16]

Participants suggested that children or older adults might be a better audience fit. Other negative feedback arose from the impression and graphic elements concerning the execution of illustrations, legibility, and lack of structure.

Discussion

Principal Findings

Qualitative analysis of participant desires for health consent revealed 3 archetypes: the Fully Informed, the Record Keeper, and the Trust Seeker. All participants wanted a high level of understanding before the consent decision, with some valuing additional elements such as obtaining copies of their decisions for their records and the trustworthiness of institutions like hospitals. The participants greatly stressed the need for short, concise, and direct consent forms that should not be longer than a page. Our results support the results of other authors, who have found that participants want to skim consent forms because consent documents are all the same, they want to save time, or they trust the ethical review of the related study [40]. In other words, individuals often engage in a form of strategic reading [26] instead of relying on attentive reading. This is why consent should contain elements that allow the visual prioritization of certain content over others, like headings, bullet points, and highlights (ie, "surface-level cues") [26] that allow individuals to skim the document effectively and discern the most important information at first sight.

On the basis of the ranking of engaging elements, the participants preferred step-by-step documents (eg, linearly numbered lists with clear headings) instead of open or story-based formats. Structure, readability, and step-by-step elements were the top 3 engaging elements and could be easily integrated into most mediums. While our study only designed the infographic using 4 of the top engaging elements (ie, structure, readability, color, and step-by-step element), other mediums like text could also use color and step-by-step elements instead of the open-format element. However, the tone and audience fit of mediums greatly influenced participant rankings, even if some mediums enhanced understanding or visual interest (eg, comics and newsletters). The negative connotations of the newsletter with marketing and comics with childishness contributed to their low rankings, while text was seen as routine and acceptable, if boring. Instead of prioritizing one medium over another, there could be a greater focus on including the most important engaging elements within mediums (eg, adding step-by-step elements in all possible mediums).

Implications for Practice

First, the creation of data-informed archetypes can be used for better understanding, and therefore accommodating, the diverse needs of a population. To leverage information describing the use of one's own personal data as a self-determination instrument, individuals can receive contextualized information and concrete examples that are relevant to their specific needs (eg, the Fully Informed and Trust Seeker archetypes), rather than one-size-fits-all terms. Archetypes also support general audience tailoring for different goals. Different approaches to consent notices may reflect strategies to cope with the 2-fold reality stemming from the fact that the risks of consent decisions are individual, while the data sharing and processing are networked across the individual, responsible institutions, and beyond [41]. For example, the Fully Informed archetype may be more concerned with individual responsibility and the personal data processing, while the Trust Seeker wants information about the organizations involved and their security and privacy measures. Using archetypes to base user profiles could also be a way to customize their experience in meaningful macrocategories without needing to customize every possibility for individual preferences. However, more research is needed to balance the actual benefit of tailoring information to different learning styles [42] against the increased costs of its creation and implementation.

Second, different mediums can be targeted based on needed affordances (Table 3) and layered to reinforce the understanding of complex information, for example, through a multifold presentation of the same content through text, video, and infographics. Official guidance about transparency requirements' implementation [15] portrays layering techniques as an appropriate means to achieve the requirement of full disclosure while allowing for prioritization and brevity. For example, summaries containing an overview of the main clauses can accompany the more comprehensive version and can be more easily browsed while consenting, with short videos and privacy icons constituting the first layer of a written notice [20]. Distributing information on separate mediums can additionally contribute to presenting the relevant information at an appropriate time. For instance, the first layer with essential information can be displayed at the moment of the consent decision, while detailed information can always remain accessible on request [43]. However, as more guidelines for multimedia consent design arise [14,20], testing and co-designing with the intended audience is key; otherwise, a medium with a negative audience fit for certain contexts may be less effective than plain text consent (eg, the unsuitable comic medium for German adults). This can be important to test for among the intended audience, especially as comics have been a case study for cultural stigmas [44]. While they have been suitable for Indigenous populations [15], some researchers are pushing for more serious comics (similar to serious games for education) [45] and the comic co-design process itself as a research practice [46].

In terms of implementation, layering has been integrated with dynamic consent platforms. Dynamic consent was built to leverage the benefits of digital communication for health research by using digital platforms to connect people and researchers and allow participants to view, update, and change their data sharing permissions dynamically. Australia's CTRL [36], a dynamic consent platform based on open-source code, incorporates multimedia (video, illustrated text, and infographics); personalization options; and informational layering techniques. Building upon this, the layering could incorporate archetypes of general profiles to be tailored for different goals. Users of different ages may prefer different mediums, such as comics for younger audiences and videos for older audiences; similarly, users with domain expertise could choose content explained with jargon.

Finally, although we did not explicitly ask about undue influence of design elements on consent decisions and trust implications, participants clearly connected the 2, and more research is needed to better understand the deep connection. The infographic had a few complaints about specific graphics, with participant 14 saying that showing a "thumbs-up" icon was perceived as a manipulative way to preempt one from making an informed decision. Similarly, participant 17 stated that participants might believe anything shown in a video and be inclined to give consent. While guidance on ethical nudging design [47-49], as well as research on dark patterns that are to be avoided [7,8], can help shed light on such thorny issues–the issue should be more deeply studied. Considering how often human beings take decisions that are not completely rational [50,51], adding elements such as icons, color, or audio may increase the potential for manipulation of choice.

Limitations

Although we strove to obtain balanced age, education, and sex representation in our sample, they cannot be fully representative of the population. More research should be done on populations other than German adults with a larger sample size. It should also be replicated in the specific consent context of interest, as our study focused on consent to share personal data for further contact; replicating the study for clinical trial consent is important, as it may offer new archetypes, rankings, and contextual concerns. More research should also be done to study how to refine and apply archetypes in practice, as it can be insufficient or biased without continued user, expert, or patient input. Our methods only concern self-reported opinions, so there may be a discrepancy between reported preferences and observed behaviors. The study materials may have influenced rankings and preferences, as they were generated by XD, who is not a professional designer. Therefore, certain choices (eg, the simple comic style) could have influenced participants' attitudes, making it difficult to determine the exact stimulus based on self-reported answers. While out of scope of this work, future studies can research how specific design elements or stereotypes impact rankings, for example, showing a comic with stick figures or realistic figures to German adults to better understand how to design the specific element. Before implementing consent mediums in line with applicable constraints, relevant expertise should be included in the design and evaluation of each medium.

Conclusions

To better understand the diversity of participant preferences, opinions, and emotions for IC in a health care scenario and the relevance of specific document criteria for engagement with various mediums (ie, infographic, video, text, newsletter, and comic), this study interviewed 24 individuals. The results not only have informed the generation of archetypes based on desired document features and goals but can also help create standardized consent documents that use layering to help address varying needs identified via archetypes. We also proposed recommendations for designing multimedia consent forms with a structure that promotes prioritization, such as headers, bullet points, and bold type within a contextually appropriate medium, such as an infographic or a video, so that the forms are seen by our participants as more attention-grabbing and serious than comics. It would be important to replicate this study setting in other countries, and the results could lead to contextually designed consents that align with the GDPR and other EU regulations. The findings reported here are meant to encourage further research to determine how to better involve individuals in designing useful, engaging consent forms to facilitate informed decisions concerning data sharing.

Acknowledgments

XD and MB contribution to this research work was funded by the European Union's Horizon 2020 Innovative Training Networks, Legality Attentive Data Scientists under grant agreement ID 956562. AS's contribution to this research work was partially funded by the German Federal Ministry of Education and Research and the Hessian Ministry of Higher Education, Research, Science, and the Arts with the joint support of the National Research Center for Applied Cybersecurity, ATHENE. AS's contribution to this research work was conducted partially during a research visit at the University of Luxembourg in 2021. AR's contribution to this research was financially supported by the project DECEPTICON Deceptive Patterns Online, funded by the Luxembourg National Research Fund under grant agreement IS/14717072 while she was employed at the University of Luxembourg. As an employee of the Sant'Anna School of Advanced Studies, her contribution was supported by the Biorobotics Research and Innovation Engineering Facilities" project (project identification code IR0000036) funded under the National Recovery and Resilience Plan, mission 4, component 2, investment 3.1 of the Italian Ministry of University and Research funded by the European Union—NextGenerationEU.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guidelines in English.
[DOCX File , 130 KB - humanfactors_v11i1e53113_app1.docx]

Multimedia Appendix 2
Full video used in the study.
[MP4 File (MP4 Video), 18842 KB - humanfactors_v11i1e53113_app2.mp4]

Multimedia Appendix 3
English version of the infographic used.
[PDF File (Adobe PDF File), 1770 KB - humanfactors_v11i1e53113_app3.pdf]

Multimedia Appendix 4
English translation of the comic shown in the study.
[PDF File (Adobe PDF File), 2043 KB - humanfactors_v11i1e53113_app4.pdf]

Multimedia Appendix 5
English version of the newsletter shown in the study.
[PDF File (Adobe PDF File), 990 KB - humanfactors_v11i1e53113_app5.pdf]

Multimedia Appendix 6
Full codebook used in analysis.
[DOCX File , 31 KB - humanfactors_v11i1e53113_app6.docx]

References

1. Gray CM, Kou Y, Battles B, Hoggatt J, Toombs AL. The dark (patterns) side of UX design. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18; April 21-26, 2018; Montreal, QC. [doi: 10.1145/3173574.3174108]
2. Bauer JM, Bergstrøm R, Foss-Madsen R. Are you sure, you want a cookie? – The effects of choice architecture on users' decisions about sharing private online data. Comput Hum Behav 2021 Jul;120:106729. [doi: 10.1016/j.chb.2021.106729]
3. Santos C, Rossi A, Chamorro LS, Bongard-Blanchy K, Abu-Salma R. Cookie banners, what's the purpose?: analyzing cookie banner text through a legal lens. In: Proceedings of the 20th Workshop on Workshop on Privacy in the Electronic Society. 2021 Presented at: WPES '21; November 15, 2021; Virtual event. [doi: 10.1145/3463676.3485611]
4. Kampanos G, Shahandashti SF. Accept all: the landscape of cookie banners in Greece and the UK. In: Proceedings of the 36th IFIP TC 11 International Conference on Information Security and Privacy Protection. 2021 Presented at: 36th IFIP TC 11 International Conference on Information Security and Privacy Protection; June 22-24, 2021; Oslo, Norway. [doi: 10.1007/978-3-030-78120-0_14]
5. Utz C, Degeling M, Fahl S, Schaub F, Holz T. (Un)informed consent: studying GDPR consent notices in the field. In: Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security. 2019 Presented at: CCS '19; November 11-15, 2019; London, UK. [doi: 10.1145/3319535.3354212]
6. Matte C, Bielova N, Santos C. Do cookie banners respect my choice? : measuring legal compliance of banners from IAB Europe’s transparency and consent framework. In: Proceedings of the IEEE Symposium on Security and Privacy (SP). 2020 Presented at: SP 2020; May 18-21, 2020; San Francisco, CA. [doi: 10.1109/sp40000.2020.00076]
7. Graßl P, Schraffenberger H, Zuiderveen Borgesius F, Buijzen M. Dark and bright patterns in cookie consent requests. J Digit Soc Res 2021 Feb 08;3(1):1-38. [doi: 10.33621/jdsr.v3i1.54]
8. Nouwens M, Liccardi I, Veale M, Karger DR, Kagal L. Dark patterns after the GDPR: scraping consent pop-ups and demonstrating their influence. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20; April 25-30, 2020; Honolulu, HI. [doi: 10.1145/3313831.3376321]
9. Montalvo W, Larson E. Participant comprehension of research for which they volunteer: a systematic review. J Nurs Scholarsh 2014 Nov 15;46(6):423-431. [doi: 10.1111/jnu.12097] [Medline: 25130209]
10. Niemiec E, Vears DF, Borry P, Howard HC. Readability of informed consent forms for whole-exome and whole-genome sequencing. J Community Genet 2018 Apr 31;9(2):143-151 [FREE Full text] [doi: 10.1007/s12687-017-0324-6] [Medline: 28856579]
11. McBride CM, Bowen D, Brody LC, Condit CM, Croyle RT, Gwinn M, et al. Future health applications of genomics: priorities for communication, behavioral, and social sciences research. Am J Prev Med 2010 May;38(5):556-565 [FREE Full text] [doi: 10.1016/j.amepre.2010.01.027] [Medline: 20409503]
12. D'Ignazio C. Creative data literacy: bridging the gap between the data-haves and data-have nots. Inf Design J 2017 Jul 5;23(1):6-18. [doi: 10.1075/idj.23.1.03dig]
13. Ferrer RA, Stanley JT, Graff K, Klein WM, Goodman N, Nelson WL, et al. The effect of emotion on visual attention to information and decision making in the context of informed consent process for clinical trials. Behav Decis Making 2016 Mar 28;29(2-3):245-253. [doi: 10.1002/bdm.1871]
14. Guidelines on transparency under regulation 2016/679 (wp260rev.01). European Commission. 2018 Aug 22. URL: https://ec.europa.eu/newsroom/article29/items/622227 [accessed 2024-04-23]
15. Steenberg WM. Visual communication as a legal-ethical tool for informed consent in genome research involving the San community of South Africa. University of South Africa. 2017. URL: https://orbilu.uni.lu/bitstream/10993/48262/1/published.pdf [accessed 2024-04-25]
16. Botes MW, Rossi A. Visualisation techniques for consent: finding common ground in comic art with indigenous populations. In: Proceedings of the IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). 2021 Presented at: EuroS&PW; September 6-10, 2021; Vienna, Austria. [doi: 10.1109/eurospw54576.2021.00037]
17. Antal H, Bunnell HT, McCahan SM, Pennington C, Wysocki T, Blake KV. A cognitive approach for design of a multimedia informed consent video and website in pediatric research. J Biomed Inform 2017 Feb;66:248-258 [FREE Full text] [doi: 10.1016/j.jbi.2017.01.011] [Medline: 28109951]
18. Jimison HB, Sher PP, Appleyard R, LeVernois Y. The use of multimedia in the informed consent process. J Am Med Inform Assoc 1998 May 01;5(3):245-256 [FREE Full text] [doi: 10.1136/jamia.1998.0050245] [Medline: 9609494]
19. Doan X, Selzer A, Rossi A, Botes WM, Lenzini G. Context, prioritization, and unexpectedness: factors influencing user attitudes about infographic and comic consent. In: Proceedings of the Companion Proceedings of the Web Conference 2022. 2022 Presented at: WWW '22; April 25-29, 2022; Virtual event. [doi: 10.1145/3487553.3524632]
20. Rossi A, Ducato R, Haapio H, Passera S. When design met law: design patterns for information transparency. Droit de la consommation 2019;122-123(5):79-121.

21. European data strategy Making the EU a role model for a society empowered by data info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en 2020. European Commission. URL: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en [accessed 2024-04-16]
22. Shaping Europe’s digital future. European Commission. URL: https://digital-strategy.ec.europa.eu/en [accessed 2024-04-23]
23. Regulation (EU) 2016/679 of the European Parliament and of the council of 27 April 2016. Official Journal of the European Union. URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj [accessed 2024-04-23]
24. Rossi A, Lenzini G. Transparency by design in data-informed research: a collection of information design patterns. Comput Law Secur Rev 2020 Jul;37:105402. [doi: 10.1016/j.clsr.2020.105402]
25. Potel-Seville M, Talbourdet E. Empowering children to understand and exercise their personal data rights. Aalto University. URL: https://acris.aalto.fi/ws/portalfiles/portal/78736878/9788855265751_Legal_Design.pdf#page=254 [accessed 2024-04-29]
26. Waller R. What makes a good document? The criteria we use. University of Reading. 2011 Apr. URL: https://uploads-ssl.webflow.com/5c06fb475dbf1265069aba1e/6090fb24421b454c776fd1a8_SC2CriteriaGoodDoc_v5.pdf [accessed 2024-04-23]
27. Pruitt J, Grudin J. Personas: practice and theory. In: Proceedings of the 2003 Conference on Designing for User Experiences. 2003 Presented at: DUX '03; June 6-7, 2003; San Francisco, CA. [doi: 10.1145/997078.997089]
28. Salminen J, Guan KW, Jung SG, Jansen B. Use cases for design personas: a systematic review and new frontiers. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022 Presented at: CHI '22; April 29-May 5, 2022; New Orleans, LA. [doi: 10.1145/3491102.3517589]
29. Hendriks N, Truyen F, Duval E. Designing with dementia: guidelines for participatory design together with persons with dementia. In: Proceedings of the Human-Computer Interaction -- INTERACT 2013. 2013 Presented at: INTERACT 2013; September 2-6, 2013; Cape Town, South Africa. [doi: 10.1007/978-3-642-40483-2_46]
30. Nehme J, El-Khani U, Chow A, Hakky S, Ahmed AR, Purkayastha S. The use of multimedia consent programs for surgical procedures: a systematic review. Surg Innov 2013 Feb 14;20(1):13-23. [doi: 10.1177/1553350612446352] [Medline: 22589017]
31. Hoffner B, Bauer-Wu S, Hitchcock-Bryan S, Powell M, Wolanski A, Joffe S. "Entering a Clinical Trial: Is it Right for You?": a randomized study of The Clinical Trials Video and its impact on the informed consent process. Cancer 2012 Apr 01;118(7):1877-1883 [FREE Full text] [doi: 10.1002/cncr.26438] [Medline: 22009665]
32. Kraft SA, Constantine M, Magnus D, Porter KM, Lee SS, Green M, et al. A randomized study of multimedia informational aids for research on medical practices: implications for informed consent. Clin Trials 2017 Feb 23;14(1):94-102 [FREE Full text] [doi: 10.1177/1740774516669352] [Medline: 27625314]
33. Wang Z, Wang S, Farinella M, Murray-Rust D, Riche NH, Bach B. Comparing effectiveness and engagement of data comics and infographics. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19; May 4-9, 2019; Glasgow, Scotland. [doi: 10.1145/3290605.3300483]
34. Cook DJ, Manning DM, Holland DE, Prinsen SK, Rudzik SD, Roger VL, et al. Patient engagement and reported outcomes in surgical recovery: effectiveness of an e-health platform. J Am Coll Surg 2013 Oct;217(4):648-655. [doi: 10.1016/j.jamcollsurg.2013.05.003] [Medline: 23891066]
35. Ivanova D, Katsaounis P. Real-time dynamic tiered e-consent: a novel tool for patients' engagement and common ontology system for the management of medical data. Innov Digit Health 2021;1(2):45-49. [doi: 10.36401/IDDB-21-01]
36. Haas MA, Teare H, Prictor M, Ceregra G, Vidgen ME, Bunker D, et al. 'CTRL': an online, Dynamic Consent and participant engagement platform working towards solving the complexities of consent in genomic research. Eur J Hum Genet 2021 Apr 06;29(4):687-698 [FREE Full text] [doi: 10.1038/s41431-020-00782-w] [Medline: 33408362]
37. Simon HA. Designing organizations for an in formation-rich world. In: Greenberger M, editor. Computers, Communications, and the Public Interest. Baltimore, MD: The Johns Hopkins Press; 1971.
38. Ciampaglia GL, Flammini A, Menczer F. The production of information in the attention economy. Sci Rep 2015 May 19;5(1):9452 [FREE Full text] [doi: 10.1038/srep09452] [Medline: 25989177]
39. Street J, Duszynski K, Krawczyk S, Braunack-Mayer A. The use of citizens' juries in health policy decision-making: a systematic review. Soc Sci Med 2014 May;109:1-9 [FREE Full text] [doi: 10.1016/j.socscimed.2014.03.005] [Medline: 24657639]
40. Varnhagen CK, Gushta M, Daniels J, Peters TC, Parmar N, Law D, et al. How informed is online informed consent? Ethics Behav 2005 Apr;15(1):37-48. [doi: 10.1207/s15327019eb1501_3] [Medline: 16127857]
41. Hargittai E, Marwick A. “What can I really do?” Explaining the privacy paradox with online apathy. Int J Commun 2016;10:21 [FREE Full text]
42. Pashler H, McDaniel M, Rohrer D, Bjork R. Learning styles: concepts and evidence. Psychol Sci Public Interest 2008 Dec 01;9(3):105-119. [doi: 10.1111/j.1539-6053.2009.01038.x] [Medline: 26162104]
43. Schaub F, Balebako R, Durity AL, Cranor LF. A design space for effective privacy notices. In: Selinger E, Polonetsky J, Tene O, editors. The Cambridge Handbook of Consumer Privacy. Cambridge, MA: Cambridge University Press; 2018.
44. Lopes P. Culture and stigma: popular culture and the case of comic books. Sociol Forum 2006 Oct 17;21(3):387-414. [doi: 10.1007/s11206-006-9022-6]

45. Linek S, Huff M. Serious comics: a new approach for science communication and learning. In: Proceedings of the 12th International Technology, Education and Development Conference. 2018 Presented at: INTED 2018; March 5-7, 2018; Valencia, Spain. [doi: 10.21125/inted.2018.0755]
46. Tavares R, Alemany-Pagès M, Araújo S, Cohn N, Ramalho-Santos J, Azul AM. Comics in science and health communication: insights from mutual collaboration and framing a research practice. Int J Qual Methods 2023;22. [doi: 10.1177/16094069231183118]
47. Engelen B. Ethical criteria for health-promoting nudges: a case-by-case analysis. Am J Bioeth 2019 May 08;19(5):48-59. [doi: 10.1080/15265161.2019.1588411] [Medline: 31068115]
48. Renaud K, Zimmermann V. Ethical guidelines for nudging in information security and privacy. Int J Hum Comput Stud 2018 Dec;120:22-35. [doi: 10.1016/j.ijhcs.2018.05.011]
49. Meske C, Amojo I. Ethical guidelines for the construction of digital nudges. In: Proceedings of the 53rd Hawaii International Conference on System Sciences. 2020 Presented at: HICSS 2020; January 7-10, 2020; Maui, HI. [doi: 10.24251/hicss.2020.480]
50. Schuck PH. Rethinking informed consent. Yale Law J 1994 Jan;103(4):899-959. [doi: 10.2307/797066]
51. Heinrichs B. Myth or magic? Towards a revised theory of informed consent in medical research. J Med Philos 2019 Jan 14;44(1):33-49. [doi: 10.1093/jmp/jhy034] [Medline: 30649452]

Abbreviations

EU: European Union
GDPR: General Data Protection Regulation
IC: informed consent
RQ: research question

Edited by K Adapa; submitted 13.10.23; peer-reviewed by Z Wang, T Church; comments to author 27.01.24; revised version received 15.02.24; accepted 12.03.24; published 30.04.24.

Please cite as:
Doan X, Rossi A, Botes M, Selzer A
Comparing Attitudes Toward Different Consent Mediums: Semistructured Qualitative Study
JMIR Hum Factors 2024;11:e53113
URL: https://humanfactors.jmir.org/2024/1/e53113
doi:10.2196/53113
PMID:38687983

Original Paper

Knowledge, Skills, and Experience With Technology in Relation to Nutritional Intake and Physical Activity Among Older Adults at Risk of Falls: Semistructured Interview Study

Julie Kikkenborg[1*], BSc; Emma Magelund[1*], BSc; Maria Silke Riise[1*], BSc; Lars Kayser[2], MD, PhD; Rikke Terp[3], MHSc, PhD

[1]Department of Clinical Physiology & Nuclear Medicine, Rigshospitalet, Copenhagen, Denmark

[2]Section of Health Service Research, Department of Public Health, University of Copenhagen, Copenhagen, Denmark

[3]Department of Internal Medicine, Herlev and Gentofte Hospital, Copenhagen University Hospital, Hellerup, Denmark

[*]these authors contributed equally

Corresponding Author:
Lars Kayser, MD, PhD
Section of Health Service Research
Department of Public Health
University of Copenhagen
Øster Farimagsgade 5
Copenhagen
Denmark
Phone: 45 28757291
Email: lk@sund.ku.dk

Abstract

Background: More than one-third of older adults (aged ≥65 y) experience falls every year. The prevalent modifiable risk factors for falling are malnutrition and physical inactivity, among others. The involvement of older adults in the prevention of falls can decrease injuries, hospitalizations, and dependency on health care professionals. In this regard, eHealth can support older adults' self-management through more physical activity and adequate food intake. eHealth must be tailored to older adults' needs and preferences so that they can reap its full benefits. Therefore, it is necessary to gain insight into the knowledge, skills, and mindset of older adults living at home who are at risk of falls regarding eHealth.

Objective: This qualitative study aims to explore older adults' use of everyday digital services and technology and how they acquire knowledge about and manage their nutritional intake and physical activity in relation to their health.

Methods: Semistructured interviews were conducted with 15 older adults (n=9, 60% women; n=6, 40% men; age range 71-87 y) who had all experienced falls or were at risk of falling. These individuals were recruited from a geriatric outpatient clinic. The interviews were analyzed using deductive content analysis based on a modification of the Readiness and Enablement Index for Health Technology framework.

Results: The qualitative data showed that the informants' social networks had a positive impact on their self-management, use of technology, and mindset toward nutritional intake and physical activity. Although the informants generally lived active lives, they all lacked knowledge about how their food intake influenced their physical health, including their risk of falling. Another finding was the large diversity in the use of technology among the informants, which was related to their mindset toward technology.

Conclusions: Older adults can use technology for everyday purposes, but some need additional introduction and support to be able to use it for managing their health. They also need to learn about the importance of proper nutritional intake and physical activity in preventing falls. Older adults need a more personalized introduction to technology, nutrition, and physical activity in their contact with health professionals.

(JMIR Hum Factors 2024;11:e52575) doi:10.2196/52575

KEYWORDS

eHealth; self-management; fall prevention; older adults; physical activity; nutritional intake; Readiness and Enablement Index for Health Technology; READHY; social support; support; management; fall; nutrition; diet; qualitative study; malnutrition; physical inactivity; injury; injuries; food; food intake; nutritional needs; outpatient clinic; social network; mobile phone

Introduction

Background

Among older adults, falls are common occurrences, with one-third of the population aged ≥65 years experiencing falls every year [1]. Falls are among the major causes of mortality and morbidity in older adults [2,3], and they contribute to social and economic costs as well as create a dependency on health care professionals [4]. A reduction in the incidence of falls would increase older adults' quality of life [5], relieve the pressure on health care professionals by reducing hospitalizations [6], and reduce health care system costs [4].

Malnutrition and physical inactivity are well-known modifiable behavioral risk factors for falls in older adults [7-9]. Older adults who are malnourished have a 45% higher risk of falling at least once [7]. Increased daily activity and moderate strength and intensity training 3 times a week can reduce the risk of falling by 30% [8]. Therefore, behavior change regarding nutritional intake and physical activity is important to prevent falls, which is also recommended in international clinical guidelines [8,10]. Behavior change in individuals is influenced by their self-management abilities [11,12].

eHealth, defined as "an emerging field at the intersection of medical informatics, public health, and business, referring to health services and information delivered or enhanced through the Internet and related technologies" [13], is often considered an effective tool for supporting self-management [14]. It may serve as an important "copilot" for older adults to increase their awareness of their nutritional needs and motivate them to engage in higher levels of physical activity [15-17]. To ensure that older adults adopt eHealth and find it useful, their level of knowledge, skills, and experience with eHealth must be addressed when they are introduced to it [18,19]. This includes the level of digital health literacy because it is a determinant considered important for technology adoption [20] and self-management [14].

Another important factor is the existence of social support from relatives and friends to successfully find and understand health information [21,22]. Support and encouragement from health care professionals are also essential for the sustainable adoption of health information and eHealth [22,23].

Older adults' perceptions and mindset may affect their engagement with health initiatives [24,25]. We have previously shown that older adults' perceptions and mindset affect their use of technology [6] and self-management [26]. Among older hospitalized patients, we found that a lack of awareness regarding meeting their nutritional needs was related to limited knowledge of their nutritional requirements and the impacts of food intake on their physical function [26]. Consequently, many older adults may not eat adequately [26].

The use of eHealth, the existence of social relations, and the capability to manage one's own condition, as well as perceptions and mindset, are all intermingled factors and important to include to obtain a well-functioning sociotechnical ecosystem that enables healthy behavior. In this study, we explore these aspects in the context of nutritional intake and physical activity in older adults living at home who are at risk of falling.

Objectives

The aim of this study is to explore older adults' use of everyday digital services and technology and how they acquire knowledge about, and manage, their nutritional intake and physical activity in relation to their health. These data may help differentiate among users and address specific gaps in relation to knowledge, skills, and mindset. The findings will provide insight into whether eHealth may be a feasible approach to enable and engage older adults living at home who are at risk of falling.

Methods

Overview

This is an explorative qualitative study based on semistructured interviews, which were analyzed using content analysis with a deductive approach. The study is part of a larger research program exploring how, through the use of eHealth, we can optimize older adults' self-management regarding nutritional intake and physical activity to prevent functional decline and reduce the risk of falling. An intervention outlining how health care professionals can assist and support older adults in self-management through the use of eHealth will be designed, developed, and tested. The findings of the study reported in this paper will inform the design of the intervention.

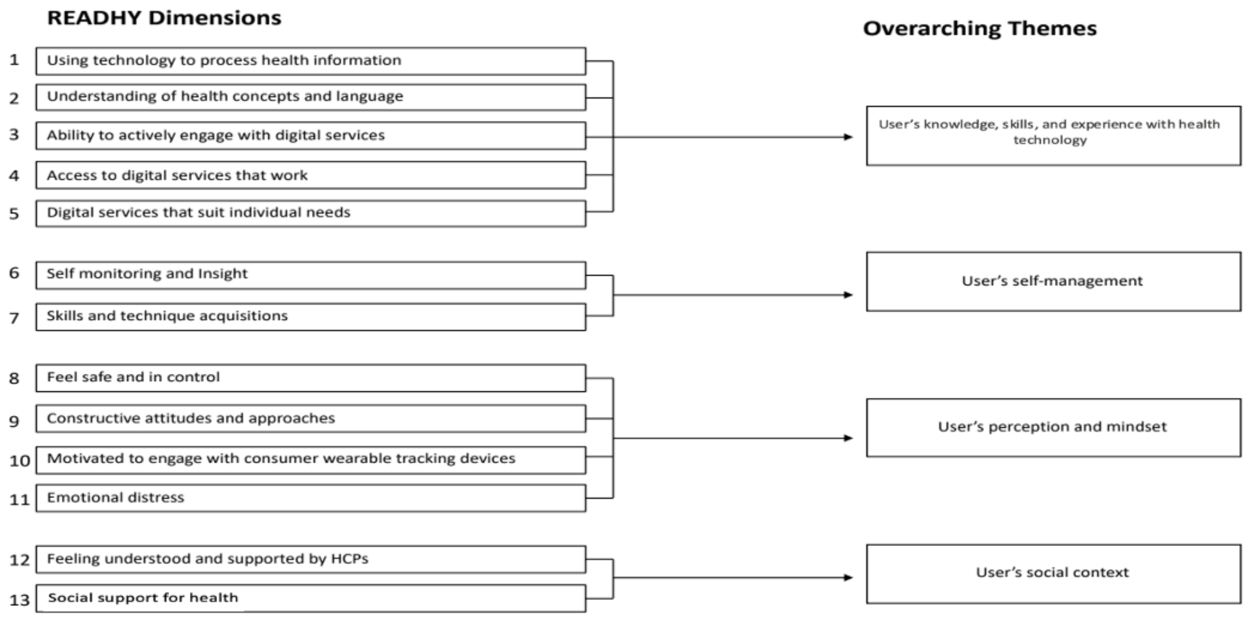
Participants and Setting

All informants were recruited through convenience sampling from a geriatric outpatient clinic in a university hospital in the capital region of Denmark. An exercise physiologist employed at the clinic obtained informed consent, allowing the first authors (JK, EM, and MSR) to contact and inform the informants before their inclusion in the study. The inclusion criteria were age ≥65 years, good comprehension of the Danish language, and being at risk of falling. The exclusion criterion was the inability to provide informed consent because of cognitive impairment. Before being interviewed, 16 informants were contacted by telephone by one of the first authors to inform them about the study and arrange the interviews. Of these 16 informants, 1 (6%) retracted their earlier decision to participate after the telephone call. Of the 15 informants, 12 (80%) were interviewed in February and March 2022; the remaining 3 (20%; all male informants to ensure information power by gaining varied experiences from both sexes [27]) were interviewed in February and March 2023 [27].

Theoretical Framework

This study is based on a sociotechnical perspective, meaning that technologies are seen as actors that interact with the users in specific contexts instead of being passive tools [28]. To explore our informants' ability to engage with eHealth, we used the Readiness and Enablement Index for Health Technology (READHY) as the theoretical framework for the interview guide and analysis [22]. The framework can be used to describe individuals' readiness for, and enablement by, eHealth. It consists of 13 dimensions within 3 main themes: self-management, social support, and digital health literacy [18]. In this study, we applied the modified READHY framework proposed by M Blaauwhof (personal communication, November 2020). The 13 dimensions were aggregated into the following four main categories: (1) the user's knowledge, skills, and experience with eHealth; (2) the user's self-management; (3) the user's perception and mindset; and (4) the user's social context (Figure 1; M Blaauwhof, personal communication, November 2020). The main categories were used to understand how eHealth could be applied to support older adults' health behaviors regarding nutritional intake and physical activity.

Figure 1. Aggregated themes. HCP: health care professional; READHY: Readiness and Enablement Index for Health Technology.



Data Collection

Data were collected from 15 semistructured interviews using an interview guide (Multimedia Appendix 1) with open-ended questions about nutritional intake and physical activity, self-management, social context, and the use of technology. All interviews were conducted by the first authors and took place in the informants' homes. A minimum of 2 of the 3 first authors were present during the interviews. Of the 15 informants, 4 (27%) had spouses present during the interviews. The interviews lasted from 23 to 60 minutes, with an average of 42 (SD 12) minutes.

Data Analysis

The interviews were recorded using Microsoft Teams, transcribed by the first authors, and analyzed using deductive content analysis [29] based on the 4 categories from the modified READHY framework. The codes for the deductive content analysis were identified by the first authors by dividing the 4 categories from the modified READHY framework into subcategories and afterward into codes. The first authors coded the interviews using the data management software program NVivo (version 14; Lumivero). The codes were revised by the first authors and the second author (LK) to ensure reliability and to provide the option of adding newly identified codes when appropriate. The codebook consisted of 25 unique codes with a matching description (Multimedia Appendix 2). First authors JK, EM, and MSR have BSc degrees in health informatics and have experience in using qualitative and quantitative methods during their studies. JK, EM, and MSR were supervised by LK and RT, who have experience in using qualitative methods. ChatGPT (GPT-3.5; OpenAI) and Grammarly (Grammarly Inc), which offers artificial intelligence–powered writing assistance, were used to support the translation of selected anonymous quotes, the interview guide, and the codebook to improve readability and language.

Ethical Considerations

Verbal and written information about the study was provided to all informants, and written consent was obtained from all of them. The study was conducted in accordance with the Helsinki Declaration. According to Danish regulations, health science questionnaire surveys and interview studies that do not involve human biological material (section 14(2) of the Danish Act on Committees) do not require reporting to, or approval from, the Danish National Centre for Ethics [30]. All data are stored in accordance with Danish legislation (General Data Protection Regulation). The informants were not reimbursed for their participation.

Results

Informant Characteristics

A total of 15 informants (n=9, 60% women; n=6, 40% men) were included. Their mean age was 80 (range 71-87, SD 5.3) years. Of the 15 informants, 12 (80%) had experienced falls within the last year, and 3 (20%) had experienced balance issues and dizziness. Other characteristics of the informants are summarized in Table 1, including sex, age, cohabitating status, and highest educational level attained.

Table 1. Informants' characteristics.

ID	Sex	Age (y)	Living alone or cohabiting	Highest level of education attained[a]
#1	Female	81	Cohabiting	Medium
#2	Male	72	Cohabiting	Short
#3	Female	85	Living alone	Medium
#4	Female	80	Living alone	Medium
#5	Female	86	Living alone	Short
#6	Female	80	Cohabiting	Short
#7	Male	83	Cohabiting	Medium
#8	Female	87	Cohabiting	Comprehensive school
#9	Female	79	Living alone	Medium
#10	Female	71	Living alone	Medium
#11	Female	87	Living alone	Short
#12	Male	84	Living alone	Medium
#13	Male	74	Cohabiting	Long
#14	Male	75	Cohabiting	Short
#15	Male	80	Cohabiting	Short

[a]The education variable is aggregated from the 8 levels of the International Standard Classification of Education 2011 [31] and classified into 4 categories as follows: comprehensive school (typically up to lower secondary education), short education (including upper secondary and some postsecondary programs), medium education (encompassing bachelor's and master's degrees), and long education (referring to doctoral studies).

The findings are presented in four categories: (1) the user's knowledge, skills, and experience with eHealth; (2) the user's self-management; (3) the user's social context; and (4) the user's perception and mindset.

The User's Knowledge, Skills, and Experience With eHealth

There was a large diversity in how experienced the informants were with technology and how often they used it in their everyday lives. The informants were divided into three groups based on their knowledge, skills, and experience with technology: (1) those who used technology daily (experienced informants); (2) those who used technology to some extent (partially experienced); and (3) those who had limited use of, and skills with, technology (inexperienced informants). More than half of the informants (8/15, 53%) were experienced users of technology, and 5 (83%) of the 6 male informants were experienced users of technology. The experienced informants were frequent internet searchers and daily social media users. The partially experienced informants only used technology for necessary purposes, such as checking email. Finally, the inexperienced informants had, to some extent, given up on technology either because they had no interest in using it or because they found it too difficult to use; these were also the oldest of the informants.

The experienced informants were aware of the fluctuating credibility of web pages on the internet:

> *It must be kind of random what you can do, what you find, and how you're loaded with the information you're seeking. So, you can risk finding something, I wouldn't say a lie, but something that might not be what you were searching for.* [Informant #2, male, aged 72 years]

All informants had mobile phones, and most of them (14/15, 93%) also had computers. The informants, usually the experienced ones, had smartphones (11/15, 73%), while some had nonsmartphones (4/15, 27%). The evolution of technology was unwelcome to some of the informants, and they were more comfortable using familiar technologies:

> *I don't have a modern cell phone, nor do I want one. I have this old Nokia, and it can send a text [message] and tell me what time it is; I can forward my landline calls to it whenever I'm out for a walk, so I just have it with me in my pocket.* [Informant #3, female, aged 85 years]

> *It's [the cell phone] just this one you can call from. It [the cell phone] can only do what I need it to do.* [Informant #11, female, aged 87 years]

Several informants used health technologies, such as pedometers, click-ons for hearing aids, and health information web pages. Experienced users were the ones who used technology to search for health information, but only a few used the Danish national health portal:

> *It was not long ago that I was on it [the Danish National health portal] to see all my diseases and look at my test results.* [Informant #7, male, aged 83 years]

The User's Self-Management

The data indicate that the informants in general were interested in taking care of their health, and several were making efforts to do so, including being physically active, but seemingly they paid limited attention to their nutritional intake. One informant expressed how important it was for her to be physically active for fear of losing the ability to move around freely one day. The same informant mentioned that she does not pay much attention to what she eats, arguing that she feels good now, so what she is eating must be fine:

> *Can you tell me about how you try to take care of your health?* [Interviewer]
>
> *Well, first and foremost, I do that by being physically active. And I'm so scared actually, it's something I'm afraid of, that one day I won't be able to move anymore.* [Informant #11, female, aged 87 years]
>
> *Do you ever think about what you should eat in relation to your health?* [Interviewer]
>
> *Not so much. I must admit.... I tell myself that I feel so good and I can still do so much, so what I eat cannot be completely wrong.* [Informant #11, female, aged 87 years]

Many of the informants understood and adhered to the health information they received from the internet, their relatives, health care professionals, and other sources of information, whereas only a few informants searched for health-related information themselves. Those who searched for health information typically used the internet or consulted their friends and relatives. Seemingly, the informants living with chronic conditions were more aware of seeking and appraising health information to minimize the impact of their conditions:

> *And we [friend] talked a lot about what she did...Well, but she recommended the anti-inflammatory diet, and so I did that for a longer period.* [Informant #4, female, aged 80 years]
>
> *And specifically with Parkinson's, you stiffen up, so it's especially good to stay active.* [Informant #9, female, aged 79 years]

Regarding nutrition, it also seemed that primarily those informants with diet-related conditions either sought or received information from health care professionals. Some mentioned receiving and adhering to nutritional advice from their general practitioners to better manage their conditions, such as high cholesterol, imbalanced salt concentration, and celiac disease.

Several informants set health-related goals for themselves to stay physically active in their everyday lives, such as taking daily walks, achieving a certain number of steps, and losing weight. In addition, they focused on maintaining mental well-being. Two of the informants, whose goals were taking daily walks and achieving 10,000 steps a day, expressed how they used pedometers to track their progress:

> *We walked, well, over 10,000 steps per day.* [Informant #1, female, aged 81 years]

Some of the female informants had goals related to losing weight or avoiding weight gain. Several had tried various weight reduction diets that they had heard about from relatives or the media, including commercial television programs. Diets included the ketogenic diet and intermittent fasting. In general, all informants who had health-related goals focused on achieving these goals. Some also expressed how their goal of having positive attitudes toward physical activity helped them become more physically active in their everyday lives:

> *Then I have a good [motivational] phrase: I always tell myself that what I could do yesterday, I can also do today.* [Informant #11, female, aged 87 years]

The User's Social Context

In general, the informants' relatives, particularly their children and grandchildren, were involved in, and supported them in, managing their health in terms of their nutritional intake and physical activity. Relatives inquired about the informants' health and took the initiative to help them improve it. These initiatives were often in the form of encouragement and incentives to be physically active, such as walking or using workout equipment or exercise bikes:

> *[T]hen my daughter and my son-in-law came by last Sunday and asked whether I wanted to go for a walk, and so, of course, I went with them. That happens sometimes.* [Informant #3, female, aged 85 years]
>
> *ago that I can use at home.* [Informant #12, male, aged 84 years]

The informants expressed that their relatives also played a role in shaping their eating habits, such as by introducing them to new diets or suggesting nutritional changes. However, only a few informants expressed involvement from their relatives regarding nutritional intake (compared to involvement in their physical activities):

> *[My son and daughter-in-law] have switched to a vegetarian diet 3 times a week, and we are also on board with that.* [Informant #2, male, aged 72 years]

Friends were also important sources of support. For the informants, talking with friends about their disease was a way to feel supported by discussing and sharing their experiences with peers, whereas relatives were the most important sources of support when it came to encouragement and incentives to improve their health:

> *[W]e ache and wonder how much one can actually become afflicted, as it gradually hits you. So, naturally, we discuss it a bit [with friends].* [Informant #9, female, aged 79 years]

Relatives, particularly the informants' children, were also perceived as important sources of support for technology use. They helped with the installation of new technologies and with other technical difficulties, such as pressing the wrong buttons or understanding how to use new apps:

> *[A]nd then maybe I've got a hold of something, and without knowing what it is, I fiddle and press various buttons, you know. And then it's good that I have him [his son].* [Informant #2, male, aged 72 years]

Only one informant had neither a child nor a spouse, and she experienced less social support than the other informants. This informant had no interest in physical activity, did not receive incentives from others to improve her health, and did not talk to friends about health-related topics to avoid burdening them:

> *It's not that I don't have good friends. But now, for instance, I have a very good friend who has been sick, with ataxia, and I don't want to burden her with that, you know. Because she's very sick already, so I don't want her to worry about me.* [Informant #10, female, aged 71 years]

If the informant had problems with technology, she sought help professionally, not from friends:

> *I have had some issues with my mail because YouSee and TDC [telecommunication providers] had some spam filters that hid all of my emails. I couldn't enter my email then...then they wrote to me and told me how to fix it.* [Informant #10, female, aged 71 years]

Most of the informants had positive experiences with support from health care professionals from the geriatric outpatient clinic. They felt supported by the exercise programs provided by physiotherapists from the clinic. In many cases, the positive experiences of support were related to being provided with individualized physical exercises and advice relating to their health conditions:

> *And she [the physiotherapist] has shown me some exercises to help with my balance, among other things. And I've gotten one of those round cushions that are soft, you know, so I can stand and maintain my balance.* [Informant #11, female, aged 87 years]

However, a few informants felt less supported because they had to perform a large part of the physical exercises on their own at home, and they lacked detailed instructions for these exercises:

> *Sometimes, I feel like I need an instructor.* [Informant #2, male, aged 72 years]
>
> *It [exercise instructions on paper] doesn't say; there are no measurements for the width of the board or how long you need to walk and how much you can deviate. I think it's very unspecific.* [Informant #7, male, aged 83 years]

Most of the informants listened to, and followed, the advice provided by the health care professionals from the geriatric outpatient clinic regarding their health and how to minimize the risk of falling. The advice included drinking enough fluid during the day and using training equipment at home correctly. Despite the informants' history of being at risk of falls, they generally did not report receiving information or advice from health care professionals regarding nutrition to maintain their physical function and thus minimize their risks of falling.

The User's Perception and Mindset

The informants' physical activities varied and could be classified into three groups: (1) structured physical activities, (2) incidental physical activities, and (3) inactive. Approximately half (7/15, 47%) belonged to the group that engaged in structured physical activities, such as daily physical exercises or weekly planned exercises with peers [32,33]. One-third (5/15, 33%) belonged to the group that engaged in incidental physical activities, such as house cleaning, short walks, and gardening [33,34]. Only a few (3/15, 20%) belonged to the inactive group of informants, who only did what was necessary because they seemingly had little interest in physical activities.

The informants' motivation for being physically active stemmed from the desire to remain physically mobile and avoid relying on others, as well as the opportunity for social interactions. Several informants found it enjoyable to participate in various structured physical activities together with their peers, such as going on excursions or joining training groups, rather than performing exercises alone:

> *And I'm so afraid; it's actually one of the things I'm most afraid of, that one day, I can't move, walk, or take care of myself anymore.* [Informant #11, female, aged 87 years]
>
> *And then I made a club down here, where we meet every Wednesday. And there [in the club], you're motivated to engage in a lot of events through songs, or someone comes and gives lectures. We'd go for a walk in the forest, or take trips on a bus, where we have lunch, or take some trips in which we'd have lunch out. All such events.* [Informant #8, female, aged 87 years]

A few informants from the structured physical activities group described how they did not perceive themselves as physically active. Some even expressed that they were lazy:

> *I must say that I'm living with the effects of my broken shoulder. And I'm completely lazy.* [Informant #4, female, aged 80 years]

The informants from the incidental physical activities group were familiar with how performing daily physical activities improved their health, but despite this knowledge, their motivation to perform physical activities was limited. However, there was a tendency for social interactions to be motivating factors for the informants belonging to this group:

> *Yes, and I can say that I don't think I'm fulfilling my responsibility [performing exercises] if I have to do it alone. I'm probably better at doing it with others.* [Informant #8, female, aged 87 years]

For the inactive informants, a primary reason for not exercising was the difficulty in finding personal meaning and purpose in physical activities. For some of these informants, the potential opportunity to engage in social interactions did not increase their motivation to be physically active:

> *I've always hated going on walks without a purpose. I've never played sports.* [Informant #12, male, aged 84 years]
>
> *Well, I don't like to do anything when there are a lot of people around, and when I need to stand there and do one thing after another.* [Informant #6, female, aged 80 years]

Some informants who were physically impaired because of health-related conditions found it difficult to be physically active. These informants were also either in the inactive or incidental physical activity group:

> *Then, I'll take a walk in the forest. I haven't done that very much lately, perhaps because I've had an operation on my knee. I'm not particularly fond of it, but right now, I can't go on my long walks.* [Informant #3, female, aged 85 years]

By contrast, the informants from the structured physical activities group had more knowledge about, and were more attentive to, staying active during and after an illness or injury.

Regarding the informants' perception and mindset toward their food intake, many expressed uncertainty about whether there were particular foods that would benefit their health. However, the majority expressed that they should consume more vegetables, and some also expressed that they strived to do so. Vitamins, both as supplements and in food, were highlighted as important for staying healthy. One informant expressed that protein intake was beneficial for weight loss. No informant mentioned the importance of protein intake for maintaining muscle strength and thus preventing falls. Several informants expressed acceptance regarding not necessarily adhering to recommendations for healthy eating (generally referred to as the intake of vegetables and foods low in fat and sugar) because most of them were under the impression that they ate according to their nutritional needs:

> *I actually think I eat as I should.* [Informant #3, female, aged 85 years]

Despite the informants' narratives revealing their perception of the concept of healthy food, it was evident that their food intake was largely guided by their preferences for the foods they liked. Some informants revealed positive attitudes toward allowing their preferences for tasty food to guide their intake:

> *I think about having some vegetables because it's supposed to be good for the stomach. But otherwise, I don't really think about whether it's healthy or not. I think all food is healthy if you feel like it. Unless you overdo it, of course.* [Informant #7, male, aged 83 years]

Advanced age was also mentioned as a justification for not making any changes to their eating behavior just to adhere to the consumption of "healthy food":

> *I suppose now that I've gotten so old, so...* [Informant #6, female, aged 80 years]

In general, the informants had a limited focus on their nutritional intake and on fulfilling their dietary needs. However, some female informants were more likely to report avoiding fatty food to avoid weight gain:

> *No, I don't eat fatty food, I don't eat butter on bread...It's a waste of calories because you don't need it; there's liver pâté.* [Informant #10, female, aged 71 years]

Several informants perceived technology as a necessity in their daily lives; however, some expressed that they thought technology had too much of an impact on their lives:

> *And the worst part of it is this damn computer. It starts in the morning during breakfast when I open it up, and then I'll check my email, then Ekstra Bladet [a Danish news magazine], and then Facebook, and only then am I ready for the day.* [Informant #14, male, aged 75 years]

Many of the informants had positive attitudes toward using a mobile phone because it made it easier for them to reach others and stay in contact with them. The experienced users preferred using computers or iPads because of their large screens, which made them easier to use:

> *Well, now if I should look at it [exercises on the computer] right, then I don't have to deal with the small font. It's nice with a big screen.* [Informant #1, female, aged 81 years]

Some informants expressed negative attitudes toward technology because they felt that it hindered personal communication and was challenging to manage. Several informants mentioned that negative experiences mainly occurred when the technology changed or did not work. These challenges were demotivating for the informants because it required time to resolve the problems and caused great frustration:

> *I think that, at least for us, there are too many times when it [the technology] doesn't work...and it takes too long.* [Informant #8, female, aged 87 years]

Discussion

Principal Findings

We explored older adults' use of everyday digital services and technology and how they acquire knowledge about, and manage, their nutritional intake and physical activity in relation to their health. A main finding in this study was the great diversity in the informants' experiences, mindset, and use of technology: some perceived technology as a necessity in their daily lives, whereas others viewed it as a source of frustration. Feelings of frustration were more prevalent among the oldest informants. Another main finding was that, although the informants were at risk of falls and had been referred to a geriatric outpatient clinic for the assessment and management of their fall risk, they possessed limited focus on, and knowledge about, nutritional needs to promote or maintain good physical function and thereby decrease their risk of falling. An important finding was the positive impact of the social network of relatives and health care professionals on the informants' use of technology and motivation for the self-management of nutritional intake and physical activity. There seemed to be a relationship between the informants' levels of physical activity and having a positive mindset, whereby the most active informants seemed to be more aware of the benefits of staying active and thinking positive. By contrast, there was limited focus on the importance of their food intake, which seemed to be related to the informants' limited knowledge about this topic and not to a lack of motivation to look after their health and well-being.

Comparison With Other Work

The great diversity in the informants' experiences and use of technology, with inexperienced informants belonging to the older age group, is also described in previous studies [6,35,36]. Rossen et al [35] examined readiness for technology and showed that the older age group had the lowest readiness for technology. However, the findings of Goyal et al [37] indicated that older adults may have greater adherence than younger adults once they adopt technology. This suggests that it is important to provide older adults with sufficient support to help them adopt technology. Although most of the informants (14/15, 93%) used technology, only a few (6/15, 40%) used it for health-related purposes. It was mainly the informants with chronic diseases (6/15, 40%) who searched for information on treatment and medication and generally focused on information that could minimize the impact of their chronic diseases. This is in alignment with 2 other studies and a recent review [6,23,38], which found that when older adults seek health-related information, it is often related to information about specific diseases, treatments, and medicines. Interestingly, those informants who used eHealth to find health-related information (6/15, 40%) mainly searched for information about their chronic diseases but did not search for information about how to prevent falls. This signifies that they do not consider information about preventive measures in relation to falls as being health related.

To understand individuals' motivation and ability for self-management, it is important to gain insight into their perception and mindset because these affect their motivation to engage in a specific behavior [39,40]. Our finding that older adults generally have positive attitudes toward technology is supported by the results of other studies [6,41]. The main reason for not using technology was frustration with technological challenges and because it hindered personal communication. This aligns with the results in the review by Wilson et al [19], who found that barriers to using eHealth included dislike of the technology and problems with functionality. Perceiving technological challenges may affect individuals' perception of ease of use, which is an important determinant for the intention to use technology [42]. Therefore, these potential barriers need to be addressed in future interventions. Furthermore, Wilson et al [19] found that a lack of knowledge and experience with using technology hindered use, whereas a belief in its benefits facilitated use. These findings are consistent with those of Terp et al [6], who found that the perception of technology as being useful facilitated its use among older hospitalized patients and that nonuse was mainly due to a lack of knowledge about the derived benefits from the technology. These findings suggest a need to provide older adults with knowledge about the advantages of using eHealth.

We found that the informants who were physically active (structured and incidental physical activities groups) were motivated to perform physical activities because they experienced these as being fun, they enjoyed the social element, they had been active throughout their lives, and they wished to remain physically mobile and avoid relying on others. These motivational factors for physical activity have also been described in other studies [43,44]. The reviews by Sandlund et al [43] and Bunn et al [44] found that important facilitators for commencing fall exercise programs were previous exercise habits, social support and interaction, the ability to remain independent, and the fun element. For the inactive informants, a primary reason for not exercising was the difficulty in finding personal meaning and purpose in physical activities. These findings also align with those of Sandlund et al [43] and Bunn et al [44], who found that the barriers to commencing fall exercise programs were lack of support and interest, concerns about the exercises, and unawareness of the benefits.

Our findings indicate that older adults, despite being at risk of falling, may have a limited focus on eating adequately to maintain or improve their physical health. Among several of the informants, the behavior and mindset toward food intake were focused on society's notion of an ideal slim body, which corresponds to the findings of previous studies [26,45]. Our findings suggest that this may not be due to a lack of motivation but merely due to limited knowledge; the majority were motivated and focused on maintaining their health because of their fear of losing physical function. Despite the importance of adequate nutrition, previous studies have reported limited nutritional knowledge among older adults [45,46]. Our findings indicate that older adults are more likely to receive, rather than actively seek, health-related information, a result also supported by our previous study among Danish hospitalized patients [26]. Therefore, health care professionals play an important role in providing older adults with relevant information, including information on their nutritional needs and the risk of falling. Our data revealed that older adults, in general, trusted health care professionals and adhered to the advice they provided. Our findings indicate potential benefits in ensuring that older adults receive relevant information and advice in future fall prevention interventions.

Social support from family, especially from their adult children, had a positive impact on the informants' technology use, self-management, and mindset concerning physical activity and nutritional intake. This positive influence on older adults' technology use was reported in a review by Levin-Zamir and Bertschi [21], who found that social support is paramount for many older adults in executing tasks related to health information from media sources. This finding is also corroborated by Takemoto et al [41], who found that human support increases accountability and enhances the use of technological devices. The positive impact of support from the family on the informants' mindset for nutritional intake and physical activity was also established in other studies [2,26,47]. Spiteri et al [47] reported that family support was one of the key motivators for physical activity among older adults, and Terp et al [26] found that relatives were an important resource for older adults' food intake. The positive impact of social support on older adults' self-management is also corroborated by Schnock et al [2], who found that being married or living with someone had a positive impact on engagement in fall prevention interventions and self-management among older adults. Although almost half of our informants (7/15, 47%) lived alone, our data showed that most of them (6/7, 86%) experienced social support. This is contradictory to the findings of another study [48] in which older adults who lived alone experienced less social support. Overall, in our study, relatives were seen to have a higher impact on the informants' levels of physical activity than on their nutritional intake. This may be explained by a lack of knowledge among the relatives about the nutritional needs of older adults. In future fall prevention interventions, it is important that information about nutrition and its impact on physical function and the prevention of falls is also provided to relatives.

Several findings of this study correspond to those in our previous study, which we conducted among older hospitalized patients [6,26]. The population in the study reported in this paper differs because all informants were at risk of falling and underwent evaluation and treatment in a geriatric outpatient clinic. International guidelines [10] recommend fall prevention interventions that enhance health behavior to reduce the risk factors for falls, such as inadequate nutrition and a lack of physical activity. The informants were recruited from the outpatient clinic and had received information and training from the clinic before the interviews. We therefore expect that they were provided with information on risk factors and advice regarding optimal nutritional intake and physical activity. However, we cannot conclude from this study whether the informants had been offered such interventions. This study provides important knowledge from the perspective of older adults about their needs in terms of developing an eHealth-based intervention aimed at supporting and motivating better health behavior, thus preventing functional decline and its consequences, such as falls.

Strengths and Limitations

A strength of this study was the use of READHY as a theoretical framework because it provided a conceptual understanding of relevant aspects of individuals' readiness to engage with eHealth, such as digital health literacy, social support, and the capability to manage their own condition. The use of a modified version of the READHY framework with the addition of a fourth theme (perception and mindset) is better suited for a qualitative analysis process. Another strength is that the informants live in a country that is among the most digitalized in the world [49]. The use of digital services and technology is therefore common in the general population, and a lack of resources or difficulties in accessing the internet and eHealth are usually not barriers, as our data also indicate. Thus, as the lack of access to digital technology is not a barrier among this group of informants, it enabled us to explore how the informants acquire knowledge about their nutritional intake and physical activity in relation to their health. However, the transferability of the study is worth considering, given that the majority of the informants (11/15, 73%) lived in the same geographic area of Denmark (Nordsjælland). In future studies, the inclusion of informants from other geographic areas can help achieve greater heterogeneity.

Conclusions

This study demonstrates the potential of eHealth to support self-management in older adults, as most of them already use digital technology in their everyday lives. Older adults' age, social context, and mindset should be considered when implementing and supporting eHealth. They must be provided with knowledge about the benefits of using eHealth to improve their motivation to use it for the self-management of their nutritional intake and physical activity. Furthermore, health professionals must be aware of the need to educate older adults about the impact of nutritional intake and physical activity in fall prevention, particularly for those who lack social support.

Acknowledgments

The authors would like to thank the informants who participated in this study. The authors also thank exercise physiologist Pernille Hansen from the geriatric outpatient clinic, Herlev-Gentofte Hospital for the assistance in recruiting informants, and Marijke Blaauwhof for sharing her master's thesis work with us and allowing us to use her modified model of the Readiness and Enablement Index for Health Technology framework. The authors used the ChatGPT (GPT-3.5; OpenAI) [50] and Grammarly (Grammarly Inc), which offers artificial intelligence–powered writing assistance, to support the translation of selected anonymous quotes, the interview guide, and the codebook to improve readability and language.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.
[DOCX File , 18 KB - humanfactors_v11i1e52575_app1.docx]

Multimedia Appendix 2
Codebook.
[DOCX File , 14 KB - humanfactors_v11i1e52575_app2.docx]

References

1. Visser M, Volkert D, Corish C, Geisler C, de Groot LC, Cruz‐Jentoft AJ, et al. Tackling the increasing problem of malnutrition in older persons: the malnutrition in the elderly (MaNuEL) knowledge hub. Nutr Bull 2017 May 08;42(2):178-186. [doi: 10.1111/nbu.12268]

2. Schnock KO, P Howard E, Dykes PC. Fall prevention self-management among older adults: a systematic review. Am J Prev Med 2019 May;56(5):747-755. [doi: 10.1016/j.amepre.2018.11.007] [Medline: 30885516]
3. Clemson L, Cumming RG, Heard R. The development of an assessment to evaluate behavioral factors associated with falling. Am J Occup Ther 2003;57(4):380-388. [doi: 10.5014/ajot.57.4.380] [Medline: 12911079]
4. Phelan EA, Ritchey K. Fall prevention in community-dwelling older adults. Ann Intern Med 2018 Dec 04;169(11):ITC81-ITC96. [doi: 10.7326/AITC201812040] [Medline: 30508457]
5. Hawley-Hague H, Tacconi C, Mellone S, Martinez E, Yang F, Su TL, et al. Using smartphone TechnolOGy to support an EffecTive home ExeRcise intervention to prevent falls amongst community-dwelling older adults: the TOGETHER feasibility RCT. Gerontology 2023;69(6):783-798 [FREE Full text] [doi: 10.1159/000528471] [Medline: 36470216]
6. Terp R, Kayser L, Lindhardt T. Older patients' competence, preferences, and attitudes toward digital technology use: explorative study. JMIR Hum Factors 2021 May 14;8(2):e27005 [FREE Full text] [doi: 10.2196/27005] [Medline: 33988512]
7. Trevisan C, Crippa A, Ek S, Welmer AK, Sergi G, Maggi S, et al. Nutritional status, body mass index, and the risk of falls in community-dwelling older adults: a systematic review and meta-analysis. J Am Med Dir Assoc 2019 May;20(5):569-82.e7. [doi: 10.1016/j.jamda.2018.10.027] [Medline: 30554987]
8. Global recommendations on physical activity for health. World Health Organization. 2010 Jan 01. URL: https://www.who.int/publications/i/item/9789241599979 [accessed 2024-04-25]
9. Gusdal AK, Johansson-Pajala RM, Arkkukangas M, Ekholm A, Zander V. Preventing falls and malnutrition among older adults in municipal residential care in Sweden: a registry study. SAGE Open Nurs 2021 Jul 07;7:23779608211026161 [FREE Full text] [doi: 10.1177/23779608211026161] [Medline: 34286078]
10. Montero-Odasso M, van der Velde N, Martin FC, Petrovic M, Tan MP, Ryg J, et al. World guidelines for falls prevention and management for older adults: a global initiative. Age Ageing 2022 Sep 02;51(9):afac205 [FREE Full text] [doi: 10.1093/ageing/afac205] [Medline: 36178003]
11. Van de Velde D, De Zutter F, Satink T, Costa U, Janquart S, Senn D, et al. Delineating the concept of self-management in chronic conditions: a concept analysis. BMJ Open 2019 Jul 16;9(7):e027775 [FREE Full text] [doi: 10.1136/bmjopen-2018-027775] [Medline: 31315862]
12. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48(2):177-187. [doi: 10.1016/s0738-3991(02)00032-0] [Medline: 12401421]
13. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
14. Marklund S, Tistad M, Lundell S, Östrand L, Sörlin A, Boström C, et al. Experiences and factors affecting usage of an eHealth tool for self-management among people with chronic obstructive pulmonary disease: qualitative study. J Med Internet Res 2021 Apr 30;23(4):e25672 [FREE Full text] [doi: 10.2196/25672] [Medline: 33929327]
15. Robert C, Erdt M, Lee J, Cao Y, Naharudin NB, Theng YL. Effectiveness of eHealth nutritional interventions for middle-aged and older adults: systematic review and meta-analysis. J Med Internet Res 2021 May 17;23(5):e15649 [FREE Full text] [doi: 10.2196/15649] [Medline: 33999005]
16. Muellmann S, Forberger S, Möllers T, Bröring E, Zeeb H, Pischke CR. Effectiveness of eHealth interventions for the promotion of physical activity in older adults: a systematic review. Prev Med 2018 Mar;108:93-110. [doi: 10.1016/j.ypmed.2017.12.026] [Medline: 29289643]
17. Terp R, Kayser L, Lindhardt T. An educative nutritional intervention supporting older hospital patients to eat sufficiently using eHealth: a mixed methods feasibility and pilot study. BMC Geriatr 2024 Jan 04;24(1):22 [FREE Full text] [doi: 10.1186/s12877-023-04582-x] [Medline: 38177992]
18. Kayser L, Karnoe A, Furstrand D, Batterham R, Christensen KB, Elsworth G, et al. A multidimensional tool based on the eHealth literacy framework: development and initial validity testing of the eHealth Literacy Questionnaire (eHLQ). J Med Internet Res 2018 Feb 12;20(2):e36 [FREE Full text] [doi: 10.2196/jmir.8371] [Medline: 29434011]
19. Wilson J, Heinsch M, Betts D, Booth D, Kay-Lambkin F. Barriers and facilitators to the use of e-health by older adults: a scoping review. BMC Public Health 2021 Aug 17;21(1):1556 [FREE Full text] [doi: 10.1186/s12889-021-11623-w] [Medline: 34399716]
20. Monkman H, Kushniruk AW. The consumer health information system adoption model. Stud Health Technol Inform 2015;218:26-31. [Medline: 26262522]
21. Levin-Zamir D, Bertschi I. Media health literacy, eHealth literacy, and the role of the social environment in context. Int J Environ Res Public Health 2018 Aug 03;15(8):1643 [FREE Full text] [doi: 10.3390/ijerph15081643] [Medline: 30081465]
22. Kayser L, Rossen S, Karnoe A, Elsworth G, Vibe-Petersen J, Christensen JF, et al. Development of the multidimensional readiness and enablement index for health technology (READHY) tool to measure individuals' health technology readiness: initial testing in a cancer rehabilitation setting. J Med Internet Res 2019 Feb 12;21(2):e10377 [FREE Full text] [doi: 10.2196/10377] [Medline: 30747717]
23. Brainin E, Neter E. Refined analysis of older eHealth users from an agency perspective: quantitative telephone interview study. JMIR Aging 2023 Apr 26;6:e40004 [FREE Full text] [doi: 10.2196/40004] [Medline: 37121572]

24. Archibald M, Lawless M, Ambagtsheer RC, Kitson A. Older adults' understandings and perspectives on frailty in community and residential aged care: an interpretive description. BMJ Open 2020 Mar 18;10(3):e035339 [FREE Full text] [doi: 10.1136/bmjopen-2019-035339] [Medline: 32193272]
25. Chan EY, Samsudin SA, Lim YJ. Older patients' perception of engagement in functional self-care during hospitalization: a qualitative study. Geriatr Nurs 2020;41(3):297-304. [doi: 10.1016/j.gerinurse.2019.11.009] [Medline: 31787364]
26. Terp R, Kayser L, Lindhardt T. "It is not rocket science.." - older peoples' understanding of nutrition - a qualitative study. Appetite 2021 Jan 01;156:104854. [doi: 10.1016/j.appet.2020.104854] [Medline: 32871203]
27. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016 Nov;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
28. Doolin B. Implementing e-health. In: Ferlie E, Montgomery K, Reff Pedersen A, editors. The Oxford Handbook of Health Care Management. Oxford, UK: Oxford University Press; 2016.
29. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
30. Overview of mandatory reporting. Danish National Center for Ethics. URL: https://researchethics.dk/information-for-researchers/overview-of-mandatory-reporting [accessed 2024-01-25]
31. OECD, Eurostat, UNESCO Institute for Statistics. ISCED 2011 Operational Manual: Guidelines for Classifying National Education Programmes and Related Qualifications. Paris, France: OECD Publishing; 2015.
32. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Rep 1985;100(2):126-131 [FREE Full text] [Medline: 3920711]
33. Strath SJ, Kaminsky LA, Ainsworth BE, Ekelund U, Freedson PS, Gary RA, et al. Guide to the assessment of physical activity: clinical and research applications: a scientific statement from the American heart association. Circulation 2013 Nov 12;128(20):2259-2279. [doi: 10.1161/01.cir.0000435708.67487.da] [Medline: 24126387]
34. Stamatakis E, Johnson NA, Powell L, Hamer M, Rangul V, Holtermann A. Short and sporadic bouts in the 2018 US physical activity guidelines: is high-intensity incidental physical activity the new HIIT? Br J Sports Med 2019 Sep;53(18):1137-1139 [FREE Full text] [doi: 10.1136/bjsports-2018-100397] [Medline: 30786998]
35. Rossen S, Kayser L, Vibe-Petersen J, Ried-Larsen M, Christensen JF. Technology in exercise-based cancer rehabilitation: a cross-sectional study of receptiveness and readiness for e-Health utilization in Danish cancer rehabilitation. Acta Oncol 2019 May;58(5):610-618. [doi: 10.1080/0284186X.2018.1562213] [Medline: 30698060]
36. Reiners F, Sturm J, Bouw LJ, Wouters EJ. Sociodemographic factors influencing the use of eHealth in people with chronic diseases. Int J Environ Res Public Health 2019 Feb 21;16(4):645 [FREE Full text] [doi: 10.3390/ijerph16040645] [Medline: 30795623]
37. Goyal S, Morita PP, Picton P, Seto E, Zbib A, Cafazzo JA. Uptake of a consumer-focused mHealth application for the assessment and prevention of heart disease: the Under 30 Days Study. JMIR Mhealth Uhealth 2016 Mar 24;4(1):e32 [FREE Full text] [doi: 10.2196/mhealth.4730] [Medline: 27012937]
38. Zhao YC, Zhao M, Song S. Online health information seeking behaviors among older adults: systematic scoping review. J Med Internet Res 2022 Feb 16;24(2):e34790 [FREE Full text] [doi: 10.2196/34790] [Medline: 35171099]
39. Tzeng HM, Okpalauwaekwe U, Lyons EJ. Barriers and facilitators to older adults participating in fall-prevention strategies after transitioning home from acute hospitalization: a scoping review. Clin Interv Aging 2020 Jun 25;15:971-989 [FREE Full text] [doi: 10.2147/CIA.S256599] [Medline: 32612356]
40. Vincenzo JL, Patton SK, Lefler LL, McElfish PA, Wei J, Curran GM. Older adults' perceptions and recommendations regarding a falls prevention self-management plan template based on the health belief model: a mixed-methods study. Int J Environ Res Public Health 2022 Feb 09;19(4):1938 [FREE Full text] [doi: 10.3390/ijerph19041938] [Medline: 35206125]
41. Takemoto M, Manini TM, Rosenberg DE, Lazar A, Zlatar ZZ, Das SK, et al. Diet and activity assessments and interventions using technology in older adults. Am J Prev Med 2018 Oct;55(4):e105-e115 [FREE Full text] [doi: 10.1016/j.amepre.2018.06.005] [Medline: 30241621]
42. Venkatesh V, Bala H. Technology acceptance model 3 and a research agenda on interventions. Decis Sci 2008 May 09;39(2):273-315. [doi: 10.1111/j.1540-5915.2008.00192.x]
43. Sandlund M, Pohl P, Ahlgren C, Skelton DA, Melander-Wikman A, Bergvall-Kåreborn B, et al. Gender perspective on older people's exercise preferences and motivators in the context of falls prevention: a qualitative study. Biomed Res Int 2018 Jul 18;2018:6865156 [FREE Full text] [doi: 10.1155/2018/6865156] [Medline: 30112416]
44. Bunn F, Dickinson A, Barnett-Page E, Mcinnes E, Horton K. A systematic review of older people's perceptions of facilitators and barriers to participation in falls-prevention interventions. Ageing Soc 2008 Nov 03;28(4):449-472. [doi: 10.1017/S0144686X07006861]
45. Avgerinou C, Bhanu C, Walters K, Croker H, Liljas A, Rea J, et al. Exploring the views and dietary practices of older people at risk of malnutrition and their carers: a qualitative study. Nutrients 2019 Jun 05;11(6):1281 [FREE Full text] [doi: 10.3390/nu11061281] [Medline: 31195731]
46. Wegener EK, Kayser L. Smart health technologies used to support physical activity and nutritional intake in fall prevention among older adults: a scoping review. Exp Gerontol 2023 Oct 01;181:112282 [FREE Full text] [doi: 10.1016/j.exger.2023.112282] [Medline: 37660762]

47. Spiteri K, Broom D, Bekhet AH, de Caro JX, Laventure B, Grafton K. Barriers and motivators of physical activity participation in middle-aged and older-adults - a systematic review. J Aging Phys Act 2019 Sep 01;27(4):929-944. [doi: 10.1123/japa.2018-0343] [Medline: 31141447]
48. Elliott S, Painter J, Hudson S. Living alone and fall risk factors in community-dwelling middle age and older adults. J Community Health 2009 Aug;34(4):301-310. [doi: 10.1007/s10900-009-9152-x] [Medline: 19333744]
49. United Nations E-Government Survey 2022: The Future of Digital Government. New York, NY: United Nations; 2022.
50. ChatGPT 3.5. OpenAI. URL: https://chat.openai.com/ [accessed 2024-04-29]

Abbreviations

READHY: Readiness and Enablement Index for Health Technology

Edited by P Santana-Mancilla; submitted 08.09.23; peer-reviewed by M Sääskilahti, J Wilson; comments to author 08.01.24; revised version received 04.03.24; accepted 15.03.24; published 08.05.24.

Please cite as:
Kikkenborg J, Magelund E, Riise MS, Kayser L, Terp R
Knowledge, Skills, and Experience With Technology in Relation to Nutritional Intake and Physical Activity Among Older Adults at Risk of Falls: Semistructured Interview Study
JMIR Hum Factors 2024;11:e52575
URL: https://humanfactors.jmir.org/2024/1/e52575
doi:10.2196/52575
PMID:38717810

Original Paper

Health Technology Access and Peer Support Among Digitally Engaged People Experiencing Homelessness: Qualitative Study

Nóra Radó[1,2*], BA, MA; Sándor Békási[1,2*], MD; Zsuzsa Győrffy[1*], PhD

[1]Institute of Behavioral Sciences, Faculty of Medicine, Semmelweis University, Budapest, Hungary

[2]Health Center, Hungarian Charity Service of the Order of Malta, Budapest, Hungary

[*]all authors contributed equally

Corresponding Author:
Nóra Radó, BA, MA
Institute of Behavioral Sciences
Faculty of Medicine
Semmelweis University
Üllői út 86
Budapest, 1085
Hungary
Phone: 36 1 210 2953
Email: rado.nora@phd.semmelweis.hu

Abstract

Background: Although the effects of digital health are receiving wide scientific attention, very little is known about the characteristics of digitally engaged people experiencing homelessness, especially in Central and Eastern Europe. Our previous research revealed a considerable level of internet use in the homeless population of Budapest, Hungary, for general purposes (350/662, 52.9%) and medical purposes (229/664, 34.6%). Moreover, a digitally engaged subgroup was identified (129/662, 19.5%).

Objective: The aim of this exploratory study was to map out the resources, attitudes, and behaviors of digitally engaged homeless individuals in relation to digital technology to set the basis for potential health policy interventions, which will enable better access to health services through strengthening of the digital components of the existing health care system.

Methods: Between August 18, 2022, and October 27, 2022, a total of 12 in-depth semistructured interviews were conducted in 4 homeless shelters in Budapest, Hungary. Upon analysis by 3 independent evaluators, 2 interviews were excluded. The interviewees were chosen based on purposive sampling with predefined inclusion criteria. Thematic analysis of the transcripts was conducted.

Results: In the thematic analysis, 4 main themes (attitude, access, usage patterns, and solutions for usage problems) emerged. Health-related technology use mostly appeared in health information–seeking behavior. Online search for prescribed medications (5 interviews), active ingredients of medications (4 interviews), medicinal herbs believed to replace certain pills (2 interviews) or foods, and natural materials (1 interview) were mentioned. Moreover, mobile health app use (3 interviews) was reported. The intention to circumvent or check on mainstream health care solutions was mainly associated with previous negative experiences in the health care system. Several gaps in the daily use of technology were identified by the interviewees; however, more than half of the interviewees (6/10) turned out to be contact points for their peers for digital problem-solving or basic digital literacy skill enhancement in the homeless shelters. Furthermore, a lack of institutional support or special programs targeting senior clients was noted.

Conclusions: Digitally engaged homeless individuals might become mediators between their peers and comprehensive digital health programs. They have the trust of their peers, can recognize and harness the benefits of digital technology, and are able to provide meaningful help in technology- and usage-related issues through experience. Digital health services have great promise in community shelters for managing and preventing health issues, and digitally engaged individuals might be important for the success of such services.

(JMIR Hum Factors 2024;11:e55415) doi:10.2196/55415

KEYWORDS
digital health; homelessness; digital technology; internet; access; health equity

Introduction

The Digital Health Paradox

By the end of 2022, the number of mobile service subscribers climbed to over 5.4 billion people globally, including 4.4 billion people who also used mobile internet, and the usage gap has narrowed markedly in the last 5 years (from 50% in 2017 to 41% in 2022 on average) [1]. As of October 2023, around 5.3 billion people use the internet worldwide, which is equivalent to 65.7% of the total population of the world, and in the last year, 189 million new members joined the global community of internet users [2]. These are unprecedented numbers. Digitization, especially the adoption of digital health technologies at scale, has been boosted by the COVID pandemic since 2020, promising access to health care systems and beneficial health outcomes.

However, there is a growing body of evidence indicating that greater reliance on digital tools has the potential to widen the gap between those who have digital skills and access to digital tools and those who do not, thereby increasing already existing health inequalities [3]. Although digital solutions might be designed following guidelines, such as the World Health Organization (WHO) Global Strategy on Digital Health 2020-2025, which states that "digital health should be an integral part of health priorities and benefit people in a way that is ethical, safe, secure, reliable, equitable, and sustainable," certain groups are unintentionally left out of the digitization boom [4]. Paradoxically, these groups often represent patients with complex psychosocial needs, specific sociodemographic characteristics, and multiple chronic conditions, and they would benefit the most from the use of digital health technologies [5-8]. van Kessel et al [6] have referred to this as the digital health paradox.

Vulnerable Groups, Homelessness, and Health Disparities

The abovementioned groups might represent vulnerable populations that are already experiencing negative health outcomes due to their detrimental social determinants of health. This has been defined by the WHO as "the forces and systems shaping the collective conditions in which people are born, grow, work, live, and age, as well as the conditions of their daily lives" [9], and they are shaped by the distributions of money, power, and other resources [10]. Emerging research shows that there is a strong relationship between socioeconomic factors, geography, demographics, and health, with poverty, housing problems, food insecurity, abuse, gender, and ethnicity creating chronic stress, which can leave the human organism with maladaptive mechanisms that result in damage to the body's functioning systems [11,12]. These have been linked to hypertension, premature aging, cardiovascular disease, type 2 diabetes, stroke, cancer, pulmonary disease, kidney disease, and many other health problems [10,13].

In the case of people experiencing homelessness, a complex set of social determinants of health are at play, which amplify each other's impacts and leave this vulnerable group at the extreme low end of health outcomes, health care access, and health literacy. According to previous research, living without adequate housing options is associated with significantly higher rates of bacterial and viral infections, diabetes, hypertension, cardiovascular disease, mental health issues, and problematic substance use compared to populations with adequate housing options [14-16]. The COVID-19 pandemic has also increased the vulnerabilities and health risks of people experiencing homelessness [17].

Life expectancy data for people experiencing homelessness compared to the general population also support these findings. In a systematic review, Aldridge et al [18] found that socially excluded populations have an 8 times higher mortality rate for men and 12 times higher rate for women than the average population. In Western high-income countries, studies have shown that homelessness is an independent risk factor for mortality, and life expectancy varies between 50 and 65 years on average [19].

When considering health care access, homeless populations frequently experience structural barriers to obtain health care, including lack of health insurance in countries without universal health insurance, as well as competing interests in health care settings to their disadvantage alongside their own financial difficulties and competing priorities, which might lead them to secure food and accommodation before health care [17,20]. Research has also shown mistrust of health care systems and experiences of discrimination in care settings. Poorer health literacy measured among people experiencing homelessness compared to the general population might also lead to a poor self-rated health status and less adherence to medical recommendations and prescription medicines [21].

Digital Health and People Experiencing Homelessness

Previous research has shown that people with lower socioeconomic status are slower to adopt new technology, and the rates of smartphone and internet use among people experiencing homelessness were lower than the rates among those with similarly low socioeconomic status but more stable housing [22]. VonHoltz et al [23] found that while experiencing homelessness, study participants showed a 68% reduction in their likelihood to access the internet compared to when they were housed. However, in terms of preferences, it was found that low-income populations, including people experiencing homelessness, rely on smartphones rather than computers for internet access owing to cost considerations, portability, and storage issues [24]. Populations at risk for limited health literacy, as indicated in the case of the homeless populations above, are also at risk for having challenges with digital technology [25].

Previous research has mentioned that it would be beneficial to equip people experiencing homelessness with the necessary tools to get them involved in digital health ecosystems as the costs of inclusion are significantly lower than the costs of treatment of health conditions, and the overall benefits show significance and persistence [3].

Digital Health and Homelessness: Research in Hungary

While the associations between people experiencing homelessness and their health status are well researched, especially in English-speaking countries, such as Canada, the United Kingdom, and the United States, a lot less is known about the access of people experiencing homelessness to digital health tools, their digital health literacy, their attitudes toward digital technologies, or their overall characteristics in different local settings, such as Hungary, and about the specific groups existing within homeless populations [26,27].

For these reasons, the Digital Health Research Group at Semmelweis University and the Hungarian Charity Service of the Order of Malta (HCSOM) have undertaken an overarching research agenda aiming to uncover the relations between digital health and homeless populations in Hungary. Digital health technologies are defined as "technologies which use computing platforms, connectivity, software, and sensors for health care and related uses" [5]. Previous research has mapped out the attitudes of people experiencing homelessness in Budapest, Hungary, toward telecare services, with the main finding that trust in the general health care system is the central issue when it comes to the decision of homeless populations about whether they have trust in telecare services as well [28]. This study served as a starting point for a pilot project assessing the viability of a telecare system for homeless populations [29].

Access to digital tools and digital health literacy were measured in another survey (n=662), where the results demonstrated that a significant proportion of people experiencing homelessness in Budapest, Hungary, were using the internet (52.9%), while the proportion was 81.3% in a representative sample of the Hungarian population that was used as a reference group [30]. Moreover, 69.6% of people experiencing homelessness reported mobile phone ownership, with 39.9% adding that their phone had a smartphone function and 34.6% mentioning that they have already used the internet for medical purposes [30]. In terms of self-rated digital health literacy, 24.5% rated themselves as experienced or very experienced regarding internet use, while 21.5% self-reported having mediocre experience [30].

Based on these access and skill-related characteristics, we were able to filter out a broadly defined digitally engaged group (n=129, 19.5%). This subgroup possessed their own digital tools, had some level of digital health literacy, and was partly using these digital tools for health-related reasons. When we analyzed the group and ran chi-square tests for gender, age, education, frequency of medical visits, prevalence of chronic illnesses, shelter type, and social services, the prevalence of chronic illnesses (P=.047) was found to be an associative factor in this subgroup for the likelihood of using the internet frequently for health-related reasons. However, the quantitative survey could not discern more relevant information [30].

Thus, the main aim of this study was to map out the characteristics of this specific subgroup in order to determine (1) for what purposes and (2) how the individuals in this subgroup are using digital health technologies in the framework of an exploratory qualitative analysis.

Methods

Checklist

Our methodology is based on the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist as well as the methodological framework of Győrffy et al [31] (Multimedia Appendix 1). For data collection, 12 semistructured interviews were conducted.

Ethics Approval

For all interviews, written informed consent statements were obtained, and ethics approval for the study was issued by the Scientific Research Ethics Committee of the Medical Research Council of Hungary (TUKEB 133/2020 and IV/10927/2020/EKU). In terms of the analytical framework, thematic analysis was chosen.

Recruitment

Purposive sampling was based on the following criteria: (1) presence in the social care system of the Charity Service of the Order of Malta, (2) use of the internet every second week or more frequently, (3) internet access with own smartphone, computer, or tablet or another device with a data contract, a pay-as-you-go facility, or free Wi-Fi, (4) self-rating of an average or more competent internet user, and (5) ever use of the internet for health-related reasons. The sampling criteria of this research and the filtering criteria for the broadly defined digitally engaged subgroup in our previous research matched [30]. However, the previous research involved anonymous data collection, and the present purposive sampling did not use the previous data pool as a starting point. Thus, there may or may not be an overlap between the 2 groups.

Malterud et al [32] theorized that information power can determine the ideal sample size for qualitative studies, with a sample holding more information requiring a lower number of participants. They enlisted the following 5 criteria for analyzing information power: (1) aim of the study, (2) sample specificity, (3) use of an established theory, (4) quality of dialogue, and (5) analysis strategy. In this case, the aim of the study was to assess the specific characteristics of a subgroup of people experiencing homelessness who have a digital skillset and usage pattern (see Multimedia Appendix 2 for the interview guide), thus creating a very specific sample with limited prevalence in the overall population as measured in our previous study [30]. As a result, a smaller sample size was chosen.

In the research process, 12 interviews were conducted, but in the final analysis, 10 interviews were included, which presented all the criteria of the purposive sampling specified above. Two interviews did not contain any reference to digital health usage. At this point, this might seem as a contradiction, but people experiencing homelessness may experience literacy issues, may have somewhat limited understanding due to health issues, and may have a risk of social desirability bias in relation to interview situations, which may result in self-contradictory statements, opinions, and behaviors, in line with previous methodological findings in relation to this vulnerable population [33].

Data Collection

Interviewees were contacted by social workers or institutional assistants at 4 shelters in the social care system of HCSOM or partner institutions. These shelters either served as a night shelter (n=1) or provided accommodation on a 24/7 basis (n=3) in Budapest, Hungary.

Based on the recommendations of the social workers or institutional assistants, one-on-one semistructured interviews were conducted between August 18 and October 27, 2022.

The interview guide was developed from experiences of the previous research, the specific study aims, and a literature review. The interviews were conducted in Hungarian with a trained interviewer. The interview guide was checked on a smaller sample of the specific subgroup (n=2) and modified based on their initial feedback.

The interview guide was based on the following topics: access to and attitude toward the health care system in general, access to and attitude toward digital tools in general and usage patterns of the internet and digital tools, and access to and attitude toward digital health and usage patterns of the internet and digital tools for health-related reasons (see Multimedia Appendix 2 for the complete interview guide).

Interviews were audio recorded in person, with an average interview length of 30 minutes. All audio-recorded interviews were transcribed verbatim, and each transcript was anonymized and assigned a unique code. The interviewer checked the transcriptions for accuracy. They were not sent back to the interviewees because people experiencing homelessness struggle with literacy challenges and Thomas et al [34] argued that evidence does not support the idea that member checking increases the credibility or trustworthiness of qualitative data [34].

Analysis

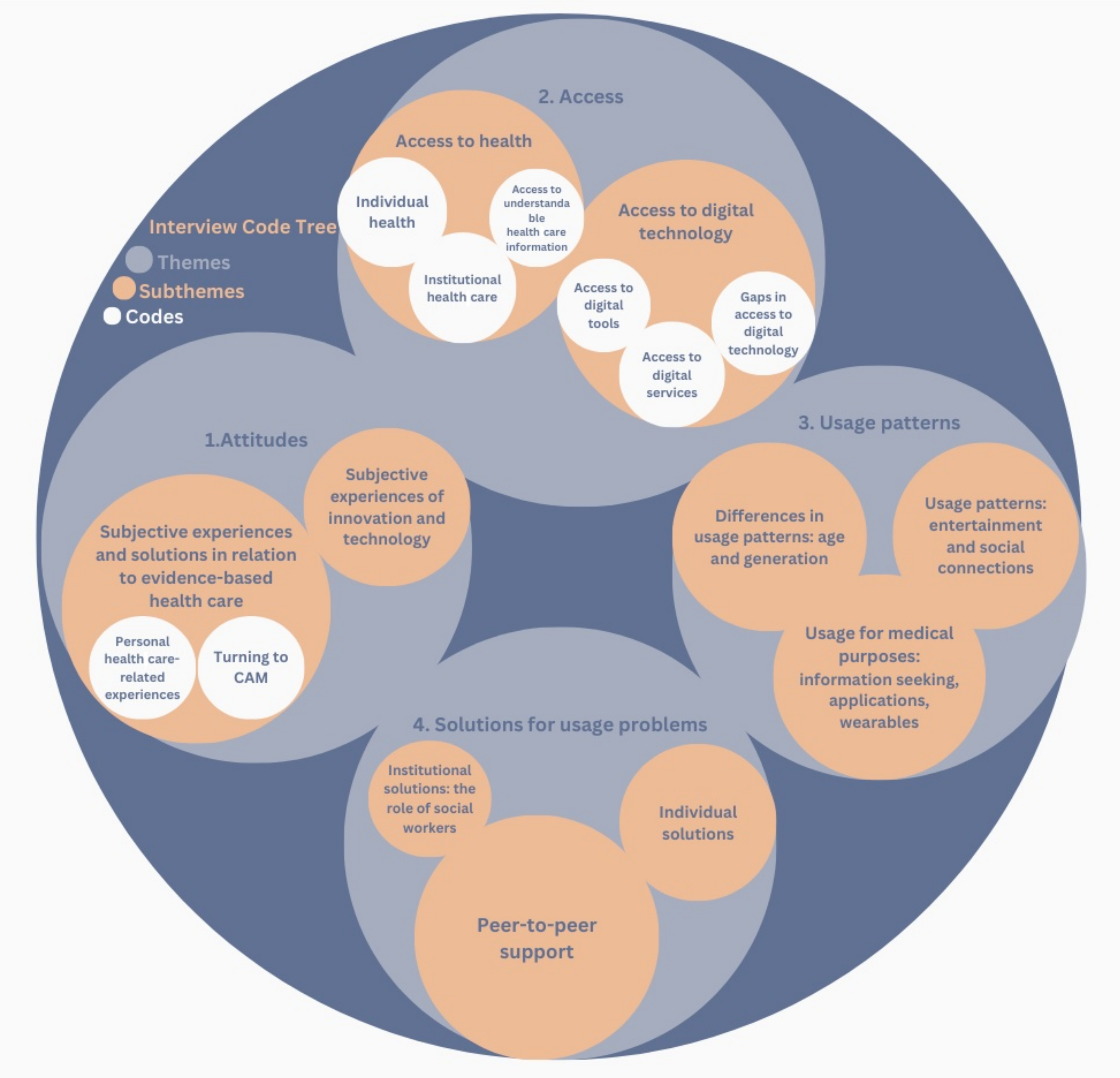
Thematic analysis as described by Braun and Clarke was chosen as an analytical and theoretical framework [35]. In coding, we followed the “theoretical” technique in an essentialist or realist method, driven by the analytic interest to report about the experiences and realities of the study participants in relation to their engagement in a digital health ecosystem. In coding, we followed the deductive technique, that is, we worked with predetermined assumptions and themes, which followed the interview guide; however, clearly characterizable subthemes emerged around the previously identified main themes. Three independent researchers (ZG, SB, and NR) read and analyzed the data and discussed their findings.

A theoretical thematic approach was used to analyze the data and identify patterns of themes based on the checklist elaborated by Braun and Clarke [35]: (1) familiarizing with the content of the data, taking notes, and making ideas for coding based on previous assumptions and following the interview guide, (2) generating initial codes manually, (3) identifying and indexing different codes across the data set manually, (4) creating relationships between the themes and subthemes, (5) defining, mapping, and naming themes, and (6) interpreting the results.

The 3 researchers discussed and developed all themes and subthemes and clarified any discrepancies during the coding. Afterwards, they laid out the final thematic map in mutual agreement. The results are supported by participants’ anonymized quotes. Interview IDs are provided for all quotes. For each interview ID, the letter indicates the first letter of the shelter where the interview was conducted (M, Miklós utca; F, Feszty; B, Budaörs; R, REVIP) and the number indicates the serial number of the interview.

For an overview of the themes, see Figure 1.

Figure 1. Interview code tree. CAM: complementary and alternative medical solutions.



Results

Demographic Characteristics

General demographic characteristics of the sample are presented in Table 1. In terms of gender, 6 male and 4 female participants were interviewed. Older age groups were overrepresented in the sample: 1 person was <40 years old, 4 people were 40-49 years old, 2 people were 50-59 years old, and 3 people were ≥60 years old. In terms of education, high school (4 people) and vocational school (3 people) were overrepresented, while 1 person had a university education and 2 people completed primary school or below.

Table 1. Demographic composition of the sample.

Characteristic	Value (N=10), n (%)
Gender	
Male	6 (60)
Female	4 (40)
Age (years)	
<40	1 (10)
40-49	4 (40)
50-59	2 (20)
≥60	3 (30)
Education	
Primary school or below	2 (20)
Vocational school	3 (30)
High school	4 (40)
University	1 (10)
Shelter	
HCSOM[a] Temporary Shelter (Feszty)	2 (20)
HCSOM Integrated Shelter (Miklós utca)	4 (40)
Shelter House Foundation's Night Shelter (Budaörs)	2 (20)
REVIP Baptist Integration Center (REVIP)	2 (20)

[a]HCSOM: Hungarian Charity Service of the Order of Malta.

Theme 1: Attitudes

Subtheme 1: Subjective Experiences and Solutions in Relation to Evidence-Based Health Care

Code 1: Personal Health Care–Related Experiences

Experiences with mainstream health care systems (hospitals, doctors, nurses, pharmacists, other medical personnel, prescription medicines, and pills) were mixed. In a minority of interviews (2/10), positive experiences with regard to access to care, quality of care, and how one was treated by the medical personnel were noted. However, the majority (7/10) reported negative experiences largely due to a negative attitude, stigmatization and mistreatment coming from the medical staff, and inadequacy of care. These signaled an overall negative attitude toward the general health care system.

> *I'm completely okay to be honest, I experienced that there are differences between the hospitals, I can only say that.* [Interview M05]
>
> *(…)…they notice where they have to go and then they have a completely different stance. Also, the emergency medical doctor, who is here, or if the ambulance services come. They behave completely differently. (…) They are condescending. Okay, we'll do it later. Okay, come back later. And another one: do pack your stuff already, we are set to go. So, they (…) are not helpful.* [Interview B09]

Code 2: Turning to Complementary and Alternative Medicine

In a minority of interviews (3/10), turning to complementary and alternative medical solutions, medicinal herbs, or Chinese medicine was noted, which was considered as an equivalent alternative of traditional Western medicine. In parallel, in 2 interviews, a negative stance toward drugs and medicines (mentioned within the same textual context) was noted.

These shed light on the fact that interviewees sought out different potential solutions to their medical problems as some of them experienced that health care systems and traditionally produced drugs cannot and have not so far provided them with appropriate solutions. They had taken medicinal herbs or trusted ingredients, which were recommended by a trusted person or were found online.

> *I am aware of that, I looked up the side effects, the medicines, I will not take what they prescribe. I have already played along for long. I rather drink an herbal tea.* [Interview F01]
>
> *I can feel if something's off in my body, and then I look up certain things, but to be honest, I always start with medicinal herbs, and not with pills. I go to the pharmacy, and I look up on the internet what is recommended for example for lower abdominal pain or for a story with joints.* [Interview M07]

Subtheme 2: Subjective Experiences of Innovation and Technology

Subjective experiences and attitudes toward novelties and technology were mixed. In almost half of the interviews (4/10), openness toward trying new programs and applications appeared, while in 2 interviews, a complete lack of interest was reported.

Attitudes toward the use of digital tools and the internet were also mixed. In part of the interviews, lack of trust and negative experiences were reported, for example, the risk of data misuse (1 interview), risk of making mistakes due to the autocomplete function and the speed of digital tools (1 interview), and inaccuracy of step counting (2 interviews). In another set of 4 interviews, openness toward trying new programs and applications appeared, while in a minority of interviews (2/10), lack of interest in this area was reported.

> *As I'm homeless at the moment, I don't have enough (money) on my pay as you go facility that I could use the internet unlimited. Where there is free Wi-Fi, I certainly search for things I think of or what I gather from my environment, or from my godchildren. So, I want to keep up with today's world in spite of the fact that I'm now a little bit on the brink of it.* [Interview M07]

> *You can really misuse data. I had that now, as well. Someone tapped into my bank account, abroad. I had to block access to my debit card, and I will have it done at some point.* [Interview F01]

> *As this device (tablet) works so that if my hand starts to shake just a little bit, and it gets close to it, it pulls in. And then, it writes something that I don't want to. So, I don't think that it is so reliable.* [Interview R12]

Theme 2: Access

Subtheme 1: Access to Health

Code 1: Individual Health

The majority of interviewees (7/10) self-evaluated their health status as average or worse. Chronic diseases (cardiovascular and heart problems, and type 2 diabetes), cancerous tumors, and lasting harm from injuries were characteristic for the group. In 4 interviews, managed alcohol problems were reported. Drug abuse was not mentioned, and in 2 interviews, aversion to drugs was noted. Diagnosed mental health problems were not mentioned.

Some interviewees regularly took medicines or mentioned that their doctors prescribed them certain types of medications, which they did not take. Some interviewees made decisions in medical matters based on their own opinions and beliefs without any professional evidence.

> *My troubles look like heart, liver, kidney, arterial obstructions. I had deep vein thrombosis in both legs but I carried that for long. I have a very high tolerance for pain. I usually operated on myself. I froze both of my legs and I cut the ulcer out as deep as I could. Then I put herbs into the wound. It recovered within 2 weeks.* [Interview F01]

Code 2: Institutional Health Care

Interviewees were clients of 4 homeless shelters in Budapest, meaning that they had institutional access to basic health care. Their legal social security status could be provided by social institutions on the grounds of homelessness under Hungarian law. Accessible health care services included primary care (prescription and dispensing of medicines, referral to specialists, and care work), publicly funded specialized outpatient care, inpatient (hospital) care, and rescue in case of emergency.

> *If I have any problem, the Maltesers (HCSOM) have a doctor's office. And if I can go there on my own feet, then I go there. If you can't, then you will be transported to the hospital by default. There are decent people who help or call an ambulance. In the doctor's office, they refer you to any specialist, no matter whether it's dermatology or cardiology. Thus, they can get you to any kind of specialist.* [Interview F03]

Code 3: Access to Understandable Health Care Information

In at least one interview, a lack of access to understandable health care information was reported, and several interviewees pointed out that they were seeking out medications and ingredients online with the help of digital tools in order to understand what impacts those materials had on their body. The need for understanding health-related information was noted in at least half of the interviews in certain forms, for example, they looked up prescription medicines (5 interviews) and their ingredients (4 interviews) online, and in at least one case, they did that for their family members as well.

> *(….) most of the time, physicians use such Latin words in general, as lawyers do. Make it simple! No one is that much overeducated to know these. For example, laboratory tests. They should include what does this mean, sodium was X. There are some apps where you can look that up.* [Interview F01]

Subtheme 2: Access to Digital Technology

Code 1: Access to Digital Tools

An overwhelming majority of interviewees (7/10) used smartphones. Notebook use was reported in 1 case, and tablet use was reported in further 2 cases. One interviewee reported power bank use to charge the device.

In a minority of interviews (2/10), it was reported that in times of need, phones, tablets, and computers were sold; thus, these were not permanently accessible tools.

> *In this living situation, people get such digital devices much easier off their hands, if they are not in such a whacking need of them, simply to be able to make money out of it.* [Interview F04]

Code 2: Access to Digital Services

In homeless shelters, interviewees had access to the computers in possession of the shelters, and through those devices, they could get access to the internet. In certain shelters, free Wi-Fi and the option to charge their phones were available.

The majority (6/10) used free Wi-Fi inside and outside of the shelters and looked actively for options of free Wi-Fi. They could afford subscription (3 interviews) or pay-as-you-go facilities (5 interviews) less frequently. In some cases, the interviewees reported that they visited cafes in order to be able to charge their phones or use the internet (3 interviews).

> *(…) the Wi-Fi is so strong that you don't have to go in and consume something, or if you go in and drink a cup of coffee or water, you get the Wi-Fi password, then sit in front of it on a bench, and it has such a strong signal that you can use it there as well, until it is open.* [Interview M07]

Code 3: Gaps in Access to Digital Technology

The interviewees reported both tool supply and network coverage as existing problems. Several interviewees mentioned the need for securing a device (smartphone) or asked about whether there was potential for decreasing the price of subscriptions and pay-as-you-go facilities. The presence of smart benches in public spaces was mentioned in 1 interview, and free Wi-Fi on trams and busses in Budapest was mentioned in another interview.

The computer park and Wi-Fi network coverage in the shelters were not mentioned as problems in the majority of interviews (7/10), and the idea of having more connectors in the building to allow easier charging surfaced in 1 interview.

> *Some support would be great so that a basic device could be ensured for them. And a separate health network, which is for free. For people who are ill. As there are these crisis helplines and these have green numbers.* [Interview F01]

> *… prices could be reduced (…) and for example such benches could be installed where phones can also be charged. And then you could use the Wi-Fi there.* [Interview F03]

> *It's very difficult, I would say there could be more charging stations. The bigger shopping malls are covered, that's fine, but what if you suddenly notice your phone is dead and you cannot go into any such places, or you are far (from the charging station), and a homeless cannot buy a ticket… How do you go there?* [Interview M05]

> *I would tell you the truth… I'm sure it would be feasible to have free Wi-Fi on busses and low-floor trams. So here, we have Wi-Fi, since this is a shelter but when we go 20 meters further, there isn't any, the network disconnects.* [Interview M05]

Theme 3: Usage Patterns

Subtheme 1: Differences in Usage Patterns: Age and Generations

Every interviewee used the internet at a measurable frequency on their own device. The age of the participants ranged from 35 to 69 years. The interviewer did not explicitly ask about usage characteristics by age, and the topic came up spontaneously in the case of 6 interviewees when talking about attitudes toward novelties.

In several cases, the interviewees mentioned generational differences in usage, characterizing the older generation as less involved in the digital world and less interested in novelties, while younger people were considered to be already born with digital devices, and their usage seemed to be self-evident. In 1 interview, it appeared that if there was individual motivation, then age would not pose a hinderance with regard to usage.

> *This is a fundamental thing, really, but many don't know, especially the older generation. (…) So, I'm quite digital, but I'm only 40 years old for that matter. We grew up on these devices more or less already.* [Interview F03]

> *(…) I think this is age-dependent, thus generation-dependent. The elderly are okay with their basic phones. When it rings, they pick it up, then put it down. My generation already needs it more, we use it more often and the younger even more, they don't even put it down.* [Interview M05]

Subtheme 2: Usage Patterns: Entertainment and Social Connections

Interviewees mainly used the internet for entertainment and maintaining their social relationships. Watching movies, listening to music, reading e-books, and playing phone-based games were also reported. Seven interviews mentioned Facebook and 1 interview also mentioned X (or formally Twitter) as frequently used social media sites. A minority of interviews mentioned information gathering through Wikipedia (1 interview), reading news (1 interview), and online banking (1 interview) as use cases.

> *I watch movies, and look up e-books, in a topic that I'm interested in. Mostly self-healing, quantum healing and such banalities.* [Interview F01]

> *I had a smartphone, so not only the music, YouTube, Facebook page is important to me, but also Wikipedia, where I can look up everything, or for example, I read a lot about various things, and the disease that I had. This is very important to me.* [Interview M05]

> *Interesting that I also keep in touch with my physician via e-mail. I had for example a CT scan, and then everything worked entirely online. I received my appointment and also the findings online. I also consider this a very positive thing, so that it is also in the cloud, and they can see it, the whole thing is much easier… I just give them my social security card (TAJ-card), and then I tell them what prescribed medication I want to have. So, I consider this absolutely positive.* [Interview M05]

Subtheme 3: Usage for Medical Purposes: Information Seeking, Applications, and Wearables

In several interviews, information seeking for medical purposes was reported. For example, interviewees looked up prescribed medications (5 interviews), active ingredients of medications

(4 interviews), medicinal herbs believed to replace certain pills (2 interviews) or foods, and natural materials (1 interview). One interviewee mentioned purchasing a product believed to have medicinal value online on the basis of a Facebook advertisement.

One interviewee in their 30s communicated with the doctor about health problems via email, provided information about their illness and the prescribed medicines online, and used a health app and a step counter. These tools (health app and step counter) were also mentioned by 2 other interviewees, but one of them stopped using the step counting option as they believed it was inaccurate.

> *I look up the active ingredient of a pill, for example when before chemotherapy certain medicines were prescribed for me, and I looked up what kind of active ingredients they have, what side effects could they have, because a package leaflet is one thing and a real person who already had this experience and took the medicine, and what is their opinion, is another thing.* [Interview M05]

> *I already had this step counting thing, this daily fitness thing. And I remember I had a heart rate monitor in my old Samsung S5, and now I really miss that my current phone doesn't have that anymore. (…) I also use a menstruation tracking app.* [Interview M05]

> *I usually look up online for my partner what kind of cremes and medicines there are … if they are interested what kind of ingredients the pill has, and due to his blood pressure.* [Interview M06]

Theme 4: Solutions for Usage Problems

Subtheme 1: Individual Solutions

We included interviewees in this study who previously stated that they frequently used digital tools and self-evaluated their skills as at least average. The majority of interviewees (8/10) themselves did not mention usage problems, and when they had problems, 1 interviewee asked their family members for help but added that they preferred to solve their problems on their own.

Subtheme 2: Peer-to-Peer Support

It was frequently (6 interviews) reported that the interviewees offered their help to other clients who lived with them in the same shelter if they had trouble around the usage of digital tools or the internet. They solved usage-related problems for their peers, such as registration of SIM cards, activation of pay-as-you-go facilities, antivirus actions for devices, problems around online programs like Facebook and Messenger, and questions around online purchases. These user troubles represented basic problems, and the majority of interviewees (8/10) had the knowledge and skills to solve them.

> *Last time they wanted to buy something online, and they asked my help in that. (…) Now one of the guys from the shelter came up to me how to activate the SIM card. And then I activated it for them. Such issues are always in need.* [Interview F04]

> *There were some who asked me how to log in, how to register with an email address, how can they make a Facebook profile. Then I helped first to make an email account and then to register with that. (…) I was happy that I could help and they accepted it gladly. And then I saw that they were using it very well, they were glued onto their screens and were happy about it.* [Interview M05]

> *Usually Facebook, Messenger, or when they cannot download a game. And there is an antivirus program on every smartphone with a broom icon but they don't know what that is. So, I tell them, pick it up and swipe with it. Clean it. And then they look at me confused. Okay, give it to me. So, then I do it, and they look. Wow, then they say, it went down to zero. Yeah, and then I say that's the point, not to have anything on it. So there are always things like this.* [Interview B09]

Subtheme 3: Institutional Solution: Role of Social Workers

Interviewees did not report institutional solutions aiming at the development of digital skills. In 1 interview, a social worker was mentioned who provided the client with basic information on tablet use. In this case, it was the individual initiative of the social worker and not an element built into the given institution's services.

> *(…) then the social worker came up to me, and taught me the basics, and then they said that I should now keep pressing the buttons around nicely, and then I'll figure everything out by myself.* [Interview R12]

Discussion

Digital Technology and People Experiencing Homelessness

Digital technologies show a general potential for improving patient outcomes. For example, Bruce et al [36] showed that both clinical and patient-centered care outcomes were significantly better with the use of mobile health technology among 2059 orthopedic patients. However, according to a systematic catalog on digital health systematic and scoping reviews, there is less specific evidence on equitable health care (16.7%) [37].

In relation to the homeless population and digital technology, Heaslip et al [26] identified in their systemic review that mobile technology has a measurable health impact on the homeless population directly and indirectly. As an indirect impact, maintaining relations with relatives and friends as well as the outside world through entertainment, movies, and music strengthened their social connectedness and elevated their self-esteem, which in turn can have a positive impact on their personal health [38]. For the direct health impact of digital technology, they found limited evidence, with the main areas being reminders for repeat prescriptions or health care appointments. However, Heaslip et al [26] mentioned that the homeless population appears to consider that digital technology has potential health benefits, mostly in terms of online health information support and appointment reminders.

Our results partly strengthen these findings. The interviewees in our digitally engaged homeless subgroup used their digital tools primarily for entertainment purposes and to maintain their personal relationships. In terms of health care, they used their devices as new channels to reach solutions for their health problems outside the conventional health care system and to search for health-related information. However, most interestingly and most importantly, the majority of interviewees (6/10) shared that this subgroup is supporting their peers in taking up digital skills and is helping them solve their usage- and device-related problems, and this behavior has a lot of untapped potential for widening digital health usage in the homeless population.

Health Care Needs and Personal Experiences

As indicated by the demographic characteristics, older and predominantly male interviewees shared their experiences. Consistent with the results from our previous studies [28-30], the majority of interviewees (6/10) reported multimorbidities [39,40] and having chronic diseases, such as cardiovascular diseases [41], type 2 diabetes, cancer, and permanent injuries. Older age (≥50 years) was associated with worse physical health in the homeless population, which was noted in the interviews, as the self-reported health status was regarded as average or worse [19].

In our small sample, there was no mention of mental health problems other than addictions. Previous research found that the ratio of serious mental disorders among people experiencing homelessness in Hungary was very high [42], which is in line with findings from Western countries [43]. Underdiagnosis and undertreatment of mental health problems caused by stigmatization and underperformance of the Hungarian care system might be prevalent among our interviewees as well [44]. Moreover, in line with previous studies, which estimated the prevalence of alcohol abuse at 8.5%-58.1% [45], treated alcohol problems were noted in 4 interviews; however, illicit drug use or treated drug abuse problems were not mentioned. A systematic review found that alcohol abuse is more prevalent in mainland Europe [43].

Issues of Access to Health Care and Digital Tools

Access to primary care is resolved via the care settings of the Health Center of the HCSOM, which includes prescribing drugs, providing basic care services, and referring clients to specialists. In line with previous studies [17,20], the experiences of interviewees with accessing health care were mixed.

When looking at access to digital tools and digital services, in line with previous research, the majority of interviewees (7/10) had smartphones, which are more accessible to people with a low socioeconomic status [24]. The partial accessibility of digital devices and their use as assets in times of need as described in a minority of interviews (2/10) have been mentioned by Heaslip et al [26]. As a need, device supply was primarily mentioned by the participants, and this is in line with our previous study where 21.4% of respondents mentioned lack of a smartphone as the main barrier for not using the internet and 24.1% mentioned that availability of an appropriate device would help them use the internet more [30].

Digital services, such as computers of the shelters, were available to the participants, and in some shelters, free Wi-Fi or charging was also provided. The majority of participants (6/10) looked for free Wi-Fi options outside the shelters as well. One interviewee mentioned the lack of free Wi-Fi on public transport services and the lack of installation of smart banks in Budapest as barriers to usage. Such infrastructural problems were mentioned as causes of nonusage by 7.6% of respondents in our previous study [30]. On the other hand, several interviewees mentioned using the paid services of cafes to charge their phones or use Wi-Fi.

Several interviewees also mentioned the need for a potential decrease in internet service prices or device prices, which is in line with the finding of our previous study where 18.4% of participants said that better access to free Wi-Fi, pay-as-you-go facilities, or data contracts would help them use the internet more [30].

Problems Around Trust

Some interviewees mentioned the feeling of being unwelcome in conventional health care settings, which is in line with previous research [41]. Some of them mentioned difficulties in getting appropriate treatment and a negative attitude from health care personnel, which might negatively influence their desire to seek health care in the future and their overall trust in the health care system, and this might explain their turn away from mainstream health care solutions.

These aspects might include a negative impact on medication adherence and an overall mistrust in mainstream medical solutions, such as taking antibiotics and chronic disease drugs, with a turn to alternative solutions. From the interviews, it was found that managing treatment themselves instead of relying on medical personnel based on their own beliefs without medical evidence was a solution. Moreover, turning to alternative and complementary medical solutions, such as homeopathy, herbal medicine, and Chinese medicine, was a way to express mistrust in conventional care settings, and digital solutions can open up a channel outside of the conventional health care system to reach such alternative solutions.

Mistrust and negative attitudes toward the health care system coupled with the need for understanding health-related language, prescription drugs, and active ingredients were associated with the main health-related use of digital tools and services in the majority of interviewees (8/10).

Age as a Predictor for Usage and Openness

When asked about usage patterns, several interviewees spontaneously shared their views on how age differences matter in usage prevalence, outlining that older generations might be less involved and less interested in novel technologies. Several studies, including our previous quantitative research, support that age is a key sociodemographic variable that has an impact on use [29,30,46,47]. Our quantitative data showed that in access to technology, age did not seem to be a key factor; however, it might be considered as a significant factor when self-evaluating competence in digital literacy skills. This appeared in at least one of the interviews, with the respondent explaining less elevated technological skills with age.

At least three interviews indicated that age was associated with openness toward or willingness to try new technologies, which might be in line with the findings of a representative questionnaire survey (n=1500) on digital health–related knowledge, attitudes, and needs [46]. This survey was completed in 2021 and found that a quarter (26.5%) of individuals aged 65-74 years and a third (31.9%) of individuals aged older than 75 years would not like to try digital technologies in the coming years [46].

Lack of Systematic Support Results in Peer Support for Skill-Related Problems

While interviewees recognized some support from shelters in solving infrastructural and service-related technology issues, there was a perceivable lack of systematic solutions when it came to usage-related problems and digital literacy issues. Only 1 interviewee mentioned that a social worker helped them set up their tablet and navigate through basic usage scenarios.

As we selected interviewees based on at least average self-reported digital health literacy skills and aptitude toward digital technology, with some demonstrating previous educational or professional background in IT services, their less digitally skilled peers turned to them for help.

The majority of interviewees (6/10) provided unintentional peer support in relation to technology usage issues, solved technology-related problems, and provided guidance for future scenarios. Peer support, also in this context, is defined in the literature as a process whereby individuals with lived experiences of a particular phenomenon provide support to others by explicitly drawing on their personal experiences [48]. Intentional peer support works as a formalized framework of this process that is fostered and developed by institutions, while unintentional peer support remains under the radar of institutions. The literature recognizes the potential of peer support and peer support workers, who have the necessary training and provide intentional support to their homeless peers by sharing their lived experiences in different areas of life, and members of this digitally engaged subgroup might show potential for offering peer support in digital upskilling [48,49]. Moreover, anyone considering a comprehensive digital health program for homeless groups in Hungary that concentrates on offering solutions to infrastructure and skill-related problems should take into account the untapped potential of members of digitally engaged subgroups. These individuals, through their elevated trust levels among peers, might provide better outcomes in digital upskilling than official and institutionalized digital health literacy programs. A systematic review found that empowerment and self-esteem in the homeless population increased when working with homeless peers as mentors and educators, and that peer support in general facilitates acceptance of illness and recovery and increases efficacy, social skills, and coping [50].

Strengths

Through the qualitative analytical framework, the characteristics of a unique subgroup of digitally engaged people experiencing homelessness could be explored in a less studied area of digital health for equitable health care, where systematic mapping of review studies showed notable gaps in evidence [37].

The study aimed to enrich the still relatively small body of research concerning the characteristics, including the digital health–related characteristics, of the homeless population in Central and Eastern Europe. In North America and Western Europe, where the majority of studies involving the homeless population are conducted, the demographic composition of such populations as well as the health care system may differ significantly from Hungarian experiences, with different problems and solutions at individual and systemic levels.

Limitations

Our study has certain limitations. As a qualitative study using in-depth semistructured interviews, the sample size was small, and this should be taken into account when drawing inferences. The study participants represented the urban homeless population from Budapest, Hungary, where socioeconomic conditions might differ from those in the countryside. The recruited homeless people had a living connection to the social infrastructure; therefore, rough sleepers and other people who were not connected to any social initiatives were not represented. The research team exclusively relied on self-reporting of digital tool access and use, and did not attempt in any way to verify these reports (eg, via phone bills, direct observation, and other methods).

In relation to people experiencing homelessness, there is an increased risk of social desirability bias when conducting interviews, meaning that respondents tend to modify their responses in the presence of an interviewer perceived to be in a different socioeconomic and overall social situation than their own [51].

Conclusions

People experiencing homelessness can face many barriers when accessing digital technologies, including lack of appropriate devices, lack of operating infrastructure (eg, free Wi-Fi hotspots), some blind spots regarding digital skills, and a general lack of interest due to the prioritization of other basic life-supporting drives. However, in spite of all these barriers, our previous research identified a digitally engaged homeless subgroup in Budapest, Hungary, whose behaviors, usage, and access patterns were mapped in this study [30].

We found that the majority of participants (7/10) possessed a smartphone and used the often scarce pool of free Wi-Fi and the infrastructural capabilities of the shelters. Based on their articulated needs, various policy recommendations might be formulated for telephone companies and government agencies or support services. Telephone companies may consider subsidy programs to support mobile ownership and data services for this vulnerable population, as well as specific discount packages and more publicly available recharge options, as these would greatly support this group that is often in crisis and need. Government agencies may consider strengthening the infrastructural background of shelters and making free Wi-Fi accessibility an option in more public places, such as busses and piazza places, which could greatly reduce the access issues of this population. Institutional aid for accessing services and digital tools may also offer a viable option for people experiencing homelessness. A higher digital accessibility of an institution in terms of both infrastructure and digital literacy is associated with a greater likelihood of an increase in the number of digitally engaged people experiencing homelessness.

In terms of usage patterns, digitally engaged people experiencing homelessness use digital tools as an alternative information point beyond mainstream health care channels, which gives them access to check information originating from mainstream health care personnel and to seek out complementary and alternative medical solutions. These might be related to low trust in mainstream health care solutions, which might be enhanced through appropriately tailored comprehensive digital health programs. These programs could include awareness raising programs on trusted online health information sources, digital literacy and health literacy enhancing programs, and other programs to enhance their general trust in evidence-based health and the health care system.

Our most important finding is that digitally engaged homeless individuals have an aptitude for technology, and they are ready and eager to share their knowledge with their peers. This could elevate them to the role of a mediator between their peers and any potential comprehensive digital health program. Digitally engaged individuals have the trust of their peers, recognize the benefits of digital technology, and are able to provide meaningful help in technology- and usage-related issues. Thus, with appropriate training, they might become tutors for upskilling people experiencing homelessness, building a bridge between their peers and digital technologies as well as digital health ecosystems. These well-informed technologically able peers might also help enhance trust in the general health care system if their peer-to-peer support could be steered toward peer-to-peer recommendations of trusted health information sources via a specific institutional program.

Overall, our previous research showed that digital health services have great promise in community shelters for managing and preventing health issues [29,30], and this study found that digitally engaged individuals might be important for the success of such services.

Acknowledgments

This research project was supported by the Hungarian National Research, Development and Innovation Office (OTKA-FK 134372).

Conflicts of Interest

None declared.

Multimedia Appendix 1

COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.

[PDF File (Adobe PDF File), 126 KB - humanfactors_v11i1e55415_app1.pdf]

Multimedia Appendix 2

Interview guide.

[PDF File (Adobe PDF File), 57 KB - humanfactors_v11i1e55415_app2.pdf]

References

1. The mobile economy 2023. GSMA. URL: https://www.gsma.com/solutions-and-impact/connectivity-for-good/mobile-economy/wp-content/uploads/2023/03/270223-The-Mobile-Economy-2023.pdf [accessed 2023-11-02]
2. Kemp S. Digital 2023 october global statshot report. DataReportal. URL: https://datareportal.com/reports/digital-2023-october-global-statshot [accessed 2024-04-30]
3. Sieck CJ, Sheon A, Ancker JS, Castek J, Callahan B, Siefer A. Digital inclusion as a social determinant of health. NPJ Digit Med 2021 Mar 17;4(1):52 [FREE Full text] [doi: 10.1038/s41746-021-00413-8] [Medline: 33731887]
4. Global strategy on digital health 2020-2025. World Health Organization. 2021. URL: https://www.who.int/publications-detail-redirect/9789240020924 [accessed 2024-04-30]
5. Ziebland S, Hyde E, Powell J. Power, paradox and pessimism: On the unintended consequences of digital health technologies in primary care. Soc Sci Med 2021 Nov;289:114419. [doi: 10.1016/j.socscimed.2021.114419] [Medline: 34619631]
6. van Kessel R, Hrzic R, O'Nuallain E, Weir E, Wong BLH, Anderson M, et al. Digital health paradox: International policy perspectives to address increased health inequalities for people living with disabilities. J Med Internet Res 2022 Feb 22;24(2):e33819 [FREE Full text] [doi: 10.2196/33819] [Medline: 35191848]
7. Sanders R. Digital inclusion, exclusion and participation. ESSS Outline. 2020. URL: https://www.iriss.org.uk/resources/esss-outlines/digital-inclusion-exclusion-and-participation [accessed 2024-04-30]
8. Menendez ME, Moverman MA, Puzzitiello RN, Pagani NR, Ring D. The telehealth paradox in the neediest patients. J Natl Med Assoc 2021 Jun;113(3):351-352 [FREE Full text] [doi: 10.1016/j.jnma.2020.09.144] [Medline: 33092857]

9. WHO Regional Office for South-East Asia. Social determinants of health. World Health Organization. 2008. URL: https://iris.who.int/handle/10665/206363 [accessed 2024-04-30]
10. Lathrop B. Moving toward health equity by addressing social determinants of health. Nurs Womens Health 2020 Feb;24(1):36-44. [doi: 10.1016/j.nwh.2019.11.003] [Medline: 31911097]
11. Braveman P, Gottlieb L. The social determinants of health: it's time to consider the causes of the causes. Public Health Rep 2014;129 Suppl 2(Suppl 2):19-31 [FREE Full text] [doi: 10.1177/00333549141291S206] [Medline: 24385661]
12. Steptoe A, Marmot M. The role of psychobiological pathways in socio-economic inequalities in cardiovascular disease risk. Eur Heart J 2002 Jan;23(1):13-25. [doi: 10.1053/euhj.2001.2611] [Medline: 11741358]
13. Cockerham WC, Hamby BW, Oates GR. The social determinants of chronic disease. Am J Prev Med 2017 Jan;52(1S1):S5-S12 [FREE Full text] [doi: 10.1016/j.amepre.2016.09.010] [Medline: 27989293]
14. Schreiter S, Bermpohl F, Krausz M, Leucht S, Rössler W, Schouler-Ocak M, et al. The prevalence of mental illness in homeless people in Germany. Dtsch Arztebl Int 2017 Oct 06;114(40):665-672 [FREE Full text] [doi: 10.3238/arztebl.2017.0665] [Medline: 29070426]
15. Zhang L, Norena M, Gadermann A, Hubley A, Russell L, Aubry T, et al. Concurrent disorders and health care utilization among homeless and vulnerably housed persons in Canada. J Dual Diagn 2018;14(1):21-31. [doi: 10.1080/15504263.2017.1392055] [Medline: 29494795]
16. Graffy P, McKinnon S, Lee G, Remington P. Life outside: A narrative ethnographic inquiry into the determinants of homelessness. J Poverty 2018 Dec 13;23(3):202-228. [doi: 10.1080/10875549.2018.1550133]
17. Bedmar MA, Bennasar-Veny M, Artigas-Lelong B, Salvà-Mut F, Pou J, Capitán-Moyano L, et al. Health and access to healthcare in homeless people: Protocol for a mixed-methods study. Medicine (Baltimore) 2022 Feb 18;101(7):e28816 [FREE Full text] [doi: 10.1097/MD.0000000000028816] [Medline: 35363172]
18. Aldridge RW, Story A, Hwang SW, Nordentoft M, Luchenski SA, Hartwell G, et al. Morbidity and mortality in homeless individuals, prisoners, sex workers, and individuals with substance use disorders in high-income countries: a systematic review and meta-analysis. Lancet 2018 Jan 20;391(10117):241-250 [FREE Full text] [doi: 10.1016/S0140-6736(17)31869-X] [Medline: 29137869]
19. van Dongen SI, van Straaten B, Wolf JRLM, Onwuteaka-Philipsen BD, van der Heide A, Rietjens JAC, et al. Self-reported health, healthcare service use and health-related needs: A comparison of older and younger homeless people. Health Soc Care Community 2019 Jul;27(4):e379-e388 [FREE Full text] [doi: 10.1111/hsc.12739] [Medline: 31020738]
20. Liu M, Hwang SW. Health care for homeless people. Nat Rev Dis Primers 2021 Jan 14;7(1):5. [doi: 10.1038/s41572-020-00241-2] [Medline: 33446661]
21. Odoh C, Vidrine JI, Businelle MS, Kendzor DE, Agrawal P, Reitzel LR. Health literacy and self-rated health among homeless adults. Health Behav Res 2019 Oct;2(4):13 [FREE Full text] [doi: 10.4148/2572-1836.1055] [Medline: 34164607]
22. Raven MC, Kaplan LM, Rosenberg M, Tieu L, Guzman D, Kushel M. Mobile phone, computer, and internet use among older homeless adults: Results from the HOPE HOME cohort study. JMIR Mhealth Uhealth 2018 Dec 10;6(12):e10049 [FREE Full text] [doi: 10.2196/10049] [Medline: 30530464]
23. VonHoltz LAH, Frasso R, Golinkoff JM, Lozano AJ, Hanlon A, Dowshen N. Internet and social media access among youth experiencing homelessness: Mixed-methods study. J Med Internet Res 2018 May 22;20(5):e184 [FREE Full text] [doi: 10.2196/jmir.9306] [Medline: 29789281]
24. Woods SS, Forsberg CW, Schwartz EC, Nazi KM, Hibbard JH, Houston TK, et al. The association of patient factors, digital access, and online behavior on sustained patient portal use: A prospective cohort of enrolled users. J Med Internet Res 2017 Oct 17;19(10):e345 [FREE Full text] [doi: 10.2196/jmir.7895] [Medline: 29042345]
25. Smith B, Magnani JW. New technologies, new disparities: The intersection of electronic health and digital health literacy. Int J Cardiol 2019 Oct 01;292:280-282 [FREE Full text] [doi: 10.1016/j.ijcard.2019.05.066] [Medline: 31171391]
26. Heaslip V, Richer S, Simkhada B, Dogan H, Green S. Use of technology to promote health and wellbeing of people who are homeless: A systematic review. Int J Environ Res Public Health 2021 Jun 25;18(13):6845 [FREE Full text] [doi: 10.3390/ijerph18136845] [Medline: 34202242]
27. Grech E, Raeburn T. Experiences of hospitalised homeless adults and their health care providers in OECD nations: A literature review. Collegian 2019 Feb;26(1):204-211 [FREE Full text] [doi: 10.1016/j.colegn.2018.04.004]
28. Győrffy Z, Békási S, Döbrössy B, Bognár V, Radó N, Morva E, et al. Exploratory attitude survey of homeless persons regarding telecare services in shelters providing mid- and long-term accommodation: The importance of trust. PLoS One 2022 Jan 6;17(1):e0261145 [FREE Full text] [doi: 10.1371/journal.pone.0261145] [Medline: 34990458]
29. Békási S, Girasek E, Győrffy Z. Telemedicine in community shelters: possibilities to improve chronic care among people experiencing homelessness in Hungary. Int J Equity Health 2022 Dec 17;21(1):181 [FREE Full text] [doi: 10.1186/s12939-022-01803-4] [Medline: 36528777]
30. Radó N, Girasek E, Békási S, Győrffy Z. Digital technology access and health-related internet use among people experiencing homelessness in Hungary: Quantitative survey. J Med Internet Res 2022 Oct 19;24(10):e38729 [FREE Full text] [doi: 10.2196/38729] [Medline: 36260379]

31. Győrffy Z, Radó N, Pölczman L, Sükösd A, Boros J. Creating work-life balance among physicians in the age of digitalization: the role of self-consciousness and communication - a qualitative study. BMC Health Serv Res 2023 Oct 24;23(1):1141 [FREE Full text] [doi: 10.1186/s12913-023-10101-w] [Medline: 37875908]
32. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: Guided by information power. Qual Health Res 2016 Nov;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
33. Steward JL, Stringfellow E, Schwartz S, Johnson-Roe NK, Holt CH, Pollio D, et al. Qualitative interviews with persons who have had experience with homelessness: Lessons Learned. 2010 Presented at: American Public Health Association 138th Annual Meeting And Exposition; November 6-10, 2010; Denver, CO.
34. Thomas DR. Feedback from research participants: are member checks useful in qualitative research? Qualitative Research in Psychology 2016 Aug 02;14(1):23-41. [doi: 10.1080/14780887.2016.1219435]
35. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
36. Bruce CR, Harrison P, Nisar T, Giammattei C, Tan NM, Bliven C, et al. Assessing the impact of patient-facing mobile health technology on patient outcomes: Retrospective observational cohort study. JMIR Mhealth Uhealth 2020 Jun 26;8(6):e19333 [FREE Full text] [doi: 10.2196/19333] [Medline: 32589161]
37. Ibrahim MS, Mohamed Yusoff H, Abu Bakar YI, Thwe Aung MM, Abas MI, Ramli RA. Digital health for quality healthcare: A systematic mapping of review studies. Digit Health 2022;8:20552076221085810 [FREE Full text] [doi: 10.1177/20552076221085810] [Medline: 35340904]
38. Uchino BN, Cacioppo JT, Kiecolt-Glaser JK. The relationship between social support and physiological processes: a review with emphasis on underlying mechanisms and implications for health. Psychol Bull 1996 May;119(3):488-531. [doi: 10.1037/0033-2909.119.3.488] [Medline: 8668748]
39. Bensken WP, Krieger NI, Berg KA, Einstadter D, Dalton JE, Perzynski AT. Health status and chronic disease burden of the homeless population: An analysis of two decades of multi-institutional electronic medical records. J Health Care Poor Underserved 2021;32(3):1619-1634 [FREE Full text] [doi: 10.1353/hpu.2021.0153] [Medline: 34421052]
40. Vallesi S, Tuson M, Davies A, Wood L. Multimorbidity among people experiencing homelessness-Insights from primary care data. Int J Environ Res Public Health 2021 Jun 16;18(12):6498 [FREE Full text] [doi: 10.3390/ijerph18126498] [Medline: 34208580]
41. Baggett TP, Liauw SS, Hwang SW. Cardiovascular disease and homelessness. J Am Coll Cardiol 2018 Jun 05;71(22):2585-2597 [FREE Full text] [doi: 10.1016/j.jacc.2018.02.077] [Medline: 29852981]
42. Braun E, Gazdag G. Pszichiátriai zavarok előfordulása hajléktalanok között [Prevalence of psychiatric disorders in homeless population]. Psychiatria Hungarica 2015;30(1):60-67 [FREE Full text]
43. Fazel S, Khosla V, Doll H, Geddes J. The prevalence of mental disorders among the homeless in western countries: systematic review and meta-regression analysis. PLoS Med 2008 Dec 02;5(12):e225 [FREE Full text] [doi: 10.1371/journal.pmed.0050225] [Medline: 19053169]
44. Tünde B. Mentális zavarral, hajléktalanságban élő kliensek ellátása. Budapest Methodological Centre of Social Policy and Its Institutions. 2022. URL: https://tinyurl.com/3myhs8je [accessed 2024-04-30]
45. McCarty D, Argeriou M, Huebner RB, Lubran B. Alcoholism, drug abuse, and the homeless. Am Psychol 1991 Nov;46(11):1139-1148. [doi: 10.1037//0003-066x.46.11.1139] [Medline: 1772151]
46. Győrffy Z, Döbrössy B, Radó N, Boros J, Békási S. State of vulnerable populations in the techquity framework in Hungary. Front Public Health 2023;11:1215325 [FREE Full text] [doi: 10.3389/fpubh.2023.1215325] [Medline: 37483940]
47. Harris J. The digitization of advice and welfare benefits services: re-imagining the homeless user. Housing Studies 2019 May 09;35(1):143-162. [doi: 10.1080/02673037.2019.1594709]
48. Miler JA, Carver H, Foster R, Parkes T. Provision of peer support at the intersection of homelessness and problem substance use services: a systematic 'state of the art' review. BMC Public Health 2020 May 07;20(1):641 [FREE Full text] [doi: 10.1186/s12889-020-8407-4] [Medline: 32381086]
49. Expanding peer support roles in homeless services delivery: A toolkit for service providers. Homeless and Housing Resource Center. 2023. URL: https://hhrctraining.org/system/files/paragraphs/download-file/file/2023-08/HHRC_Peer_Support_Toolkit-508.pdf [accessed 2024-04-30]
50. Barker SL, Maguire N. Experts by experience: Peer support and its use with the homeless. Community Ment Health J 2017 Jul;53(5):598-612 [FREE Full text] [doi: 10.1007/s10597-017-0102-2] [Medline: 28176210]
51. Stefkovics A, Sik E. What drives happiness? The interviewer's happiness. J Happiness Stud 2022;23:2745-2762. [doi: 10.1007/s10902-022-00527-0]

Abbreviations

HCSOM: Hungarian Charity Service of the Order of Malta
WHO: World Health Organization

Edited by C Jacob; submitted 12.12.23; peer-reviewed by N Karnik, S Markham, G Sebastian; comments to author 19.02.24; revised version received 02.03.24; accepted 10.03.24; published 14.05.24.

Please cite as:
Radó N, Békási S, Győrffy Z
Health Technology Access and Peer Support Among Digitally Engaged People Experiencing Homelessness: Qualitative Study
JMIR Hum Factors 2024;11:e55415
URL: https://humanfactors.jmir.org/2024/1/e55415
doi:[10.2196/55415](https://doi.org/10.2196/55415)
PMID:[38743937](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=38743937&dopt=Abstract)

Original Paper

Digital Lifestyle Interventions for Young People With Mental Illness: A Qualitative Study Among Mental Health Care Professionals

Chelsea Sawyer[1], PhD; Rebekah Carney[2], PhD; Lamiece Hassan[1], PhD; Sandra Bucci[1], DClinPsy; John Sainsbury[2], MSc; Karina Lovell[3], PhD; John Torous[4], MD; Joseph Firth[1], PhD

[1]Division of Psychology and Mental Health, University of Manchester, Manchester, United Kingdom

[2]Greater Manchester Mental Health NHS Foundation Trust, Manchester, United Kingdom

[3]Division of Nursing, Midwifery & Social Work, University of Manchester, Manchester, United Kingdom

[4]Beth Israel Deaconness Medical Centre, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Joseph Firth, PhD
Division of Psychology and Mental Health
University of Manchester
3.005 Jean Mcfarlane Building
Manchester, M13 9PL
United Kingdom
Phone: 44 161 306 7811
Email: Joseph.firth@manchester.ac.uk

Abstract

Background: Given the physical health disparities associated with mental illness, targeted lifestyle interventions are required to reduce the risk of cardiometabolic disease. Integrating physical health early in mental health treatment among young people is essential for preventing physical comorbidities, reducing health disparities, managing medication side effects, and improving overall health outcomes. Digital technology is increasingly used to promote fitness, lifestyle, and physical health among the general population. However, using these interventions to promote physical health within mental health care requires a nuanced understanding of the factors that affect their adoption and implementation.

Objective: Using a qualitative design, we explored the attitudes of mental health care professionals (MHCPs) toward digital technologies for physical health with the goal of illuminating the opportunities, development, and implementation of the effective use of digital tools for promoting healthier lifestyles in mental health care.

Methods: Semistructured interviews were conducted with MHCPs (N=13) using reflexive thematic analysis to explore their experiences and perspectives on using digital health to promote physical health in youth mental health care settings.

Results: Three overarching themes from the qualitative analysis are reported: (1) motivation will affect implementation, (2) patients' readiness and capability, and (3) reallocation of staff roles and responsibilities. The subthemes within, and supporting quotes, are described.

Conclusions: The use of digital means presents many opportunities for improving the provision of physical health interventions in mental health care settings. However, given the limited experience of many MHCPs with these technologies, formal training and additional support may improve the likelihood of implementation. Factors such as patient symptomatology, safety, and access to technology, as well as the readiness, acceptability, and capability of both MHCPs and patients to engage with digital tools, must also be considered. In addition, the potential benefits of data integration must be carefully weighed against the associated risks.

(JMIR Hum Factors 2024;11:e53406) doi:10.2196/53406

KEYWORDS

digital health; behavior change; mental health care professionals; physical health; lifestyle intervention; qualitative; thematic analysis; service optimization; mobile phone

Introduction

Background

People with "severe mental illness" (SMI), such as schizophrenia, bipolar disorder, and associated psychotic or mood disorders, experience poorer physical health outcomes, which negatively affects their well-being across the life course and reduces life expectancy by up to approximately 15 years [1-3]. To reduce health disparities, it is crucial to adopt a preventative approach and intervene early. Adolescence or young adulthood presents a key opportunity as this is when most enduring mental health conditions are first diagnosed [4]. Young people with SMI and those at risk of SMI exhibit signs of poor cardiometabolic health and are more likely to engage in behaviors that are detrimental to their physical health; yet, much of this risk is modifiable [4,5].

Mental health care professionals (MHCPs) play a crucial role in supporting the mental and physical health needs of people with psychosis. However, MHCPs face significant barriers to delivering physical health interventions in practice [6]. This includes inadequate time and training in delivering evidence-based physical health interventions, difficulty reaching people in rural or remote areas, financial implications of delivering face-to-face interventions (particularly one-to-one), and limited National Health Service (NHS) resources for implementation [6].

Given these barriers and the increasing demand on the NHS, there is a growing focus on digital lifestyle interventions (DLIs), for example, using smartphones and websites to provide low-cost, scalable, and flexible interventions to promote healthier lifestyles [6,7]. The delivery of DLIs will require behavior changes among MHCPs. One model that can explain behavior change is the Capability, Opportunity, and Motivation–Behavior (COM-B) model [8]. According to this model, MHCP capability and opportunity to perform the behavior will influence their motivation to use DLIs and impact their delivery of DLIs in mental health care (MHC) settings. Capability refers to whether a person has the psychological (knowledge) or physical (skills) capability to perform the behavior. Alongside capability, an individual must have the opportunity to perform the behavior, and this refers to both physical (this includes the environment where the behavior will be performed and resources such as money and time) and social (the behavior of others) opportunity. Both reflective (reflective processes such as beliefs, goals, and values) and automatic (habitual and emotional responses) motivation also influence our behavior. The COM-B model can be used to inform future interventions [8].

Previous research suggests that MHCPs see the benefits of physical activity interventions [9]. However, MHCPs report barriers to implementation such as concerns about patient motivation and safety and logistical concerns on behalf of the patient, such as having equipment, clothes, and space. MHCPs have also reported personal barriers such as low confidence and capability to deliver interventions, lack of time and resources, and the belief that MHC should be a priority [10].

It is likely that MHCP attitudes toward and perceived barriers to using DLIs in MHC settings will vary from those for in-person interventions. According to actor-network theory, technology is not simply a tool or passive instrument that humans use to accomplish their goals [11]. Instead, technology can shape human behavior by creating new opportunities, alleviating constraints, and providing affordances that shape the way in which people think, communicate, and interact. Previous research has found that MHCPs believe that digital tools that support patient self-management would change their own roles and responsibilities [12]. Numerous studies have shown that, while MHCPs see the potential benefits, they are concerned about issues of liability, harm to patients, and lack of training regarding using DLIs [12-14]. As the MHCP role primarily focuses on treating mental health difficulties [15], it is important to explore and compare MHCP beliefs about using digital health for managing symptoms versus delivering lifestyle interventions.

Objectives

Therefore, this study aimed to explore MHCP perspectives, including barriers to and facilitators of using DLIs in MHC settings, with a particular focus on young people. These insights will provide key considerations for the implementation and use of DLIs in MHC settings.

Methods

Study Design

A mixed methods design was used, including a web-based survey and qualitative interviews, to examine the attitudes of MHCPs toward digital health in young people's MHC. An overview of the aligned findings of the combined survey and interview components of the project has been presented elsewhere [15]. While the previous mixed methods analysis provides a foundation for the research presented in this paper, this paper focuses on presenting the results of an in-depth qualitative examination of all the interview data, presenting the subjective experiences and perspectives of MHCPs regarding digital technology. Using a reflexive thematic analysis, we offered a more comprehensive understanding of subjective factors affecting the barriers and implementation of DLIs in clinical practice beyond the scope of the previous descriptive results to present useful recommendations for facilitating uptake of novel technology in health care settings.

The inclusion criteria were NHS MHCPs (1) working with young adults aged 16 to 35 years with mental illness, including specialist mental health services or in the context of broader primary care, and (2) working with young adults with mental illness for at least 6 months. The exclusion criteria were MHCPs working primarily with eating disorders due to differing treatment needs on nutrition and exercise [16-18]. The COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist was used to ensure a comprehensive and explicit report of the interview process (Multimedia Appendix 1).

Participants

All 13 participants were MHCPs working within NHS services with young adults aged 16 to 35 years, including specialist mental health services. Purposive sampling was used to recruit potential participants of a variety of occupational backgrounds and years of experience. Participants were recruited through emailed flyers. While interview participants were not reimbursed or rewarded for taking part in the interview, survey respondents were offered the choice to enter a prize draw to win a £50 (US $62.62) voucher. It was made clear to participants that completing the interview did not increase their chances of winning.

Data Collection

Overview

Semistructured interviews were conducted remotely using Microsoft Teams (Microsoft Corp) and audio recorded with participant consent. The interviews lasted 26 to 79 minutes and followed a topic guide (Multimedia Appendix 2) developed with input from our Patient and Public Involvement group and research team. In total, 2 researchers (CS and JF) conducted the interviews, which consisted of questions about participants' experience using digital health, the potential use of mobile health and DLIs in MHC, barriers to integration and use, and ways to boost engagement. Interview guides were flexible, using prompts and open questions to encourage participants to talk in depth about their experiences. All interviews were recorded and transcribed verbatim. Participants were assigned pseudonyms to maintain anonymity.

Data Analysis

Interviews were analyzed using a reflexive thematic approach [19-21]. Reflexive thematic analysis involves the researcher reflecting on how their experiences, personal assumptions, and background shape their analysis and interpretation of the data [20]. An inductively orientated experiential approach underpinned by critical realism was used [20]. This means that our themes were generated from the interviewees' direct experiences and observations while also recognizing that their understanding of reality is shaped by social and cultural factors. Critical realism was used as it allows the researcher to analyze participants' experiences while allowing the analysis to be informed by theory. The COM-B model of behavior change [21], which proposes that behavior is defined by our capability, opportunity, and motivation, was used as a theoretical underpinning and a prespecified area of interest. The model was used to identify potential barriers to the implementation of DLIs for young people with mental health conditions and any potential solutions to overcome these barriers.

Thematic analysis is a systematic approach whereby patterns and common themes are identified to describe a data set and understand a phenomenon [19,20]. The 6-phase guidelines by Clarke and Braun [20] were used to guide the analysis. These phases are recursive: (1) transcripts were read and reread so that the researcher (CS) could become familiar with the data, (2) systemic line-by-line coding was conducted to identify common features in the data, (3) codes were reviewed to determine themes, (4) themes were reviewed by 3 researchers (JF, CS, and LH) for homogeneity and heterogeneity to ensure that they were distinctive and coherent, (5) themes were defined and names were generated, and (6) findings were reported.

A primarily inductive approach was adopted with the interviews, but a deductive approach was taken when examining the barriers and facilitators informed by the COM-B model (based on previous research). Interview extracts related to barriers to and facilitators of implementing DLIs were mapped to the components of the COM-B model, whereas an inductive approach was taken for the remaining data. At the time of analysis, the researcher was not familiar with the current literature on digital health and in particular in the context of mental health, allowing them to analyze the data without preconceived themes or experiences.

To reduce the risk of bias, all researchers were involved in the analysis through regular meetings to discuss codes and themes. Themes and subthemes were generated and finalized using the NVivo software (version 12; QSR International) and MindView (version 7.0; Matchware). The team discussed the themes until consensus was reached within the team. During manuscript writing, subthemes with overlap were combined to avoid repetition. The final theme structure presented in the manuscript was reviewed and agreed upon by all coauthors.

Reflexivity

The two researchers who conducted the interviews (JF and CS) do not work in the NHS and made this clear to interviewees who were NHS employees. Both researchers have experience interviewing people on a variety of sensitive topics (self-harm, cancer [CS], and mental health [JF and CS]). All authors have an interest in digital health and promoting physical health in MHC settings. First author CS personally uses digital health apps. These views and experiences may have influenced our analysis; therefore, author CS kept a reflexive journal throughout the study. The author routinely reflected in the journal during data collection and analysis to reduce the possibility of their personal experiences and beliefs biasing their interpretations. A potential influencing factor in the interviews could be attributed to age. Some of the interviewees remarked that they did not have the same familiarity with apps as the interviewer (CS); this assumption could have influenced the views and opinions that participants expressed to this author.

Due to the COVID-19 pandemic, the interviews were conducted remotely using Microsoft Teams. Most participants joined the interviews from their own homes, providing a private setting that potentially fostered comfort and openness. However, working from home could have affected their work mindset and introduced distractions, such as pets, deliveries, or background noises. During an interview held in a private room at an interviewee's workplace, a team member interrupted the participant. While this interruption may have had an impact, it did not seem to alter their perceived barriers, and they continued discussing barriers, including those related to NHS staff, possibly influenced by their senior role. The remote format of Microsoft Teams interviews might have resulted in the interviewer missing out on subtle body language and facial responses, especially in the case of one participant who opted to keep their camera off. Nonetheless, Microsoft Teams provided flexibility, enabling participants to join at their preferred times.

Patient and Public Involvement

The overall study protocol had patient and carer involvement to ensure that all materials were appropriate and the content discussed about patients was appropriate and meaningful. The topic guide was developed using lived experience input. RC is a research fellow at a research unit embedded within clinical services at the Greater Manchester Mental Health NHS Foundation Trust (JUICE Youth Mental Health Research Unit). JUICE consists of academics as well as current practicing clinicians such as ward managers, lead psychiatrists, therapists, physiotherapists, dietitians, occupational therapists, experts by experience, and carers. A weekly consultation is held on the Child and Adolescent Mental Health Services inpatient units, where the topic guide was discussed.

Ethical Considerations

Ethics approval was granted by the University of Manchester Research Ethics Committee (2020-10603-17104) and the Health Research Authority (288734). Participants were briefed on the purpose of the study, and written informed consent was obtained.

Results

Participant Characteristics

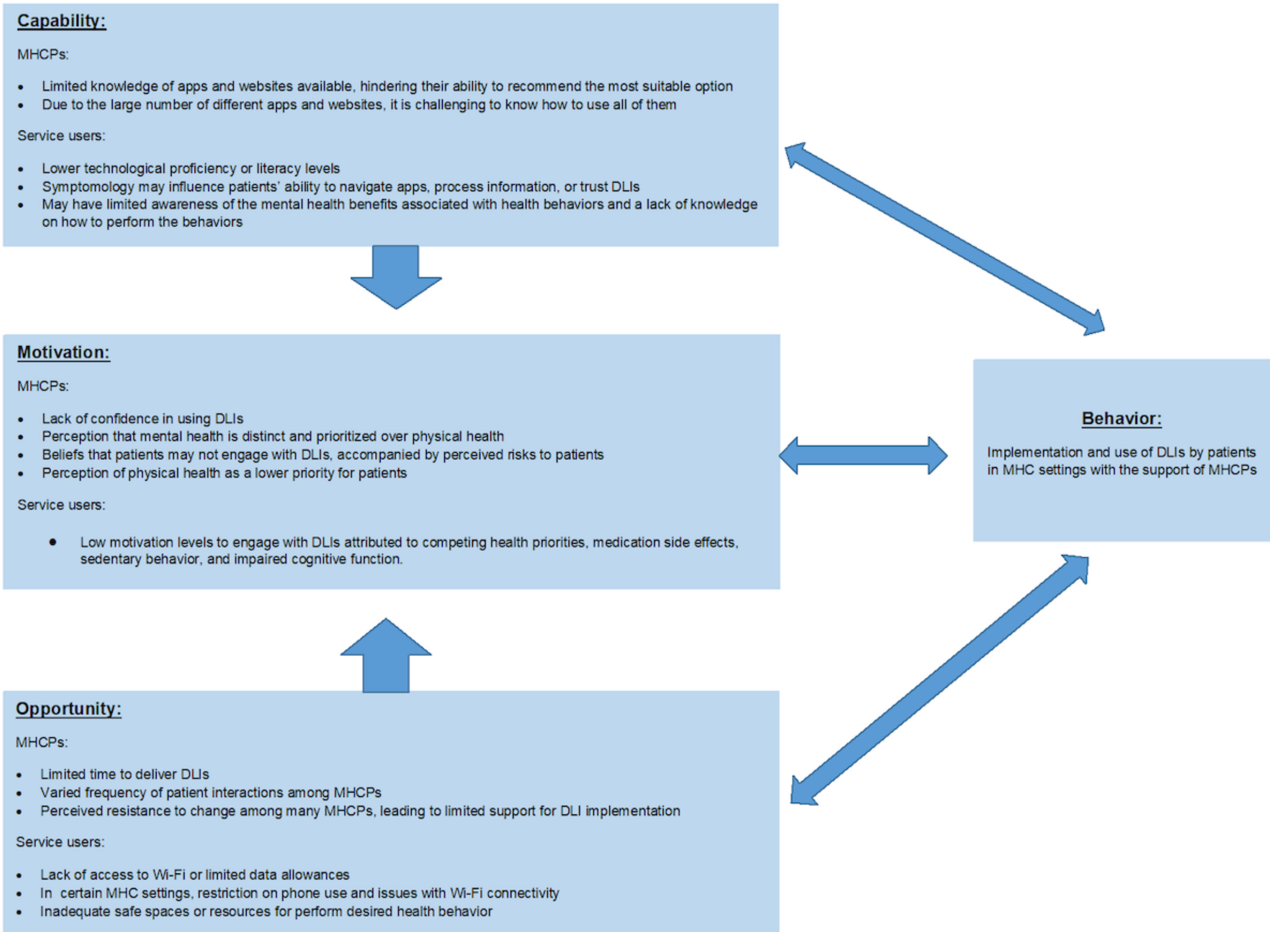
A total of 13 MHCPs were recruited from various MHC settings and roles, including MHC and research nurses (n=4, 31%), trainee psychiatrists (n=2, 15%), support workers (n=2, 15%), occupational therapists (n=1, 8%), physical health support workers (n=1, 8%), service managers (n=1, 8%), operational leads (n=1, 8%), and trainee advanced practitioners (n=1, 8%). Of the participants, 85% (11/13) were female, and 100% (13/13) were White British, with experience varying from at least 6 months to 20 years working in MHC service.

Analysis

Overview

Three main themes were identified: (1) motivation will affect implementation, (2) patients' readiness and capability, and (3) reallocation of staff roles and responsibilities. The themes and subthemes are described in Multimedia Appendix 3. Figure 1 presents how our findings also fit with the domains of the COM-B model [8].

Figure 1. Barriers to implementation of DLIs in MHC using the COM-B model_v2. COM-B: Capability, Opportunity, and Motivation–Behavior; DLI: digital lifestyle intervention; MHC: mental health care; MHCP: mental health care professional.



Theme 1: Motivation Will Affect Implementation

Overview

MHCP and patient motivation was perceived as the largest potential barrier to implementing DLIs in MHC settings. All interviewees felt that it would be important to emphasize the value and benefit of DLIs to patients, whereas acknowledging their potential risk is crucial for implementing DLIs in MHC settings:

> *I think some of them [MHCP], don't necessarily promote the apps, erm, because of the motivation...it can be quite problematic getting them [patients] to engage.* [INT5]

Subtheme 1: Individuals Are Motivated but Others Are Resistant

Most participants had a positive personal view of using digital health in MHC, in particular for young adults. However, some interviewees said that, although they were personally motivated to introduce and implement DLIs in the context of MHC delivery, their colleagues were reluctant to change, which posed a barrier to rolling out DLIs. MHCPs who are unaware of the benefits of DLIs or, for example, who are accustomed to established ways of working, were perceived by interviewees to be more resistant to adopting new practices:

> *...lots of people don't like change do they, and I think whenever you try and do anything new they'll be somebody that, will have something to say about it. I can't think of any specific kind of negative view of it, other than that, that just, traditional view that, oh it won't work, they won't use it.* [INT10]

Furthermore, several participants expressed concerns about time constraints and competing priorities faced by MHCPs. With an already demanding workload, interviewees said that MHCPs tend to prioritize mental health–focused care over physical health approaches or interventions. Interestingly, most people felt that their role was not best suited to the introduction of DLIs due to limited interactions with patients compared to their care coordinators despite being supportive of DLIs and their benefit more generally:

> *...it's [digital interventions] probably one of those things it's quite a good idea for the whole team to have an awareness of, but maybe, erm, you know, particularly care coordinators who are having the most contact with service users.* [INT9]

Subtheme 2: Patients Have Other Priorities

Interviewees perceived low motivation among young adults as a common barrier to engaging with DLIs. Low motivation was often attributed to medication side effects, sedentary behavior, and impaired cognitive function. Many interviewees perceived physical health to be a low priority for patients and that patients' main priority during their involvement with an MHC team was their mental health:

> *I think it depends on where they're at within their illness and how engaged they are with treatment and especially during those initial phases, it can be quite problematic, erm, and getting them to...engage.* [INT5]

Some interviewees felt that it was important to highlight the link between patients' physical and mental health and how changes to behavior could lead to changes in medication or treatment as well as improving physical and mental health:

> *Well talk about how, erm, maybe weight changes affect their mental health, how their medication has, the amount of medication that they've had has changed, how their, erm, you know, their diabetes diagnosis was reversed because they engaged in exercise and speak to them about what significance it would be to them, it's not about a size 8 jeans, it's about, I don't need to take as much Clozapine, and plus if I don't take as much Clozapine at night, then in a morning I'm not as knackered.* [INT5]

Interviewees also felt that the service setting might influence the uptake of DLIs. For example, young adults in inpatient settings may be hesitant to change due to the various restrictions in place, such as limiting takeaways or access to unhealthier food. DLIs that provided young adults with something such as promoting physical activity in inpatient settings were viewed as acceptable by interviewees. However, DLIs viewed as restrictive, such as those concerning diet or smoking, were assumed as not being well received by patients:

> *...inpatient mental health unit, erm, you've had so many bits of your identity taken away from you, in terms of like being able to access outside, and, I think, when you do try and have those conversations about, you know, well try and eat a bit healthier, it's gotten very angry really quickly because there's been so much of their liberties taken away, the fact that we try and take away the little things that they do enjoy like, staying in bed, erm, eating junk food, smoking, it, it, yeah, it can be quite a difficult subject.* [INT11]

Therefore, these preassumptions regarding what types of DLIs young adults are resistant to may reduce MHCP enthusiasm for undertaking the actions required to implement specific DLIs in inpatient settings. One MHC service that several MHCPs felt could work well to integrate DLIs in was early intervention services:

> *I think the young, younger people are more likely to use apps and You know when people first present to services, such as early intervention team. Uh, I think that would would work very, very well.* [INT1]

Subtheme 3: DLIs Need to Be Intuitive and Engaging

To overcome the perceived low motivation of patients to engage with DLIs, MHCPs felt that DLIs need to capture patients' attention and be engaging, which involved being visually appealing, interactive, and user-friendly, particularity due to the patients' age. Gamification, linking changes in behavior to patient-valued outcomes, and intuitive interfaces were perceived as being important in promoting engagement. Simple designs and usability were considered key, ensuring that both MHCPs and patients can navigate through the intervention:

> *...if you're focussing on young people, I think, I don't really see many barriers, if it's free, and, you know it's easy to use, I think it's just, it'd just be about that initial engagement, that initial kind of, them trying and it being good enough to keep them, er, interested.* [INT10]

> *...anything that's simple, straightforward and just easy to use would be probably the best starting point for now, for us [Mental Health are professionals].* [INT12]

> *I've never used gamified apps to be honest up, but yeah, it sounds great, It will make the younger people engage.* [INT2]

Several MHCPs shared the perspective that involving young adults with a wide spectrum of mental health conditions and literacy skills in the app design process is crucial to ensure both intuitiveness and engagement for patients:

> *...the important thing would be that if you were gonna design an app to, to genuinely have young people with a variety of mental health conditions, neurodevelopmental disorders, etc, all having input on what it looks like, how it works and how you engage with it.* [INT13]

Theme 2: Need to Consider Patients' Readiness and Capability to Use Digital Health

Subtheme 1: Patient Safety

A prominent theme was the need for patient benefit from digital health to outweigh potential risks or harm. Interviewees expressed the need to consider each patient and when was best to introduce DLIs. Initially, interviewees focused on what MHC settings would be the most suitable to introduce DLIs, with mixed views on the appropriateness of implementing DLIs in inpatient settings. On further reflection, the severity of the patients' symptoms rather than the MHC setting was considered most important when deciding on an appropriate time to introduce patients to DLIs:

> *...people with schizophrenia aren't unwell constantly, so you would use it and if they started to become paranoid or unwell, that's probably when you'd be able to just say, let's just remove it.* [INT12]

Concerns for patient safety were raised if DLIs required the young person to self-monitor physical activity or health data. For example, there were concerns that patients might misunderstand and assume that their MHC team was also monitoring their data, potentially leading them to not inform their MHC team of important changes. Alternatively, if health data were integrated and monitored by their MHC team, there were concerns that important data such as irregular heartbeat or reduced smoking while on clozapine could be missed, which could lead to harm:

> *I'm just thinking about difficulty I'd, I'd, feel terrible. If I had somebody on my caseload as a care coordinator and this information was there and I didn't pick up on it and I didn't notice that and then something happened, it's. Then there's a kind of risk factor there of Case negligence, maybe potentially, and who's going to be overseeing that.* [INT1]

Interviewees were also concerned that DLIs could worsen young adults' mental health symptoms, including paranoia, particularly for those who used phones or wearable devices. MHCPs articulated concerns that the tracking of patients' location and behavior via these wearable devices and phones could intensify patients' anxieties regarding perceived surveillance:

> *Consideration for people who could be psychotic, paranoid, suspicious, you know, and whether that might increase some of their Symptomatology, illusions or paranoia. You know if they were wearing a watch. For instance, knowing that I had access to that information.* [INT1]

MHCPs expressed concerns about potential risks that DLIs could pose regarding social interactions for young adults with psychosis. They were worried about inaccurate or harmful advice, exploitation, or negative interactions. Therefore, ensuring monitoring and moderation of social interactions within DLIs was necessary. However, interviewees also recognized the benefits of social support in boosting engagement:

> *...you would need some level of moderation in that community, 'cos you don't want unhelpful comments and views and, and unfortunately with online systems you get a lot of that. erm, because people are anonymous...there could be like a positive element of that and it could increase engagement.* [INT13]

Subtheme 2: Patients' Capability and Opportunity to Use Digital Technology

Digital health was perceived to be acceptable in MHC, especially among younger patients who were likely to have higher smartphone use. However, symptoms (eg, paranoia), reduced cognitive abilities (partly due to medication), and limited technological skills were seen as barriers to patients using digital technology:

> *A lot of the younger people are, erm [have technical skills], but, but also even, even the younger people when they're unwell, they have information processing problems.* [INT6]

Some MHCPs commented that paranoia may also hinder engagement. To address this, collecting minimal personal data through apps or websites was recommended:

> *I think just, the simpler the better maybe, you know, not having to gather as, as much information, just going off, you know, erm, individuals who, who can paranoid or, or the barriers, I think it, just something that's simple, that's easy, you know, you're not having to put a lot of data in.* [INT7]

Access was a potential barrier as some patients may lack access to phones, data, or apps, especially during periods of psychosis or as inpatients due to restrictions or poor Wi-Fi connections. Interviewees suggested that data-free apps that synchronize to Wi-Fi or apps with lower data requirements might overcome this barrier. To reduce the digital divide, interviewees said that services (ie, the NHS) should provide smartphones, pay for subscription costs or data network charges, or provide wearables to ensure that recommended DLIs can be accessed fairly:

> *If something's gonna be effective and there's evidence base into it, I don't think that people should [pay], if it's about health and it's actually gonna reduce our costs in the long term, then it should be free.* [INT13]

Theme 3: Integrating Digital Health Will Require the Reallocation of Staff Roles and Responsibilities

Subtheme 1: Technology Changes Our Roles and Responsibilities

Despite attitudes being largely positive, there were differences in interviewees' perspectives on the impact that implementing DLIs into routine care would have on MHCP workload. Some interviewees who felt that their role should mainly focus on the mental health needs of a patient believed that DLIs would increase their workload. They suggested that the care coordinator (case manager) role would be best placed to implement DLIs as they have more patient contact and involvement and, therefore, have the time and a preexisting relationship:

> *...particularly care coordinators who are having the most contact with service users.* [INT9]

In contrast, interviewees in senior or physical health–focused professions viewed digital health as a way to reduce staff workload and enable frequent physical health monitoring (eg, remote blood pressure monitoring), resulting in better patient care; these interviewees believed that it was everyone's role and responsibility to implement DLIs in MHC:

> *...anything that's from that, the core components of the physical health check that they can input, would save everybody a lot of time and, erm, effort and money and it would also make it much more up to date.* [INT10]

Interviewees expressed concerns that the use of DLIs in MHC could affect their interactions with patients. Some interviewees believed that building a strong relationship with patients was crucial and DLIs entail a loss of face-to-face nuances, resulting in difficulties detecting physical symptoms, which in turn would be detrimental to patients' mental health. While some interviewees felt that DLIs may lead to people "becoming more isolated" (INT3), others felt that they would improve access to services for patients, particularly those with social anxieties and those who, due to their younger age, are more comfortable interacting digitally:

> *...with our cohort who might socially find things challenging and difficult actually a screen is quite familiar to them, so they will often prefer that.* [INT8]

Subtheme 2: MHCPs Will Need to Acquire Additional Skills

Interviewees said that staff involvement was crucial for successful implementation of DLIs in MHC settings. However, interviewees recognized that not all staff members have the necessary skills or knowledge to do this. Tailored training, including interactive sessions and MHCPs using the apps themselves, was recommended to enhance confidence, motivation, and psychological capability to use and recommend digital technology in clinical settings:

> *There's always room for training I find that interesting to go through it and if I was clinician, finding out what apps there are out there, how we can use them, how we can recommend.* [INT1]

One interviewee also felt that digital health education could be provided in health care degrees such as nursing and occupational therapy. Several MHCPs lacked awareness of available apps or websites. To address this, interviewees suggested that the NHS could provide a list of approved apps, improving trust and credibility while improving their knowledge and reducing guilt or personal blame in case of adverse events. Opinions differed on recommending apps without previous experience. Some expressed concerns about patient safety when recommending potentially ineffective or harmful apps, such as exercise apps that result in injury, whereas others felt comfortable if the apps were available to the general population:

> *...some people, if it's not NHS approved, might be a bit more nervous. I think that's a thing with the YouTube videos as well, like if it's just a person who's put together a, an exercise for somebody to follow, and it's not to do with the NHS, I think it does make people a bit more nervous about engaging people in that, but I personally have done, if, if I feel confident.* [INT11]

Subtheme 3: Who Is Responsible for Managing the Risk?

A concern regarding the implementation of DLIs was the issue of responsibility. All interviewees questioned who would be responsible if patients experienced negative outcomes as a result of using digital technology in the context of their health care. A particular concern was related to data monitoring, such as what data MHCPs should have access to and who should monitor them (patients or MHCPs). Interviewees expressed concerns about who was responsible for the oversight of digital data collection, especially if fluctuations in health or changes in behavior were missed. Interviewees wondered about their potential liability and described the guilt that they would feel if important changes in mental or physical health were missed. Some MHCPs strongly believed that data should never be integrated into NHS systems but patients could share and discuss their health with MHCPs if desired, empowering patients to take responsibility of their health:

> *I was thinking about the patient doing it for themselves rather than anybody having anybody sitting behind the scenes monitoring it.* [INT2]

On the other hand, some interviewees who routinely collected physical health data felt that, if DLIs collected data without MHCPs acknowledging or providing feedback in response, this could demotivate younger patients to use apps that track or record behavior:

> *I think for this kind of age group and even older, you need that, well done, you're doing well there, and not, relentless, I think having a barrier would be if there wasn't any kind of short-term goals that you could say, right we're making progress here, we're*

> *doing well, or this is what you need to work on.* [INT8]

> *...like the feedback, you know, 'cos a lot of the young people that, that we work with on the wards, they really, are seeking time with people and if you can provide that time that's quite focused on something and provide lots of positive reinforcement if something's gone really well, I think it might motivate people to, to get sustained use from an app like that, you probably need somebody who's on your side and really supporting you to use it well.* [INT11]

In addition, interviewees believed that young adults with mental health conditions may need support and advice from MHCPs in cases in which they implemented minor changes but did not observe any discernible outcomes, for example, if they made changes to their diet but did not lose weight. To overcome the burden of data monitoring and potential liability, one interviewee suggested implementing automated systems that notify clinicians of changes in heart rate, smoking behavior, and so on:

> *I prefer the idea of, I think it's a fantastic idea for them [young adult service users] to go away and when they come back and you say, your weight, you've been putting weight on, let's have a look on your app what you actually have been eating, so you can have those sort of discussions with them.* [INT6]

Discussion

Principal Findings

The aim of this study was to explore MHCP perspectives, including barriers to and facilitators of using DLIs to support young people with mental illness. To our knowledge, this is the first study to explore MHCP views on using and integrating digital health in MHC for young people. Overall, MHCPs felt that digital health care is acceptable when delivered alongside face-to-face care and has the potential to enhance the current care that patients receive. However, they also identified barriers to implementation, including staff and patient motivation and capability to deliver or use DLIs, concerns regarding patient safety, the digital divide, and the privacy of data.

Relevance to Previous Research and Existing Theory

Overview

Similar to previous research on lifestyle interventions [9,10,22], MHCPs expressed concerns with patient motivation and safety as well as time constraints and staff motivations and capability to deliver interventions. Notably, there were differences in concerns about resource availability as, while there were concerns about insufficient phone data interfering with DLIs, a lack of other resources (such as home exercise equipment, healthy eating ingredients, or staffing and clothing [9,22]) was not commonly raised as an issue in this study. The remote nature of DLIs raised concerns about missing important data, and key considerations for implementation included ensuring the credibility and trustworthiness of the apps or websites used in DLIs, prioritizing patient safety, and effective data monitoring.

In some cases, MHCPs may differentiate between the promotion of physical health and their core responsibilities. A decreased enthusiasm among MHCPs for addressing physical well-being may be tied to factors such as apprehension toward assuming personal responsibility and diminished patient motivation [10].

Our findings were in line with actor-network theory [11]—MHCPs did not perceive the implementation of DLIs as an additional resource to use. Instead, they felt that the implementation of DLIs would change their roles, bring new risks to patients, and affect rapport. However, a recent study exploring the views of patients with SMI on digital health found that they also believed that such technologies could change the relationship with MHCPs but in a positive way by empowering them to manage their health and providing a source of help other than their health care providers [18,23]. They also valued the ability to self-monitor and share their progress or behavior with their MHCPs to obtain additional support or positive feedback [23]. These results are also in line with those of broader digital health implementation studies suggesting a need for support for patients and clinicians as well as systems-related issues such as regulation, workflow, and safety [24]. Our findings also fit with the domains of the COM-B model [8] (Figure 1). Barriers identified from the interviews related to capability, opportunity, and motivation, along with potential solutions to overcome these barriers generated by the researchers, are presented in Table 1. The potential solutions were then mapped to potential intervention functions (broad categories to change behavior), and these potential barriers and solutions are discussed in more detail in the following sections.

Table 1. Potential problems and solutions for the implementation of digital lifestyle interventions (DLIs) in mental health care settings.

Domain and problem	Solution	Intervention function
Capability		
Patients may have poorer digital literacy and skills to use technology or phones (perceived as less of an issue for younger populations)	Instructions on how to effectively use phones and apps	Education
Symptoms that may limit patients' cognitive ability to use digital technology (eg, side effects of medication)	Introduce DLIs only when patients have the mental capacity to consent	Restriction
MHCPs[a] do not know which apps are available and effective and how to use them	Training on how to find and use apps and knowledge sharing within clinical teams	Training or education
Opportunity		
Patients not having access to phones, internet, or data	Provision of phones, payment for data and wearables, and use of apps that are usable offline and resynchronize or update once connected to Wi-Fi or data	Enablement or environmental restructuring
Cost of wearables, data, and app subscriptions	Provision of wearables, NHS[b] covering the cost of app subscriptions, and having an iPad or device that can be used by an entire inpatient ward or service	Enablement or environmental restructuring
Restrictions (eg, no phones and no space) and restricted use or functionality when in inpatient units	Having a shared device on the ward or having supervised access to use the app and being shown how to use it before discharge	Service provision
MHCPs do not have the time to deliver interventions	Peer coaches or digital navigators to help patients install apps and deal with issues	Environmental restructuring
Integrating health data to ensure patient safety	Automated notifications if there are high risks or behavior changes that need to be addressed (ie, changes in smoking on clozapine)	Enablement
Secure storage of data	Improvement in the current technology infrastructure of the NHS to allow app data to be securely stored on the system	Enablement
Motivation		
Patients having low motivation	To boost motivation, use rewards, games, self-monitoring of behaviors or health outcomes, and provision of feedback	Incentives
MHCPs lack confidence with digital technology	Training and education and having regular drop-in clinics to problem solve any issues	Training
MHCPs' competing priorities (focus on treating mental health)	When implementing DLIs, services will need to consider which MHCP roles would be best to implement this and could use individual or service-based targets regarding delivering DLIs	Incentives or coercion
MHCP beliefs that patients do not want to change behavior	Involve patients in the integration of DLIs; this will demonstrate that patients are willing to use DLIs	Persuasion
MHCPs do not see the benefit of digital interventions	Educating MHCPs on the benefit of treating physical and mental health together and how digital health can play a role in this	Education

[a]MHCP: mental health care professional.

[b]NHS: National Health Service.

Capability

MHCPs voiced uncertainties about their own confidence and ability to effectively deliver DLIs. According to our previous work, MHCPs have limited opportunities to use DLIs in their current role [15], and this is consistent with current literature [13,25]. This may have led MHCPs to perceive a lack of psychological capability (knowledge) to recommend and use apps. Therefore, appropriate training is required, such as interactive training sessions that allow MHCPs to trial various apps and websites, thereby boosting their self-efficacy (the belief in their ability to perform the behavior) and motivation [26]. Alternatively, some MHCPs suggested that participating in DLIs themselves could enhance their motivation. This notion is supported by recent research demonstrating that MHCPs who actively participated in exercise were more inclined to encourage its adoption among inpatients [10].

A lack of knowledge of apps available was perceived as a barrier. MHCPs felt that receiving a list of approved apps to recommend would reinforce the credibility and trustworthiness of the apps, remove personal liability, and improve their knowledge about what is available. Interestingly, very few MHCPs were aware of organizations such as the Organisation for the Review of Care and Health Apps or SilverCloud, which provide a list of NHS-approved apps. Awareness of approved and endorsed apps could reduce the burden on MHCPs in selecting appropriate apps for patients. For example, the MINDapps database allows users to search for mental health apps using specific criteria and provides a description, a rating, and reviews of each app.

MHCPs emphasized the need for considering the clinical presentation of young adults with a mental health condition before the initiation of DLIs and continuously monitoring it should any clinical deterioration occur. They stressed the importance of evaluating symptom severity and diagnosis, with particular care taken for individuals with eating disorders (when recommending diet or physical activity interventions). The focus was on prioritizing patient safety and avoiding potential harm while acknowledging that symptom severity influences patient engagement and motivation regarding physical health interventions. In addition, patients may have limited technology skills or literacy and may need additional support or training on how to use apps [27]. Promisingly, previous research [28,29] has shown that individuals who are less familiar with DLIs can use them after minimal training or with support from peer coaches. Incorporating peer coaches to train and support patients could help lighten the workload burden on MHCPs and contribute to reducing the digital divide [30].

Opportunity

Competing interests and limited time, limited patient contact, and inadequate technology infrastructure in the NHS were perceived as barriers to implementing DLIs in MHC settings for MHCPs. This was in line with a recent study that found that MHC settings lacked effective integration of digital technologies [31]. The study also found significant differences among MHCPs regarding whether it falls under their role to implement digital MHC [31]. Therefore, ensuring successful integration of DLIs in MHC requires a thoughtful evaluation of their alignment with existing services, restrictions in inpatient settings, the optimal MHCP role or roles to deliver DLIs, and ways to minimize unnecessary burden on MHCPs.

One solution alongside improvements in technology infrastructure is new MHCP roles dedicated to implementing DLIs and monitoring patient data in MHC settings via the digital navigator pathway [30]. In addition to alleviating burdens, digital navigators could provide feedback to patients on their data. Berry et al [32] discovered that patients expressed a keen interest in engaging with MHCPs to review digitally collected outcome data related to their mental health. This notion was mirrored in this study, with MHCPs emphasizing the need for feedback on health data to gain deeper insights into causal links between behavior and health or to underscore the effectiveness of subtle changes that might take time to be realized more broadly. Although concerns were expressed about collecting and monitoring health data, there are several benefits to using near–real-time data—they can improve the quality of care received, detect health deterioration earlier, and reduce staff demands through more efficient monitoring [33-35].

In contrast with other research exploring MHCPs' views on digital health for the self-management of severe mental health conditions [32], we found that access to phones was not perceived as a barrier, likely due to the young age of the patient population (18-35 years) and increase in mobile phone ownership in the last decade [36]. However, limited data and app subscriptions were perceived as significant barriers. Staff felt that the NHS should provide phones and wearables and cover the cost of app subscriptions and data allowance. If the NHS is to cover app subscriptions and data allowance, the cost-effectiveness of DLIs needs to be considered to reduce the digital divide [30,37,38].

Motivation

In line with previous research, staff were reluctant to promote physical health [10,39,40] because it was perceived as not part of their role. MHCPs also lacked confidence in delivering DLIs and perceived that young people with mental illness have other priorities than their physical health and that risk of harm to patients may outweigh potential benefit. This perceived distinction between physical and mental health needs to be addressed through wider changes in education or service provision to highlight the need to take a holistic approach and treat mental and physical health concurrently [10]. Staff were particularly reluctant to implement DLIs for young people in inpatient settings, where the environment may be more unpredictable and access to technologies may be restricted in some cases. Furthermore, MHCPs felt that young people may be difficult to engage in this environment due to feeling out of control and to restrictions on movement [41]. However, introducing DLIs may provide patients with the autonomy to look after their own physical health when they feel out of control [42] as well as the potential therapeutic effects from physical activity [43]. To address MHCPs' concerns regarding professional liability and patient safety, clear guidelines and frameworks will need to be in place. MHCPs also stressed the significance of engaging both MHCPs and patients in the development and implementation of DLIs in MHC, believing that this collaborative approach can enhance overall buy-in from MHCPs and patients alike.

Strengths and Limitations

A strength of this research is the potential real-world impact on patient care by providing a rich understanding of MHCP experiences and clinical recommendations to implement digital health in MHC settings. Our sample comprised primarily White British and female individuals; this means that the findings may not be transferable across cultures, gender, and ethnicities. As with all qualitative research, this study was shaped by the researchers' personal experiences and views stemming from their own use of digital health and views on DLIs. A reflexive approach among the research team reflecting on the researchers' personal experiences, assumptions, and perspectives was used throughout the research process.

Recommendations for Implementation

We make 5 recommendations based on our findings. First, clear guidelines for recommending apps, handling data, and monitoring safety are needed. These guidelines should be developed with stakeholders, patients, and MHCPs to ensure that their concerns and needs are addressed. Second, DLIs should be coproduced, to some extent, with their intended end users to ensure that they are appropriate, engaging, and user-friendly and with those who will be delivering the DLIs to ensure that MHCPs have the skills and these steps may lead to better engagement. While coproduction of DLIs could involve full cocreation of bespoke technologies, this may not always be required, and instead, coproducing the way in which an existing technology is implemented in MHC and provided to patients could be sufficient. Third, to boost MHCPs' psychological capability and skills in using digital technology, training should be provided. This training could be provided by colleagues who were involved in implementing DLIs to gain clinician buy-in or as part of their professional training, which could lead to increased adoption, better implementation, and improved outcomes for patients. Alternatively, the delivery of this training by patients could help dispel certain preconceived notions, such as the belief that service users may be unable or unwilling to effectively use apps. Fourth, MHCPs need to be made aware of evidence-based digital resources available to them. Finally, a patient-centered approach should be used when implementing DLIs that considers patients' current symptomatology, particularly regarding paranoia, and clear criteria should be set to reduce any potential risk.

Future Research

Areas for future consideration toward real-world implementation are (1) the provision of apps, data, or phones to reduce the digital divide; (2) the practical, ethical, and security issues regarding the collection and monitoring of health data; and (3) which MHCP role would be best to implement DLIs or whether a new role is required. First, bridging the digital divide is crucial, and we need to ensure equal opportunities for those with mental health conditions to engage with DLIs. Therefore, future work needs to determine the preventative cost and impact of providing DLIs in MHC settings, focusing on patients' safety, efficacy, the quality of care received, and the impact on patients' physical health and their use of other health services. Second, it is important to determine the parameters for data collection. This includes working to determine the optimal type and frequency of health or behavior data and how this can be applicable beyond research purposes and actually become clinically useful for improving patient outcomes. In addition, there is a need to address the ethical considerations surrounding data access in clinical settings, ensuring that patients are able to provide informed consent and outlining what happens when patients lose the capacity to provide informed consent. Finally, further efforts are required to establish whether MHCPs can adequately deliver DLIs in their current roles or whether the development of a new role is necessary to maximize the effectiveness of DLIs in MHC settings. In summary, future research is required to examine the sustainability and cost-effectiveness of the implementation of DLIs, the optimal times and means for introducing DLIs to patients, and how this can be tailored to suit the individual needs of people diagnosed with a mental health condition.

Conclusions

Implementing lifestyle interventions for individuals with SMI is imperative, and the incorporation of DLIs may overcome some of the barriers faced by in-person interventions. Alongside this, implementation during the early intervention period could present a pivotal opportunity for timely approaches to preventing physical comorbidities from arising. The findings from this study suggest that, while digital health has the potential to enhance MHC and the quality of care that patients receive, there are important concerns that MHCPs hold that need to be addressed when considering implementation. Efforts are required to work with patients, MHCPs, and other stakeholders to identify appropriate content and delivery of DLIs, along with the types and content of training required to facilitate their implementation in routine clinical practice.

Acknowledgments

The authors would like to thank all the mental health care professionals who participated and helped recruit participants for this study. This study was funded by the University of Manchester Presidential Fellowship (P123958) and a UK Research and Innovation Future Leaders Fellowship (MR/T021780/1). SB is funded by a research professorship (NIHR300794).

Data Availability

Data sharing is not applicable to this paper as no data sets were generated or analyzed during this study.

Authors' Contributions

JF, RC, JS, and SB developed the study design. JF supervised the study. CS and JF collated and analyzed the data. CS, LH, and JF discussed the themes. CS and JF wrote the manuscript. JF, SB, LH, RC, and JS provided substantial input throughout the development and writing of the manuscript.

Conflicts of Interest

JF has received honoraria and consultancy fees from Atheneum, Informa, Gillian Kenny Associates, Big Health, Nutritional Medicine Institute, ParachuteBH, Richmond Foundation, and Nirakara independent of this work. SB is the director and a shareholder of CareLoop Health Ltd, a spinout from the University of Manchester to develop and market digital solutions for remote monitoring using smartphones for mental health conditions (currently, schizophrenia and postnatal depression). JT is the editor of JMIR

Mental Health and is on the scientific board of Precision Mental Wellness; however, this is unrelated to this work. The remaining authors (CS, RC, and LH) declare no other conflicts of interest.

Multimedia Appendix 1
COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.
[DOCX File , 17 KB - humanfactors_v11i1e53406_app1.docx]

Multimedia Appendix 2
Interview schedule.
[DOCX File , 15 KB - humanfactors_v11i1e53406_app2.docx]

Multimedia Appendix 3
Theme summary table.
[DOCX File , 25 KB - humanfactors_v11i1e53406_app3.docx]

References

1. Firth J, Siddiqi N, Koyanagi A, Siskind D, Rosenbaum S, Galletly C, et al. The Lancet Psychiatry Commission: a blueprint for protecting physical health in people with mental illness. Lancet Psychiatry 2019 Aug;6(8):675-712. [doi: 10.1016/S2215-0366(19)30132-4] [Medline: 31324560]
2. Solmi M, Fornaro M, Ostinelli EG, Zangani C, Croatto G, Monaco F, et al. Safety of 80 antidepressants, antipsychotics, anti-attention-deficit/hyperactivity medications and mood stabilizers in children and adolescents with psychiatric disorders: a large scale systematic meta-review of 78 adverse effects. World Psychiatry 2020 Jun 11;19(2):214-232 [FREE Full text] [doi: 10.1002/wps.20765] [Medline: 32394557]
3. Perry BI, Stochl J, Upthegrove R, Zammit S, Wareham N, Langenberg C, et al. Longitudinal trends in childhood insulin levels and body mass index and associations with risks of psychosis and depression in young adults. JAMA Psychiatry 2021 Apr 01;78(4):416-425 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.4180] [Medline: 33439216]
4. Carney R, Cotter J, Bradshaw T, Firth J, Yung AR. Cardiometabolic risk factors in young people at ultra-high risk for psychosis: a systematic review and meta-analysis. Schizophr Res 2016 Feb;170(2-3):290-300 [FREE Full text] [doi: 10.1016/j.schres.2016.01.010] [Medline: 26794596]
5. Carney R, Firth J, Pedley R, Law H, Parker S, Lovell K. The clinical and behavioral cardiometabolic risk of children and young people on mental health inpatient units: a systematic review and meta-analysis. Gen Hosp Psychiatry 2021;70:80-97 [FREE Full text] [doi: 10.1016/j.genhosppsych.2021.03.007] [Medline: 33773375]
6. Weightman M. Digital psychotherapy as an effective and timely treatment option for depression and anxiety disorders: implications for rural and remote practice. J Int Med Res 2020 Jun;48(6):300060520928686 [FREE Full text] [doi: 10.1177/0300060520928686] [Medline: 32527170]
7. Stefanopoulou E, Lewis D, Taylor M, Broscombe J, Larkin J. Digitally delivered psychological interventions for anxiety disorders: a comprehensive review. Psychiatr Q 2019 Mar;90(1):197-215. [doi: 10.1007/s11126-018-9620-5] [Medline: 30488330]
8. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
9. Martland R, Gaughran F, Stubbs B, Onwumere J. Perspectives on implementing HIIT interventions for service users in inpatient mental health settings: a qualitative study investigating patient, carer and staff attitudes. J Affect Disord 2021 Mar 15;283:198-206. [doi: 10.1016/j.jad.2021.01.065] [Medline: 33561800]
10. Ball H, Yung A, Bucci S. Staff perspectives on the barriers and facilitators to exercise implementation in inpatient mental health services: a qualitative study. Ment Health Phys Act 2022 Mar;22:100452. [doi: 10.1016/j.mhpa.2022.100452]
11. Latour B. Reassembling the Social: An Introduction to Actor-Network-Theory. Oxford, UK: Oxford University Press; 2007.
12. Berry N, Bucci S, Lobban F. Use of the internet and mobile phones for self-management of severe mental health problems: qualitative study of staff views. JMIR Ment Health 2017 Nov 01;4(4):e52 [FREE Full text] [doi: 10.2196/mental.8311] [Medline: 29092809]
13. Sterling WA, Sobolev M, Van Meter A, Guinart D, Birnbaum ML, Rubio JM, et al. Digital technology in psychiatry: survey study of clinicians. JMIR Form Res 2022 Nov 10;6(11):e33676 [FREE Full text] [doi: 10.2196/33676] [Medline: 36355414]
14. Schmidt S, D'Alfonso S. Clinician perspectives on how digital phenotyping can inform client treatment. Acta Psychol (Amst) 2023 May;235:103886 [FREE Full text] [doi: 10.1016/j.actpsy.2023.103886] [Medline: 36921359]
15. Sawyer C, Hassan L, Sainsbury J, Carney R, Bucci S, Burgess H, et al. Using digital technology to promote physical health in mental healthcare: a sequential mixed-methods study of clinicians' views. Early Interv Psychiatry 2024 Feb;18(2):140-152. [doi: 10.1111/eip.13441] [Medline: 37318221]

16. Eikey EV, Reddy MC, Booth KM, Kvasny L, Blair JL, Li V, et al. Desire to be underweight: exploratory study on a weight loss app community and user perceptions of the impact on disordered eating behaviors. JMIR Mhealth Uhealth 2017 Oct 12;5(10):e150 [FREE Full text] [doi: 10.2196/mhealth.6683] [Medline: 29025694]
17. Eikey EV, Reddy MC. "It's definitely been a journey": a qualitative study on how women with eating disorders use weight loss apps. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI '17; May 6-11, 2017; Denver, CO p. 642-654 URL: https://dl.acm.org/doi/10.1145/3025453.3025591 [doi: 10.1145/3025453.3025591]
18. Torous J, Bucci S, Bell IH, Kessing LV, Faurholt-Jepsen M, Whelan P, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021 Oct 09;20(3):318-335 [FREE Full text] [doi: 10.1002/wps.20883] [Medline: 34505369]
19. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
20. Braun V, Clarke V. Thematic Analysis: A Practical Guide. Thousand Oaks, CA: SAGE Publications; 2021.
21. Braun V, Clarke V. Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. Couns Psychother Res 2020 Oct 18;21(1):37-47. [doi: 10.1002/capr.12360]
22. Happell B, Platania-Phung C, Scott D. Placing physical activity in mental health care: a leadership role for mental health nurses. Int J Ment Health Nurse 2011 Oct;20(5):310-318. [doi: 10.1111/j.1447-0349.2010.00732.x] [Medline: 21896121]
23. Williams J, Gillis C, Williams J, Gillett G, McGrath R, Ang K, et al. Exploring the use of digital interventions by people with severe mental illness to support their physical health: a mixed methods study. Issues Ment Health Nurs 2024 Jan;45(1):9-26. [doi: 10.1080/01612840.2023.2279207] [Medline: 38190422]
24. van Kessel R, Roman-Urrestarazu A, Anderson M, Kyriopoulos I, Field S, Monti G, et al. Mapping factors that affect the uptake of digital therapeutics within health systems: scoping review. J Med Internet Res 2023 Jul 25;25:e48000 [FREE Full text] [doi: 10.2196/48000] [Medline: 37490322]
25. Mahmoud K, Jaramillo C, Barteit S. Telemedicine in low- and middle-income countries during the COVID-19 pandemic: a scoping review. Front Public Health 2022 Jun 22;10:914423 [FREE Full text] [doi: 10.3389/fpubh.2022.914423] [Medline: 35812479]
26. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Mar;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191] [Medline: 847061]
27. Spanakis P, Wadman R, Walker L, Heron P, Mathers A, Baker J, et al. Measuring the digital divide among people with severe mental ill health using the essential digital skills framework. Perspect Public Health 2024 Jan 05;144(1):21-30 [FREE Full text] [doi: 10.1177/17579139221106399] [Medline: 35929589]
28. Olmos-Ochoa TT, Niv N, Hellemann G, Cohen AN, Oberman R, Goldberg R, et al. Barriers to participation in web-based and in-person weight management interventions for serious mental illness. Psychiatr Rehabil J 2019 Sep;42(3):220-228 [FREE Full text] [doi: 10.1037/prj0000363] [Medline: 31081651]
29. Young AS, Cohen AN, Goldberg R, Hellemann G, Kreyenbuhl J, Niv N, et al. Improving weight in people with serious mental illness: the effectiveness of computerized services with peer coaches. J Gen Intern Med 2017 Apr 7;32(Suppl 1):48-55 [FREE Full text] [doi: 10.1007/s11606-016-3963-0] [Medline: 28271427]
30. Wisniewski H, Gorrindo T, Rauseo-Ricupero N, Hilty D, Torous J. The role of digital navigators in promoting clinical care and technology integration into practice. Digit Biomark 2020 Nov 26;4(Suppl 1):119-135 [FREE Full text] [doi: 10.1159/000510144] [Medline: 33442585]
31. LaMonica HM, Iorfino F, Lee GY, Piper S, Occhipinti JA, Davenport TA, et al. Informing the future of integrated digital and clinical mental health care: synthesis of the outcomes from project synergy. JMIR Ment Health 2022 Mar 09;9(3):e33060 [FREE Full text] [doi: 10.2196/33060] [Medline: 34974414]
32. Berry N, Lobban F, Emsley R, Bucci S. Acceptability of interventions delivered online and through mobile phones for people who experience severe mental health problems: a systematic review. J Med Internet Res 2016 May 31;18(5):e121 [FREE Full text] [doi: 10.2196/jmir.5250] [Medline: 27245693]
33. Torous J, Bucci S, Bell IH, Kessing LV, Faurholt-Jepsen M, Whelan P, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021 Oct 09;20(3):318-335 [FREE Full text] [doi: 10.1002/wps.20883] [Medline: 34505369]
34. Marzano L, Bardill A, Fields B, Herd K, Veale D, Grey N, et al. The application of mHealth to mental health: opportunities and challenges. Lancet Psychiatry 2015 Oct;2(10):942-948. [doi: 10.1016/s2215-0366(15)00268-0]
35. Macrynikola N, Nguyen N, Lane E, Yen S, Torous J. The digital clinic: an innovative mental health care delivery model utilizing hybrid synchronous and asynchronous treatment. NEJM Catal Innov Care Deliv 2023 Aug 16;4(9). [doi: 10.1056/cat.23.0100]
36. Firth J, Cotter J, Torous J, Bucci S, Firth JA, Yung AR. Mobile phone ownership and endorsement of "mHealth" among people with psychosis: a meta-analysis of cross-sectional studies. Schizophr Bull 2016 Mar 22;42(2):448-455 [FREE Full text] [doi: 10.1093/schbul/sbv132] [Medline: 26400871]

37. Sylvia LG, Faulkner M, Rakhilin M, Amado S, Gold AK, Albury EA, et al. An online intervention for increasing physical activity in individuals with mood disorders at risk for cardiovascular disease: design considerations. J Affect Disord 2021 Aug 01;291:102-109. [doi: 10.1016/j.jad.2021.04.094] [Medline: 34029880]
38. Wisniewski H, Torous J. Digital navigators to implement smartphone and digital tools in care. Acta Psychiatr Scand 2020 Apr 26;141(4):350-355 [FREE Full text] [doi: 10.1111/acps.13149] [Medline: 31930477]
39. Kohn L, Christiaens W, Detraux J, De Lepeleire J, De Hert M, Gillain B, et al. Barriers to somatic health care for persons with severe mental illness in Belgium: a qualitative study of patients' and healthcare professionals' perspectives. Front Psychiatry 2021 Jan 26;12:798530 [FREE Full text] [doi: 10.3389/fpsyt.2021.798530] [Medline: 35153863]
40. Happell B, Scott D, Platania-Phung C. Perceptions of barriers to physical health care for people with serious mental illness: a review of the international literature. Issues Ment Health Nurs 2012 Nov 12;33(11):752-761. [doi: 10.3109/01612840.2012.708099] [Medline: 23146009]
41. Tully SM, Bucci S, Berry K. "My life isn't my life, it's the systems": a qualitative exploration of women's experiences of day-to-day restrictive practices as inpatients. J Psychiatr Ment Health Nurs 2023 Feb 18;30(1):110-122 [FREE Full text] [doi: 10.1111/jpm.12855] [Medline: 35771190]
42. Szinay D, Perski O, Jones A, Chadborn T, Brown J, Naughton F. Perceptions of factors influencing engagement with health and well-being apps in the United Kingdom: qualitative interview study. JMIR Mhealth Uhealth 2021 Dec 16;9(12):e29098. [doi: 10.2196/29098] [Medline: 34927597]
43. Bull ER, McCleary N, Li X, Dombrowski SU, Dusseldorp E, Johnston M. Interventions to promote healthy eating, physical activity and smoking in low-income groups: a systematic review with meta-analysis of behavior change techniques and delivery/context. Int J Behav Med 2018 Dec 12;25(6):605-616 [FREE Full text] [doi: 10.1007/s12529-018-9734-z] [Medline: 30003476]

Abbreviations

COM-B: Capability, Opportunity, and Motivation–Behavior
COREQ: Consolidated Criteria for Reporting Qualitative Research
DLI: digital lifestyle intervention
MHC: mental health care
MHCP: mental health care professional
NHS: National Health Service
SMI: severe mental illness

Edited by A Kushniruk, A Parush; submitted 17.11.23; peer-reviewed by B Lambe, N Josifovski, C Baxter; comments to author 11.01.24; revised version received 04.03.24; accepted 16.03.24; published 05.06.24.

Please cite as:
Sawyer C, Carney R, Hassan L, Bucci S, Sainsbury J, Lovell K, Torous J, Firth J
Digital Lifestyle Interventions for Young People With Mental Illness: A Qualitative Study Among Mental Health Care Professionals
JMIR Hum Factors 2024;11:e53406
URL: https://humanfactors.jmir.org/2024/1/e53406
doi:10.2196/53406
PMID:38837191

Viewpoint

Implementation of Anxiety UK's Ask Anxia Chatbot Service: Lessons Learned

Luke Collins[1], PhD; Niamh Nicholson[2], BSc; Nicky Lidbetter[2], PhD; Dave Smithson[2], HNC; Paul Baker[1], Prof Dr

[1]Linguistics and English Language, Lancaster University, Lancaster, United Kingdom

[2]Anxiety UK, Manchester, United Kingdom

Corresponding Author:
Luke Collins, PhD
Linguistics and English Language
Lancaster University
Economic and Social Research Council Centre for Corpus Approaches to Social Science
Bailrigg
Lancaster, LA1 4YW
United Kingdom
Phone: 44 1524 65201
Email: l.collins3@lancaster.ac.uk

Abstract

Chatbots are increasingly being applied in the context of health care, providing access to services when there are constraints on human resources. Simple, rule-based chatbots are suited to high-volume, repetitive tasks and can therefore be used effectively in providing users with important health information. In this Viewpoint paper, we report on the implementation of a chatbot service called Ask Anxia as part of a wider provision of information and support services offered by the UK national charity, Anxiety UK. We reflect on the changes made to the chatbot over the course of approximately 18 months as the Anxiety UK team monitored its performance and responded to recurrent themes in user queries by developing further information and services. We demonstrate how corpus linguistics can contribute to the evaluation of user queries and the optimization of responses. On the basis of these observations of how Anxiety UK has developed its own chatbot service, we offer recommendations for organizations looking to add automated conversational interfaces to their services.

(JMIR Hum Factors 2024;11:e53897) doi:10.2196/53897

KEYWORDS

chatbots; anxiety disorders; corpus linguistics; conversational agents; web-based care

Introduction

In the context of developing technologies, many businesses and services are turning to automated systems to provide users with information and accessible customer service. Among such tools, we find natural language processing systems, such as chatbots, that act as conversational interfaces, typically in lieu of interactions with human professionals. In health care, chatbots have a meaningful role to play, alongside other provisions, in increasing access to services, particularly in instances where there are restrictions in accessing face-to-face services [1,2]. Medical chatbots are already being used to provide and elicit information, create patient records, and discuss the results of clinical tests [3]. Furthermore, as Amiri and Karahanna [4] argue, health chatbots were shown to be particularly valuable in periods of quarantine as a response to the COVID-19 pandemic, in that "[t]heir scalability, wide accessibility, fast information dissemination, and substitution for in-person contact provide the functionality required to address the capacity expansion, social distancing requirements, and quick accurate information transmission needs of the public health response." Ultimately, chatbot tools and similar automated systems can make an important contribution to the provision of health information and support in the context of time and resource restraints.

There is a wide range of capabilities demonstrated in the deployment of conversational interfaces of varying complexity, from rule-based chatbots that produce prewritten responses based on recognizing programmed terms and phrases to embodied conversational agents manifesting as a computer-generated avatar and smart conversational interfaces such as Apple's Siri or Amazon's Alexa [1]. Nevertheless, simple conversational agents are increasingly used in executing tasks without the need for human involvement, including booking appointments, purchasing merchandise, ordering food, and sharing information [5].

Research has shown that users respond positively toward the perceived convenience of medical chatbots, showing appreciation for swift information retrieval as an alternative to delays in scheduling a consultation, queuing in a phone service, or waiting for an email response [3]. There is wide acceptance of automated systems providing general health advice [2]. Furthermore, there are indications that computer services can reduce perceived stigma, in that users are more willing to disclose details about their health concerns to an automated system on the basis that they are regarded to be more trustworthy and nonjudgmental, reducing the potential for embarrassment [6,7]. In the mental health domain, chatbots and other kinds of conversational agents have been shown to assist in the diagnosis and reduction of symptoms among individuals with major depressive disorder, promote adherence, provide cognitive behavioral therapy, and cultivate a stronger therapeutic alliance compared with users' interactions with a clinician [7].

This study details the development and learning from the implementation of a chatbot service through the website of the mental health charity Anxiety UK, the largest national charity in the United Kingdom to offer support for anxiety disorders. The charity has created a chatbot service called Ask Anxia to complement its other support and information services. In this study, we summarize the patterns of queries submitted to the Ask Anxia service after approximately 18 months of its activation using procedures from corpus linguistics, which involves using software tools to compute frequency-based measures of naturally occurring language data [8]. In addition, we review the quality of Ask Anxia's responses based on manual coding. We offer some reflections on the development of the Ask Anxia service as "lessons learned," with the intention that these will be instructive to others seeking to incorporate a conversational agent into their provision of information and support.

Anxiety UK and Ask Anxia

Anxiety UK was established in 1970 and provides a wide range of support services and information for those affected by anxiety, stress, and anxiety-based depression. Anxiety disorders are characterized by excessive worry and fear [9] and are included among the "Common mental disorders" that are recorded as becoming increasingly prevalent in the United Kingdom [10]. The charity supports individuals from all over the United Kingdom and, in some cases, the rest of the world and has recently led on the development of an informal global alliance of not-for-profit anxiety organizations. Anxiety UK has a strong service delivery arm offering support via their helpline, therapy, peer support groups, and anxiety management courses. Most of, if not all, its volunteers, staff, and trustees have some experience of anxiety disorders. Anxiety UK states through all its communications, including the chatbot service, that it does not provide crisis support and directs those in need of such support to urgent care services such as the National Health Service and the charity, Samaritans.

Anxiety UK introduced an automated chatbot service, Ask Anxia, with the principal aim of offering an out-of-office-hours service to users, helping them to navigate more quickly to information that was already available, for example, through Anxiety UK's web pages. Furthermore, the Anxiety UK team found that a high number of user queries received by phone or email concerned administration issues, and so, providing such information through an automated chatbot was seen as a way to release staff members and helpline volunteers to attend to other responsibilities that demanded more critical and engaged attention, including providing real-time interactional support via the helpline. Anxia is now a registered trademark that includes but is not limited to computer and application software provided by Anxiety UK as part of their mental health services.

Ask Anxia is a simple, pattern-matching chatbot that has been programmed to recognize certain stimuli (specific terms or phrases) and generate a response, which has been composed by the Anxiety UK team. At the time of writing, Ask Anxia had a content bank of 315 unique responses that has been developed and refined since the service has been operational, and the Anxiety UK team continues to monitor these response options based on the range of queries that users submit. An overview of the categories of responses is provided in Table 1, indicating the types of terms that Ask Anxia has been programmed to recognize in user queries.

Ask Anxia was launched in the beginning of July 2021, and we have applied procedures from corpus analysis (discussed in the Developments section) to 56 weeks' worth of anonymized, aggregated user queries submitted to the service (up until the end of July 2022). This amounted to 14,359 queries consisting of 139,286 words.

Table 1. Recurring themes in queries to Ask Anxia and examples of terms used to determine a response.

Theme	Examples of pattern-matched terms
Request for help	help; support; advice; guidance
Looking for information on a specific anxiety type	GAD; health anxiety; OCD; PTSD; emetophobia; phobia
Physical symptoms	headache; chest pain; breathing; appetite; nausea; feel sick; dizzy
Psychological symptoms	intrusive thoughts; negative thinking; overthinking; constant worry
Information on a service	therapy; group; course; class; counselling; CBT; EMDR; resources
How to access a service	membership; cost; referral; book; sign up; join
Wanting to connect	talk; human; chat; agent
How to support others	family; partner; son; daughter; child; colleague
Getting involved with Anxiety UK	volunteering; approved therapist; fundraising; donate; placements
Diagnosis	do I have anxiety; diagnosis; symptoms
Medication	antidepressants; tablets; medication
Coping techniques	can't cope; what can I do; panic; now; thoughts; relax; sleep
Location	in person; areas; UK; Europe; face-to-face; online
Crisis	suicidal; self-harm; die

We recognize that there are important ethical considerations pertaining to the data, given that queries submitted to the Ask Anxia service are highly personal and relate to individuals' well-being. In the privacy notice that is posted on the Anxiety UK website, users are informed that interactions with Ask Anxia are reviewed as part of the procedures for improving the quality of the service and that these may be shared with third parties for the purposes of research. Participants are discouraged from including personal information in their queries, and any such information that appears in the original message has been redacted. To protect the personal experiences of those who have accessed the service, we have provided generic examples, where cited, to demonstrate the interactional dynamics between constructed user queries and Ask Anxia's (authentic) responses. Reported figures for word frequencies are based on original user queries. As part of a more recent update to the service (July 2022), a message encouraging users not to disclose personal information (such as name, address, and place of work) was added to the Ask Anxia header to ensure that this is visible. Furthermore, such information does not inform Ask Anxia's pattern-matching programming, and so, it will only hinder the identification of an appropriate response.

In addition to reporting commonly used terms and phrases, we refer to the quality coding carried out by the Anxiety UK team, which is explained in the next section. Our study, then, offers a critical evaluation of the contribution of the chatbot Ask Anxia to Anxiety UK's wider provision of services and helps us to understand the general patterns of what visitors to the site collectively seek, in terms of information and support. In the next section, we summarize the insights that we have gained through developing Ask Anxia's programming, as the service has evolved over time.

Developments

Overview

In this section, we summarize the developments that have been applied to the Ask Anxia service based on observations of user queries, including where potential misunderstandings in queries asked by users arose. We present these developments as the lessons we have learned through reviewing the various updates that have been applied to the service since its launch, which are likely to be informative to those looking to implement similar tools. The continued monitoring of the service has contributed to its optimization and generated insights into user expectations. The time stamps for user queries indicate that 57.2% (8213/14,359) of the queries were submitted outside of Anxiety UK's office hours (9:30 AM-5:30 PM), demonstrating that the Ask Anxia service is used when other contact services, such as the helpline, are closed. Indeed, one of the earliest modifications to Ask Anxia, in August 2021, was to remove the cap on how many queries it responded to, given its popularity.

Updates Based on Frequent Terms in User Queries

The pattern-matched terms presented in Table 1 were largely informed by the Anxiety UK team's own long-standing experiences of working with people seeking support for their, or a loved one's, experiences of anxiety. Of course, the queries submitted to Ask Anxia provide further indications of what users seek from the service. As such, alongside the Anxiety UK team's expert judgment, procedures from corpus linguistics can be drawn on to help identify topics and terms that are commonly cited by users, which can potentially highlight important areas for extending the existing information provision.

Corpus linguistics refers to a set or procedures for making quantitative and qualitative observations of the patterns of natural language use and can straightforwardly tell us, by way of a wordlist, for example, what the most common terms in our data are and how often they occur. We used the corpus analysis tool *#LancsBox* [11] to examine the user queries. Researchers have found, however, that because of how the English language is structured, often the most common words largely remain the same across data sets (typically, *I*, *the*, *you*, *and*, *it*, etc). Indeed, the 5 most frequent terms occurring in the user queries of the Ask Anxia service, were *I*, *to*, *a*, *and*, and *hi*. As such, corpus linguists have developed the concept of keyness, enabling us to determine which words appear in our data more frequently, to a statistically significant degree, when compared with a corpus of larger or equal size [8]. A keyness analysis of the queries submitted to Ask Anxia through comparison with a 10 million–word corpus of general English spoken language, the British National Corpus 2014 [12], identified the keywords that are particularly characteristic of the language used by contributors in this context. The statistical measure used in this case was log likelihood, which established a confidence score indicating that the observed differences are not the result of chance. A threshold value of 15.13 was applied, which equates with a *P* value of <.001. The top 20 keywords are shown in Table 2 and ranked according to log likelihood value (not reported).

Table 2. Keywords in user queries to the Ask Anxia service (n=139,286).

Rank	Keyword	Frequency, n (%)
1	anxiety	2461 (1.77)
2	hi	2670 (1.92)
3	help	1588 (1.14)
4	hello	1332 (0.96)
5	yeah	13 (0.01)
6	am	1123 (0.81)
7	anxious	525 (0.4)
8	therapy	451 (0.3)
9	my	2358 (1.69)
10	panic	373 (0.3)
11	support	411 (0.3)
12	im	302 (0.2)
13	membership	310 (0.2)
14	struggling	334 (0.2)
15	i	7805 (5.6)
16	ok	259 (0.2)
17	how	1370 (0.98)
18	oh	47 (0.03)
19	feeling	394 (0.3)
20	can	1606 (1.15)

What is clear from the keywords is the topical focus on *anxiety* and the prevalence of appeals for *help* and *support* on the basis that users are *struggling*, *feelinganxious*, or experiencing *panic* (attacks), for instance. We can also see that queries are typically written in the first person (*I*, *my,* and *im*), take a question form (*how* and *can*), and have a relatively informal style (*hi*, *yeah*, *ok*, and *im*), that is, consistent with the instant messaging–like format through which users interact with Ask Anxia.

The prevalence of the terms *help* and queries about *therapy* and *membership* indicate that the Anxiety UK team had largely anticipated the themes most often captured in user queries, as indicated in Table 1. Nevertheless, the recurrence of particular terms, including at specific moments, has informed the continued refinement of Ask Anxia's responses and the information that is made available through the website. For example, in the week beginning September 13, 2021, keyness analysis showed that there was an increase in references to *fear* and *needle*, which coincided with booster doses of the COVID-19 vaccine being made available (to certain groups) and vaccines being approved for 12- to 15-year-olds, in anticipation of a new school term. Subsequently, the terms *fear* and *needle* appeared much more frequently in user queries. In response, Anxiety UK produced specific information concerning COVID-19 and related vaccines.

The Anxiety UK team has continued to extend Ask Anxia's response options since it was launched in July 2021. In addition, because of identifying themes arising from user enquires, the team has carried out the following activities:

- created factsheets specifically on perinatal anxiety, peri- and postmenopausal anxiety, and negative thoughts and catastrophizing
- created additional web content such as adding a do-it-yourself self-diagnosis section to the "About Anxiety"

page, extending the list of associated symptoms, explaining additional types of anxiety disorder such as dermatillomania, and adding further detail to the process of becoming a volunteer

- written and posted blogs on the topics of older people and anxiety, high functioning anxiety, highly sensitive people and anxiety, work anxiety, anxiety and appetite, autism and anxiety, anger and irritability with anxiety, returning to work post lockdown, and placements for students
- added entries to the frequently asked questions section relating to costs and arrangements for therapy
- extended member benefits, including researching the provision of fidget toys and fidget jewelry

As the updates have been informed by recurring user queries, we can expect that they will be of value to users generally, and by linking the updated information to Ask Anxia's responses, a greater number of queries can be addressed automatically, out of hours and without the need for human intervention. Being able to identify trends in information-seeking requests has enabled Anxiety UK to respond operationally and strategically to meet the needs of its beneficiaries.

Quality Coding

Each week, the Anxiety UK team manually coded a sample of Ask Anxia's responses to monitor quality, which we have labelled Good, Okay, Bad, or Puzzled. On average, the Anxiety UK team would code 155 queries per week (ranging between 0 and 408). Good responses provided the appropriate information based on the query and constituted the response option that the human coder would have selected. The following example shows how Ask Anxia responds to the mention of "social anxiety" and directs users to the appropriate information:

> *Details about social phobia/social anxiety can be found here* [link provided]

Responses coded as Okay were not necessarily the optimal response option but were still topically relevant. For instance, in the case of a user posting a query that indicated that they wanted to talk to someone about dealing with anxiety, a human reader is likely to recognize the importance of talking to *someone*, whereas a response from Ask Anxia, which would subsequently be coded as Okay, might respond to the mention of anxiety as follows:

> *We provide a wide range of services and information for those dealing with anxiety, stress, or anxiety-based depression. Check out our homepage as a start here: [link provided] to see our calendar of upcoming events and our latest news.*

In this instance, the user is still directed to information that is likely to be useful to them, even if this was not the primary purpose of their query. Where this points to a potential recalibration of the service is that Ask Anxia was already programmed with a response that more directly attends to the question of speaking with a (human) member of the team.

Bad responses appeared when there was misalignment with what the coders, and we, can perceive as the user's intended meaning, with Ask Anxia generating an irrelevant or inappropriate response when a more pertinent option was available. For example, following a query that mentioned "joining," that is, membership with Anxiety UK, Ask Anxia generated the following response pertaining to joining a webinar:

> *You can book on to our next webinar here:* [link provided]

Again, reviewing the cause of the misalignment highlights ways in which Ask Anxia can be improved.

Finally, certain responses are generated when a more specific alternative, relating to the topic or passage of interaction, is not available, such as:

> *Sorry, I am not sure how to answer that, do feel free to use our website search bar which may find the answer for you or contact our team directly; we are open Mon-Fri 09:30-17:30 (excluding bank holiday) and our contact list can be found here:* [link provided]

Such responses were coded as Puzzled.

Of the 14,359 queries, 8669 (60.37%) were subject to quality coding, and the distribution of these according to the different quality labels is shown in Table 3.

The quality coding figures provided in Table 3 indicate that Ask Anxia generally performed well, providing a Good response in two-thirds of cases. Furthermore, we can see how this coding was applied weekly, given that the Anxiety UK team made adjustments to Ask Anxia's response options based on what they observed in their coding. Figure 1 indicates how the queries were coded between October 2021 and December 2022, showing the proportion of Ask Anxia responses that received the codes Good, Okay, Bad, and Puzzled.

Table 3. Number of queries coded according to each quality code (N=14,359).

Code	Queries, n (%)
Good	5801 (66.92)
Okay	911 (10.51)
Bad	1537 (17.73)
Puzzled	420 (4.84)

Figure 1. Quality coding of Ask Anxia responses as a percentage during each week from September 27, 2021, to December 26, 2022.

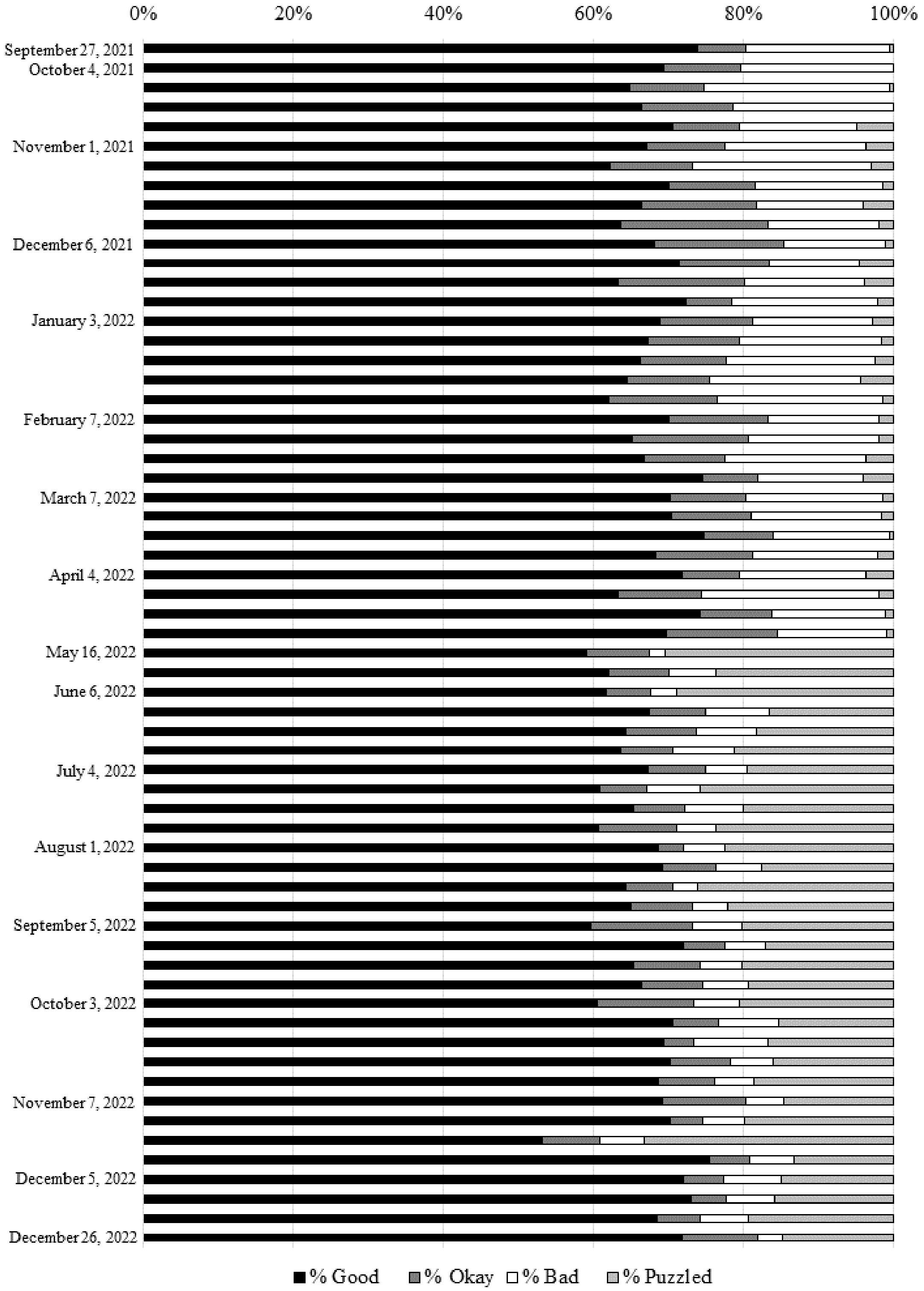


Figure 1 shows that the proportion of responses coded as Good was consistently >60% and that typically, Bad responses accounted for <30%. From May 2022 onward, there is a notable shift toward a smaller proportion of Bad and Okay responses and, instead, a greater number of Puzzled responses. As opposed to any change in the human evaluation of the responses or specific update to the software, this improvement reflects the ongoing work that the Anxiety UK team had been doing to calibrate the responses. This finding indicates that over time, Ask Anxia is less likely to generate an inappropriate or irrelevant response that could lead to disengagement from the user and more likely to provide a response that facilitates further engagement and opportunities for the user to find the relevant support.

The example of a Puzzled response provided earlier in this section on Quality coding was not initially one of the preprogrammed options but was added shortly after the launch (in September 2021), as the Anxiety UK team recognized a need to indicate when an optimal response could not be offered. The necessity of a Puzzled response is to be expected, given that the Anxiety UK team is not reasonably going to be able to anticipate the full range of queries users could conceivably submit. Furthermore, in many cases, a Puzzled response is preferable to a Bad response because it encourages the participant to remain engaged and try again. In work analyzing approximately 20,000 conversational exchanges between customers and a task-oriented chatbot for a Taiwanese banking firm, Li et al [13] focused on the problem of "nonprogress" responses, where users abandoned the dialogue. They identified a number of "reformulation" strategies when progress was halted, including rephrasing; adding different words; repeating the same words; and to a lesser extent, removing words [13]. This suggests that prompting the user to reformulate their query or to try an alternative mode of engagement, which the Puzzled response does, is preferable to closing down the exchange. Often, users simplify their reformulated messages [14], which increases the probability for pattern matching and Ask Anxia finding a relevant response.

While a Puzzled response can be the appropriate response, for example, when there is no suitable prewritten response or information provision, monitoring the instances when such a response is elicited highlights areas where Anxiety UK can consider extending the response options or the information and services they provide through their website.

Pattern Matching

In this section, we report some of the modifications made to Ask Anxia designed to attend to features of user queries that can potentially disrupt the pattern-matching mechanism of simple automated chatbot systems. For instance, the Anxiety UK team became aware that the use of certain punctuation affected the ability of the bot to respond correctly to the query and duly updated the program to navigate around such characters.

The simplicity of a pattern-matching procedure is demonstrated when the input (the user query) is not identical to the stimulus the chatbot is programmed to recognize, which can occur with misspellings. In addition to informing us that the term *anxiety* appeared 2461 times in the user queries, the wordlist generated in *#LancsBox* also indicates that the following (likely) misspellings of "anxiety" occurred: *aniety, aniexty, aniexy, anixety, anixity, anixtey, anixty, anxciety, anxeity, anxety, anxiatey, anxiery,* and *anxiey.*

Recognizing common misspellings of relevant terms can help to minimize the number of cases in which the chatbot cannot identify an appropriate response, and while it may be unfeasible to program the service to recognize all possible variants, the wordlist allows us to identify the most common.

The use of negation can result in false negatives, in cases where users produce the relevant stimulus but deny or distance themselves from the concept in their proposition, for example, "not needle phobia." In such instances, while a chatbot can be programmed to recognize negation (in terms such as *not*, *isn't*, or *no*), the query does not provide the input to determine what is the impetus of the query, and so recognizing negation would not then help to identify a suitable response. In such cases, the onus may be on the user to deduce how the inappropriate response has been generated (ie, seeing the pattern matching with their original query) and to try reformulating their message. A more proactive response, on the part of the service provider, would be to program the chatbot to recognize negation and to generate a Puzzled-type response that prompts the user to reformulate their query.

Users' queries might also include additional pattern-matching terms that do not constitute the primary focus of their message but which nevertheless prompt a response This was often the case with longer, more complex query formulations in which multiple competing trigger terms appeared. In most cases, one of the terms would elicit a corresponding response, but this might simply be a greeting to a query that happened to begin with the word *hello*. AbuShawar and Atwell [15] compare a "first-word" approach to a "most significant word" approach with respect to programming chatbots; they explain that the "most significant" word is determined according to low frequency, on the basis that a low-frequency word is what distinguishes an utterance and will favor informational content over high-frequency function words, such as *a*, *to*, *in*, etc. This approach increases the probability that the tool is responding to a "topic" word rather than, say, a grammatical word; however, implementation as part of a simple, pattern-matching chatbot would require additional programming. In the case of Ask Anxia, the Anxiety UK team introduced a prompt in August 2022 that advised users to construct their queries in a simple and direct manner, thereby maximizing the potential for Ask Anxia to recognize a relevant term. Such a response can be generated on the basis of the length of the query (ie, character or line count).

Managing Expectations

In the previous section, we have seen that optimizing a chatbot service relies, to some extent, on the understanding of the user that, for example, simple direct queries are likely to produce the best results. As such, there is a degree of familiarity, or "literacy," that can help to ensure that users find the support and information that is of most benefit to them. Working toward this alignment between user goals and service is also a case of managing expectations, first and foremost in relation to what the Ask Anxia service is and can do.

A series of updates applied to Ask Anxia reflected the increasing explicitness with which the Anxiety UK team described the automated nature of the service. Shortly after its initial launch (July 2021), the team supplemented the initial "Hello" response with the following message:

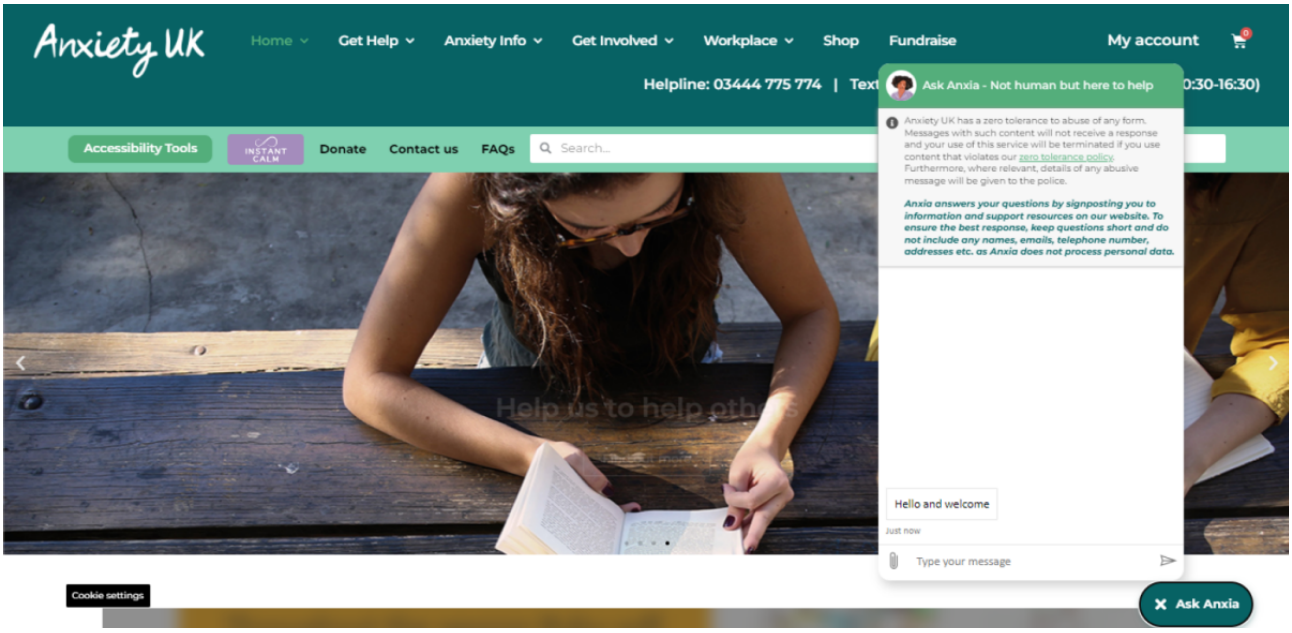
> *Hello, I am Anxia the Anxiety UK chat bot. I am here to provide you with advice and information. A brief disclaimer: this is not a crisis service, if you feel you are at risk, please contract 111 or 999. Now, how may I help?*

Subsequently, Ask Anxia has been relabeled an "eHelper" (August 2021) to avoid potential stigma associated with "chatbots" [3], then later (February 2022), the introductory prompt was rephrased to read the following: "Ask Anxia—Not human but here to help." Reviewing user queries, it becomes apparent why this clarification that the service is not operated by a human was required.

In Table 2, we say that one of the keywords for user queries was *oh*. Heritage [16] asserts that "where oh is produced as a response to information of some kind, it functions as a 'change of state' token; it registers, or at least enacts the registration of, a change in its producer's state of knowledge or information." In other words, the use of *oh* can indicate a degree of surprise or unexpectedness on the part of the recipient. When we refer to the queries, we see that often, this interjection reflected a realization on the part of the user that they were interacting with an automated service. The wordlist for the queries demonstrates the number of references to *bot* (87), *robot* (62), and *chatbot* (3), and we can extend our analysis in *#LancsBox* to determine frequencies of fixed phrases that include these terms, namely, *are you a robot* (21), *is this a bot* (16), *is this a robot* (14), *are you a bot* (12), etc.

On the one hand, this realization indicates a prior belief that the service was operated by a human and thereby might attest to the verisimilitude of the responses. On the other hand, the fact that this realization has come about indicates that such an illusion has been shattered, that is, because of an inappropriate response or perhaps because of repetition of the kind that is associated with pattern-matched chatbots with a limited number of responses [17]. Thus, with the aim of managing expectations and minimizing the potential for interactional trouble, the Anxiety UK team has worked toward more explicit signaling of the automated nature of the service (Figure 2). With this transparency, users can design their queries appropriately, and Anxiety UK can avoid too many instances where users become disillusioned by the potentially jarring realization that the interaction is not what they had presumed. Furthermore, we have established that some users may be more forthcoming knowing they are interacting with a nonhuman automated service [6].

Figure 2. The Ask Anxia chat window as it appears on the Anxiety UK website.



The Anxiety UK team had initially programmed Ask Anxia with a small number of light-hearted, conversational responses that contributed to a kind of persona, such as "I'm great, thanks for asking!" and "I don't have an age." Many of these responses were removed around September 2021 to October 2021, as they were often generated in inappropriate situations and there was a danger that they undermined the serious nature of the user query. Following their survey of motivations for using medical chatbots, Chang et al [3] determined that helping users acquire critical health information should take precedence over whether or not the chatbot appears empathetic or personable. Furthermore, when chatbots appear "humanlike," this can raise expectations about interactive capabilities, which in turn can negatively impact the interaction when the service's limitations are exposed [18]. Ultimately, transparency around the service's purpose and capabilities can help to avoid communicative misalignments and, given the high risk for responses that are designed to appear person like to actually appear flippant, such responses are arguably best avoided when user disengagement could potentially give rise to detrimental health consequences.

Finally, while Anxiety UK encourages users to be candid in their queries, on the basis that being direct will most likely mean that they can get the appropriate support, another series of developments to the Ask Anxia service has been to communicate a zero-tolerance policy for the use of profanities or abuse. A predefined response expressing this position was introduced in December 2021 and has since been subject to minor edits (also shown in Figure 2). Whether the use of profane language is motivated by frustration or is a more facetious "test" of the chatbot's capabilities (and there is evidence in the queries to indicate both), this does not include terms that are likely to be pattern matched to an informational resource. As such, there is no more preferable response for Ask Anxia to provide, other than to restate the zero-tolerance policy for such language, or alternatively, a response that encourages the user to reformulate their query. A summary of the updates described here is provided in Table 4.

Table 4. Summary of updates to the Ask Anxia service.

Date	Action
July 2021	• Extended welcome message clarifying that the service is operated by a bot as opposed to a human and to advise users that the service is not a crisis service.
August 2021	• Removed the limit cap on messages.
September 2021	• Removed a selection of prebuilt answers that were designed to make the bot seem more friendly and human-like but were giving inappropriate responses to sensitive queries. • Added a response advising users that the bot did not know how to answer their query ("Sorry I'm not sure how to answer that...") as an alternative to generating poorly matched responses from the existing content bank.
December 2021	• Added a "zero-tolerance policy" response due to a minority of users using profanities. Further clarification of this response was enacted through minor amendments in January 2022 and also February 2022.
July 2022	• Added a response discouraging users from including personal data in their queries.
August 2022	• Added a response advising users to keep messages brief, on the basis that longer queries gave rise to confusion and poor responses from the chatbot.
October 2022	• Updated the content bank to recognize regular typing errors related to existing prompt terms.

Future Developments

It is worth highlighting some of the anticipated developments that will be implemented to continue to optimize the Ask Anxia service. These developments primarily orient around connecting users to the appropriate mode of service, for instance, providing the connection to contact a (human) operator when this is recognized in the user query. There are also instances in which a more informed response, beyond the level of detail provided in the preprogrammed replies, is required; in such cases, where the user query seems to rely upon more specific contextual or personal circumstances, Ask Anxia can direct users toward the helpline. The Anxiety UK team continues to refine the Puzzled responses to encourage further engagement from the user, for instance, providing the prompt, "Can you phrase this differently?" Finally, the Anxiety UK team is working on developing a mobile app that has the chatbot functionality embedded within it, thereby providing another arm of support and format to use the Ask Anxia service to reach a wider audience and attend to different user preferences.

Discussion

Organizations implementing pattern-matching chatbots for the purposes of providing information and support will benefit from continuous review of the response options and queries that users submit to their service. Furthermore, an initial set of programmed responses will likely need to be extended, and this will be informed by the nature of the queries that users submit. Our corpus analysis of frequently used terms in user queries to Ask Anxia demonstrated that the initial set of programmed responses was well aligned with the concerns of users but nevertheless helped to highlight areas where additional materials could prove to be useful. The manual quality coding of responses showed that Ask Anxia performs well, offering Good responses at a rate consistently >60%, and this procedure helped to identify areas where responses could be developed to address information gaps or otherwise refined to discern, for example, queries about needles generally and questions about specific vaccinations.

With respect to lessons learned through the implementation and review of the service, first, we have highlighted the informal nature of user queries, which often included ritualized greetings (*Hi* and *hello*). As such, it is useful to have a chatbot response that simply provides a greeting in kind. However, it is important to note that if a user greeting appears at the beginning of a more elaborate query, a response that attends to the topic of the query would be more appropriate.

Second, we have recommended that when an appropriate response cannot be readily identified, there is value in continuing the exchange, that is, encouraging the user to reformulate their query and thereby create additional input from which the chatbot can match an appropriate response. Researchers have highlighted the dangers of "nonprogress" responses that result in user disengagement [13]. Thus, while service providers are unlikely to be able to anticipate the full range of queries their users will submit, they can at least work to facilitate further engagement and use a preprogrammed, albeit uncertain reply to instruct participants on how best to elicit an acceptable response.

Third, we have seen that it is important to manage users' expectations about what the tool can provide, which includes being explicit that the service is not provided by a human. Relatedly, responses that presented humanlike qualities proved to be of limited value, potentially raising expectations that the tool could offer humanlike judgments. Simple, pattern-matching chatbots such as Ask Anxia are best suited to "frequently asked questions"–type services, rather than more interactional, relationship-building tasks [17]. The benefit of these less-complex systems is that they are easier to program and implement and so can be adopted by service providers with minimal knowledge of the computational systems involved. It is important, nevertheless, to be cognizant of the limitations of such services. For instance, Ask Anxia does not track conversations over multiple turns but rather treats each post as a new query; as such, any pertinent information provided at a previous turn is lost, and users may find themselves having to restate the fundamental purpose of their query. Similarly, the quality of Ask Anxia's performance is likely to diminish with longer, more complex queries, as it becomes more difficult to discern a singular, relevant prompt. Subsequently, users will be discouraged from providing contextual information (Figure 2) and are unlikely to receive personalized support in this mode. Simple chatbots, therefore, are arguably best used as part of an array of support options, including those which allow for more nuanced exchanges, for example, with a human provider over the telephone.

Laranjo et al [1] assert, based on a systematic review, that applications of chatbots in health care are in the early stages of development and evaluation. Furthermore, the systems used in health care lag behind those used in domains such as travel information and restaurant selection. As their deployment can have consequences for health outcomes, it is appropriate that such systems are continuously tested and evaluated. Language analysis is key to understanding both how users express themselves in queries to chatbots and the design of appropriate responses, and so, we advocate for the continued application of procedures such as those of corpus linguistics to support the extended use and performance of chatbots in health care.

Conclusions

The launch of the chatbot Ask Anxia was designed to support Anxiety UK in delivering information and support services to people concerned with anxiety disorders. The number of queries submitted to Ask Anxia, particularly out of hours, attests to the value of the service. In this study, we have demonstrated that procedures from corpus linguistics can help to identify patterns in user queries that reflect their needs and expectations of the service as well as direct us to where potential breakdowns in communication occur. For chatbot services to achieve optimal performance, human oversight is required, particularly during the first 6 to 12 months. Thereafter, less staff intervention is likely to be needed.

Acknowledgments

The authors would like to acknowledge the technical support provided by Roger Kadama in developing the Ask Anxia service. LC and PB were supported by funding from the Economic and Social Research Council (ES/R008906/1). The funding source had no role in the design of this study.

Authors' Contributions

All authors participated in the conceptualization of the paper and reviewed the manuscript. NN, NL, and DS carried out the manual coding of Ask Anxia's responses. LC and PB carried out the corpus analysis of user queries. LC wrote the manuscript with contributions from NN and NL on developments to the Ask Anxia service. NL conceptualized the Ask Anxia service. NN, NL, and DS contributed to the initial development of Ask Anxia and each continues to shape its development.

Conflicts of Interest

NN, NL, and DS are employed by the national charity, Anxiety UK, and were involved in the development of the Ask Anxia chatbot service.

References

1. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
2. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: a mixed-methods study. Digit Health 2019 Aug 21;5:2055207619871808 [FREE Full text] [doi: 10.1177/2055207619871808] [Medline: 31467682]
3. Chang IC, Shih YS, Kuo KM. Why would you use medical chatbots? Interview and survey. Int J Med Inform 2022 Sep;165:104827. [doi: 10.1016/j.ijmedinf.2022.104827] [Medline: 35797921]
4. Amiri P, Karahanna E. Chatbot use cases in the COVID-19 public health response. J Am Med Inform Assoc 2022 Apr 13;29(5):1000-1010 [FREE Full text] [doi: 10.1093/jamia/ocac014] [Medline: 35137107]
5. Tudor Car L, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng YL, et al. Conversational agents in health care: scoping review and conceptual analysis. J Med Internet Res 2020 Aug 07;22(8):e17158 [FREE Full text] [doi: 10.2196/17158] [Medline: 32763886]
6. Palanica A, Flaschner P, Thommandram A, Li M, Fossat Y. Physicians' perceptions of chatbots in health care: cross-sectional web-based survey. J Med Internet Res 2019 Apr 05;21(4):e12887 [FREE Full text] [doi: 10.2196/12887] [Medline: 30950796]
7. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry 2019 Jul;64(7):456-464 [FREE Full text] [doi: 10.1177/0706743719828977] [Medline: 30897957]
8. Baker P, Hardie A, McEnery T. A Glossary of Corpus Linguistics. Edinburgh, UK: Edinburgh University Press; 2006.

9. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th edition. New York, NY: American Psychological Association; 2013.
10. McManus S, Bebbington P, Jenkins R, Brugha T. Mental health and wellbeing in England: adult psychiatric morbidity survey 2014. NHS Digital. 2016. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/556596/apms-2014-full-rpt.pdf [accessed 2024-04-29]
11. Brezina V, Weill-Tessier P, McEnery A. #LancsBox v.6. Lancaster University. 2021. URL: http://corpora.lancs.ac.uk/lancsbox/ [accessed 2024-04-29]
12. Love R, Dembry C, Hardie A, Brezina V, McEnery T. The Spoken BNC2014: designing and building a spoken corpus of everyday conversations. Int J Corpus Linguist 2022 Jul 5;22(3):319-344. [doi: 10.1075/ijcl.22.3.02lov]
13. Li CH, Yeh SF, Chang TJ, Tsai MH, Chen K, Chang YJ. A conversation analysis of non-progress and coping strategies with a banking task-oriented chatbot. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20; April 25-30, 2020; Honolulu, HI p. 1-12 URL: https://dl.acm.org/doi/abs/10.1145/3313831.3376209 [doi: 10.1145/3313831.3376209]
14. Myers C, Furqan A, Nebolsky J, Caro K, Zhu J. Patterns for how users overcome obstacles in voice user interfaces. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18; April 21-26, 2018; Montreal, QC p. 1-7 URL: https://dl.acm.org/doi/10.1145/3173574.3173580 [doi: 10.1145/3173574.3173580]
15. AbuShawar B, Atwell E. ALICE chatbot: trials and outputs. Comp Y Sist 2015 Dec 27;19(4):625-632. [doi: 10.13053/cys-19-4-2326]
16. Heritage J. Oh-prefaced responses to inquiry. Lang Soc 1998 Jul 1;27(03):291-334. [doi: 10.1017/s0047404598003017]
17. Boucher EM, Harake NR, Ward HE, Stoeckl SE, Vargas J, Minkel J, et al. Artificially intelligent chatbots in digital mental health interventions: a review. Expert Rev Med Devices 2021 Dec 31;18(sup1):37-49 [FREE Full text] [doi: 10.1080/17434440.2021.2013200] [Medline: 34872429]
18. Gnewuch U, Morana S, Mädche A. Towards designing cooperative and social conversational agents for customer service. In: Proceedings of the 2017 International Conference on Information Systems - Transforming Society with Digital Innovation. 2017 Presented at: ICIS '17; December 10-13, 2017; Seoul, South Korea p. 1-13 URL: https://aisel.aisnet.org/icis2017/HCI/Presentations/1/ [doi: 10.17705/4icis]

Edited by A Kushniruk; submitted 23.10.23; peer-reviewed by M Chatzimina, AL MacNeill; comments to author 20.03.24; accepted 05.05.24; published 17.06.24.

Please cite as:
Collins L, Nicholson N, Lidbetter N, Smithson D, Baker P
Implementation of Anxiety UK's Ask Anxia Chatbot Service: Lessons Learned
JMIR Hum Factors 2024;11:e53897
URL: https://humanfactors.jmir.org/2024/1/e53897
doi:10.2196/53897
PMID:38885016

Original Paper

Alarm Management in Intensive Care: Qualitative Triangulation Study

Lina Mosch[1,2], MD; Meltem Sümer[3], MSc; Anne Rike Flint[1], MSc; Markus Feufel[3], MSc, PhD; Felix Balzer[1], MSc, MD, Dr rer nat; Frauke Mörike[3,4*], MA, PhD; Akira-Sebastian Poncette[1,2*], MD

[1]Institute of Medical Informatics, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt Universität zu Berlin, Berlin, Germany

[2]Department of Anesthesiology and Intensive Care Medicine, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt Universität zu Berlin, Berlin, Germany

[3]Division of Ergonomics, Department of Psychology and Ergonomics (IPA), Technische Universität Berlin, Berlin, Germany

[4]Department of Rehabilitation Sciences, Research Unit Work, Inclusion and Technology, Technische Universität Dortmund, Dortmund, Germany

[*]these authors contributed equally

Corresponding Author:
Lina Mosch, MD
Institute of Medical Informatics
Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt Universität zu Berlin
Charitéplatz 1
Berlin, 10117
Germany
Phone: 49 30450581 016
Email: lina.mosch@charite.de

Abstract

Background: The high number of unnecessary alarms in intensive care settings leads to alarm fatigue among staff and threatens patient safety. To develop and implement effective and sustainable solutions for alarm management in intensive care units (ICUs), an understanding of staff interactions with the patient monitoring system and alarm management practices is essential.

Objective: This study investigated the interaction of nurses and physicians with the patient monitoring system, their perceptions of alarm management, and smart alarm management solutions.

Methods: This explorative qualitative study with an ethnographic, multimethods approach was conducted in an ICU of a German university hospital. Using triangulation in data collection, 102 hours of field observations, 12 semistructured interviews with ICU staff members, and the results of a participatory task were analyzed. The data analysis followed an inductive, grounded theory approach.

Results: Nurses and physicians reported interacting with the continuous vital sign monitoring system for most of their work time and tasks. There were no established standards for alarm management; instead, nurses and physicians stated that alarms were addressed through ad hoc reactions, a practice they viewed as problematic. Staff members' perceptions of intelligent alarm management varied, but they highlighted the importance of understandable and traceable suggestions to increase trust and cognitive ease.

Conclusions: Staff members' interactions with the omnipresent patient monitoring system and its alarms are essential parts of ICU workflows and clinical decision-making. Alarm management standards and workflows have been shown to be deficient. Our observations, as well as staff feedback, suggest that changes are warranted. Solutions for alarm management should be designed and implemented with users, workflows, and real-world data at the core.

(JMIR Hum Factors 2024;11:e55571) doi:10.2196/55571

KEYWORDS

digital health; transdisciplinary research; technological innovation; patient-centered care; qualitative; ethnographic; ethnography; intensive care unit; ICU; intensive care; German; Germany; Europe; European; interview; interviews; alarm; alarms; intelligent; artificial intelligence; grounded theory; experience; experiences; attitude; attitudes; opinion; opinions; perception; perceptions; perspective; perspectives

Introduction

Background

In intensive care units (ICUs), continuous monitoring of multiple vital signs results in a high number of alarms [1] intended to inform staff of critical patient conditions, thus ensuring patient safety [2]. However, a large proportion of alarms do not require medical intervention [3], which, along with the sheer amount of alarms, disturb patient care [4]. In addition, the desensitization of ICU staff to alarms, "alarm fatigue" [5-7], is a threat to patient safety as it causes slower or no reactions to alarms [8,9]. Alarm fatigue is associated with working conditions and individual staff characteristics in deteriorating alarm monitoring performance [10]. Thus, for an improved monitoring performance, an understanding of workflows and the inner setting of a unit conducting vital sign monitoring is essential.

Alarm management standards have been proven effective in reducing unnecessary alarms [11,12]. Individual alarm thresholds, tailored to a patient based on previous monitoring data, alarm logs, and the medical data stored in the patient data management system (PDMS), would further improve alarm management [13,14]. One approach to this is using artificial intelligence (AI) to integrate these data and suggest alarm thresholds for an individual patient or actions to take after an alarm [15,16]. The research project *Intelligent Alarm Optimizer for the Intensive Care Unit* (INALO) follows this approach by creating a data set with semiautomatically annotated data (ie, alarms and the reactions to it) and using machine learning to predict the probability of an alarm to be actionable or nonactionable [17].

Before introducing new procedures or technologies in a complex sociotechnical environment such as an ICU, the characteristics of this setting and the individuals working there should be assessed [18,19], especially focusing on lived everyday work practices [20]. Since many alarm reduction solution approaches known in research have failed because of being too general [5], the interaction of the ICU staff with the patient monitoring system and the practiced alarm management should be investigated.

Objective

This qualitative study aimed:

1. to investigate the sociotechnical system ICU, with regard to the interaction of staff with the patient monitoring system and its alarms and
2. to understand the staff's perceptions of alarm management and the potential they see in the use of AI in this context.

We aimed to address the following research questions:

1. What is the sociotechnical role of patient monitoring in the work practice of nurses and physicians of the research site?
2. How does the staff, including nurses and physicians, of an ICU of a university hospital interact with the patient monitoring system, especially their interaction with alarms?
3. What is the attitude of physicians and nurses toward intelligent alarm management systems?

Methods

We consulted the Standards for Reporting Qualitative Research for reporting this research [21].

Qualitative Approach and Research Paradigm

This ethnographic study follows an explorative approach with (1) observation-based data collection triangulated with (2) semistructured interviews supported by (3) a self-reported overview of activities mapped out by each participant [22-24]. The research paradigm is postpositivist. We aimed to maintain a holistic, observing perspective while acknowledging that all the collected data and the derived findings are fallible, value laden, and need to be reflected from different researchers' perspectives [25]. The use of ethnographic methods fulfills the requirements of the postpositivist paradigm; it can create an understanding of social structures and capture the role and interaction with a technical system within those structures, revealing underlying sociotechnical relationships and patterns [26]. In order to build this comprehensive understanding, ethnographic research usually uses other methods in addition to observation [27]. Therefore, in addition to observation, semistructured interviews were conducted, and a participatory task was set for this work.

Researcher Characteristics and Reflexivity

The interdisciplinary research team consisted of a physician with work experience in medical informatics and anesthesiology and a research focus in implementation science and patient monitoring (LM); a psychologist enrolled in the master's program, *Human Factors*, exploring the interface between humans and technology (MS); a digital clinician scientist with work experience in medical informatics, anesthesiology, and internal medicine and a research focus on patient monitoring and alarm management (ASP); a professor of medical informatics (FB); a professor of inclusive work systems with research focus on ethnographic methods for technology design (FM); and a professor of ergonomics with a focus on human factors research in medical work environments (MF). MS was methodologically trained by FM and conducted the data collection from the etic perspective, that is, from the perspective of a person foreign to the field and the field of activity, supervised closely by FM throughout the field work. LM's and ASP's clinical perspectives provided a strong interdisciplinary contextualization of the data and informed the analysis. Throughout all phases of the research, FM, MS, and LM met regularly to discuss and reflect on the data collection process and preliminary findings. None of the research team members had a direct professional relationship with the research field.

Context

The motivation for this research, along with literature evidence of alarm fatigue, was a previous study that we had conducted to identify clinical requirements of future patient monitoring. We found that staff perceived alarm management as insufficient, threatening patient safety and disturbing workflows [28].

Setting

The research field of this study was an ICU of a German university hospital, where up to 21 patients can be treated in nine 2-bed rooms and six 1-bed (isolation) rooms. The devices used in the patient rooms were a continuous monitoring system (Philips IntelliVue) with sensors monitoring multiple vital parameters (eg, electrocardiogram, blood pressure, temperature, oxygen saturation, intracranial pressure, and electroencephalogram); a ventilator (Dräger Evita V800); infusion pumps (Agilia connect by Fresenius Kabi); and, if needed, a temperature management system (Arctic Sun by BD) or a dialyzer (various manufacturers). A table near the patient's bed with a computer enabled the caregivers to call up the PDMS to retrieve patient data and to document examination results. The PDMS used in this setting was COPRA 6 (COPRA System GmbH) [29], while the hospital information system used was i.s.h.med by Cerner [30]. The patient monitoring system installed on the ward included screens at the nurses' workstations directly outside the patient rooms and in the ward pulpit, the conference room, and a medication room. These screens displayed an overview of the vital signs of patients in adjacent rooms or of all patients in the ward.

Intelligent Alarm Optimizer Project

This study was conducted before the implementation and testing of INALO [31]. This project aims to improve alarm management for medical personnel using AI to decrease alarm loads, therefore easing the burden of alarm fatigue. The methodological approach included integrating and annotating data from multiple sources, including the hospital-wide patient monitoring system, the PDMS, and the hospital information system.

Access to the Field and Sampling Strategy

Access to the field was provided by the medical supervisor of this work (ASP) as well as the nursing management of the researched unit. The observations were announced to the entire ward staff by the ward manager via email, and the staff members were provided with the opportunity to address any concerns or queries they might have. A total of 12 observations were conducted by MS, each of which involved shadowing a nurse or physician for the duration of a shift (Table 1), for a total of 102 hours of observational data. The first 2 observations, shadowing both a nurse and a physician, were considered pilot observations. Their purpose was to become familiar with the environment of an ICU, test the methodology of observation and interviewing, and adapt the interview questions. The data from pilot observations were not used for later analysis.

Table 1. Number of observations divided by shift and profession and level of expertise, respectively[a].

Profession and level of experience	Early shift	Late shift	Night shift
Nurse (1.5-30 years of work experience in the ICU[b])	3	2	1
Resident physician (1-2 years of work experience in the ICU)	2	1	—[c]
Senior physician (7 years of work experience in the ICU)	1	—	—

[a]One participant was shadowed twice.

[b]ICU: intensive care unit.

[c]Not applicable.

The 5 nurses shadowed were suggested by the ward manager and scheduled with MS for different observation days after they had given written consent to participate in the study. Physicians were contacted by MS partly with the help of the medical supervisor or directly in the course of a field observation. In total, 4 physicians agreed to participate in the study.

The aim was to shadow a sample of staff members representative of the whole team, thus covering the greatest possible variety of perspectives in the research. Nurses and physicians of different ages and experience levels in an ICU were shadowed in all possible work shifts (early, late, and night shifts). The number of accompanying nurses and physicians per shift is listed in Table 1, where the experience level of working on an ICU is shown as the duration of work in intensive care.

Ethical Considerations

The study was approved by the ethics committee of the Charité Universitätsmedizin Berlin (EA4/218/20). Field notes and transcripts were anonymized or, if complete anonymization was not possible, pseudonymized. The data are stored on an internal institutional server, encrypted so that only the research team had access to it. Audio recordings were deleted after transcription. All staff members shadowed and interviewed for this study were informed about the research project and its aims and gave their written consent to participate. There was no compensation given to the staff members who were shadowed or interviewed.

Data Collection Methods and Instruments

Field Observation

The main data collection method in this study was field research in the form observation. In ethnographic research, the observation part varies by the degree of participation of the researcher in the field [32]. The setting of the field observations and the nonmedical background of the researcher suggest a nonparticipatory observer role (shadowing), in which events are observed in the background without involvement in any work processes. This methodology has been previously used in clinical settings [33-36]. Although the researcher observed the field from the background most of the time, she was also involved in conversations during the shifts by the shadowees. This gave her the opportunity to ask clarifying questions, which was beneficial in gaining insights. The observations were performed from April to October 2021. Breaks of up to 3 months were taken for interim analysis of the data. This allowed for the readjustment of focus for the remaining observations and the evaluation of inductive thematic saturation [37], ultimately leading to the determination of the number of field observations as 7.

Semistructured Interviews

The interviews serve to explore the nonobservable topics relevant to the research question. They add to the observation method in the sense of methodological triangulation [22]. In particular, reactions to the planned AI-based INALO system and specific inquiries about the handling of the monitoring and its alarms were assessed. By not including the explicit term *AI* in the interview guide, the goal was to avoid potential biases or preconceived notions associated with the term, allowing for a more neutral and open interview discussion. *AI* was considered a polarized term, evoking personal opinions, as people may have subjective and varied views on the topic [38-40]. Consequently, the interviewees were provided with an explanation of the functionality of the INALO system, leaving out the specification *AI* (refer to question 5 in the interview guide in Textbox 1). The interview guide was developed through discussions of the research team and informed through the first 2 pilot observations

Textbox 1. Interview guide.

Questions

- Are there certain situations in which the sounding of an alarm is particularly inconvenient for you? Which ones?
- What role does managing alarms and especially their thresholds currently play during your work?
- How do you know it is time to adjust the alarm thresholds of certain parameters?
- When do you take the time to adjust an alarm threshold?
- Imagine a system, in addition to patient monitoring, that controls the alarm thresholds for oxygen saturation, arterial blood pressure and heart rate. If it registers that the current thresholds are no longer suitable for the patient in question, it suggests suitable alarm thresholds. What do you think about such a system?

In total, 7 interviews were performed individually with the 7 shadowed staff members (Table 1) after the end of a work shift in an open space to minimize outside distractions. The mean length of the interviews was 37:57 (range 56:22-15:05) minutes. We considered data saturation [37] including notes from field observations and interview transcripts. It was achieved after 6 field observations and, correspondingly, 6 interviews, which included early, late, and night shifts as well as nurses, resident, and senior physicians.

Participatory Task

In addition to the observation and the interview, following the triangulation strategy [41], a participatory task was conducted with the study participants. This helped gain insight into the representation of patient monitoring in the daily work routine and workflows of ICU staff members.

In a classic note-and-pen task, participants were asked to draw a pie chart of all the thoughts and tasks they think about during a work shift, with each piece of pie representing a thought, a task, or a task area. The size of the pie pieces was intended to show how the respondent’s cognitive resources were divided during a shift. The participants were requested to prioritize the tasks that they had visualized based on their subjective perception. This was done to compare the estimated time allocated to each task in their workday with its perceived importance. The goal was note only to obtain indirect evidence on the importance of monitoring and alarms but also to understand the work routine in general from the respondent’s perspective.

Data Collection Instruments and Technologies and Data Processing

Handwritten field notes were prepared following the methods of Emerson et al [42], and the date, location, and content of the observation were taken and promptly consolidated in MAXQDA (VERBI GmbH) [43] after each observation session. Interviews were recorded and transcribed according to the minimal transcript based on GAT 2 [44] and consolidated in MAXQDA. The participatory task was recorded in order to be able to use follow-up questions about the processing of the task as well as incidental comments for the later analysis.

Data Analysis

For inductive analysis, a grounded theory approach was followed [45]. The consolidated field notes and transcribed interview data from 6 observations were analyzed in MAXQDA 2022 and coded line by line in a first step, deriving thematic codes from the data while trying to neglect any relationship of the codes to the overarching research questions. In a second step, memoing was started to extract meaning from the data and to create and map an initial overview of themes [46]. This overview represented a preliminary code system after 6 observations. After all the gathered data were collected, coded, and tagged with open memos, the result consisted of 1336 coded data sections assigned to 47 parent codes and 70 thematically subordinate codes and tagged with 207 memos. Memos (ie, field notes with headlines) were organized thematically and considered in the context of their associated data sections (eg, interactions with the monitoring systems and alarm management). To ensure this thematic clustering according to the memo headline was rightful, we went back into the raw data (field notes) to evaluate whether the groups of memos with a similar headline could indeed be grouped together based on their raw data points. This helped to reevaluate and refine the derived themes from the memos (eg, integrating memos for more specific details and adding key points from field notes to memos from interviews and vice versa). When analyzing the data collected through the participatory task, the recorded topics were written down, divided according to occupational group (nurses and physicians), and their occurrence was counted. As a result, an overview of superficial and seemingly less relevant task areas was created. The resulting themes from memos, codes, and topics from the participatory task were summarized and put in writing.

Techniques to Enhance Trustworthiness

Regular research meetings took place, where LM and MS discussed the findings and reflected on them. The code system (Multimedia Appendix 1) and memos were checked by both researchers under the supervision of FM. The combination of a psychologically trained Human Factors graduate student (MS) together with an expert in ethnographic research for work systems design (FM) and a physician with professional experience in an ICU and expertise in qualitative methods and implementation science (LM) was chosen to achieve the best balance of perspectives and topic prioritization. Interdisciplinary approaches are important to leverage the potential of research on the intersection of human-computer interaction, information systems, and health [47]. In addition, the multimethod study design with triangulation of complex data allowed for an increased credibility and trustworthiness of the results.

Results

Overview

The following 3 topics were identified from the data, following the research questions:

- the perception of the role of monitoring in the ICU
- the management and communication of vital sign limits (*dealing with alarms*)
- wishes and concerns regarding the intelligent alarm management system (eg, INALO)

An overview of nurses' and physicians' perceptions can be found in Table 2.

Table 2. Perceptions of monitoring and alarm management divided by professions.

Categories	Nurses	Physicians
The role of monitoring in the intensive care unit	• Use monitors at the bedside • Direct monitoring of vital signs at the screens	• Use monitors in the pulpit • Use nurses as monitoring filter
Dealing with alarms	• Subconscious adaptation to alarm patterns • First reaction to alarms • Implement alarm thresholds	• Subconscious adaptation to alarm patterns • Set and adjust alarm thresholds
Intelligent alarm management	• Concerned about the system's functionality • Highlighted the importance of understanding the system's operational principles • Concerned about excessive confidence	• Perceived as positive for lowering alarm burden • Perceived as valuable for new patients in the ward or upon returning from leave

The Perception of the Role of Monitoring in the ICU

Continuous monitoring of vital signs is an essential component of intensive care management of critically ill patients and a ubiquitous part of the ICU under study. According to a senior physician, even a short period without monitoring would be "grossly negligent."

The nursing staff primarily used the monitors located directly at the patient's bedside (ie, *direct* monitoring), as well as the overview screens at their workstations outside the patient rooms. During lengthy nursing activities in a room, it was observed that nurses used the monitoring system's function to display other patients' vital sign data on the screen of the bed where they were busy. Due to the larger number of patients under their care, physicians relied on the feedback from the nursing staff in the event of a critical situation (nurses as *monitoring filter*). They also paid closer attention to alarms on the overview monitors at the pulpit, and it was important that alarm thresholds were well set so that potentially life-threatening situations did not go unnoticed for long:

> *Especially when patients deteriorate, you notice that very often because the alarms are triggered all the time...These alarms...also signal something to us[...], without us seeing the patient directly[...], and if [the thresholds] were always set lower and it simply no longer alerts me, I'd only notice it when it's really bad or too bad or the catastrophe is there, and I wouldn't want that.* [Interview, physician]

When making therapy decisions, physicians considered the target values for specific parameters that had to be adhered to, based on therapy plans. Continuous monitoring of vital signs was crucial for physicians in ensuring that these target values are met.

Although monitoring was used in various areas of activity, it was not considered a central activity in and of itself. In the participatory task, monitoring was not mentioned by anyone. Alarms occurred in 2 of the 5 participatory tasks completed. However, when asked to indicate the role of monitoring in all the tasks written down, interviewees assigned monitoring to most of them (Multimedia Appendix 2).

Dealing With Alarms

The predominant observation regarding the handling of alarms was that they were directly paused or completely ignored, unless it was a red alarm. Often, there was no further reaction or intervention following an alarm.

However, as described in the previous section on monitoring, even an alarm that did not require intervention could provide helpful information about the progress of an individual patient's condition. Both physicians and nurses showed a reliance on the alarms' feedback to be passively informed of the change in their patients' condition (subconscious adaptation to alarm patterns). In contrast, the alarms interfered with their work:

> *[...]as you may have noticed, and if this alarm comes up all the time...I actually look at it [the overview monitor] a lot. That disturbs me in my regular workflow.* [Interview, physician]

The alarm thresholds of vital parameters were set and adjusted either based on a patient's admission to the ward or return from the operating room or based on evaluations of the thresholds during the course of therapy. In the case of newly admitted patients, the threshold values were determined directly on admission, with the parameters being set according to standard values or having to be determined individually depending on the clinical picture. Nurses, in consultation with physicians, adjusted alarm thresholds according to the patients' condition to avoid unnecessary alarms. The decision to adjust thresholds varied based on the nurse's experience and familiarity with the patient. However, these informal procedures sometimes led to inappropriate thresholds on the next shift, as they were not always updated in the COPRA PDMS [29]:

> *...I come to the ward, and I don't know the patient. COPRA is binding, what is documented there should be correct and implemented at the bedside. If you're lucky, it says everything about which blood pressure values are desired...You have to ask about all that [if it is not entered correctly] and that's sometimes a bit annoying, until you find out which doctor is responsible for your room, where he is and what his telephone number is. It doesn't have to be like that.* [Interview, nurse]

Ideally, the determined threshold values were reevaluated daily in the morning shift. Often this was not the case due to a lack of time. Rather, this reevaluation only happened, if a new test result or diagnosis was found or in the case of disturbing, conspicuous, and unnecessary alarms. The chain of communication regarding the change of alarm thresholds from their determination to their implementation in monitoring ran hierarchically downwards, from senior physicians to assistant physicians to nursing.

If the alarm of a certain parameter would continue despite intervening measures or if no measures had been necessary at all, the alarm thresholds would be adjusted. Physicians and nurses frequently sought each other out for direct and immediate instructions. The official way of issuing instructions was via the COPRA PDMS, but according to staff, this was often too time-consuming, with delays for actual implementation:

> *And this adjustment [of the thresholds]...doesn't take place so much on this piece of paper, but rather on the patient himself, that I say, ok, we have to change the threshold here[...], then you discuss it next to the device itself with the nursing staff.* [Interview, physician]

Staff Perceptions of Intelligent Alarm Management

There were mixed reactions regarding an automated system to improve alarm management on the part of the nursing staff interviewed. Some expressed reservations, while others were open to the system described but said it would have to work well. There were concerns about the functionality of an intelligent system. The importance of understanding how such a system operates was highlighted. Staff members argued that currently a lot of responsibility was being handed over to technology anyway and that it would be more appropriate to be cautious about overconfidence:

> *I probably rely more on my own experience and won't put it in the hands of a machine that I don't know how it's programmed, what kind of things it reacts to, or what it takes as a basis for the recommendations. That would make a lot of sense for young colleagues who don't have a lot of experience yet, but you end up relying more on your own experience.* [Interview, nurse]

> *The patients, the clinical pictures, and that's all so individual and different that I can't imagine that a machine can [suggest thresholds].* [Interview, nurse]

Second, the statement was made that with clinical experience, one would notice when a threshold should be changed, and therefore, no system would be needed to do so:

> *I know already by my experience that I must intervene starting from certain values...you observe the patient, you observe the monitor and you already see...that you have to do something.* [Interview, nurse]

Thus, staff members saw the added value of an intelligent alarm system for less experienced colleagues and for leasing nurses, as the former would not yet have developed a sense of threshold assessment and the latter would not be familiar with patients. In addition, staff members saw a benefit in the use of the system to ensure regular evaluation and adjustment of alarm thresholds, preventing them from being forgotten.

Physicians reported that any solutions leading to a reduction in alarms would be welcome due to frequent alarms disturbing their concentration during work, especially in the ward pulpit:

> *[...]If it leads to the fact that the alarms are optimized, then I think it is already a relief, because I get fewer false alarms and thus, I can do the rest of the work more concentrated. So, that would already be an advantage. Certainly, it doesn't take away my cognitive work whether these alarms are, so to speak, suitable or not, I have to do that, but that's also my job.* [Interview, physician]

Finally, the advantage of such a system was observed by physicians for patients who were new to the ward and whose condition they did not yet know. The physicians said that it was also helpful to receive threshold suggestions when returning to the ward after a leave of absence, as they also needed to reacquaint themselves with patients.

Discussion

Principal Findings

Patient monitoring is an integral part of the work routine in the ICU. However, standards for working with the system were not implemented in clinical routine in the studied ICU. Alarm management is one of the core interactions with the monitoring system. Most alarms were confirmed without a reaction or intervention. The setting and adjusting of alarm thresholds were performed upon the arrival of a patient in the ICU or over the course of treatment by (senior) physicians and implemented by nursing staff. Perceptions of an intelligent system to suggest alarm thresholds varied: physicians saw potential advantages in a relief of the flood of (nonactionable) alarms by individualized alarm thresholds. In contrast, both nurses and physicians were skeptical about the capability of an automated system to perform the complex task of interpreting alarms and suggesting thresholds, encompassing the integration of different data and information. The importance of knowing how such a system worked internally and how it took decisions was highlighted.

Guidelines for Patient Monitoring and Alarm Management in the ICU

This work highlights the essential role of vital sign monitoring in the daily routines of all health care professionals in the ICU. Yet, to the authors' knowledge, interprofessional, systematic standard, and evidence-based guidelines for patient monitoring in clinical practice remain limited. Alarm management, a fundamental aspect of interacting with monitoring systems, warrants special attention, since alarm overload in ICUs hampers adequate response by personnel [48]. The American Association of Critical Care Nurses published recommendations for clinical alarm management, focusing on personalization of alarms and interprofessional alarm management strategies and highlighting the potential of smart alarms [49,50]. Our research supports the need for a wider clinical implementation of such recommendations, as we saw that a majority of alarms are disregarded or confirmed without any medical response or intervention [51,52].

Take Aways for Future Alarm Management

It is crucial to understand user perspectives when developing intelligent (AI-based) systems for alarm management in intensive care medicine [49]. The following aspects should be considered.

- Skepticism exists among staff members regarding the ability of an intelligent system to integrate diverse data formats and information to effectively interpret alarms and propose suitable thresholds. It might stem from a lack of knowledge and understanding of the internal workflows and decision-making processes of such systems. The potential benefits of integrating intelligent systems to automatically suggest personalized alarm thresholds and alleviate alarm fatigue [16,53] were acknowledged by physicians in our study.
- Standardized workflows for alarm management were not existent, and alarm thresholds and their adaptation were communicated irregularly and in a variety of ways. Well-defined and clearly communicated standard operating procedures (SOPs) for alarm management could address some of the challenges faced by health care professionals in the ICU [54,55]. Intelligent alarm management solutions should be implemented in clinical environment with well-established alarm management SOPs.
- Alarm management was performed mainly after new test results for the patients were received, a diagnosis was confirmed, or when the alarms were frequent and obviously unnecessary. This indicates that the data used for future alarm management systems based on AI need to mirror highly individual and complex medical conditions. Patients and clinical routines can differ even for various wards in the same clinic.

Extrapolating from these findings, we advise the following for ICU AI projects.

- For a truly effective implementation of AI systems, the ICU staff must be integrated in the design and implementation process, as well as possess adequate AI literacy. Encouraging and providing training to understand and use AI can empower ICU staff to embrace AI technologies confidently.
- As we need standards for alarm management workflows today, standardization is all the more essential in the context of integrating intelligent alarm management (AI-based) technologies. Significant workflow changes could be evoked and need to be considered by ICU leaders and implementation managers.

Limitations

In this study, the research focus lay on a single setting (ICU), which was suggested as a methodological approach by Wilken et al [5], allowing for an in-depth analysis of the contributing factors to an ICU's alarm management and dealing with the patient monitoring system. In addition to the stated benefits of this approach by the authors, it comes with limitations. On the one hand, the focus on a single ICU's cultural peculiarities restricts the transferability of the derived recommendations to other settings. The results and methods may serve, however, as inspiration and study protocol for evaluating an ICU's alarm management. By contrast, following a qualitative research approach, the number of shadowees and the number of shadowing days had to fit in a reasonable time schedule for the research conduction. The unit manager's suggestion of shadowees may have introduced a selection bias, despite being informed about the goal of diverse perspective shadowing. The results presented here are therefore only an excerpt of the reality in the ICU, which must be taken into account when interpreting them.

Conclusions

Our study highlights that interactions with the patient monitoring system and its alarms are a core part of tasks and workflows in the ICU. Alarm management tasks are performed based on ad hoc responses to clinical events; responsibilities are not well defined, and there is no standardized workflow or an SOP. Staff members were not satisfied with the current alarm management, which emphasizes a need for standard and clinician-centered guidelines in this field. Establishing SOPs for configuring and responding to alarms and considering local patient and workflow characteristics can streamline tasks and enhance the overall efficiency of care delivery. Systems that enable an intelligent alarm management to reduce alarm fatigue among staff members should be designed to make understandable and traceable suggestions, while health care professionals should be empowered to use them meaningfully through digital health literacy. By establishing these standards and thoughtfully incorporating AI into clinical workflows, health care institutions could enhance patient safety and relieve staff and patients from alarm-induced stress. To explore this effect on outcomes, more research in this field is needed.

Acknowledgments

The authors express their gratitude to the volunteers and the ward staff members. The project *Intelligent Alarm Optimizer for the Intensive Care Unit* (INALO) is funded by the German Federal Ministry of Education and Research under grant 16SV8559. We acknowledge financial support from the Open-Access Publication Fund of Charité Universitätsmedizin Berlin and the German Research Foundation (DFG). We used the generative AI tool, ChatGPT by OpenAI, after the manuscript was written to rephrase single (approximately 12) sentences, to improve readability. The original ChatGPT transcripts can be obtained upon request from the first author (LM).

ASP and ARF received funding by the German Federal Ministry of Education and Research under grant 16SV8559 as part of the project INALO. The authors acknowledge financial support from the Open-Access Publication Fund of Charité Universitätsmedizin Berlin and the German Research Foundation (DFG).

Data Availability

The data sets generated and analyzed during this study are not publicly available because of data privacy; however, they are available from the corresponding author (LM) upon reasonable request.

Authors' Contributions

The study was conceived by ASP, LM, FM, and MS. FM designed the ethnographic study. MS conducted the data acquisition, supported by LM and ASP and under the methodological supervision of FM. Data analysis was performed by MS, LM, and FM. LM, MS, and ARF wrote the manuscript. ASP, FB, and MF supervised all parts of the study. All authors critically reviewed and approved the manuscript.

Conflicts of Interest

ASP and ARF received funding by the German Federal Ministry of Education and Research under grant 16SV8559 as part of the project INALO.

Multimedia Appendix 1
Codes and subcodes derived during data analysis from field notes, interview documentation, and documentation of the participatory task.
[DOCX File , 31 KB - humanfactors_v11i1e55571_app1.docx]

Multimedia Appendix 2
Indicated activities of a physician and their perceived proportion of the workday adapted from the results of the participatory task performed by a physician shadowee. The tasks with monitoring influence are indicated with the monitor icon.
[PNG File , 173 KB - humanfactors_v11i1e55571_app2.png]

References

1. Jones K. Alarm fatigue a top patient safety hazard. CMAJ 2014 Feb 18;186(3):178 [FREE Full text] [doi: 10.1503/cmaj.109-4696] [Medline: 24418978]
2. Moreno RP, Rhodes A, Donchin Y, European Society of Intensive Care. Patient safety in intensive care medicine: the Declaration of Vienna. Intensive Care Med 2009 Oct;35(10):1667-1672 [FREE Full text] [doi: 10.1007/s00134-009-1621-2] [Medline: 19697007]
3. Honan L, Funk M, Maynard M, Fahs D, Clark JT, David Y. Nurses' perspectives on clinical alarms. Am J Crit Care 2015 Sep;24(5):387-395. [doi: 10.4037/ajcc2015552] [Medline: 26330431]
4. Poncette AS, Mosch L, Spies C, Schmieding M, Schiefenhövel F, Krampe H, et al. Improvements in patient monitoring in the intensive care unit: survey study. J Med Internet Res 2020 Jun 19;22(6):e19091 [FREE Full text] [doi: 10.2196/19091] [Medline: 32459655]

5. Wilken M, Hüske-Kraus D, Klausen A, Koch C, Schlauch W, Röhrig R. Alarm fatigue: causes and effects. In: Proceedings of the 62nd Annual Meeting of the German Association of Medical Informatics, Biometry and Epidemiology. 2017 Presented at: GMDS '17; August 29, 2017; Oldenburg, Germany p. 107-111 URL: https://ebooks.iospress.nl/publication/47517 [doi: 10.3233/978-1-61499-808-2-107]
6. Sendelbach S. Alarm fatigue. Nurs Clin North Am 2012 Sep;47(3):375-382. [doi: 10.1016/j.cnur.2012.05.009] [Medline: 22920428]
7. Despins LA. Factors influencing when intensive care unit nurses go to the bedside to investigate patient related alarms: a descriptive qualitative study. Intensive Crit Care Nurs 2017 Dec;43:101-107. [doi: 10.1016/j.iccn.2017.04.003] [Medline: 28552259]
8. Paine CW, Goel VV, Ely E, Stave CD, Stemler S, Zander M, et al. Systematic review of physiologic monitor alarm characteristics and pragmatic interventions to reduce alarm frequency. J Hosp Med 2016 Feb;11(2):136-144 [FREE Full text] [doi: 10.1002/jhm.2520] [Medline: 26663904]
9. Joint Commission. Medical device alarm safety in hospitals. Sentinel Event Alert 2013 Apr 08(50):1-3. [Medline: 23767076]
10. Deb S, Claudio D. Alarm fatigue and its influence on staff performance. IIE Trans Healthc Syst Eng 2015 Jul 29;5(3):183-196. [doi: 10.1080/19488300.2015.1062065]
11. Hüske-Kraus D, Wilken M, Röhrig R. Measuring alarm system quality in intensive care units (work in progress). EasyChair. Preprint posted online May 6, 2018 [FREE Full text] [doi: 10.29007/5vpj]
12. Bach TA, Berglund LM, Turk E. Managing alarm systems for quality and safety in the hospital setting. BMJ Open Qual 2018 Jul 25;7(3):e000202 [FREE Full text] [doi: 10.1136/bmjoq-2017-000202] [Medline: 30094341]
13. Cosper P, Zellinger M, Enebo A, Jacques S, Razzano L, Flack MN. Improving clinical alarm management: guidance and strategies. Biomed Instrum Technol 2017;51(2):109-115. [doi: 10.2345/0899-8205-51.2.109] [Medline: 28296432]
14. Balzer F, Agha-Mir-Salim L, Ziemert N, Schmieding M, Mosch L, Prendke M, et al. Staff perspectives on the influence of patient characteristics on alarm management in the intensive care unit: a cross-sectional survey study. BMC Health Serv Res 2023 Jul 05;23(1):729 [FREE Full text] [doi: 10.1186/s12913-023-09688-x] [Medline: 37407989]
15. Huo J, Wung S, Roveda J, Li A. Reducing false alarms in intensive care units: a scoping review. Explor Res Hypothesis Med 2022 Jun 2;8(1):57-64. [doi: 10.14218/erhm.2022.00026]
16. Chromik J, Klopfenstein SA, Pfitzner B, Sinno ZC, Arnrich B, Balzer F, et al. Computational approaches to alleviate alarm fatigue in intensive care medicine: a systematic literature review. Front Digit Health 2022;4:843747 [FREE Full text] [doi: 10.3389/fdgth.2022.843747] [Medline: 36052315]
17. INALO: Intelligenter Alarmoptimierer für die Intensivstation. Miteinander durch Innovation. URL: https://www.interaktive-technologien.de/projekte/inalo [accessed 2023-12-04]
18. Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, Matthieu MM, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. Implement Sci 2015 Feb 12;10(1):21 [FREE Full text] [doi: 10.1186/s13012-015-0209-1] [Medline: 25889199]
19. Mosch LK, Poncette AS, Spies C, Weber-Carstens S, Schieler M, Krampe H, et al. Creation of an evidence-based implementation framework for digital health technology in the intensive care unit: qualitative study. JMIR Form Res 2022 Apr 08;6(4):e22866 [FREE Full text] [doi: 10.2196/22866] [Medline: 35394445]
20. Mörike F. Inverted hierarchies on the shop floor: the organisational layer of workarounds for collaboration in the metal industry. Comput Supp Coop Work 2021 Dec 02;31(1):111-147. [doi: 10.1007/S10606-021-09415-2]
21. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
22. Flick U. Triangulation in der Ethnographie. In: Flick U, editor. Triangulation: Eine Einführung. Cham, Switzerland: Springer; 2011:51-74.
23. Patton MQ. Qualitative Research & Evaluation Methods: Integrating Theory and Practice. Thousand Oaks, CA: SAGE Publications; 2014.
24. Vocke M, Mörike F. E-Mails and teamrooms: immaterielle arbeit und digitale kommunikationstechnologien bei wissensarbeiter*innen in multinationalen unternehmen. In: Proceedings of the 66th Congress of the Society for Ergonomics on Digital Work, Digital Change, Digital People. 2020 Presented at: DWCP '20; March 16-18, 2020; Berlin, Germany URL: https://www.researchgate.net/publication/340005897_Chats_E-Mails_Teamrooms_Immaterielle_Arbeit_und_digitale_Kommunikationstechnologien_bei_Wissensarbeiterinnen_in_multinationalen_Unternehmen
25. Kankam PK. The use of paradigms in information research. Libr Inform Sci Res 2019 Apr;41(2):85-92. [doi: 10.1016/j.lisr.2019.04.003]
26. Dourish P. Implications for design. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2006 Presented at: CHI '06; April 22-27, 2006; Montréal, QC p. 541-550 URL: https://dl.acm.org/doi/10.1145/1124772.1124855 [doi: 10.1145/1124772.1124855]
27. Hammersley M. Reading Ethnographic Research. New York, NY: Routledge; 2016.

28. Poncette AS, Spies C, Mosch L, Schieler M, Weber-Carstens S, Krampe H, et al. Clinical requirements of future patient monitoring in the intensive care unit: qualitative study. JMIR Med Inform 2019 Apr 30;7(2):e13064 [FREE Full text] [doi: 10.2196/13064] [Medline: 31038467]
29. COPRA Patient Data Management Systems (PDMS). COPRA System GmbH. URL: https://copra-system.de [accessed 2023-12-16]
30. i.s.h.med: das universelle, in die SAP-welt integrierte KIS. Oracle Cerner. URL: https://www.cerner.com/de/de/loesungen/ishmed [accessed 2023-12-16]
31. Home page. INALO Intelligenter Alarmoptimierer für die Intensivstation. URL: https://www.inalo.ai [accessed 2023-02-25]
32. Czarniawska-Joerges B. Shadowing: And Other Techniques for Doing Fieldwork in Modern Societies. Copenhagen, Denmark: Copenhagen Business School Press; 2012.
33. Mörike F, Spiehl HL, Feufel MA. Workarounds in the shadow system: an ethnographic study of requirements for documentation and cooperation in a clinical advisory center. Hum Factors 2024 Mar;66(3):636-646. [doi: 10.1177/00187208221087013] [Medline: 35471848]
34. Leslie M, Paradis E, Gropper MA, Kitto S, Reeves S, Pronovost P. An ethnographic study of health information technology use in three intensive care units. Health Serv Res 2017 Aug 25;52(4):1330-1348 [FREE Full text] [doi: 10.1111/1475-6773.12466] [Medline: 28124443]
35. Smith A, Goodwin D, Mort M, Pope C. Expertise in practice: an ethnographic study exploring acquisition and use of knowledge in anaesthesia. Br J Anaesth 2003 Sep;91(3):319-328 [FREE Full text] [doi: 10.1093/bja/aeg180] [Medline: 12925468]
36. Schmidt C, Mörike F, Feufel MA. Ein digitaler Dritter im Ärzt*in-Patient*in-Gespräch: wie ein online-gestütztes tool beratungen standardisiert und personalisiert. Curare 2021;45(2022):77-100 [FREE Full text]
37. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
38. Fast E, Horvitz E. Long-term trends in the public perception of artificial intelligence. Proc AAAI Conf Artif Intell 2017 Feb 12;31(1):10635. [doi: 10.1609/aaai.v31i1.10635]
39. Liehner G, Hick A, Biermann H, Brauner P, Ziefle M. Perceptions, attitudes and trust toward artificial intelligence — an assessment of the public opinion. In: Ahram T, Kalra J, Karwowski W, editors. Artificial Intelligence and Social Computing. Nice, France: AHFE Publications; 2023.
40. Chuan CH, Tsai WH, Cho SY. Framing artificial intelligence in American newspapers. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. 2019 Presented at: AIES '19; January 27-28, 2019; Honolulu, HI p. 339-344 URL: https://dl.acm.org/doi/10.1145/3306618.3314285 [doi: 10.1145/3306618.3314285]
41. Carter N, Bryant-Lukosius D, DiCenso A, Blythe J, Neville AJ. The use of triangulation in qualitative research. Oncol Nurs Forum 2014 Sep;41(5):545-547. [doi: 10.1188/14.ONF.545-547] [Medline: 25158659]
42. Emerson RM, Fretz RI, Shaw LL. Writing Ethnographic Fieldnotes. 2nd edition. Chicago, IL: University of Chicago Press; 2011.
43. MAXQDA 2022. VERBI Software. URL: https://www.maxqda.com/about [accessed 2024-04-29]
44. Selting M, Auer P, Barth-Weingarten D, Bergmann JR, Bergmann P, Birkner K, et al. Gesprächsanalytisches transkriptionssystem 2 (GAT 2). Gesprachsforschung 2009:1617-1837 [FREE Full text]
45. Strauss A, Corbin JM. Basics of Qualitative Research: Grounded Theory Procedures and Techniques. Thousand Oaks, CA: Sage Publications; 1990:978.
46. Birks M, Chapman Y, Francis K. Memoing in qualitative research: probing data and processes. J Res Nurs 2008 Jan 01;13(1):68-75. [doi: 10.1177/1744987107081254]
47. Blandford A, Gibbs J, Newhouse N, Perski O, Singh A, Murray E. Seven lessons for interdisciplinary research on interactive digital health interventions. Digit Health 2018;4:2055207618770325 [FREE Full text] [doi: 10.1177/2055207618770325] [Medline: 29942629]
48. Poncette AS, Wunderlich MM, Spies C, Heeren P, Vorderwülbecke G, Salgado E, et al. Patient monitoring alarms in an intensive care unit: observational study with do-it-yourself instructions. J Med Internet Res 2021 May 28;23(5):e26494 [FREE Full text] [doi: 10.2196/26494] [Medline: 34047701]
49. Albanowski K, Burdick KJ, Bonafide CP, Kleinpell R, Schlesinger JJ. Ten years later, alarm fatigue is still a safety concern. AACN Adv Crit Care 2023 Sep 15;34(3):189-197. [doi: 10.4037/aacnacc2023662] [Medline: 37644627]
50. Managing alarms in acute care across the life span: electrocardiography and pulse oximetry. American Association of Critical Care Nurses. URL: https://www.aacn.org/clinical-resources/practice-alerts/managing-alarms-in-acute-care-across-the-life-span [accessed 2024-01-14]
51. Sendelbach S, Funk M. Alarm fatigue: a patient safety concern. AACN Adv Crit Care 2013;24(4):378-387. [doi: 10.1097/NCI.0b013e3182a903f9] [Medline: 24153215]
52. Oliveira AE, Machado AB, Santos ED, Almeida É. Alarm fatigue and the implications for patient safety. Rev Bras Enferm 2018;71(6):3035-3040 [FREE Full text] [doi: 10.1590/0034-7167-2017-0481] [Medline: 30517409]

53. Sinno ZC, Shay D, Kruppa J, Klopfenstein SA, Giesa N, Flint AR, et al. The influence of patient characteristics on the alarm rate in intensive care units: a retrospective cohort study. Sci Rep 2022 Dec 16;12(1):21801 [FREE Full text] [doi: 10.1038/s41598-022-26261-4] [Medline: 36526892]
54. Winters BD, Cvach MM, Bonafide CP, Hu X, Konkani A, O'Connor MF, Society for Critical Care Medicine Alarm and Alert Fatigue Task Force. Technological distractions (part 2): a summary of approaches to manage clinical alarms with intent to reduce alarm fatigue. Crit Care Med 2018 Jan;46(1):130-137. [doi: 10.1097/CCM.0000000000002803] [Medline: 29112077]
55. Turmell JW, Coke L, Catinella R, Hosford T, Majeski A. Alarm fatigue: use of an evidence-based alarm management strategy. J Nurs Care Qual 2017;32(1):47-54. [doi: 10.1097/NCQ.0000000000000223] [Medline: 27500694]

Abbreviations

AI: artificial intelligence
ICU: intensive care unit
INALO: Intelligent Alarm Optimizer for the Intensive Care Unit
PDMS: patient data management system
SOP: standard operating procedure

Edited by C Jacob; submitted 17.12.23; peer-reviewed by J Schlesinger, C Lehet; comments to author 10.01.24; revised version received 24.01.24; accepted 08.04.24; published 18.06.24.

Please cite as:
Mosch L, Sümer M, Flint AR, Feufel M, Balzer F, Mörike F, Poncette AS
Alarm Management in Intensive Care: Qualitative Triangulation Study
JMIR Hum Factors 2024;11:e55571
URL: https://humanfactors.jmir.org/2024/1/e55571
doi:10.2196/55571
PMID:38888941

Original Paper

Probing the Role of Digital Payment Solutions in Gambling Behavior: Preliminary Results From an Exploratory Focus Group Session With Problem Gamblers

Nathan Lakew[1], PhD; Jakob Jonsson[1], PhD; Philip Lindner[1,2], PhD

[1]Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

[2]Center for Dependency Disorders, Stockholm Health Care Services, Stockholm, Sweden

Corresponding Author:
Nathan Lakew, PhD
Department of Clinical Neuroscience
Karolinska Institutet
Norra Stationsgatan 69
Plan 7
Stockholm, 113 64
Sweden
Phone: 46 852483391
Email: nathan.lakew@ki.se

Abstract

Background: Technology has significantly reshaped the landscape and accessibility of gambling, creating uncharted territory for researchers and policy makers involved in the responsible gambling (RG) agenda. Digital payment solutions (DPS) are the latest addition of technology-based services in gambling and are now prominently used for deposit and win withdrawal. The seamless collaboration between online gambling operators and DPS, however, has raised concerns regarding the potential role of DPS platforms in facilitating harmful behavior.

Objective: Using a focus group session with problem gamblers, this study describes a preliminary investigation of the role of DPS in the online gambling context and its influence on players' gambling habits, financial behavior, choices of gambling environment, and the overall outcome of gambling subjective experiences.

Methods: A total of 6 problem gamblers participated in a one-and-half-hour focus group session to discuss how DPSs are integrated into their everyday gambling habits, what motivates them to use DPS, and what shifts they observe in their gambling behavior. Thematic analysis was used to analyze the empirical evidence with a mix of inductive and deductive research approaches as a knowledge claim strategy.

Results: Our initial findings revealed that the influence of DPSs in online gambling is multifaced where, on the one hand, their ability to integrate with players' existing habits seamlessly underscores the facilitating role they play in potentially maximizing harm. On the other hand, we find preliminary evidence that DPSs can have a direct influence on gambling outcomes in both subtle and pervasive ways—nudging, institutionalizing, constraining, or triggering players' gambling activities. This study also highlights the increasingly interdisciplinary nature of online gambling, and it proposes a preliminary conceptual framework to illustrate the sociotechnical interplay between DPS and gambling habits that ultimately capture the outcome of gambling's subjective experience.

Conclusions: Disguised as a passive payment enabler, the role of DPS has so far received scant attention; however, this exploratory qualitative study demonstrates that given the technological advantage and access to customer financial data, DPS can become a potent platform to enable and at times trigger harmful gambling. In addition, DPS's bird's-eye view of cross-operator gambling behavior can open up an opportunity for researchers and policy makers to explore harm reduction measures that can be implemented at the digital payment level for gambling customers. Finally, more interdisciplinary studies are needed to formulate the sociotechnical nature of online gambling and holistic harm minimization strategy.

(JMIR Hum Factors 2024;11:e54951) doi:10.2196/54951

KEYWORDS
digital payment solutions; online gambling behavior; sociotechnical; subjective experience; focus group

Introduction

As gambling became a readily available digital commodity, examining and regulating the influence of technology on gambling behavior has emerged as an important agenda for harm reduction research [1]. In the early days of the internet, Shaffer [2] noted that the innate nature of the "new" technology has the potential to become a potent vehicle to introduce certain biases in gambling, resulting in a shift in the subjective experiences of gamblers. Almost 30 years later, and with online gambling becoming commonplace, the extant literature has seemingly reached a consensus that technology-enabled gambling platforms with engaging and persuasive features can facilitate attention bias, impulsive betting, or longer time-on-devices, which can ultimately trigger or maintain addictive behavior [3].

The online gambling landscape has also created a demand for more enabling technologies, such as digital payment solutions (DPSs), for gambling activities. Under the umbrella of a general term known as financial technology, DPSs come in different forms, including digital wallets, payment gateways, or processors that can easily be integrated with online gambling sites or points of sale. Historically, the transfer and funding process of gambling has been perceived as cumbersome practicalities with unacceptable downtime in the eyes of gambling operators [4]. The fast adoption of DPS among the general population in other consumer domains (eg, digital shopping), its convenient account-to-account (direct bank) solution, and merchant-centric flexibility have quickly attracted gambling operators to integrate it into their digital platforms.

DPSs are seen as having a multifaceted advantage for the gambling industry [5]. First, DPS's fast and seamless deposit and withdrawal solution facilitates longer gambling periods (ie, time-on-device) while solving the problem of downtime [4]. Second, DPS connects multimodal payment options to a one-click account [6]. Depending on local jurisdictions, a user can, for example, connect different payment modes, such as credit or invoice or several bank accounts, to their DPS account, to achieve multiple funding paths. Third, DPSs enable a unison provision of multiple services in one speedy process. For example, gambling operators can use DPS providers to both process payments and conduct Know-Your-Customer responsibilities of authenticating users [7], hence reducing the time and inconveniences associated with starting gambling. With the removal of lengthy registration and authentication processes, options, such as pay-and-play and no-account casinos, are emerging as the new standard of online gambling [8]. Other factors, including general familiarity with DPS in other e-commerce settings in past purchases and trust, can play a role in easy implementation, as well as the adoption of DPS in online gambling [9].

Apart from the technical aspect, DPS providers have benefited from recent policy relaxation of financial regulations, such as open banking [10]. Consequently, digital payment companies have the advantage of both agility and unprecedented access to financial data to influence consumer behavior (provided users' consent). The nature of open banking allows more transparency and access to individual payment behavior; however, there is currently no strong incentive for DPS providers with gambling customers to be involved in the effort to promote responsible gambling (RG). Traditional policies and regulatory bodies that govern financial institutions, including DPS providers, mainly focus on ensuring that financial legal requirements are met and that a "well-functioning" marketplace is viable [11]. Financial decisions and well-being are broadly seen as individual responsibilities [12]. In addition, gambling policy making has historically focused on either RG measures implemented at gambling operator sites or consumers' responsible behavior agendas.

Similar to other technological products, digital payments have been shown to influence subjective gambling experiences, even promoting harmful gambling. For instance, DPS enables an account-based registry in which consumers can connect different sources of funds (eg, multiple bank accounts) into a one-click account, which creates (an illusion of) "uninterrupted availability" of funds. Their fast-deposit mechanism and auto-deposit functionalities can result in lowering the awareness of money spent, leading to prolonged gambling consumption, and dark gambling flow—a state of intense immersion experienced by gamblers leading to a loss of awareness of money and time spent in gambling [13,14]. Previous studies have also theoretically explored DPS in the context of gambling, where payment technologies can be used to augment harm reduction measures [13].

Although the significance of DPS in gambling is beginning to gain attention [5,15-17], we are unaware of any past research that has examined players' perspective of digital payment in an online gambling setting. In addition, and perhaps due to its interdisciplinary nature, there exists no conceptualization of the mechanism through which technology-enabled payment platforms can shift players' subjective experiences toward harmful gambling behavior. As the current harm-minimization efforts in online gambling are heavily dependent on understanding, predicting, and promoting healthy gambling behavior, the research field needs to have a holistic understanding of how these tools are integrated into players' everyday gambling experience and possibly facilitate harmful behavior.

This study aims to supplement the research area by conducting explanatory qualitative research involving first-hand accounts of DPS use among problem gamblers. Exploratory qualitative methods provide a practical approach for laying a strong foundation to further investigate emerging research areas [18]. In addition, qualitative methods, in general, are effective in understanding complex social phenomena of interest and human stories that involve probing participants regarding their experiences with open-ended questions [19]. In this study, the following explorative research questions were examined:

- How are DPSs used in gambling activities among problem gamblers, and what do they think about the influence of DPS on their gambling behavior?
- How can we conceptualize the role of DPS on the outcome of gambling behavior?

Methods

Ethical Considerations

The Swedish Ethics Authority approved this qualitative focus group study (DNR 2022-05843-01). All focus group participants gave informed consent before joining the sessions, and pseudonyms were used to protect their identities during the analysis. Participants were not compensated for their involvement in the study.

Design and Data Collection

The research uses an exploratory qualitative design approach [20]. The approach provides methodological and theoretical flexibility to investigate topical research domains, aiming to identify important concepts and potential conceptual constructs [21,22]. Qualitative exploratory inquiries typically entail gathering primary data using techniques, such as interviews and focus groups, followed by a grounded theory approach analysis, occasionally leveraging on deductive insights from existing literature [23]. In this study, a focus group session serves as the primary data collection method. Following an explanatory research approach, the questions posed during the session were semistructured yet exploratory, providing participants with an opportunity to share their experiences freely.

Current studies on the adverse impacts of gambling indicate that harmful consequences, including financial ones, are prevalent in the problem gambling (PG) community [24]. Our study participants were selected from this group, allowing us to observe a wide range of harmful gambling behaviors. In addition, the homogeneity of the group has provided an environment where all participants felt comfortable to freely express their past experiences and views. However, it is important to note that individuals who do not have a gambling problem are not necessarily immune to the impact of DPS.

A total of 6 participants with a history of PG were recruited for a focus group interview. The participants were recruited through the Swedish National Association for Gambling Addicts—a national support group for problem gamblers. With the help of the support group administration, a recruitment leaflet was prepared and distributed during support group sessions. The leaflet outlined the goals of the focus group discussion and criteria for participating in the session: (1) experience in online gambling using 1 or more DPS, (2) having played in land-based gambling avenues using nondigital payment methods, and (3) being able to speak English. The participants accepted and returned the consent statement to the support group administration before the focus group session.

Participants were asked to complete brief questionnaires regarding demographics and gambling history before the focus group session started (Multimedia Appendix 1). The most common gambling modality reported was online gambling, including casinos and sports betting, with excitement chasing ranked as the main reason for gambling. The participants' average gambling experience was more than 10 years. All participants were engaged in a biweekly support session. Table 1 provides supplementary profile details; however, following the ethical approval obtained and the sensitive nature of the research, we are only able to provide aggregate information regarding the participants.

Table 1. Participant characteristics.

Variables and options	Participants, n (%)
Age (years)	
26-34	1 (17)
35-50	3 (50)
51-70	2 (33)
Sex	
Male	5 (83)
Female	1 (17)
Marital status	
Married	2 (33)
In relationship	2 (33)
Separated	1 (17)
Single	1 (17)
Gambling type preference	
Land-based gambling	0 (0)
Online gambling	4 (67)
Both	2 (33)
Experience of problem gambling (years)	
3-5	2 (33)
7-10	1 (17)
10+	3 (50)

Procedure

The focus group discussion was conducted at the premises of the association after one of the support group sessions was completed. The focus group study format enables researchers to collect empirical evidence from multiple sources simultaneously while providing a sense of cohesiveness and a safe environment for participants to share their opinions [25]. In addition, live interaction between participants who share the same background (here, problem gamblers) provided an opportunity for honest and personal story discussions about their struggles and reflection on each other's experiences. The first author (NL) led the focus group discussion, while the second author (JJ) played a supporting role during the meeting, including time management and follow-up questions. Additionally, the second author (JJ) participated as a professional psychologist with experience in gambling therapy to ensure safety in case the group discussions triggered any distressing memories among the participants.

The one-and-a-half-hour session started with an introduction of the aim of the focus group and the procedure. Participants were informed that they had the right to leave the session at any given time, refrain from responding to questions, and provide other relevant practical information. A semistructured set of questions was prepared in advance to conduct the discussion, and this material was used to organize the discussion around DPS's influence on financial behavior, user gambling behavior, time spent on a device, gambling environment, and segment choices. The focus group protocol can be found in Multimedia Appendix 2.

Analysis of Empirical Evidence

The focus group session was recorded and transcribed verbatim by NL. Thematic content analysis was then applied to analyze the material [26]. The analysis was conducted to develop a "ground-up" understanding of problem gamblers' experience with DPS, which was later used to conceptualize the influence of DPS on the subjective experience of gambling. As such, thematic coding was performed using a grounded theory framework starting with "open coding" or manifest data analysis [27]. The transcribed focus group session was coded using QualCoder 3.2 (MIT-based open-source qualitative data analysis tool), with each participant assigned a pseudonym of P1-P6. The first 2 authors read and reread the empirical material for conversational-level open coding [28]. NL used the first coding, which was later reviewed by JJ for interpretation and coding validity. The manifest analysis followed an inductive approach where efforts were made to code block the material based on "first order value data," which is what is obviously observable in the transcription [29]. With iterative coding and merging of codes at hand, a further validity review of codes was performed by JJ.

After organizing the merged codes into meaningful subthemes, potential themes emerged to categorize the subthemes. At the stage of building subthemes and themes, a deductive research approach was comparatively used as a lens to supplement theme building using existing relevant theoretical concepts in gambling and sociotechnical research, such as socio-materiality, imbrication, and reciprocal interaction theoretical frameworks [30-32]. The final themes and subthemes were then discussed by all 3 authors (NL, JJ, and PL) with the original text of transcription used as a confirmation of concept summary and naming before the final reporting.

Results

Overview

Three main themes emerged from the explorative analysis: (1) existing addictive needs dictate DPS's "placement" in gambling, (2) DPS changes gambling subjective experience, and (3) problem gamblers' perspectives on DPS. The first theme focuses on players' way of integrating DPS into their gambling activities, covering three subthemes, each developing players' appropriation of DPS into their gambling context to (1) workaround RG measures, (2) fit into gambling habit, and (3) shorter time to start and get a reward. With four subthemes, the second theme sheds light on DPS's influence on gambling habits in the form of (1) subtle nudge on behavior, (2) opens up new gambling habits, (3) introduce bias, and (4) facilitate intangible qualities, such as trust to sanction different forms of illegal gambling.

Additionally, we find three subthemes with a nature of reflective content summarized in the third theme where problem gamblers react to DPS: (1) DPS is not for us, (2) pessimistic about RG tools in DPS, and (3) recommendations and opinions where the DPS role is conceived as having a minimal role in harmful gambling. Table 2 presents the analysis summary, and a more detailed theme development process can be found in Multimedia Appendix 3.

Table 2. Summary of themes and subthemes.

Main themes and subthemes	Summary of subthemes
Existing addictive needs dictate DPS's[a] "placement" in gambling	
Workaround RG[b] measures	• DPS can be used to bypass RG tools. • It can also enable players to maximize gambling "perks."
Fit into gambling habit	• DPS fits into a habit of a gentle start of the day with gambling. • Gambling habits direct both the choices and ways of DPS use.
Shorter time to start and get a reward	• DPS is used to make the "chores" of depositing "invisible." • DPS's speedy withdrawal function is seen as a reward not just in terms of earning a prize but also for continuing to play.
DPS changes gambling subjective experience	
Subtle nudge toward harmful behavior	• DPS can potentially facilitate dark flow in gambling. • Subtle nudging toward longer "time on device" and harmful games, such as casinos.
Opens up new subjective experience	• DPS enables one to develop new habits and ways of being a gambler. • Rewards do not feel like wins anymore as they just represent numbers.
Introduces bias	• DPS can introduce multiple biases, such as the illusion of control or the availability of unlimited funds. • DPS can create a desensitizing feeling toward the value of money and gambling losses.
Facilitate intangible attributes	• DPS can facilitate confirmation bias and increase trust in unregulated gambling sites.
Problem gamblers' perspectives on DPS	
DPS is not for "us"	• The integration of DPS into the gambling context needs scrutiny. • The payment tool is used in gambling with an advantage for the industry in mind.
Pessimistic about RG tool in DPS	• The effect of DPS on gambling behavior is minimal. • Integrating "passive" RG tools, such as red warning with DPS, will not help problem gamblers.
Recommendations	• Wins withdraw-focused RG measures in DPS can be more effective. • DPS should be banned from operating in unregulated gambling sites.

[a]DPS: digital payment solution.

[b]RG: responsible gambling.

Existing Addictive Needs Dictate DPS Placement in Gambling

Workaround RG Measures

The participants noted that DPS provides opportunities to engage in gambling behaviors that would not be easily possible otherwise. In some instances, for example, participants reported that access to gambling was intact even though their names were included in the Swedish national self-exclusion registry because DPS enabled them to access unlicensed gambling operators. One participant described the use of DPS as a "negotiating system," effectively offsetting the goal of RG tools implemented nationwide:

> *The bigger factor for players to go play on unlicensed (sites) is payment solutions. They (players) are in Spelpaus (self-exclusion registry). Therefore, they negotiate the ban with (a) payment solution in Sweden.* [Participant 1]

In other instances, some participants describe using DPS to circumvent specific RG features aimed at preventing harmful gambling, such as bonus offers and free spins on online casinos. Consequently, most participants recognized how DPS can constitute a ready-to-hand tool to overcome the "challenges" they face in satisfying gambling cravings and maximizing rewards.

> *Swedish casino is super easy, but I get (a) bonus in the unlicensed sites. Sometime(s) also find it cheaper to deposit in the unlicensed ones. I used the payment solution to do that. They (unlicensed sites) also offer free spins.* [Participant 4]

Fit Into Existing Gambling Habits

We also found players to appropriate DSP in line with their usual gambling habits. As they became comfortable with different payment solutions, they picked "favorites" that seamlessly fit into their gambling style. For example, 1 participant noted that his choice of a specific DPS—a popular Swedish mobile payment system—allowed him to start the day the way he "wanted it":

> *I always start with the Swish [a mobile payment solution] on casinos, and when I reach my limits, I move on to the ones that I have my card saved and so on. For me, the easiest one to start the day with and then just push it forward.* [Participant 4]

Furthermore, the participant also mentioned that his choice of DPS has enabled him to make a smooth transition to other casino games after a win without worrying about lengthy fund transfer processes.

> *I want to change casino(s) because I've been lucky for a while and then I need to try the other sixteen so just moving around the money. Swish is the best one because if you played with the Swish, you could have the money quite quick(ly) back and move it forward.* [Participant 4]

The same sentiment was shared by other participants: different payment solutions were seen to satisfy different purposes and hence apt for different gambling routines.

Shorter Time to Start and Get a Reward

The most prominent reason for integrating DPS noted among all participants was related to the speed and "invisibility" a DPS affords in transferring funds during both deposit and win withdrawal. Most participants noted that they always wanted to quickly start gambling and perceived the process of depositing funds as a cumbersome chore to deal with before the fun started. DPSs that are already integrated with operators and commonly used among the players for other purposes were considered the best to make depositing "task invisible." One participant noted that he chose a payment solution that makes betting and deposit almost synonymous:

> *When I'm thinking about payment solutions, for me I never used a solution where you had to make an account on one and then deposit. Everything should be direct. I never used the systems that were not directly connected to the account. Because it was too much of a hassle and then I had to see the money two times I would say (laugh). Yeah, but yeah, I had to see the money going in one then going away to another account. I didn't want that.* [Participant 5]

Participants saw a speedy win withdrawal not just as receiving a prize but mainly as a resource to continue gambling. As such, win withdrawals may stay in the DPS account for fast redepositing. In addition, the need for speedy withdrawal seemed to correlate with the progressive worsening of PG.

> *The withdrawal time is important, especially when you need the money. When I started gambling, it was OK for me to wait for a couple of days. But when I got more and more addicted and intense, it became very important for me then. I remember I used to select Poker sites on stuff like how they looked; but that all went away once I got addicted, and payment solution and withdrawal time became important factors for me.* [Participant 1]

DPS Changes Gambling Subjective Experiences

Subtle Nudges Toward Harmful Behavior

Most participants noted that DPS seamlessly fit into, and contributed to, the development of an uninterrupted "fictional world of gambling," potentially affecting one's financial well-being. Recounting such an experience, one participant described a dark flow of gambling where even getting a payday loan amid intense gambling activity was perceived as winning a jackpot:

> *You keep swishing; it is a fantasy world. When you are out, you (use) SMS loan money, it takes only 5 minutes. It's as if you got a jackpot in (a) casino. It is the same feeling. It is a success as if you gamble and win.* [Participant 5]

In the same vein, DPS has been shown to subtly facilitate longer time-on-device by enabling operators' actions, such as a suggestion to refill deposit accounts via a DPS one-click link and a sign-off. At times, refill amounts were suggested.

> *When money is almost gone, not completely gone almost gone always pops up "OK do you want to refill" ... that's bad because then it's like a quick link ... When you're in a game and then it's just OK, it feels quick and it's like one press. That's really stupid.* [Participant 3]

As noted in the first theme in passing, the speedy solution among payment methods was revealed to subtly nudge players toward harmful types of games. In many instances, participants were heard associating DPS with casino games where at times, the payment method is used as a reason for choosing gaming platforms.

> *You always choose those casinos that are easy to deposit. I used quick casinos to put small amounts of money many times ... I use the easiest, Swish, to deposit.* [Participant 4]

Open Up New Subjective Experience

In addition to subtle behavioral nudging, DPS can enable a new subjective experience of gambling behavior. Some participants reported that they developed new habits, such as gambling at workplaces or playing while driving as the adoption of DPS became widespread.

> *I started to play a lot more at (the) workplace when Swish came into (the) picture, and it was so easy. I have what you call blue-collar work that I must work (with) my hands, but I've seen when it was so easy you managed to work and play.* [Participant 6]

> *After 6 hours I could drive 300 km, and I use the easiest one; Swish to put the deposit while driving. One time I was about (to) crash: I panicked.* [Participant 4]

Introduces Bias

We also found some evidence of DPS's nature in introducing multiple biases in gambling activities due to its technological makeup. In line with past research, the most obvious observation was the effect of digitalization on diminishing the value of money [33]. Some participants have also discussed how DPS enables them to connect different sources of funds (eg, different bank accounts, cards, or at times, invoices) into 1 DPS account to select deposit sources as they see fit. In addition to increasing money spent on gambling, such a setting created the illusion of unlimited funds during gambling activities.

> *They're very easy because when I played, I had a few cards here which are used and they just picked up one of them, and (that) is OK.* [Participant 4]

> *You lose all the picture and it's just a figure. You have money then out of money and done.* [Participant 5]

DPS was also shown to preface the illusion of control by enabling both a "controlled" amount of spending and frictionless payment. One participant noted that since his choice of DPS asks him to sign off each deposit amount, it felt like it was controlled spending. However, since the flow of signing off was frictionless and almost invisible, in practice, it was not so controlled.

> *For me, I used quick casinos to put in small amounts of money many times to not feel bad and lie to myself. I put 500 SEK again and again, but how many times? (laugh).* [Participant 4]

Facilitate Intangible Attributes

Finally, participants were heard discussing how DPS can influence their choice of games, gambling segments, and environment (ie, licensed vs unlicensed operators) by creating confidence and trust in the site where they are being integrated. The trust relegating effect was stronger for DPS, which is also used daily for other digital purchases.

> *It mattered to me if I knew the payment solution to play in the unlicensed sites, and it is a factor in selecting unlicensed sites.* [Participant 2]

Problem Gamblers' Perspectives on DPS

DPS Is Not for Us

The majority agreed that DPSs have negatively influenced their gambling behavior. There was a general sentiment that DPS are being swiftly integrated into gambling platforms without sufficient regard for the challenges they present to problem gamblers.

> *I think it's scary that the Swish doesn't have some red button. No one asks you if you do it again and again.* [Participant 4]

> *Yeah, or at least red flags ... The system wasn't designed for us. It was designed for the general population.* [Participant 1]

Another reaction from the participants emphasizes that the DPS industry is taking advantage of the transition to a cashless society, which in general, gives an advantage to the financial industry.

Pessimistic About RG Tools in DPS

Some participants expressed pessimism about whether DPS-specific RG tools would even have helped them. More specifically, RG measures that are more reactive to players' behavior, such as warnings or even bans based on gambling behavior, would not have helped. They noted that as long as the resource exists to spend on gambling, such measures can easily be evaded.

> *For me, the payment solution didn't (matter), I mean, I would have gambled. It doesn't matter if it took five minutes or ten, I would still have done it ... For the people who do it on a smaller scale then maybe it's a little bit different. For me, it didn't matter. If I just can get money in and out; that is the way, it is.* [Participant 5]

> *I'm just thinking that a limit or ban can be good in some ways in Fintech, but I also think that for me I wouldn't stop there if there was a red button or something like that it wouldn't have helped me.* [Participant 6]

Recommendations

Toward the end of the session, participants made a few proposals on how to make DPS a more RG-oriented product. This included measures concerning withdrawals, including making it harder to withdraw, removing cancel withdrawal options, and rethinking the design of win redepositing options. Other proposals include flagging deposit intensity as opposed to amount and the banning of DPS at unlicensed gambling operators.

> *It is easier to deposit than to withdraw your money, but it should be the other way around ... I will make it harder for withdrawal.* [Participant 3]

> *So, the amount might not be a good sign, but how many times I did (deposit) should be the sign for them to signal ... no one asks you if you do it again and again.* [Participant 4]

Discussion

General Findings

Our preliminary findings revealed DPS's ability to seamlessly integrate into existing habits and enhance players' ability to achieve their goals. We also found evidence that the harmful influence of DPS can, at times, originate from its technological makeup. Consequently, changes in subjective experiences that emerged in the use of DPS were multifaceted. First, existing gambling routines, sensemaking of DPS features, and variation of individual habits and behaviors influence how and to what extent DPS affects gambling behavior. This aligns with existing literature that shows how various users interpret and use DPS features in diverse ways. Some problem gamblers, for example, were shown to perceive DPS's budget tools as a means to address their gambling issues while players who identify themselves as healthy gamblers considered such features as merely a "safety net" [5] or found them restrictive and unnecessary [17].

Second, our findings indicate that the nature of DPS can influence players' gambling behaviors in a variety of ways. For instance, participants felt that they were in control of their spending since they authorized the deposit "button" at all times, while the frictionless payment flow diminished the money spent. As such, the inherent flexibility and "invisibility" of DPS in players' gambling experiences enabled the mediation of conflicting values and facilitated impulsive deposits over rational decision-making. In another respect, we observed a more pervasive nature of DPS where it can trigger new gambling behaviors. For instance, participants reported gambling more often in situations that they would not typically gamble in (eg, workplace or driving), while others noted that they developed trust in illegal operator sites because of their familiarity with the DPS used by those operators.

Finally, participants' discussion on their general perception of DPS surfaced different perspectives that span from an unfavorable view of DPS to an agnostic attitude to recommendations on how to make payment solutions better. The findings underline players' keen awareness of digital payment pitfalls in their gambling behavior and disappointment in the lack of RG effort in the area. In what follows, some theoretical and practical implications of this study are outlined.

Theoretical Implications

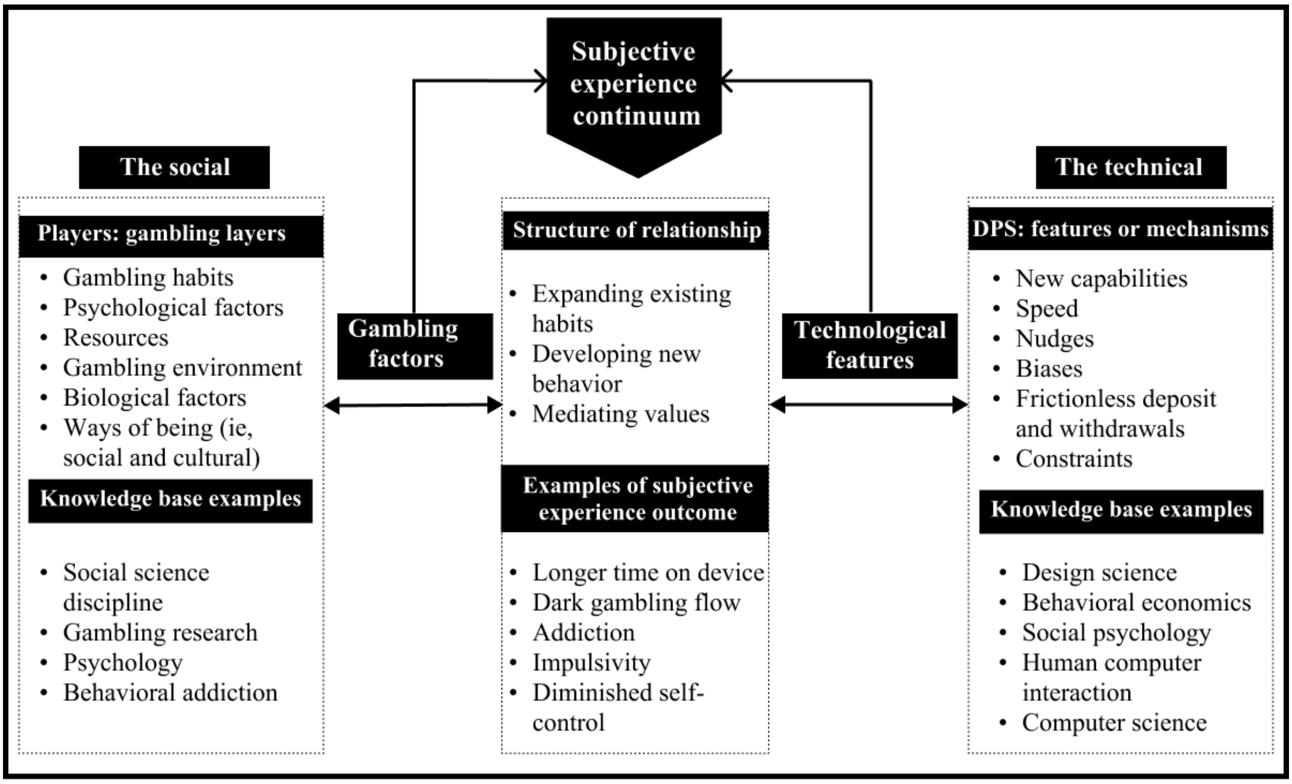
This study highlights the need for interdisciplinary research to conceptualize the role of DPS in gambling activities. On the one hand, the evidence of how existing gambling behavior influences players' way of integrating DPS for gambling is in line with research that underscores the "social imperative on the technical"—the argument that technological features or their use are dependent on, and a product of human choices [34]. Typically, characterized as a sociocentric framing, users and their existing habits influence how technology is used, hence the outcome of subjective experience [35]. From this perspective, DPS can be seen as a tool to enhance one's ability to work around, augment, or enable existing habits in gambling activities.

On the other hand, the findings that emphasize DPS's both subtle and pervasive influence in determining the outcome of gambling, underscore a techno-centric perspective where technology has a more "deterministic" demeanor and direct effect on subjective experience. From this contrasting perspective, a more dominating role can be manifested where DPS can constrain, institutionalize, or directly shape players' gambling activities. In addition to being visibly deterministic, technology can have an "imposter" nature by presenting itself as a response to human needs while subtly nudging players toward harmful experiences [36]. With so-called user-centric design, for instance, gambling technologies have been shown to trigger greater betting intention, an illusion of control, and attention bias [3,37]. In both the "subtle and pervasive" nature of technology influence, our analysis has shown that DPS can facilitate gambling dark flow, longer time-on-device, access to unsafe gambling environments, diminish the value of spending, and nudge toward harmful games, such as casinos. Finally, the innate nature of technology, such as its transparency, is associated with lessening the pain of paying, resulting in impulsive consumption [33].

As such, conceptualizing the influence of DPS on gambling behavior constitutes a confluence of multifactorial structures. Put another way, capturing the interplay between the socio and technical aspects, symbolized by the hyphen that connects the 'socio-technical' term, is imperative to conceptualize the role of DPS in gambling activities [38]. The societal context on the left side of the hyphen encompasses cultural factors, sociodemographics, psychology, and broader societal influences, all of which shape habits and behaviors [39]. Representing the technical, the right side of the hyphen encompasses DPS's technological features and mechanisms that facilitate players in achieving their goals or directly enable new capabilities and forms of gambling activities. A preliminary conceptual framework illustrating the interaction between the social and technical aspects within the context of DPS is depicted in Figure 1. DPS-related changes in gambling subjective experiences (eg, addiction) can be seen as an outcome of various factors from the social and the technical. In terms of their scale of influence, these factors are not equally distributed and vary based on contexts, such as individuals and time (eg, severity of gambling addiction).

As players integrate DPS into their gambling activities in the context of existing habits, their subjective experience can shift toward harmful gambling based on what the technology affords them to achieve. In addition, the framework suggested that layers of contextual factors that can affect an individual's use of DPS in gambling are best developed in social disciplines, such as gambling research, psychology, and behavioral addiction [39]. DPS can also directly influence gambling experience outcomes in the form of technological features using mechanisms, such as nudging, pervasive enframing, and constraining options. These features can introduce harm to gambling activities in the form of speed and biases or enable new ways of harmful gambling, which in turn can be translated to longer time-on-device, excessive spending, or dark flow of gambling. The development of these features and mechanisms can be both technical and psychological; hence, they apply design and behavioral science principles.

Figure 1. Preliminary conceptual framework to examine the influence of DPS in gambling subjective experience. DPS: digital payment solution.



Practical Implications

Historically, intervention measures, policies, and RG tools predominantly focus on gambling operators since they are responsible for how the gambling product is ultimately presented to the customer. However, the findings from this study illustrate that DPS can play an active role in facilitating harmful gambling experiences. Account-level protection, such as a "universal" budget limit and single customer view options with high spending and gambling intensity data sharing, are some of the opportunities that exist for policy-anchored implementations. In addition, it is important to detect and disable technology-enabled workarounds, such as gambling on unlicensed sites to rein in DPS's offsetting nature of RG efforts. In line with policies, effective design interventions can also be used to compensate for "harmful" affordances. For instance, as noted by the participants, intentional delays in win withdrawal and monthly gambling spending reports across operators are within reach for any financial institution given their data set.

Finally, there have been calls for a more comprehensive and cross-sectoral approach, incorporating measures, such as data interoperability among industry stakeholders and integrating public health perspectives [40,41]. With appropriate regulations, DPS with gambling customers can play a pivotal role in data-sharing efforts that enable an overview of a customer's gambling activities across different operators. In addition, public health preventive measures, such as affordability testing, which takes into account both financial state and gambling behavior, can be implemented in DPS platforms as a deterrent against unaffordable gambling [42]. These preventive measures also serve as a "safety net" for nonproblem gamblers, thereby facilitating the implementation of a population-oriented harm minimization strategy [5].

Limitations and Future Research

Given the exploratory nature of this study, several limitations have been identified, highlighting both the necessity for further investigation and new research opportunities. First, due to its limited size of empirical evidence, there is a need for more studies with a larger number of participants for the generalizability of this study's results. Second, we decided to select problem gamblers as focus group participants, as financial decision-related gambling harm is more visible among high-intensity gambling user groups. However, the associations made between gambling behaviors and DPS can benefit from future diverse groups of gambling participants. Furthermore, there were no notable demographic-based differences observed in the use of DPSs, possibly due to the group's uniformity in terms of payment methods used, types of gambling, or the regular attendance of participants in biweekly gambling support sessions. We also want to stress that the proposed conceptual framework presented here should be seen as a preliminary effort to start the conversation of how DPS, in its capacity as a technological product, can potentially facilitate harmful behavior. Using the corresponding knowledge base noted on either side of the hyphen, future research can identify and conceptualize key sociotechnical elements and their mode of interaction in shaping players' behavior.

Finally, technological products can be intentionally designed to mediate certain values—positive interventions or self-interest goals—that can shift subjective experiences for better or worse [43]. As such, more research is needed on the design aspect of technology that has a target audience in gambling. Due to space limitations and the scope of this paper, we will not go further into the design research field. However, the concepts of user-centered design, design for appropriation, persuasive system design methods, and overall intentional design research need to be part of future research to examine technological antecedents and their influence on gambling behavior [44-48].

Conclusions

Given that online gambling will continue to optimize its edge with the support of technological products, new challenges are emerging for researchers and policy makers in the effort to reduce gambling harm. This study focused on problem gamblers' first-hand experience of DPS to examine its role in the overall setup of online gambling activities. Our preliminary findings suggest that seamless collaboration between operators and DPSs can create a potential environment conducive to facilitating harmful gambling behavior. Given their central place in gambling activities, however, DPS providers are also in a unique position to implement umbrella-like harm reduction measures, such as limit settings applicable across gambling sites. As such, there exist avenues for policy makers to onboard DPS to positively contribute to the effort of the RG agenda, as they have a bird's-eye view of gamblers across different operators.

Finally, as digital payment providers continue to attune their products to gambling merchants using "user-centric" design and consumer behavior data, the field needs to examine DPS's platform-like role in transforming players' subjective experiences in online gambling. This study highlights possible interdisciplinary avenues, such as sociotechnical-oriented research, to fully grasp the role of technological products in general and DPS in particular in gambling behavior. The preliminary framework demonstrates the interplay between the sociotechnical components—from privileging the social while acknowledging the technical to giving equal agency for both sides of the hyphen to tech-dominated social experiences. More research is needed to further delineate the structure of the relationship between enabling technological products and the characteristics of gamblers that will ultimately influence the outcome of gambling.

Acknowledgments

The authors would like to thank their focus group participants for sharing their stories and willingness to spend time with us. This study was funded by a grant from the Independent Research Council (forskningsråd) of Svenska Spel in Sweden (fund ID FO2022-0004).

Data Availability

The data sets generated and analyzed during this study are not publicly available due to the sensitivity of the data and per ethical approval guidelines.

Conflicts of Interest

NL has received funding from the independent research council of Svenska Spel and is currently employed by Karolinska Institute. He has not received any direct reimbursement from the gambling industry, nor holds any shares. JJ is an employee at Sustainable Interaction, a private company working with responsible gambling and online training with the gambling industry, reports past and ongoing industry-academia collaborations with several gambling providers, yet has no personal financial ties to the industry. PL is employed by the public-operated addiction clinic in Stockholm and has past and ongoing industry-academia collaborations (with full academic freedom) with several gambling operators, including project-specific financing, albeit not on the topic of the manuscript in question. PL has no personal financial ties to the gambling industry (including no honoraria) nor any other ties.

Multimedia Appendix 1

Brief demographics and gambling background questionnaires.

[DOCX File , 13 KB - humanfactors_v11i1e54951_app1.docx]

Multimedia Appendix 2

Focus group material: summary of discussion questions.

[DOCX File , 15 KB - humanfactors_v11i1e54951_app2.docx]

Multimedia Appendix 3
Empirical material analysis and theme development process.
[DOCX File , 24 KB - humanfactors_v11i1e54951_app3.docx]

References

1. Cemiloglu D, Naiseh M, Catania M, Oinas-Kukkonen H, Ali R. The fine line between persuasion and digital addiction. 2021 Presented at: International Conference on Persuasive Technology; June 23, 2021; United States p. 289-307. [doi: 10.1007/978-3-030-79460-6_23]
2. Shaffer HJ. Understanding the means and objects of addiction: technology, the internet, and gambling. J Gambling Stud 1996 Dec;12(4):461-469. [doi: 10.1007/bf01539189]
3. Flayelle M, Brevers D, King DL, Maurage P, Perales JC, Billieux J. A taxonomy of technology design features that promote potentially addictive online behaviours. Nat Rev Psychol 2023;2(3):136-150. [doi: 10.1038/s44159-023-00153-4]
4. Schüll ND. Addiction by Design. United States: Princeton University Press; 2012.
5. Lakew N. "Show me the money": preliminary lessons from an implementation of intervention tools at the payment gateway level. J Gambl Stud 2022;38(1):297-317. [doi: 10.1007/s10899-021-10023-4] [Medline: 33761066]
6. Rao S. E-wallet-a 'pay'volution. 2015 Presented at: Business Dynamics—A Paradigm Shift in Policies for Sustainability; August 23, 2015; India p. 1-12.
7. Kiayias A, Kohlweiss M, Sarencheh A. Peredi: privacy-enhanced, regulated and distributed central bank digital currencies. 2022 Presented at: Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security; November 7, 2022; Los Angeles, CA, USA p. 1739-1752. [doi: 10.1145/3548606.3560707]
8. Gustafsson T. Designing a European gambling self-exclusion registry. Tallinn University of Technology. 2021. URL: https://tinyurl.com/ybj8rwbx [accessed 2021-05-10]
9. Archer N, Wang S, Pei Y. Reputation, familiarity and use intention for e-payment services: a comparison of pure-play and click-and-mortar e-payment services. IJSTM 2021;27(1-2):72-103. [doi: 10.1504/ijstm.2021.10035938]
10. Farrow GS. Open banking: the rise of the cloud platform. J Paym Strategy Syst 2020;14(2):128-146.
11. Baker TH, Stone C. Making outcomes matter: an immodest proposal for a new consumer financial regulatory paradigm. SSRN J 2020 Nov 1;4:1-56. [doi: 10.2139/ssrn.3607308]
12. Financial well-being: What it means and how to help. Consumer Financial Protection Bureau. 2017. URL: https://files.consumerfinance.gov/f/201501_cfpb_digest_financial-well-being.pdf [accessed 2024-06-28]
13. Gainsbury SM, Blaszczynski A. Digital gambling payment methods: harm minimization policy considerations. Gaming Law Rev 2020;24(7):466-472. [doi: 10.1089/glr2.2020.0015]
14. Lavoie RV, Main KJ. When losing money and time feels good: the paradoxical role of flow in gambling. JGI 2019;41:53-72. [doi: 10.4309/jgi.2019.41.4]
15. Ghaharian K, Abarbanel B, Kraus SW, Singh A, Bernhard B. Players gonna pay: characterizing gamblers and gambling-related harm with payments transaction data. Comput Human Behav 2023;143:107717. [doi: 10.1016/j.chb.2023.107717]
16. Limbrick-Oldfield EH, Chua C, Cringle N, MacDonald K, Ferrari MA, Zhang K, et al. Cashless gambling and the pain of paying: effects of monetary format on slot machine gambling. Addict Res Theory 2021;30(3):220-230. [doi: 10.1080/16066359.2021.2009465]
17. Swanton TB, Tsang S, Collard SB, Garbarino E, Gainsbury SM. Cashless gambling: qualitative analysis of consumer perspectives regarding the harm minimization potential of digital payment systems for electronic gaming machines. Psychol Addict Behav 2024;38(4):451-464. [doi: 10.1037/adb0000962] [Medline: 37768594]
18. Swedberg R. Exploratory research. In: The Production of Knowledge Enhancing Progress in Social Science. Cambridge: Cambridge University Press; 2020:17-41.
19. Farrelly P. Choosing the right method for a qualitative study. Br J School Nurs 2013;8(2):93-95. [doi: 10.12968/bjsn.2013.8.2.93]
20. Stevens RE, Loudon DL, Cole H, Wrenn B. Exploratory (qualitative) research. In: Concise Encyclopedia of Church and Religious Organization Marketing. 1st Edition. New York, NY: Routledge; Apr 15, 2013.
21. Stebbins RA. Exploratory Research in the Social Sciences. London: Sage; 2001.
22. Rendle KA, Abramson CM, Garrett SB, Halley MC, Dohan D. Beyond exploratory: a tailored framework for designing and assessing qualitative health research. BMJ Open 2019;9(8):e030123 [FREE Full text] [doi: 10.1136/bmjopen-2019-030123] [Medline: 31462482]
23. Mansourian Y. Exploratory nature of, and uncertainty tolerance in, qualitative research. New Libr World 2008;109(5/6):273-286. [doi: 10.1108/03074800810873614]
24. Langham E, Thorne H, Browne M, Donaldson P, Rose J, Rockloff M. Understanding gambling related harm: a proposed definition, conceptual framework, and taxonomy of harms. BMC Public Health 2016;16:80 [FREE Full text] [doi: 10.1186/s12889-016-2747-0] [Medline: 26818137]
25. Onwuegbuzie AJ, Dickinson WB, Leech NL, Zoran AG. A qualitative framework for collecting and analyzing data in focus group research. IJQM 2009;8(3):1-21. [doi: 10.1177/160940690900800301]

26. Morgan DL. Qualitative content analysis: a guide to paths not taken. Qual Health Res 1993;3(1):112-121. [doi: 10.1177/104973239300300107] [Medline: 8457790]
27. Corbin JM, Strauss A. Grounded theory research: procedures, canons, and evaluative criteria. Qual Sociol 1990;13(1):3-21. [doi: 10.1007/bf00988593]
28. Myers G. 'Where are you from?': identifying place. J Socioling 2006;10(3):320-343. [doi: 10.1111/j.1360-6441.2006.00330.x]
29. Cash P, Snider C. Investigating design: a comparison of manifest and latent approaches. Des Stud 2014;35(5):441-472. [doi: 10.1016/j.destud.2014.02.005]
30. Cecez-Kecmanovic D, Galliers RD, Henfridsson O, Newell S, Vidgen R. The sociomaterialty of information systems: current status, future directions. MIS Q 2014;38(3):809-830. [doi: 10.25300/misq/2014/38:3.3]
31. Goh JM, Gao G, Agarwal R. Evolving work routines: adaptive routinization of information technology in healthcare. ISR 2011;22(3):565-585. [doi: 10.1287/isre.1110.0365]
32. Leonardi PM. When flexible routines meet flexible technologies: affordance, constraint, and the imbrication of human and material agencies. MIS Q 2011;35(1):147-167. [doi: 10.2307/23043493]
33. Mazar N, Plassmann H, Robitaille N, Lindner A. Pain of paying? A metaphor gone literal: evidence from neural and behavioral science. Rotman School of Management. 2016. URL: https://marketing.wharton.upenn.edu/events/colloquia-2016-2017/ [accessed 2024-07-04]
34. Sarker S, Chatterjee S, Xiao X, Elbanna A. The sociotechnical axis of cohesion for the IS discipline: its historical legacy and its continued relevance. MIS Q 2019;43(3):695-720. [doi: 10.25300/MISQ/2019/1374]
35. de Guinea AO, Markus ML. Why break the habit of a lifetime? Rethinking the roles of intention, habit, and emotion in continuing information technology use. MIS Q 2009;33(3):433-444. [doi: 10.2307/20650303]
36. Feenberg A. Critical theory of technology: an overview. In: Information Technology in Librarianship: New Critical Approaches. Westport: Libraries Unlimited; Mar 04, 2008:31-46.
37. Newall PWS. Reduce the speed and ease of online gambling in order to prevent harm. Addiction 2022 Aug 24;118(2):204-205. [doi: 10.1111/add.16028] [Medline: 36002979]
38. Meyer ET. Examining the hyphen: the value of social informatics for research and teaching. In: Fichman P, Rosenbaum F, editors. Social Informatics: Past, Present and Future. Newcastle Upon Tyne: Cambridge Scholars Publishing; 2014.
39. Hilbrecht M, Baxter D, Abbott M, Binde P, Clark L, Hodgins DC, et al. The conceptual framework of harmful gambling: a revised framework for understanding gambling harm. J Behav Addict 2020;9(2):190-205 [FREE Full text] [doi: 10.1556/2006.2020.00024] [Medline: 32554839]
40. Wardle H, Reith G, Langham E, Rogers RD. Gambling and public health: we need policy action to prevent harm. BMJ 2019;365:1807 [FREE Full text] [doi: 10.1136/bmj.l1807] [Medline: 31068335]
41. Ladouceur R, Blaszczynski A, Shaffer HJ, Fong D. Extending the Reno model: responsible gambling evaluation guidelines for gambling operators, public policymakers, and regulators. Gaming Law Rev Econ 2016;20(7):580-586. [doi: 10.1089/glre.2016.2074]
42. Marionneau V, Ruohio H, Karlsson N. Gambling harm prevention and harm reduction in online environments: a call for action. Harm Reduct J 2023;20(1):92 [FREE Full text] [doi: 10.1186/s12954-023-00828-4] [Medline: 37481649]
43. Cemiloglu D, Arden-Close E, Hodge SE, Ali R. Explainable persuasion for interactive design: the case of online gambling. J Syst Software 2023;195:111517. [doi: 10.1016/j.jss.2022.111517]
44. Abras C, Maloney-Krichmar D, Preece J. Encyclopedia of Human-Computer Interaction. Thousand Oaks, CA: Sage Publications; 2004.
45. Dix A. Designing for appropriation. 2007 Presented at: Proceedings of HCI 2007 The 21st British HCI Group Annual Conference University of Lancaster; September 3, 2007; United Kingdom p. 1-4. [doi: 10.14236/ewic/hci2007.53]
46. Goodman E, Stolterman E, Wakkary R. Understanding interaction design practices. 2011 May 01 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; May 7, 2011; Canada p. 1061-1070. [doi: 10.1145/1978942.1979100]
47. Gregory J. Scandinavian approaches to participatory design. 2003 Presented at: Proceedings of the 4th International Conference on Persuasive Technology; April 26, 2009; Claremont, CA, USA p. 74.
48. Torning K, Oinas-Kukkonen H. Persuasive system design: state of the art and future directions. 2009 Apr 26 Presented at: Proceedings of the 4th International Conference on Persuasive Technology; April 26, 2009; United States p. 1-8. [doi: 10.1145/1541948.1541989]

Abbreviations

DPS: digital payment solution
PG: problem gambling
RG: responsible gambling

Edited by K Adapa; submitted 28.11.23; peer-reviewed by M Bellringer, V Velasco; comments to author 12.03.24; revised version received 28.03.24; accepted 23.05.24; published 23.07.24.

Please cite as:
Lakew N, Jonsson J, Lindner P
Probing the Role of Digital Payment Solutions in Gambling Behavior: Preliminary Results From an Exploratory Focus Group Session With Problem Gamblers
JMIR Hum Factors 2024;11:e54951
URL: https://humanfactors.jmir.org/2024/1/e54951
doi:10.2196/54951
PMID:

Original Paper

Digital Adherence Technologies and Differentiated Care for Tuberculosis Treatment and Their Acceptability Among Persons With Tuberculosis, Health Care Workers, and Key Informants in the Philippines: Qualitative Interview Study

Chung Lam Leung[1], BSSc, MA; Jason Alacapa[2], BSc, MBA, MHM, MPH; Bianca Gonçalves Tasca[1], MSc, MPH; Andre Daniel Villanueva[2], MD; Saniata Masulit[2], MA; Marvin Louie Ignacio[2], MOH; Kathleen Nicole Uy[2], RN; Christopher Pell[1], PhD; Kristian van Kalmthout[1], MSc; Rachel Powers[1], BA, MSc; Katherine Fielding[3], BSc, MSc, PhD; Degu Jerene[1], MD, PhD

[1]KNCV Tuberculosis Foundation, Den Haag, Netherlands

[2]KNCV Tuberculosis Foundation Philippines, Metro Manila, Philippines

[3]TB Centre and Department of Infectious Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, United Kingdom

Corresponding Author:
Chung Lam Leung, BSSc, MA
KNCV Tuberculosis Foundation
Maanweg 174
Den Haag, 2516AB
Netherlands
Phone: 31 070 416 7222
Email: adrian.leung@kncvtbc.org

Abstract

Background: Digital adherence technologies (DATs) are being studied to determine their potential to support tuberculosis (TB) treatment and address the shortcomings of directly observed therapy. Previous research has shown inconclusive results on whether DATs can enhance medication adherence among persons with TB.

Objective: This study aims to understand the acceptability of DATs, namely, medication labels and smart pillboxes, among persons with TB, health care workers (HCWs), and key informants (KIs) in the Philippines. The objective is to gain valuable insights that can inform the design and implementation of DATs in the Southeast Asian region, which meet the needs and preferences of end users.

Methods: Persons with TB, HCWs, and KIs were recruited from intervention facilities to participate in in-depth interviews conducted between March 2022 and January 2023. These interviews were transcribed and translated into English. A thematic analysis was carried out using NVivo software (Lumivero) to identify and analyze themes. Themes were then structured within a modified social-ecological model.

Results: A total of 25 persons with drug-sensitive TB and 20 HCWs or KIs were interviewed. Both groups emphasized that users' technology literacy level, financial conditions, and motivation to be cured determined how they interacted with the DAT. They also acknowledged that DATs helped foster their relationship with HCWs and enabled efficient treatment support. Concerning technology, persons with TB found DATs easy to use and able to reduce clinic visits. HCWs mentioned that DATs added to their workload but also allowed them to support users who missed doses. However, both groups experienced technical challenges with DATs. Regarding program implementation, users appreciated the clear explanations and demonstrations provided by HCWs. Yet, some users reported inconsistencies between DAT settings and the information provided. HCWs stressed the importance of comprehensive training and sufficient resources for effective program implementation in the future. At the community level, both groups noted that DATs and program design protected users' privacy and reduced the risk of stigma. Finally, users and HCWs shared various contextual factors that influenced their experience with DAT, including infrastructure challenges and the impact of the COVID-19 pandemic.

Conclusions: In the Philippines, persons with TB and HCWs showed a high level of acceptance and satisfaction with the impact of DAT and program design. They expressed a desire for the continuation of DATs. The challenges encountered underscore the need for ongoing technological development to minimize malfunctions, enhance the capacity of health facilities, and improve infrastructure. DATs have demonstrated their ability to strengthen user-HCW relationships and protect users from stigmatization. Additional efforts are required to scale up the DAT program in the Philippines.

(JMIR Hum Factors 2024;11:e54117) doi:10.2196/54117

KEYWORDS

tuberculosis; digital adherence technologies; implementation; acceptability; qualitative research; Philippines; digital health; tuberculosis treatment; support strategy; support; medication adherence; health care workers; interview; interviews; user; user privacy; privacy; digital adherence

Introduction

Endorsed by the World Health Organization (WHO), directly observed therapy (DOT) is part of tuberculosis (TB) programs worldwide at varying degrees of implementation. Despite its widespread implementation, shortcomings in the DOT approach have emerged. Frequent health facility visits disrupt patients' daily routines, increase financial burdens, and impact patient autonomy [1,2]. Hence, alternative treatment support strategies such as digital adherence technologies (DATs) are being evaluated to determine their effectiveness in supporting TB treatment and addressing the limitations of DOT [3-5]. These technologies leverage mobile phones, computers, web-based platforms, and smart pillboxes to collect real-time data on patient treatment adherence.

Studies conducted in Kenya, Vietnam, Peru, and China have shown that DATs have the potential to improve medication adherence among persons with TB [4,6-9]. By empowering persons with TB to take a more active role in managing their health conditions alongside health care providers, these technologies have the potential to reduce the financial burden and disruption associated with DOT and promote autonomy [10]. Additionally, DATs allow health care workers (HCWs) to identify individuals who may require additional support during their treatment [11]. Studies conducted elsewhere have yielded inconsistent evidence regarding the impact of DATs on improving adherence outcomes compared to the standard of care, which includes DOT [4,12-17].

The multicountry Adherence Support Coalition to End TB (ASCENT) project was initiated to add to the evidence base around DATs [18]. This multidisciplinary project compared the impact of DATs against the standard of care in terms of treatment outcomes and sought to determine the feasibility and acceptability of DATs for TB treatment support. Understanding the acceptability of DATs among persons with TB and HCWs implementing DATs are key to determining the technology's long-term impact and viability. According to Sekhon et al [19], the acceptability of a health care intervention is determined by the "people delivering or receiving a health care intervention consider it to be appropriate, based on anticipated or experienced cognitive and emotional responses to the intervention." Currently, there is limited research examining the acceptability of DATs among persons with TB and HCWs in the Southeast Asian region; further research is needed to determine their long-term impacts and viability [20-22].

Drawing on in-depth interviews conducted as part of the ASCENT study in the Philippines, this paper describes the perceptions and experiences of persons with TB, HCWs, and key informants (KIs) regarding DATs. It seeks to address the evidence gap around the acceptability of DATs and ultimately seeks to garner insights to design and implement DATs in the Southeast Asian region that are responsive to the needs and preferences of the end users.

Methods

Study Setting

This study is part of the ASCENT: 4-country evaluation of DATs for TB treatment, aimed at evaluating the effectiveness of DATs across 4 countries [18]. The DAT intervention included the provision of medication labels or smart pillboxes (Multimedia Appendix 1), a digital adherence platform, and differentiated care among adults with drug-sensitive tuberculosis (DS-TB) in the Philippines, South Africa, Tanzania, and Ukraine.

According to the WHO [23], the total TB incidence rate in the Philippines was 638 per 100,000 population in 2022. In the 2000s, directly observed treatment short-course achieved full coverage in the Philippines, being available in public and private health care facilities [24]. As per the National Tuberculosis Control Program (NTP) Manual of Procedures 6th edition of the Philippines [25], DOT can be administered within a health care facility on a daily basis by an HCW or outside the facility by a trained treatment supporter, typically receiving a 1-week supply of medication. The treatment approach is determined jointly by the persons with TB and the health care provider.

In the Philippines, the ASCENT study used a cluster randomized trial, using health facilities as the unit of randomization. In total, 64 health facilities from the Bulacan and Pampanga provinces in Region III of the Philippines were included. Of the 32 facilities in the intervention arm, 16 used medication labels, while the remaining ones implemented smart pillboxes. Adults (18 years or older) starting the DS-TB regimen were invited to enroll. In medication label facilities, DAT was offered to consenting persons with TB who had access to a mobile phone, either their own or that of their treatment companion. Persons with TB who did not have access to or had difficulties using mobile phones were offered a smart pillbox while supplies lasted in that facility. In smart pillbox facilities, consenting persons with TB were offered a smart pillbox. Individuals younger than

18 years and with drug-resistant TB were excluded from the study according to the study protocol [18].

The overarching digital adherence platform was used to track adherence information for both DATs. Medication label users were asked to take their daily medication and then log their intake by sending an SMS text message using the toll-free number on the medication blister packs. Some telecom operators in the Philippines require a positive credit balance (Php 1.00 or approximately US $0.02) to send free SMS text messages. Smart pillbox users received a pillbox to store their medication. When the user opens the box to take their medication, the box's sensor sends a notification to the platform registering the medication intake. In cases of an unstable network connection, the box records the opening and transmits the timestamped data once the network is available again. If a user fails to send a code or open the box before 4 PM, the system triggers a reminder message sent to the user's phone to prompt the patient to take and log their daily dose. If no dose is subsequently logged before midnight, the system triggers a notification message sent to the user at 8 AM the following morning, alerting them of their missed logged dose and reminding them to complete the current day's intake.

The digital adherence platform allows HCWs to remotely view users' treatment adherence. Each person with TB is linked to an adherence calendar, visualized by 1 colored box per day. Individuals who have logged their dose by opening their smart pillbox or sending a message will have that day marked in green on the platform, whereas those who have missed a dose will have it marked in red. If a person with TB missed 1 daily dose, the HCW was expected to make a follow-up call to provide support. If a person with TB missed doses for 2 consecutive days, community-level HCWs were expected to visit the user. Persons with TB were only required to visit the facility to refill their medication or seek medication advice. HCWs were expected to check the system daily, wherein they viewed the adherence calendars and accessed Task Lists—consolidated lists per support action (eg, call and home visit) indicating which persons with TB required follow-up support. This full pathway of support escalation, or differentiated care, was determined together by in-country health system stakeholders, implementers, and patient advocacy groups.

User adherence records were captured by the adherence platform and then transferred to the local Department of Health (DoH) Integrated Tuberculosis Information System. Prior to the implementation, training was provided to all HCWs from the 32 intervention facilities involved in the project on platform use, patient orientation, knowledge about DATs, and treatment counseling.

Participant Recruitment and Data Collection

We used semistructured, in-depth interview methods. DAT users (medication label and smart pillbox), HCWs, and KIs were purposively recruited from up to 10 ASCENT DAT facilities in the Philippines. HCWs and KIs were selected based on their role and level of involvement in the project. Interviews were conducted in enclosed meeting rooms within the facilities, and before the interviews, all participants provided informed consent for the interview and it being recorded. Data collection took place between March 2022 and January 2023.

For user interviews, we aimed to recruit a minimum of 20 participants, comprising 10 male and 10 female participants, who were currently undergoing TB treatment in DAT facilities (offering either DAT). Users were selected based on either demonstrating documented good adherence, receiving differentiated care, or having chosen not to use DAT. Furthermore, we sought to purposively recruit a minimum of 20 DAT HCWs and KIs from the DoH or the NTP of the Philippines.

Data Collection

The interviews were conducted by 1 researcher and 2 research assistants from the ASCENT Philippines team. These individuals received training in qualitative research methodologies prior to data collection. An open-ended interview guide was used to facilitate the interview with persons with TB. The guide covered topics, such as users' experiences with using DATs, including unintentional disclosure, factors that facilitated adherence or nonadherence to the treatment, and cultural or other barriers that deterred them from using DATs. Due to COVID-19 restrictions in the Philippines, interviews were conducted via phone calls. Due to language barriers and age considerations, some interviews were conducted in the presence of the user's treatment companion. The duration of the interviews ranged from 30 to 50 minutes. The same research team used a separate interview guide developed for HCWs and KIs. They were queried about their perceptions and the feasibility of implementing DATs in the Philippines. The duration of these interviews ranged from 40 to 60 minutes. Interviews were conducted in either Filipino or English, depending on the participants' preference, and were audio recorded.

Data Processing and Analysis

The recorded interviews were transcribed verbatim in the local language and subsequently translated into English. To ensure accuracy and quality, an independent staff member translated the English transcript back into Filipino for cross-checking. This process helped validate the quality of the translation and ensure that the original meaning was preserved.

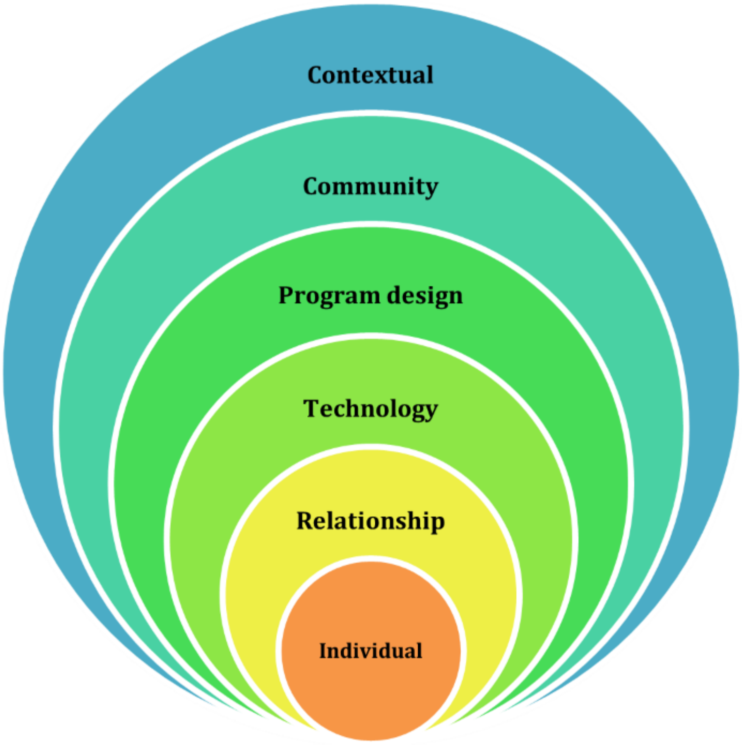
Thematic analysis was conducted to identify and analyze themes emerging from interviews, capturing the richness of the data while remaining independent of existing theory [26]. Using critical realism, this analysis describes participants' experiences with DAT and their interpretation of those experiences in relation to the local context [27].

A total of 2 coders, CLL and BGT, based in the Netherlands, independently coded the transcripts using NVivo software (Lumivero). Regular meetings were held between the coders and other coinvestigators to discuss the coding process and identify emerging themes. Following the establishment of the initial codebook, the coders collaborated with the interviewers from the Philippines to gain a deeper understanding of the interview process, project implementation, and the contextual factors specific to the country. This collaboration aimed to further refine the codebook.

The coding process was conducted separately, categorizing the type of DAT used by persons with TB or overseen by HCWs. After refining the individual codebook for each DAT, a decision was made to consolidate the analysis of the 2 DATs featured in this study. This decision was supported by the observation of overlapping factors influencing persons with TB and HCWs in their use of the 2 DATs.

Following the coding process, the social-ecological model has been modified to present the identified themes into 6 levels (Figure 1). The individual level considers personal factors (eg, age, education, and income) that influence DAT interaction [28,29]. The relationship level focuses on social ties impacting DAT engagement [29]. The technology level explains how the function of DATs impacts user use. Program design levels relate to how the implementation process influences DAT use. The community level encompasses aspects of social culture, norms, and stigma, explaining how these factors guide interactions involving DATs [30]. Finally, the contextual level considers country-specific conditions (eg, infrastructure and policy) affecting DAT implementation.

Figure 1. The modified socioecological model (adapted from the findings of Sazzad et al [28] and McLeroy et al [29]).



Ethical Considerations

The main study has been approved by the WHO Ethical Review Committee (0003296), the Single Joint Research Ethics Board (the Ethics Committee for the Philippines), and the London School of Hygiene & Tropical Medicine Ethics Committee, United Kingdom. Written informed consent was obtained from all participants prior to their interviews. All interview transcripts were deidentified to preserve the privacy of the participants.

Results

Characteristics of Participants

Interviews were conducted with 25 persons with DS-TB (12 male and 13 female participants; ages ranged from 22 to 86 years; 14 bacteriologically confirmed TB and 11 were clinically diagnosed) from DAT implementation facilities. Overall, 19 persons used medication labels, and 6 used smart pillboxes. Notably, 2 users who chose not to use DAT were invited to participate in the interview but did not provide consent. As a result, this study included only individuals who used DATs in their treatment.

A total of 20 HCWs and KIs (5 male and 15 female participants) were interviewed. In total, 10 HCWs worked in a medication label facility and 10 in a smart pillbox facility. In terms of professions, 13 of the participants were nurses in health care facilities, 3 were NTP medical coordinators, 2 were midwives, 1 was a medical technologist, and 1 was a data encoder. Due to a shortage of HCWs, midwives were assisting with TB treatment; thus, they were included in this study. The 3 KIs included an NTP medical coordinator, an NTP city health officer from Bulacan, and an NTP municipal health officer from Pampanga. In terms of the geographical distribution of HCWs, 11 were from Bulacan, 8 were from Pampanga, and 1 was from Region III province office (Table 1).

Table 1. Interview participant characteristics.

	Values, n (%)
Persons with TB[a] (n=25)	
Sex	
Male	12 (48)
Female	13 (52)
Age (years)	
18-29	7 (28)
30-39	6 (24)
40-49	5 (20)
50-59	3 (12)
≥60	4 (16)
Disease classification	
Bacteriologically confirmed	14 (56)
Clinically diagnosed	11 (44)
DAT[b] type	
Medication label	19 (76)
Smart pillbox	6 (24)
Health care workers and KIs[c] (n=20)	
Sex	
Male	5 (25)
Female	15 (75)
Role	
Nurse	13 (65)
Midwife	2 (10)
Medical technologist	1 (5)
Data encoder	1 (5)
NTP[d] medical coordinator	3 (15)
Facility	
Medication label	10 (50)
Smart pillbox	10 (50)
Province	
Bulacan	11 (55)
Pampanga	8 (40)
Other	1 (5)

[a]TB: tuberculosis.
[b]DAT: digital adherence technology.
[c]KI: key informant.
[d]NTP: National Tuberculosis Control Program.

DAT User and HCW Perceptions and Experiences

Individual Level

At the individual level, users' technology literacy, financial conditions, and motivation to be cured influenced their interactions with the DAT. Prior to implementation, HCWs expressed concerns about the potential technological complexity of DAT, which could pose usability and adaptability challenges for users.

> *... it seems like it's easier for the patient to have face-to-face interaction with the HCW.* [Nurse, female]

Users with low technology literacy and older people encountered difficulties in sending the code using the mobile phone. Without the assistance of family members or treatment companions, older people struggled to comprehend the instructions for using DAT. HCWs also found it challenging to explain the system to individuals with limited technology literacy.

> *Interviewer: When Grandma just started taking medicine, did the nurse teach her as well?*

> *Participant: No, just me because she wouldn't be able to understand the instructions.* [Treatment companion of a medication label user, age ≥60 years, female]

> *So, the challenge when we introduce it to the elderly, they feel like they can't understand, and it's hard to use. Then they refuse.* [Nurse, female]

HCWs observed that many potential DAT users lacked access to a phone or had insufficient credit on their SIM cards, causing fewer individuals to use the medication label. Some HCWs working in the medication label facility mentioned that if the facility ran out of pillboxes, eligible users for joining the DAT program would be excluded from the intervention.

> *That's when we have a problem when the patient doesn't have a cell phone and then they don't have a support system, so we give them a pillbox.* [Nurse, female]

Despite the above personal circumstances, many users showed a strong intrinsic motivation to be cured, which drove them to embrace the DAT and remain committed to their treatment. Some users even used supplementary methods, such as paper calendars and phone alarms, to stay on track with their treatment.

> *It was his willingness to get better. That was his main motivation to continue taking the medications given by the nurse or the doctor.* [Treatment companion of a medication label user, age ≥60 years, male]

Relationship Level

Interpersonal relationships shaped how users interacted with DATs, potentially facilitating their use and potentially being transformed by DATs. Many users disclosed their use of DAT to their family members but did not share this information further. This is because, during their initial facility visit, persons with TB in the Philippines are required to bring along a family member as their treatment companion. Multiple participants mentioned that their family members or treatment companions reminded them to take their medication.

> *Yes, because aside from me, I was also being reminded by my husband.* [Smart pillbox user, age 50-59 years, female]

Users and HCWs both acknowledged that using the DAT helped foster trust between them. It enabled efficient and less intrusive treatment support.

> *there is another monitor on their end where they could see if you took your medicine or not. Doesn't it make you happy and proud to have something like that?* [Medication label user, age 30-39 years, female]

> *Yes, it will help me because there are patients who are compliant and we do not supervise them that much. We have built trust with each other.* [Nurse, female]

> *I can say that I have become closer to the patients but it really depends on the person. They know that I am checking their adherence, so they can approach me with more ease.* [Nurse, female]

Technology Level

At the technology level, both groups experienced positive effects and technical challenges of the DAT that affected TB care. Among the positive effects, many DAT users reported that it reduced the burden of treatment by decreasing the frequency of clinic visits and the associated transportation costs. They described how this allowed them to maintain their regular lives while adhering to their treatment. In contrast to traditional DOT, HCWs agreed DAT can reduce user burdens and increase autonomy considerably.

> *... if we go to the centre daily, the fare back and forth will cost us 200 pesos ... it's difficult if I must go there daily because I can't work and do anything.* [Medication label user, age 18-29 years, male]

> *You have to think economically. If the house is far away, the concern is not just the relationship of having a face-to-face talk, but economically it would burden the client.* [NTP medical coordinator, female]

Moreover, users found DATs easy to use, describing the experience as trouble-free. They appreciated the specific features of each DAT that helped them stay on their treatment. Medication label users valued the SMS text message reminders, while pillbox users relied on the alarms. One user mentioned that simply seeing the smart pillbox at home served as a reminder to take medication.

> *I work the night shift, and sometimes the morning shift, sometimes I don't know the time anymore ... it reminds me that I need to take my medications.* [Medication label user, age 30-39 years, female]

> *... it's good because every time I forget to take medications ... it has an alarm that rings at a quarter to 8 in the morning. So, this pillbox reminds me to take my medications.* [Smart pillbox user, age 30-39 years, male]

Prior to the implementation, HCWs were concerned that DAT would increase their workload, but once the program was in place, some of them changed their perspective. They reported that DATs added to their tasks but also allowed them to view users' adherence and helped them adhere to treatment. This enabled HCWs to provide differentiated care to those who missed doses.

> *At first, I thought it was extra work because you have a lot [of forms] to fill out and with the number of patients we have, a few minutes for each TB patient ...* [Midwife, female]

> *I'm able to know immediately with the use of the apps from Everwell who missed their intake, and I would call right after.* [Nurse, male]

However, both DATs and the adherence platform experienced glitches. Medication label users sometimes received reminders after sending the code or did not receive SMS text message reminders at all. Although a small number of users expressed worry, most users disregarded incorrect reminders, confident that they had taken their medication.

> *A bit dissatisfied because then I get confused. Of course, at my age, I can be forgetful, but I know in myself that I have taken my medications. When they send that type of text, it makes me think again.* [Medication label user, age 40-49 years, female]

Pillbox users and HCWs reported pillbox alarm malfunction, and the adherence platform also failed to show user adherence records correctly.

> *Interviewer: Was there a time when you were not sure what would happen to the pillbox while you were using it?*
>
> *Participant: Yes, because sometimes it suddenly makes a sound, but sometimes we've already taken the medicine but it still makes a sound.* [Smart pillbox user, age 50-59 years, female]
>
> *When the pillbox is opened, sometimes they don't even turn green.* [Midwife, female]

In response to the technical challenge, HCWs highlighted the importance of improving the adherence platform by solving the existing glitches and integrating it with the local digital health information system.

> *Maybe what we mentioned is to integrate the KNCV program with ITIS so that there is less workload for HCWs so that they don't have to work twice.* [Nurse, female]

Program Design Level

Similarly, users and HCWs shared their positive and negative experiences regarding the program implementation, which had direct implications on the effectiveness and user perception of the intervention. Users generally appreciated the clear explanations and demonstrations provided during their initial DAT consultations with HCWs. HCWs remain supportive throughout the treatment was one of the reasons why users adhered to using DATs and following the treatment.

> *they will find out what your concern is ... The people are kind. You won't feel that you are treated like someone who's sick. You will not feel discriminated that you are sick, that you have TB. They are accommodating ... Ahhh caring.* [Medication label user, age 40-49 years, male]

Nevertheless, some users reported that the DAT setting did not align with what HCWs had initially informed them during the introduction session, causing confusion. Additionally, individuals working in environments where cell phones were prohibited faced challenges with the use of DAT.

> *I'm just wondering, the medications are reminded to be taken at 12 midnight in the text message. But the instructions given to me before at the centre are in the morning.* [Medication label user, age 30-39 years, female]
>
> *Cellphones are prohibited in our factory ... Yes, so I get to text during trips.* [Medication label user, age 18-29 years, male]

Regarding HCWs' experience, they acknowledged the value of DAT training in understanding its purpose and operation to alleviate misconceptions and misconceived notions of DAT complexity. However, some HCWs expressed concerns about sporadic or insufficient training opportunities, and others mentioned a lack of follow-up or refresher training. In some instances, HCWs were tasked with implementing DAT without adequate training. Hence, HCWs require comprehensive training and adequate resources to effectively implement the program in the future.

> *I wish there would be another training so that I can really know the program. I was just able to do that by fiddling with my cellphone.* [Nurse, male]
>
> *When she gave it to me, I was not yet trained. There was only one time that she told me to train but that time I couldn't train either because I was alone.* [Midwife, female]

HCWs also indicated that there was a shortage of pillboxes in their facilities, leading to the exclusion of persons with TB from the DAT program when they could not use medication labels for whatever reasons.

> *So, there are times when we can't enrol the patient to DAT because we don't have any pillbox available anymore.* [Nurse, female]

In summary, users and HCWs were satisfied with the result, and they expressed a strong desire for the continuation and expansion of the DAT program in the future.

> *I wish they continue to give me medicine and always remind me.* [Medication label user, age 18-29 years, male]
>
> *I hope this DAT can used by TB patients all over the Philippines because I think it is really useful.* [Nurse, female]

Community Level

Many users highlighted that TB stigma remained an important issue within their communities. They expressed concerns about discrimination and stigmatization, which often deterred them from disclosing their use of the DAT to persons outside their immediate family.

> *I told ma'am that there are a lot of gossip mongers here and they might ask why the centre is bringing medicines to me, right?* [Smart pillbox user, age 40-49 years, male]
>
> *Not anymore, but of course there is discrimination, isn't there? People are not that open about it yet, like,*

> *"Oh my gosh, [she has] TB."* [Medication label user, age 30-39 years, female]

Nonetheless, both groups stated that DATs and the program design helped mitigate the risk of unintentionally disclosing users' health conditions through daily visits to TB clinics. Users viewed DAT as a way to protect their privacy and reduce the chances of experiencing stigma. As one medication label user expressed:

> *Because you stick to texting ... You will not be ashamed to go out as if the stigma you are talking about is already gone ... They don't know you're sick ... so you are not immersed in such an issue ... then your treatment is monitored so you come out of the treatment episode quietly.* [Sleeve label user, age 40-49 years, male]

> *If there is someone that you know saw you at a TB facility, the perception is you have TB, and they will avoid you. So, by means of texting, and having their medicines daily, their neighbors don't have to know that they are having checkups with DOTs.* [Nurse, female]

In this context, stigma not only acts as a factor for users to keep their health condition undisclosed but also serves as an incentive for them to use DAT to avoid revealing their health condition through HCW visits or daily visits to the facilities.

Contextual Level

Users and HCWs also shared various contextual factors that influenced their experience with DAT. Medication label users encountered challenges related to the availability of mobile network signals and power in their areas. Users were not able to charge their phones when there was a power outage in the area. These issues prevented them from instantly sending SMS text messages as required. Some service providers in the Philippines also required users to have a positive balance on their SIM to send the code. When users depleted their credit, they were unable to send the code, which then led to receiving reminder SMS text messages from the adherence platform.

> *It's just really difficult, for example, when the signal is affected ... What happens is when there is no electricity ... you can't feedback.* [Medication label user, age 40-49 years, male]

> *Interviewer: But you mentioned that there were issues with the network sim card?*

> *Participant: Because I am unable to chat when I don't have load. That is why I load up whatever amount.* [Medication label user, age 18-29 years, male]

For HCWs, the challenges posed by the COVID-19 pandemic added to their workload, which, in turn, affected the implementation of the program. Some HCWs were diverted to COVID-19–related tasks, limiting their ability to handle DAT-related tasks.

> *However, because of COVID-19, there were many cases. My job then was to swab patients, so I wasn't able to visit the RHU and handle DAT-related tasks.* [Medical Technologist, female]

HCWs recognized the need for additional resources to address these challenges effectively. They emphasized the importance of improving the network infrastructure to ensure reliable communication for DAT users. Additionally, they called for dedicated staff members to manage and support the DAT program.

> *I think the system improvement with the network because some clients are complaining that they cannot send a message and they didn't receive a reply.* [Nurse, female]

> *There should be staff handling this program ... I have many programs under me. So really there should be staff who will handle that specific program, correct?* [Nurse, female]

Discussion

Findings

This study illustrates the perceptions and experiences of persons with TB, HCWs, and KIs regarding the use of DATs. By analyzing their responses, we can assess the acceptability of DAT as described by Sekhon et al [19]. As previously highlighted, user and HWC use of the 2 DATs are influenced by overlapping factors. Our analysis also observes DAT-specific strengths and weaknesses, which will be illustrated using a modified social-ecological framework.

At the individual level, the results highlight the critical role of user motivation in achieving successful treatment outcomes. Users demonstrated their commitment to treatment through various means, not solely relying on DAT, HCWs, or family support. This observation aligns with prior research conducted in Pakistan [31]. HCWs noted that DATs assisted users adhere to treatment and fostered a sense of accountability for timely medication intake. This finding is consistent with similar results from Uganda and a meta-analysis [12,22]. However, users with lower technology literacy and financial constraints faced more challenges in using the medication label, potentially leading to exclusion from the program if a smart pillbox was unavailable in the facility. Within this group, acceptability appears to be slightly lower. Increasing the supply of smart pillboxes in the future might assure more users can benefit from the program.

Regarding relationships, DAT significantly alters the user-HCW relationships, even in the context of limited in-person interaction when compared to traditional DOT. Through DAT, persons with TB felt supported by knowing that HCWs were supporting them remotely, even when care delivery had been substantially impacted by COVID-19 restrictions. This finding aligns with similar research conducted in Africa, South Asia, and Southeast Asia contexts [4,12,20,21,32,33]. Technological advancements enabled users and HCWs in the Philippines to stay connected via mobile communication apps such as WhatsApp and Facebook. These communication channels contributed to the enhancement of their relationship, requiring minimal effort on both sides. In this aspect, both groups exhibited a high level of acceptability.

At the technology level, users and HCWs showed high levels of acceptability toward DATs due to their ease of use, assistance in maintaining treatment adherence, and reduced treatment burden. HCWs valued the adherence platform for its ability to assist them in providing adherence support despite the additional workload it entails. However, the interviews revealed that medication label users faced more challenges stemming from external factors such as network issues, power supply, user technology literacy, phone ownership, and SIM card credit. Similar challenges were observed in previous research [20,22]. The smart pillbox on the other hand also encountered alarm malfunction issues. Research in Vietnam and India [13,21] showed that users consider pillbox alarms too loud and inconvenient for travel and might attract unnecessary attention when they visit the clinic to refill their medications. Yet, users from the Philippines did not express these concerns. In addition, the adherence platform sometimes failed to show correct user adherence records. Despite both groups expressing worries when the technology failed and stressing that these glitches need to be addressed, they wanted the DAT program to continue in the future, which shows they have a high degree of acceptance in this domain.

At the community level, TB stigma remained a significant concern causing hesitancy among persons with TB to seek help. The 2 groups stressed how DATs and its operating model in the Philippines effectively protected user privacy, reducing the risk of stigmatization. In contrast to certain other countries, users in this context reported no worries about stigma within their families, given the mandatory requirement for them to be accompanied by a family member or treatment companion during their initial consultation. However, they remained concerned regarding potential stigma at the community and societal level. Users refrained from sharing their use of DAT beyond their immediate family and fearing stigmatization if observed visiting a TB clinic. Research conducted in Ethiopia, Thailand, South Africa, and Zambia found that persons with TB do not want to be seen at TB clinics because of perceived stigmatization [34-37]. As previously outlined, in the Philippines, DATs substantially reduce the frequency of facility visits, from daily to every 2 to 4 weeks, depending on the user's circumstances, HCW arrangements, and local COVID-19 restrictions of the time. In some instances, community HCWs deliver medication to the person with TB if one lives in remote areas or without a means of transportation. During COVID-19, some local health care facilities also deliver medication to users because of the government's "no contact policy." The DAT operating model in the Philippines decreases the frequency of clinic visits, protects users from unintended disclosure, and thus reduces their perceived stigma.

Despite COVID-19 restrictions limiting home visits, users were motivated to adhere to their treatment to avoid being visited by HCWs, which might expose their condition within their community. This could explain why the interview participants demonstrated a high level of acceptance toward DAT, as it can protect them from stigmatization.

Program implementation is closely intertwined with contextual factors. The outbreak of COVID-19 significantly influenced DAT implementation in the Philippines, leading the government to enforce policies aimed at containing the virus's spread. The pandemic further strained the already burdened health care system in the Philippines. HCWs stationed at research facilities were required to conduct COVID-19 testing and vaccination, diverting their attention and resources from providing differentiated care. This further amplified the workload of HCWs who were enrolling DAT users and managing 2 parallel health information systems without additional staff. Despite these challenges, they acknowledged that DAT enabled real-time monitoring and facilitated prompt follow-up with nonadherent users, ultimately enhancing their efficiency. The consensus among both groups underscored DAT's capacity to function in regions facing health emergencies and operating in stressed health care systems.

The issue of training remained a persistent concern for HCWs involved with DAT across various settings. Previous research carried out in India revealed that HCWs emphasized the necessity for more DAT training due to the frequent turnover of staff [21]. The phenomenon was also reported by the ASCENT Philippines team. The impact of COVID-19 further heightened matters, with newly recruited HCWs being focused on enrolling persons with TB without sufficient training. This issue is rooted within the health care system, and simply increasing the frequency of DAT training may not fully address the underlying problem. These observations emphasize that the acceptability of DAT is heavily reliant on the local health care system. Expanding the DAT program effectively requires addressing systemic health care issues. Failure to do so may compromise the program's impact and effectiveness.

Limitations

This study has successfully captured the perspectives and experiences of persons with TB, HCWs, and KIs in the Philippines who engaged with DATs. It is the first DAT acceptability research conducted in the Southeast Asian context and provides insights on how to improve DAT adoption in the Philippines from a bottom-up approach. However, this study only included individuals who used DATs in their treatment because some persons with TB who demonstrated low levels of treatment adherence or opted out of using DAT declined participation in this research. It was not possible, therefore, to examine the reasons underlying the low levels of adherence or why persons opted out of DATs. Their experiences and perceptions are extremely important for DoH and NTP to formulate a better DAT implementation. Hence, further research is needed to study users who opted out of DATs. This study included only 6 smart pillbox users, so we might risk underreporting issues encountered by smart pillbox users during the treatment.

Conclusions

In conclusion, this study reveals the experiences, perceptions, and challenges encountered by persons with TB, HCWs, and KIs during the research period. Both groups demonstrated a high level of acceptance and satisfaction toward the positive impact brought by DAT. The challenges encountered underscore the need for ongoing technological development to minimize malfunctions and improve local adoption. The majority of interviewees expressed a desire for the scaling up of DATs in the Philippines.

To achieve this, enhancing the capacity of health facilities through increased staffing, particularly with personnel dedicated to enrollment and user support, is crucial. Regular training can contribute to manageable user monitoring. External factors, such as network availability, SIM card credit, and power supply, need to be fixed to ensure smooth DAT implementation, especially for the medication label, in the future.

Finally, integrating the parallel digital health systems is essential to alleviate the health care workload. Beyond the potential to improve treatment outcomes and reduce loss to follow-up, DAT has demonstrated its ability to enhance user-HCW relationships and prevent unwanted TB treatment disclosure or even stigmatization. Additional efforts are required to enhance the user experience for both persons with TB and HCWs, with the aim of expanding the program to other regions in the Philippines. Although the experiences gained in the Philippines can serve as a reference for other countries in the region considering implementing DAT in the future, it is crucial to recognize that DAT programs need to be tailored to fit within the local health systems and context of each specific region.

Acknowledgments

This research and the main Adherence Support Coalition to End TB project were primarily supported by grants from Unitaid. The funding organization had no role in the study design, data collection, data analysis, data interpretation, or manuscript writing.

Conflicts of Interest

The authors of this study received funding to implement the digital adherence technology (DAT) program in the Philippines as part of the Adherence Support Coalition to End TB project led by KNCV. The Adherence Support Coalition to End TB project aims to establish the evidence base for DATs and oversees their implementation across Tanzania, Ethiopia, and the Philippines. In addition, KNCV receives funding to facilitate the Global DAT Task Force secretariat and by extension to provide technical assistance to countries planning or implementing DAT.

Multimedia Appendix 1
Information on the medication label and smart pillbox.
[DOCX File , 270 KB - humanfactors_v11i1e54117_app1.docx]

References

1. Sagbakken M, Frich JC, Bjune GA, Porter JDH. Ethical aspects of directly observed treatment for tuberculosis: a cross-cultural comparison. BMC Med Ethics 2013 Jul 02;14:25 [FREE Full text] [doi: 10.1186/1472-6939-14-25] [Medline: 23819555]
2. Yellappa V, Lefèvre P, Battaglioli T, Narayanan D, Van der Stuyft P. Coping with tuberculosis and directly observed treatment: a qualitative study among patients from South India. BMC Health Serv Res 2016;16:283 [FREE Full text] [doi: 10.1186/s12913-016-1545-9] [Medline: 27430557]
3. Cattamanchi A, Crowder R, Kityamuwesi A, Kiwanuka N, Lamunu M, Namale C, et al. Digital adherence technology for tuberculosis treatment supervision: a stepped-wedge cluster-randomized trial in Uganda. PLoS Med 2021 May;18(5):e1003628 [FREE Full text] [doi: 10.1371/journal.pmed.1003628] [Medline: 33956802]
4. Liu X, Thompson J, Dong H, Sweeney S, Li X, Yuan Y, et al. Digital adherence technologies to improve tuberculosis treatment outcomes in China: a cluster-randomised superiority trial. Lancet Glob Health 2023 May;11(5):e693-e703 [FREE Full text] [doi: 10.1016/S2214-109X(23)00068-2] [Medline: 37061308]
5. Manyazewal T, Woldeamanuel Y, Holland DP, Fekadu A, Marconi VC. Effectiveness of a digital medication event reminder and monitor device for patients with tuberculosis (SELFTB): a multicenter randomized controlled trial. BMC Med 2022 Sep 28;20(1):310 [FREE Full text] [doi: 10.1186/s12916-022-02521-y] [Medline: 36167528]
6. Boutilier JJ, Yoeli E, Rathauser J, Owiti P, Subbaraman R, Jónasson JO. Can digital adherence technologies reduce inequity in tuberculosis treatment success? Evidence from a randomised controlled trial. BMJ Glob Health 2022;7(12):e010512 [FREE Full text] [doi: 10.1136/bmjgh-2022-010512] [Medline: 36455988]
7. Wei X, Hicks JP, Zhang Z, Haldane V, Pasang P, Li L, et al. Effectiveness of a comprehensive package based on electronic medication monitors at improving treatment outcomes among tuberculosis patients in Tibet: a multicentre randomised controlled trial. Lancet 2024 Mar 09;403(10430):913-923. [doi: 10.1016/S0140-6736(23)02270-5] [Medline: 38309280]
8. Acosta J, Flores P, Alarcón M, Grande-Ortiz M, Moreno-Exebio L, Puyen Z. A randomised controlled trial to evaluate a medication monitoring system for TB treatment. Int J Tuberc Lung Dis 2022 Jan 01;26(1):44-49 [FREE Full text] [doi: 10.5588/ijtld.21.0373] [Medline: 34969428]
9. Velen K, Nguyen TA, Pham CD, Le HT, Nguyen HB, Dao BT, et al. The effect of medication event reminder monitoring on treatment adherence of TB patients. Int J Tuberc Lung Dis 2023 Apr 01;27(4):322-328. [doi: 10.5588/ijtld.22.0500] [Medline: 37035979]
10. Metcalfe JZ, O'Donnell MR, Bangsberg DR. Moving beyond directly observed therapy for tuberculosis. PLoS Med 2015 Sep;12(9):e1001877 [FREE Full text] [doi: 10.1371/journal.pmed.1001877] [Medline: 26372602]

11. Subbaraman R, de Mondesert L, Musiimenta A, Pai M, Mayer KH, Thomas BE, et al. Digital adherence technologies for the management of tuberculosis therapy: mapping the landscape and research priorities. BMJ Glob Health 2018;3(5):e001018 [FREE Full text] [doi: 10.1136/bmjgh-2018-001018] [Medline: 30364330]
12. Musiimenta A, Tumuhimbise W, Mugaba A, Muzoora C, Armstrong-Hough M, Bangsberg D, et al. Digital monitoring technologies could enhance tuberculosis medication adherence in Uganda: mixed methods study. J Clin Tuberc Other Mycobact Dis 2019 Dec;17:100119 [FREE Full text] [doi: 10.1016/j.jctube.2019.100119] [Medline: 31788561]
13. Drabarek D, Anh NT, Nhung NV, Hoa NB, Fox GJ, Bernays S. Implementation of Medication Event Reminder Monitors among patients diagnosed with drug susceptible tuberculosis in rural Viet Nam: a qualitative study. PLoS One 2019;14(7):e0219891 [FREE Full text] [doi: 10.1371/journal.pone.0219891] [Medline: 31329610]
14. Thekkur P, Kumar AN, Chinnakali P, Selvaraju S, Bairy R, Singh AR, et al. Outcomes and implementation challenges of using daily treatment regimens with an innovative adherence support tool among HIV-infected tuberculosis patients in Karnataka, India: a mixed-methods study. Glob Health Action 2019;12(1):1568826 [FREE Full text] [doi: 10.1080/16549716.2019.1568826] [Medline: 30712507]
15. Liu X, Blaschke T, Thomas B, De Geest S, Jiang S, Gao Y, et al. Usability of a medication event reminder monitor system (MERM) by providers and patients to improve adherence in the management of tuberculosis. Int J Environ Res Public Health 2017;14(10):1115 [FREE Full text] [doi: 10.3390/ijerph14101115] [Medline: 28946683]
16. Bediang G, Stoll B, Elia N, Abena JL, Geissbuhler A. SMS reminders to improve adherence and cure of tuberculosis patients in Cameroon (TB-SMS Cameroon): a randomised controlled trial. BMC Public Health 2018 May 02;18(1):583 [FREE Full text] [doi: 10.1186/s12889-018-5502-x] [Medline: 29720146]
17. Mohamed MS, Zary M, Kafie C, Chilala CI, Bahukudumbi S, Foster N, et al. The impact of digital adherence technologies on health outcomes in tuberculosis: a systematic review and meta-analysis. medRxiv (forthcoming) 2024:1. [doi: 10.1101/2024.01.31.24302115]
18. Jerene D, Levy J, van Kalmthout K, van Rest J, McQuaid CF, Quaife M, et al. Effectiveness of digital adherence technologies in improving tuberculosis treatment outcomes in four countries: a pragmatic cluster randomised trial protocol. BMJ Open 2023 Mar 14;13(3):e068685 [FREE Full text] [doi: 10.1136/bmjopen-2022-068685] [Medline: 36918242]
19. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017;17(1):88 [FREE Full text] [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]
20. Thomas BE, Kumar JV, Onongaya C, Bhatt SN, Galivanche A, Periyasamy M, et al. Explaining differences in the acceptability of 99DOTS, a cell phone-based strategy for monitoring adherence to tuberculosis medications: qualitative study of patients and health care providers. JMIR Mhealth Uhealth 2020 Jul 31;8(7):e16634 [FREE Full text] [doi: 10.2196/16634] [Medline: 32735220]
21. Thomas BE, Kumar JV, Periyasamy M, Khandewale AS, Hephzibah Mercy J, Raj EM, et al. Acceptability of the medication event reminder monitor for promoting adherence to multidrug-resistant tuberculosis therapy in two Indian cities: qualitative study of patients and health care providers. J Med Internet Res 2021 Jun 10;23(6):e23294 [FREE Full text] [doi: 10.2196/23294] [Medline: 34110300]
22. Guzman K, Crowder R, Leddy A, Maraba N, Jennings L, Ahmed S, et al. Acceptability and feasibility of digital adherence technologies for drug-susceptible tuberculosis treatment supervision: a meta-analysis of implementation feedback. PLOS Digit Health 2023 Aug;2(8):e0000322 [FREE Full text] [doi: 10.1371/journal.pdig.0000322] [Medline: 37582066]
23. Tuberculosis profile: Philippines. World Health Organization. 2023. URL: https://worldhealthorg.shinyapps.io/tb_profiles/?_inputs_&entity_type=%22country%22&lan=%22EN%22&iso2=%22PH%22 [accessed 2023-10-30]
24. Romualdez AG. TB-DOTS in the Philippines: impact of decentralization and health sector reform. Bull World Health Organ 2007 May 1;85(5):399-400 [FREE Full text] [doi: 10.2471/blt.06.036285]
25. NTP Manual of Procedures 6th Edition. National Tuberculosis Control Program-Department of Health. 2021. URL: https://ntp.doh.gov.ph/download/ntp-mop-6th-edition/ [accessed 2024-06-21]
26. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
27. Houston S. Beyond social constructionism: critical realism and social work. Br J Social Work 2001;31(6):845-861 [FREE Full text] [doi: 10.1093/bjsw/31.6.845]
28. Sazzad HMS, McCredie L, Treloar C, Lloyd AR, Lafferty L. Violence and hepatitis C transmission in prison—a modified social ecological model. PLoS One 2020;15(12):e0243106 [FREE Full text] [doi: 10.1371/journal.pone.0243106] [Medline: 33259565]
29. McLeroy KR, Bibeau D, Steckler A, Glanz K. An ecological perspective on health promotion programs. Health Educ Q 1988;15(4):351-377. [doi: 10.1177/109019818801500401] [Medline: 3068205]
30. Rhodes T, Singer M, Bourgois P, Friedman SR, Strathdee SA. The social structural production of HIV risk among injecting drug users. Soc Sci Med 2005 Sep;61(5):1026-1244. [doi: 10.1016/j.socscimed.2004.12.024] [Medline: 15955404]
31. Mohammed S, Glennerster R, Khan A. Impact of a daily SMS medication reminder system on tuberculosis treatment outcomes: a randomized controlled trial. PLoS One 2016;11(11):e0162944 [FREE Full text] [doi: 10.1371/journal.pone.0162944] [Medline: 27802283]

32. Hoffman JA, Cunningham JR, Suleh AJ, Sundsmo A, Dekker D, Vago F, et al. Mobile direct observation treatment for tuberculosis patients: a technical feasibility pilot using mobile phones in Nairobi, Kenya. Am J Prev Med 2010 Jul;39(1):78-80. [doi: 10.1016/j.amepre.2010.02.018] [Medline: 20537846]
33. Leddy A, Ggita J, Berger C, Kityamuwesi A, Sanyu AN, Tinka LK, et al. Barriers and facilitators to implementing a digital adherence technology for tuberculosis treatment supervision in Uganda: qualitative study. J Med Internet Res 2023 May 30;25:e38828 [FREE Full text] [doi: 10.2196/38828] [Medline: 37252774]
34. Demissie M, Getahun H, Lindtjørn B. Community tuberculosis care through "TB clubs" in rural North Ethiopia. Soc Sci Med 2003 May;56(10):2009-2018. [doi: 10.1016/s0277-9536(02)00182-x] [Medline: 12697193]
35. Chaychoowong K, Watson R, Barrett D. Perceptions of stigma among pulmonary tuberculosis patients in Thailand, and the links to delays in accessing healthcare. J Infect Prev 2023 Mar;24(2):77-82. [doi: 10.1177/17571774231152720] [Medline: 36815061]
36. DeSanto D, Velen K, Lessells R, Makgopa S, Gumede D, Fielding K, et al. A qualitative exploration into the presence of TB stigmatization across three districts in South Africa. BMC Public Health 2023 Mar 15;23(1):504 [FREE Full text] [doi: 10.1186/s12889-023-15407-2] [Medline: 36922792]
37. Cremers AL, de Laat MM, Kapata N, Gerrets R, Klipstein-Grobusch K, Grobusch MP. Assessing the consequences of stigma for tuberculosis patients in urban Zambia. PLoS One 2015;10(3):e0119861 [FREE Full text] [doi: 10.1371/journal.pone.0119861] [Medline: 25806955]

Abbreviations

ASCENT: Adherence Support Coalition to End TB project
DAT: digital adherence technology
DoH: Department of Health
DOT: directly observed therapy
DS-TB: drug-sensitive tuberculosis
HCW: health care worker
KI: key informant
NTP: National Tuberculosis Control Program
TB: tuberculosis
WHO: World Health Organization

Edited by C Jacob; submitted 18.11.23; peer-reviewed by S Alghamdi, X Li, B Thies; comments to author 29.02.24; revised version received 25.04.24; accepted 01.05.24; published 23.07.24.

Please cite as:
Leung CL, Alacapa J, Tasca BG, Villanueva AD, Masulit S, Ignacio ML, Uy KN, Pell C, van Kalmthout K, Powers R, Fielding K, Jerene D
Digital Adherence Technologies and Differentiated Care for Tuberculosis Treatment and Their Acceptability Among Persons With Tuberculosis, Health Care Workers, and Key Informants in the Philippines: Qualitative Interview Study
JMIR Hum Factors 2024;11:e54117
URL: https://humanfactors.jmir.org/2024/1/e54117
doi:10.2196/54117
PMID:

Original Paper

Co-Designing Digital Health Intervention for Monitoring Medication and Consultation Among Transgender People in Underserved Communities: Collaborative Approach

Emmanuel Oluwatosin Oluokun[1*], PhD; Festus Fatai Adedoyin[1*], PhD; Huseyin Dogan[1*], PhD; Nan Jiang[1*], PhD

Department of Computing and Informatics, Faculty of Science and Technology, Bournemouth University, Poole, Dorset, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Emmanuel Oluwatosin Oluokun, PhD
Department of Computing and Informatics
Faculty of Science and Technology
Bournemouth University
Fern Barrow
Poole, Dorset, BH12 5BB
United Kingdom
Phone: 44 01202524111
Email: eoluokun@bournemouth.ac.uk

Abstract

Background: In many parts of the world, men who have sex with men and transgender individuals face criminalization and discrimination. As a result, they are less likely to seek medical help, despite experiencing higher rates of HIV/AIDS, mental health issues, and other health problems. Reaching key populations (KPs) with essential testing, care, and treatment services can be challenging, as they often have a higher likelihood of contracting and spreading the virus. They have limited access to antiretroviral (ARV) therapy (ART) services, which means that KPs may continue to serve as reservoirs for new HIV infections if they do not receive effective HIV programming. This ongoing issue complicates efforts to control the epidemic. Therefore, modeling a digital health system to track ARV medication access and use is crucial. This paper advocates for the use of digital interventions to manage the health of KPs in underserved regions, using Nigeria as a case study.

Objective: This study aims to assess digital health interventions for monitoring medication and consultations among transgender people in underserved communities. It also sought to determine whether a system exists that could support ART adherence in Nigeria. Additionally, the study evaluated design strategies to address privacy and confidentiality concerns, aiming to reduce nonadherence to ARV medications among KPs in Nigeria.

Methods: A qualitative approach was adopted for this research, involving a thematic analysis of information collected from interviews with clinicians and other health practitioners who work directly with these communities, as well as from an interactive (virtual) workshop.

Results: The findings from the thematic analysis indicate a need to increase attendance at ART therapy sessions through the implementation of an intensive care web app. Unlike previous solutions, this study highlights the importance of incorporating a reminder feature that integrates with an in-app telemedicine consultancy platform. This platform would facilitate discussions about client challenges, such as adverse drug effects, counseling sessions with clinical psychologists, and the impact of identity discrimination on mental health. Other data-driven health needs identified in the study are unique drug request nodes, client-led viral load calculators, remote requests, and drug delivery features within the web app. Participants also emphasized the importance of monitoring medication compliance and incorporating user feedback mechanisms, such as ratings and encouragement symbols (eg, stars, checkmarks), to motivate adherence.

Conclusions: The study concludes that technology-driven solutions could enhance ART adherence and reduce HIV transmission among transgender people. It also recommends that local governments and international organizations collaborate and invest in health management services that prioritize health needs over identity.

(JMIR Hum Factors 2024;11:e45826) doi:10.2196/45826

KEYWORDS

digital health; HIV/AIDS medication; data-driven health care; ART; transgender; LGTBQI+; interactive management

Introduction

Background

The HIV infection rate in Nigeria is the third highest in the world [1]. As of 2018, 1.9 million people in Nigeria were living with HIV/AIDS, according to the 2019 Nigeria National HIV/AIDS Indicator and Impact Survey [2]. HIV and AIDS are much more prevalent among incarcerated individuals, high-risk drug users, sex workers, transgender people, men who have sex with men (MSM), and people who inject drugs. Studies by the United Nations Office on Drugs and Crime found that 9% of people who inject drugs [3] and 2.8% of incarcerated individuals in Nigeria are living with HIV/AIDS [1]. These rates are significantly higher than the estimated 1.4% prevalence in the general population [1].

HIV/AIDS is undoubtedly a challenging condition, as it requires lifelong therapy and may lead to the emergence of new HIV-associated complications, as noted by Deeks et al [4]. Although combination antiretroviral (ARV) therapy (ART) can improve the health of patients infected with HIV, barriers to effective ARV treatment can negatively affect outcomes for those living with the disease [5]. HIV/AIDS is a highly stigmatized condition, and individuals with HIV/AIDS are more likely to be diagnosed with other health issues, such as poor mental health, psychiatric disorders, and substance use disorders. These additional, negatively perceived conditions can increase stigma and hinder adherence to care [6,7]. Therefore, it is crucial to explore the benefits of digital health interventions for monitoring medication and consultations among people living with HIV/AIDS.

It is crucial to provide key populations (KPs) with necessary testing, care, and treatment services, as they are at higher risk of contracting and spreading the virus. Unfortunately, these groups often have limited access to ART services. Without proper HIV programming, KPs may continue to serve as a reservoir for new infections, hindering efforts to contain the pandemic [8]. There is a need to expand knowledge on the use of digital health innovations in ART delivery. However, there is a lack of understanding regarding the ethical creation and use of these digital health solutions [9]. This study aims to cocreate a digital health solution to track ARV drug access among transgender people in underserved communities. It also seeks to determine if a system exists that can be utilized to support ARV adherence in Nigeria. The study also explored design strategies to address privacy and confidentiality concerns among KPs, aiming to reduce nonadherence to ARV medications in Nigeria.

This paper advocates for the role of digital interventions in managing the health of KPs in underserved regions, using Nigeria as a case study. The study makes several innovative contributions: first, it documents for the first time in the literature the challenges faced by health care professionals working with KPs in Nigeria; second, it introduces a trigger question on the role of digital health interventions through an interactive management (IM) approach; and third, it outlines the cocreation process with users in developing a digital solution, presenting this as an agenda for future research. The “Literature Review” section reviews the literature on ART and digital interventions, followed by the “Methods” section, which outlines the research approach and methods used in this study. The key findings (the “Results” section) and discussions (the “Discussion” section) are then presented, followed by vital recommendations and an agenda for future research in the “Conclusions” section.

Literature Review

ART has transformed HIV infection from a fatal illness into a manageable chronic condition [10]. ART can also reduce viral load (ie, levels of HIV RNA) and the risk of secondary transmission, establishing a new preventive paradigm where effective scaling of therapy could contribute to the end of AIDS [11]. Iyun [12] found that the majority of the 35 million individuals testing positive for HIV live in resource-limited settings. In 2016, an estimated 17 million of them were receiving ART, up from 1.3 million in 2006. Goals for 2020 projected treatment for an additional 20 million patients [13]. Although this expansion of ART is one of the greatest public health achievements of our time, significant challenges remain. Patients require near-perfect adherence of at least 95% to maintain undetectable viral loads and support immune system activity, making adherence to ART a considerable challenge.

In Nigeria, HIV prevalence rates among lesbian, gay, bisexual, transgender, queer, and intersex (LGBTQI) communities and MSM are up to 19 times higher than those in other populations [14]. Analyzing these groups is particularly challenging due to the limited number of studies and the insufficient funding allocated to this community. In a study by Liu et al [15], only 45% of MSM reported good adherence to ART. Low adherence rates are attributed to various issues, including HIV stigma, social exclusion, limited access to health care programs, fear of seeking care or being denied it, depression, and insufficient information about drug interactions between hormonal therapy and ART [16]. A recent meta-analysis [17] revealed that only 77% of sub-Saharan Africans on ARV medications adhered to the recommended dosage schedule. Overall, there is limited information available on the adherence levels achieved by medical facilities providing regular ART services.

Individuals who use ARV medication erratically may experience side effects, minimal benefits, and fewer treatment options in the future. It is crucial that every patient understands this before starting treatment. If a patient discontinues ART entirely, they will quickly lose any gains in immunity due to the continued spread of the virus and destruction of CD4+ cells [18]. It is essential to emphasize to patients that ART is a lifelong commitment. Effective patient education and adherence assessments require significant time and effort, but this investment is worthwhile. Clinical negligence occurs when a prescription is provided only at the initial visit without adequate adherence counseling, a practice that unfortunately remains common [19].

Transportation issues, such as difficulties reaching the nearest medical facility, combined with fears of job loss from taking time off work, have made consistent adherence to ART challenging [20-23]. Additionally, several interview-based studies have identified patients' attitudes toward ART as a factor contributing to the rejection of HIV/AIDS antiretroviral therapy (HAART) [24,25]. People with HIV/AIDS have expressed various concerns about HAART, including worries about side effects, the need for strict adherence, inconvenience and practical issues related to the regimen, mistrust of conventional medications, fears of long-term organ damage, and the belief that treatment is unnecessary in the absence of symptoms. Additionally, concerns about the impact of HAART on self-identity and the potential for treatment to reveal their HIV status contribute to these negative opinions [26,27]. These concerns highlight the need to explore digital health solutions.

The potential of digital interventions in ART has yet to be fully realized, partly because it is challenging to build a comprehensive body of knowledge to guide decisions about digital health solutions. Information and communications technology is utilized in digital health interventions to improve health outcomes [28]. Digital innovations can be categorized into noninternet technologies (eg, SMS text messages and phone calls) and internet-based technologies (eg, social media, mobile apps, and websites) [29]. Internet-based digital interventions allow users to publish content and share information on sensitive topics at any time and from any location, potentially minimizing the risk of unintended disclosure of private behaviors. The recent globalization of instant messaging platforms has laid the foundation for internet-based digital interventions. For example, a widely used instant messaging service, which includes nonsmartphone options, has 2 billion monthly active users across 180 countries [30]. SMS text messages and real-time medication monitoring through such platforms have been promoted to enhance ART adherence [31-34].

Correspondingly, the World Health Organization (WHO) and other agencies have recommended using digital technologies to deliver adherence interventions [35]. The rapid advancement of the technological landscape requires the continuous evolution and updating of digital interventions to improve adherence to ARV medications. These digital technologies can promote healthy behaviors; enhance treatment outcomes for chronic conditions such as diabetes, heart disease, and mental health issues; and provide remote access to effective treatments (eg, computerized cognitive behavioral therapy for patients with HIV) [36]. Digital interventions are often complex, involving multiple components and goals. These can include empowering users to learn more about their health, share experiences with others in similar situations, change perceptions and beliefs about health, evaluate and track health states or behaviors, adjust medication, identify health priorities, make informed treatment decisions, and enhance communication between patients and providers.

However, there have been reports of negative effects associated with digital health interventions. Research conducted in some parts of Africa revealed concerns about unintended HIV disclosure [37-40], while others anticipated stigma from SMS text message content that included terms such as "medication" and "HIV" [41,42]. Systematic evaluations [43,44] have also found that low-resource environments, poor internet connectivity, and the high cost of smartphones and their maintenance have contributed to the failure of digital health interventions. Logistics issues, such as frequent sharing of mobile phones among family members, intermittent electricity availability, and mobile phone malfunctions, were also highlighted [45]. These factors can complicate the adoption of digital innovations by clinic personnel and end users [46]. Given the recognized implementation challenges associated with digital innovations, further research is needed to identify and address barriers that may limit the acceptance of these interventions in practical settings.

Methods

Research Design

A qualitative approach was adopted for the research, utilizing thematic analysis of data collected from an IM workshop and an interview session. This approach was chosen because it facilitates in-depth probing and questioning of participants based on their responses, allowing both participants and researchers to explore underlying reasons and sentiments. The study aims to understand the perspectives of health professionals on how digital health care interventions can best enhance the monitoring of ARV medications among KPs. Interaction among focus group participants can often yield more insights than one-on-one interviews. Therefore, relevant data are collected through an IM workshop and a follow-up interview.

Ethics Approval

This study was approved by Bournemouth University (approval number 44608). The authors understand that their data may be used in an anonymized form by research teams to support other ethically approved research projects in the future, including future publications, reports, or presentations.

Population and Sample

Health care practitioners participating in the study included doctors, pharmacists, nurses, and clinical psychologists from Lagos State, Nigeria. Two sampling techniques were used: purposeful and convenience sampling. The purposeful sampling technique was used to select health care professionals working within the KP community. The convenience sampling technique was then used to select 1 doctor, 1 nurse, 1 pharmacist, and 1 clinical psychologist, with additional participants being social workers who work with the KP. The 2 sampling techniques were chosen because they help identify participants who are well-suited for the study [47,48]. Because of the extensive ethical clearance required, the study could not include all relevant participants. The Initiative for Equal Rights (TIERs) was selected as the venue for the IM workshop, while the interview was conducted via Zoom (Zoom Video Communications/Qumu Corporation) after securing participant agreements for both the workshop and interview. The study sample consisted of 13 participants in total, with all participating in the IM workshop and 7 participating in the interview.

Instrument

Initially, an open-ended survey was planned, accompanied by an interview guide with 5 questions aligned with the study's objectives. Instructions were provided on how to respond to the questions and an explanation was given about the features of digital health care interventions for monitoring ARV medications among KPs. This was done to ensure participants had a clear understanding of the study and how to provide their responses. The interview questions are crucial for gathering information on factors pertinent to the study's objectives.

The Procedure of Administration

A formal agreement was established with each contact person to schedule an interview following the IM workshop, conducted as a focused group discussion. The interviews lasted between 20 and 30 minutes, with responses recorded for transcription. Participants provided informed consent by signing agreement forms after carefully reviewing the participant information page, which included all relevant details about the project, as required by ethical standards. During the interview, 5 questions were posed, and 7 participants actively responded. Their answers were recorded, transcribed by a researcher (FFA), and reported accordingly.

Trustworthiness

Trustworthiness is a critical consideration in research as it allows researchers to demonstrate the value of findings beyond typical qualitative research parameters [47]. In this study, trustworthiness is aimed at reinforcing the significance of the findings. To achieve this, a pilot study was conducted to assess the reliability of the data. Additionally, an interview was conducted with 1 pharmacist and 1 doctor, and their responses were transcribed and returned to them for verification to ensure the accuracy of the content.

Interactive Management Session

Group work on complex issues is facilitated using computer assistance through IM [49]. IM can be viewed as a structured focus group method applied across various research fields, including cybersecurity. It uses human factor approaches [50] and requirements engineering [51] to support consensus decision-making through idea generation, structuring, and design. Typically, an IM session involves 8-12 participants who are knowledgeable about the topic and represent diverse viewpoints. The group usually convenes for 3-5 days, with follow-up sessions often occurring as needed. Before the working sessions, a detailed work plan is developed through collaboration between the workshop planner and an organizational representative. Participants are supported in generating, clarifying, and structuring concepts using well-established and effective approaches. The flowchart for conducting a successful IM session, as adopted for this research, is presented in Figure 1.

Figure 1. Interactive management flowchart.

The Planning Phase

Overview

The first and most crucial step in this phase is to understand the current situation. This involves defining the state, identifying the key actors, and drafting scope and context statements. These techniques help participants explore who is affected, what issues arise, and how they are impacted. By clearly defining the problem, participants can identify key questions that, when answered, could significantly expedite the development of effective solutions [49].

The Workshop Phase

Next is the workshop phase, where participants convene to address planning-related concerns and implement consensus decisions. The session focuses on 3 key concepts: context, content, and process [49]. The facilitator begins by guiding the discussions and providing context from the planning phase. The group then contextualizes the information through discussion and idea sharing. The facilitator manages the workshop flow to ensure discussions remain focused and participants make efficient use of their time.

Idea Writing

Nominal group technique and interpretive structural modeling are the techniques covered in the IM workshop. Participants start by responding to a trigger question through idea writing. After exchanging written ideas with others, additional ideas are integrated. The compiled information is then categorized and presented to the group. Following this, the nominal group technique is applied, where participants generate further ideas based on the enhanced understanding of the issue gained from the initial idea writing. Additionally, the workshop facilitates the editing and clarification of problem statements. Participants assign priority ratings to each idea. The final stage of the workshop aims to convert idea statements into objectives, which are then used to construct an interpretive structural model to reveal connections between different aspects of the issue [52].

The Follow-Up Phase

This stage marks the beginning of the solution implementation planning while putting the workshop's goals into practice. If it is discovered during this phase that the problem was misunderstood or if new issues have emerged that were not previously considered, a new planning phase would be initiated [49].

Thematic Analysis

A thematic analysis was conducted on the core questions asked during interviews with key stakeholders working with KPs. The data were collected between July 2022 and August 2022. The process and core issues identified from the thematic analysis are presented in Figure 2 and are derived from the interview questions listed in Table 1.

Figure 2. Thematic map capturing the core themes from the qualitative analysis. ARV: antiretroviral.

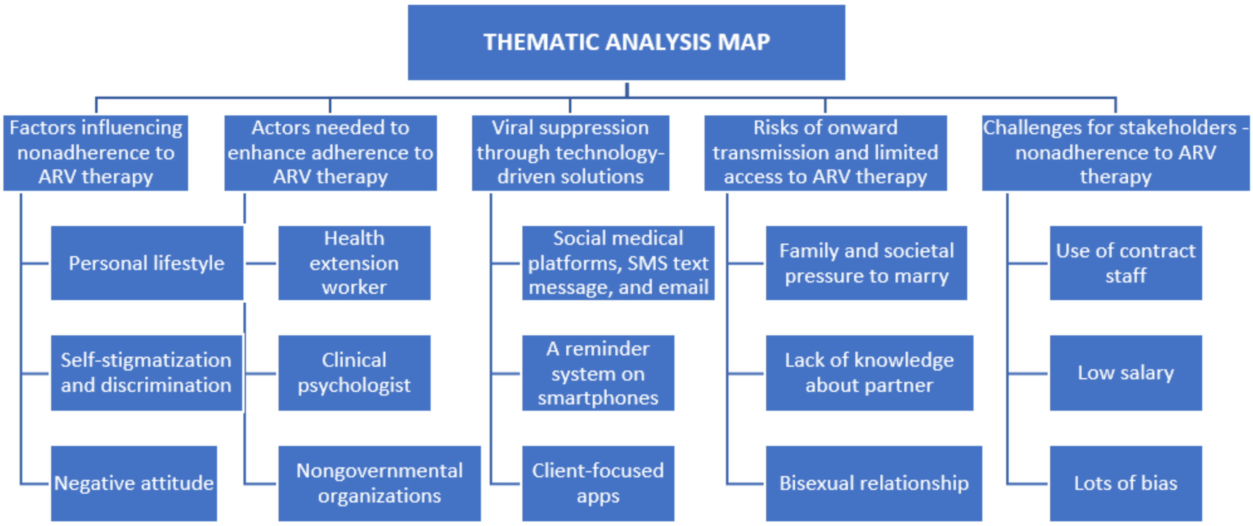


Table 1. Interview questions and codes for creating themes from interview responses.

Question number	Interview questions	Codes for creating themes from interview responses
1	What are the key factors influencing nonadherence to ARV[a] therapy among key populations?	Self-stigmatization, condition of health status, poverty, lack of sensitization and information, poor attitude, personal lifestyle, substance abuse, state of hopelessness, denial of diagnosis, cultural disposition, psychological distress, migration, logistics and financial constraint, privacy and confidentiality, the proximity of medical care facility, religious belief, hostile environment, drug side effect.
2	What category of actors do you consider relevant in the suppression of nonadherence to ARV therapy among key populations through technology-driven solutions?	International organizations, national organizations, NGOs[b], health practitioners, religious organizations/leaders, and health extension workers are the key actors that are relevant in the suppression of nonadherence to ARV therapy.
3	How can we maintain viral suppression through technology-driven solutions among key populations?	Social media, SMS text messages, email, Google Maps, a reminder system on smartphones, embedded app software powered with AI[c], client-focused apps, mobile apps, and interactive-driven systems are identified by the participants as a way of maintaining the viral suppression.
4	How do you consider the high risk of onward transmission and limited access to ARV therapy?	Key populations are not ready to share their problems with anyone, and at the same time, they engage in unprotected sex. The key populations also lack sufficient information to help prevent the spread of the virus, as well as awareness campaigns from key actors to address nonadherence to ARV therapy. Additionally, there is a lack of knowledge about one's partner, societal and family pressure to marry at all costs, and self-denial of a diagnosis.
5	What are the challenges associated with working with the key stakeholders in the suppression of nonadherence to ARV therapy?	The workers working with the NGOs are contract staff; therefore, their salaries are not encouraging. Lack of collaboration between and among stakeholders as well as a lack of passion to work and lots of bias.

[a]ARV: antiretroviral.

[b]NGO: nongovernmental organization.

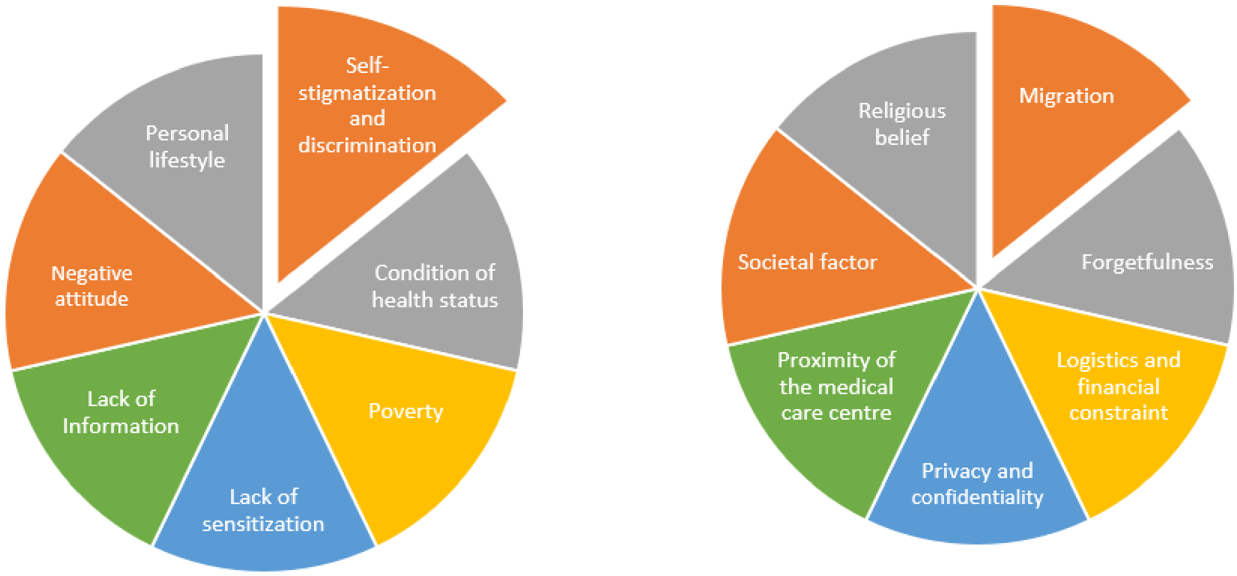
[c]AI: artificial intelligence.

Results

Factors Influencing Nonadherence to ART

The participants were asked about the key factors likely to influence nonadherence to ART among KPs. The findings revealed that factors such as self-stigmatization, health status, poverty, lack of sensitization and information, poor attitude, personal lifestyle, substance abuse, hopelessness, denial of diagnosis, cultural disposition, psychological distress, migration, logistical and financial constraints, privacy and confidentiality concerns, proximity to medical care facilities, religious beliefs, hostile environments, and drug side effects all contribute to nonadherence to ART. Participants provided various justifications for the factors influencing nonadherence to ART. However, self-stigmatization and discrimination, economic status, attitude, religious beliefs, and proximity to medical care facilities were ranked as the most significant factors, as shown in Figure 3.

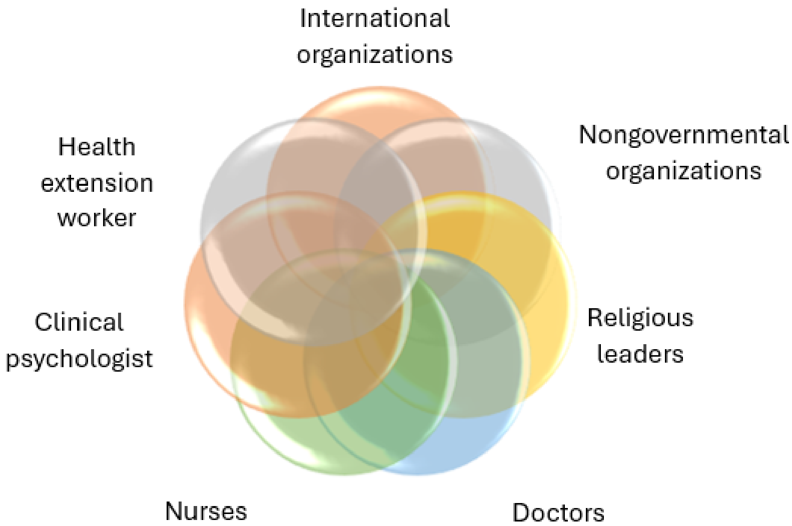
Figure 3. Thematic description of key factors influencing non-adherence to antiretroviral therapy.



Actors Needed to Enhance Adherence to ART

Participants were asked to identify and categorize the key actors relevant to addressing nonadherence to ART among KPs. The findings indicated that international organizations, national organizations, nongovernmental organizations (NGOs), health practitioners, religious organizations/leaders, and health extension workers are crucial actors. Their categorization and ranking are as follows: NGOs, international organizations, national organizations, doctors, nurses, clinical psychologists, and religious organizations. This is illustrated in Figure 4, highlighting the significant influence of international organizations in financing health care interventions for KPs in Nigeria.

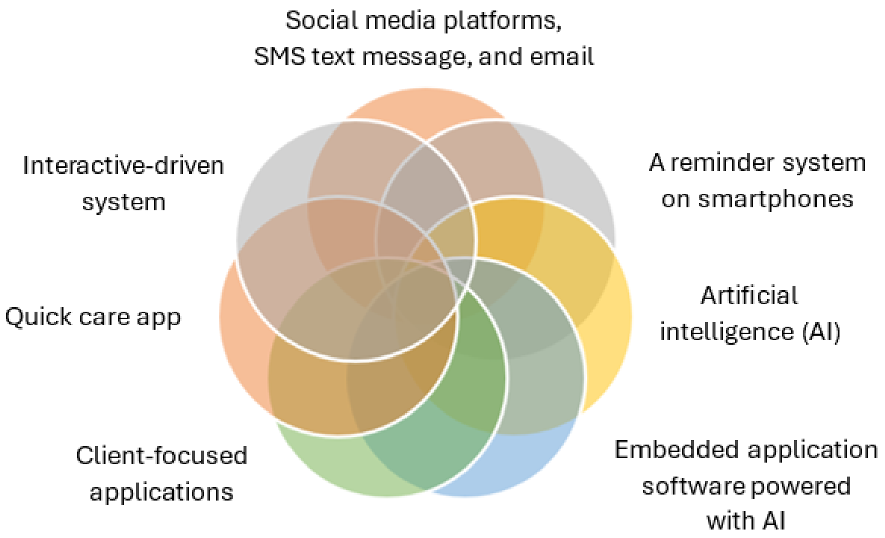
Figure 4. Thematic description of actors needed to enhance adherence to antiretroviral therapy.



Maintaining Viral Suppression Through Technology-Driven Solutions

Participants were asked to suggest technological solutions for maintaining viral suppression among KPs. The results revealed that social media, SMS text messages, email, Google Maps, smartphone reminder systems, artificial intelligence–powered app software, client-focused apps, mobile apps, and interactive systems were identified as effective methods for maintaining viral suppression, as shown in Figure 5.

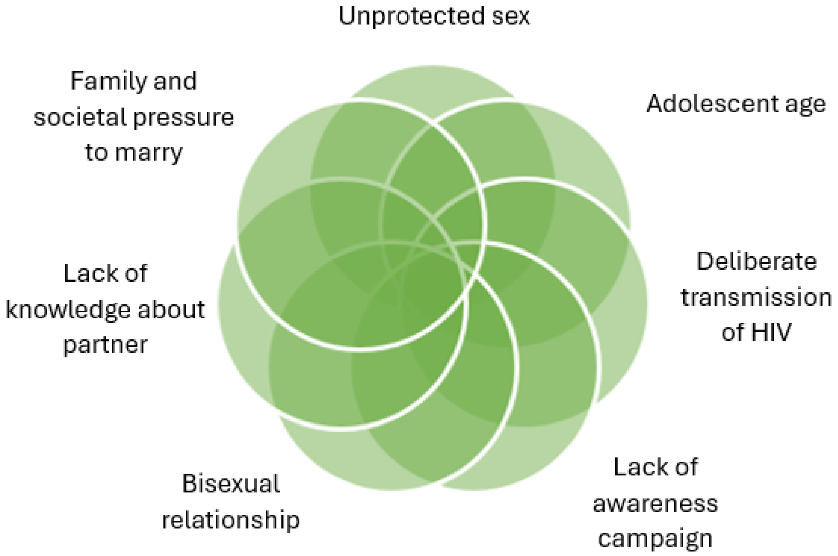
Figure 5. Thematic description of maintaining viral suppression through technology-driven solutions.



Risks of Onward Transmission and Limited Access to ART

Participants were asked about factors contributing to the high risk of onward transmission and limited access to ART. The findings indicated that many KPs are unwilling to share their issues, engage in unprotected sex, lack information to prevent virus spread, experience insufficient awareness campaigns, engage in bisexual relationships, have limited knowledge about partners, face family and societal pressure to marry, and self-deny their diagnosis, as presented in Figure 6.

Figure 6. Thematic description of risks of onward transmission and limited access to antiretroviral therapy.



Challenges With Stakeholders in the Suppression of Nonadherence to ART

Participants were asked to identify challenges in working with stakeholders to address nonadherence to ART. Findings, presented in Figure 7, highlight issues such as contract staff with inadequate salaries, lack of collaboration among stakeholders, lack of passion for the work, and prevalent biases.

Figure 7. Thematic description of challenges with stakeholders.

Results of the IM Workshop Session

The IM workshop (virtual) was facilitated by one of the authors (FFA) and took place in a conference room at Bournemouth University for the attendees' convenience. A total of 8 participants were selected based on the following criteria: they must be a member of the social care community, volunteer within the LGBTQ community in Lagos, and have expertise in HIV/AIDS care provision. The exclusion criteria for the workshop included health care workers not involved in ARV medication provision and care. Of the 8 participants, 7 participated in the Zoom session. To facilitate idea writing, a trigger question was provided for participants to record their responses and exchange them. As a result of time constraints imposed by participants' schedules, the phase of sharing ideas aloud began immediately. A computer paired with a projector was used to record and display the ideas.

The following are the trigger queries:

- What digital intervention can be used to improve ARV medication delivery and usage?
- How can we enhance viral suppression through technology-driven solutions?

Tables 2-5 present the ideas and responses generated from the trigger questions. These ideas were categorized as shown in Table 2, which outlines the different categories of responses. Table 3 highlights the significant concern for health care professionals regarding the development of a drug interaction monitoring system, as this was a major point of discussion among doctors, clinical psychologists, social workers, and other health care professionals working with KPs. The categories are ranked based on the number of ideas in each category. The IM session also identified the development of a web app, ideally named a KP Intensive Care App, and a medication use tracking system as crucial digital solutions or digital health care interventions.

Table 2. Results of idea generation.

Number	Idea
1	Development of the KP[a] Intensive Care App that would enable easy access to and delivery of ARV[b] medication.
2	A reminder feature to notify the client to take medication at scheduled times, and to remind them of upcoming appointments and drug pickups.
3	In-app telemedicine consultation to address client challenges, such as drug side effects and counseling needs.
4	Request for medication and viral load tests remotely, with delivery arranged through the app.
5	Pop-up messages to notify clients about potential drug interactions.
6	Avatar-guided animation videos demonstrating how to use medications, collect samples for viral load testing, and present samples within the app.
7	Section for booking appointments, including next drug pickup and doctor visits.
8	Monitor medication compliance and provide feedback for encouragement, such as stars or okay signs.
9	A feature that displays up-to-date tracking data, such as consultation records and viral load results, for both backend management and client benefits.
10	Development of a KP-friendly app, similar to a chatbot, to remind clients about medication schedules and clinic appointments.
11	Manufacturing medications or injectables with less frequent dosing to reduce pill burden.
12	Using dried blood spot collection for viral load testing can encourage clients, especially those with trypanophobia, to undergo viral load testing and adhere to their testing schedule.

[a]KP: key population.

[b]ARV: antiretroviral.

Table 3. Categorization of ideas.

Category	Ideas[a]	Ranking
KP[b] Intensive Care App	1, 3, 10	2
Viral load collection system	4, 9	3
Medication use tracking device	2, 6, 8	2
Drug interaction monitoring system	5, 7, 11, 12	1

[a]Refers to the categorization of results of idea generation in Table 2.
[b]KP: key population.

Additionally, the nominal group technique was used to enable participants to prioritize the top 5 issues from the generated ideas, as shown in Table 4. Participants ranked each idea from 1 (most important) to 5 (least important). The results indicate that the development of a KP Intensive Care App, which facilitates easy access to ARV medication, provides a digital platform for monitoring medication compliance, and offers motivational feedback (such as stars or okay signs), is considered the most important. Moreover, within the web app, features such as reminders for drug use and appointments, management of drug requests, calculation of viral load, and remote ordering and delivery were also ranked as significant.

Table 4. Participant's ranking of ideas.

Idea	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5	Total score
1	1	3	1	5	1	11
2	—[a]	5	2	1	2	10
3	3	—	—	—	—	3
4	—	1	—	—	—	1
5	—	—	—	5	—	5
6	4	—	3	—	2	9
7	—	4	5	—	5	14
8	1	—	1	—	—	2
9	—	2	—	—	—	2
10	—	—	—	4	—	4
11	2	5	—	—	3	10
12	—	3	—	5	—	8

[a]Not available.

According to Table 4, ideas 7, 1, 2, 11, and 6 are ranked as the top 5 based on the total score.

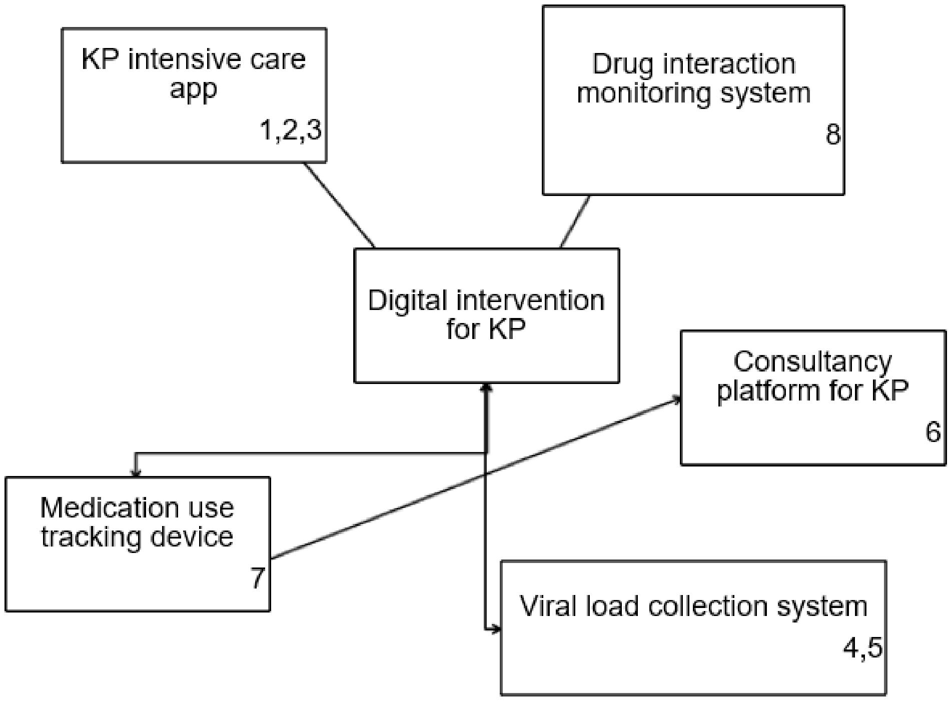
Table 5 displays the objective statements derived from the idea statements during the workshop. These objective statements were used to construct an interpretive structural model, as illustrated in Figure 8. The figure shows the grouping of similar objective statements, with numbers assigned to each box to reflect their categorization.

Table 5. Objective statements.

Number	Objective statements
1	Develop a KP[a]-focused intensive care app that provides easy access to and delivery of ARV[b] medication.
2	Include unique reminder features to notify clients to take their medication at scheduled times, as well as to remind them of appointments and drug pickups.
3	Develop an in-app telemedicine consultancy platform for clinical psychologists to conduct private sessions with KPs.
4	Incorporate a drug request framework that depends on the accurate calculation of viral load from the app.
5	Provide pop-up messages to alert clients about potential drug interactions.
6	Include avatar-guided demonstrations and animation videos on how to use medications.
7	Establish a partnership with HIV/AIDS medication manufacturers to explore options for less frequent usage or dosage of medicines.
8	Develop a systematic and noninvasive data collection approach to encourage clients to participate in testing and adhere to their treatment regimen.

[a]KP: key population.
[b]ARV: antiretroviral.

Figure 8. Interpretive structural model. KP: key population.



Discussion

Summary

The study focused on utilizing digital health care interventions for the consultation and monitoring of ARV medication among transgender individuals in underserved regions. To achieve the study's objectives, the following topics were investigated: factors influencing nonadherence to ART; key actors needed to improve adherence; methods to sustain viral suppression through technology-driven solutions; reasons for the high risk of onward transmission and limited access to ART; and challenges associated with collaborating with stakeholders to address nonadherence to viral treatment.

Principal Findings

Indeed, the study's findings revealed that several factors influence nonadherence to ART among MSM. These factors include self-stigmatization, health status, poverty, lack of awareness and information, poor attitude, denial of diagnosis, cultural disposition, psychological distress, migration, logistical and financial constraints, issues of privacy and confidentiality, proximity to medical facilities, and drug side effects. Indeed, these results align with a previously conducted study [15], which indicated that ART adherence was below average among MSM at risk for HIV infection in the United States. Several factors contribute to these low adherence rates, including HIV stigma, social exclusion, difficulties accessing health care programs, anxiety about seeking and potentially being denied care, depression, and a lack of knowledge about the interactions between hormonal therapy and various medications.

The findings also identified key players crucial for addressing nonadherence to ART, including international organizations, national organizations, NGOs, health professionals, religious leaders, and health extension workers. This outcome is supported by recent research conducted in South Africa, which concluded that a shortage of health professionals at medical centers is a significant barrier to adherence. This suggests that medical staff are essential actors in efforts to reduce nonadherence to ART [53]. Furthermore, this study aligns with the perspective of Murungi et al [54], who argued that religious leaders play a crucial role in HIV/AIDS prevention and ART adherence [55].

In addition, our study revealed how viral suppression can be maintained through technology-driven solutions for KPs. Participants identified several tools, including social media, SMS text messages, email, Google Maps, reminder systems on smartphones, artificial intelligence–powered embedded apps, client-focused apps, mobile apps, and interactive systems, as effective means for sustaining viral suppression among these groups. This aligns with Phan et al [56], who demonstrated the potential benefits of digital interventions in promoting adherence to health solutions among people living with HIV/AIDS.

Furthermore, our data revealed that many key demographics are unwilling to disclose their challenges while continuing to engage in unprotected sex. KPs also face a lack of information to help combat virus transmission, as well as insufficient awareness campaigns, bisexual relationships, inadequate knowledge about partners, family and societal pressure to marry at all costs, and self-denial of diagnosis. Several studies corroborate this finding. For instance, Cherutich et al [57] identified a lack of HIV status knowledge as a significant barrier to HIV prevention, care, and treatment activities. In Kenya, high rates of undiagnosed HIV infection are prevalent among gay, bisexual, and other MSM, as well as transgender women [58]. The African Union Commission [59] has demonstrated that high rates of child marriage often coincide with high rates of HIV infection in many countries. Acceptance of HIV status is critical for the effectiveness of HIV tests and related activities, whereas self-denial following a positive diagnosis can hinder adherence [60].

Our findings also identified several significant challenges among stakeholders, including insufficient salary payments, lack of collaboration between stakeholders, lack of motivation, and prevalent biases. Proper stakeholder education and accurate information can positively impact adherence, which supports the conclusions of this research [61].

Evaluating Relationship With Previous Findings

These findings align with the suggestions of Labrique et al [35], who indicated that the WHO and other international organizations advocate for the use of digital technologies to deliver adherence interventions and suppress viral infections. Previous studies on digital innovations in health care (eg, [22,23,29,62]) demonstrate the availability of various digital solutions. These include real-time medication monitoring and SMS text message reminders, which signal medication events, as well as real-time digital interventions through SMS text message reminders and remote monitoring solutions. The study also confirmed that adolescents are at high risk of transmitting HIV, partly because many individuals with the infection are reluctant to discuss their health openly, thereby contributing to its spread. Additionally, challenges in working with stakeholders include the prevalence of contract employees with unsatisfactory pay, lack of collaboration among stakeholders, a lack of passion for the job, and significant bias. These findings are strongly supported by the study conducted by Ledda et al [62].

Limitations

This study investigated digital health care interventions for monitoring ARV medications among Nigeria's KPs through empirical analysis. It is important to acknowledge some limitations, as no research is without flaws. The study focused exclusively on Lagos, and thus, expanding it to include other regions across Nigeria could provide valuable and practical insights applicable to health care facilities nationwide. Furthermore, the study's sample size was limited to a maximum of 7 key medical actors, and data were collected using an IM approach. Future research should consider conducting mixed methods or strictly quantitative assessments to provide a more comprehensive evaluation.

Conclusions

The study investigates health professionals' perceptions of digital health care interventions for monitoring ARV medications among KPs in Nigeria. The study concluded that technology-driven solutions could improve adherence to ART and reduce HIV transmission among KPs. Based on the findings of this study, the following recommendations are made: The government should implement improved policies to encourage positive attitudinal changes among stakeholders, promoting the use of technology-driven solutions to maintain viral suppression among the HIV population in the country; additionally, local governments and international agencies should facilitate awareness campaigns to improve ART sessions among KPs. Strategies should also be developed to help Nigerian health facilities integrate suitable channels for cocreating digital solutions that will increase attendance at ART sessions.

Acknowledgments

The authors acknowledge the Department of Computing and Informatics at Bournemouth University for their 2022 Quality Research funding, which supported this study.

Conflicts of Interest

None declared.

References

1. United Nations Office on Drugs and Crime. Drug use in Nigeria. United Nations Office on Drugs and Crime. 2018. URL: https://www.unodc.org/documents/data-and-analysis/statistics/Drugs/Drug_Use_Survey_Nigeria_2019_BOOK.pdf [accessed 2023-05-05]
2. Federal Ministry of Health, Nigeria. Nigeria HIV/AIDS Indicator and Impact Survey (NAIIS) 2018: preliminary findings. NAIIS. 2018. URL: https://www.naiis.ng/resource/factsheet/NAIIS%20PA%20NATIONAL%20FACTSHEET%20FINAL.pdf [accessed 2023-05-05]
3. United Nations Office on Drugs and Crime (UNODC), National Agency for the Control of AIDS, Nigerian Prisons Service. The national situation and needs assessment of HIV and AIDS, drug use and related health services in Nigerian prisons; study report. UNODC. 2023 May. URL: https://www.unodc.org/conig/uploads/documents/HIV_Prisons_Full_Study_Report_.pdf [accessed 2023-05-05]
4. Deeks S, Lewin S, Havlir D. The end of AIDS: HIV infection as a chronic disease. The Lancet 2013 Nov;382(9903):1525-1533 [FREE Full text] [doi: 10.1016/s0140-6736(13)61809-7]
5. Chenneville T, Gabbidon K, Hanson P, Holyfield C. The impact of COVID-19 on HIV treatment and research: a call to action. Int J Environ Res Public Health 2020 Jun 24;17(12):4548 [FREE Full text] [doi: 10.3390/ijerph17124548] [Medline: 32599783]
6. St Clair-Sullivan N, Mwamba C, Whetham J, Bolton Moore C, Darking M, Vera J. Barriers to HIV care and adherence for young people living with HIV in Zambia and mHealth. Mhealth 2019;5:45 [FREE Full text] [doi: 10.21037/mhealth.2019.09.02] [Medline: 31620472]
7. Henny KD, Wilkes AL, McDonald CM, Denson DJ, Neumann MS. A rapid review of eHealth interventions addressing the continuum of HIV care (2007-2017). AIDS Behav 2018 Jan;22(1):43-63 [FREE Full text] [doi: 10.1007/s10461-017-1923-2] [Medline: 28983684]
8. United Nations Programme on HIV/AIDS. Inequalities are blocking the end of the AIDS pandemic, say UN. United Nations. 2022. URL: https://www.un.org/africarenewal/magazine/december-2022/inequalities-are-blocking-end-aids-pandemic-say-un [accessed 2023-05-05]
9. Perez Vallejos E, Koene A, Carter CJ, Hunt D, Woodard C, Urquhart L, et al. Accessing online data for youth mental health research: meeting the ethical challenges. Philos Technol 2017 Oct 12;32(1):87-110 [FREE Full text] [doi: 10.1007/s13347-017-0286-y]
10. Paterson DL, Swindells S, Mohr J, Brester M, Vergis EN, Squier C, et al. Adherence to protease inhibitor therapy and outcomes in patients with HIV infection. Ann Intern Med 2000 Jul 04;133(1):21-30. [doi: 10.7326/0003-4819-133-1-200007040-00004] [Medline: 10877736]
11. Haberer JE, Sabin L, Amico KR, Orrell C, Galárraga O, Tsai AC, et al. Improving antiretroviral therapy adherence in resource‐limited settings at scale: a discussion of interventions and recommendations. Journal of the International AIDS Society 2017 Mar 22;20(1):21371. [doi: 10.7448/ias.20.1.21371]
12. Iyun V. Universal antiretroviral therapy (ART) for infants and young children living with HIV: assessing the effect of guideline changes on ART initiation characteristics and treatment outcomes in resource-limited settings. Doctoral Thesis. University of Cape Town. 2022. URL: http://hdl.handle.net/11427/36782 [accessed 2023-05-05]
13. Marsh K, Eaton J, Mahy M, Sabin K, Autenrieth C, Wanyeki I. Global, regional and country-level 90-90-90 estimates for 2018: assessing progress towards the 2020 target. AIDS (London, England) 2019;33(Suppl 3):S213. [doi: 10.1097/qad.0000000000002355]
14. Beyrer C, Baral SD, Collins C, Richardson ET, Sullivan PS, Sanchez J, et al. The global response to HIV in men who have sex with men. The Lancet 2016 Jul;388(10040):198-206. [doi: 10.1016/s0140-6736(16)30781-4]
15. Liu AY, Hessol NA, Vittinghoff E, Amico KR, Kroboth E, Fuchs J, et al. Medication adherence among men who have sex with men at risk for HIV infection in the United States: implications for pre-exposure prophylaxis implementation. AIDS Patient Care STDS 2014 Dec;28(12):622-627. [doi: 10.1089/apc.2014.0195] [Medline: 25396706]
16. Iacob SA, Iacob DG, Jugulete G. Improving the adherence to antiretroviral therapy, a difficult but essential task for a successful HIV treatment-clinical points of view and practical considerations. Front Pharmacol 2017 Nov 23;8:831 [FREE Full text] [doi: 10.3389/fphar.2017.00831] [Medline: 29218008]
17. Mills EJ, Nachega JB, Buchan I, Orbinski J, Attaran A, Singh S, et al. Adherence to antiretroviral therapy in sub-Saharan Africa and North America: a meta-analysis. JAMA 2006 Aug 09;296(6):679-690. [doi: 10.1001/jama.296.6.679] [Medline: 16896111]
18. Sonia M. Measurement of immune response to hepatitis B vaccine in HIV 1 infected and healthy children. Doctoral dissertation. Christian Medical College, Vellore. 2019. URL: https://core.ac.uk/reader/235670722 [accessed 2024-08-29]

19. Obare F, Abuya T, Matanda D, Bellows B. Assessing the community-level impact of a decade of user fee policy shifts on health facility deliveries in Kenya, 2003-2014. Int J Equity Health 2018 May 25;17(1):65-13 [FREE Full text] [doi: 10.1186/s12939-018-0774-4] [Medline: 29801485]
20. Wilkinson R, Garden E, Nanyonga RC, Squires A, Nakaggwa F, Schwartz JI, et al. Causes of medication non-adherence and the acceptability of support strategies for people with hypertension in Uganda: a qualitative study. Int J Nurs Stud 2022 Mar;126:104143. [doi: 10.1016/j.ijnurstu.2021.104143] [Medline: 34953374]
21. Evans KN, Hassan R, Townes A, Buchacz K, Smith DK. The potential of telecommunication technology to address racial/ethnic disparities in HIV PrEP awareness, uptake, adherence, and persistence in care: a review. AIDS Behav 2022 Dec 25;26(12):3878-3888 [FREE Full text] [doi: 10.1007/s10461-022-03715-4] [Medline: 35614366]
22. Bennett DE, Myatt M, Bertagnolio S, Sutherland D, Gilks CF. Recommendations for surveillance of transmitted HIV drug resistance in countries scaling up antiretroviral treatment. Antivir Ther 2008;13 Suppl 2:25-36. [Medline: 18575189]
23. Mills E, Nachega J, Bangsberg D, Singh S, Rachlis B, Wu P, et al. Adherence to HAART: a systematic review of developed and developing nation patient-reported barriers and facilitators. PLoS Med 2006 Nov;3(11):e438 [FREE Full text] [doi: 10.1371/journal.pmed.0030438] [Medline: 17121449]
24. Kasumu L. Knowledge and attitude towards antiretroviral therapy and adherence pattern of HIV patients in southwest Nigeria. International Journal of Infection Control 2014;10(3):1-8 [FREE Full text] [doi: 10.3396/ijic.v10i3.12971]
25. Nutor J, Slaughter-Acey J, Marquez S, DiMaria-Ghalili RA, Momplaisir F, Oladimeji KE, et al. Impact of attitudes and beliefs on antiretroviral treatment adherence intention among HIV-positive pregnant and breastfeeding women in Zambia. BMC Public Health 2020 Sep 16;20(1):1410-1419 [FREE Full text] [doi: 10.1186/s12889-020-09505-8] [Medline: 32938415]
26. Proctor V, Tesfa A, Tompkins DC. Barriers to adherence to highly active antiretroviral therapy as expressed by people living with HIV/AIDS. AIDS Patient Care STDS 1999 Sep;13(9):535-544 [FREE Full text] [doi: 10.1089/apc.1999.13.535] [Medline: 10813033]
27. Bukenya D, Mayanja B, Nakamanya S, Muhumuza R, Seeley J. What causes non-adherence among some individuals on long term antiretroviral therapy? Experiences of individuals with poor viral suppression in Uganda. AIDS Res Ther 2019 Jan 21;16(1):2-9 [FREE Full text] [doi: 10.1186/s12981-018-0214-y] [Medline: 30665440]
28. World Health Organization (WHO). WHO guideline: recommendations on digital interventions for health system strengthening. World Health Organization. Geneva, Switzerland: WHO; 2019. URL: https://iris.who.int/bitstream/handle/10665/311980/WHO-RHR-19.10-eng.pdf?sequence=1&isAllowed=y [accessed 2023-05-05]
29. Jiao K, Wang C, Liao M, Ma J, Kang D, Tang W, et al. A differentiated digital intervention to improve antiretroviral therapy adherence among men who have sex with men living with HIV in China: a randomized controlled trial. BMC Med 2022 Oct 10;20(1):341-315 [FREE Full text] [doi: 10.1186/s12916-022-02538-3] [Medline: 36210434]
30. Weaver N, Roy A, Martinez S, Gomanie N. How WhatsApp is transforming health care services and empowering health workers in low- and middle-income countries. New York, NY: IEEE; 2022 Presented at: 2022 IEEE Global Humanitarian Technology Conference (GHTC); September 8-11, 2022; Santa Clara, CA p. 234-241. [doi: 10.1109/ghtc55712.2022.9911048]
31. Nyoni T, Sallah Y, Okumu M, Byansi W, Lipsey K, Small E. The effectiveness of treatment supporter interventions in antiretroviral treatment adherence in sub-Saharan Africa: a systematic review and meta-Analysis. AIDS Care 2020 May;32(Suppl 2):214-227. [doi: 10.1080/09540121.2020.1742870] [Medline: 32196385]
32. Pintye J, Rogers Z, Kinuthia J, Mugwanya KK, Abuna F, Lagat H, et al. Two-way short message service (SMS) communication may increase pre-exposure prophylaxis continuation and adherence among pregnant and postpartum women in Kenya. Glob Health Sci Pract 2020 Mar 05;8(1):55-67. [doi: 10.9745/ghsp-d-19-00347]
33. Haberer J, Musiimenta A, Atukunda E, Musinguzi N, Wyatt M, Ware N. SMS reminders plus real-time adherence monitoring improve adherence to antiretroviral therapy in rural Uganda. AIDS 2016 May;30(8):1295-1299. [doi: 10.1097/qad.0000000000001021]
34. Bionghi N, Daftary A, Maharaj B, Msibi Z, Amico KR, Friedland G, et al. Pilot evaluation of a second-generation electronic pill box for adherence to Bedaquiline and antiretroviral therapy in drug-resistant TB/HIV co-infected patients in KwaZulu-Natal, South Africa. BMC Infect Dis 2018 Apr 11;18(1):171-179 [FREE Full text] [doi: 10.1186/s12879-018-3080-2] [Medline: 29642874]
35. Labrique AB, Wadhwani C, Williams KA, Lamptey P, Hesp C, Luk R, et al. Best practices in scaling digital health in low and middle income countries. Global Health 2018 Nov 03;14(1):103-108 [FREE Full text] [doi: 10.1186/s12992-018-0424-z] [Medline: 30390686]
36. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
37. Nordberg B, Gabriel E, Were E, Kaguiri E, Ekström AM, Kågesten A, et al. Social concerns related to HIV status disclosure and participation in the prevention of mother-to-child transmission of HIV care among pregnant women in Kenya. BMC Pregnancy Childbirth 2020 Apr 16;20(1):225-229 [FREE Full text] [doi: 10.1186/s12884-020-02907-x] [Medline: 32299386]
38. Ngonzi J, Mugyenyi G, Kivunike M, Mugisha J, Salongo W, Masembe S, et al. Frequency of HIV status disclosure, associated factors and outcomes among HIV positive pregnant women at Mbarara Regional Referral Hospital, southwestern Uganda. Pan Afr Med J 2019;32:200. [doi: 10.11604/pamj.2019.32.200.12541] [Medline: 31312312]

39. Ismail N, Matillya N, Ratansi R, Mbekenga C. Barriers to timely disclosure of HIV serostatus: a qualitative study at care and treatment centers in Dar es Salaam, Tanzania. PLoS One 2021;16(8):e0256537 [FREE Full text] [doi: 10.1371/journal.pone.0256537] [Medline: 34437597]
40. Madiba S, Josiah U. Perceived stigma and fear of unintended disclosure are barriers in medication adherence in adolescents with perinatal HIV in Botswana: a qualitative study. Biomed Res Int 2019;2019:9623159 [FREE Full text] [doi: 10.1155/2019/9623159] [Medline: 31886271]
41. Yang F, Shah D, Tahk A, Vjorn O, Dietz S, Pe-Romashko K, et al. mHealth and social mediation: mobile support among stigmatized people living with HIV and substance use disorder. New Media & Society 2023 Apr 17;25(4):702-731 [FREE Full text] [doi: 10.1177/14614448231158653]
42. Iacoella F, Gassmann F, Tirivayi N. Impact of mobile phones on HIV public stigma: a cross-sectional and pseudo-panel analysis from Ghana. BMJ Open 2022 Nov 08;12(11):e062594 [FREE Full text] [doi: 10.1136/bmjopen-2022-062594] [Medline: 36351734]
43. Sarna A, Saraswati LR, Okal J, Matheka J, Owuor D, Singh RJ, et al. Cell phone counseling improves retention of mothers with HIV infection in care and infant HIV testing in Kisumu, Kenya: a randomized controlled study. Glob Health Sci Pract 2019 May 29;7(2):171-188. [doi: 10.9745/ghsp-d-18-00241]
44. Kebaetse M. The use of mobile devices in HIV and AIDS communication: opportunities for Botswana. Botswana Journal of African Studies 2016;30(2):298-311 [FREE Full text]
45. Ngowi K, Pima F, Mmbaga BT, Aarnoutse RE, Reiss P, Nieuwkerk PT, et al. “I Wish to Continue Receiving the Reminder Short Messaging Service”: A mixed methods study on the acceptability of digital adherence tools among adults living with HIV on antiretroviral treatment in Tanzania. PPA 2021 Mar;Volume 15:559-568. [doi: 10.2147/ppa.s290079]
46. Reid M, Dhar S, Cary M, Liang P, Thompson J, Gabaitiri L. Opinions and attitudes of participants in a randomized controlled trial examining the efficacy of SMS reminders to enhance antiretroviral adherence: a cross-sectional survey. JAIDS Journal of Acquired Immune Deficiency Syndromes 2014;65(2):e86-e88. [doi: 10.1097/qai.0b013e3182a9c72b]
47. Weyant E. Research design: qualitative, quantitative, and mixed methods approaches, 5th Edition. Journal of Electronic Resources in Medical Libraries 2022 Mar 14;19(1-2):54-55. [doi: 10.1080/15424065.2022.2046231]
48. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm Policy Ment Health 2015 Sep;42(5):533-544 [FREE Full text] [doi: 10.1007/s10488-013-0528-y] [Medline: 24193818]
49. Warfield J, Cardenas AR. A Handbook of Interactive Management. Palm Harbor, FL: Ajar Publishing Company; 2002.
50. Ward J, Dogan H, Apeh E, Mylonas A. Using human factor approaches to an organisation's bring your own device scheme. In: Human Aspects of Information Security, Privacy and Trust (Vol. 10292). Cham, Switzerland: Springer; 2017:396-413.
51. Dogan H, Henshaw M. A ‘soft’ approach to TLM requirements capture to support through-life management. Berlin, Germany: Springer; 2009 Presented at: Knowledge Science, Engineering and Management (KSEM 2009); November 25-27, 2009; Vienna, Austria p. 458-469. [doi: 10.1007/978-3-642-10488-6_44]
52. Attri R, Dev N, Sharma V. Interpretive structural modelling (ISM) approach: an Overview. Res J Management Sci 2013 Feb;2(2):3-8 [FREE Full text]
53. Simelane P, Simelane M, Amoateng AY. Barriers and facilitators to adherence for antiretroviral therapy: the perspectives of patients from a wellness center in the Mpumalanga Province, South Africa. Afr Health Sci 2022 Sep;22(3):455-462 [FREE Full text] [doi: 10.4314/ahs.v22i3.49] [Medline: 36910408]
54. Murungi T, Kunihira I, Oyella P, Mugerwa M, Gift P, Aceng M, et al. The role of religious leaders on the use of HIV/AIDS prevention strategies among young people (15-24) in Lira district, Uganda. PLoS One 2022;17(10):e0276801 [FREE Full text] [doi: 10.1371/journal.pone.0276801] [Medline: 36301999]
55. Azia IN, Carelse S, Nyembezi A, Mukumbang FC. Role of religious beliefs in adherence to antiretroviral therapy in the Cape Town metropole: a study protocol. BMJ Open 2022 Sep 14;12(9):e062464 [FREE Full text] [doi: 10.1136/bmjopen-2022-062464] [Medline: 36104142]
56. Phan JM, Kim S, Linh, Cosimi LA, Pollack TM. Telehealth interventions for HIV in low- and middle-income countries. Curr HIV/AIDS Rep 2022 Dec 26;19(6):600-609 [FREE Full text] [doi: 10.1007/s11904-022-00630-0] [Medline: 36156183]
57. Cherutich P, Kaiser R, Galbraith J, Williamson J, Shiraishi RW, Ngare C, KAIS Study Group. Lack of knowledge of HIV status a major barrier to HIV prevention, care and treatment efforts in Kenya: results from a nationally representative study. PLoS One 2012;7(5):e36797 [FREE Full text] [doi: 10.1371/journal.pone.0036797] [Medline: 22574226]
58. Dijkstra M, Mohamed K, Kigoro A, Mumba T, Mahmoud S, Wesonga A, et al. Peer mobilization and human immunodeficiency virus (HIV) partner notification services among gay, bisexual, and other men who have sex with men and transgender women in coastal Kenya identified a high number of undiagnosed HIV infections. Open Forum Infect Dis 2021 Jun;8(6):ofab219 [FREE Full text] [doi: 10.1093/ofid/ofab219] [Medline: 34113688]
59. African Union Commission. Ending child marriage and stopping the spread of HIV opportunities and challenges for action. African Union Commission. 2016. URL: https://au.int/sites/default/files/documents/31788-doc-desk_review_on_the_linkages_of_child_marriage_and_hiv.pdf [accessed 2023-05-16]

60. Horter S, Thabede Z, Dlamini V, Bernays S, Stringer B, Mazibuko S, et al. "Life is so easy on ART, once you accept it": acceptance, denial and linkage to HIV care in Shiselweni, Swaziland. Soc Sci Med 2017 Mar;176:52-59 [FREE Full text] [doi: 10.1016/j.socscimed.2017.01.006] [Medline: 28129547]
61. Peña S, Bravo M, Tomás M, Martínez P, Guillén C, Jiménez-Ruiz I. Factors affecting adherence to antiretroviral treatment in women with HIV: a sequential mixed model research design. Enfermeria Global 2021;20(2):18-34 [FREE Full text] [doi: 10.6018/eglobal.437711]
62. Ledda C, Cicciù F, Puglisi B, Ramaci T, Nunnari G, Rapisarda V. Attitude of health care workers (HCWs) toward patients affected by HIV/AIDS and drug users: a cross-sectional study. Int J Environ Res Public Health 2017 Mar 09;14(3):284 [FREE Full text] [doi: 10.3390/ijerph14030284] [Medline: 28282937]

Abbreviations

ART: antiretroviral therapy
ARV: antiretroviral
HAART: HIV/AIDS antiretroviral therapy
IM: interactive management
KP: key population
LGBTQI: lesbian, gay, bisexual, transgender, queer, and intersex
MSM: men who have sex with men
NGO: nongovernmental organization
TIERs: The Initiative for Equal Rights
WHO: World Health Organization

Edited by A Kushniruk, Y Quintana; submitted 18.01.23; peer-reviewed by G Rebchook, C Mbotwa; comments to author 15.04.23; revised version received 30.06.23; accepted 11.07.24; published 12.09.24.

Please cite as:
Oluokun EO, Adedoyin FF, Dogan H, Jiang N
Co-Designing Digital Health Intervention for Monitoring Medication and Consultation Among Transgender People in Underserved Communities: Collaborative Approach
JMIR Hum Factors 2024;11:e45826
URL: https://humanfactors.jmir.org/2024/1/e45826
doi:10.2196/45826
PMID:

Original Paper

Older Adults' Perspectives and Experiences With Digital Health in Singapore: Qualitative Study

Qiao Ying Leong[1,2*], MPH; V Vien Lee[1,2*], PhD; Wei Ying Ng[1,2], BSocSci; Smrithi Vijayakumar[1,2], PhD; Ni Yin Lau[1,2], BSocSci; Ingela Mauritzon[3], MSc; Agata Blasiak[1,2,4,5], PhD; Dean Ho[1,2,4,5,6], PhD

[1]The Institute for Digital Medicine (WisDM), Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[2]The N.1 Institute for Health, National University of Singapore, Singapore, Singapore

[3]IVV Labs AB, Lund, Sweden

[4]Department of Biomedical Engineering, National University of Singapore, Singapore, Singapore

[5]Department of Pharmacology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[6]The Bia-Echo Asia Centre for Reproductive Longevity and Equality, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[*]these authors contributed equally

Corresponding Author:
Dean Ho, PhD
The Institute for Digital Medicine (WisDM), Yong Loo Lin School of Medicine, National University of Singapore
Centre for Life Sciences, 28 Medical Drive, #05-COR
Singapore, 117456
Singapore
Phone: 65 6601 7515
Email: biedh@nus.edu.sg

Abstract

Background: Technology use among older adults is increasingly common. Even though there is potential in leveraging technology to help them manage their health, only a small fraction of them use it for health-related purposes.

Objective: This study seeks to understand the perspectives of and experiences with digital health (DH) among older adults in Singapore.

Methods: A total of 16 participants (age range 60-80 years; n=11, 69% female) were interviewed for approximately an hour (range 27-64 minutes) about their health, DH use, and DH experiences. The interviews were recorded, transcribed verbatim, and thematically analyzed.

Results: Five main themes emerged from the interview: support in developing DH literacy, credibility, cost and benefit considerations, intrinsic drive to be healthy, and telehealth. Older adults need support in familiarizing themselves with DH. When considering DH options, older adults often relied on credible sources and preferred DH to be free. Monetary incentives were brought up as motivators. The intrinsic drive to live longer and healthily was expressed to be a huge encouragement to use DH to help obtain health-related knowledge and achieve healthy living goals. The idea of telehealth was also appealing among older adults but was seen to be more suited for individuals who have issues accessing a physical clinic.

Conclusions: Our findings offer insights into the various aspects that matter to older adults in the adoption of DH, which in turn can help reshape their health-seeking behavior and lifestyle. As such, policy makers and DH implementors are encouraged to take these into consideration and align their strategies accordingly.

(JMIR Hum Factors 2024;11:e58641) doi:10.2196/58641

KEYWORDS

digital health; gerontology; geriatrics; elder; aging; Singapore; qualitative; mHealth; mobile health; experience; technology use; interview; perspective; acceptance; technology adoption

Introduction

The global population of older adults aged 65 years and older is expected to increase in the coming years. In 2015, there was an estimated 8.5% of individuals aged 65 years and older globally, and this proportion has been projected to grow to 16.7% by 2050 [1]. In Singapore, the rapid increase of an aging population is evident given the growth of the proportion of older residents from 11.7% in 2013 to 19.1% in 2023 [2]. As the median age of Singapore continues to increase (median age is 43 years in 2023), it is projected that 24.1% of the Singaporean population will be aged 65 years and older by 2030 [2]. While increased longevity can be attributed to the success of medical advancements and improvements in public health practices [3], longevity is often accompanied by an increased risk of chronic diseases and a decrease in physical and cognitive function [4]. Beyond disease burden, multiple chronic diseases in older adults have been associated with greater health care use and costs [5]. Thus, it further contributes to the increasing load on the health care system and patients themselves, which is evident in Singapore as the acute hospital admissions [6] and duration for hospitalization episodes [7] were both reported to be rising.

To meet the health demands of a growing aging population presented within Singapore, the application of digital health (DH) to encourage these healthy behaviors serves as a valuable tool to assist in the journey toward a healthy aging population. DH for an aging population can include wearable devices for health monitoring, telemedicine, and mobile phone apps [8]. While the potential of DH for health behavioral change is still being tested, findings to date have shown promising results in promoting health behaviors and improving health-related outcomes [9-11]. Such an approach can serve as an invaluable tool to assist older adults in their journey toward healthy aging through preventive health, active aging programs, and care services. Accordingly, Singapore's 2023 action plan for successful aging aims to empower seniors to take charge of their physical and mental well-being through the engagement of regular group exercises, adopt a healthy and balanced diet, and go for regular health screenings and follow-ups [12]. Nevertheless, the success of a digital tool is dependent on its uptake and sustained use by the intended population. Despite a high proportion of older adults in Singapore owning at least 1 digital device (38.1% have 1 digital device and 55.82% have 2 or more digital devices), only a small number of older adults use their devices for functional purposes such as telehealth services (10.5%) [13].

Expanding the use of DH to older adults will benefit from significant input from older adults themselves to ensure that the tool is designed around their needs, capabilities, and preferences [8]. Through engaging older adults during the design process, DH innovators can better understand what drives DH use, which can then inform downstream design considerations and implementation strategies when rolling out the tool. While Singaporean older adults consider DH useful in health management, they do not consider health care–related technology critical for their daily lives [14]. Furthermore, there is generally a strong preference for direct contact with health care professionals among this group [15].

This study aimed to explore Singaporean older adults' perspectives of DH and their experiences with existing DH tools. Through this, the study sought to understand what matters to Singaporean older adults when using DH to encourage improved health behaviors.

Methods

Recruitment

We recruited individuals through purposive sampling from the community in Singapore through outreach sessions and social media platforms (ie, Telegram and Facebook). Interested participants responded to advertisements by reaching out to the study team via email or text message, and they were provided with the participant information sheet and informed consent form. Participants were screened based on the following inclusion criteria: (1) age of 60 years and older and (2) have experience using health-related technologies. Recruitment took place over a period of 1 year (November 2021 to October 2022). None of the participants refused to participate or dropped out after consenting to the study.

Data Collection

This study used semistructured interviews to understand the older adults' perspectives of and experiences with DH in Singapore. Each interview was approximately 60 minutes and conducted either in person or via Zoom depending on the participants' preference. In-person interviews were conducted in a quiet public space (eg, conference room, cafes, and void decks). All interviews were conducted in English with at least 2 researchers from the data collection team—QYL, VVL, SV, WYN, and NYL—present. The all-female data collection team consists of 2 postdoctoral fellows and 3 research assistants (with bachelor's degrees). All researchers who collected interviews had prior training in qualitative data collection and analyses. To minimize interview bias, a standardized interview protocol and guide were used (Multimedia Appendix 1). All researchers in the data collection team underwent pilot sessions before the interviews. Peer feedback was provided during the pilot sessions. Relationships with the participants prior to the interviews were established through email, call, or messages. Participants were told about the researchers' involvement in the study and the reasons for doing the research. At the beginning of the session, informed consent was obtained, and all participants consented to audio recording of the interview for transcription purposes. Open-ended questions were asked during the interview to understand the participants' health care journey, their perspectives, and their experiences with DH. This paper focuses on the DH aspect of the interviews. At the end of the session, all participants were reimbursed for their time. All data collected, including signed consent forms and interview recordings, were deidentified, encrypted, and stored in a secure database. No repeat interviews were conducted, and transcripts were not returned to participants for comments.

Data collection ended when data saturation was achieved, and the sample had reached maximum variation by age, gender, and ethnicity.

Data Analysis

All of the interview recordings were transcribed verbatim and thematically analyzed using the method described by Braun and Clarke [16]. Each transcript was randomly assigned to a member of the data collection team for inductive coding. All coders familiarized themselves with all interview transcripts and initial coding was conducted through a process of descriptively labeling segments of data. Each coder independently analyzed their assigned transcript. As the coding process progressed, the data collection team consistently reviewed and agreed on the coded data collectively and subsequently resolved any discrepancies through discussion. Categorization of the codes was conducted by QYL, who grouped the labeled data into categories and subsequently into broader themes. Data saturation was achieved when no new theme was identified. Grouping of themes was tracked on a master Excel workbook, and data saturation was confirmed through discussion with all researchers who were involved in data collection and analysis. The final codes and themes were based on the discussions and iterations with VVL, and a final review with the other members of the data collection team. The transcribing and analyzing steps were conducted using Microsoft Word and Excel respectively [17].

To ensure credibility and confirmability, reflexivity was applied throughout the data collection process. Team discussions were conducted at the end of each week. During team discussions, the researcher's interaction with participants and thoughts on the interview was discussed and noted down. Furthermore, peer debriefing was conducted throughout the interview and coding process to obtain feedback from other members of the data collection team and minimize personal biases. As the study team had specific interests in DH, it may have influenced the team to engage with participants about DH in a positive manner. During each discussion, the study team made note of the positives and negatives in relation to DH and were consistently reminded to prompt participants about both sides. To ensure transferability and dependability, sampling strategies, research procedures, and thick descriptions are provided in the manuscript.

Ethical Considerations

The National University of Singapore Institutional Review Board approved this study (NUS-IRB reference: NUS-IRB-2021-12). All participants provided informed consent and were informed of their right to withdraw from the study at any time. All data collected from participants were de-identified. Participants were reimbursed SGD 30 (US $22.60) for their time and participation.

Results

Participant Demographics

The study sample consisted of 16 participants, with an age range of 60 to 80 (mean age 66.9, SD 5.9) years. They were of Chinese (9/16, 56%), Indian (5/16, 31%), and Malay (2/16, 13%) descent and were fluent in English. A total of 11 (69%) participants were female. Hypertension, hyperlipidemia, diabetes, and cardiovascular diseases were some common medical conditions reported by the participants. Other medical conditions brought up less frequently included history of breast cancer, psychosis, ulcerative colitis, and shingles. Participant demographic data are presented in Multimedia Appendix 2.

During the interview, participants described their experiences with a variety of technologies. These technologies, consisting of mobile apps, web-based platforms, and DH devices, were used for multiple purposes as listed in Table 1.

Table 1. Technology use of participants.

Category	Technologies used
General	Zoom, Google, YouTube, Lazada App, Singtel App, and Amazon Alexa
Social media	Facebook, WhatsApp, and Telegram
Health-specific mobile apps and webpages	HealthXchange.sg, Healthy365 App, ActiveSG App, SAFRA Mobile App, HealthHub App, NUHS App, and Mayo Clinic
Health-specific devices	Bluetooth-connected Glycoleap App, Abbott Freestyle Glucose Monitoring device, FitBit, blood pressure monitoring device, and oximeter

Interview Data

Five themes emerged from the interview data. The themes consist of support in developing DH literacy, credibility, cost and benefit considerations, intrinsic drive to be healthy, and telehealth.

Support in Developing DH Literacy

Overview

The first theme—support in developing DH literacy—refers to the endeavor to improve DH literacy among older adults. This support builds upon the foundation laid by the existing awareness of DH through education and accessibility. The participants often learned about the existence of DH by word of mouth through family, friends, and health care professionals. Another way in which DH was made aware among older adults was through promotional materials disseminated through mass media namely newspapers and social media platforms (eg, Facebook).

Education

At the initial stage of DH adoption, education was crucial to ensure participants familiarized themselves with the DH interface, functionality, and navigation. Besides familiarization, education provided the support needed to sustain DH use. The overview of how the older adults might go from not knowing to knowing was described by participant 15, "Now the old people also trying to catch up, but someone [has] to educate them right...how to use it. Nowadays, there are the workshop[s] for elderly, how to use the app[s], how to use Tiktok, Facebook and all that. So, it [will] take some time to...give them teaching, [it] will take time [for them] to learn."

When asked about thoughts on creating awareness and bridging that to actual DH use through education, participant 5 described, "I think it involves...some outreach, educational programs, so people will come to know." The participants added that outreach sessions, education programs, and workshops at community centers and hospitals are good starting points. Participant 16 shared that community centers were an ideal point of contact "because they come to the community centre to spend time sometime[s]" but also specified that this option was only available for "those who can walk." Upon further elaboration, the same participant also shared, "The best place [to spread DH awareness] is the hospital. Hospital is the only place you can get that in...The nurses can educate them that these avenues are also open."

These sessions should offer guidance and troubleshooting support through step-by-step instructions and video tutorials. Each session should preferably be one-to-one and in person "to get seniors used to the concept...[and] to explain to them...it's non-threatening, it's helpful," as shared by participant 11. With repeated familiarization over time and easier access to avenues for education, older adults could become more receptive to DH. However, this process can be very time-consuming as it can be repetitive in nature and would require immense patience and time.

Accessibility

Accessibility reflects the ability of older adults to reach and use DH. Even though there was an intrinsic drive to learn and explore DH platforms available, participants expressed discomfort over unfamiliarity and poorer cognitive abilities. Participant 10 described, "No chance to learn [about DH]. If there is opportunity, then we pick here and there. So, sometimes we are struggling with it...The young ones will be very good. You need to start playing since young. For us...[we are] a bit slower, memory also weaker. So, what we have done earlier, we may not be able to remember. So, it's very difficult to do it on your own to try to learn." Concerns about slower cognition and poorer memory led to lower confidence in their ability to use DH, thereby creating barriers to DH adoption.

Support and guidance from "someone at home or...some neighbours" could act as a push factor "to help them [the older adults] understand certain things [DH platforms]" to overcome the aforementioned barriers to DH adoption, as shared by participant 11. Support beyond their own family members was especially important as pointed out by participant 6, as "the singles [elderly]...have no one to help them, they can go to the digital ambassadors." These digital ambassadors could potentially be the younger generation. To illustrate this point, participant 11 shared, "I got [tech] problem here. How? Anybody know? Things like that. And if you have a handphone, you look for a younger colleague...So, sometimes you can [ask them], 'help me'."

Credibility of DH Platforms

What and Who Do They Trust

Trust emerged to be essential in the health management of older adults. Although the information was described to be readily available, the participants expressed the need to sieve through it and had to rely on trusted sources to assist in their decision-making regarding their own health. Rather than having just 1 trusted individual or organization, it was common for older adults to seek health-related information or DH platform recommendations from multiple sources such as social networks or health care professionals. For instance, participant 1 expressed, "I always trust my friend. What they tell me to do, usually I will do it" while adding on, "If doctor say that I need this, of course I will trust the doctor". The mention of recommendations from friends emerged repeatedly. This was attributed to the trust in their friends' testimonials to ascertain the reliability of DH platforms. In participant 15's case, concern from friends encouraged them to take their health seriously and that included adopting DH to manage their health. "I just wanted to try out...They are so concerned [of me], they say, you are putting on weight...you must check up your health...So, your friends are so concerned about you...why not I take control for myself, that's why I take it seriously."

While information or suggestions from doctors or health care institutions such as hospitals and clinics were perceived as trusted sources, other individuals or organizations that made up the trusted circle of the older adults also included family members, friends, neighbors, governmental organizations (eg, Ministry of Health) and higher learning or research institutions (eg, National University of Singapore).

What Matters in Establishing the Trust

The DH recommendations from health care professionals were built on the existing doctor-patient relationship. Several considerations were brought up as crucial in forging and strengthening that bond. For instance, participant 7 expressed, "I think if the doctor has been with me for quite a while and I trust the doctor, it should not be an issue". Beyond that, the type of artificial intelligence used in the DH, the length of medical practice as well as the level of expertise (eg, junior doctors and consultants) of doctors recommending the DH also mattered while forging this sense of trust. Participant 7 further expanded, "It depends on the doctor himself, the level of expertise. I mean you can't compare between a houseman or a student doctor, to a consultant, right? So, to have my level of assurance that I can trust them fully, I think it depends on the level of expertise and the experience—the experience of the doctor."

Apart from the source of DH recommendations, the mode of information delivery was also crucial. For example, vital pieces of information to establish trust could be delivered in the form of statistics. Participant 14 described a possible way of conveying this information, "Based on your precedence [medical history] and all that kind of thing, [the DH] will tell them [the users], how many percent has actually...done this and they have recovered...Some stats based on the recovery rate, that kind of thing...I think that helps."

Furthermore, there were other considerations that cemented this trust. For example, participant 9 mentioned, "I'll ask him [the doctor], is there any risk [of] like you go on an app and do it yourself? That way I don't mind". The participants also emphasized the need for DH platforms to be validated and substantiated by evidence in order to be deemed as trustworthy tools in health management.

What Makes Them Lose Trust

"Having to see the doctor personally, and he diagnose...and do it for us, that's all I want. I would rather do that," said participant 9, who felt that DH platforms were not as trustworthy as doctors due to the lack of human touch. Cybersecurity concerns also emerged as a factor that would erode trust in DH use. For instance, participant 13 expressed, "I don't find [it] comfortable...when you go to the app, [and] they ask for Singpass...sometime, you may click the wrong button or you go to URL, they might take...all your information...then from there phish."

Cost and Benefit Considerations

Types of Incentives

Two types of incentives were described by the older adults as follows: monetary incentives and results from DH engagement.

"Number one of course monetary...Even if you give me an NTUC [a local supermarket chain] voucher, it is still money", said participant 10. The participants described that the use of gamification and financial incentives while using the DH platforms had been an effective and creative way for them to engage in DH use. The distribution of free DH devices, such as wearables, was also quoted to have similar motivating effects as financial incentives. For example, ActiveSG (Singapore government health promotion program) has 100 Singapore dollars preloaded to incentivize the onboarding of potential users and promote sustained use. As described by participant 6, "If you don't learn...it's forfeited from you. Then you are forced to learn."

Without incentives, the participants were more inclined to use DH platforms simply as a means of tracking health metrics. After the initial engagement, observable health outcomes from their DH use seemed to be the driver for sustained use. To illustrate this point, participant 12 described, "I have to see the result...maybe after a month or two. If they say do this, do that, you will lose your weight, I have to see whether it happens. After that, yes, I'll use. But I don't mind taking the first step, yes. But after that time, if I feel it's of no use, I will stop it."

What Incentives Enable Them to Do

Two health-seeking behaviors consistently emerged in the data when incentives were involved—staying active and opting for healthier dietary options.

Regarding the use of incentives to encourage users to stay active, participant 6 mentioned, "They put some codes around different places for you to go and find and scan...If I am passing by [and] I know that there is a place to scan to get points, then I just make a detour...So, they got all these challenges inside [the app] to keep you more active and maybe more mentally alert."

Furthermore, the participants were highly motivated to opt for healthier food options when incentivized by a point system through a DH platform. For example, participant 7 expressed, "That's how I accumulated [points], this QR code from buying food...with the healthier choice logo."

Cost

Generally, participants preferred DH use to be free. Participant 6 shared, "[The DH] must be available free, if you ask people to pay, they will be put off" and further elaborated, "I go for all the free things. Because this cost, sometimes I don't feel is justified." While there was a clear preference for free DH platforms, some participants expressed that they did not mind paying a one-off or subscription fee, but only if the cost was reasonable and the DH proved to be beneficial.

Intrinsic Drive to be Healthy

Overview

"Health is already a motivating factor, nobody wants to be unhealthy." Participant 6 succinctly summarized that this intrinsic drive, along with the aspiration to live longer and healthier, motivated older adults to adopt health-seeking behaviors. DH platforms were described to have the potential to help participants in their pursuit of realizing their healthy living goals and gaining knowledge out of necessity, or curiosity.

Healthy Living Goals

Participant 11 emphasized that addressing health-related concerns should be approached holistically - catering to both physical and mental well-being. While concomitant conditions were common, participant 7 shared, "You must treat the patient as a whole, not compartmentalize them into health issues like mental, diabetes, heart." The idea of personalized health-related tracking also appealed to older adults. For instance, participant 10 shared that diabetic patients were motivated to use glucometers to understand how different food choices affected their blood sugar levels. Furthermore, the relevance of the materials offered by DH platforms heavily dictated their use among older adults. Although the idea of relevance varied according to each older adult's area of interest (eg, diet, cultural practices, sleep, and medical concerns), there was a consensus that DH platforms available were too based on Western culture, too general, and lacked the touch of personalization. When seeking knowledge-based content, participant 8 shared, " [The DH should be] a one-to-one consultation and give you information that is relevant to you, specific to you, not so general that I might as well go Google you know". This highlighted the preference for more personalized content over generic information. Another example of relevance was based on the idea of localization. Participant 14 described that "a lot of [DH platforms] are...catered to Western kind of...audience." Accordingly, they may not be relatable to the Asian context and users. These examples demonstrate that irrelevance can impact the perceived relatability of DH platforms.

The participants expressed various aspects where DH can assist in health management and attainment of healthy living goals. First, DH platforms should have provided feedback especially if gamification was incorporated. Such feedback could be a report of their performances based on a benchmark, or actionable recommendations based on individual health concerns. Second, participant 16 suggested that DH platforms could serve as the "one tool for common disease[s] like diabetes or high blood pressure." Third, nudging as a form of reminder, as described by participant 15, could be "good, very helpful, like someone taking care and telling me...you must take this, you must do this."

Knowledge

While achieving one's health-related goals is important, participants also expressed that this pursuit needed to be supplemented by knowledge. Most described using DH out of necessity or curiosity. Participant 11 expressed, "I think that the onus is on me to do my own reading and find things also, to be more knowledgeable about my own issues." This motivation for knowledge was not only described to be driven by the goodness of oneself, but it also extended to their partners. Such motivation was described by participant 12 who mentioned, "My husband is diabetic, so I try to look for items which is better for him".

As such, the older adults generally appreciated health information and recommendations deployed through DH as supportive of their aspirations to live healthily. For example, participant 7 shared, "The [web-based] articles from SingHealth (health system in Singapore), when I got my treatment for my oral health...so they started giving me all these articles every now and then which I would read up." Participant 8 also described, "Now, [the] internet [is] so convenient. A lot of YouTube [videos] advise how to...take care of diabetes...fasting...that kind of thing. So, I try everything."

Beyond knowledge-based information, DH was also described as an ideal avenue to broadcast events that support healthy and active living. For example, participant 15 shared that happenings of interest including "a weekend activity...group walking activity, or running, simple exercise, Tai Chi...all kinds of simple exercises" could be broadcasted through DH, which could subsequently enable "most [in] my age group [to] get together and then become more active, then you get to know people". A similar sentiment was expressed by participant 1, "I will use that [DH] if they got social [events], they will tell me, where can [I] go."

Furthermore, the accessibility of web-based information also enabled health-seeking behaviors, such as in the case of participant 15 who shared, "During the COVID time, I was afraid to take the COVID injection, so I read through more [on websites], now I'm happy [to take the injection]."

Convenience of Telehealth

The appeal of telehealth is centered around the convenience that it offers. Telehealth not only allowed the participants to seek medical attention remotely, but it also eliminated the need and hassle of traveling to a clinic or health care institution. Participant 5 shared that their interaction with the polyclinic through telehealth made him "feel better because [there is] no need to walk all the way there [the polyclinic]."

Furthermore, the prompt responses by the medical team also made telehealth an ideal feature in DH. Participant 15 described, "It's a good idea that nowadays...to interact with the phone, and anything you can talk faster. I think during the COVID time, it started very well...I got COVID also, so they consult me to tell me [via] this thing [telehealth], then advise me the medication, they send me through phone, then they deliver to my house. It's quite [a] good idea." This also demonstrates how the recent COVID-19 pandemic accelerated the shift in favor of telehealth adoption.

Nevertheless, despite the boons that telehealth offers, the participants generally still preferred to consult the doctors in person. However, they also acknowledged that under some circumstances, telehealth might be more ideal. For example, participant 1 shared, "Last time is okay...[I] don't mind going out. But nowadays, [I am] usually very scared to go out, especially for the elderly because we are more prone to the disease. So, [it] is good that...I make appointment and zoom doctor. [The doctor] tell us, your results...you must take care, then is good. Better than we travel there for the results." Other groups of individuals who were identified to potentially benefit from telehealth included those who "are really very sick...and are not able to move", as shared by participant 12.

Discussion

While it is usual for older adults to use technology, using DH is not as common [18]. With the added challenges stemming from the natural progression of aging—slower cognition [19,20] and poorer memory [21]—overcoming the learning curve of DH can be daunting. Awareness and accessibility form the foundation to overcome the learning curve of DH, which in and of itself is a multi-step process. The building blocks leading to DH adoption and subsequently sustained use can be facilitated through various factors, some of which are unique to the Singaporean context, and will be discussed in this section.

While awareness of DH can be achieved through DH exposure, it is not always sufficient for obtaining access to DH. Accessibility issues expressed by the participants in this study generally centered around the need for support and assistance to help them to familiarize themselves and use the different aspects of DH (eg, interface, navigation, and functionality). Whether it was contextualized learning materials targeted to the masses or personalized guidance by DH ambassadors, participants described education as a necessity to overcome the steep learning curve—a barrier for DH uptake. Similar findings are echoed in a review that reported difficulty in technology use as the most common patient-level barrier to DH technology uptake, while facilitators included empowerment, education, and training sessions [22]. Overcoming this barrier is a mammoth effort, which was also echoed by the older adults in this study since it would require significant resources (ie, manpower, time, and funding) to ensure continual DH support ranging from step-by-step guidance to troubleshooting expertise. This understanding from our study, along with the existing qualitative work that provides invaluable insights regarding the general perception [14,22,23] and the user preferences [24] of DH among the older adults in Singapore, are critical in informing policymakers and DH implementers, and in aligning strategies that promote DH adoption for its potential to be realized and benefits reaped.

Trust repeatedly emerges as an important factor in technology adoption, including in other populations such as women [25,26], Korean Americans [27], and patients with cancer [28]. Like the findings in this study, individuals in those studies tended to seek health-related information from multiple sources—often from health care professionals, family, and friends [26,29], which were reported to be credible. Different sources of information are used to compare, evaluate, and triangulate the information as a means of filtering the vast body of data available. The older adults in this study expressed trust in physicians and preferred to cross-check health information with them. Interestingly, even though the older adults expressed trust in physicians, several nuances mattered in establishing this trust. For instance, the older adults in this study expressed that if a physician was recommending a novel technology, they were more comfortable following the recommendation if the physician had vast experience in the technology and was a seasoned medical practitioner (eg consultant over a junior doctor). Such nuanced behavior associated with trust was also reported in a study involving Singaporean older women [29], where information regarding health-related products was not taken at face value but was closely judged. Furthermore, it is worthwhile considering the influences of ethnic norms, linguistic capabilities, and level of education in instilling trust in DH among older adults in a culturally and racially diverse country like Singapore. In 2 studies—1 in Singapore [29] and 1 in the United States [27]—ethnic language, dialects, and linguistic capabilities might have played a role in establishing trust. Additionally, it is worthwhile noting that Singaporeans are generally trusting and responsive towards most government-related directives. This is especially evident in the COVID-19 vaccination uptake during the recent pandemic. Vaccination status was shown to be positively predicted by trust in formal sources (ie, government services, local television, and radio), but not by trust in informal sources (ie, family and friends) [30].

Older adults in this study emphasized the importance of having both extrinsic and intrinsic motivation to drive DH adoption and sustained use. Extrinsically, older adults mentioned financial incentives and tangible feedback about the results from their behavior change as drivers that would encourage greater DH use. This is in line with one of Singapore's nationwide mobile health physical activity programs, The National Steps Challenge deployed via the Healthy365 (Health Promotion Board) app, which was able to successfully reach about 26% of the Singapore population (ie, 1.3 million out of 5.0 million as of 2019) [31]. Key factors that were cited to contribute towards the success of the program's reach include monetary reward in exchange for Healthpoints (collected through steps) and free fitness wristbands offered to participants [31]. Beyond the need for external motivators, older adults in this study also demonstrated an intrinsic drive to be healthy and highlighted the potential of DH in aiding them to achieve health-related goals through education and actionable recommendations. This need for both intrinsic and extrinsic motivation in DH tools targeted towards older adults was also described in a recent qualitative systematic review, which emphasized the need for users to first be motivated to make a change in their life with the use of extrinsic motivation to aid in device adoption and value adding towards the users' life [32]. Accordingly, DH that is being developed to induce healthy behavior change in older adults should consider tapping into both types of motivation to strengthen adoption and engagement rates.

When asked about telehealth, older adults in this study were generally receptive and brought up convenience as its main benefit. While the COVID-19 pandemic may have facilitated the shift in perception in favor of telehealth in the eyes of older adults, participants in this study still preferred in-person consultation. The preference for in-person consultation was also reported in another Singaporean qualitative study exploring the attitudes of older Singaporeans towards telehealth [33]. Furthermore, the overall awareness among Singaporean older adults about the range of telehealth services available remains low with more than 50% of participants not knowing about telehealth [33]. A consideration suggested by older adults in this study was to focus telehealth efforts on individuals who would benefit from telehealth the most (eg, immunocompromised individuals and individuals with mobility concerns). Targeted strategies to reach older adults who face barriers to in-person consultations could ensure that the technology is being used by groups of older adults whose health care experience may be enhanced the most through the adoption of telehealth.

Due to the linguistic constraints of the data collection team, the study only recruited older adults who were English speakers. As such, older adults whose first language was not English (eg Mandarin, Malay, Tamil, Hokkien, and Cantonese) were excluded. Accordingly, the study may not capture nuances that may be specific to non-English speakers. Furthermore, the recruitment strategies of using social media platforms and community outreach might miss segments of the population not active in these areas. For instance, older adults with lower socioeconomic status or socially isolated may not be captured through these recruitment strategies. Nevertheless, this study was able to explore the perspectives of older adults from the 3 major ethnicities in Singapore (Chinese, Indian, and Malay) with varying levels of health conditions. As the aim and the design of this study were to understand the perspectives of older adults in Singapore through a qualitative study, the findings contribute to the knowledge about DH in the Singapore context. To further strengthen the findings from this study, next steps can include the use of data triangulation to corroborate current findings using a different data source (eg, focus groups or observational data), and increasing the sample size or interview length to enhance comparability and generalizability of the study findings. Moreover, strategies toward the development and implementation of DH should be context-specific, and an application of the findings outside of Singapore may require additional validation.

The process of aging creates an inevitable burden not only on an individual but also on the health care system and society. A potential way to help ease this burden is to harness the potential of technology, specifically through DH. This study demonstrated the various perspectives of and experiences with DH among older adults in Singapore. DH developers and implementors are encouraged to take these into consideration to improve DH adoption and implementation within this population, and to align their strategies accordingly.

Acknowledgments

DH gratefully acknowledges funding from the Institute for Digital Medicine (WisDM) Translational Research Programme at the Yong Loo Lin School of Medicine. The funder played no role in the study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

Conflicts of Interest

AB and DH are co-inventors or previously filed pending patents on artificial intelligence-based therapy development. DH is a shareholder of KYAN Therapeutics, which has licensed intellectual property pertaining to artificial intelligence-based drug development. IM is a co-founder and shareholder of IVV Labs AB, a Swedish lab diagnostics company. All other authors declare no financial or non-financial competing interests.

Multimedia Appendix 1
Study recruitment and interview protocol.
[DOCX File , 26 KB - humanfactors_v11i1e58641_app1.docx]

Multimedia Appendix 2
Demographic data of participants.
[DOCX File , 21 KB - humanfactors_v11i1e58641_app2.docx]

References

1. Xie L, Zhang S, Xin M, Zhu M, Lu W, Mo PK. Electronic health literacy and health-related outcomes among older adults: a systematic review. Prev Med 2022;157:106997. [doi: 10.1016/j.ypmed.2022.106997] [Medline: 35189203]
2. Population in brief 2023. National Population and Talent Division. 2023. URL: https://www.population.gov.sg/files/media-centre/publications/population-in-brief-2023.pdf [accessed 2024-10-18]
3. Costa D. Causes of improving health and longevity at older ages: a review of the explanations. In: Increasing Longevity: Causes, Trends, and Prospects. Rome, Italy: Università degli Studi di Roma “La Sapienza”; 2005:21-38.
4. Jaul E, Barron J. Age-related diseases and clinical and public health implications for the 85 years old and over population. Front Public Health 2017;5:335 [FREE Full text] [doi: 10.3389/fpubh.2017.00335] [Medline: 29312916]
5. Lehnert T, Heider D, Leicht H, Heinrich S, Corrieri S, Luppa M, et al. Review: health care utilization and costs of elderly persons with multiple chronic conditions. Med Care Res Rev 2011;68(4):387-420. [doi: 10.1177/1077558711399580] [Medline: 21813576]
6. Soong JTY, Bell D, Ong MEH. Meeting today's healthcare needs: medicine at the interface. Ann Acad Med Singap 2022;51(12):787-792 [FREE Full text] [doi: 10.47102/annals-acadmedsg.2022196] [Medline: 36592147]
7. Hospital admission rate by sex and age. Department of Statistics Singapore. 2023. URL: https://tablebuilder.singstat.gov.sg/table/TS/M870321 [accessed 2024-09-15]
8. Chen C, Ding S, Wang J. Digital health for aging populations. Nat Med 2023;29(7):1623-1630. [doi: 10.1038/s41591-023-02391-8] [Medline: 37464029]
9. Solis-Navarro L, Gismero A, Fernández-Jané C, Torres-Castro R, Solá-Madurell M, Bergé C, et al. Effectiveness of home-based exercise delivered by digital health in older adults: a systematic review and meta-analysis. Age Ageing 2022;51(11):afac243 [FREE Full text] [doi: 10.1093/ageing/afac243] [Medline: 36346736]
10. Roberts AL, Fisher A, Smith L, Heinrich M, Potts HWW. Digital health behaviour change interventions targeting physical activity and diet in cancer survivors: a systematic review and meta-analysis. J Cancer Surviv 2017;11(6):704-719 [FREE Full text] [doi: 10.1007/s11764-017-0632-1] [Medline: 28779220]
11. McKinley A. Investigating the Efficacy of Gamification in an Online Positive Psychology Intervention (OPPI) Aimed at Increasing Employee Well-being. Stellenbosch, South Africa: Stellenbosch University; 2022.
12. Living life to the fullest: 2023 action plan for successful ageing. Ministry of Health Singapore. 2023. URL: https://www.moh.gov.sg/docs/librariesprovider3/action-plan/2023-action-plan.pdf [accessed 2024-10-18]
13. Tadai M, Tan M. Digital literacy among older adults in Singapore. Singapore Management University. 2023. URL: https://rosa.smu.edu.sg/sites/rosa.smu.edu.sg/files/Briefs/May%20Jun23/Digital%20Literacy_Final2.pdf [accessed 2024-10-18]
14. Low STH, Sakhardande PG, Lai YF, Long ADS, Kaur-Gill S. Attitudes and perceptions toward healthcare technology adoption among older adults in Singapore: a qualitative study. Front Public Health 2021;9:588590 [FREE Full text] [doi: 10.3389/fpubh.2021.588590] [Medline: 33659232]

15. Bujnowska-Fedak M, Grata-Borkowska U. Use of telemedicine-based care for the aging and elderly: promises and pitfalls. Smart Homecare Technol Telehealth 2015:91-105. [doi: 10.2147/shtt.s59498]
16. Clarke V, Braun V. Thematic analysis. J Posit Psychol 2016;12(3):297-298. [doi: 10.1080/17439760.2016.1262613]
17. Ose SO. Using excel and word to structure qualitative data. J Appl Soc Sci 2016;10(2):147-162. [doi: 10.1177/1936724416664948]
18. Visaria A, Aithal S, Malhotra R. Digital technology use, in general and for health purposes, by older adults in Singapore. Aging Health Res 2023;3(1):100117. [doi: 10.1016/j.ahr.2023.100117]
19. Glisky E. Changes in cognitive function in human aging. In: Brain Aging: Models, Methods, and Mechanisms. Boca Raton, FL: CRC Press/Taylor & Francis; 2007:3-20.
20. Murman DL. The impact of age on cognition. Semin Hear 2015;36(3):111-121 [FREE Full text] [doi: 10.1055/s-0035-1555115] [Medline: 27516712]
21. Light LL. Memory and aging: four hypotheses in search of data. Annu Rev Psychol 1991;42:333-376. [doi: 10.1146/annurev.ps.42.020191.002001] [Medline: 2018397]
22. Whitelaw S, Pellegrini D, Mamas M, Cowie M, van Spall HGC. Barriers and facilitators of the uptake of digital health technology in cardiovascular care: a systematic scoping review. Eur Heart J Digit Health 2021;2(1):62-74 [FREE Full text] [doi: 10.1093/ehjdh/ztab005] [Medline: 34048508]
23. Lin TTC, Bautista JR, Core R. Seniors and mobiles: a qualitative inquiry of mHealth adoption among Singapore seniors. Inform Health Soc Care 2020;45(4):360-373. [doi: 10.1080/17538157.2020.1755974] [Medline: 32484720]
24. Haldane V, Koh JJK, Srivastava A, Teo KWQ, Tan YG, Cheng RX, et al. User preferences and persona design for an mHealth intervention to support adherence to cardiovascular disease medication in Singapore: a multi-method study. JMIR Mhealth Uhealth 2019;7(5):e10465 [FREE Full text] [doi: 10.2196/10465] [Medline: 31140445]
25. Brown JB, Carroll J, Boon H, Marmoreo J. Women's decision-making about their health care: views over the life cycle. Patient Educ Couns 2002;48(3):225-231. [doi: 10.1016/s0738-3991(02)00175-1] [Medline: 12477607]
26. Gollop CJ. Health information-seeking behavior and older African American women. Bull Med Libr Assoc 1997;85(2):141-146 [FREE Full text] [Medline: 9160150]
27. Oh KM, Kreps GL, Jun J, Chong E, Ramsey L. Examining the health information-seeking behaviors of Korean Americans. J Health Commun 2012;17(7):779-801. [doi: 10.1080/10810730.2011.650830] [Medline: 22642692]
28. Czaja R, Manfredi C, Price J. The determinants and consequences of information seeking among cancer patients. J Health Commun 2003;8(6):529-562. [doi: 10.1080/716100418] [Medline: 14690888]
29. Chang L, Basnyat I, Teo D. Seeking and processing information for health decisions among elderly Chinese Singaporean women. J Women Aging 2014;26(3):257-279. [doi: 10.1080/08952841.2014.888881] [Medline: 24919105]
30. Tan M, Straughan PT, Cheong G. Information trust and COVID-19 vaccine hesitancy amongst middle-aged and older adults in Singapore: a latent class analysis approach. Soc Sci Med 2022;296:114767 [FREE Full text] [doi: 10.1016/j.socscimed.2022.114767] [Medline: 35144226]
31. Yao J, Tan CS, Chen C, Tan J, Lim N, Müller-Riemenschneider F. Bright spots, physical activity investments that work: national steps challenge, Singapore: a nationwide mHealth physical activity programme. Br J Sports Med 2020;54(17):1047-1048. [doi: 10.1136/bjsports-2019-101662] [Medline: 31857340]
32. Moore K, O'Shea E, Kenny L, Barton J, Tedesco S, Sica M, et al. Older adults' experiences with using wearable devices: qualitative systematic review and meta-synthesis. JMIR Mhealth Uhealth 2021;9(6):e23832 [FREE Full text] [doi: 10.2196/23832] [Medline: 34081020]
33. Man REK, Ho AXY, Lee EPX, Fenwick EKD, Aravindhan A, Ho KC, et al. Awareness and attitudes of elderly Southeast Asian adults towards telehealth during the COVID-19 pandemic: a qualitative study. Singapore Med J 2023 [FREE Full text] [doi: 10.4103/singaporemedj.SMJ-2022-117] [Medline: 37675683]

Abbreviations

DH: digital health

Edited by A Kushniruk; submitted 21.03.24; peer-reviewed by F Özsungur; comments to author 19.04.24; revised version received 29.07.24; accepted 14.10.24; published 11.11.24.

Please cite as:
Leong QY, Lee VV, Ng WY, Vijayakumar S, Lau NY, Mauritzon I, Blasiak A, Ho D
Older Adults' Perspectives and Experiences With Digital Health in Singapore: Qualitative Study
JMIR Hum Factors 2024;11:e58641
URL: https://humanfactors.jmir.org/2024/1/e58641
doi:10.2196/58641
PMID:

Original Paper

Improving Social Media-Based Support Groups for the Rare Disease Community: Interview Study With Patients and Parents of Children with Rare and Undiagnosed Diseases

Tom A Doyle[1], PhD; Samantha L Vershaw[1], BS; Erin Conboy[2], MD; Colin M E Halverson[1], PhD

[1]Center for Bioethics, Indiana University School of Medicine, Indianapolis, IN, United States

[2]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, United States

Corresponding Author:
Tom A Doyle, PhD
Center for Bioethics
Indiana University School of Medicine
410 W. 10th Street
Suite 3100
Indianapolis, IN, 46202
United States
Phone: 1 (317) 278 4052
Email: doylet@iu.edu

Abstract

Background: The rarity that is inherent in rare disease (RD) often means that patients and parents of children with RDs feel uniquely isolated and therefore are unprepared or unsupported in their care. To overcome this isolation, many within the RD community turn to the internet, and social media groups in particular, to gather useful information about their RDs. While previous research has shown that social media support groups are helpful for those affected by RDs, it is unclear what these groups are particularly useful or helpful for patients and parents of children with RDs.

Objective: This study aimed to identify what specific features of disease-related support groups (DRSGs) the RD community finds particularly useful or supportive and provide a set of recommendations to improve social media–based RD support groups based on this information.

Methods: Semistructured qualitative interviews were performed with patients and parents of patients with RDs. Interview participants had to be at least 18 years of age at the time of the interview, be seen by a genetics specialist at a partner health care institution and be proficient in the English language. Social media use was not a prerequisite for participation, so interview participants ranged from extensive users of social media to those who chose to remain off all social media. All interviews were conducted by phone, recorded, and then transcribed. Interview transcripts were then coded using the 6 steps outlined by Braun and Clarke. Three researchers (TAD, SLV, and CMEH) performed initial coding. After this, the study team conducted a review of themes and all members of the team agreed upon a final analysis and presentation of data.

Results: We conducted 31 interviews (mean age 40, SD 10.04 years; n=27, 87% were women; n=30, 97% were non-Hispanic White). Thematic analysis revealed that social media DRSG users identified the informational usefulness of these groups as being related to the gathering and sharing of specific information about an RD, clarification about the importance and meaning of certain symptoms, and obtaining insight into an RD's progression and prognosis. Participants also identified that DRSGs were useful sources of practical information, such as tips and tricks about managing RD-related issues and concerns. In addition, participants found DRSGs to be a useful space for sharing their disease-related stories but also highlighted a feeling of exhaustion from overexposure and overuse of DRSGs.

Conclusions: This study identifies the usefulness of DRSGs for the RD community and provides a set of recommendations to improve future instances of DRSGs. These recommendations can be used to create DRSGs that are less prone to splintering into other DRSGs, thus minimizing the risk of having important RD-related information unhelpfully dispersed amongst a multitude of support groups.

(JMIR Hum Factors 2024;11:e57833) doi:10.2196/57833

KEYWORDS

social media; rare disease; support groups; pediatric rare disease; Ehlers-Danlos syndrome; collagen disease; fibrillar collagen; cutis elastica; connective tissue disorders; hyperelasticity; hypermobility of joints, inherited; genetic disorder; genetics; pediatric

Introduction

The National Institutes of Health estimates that 25-30 million Americans, roughly 1 in 10, are affected by rare diseases (RDs). Therefore, while a single RD may impact only a small percentage of people, the culminative sum of all RDs impacts a sizable portion of the overall population. Despite this, patients with RDs often face significant barriers to appropriate care. The rarity of their conditions means that substantive information about their health is often lacking, and patients encounter significant challenges in finding and understanding the little information that is available [1]. Coupled with a paucity of appropriate diagnostic services, they face exceptionally long and difficult diagnostic odysseys [2]. This means that patients with RD find themselves on “the journey of experiencing unexplained symptoms, seeking evaluation, experiencing symptom evolution, and seeking further evaluation, all in an attempt to obtain an accurate diagnosis” [3]. Another consequence of the rarity of these conditions is that opportunities for socializing and networking between patients are few and far between. Patients therefore often feel isolated, unprepared, and unsupported in managing their care [2,4,5], as though they had been left to undertake their medical journey alone.

Given these challenges, for many patients with RDs, the internet may provide the only medium through which they can connect with others who have the same or similar diseases. Previous research has shown that patients and guardians of children with RD often use the internet, and social media in particular, as a source of informational, emotional, and social support [5-9]. Platforms that allow users to create disease-specific groups, such as Facebook (Meta) or Reddit, have been found to be especially useful for patients and parents of children with RD. Notably, a recent study found that Facebook had over 6000 user-created groups related specifically to RDs [10].

Yet, while it is evident that the members of the RD community use social media groups to engage with one another, it remains unclear what specific features of these disease-related support groups (DRSGs) make them helpful for patients and guardians of children with RD. How do patients find support within the various RD communities on social media? In order to fill this knowledge gap, we conducted a series of in-depth, qualitative interviews with patients and their guardians who were being seen in the clinic for rare genetic disorders. The results of our study are of significance for RD advocates and organizations that hope to design, establish, and cultivate social media–based RD communities.

Methods

Participants and Procedures

The Indiana University institutional review board approved this qualitative study. The study team consisted of three researchers: (1) one male, faculty investigator (TAD) with experience in philosophy of medicine and bioethics and doctoral-level training in qualitative research methods; (2) one male, with a doctorate in medical anthropology, advanced training in bioethics (CMEH), and extensive experience in qualitative research methods; and (3) one female research assistant (SLV) whose work focuses on improving care for the RD population.

Involvement in the study required participants to be older than 18 years of age and to have proficiency in the English language. The candidate needed either to be actively followed at Indiana University Health or to be the guardian of a minor who was being followed at Indiana University Health. Candidates were identified by genetics specialists using convenience sampling, and those individuals who expressed verbal interest in participation were scheduled for a 1-time phone interview. The recruitment strategy was agnostic as to whether candidates were intensive or casual users of social media, and even individuals who preferred not to use DRSGs at all were invited to participate. This design was meant to capture the perspectives of both individuals who do and those who do not find value in engaging in social media–based RD communities. Participants provided verbal informed consent before participating in this study. Interviews ran from April to September 2023, at which point team members agreed that the sufficient depth and richness of the data had been achieved to generate meaningful insights and themes, as is appropriate for interpretive qualitative research [11,12].

Instruments and Analysis

The interviews were structured to learn about participants’ social media practices and values in relation to their RD needs and interests. The team jointly developed the interview guide based on a thorough literature review. The interview template consisted of approximately 21 questions outlining general uses of social media, including account platforms and perceptions of their typical social media usage; attitudes toward interviewees’ use of social media for disease information; and how social media has influenced their social lives, both as a family and individually. It also addressed any advocacy efforts in which interviewees had participated and their interactions with peers on social media through RD groups and informational sites. In order to avoid prompting participants to discuss a particular social media site or group, these sites and groups were introduced to participants in a general, nonspecific manner (eg, “on what platforms do you have a social media account?”; “are you an active member in any social media groups?”). The interview guide concluded with basic demographic questions, including gender, race, age, religious affiliation, and highest level of education. Participants could complete the interview in whatever space they felt comfortable, while the interviewer conducted the calls from a private office in order to maintain participant confidentiality. The complete interview guide used for these interviews can be found in the Multimedia Appendix 1 of this article.

One researcher (TAD) conducted the interviews while completing detailed field notes. Each interview was audio-recorded and transcribed through MacWhisper (Sindre Sorhus) or audio transcription through Microsoft Teams, both HIPAA (Health Insurance Portability and Accountability Act)-compliant, artificial intelligence–based programs. The generated transcripts were then reviewed by a member of the research team, who compared the audio recording against the transcript in order to ensure accuracy. In cases where there was a discrepancy between the audio recording and the generated transcript, the team member who conducted the interview was consulted to clarify what was said in the interview. After this, the researchers uploaded transcripts to Dedoose (SocioCultural Research Consultants, LLC), a mixed methods analysis tool especially useful for qualitative research. It was determined that reflexive thematic analysis based on the 6 steps outlined by Braun and Clarke [11] was the best method to investigate these interviews both rigorously and yet flexibly. Members of the study team familiarized themselves with the conversations and constructed a coding tree outlined by overall platform use, uses and goals of social media, information behavior, social media's impact of perception of RD, behavior on social media, advocacy and support, mental health, and portrayal of clinicians. The codes were further enhanced, as the study team met periodically to refine and clarify themes emerging from the data, thus ensuring they were used consistently by all team members. The study team conducted a review of themes, and all members approved the final analysis and presentation of data.

Ethical Considerations

Study procedures were approved by the Indiana University institutional review board (protocol 18779). The study was determined to be exempt research and thus consent was obtained verbally after a member of the study team thoroughly explained to participants the study procedure, the risks and benefits of participating, and how their data will be protected. Participants were also provided with this information in a study information sheet. All data presented in this manuscript has undergone proper deidentification.

Results

Overview

A total of 31 participants completed an interview. The average length of an interview was 45 minutes (range 21-73 minutes). After coding the data, 3 themes were identified regarding how and why DRSGs were considered valuable to patients with RDs and families, and the themes are (1) for general information about a disease and its symptoms, (2) as a source of practical advice, and (3) for accessing and managing emotional and social support. Each of these themes is elucidated in their respective sections below.

Participant Characteristics and Demographics

Of the 31 participants who completed an interview, 8 (26%) were parents of patients with an undiagnosed or RD, and 23 (74%) were patients with an undiagnosed or RD. Overall, differences between these 2 groups were not observed; both groups shared congruent experiences that mapped onto a common set of themes and concerns.

All participants were residents of the United States. The majority identified as female (87%, 27/31) and White (97%, 30/31). The average age of participants was 40 years. Roughly half of our participants identified as Christian. Regarding educational attainment, about half had obtained a bachelor's degree. More information about their demographics is found in Table 1.

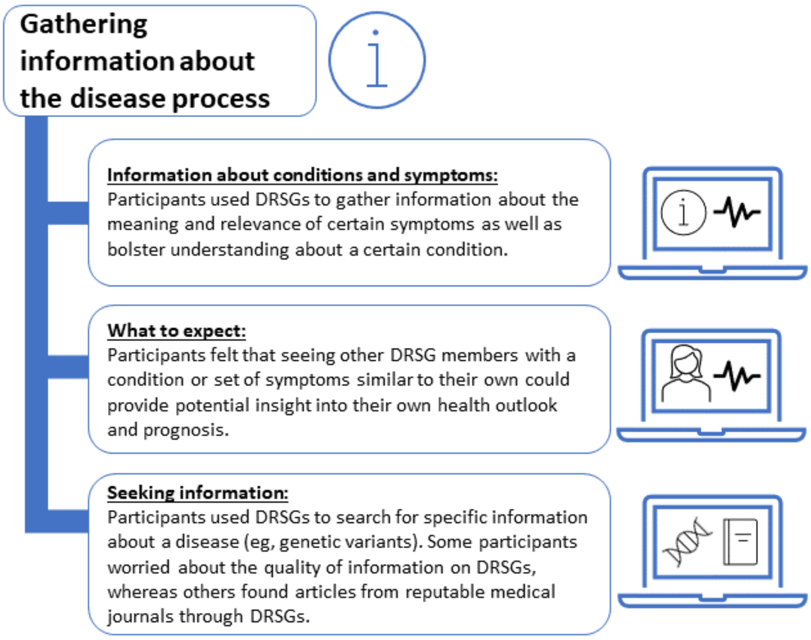
Table 1. Participant demographics.

Demographics variables	Interviews (N=31), n (%)
Gender	
Female	27 (87)
Male	2 (6)
Nonbinary	2 (6)
Religion	
Christian	16 (52)
Spiritual	5 (16)
None	6 (19)
Other	4 (13)
Education	
High school	5 (16)
Some college	4 (13)
Bachelor's degree	14 (45)
Graduate degree	8 (26)
Race or Ethnicity	
Non-Hispanic White	30 (97)
More than one race	1 (3)
Age (years)	
Average	40
Range	19-63

Gathering Information About the Disease Process

Participants often used DRSGs to gather information about their disease. In particular, they used DRSGs to gather more information about conditions and symptoms, a disease prognosis, or to seek further information about a disease (Figure 1).

Figure 1. Gathering information about the disease process visual summary. DRSG: disease-related support group.



Information About Conditions and Symptoms

Participants used DRSGs to gather information about the relevance or meaning of certain symptoms as well as general information about their medical conditions. One participant remarked that social media was “a good place to start your search” [P27], and another participant stated that it could be used for “crowdsourcing medicine” [P23] insofar as one could leverage the insights offered by members of a particular DRSG in pursuit of effective diagnosis and care.

Participants recounted how joining groups for a particular condition greatly bolstered their understanding of that condition. For instance, one parent recalled the usefulness of a DRSG for gathering information about her daughter’s disorder: “It was super helpful because for us, it was kind of like, ‘Here you go, this is it.’” The group members, she said, “were pretty knowledgeable of the condition and how to manage it and work with the symptoms” [P06]. In another instance, a parent remarked that “we found out that [our daughter] had hypotonia and we didn’t know what that was, [so] I was like, ‘I’m going to look that up’, and then I saw on Facebook they have a hypotonia support group, so I did join that” [P03].

Participants also used social media groups to discover the meaning behind certain symptoms. This allowed them to relate a seemingly disparate cluster of symptoms to a particular rare condition. As one of our participants put it, “we don’t even know that our symptoms are symptoms, so you can be like, ‘Oh, yeah, that’s not normal.’ It’s validating” [P29]. In one instance, a parent of a child with an ultra-RD credited her DRSG use with linking her child’s symptoms to a specific diagnosis: “A lady […] was telling her child’s story, and I was half-listening. But then she started talking about the [symptom] and I was like, ‘What did she just say?’ and I rewound it, and then everything she said about her kid, and even a video she had of him crawling away, looked like my son” [P07]. In another case, a woman explained how she was able to use a social media group to trace “a really, really bizarre sensation” in her chest to costochondritis, a condition that is commonly experienced by those with her particular RD but one about which she had never previously heard [P12].

What to Expect

Through their DRSG usage, our participants sought information about what they should expect when living with a certain RD or caring for a child with such a condition. One participant remarked that members of DRSGs often asked whether it was normal for their child not to reach certain developmental milestones [P01], to which other members reportedly responded by describing their own personal experiences with the condition. When asked what participants would want to get out of a DRSG, multiple interviewees cited information regarding how their particular RD progresses: “It doesn’t feel like a stagnant disease,” one participant explained, describing the unexpectedness of her child’s evolving symptoms, so she turned to social media “to just kind of compare stories” [P02]. Another participant explained that “seeing other kids, maybe that are older than [her daughter], how they’re doing and maybe the outlook on that” had been beneficial for her [P03]. Some participants, however, cautioned that hearing too much about more severe prognoses can be harmful: “You have to be in the right mindset, because you can come across very scary stories about people” [P29].

Seeking Information

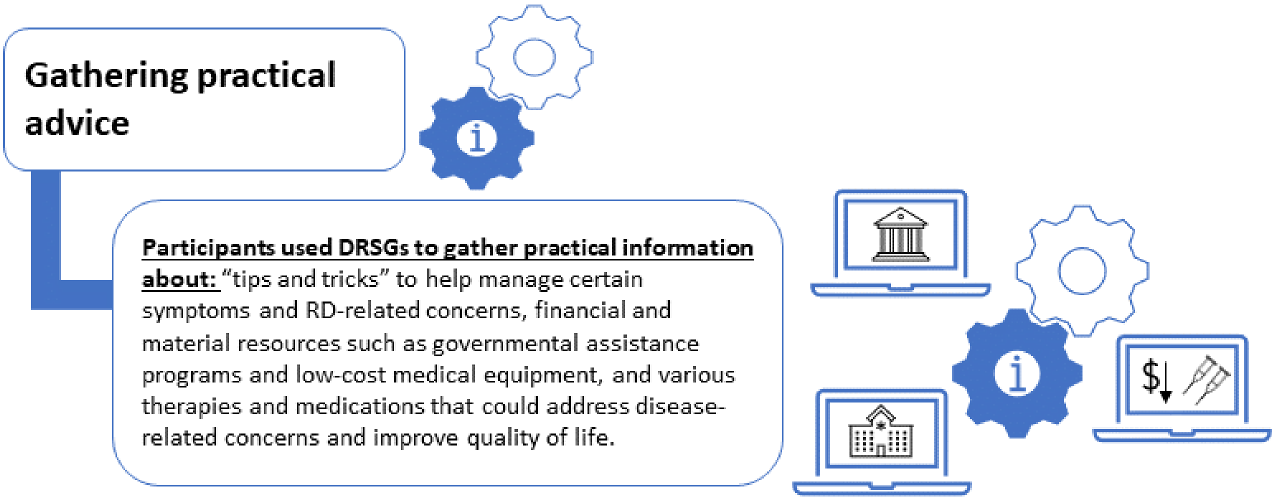
In general, our participants recalled using DRSGs to seek information about their or their child’s condition. In particular, some searched social media sites in an attempt to learn about rare genetic variants: “I search every once in a while, [for] that specific gene and tag to see if anyone else pops up” [P02]. Another participant recounted that “I did post or search for [genetic variant] and see what posts came up with that” [P01].

In contrast, participants with either medical training or a graduate degree tended to conduct research without the aid of DRSGs: “I would just go to Google Scholar. When I was at the university and I had access to the library, I would use the library to look at different articles. I might still see about access though PubMed” [P08]. Another participant, who had a background in nursing, recommended to her daughter after her diagnosis “that she had to look at more medical journal type things” rather than rely on what she could find on social media [P09]. This reticence was due to an expressed skepticism on the part of many regarding the quality of information available on such platforms. Some groups are organized for the sharing of peer-reviewed research articles, which for medically literate patients was seen as very beneficial and a way to avoid dubious information: “I can’t tell you how much I’ve learned through the [DRSG],” one young woman explained [P15].

Gathering Practical Advice

Participants used DSRGs to gather practical advice relating to their RD and related concerns. In particular, participants used DRSGs to gather various “tips and tricks” that were useful in addressing their disease-related concerns (Figure 2).

Figure 2. Gathering practical advice visual summary. RD: rare disease.



Participants often praised DRSGs as sources for useful "tips and tricks" to help manage disease-related issues. One participant explained that she used social media to gather practical rather than medical information: "It's more about the giving and sharing of advice and tips" [P12]. This advice often took the form of recommendations for symptom management based on personal experience. Other group members regularly provide testimonials and appraisals of resources for managing symptoms, including everything from braces and topical creams to supportive mattresses and pillows. One parent of a child requiring a feeding tube mentioned that joining a DRSG was helpful initially after diagnosis because it allowed her to collect "a bunch of different tricks, or different clothing" [P04]. One parent of a child with ankle braces remarked: "I put on [the DRSG], 'who would have thought that buying a pair of shoes for a two-year-old would be so difficult?' and there was a couple of other moms who were like 'Oh, go with this brand'" [P03]. A challenging task with the potential for significant trial-and-error related cost and frustration was averted by accessing others' lived experiences through social media.

In addition to advice, our participants used DRSGs to gather practical resources to address disease-specific concerns. Participants mentioned DRSGs that focused narrowly on enabling users to acquire medical equipment and other materials. For instance, one interviewee described a group in which "families who have extra of different things" provided valuable supplies to families with limited financial means and access [P10]. One parent mentioned a local group whose sole purpose was to allow parents to provide others with "used equipment, or things that their kids have outgrown, or that they don't need anymore […] and you don't know what to do with them […] You post them on there so local people can either buy it from you or get it from you" [P07]. In other cases, resource sharing occurred through posts within general DRSGs: "People post on there if they have Aspen collars or things that they don't need anymore," noted one participant of a group broadly dedicated to her particular RD [P288].

Information about programs designed to assist those with disabilities was also commonly shared on DRSGs. One participant recounted that she "learned about the Medicaid waiver" for home and community-based services through DRSGs, adding that "I'm so thankful [my daughter] has that now" [P02]. This participant was also able to gather information about a program offering reduced admission to local museums for children with disabilities, and she even obtained a medical stroller and car seat through a program she learned about through a DRSG.

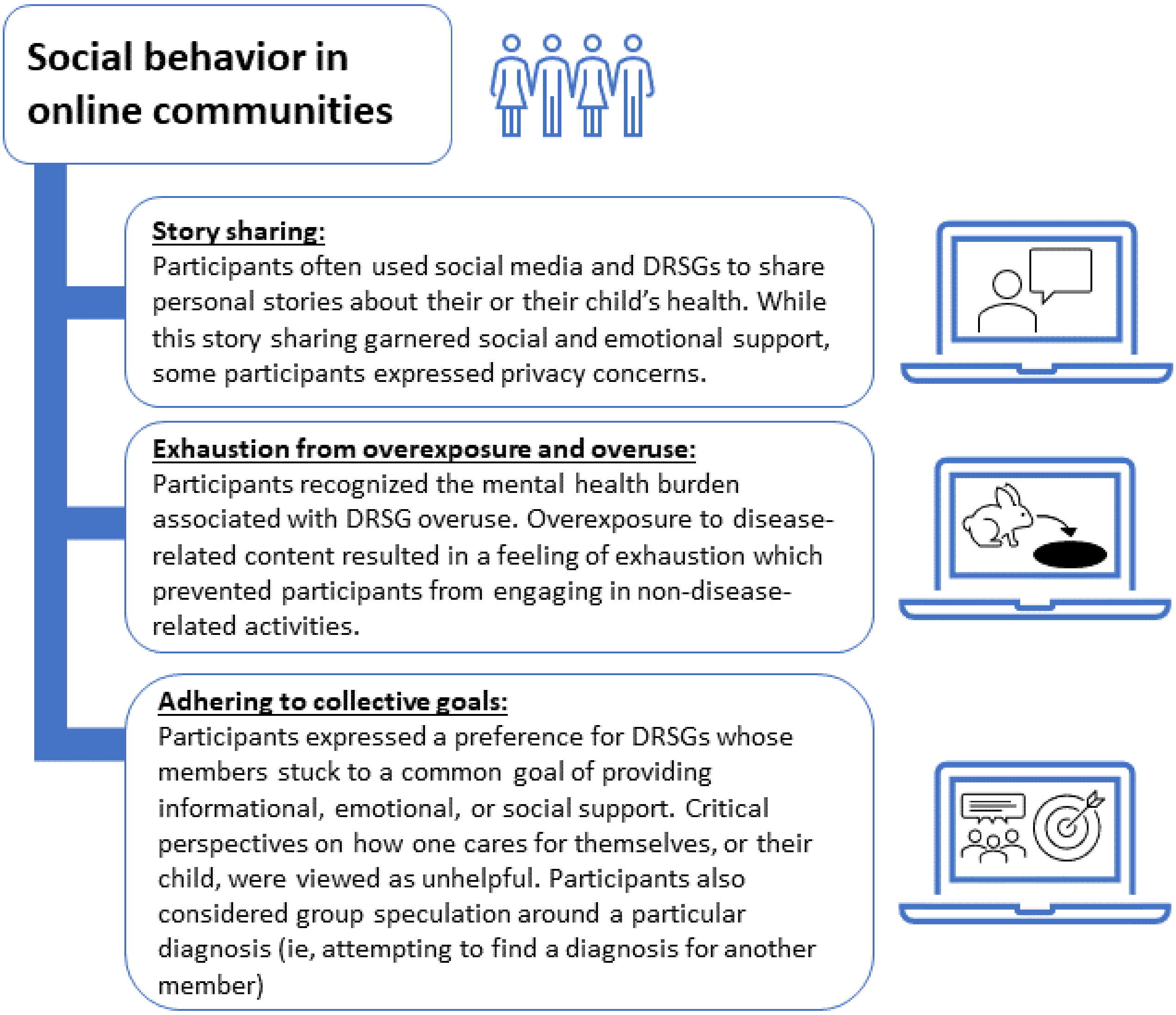
A large portion of our participants (74%, 23/31) reported using DRSGs to obtain or provide disease-related guidance. While the type of guidance they sought varied, participants used this information in a practical manner, either to discover an actionable therapy or to judge the usefulness of a particular medication. Some of our participants used DRSGs to collect information about therapies of which they had previously been unaware: "I hadn't known about, for example, horseback riding therapy. I didn't know that was a thing" [P01].

Our participants also found value in learning about others' experiences with clinicians. One participant described the use of DRSGs to ascertain "people's opinions about a doctor at a hospital or within a group: 'Who has had this doctor? What is your feedback' or 'Have you tried this medication for [something]?'" [P02]. Such advice was seen as particularly useful, as interviewees felt that many clinicians whom they encountered were not sufficiently knowledgeable or otherwise disposed to care for their particular condition.

Social Behavior in Social Media–Based Communities

Participants used DRSGs to engage in social behaviors. In particular, participants used DRSGs to share stories regarding their RD experiences. However, participants also expressed concerns regarding the potential for overuse of DRSGs. Participants also believed that behavior on DRSGs should adhere to the collective goals of providing emotional, informational, and social support (Figure 3).

Figure 3. Social behavior in social media–based communities visual summary. DRSG: disease-related support group.



Story Sharing

About half of our participants (54%, 17/31) reported that they have used social media to tell their personal stories or provide updates about their own or their child's health. Motivations for sharing personal stories and information about health status varied. One woman used Facebook to describe her experiences in a series of posts that she had intended "for educational purposes, for awareness purposes" [P25]. A mother cited social media as "a way to keep everybody in the loop" about her child's health, which saved her the trouble of "texting 18,000 people" to provide health updates. As another participant put it, "It's just so much easier to put out updates like that online than it is to try to explain it 30 different times to 30 different people" [P02]. However, regardless of the participants' motivations for sharing their personal stories, they expressed that they had received sympathetic and reassuring responses to those posts. For instance, one mother remarked, "[My daughter] had eye muscle surgery, and I did a little post, and people were saying 'thoughts and prayers' for her and a quick recovery" [P03].

Contrarily, some participants hesitated to share their stories so openly. "I'm very careful about what I post because I just don't want people seeing everything. I'm private," one participant stated [P15]. "Oversharing could be very dangerous to their own privacy as well as their own health," noted another woman [P23].

Exhaustion from Overexposure and Overuse

Our participants appraised social media as a generally supportive environment. However, many remarked that DRSGs can become emotionally burdensome when members spend too much time on them. They often mentioned limiting their consumption of DRSG content for that reason. One woman, for instance, stated that "there was a group of moms from the [DRSG] where we had a private group chat that would occur daily, all the time. I had to mute it. It was too much" [P06]. Another woman limited her participation in social media groups because "if I spend all my time in a chat room, that's where my energy is going, and I had other things I wanted to do" [P24]. In fact, after leaving social media, she returned to school and graduated shortly after our interview. One parent highlighted how her DRSG use led to maladaptive and obsessive behavior. Particularly, she felt compelled to follow every lead related to her child's condition:

"It gets me looking into things to see if maybe I can find that, and I'm just like, 'Oh, maybe I can bring this to the doctor,' you know? And sometimes it just really takes over" [P10].

Adhering to Collective Goals

Our participants often asserted that DRSGs have the best cohesion when members keep to the specific goal of providing informational, emotional, or social support to other members. One of our participants stated, "We're all moms that are trying to figure out what the heck is going on with our kids. We don't need anybody beating us up over a decision that we made or didn't make [...] Just make sure [the group is] really monitored and positive. And remember that we're all in situations together" [P04]. Another participant expressed frustration at other members of these groups who tried to play the role of a clinician rather than providing social support or practical advice: "You have a lot of parents in these groups that I guess are pushing what they feel your child's diagnosis is, and let's face it: they're not the doctor. So, I can give my information on my child, and they can be like, 'Oh, that sounds like this.' And then they stick to that point of they strongly feel that this is what is wrong with your child." [P10].

Discussion

Principal Findings

In many cases, DRSGs provided our participants with an understanding of their RD that was informed by others' lived experiences. This allowed them to gather practical advice regarding how others had dealt with particular symptoms, obtain general information about a disease, and socialize with those who share similar experiences. These findings support previous research on DRSGs, which has shown that online groups provide both informational and emotional support for those who share a similar diagnosis [13]. In addition, past studies have found that DRSGs provide users a space to ask and answer specific questions about medication management, hospitalizations, and clinic visits as well as provide a supportive environment where users can disclose personal stories about their disease experience [14].

Due to random sampling used during the recruitment process, our study was also able to capture the perspectives of those who do not regularly use social media. This has been an oversight in much previous research [5-7,15,16]. We found that these participants avoided social media either due to privacy concerns or a perceived lack of benefit. Privacy concerns have been found in other studies, though they do not always rise to the level of stopping patients from participating in DRSGs [5,15,16].

Information About the Disease Process

Within the existing literature on RD and social media use, it has been found that users often turn to DRSGs for information about a diagnosis or test results or to participate in conversations about a specific diagnosis [6,7]. Another study reported that patients with RDs use DRSGs primarily as a source to share medical information about symptoms, treatments, and diagnoses [8].

Our study likewise found that participants regularly made use of social media for informational needs. However, they did not expect this information to replace communication with trusted medical professionals. Many participants understood the role of the community in these groups not as that of experts, but rather as individuals with unique and personal insight into what it is like to live with and care for someone with an RD. This aligns with previous research which has found that participants often regarded clinical advice accessed through social media as dubious and to be met with skepticism [5].

DRSGs, then, were seen not as a replacement for recommendations offered by a participant's clinician, but rather as a complement to these clinical recommendations. This finding is congruent with a review of previous research which has found that patients often used social media–based communities to receive information and advice on how to improve their medical care or explore treatment options [17]. One study found that patients often turned to online support groups because they believed their providers lacked a sufficient amount of time to discuss each treatment option in depth. These groups allowed patients to gather information from members of the community who were actively undergoing these treatments [18]. Our participants engaged in similar behavior, which indicates that the patient perspectives gathered from DRSGs parallel information discussed in a clinical setting.

Many participants preferred internet-based databases sources such as Google Scholar or PubMed as a source of information rather than social media, due to the credibility they attached to them. Information from medical journals was generally felt to be more reliable and trustworthy than information found through the self-report of DRSG members. Such a finding corroborates a recent review of the literature regarding online health information–seeking behaviors which highlighted the central role of trust in determining patients' information-seeking behavior [19].

One critical finding from our study is that information-seeking through social media was often seen as burdensome and could even become maladaptive. Our participants reported using the search feature on DRSGs to find information about particular genetic findings, but they also highlighted that such searching could also become obsessive and result in mental exhaustion.

Practical Advice and Lived Experience

While it is evident that our participants engaged in information sharing and gathering, it is important to emphasize that our participants often used DRSGs as more than just an informational resource. In particular, they relayed these groups' usefulness for obtaining practical advice, for providing and seeking appraisal of medical providers, as well as for sharing and gathering guidance on how to enroll in certain government programs or to obtain certain government benefits. These "tips and tricks" allowed them to navigate and address their many concerns that fall outside of the scope of traditional clinical care. They additionally provide information related to specific products that particular RD communities have found useful, such as brands of clothing or specific kinds of medical braces. Moreover, participants sought information about others' lived experiences in order to forecast what to expect regarding disease progression, assisting in personal prognostication.

While previous research has identified RD-specific DRSGs as a useful resource for such general practical advice [5], one of our study's novel findings is that DRSGs are important sources of material exchange. While much of the literature focuses on social media as a place to obtain information and social support, our participants relayed the importance of DRSGs in the selling, buying, and trading of specific medical equipment. Similar to this, an additional novel finding of our study was the use of DRSGs to discover and help obtain various material and service benefits offered by federal and state governments. Previous cost-of-illness research has established RDs as a major economic burden for patients and their families [20-22]. Hence, it is noteworthy that DRSGs have the potential to alleviate particular disease-related economic burdens. Conceivably, DRSGs allow users to recuperate some of the costs associated with medical equipment by selling them directly to other patients and their guardians; in turn, those individuals may do so at a reduced price and thus do not incur the often-exorbitant cost of new equipment.

Socializing and Networking

As made evident by our study, social media serves as one possible venue for patients and their guardians to establish and maintain social connections despite physical, geographic, and medical barriers. Our findings concur with these previous findings that social media sites that allow users to create disease-specific support groups, such as Facebook or Reddit, can be helpful resources for those with chronic diseases [23,24]. Social support has been shown to have a positive impact on one's health [25]. In particular, social support can reduce psychological distress, mitigate feelings of loneliness, improve quality of life, and decrease mortality risk [26]. In light of this fact, many health organizations, such as the World Health Organization (WHO) and the US Department of Health and Human Services (DHHS), now emphasize the importance of a person's ability to establish and maintain social relationships with members of their community [27,28]. For this reason, the WHO has recently established a Commission on Social Connection in order to address social isolation and loneliness and the DHHS has established the objective of increasing social and community support as part of its Healthy People 2030 longitudinal-objective program [29,30]. DRSGs therefore have the ability to play a critical role in improving the overall health and connectedness of RD communities.

Another novel finding of our study was participants' preference for DRSGs that adhere to the perceived objectives or goals of a support group. That is, many of our participants understood the overall purpose of these groups as providing a supportive and non-judgmental space for patients and caregivers to engage in discussions about a particular disease and its symptoms, practical tips and tricks that could improve the overall quality of life, and an environment in which they could share their personal experience with a disease if they so choose. Our participants found it off-putting when discussions deviated from these goals.

Recommendations

Quantitative analysis has shown that thousands of RD social media groups now exist [10]. While this staggering number may provide a diverse range of opportunities for informational, emotional, and social support, it also suggests that valuable resources are unhelpfully distributed across an unmanageable number of virtual locations. For instance, the aforementioned quantitative analysis found 93 unique support groups specifically for a single rare connective tissue disorder, and that number accounts only for groups dedicated to its pediatric form [10]. This suggests that patients are faced with the burden of having to figure out which specific support groups to join, how to resolve possible conflicts in information gathered from these various groups, and how to avoid oversaturating their social media feeds with disease-related information from disparate groups, which we have found is often emotionally burdensome.

It is evident, then, that the duplication of efforts is a serious problem facing DRSG administrators and their users. We contend that a solution to this substantial problem is to ensure that DRSGs are designed and operated in a manner that ensures users' informational and social needs are fulfilled by a single group. It is likely that they will then be less likely to create new, schismogenic DRSGs, thus alleviating the problem of duplicating efforts and stretching virtual resources too thin.

Based on the data from our interviews, we propose the following recommendations for creating a DRSG that patients with RDs and their guardians will find valuable, usable, and worthwhile (Table 2). While the scope of our study was limited to RD, it is possible that these recommendations may be generalizable to other, non-RD social media–based communities.

Table 2. Recommendations to Improve DRSGs[a].

DRSG usability issue	Potential solution	Explanation	Usefulness
Our results show that DRSGs are commonly used to gather practical or general advice related to a particular rare disease. This Important information may become "buried" by numerous new posts which may lack such practical relevance.	Establish and make use of pinned posts.	Posts that contain information, practical advice, or resources that the community finds particularly helpful should be "pinned" at the top of the group page.	Previous research has shown pinned posts to be a useful way to highlight and propagate important information [31]. Pinned posts can prevent members from repeatedly asking the same important questions. This allows new members to find the most helpful and important information as soon as they become a part of the group.
Certain materials, such as applications for governmental benefits, information about disability assistance programs, and other similar documents may not get proper attention when simply shared through a single post. DRSG users have identified these documents and resources as being particularly helpful to them.	Establish and make use of a "Files" section.	Official resources and documents should be archived in a set of group files. Group files can either be integrated through the social platform itself (eg, through the "Files" feature on Facebook) or through a third-party document hosting service (eg, Google or Dropbox).	A "Files" section allows users to share and store documents related to various assistance programs. This ensures that users can easily find and access important documents or resources. In addition, a "Files" section can allow DRSG users to share articles from journals for members of the community who possess a high health literacy.
In order to find relevant information related to a specific medication, therapy, symptom, or medical concern, DRSG users are often forced to conduct an imprecise search of all posts in a particular group and decipher them for potential relevance.	Introduce and encourage the use of posttags.	To enhance the searchability of the group on a particular topic, users should be encouraged to put particular keywords at the beginning of their posts that summarize the overall theme or topic of the post.	Prior research [32] has highlighted the usefulness of tags in enhancing searchability [33,34]. Posttags can assist users in searching quickly for a particular symptom, medication, or disease-related concern. This prevents repeated posting about similar issues or symptoms.
Our participants described the overuse of DRSGs as leading to exhaustion and emotional burden.	Creation and amplification of a community guide.	A community guide directs users on how to navigate the DRSG and how to make use of its various features and materials. This guide can also indicate the dangers of overuse (ie, mental exhaustion) of social media.	A community guide can help to ensure that users are "getting the most" out of a DRSG while ensuring that they do not feel overwhelmed by an abundance of information. In addition, including a section on the importance of taking breaks from social media can help users maintain healthy DRSG use.
DRSG users expressed frustration when other group members did not adhere to the perceived goal of the group (eg, to provide a supportive environment for patients and caregivers).	Creation of a group mission statement and a clear explanation of moderation decisions based on that mission statement.	In conjunction with a set of standard rules to maintain proper group decorum, a mission statement establishes the specific goals of the group and allows moderators to ensure the community focuses on these goals.	Establishing a mission statement for a DRSG can ensure that users are mindful of the specific goals and objectives of a group, preventing them from sharing irrelevant or disruptive content in an otherwise supportive environment.

[a]DRSG: disease-related support group.

Aside from these more targeted improvements for DRSGs, clinicians and RD advocates should also be aware of the importance of patient autonomy and community sovereignty when creating or recommending support groups. While our participants expressed a preference for competent moderation and proper group governance, it was also evident that they appreciated the ability to behave in an authentic or organic manner within these groups. While clinicians and advocacy organizations have a role to play in patient-oriented social media, it is important to recognize that this role should not impede a community's or patient's ability to express concerns or criticisms about specific clinicians or advocacy organizations in these virtual spaces. Therefore, clinicians and advocacy organizations that either create or engage in DRSGs must recognize that a social media–based support group truly belongs to a community of patients rather than to their organizational or clinical agenda. By asserting that a group belongs to the community and its members, clinicians and advocates help to empower patients and caregivers [35,36].

Therefore, we recommend that clinicians and advocacy organizations work to develop plans to ensure that DRSGs foster patient autonomy and community sovereignty rather than diminish it. With respect to group governance, this means giving community members a say in group moderation. Kiesler and colleagues offer useful guidance on this point [37]: They provide a total of 33 "design claims" that offer insight into the effective regulation of online spaces. Of particular relevance, they highlight that moderation decisions from members of the community are viewed as more legitimate and more effective, and that rule making which involves and incorporates the community's feedback allows for better compliance with rules.

Conclusion

Our study provides insight into the specific ways that patients and parents of children with a RD interact with DRSGs. We found that DRSGs were used to collect and seek medical information about a disease and its symptoms, gather practical resources and advice, and socialize and network with those who share a similar experience to their own regarding a particular disease. Importantly, these results uncover and highlight the particular features and elements of DRSGs that members of the RD community find helpful, usable, and worthwhile. Through considering the practical importance of these results, we established a set of recommendations that are tailored specifically to DRSGs. These recommendations provide guidance to RD advocates and organizations that manage DRSGs, which may reduce the reduplication of efforts and the splintering of disease-related support across an unmanageable multitude of groups, which at present is often experienced as time-consuming and emotionally burdensome.

Limitations

Our sample population featured a significant gender and race skew, with participants predominantly identifying as female and White. The collected data may therefore not necessarily represent the diverse perspectives present in the RD community. This skew could also impact the generalizability of our results to the many DRSGs found across various social media platforms. In addition, limitations related to scope and interview length prevented our study from examining how participants discovered support groups and whether certain search strategies or terms yielded information about potentially relevant social media–based support groups.

Data Availability

The datasets generated during and/or analyzed during this study are not publicly available due to the private and sensitive nature contained in the qualitative data. Relevant sections of deidentified data and/or coding trees are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview Questions.
[DOCX File , 23 KB - humanfactors_v11i1e57833_app1.docx]

References

1. Stoller JK. The challenge of rare diseases. Chest 2018;153(6):1309-1314. [doi: 10.1016/j.chest.2017.12.018] [Medline: 29325986]
2. Survey of the Delay in Diagnosis for 8 Rare Diseases in Europe (EurordisCare2. EURODIS. 2007. URL: https://www.eurordis.org/survey-reveals-lengthy-diagnostic-delays/ [accessed 2024-11-22]
3. Bordini BJ, Walsh RD, Basel D, Deshmukh T. Attaining diagnostic excellence: how the structure and function of a rare disease service contribute to ending the diagnostic odyssey. Med Clin North Am 2024;108(1):1-14. [doi: 10.1016/j.mcna.2023.06.013] [Medline: 37951644]
4. Zurynski Y, Frith K, Leonard H, Elliott E. Rare childhood diseases: how should we respond? Arch Dis Child 2008;93(12):1071-1074. [doi: 10.1136/adc.2007.134940] [Medline: 18684747]
5. Deuitch NT, Beckman E, Halley MC, Young JL, Reuter CM, Kohler J, Undiagnosed Diseases Network, et al. "Doctors can read about it, they can know about it, but they've never lived with it": how parents use social media throughout the diagnostic odyssey. J Genet Couns 2021;30(6):1707-1718 [FREE Full text] [doi: 10.1002/jgc4.1438] [Medline: 34096130]
6. Nicholl H, Tracey C, Begley T, King C, Lynch AM. Internet use by parents of children with rare conditions: findings from a study on parents' web information needs. J Med Internet Res 2017;19(2):e51 [FREE Full text] [doi: 10.2196/jmir.5834] [Medline: 28246072]
7. Rocha HM, Savatt JM, Riggs ER, Wagner JK, Faucett WA, Martin CL. Incorporating social media into your support tool box: points to consider from genetics-based communities. J Genet Couns 2018;27(2):470-480 [FREE Full text] [doi: 10.1007/s10897-017-0170-z] [Medline: 29130143]
8. Vasilica C, Oates T, Clausner C, Ormandy P, Barratt J, Graham-Brown M. Identifying information needs of patients with IgA nephropathy using an innovative social media-stepped analytical approach. Kidney Int Rep 2021;6(5):1317-1325 [FREE Full text] [doi: 10.1016/j.ekir.2021.02.030] [Medline: 34013110]
9. Long JC, Best S, Nic Giolla Easpaig B, Hatem S, Fehlberg Z, Christodoulou J, et al. Needs of people with rare diseases that can be supported by electronic resources: a scoping review. BMJ Open 2022;12(9):e060394 [FREE Full text] [doi: 10.1136/bmjopen-2021-060394] [Medline: 36581982]
10. Titgemeyer SC, Schaaf CP. Facebook support groups for rare pediatric diseases: quantitative analysis. JMIR Pediatr Parent 2020;3(2):e21694 [FREE Full text] [doi: 10.2196/21694] [Medline: 33211019]

11. Braun V, Clarke V. To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. Qualitative Research in Sport, Exercise and Health 2019;13(2):201-216. [doi: 10.1080/2159676x.2019.1704846]
12. Sim J, Saunders B, Waterfield J, Kingstone T. Can sample size in qualitative research be determined a priori? International Journal of Social Research Methodology 2018;21(5):619-634. [doi: 10.1080/13645579.2018.1454643]
13. Britt RK. Online social support for Participants of crohn's and ulcerative colitis groups. Health Commun 2017;32(12):1529-1538. [doi: 10.1080/10410236.2016.1234539] [Medline: 27819486]
14. Apperson A, Stellefson M, Paige SR, Chaney BH, Chaney JD, Wang MQ, et al. Facebook groups on chronic obstructive pulmonary disease: social media content analysis. Int J Environ Res Public Health 2019;16(20):3789 [FREE Full text] [doi: 10.3390/ijerph16203789] [Medline: 31600907]
15. Iyer AA, Barzilay JR, Tabor HK. Patient and family social media use surrounding a novel treatment for a rare genetic disease: a qualitative interview study. Genet Med 2020;22(11):1830-1837 [FREE Full text] [doi: 10.1038/s41436-020-0890-6] [Medline: 32601388]
16. Titgemeyer SC, Schaaf CP. Facebook support groups for pediatric rare diseases: cross-sectional study to investigate opportunities, limitations, and privacy concerns. JMIR Pediatr Parent 2022;5(1):e31411 [FREE Full text] [doi: 10.2196/31411] [Medline: 34989690]
17. Chen J, Wang Y. Social media use for health purposes: systematic review. J Med Internet Res 2021;23(5):e17917 [FREE Full text] [doi: 10.2196/17917] [Medline: 33978589]
18. Rupert DJ, Gard Read J, Amoozegar JB, Moultrie RR, Taylor OM, O'Donoghue AC, et al. Peer-generated health information: the role of online communities in patient and caregiver health decisions. J Health Commun 2016;21(11):1187-1197 [FREE Full text] [doi: 10.1080/10810730.2016.1237592] [Medline: 27805496]
19. Wang X, Shi J, Kong H. Online health information seeking: a review and meta-analysis. Health Commun 2021;36(10):1163-1175. [doi: 10.1080/10410236.2020.1748829] [Medline: 32290679]
20. Angelis A, Tordrup D, Kanavos P. Socio-economic burden of rare diseases: a systematic review of cost of illness evidence. Health Policy 2015;119(7):964-979. [doi: 10.1016/j.healthpol.2014.12.016] [Medline: 25661982]
21. Navarrete-Opazo AA, Singh M, Tisdale A, Cutillo CM, Garrison SR. Can you hear us now? The impact of health-care utilization by rare disease patients in the United States. Genet Med 2021;23(11):2194-2201 [FREE Full text] [doi: 10.1038/s41436-021-01241-7] [Medline: 34183788]
22. Chung CCY, Ng NYT, Ng YNC, Lui ACY, Fung JLF, Chan MCY, et al. Socio-economic costs of rare diseases and the risk of financial hardship: a cross-sectional study. Lancet Reg Health West Pac 2023;34:100711 [FREE Full text] [doi: 10.1016/j.lanwpc.2023.100711] [Medline: 37283971]
23. Merolli M, Gray K, Martin-Sanchez F. Health outcomes and related effects of using social media in chronic disease management: a literature review and analysis of affordances. J Biomed Inform 2013;46(6):957-969 [FREE Full text] [doi: 10.1016/j.jbi.2013.04.010] [Medline: 23702104]
24. Patel R, Chang T, Greysen SR, Chopra V. Social media use in chronic disease: a systematic review and novel taxonomy. Am J Med 2015;128(12):1335-1350. [doi: 10.1016/j.amjmed.2015.06.015] [Medline: 26159633]
25. Reblin M, Uchino BN. Social and emotional support and its implication for health. Curr Opin Psychiatry 2008;21(2):201-205 [FREE Full text] [doi: 10.1097/YCO.0b013e3282f3ad89] [Medline: 18332671]
26. Drageset J. Social support. In: Haugan G, Eriksson M, editors. Health Promotion in Health Care -Vital Theories and Research. Cham: Springer International Publishing; 2021:137-144.
27. Social determinants of health - healthy people 2030. Health.Gov. URL: https://health.gov/healthypeople/priority-areas/social-determinants-health [accessed 2022-11-10]
28. WHO. Social Determinants of Health. URL: https://www.who.int/health-topics/social-determinants-of-health [accessed 2022-11-10]
29. Krug E. It's Time to Harness the Power of Connection for Our Health and Well-Being.: World Health Organization; 2023. URL: https://www.who.int/news-room/commentaries/detail/it-s-time-to-harness-the-power-of-connection-for-our-health-and-well-being [accessed 2024-01-08]
30. U.S. Department of Health and Human Services. Social and Community Context - Healthy People 2030. Health.gov. URL: https://health.gov/healthypeople/objectives-and-data/browse-objectives/social-and-community-context [accessed 2024-01-08]
31. Raza AA, Naseem M, Qasmi NH, Randhawa S, Malik F, Taimur B, et al. Fostering engagement of underserved communities with credible health information on social media. : ACM; 2022 Presented at: Proceedings of the ACM Web Conference 2022; 2022 April 25-29; Lyon France p. 3718-3727 URL: https://dl.acm.org/doi/10.1145/3485447.3512267 [doi: 10.1145/3485447.3512267]
32. Vander Wal T. Folksonomy. 2007. URL: https://www.vanderwal.net/folksonomy.html [accessed 2024-01-26]
33. Jason Morrison P. Tagging and searching: search retrieval effectiveness of folksonomies on the world wide web. Information Processing & Management 2008;44(4):1562-1579. [doi: 10.1016/j.ipm.2007.12.010]
34. DeFrancis Sun B. Folksonomy and health information access: how can social bookmarking assist seekers of online medical information? Journal of Hospital Librarianship 2008;8(1):119-126. [doi: 10.1080/15323260801943570]

35. Fayn MG, des Garets V, Rivière A. Collective empowerment of an online patient community: conceptualizing process dynamics using a multi-method qualitative approach. BMC Health Serv Res 2021;21(1):958 [FREE Full text] [doi: 10.1186/s12913-021-06988-y] [Medline: 34511115]
36. van Uden-Kraan CF, Drossaert CHC, Taal E, Seydel ER, van de Laar MAFJ. Participation in online patient support groups endorses patients' empowerment. Patient Educ Couns 2009;74(1):61-69. [doi: 10.1016/j.pec.2008.07.044] [Medline: 18778909]
37. Kiesler S, Kraut RE, Resnick P, Kittur A, Burke M, Chen Y, et al. Regulating behavior in online communities. In: Building Successful Online Communities. Cambridge, Massachusetts: The MIT Press; 2011.

Abbreviations

DHHS: Department of Health and Human Services
DRSG: disease-related support group
HIPAA: Health Insurance Portability and Accountability Act
RD: rare disease
WHO: World Health Organization

Edited by A Kushniruk; submitted 27.02.24; peer-reviewed by A Mallett, J Long; comments to author 29.05.24; revised version received 24.06.24; accepted 08.11.24; published 30.12.24.

Please cite as:
Doyle TA, Vershaw SL, Conboy E, Halverson CME
Improving Social Media-Based Support Groups for the Rare Disease Community: Interview Study With Patients and Parents of Children with Rare and Undiagnosed Diseases
JMIR Hum Factors 2024;11:e57833
URL: https://humanfactors.jmir.org/2024/1/e57833
doi:10.2196/57833
PMID:

Older Adults' Acceptance of a Virtual Reality Group Intervention in Nursing Homes: Pre-Post Study Under Naturalistic Conditions

Yijun Li[1], MSc; Irina Shiyanov[2], BSc; Beate Muschalla[1], Prof Dr

[1]Department of Psychotherapy and Diagnostics, Institute of Psychology, Technische Universität Braunschweig, Humboldtstraße 33, Braunschweig, Germany

[2]VirtuaLounge GmbH, Braunschweig, Germany

Corresponding Author:
Yijun Li, MSc
Department of Psychotherapy and Diagnostics, Institute of Psychology, Technische Universität Braunschweig, , Humboldtstraße 33, Braunschweig, , Germany

Abstract

Background: Virtual reality (VR) group activities can act as interventions against inactivity and lack of meaningful activities in nursing homes. The acceptance of VR among older adults has been explored from different perspectives. However, research on the impact of older adults' individual characteristics on the acceptance of VR group activities in nursing homes is necessary.

Objective: This study investigates the impact of individual characteristics (eg, psychosocial capacities) on VR acceptance among older adults in nursing homes, as well as this group's perceptions of VR after participating in a VR intervention.

Methods: In this pre-post study conducted in nursing homes, we applied a VR group intervention with 113 older adult participants. These participants were categorized into two groups based on their naturalistic choice to join the intervention: a higher VR acceptance group (n=90) and a lower VR acceptance group (n=23). We compared the two groups with respect to their sociodemographic characteristics, psychosocial capacities, and attitudes toward new technologies. Additionally, we examined the participants' perceptions of VR.

Results: The results show that those with lower acceptance of VR initially reported higher capacities in organizing daily activities and stronger interpersonal relationships compared to older adults with higher VR acceptance. The VR group activity might hold limited significance for the latter group, but it offers the chance to activate older adults with lower proactivity. Openness to new technology was associated with a favorable perception of VR. After the VR intervention, the acceptance of VR remained high.

Conclusions: This study investigates the acceptance of VR group events as meaningful activities for older adults in nursing homes under naturalistic conditions. The results indicate that the VR group intervention effectively addressed low proactivity and interpersonal relationship issues among older adults in nursing homes. Older adults should be encouraged to experience VR if the opportunity to participate is offered, potentially facilitated by caregivers or trusted individuals.

(JMIR Hum Factors 2024;11:e56278) doi:10.2196/56278

KEYWORDS

virtual reality; VR; computer-generated simulation; simulation; technology acceptance; nursing home; nursing facility; long-term care center; long-term care facility; older adult; elder; elderly; older person; older people; senior; understanding human behavior; meaningful activity; group intervention; human behavior

Introduction

The absence of meaningful activities and a lack of social interaction can lead to loneliness in older adults in nursing homes [1,2]. Meaningful activities are defined as those that hold personal significance or offer enjoyment to individuals, aligning with their present and past interests, routines, habits, and roles [3,4]. In the context of nursing homes, a meaningful activity could include household tasks like cooking, which older adults may no longer be able to do on their own. A lack of meaningful activity leads to older adults' inactivity, resulting in mental and physical impairment, as well as an increased risk of mortality [5-8]. Furthermore, several studies show that older adults in nursing homes experience more loneliness compared to those living in the community, even though they are often surrounded by other residents and caregivers [9,10]. Addressing these issues requires innovative solutions [8,11,12]. Virtual reality (VR) has emerged as a promising solution in rehabilitation [13,14], including for older adults [1,15,16]. VR technology, especially fully immersive VR technology, delivers a comprehensive and lifelike experience, creating a strong sense of presence for users by using head-mounted displays and motion-tracking controllers. These devices work in tandem to simulate a realistic, interactive environment, allowing users to see, hear, and interact with the virtual world in a manner similar to real-world experiences [17].

Various studies have shown that VR interventions can enhance cognitive capacities [18,19], improve physical strength [16,18,20]—for example, via walking training [21]—and enhance the overall well-being of older adults [16,19,21-23]. Moreover, the immersive nature of VR not only conserves resources and reduces costs but also ensures safety [24-26]. Thus, VR can be a viable choice for providing meaningful activities within nursing homes to enhance the daily activities, capacities, social activities, and well-being of older adults in nursing homes [27].

In a study on the acceptance of VR technology that included 76 older adults, it was found that participants developed a positive attitude toward VR after using it [17]. A systematic review and meta-analysis found that in most of the 54 relevant studies, older adults reported pleasant experiences with VR and expressed a desire to use it again [28]. In addition to assessing the level of acceptance of VR among older adults, it is also necessary to understand the factors that could influence this acceptance [29]. According to the Technology Acceptance Model (TAM) [30], the acceptance of a technology can be explained and positively predicted by perceived usefulness and perceived ease of use. In the context of our study, the implementation of VR technology should address specific needs, such as providing meaningful activities and combating loneliness, and the design of the VR intervention should be in alignment with the capacities of nursing home residents to ensure the perceived ease of use of the intervention. However, this requires an in-depth understanding of the characteristics of the target population. The field of gerontechnology has advanced in developing frameworks to address the unique capacities and limitations of older adults. A notable gap in the previous acceptance models [31] is the insufficient exploration of personal characteristics. In addition to collecting demographic factors, research should explore participants' cognitive capacities, social relationships, environmental influences, and psychosocial traits. Previous research about VR acceptance among older adults has similarly highlighted the importance of individual characteristics such as physical constraints, educational attainment, and socioeconomic status as predictive factors [32,33]. A qualitative analysis about the acceptance of technology among older adults [34] found noticeable differences in attitude linked to participants' educational background and work experiences, but no definitive differences were found with regard to gender and ages. Older age negatively impacted the willingness of participants to use robots [35]. Further studies are needed to confirm the acceptability of different types of immersive technology devices [28], including a detailed report on VR interventions [29]. An in-depth investigation into personal characteristics, including how the sociodemographic factors and psychosocial capabilities of nursing home residents affect their acceptance of VR group activities, will be valuable for understanding these dynamics and contributing to future design improvements.

Besides sociodemographic characteristics, technological engagement is another important factor for individual technological acceptance. Empirical findings underscore the pivotal role of openness to new experiences as a predictor of improved digital acceptance [36]. Individuals with a higher motivation for self-actualization tend to be more receptive to new encounters and the acquisition of novel skills and ideas [37]. This propensity for continuous learning aligns with the potential for individuals to actualize their personal capacities through the acquisition of new technological skills [38]. In addition, prior knowledge about VR has been identified as a decisive factor in VR acceptance [32]. This finding resonates with the continuity theory, which posits that older adults make adaptive choices to maintain ties with their past experiences [39]. Adhering to established habits rooted in past experiences serves to mitigate the uncertainty that may accompany new environments. This preference for continuity extends to the perpetuation of personal cognitive frameworks shaped by past preferences. Attitudes toward immersive VR have been observed to change from neutral to positive after the first exposure [17]. Therefore, older adults with higher technical engagement prior to the study may have a higher acceptance of VR-based group activities.

Anxiety toward new technologies, which significantly influences technology acceptance, is rooted in factors like unfamiliarity with computers, perceived uncertainty, fear, and a general apprehension toward making mistakes [40,41]. There is a significant correlation between elevated levels of self-efficacy and reduced anxiety with increased utilization of gerontechnology [42]. In essence, this implies that for older adults to embrace novel technology, they must overcome the fear of uncertainty by taking the initial step of trying it out, preferably with sufficient support. Users tend to experience contentment with new technology once they start using it [43-45]. Therefore, after participating in a VR intervention, the acceptance of new technologies, especially VR technology and VR group events, should be maintained or even improved.

This study focuses especially on the impact of individual characteristics of older adults on the acceptance of VR-based group activities in nursing homes. A series of VR interventions was conducted to address the residents' lack of meaningful activities and loneliness. Over the course of 4 VR intervention sessions, older adults engaged in serious games involving tasks that they may no longer be able to perform within the nursing home environment, such as cooking and gardening. This aligns with the concept of meaningful activities. This study aimed to investigate whether acceptance of VR is linked to personal characteristics, such as sociodemographic background, psychosocial capacities for nursing home residency, and technological engagement. Therefore, the research question and hypotheses of this paper are the following:

- Research question: What are the differences in sociodemographic status and psychosocial capacities between older adults with higher and lower acceptance of using a VR intervention for meaningful activity in nursing homes under naturalistic conditions?
- Hypothesis 1: The group exhibiting higher acceptance of a VR group activity reports a higher level of technological engagement compared to the group with lower VR acceptance.
- Hypothesis 2: After the VR group intervention, technological engagement among older adults will either be maintained or improved.

- Hypothesis 3: After the VR group intervention, the acceptance of VR technology and willingness to participate will both align at a high level.

Our research addresses critical gaps in the existing literature. First, our investigation focused on a specific population and implementation of VR technology: older adults living in nursing homes and VR group events as meaningful activities, respectively. Second, the study focused on differences in personal capacities and was not limited to demographic background. Third, the study was done under naturalistic conditions with high ecological validity, that is, using VR in a group event in a nursing home and collecting naturalistic data. This investigation holds the potential to provide a more nuanced classification and predictive understanding of the specific demographics within nursing homes. It can pave the way for the future development of VR programs that are precisely tailored to the individual needs and preferences of older adults in nursing homes.

Methods

Research Design and Recruiting

We conducted a VR group intervention study as a pre-post observation study in naturalistic settings with older adults in 14 nursing homes in a city of 250,000 inhabitants in Germany.

In a first step, the nursing homes were contacted via telephone. After receiving confirmation of willingness to participate from the institution, an email with the participant selection criteria was sent to the social coordinator of each nursing home. The older adults were selected by the nursing home caregivers based on the following criteria: (1) the participant should be older than 60 years; (2) they can use at least one arm and hand for interacting with the system; (3) they can see and hear, with assistive devices such as glasses allowed; (4) their cognitive and mental capacity are sufficient for individual interviews; (5) they do not have conditions that could be triggered by VR, such as epilepsy. In addition to meeting these criteria, caregivers asked the older adults about their willingness to participate in the VR intervention. Furthermore, if needed, permission was sought from legal guardians.

Procedure

In the initial study week, a structured individual interview was conducted with the older adults by a psychologist. Baseline sociodemographic characteristics, capacities, and technological inclination of the older adults were assessed (T1). Following a warm-up interview in the second week (T2), the subsequent 4 weeks included repeated exposure to VR group activities (T3-T6). During these sessions, the older adults were organized into groups, with each group consisting of a maximum of 5 members. These groups remained consistent throughout the entire VR intervention. The older adults were expected to accomplish designated tasks individually (Figure 1). After the VR intervention, participants exchanged thoughts about their experiences in small groups. Moreover, an immersive VR video was provided for relaxing after each intervention. After completion of all VR interventions, a posttest was used to assess the older adults' technological engagement (T7).

Participants declining VR intervention involvement were offered participation in a control group that underwent an identical measurement and interview procedure excluding the VR intervention (Table 1).

Figure 1. Older adults participating in a virtual reality group event.

Table . Procedure of the VR[a] group intervention study in nursing homes.

	Week 1	Week 2	Weeks 3 - 6	Week 7	Week 10
Event	Baseline (T0)	Warm-up (T1)	VR interventions (T2-T5)	Posttest (T6)	Follow-up (T7)
Content	Interview: • Demographic information • Mini-ICF-APP • Technological engagement • Perception of VR	Short interview	VR interventions	Interview: • Technological engagement • Perception of VR	Interview

[a]VR: virtual reality.

Participants

To answer our research questions, we needed to compare participants with higher and lower acceptance of the VR intervention. The necessary sample size for group comparison with a t test for independent groups, with a medium effect size d=0.5, an α level of .05, and a power of $1–\beta$=0.80, was calculated to be 102 participants, with 51 participants in each group.

We collected data from 129 participants, of which 113 were relevant to this analysis. The composition of the groups, based on the naturalistic characteristics of the participants (higher or lower VR intervention acceptance), was unpredictable, resulting in an unequal group size.

Initially, a total of 129 older adults in nursing homes participated in the VR intervention study, of which 12 opted for the control group and 117 opted for the intervention group. Among these participants, 27 individuals in the intervention group discontinued their involvement in the VR intervention. The reasons for dropout and nonparticipation in the intervention group were categorized as non–motivation-related (eg, health, life status; n=16) or motivation-related (n=11). Non–motivation-related reasons included illness (switch to hospital stay), limited station participation (no group established), and cybersickness. The non–motivation-related dropouts were not investigated further in this study. Motivation-related reasons included diminished interest after the initial interview and having a preference for an alternate activity. In the end, there were 90 older adults in the intervention group who completed the posttest.

The sample that was analyzed to answer our research questions included 113 participants, that is, the initial sample (n=129) reduced by the participants who dropped out due to nonmotivational reasons (n=16). The investigated sample was categorized into two groups: individuals with higher or lower acceptance of the VR intervention. The higher acceptance group encompassed the participants (n=90) who actively participated in the VR interventions and successfully completed the subsequent postinterview. The lower acceptance group (n=23) comprised those who dropped out of the intervention group due to motivation-based reasons and older adults who initially were not interested (control group). The higher VR acceptance group was younger (mean 80.13, SD 8.39 years) than the lower VR acceptance group (mean 83.74, SD 8.31 years), but this was not significant (t_{111}=–1.84; *P*=.07; *d*=–.43). For comprehensive sociodemographic details of these two groups, refer to Table 2.

Table . Sociodemographic data of older adults who participated in the VR[a] intervention.

		Higher VR acceptance[b] (n=90), n (%)	Lower VR acceptance[c] (n=23), n (%)	Chi-square (*df*)	*P* value
Sex (female)		59 (65.6)	18 (78.3)	1.36 (1)	.24
Education				8.61 (5)	.13
	None	6 (6.7)	3 (13)		
	Primary school	1 (1.1)	0 (0)		
	Lower secondary school (ninth or tenth grade)	57 (63.3)	11 (47.8)		
	Upper secondary school	16 (17.8)	9 (39.1)		
	A-levels	10 (11.1)	0 (0)		
Professional qualification				1.61 (3)	.68
	None	27 (30)	7 (30.4)		
	Apprenticeship or skilled work	52 (57.8)	15 (65.2)		
	Master craftsman	6 (6.7)	1 (4.3)		
	University studies	5 (5.6)	0 (0)		
Longest professional activity in working life				3.04 (6)	.80
	Crafts, industry, production	30 (33.3)	6 (27.3)		
	Research and development	3 (3.3)	0 (0)		
	Agriculture	1 (1.1)	1 (4.5)		
	Office, management	21 (23.3)	5 (22.7)		
	Service, gastronomy, customer service	20 (22.2)	7 (31.8)		
	Practical health care (eg, nurse, doctor, therapist)	7 (7.8)	1 (4.5)		
	Housewife	8 (8.9)	2 (9.1)		
Frequency of visits from trusted people				4.85 (5)	.43
	Several times per week	44 (48.9)	16 (69.9)		
	Weekly	23 (25.6)	2 (8.7)		
	Every 2-3 weeks	8 (8.9)	2 (8.7)		
	Monthly	2 (2.2)	0 (0)		
	Less than monthly	2 (2.2)	1 (4.3)		
	No regular contacts	11 (12.2)	2 (8.7)		
No previous experience with VR		82 (91.1)	17 (94.4)	0.218 (1)	.64

[a]VR: virtual reality.

[b]Higher VR acceptance group included the older adults who participated in at least the VR intervention and the posttest.

[c]Lower VR acceptance group included the older adults who did not want to join the VR intervention at the beginning (control group) and motivation-related dropouts; one of the motivation-related dropouts was excluded due to an incomplete baseline interview.

Stimuli and Equipment

The VR intervention was done with a virtual vacation home scenario. During each VR session, the participants' tasks were to complete 4‐5 activities such as gardening, crafting, and baking (see Figures 2 and 3 for examples). Participants started by crafting a chair, then arranged a garden around it, and finally, they could enjoy sitting on the chair while tending to the garden's plants. In the virtual environment, participants worked on tasks individually. After removing their VR headsets, they had the opportunity to discuss their experiences with each other in a group setting. Every VR event lasted approximately 20 - 30 minutes, depending on the tasks assigned for each session.

The Pico Neo 3 Pro VR headset and Pico Neo 3 controller were used. The resolution of the head-mounted display was 1832 × 1920 pixels per eye, with a refresh rate of 72 Hz and 6-degrees-of-freedom inside-out tracking.

Figure 2. Virtual reality gardening task [27].

Figure 3. Virtual reality task in the kitchen: baking a pizza [27].

Data Collection

Psychosocial and cognitive capacities were assessed with the Mini-ICF-APP scale [46]. In this study, the reference context for capacity assessment was daily life in the nursing home. The Mini-ICF-APP covers the following capacity dimensions: (1) adherence to regulations, (2) planning and structuring of tasks, (3) flexibility, (4) competence and knowledge application, (5) ability to make decisions and judgments, (6) proactivity and spontaneous activities, (7) endurance, (8) self-assertiveness, (9) contact with others, (10) group integration, (11) intimate relationships, (12) self-care, and (13) mobility. Each dimension is rated on an 8-point scale (from 0 being "This is a strength of mine," to 7 indicating "This is impossible for me"). The Mini-ICF-APP is a heterogeneous scale, covering different psychosocial capacities. Each item can be interpreted individually because all items reflect different capacities. If a mean score across the 13 items is calculated, this can be interpreted as a global capacity impairment level. The Mini-ICF-APP is a standard scale for measuring psychosocial capacities that has already been validated in different languages [46-48]. The interrater reliability in our study ranged from r=0.446 (untrained) to r=0.910 (trained).

The structured questionnaire addressing technology engagement asked for the frequency of engagement with 4 distinct technologies: televisions, smartphones, PCs or tablets, and other novel technologies. Responses were recorded on a 5-point Likert scale, spanning from 1=never/not at all to 5=several times every day/very much.

To assess the acceptance of VR technology, two questions were asked, one about the participant's overall perspective regarding VR and another about the inclination to participate in a VR group activity again. Each item was rated on a scale from 1=not at all to 5=very good/very much.

The VR group activities and the questionnaires were conducted in the nursing home, integrated as a daily event for the participants. This is the basis for ecological validity of the data.

Statistical Analysis Plan

Older adults with higher and lower acceptance of using a VR intervention were compared. Initially, the higher and lower acceptance groups were defined. Participants were categorized into four groups: (1) no interest in the VR intervention but willing to participate in the interview (lower VR acceptance), (2) dropped out due to motivational reasons (lower VR acceptance), (3) dropped out due to nonmotivational reasons, and (4) completed the VR intervention program and the postinterview (higher VR acceptance). Participants who dropped out for nonmotivational reasons were excluded from the study. Categories 1 and 2 were allocated to the "lower VR acceptance" group, while category 4 represented the "higher VR acceptance" group.

Sociodemographic data, psychosocial capacities, and technological engagement of the older adults were collected from both groups to enable us to answer the research question and hypothesis 1. Additionally, data on technological engagement and attitudes toward VR technology and VR intervention as a group activity were collected from the posttests of the higher VR acceptance group to analyze changes in acceptance within this group to provide answers to hypotheses 2 and 3.

Data were analyzed with the statistical software SPSS (version 29; IBM Corp [49]). We conducted t tests and χ^2 tests for group comparisons of sociodemographic factors. Psychosocial capacities and technological engagement between the higher and lower VR acceptance group were analyzed with a two-tailed t test. We used a t test for paired samples to compare the degree of technological engagement within the higher acceptance group at baseline and post interview.

Ethical Considerations

This research was funded by the German Federal Ministry of Education and Research (BMBF), project number 16SV8561 VRalive. This research was approved by the ethics committee at Technische Universität Braunschweig (FV-2020 - 18). The study was preregistered in Deutsches Register Klinischer Studien on December 11, 2020. Informed consent, confidentiality, and data protection agreements were obtained from older adults or their legal guardian under the supervision of a caregiver. There was no compensation for participation. All activities conducted in the nursing home strictly adhered to the prevailing nursing home COVID-19 prevention and treatment policy. The VR goggles were diligently disinfected after each use.

Results

There were no statistically significant differences (all P values >.05) in sociodemographic characteristics between both groups (Table 2).

Concerning psychosocial capacities (Table 3), older adults with lower VR acceptance were more proactive and had more robust social relationships in comparison with the older adults in the higher acceptance group.

Table . Self-reported impairment in psychosocial capacities according to the Mini-ICF-APP scale (0=that is clearly a strength of mine, 7=I am not able at all).

	Higher virtual reality acceptance group (n=90)		Lower virtual reality acceptance group (n=23)		*t* test (*df*)	*P* value	Effect size (*d*)
	Rating, mean (SD)	Impaired, % (rating≥5)	Rating, mean (SD)	Impaired, % (rating≥5)			
Adjustment to rules and routines	2.59 (0.89)	2 (2.2)	2.30 (0.97)	1 (4.3)	1.35 (111)	.18	0.31
Planning and structuring tasks	3.14 (1.83)	19 (21.1)	3.17 (1.85)	6 (26.1)	–0.07 (111)	.95	–0.02
Flexibility and adaptability	2.40 (0.92)	2 (2.2)	2.35 (1.30)	1 (4.3)	0.18 (27.89)	.86	0.05
Competence and knowledge application	2.27 (1.23)	5 (5.6)	2.04 (1.33)	1 (4.3)	0.76 (111)	.45	0.18
Ability to make decisions and judgments	2.63 (1.19)	8 (8.9)	2.39 (0.89)	0 (0)	0.91 (111)	.37	0.21
Proactivity and spontaneous activities	2.43 (1.20)	5 (5.6)	1.78 (1.44)	1 (4.3)	2.22 (111)	.03	0.52
Resilience and perseverance	2.56 (1.11)	3 (3.3)	2.57 (1.56)	3 (13)	–0.03 (111)	.97	–0.01
Self-assertiveness	2.60 (1.07)	5 (5.6)	2.13 (1.14)	1 (4.3)	1.86 (111)	.07	0.43
Contact with others	2.44 (1.43)	9 (10)	2.39 (1.44)	2 (8.7)	0.16 (111)	.87	0.04
Group integration	2.68 (1.36)	9 (10)	2.65 (1.77)	2 (8.7)	0.06 (28.99)	.95	0.02
Dyadic relationships	2.62 (1.63)	13 (14.4)	1.61 (1.56)	1 (4.3)	2.69 (111)	.008	0.63
Self-care	3.27 (1.71)	19 (21.1)	3.78 (2.43)	11 (47.8)	–0.96 (27.84)	.35	–0.27
Mobility	2.47 (1.49)	7 (7.8)	2.17 (1.30)	1 (4.3)	0.86 (111)	.39	0.20
Average score	2.62 (0.77)	5 (5.6)	2.41 (0.89)	1 (4.3)	1.14 (111)	.26	0.27

Older adults with low VR acceptance experience more self-care impairments and report a conservative stance toward adopting new technology (Tables 3 and 4). In contrast, those with higher VR acceptance report being less active, with less meaningful social connections, but with higher self-care capacity and a more open attitude toward new technology (Tables 3 and 4). There is no statistically significant difference in engagement with televisions (*P*=.62) and laptops (*P*=.25) between higher and lower acceptance groups (Table 4). In both groups, over 90% of participants use a television at least once per day. A contrast emerged in smartphone use, with 23.3% (21/90) of the higher acceptance group using it daily, while merely 4.3% (1/23) of the lower acceptance group do the same. These results support hypothesis 1, which postulated that the group exhibiting higher VR acceptance would report a higher level of technological engagement compared to the group with lower VR acceptance.

Table . Technological engagement of lower and higher VR[a] acceptance groups at baseline.

	Higher VR acceptance (n=90)		Lower VR acceptance (n=23)		*t* test (*df*)	*P* value	Effect size (*d*)
	Rating, mean (SD)	Frequent use and high willingness, n (%) with rating≥4	Rating, mean (SD)	Frequent use and high willingness, n (%) with rating≥4			
Television[b]	4.46 (0.91)	84 (93.3)	4.35 (1.03)	21 (91.3)	0.49 (111)	.62	0.11
Smartphone[b]	1.91 (1.60)	21 (23.3)	1.17 (0.83)	1 (4.3)	3.04 (67.88)	.003	0.50
Laptop/PC[b]	1.68 (1.45)	16 (17.8)	1.35 (1.15)	2 (8.7)	1.16 (41.77)	.25	0.24
Other technologies[c]	3.31 (1.72)	50 (55.6)	1.70 (1.29)	3 (13)	4.97 (44.11)	<.001	0.98

[a]VR: virtual reality.

[b]Participants answered the question “How often do you use [technology]?” (1=never, 5=several times per day).

[c]Participants answered the question “Do you want to try other new technologies?” (1=not at all, 5=very much).

After their participation in the VR intervention (Tables 5 and 6), older adults’ perspectives toward VR remained very positive and even increased, with many participants reporting looking forward to the next VR group event; therefore, hypothesis 2 (“After the VR group intervention, technological engagement among older adults will either be maintained or improved”) and hypothesis 3 (“After the VR group intervention, the acceptance of VR technology and willingness to participate will both align at a high level”) were confirmed as well.

Table . Technological engagement before and after a virtual reality intervention among higher virtual reality acceptance group participants (n=90).

	Baseline		Posttest		*t* test (*df*)	*P* value	Effect size (*d*)
	Rating, mean (SD)	Frequent use and high willingness, n (%) with rating≥4	Rating, mean (SD)	Frequent use and high willingness, n (%) with rating≥4			
Television[a]	4.46 (0.91)	84 (93.3)	4.49 (0.91)	84 (93.3)	–0.48 (89)	.63	–0.05
Smartphone[a]	1.91 (1.60)	21 (23.3)	1.77 (1.50)	16 (17.8)	1.37 (89)	.17	0.14
Laptop/PC[a]	1.68 (1.45)	16 (17.8)	1.60 (1.36)	14 (15.6)	1.26 (89)	.21	0.13
Other new technologies[b]	3.31 (1.72)	50 (55.6)	3.31 (1.67)	49 (54.4)	0.0 (89)	>.99	0.00

[a]Participants answered the question “How often do you use [technology]?” (1=never, 5=several times per day).

[b]Participants answered the question “Do you want to try other new technologies?” (1=not at all, 5=very much).

Table . Acceptance of VR[a] following a VR intervention among higher VR acceptance group participants (n=88).

	Rating, mean (SD)	Positive responses, n (%) with rating≥4
How do you like VR technology in general? (1=not at all, 5=very much)	4.42 (1.07)	75 (85.2)
Would you like to participate in another VR activity in the future? (1=not at all, 5=very much)	4.22 (1.39)	68 (77.3)

[a]VR: virtual reality.

Discussion

Demographic Status

There were no differences in demographic status between the groups with higher and lower VR acceptance. This finding may seem to contradict previous research that suggested that factors such as age and education could predict the acceptance of new technology or VR [32,50]. The divergence in outcomes can be explained by the homogeneous sample in our study [51]. The participants in our study were selected based on their care requirements, creating a convergence of demographic and life status. The data from this naturalistic explorative study suggest that demographic factors have no significant impact on older adults’ VR group event participation.

Psychosocial Capacities

Individuals within the lower acceptance group had a higher reported level of organizing daily activities and building dyadic relationships with trusted individuals. A potential reason for nonparticipation in VR activities among older adults in nursing homes is motivation-related. This supports the idea that new technology should align well with users' life statuses and requests [52]. Among older adults, the adoption of technology for health purposes often aims to compensate for deficiencies [53]. Older adults with robust relationships may be better integrated with others and can meet their need for meaningful activities in their daily lives. Additionally, their capacity to organize activities may contribute to their lower intrinsic motivation for new activities, leading to a reduced perception of VR usefulness and subsequently decreased acceptance. The Senior Technology Acceptance Model [50] states that older adults with strong social relationships believed in the utility of technologies and were more inclined to use them compared to older adults with weaker social networks. Moreover, a study involving 31 participants conducted semistructured interviews, indicating that socially well-integrated older adults may not perceive a necessity to use socially assistive augmented reality systems [54]. The VR intervention in our project is oriented toward meaningful activities and entertainment, which may be less attractive for those older adults who are already engaged in social relationships within the nursing home. Therefore, the target group of VR intervention could be older adults with a lack of proactivity. These results refine the target group for VR group interventions in nursing homes, that is, older adults who are less active and have limited meaningful social networks. This aligns with the core objective of our VR study—to offer meaningful activities that enhance the activity level, capacities, and overall well-being of less active older adults.

Technological Engagement and Attitude After VR Activities

Older adults with higher acceptance of VR technology also tend to demonstrate a stronger willingness to embrace new technological advancements. This indicates that, even in the context of VR as a group activity for older adults in nursing homes, an openness to new technology can be a predictor of higher acceptance.

Considering device usage, several factors warrant attention. First, television is already widely accepted throughout society, including among older adults in nursing homes [2]. Therefore, the acceptance of television was high in both groups.

Second, costs play a significant role [55]. Although the higher VR acceptance group showed greater engagement with smartphones, the actual usage rate remained minimal. One explanation may be that some older adults are facing financial constraints that hinder access to smartphones and PCs. Socioeconomic status is a factor that should not be neglected as a potential contributor to the acceptance of VR activities among older adults [32]. The Unified Theory of Acceptance and Use of Technology model by Venkatesh et al [56] expands on the TAM [30] by incorporating additional external factors such as social influence and facilitating conditions. Social influence refers to the extent to which an individual perceives that people they are close with believe he or she should adopt or utilize the new system. Facilitating conditions are characterized by the individual's perception of the organizational and technical infrastructure available to support system usage. Social support is crucial in affording older adults the opportunity to access new technology and in facilitating their learning process, consequently bolstering their technological engagement [44,50]. If nursing homes could provide access to new technologies such as VR, this could potentially enhance technology engagement among older adults.

Third, the difference in engagement between laptops and smartphones may be attributed to perceptions of usability, ease of use, and cost considerations. Smartphones, with their convenient accessibility and added functionality (such as phone calls), are perceived as more useful among older adults in nursing homes. This aligns with one of the key concepts of the TAM: the perception of usefulness. Therefore, it is crucial to reiterate that to enhance the acceptance of VR group activities as meaningful experiences for nursing home residents, these activities should be tailored to meet the needs and preferences of older adults.

Furthermore, after the intervention, a positive attitude toward VR was evident in participants' responses. This aligns with the idea that users often experience contentment with new technology once they use it [44,45], as well as with the Senior Technology Acceptance Model framework [50]. This suggests that more older adults, even those with lower VR acceptance, should be encouraged to give VR a chance. Such efforts could be facilitated by trusted individuals such as long-term social workers within nursing homes.

Limitations, Potential Applications, and Future Research

There are several limitations to this study. First, there was a small group of older adults with lower technological acceptance. The unequal sample sizes of the two groups resulted from the naturalistic group formation. The problem with unequal group sizes in research studies is that this can introduce bias and affect the statistical validity of the results. Additionally, unequal group sizes may impact the power of statistical tests, potentially making it more challenging to detect true differences between groups if the sample sizes are not balanced.

Second, there could be selection bias during recruitment. The initial selection of older adults was done by the caregivers, and their choices could be influenced by their expectations about the older adults. Additionally, the older adults who chose to participate in the study already had a willingness to experience the VR activity. Therefore, the results may not represent the entire population of nursing home residents, particularly those who declined to participate. This limits the external validity of the study.

Third, the behavior and responses of the older adults could be influenced by the Hawthorne effect, leading individuals to alter their actions based on the perceived expectations of the researchers. Moreover, the older adults might have enjoyed talking to the interviewer and could have overrated their attitudes toward the VR activity.

Lastly, a methodologically robust randomized controlled trial should be conducted to make conclusions about intervention effects possible.

On the other hand, this study offers a perspective from the implementation of VR group activities in a real-life setting, thus boasting high ecological validity. It specifically focuses on the acceptance of defined VR applications for residents in nursing homes as meaningful activities. This enables a better understanding of the factors influencing attitudes toward VR group activities and provides an opportunity to address the needs of the target population.

This study provides several insights regarding future VR intervention acceptance among older adults in nursing homes and rehabilitation facilities. First, demographic status does not impact the acceptance of a VR group activity among older adults in nursing homes. Second, this study underscores the importance of targeting specific groups and acknowledging individual differences in characteristics and needs. This research suggests that the target population for VR group activities in nursing homes should be residents lacking proactive capacity and social relationships. Future studies should address the particular needs of this population. Moreover, the findings emphasize the need for enhanced social support to boost technological engagement among older adults, thereby promoting greater acceptance of digital interventions to address their needs. This may involve nursing homes providing increased access to VR and other new technologies, or nursing staff fostering trust and offering encouragement to residents to participate in VR activities. Finally, future research in the domain of older adults' technology acceptance could specify capacities such as activities of daily living or focus on specific subgroups within nursing homes.

Conclusion

This study explored characteristics of older adults in nursing homes with varying levels of VR acceptance and their perceptions of VR after participation in a VR group activity. The study found no sociodemographic differences between older adults with higher or lower acceptance of VR activities; however, the findings suggest tailoring VR interventions to older adults who are less proactive. This is in line with the purpose of the VR group event in this study—to improve activity, psychosocial capacities, and well-being for inactive older adults in nursing homes. The data suggest that it can be fruitful to motivate older adults, including those with apparently lower technology acceptance, to use the opportunity to experience VR, potentially facilitated by the support of their trusted individuals.

Acknowledgments

We thank Carlotta Wilke, MSc, for her contributions to the intervention and data collection. This research was funded by the German Federal Ministry of Education and Research (BMBF), project number 16SV8561 VRalive.

Data Availability

The data were originally collected from 14 nursing homes. The data are available from the authors upon request.

Authors' Contributions

YL conducted the study in nursing homes with technical support from IS. YL collected and analyzed the data, prepared the tables, and wrote the manuscript. BM designed the research question, supervised the research process, and contributed to writing and revision of the manuscript.

Conflicts of Interest

IS is the cofounder of the company VirtuaLounge, which developed the virtual reality program used in this study. It is possible that the developed program will be used commercially in the future.

References

1. Benjamin K, Edwards N, Ploeg J, Legault F. Barriers to physical activity and restorative care for residents in long-term care: a review of the literature. J Aging Phys Act 2014 Jan;22(1):154-165. [doi: 10.1123/japa.2012-0139] [Medline: 23434919]
2. Tak SH, Kedia S, Tongumpun TM, Hong SH. Activity engagement: perspectives from nursing home residents with dementia. Educ Gerontol 2015 Mar;41(3):182-192. [doi: 10.1080/03601277.2014.937217] [Medline: 25489122]
3. Groenendaal M, Loor A, Trouw M, Achterberg WP, Caljouw MAA. Perspectives of healthcare professionals on meaningful activities for persons with dementia in transition from home to a nursing home: an explorative study. Healthcare (Basel) 2019 Aug 19;7(3):98. [doi: 10.3390/healthcare7030098] [Medline: 31430855]
4. Tierney L, Beattie E. Enjoyable, engaging and individualised: a concept analysis of meaningful activity for older adults with dementia. Int J Older People Nurs 2020 Jun;15(2):e12306. [doi: 10.1111/opn.12306] [Medline: 32022414]
5. Maher JP, Conroy DE. Daily life satisfaction in older adults as a function of (in)activity. J Gerontol B Psychol Sci Soc Sci 2017 Jul 1;72(4):593-602. [doi: 10.1093/geronb/gbv086] [Medline: 26405078]

6. Leigh-Hunt N, Bagguley D, Bash K, et al. An overview of systematic reviews on the public health consequences of social isolation and loneliness. Public Health (Fairfax) 2017 Nov;152:157-171. [doi: 10.1016/j.puhe.2017.07.035] [Medline: 28915435]
7. Vossius C, Selbæk G, Šaltytė Benth J, Bergh S. Mortality in nursing home residents: a longitudinal study over three years. PLoS One 2018;13(9):e0203480. [doi: 10.1371/journal.pone.0203480] [Medline: 30226850]
8. Wu B. Social isolation and loneliness among older adults in the context of COVID-19: a global challenge. Glob Health Res Policy 2020;5:27. [doi: 10.1186/s41256-020-00154-3] [Medline: 32514427]
9. Victor CR. Loneliness in care homes: a neglected area of research? Aging Health 2012 Dec;8(6):637-646. [doi: 10.2217/ahe.12.65]
10. Gardiner C, Laud P, Heaton T, Gott M. What is the prevalence of loneliness amongst older people living in residential and nursing care homes? A systematic review and meta-analysis. Age Ageing 2020 Aug 24;49(5):748-757. [doi: 10.1093/ageing/afaa049] [Medline: 32396600]
11. Balki E, Hayes N, Holland C. Effectiveness of technology interventions in addressing social isolation, connectedness, and loneliness in older adults: systematic umbrella review. JMIR Aging 2022 Oct 24;5(4):e40125. [doi: 10.2196/40125] [Medline: 36279155]
12. MacNeil M, Hirslund E, Baiocco-Romano L, Kuspinar A, Stolee P. A scoping review of the use of intelligent assistive technologies in rehabilitation practice with older adults. Disabil Rehabil Assist Technol 2024 Jul;19(5):1817-1848. [doi: 10.1080/17483107.2023.2239277] [Medline: 37498115]
13. Gamito P, Oliveira J, Coelho C, et al. Cognitive training on stroke patients via virtual reality-based serious games. Disabil Rehabil 2017 Feb;39(4):385-388. [doi: 10.3109/09638288.2014.934925] [Medline: 25739412]
14. Hsu SY, Fang TY, Yeh SC, Su MC, Wang PC, Wang VY. Three-dimensional, virtual reality vestibular rehabilitation for chronic imbalance problem caused by Ménière's disease: a pilot study. Disabil Rehabil 2017 Aug;39(16):1601-1606. [doi: 10.1080/09638288.2016.1203027] [Medline: 27418422]
15. D'Cunha NM, Nguyen D, Naumovski N, et al. A mini-review of virtual reality-based interventions to promote well-being for people living with dementia and mild cognitive impairment. Gerontology 2019;65(4):430-440. [doi: 10.1159/000500040] [Medline: 31108489]
16. Skurla MD, Rahman AT, Salcone S, et al. Virtual reality and mental health in older adults: a systematic review. Int Psychogeriatr 2022 Feb;34(2):143-155. [doi: 10.1017/S104161022100017X] [Medline: 33757619]
17. Huygelier H, Schraepen B, van Ee R, Vanden Abeele V, Gillebert CR. Acceptance of immersive head-mounted virtual reality in older adults. Sci Rep 2019 Mar 14;9(1):4519. [doi: 10.1038/s41598-019-41200-6] [Medline: 30872760]
18. Chau PH, Kwok YYJ, Chan MKM, et al. Feasibility, acceptability, and efficacy of virtual reality training for older adults and people with disabilities: single-arm pre-post study. J Med Internet Res 2021 May 4;23(5):e27640. [doi: 10.2196/27640] [Medline: 33944795]
19. Liao YY, Tseng HY, Lin YJ, Wang CJ, Hsu WC. Using virtual reality-based training to improve cognitive function, instrumental activities of daily living and neural efficiency in older adults with mild cognitive impairment. Eur J Phys Rehabil Med 2020 Feb;56(1):47-57. [doi: 10.23736/S1973-9087.19.05899-4] [Medline: 31615196]
20. Lima Rebêlo F, de Souza Silva LF, Doná F, Sales Barreto A, de Souza Siqueira Quintans J. Immersive virtual reality is effective in the rehabilitation of older adults with balance disorders: a randomized clinical trial. Exp Gerontol 2021 Jul 1;149:111308. [doi: 10.1016/j.exger.2021.111308] [Medline: 33744393]
21. Yamagami M, Imsdahl S, Lindgren K, et al. Effects of virtual reality environments on overground walking in people with Parkinson disease and freezing of gait. Disabil Rehabil Assist Technol 2023 Apr;18(3):266-273. [doi: 10.1080/17483107.2020.1842920] [Medline: 33155870]
22. Fan CC, Choy CS, Huang CM, et al. The effects of a combination of 3D virtual reality and hands-on horticultural activities on mastery, achievement motives, self-esteem, isolation and depression: a quasi-experimental study. BMC Geriatr 2022 Sep 12;22(1):744. [doi: 10.1186/s12877-022-03431-7] [Medline: 36096746]
23. Fiocco AJ, Millett G, D'Amico D, et al. Virtual tourism for older adults living in residential care: a mixed-methods study. PLoS One 2021;16(5):e0250761. [doi: 10.1371/journal.pone.0250761] [Medline: 34014948]
24. Dascal J, Reid M, IsHak WW, et al. Virtual reality and medical inpatients: a systematic review of randomized, controlled trials. Innov Clin Neurosci 2017;14(1-2):14-21. [Medline: 28386517]
25. Emmelkamp PMG, Meyerbröker K. Virtual reality therapy in mental health. Annu Rev Clin Psychol 2021 May 7;17:495-519. [doi: 10.1146/annurev-clinpsy-081219-115923] [Medline: 33606946]
26. Wilson PN, Foreman N, Stanton D. Virtual reality, disability and rehabilitation. Disabil Rehabil 1997 Jun;19(6):213-220. [doi: 10.3109/09638289709166530] [Medline: 9195138]
27. Li Y, Wilke C, Shiyanov I, Muschalla B. Impact of virtual reality-based group activities on activity level and well-being among older adults in nursing homes: longitudinal exploratory study. JMIR Serious Games 2024 Mar 29;12:e50796. [doi: 10.2196/50796] [Medline: 38551635]
28. Doré B, Gaudreault A, Everard G, et al. Acceptability, feasibility, and effectiveness of immersive virtual technologies to promote exercise in older adults: a systematic review and meta-analysis. Sensors (Basel) 2023 Feb 24;23(5):2506. [doi: 10.3390/s23052506] [Medline: 36904709]

29. Hosseini M, Thomas R, Pilutti L, Fallavollita P, Jutai JW. Acceptance of physical activity virtual reality games by residents of long-term care facilities: a scoping review. Disabil Rehabil Assist Technol 2023 Dec 26:1-9. [doi: 10.1080/17483107.2023.2298839] [Medline: 38146958]
30. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Management Science 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
31. Chen K, Chan AHS. A review of technology acceptance by older adults. Gerontechnology 2011;10(1):1-12 [FREE Full text] [doi: 10.4017/gt.2011.10.01.006.00]
32. Pang WYJ, Cheng L. Acceptance of gamified virtual reality environments by older adults. Educ Gerontol 2023 Sep 2;49(9):830-841. [doi: 10.1080/03601277.2023.2166262]
33. Xu W, Liang HN, Yu K, Wen S, Baghaei N, Tu H. Acceptance of virtual reality exergames among Chinese older adults. Int J Hum-Comput Interact 2023 Mar 16;39(5):1134-1148. [doi: 10.1080/10447318.2022.2098559]
34. Weck M, Helander N, Meristö T. Active DigiAge - technology acceptance by ageing people. IJTMCP 2020;3(3):223. [doi: 10.1504/IJTMCP.2020.10026516]
35. Broadbent E, Stafford R, MacDonald B. Acceptance of healthcare robots for the older population: review and future directions. Int J Soc Robotics 2009 Nov;1(4):319-330. [doi: 10.1007/s12369-009-0030-6]
36. Apolinário-Hagen J, Hennemann S, Kück C, et al. Exploring user-related drivers of the early acceptance of certified digital stress prevention programs in Germany. Health Serv Insights 2020;13:1178632920911061. [doi: 10.1177/1178632920911061] [Medline: 32206013]
37. Heylighen F. A cognitive-systemic reconstruction of Maslow’s theory of self-actualization. Syst Res 1992 Jan;37(1):39-58 [FREE Full text] [doi: 10.1002/bs.3830370105]
38. Phang CW, Sutanto J, Kankanhalli A, Li Y, Tan BCY, Teo HH. Senior citizens’ acceptance of information systems: a study in the context of e-government services. IEEE Trans Eng Manage 2006;53(4):555-569. [doi: 10.1109/TEM.2006.883710]
39. Atchley RC. A continuity theory of normal aging. Gerontol 1989 Apr;29(2):183-190. [doi: 10.1093/geront/29.2.183] [Medline: 2519525]
40. Lin CIC, Tang WH, Kuo FY. Mommy wants to learn the computer. Adult Educ Q (Am Assoc Adult Contin Educ) 2012 Feb;62(1):73-90. [doi: 10.1177/0741713610392760]
41. Scott JE, Walczak S. Cognitive engagement with a multimedia ERP training tool: assessing computer self-efficacy and technology acceptance. Inf & Manag 2009 May;46(4):221-232. [doi: 10.1016/j.im.2008.10.003]
42. Chen K, Chan AHS. Predictors of gerontechnology acceptance by older Hong Kong Chinese. Technov 2014 Feb;34(2):126-135. [doi: 10.1016/j.technovation.2013.09.010]
43. Appel L, Appel E, Bogler O, et al. Older adults with cognitive and/or physical impairments can benefit from immersive virtual reality experiences: a feasibility study. Front Med (Lausanne) 2019;6:329. [doi: 10.3389/fmed.2019.00329] [Medline: 32010701]
44. Brown JA. An exploration of virtual reality use and application among older adult populations. Gerontol Geriatr Med 2019;5:2333721419885287. [doi: 10.1177/2333721419885287] [Medline: 31723574]
45. Tacken M, Marcellini F, Mollenkopf H, Ruoppila I, Széman Z. Use and acceptance of new technology by older people. Findings of the international MOBILATE survey: ‘Enhancing mobility in later life'. Gerontechnology 2005;3(3). [doi: 10.4017/gt.2005.03.03.002.00]
46. Linden M, Baron S, Muschalla B. Mini-ICF-Rating Für Psychische Störungen (Mini-ICF-APP): Ein Kurzinstrument Zur Beurteilung Von Fähigkeits- Bzw. Kapazitätsstörungen Bei Psychischen Beschwerden [Book in German]: Hans Huber; 2009.
47. Burri M, Werk LP, Berchtold A, Pugliese M, Muschalla B. Mini-ICF-APP inter-rater reliability and development of capacity disorders over the course of a vocational training program—a longitudinal study. J Psychosoc Rehabil Ment Health 2021 Aug;8(2):159-169. [doi: 10.1007/s40737-021-00216-w]
48. Linden M, Baron S, Muschalla B, Ostholt-Corsten M. Fähigkeitsbeeinträchtigungen Bei Psychischen Erkrankungen: Diagnostik, Therapie Und Sozialmedizinische Beurteilung in Anlehnung an Das Mini-ICF-APP [Book in German]: Huber; 2015.
49. IBM SPSS software (version 29). IBM. URL: https://www.ibm.com/spss [accessed 2024-09-10]
50. Chen K, Chan AHS. Gerontechnology acceptance by elderly Hong Kong Chinese: a senior technology acceptance model (STAM). Ergonomics 2014;57(5):635-652. [doi: 10.1080/00140139.2014.895855] [Medline: 24655221]
51. Wandke H, Sengpiel M, Sönksen M. Myths about older people’s use of information and communication technology. Gerontology 2012;58(6):564-570. [doi: 10.1159/000339104] [Medline: 22739502]
52. Harris MT, Rogers WA. Developing a healthcare technology acceptance model (H-TAM) for older adults with hypertension. Ageing Soc 2023 Apr;43(4):814-834. [doi: 10.1017/s0144686x21001069] [Medline: 37007645]
53. Kohlbacher F, Herstatt C, Schweisfurth T. Product development for the silver market. In: Kohlbacher F, editor. The Silver Market Phenomenon: Springer Berlin Heidelberg; 2011:3-13. [doi: 10.1007/978-3-642-14338-0_1]
54. Mikhailova V, Conde M, Döring N. “Like a virtual family reunion”: older adults defining requirements for an augmented reality communication system. Inf 2024;15(1):52. [doi: 10.3390/info15010052]

55. Lee C, Coughlin JF. PERSPECTIVE: older adults' adoption of technology: an integrated approach to identifying determinants and barriers. J of Prod Innov Manag 2015 Sep;32(5):747-759 [FREE Full text] [doi: 10.1111/jpim.12176]
56. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]

Abbreviations

TAM: Technology Acceptance Model
VR: virtual reality

Edited by E Borycki; submitted 12.01.24; peer-reviewed by N Adi, N Döring, P Valkonen; revised version received 06.06.24; accepted 16.06.24; published 04.10.24.

Please cite as:
Li Y, Shiyanov I, Muschalla B
Older Adults' Acceptance of a Virtual Reality Group Intervention in Nursing Homes: Pre-Post Study Under Naturalistic Conditions
JMIR Hum Factors 2024;11:e56278
URL: https://humanfactors.jmir.org/2024/1/e56278
doi:10.2196/56278

Original Paper

Usability, Ergonomics, and Educational Value of a Novel Telestration Tool for Surgical Coaching: Usability Study

Parmiss Kiani[1], BSc; Roberta Dolling-Boreham[1,2], BACS; Mohamed Saif Hameed[1], DNB, MBBS; Caterina Masino[1], MA; Andras Fecso[3], MD, PhD; Allan Okrainec[3], MHPE, MD; Amin Madani[1,3], MD, PhD

[1]Surgical Artificial Intelligence Research Academy, University Health Network, Toronto, ON, Canada

[2]Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[3]Department of Surgery, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Amin Madani, MD, PhD
Surgical Artificial Intelligence Research Academy
University Health Network
399 Bathurst Street
Toronto, ON, M5T 2S8
Canada
Phone: 1 4163403843
Email: amin.madani@uhn.ca

Abstract

Background: Telementoring studies found technical challenges in achieving accurate and stable annotations during live surgery using commercially available telestration software intraoperatively. To address the gap, a wireless handheld telestration device was developed to facilitate dynamic user interaction with live video streams.

Objective: This study aims to find the perceived usability, ergonomics, and educational value of a first-generation handheld wireless telestration platform.

Methods: A prototype was developed with four core hand-held functions: (1) free-hand annotation, (2) cursor navigation, (3) overlay and manipulation (rotation) of ghost (avatar) instrumentation, and (4) hand-held video feed navigation on a remote monitor. This device uses a proprietary augmented reality platform. Surgeons and trainees were invited to test the core functions of the platform by performing standardized tasks. Usability and ergonomics were evaluated with a validated system usability scale and a 5-point Likert scale survey, which also evaluated the perceived educational value of the device.

Results: In total, 10 people (9 surgeons and 1 senior resident; 5 male and 5 female) participated. Participants strongly agreed or agreed (SA/A) that it was easy to perform annotations (SA/A 9, 90% and neutral 0, 0%), video feed navigation (SA/A 8, 80% and neutral 1, 10%), and manipulation of ghost (avatar) instruments on the monitor (SA/A 6, 60% and neutral 3, 30%). Regarding ergonomics, 40% (4) of participants agreed or strongly agreed (neutral 4, 40%) that the device was physically comfortable to use and hold. These results are consistent with open-ended comments on the device's size and weight. The average system usability scale was 70 (SD 12.5; median 75, IQR 63-84) indicating an above average usability score. Participants responded favorably to the device's perceived educational value, particularly for postoperative coaching (agree 6, 60%, strongly agree 4, 40%).

Conclusions: This study presents the preliminary usability results of a novel first-generation telestration tool customized for use in surgical coaching. Favorable usability and perceived educational value were reported. Future iterations of the device should focus on incorporating user feedback and additional studies should be conducted to evaluate its effectiveness for improving surgical education. Ultimately, such tools can be incorporated into pedagogical models of surgical coaching to optimize feedback and training.

(JMIR Hum Factors 2024;11:e57243) doi:10.2196/57243

KEYWORDS

augmented reality; AR; surgical training; telestration; tele-stration; surgical training technology; minimally invasive surgery; surgery; surgeon; surgeons; surgical; surgical coaching; surgical teaching; surgical training; telemonitoring; telemonitor; tele-monitoring; tele-monitor; usability; usable; usableness; usefulness; utility; digital health; digital technology; digital intervention; digital interventions

Introduction

Telementoring studies found technical challenges in achieving accurate and stable annotations using commercially available telestration software intraoperatively [1-3].

The first challenge is the dynamic nature of the video feed; there are frequent laparoscopic camera movements, field of view changes, and deformation of anatomic structures due to the mobilization and retraction of anatomical structures, and maneuvering of the camera [4,5].

Mitigation strategies during coaching activities included freezing the video and converting it to still images [6]. This is not practical for real-time intraoperative coaching by surgeons and greatly increases the time spent on the activity to stop and restart the session during annotation.

Previous usability studies that used telestration [6-8] used a trackpad, mouse, or touchscreen during annotation mode and found that the trackpad or mouse performed best in the delineation of structures, while the touch screen was superior in conveying directional information [7]. Further, 1 study compared the usability of similar telestration devices with conventional interfacing devices such as a computer and mouse, and a tablet and stylus [4]. With the advancement of technology in the gaming world, new virtual reality (VR) or augmented reality systems have emerged as viable solutions for the development of systems for telestration.

To address the educational gap in teaching minimally invasive procedures, a wireless handheld telestration device was developed to facilitate dynamic user interaction with live video streams. This study examines the usability of a first-generation handheld wireless telestration platform.

Continuing professional education activities such as surgical coaching provide opportunities for the continued acquisition of new techniques and professional expertise.

Telestration is a technique for teaching whereby instructors annotate images or videos to enhance the learning experience for surgical trainees [9]. This technique has shown promise for improving surgical skills more effectively than traditional verbal coaching across a broad range of metrics including faster task completion, reduced coaching time, better surgical performance, and greater trainee confidence [7,10-12]. Previous telestration studies have highlighted the importance of mentor-mentee communication in surgical training [7,11,13].

In the setting of laparoscopic surgery, the learning curve for trainees tends to be greater as the surgical field cannot be directly visualized or palpated as in open surgeries and directly pointing at surgical display screens for teaching and coaching activity raises concerns of sterility. Instead, feedback and guidance are typically verbally described without the ability to make direct references, which may lead to greater trainee confusion, miscommunication, and ultimately reduced efficacy of the teaching process. This is especially relevant with the shift to minimally invasive surgeries, which provide numerous benefits to patients over open surgeries [14]. Given the benefits of telestration seen for trainees and the adoption of minimally invasive surgery for increasingly more complex procedures, it is critical to develop better tools to improve the acquisition of these skills.

To augment the coaching experience, a wireless handheld telestration tool was developed to better address dynamic on-demand teaching requirements both intraoperatively during a procedure and postoperatively on a recorded surgical video. This study presents the usability results of a first-generation handheld wireless telestration platform.

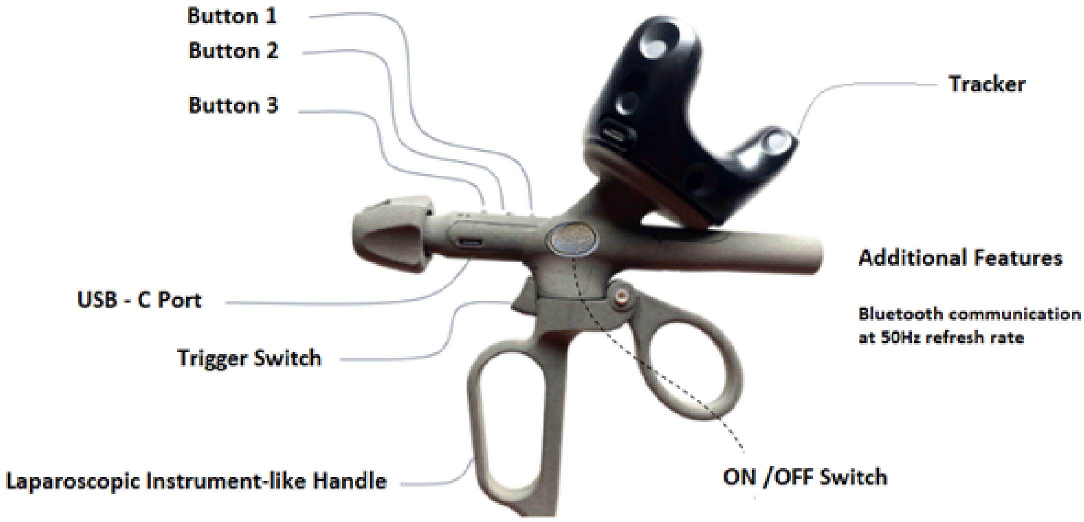
Methods

Hardware Design

A wireless handheld telestration device prototype, hereon referred to as the pen, was designed and manufactured to enable the user (ie, surgeon coach) to interact with the surgical display field during coaching activities in both intraoperative and postoperative settings. This prototype uses a proprietary augmented reality platform.

The pen (Figure 1) consists of two main parts: (1) a motion tracker mounted to (2) a 3D printed controller. The motion tracker and the controller have separate charging systems and on/off buttons. The controller has 3 buttons with various functions, a trigger, and a Maryland grasper handle to create a more realistic user experience. The pen weighs 152.76 g.

Figure 1. Telestration device.

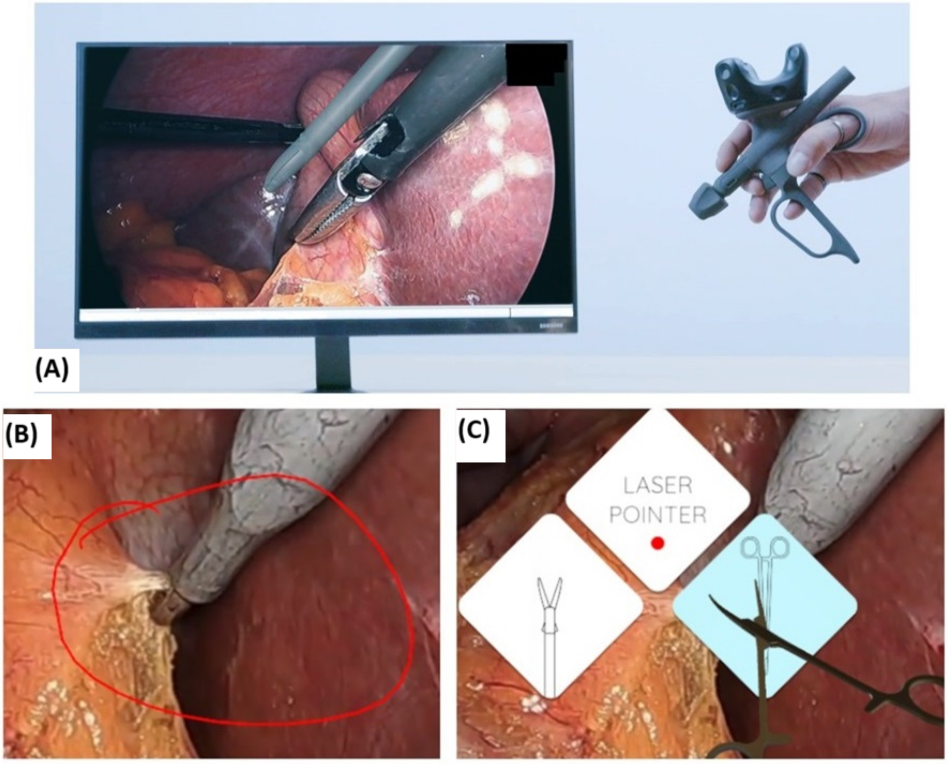


Software Design

The four core functions of this video-based coaching platform are (1) free hand annotation, (2) cursor navigation, (3) overlay and manipulation of ghost (avatar) instruments, and (4) hand-held video feed navigation.

These functions may be completed on a live or recorded video feed. To achieve these core functions, the telestration software menu (Figure 2C) allows the user to choose from 3 interactive tools: a laser pointer for cursor navigation and annotation (Figure 2B), and 2 digital avatar instruments for dynamic coaching and positioning instructions: 1 for laparoscopy (designed to resemble a Maryland Grasper) and 1 for open surgery (designed to resemble a needle driver; Figure 2).

Figure 2. (A) Telestration device used for video-based coaching, (B) annotation feature demonstration, and (C) telestration software menu options.



System Platform Integration

The telestration platform integrates a legacy guidance and tracking system using: Vive (HTC Corp) and SteamVR (Valve Corp).

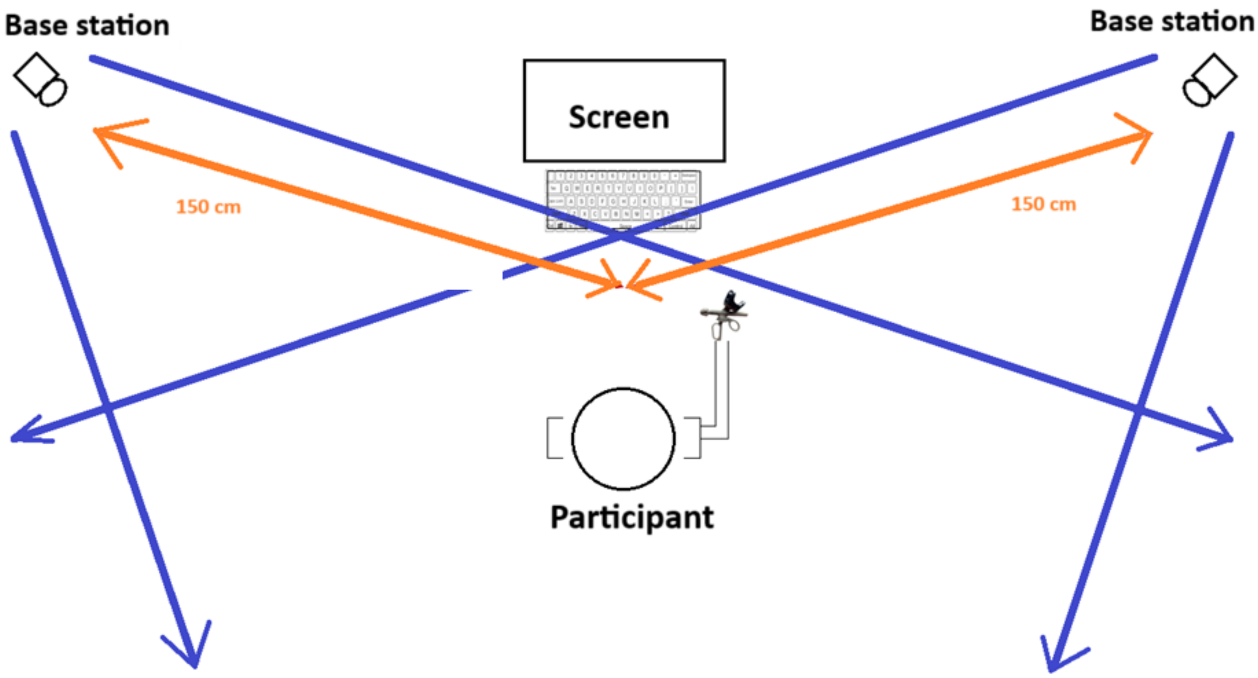
The tracker unit is mounted onto the telestration pen's hardware (Figure 1). The tracker has multiple diodes that detect the infrared signals emitted by the SteamVR lighthouse units. Lighthouse units are externally powered devices mounted at locations of high visibility around the physical space to detect and track the telestration device's movements by the user. The number of lighthouse units is determined by the size of the room where the telestration pen is implemented. A total of 2 lighthouse units were placed 1.5 m apart to achieve optimal performance as per the manufacturer's instructions (Figure 3). The SteamVR system appears to designate 1 lighthouse as the primary source of data and another as a secondary source [15]. Figure 3 illustrates the room setup.

The orientation of the lighthouses per the object being tracked is crucial to the accuracy of the system. The tracking accuracy is greatest with the pen positioned orthogonally to it. The maximum distance between the lighthouse and the tracker is the maximum distance to work effectively which is 7 m [15].

Therefore, for this study, a total of 2 lighthouse units were placed 1.5 m apart to achieve optimal performance as per the manufacturer's instructions (Figure 3).

The lighthouse units ("base stations") emit infrared light which is detected by diodes present on the tracker and this information is converted to positional data by SteamVR software.

Figure 3. Usability study room set-up.



Usability Study Design

The International Organization of Standardization defines usability as "the extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" [16]. Testing a device with end users is essential for a comprehensive evaluation of its usability. Therefore, the usability study was conducted in 3 phases (presurvey, usability testing, and post survey) and included an informal debrief at the end of the session.

Participants

All physicians who play a role in surgical education were allowed to participate. Participants were surgeons and surgical trainees recruited from a multi-cite academic teaching hospital in Toronto, Ontario, Canada.

Data Collection

Phase 1 (Presurvey)

A short questionnaire was administered to gather baseline information regarding demographics and prior experience with VR and telestration technologies.

Phase 2 (Usability Testing)

The functional capability of the telestration device was designed by surgeon educators based on the needs they noted in real-world coaching situations. Scenarios were then prepared to evaluate the 4 core functions of the device in those situations. A study facilitator guided participants through the scenarios in a representative simulated environment. Examples of tasks include starting the device, menu selection, tool navigation and overlay, annotating and drawing, moving video playback or pause, and so on. A complete list of tasks is outlined in Textbox 1.

Textbox 1. Tasks for telestration function assessment.

Standard tasks

- App launch and device calibration: participants were asked to start a preselected video displaying certain clips from a laparoscopic cholecystectomy.
- Play or pause video feed: participants were asked to play and pause the video at certain points throughout the session.

Video feed navigation

- Fast forward and rewind: participants were asked to fast forward and rewind the video to predetermined points of the video throughout the session.

Overlay and manipulation of ghost avatar instrument

- Menu-instrument selection: participants were asked to open the menu and select either the laser pointer or the grasper multiple times throughout the session.
- Rotate grasper tip: participants were asked to rotate the grasper tip.
- Open and close grasper tip: participants were asked to open and close the grasper tip.

Freehand annotation

- Annotation: participants were asked to open and close the grasper tip.
- Erase annotation: participants were asked to annotate where they would perform a dissection at a specific moment of a procedure and to circle the cystic duct.

Phase 3 (Postsurvey)

Participants completed the system usability scale (SUS) after testing was completed [17]. This scoring system compiles the responses from a series of 10 questions that cover topics including device complexity, ease of use, the learning curve required, and user confidence [18]. In addition to this, a questionnaire was developed by the researchers to gather the attitudes and opinions of the participants on usability, satisfaction, ergonomics, ease of task completion, confidence, and the perceived educational value of the device, using a 5-point Likert scale.

Procedure

Before the start, this study's room was set up as shown in Figure 3, the VIVE tracker was calibrated to the room using SteamVR. A member of the research team was this study's facilitator. Participants began the session by completing the presurvey (phase 1). Next, the participants were asked to hold a device such as a laparoscopic grasper. The facilitator trained the participants on the functionality of the device and provided the participants with a quick tip sheet to reference during the usability test.

Participants were asked to select a video of a laparoscopic cholecystectomy that was provided for this study's purpose. Participants were then asked to calibrate the controller by following the instructions on the screen. Participants were first asked to use the telestration device while standing to emulate its use in an operating room, followed by a seated position for an office or boardroom setting. The facilitator guided the participant to complete tasks in Textbox 1, testing the 4 core functions of the device in a standing position and ending the video on completion. Upon study completion, participants completed phase 3 of this study.

Outcomes

Outcomes measured in this study were perceived usability, ergonomics, overall satisfaction, and the perceived educational value of the telestration device.

Data Analysis

Descriptive statistics were used, and qualitative variables were reported as frequencies and percentages.

The results of the SUS were analyzed according to the scoring procedure documented by Brooke [17]. A product with a SUS score greater than 70 is considered to have above-average usability [18].

Ethical Considerations

Informed consent was taken before the start of this study. This study was approved by the University Health Network's (UHN) Research Ethics Board (22-5556; education research protocol dated September 14, 2022).

Results

Phase 1 Demographics

A total of 9 surgeons and 1 senior resident (n=10) participated (5 males and 5 females). The average age of participants was 36.4 (SD 6) years with a mean of 7 (SD 6.41) years of practice. All participants reported being right-handed. The majority of participants (7 out of 10, 70%) reported no previous telestration system experience. Only 1 participant reported having received training with a surgical VR system before this study. Most participants (5 out of 10, 50%) reported not using VR and gaming consoles in the last 12 months (3 out of 10, 30% monthly and 2 out of 10, 20% rarely).

About SUS

This study's average SUS was 70 (SD 12.5) with a median of 75 (IQR 63-84).

Overall Satisfaction

Participants were asked to rate their overall satisfaction using a 5-point Likert scale where 10% (1 out of 10) of participants were completely satisfied, 50% (5 out of 10) were very satisfied, 20% (2 out of 10) were moderately satisfied, and 20% (2 out of 10) slightly satisfied.

Ergonomics

When asked about the intuitiveness of the device, most participants strongly agreed or agreed (SA/A) that the device was intuitive (5 out of 10, 50%), 40% (4 out of 10) felt neutral, and 10% (1 out of 10) disagreed or strongly disagreed (D/SD). Ergonomics was further assessed by asking participants to respond specifically regarding the pen's physical comfort (4 out of 10, 40% SA/A; 4 out of 10, 40% neutral; and 2 out of 10, 20% D/SD), the weight of the pen (6 out of 10, 60% SA/A and 4 out of 10, 40% D/SD), and the ability to use the physical features (buttons and trigger; 4 out of 10, 40% SA/A; 3 out of 10, 30% neutral; and 3 out of 10, 30% D/SD).

Participants were also asked if they preferred completing this study's tasks while in a seated or standing position. In total, 4 participants preferred to be seated, 3 preferred to be standing, and 3 had no preference at all. When asked about their confidence in completing the tasks correctly, 90% (9 out of 10) and 80% (8 out of 10) SA/A that they correctly completed the tasks seated and standing, respectively.

Open Ended Survey Responses

Participants were asked to discuss features of the pen and telestration system that they felt were design strengths as well as areas of improvement (Textbox 2). Of the 8 participants who responded, 3 participants reported the ability to annotate on the screen as a good feature of the device. Regarding areas for improvement, 6 participants indicated to have the location of the buttons moved.

Textbox 2. Open feedback responses from the participants regarding the virtual reality system.

What 3 things are good about the virtual reality system?

- Hands-on teaching, not invasive, and may be widely used.
- It allows you to draw on the image on the screen, it allows you to demonstrate the orientation of instruments, and it allows you to fast forward and rewind.
- Felt accurate in terms of location of the pointer, intuitive to use, and realistic feeling.
- Great response time, useful for annotation, and innovative.
- Teaching.
- Easy to use.
- Ability to annotate, erase, and select instruments.
- Relatively easy to use after a short guidance, it is cool, innovative, and less stressful.

What 3 things can be improved in the virtual reality system?

- Button placement, precision, and ergonomics.
- The actual instrument itself (the weight of it and location of buttons), the directionality of fast forward or rewind (to make it more intuitive), and the function of the hand holds.
- Location of buttons.
- Calibration of pen not in line with pen, instrument heavy, and hard to hold for small hands to reach the top buttons.
- Fulcrum on table.
- Calibration, weight, and function of finger loops.
- Ergonomics (handle is not needed), buttons are not easy to use, and tracking sensor drift.
- The [pen] is a little uncomfortable, the pen is heavy, and when sitting it is hard to see the animation for the Maryland.

Postsurvey Satisfaction Responses

Participants were asked to rate their level of agreement with statements on the standard tasks performed in this study (Table 1).

The majority of participants (6 out of 10, 60%) found the setup tasks (launching the video and device calibration) easy to complete, while 30% (3 out of 10) of participants found it difficult and 10% (1 out of 10) felt neutral. On the other hand, 90% (9 out of 10) of participants agreed that fast-forwarding and rewinding, as well as annotating, were easy, and 100% (10 out of 10) of participants agreed that erasing annotations was easy.

Table 1. Ease of use in completing tasks testing the device's various functions.

Function	Agree or strongly agree, n (%)	Neutral, n (%)	Disagree or strongly disagree, n (%)
Standard tasks			
Difficulty completing initial setup: launch and calibration	3 (30)	1 (10)	6 (60)
Play or pause the video feed	8 (80)	2 (20)	0 (0)
Video feed navigation			
Fast forward and rewind	9 (90)	1 (10)	0 (0)
Overlay and manipulation of ghost instrument			
Menu-instrument selection	7 (70)	3 (30)	0 (0)
Rotate grasper tip	4 (40)	5 (50)	1 (10)
Open and close the grasper tip	7 (70)	2 (20)	1 (10)
Freehand annotation			
Annotation	9 (90)	0 (0)	1 (10)
Erase annotation	10 (100)	0 (0)	0 (0)

Confidence Levels (Pre- Versus Poststudy Comparison)

Overview

Participants were asked to rate their confidence in the system features at baseline and post study using a scale of 1 to 5, where 1 is not confident and 5 is very confident.

Set Up and Training

Participants were also asked to rate their level of confidence in their technical ability to independently set up a VR or gaming system in the post study. The majority of participants (6 out of 10, 60%) rated their confidence in system setup at a 4 out of 5, while 20% (2 out of 10) rated it a 5 out of 5, and another 20% (2 out of 10) rated it a 3 out of 5. In addition, the majority of participants felt that the system training completed by the facilitator was adequate (8 out of 10, 80%), with 20% (2 out of 10) of participants feeling the training period was too short. Lastly, the majority rated the quality of the training provided for study purposes as excellent 1 out of 10, 10%; very good 6 out of 10, 60%; and good 3 out of 10, 30%.

Navigation

The majority of participants reported a confidence rating of 4 at baseline for navigating accurately and realistically (confidence of 5: 1 out of 10, 10%; confidence of 4: 8 out of 10, 80%; confidence of 3: 1 out of 10, 10%), which then increased to a rating of 5 post study (confidence of 5: 5 out of 10, 50%; confidence of 4: 4 out of 10, 40%; and confidence of 3: 1 out of 10, 10%). No participant reported a confidence of 2 or less in either the pre or post study.

Instrument Overlay

The greatest positive change (40% increase) between pre and post study in confidence rating was present for the overlay of the digital tool. Confidence ratings for this category prestudy were (confidence of 5: 1 out of 10, 10%; confidence of 4: 4 out of 10, 40%; confidence of 3: 4 out of 10, 40%; confidence of 2: 0 out of 10, 0%; and confidence of 1: 1 out of 10, 10%), while post study were (confidence of 5: 1 out of 10, 10%; confidence of 4: 8 out of 10, 80%; confidence of 3: 1 out of 10, 10%; confidence of 2: 0 out of 10, 0%; and confidence of 1: 0 out of 10, 0%).

Annotations

The confidence levels for performing annotations were highly rated for both pre- (confidence of 5: 2 out of 10, 20%; confidence of 4: 7 out of 10, 70%; and confidence of 3: 1 out of 10, 10%) and post study (confidence of 5: 4 out of 10, 40% and confidence of 4: 6 out of 10, 60%).

Video Feedback

Regarding video feed playback or pause reviewing functions, confidence levels remained unchanged pre to post study (confidence of 5: 3 out of 10, 30%; confidence of 4: 6 out of 10, 60%; and confidence of 3: 1 out of 10, 10%).

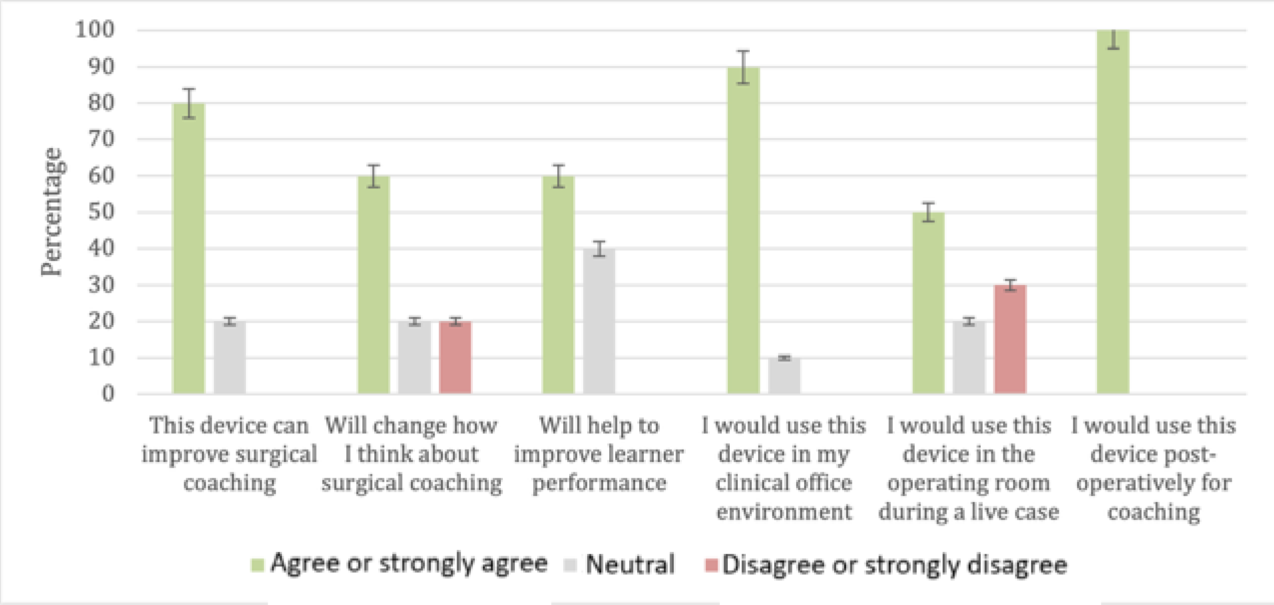
Select and Change Tools

Lastly, while no participant felt very confident selecting or changing tools in the prestudy (confidence of 5: 0 out of 10, 0%; confidence of 4: 7 out of 10, 70%; and confidence of 3: 3 out of 10, 30%), 30% (3 out of 10) of participants rated that they felt very confident to select or change tools in the post study phase (confidence of 5: 3 out of 10, 30%; confidence of 4: 5 out of 10, 50%; and confidence of 3: 2 out of 10, 20%). When participants were also asked for their level of agreement about the ease of switching between digital tools, 70% (7 out of 10) of participants SA/A that it was easy to do and 30% (3 out of 10) of participants felt neutral about it.

Educational Value

Participants were asked to rate the perceived educational value of the device using a 5-point Likert agreement scale. All participants agreed that they would use this device for postoperative coaching, while only 50% (5 out of 10) of participants agreed that they would use the device in an intraoperative setting. Additional questions were asked regarding the use for educational purposes described in Figure 4.

Figure 4. Participant agreement responses on perceived educational value.



Technical Difficulties

Lastly, participants were asked to report on technical difficulties experienced while participating in this study. While 40% (4 out of 10) of participants reported experiencing a delay or lag with the device positioning, only 10% (1 out of 10) of participants reported experiencing difficulties loading this study's video and 10% (1 out of 10) also noticed instructional text not displaying correctly. When participants were asked about the device's ability to track their hand movements accurately, 60% (6 out of 10) SA/A that it did and 40% (4 out of 10) of participants felt neutral about it.

Discussion

Principal Findings

A novel telestration device for surgical coaching was designed to enhance the surgeon coach and learner experience in the context of laparoscopic surgeries. This device enables a dynamic interaction with surgical display monitors with a free-hand annotation function and live overlay of 3D digital laparoscopic tool avatars.

Previously described devices and systems for telementoring had usability challenges that this device aims to mitigate. This study aimed to evaluate the overall satisfaction and usability of the first-generation prototype telestration device for use in surgical coaching activities. In terms of demographics, our study had an equal distribution of women and men, which strengthens the validity of our results.

The use of evaluation tools such as SUS during the development and testing process of user interface apps is commonly recommended in the literature [7,10]. The SUS reported an average score of 70 (SD 12.5) with a median score of 75 (IQR 63-84) indicating an above-average usability rating in comparison to thousands of other devices and systems [18]. Additionally, 60% (6 out of 10) of participants were either completely or very satisfied with the device overall. This is an encouraging result of our first iteration prototype.

On the other hand, those who rated their overall satisfaction less than this, commented on the device's ergonomics, including the button placement, finger loops, and weight, as well as the device's precision and lagging experience. These comments on ergonomics are likely why participants had lower confidence in the post study to select or change tools compared to the other tasks. Additionally, while most users agreed that the device's weight was comfortable, only 40% (4 out of 10) felt the device's shape was physically comfortable to hold. Furthermore, 40% (4 out of 10) felt that the buttons were hard to reach; a theme that also emerged in the open feedback responses where 1 user commented "instrument … hard to hold for small hands to reach the top buttons." Comments were also made about the usefulness of the handle piece. While confidence levels of completing study tasks accurately were similar in either position, participants did prefer using the device while sitting. The ergonomic feedback is in line with the SUS scores reported. With ergonomics being a priority for surgeons, future iterations of this device will aim at improving these scores.

With regards to task completion, including video stream controls (play or pause), annotation, and tool avatar manipulation, was generally very positive; the majority of them were considered "easy" to complete. All tasks were completed successfully with the provided training even though participants had never interacted with the device before this study. In addition, many of the participants did not experience any technical difficulties performing all the tasks—no major bugs were identified in this study. Only 40% of participants reported experiencing a lag during the usability testing period, and the time required to complete a task was comparable to that of this study's facilitator (trainer). Therefore, in future iterations of the device and software, addressing the lag experienced by users is of importance.

With regards to task completion, including video stream controls (play or pause), annotation, and tool avatar manipulation, was generally very positive; the majority of them were considered "easy" to complete. All tasks were completed successfully with the provided training even though participants had never interacted with the device before this study. In addition, the majority of the participants did not experience any technical difficulties performing all of the tasks listed in Textbox 1, as only 40% (4 out of 10) participants reported experiencing a lag during the usability testing period, and the time required to complete a task was comparable to that of this study's facilitator (trainer). Therefore, in future iterations of the device and software, addressing the lag experienced by users is of importance.

Furthermore, participants found the majority of tasks easy to complete, most notably video manipulation (pause or play at 8 out of 10, 80% and fast-forward or rewind at 9 out of 10, 90%) and annotation (draw at 9 out of 10, 90% and erase at 10 out of 10, 100%). Overall, about task completion, this device demonstrates an acceptable level of usability; all tasks were completed successfully, and the majority of them were considered easy to complete.

In analyzing the open feedback responses, participants used language including "easy to use," "great response time," "accurate," "intuitive," and "realistic" to describe the device when allowed to provide open and anonymous feedback. This positive feedback is highly encouraging and highlights important themes relevant to usability including an acceptable level of complexity and realistic experience.

Lastly, participants evaluated the perceived educational value of the device with an overwhelming majority (8 out of 10, 80%) of users agreeing that this device can improve surgical coaching, especially in the postoperative setting (10 out of 10, 100%). However, only 50% (5 out of 10) of participants agreed that they would use the device live in an operating room, the main setting in which we intended this device to be used. Thus, participant hesitancy is an important goal of future usability assessments of the telestration device and perhaps would be better understood with testing in an operating room.

Study Limitations

While a sample size of 4 to 5 participants in usability studies is usually adequate in detecting 80% of system issues, a limitation of our study was its smaller sample size [19]. Another limitation of our study was that the participants were asked about their thoughts on the usability of the device in different settings, specifically intraoperatively. As this study was conducted within an office setting and not within the operating room, it limits the applicability of the answers to being a preliminary thought rather than an actual observation in the asked-about scenario.

Future Considerations

Overall, the results of this first iteration study indicate our novel telestration device has a strong degree of usability, general user satisfaction, and potential concerning surgical skills coaching. Therefore, we have determined the prototype to have met a satisfactory threshold to merit further development and refinement.

Further improvements will focus on ergonomics with effort dedicated to making the device lighter and relocating the buttons to a more accessible location. Based on participant input, future iterations should also investigate either adding functionality to the device's handle or potentially removing it altogether. This would address the majority of constructive criticisms from users. From the software perspective, improvements of priority include refinements in the software to allow for a more simplified app launch and calibration as this was the main task of difficulty for our participants.

Future studies should evaluate the educational value of the device in the operating room setting and further evaluate its effectiveness in enhancing the surgeon coach and trainee experience.

Conclusion

In conclusion, preliminary usability testing of a prototype telestration device for surgical coaching has demonstrated above-average usability and positive feedback regarding the perceived educational value and task completion. Future improvements should focus on ergonomics and design, namely weight and button location, as well as app launch and calibration. The next steps following usability testing can include the assessment of the educational value of the telestration device.

Acknowledgments

This work was funded by the UHN Foundation. The UHN Foundation and Haply Robotics Inc were not involved in this study's design; in the collection, analysis, and interpretation of data; in the writing of this paper; and in the decision to submit this paper for publication.

Conflicts of Interest

The development of the telestration platform was provided through a collaborative research agreement between Haply Robotics Inc and the coinvestigators AO and AM. The development work provided by Haply Robotics Inc consisted of the hardware development of the telestration tool prototype and the control software and user interface development.

References

1. Budrionis A, Augestad K, Bellika J. Telestration in Mobile Telementoring. 2013. URL: https://munin.uit.no/bitstream/handle/10037/8287/paper_VI.pdf?isAllowed=y&sequence=8 [accessed 2024-07-25]

2. Budrionis A, Augestad KM, Patel HRH, Bellika JG. Towards requirements for telementoring software. 2012 Presented at: Scandinavian Conference on Health Informatics; 2012 October 2-3; Linköping, Sweden.
3. Gambadauro P, Magos A. NEST (network enhanced surgical training): a PC-based system for telementoring in gynaecological surgery. Eur J Obstet Gynecol Reprod Biol 2008;139(2):222-225. [doi: 10.1016/j.ejogrb.2007.12.004] [Medline: 18243486]
4. Ereso AQ, Garcia P, Tseng E, Dua MM, Victorino GP, Guy LTS. Usability of robotic platforms for remote surgical teleproctoring. Telemed J E Health 2009;15(5):445-453. [doi: 10.1089/tmj.2008.0164] [Medline: 19548825]
5. Shin DH, Dalag L, Azhar RA, Santomauro M, Satkunasivam R, Metcalfe C, et al. A novel interface for the telementoring of robotic surgery. BJU Int 2015;116(2):302-308 [FREE Full text] [doi: 10.1111/bju.12985] [Medline: 25381917]
6. Budrionis A, Hartvigsen G, Lindsetmo R, Bellika JG. What device should be used for telementoring? Randomized controlled trial. Int J Med Inform 2015;84(9):715-723. [doi: 10.1016/j.ijmedinf.2015.05.004] [Medline: 26048739]
7. Budrionis A, Hasvold P, Hartvigsen G, Bellika JG. Assessing the impact of telestration on surgical telementoring: a randomized controlled trial. J Telemed Telecare 2016;22(1):12-17. [doi: 10.1177/1357633X15585071] [Medline: 26026177]
8. Patel E, Mascarenhas A, Ahmed S, Stirt D, Brady I, Perera R, et al. Evaluating the ability of students to learn and utilize a novel telepresence platform, Proximie. J Robot Surg 2022;16(4):973-979 [FREE Full text] [doi: 10.1007/s11701-021-01330-4] [Medline: 34741714]
9. Telestration in minimally invasive surgery. JAMA Surgery. 2021. URL: https://edhub.ama-assn.org/jn-learning/video-player/18644961 [accessed 2023-08-30]
10. Wild C, Lang F, Gerhäuser AS, Schmidt MW, Kowalewski KF, Petersen J, et al. Telestration with augmented reality for visual presentation of intraoperative target structures in minimally invasive surgery: a randomized controlled study. Surg Endosc 2022;36(10):7453-7461 [FREE Full text] [doi: 10.1007/s00464-022-09158-1] [Medline: 35266048]
11. Nickel F, Cizmic A, Chand M. Telestration and augmented reality in minimally invasive surgery: an invaluable tool in the age of COVID-19 for remote proctoring and telementoring. JAMA Surg 2022;157(2):169-170. [doi: 10.1001/jamasurg.2021.3604] [Medline: 34705030]
12. Jarc AM, Stanley AA, Clifford T, Gill IS, Hung AJ. Proctors exploit three-dimensional ghost tools during clinical-like training scenarios: a preliminary study. World J Urol 2017;35(6):957-965 [FREE Full text] [doi: 10.1007/s00345-016-1944-x] [Medline: 27671899]
13. Jarc AM, Shah SH, Adebar T, Hwang E, Aron M, Gill IS, et al. Beyond 2D telestration: an evaluation of novel proctoring tools for robot-assisted minimally invasive surgery. J Robot Surg 2016;10(2):103-109 [FREE Full text] [doi: 10.1007/s11701-016-0564-1] [Medline: 26914650]
14. Buia A, Stockhausen F, Hanisch E. Laparoscopic surgery: a qualified systematic review. World J Methodol 2015;5(4):238-254 [FREE Full text] [doi: 10.5662/wjm.v5.i4.238] [Medline: 26713285]
15. Bauer P, Lienhart W, Jost S. Accuracy investigation of the pose determination of a VR system. Sensors (Basel) 2021;21(5):1622 [FREE Full text] [doi: 10.3390/s21051622] [Medline: 33669148]
16. Ergonomics of human-system interaction? Part 11: usability: definitions and concepts. In: International Organization for Standardization. Geneva, Switzerland: ISO; 2018.
17. Brooke J. SUS: a quick and dirty usability scale. In: Usability Eval Ind. US: CRC Press; 1995:189.
18. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123.
19. Lewis JR. Sample sizes for usability studies: additional considerations. Hum Factors 1994;36(2):368-378. [doi: 10.1177/001872089403600215] [Medline: 8070799]

Abbreviations

D/SD: disagreed or strongly disagreed
SA/A: strongly agreed or agreed
SUS: system usability scale
UHN: University Health Network
VR: virtual reality

Edited by C Jacob; submitted 20.02.24; peer-reviewed by A Hassan, SK Yadav, LV Selby, N Chaniaud; comments to author 01.05.24; revised version received 26.06.24; accepted 17.07.24; published 10.09.24.

Please cite as:
Kiani P, Dolling-Boreham R, Hameed MS, Masino C, Fecso A, Okrainec A, Madani A
Usability, Ergonomics, and Educational Value of a Novel Telestration Tool for Surgical Coaching: Usability Study
JMIR Hum Factors 2024;11:e57243
URL: https://humanfactors.jmir.org/2024/1/e57243
doi:10.2196/57243
PMID:

Original Paper

A Handheld Tool for the Rapid Morphological Identification of Mosquito Species (VectorCam) for Community-Based Malaria Vector Surveillance: Summative Usability Study

Saisamhitha Dasari[1*], MSE; Bhavya Gopinath[1*], MSE; Carter James Gaulke[1], MSE; Sunny Mahendra Patel[1], MSE; Khalil K Merali[1], MSE, MD; Aravind Sunil Kumar[1], MSE; Soumyadipta Acharya[1], MSE, MD, PhD

Johns Hopkins University Center for Bioengineering Innovation and Design, Baltimore, MD, United States

[*]these authors contributed equally

Corresponding Author:
Saisamhitha Dasari, MSE
Johns Hopkins University Center for Bioengineering Innovation and Design
3400 Charles Street
Baltimore, MD, 21218
United States
Phone: 1 17707785272
Email: dasari.sam21@gmail.com

Abstract

Background: Malaria impacts nearly 250 million individuals annually. Specifically, Uganda has one of the highest burdens, with 13 million cases and nearly 20,000 deaths. Controlling the spread of malaria relies on vector surveillance, a system where collected mosquitos are analyzed for vector species' density in rural areas to plan interventions accordingly. However, this relies on trained entomologists known as vector control officers (VCOs) who identify species via microscopy. The global shortage of entomologists and this time-intensive process cause significant reporting delays. VectorCam is a low-cost artificial intelligence–based tool that identifies a mosquito's species, sex, and abdomen status with a picture and sends these results electronically from surveillance sites to decision makers, thereby deskilling the process to village health teams (VHTs).

Objective: This study evaluates the usability of the VectorCam system among VHTs by assessing its efficiency, effectiveness, and satisfaction.

Methods: The VectorCam system has imaging hardware and a phone app designed to identify mosquito species. Two users are needed: (1) an imager to capture images of mosquitos using the app and (2) a loader to load and unload mosquitos from the hardware. Critical success tasks for both roles were identified, which VCOs used to train and certify VHTs. In the first testing phase (phase 1), a VCO and a VHT were paired to assume the role of an imager or a loader. Afterward, they swapped. In phase 2, two VHTs were paired, mimicking real use. The time taken to image each mosquito, critical errors, and System Usability Scale (SUS) scores were recorded for each participant.

Results: Overall, 14 male and 6 female VHT members aged 20 to 70 years were recruited, of which 12 (60%) participants had smartphone use experience. The average throughput values for phases 1 and 2 for the imager were 70 (SD 30.3) seconds and 56.1 (SD 22.9) seconds per mosquito, respectively, indicating a decrease in the length of time for imaging a tray of mosquitos. The loader's average throughput values for phases 1 and 2 were 50.0 and 55.7 seconds per mosquito, respectively, indicating a slight increase in time. In terms of effectiveness, the imager had 8% (6/80) critical errors and the loader had 13% (10/80) critical errors in phase 1. In phase 2, the imager (for VHT pairs) had 14% (11/80) critical errors and the loader (for VHT pairs) had 12% (19/160) critical errors. The average SUS score of the system was 70.25, indicating positive usability. A Kruskal-Wallis analysis demonstrated no significant difference in SUS (*H* value) scores between genders or users with and without smartphone use experience.

Conclusions: VectorCam is a usable system for deskilling the in-field identification of mosquito specimens in rural Uganda. Upcoming design updates will address the concerns of users and observers.

(JMIR Hum Factors 2024;11:e56605) doi:10.2196/56605

KEYWORDS

malaria; vector surveillance; usability study testing; usability study; usability studies; usability; usable; usableness; usefulness; utility; convolution neural networks; artificial intelligence; AI; human factors design; human factors; mHealth; mobile health; app; apps; digital health; digital technology; digital intervention; digital interventions; smartphone; smartphones; mobile phone

Introduction

Background

Malaria infects an estimated 249 million individuals annually, causing >600,000 deaths worldwide [1]. The global incidence of malaria has increased since the COVID-19 pandemic, with 13 million more cases and 63,000 more deaths [2]. According to the World Health Organization, regions in Sub-Saharan Africa are particularly susceptible to malarial infection, accounting for >95% of global malaria cases and deaths; children aged <5 years account for nearly 80% of malaria deaths in the region [3]. Specifically, Uganda has one of the highest global burdens of malaria cases, costing the country US $500 million annually [4]. In 2021, the World Health Organization estimated that there were 13 million malaria cases and nearly 20,000 malaria deaths in Uganda [4].

Currently, efforts to eliminate malaria rely on monitoring vector species composition, abundance, distribution, and behavior across different transmission geographies. Vector control strategies, such as distributing long-lasting insecticidal nets and performing indoor residual spraying of insecticide in high transmission regions, have proven to be highly effective in preventing infection [5]. However, mosquito vectors vary in their biting patterns (some may bite outdoors more frequently, and others may bite indoors more frequently) and subsequently necessitate different vector control strategies. Therefore, the Ministries of Health and their National Malaria Control Programs in almost all African countries have emphasized the need for vector surveillance, a system that allows for the collection of vector species prevalence and density data to target vector control interventions and resource allocation decisions based on the mosquito's species-specific strategies [6]. Routine vector surveillance strategies begin with mosquito collection, where mosquito specimens are collected at sentinel sites using various vector collection methods, including Centers for Disease Control and Prevention light traps, pyrethrum spray catches, and human landing catches [7]. These specimens, collected daily, are then transported to a central laboratory. There, they are morphologically identified through microscopic examination by individuals highly trained in entomology and vector surveillance, known as vector control officers (VCOs), for their species, sex, and abdominal status. A subset of the specimens is then sent for molecular analysis through polymerase chain reactions (PCRs) and DNA sequencing when resources are available. This subset is treated as a gold standard for further confirmation and quality assurance of the species identified through vector surveillance [8]. Unfortunately, the global shortage of entomologists hinders large-scale surveillance efforts, especially in areas with a high burden of malaria, where vector surveillance is needed the most [9,10]. As a result, the sites where specimens are collected and analyzed are sparsely distributed across a target region and treated as a representation of the entire country. This causes inaccuracies in the interventions deployed, which is further worsened given the time lag between the capturing of specimens and the reporting time of usable data for decision-making [11].

To address this unmet need arising from the global shortage of entomologists, our research group at the Johns Hopkins Center for Bioengineering Innovation and Design has developed VectorCam. VectorCam is a low-cost artificial intelligence (AI)–based tool that morphologically identifies a mosquito's species, sex, and abdomen status based on a simple photograph and concurrently uploads the summary data to a central electronic dashboard, thereby deskilling the identification and reporting process. Such a tool helps share the efforts of trained entomologists with rural community health workers, termed village health teams (VHTs) in Uganda, thereby enabling the implementation of widespread vector surveillance programs and driving better-informed, data-driven malaria intervention decisions in a cost-effective manner.

The VectorCam System: An Overview

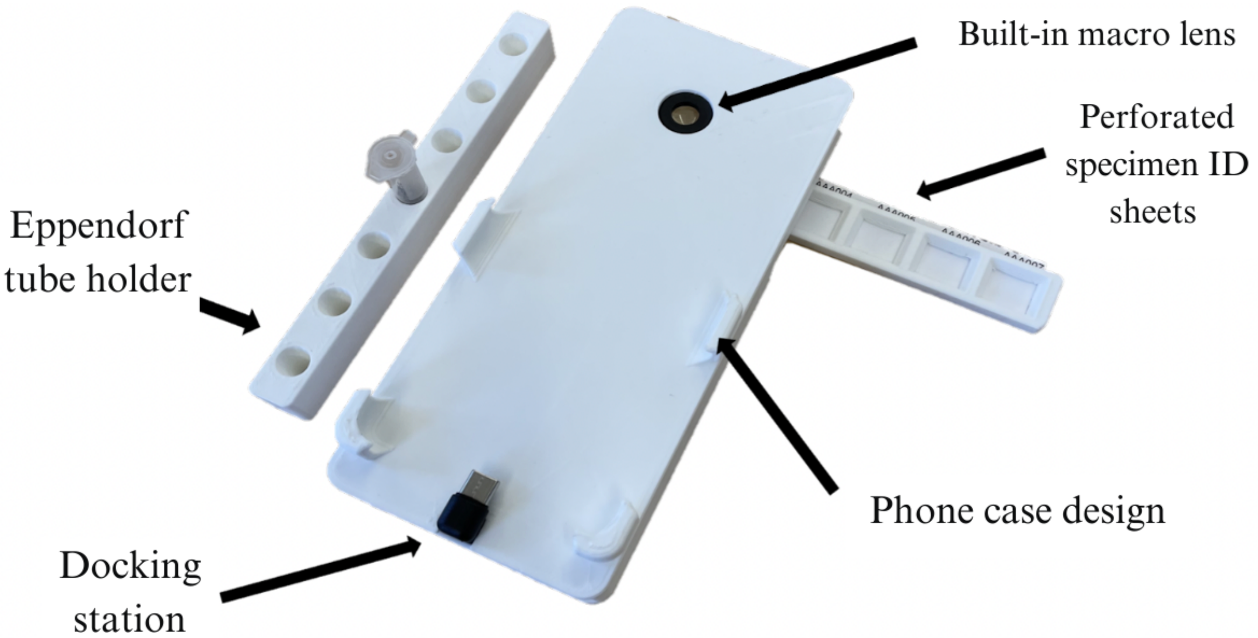
VectorCam enables the accurate morphological identification of mosquito species using 3 core components: specialized imaging hardware, a computer vision algorithm, and a mobile app. The hardware consists of a simple light box with a built-in 15x macro lens, paired with a basic smartphone and several mosquito trays, enabling the rapid identification of multiple mosquitos with high throughput. The VectorCam Android app is powered by an optimized version of a previously published computer vision algorithm from our laboratory, which can identify >39 species across multiple genera [12]. This adapted version of the algorithm, used in VectorCam, identifies major genera (Anopheles, Culex, Aedes, Mansonia, and nonmosquitos) and 4 different groups of species within the Anopheles genera of interest in Uganda (*Anopheles funestus* sl, *Anopheles gambiae* sl, *Anopheles stephensi*, and other Anopheles species), as Anopheles is the only malaria-carrying mosquito genera. This computationally efficient version can run locally on low-end Android phones without access to the internet, thereby allowing users to capture images of collected mosquito specimens and instantaneously obtain the classification. The app also consists of a workflow for data inputs from the user, including logistical information about the collection site, date, and other notes for data tracing. Finally, the user can capture an image of the mosquito specimen directly within the app. The algorithm then processes this image and provides the predicted species, sex, and abdominal status of the mosquito.

The imaging hardware and software user interface and user experience have been cocreated through a multiyear human-centered design approach involving inputs from >50 VHT members and VCOs from Uganda, Zambia, and Ghana. Multiple formative usability studies have been performed with nearly 40 VHTs and 10 VCOs in Uganda to finalize the hardware and app design, as well as essential performance criteria. This study serves as a summative usability assessment of a 2-person task-shifted vector identification strategy using the VectorCam app. This system was designed to fit easily into the current vector surveillance practices. For quality assurance purposes, careful consideration was taken to ensure that each mosquito was uniquely tagged for verification and molecular and PCR identification, as needed. The results of the study will inform future design iterations of the VectorCam system for field deployment in vector surveillance programs.

VectorCam Hardware and Software Details

VectorCam's hardware consists of a light box with a 15x macro lens attached to it. The top side of this light box comprises of a phone case, where the phone slides on, plugging into the docked charger at the bottom end of this light box (Figure 1). The charger is connected to circuitry on the interior of the box, which turns on the LEDs as soon as the phone is docked, allowing uniform lighting between each specimen. The phone's camera is then aligned with the 15x macro lens, allowing uniform magnification. The box itself also has storage compartments for 2 mosquito trays on which mosquitos are placed for ease of imaging, as well as an associated Eppendorf tube holder to hold mosquitos with their unique tag after they have been identified on the app. Within the VectorCam workflow, each mosquito must not touch one another, as this can impact DNA contamination. Each mosquito tray has a slit where a disposable piece of paper can be inserted to limit any DNA contamination. Furthermore, these slits of paper have preprinted and perforated unique specimen ID tags attached for each well. Therefore, when storing these mosquitos for PCR identification, the user can simply tear off the specimen ID and place it as well as the corresponding mosquito specimen in an associated Eppendorf tube. The Eppendorf tube holder allows for ease of packing and labeling these mosquitos for subsequent molecular identification and maintaining its traceability to the identification generated by the VectorCam system.

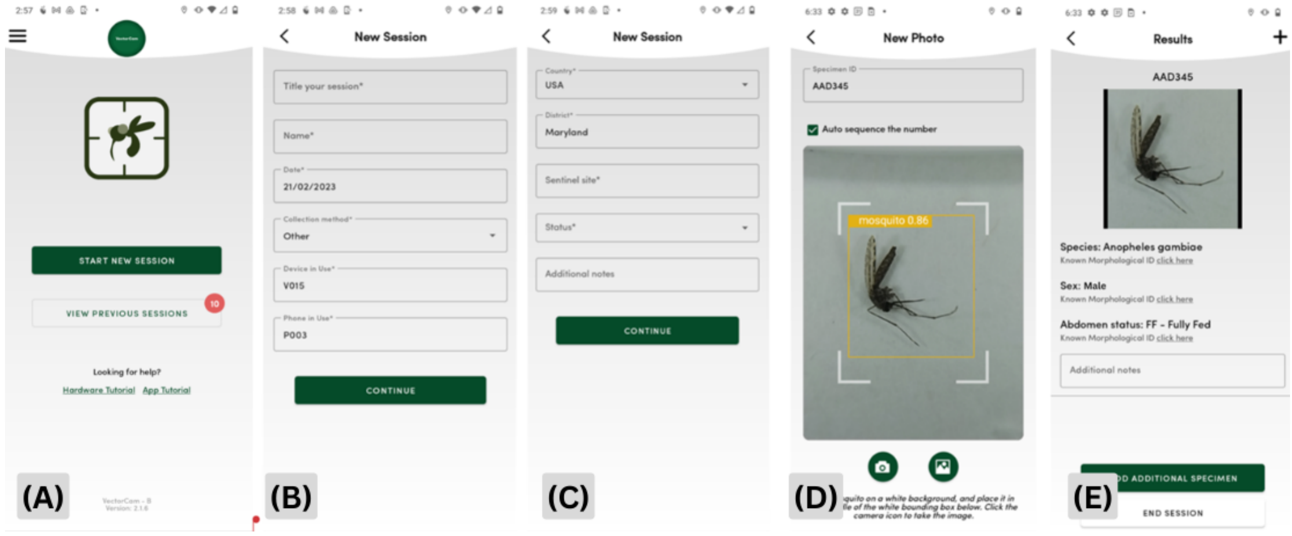
Figure 1. The hardware components of the VectorCam system include the light box with a built-in 15x macro lens, a phone case design, and a docking station. Hardware also includes an Eppendorf tube holder and mosquito trays and perforated specimen ID sheets to adequately pack and store these mosquitos for molecular identification after imaging.



PCR identification will be facilitated after imaging the collected samples with VectorCam to grow the continued database of true labels of specimens. These true labels will be used to continually train the VectorCam computer vision algorithm, as it is still in development. In the ideal setting, a finalized VectorCam product will replace the need for PCRs. However, the specimen ID tear-off step will still be included in the workflow to allow for traceability and postimplementation verifications. Figure 1 shows how all these hardware components interact.

The app itself is hosted on a Motorola G Play (2021; Motorola, Inc) smartphone. The app consists of three primary processes: (1) entering background information, (2) imaging mosquitos, and (3) submitting a session. Entering background information effectively stores all the information that entomologists usually write when generating reports for the Ministry of Health. This includes how these mosquitos were collected, the date they were collected, and the location where they were collected. Figure 2 shows details of each screen on the app. The goal of this summative usability study was to evaluate the usability of the VectorCam system among village health workers by assessing its efficiency, effectiveness, and satisfaction. We hypothesized that the VectorCam system will overall be usable with no significant differences between sex or users with and without smartphone use experience.

Figure 2. User interface screen sequences for entering background information and imaging mosquitos in order of appearance as the imager starts interacting with the app (from left to right). (A) The home screen of the VectorCam app allowing the imager to choose between starting the imaging session or viewing previous sessions already submitted in the app. (B) Under "entering background information," the first screen allows users to enter the title of the session, the name of the user, the date of the imaging session, and information on the device and phone in use. (C) The second screen allows the user to input additional information, including country, district, and sentinel site of the current catch and imaging process as well as the status of the mosquito specimen (freshly collected specimens or desiccated). (D) The imaging screen for the imager to enter the specimen ID of the mosquito and image the mosquito using zooming and capture. (E) The results screen of the previous image of the specimen captured showing the species, sex, and abdomen status. This screen allows the imager to choose to add an additional specimen or end the imaging session.



Methods

VectorCam Device Workflow

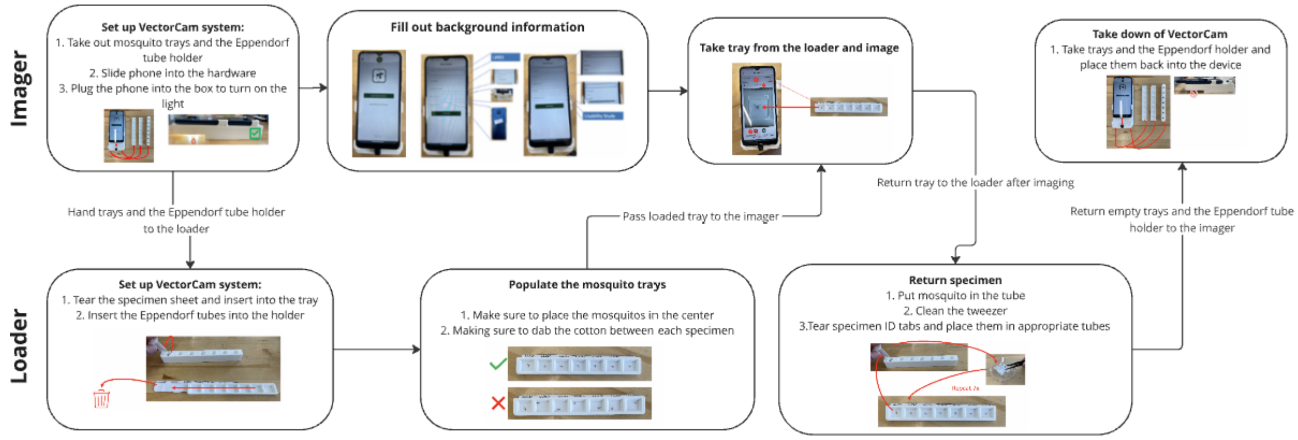
The VectorCam system involves placing collected mosquitos into the VectorCam hardware; capturing a magnified image of the mosquito using a smartphone-based mobile app for morphological identification; and storing the mosquito in an Eppendorf tube with a unique label for subsequent molecular verification, if needed.

The VectorCam system was designed to be operated by a group of 2 VHT members working in collaboration by sharing the critical pieces of the workflow, whereby one person (the loader) loads specimens and another (the imager) uses the app to capture images of the mosquitos. Having a 2-person team speeds up the morphological identification process of a large batch of mosquitos. The materials used in this process are presented in Table 1, and a workflow summary of the VectorCam system is presented in Figure 3.

Table 1. Materials needed to operate VectorCam in a 2-user system, their purpose in the workflow, and the role of the person using the equipment.

Equipment used	Purpose of the equipment	Person using the equipment
VectorCam light box	Contains a holder for the smartphone with the VectorCam app, 15x macro lens for magnification, and a phone-powered LED light source on the underside	Imager
Motorola G Play (2021; Motorola, Inc) smartphone	Contains the VectorCam app to enter relevant background information; capture images of multiple mosquitos per session; display corresponding species, sex, and abdominal status as identified by the algorithm; and submit sessions to the cloud-based server	Imager
Plastic mosquito trays	Contains designated wells to allow for the placement of several specimens in 1 tray and for rapidly imaging all the specimens loaded onto the tray	Both imager and loader
Eppendorf tube holder	Contains designated holes for placing 7 Eppendorf tubes to facilitate easy and rapid packing of each imaged specimen into respective Eppendorf tubes along with the corresponding specimen ID label	Loader
Sheet of specimen IDs	Perforated sheets with unique specimen IDs to allow for separation and placement into the mosquito trays. The labels can be removed and placed into specific Eppendorf tubes with their associated specimens	Loader
Petri dish with cotton	Contains cotton soaked in isopropyl alcohol for periodic cleansing of the tweezers handling the specimen	Loader
Tweezers	Used for handling the mosquito specimen during tasks in the workflow	Loader
Eppendorf tubes	Used to store the specimen after the imaging process is complete	Loader

Figure 3. The entire workflow of the use of VectorCam, including how the imager and loader roles work together in this process. The imager primarily interacts with the specialized imaging hardware, and the loader primarily interacts with the mosquito trays and Eppendorf tube holder.



Description of the Loader's Role

The loader has two primary responsibilities: (1) placing mosquitos onto the mosquito trays to streamline the imaging process and (2) storing already-imaged mosquitos in individually labeled Eppendorf tubes such that the mosquitos are uniquely tagged and can be sent to a laboratory for molecular identification via PCRs or sequencing. To properly place the mosquitos onto the VectorCam hardware, the loader takes the 2 mosquito trays provided in the hardware and slides the 8 inch (length) by 1 inch (width) specimen ID sheets inside the tray. Each of these sheets has a perforated tab that appears above the tray, displaying a specimen ID for each of the wells on the tray. This allows each mosquito on the tray to have a unique label. The loader then uses tweezers to add a singular mosquito to each well, decontaminating the tweezers with isopropyl alcohol–soaked cotton before handling each mosquito to prevent DNA cross-contamination. With this sterilization, the mosquitos can be sent for molecular identification via PCRs for quality control. Once the tray is filled, it is handed off to the imager.

Once the tray is imaged, the imager passes the mosquito tray back to the loader. At this point, the loader works to unload and store each mosquito and its associated specimen ID. The loader first adds a mosquito to an Eppendorf tube that can be held on the Eppendorf tube holder provided. This is done with tweezers that are sterilized with isopropyl alcohol between mosquito transfers to prevent DNA cross-contamination. Then, the specimen ID sheet is taken out, and each ID is torn off and placed in its corresponding Eppendorf tube along with the mosquito. This gives each mosquito a unique specimen ID when it is sent for molecular identification. Therefore, quality control can be established where the gold-standard assessment of each mosquito's species, sex, and abdominal status can also be received for further analysis.

Description of the Imager's Role

Once the imager receives the mosquito tray filled with specimens, they attach the phone to the light box, ensuring the LEDs are turned on, and then they go through the process of imaging the mosquitos on the app. This includes entering the background information, taking a photo of each specimen, zooming in to ensure that there is enough resolution in each photograph, and finally saving the session for later uploading to our cloud-based server.

After the imager finishes imaging all mosquitos on VectorCam, they pass the tray back to the loader, who properly stores these mosquitos. Furthermore, once the entire process of imaging is finished for the day, the imager collects all mosquito trays and Eppendorf tube holders, undocks the phone, and places the materials inside the light box for storage.

Study Design: Materials and Methods

Participants and Location Selection

This summative usability study was conducted with 20 participants located in 2 malaria-endemic districts in Uganda: Mayuge and Adjumani. A sample size of 20 participants was determined using a study conducted by Faulkner [13] that collected empirical data from a sample of 60 individuals with varying levels of computer experience and varying levels of knowledge on the software used in the evaluation. In the study, a sample of 20 people was able to find a minimum of 95% and an average of 98% of the problems. Therefore, a sample population of 20 users was sufficient to encompass most of the usability problems with a device and design and a manageable number for our group to recruit, given the time constraints.

Participant recruitment criteria included those who (1) were aged between 18 and 65 years, (2) had existing status as a VHT member within Mayuge or Adjumani, and (3) had no previous experience using the VectorCam app or its previous iterations. Participants were not excluded based on prior experience with smartphone app use or with vector surveillance within their communities. Sites were selected with the support of the Ugandan Ministry of Health to ensure the heterogeneity of experience with vector control programs. At the time of the study, Mayuge had no established vector control program, while Adjumani had an established vector control program, according to the Ugandan Ministry of Health. Adjumani currently has a high burden of malaria prevalence in Uganda, as it is situated in an area of high, stable transmission of malaria [12].

Participants were recruited randomly from pools of existing VHT members in each district using the inclusion criteria described. The feedback received from the recruited VHTs serves as a valuable resource for enhancing VectorCam's functionality and its potential impact on vector control efforts in rural Uganda.

Separating the Study Into Phases

During phase 1 of this study, 20 VHT members were paired with 1 VCO (out of a pool of 4 VCOs), and the VHT member was evaluated on the imager and loader roles in turns, with the VCO assuming the other role. This was done to limit any impact of another newly trained VHT member on the evaluation of outcome success criteria described in the *Outcomes* section. A total of 14 mosquitos were imaged or loaded during this evaluation. After 14 mosquitos were imaged, the VCO and VHT member switched roles so that the VHT member was evaluated for performance as both the loader and imager.

Phase 2 had 2 sets of trained VCOs imaging and 10 sets of paired VHT members imaging and loading a set of 28 mosquitos. After these 28 mosquitos were imaged, the 2 VCOs or VHT members swapped roles, allowing each participant to be evaluated in both roles. VCOs were evaluated using the same outcome measures as the VHT members, and this was done to benchmark their performance and compare it to that of the VHT pairs as a secondary analysis.

Ethical Considerations

This study was conducted with the approval of the Johns Hopkins University Institutional Review Board (00259683) and The Aid Support Organization Research Ethics Committee approval (2022-91) from the Makerere University School of Public Health. All participants were deidentified within the data. Each participant was compensated 10,000 Ugandan Shillings for their successful completion of their portion of the study.

Outcomes

The primary outcome of interest was the usability of the VectorCam system (device and app). This was assessed using efficiency, effectiveness, and satisfaction measurements. Each outcome of interest was observed and recorded by an observer in real time.

Effectiveness

The effectiveness of the use of the VectorCam system was defined by the error rates caused by the failure of critical tasks of the use of VectorCam by the VHT members. For instance, phase 1 had 4 critical tasks for the imager and loader (imager: setup, imaging—entering background information, imaging—taking pictures, and takedown; loader: setup 1, setup 2, takedown 1, and takedown 2) and phase 2 had 8 critical tasks for the loader (setup 1 to 4 and takedown 1 to 4) and 4 critical tasks for the imager (setup, imaging—entering background information, imaging—taking pictures, and takedown). To calculate the error rates, the total number of failures on these critical tasks was recorded (each task had a binary pass or fail) and summed up. This was then divided by the total number of critical tasks. To evaluate this on a participant pool basis, the number of critical tasks was summed across participants and then divided by the total number of critical tasks across participants as well.

To identify these critical tasks and subtasks, each of the authors did a cognitive walk-through of the VectorCam system workflow, asking 2 key questions for each step: "Will the user know what to do at this step?" and "Will the user know that s/he did the right thing?" [14]. After a collective discussion, a comprehensive list of all the critical tasks as well as the subtasks in each of these critical tasks was finalized. The critical tasks included (1) hardware setup, (2) populating regional background information into the app, (3) imaging of the specimens, and (4) takedown of the hardware. A success criterion was then determined for each task, and a binary pass or fail metric was established. Any failure of the success criteria contributed to the total failure of the larger critical task and the error rate. Table 2 includes lists of the subtasks performed by both the imager and loader and their success criteria during each of the 4 critical tasks.

Table 2. From the cognitive walk-through analysis performed by the authors of this paper, different subtasks for the process of using VectorCam for both the imager and loader roles were identified. Corresponding success criteria for both the imager's and loader's subtasks were identified during the setup of VectorCam. The subtasks and success criteria were identified for each main task of VectorCam: setup of the system, entering background information, taking pictures, and the takedown of the device.

Main task, role, and subtasks	Success criteria
Setup of VectorCam	
Imager	
Take out mosquito trays 1 and 2 and the Eppendorf tube holder and pass them to person 2.	All mosquito trays and tubes are taken out.
Slide the phone into the 3D-printed hardware.	The phone is slid into the 3D-printed box.
Place the phone into the box.	The phone is plugged into the box, and LEDs light up.
Loader	
Write specimen IDs on a sheet of paper.	All and accurate specimen IDs are written on the paper.
Place the paper in the mosquito tray and Eppendorf tubes in the tube holder and puncture the Eppendorf tube.	Paper is placed in the mosquito tray, and tubes are in the Eppendorf tube holder. Eppendorf tubes are properly punctured.
Populate the mosquito tray with 7 mosquitos (1 in each box).	All 7 mosquitos are populated into the tray.
Imaging process: entering background information	
Imager	
Create a new session on the VectorCam app and enter all the background information.	A new session is created on the VectorCam app with the correct background information and title of the session.
Enter specimen ID and click "auto-populate," if sequential.	Specimen ID is entered as stated on the mosquito tray sheet.
Loader	
Ensure mosquitos are in the center of mosquito tray 1.	All mosquitos are centered in each well in mosquito tray 1.
Pass mosquito tray 1 to the imager.	Mosquito tray 1 is passed to the imager.
Imaging process: taking pictures	
Imager	
Pinch zoom on the mosquito (pinch diagonally across the screen).	Once the mosquito is in the camera screen, it is pinch zoomed such that the entire mosquito is in frame.
Take an image by clicking the camera icon and select analyze.	The camera icon is selected, and the analyze option is selected.
Add an additional specimen and move the mosquito tray.	An additional specimen is added, and the previous specimen is not overwritten.
Once 7 mosquitos have been imaged, pass mosquito tray 1 to the loader and take mosquito tray 2 from the loader.	The mosquito tray is passed to the loader.
Add an additional specimen and place mosquito tray 2 under the camera so that the mosquito is seen.	An additional specimen from location 1 is added.
Loader	
Place the paper in mosquito tray 2.	The paper is placed in the mosquito tray, with specimen IDs visible.
Populate mosquito tray 2 with 7 mosquitos.	Mosquito tray 2 is populated with 7 mosquitos from the Eppendorf tube holder.
Ensure mosquitos are in the center of mosquito tray 2.	All mosquitos are centered in each well in mosquito tray 2.
Pass mosquito tray 2 to the imager.	Mosquito tray 2 is passed to the imager.
Move each mosquito from tray 1 to the respective Eppendorf tube. Place the specimen ID into the tube.	All mosquitos are moved from tray 2 to the correct Eppendorf tube with the correct specimen ID.
Replace the paper in mosquito tray 1 with a new sheet with the next specimen ID.	A new sheet of paper is placed in mosquito tray 1 with the correct specimen ID.
Takedown of the device	
Imager	
Click "submit session" and either upload the session or save the session to upload later.	"Submit session" is selected, and the session is uploaded.

Main task, role, and subtasks	Success criteria
Unplug the phone from the device and turn off the phone.	The phone is unplugged and turned off.
Place trays and the Eppendorf tube holder in the VectorCam hardware.	All trays and the tube holder are placed into the VectorCam hardware.
Loader	
Move each mosquito from tray 2 to the respective Eppendorf tube. Place the specimen ID into the tube.	All mosquitos are moved from tray 2 to the correct Eppendorf tube with the correct specimen ID.
Remove any remaining paper from all mosquito trays.	All pieces of paper are removed from mosquito trays.
Hand mosquito trays and the Eppendorf tube holder to the imager.	Trays and the Eppendorf tube holder are handed back to the imager.

Next, to determine potential user errors in VectorCam's workflow tasks, we performed a heuristic analysis to inspect the VectorCam's interface and identify initial usability problems using the 10 usability heuristics proposed by Nielsen and Molich [15,16], which are listed in Textbox 1. Each of the 4 authors judged the product interface separately and created a unique list of usability problems based on the usability heuristics, thereby generating a comprehensive list of user errors. After a collective discussion, a list of user errors was finalized for each subtask. These potential errors, in addition to additional errors found with VHT members during this usability study, were used to further inform future iterations of VectorCam, if needed.

Textbox 1. The 10 usability heuristics proposed by Nielsen and Molich [15,16] were used to analyze the VectorCam system's user interface to determine possible user errors and inefficiencies for future design considerations and changes.

1. Visibility of system status
2. Match between the system and the real world
3. User control and freedom
4. Consistency and standards
5. Error prevention
6. Recognitions rather than recall
7. Flexibility and efficiency of use
8. Aesthetic and minimalistic design
9. Help users recognize, diagnose, and recover from errors
10. Help and documentation

Efficiency: Phase 1

Efficiency in phase 1 was defined by how much time in minutes it took for each VHT member to both image and load 14 mosquitos while paired with a VCO doing the other role. Therefore, each VHT member performed one role with the VCO and then switched so that they were also evaluated in the other role. During phase 1, the VCO was included to limit any variation in the results due to the impact of being paired with another newly trained participant.

Efficiency: Phase 2

Comparatively, in phase 2, a "real-world" implementation of VectorCam was tested, where 2 newly trained VHTs were paired together to image a total of 28 mosquitos, switching positions afterward so that they can both be evaluated in both the loader and imager roles. Evaluators studied efficiency by timing how long the VHT member took to load or image the 28 mosquitos, depending on their role. Therefore, in phase 2, the impact of being paired with another newly trained participant was studied. The total length of time in minutes was measured, and the average length of time to image each mosquito was calculated.

Concurrently, 2 pairs of trained VCOs worked in the 2-person system as an imager and a loader to image 28 mosquitos in 4 trays each to benchmark the performance of the VHT pairs. Because these VCOs were trained on the system for multiple months, comparing the VCOs' average length of time to load or image each mosquito with that of the VHT pairs illustrated the impact of long-term training and exposure to VectorCam on efficiency.

Satisfaction

Satisfaction was defined using the System Usability Scale (SUS), a robust and validated criterion used to measure end-user usability of the product [17]. The items constituting the SUS score are listed in Textbox 2.

Textbox 2. The System Usability Scale was administered to each participant. Each item was answered on a scale of 1 (strongly disagree) to 5 (strongly agree). Items 2, 4, 6, 8, and 10 are those for which a lower numerical score is desired, and items 1, 3, 5, 7, and 9 are those for which a higher numerical score is desired. Each of these items was asked verbally by a translator.

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well-integrated.
6. I thought there was too much inconsistency in the system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

The generated score (out of 100) from the SUS was used to evaluate user satisfaction. Because the SUS score can be similarly structured to that of a 100-point academic scale, 90 to 100 points was an A, 80 to 89 points was a B, and so on, with the standard average score being 68 [15]. A score >68 was passed for a positive usability assessment, and a score >80 indicated excellent usability. To further understand the functionality of the devices and the rationale for the scores given, open-ended interviews with participants were also conducted.

Secondary outcomes of interest included the following: (1) the impact of age and previous smartphone use on satisfaction and (2) the thematic clustering of voiced barriers to usability from open-ended interviews following the completion of study procedures.

Study Execution Procedures: Materials and Methods

Overview

Study instructions and a user ID were verbally provided to each participant. After finishing the evaluation stage of this procedure, SUS and demographic questionnaires were administered verbally to local interpreters.

The study was conducted in five segments, namely (1) initial training and certification; (2) phase 1: initial testing with a VCO partner; (3) phase 2: testing in VHT pairs; (4) SUS and demographic questionnaires; and (5) planned statistical analysis.

Initial Training and Certification

All VHT members recruited were onboarded with a standardized training protocol (Multimedia Appendix 1) delivered by expert VCOs familiar with the VectorCam hardware, software, and vector surveillance workflow. VCOs trained VHT pairs to use the VectorCam hardware and software. During the training protocol, any questions asked by participants were addressed as part of the initial training session. After initial training was completed, the VCO trainers formally assessed the ability of VHT participants to use the VectorCam system in the roles of an imager and a loader. A standardized certification checklist was used to determine whether each member of the VHT pair could independently use the VectorCam system. A detailed list of certification criteria is presented in Multimedia Appendix 2. If a participant did not meet certification requirements, VCO trainers clarified and corrected errors before another attempt at certification was undertaken. The participants were then brought to the study session 24 hours after training certification to allow for sufficient natural decay of memory.

Phases 1 and 2: Evaluating Performance

Throughout the study sessions in both phases, an observer focused on either the imager or the loader by noting the time taken to image each tray of mosquitos, the number of errors in critical tasks in the workflow, and any observational artifacts worth noting about usability. In phase 1, one observer focused on the VHT member in the VHT and VCO pair, evaluating them in both the loader and imager roles. In phase 2, two observers were placed per pair, with 1 observer matched with 1 VHT member (so that the observer could evaluate 1 VHT member in both the imager or loader roles).

SUS and Demographic Questionnaires

The SUS was administered once to each participant in the study after the completion of both the roles, an imager and a loader. Each participant was administered this orally, with the question and the scoring scale being described. Their answers were then recorded, along with any qualitative feedback they had with each answer.

In this usability study, a variety of ages, genders, and technological savviness were encompassed in the participant population. A total of 8 participants were chosen for the study from Mayuge and 12 from Adjumani. Each participant completed an orally administered questionnaire, which collected data on their age, gender, experience using a smartphone, experience with vector surveillance processes, and the length of time served as a VHT member. Besides the question on the number of years of VHT experience, all other questions were answered using a standard yes or no format for ease of standardization. This questionnaire is presented in Multimedia Appendix 3.

Planned Statistical Analysis

A Kruskal-Wallis 1-way analysis (α=.05) was performed to test for differences in SUS scores based on (1) the gender of the participant and (2) the participant's experience with a smartphone. This test was selected as it would prove to be more stable in case of any outliers in the samples.

After using VectorCam, each participant was questioned individually by both a translator and the study designers (authors BG, SD, and KKM). There was no time limit set for answering questions, and all users were also able to provide qualitative comments at the end of the survey.

Results

Participant Demographics

Of the 20 participants, 14 (70%) were male and 6 (30%) were female, with a mean age of 36 (SD 9.8) years. A total of 8 (40%) participants were aged 30 to 40 years, 6 (30%) participants were aged 20 to 30 years, 4 (20%) participants were aged 40 to 50 years, and only 2 (10%) participants were aged >50 years. A total of 12 (60%) participants had experience using a smartphone, and 4 (20%) participants had experience with vector control strategies (notably, all these participants were from Adjumani).

Phase 1: Partner With VCO

Efficiency

Efficiency was defined as the time taken (in seconds) to image 14 mosquitos in phase 1. During phase 1 of the study, where each VHT member was paired with a trained VCO, each VHT member both imaged and loaded 14 mosquitos while in their 2 different roles. The average time spent imaging per mosquito for 20 VHT members was 70 (SD 30.3) seconds, and the average time spent loading per mosquito was 50 (SD 12.4) seconds.

Effectiveness

Effectiveness was analyzed based on the failure rates demonstrated by the VHT members when using VectorCam, which were categorized as a failure in the critical tasks outlined for the imager and the loader. During phase 1, for VHT members, there was an 8% (6/80 critical errors) error rate for the imager and a 13% (10/80 critical errors) error rate for the loader. The most frequently failed tasks for the imager and the loader were during imaging: entering the background information and setting up trays (3/20, 15% and 8/40, 20%, respectively). A larger breakdown of the error rates of both the imager and loader is provided in Tables 3 and 4, respectively.

Table 3. Effectiveness and error rates and number of errors for each critical task for the imager role in both phases 1 and 2, including the error rates for vector control officers (VCOs) in phase 2. Overall, 20 village health teams were evaluated in each phase, and 4 VCOs were evaluated in phase 2.

Imager	Setup, n (%)	Imaging		Takedown, n (%)
		Entering background information, n (%)	Taking pictures, n (%)	
Phase 1 (n=20)	0 (0)	3 (15)	2 (10)	1 (5)
Phase 2 (n=20)	0 (0)	3 (15)	5 (25)	3 (15)
VCO phase 2 (n=4)	0 (0)	1 (25)	1 (25)	0 (0)

Table 4. Effectiveness and error rates and number of errors for each critical task for the loader role in both phases 1 and 2, including the error rates for vector control officers (VCOs) in phase 2. Overall, 20 village health teams were evaluated in each phase, and 4 VCOs were evaluated in phase 2.

Loader	Setup 1, n (%)	Setup 2, n (%)	Setup 3, n (%)	Setup 4, n (%)	Takedown 1, n (%)	Takedown 2, n (%)	Takedown 3, n (%)	Takedown 4, n (%)
Phase 1 (n=20)	3 (15)	5 (25)	—[a]	—	1 (5)	1 (5)	—	—
Phase 2 (n=20)	7 (35)	3 (15)	3 (15)	2 (10)	2 (10)	1 (5)	0 (0)	1 (5)
VCO phase 2 (n=4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

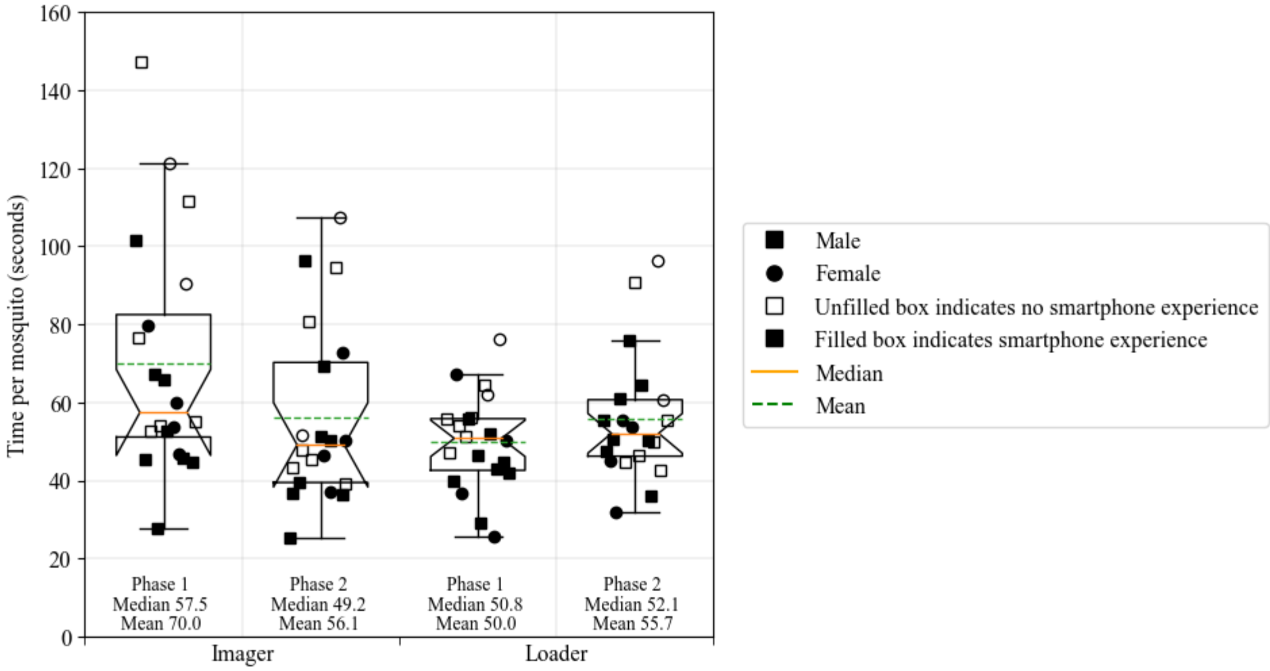
[a]Not applicable.

Phase 2: Two VHT Participants Working Independently

Efficiency

During phase 2, a total of 48 mosquitos were imaged by the VHT member in the imager role, and the average time spent imaging per mosquito was 56.1 (SD 22.9) seconds for the VHT members. The average time spent loading per mosquito was 55.7 (SD 16.3) seconds. When phase 2 was performed by 4 VCOs with much higher training and previous smartphone experience, they were able to accomplish a throughput of 22.4 (SD 4.0) seconds per mosquito. The box-and-whisker plots for the 2 phases of efficiency and both roles are consolidated in Figure 4.

Figure 4. The box-and-whisker plots for efficiency metric in both phases and roles. The first 2 plots show the efficiency plots for the imager role in phase 1 and phase 2, respectively. Both the median and the mean time per mosquito decreased from phase 1 to phase 2. The third and fourth plots are the efficiency plots for the loader role in phase 1 and phase 2, respectively. Conversely, for the imager role, both the median and mean time per mosquito increased from phase 1 to phase 2.



Effectiveness

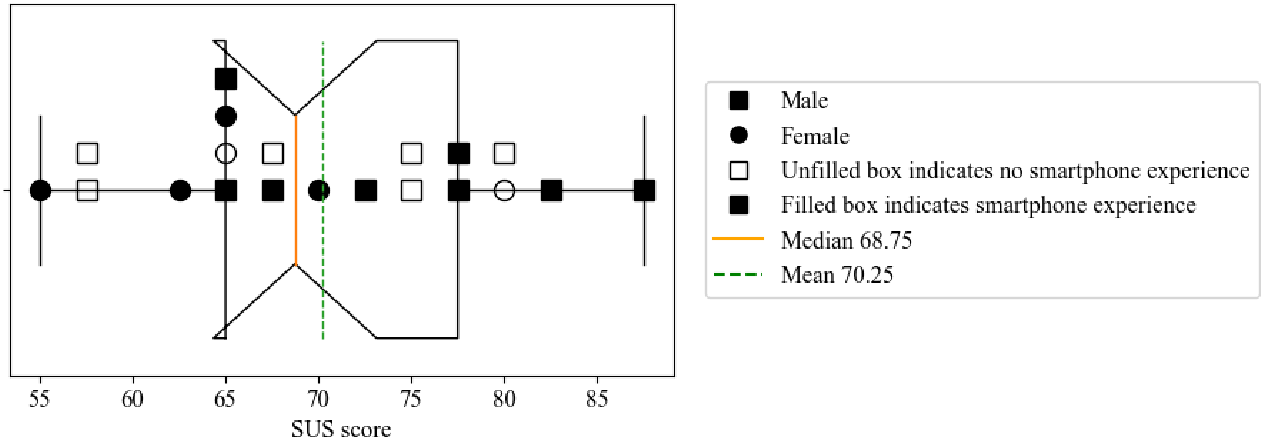
During phase 2, there was a 14% (11/80 critical errors) error rate for the imager (for VHT pairs) and a 12% (19/160 critical errors) error rate for the loader (for VHT pairs). The most frequently failed task for the imager was imaging: taking pictures. The specific subtask that was the most erroneous was the zooming in of the specimen. For the loader, the most frequently failed task was during the setting up of trays (most notably, on the first tray, as seen in Table 4), with the most erroneous subtasks being the puncturing of the Eppendorf tubes and the transfer of the mosquito specimen to the tray. When phase 2 was performed by the 4 VCOs, the total failure rates for the imager role and loader role were 13% (2/16 critical errors) and 0% (0/32 critical errors), respectively, with "imaging: taking pictures" being the most frequently failed task for the imager role. It is important to acknowledge the smaller sample size of VCOs during this evaluation, and for a one-to-one comparison with VHT pairs, a larger sample size must be obtained.

A larger breakdown of the error rates of both the imager and loader is provided in Tables 3 and 4, respectively.

Comfort

At the end of the training and validation period, the SUS score and respective background information were collected. On average, the SUS score reported by participants was 70.25 (SD 8.99). The average scores across all positive and negative items were 4.48 (SD 0.72) and 2.86 (SD 0.71), respectively. The box-and-whisker plot for the distribution of the SUS scores is presented in Figure 5. The item that serves as a statistical outlier is item 10 ("I needed to learn a lot of things before I could get going with this system").

Figure 5. System Usability Scale (SUS) score of study participants. Each participant is stratified by sex as well as experience with a smartphone. The mean SUS score was 70.25 (SD 8.99) for all participants. Using a Kruskal-Wallis test, it was revealed that there was no significant difference in SUS scores based on smartphone experience or gender (α=.05).



An analysis of the satisfaction of the usability of the device was performed by looking at the difference in SUS rankings for the positive items versus the negative items in the SUS questionnaire. Figures 6 and 7 show the distribution and percentages of each SUS ranking stratified by the positive or negative categorization of the items.

Figure 6. The distribution of the System Usability Scale (SUS) scores for each of the positive usability questions asked in the poststudy interview of the participants. The figure demonstrates that most users answered with a score of 4 or 5 (agree or strongly agree, respectively) for each question, as seen by the higher percentages.

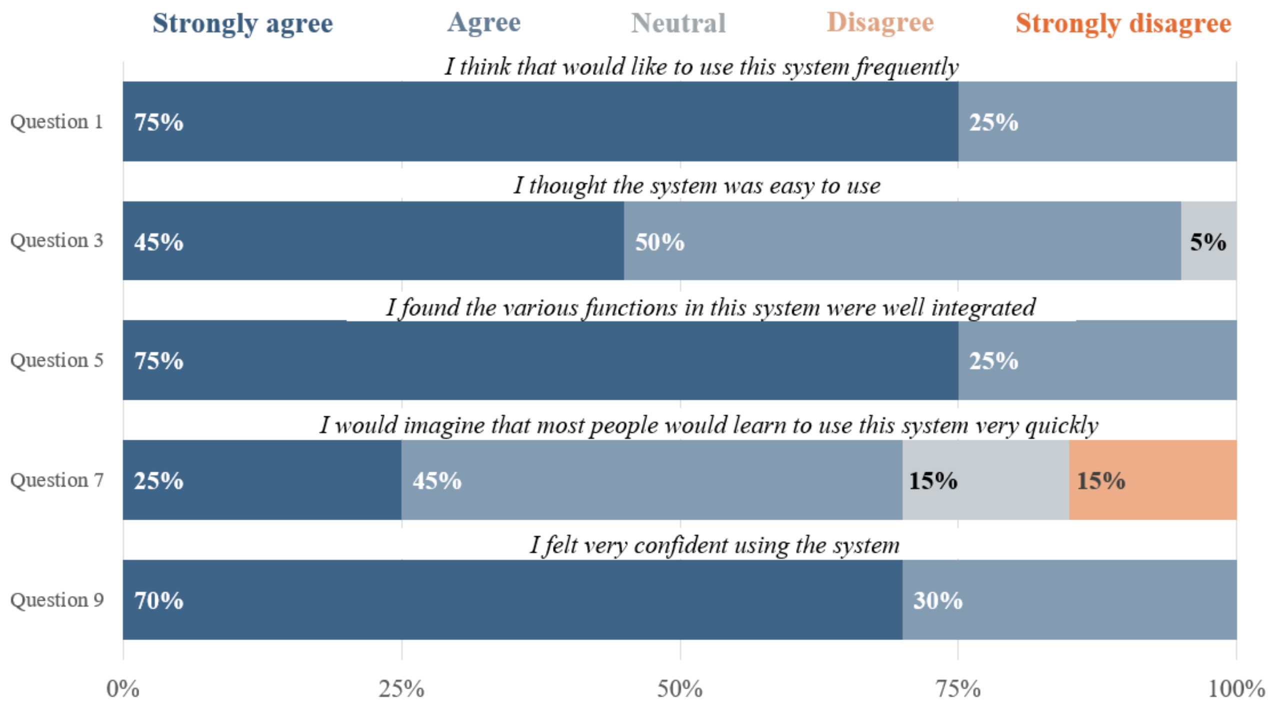
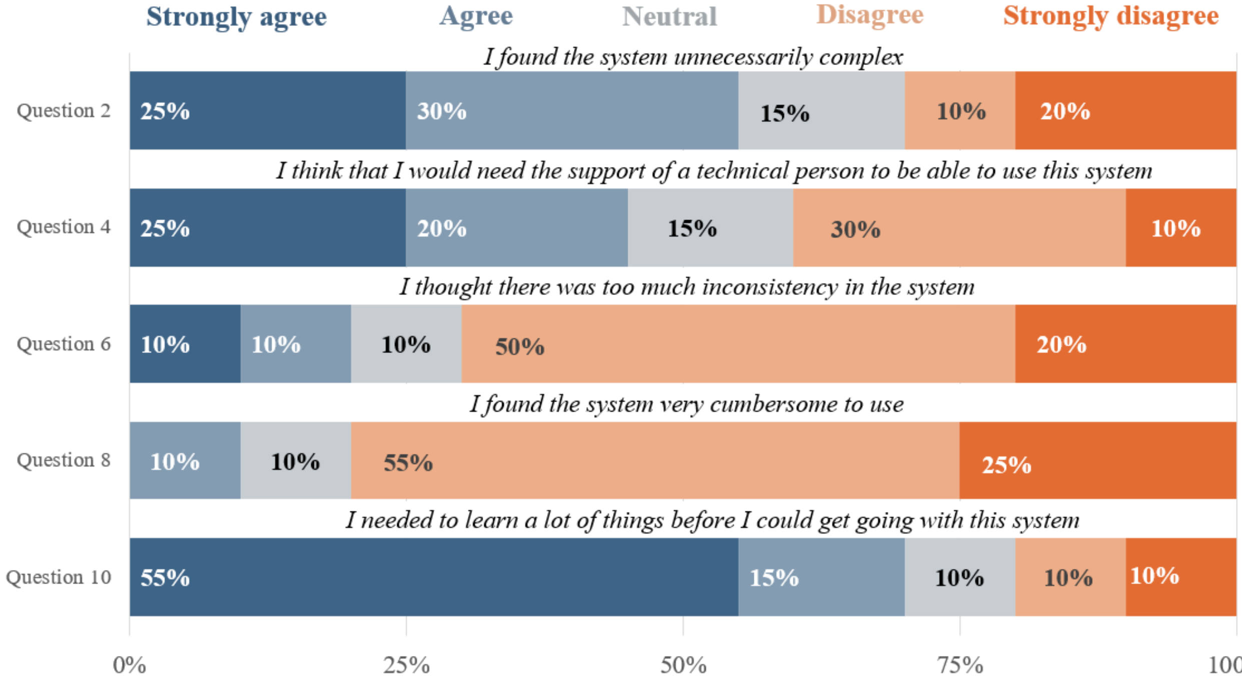


Figure 7. The distribution of the System Usability Scale (SUS) scores for each of the negative usability questions asked in the poststudy interview of the participants. The figure demonstrates that most users answered with a score of 1 or 2 (strongly disagree or disagree, respectively) for each question, as seen by the higher percentages. However, the last question demonstrates the opposite, with most users answering with 4 or 5 (agree or strongly agree, respectively), indicating that users felt that they needed extensive training before using the device properly.



Secondary Results

Kruskal-Wallis Results of the Effects of Gender and Smartphone Experience on Satisfaction

The results of this usability study furthermore revealed that the SUS scores did not exhibit significant differences between the participants across Adjumani and Mayuge (N=20) based on gender or prior experience with smartphones. Using a Kruskal-Wallis test with a significance level of α at .05, there was no statistically significant variance in SUS scores based on gender or prior experience with smartphones (*H* value=0.006).

Thematic Clustering of Participants' Qualitative Feedback

Themes from the postsession interviews were gathered from the feedback given by users about certain qualms of the workflow or interactions in the device. Textbox 3 highlights the main themes of the concerns voiced by the users, along with participant quotes.

Textbox 3. Themes from qualitative comments provided by participants after administering the System Usability Scale. All participants are anonymized and are therefore listed as U1 to U20. Quotes were translated into English for the purposes of this research.

- Need for better training to adequately develop skills needed to use VectorCam
  - "The information from the trainer needs to be clear, or there should be something we can use to know we know we are doing [the imaging process] well." [U15]
- Difficulty handling the specimens with tweezers
  - "It is hard to center the specimens [on the mosquito tray tables] because they break as you use [the tweezers]." [U13]
  - "The biggest challenge was the tweezers being hard to use." [U15]
  - "Putting specimen in tubes was difficult to do with forceps." [U20]
- Difficulty and confusion when attributing a specimen ID to each mosquito
  - "The Specimen ID is very confusing, and I do not understand how the numbers increment." [U9]
  - "The numbers are very confusing; it is confusing to have both letters and numbers [in the specimen ID]." [U18]
  - "The names of [the mosquitos] are hard to remember." [U20]
- Lack of understanding of the impact of the data being recorded and hesitancy in using VectorCam without knowing this impact
  - "I'm not sure where the data goes in general." [U18]
  - "I want to know where the data is being sent to." [U14]

Discussion

Principal Findings

This study is the first summative field usability assessment of VectorCam, an AI-based tool that automatically detects a mosquito's species, sex, and abdomen status, thereby deskilling the identification process. This tool has been designed and manufactured with multiple iterations of input from field users, VCOs, and engineers to be low cost and usable in rural regions in Africa. The results of this study demonstrated that the VectorCam system is overall generally usable in terms of its effectiveness, efficiency, and satisfaction among 20 VHTs in 2 districts in Uganda. A Kruskal-Wallis statistical test revealed that there is no significant difference in usability between sex and end users with and without smartphone user experience.

This study had inherent limitations that shaped the scope and interpretation of its findings. Due to a lower sample size and this study being underpowered, we are unable to confidently claim statistically significant differences in accuracy and throughput based on smartphone use. Furthermore, this study was conducted in an acute short-term setting, where VHT members had limited exposure to the VectorCam device and app (on the magnitude of 6 to 8 hours) before being evaluated on their use of the system. Comparatively, in a programmatic implementation of VectorCam, each VHT member would have multiday exposure to the device before using it independently. Prior usability studies involving medical or field technologies with users who have had long-term training on the device or previous experience with similar products demonstrated higher accuracy in using the device in terms of lower error rates and higher efficiency [18]. A notable example was seen in the lower error rates of medical professionals using a pulse oximeter with a new design [18]. Therefore, results on the accuracy as well as throughput using the VectorCam system may be skewed and could perhaps be improved given a longer training period.

The usability of the platform should also be reassessed following long-term use to further study its efficacy, satisfaction, and accuracy. Comparing the results of phase 1 with phase 2 of this study, the time taken to image each mosquito decreased by an average of 13 seconds per mosquito, as the participants had increased exposure to the device. Comparatively, the time taken by VCOs to image a mosquito was nearly half of that taken by the VHT members, showing that with more training and exposure to using the device, throughput could further increase. As shown in a study evaluating case studies of accessible digital technology for education in a low- and middle-income country setting, the programmatic implementation of tools over a longer period with good transfer and retention of knowledge on how to use them had more success in implementation compared with short-term training and evaluation [19]. Notably, the time taken to identify mosquito vectors using VectorCam was still a considerable improvement over prior methods of vector surveillance, which take a few days, and the time taken to image a mosquito using VectorCam (60 seconds) matched our success criteria [20].

To further improve this throughput, key design changes can be made. The 2 critical steps that took the longest time for the imager were entering the background information on the app as well as entering the specimen ID per mosquito. This study was conducted using an app with English as the primary written language. While English is the national language of Uganda, local dialects vary heavily between regions; therefore, the app would benefit from pictographic or video-based prompts rather than written ones for generalized use or use outside of urban settings. A subgroup analysis for literacy in English was not conducted but would be important to include in further studies.

Furthermore, the outcome measures were captured by a researcher observing the participants and recording the outcome measures manually in real time. Therefore, it is possible that critical errors were missed, and efficiency times may not be as precise as they would have been if they were coded post hoc using video recordings of each participant.

The error rates of the VHT members tested in the imager and loader roles reveal that VectorCam's usability issues need to be addressed with revisions to the software and hardware. An important aspect to consider during the analysis of these rates is the training protocol provided to the VHT members by the VCOs, as mentioned previously. Because of the acute nature of the training and testing of these users, longer exposure and use of the VectorCam system will demonstrate the extent of these failed tasks and the possible reduction of high error rates. This hypothesis can be observed by the 4% (2/48 critical errors) error rate of the tested VCOs compared with the 13% (30/240 critical errors) error rate of the VHT members during phase 2. With this information, we can speculate that with longer training, the error rate would further decrease.

The task with the highest failure rate for the imager was taking pictures on the app. The subtask within taking pictures that was most prone to error was the zooming in of the specimen during the imaging process, which was omitted or not performed thoroughly by the imagers in some cases. To address these design inefficiencies, some design recommendations include a built-in auto-zoom feature that prevents the need for the user to zoom into the specimen in addition to adding a user profile to prevent typing of the background information. The specimen ID can also be automated, as mentioned previously, through an optical character recognition feature. These key design changes would create improvements in both types of error rates noted, as well as increase the throughput by putting less burden on the user to enter these metrics.

For the loader role, the task with the highest failure rate for both phases was the setting up of the mosquito trays. The researchers observed that the transfer of the mosquito specimen to the tray and puncturing of the Eppendorf tube were the subtasks prone to error in this task. Therefore, some hardware design revisions were proposed, such as a sliding edge to eliminate the use of tweezers for transferring mosquito specimens. Another design idea is a clasping apparatus with a sharp edge attached to the Eppendorf tube holder to automatically puncture the standing tubes with a single closing motion. This could prevent any accidental punctures to the skin, which is a potential hazard with the current process using tweezers.

Conclusions

Vector surveillance and entomological classification are the cornerstones of malaria prevention. Through this study, it has been demonstrated that VectorCam, an AI-based tool for task-shift vector surveillance, can effectively empower VHTs to conduct vector surveillance. This study has illustrated VectorCam's usability and accessibility, showing its potential to task-shift the time-consuming and resource-intensive process of vector identification to VHTs embedded in malaria prevention strategies within their communities. Modifications to the hardware and software solutions that are currently in progress are needed to ensure optimal usability and are the current focus of ongoing efforts by our research group. Furthermore, from September 2023 to August 2024, our team is conducting a randomized controlled trial of VectorCam to evaluate the long-term use of VectorCam in the hands of VHTs over a 12-month period. Nevertheless, VectorCam serves as a promising technology that can transcend barriers in traditional vector surveillance through task-shifting malaria prevention efforts, creating a new 21st century approach to community-based malaria vector surveillance in rural Africa and beyond.

Acknowledgments

This study was supported by a grant from the Innovative Vector Control Consortium (IVCC). Additional support was provided by the Makerere University School of Public Health and the Johns Hopkins Center for Bioengineering Innovation and Design (CBID). The authors would like to thank the support of the vector control officers Kigongo Siriman, David Onanyang, Jovan Batte, and James Kaweesa in executing this study. The authors extend additional acknowledgments to the National Malaria Control Division (NMCD) within the Ministry of Health in Uganda for aid in identification of target districts used in the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The VectorCam training protocol script that each vector control officer had to follow as they trained each village health team (VHT) member in the use of VectorCam. This involves an introductory script, where all parts of VectorCam are explained. Afterward, this describes the protocol that must be followed to train each VHT member in the imager and loader roles. Before completing the usability test, they must be able to perform these steps once alone, with no errors.

[PDF File (Adobe PDF File), 85 KB - humanfactors_v11i1e56605_app1.pdf]

Multimedia Appendix 2

This checklist details the steps that the village health team (VHT) members needed to perform to demonstrate their ability to complete critical tasks independently, for both the imager and the loader roles. Only when each of these tasks were checked off by a vector control officer, the VHT members were able to participate in the usability study. This was done to ensure that all VHT members were trained to a standardized level.

[PDF File (Adobe PDF File), 177 KB - humanfactors_v11i1e56605_app2.pdf]

Multimedia Appendix 3

This questionnaire was administered to each village health team member by a vector control officer before conducting training or any part of the usability assessment. In this questionnaire, demographic questions about their age, gender, smartphone use, and vector surveillance experience were asked.

[PDF File (Adobe PDF File), 73 KB - humanfactors_v11i1e56605_app3.pdf]

References

1. Malaria. World Health Organization. 2023 Dec 4. URL: https://www.who.int/news-room/fact-sheets/detail/malaria [accessed 2024-07-22]
2. World malaria report 2021. World Health Organization. 2021. URL: https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2021 [accessed 2024-07-22]
3. Anjorin S, Okolie E, Yaya S. Malaria profile and socioeconomic predictors among under-five children: an analysis of 11 sub-Saharan African countries. Malar J 2023 Feb 14;22(1):55 [FREE Full text] [doi: 10.1186/s12936-023-04484-8] [Medline: 36788541]
4. Uganda. Target Malaria. URL: https://targetmalaria.org/about-us/where-we-operate/uganda/ [accessed 2024-07-22]
5. Pryce J, Medley N, Choi L. Indoor residual spraying for preventing malaria in communities using insecticide-treated nets. Cochrane Database Syst Rev 2022 Jan 17;1(1):CD012688 [FREE Full text] [doi: 10.1002/14651858.CD012688.pub3] [Medline: 35038163]

6. Mogeni P, Omedo I, Nyundo C, Kamau A, Noor A, Bejon P. Effect of transmission intensity on hotspots and micro-epidemiology of malaria in sub-Saharan Africa. BMC Med 2017 Jun 30;15(1):121 [FREE Full text] [doi: 10.1186/s12916-017-0887-4] [Medline: 28662646]
7. van de Straat B, Russell TL, Staunton KM, Sinka ME, Burkot TR. A global assessment of surveillance methods for dominant malaria vectors. Sci Rep 2021 Jul 28;11(1):15337 [FREE Full text] [doi: 10.1038/s41598-021-94656-w] [Medline: 34321525]
8. Erlank E, Koekemoer LL, Coetzee M. The importance of morphological identification of African anopheline mosquitoes (Diptera: Culicidae) for malaria control programmes. Malar J 2018 Jan 22;17(1):43 [FREE Full text] [doi: 10.1186/s12936-018-2189-5] [Medline: 29357873]
9. Audisio P. Insect taxonomy, biodiversity research and the new taxonomic impediments. Fragm Entomol 2017 Dec 22:121-124. [doi: 10.4081/fe.2017.252]
10. Hopkins GW, Freckleton RP. Declines in the numbers of amateur and professional taxonomists: implications for conservation. Anim Conserv 2006 Feb 28;5(3):245-249. [doi: 10.1017/S1367943002002299]
11. Ayala MJ, Valiati NC, Bastos LS, Villela DA. Notification of malaria cases in the Brazilian Amazon Basin from 2010 to 2020: an analysis of the reporting times. Malar J 2023 Feb 10;22(1):49 [FREE Full text] [doi: 10.1186/s12936-023-04464-y] [Medline: 36765345]
12. Goodwin A, Padmanabhan S, Hira S, Glancey M, Slinowsky M, Immidisetti R, et al. Mosquito species identification using convolutional neural networks with a multitiered ensemble model for novel species detection. Sci Rep 2021 Jul 01;11(1):13656 [FREE Full text] [doi: 10.1038/s41598-021-92891-9] [Medline: 34211009]
13. Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. Behav Res Methods Instrum Comput 2003 Aug;35(3):379-383 [FREE Full text] [doi: 10.3758/bf03195514] [Medline: 14587545]
14. Spencer R. The streamlined cognitive walkthrough method, working around social constraints encountered in a software development company. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2000 Presented at: CHI '00; April 1-6, 2000; The Hague, The Netherlands. [doi: 10.1145/332040.332456]
15. Nielsen J, Molich R. Heuristic evaluation of user interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1990 Presented at: CHI '90; April 1-5, 1990; Seattle, WA. [doi: 10.1145/97243.97281]
16. Coldiron ME, Lasry E, Bouhenia M, Das D, Okui P, Nyehangane D, et al. Intermittent preventive treatment for malaria among children in a refugee camp in Northern Uganda: lessons learned. Malar J 2017 May 23;16(1):218 [FREE Full text] [doi: 10.1186/s12936-017-1869-x] [Medline: 28535793]
17. Brooke J. SUS: a 'quick and dirty' usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation In Industry. London, United Kingdom: CRC Press; 1996.
18. Chaniaud N, Métayer N, Megalakaki O, Loup-Escande E. Effect of prior health knowledge on the usability of two home medical devices: usability study. JMIR Mhealth Uhealth 2020 Sep 21;8(9):e17983 [FREE Full text] [doi: 10.2196/17983] [Medline: 32955454]
19. Grönlund Å, Lim N, Larsson H. Effective use of assistive technologies for inclusive education in developing countries: issues and challenges from two case studies. Int J Educ Dev Inf Commun Technol 2010 Jan;6(4):5-26 [FREE Full text]
20. Guidelines for West Nile virus surveillance and control. Centers for Disease Control and Prevention. 2022 Jun 06. URL: https://www.cdc.gov/west-nile-virus/php/surveillance-and-control-guidelines/?CDC_AAref_Val=https://www.cdc.gov/mosquitoes/guidelines/west-nile/surveillance/environmental-surveillance.html [accessed 2024-07-11]

Abbreviations

AI: artificial intelligence
PCR: polymerase chain reaction
SUS: System Usability Scale
VCO: vector control officer
VHT: village health team

Edited by C Jacob; submitted 28.01.24; peer-reviewed by S Matin, J Brooke, T Davidson, A Hassan; comments to author 10.03.24; revised version received 03.05.24; accepted 14.06.24; published 16.08.24.

Please cite as:
Dasari S, Gopinath B, Gaulke CJ, Patel SM, Merali KK, Sunil Kumar A, Acharya S
A Handheld Tool for the Rapid Morphological Identification of Mosquito Species (VectorCam) for Community-Based Malaria Vector Surveillance: Summative Usability Study
JMIR Hum Factors 2024;11:e56605
URL: https://humanfactors.jmir.org/2024/1/e56605
doi:10.2196/56605
PMID:

Original Paper

Design and Psychometric Evaluation of Nurses' Mobile Health Device Acceptance Scale (NMHDA-Scale): Application of the Expectation-Confirmation Theory

Narjes Mirabootalebi[1*], MSc; Zahra Meidani[1,2*], PhD; Hossein Akbari[3], PhD; Fatemeh Rangraz Jeddi[1,2], PhD; Zahra Tagharrobi[4], PhD; Walter Swoboda[5], Dr med; Felix Holl[5], MSc, MPH, PhD

[1]Health Information Management Research Center (HIMRC), Kashan University of Medical Sciences, Kashan, Iran

[2]Department of Health Information Management & Technology, Allied Medical Sciences Faculty, Kashan University of Medical Sciences, Kashan, Iran

[3]Social Determinants of Health Research Center, Kashan, Iran

[4]Trauma Nursing Research Centre, Kashan University of Medical Sciences, Kashan, Iran

[5]DigiHealth Institute, Neu-Ulm University of Applied Sciences, Neu-Ulm, Germany

[*]these authors contributed equally

Corresponding Author:
Zahra Tagharrobi, PhD
Trauma Nursing Research Centre
Kashan University of Medical Sciences
Isfahan Province
Kashan, 2C74+M7F
Iran
Phone: 98 9131613899
Email: Tagharrobi_z@kaums.ac.ir

Abstract

Background: The use of mobile tools in nursing care is indispensable. Given the importance of nurses' acceptance of these tools in delivering effective care, this issue requires greater attention.

Objective: This study aims to design the Mobile Health Tool Acceptance Scale for Nurses based on the Expectation-Confirmation Theory and to evaluate it psychometrically.

Methods: Using a Waltz-based approach grounded in existing tools and the constructs of the Expectation-Confirmation Theory, the initial version of the scale was designed and evaluated for face and content validity. Construct validity was examined through exploratory factor analysis, concurrent validity, and known-group comparison. Reliability was assessed using measures of internal consistency and stability.

Results: The initial version of the scale consisted of 33 items. During the qualitative and quantitative content validity stage, 1 item was added and 1 item was removed. Exploratory factor analysis, retaining 33 items, identified 5 factors that explained 70.53% of the variance. A significant positive correlation was found between the scores of the designed tool and nurses' attitudes toward using mobile-based apps in nursing care (r=0.655, P<.001). The intraclass correlation coefficient, Cronbach α, and ω coefficient were 0.938, 0.953, and 0.907, respectively.

Conclusions: The 33-item scale developed is a valid and reliable instrument for measuring nurses' acceptance of mobile health tools.

(JMIR Hum Factors 2024;11:e55324) doi:10.2196/55324

KEYWORDS

mobile health; acceptance; psychometric evaluation; nursing; Expectation-Confirmation Theory; smartphone

Introduction

Mobile phones and other electronic devices are becoming increasingly important for health care professionals [1]. These tools, along with mobile apps, provide numerous benefits for health care professions [2], including time savings [3]; cost-effectiveness [4]; enhanced self-efficacy [5]; greater access to evidence-based resources [6]; reduced medication errors [5]; decision support [6]; medication guides and guidelines [5]; video consultations with other physicians, alerts, and patient education [7]; and improved communication [5]. Given that the majority of hospital staff are nurses [8], and with an estimated 140,000 nurses in Iran according to a 2018 report by the Ministry of Health [9], it has been reported that approximately 80% of these nurses use mobile phones [8]. Another study found that 98% of nursing students use mobile phones to access drug guidelines, and 83% use them to look up medical terminology [10]. Unfortunately, hospitals do not provide personal mobile phones, and their use may raise concerns for health care organizations, potentially prompting them to implement policies to restrict mobile phone usage [11]. These concerns include distractions for nurses [12], infection control issues [13], and patient privacy risks [14]. Given the importance of mobile phones in nursing practice and the large population of nurses [15], health care stakeholders, managers, and planners should develop policies that support the appropriate use of mobile phones in nursing care and ensure its continuity [16]. The first step in achieving this is to assess the current situation, which requires a suitable tool. This tool should propose and validate a comprehensive model that considers individuals' characteristics, technology, and tasks. The model should address factors such as user satisfaction, confirmation, mobile health (mHealth) continuance, maturity, mobility, individual performance, perceived usefulness, and personal habits [17]. In similar studies, tools specifically designed to identify the unique needs of different stakeholders—such as physicians, nurses, and patients, who have distinct priorities and requirements for using mobile tools—have not been developed. Many of these tools are general questionnaires used in health care and other industries, such as mobile banking and e-commerce, rather than being tailored specifically for nursing [18]. Despite the emphasis on the sociotechnical perspective, which highlights the importance of identifying behavioral and social factors alongside technical ones among users of information technology tools [19], the tools used in these studies often fail to adhere to appropriate psychometric principles. Moreover, they frequently overlook essential background, individual, social, organizational, and cultural factors in their design [20]. Research studies have yet to identify a tool with a suitable theoretical framework specifically designed and localized for nursing based on research and psychometric principles. Additionally, some health care organizations are actively working to establish mHealth stations. Therefore, this study aimed to design and psychometrically validate a Mobile Health Acceptance Tool for clinical nurses in Iran, based on the constructs proposed in the Expectation-Confirmation Theory. According to behavior change theories such as the Expectation-Confirmation Theory, nurses who have a positive perception of using mobile devices in health care, and who believe in their usefulness, effectiveness, and ease of use, are more likely to adopt mHealth tools. The concept of acceptance is explained through the constructs of the Expectation-Confirmation Theory. The scale measuring nurses' acceptance of mHealth tools includes questions addressing various dimensions of attitude, belief, and intention to use mobile tools in providing nursing services, based on Expectation-Confirmation Theory structures. The study's key strength was the application of the Expectation-Confirmation Theory to define acceptance of usage behavior, with a focus on the principles and stages of tool design according to the theory. Additionally, the study considered diverse research units based on individual characteristics, particularly employment context.

Methods

Overview

The researchers initially aimed to identify factors influencing nurses' adoption of mobile devices using the Expectation-Confirmation Theory. They identified key factors such as security risk, new technology anxiety, subjective norms, perceived ease of use, and approval. Following this, they reviewed other studies on mobile app evaluation and developed a preliminary list of measurement items.

This research used a tool design and psychometric validation approach, which was conducted in 2 phases.

Preliminary Phase: Designing the Initial Version of the Tool

In the preliminary phase, the tool was designed using the 4-stage approach proposed by Waltz et al [21]. First, the concept to be measured and its constructs were identified based on the Expectation-Confirmation Model. Second, measurement objectives were established based on the characteristics of the acceptance concept in the Expectation-Confirmation Theory, focusing on 8 constructs: perceived ease of use, social influence, new technology anxiety, personal habit, perceived security risk, confirmation, maturity, and perceived usefulness. In the third stage, the initial draft of the tool was developed. This involved a comprehensive search of both Persian and English articles in national and international databases, including PubMed, Ovid, Scopus, Web of Science, Magiran, IranDoc, Noor Mags, Science Direct, Jihad Daneshgahi Scientific Information Center, ProQuest, CINAHL, and SAGE, without time restrictions. The search utilized keywords related to mHealth, nursing, acceptance, scale, and attitude. All articles were critically reviewed to identify scales used to measure the acceptance of mHealth tools. Items from existing tools, including perceived usefulness, user satisfaction, and acceptance [22-25], were gathered from relevant texts, reviewed, categorized, and then integrated. Overlapping and inappropriate items were eliminated, and the research team determined the appropriate number of items for each decision domain. In the fourth phase, the tool's development involved defining the wording of items based on the conceptual constructs of the theory. The researchers formulated item wordings according to the theory's conceptual structures, revised them, and established scoring rules. Considering the cultural context of Iranian society and the organizational structure of Iranian hospitals, the research team reviewed and adjusted the items designed based on existing tools.

For item scoring, a 7-point Likert scale (ranging from 1=completely disagree to 7=completely agree) was used, in line with common principles in attitude measurement tool design. Reverse scoring was applied to negatively phrased questions [26]. The total score of the tool was calculated as the average score of all items.

Phase 2: Psychometric Validation of the Tool

Step 1: Face and Content Validity Assessment

The preliminary questionnaire was reviewed by 10 experts in nursing, tool design, psychometric validation, health information management, and health informatics to assess qualitative content validity. Initially, a qualitative approach was used to gather expert opinions on the questionnaire's comprehensibility, grammar, language, scoring, key aspects, essential components of the concept, and the clarity and simplicity of the items [27]. Based on their feedback, necessary revisions were made to the instrument.

For quantitative content validity assessment, the content validity ratio CVR_{strict}, content validity index (CVI), and modified Kappa statistic (modified Kappa) were used. CVR_{strict} was calculated for each item, considering its necessity. Additionally, CVI and modified Kappa were calculated for each item based on the relevance criterion [28]. The Lawshe table, Waltz and Bausell index, and Polit and Beck criteria were used to evaluate the results for CVI, CVR, and modified Kappa. Additionally, the overall CVI for the entire instrument was calculated using the $S\text{-}CVI_{Average}$ method [29].

To assess face validity qualitatively, in addition to the expert review, the first author (NM) read each question in the questionnaire aloud to 10 clinical nurses. Their interpretations of each question were compared with the original intent. The research team revised items in cases of ambiguity, inconsistency, or difficulty understanding the questions [30]. A professional Persian language editor was also consulted during this phase.

For the quantitative face validity assessment, nurses provided individual opinions on the importance of each item. They rated each item's importance on a 5-point Likert scale. Based on these ratings, an item impact score was calculated for each item, with scores above 1.5 considered desirable [31].

Step 2: Questionnaire Item Analysis, Construct Validity (Factor Analysis, Concurrent Validity, and Comparison of Known Groups), and Ceiling and Floor Effects

Experts recommend a sample size of 100-300 individuals for psychometric validation of a tool, regardless of the number of items [31]. Therefore, this study targeted a sample size of 250 individuals. The inclusion criteria were working in clinical settings (direct patient care), a minimum of 6 months of clinical experience, holding a university degree in nursing, no known psychological disorders, Iranian citizenship, and consent to participate. The exclusion criteria were unwillingness to continue cooperation or withdrawal from completing the questionnaire during the study.

After assessing the face and content validity of the instrument and obtaining the necessary permissions, the first author (NM) visited the nursing offices of hospitals and conducted sampling with an introduction letter. The sampling was performed in a stratified random manner between 2021 and 2022, based on the type of clinical ward. Different clinical wards within hospitals under the coverage of Kashan University of Medical Sciences were identified, and a list of qualified nurses in those wards was compiled. A simple random sample of participants was selected from each ward using a random number table, in proportion to the required sample size and the number of nurses employed in each ward. A questionnaire was used to collect demographic data, including personal and occupational information such as age, gender, education, marital status, work experience, ward of employment, predominant shift, positions in nursing management, smartphone ownership, type of smartphone operating system, daily internet usage duration, and smartphone usage duration. This was accompanied by the initial version of the Mobile Health Acceptance Tool (validated in the final phase of face and content validity assessment) and a single-item tool to assess nurses' attitudes toward using mobile-based apps in nursing care, rated on a scale from 1 (completely disagree) to 7 (completely agree). At the beginning of each shift, the ward was visited. After obtaining the nurses' consent and providing general instructions on completing the tools, the questionnaires were collected at the end of the shift. If a questionnaire was not completed on time, arrangements were made with the respective nurse. If access was impossible or cooperation was not obtained from the selected individual, a replacement was randomly chosen from the same ward.

After collecting the data, item analysis was initially performed using the loop method. Exploratory factor analysis was then conducted using the maximum likelihood method with varimax rotation. Eigenvalues greater than 1 and scree plots were used to determine the number of factors. A factor loading above 0.44 was used as the threshold for item retention. Items were assigned to the factor that conceptually aligned with them based on their common factor loads. After conducting the factor analysis, ceiling and floor effects were evaluated. The instrument's ceiling and floor effects were assessed based on the relative frequencies of samples with the highest and lowest achievable scores [32].

Concurrently, the known-groups comparison method was used to assess the construct validity of the final version of the instrument. Nurses were categorized into 7 groups, ranging from "1=completely agree" to "7=completely disagree," based on their responses to a question evaluating their attitude toward using mobile-based applications in nursing care. The acceptance scores on the Mobile Health Tool Acceptance Scale were then compared across these groups.

Step 3: Reliability Assessment

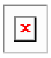
The internal consistency of the final version of the instrument and its subscales (factors extracted during factor analysis) was assessed for the entire sample using Cronbach α coefficient and McDonald ω.

The test-retest method [33] was used to assess the instrument's stability. Ten randomly selected participants from the study completed the final version of the instrument again 1 week later.

The intraclass correlation coefficient between the scores from the 2 assessments was calculated, and standard error of measurement (SEM) and smallest detectable change (SDC) were also estimated [34].

Data Analysis

Data analysis was conducted using SPSS version 16 (IBM Corp.). Quantitative variables were described using measures of central tendency and dispersion, while categorical variables were described using absolute and relative frequencies. Content validity was assessed quantitatively using CVI, CVR, and the modified Kappa statistic, and quantitative face validity was determined using the impact factor. The normality of quantitative data was assessed using skewness and kurtosis indices (for both parameters, the range of –2 to +2 was considered as the normal distribution). To check construct validity, exploratory factor analysis was performed using the maximum likelihood method with varimax rotation. The suitability of the data for exploratory factor analysis was evaluated using the Kaiser-Meyer-Olkin statistic and Bartlett test. The concurrent validity of the instrument was assessed using the Pearson correlation coefficient with the single-item attitude measurement scale. One-way ANOVA was used to compare known groups. The internal consistency of the instrument was assessed using Cronbach α and ω coefficients. Intraclass correlation coefficients were calculated to evaluate the correlation of scores between the 2 assessments in the test-retest. The SEM was computed using Equation 1, where SD represents the SD of scores and *r* is the Cronbach α coefficient.



SDC was reported based on Equation 2. A significance level of <.05 was considered in all analyses.

SDC = 1.96 × √(2×SEM) **(2)**

Ethics Approval

This study was approved by Kashan University of Medical Sciences (KAUMS), Kashan, Iran (ethical code number IR.KAUMS.NUHEPM.REC.1401.039).

Results

Preliminary Phase

The initial draft of the instrument consisted of 33 items across 8 domains: perceived ease of use, social influence, new technology anxiety, personal habit, perceived security risk, confirmation, maturity, and perceived usefulness. These domains included 5, 6, 4, 3, 3, 5, 3, and 4 items, respectively.

Psychometric Phase

Stage 1: Content and Face Validity Assessment

In the qualitative content validity assessment, some items were revised. For example, the item "Nurses can easily use mobile app–based applications in patient care" was changed to "The interaction of nurses with mobile tools for providing nursing services is a simple task." Additionally, the item "The use of mobile apps by nurses in patient care saves time" was added to the perceived ease of use domain.

In the quantitative content validity assessment, the CVR_{strict} for all items, except 1 that was removed, was equal to or higher than the acceptable value specified in the Lawshe table (the minimum acceptable CVR for 10 experts is 0.62). The CVI and the modified Kappa statistic for the 33 retained items were within the range of 0.80-1. Additionally, the $S\text{-}CVI_{Average}$ was calculated to be 0.98.

In the face validity assessment, no changes were made to the items. Additionally, in the quantitative face validity assessment, the impact score for all items was above 1.5.

In summary, following the revisions in the initial psychometric phase, the final version of the instrument retained 33 items.

Stage 2: Questionnaire Item Analysis, Construct Validity (Factor Analysis, Concurrent Validity, and Comparison of Known Groups), and Ceiling and Floor Effects

During the sampling process, 357 (44.1%) eligible nurses out of 810 were selected. Of these, 107 (30%) did not consent to participate, resulting in data analysis for 250 (70%) individuals.

The mean age of the participants was 35.6 (SD 7.6) years, with an average work experience of 11.6 (6.7) years; 231 (92.4%) participants owned mobile phones, which they had used for an average of approximately 9.5 years (Table 1).

Item analysis revealed that removing items with correlation coefficients less than 0.30 or greater than 0.70 with the total score would not significantly impact the instrument's α coefficient. Therefore, all items were retained.

The Kaiser-Meyer-Olkin measure was 0.943, and the Bartlett test of sphericity yielded a chi-square value of 8651.805 (*df*=528, *P*<.001), indicating the suitability of the 33-item instrument for factor analysis. All items had factor loadings above 0.44, and none were removed during this phase. Factor analysis extracted 5 factors that accounted for 70.539% of the total variance in the Mobile Health Tool Acceptance Scale score (see Tables 2 and 3, and Figure 1).

Table 1. Characteristics of clinical nurses working in hospitals under the coverage of Kashan University of Medical Sciences, 2022 (n=250).

Categorized variables	Values, n (%)
Gender	
Male	44 (17.6)
Female	206 (82.4)
Marital status	
Married	201 (80.4)
Single	47 (18.8)
Divorced	2 (0.8)
Widow	0 (0)
Education	
Bachelor's	217 (86.8)
Master's	33 (13.2)
Doctorate	0 (0)
Ward of employment	
Emergency	23 (9.2)
Internal	31 (12.4)
General surgery	50 (20.0)
Intensive care unit	51 (20.4)
Pediatrics	7 (2.8)
Operating room	34 (13.6)
Other	54 (21.6)
Holding a position in nursing management levels	
Yes	49 (19.6)
No	201 (80.4)
Primary shift	
Morning	114 (45.6)
Evening	29 (11.6)
Night	36 (14.4)
Rotating	71 (28.4)
Having a smartphone	
Yes	231 (92.4)
No	19 (7.6)
Mobile operating system (if smartphone; n=231)	
Android	207 (89.6)
Apple iOS	24 (10.4)
Duration of internet usage during the day	
Less than 1 hour	49 (19.6)
1-2 hours	91 (36.4)
2-4 hours	62 (24.8)
More than 4 hours	48 (19.2)

Table 2. Eigenvalue, explained variance percentage, and internal consistency coefficients of the factors extracted from the Mobile Health Tool Acceptance Scale in nurses along with their correlation coefficients with the single-item attitude assessment tool score.

Factor	Question number	Special value	Percentage of variance[a]	Internal consistency coefficient		Correlation with the single-item attitude assessment tool score	
				Cronbach α[b]	ω[c]	Pearson coefficient[d]	*P* value
Factor 1	10	6.703	20.313	0.882	0.923	0.794	<.001
Factor 2	7	4.671	14.155	0.916	0.919	0.642	<.001
Factor 3	5	4.378	13.268	0.935	0.935	0.537	<.001
Factor 4	7	3.844	11.469	0.950	0.950	0.591	<.001
Factor 5	4	3.681	11.155	0.907	0.907	-0.447	<.001

[a]The total percentage of variance explained by each factor was 70.539.

[b]The total value was 0.938.

[c]The total value was 0.953.

[d]The total value was 0.655.

Table 3. Items of the extracted factors in the factor analysis of the Mobile Health Tool Acceptance Scale and their factor loadings[a].

Item number	Item	Extracted factor number[b]				
		1	2	3	4	5
23	Using the mobile app–based tools in nursing care, beyond nurses' expectations, helps in team coordination in processing patient information and making appropriate decisions.	0.818	—[c]	—	—	—
22	Using the mobile app–based tools in nursing care, beyond nurses' expectations, contributes to the improvement of care quality.	0.803	—	—	—	—
24	Using the mobile app–based tools in nursing care, beyond nurses' expectations, accelerates the execution of therapeutic and care interventions.	0.760	—	—	—	—
26	Using the mobile app–based tools in nursing care, beyond nurses' expectations, assists in the proper and effective implementation of clinical care guidelines and instructions.	0.741	—	—	—	—
25	Using the mobile app–based tools in nursing care, beyond nurses' expectations, enhances the management of nursing services.	0.739	—	—	—	—
27[d]	Nurses can use the mobile app–based tools for their primary duties.	0.576	—	—	0491	—
19	The more confident nurses are about the security of patients' information when using the mobile app–based tools for nursing care, the more they use them.	0.573	—	—	—	—
20	Nurses are willing to use the mobile app–based tools for care, provided they are confident that patient or hospital data are not accessible to unauthorized individuals.	0.564	—	—	—	—
18	Nurses must use the mobile app–based tools for nursing care.	0.499	—	—	—	—
21	The use of the mobile app–based tools in providing nursing care increases the potential risk of unauthorized individuals tampering with patient or hospital data.	0.452	—	—	—	—
17	Nurses prefer to use the mobile app–based tools in patient care.	0.442	—	—	—	—
8	Senior hospital and university managers support the use of the mobile app–based tools by nurses.	—	0.783	—	—	—
9	Nurses recommend and emphasize the use of the mobile app–based tools for care to their colleagues.	—	0.736	—	—	—
6	Nursing managers believe that nurses should use the mobile app–based tools in patient care.	—	0.690	—	—	—
10	Physicians welcome the use of the mobile app–based tools in nursing care.	—	0.680	—	—	—
11	Messages sent through social media and group media encourage nurses to use the mobile app–based tools in patient care.	—	0.595	—	—	—
7	Higher authorities such as the Ministry of Health, Treatment, and Medical Education play a vital role in the use of the mobile app–based tools by nurses.	—	0.463	—	—	—
16	Using the mobile app–based tools in patient care is considered normal among nurses.	—	0.442	—	—	—
3	Learning how to use the mobile app–based tools for nursing care is easy.	—	—	0.797	—	—
4	Gaining skills in using the mobile app–based tools for nursing care is easily possible.	—	—	0.770	—	—
1	Nurses can easily use the mobile app–based tools for patient care.	—	—	0.755	—	—
2	Nurses can perform patient care activities more easily using the mobile app–based tools.	—	—	0.739	—	—
5	Nurses' use of the mobile app–based tools in patient care helps save time.	—	—	0.686	—	—
31	Using the mobile app–based tools in nursing care improves the process of collecting, documenting, and analyzing patients' clinical data.	—	—	—	0.738	—
33	Using the mobile app–based tools in nursing care enhances communication among nurses and other members of the health care team.	—	—	—	0.690	—

Item number	Item	Extracted factor number[b]				
		1	2	3	4	5
32	Using the mobile app–based tools in nursing care supports family-centered care and reduces nurses' direct involvement in some interventions, such as medication administration.	—	—	—	0.617	—
30	Using the mobile app–based tools in nursing care increases nurses' productivity.	—	—	—	0.612	—
28[d]	The features of the mobile app–based tools are adaptable and compatible with nurses' clinical performance.	0.525	—	—	0.543	—
29[d]	The mobile app–based tools for assisting with daily nursing activities are sufficiently adequate.	0.488	—	—	0.527	—
15	Nurses experience high stress when using the mobile app–based tools in patient care due to their inability to manage potential problems.	—	—	—	—	–0.864
14	Using the mobile app–based tools in patient care confuses and bewilders nurses, making them feel disoriented.	—	—	—	—	–0.828
12	Nurses have doubts about using the mobile app–based tools for patient care due to the fear of not being able to correct mistakes.	—	—	—	—	–0.719
13	Mandatory use of the mobile app–based tools for patient care causes fear and anxiety in nurses.	—	—	—	—	–0.715

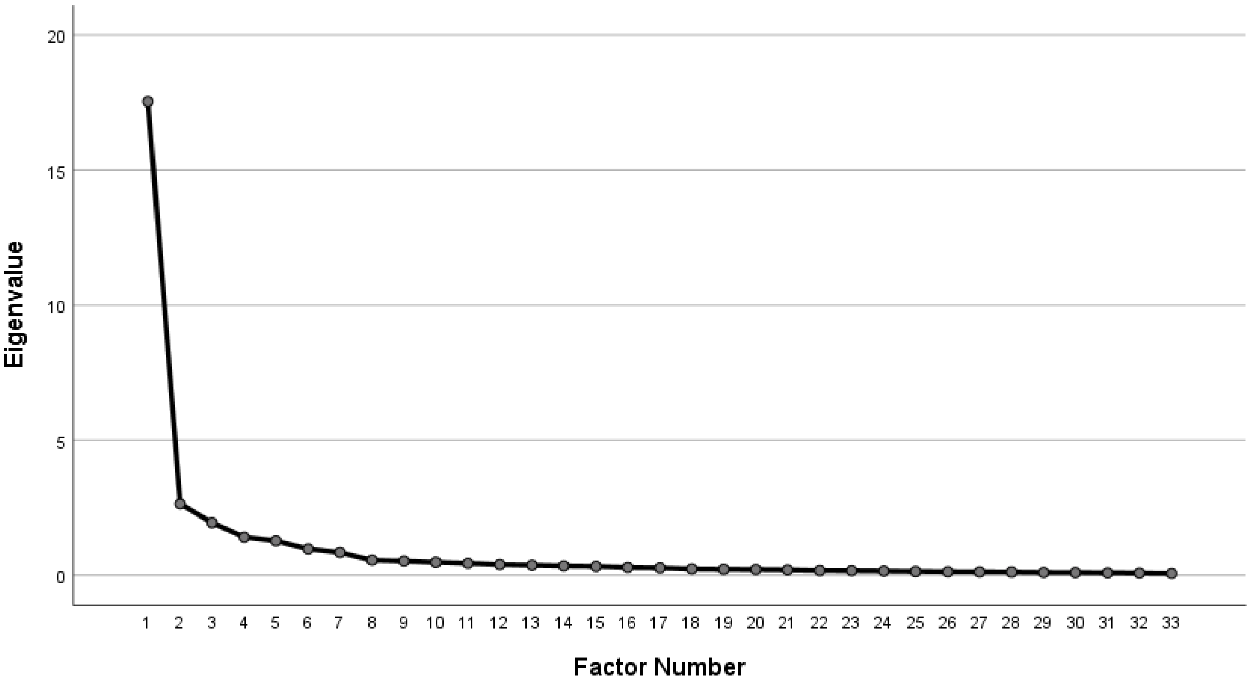
[a]Factor naming is as follows: factor 1 encompasses 10 questions consisting of items 17, 18, 19, 20, 21, 22, 23, 24, 25, and 26, which were named "application and performance"; factor 2 encompasses 7 questions consisting of items 6, 7, 8, 9, 10, 11, and 16, which were named "social impact"; factor 3 encompasses 5 questions consisting of items 1, 2, 3, 4, and 5, which were named "perceived ease of use"; and factor 4 encompasses 7 questions consisting of items 27, 28, 29, 30, 31, 32, and 33, which were named "effectiveness"; and factor 5 encompasses 4 questions consisting of items 12, 13, 14, and 15, which were named "new technology anxiety."

[b]A minimum factor loading of 0.44 was considered. Factor loadings less than 0.44 are not included in the table.

[c]Not available.

[d]For common factor loadings, the item was loaded on a factor that conceptually aligned with the item's content.

Figure 1. Scree plot of the Nurses' Mobile Health Device Acceptance Scale (NMHDA-Scale).



Based on the instrument validated in the exploratory factor analysis (33 items), the mean Mobile Health Tool Acceptance Scale score for nurses was 4.207 (SD 0.740) on a 1-7 scale. With 95% CI, this score is estimated to range between 3.29 and 5.12 for the nursing population working in hospitals under the coverage of Kashan University of Medical Sciences. The mean score for nurses' attitudes toward using the mobile app–based programs in nursing care, as measured by the single-item tool on a 1-7 scale, was 4.340 (SD 1.510). With 95% CI, this score is estimated to range between 2.468 and 6.212 for the target population. In the concurrent validity assessment, a significant positive correlation (P<.001) was found between the Mobile

Health Tool Acceptance Scale score and the single-item tool score measuring nurses' attitudes, indicating that higher acceptance scores were associated with more positive attitudes (Table 3). In the known-groups comparison, one-way ANOVA revealed a statistically significant difference in the Mobile Health Tool Acceptance Scale scores among groups based on their level of agreement or disagreement with using the mHealth tool (P<.001; Table 4).

In the floor and ceiling effect assessment, the relative frequencies of nurses' lowest and highest possible scores on the Mobile Health Tool Acceptance Scale were both less than 15%.

Table 4. Acceptance Score of the Mobile Health Tool Acceptance Scale in nurses, differentiated by their overall agreement or disagreement regarding the use of the mobile app–based nursing care (single-item tool score; n=250)[a].

One-way ANOVA results	Nurses' agreement or disagreement with the use of the mobile app–based tools in nursing care (single-item tool)							Mobile Health Tool Acceptance Scale
	Completely disagree (n=13)	Disagree (n=22)	Somewhat disagree (n=27)	Neutral (n=64)	Somewhat agree (n=67)	Agree (n=43)	Completely agree (n=14)	
Score, mean (SD)	3.263 (0.782)	3.484 (0.833)	3.678 (0.546)	4.019 (0.324)	4.489 (0.443)	4.820 (0.436)	4.869 (1.195)	Welch statistic=31.266 (P<.001)

[a]In terms of comparing the 2 groups, the Games-Howell post hoc test showed that there is a significant difference between the acceptance scores of the mobile health tool in the following group pairs: 1 and 7 (P<.006), 1 and 6 (P<.001), 1 and 5 (P<.001), 2 and 5 (P<.001), 2 and 6 (P<.001), 2 and 7 (P<.02), 3 and 5 (P<.001), 3 and 6 (P<.001), 3 and 7 (P<.03), 4 and 3 (P<.001), 4 and 5 (P<.001), and 6 and 5 (P<.004).

Stage 3: Internal Consistency and Stability

In the assessment of internal consistency, the Cronbach α coefficient and the ω total coefficient for the entire instrument were 0.938 and 0.953, respectively. The coefficients for the 5 extracted factors were also above 0.88 (Table 2).

In the tool's stability assessment, the intraclass correlation coefficient between test and retest scores was 0.907 (95% CI 0.615-0.977, P<.001).

The SEM for the designed instrument was 0.184, and the SDC was 1.19, with 95% CI.

Discussion

Principal Findings

Previous studies have used measurement tools to assess the use of mHealth tools in various health care and other groups [18,35,36]. However, none of these studies utilized a psychometrically tested tool specifically designed for nurses. Therefore, this study aimed to develop and validate the Nurses' Mobile Health Device Acceptance Scale (NMHDA-Scale) based on the Expectation-Confirmation Model.

A 33-item questionnaire was designed to assess the acceptance of mHealth tools among clinical nurses, demonstrating strong validity and reliability within the target population.

The draft of the NMHDA scale was developed using Waltz's 4-step approach and the Expectation-Confirmation Model. This study, grounded in its theoretical framework and the principles of the Expectation-Confirmation Model [37], offers a broader range of acceptance dimensions compared with other existing tools [38,39]. Experts argue that a theoretical foundation for defining the content domains of a tool leads to the creation of relevant items and is crucial for ensuring the tool's validity [40].

In the content validity assessment, revisions to the tool were made based on expert feedback from various relevant and specialized fields. The CVR, CVI, and modified Kappa statistics for all retained items were higher than 0.62, 0.80, and 0.74, respectively. Additionally, the overall S-CVI (S-$CVI_{Average}$) for the entire tool was greater than 0.90. Experts agree that assessing content validity is crucial to ensure the tool covers all essential aspects of the intended concept. The reliability of this process increases with the expertise of the individuals involved in this stage [21]. Moreover, an S-$CVI_{Average}$ greater than 0.90 is considered desirable for content validity [41]. Based on the presented information, the developed tool meets the criteria for establishing content validity.

In the face validity assessment, modifications were made to the tool based on feedback from clinical nurses, the target group. Additionally, the impact scores for all retained items were above 1.5. Connell et al [42] highlighted that face validity is crucial for addressing the needs of the target group, as what researchers consider essential may differ from the perspective of the primary group. Therefore, face validity can enhance the measurement's acceptability, relevance, and quality [42]. Thus, it can be claimed that the tool's items are well understood by the target group, confirming its face validity.

The exploratory factor analysis identified 5 factors: "application and performance," "social impact," "perceived ease of use," "effectiveness," and "new technology anxiety." These factors collectively explained more than 50% of the total score variance, with each contributing over 5%. Additionally, all items had factor loadings exceeding 0.44, and there was one common factor among them. Some experts argue that for construct validity, the identified factors should account for at least 40% of the total variance [43], with each factor explaining more than 5% of the total variability [44]. Additionally, other sources suggest that for robust construct validity, the factors should collectively account for more than 50% of the total variance [45]. Given that the identified factors in this study explained over 50% of the total variance, each contributing more than 5%, the construct validity of the tool is well-established. Therefore, the exploratory factor analysis results suggest the construct validity of the tool. The high factor loadings of the items and the existence of only one common factor further support the desirable structure of the tool [46].

The content of the items loaded onto the extracted factors aligns well with the intended acceptance concept. For instance, factors such as "technology anxiety," "social impact," and "perceived ease of use" directly correspond with theoretical expectations. The factor "usefulness and performance" aligns with the "security risk" factor, while "confirmation" and "effectiveness" correspond with "maturity" and "perceived benefit." Compared with other tools used to assess the acceptance of mHealth tools [47], the structure of this tool is notably more desirable.

A comparison of known groups revealed a significant difference in acceptance scores among nurses based on their agreement or disagreement with using mHealth tools. This finding indicates that the designed tool can effectively differentiate between nurses with varying levels of agreement regarding mHealth tools. Such a result supports the structural validity of the tool, as it is intended to distinguish between groups expected to have differences in a specific characteristic. The observed significant differences further validate the tool's structure [48].

In our study, the relative frequency of the minimum and maximum possible scores obtainable from the acceptance measurement tool was 0, indicating the absence of floor and ceiling effects. Floor and ceiling effects are observed when more than 15% of respondents achieve the highest or lowest possible score on a tool. The absence of such effects in this study suggests that the tool's items are appropriately distributed across the scale, supporting both the content validity and reliability of the instrument [49].

In our study, the total Cronbach α coefficient of the tool was 0.938. According to Shrestha [50], an acceptable lower limit for Cronbach α for reliability is 0.70. Values between 0.60 and 0.80 are considered average, while those between 0.80 and 1.00 are deemed very good [51]. Thus, the high Cronbach α value indicates that the tool demonstrates excellent internal consistency.

The correlation coefficient between the scores obtained from the 2 test sessions was 0.655. Lotfi et al [52] suggested that a correlation coefficient between scores from 2 test sessions indicates the test's stability and repeatability, with coefficients above 0.70 considered acceptable and those above 0.80 very good. Although the coefficient in this study is slightly below the optimal threshold, it still reflects the tool's satisfactory stability and demonstrates higher consistency compared with similar tools [39].

In our study, the SEM was estimated to be 0.184, with the SDC being 1.19. This means that if the test is repeated for an individual, their score may vary by up to 0.184 points. The small SEM supports the tool's stability and reliability [53]. Given the tool's score range, this SEM value indicates the tool's robustness in terms of stability, repeatability, and overall reliability.

Study Limitations

This study has several limitations: data collection during the COVID-19 crisis, which coincided with a high workload for nursing staff, led to a significant number of nurses being unwilling to cooperate. Additionally, the lack of appropriate Persian language tools for assessing convergent validity presents another constraint.

Application of Findings

The current research serves as a valuable tool across various organizational dimensions. It assists senior managers and decision makers in evaluating the potential success of new technologies. For instance, it can contribute to reshaping nursing education curricula by integrating organizational and environmental variables, developing guidelines and regulations for app design, and more. This tool assists nurses in understanding the benefits of technology use by focusing on perceived usefulness, ease of use, adoption concerns, security perceptions, user satisfaction, habit maturity, acceptance, and societal impact. The integration of mobile tools among nurses underscores the need for national policies that provide a clear framework for the design, implementation, and evaluation of mobile apps in nursing.

Conclusions

This study details the development and psychometric evaluation of the Mobile Health Tool Acceptance Scale for nurses, grounded in the Expectation-Confirmation Theory. Exploratory factor analysis confirmed a 5-factor model with 33 items. The factors—usability and performance, social influence, perceived ease of use, effectiveness, and anxiety about new technology—are distinct and collectively capture nurses' intention to accept mHealth tools for professional and job-related purposes. Future studies can utilize this tool to assess the intention of nurses and other health care professionals to use mobile phones for work purposes. This approach allows for the exploration of predictors and outcomes related to both theoretical frameworks and practical applications.

This tool provides valuable insights for managers at various levels, aiding in the creation of guidelines and strategies related to the design and use of mobile apps in the health sector. It also supports the Ministry of Health and Medical Sciences Universities in revising nursing education curricula to enhance the integration of information technology tools and expand their usage based on key efficiency factors. Given that health service providers share common missions and goals and collaborate as a team during clinical activities, it is likely that, with appropriate modifications and psychometric testing, the revised scale could be adapted for use with other target groups within the health care field.

Acknowledgments

This project was financially supported by the Deputy for Research and Technology, Kashan University of Medical Sciences (KAUMS), Kashan, Iran (ethical code number IR.KAUMS.NUHEPM.REC.1401.039; grant number 401038). The article processing charge (APC) was funded through the Open Access Fund of Neu-Ulm University of Applied Sciences (HNU).

Data Availability

The developed Mobile Health Tool Acceptance Scale for Nurses is available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. Guo P, Watts K, Wharrad H. An integrative review of the impact of mobile technologies used by healthcare professionals to support education and practice. Nurs Open 2016 Apr;3(2):66-78 [FREE Full text] [doi: 10.1002/nop2.37] [Medline: 27708817]
2. McBride DL, LeVasseur SA. Personal communication device use by nurses providing in-patient care: survey of prevalence, patterns, and distraction potential. JMIR Hum Factors 2017 Apr 13;4(2):e10 [FREE Full text] [doi: 10.2196/humanfactors.5110] [Medline: 28408359]
3. Oh YS, Yeon JJ, Ens TA, Mannion CA. A review of the effect of nurses? Use of smartphone to improve patient care. Journal of Undergraduate Research in Alberta 2017;6:32-39 [FREE Full text]
4. Cleghorn C, Wilson N, Nair N, Kvizhinadze G, Nghiem N, McLeod M, et al. Health benefits and cost-effectiveness from promoting smartphone apps for weight loss: multistate life table modeling. JMIR Mhealth Uhealth 2019 Jan 15;7(1):e11118 [FREE Full text] [doi: 10.2196/11118] [Medline: 30664471]
5. Beauregard P, Arnaert A, Ponzoni N. Nursing students' perceptions of using smartphones in the community practicum: a qualitative study. Nurse Educ Today 2017 Jun;53:1-6. [doi: 10.1016/j.nedt.2017.03.002] [Medline: 28324823]
6. Lupo R, Zacchino S, Caldararo C, Calabrò A, Carriero MC, Santoro P, et al. The use of electronical devices and relative levels of nomophobia within a group of Italian nurses: an observational study. EBPH 2022 Jan 13;17(1):e13272-1-e13272-10. [doi: 10.2427/13272]
7. Fiorinelli M, Di Mario S, Surace A, Mattei M, Russo C, Villa G, et al. Smartphone distraction during nursing care: systematic literature review. Appl Nurs Res 2021 Apr;58:151405. [doi: 10.1016/j.apnr.2021.151405] [Medline: 33745553]
8. Bautista JR, Lin TT. Sociotechnical analysis of nurses' use of personal mobile phones at work. Int J Med Inform 2016 Nov;95:71-80. [doi: 10.1016/j.ijmedinf.2016.09.002] [Medline: 27697234]
9. Raeissi P, Rajabi MR, Ahmadizadeh E, Rajabkhah K, Kakemam E. Quality of work life and factors associated with it among nurses in public hospitals, Iran. J Egypt Public Health Assoc 2019 Nov 29;94(1):25-28 [FREE Full text] [doi: 10.1186/s42506-019-0029-2] [Medline: 32813080]
10. Sheikhtaheri A, Kermani F. Use of mobile apps among medical and nursing students in Iran. Stud Health Technol Inform 2018;248:33-39. [Medline: 29726416]
11. Bautista J, Rosenthal S, Lin TTC, Theng YL. Psychometric evaluation of the smartphone for Clinical Work Scale to measure nurses' use of smartphones for work purposes. J Am Med Inform Assoc 2018 Aug 01;25(8):1018-1025 [FREE Full text] [doi: 10.1093/jamia/ocy044] [Medline: 29788375]
12. Cho S, Lee E. Distraction by smartphone use during clinical practice and opinions about smartphone restriction policies: a cross-sectional descriptive study of nursing students. Nurse Educ Today 2016 May;40:128-133. [doi: 10.1016/j.nedt.2016.02.021] [Medline: 27125162]
13. Mushabati N, Samutela M, Yamba K, Ngulube J, Nakazwe R, Nkhoma P, et al. Bacterial contamination of mobile phones of healthcare workers at the University Teaching Hospital, Lusaka, Zambia. Infect Prev Pract 2021 Jun;3(2):100126 [FREE Full text] [doi: 10.1016/j.infpip.2021.100126] [Medline: 34368743]
14. Bromwich M, Bromwich R. Privacy risks when using mobile devices in health care. CMAJ 2016 Sep 06;188(12):855-856. [doi: 10.1503/cmaj.160026] [Medline: 27221271]
15. Mayer MA, Rodríguez Blanco O, Torrejon A. Use of health apps by nurses for professional purposes: web-based survey study. JMIR Mhealth Uhealth 2019 Nov 01;7(11):e15195 [FREE Full text] [doi: 10.2196/15195] [Medline: 31682587]
16. Samadbeik M, Garavand A, Kordi M, Abtin A, Asadi H. Factors affecting the acceptance of mobile health by medical sciences students: a cross-sectional study. Iranian J Nursing Midwifery Res 2020;25(6):476. [doi: 10.4103/ijnmr.ijnmr_49_19]
17. Hsiao J, Chen R. Understanding determinants of health care professionals' perspectives on mobile health continuance and performance. JMIR Med Inform 2019 Mar 18;7(1):e12350 [FREE Full text] [doi: 10.2196/12350] [Medline: 30882353]
18. Chopdar PK, Korfiatis N, Sivakumar V, Lytras MD. Mobile shopping apps adoption and perceived risks: a cross-country perspective utilizing the Unified Theory of Acceptance and Use of Technology. Computers in Human Behavior 2018 Sep;86:109-128. [doi: 10.1016/j.chb.2018.04.017]

19. Valenta A, Berner E, Boren S, Deckard G, Eldredge C, Fridsma D, et al. AMIA Board White Paper: AMIA 2017 core competencies for applied health informatics education at the master's degree level. J Am Med Inform Assoc 2018 Dec 01;25(12):1657-1668 [FREE Full text] [doi: 10.1093/jamia/ocy132] [Medline: 30371862]
20. Barry AE, Chaney B, Piazza-Gardner AK, Chavarria EA. Validity and reliability reporting practices in the field of health education and behavior: a review of seven journals. Health Educ Behav 2014 Feb;41(1):12-18. [doi: 10.1177/1090198113483139] [Medline: 23553350]
21. Waltz CW, Strickland OL, Lenz E. Measurement in Nursing and Health Research (5th Edition). New York, NY: Springer; 2017:67.
22. Chen S, Chen J, Chang S. Understanding the antecedents of individuals intention of using cloud services. Journal of Economics and Management 2017;13(2):139-166 [FREE Full text]
23. Almaiah MA. Acceptance and usage of a mobile information system services in University of Jordan. Educ Inf Technol 2018 Feb 20;23(5):1873-1895. [doi: 10.1007/s10639-018-9694-6]
24. Hsiao J, Chen R. Critical factors influencing physicians' intention to use computerized clinical practice guidelines: an integrative model of activity theory and the technology acceptance model. BMC Med Inform Decis Mak 2016 Jan 16;16(1):3 [FREE Full text] [doi: 10.1186/s12911-016-0241-3] [Medline: 26772169]
25. Oghuma AP, Libaque-Saenz CF, Wong SF, Chang Y. An expectation-confirmation model of continuance intention to use mobile instant messaging. Telematics and Informatics 2016 Feb;33(1):34-47. [doi: 10.1016/j.tele.2015.05.006]
26. van Sonderen E, Sanderman R, Coyne JC. Ineffectiveness of reverse wording of questionnaire items: let's learn from cows in the rain. PLoS One 2013;8(7):e68967 [FREE Full text] [doi: 10.1371/journal.pone.0068967] [Medline: 23935915]
27. Torabi M, Safdari R, Nedjat S, Mohammad K, Goodarzi M, Dargahi H. Design and psychometrics of the assessment instrument for innovation capabilities of medical sciences universities using the cube model approach. Iran J Public Health 2020 Feb;49(2):323-331 [FREE Full text] [Medline: 32461940]
28. Zamanzadeh V, Ghahramanian A, Rassouli M, Abbaszadeh A, Alavi-Majd H, Nikanfar A. Design and implementation content validity study: development of an instrument for measuring patient-centered communication. J Caring Sci 2015 Jun;4(2):165-178. [doi: 10.15171/jcs.2015.017] [Medline: 26161370]
29. Melkamu Asaye M, Gelaye KA, Matebe YH, Lindgren H, Erlandsson K. Assessment of content validity for a Neonatal Near Miss Scale in the context of Ethiopia. Glob Health Action 2021 Jan 01;14(1):1983121 [FREE Full text] [doi: 10.1080/16549716.2021.1983121] [Medline: 34694977]
30. Bahariniya S, Ezatiasar M, Madadizadeh F. A brief review of the types of validity and reliability of scales in medical research. JCHR 2021 Jun 30;10(2):100-102. [doi: 10.18502/jchr.v10i2.6582]
31. Montazeri M, Tagharrobi Z, Sooki Z, Sharifi K. Development and psychometric evaluation of the Hospital Nurse Interpersonal Empathy Questionnaire. Int J Nurs Sci 2020 Jul 10;7(3):337-343 [FREE Full text] [doi: 10.1016/j.ijnss.2020.06.012] [Medline: 32817857]
32. Simbar M, Rashidi FF, Ebadi A, Safari S, Alavi MH. The Satisfaction Assessment Tool for Obstetrics Triage (SATOT): development and psychometric properties. Journal of Midwifery and Reproductive Health 2022 Jan;10(1):3075-3083 [FREE Full text] [doi: 10.22038/jmrh.2022.62273.1750]
33. Hamre C, Botolfsen P, Tangen GG, Helbostad JL. Interrater and test-retest reliability and validity of the Norwegian version of the BESTest and mini-BESTest in people with increased risk of falling. BMC Geriatr 2017 Apr 20;17(1):92-98 [FREE Full text] [doi: 10.1186/s12877-017-0480-x] [Medline: 28427332]
34. Ebadi A, Zarshenas L, Rakhshan M, Zareiyan A, Sharifnia S, Mojahedi M. Principles of Scale Development in Health Science. Tehran, Iran: Jame-e-negar; 2017:1-186.
35. Foroughi B, Iranmanesh M, Hyun SS. Understanding the determinants of mobile banking continuance usage intention. JEIM 2019 Oct 11;32(6):1015-1033 [FREE Full text] [doi: 10.1108/jeim-10-2018-0237]
36. Natarajan T, Balasubramanian SA, Kasilingam DL. Understanding the intention to use mobile shopping applications and its influence on price sensitivity. Journal of Retailing and Consumer Services 2017 Jul;37:8-22. [doi: 10.1016/j.jretconser.2017.02.010]
37. Bhattacherjee A. Understanding information systems continuance: an Expectation-Confirmation Model. MIS Quarterly 2001 Sep;25(3):351. [doi: 10.2307/3250921]
38. Wang Y, Lin K. Understanding continuance usage of mobile learning applications: the moderating role of habit. Front Psychol 2021 Nov 10;12:736051-736028 [FREE Full text] [doi: 10.3389/fpsyg.2021.736051] [Medline: 34858271]
39. Tam C, Santos D, Oliveira T. Exploring the influential factors of continuance intention to use mobile apps: extending the Expectation Confirmation Model. Inf Syst Front 2018 May 31;22(1):243-257 [FREE Full text] [doi: 10.1007/s10796-018-9864-5]
40. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, Young SL. Best practices for developing and validating scales for health, social, and behavioral research: a primer. Front Public Health 2018 Jun 11;6:149-118 [FREE Full text] [doi: 10.3389/fpubh.2018.00149] [Medline: 29942800]
41. Jazi Z, Peyrovi H, Zareiyan A. Designing and psychometric evaluation of nurses' social responsibility instrument: a mixed-method study. Iranian J Nursing Midwifery Res 2020;25(2):166. [doi: 10.4103/ijnmr.ijnmr_145_19]

42. Connell J, Carlton J, Grundy A, Taylor Buck E, Keetharuth AD, Ricketts T, et al. The importance of content and face validity in instrument development: lessons learnt from service users when developing the Recovering Quality of Life measure (ReQoL). Qual Life Res 2018 Jul 19;27(7):1893-1902 [FREE Full text] [doi: 10.1007/s11136-018-1847-y] [Medline: 29675691]
43. Watkins MW. Exploratory factor analysis: a guide to best practice. Journal of Black Psychology 2018 Apr 27;44(3):219-246. [doi: 10.1177/0095798418771807]
44. Firouzbakht M, Tirgar A, Ebadi A, Sharif Nia H, Oksanen T, Kouvonen A, et al. Psychometric properties of Persian version of the Short-Form Workplace Social Capital Questionnaire for female health workers. Int J Occup Environ Med 2018 Oct 01;9(4):184-193 [FREE Full text] [doi: 10.15171/ijoem.2018.1264] [Medline: 30325359]
45. Streiner DL. Figuring out factors: the use and misuse of factor analysis. Can J Psychiatry 1994 Apr 01;39(3):135-140. [doi: 10.1177/070674379403900303] [Medline: 8033017]
46. Kaviani N, Tagharrobi Z, Sharifi K, Sooki Z. Development of nurses' beliefs and attitudes towards Complementary and Alternative Medicine Questionnaire and its psychometric evaluation. cmja 2020 Jun 01;10(1):12-33 [FREE Full text] [doi: 10.32598/cmja.10.1.589.2]
47. Wang J, Cao Y. Factors affecting users' continuance intention toward mobile health: integration of theory of consumption value and expectation confirmation. OALib 2023;10(03):1-15. [doi: 10.4236/oalib.1109851]
48. Nikpour M, Tirgar A, Ebadi A, Ghaffari F, Firouzbakht M, Hajiahmadi M. Development and psychometric evaluation of a women shift workers' reproductive health questionnaire: study protocol for a sequential exploratory mixed-method study. Reprod Health 2018 Feb 06;15(1):22-10 [FREE Full text] [doi: 10.1186/s12978-018-0456-0] [Medline: 29409501]
49. Axboe MK, Christensen KS, Kofoed P, Ammentorp J. Development and validation of a self-efficacy questionnaire (SE-12) measuring the clinical communication skills of health care professionals. BMC Med Educ 2016 Oct 18;16(1):2-10. [doi: 10.1186/s12909-016-0798-7]
50. Shrestha N. Factor analysis as a tool for survey analysis. AJAMS 2021 Jan 20;9(1):4-11. [doi: 10.12691/ajams-9-1-2]
51. Daud K, Khidzir N, Ismail A, Abdullah F. Validity and reliability of instrument to measure social media skills among small and medium entrepreneurs at Pengkalan Datu River. International Journal of Development and sustainability 2018;7(3):1026-1037 [FREE Full text]
52. Lotfi M, Tagharrobi Z, Sharifi K, Abolhasani J. Psychometric Evaluation of the Cognitive State Test (COST) in a sample of Iranian elderly people. Iran Red Crescent Med J 2016 May 05;18(5):e23786-e23788 [FREE Full text] [doi: 10.5812/ircmj.23786] [Medline: 27478627]
53. Geerinck A, Alekna V, Beaudart C, Bautmans I, Cooper C, De Souza Orlandi F, et al. Standard error of measurement and smallest detectable change of the Sarcopenia Quality of Life (SarQoL) questionnaire: an analysis of subjects from 9 validation studies. PLoS One 2019 Apr 29;14(4):e0216065-e0216013 [FREE Full text] [doi: 10.1371/journal.pone.0216065] [Medline: 31034498]

Abbreviations

CVI: content validity index
CVR: content validity ratio
mHealth: mobile health
NMHDA-Scale: Nurses' Mobile Health Device Acceptance Scale
SDC: smallest detectable change
SEM: standard error of measurement

Edited by P Santana-Mancilla; submitted 09.12.23; peer-reviewed by R Caruso, Y Chu, O Avarez Cardenas; comments to author 29.03.24; revised version received 18.05.24; accepted 02.07.24; published 17.09.24.

Please cite as:
Mirabootalebi N, Meidani Z, Akbari H, Rangraz Jeddi F, Tagharrobi Z, Swoboda W, Holl F
Design and Psychometric Evaluation of Nurses' Mobile Health Device Acceptance Scale (NMHDA-Scale): Application of the Expectation-Confirmation Theory
JMIR Hum Factors 2024;11:e55324
URL: https://humanfactors.jmir.org/2024/1/e55324
doi:10.2196/55324
PMID:

Clinical Acceptability of a Quality Improvement Program for Reducing Cardiovascular Disease Risk in People With Chronic Kidney Disease in Australian General Practice: Qualitative Study

Caroline McBride[1], BSc, MD; Barbara Hunter[1], BA (Hons), PhD; Natalie Lumsden[1,2], BSc (Hons), PhD; Kaleswari Somasundaram[1], MPhil; Rita McMorrow[1], MB BCh BAO, PhD; Douglas Boyle[1], BSc (Hons), PhD; Jon Emery[1], MB BCh, MA, PhD; Craig Nelson[2,3,4], MBBS, PhD; Jo-Anne Manski-Nankervis[1], BSc (Hons), MBBS (Hons), PhD

[1]Department of General Practice and Primary Care, The University of Melbourne, Melbourne, Australia

[2]Western Health Chronic Disease Alliance, Western Health, Sunshine, Australia

[3]Department of Nephrology, Western Health, Sunshine, Australia

[4]Department of Medicine, The University of Melbourne, Melbourne, Australia

Corresponding Author:
Caroline McBride, BSc, MD
Department of General Practice and Primary Care, The University of Melbourne, , Melbourne, , Australia

Abstract

Background: Future Health Today (FHT) is a technology program that integrates with general practice clinical software to provide point of care (PoC) clinical decision support and a quality improvement dashboard. This qualitative study looks at the use of FHT in the context of cardiovascular disease risk in chronic kidney disease (CKD).

Objective: This study aims to explore factors influencing clinical implementation of the FHT module focusing on cardiovascular risk in CKD, from the perspectives of participating general practitioner staff.

Methods: Practices in Victoria were recruited to participate in a pragmatic cluster randomized controlled trial using FHT, of which 19 practices were randomly assigned to use FHT's cardiovascular risk in CKD program. A total of 13 semistructured interviews were undertaken with a nominated general practitioner (n=7) or practice nurse (n=6) from 10 participating practices. Interview questions focused on the clinical usefulness of the tool and its place in clinical workflows. Qualitative data were coded by 2 researchers and analyzed using framework analysis and Clinical Performance Feedback Intervention Theory.

Results: All 13 interviewees had used the FHT PoC tool, and feedback was largely positive. Overall, clinicians described engaging with the tool as a "prompt" or "reminder" system. Themes reflected that the tool's goals and clinical content were aligned with clinician's existing priorities and knowledge, and the tool's design facilitated easy integration into existing workflows. The main barrier to implementation identified by 2 clinicians was notification fatigue. A total of 7 interviewees had used the FHT dashboard tool. The main barriers to use were its limited integration into clinical workflows, such that some participants did not know of its existence; clinicians' competing clinical priorities; and limited time to learn and use the tool.

Conclusions: This study identified many facilitators for the successful use of the FHT PoC program, in the context of cardiovascular risk in CKD, and barriers to the use of the dashboard program. This work will be used to inform the wider implementation of FHT, as well as the development of future modules of FHT for other risk or disease states.

Trial Registration: Australian New Zealand Clinical Trial Registry ACTRN12620000993998; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=380119&is

(JMIR Hum Factors 2024;11:e55667) doi:10.2196/55667

KEYWORDS

clinical decision support; general practice; GP; primary care; family medicine; general medicine; family physician; implementation science; chronic kidney disease; CKD; nephrology; nephrologist; chronic disease; cardiovascular risk; cardiology; quality improvement; EHR; electronic health record; clinical software

Introduction

Background

The 2011 - 12 Australian Health Survey revealed that 10% of Australian adults had biochemical signs of chronic kidney disease (CKD) [1]. CKD is a significant risk factor for cardiovascular disease [2], which is Australia's leading cause of death [3]. Early intervention can slow the deterioration in kidney function and reduce the risk of cardiovascular complications [4]. Australian CKD guidelines recommend the use of angiotensin-converting enzyme inhibitors (ACEis) or angiotensin-receptor blockers (ARBs) in the presence of albuminuria or hypertension (defined as a blood pressure above 130/80 mm Hg), and statins are recommended for everyone with CKD over the age of 50 years or for younger individuals with CKD in the presence of comorbidities [5].

CKD is underrecognized and undertreated in Australian general practice. One study found that 76.8% of people with biochemical results consistent with CKD did not have the diagnosis recorded in their medical file [6], and another found that only 65% of people with a formal diagnosis of CKD were prescribed an ACEi/ARB while 56% were prescribed a statin [7]. A known barrier to achieving quality improvement in chronic disease is the lack of timely and straightforward access to clinical guidelines [8,9]. Electronic clinical decision support tools aim to address this barrier by integrating with clinical records to automatically provide key information from appropriate clinical guidelines.

Prior Work

The Future Health Today (FHT) program is a general practice quality improvement technology platform developed by the University of Melbourne and Western Health. FHT integrates with the electronic medical record to provide 2 components: a clinical decision support tool active at the point of care (PoC) and a web-based dashboard that facilitates practice-wide audit. Screenshots of FHT are available in Multimedia Appendix 1. FHT was developed using a service design approach, whereby clinicians were involved throughout an iterative technical development process [10]. It underwent optimization at 12 general practice clinics, using guidelines for CKD, cardiovascular disease, and cancer prevention. A qualitative optimization study for the cancer recommendations has been described elsewhere [11]. A cluster randomized controlled trial was completed in 2022 (trial registration: ACTRN12620000993998) with the aim of exploring whether the FHT platform increased the prescription of ACEis/ARBs and statins in people with CKD at high cardiovascular disease risk [12]. This paper reports on barriers and facilitators of clinical performance improvement as reported by clinicians themselves.

Methods

Study Setting

General practice provides primary care to more than 8 in 10 Australians each year [13] and operates on a fee-for-service model, with rebates available through Medicare, the health care insurance scheme funded by the Australian Government Department of Health and Aged Care. General practices may bulk bill (payment covered in full by Medicare), privately bill (patients pay a fee in excess of the Medicare rebate), or mixed bill (variably bulk bill or privately bill patients).

The FHT trial was conducted in 39 general practices in Victoria and 1 general practice in Tasmania, Australia. In summary, clinics were randomized into 2 arms: cardiovascular risk reduction in people with CKD or identification of cancer risk, with each group acting as the control for the other [12]. Twenty-one general practices were allocated to the CKD arm, with 19 practices completing the trial. The trial ran from October 2021 to September 2022. The CKD algorithm was deactivated in December 2021 after errors in algorithm deployment were noted and was reactivated in February 2022. The interviews analyzed in this paper occurred between May and July 2022. The state of Victoria was subject to a COVID-19 pandemic declaration throughout the duration of the trial.

Study Design

This qualitative implementation study focuses on the clinical experience and use of FHT for medication management to reduce cardiovascular risk in people with CKD. Semistructured interviews were conducted with general practitioners (GPs) and practice nurses (PNs) participating in the CKD intervention arm of the FHT trial.

Participants

Participating clinics were recruited via VicREN (Victorian Primary Care Practice-Based Research and Education Network) at the Department of General Practice and Primary Care, The University of Melbourne, and The University of Tasmania's Northern Tasmania General Practice-based research network. Each practice was asked to nominate a practice champion at the beginning of the trial; this tended to be a PN or practice manager. The practice champion was requested to assist with the recruitment of GPs and PNs to participate in one-on-one interviews held between 8 and 10 months into the trial. Recruitment was also facilitated by the FHT trial coordinator. Potential participants were emailed a plain-language statement and consent form to complete if they were interested in participating. Participants were reimbursed with a AUD $50 (US $33.69) debit card voucher.

Data Collection

Interviews occurred over the phone and were recorded for transcription. There were 3 interviewers: CM, NL, and KS. CM is an academic GP registrar, NL is a research fellow, and KS is a research assistant. All are female and are affiliated with the University of Melbourne. NL acted in a liaison role for practices during the trial. An interview guide (Multimedia Appendix 2) was developed to focus on questions regarding the clinical usefulness of the recommendations, impact on clinical workflows, and perceived change in clinical performance.

Data Analysis

Interview data were analyzed using Clinical Performance Feedback Intervention Theory (CP-FIT), a theory for designing, implementing, and evaluating feedback specifically in the health care context [14]. The theory proposes 42 variables that influence a feedback cycle via 7 mechanisms. Each step in the cycle is vital for successful feedback to occur. Each interview was independently coded by CM and BH (a qualitative researcher) using a combination of NVivo (Lumivero) and Microsoft Excel. Any discrepancies in coding were discussed and a consensus was reached.

Ethical Considerations

The overall FHT trial protocol was approved by the Department of General Practice University of Melbourne Human Ethics Advisory Group and the Faculty of Medicine, Dentistry, and Health Sciences Human Ethics Sub-Committee. The ethics ID is 2056564. These ethics committees use the National Statement on Ethical Conduct in Human Research published by the Australian Government National Health and Medical Research Council.

Results

Overview

A total of 13 interviews were held, with staff from 10 clinics. The demographic and clinic details of the 7 GP and 6 PN participants are summarized in Table 1. Participants were assigned a code based on their role (GP or PN) and their clinic's allocated number within the trial. All 10 clinics were located in the state of Victoria and varied in their location (regional Victoria or metropolitan Melbourne) and billing policy (mixed or bulk billing). The sample size aimed to gather perspectives from as many staff at the 19 intervention practices as possible but was limited by the response rate. A total of 5 practices did not respond to the request for interview, 1 canceled citing personal reasons, and 1 canceled citing lack of use of FHT. Duration of interviews ranged from 14 to 39 minutes (average 21 min).

All interviewees had used the FHT PoC tool (N=13), with only 7 interviewees having experience using the FHT dashboard. The 5 GPs not using the dashboard were not aware of it, were not able to access it due to technological issues, or had not had time to use it. Five of the PNs had used the dashboard, but none had used it to recall patients.

Table . Interview participant characteristics.

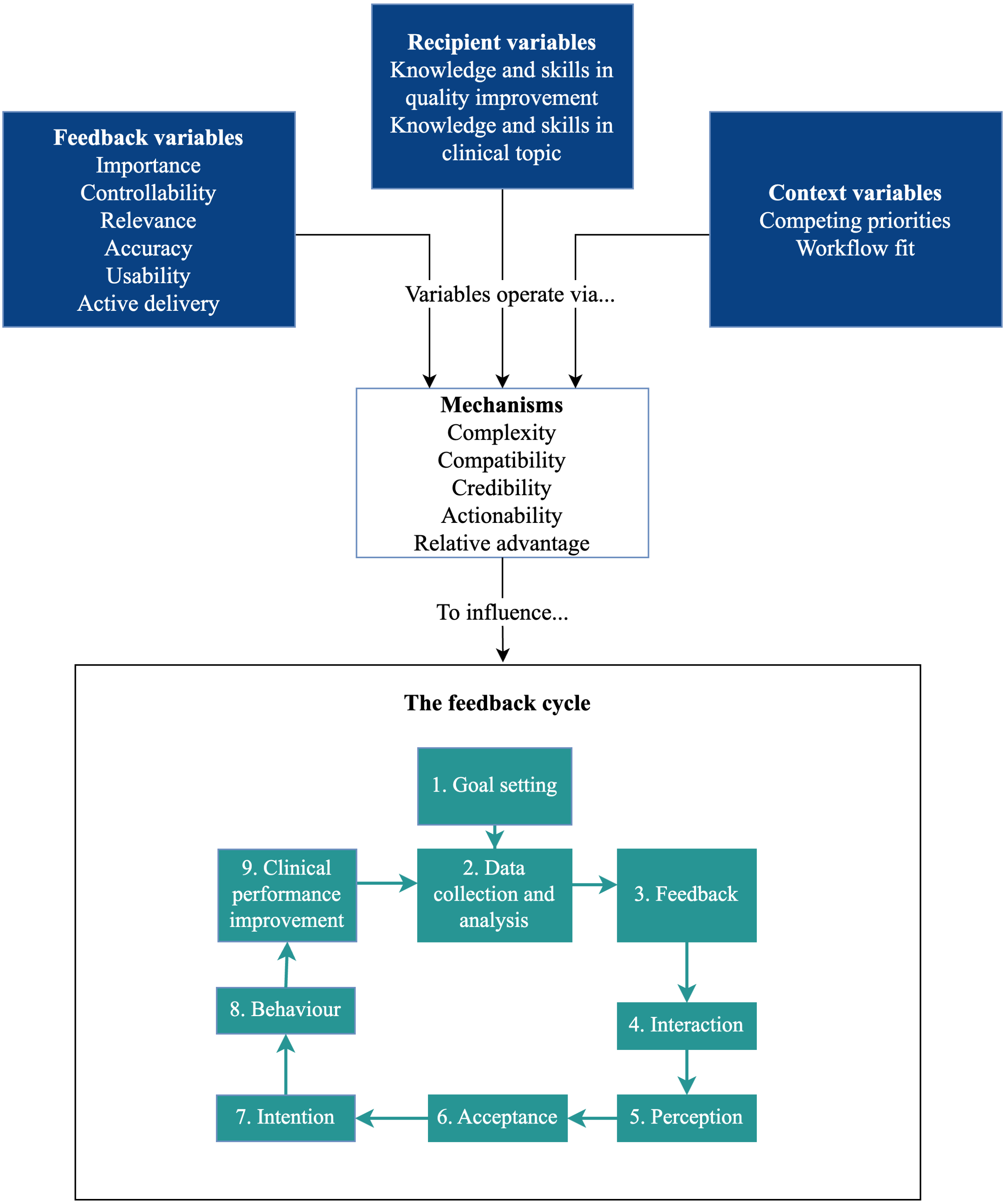
Participant	Sex	Location	Billing policy	Use of FHT[a]
GP7A[b]	Male	Regional	Mixed	PoC[c]
GP7B	Male	Regional	Mixed	PoC
GP9	Male	Metro	Bulk billing	PoC and dashboard
GP22	Female	Metro	Mixed	PoC
GP23	Male	Regional	Mixed	PoC
GP31	Female	Metro	Bulk billing	PoC and dashboard
GP38	Male	Metro	Mixed	PoC
PN5[d]	Female	Metro	Mixed	PoC and dashboard
PN6	Female	Metro	Mixed	PoC and dashboard
PN18	Male	Metro	Bulk billing	PoC
PN19	Female	Metro	Bulk billing	PoC and dashboard
PN22	Female	Metro	Mixed	PoC and dashboard
PN23	Female	Regional	Mixed	PoC and dashboard

[a]FHT: Future Health Today.
[b]GP: general practitioner.
[c]PoC: point of care.
[d]PN: practice nurse.

Interview Findings: The CP-FIT Feedback Cycle

Results were mapped to the CP-FIT feedback cycle. An adapted version of CP-FIT is shown in Figure 1, with variables and mechanisms relevant to this study included in the diagram. In total, 10 explanatory variables and 5 mechanisms were identified as influencing how clinicians used FHT.

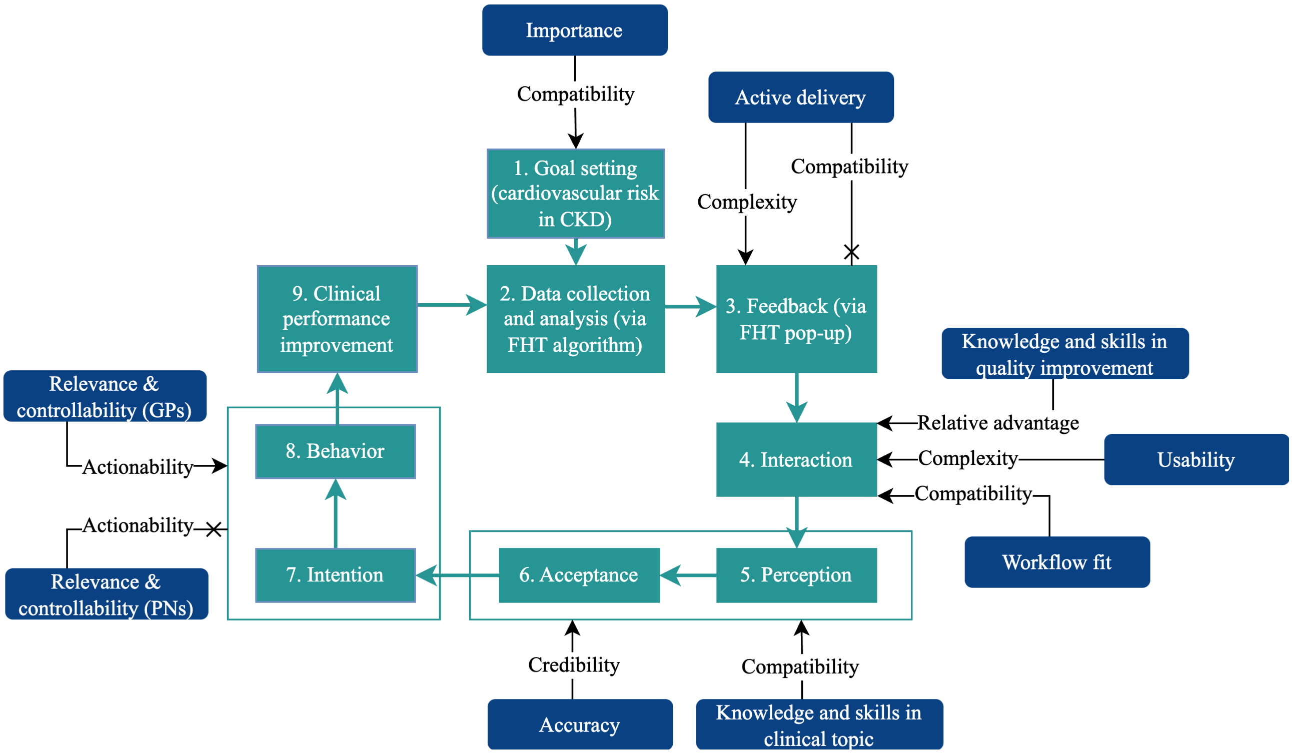
Figure 1. Clinical Performance Feedback Intervention Theory variables, explanatory mechanisms, and feedback cycle (adapted from Brown et al [14], which is published under a Creative Commons Attribution 4.0 International License [15]).



Interview Findings: The PoC Tool

Results for the PoC tool are presented in sequential order following the CP-FIT feedback cycle as shown in Figure 2, with key quotes to illustrate themes. Some steps of the feedback cycle are presented concurrently, reflecting that participants described these processes occurring simultaneously.

Figure 2. The Clinical Performance Feedback Intervention Theory feedback cycle when using the point of care tool (adapted from Brown et al [14], which is published under a Creative Commons Attribution 4.0 International License [15]). CKD: chronic kidney disease; FHT: Future Health Today; GP: general practitioner; PN: practice nurse.



Goal Setting: Importance

It was universally agreed that CKD is an area of importance in general practice. Clinicians reflected on how CKD is underrecognized and undertreated.

> *It's certainly an area that seems to be pretty poorly documented in the patient's records and then consequently not always particularly well treated either.* [GP7A]

> *It's actually really important, because it's – probably 10 per cent of the clinic patients have chronic kidney disease, according to the statistics and probably I don't have all of them already coded in my computer* [GP31]

Data Collection and Analysis

Step 2 occurred automatically via the FHT software program, without needing intervention by clinicians.

Feedback: Active Delivery

Active delivery of feedback via the PoC pop-up facilitated ease of use. Put simply:

> *It's just there. It just pops up. I read it.* [PN5]

The method of delivery was designed to not be intrusive, and this is consistent with how it was perceived.

> *It's quite a minimal popup. Well, it has to be big enough that you see it, but it's small enough that it doesn't just take over the whole of the screen.* [GP7A]

Two participants identified concerns about notification fatigue.

> *[Interviewer: have people been receptive to learning how to use it?] Yeah. But also…that's one more thing we have to deal with. There has been a little bit of reluctance.* [PN5]

> *Any decision support software…you get notification fatigue* [GP7B]

Interaction: Knowledge and Skills in Quality Improvement, Usability, and Workflow Fit

GPs demonstrated a keen interest in quality improvement and proactively reviewed information from the PoC tool to facilitate this.

> *It's always good to review your management.* [GP31]

> *There will be a patient that they are not on what's recommended just because of an oversight there. So I think that it's a good double check.* [GP23]

Clinicians found the PoC tool to be usable because of its information clarity and lack of complexity.

> *They are useful because they're clear, or clear-cut and they're not a big swath of information coming through, and that gives you an opportunity to go into the resources as you need to.* [GP7A]

For GPs, the PoC tool was compatible with existing workflows. Most GPs reported reading and verifying the information in the PoC tool prior to bringing the patient into their consulting room, as part of their usual preconsultation file review. This process was described as very quick (GP23, GP31), and the recommendations were able to be actioned within their standard consultation structure.

> *I usually familiarise myself with the patients before calling them, so I go through the history, recalls, results or whatever, including if there was any popups* [GP31]

> *It's pretty straightforward. The discussion goes something like, "so Mrs. Jones, just looking at your recent renal function, the current recommendations is that we add a statin or you should be on an ACE or ARB. what do you think about that?" They usually say, well, what do you think? And we go down that path of shared decision making.* [GP7B]

PNs described more variation in how they interacted with the PoC tool and integrated it into their workflows. Two described using the program opportunistically (PN5, PN19) during consultations or when reviewing existing recalls. One reviewed the recommendations only when the patient was booked for a long appointment such as a care plan or health assessment. Practices nurses mostly acted on the recommendations by documenting in the patient file or directly discussing with the patient's GP.

> *It's for patients who have long appointments. For example, they're in for a care plan or a health assessment wherein we actually review and check the file for a longer period of time. Unlike those who are only here for, let's say, vaccinations where they will be in and out of the treatment room and we don't really have that much time to review and study.* [PN22]

> *I usually will just alert the doctor and then…the doctor will, based on their own view and they would decide whether or not to follow the advice* [PN6]

Perception and Acceptance: Knowledge and Skills in Clinical Topic and Accuracy of Recommendations

When asked directly, each GP reported a good pre-existing knowledge of CKD; none of the information in the PoC tool was new or surprising. They expressed confidence in interpreting and actioning the recommendations in the context of each patient.

> *Guidelines are guidelines. That's exactly what they are. They're not - they don't actually always - are not always appropriate for the person sitting in front of you. we go down that path of shared decision making.* [GP7B]

> *Every patient that I have I have a look at the recommendations to see what their recommendations are and if it comes up in orange or that I consider their recommendations and follow them unless there's a reason not to. I review my medication and have a good think about it.* [GP23]

Most PNs reported less confidence with CKD management in general but were empowered by the PoC tool to learn more.

> *I don't think I know that area so well, yeah, so it is quite useful to have this pop-up icon.* [PN6]

GPs largely found the recommendations to be accurate. The process for verifying information was straightforward and they felt comfortable discounting recommendations when appropriate. The format of clinical data storage and the way data were recorded in the electronic medical record were identified as potential sources of inaccuracy.

> *I think it seems to be pretty accurate most of the time…normally I will go having a look through the most recent result eGFR and any albumins that are there and just see if there is anything that does meet the criteria for a diagnosis. Then if not, then I'll leave it.* [GP7A]

> *I think a lot of them are on it. It just wasn't documented in the right place…one of the limitations in this clinic in particular is that [GP] handwrites scripts* [PN19]

Intention and Behavior: Usability, Relevance, and Controllability

Themes related to controllability and relevance diverged according to role, due to the recommendations having a focus on medication and prescribing. GPs found that the PoC recommendations were well within their scope and control, and most GPs found them relevant to their patient population.

> *Every time, if it came up with a recommendation, I would act on it to decide whether they should be on it or whether there's a reason not to.* [GP23]

In contrast, PNs described how the PoC had limited direct relevance to their role, which excludes prescribing.

> *I can't prescribe anything, and a lot of it is about prescribing* [PN5]

Clinical Performance Improvement

Overall, these themes reflected the view that the PoC tool was seen as a "prompt" or "reminder" that reinforced pre-existing knowledge in a straightforward manner consistent with usual practice.

> *It's a nice little way of reminding people to do it really…. I think just having it coming up is always reminding me to check what are the best practice principles, things like the ACE and ARB.* [GP7A]

> *I think it's just a good prompt because it does pop up there, when you can get really busy.* [PN19]

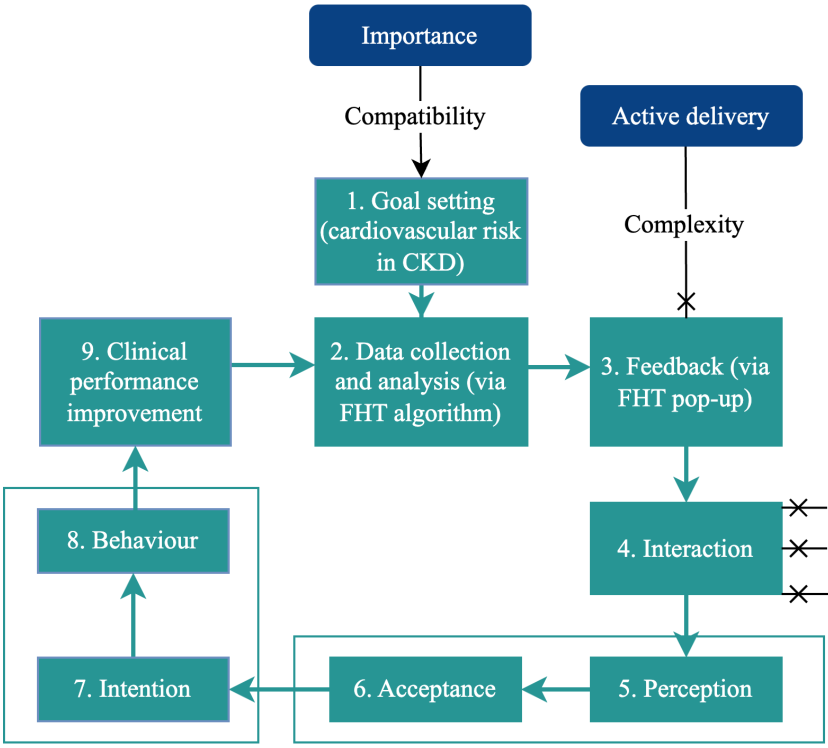
One GP reflected on how the presence of FHT had the effect of encouraging more general clinical performance improvement.

> *It certainly does focus particularly in general consults on preventative health and flags that as an issue to discuss and might then prompt me to not only look at the CKD, but make sure they've had a recent blood pressure and height and weight and make sure their cancer screening is up to date; all of those sorts of things as well.* [GP7A]

Interview Findings: The Dashboard Tool

Results for the dashboard tool are shown in Figure 3. Feedback cycle steps 1 and 2 (goal setting and data collection) operated in the same manner as the PoC tool. However, participants became "stuck" at either steps 3 or 4.

Figure 3. The Clinical Performance Feedback Intervention Theory feedback cycle when using the dashboard tool (adapted from Brown et al [14], which is published under a Creative Commons Attribution 4.0 International License [15]). CKD: chronic kidney disease; FHT: Future Health Today.



Feedback: Active Delivery

Use of the dashboard was significantly limited by its lack of active delivery; multiple clinicians were not aware it existed.

Interaction: Usability, Competing Priorities, and Workflow Fit

Clinicians who had used the dashboard reported some difficulty learning how to navigate the tool, and use of the dashboard was largely incompatible with existing workflows. This was exacerbated by time constraints associated with additional clinical tasks and workforce pressures, particularly with the impacts of the COVID-19 pandemic.

> *I get a little bit lost still in [the dashboard].* [PN19]
>
> *A lot of it is just timing. We're still so busy. I would love to have a bit more time to sit down and just work through things. We're starting to do a little bit now but, really, it's just time.* [PN5]
>
> *With the Pfizer vaccines and all that we are inundated with work…COVID has really taken over* [GP22]

Discussion

Principal Results: The PoC Tool

In this qualitative study exploring the implementation of FHT using CP-FIT, we found that the FHT PoC tool was well received by clinicians and facilitated guideline-informed care. When mapped to the CP-FIT feedback cycle in Figure 2, we could see that they were able to move rapidly through all steps of the process, facilitated by the variables identified. GPs found that the tool was compatible with their quality improvement goals and clinical workflows; was simple to use; and provided credible, clinically useful information. This was supported by their pre-existing confidence in managing CKD, so the tool was seen as a quick reminder system that prompted further personalized review and discussion. Clinicians in our study worked across a range of location and billing styles, but all described similar workflows and clinical knowledge that supported uptake of the tool. This concept of a clinical decision support tool working most effectively as an "aid-memoire" is supported by existing literature [16]. PNs had a slightly different experience given their self-reported lesser knowledge of CKD and limited role in medication prescribing, but nonetheless, they described the PoC tool as useful and simple to use.

Principal Results: Dashboard Tool

Use of the FHT dashboard tool was limited. Most GPs did not get to the "feedback" stage of accessing the dashboard, and most PNs did not progress past initial interactions with the software. During the co-design phase, clinicians had expressed a desire for FHT to facilitate both planned and spontaneous interactions [10], but in practice, most interactions occurred spontaneously. The functions available in the dashboard did align with the goals described by clinicians in both the co-design phase and this implementation study, but this was insufficient to overcome competing clinical priorities, particularly the lack of time and staff resources in the context of the COVID-19 pandemic.

The barriers to the use of the dashboard tool stood in clear contrast to the facilitators for the PoC tool. Whereas the PoC tool was simple to access by virtue of being a "pop-up," the dashboard required clicking away from the patient file. The visual design of the PoC tool and wording of the recommendations were designed with simplicity and conciseness as a priority. By contrast, the dashboard tool has multiple aspects of functionality, allowing for extensive customization

(Multimedia Appendix 1). The few clinicians who reached the tool described being overwhelmed or not having enough time to learn how to use the tool.

The nature of Australian general practice funding tends to incentivize opportunistic rather than systematic preventative health activities [17]. Medicare uses a fee-for-service model, whereby patients receive a rebate for attendance either in person or via telehealth/telephone. Alternative payment mechanisms have been proposed that allow for more of a focus on proactive, coordinated care that includes funding non–face-to-face care, such as that required to interact with tools like the FHT dashboard in any meaningful way [17].

Limitations

All clinicians we interviewed had used FHT in their workplace; by nature of agreeing to be interviewed, they were able to describe the enablers that led them to progress along the CP-FIT feedback pathway. Respondent bias is a limitation of our study, as those who feel most strongly are most likely to speak with us. We also acknowledge that the sample size may not encompass the experiences within the diverse general practice context. However, given the impact of the COVID-19 pandemic on primary care, we are grateful to the 13 participants who took the time to speak with us. The barriers remain unknown for clinicians who did not use FHT, although they can be hypothesized, as discussed below.

While this study focused on the implementation of FHT from the clinical perspective, there are of course other important considerations. In particular, other studies have identified concerns related to the financial costs of staff time and the software itself [18].

Comparison With Prior Work

Many electronic quality improvement programs in current use in Australian general practice have not been evaluated in a research context, and as such, there is a lack of data exploring implementation factors for these programs. The Australian Government Department of Health has recognized the need for clearer policy around the governance of general practice data and the development of clinical decision support tools. In particular, concerns have been raised about the variable quality of existing clinical decision support tools and lack of implementation considerations, which may create general distrust in electronic clinical decision support by GPs [19].

A commonly reported barrier to the use of clinical decision support systems is that of “too much information,” either in the content of a prompt or in the total volume of prompts [20,21]. This notification fatigue is an important consideration for any clinical decision support tool that activates automatically. There is a delicate balance between “active delivery” to facilitate easy interaction with a tool, and unobtrusiveness to the extent that clinicians are not aware of a tool, which is what occurred with the FHT dashboard. Clinicians in our study praised the physical size and placement of the PoC pop-up on the screen, as well as the simplicity of wording of the recommendations. Only 1 GP brought up concerns about notification fatigue, although did not report it as a barrier for their own use of the PoC tool. As more FHT modules are developed in the future, it will be vital to account for potential alert fatigue in how the pop-ups are deployed and displayed.

Lack of time has similarly been cited as a widespread barrier to clinical care for both GPs and nurses [9,18]; however, clinicians in our study did not describe time as a barrier to the use of the PoC tool. This is likely due to the speed at which they interacted with the tool, which itself was due to the usability of the tool and knowledge of the GPs. The marginal time required to interact with the tool, however, will inevitably become more significant as more FHT modules are added. It will be important for the sustainability of the FHT program to take this into account, to reduce the risk that lack of time becomes a barrier to the use of the entire PoC tool.

A study in 2015 in the United States reported barriers to the implementation of a clinical decision support tool specific to CKD, which included limited knowledge of CKD guidelines, lack of interest or confidence in technology, concerns about patient engagement, and time and competing demands [22]. In our study, clinicians’ knowledge of CKD and interest in technology for quality improvement were strong facilitators for FHT, and they were confident in discussing recommendations with their patients. Lack of knowledge in the identification and management of CKD has been cited as a general barrier to care [9]. Overall, it appears that our group of clinicians were particularly knowledgeable and proactive, and this greatly facilitated their engagement with FHT. The PoC tool had been designed with usability and clarity in mind, but these health professional characteristics were vital for uptake.

Acknowledgments

We would like to thank the general practices, health care professionals, and consumers who provided input to the development of Future Health Today (FHT), piloted the tool in their practices, and participated in the trial and the interviews presented in this study. We would also like to acknowledge Adrian Laughlin, Patty Chondros, Shilpanjali Jesudason, Jan Radford, Megan Prictor, Paul Amores, and An Tran-Duy for their role in the chronic kidney disease arm of the FHT trial. The Paul Ramsay Foundation has provided funding for staff costs related to the technical development of FHT and the Implementation and Evaluation team members. They have also provided funding for professional staff support and general practice and consumer advisory groups. The Melbourne Academic Centre for Health provided funding for the development of FHT, and the FHT trial is affiliated with the CanTest Collaborative (funded by Cancer Research UK C8640/A23385). CM is supported by The Royal Australian College of General Practitioners with funding from the Australian Government under the Australian General Practice Training Program. The design, management, analysis, and reporting of the study are entirely independent of the funders.

Conflicts of Interest

Future Health Today (FHT) software was developed by The University of Melbourne in collaboration with Western Health. Intellectual property related to FHT is owned by The University of Melbourne.

Multimedia Appendix 1
Screenshots of Future Health Today software.
[DOCX File, 21658 KB - humanfactors_v11i1e55667_app1.docx]

Multimedia Appendix 2
Interview guide.
[DOCX File, 36 KB - humanfactors_v11i1e55667_app2.docx]

References

1. Australian health survey: biomedical results for chronic diseases, 2011-12 financial year. Australian Bureau of Statistics. 2013. URL: https://www.abs.gov.au/statistics/health/health-conditions-and-risks/australian-health-survey-biomedical-results-chronic-diseases/latest-release [accessed 2023-12-20]
2. Tonelli M, Muntner P, Lloyd A, et al. Risk of coronary events in people with chronic kidney disease compared with those with diabetes: a population-level cohort study. Lancet 2012 Sep 1;380(9844):807-814. [doi: 10.1016/S0140-6736(12)60572-8] [Medline: 22717317]
3. Deaths in Australia. Australian Institute of Health and Welfare. 2023. URL: https://www.aihw.gov.au/reports/life-expectancy-deaths/deaths-in-australia/contents/leading-causes-of-death [accessed 2023-12-20]
4. Blood Pressure Lowering Treatment Trialists' Collaboration, Ninomiya T, Perkovic V, et al. Blood pressure lowering and major cardiovascular events in people with and without chronic kidney disease: meta-analysis of randomised controlled trials. BMJ 2013 Oct 3;347:f5680. [doi: 10.1136/bmj.f5680] [Medline: 24092942]
5. Chronic Kidney Disease (CKD) Management in Primary Care: 4th Edition. Kidney Health Australia. 2020. URL: https://kidney.org.au/health-professionals/ckd-management-handbook [accessed 2023-12-20]
6. Kitsos A, Peterson GM, Jose MD, Khanam MA, Castelino RL, Radford JC. Variation in documenting diagnosable chronic kidney disease in general medical practice: implications for quality improvement and research. J Prim Care Community Health 2019;10:2150132719833298. [doi: 10.1177/2150132719833298] [Medline: 30879383]
7. Jones JL, Lumsden NG, Simons K, et al. Using electronic medical record data to assess chronic kidney disease, type 2 diabetes and cardiovascular disease testing, recognition and management as documented in Australian general practice: a cross-sectional analysis. Fam Med Community Health 2022 Feb;10(1):e001006. [doi: 10.1136/fmch-2021-001006] [Medline: 35177470]
8. Correa VC, Lugo-Agudelo LH, Aguirre-Acevedo DC, et al. Individual, health system, and contextual barriers and facilitators for the implementation of clinical practice guidelines: a systematic metareview. Health Res Policy Syst 2020 Jun 29;18(1):74. [doi: 10.1186/s12961-020-00588-8] [Medline: 32600417]
9. Neale EP, Middleton J, Lambert K. Barriers and enablers to detection and management of chronic kidney disease in primary healthcare: a systematic review. BMC Nephrol 2020 Mar 12;21(1):83. [doi: 10.1186/s12882-020-01731-x] [Medline: 32160886]
10. Hunter B, Biezen R, Alexander K, et al. Future Health Today: codesign of an electronic chronic disease quality improvement tool for use in general practice using a service design approach. BMJ Open 2020 Dec 18;10(12):e040228. [doi: 10.1136/bmjopen-2020-040228] [Medline: 33371024]
11. Chima S, Martinez-Gutierrez J, Hunter B, Manski-Nankervis JA, Emery J. Optimization of a quality improvement tool for cancer diagnosis in primary care: qualitative study. JMIR Form Res 2022 Aug 4;6(8):e39277. [doi: 10.2196/39277] [Medline: 35925656]
12. Manski-Nankervis JA, Hunter B, Hallinan CM, et al. Future Health Today: protocol for a pragmatic stratified cluster randomised head-to-head trial of quality improvement activities in general practice compared to active control. : University of Melbourne; 2022. [doi: 10.26188/21342891.v3]
13. General practice workforce providing primary care services in Australia. Australian Government Department of Health and Aged Care. 2023. URL: https://hwd.health.gov.au/resources/data/gp-primarycare.html [accessed 2023-12-20]
14. Brown B, Gude WT, Blakeman T, et al. Clinical Performance Feedback Intervention Theory (CP-FIT): a new theory for designing, implementing, and evaluating feedback in health care based on a systematic review and meta-synthesis of qualitative research. Implement Sci 2019 Apr 26;14(1):40. [doi: 10.1186/s13012-019-0883-5] [Medline: 31027495]
15. Attribution 4.0 international (CC BY 4.0). Creative Commons. URL: https://creativecommons.org/licenses/by/4.0/ [accessed 2024-07-26]
16. Ford E, Edelman N, Somers L, et al. Barriers and facilitators to the adoption of electronic clinical decision support systems: a qualitative interview study with UK general practitioners. BMC Med Inform Decis Mak 2021 Jun 21;21(1):193. [doi: 10.1186/s12911-021-01557-z] [Medline: 34154580]

17. Primary Health Care Advisory Group. Better outcomes for people with chronic and complex health conditions. : Department of Health, Commonwealth of Australia; 2016 URL: https://apo.org.au/node/62407 [accessed 2024-10-09]
18. Sinclair PM, Day J, Levett-Jones T, Kable A. Barriers and facilitators to opportunistic chronic kidney disease screening by general practice nurses. Nephrology (Carlton) 2017 Oct;22(10):776-782. [doi: 10.1111/nep.12856] [Medline: 27436704]
19. General practice data and electronic clinical decision support: issues paper. Australian Government Department of Health and Aged Care. 2021. URL: https://consultations.health.gov.au/primary-health-network/general-practice-data-and-electronic-clinical-deci/ [accessed 2023-12-20]
20. Fletcher E, Burns A, Wiering B, et al. Workload and workflow implications associated with the use of electronic clinical decision support tools used by health professionals in general practice: a scoping review. BMC Prim Care 2023 Jan 20;24(1):23. [doi: 10.1186/s12875-023-01973-2] [Medline: 36670354]
21. Chen W, O'Bryan CM, Gorham G, et al. Barriers and enablers to implementing and using clinical decision support systems for chronic diseases: a qualitative systematic review and meta-aggregation. Impl Sci Commun 2022 Jul 28;3(1):81. [doi: 10.1186/s43058-022-00326-x] [Medline: 35902894]
22. Vest BM, York TRM, Sand J, Fox CH, Kahn LS. Chronic kidney disease guideline implementation in primary care: a qualitative report from the TRANSLATE CKD study. J Am Board Fam Med 2015;28(5):624-631. [doi: 10.3122/jabfm.2015.05.150070] [Medline: 26355134]

Abbreviations

ACEi: angiotensin-converting enzyme inhibitor
ARB: angiotensin-receptor blocker
CKD: chronic kidney disease
CP-FIT: Clinical Performance Feedback Intervention Theory
FHT: Future Health Today
GP: general practitioner
PN: practice nurse
PoC: point of care
VicREN: Victorian Primary Care Practice-Based Research and Education Network

Edited by E Borycki; submitted 21.12.23; peer-reviewed by CL Chan, D Singh; revised version received 23.05.24; accepted 02.06.24; published 13.11.24.

Please cite as:
McBride C, Hunter B, Lumsden N, Somasundaram K, McMorrow R, Boyle D, Emery J, Nelson C, Manski-Nankervis JA
Clinical Acceptability of a Quality Improvement Program for Reducing Cardiovascular Disease Risk in People With Chronic Kidney Disease in Australian General Practice: Qualitative Study
JMIR Hum Factors 2024;11:e55667
URL: https://humanfactors.jmir.org/2024/1/e55667
doi:10.2196/55667

Design and Implementation of an Opioid Scorecard for Hospital System–Wide Peer Comparison of Opioid Prescribing Habits: Observational Study

Benjamin Heritier Slovis[1,2], MD, MA; Soonyip Huang[3], PharmD; Melanie McArthur[3], RN, MSN; Cara Martino[3,4], RN, DNP; Tasia Beers[3], RN, BSN, MBA; Meghan Labella[3], RN, BSN; Jeffrey M Riggio[1,5], MD; Edmund deAzevedo Pribitkin[6,7], MD, MBA

[1]Office of Clinical Informatics, Thomas Jefferson University, 111 S 11th St, Philadelphia, PA, United States

[2]Department of Emergency Medicine, Jefferson Health Northeast, Thomas Jefferson University, Philadelphia, PA, United States

[3]Enterprise Analytics, Thomas Jefferson University, Philadelphia, PA, United States

[4]Office of Quality and Patient Safety, Thomas Jefferson University, Philadelphia, PA, United States

[5]Department of Medicine, Thomas Jefferson University, Philadelphia, PA, United States

[6]Office of the Chief Physician Executive, Thomas Jefferson University, Philadelphia, PA, United States

[7]Department of Otolaryngology, Thomas Jefferson University, Philadelphia, PA, United States

Corresponding Author:
Benjamin Heritier Slovis, MD, MA
Office of Clinical Informatics, Thomas Jefferson University, , 111 S 11th St, Philadelphia, PA, , United States

Abstract

Background: Reductions in opioid prescribing by health care providers can lead to a decreased risk of opioid dependence in patients. Peer comparison has been demonstrated to impact providers' prescribing habits, though its effect on opioid prescribing has predominantly been studied in the emergency department setting.

Objective: The purpose of this study is to describe the development of an enterprise-wide opioid scorecard, the architecture of its implementation, and plans for future research on its effects.

Methods: Using data generated by the author's enterprise vendor–based electronic health record, the enterprise analytics software, and expertise from a dedicated group of informaticists, physicians, and analysts, the authors developed an opioid scorecard that was released on a quarterly basis via email to all opioid prescribers at our institution. These scorecards compare providers' opioid prescribing habits on the basis of established metrics to those of their peers within their specialty throughout the enterprise.

Results: At the time of this study's completion, 2034 providers have received at least 1 scorecard over a 5-quarter period ending in September 2021. Poisson regression demonstrated a 1.6% quarterly reduction in opioid prescribing, and chi-square analysis demonstrated pre-post reductions in the proportion of prescriptions longer than 5 days' duration and a morphine equivalent daily dose of >50.

Conclusions: To our knowledge, this is the first peer comparison effort with high-quality evidence-based metrics of this scale published in the literature. By sharing this process for designing the metrics and the process of distribution, the authors hope to influence other health systems to attempt to curb the opioid pandemic through peer comparison. Future research examining the effects of this intervention could demonstrate significant reductions in opioid prescribing, thus potentially reducing the progression of individual patients to opioid use disorder and the associated increased risk of morbidity and mortality.

(JMIR Hum Factors 2024;11:e44662) doi:10.2196/44662

KEYWORDS

opioids; peer comparison; quality; scorecard; prescribing; design; implementation; opioid; morbidity; mortality; opioid usage; opioid dependence; drug habits

Introduction

The United States is in an opioid epidemic originally partially fueled by legitimate but inappropriate rates of opioid prescriptions [1-4], with additional waves associated with nonprescription opioids [5]. Reducing inappropriate opioid prescribing by health care providers is a key element of mitigating the risk of long-term opioid use and potential dependence for patients [6-10]. A well-established way to combat the opioid epidemic is to reduce the number of tablets and duration of opioid therapy prescribed by health care providers [9,10]. Many guidelines have been created for this purpose; yet, prescribers routinely do not adhere to them [11,12]. Peer comparison holds the potential to improve care [13]. Research on other prescribing habits has demonstrated that when individuals identified as outliers are presented with information about how they defer from their peers, they tend to appropriately alter their prescribing habits [14,15].

The Health Information Technology for Economic and Clinical Health (HITECH) act of 2009 has made electronic health records (EHRs) ubiquitous. By 2017, 95% of US hospitals have been using EHRs [16]. EHRs and the technologies associated with them have been successfully used to combat the opioid epidemic. Electronic prescribing of controlled substances (EPCS) can improve medication safety, and a 2017 study reported that an increasing number of prescribers are prescribing electronically [17]. Prescription drug monitoring programs, which require prescribers to review prior controlled substance prescriptions prior to prescribing, have shown reductions in opioid prescribing rates [18]. Passive clinical decision support at the time of order entry has also demonstrated alterations in opioid prescribing habits [19-24], including that at Thomas Jefferson University [23,25,26].

One advantage of EHR technology is that reliable and accessible data can be used for the generation of informatics-based reports and analytics. Despite the aforementioned efforts to curb the epidemic, there remains a need for health care providers to review prescribing habits and take the initiative to improve their practice [27]. Hayes and Mycnk [27] state the following: “Better understanding of our own behavior will impact those same behaviors. Mindful practice will lead to more deliberate practice and, hopefully, improved patient care.” Though peer comparison has demonstrated positive effects with prescribing of antibiotics [14,15], research on peer comparison of opioid prescribing has been limited to the emergency department (ED) setting [28-30] and urologists [31] and has been proposed for oral maxillofacial surgeons [32]. These interventions predominantly focused on the number of prescriptions ordered and do not include other metrics such as morphine equivalent daily dose (MEDD) and how these compare to established prescribing guidelines.

In this paper, we describe the process of designing and implementing an “opioid scorecard” that demonstrates key metrics in opioid prescribing (ie, number of tablets, number of prescriptions, MEDD, and calculated day supply). These metrics include a focus on established guidelines for MEDD thresholds, previously not described in other peer comparison interventions. The thresholds for these metrics are determined at the department level, thus comparing prescribers of similar clinical backgrounds. These scorecards are released electronically on a quarterly basis and present prescribing information at the provider level, and they are distributed to individual prescribers, their department chairs, service-line leaders, and chief medical officers. Here we describe the development of the scorecard, the architecture of its implementation, and plans for future research on its effects.

Methods

Setting

This implementation occurred at a single hospital system: Jefferson Health (an operating division of Thomas Jefferson University). Jefferson Health includes 18 hospitals and over 50 outpatient locations across the greater Philadelphia metropolitan area. Philadelphia is a multicultural and diverse city with large Black and Hispanic populations [33]. Unfortunately, our system is partially located in the county that includes the highest estimated frequency of overdose deaths in the state [34] and contains the neighborhood cited as the source of much of the illicit trade on the East coast [35].

Jefferson Health services over 170,000 inpatient and 6.2 million outpatient visits. The system encompasses 3 divisions: the Center City Division (Thomas Jefferson University Hospital, Jefferson Hospital for Neuroscience and Methodist Hospital), Jefferson New Jersey (Washington Township, Cherry Hill and Stratford Hospitals), and the Northern Division (Abington, Lansdale, Torresdale, Bucks, and Frankford Hospitals). Within the enterprise hospital system, there are 5231 licensed prescribers, ordering over 130,000 opioid prescriptions per year. Our medical leadership noticed variability in opioid prescribing practices across the organization, prompting a need for a novel way to provide asynchronous feedback to reduce practice variation and the impetus for development of the scorecard.

Ethical Considerations

This research was approved by the institutional review board of Thomas Jefferson University (IRB# 21E.083).

Participants

Participants included any individual health care provider who had ordered or authorized an ambulatory prescription for an opioid medication at Jefferson Health in the last 12 months prior to the study period, regardless of division or provider specialty. Health care providers who did not order an opioid prescription in the prior 12 months were excluded.

Defining Metrics

The author’s development group consisted of physicians, clinical informaticists, nurses, and analytics experts who are passionate about combating the opioid epidemic. Metrics were first established on the basis of prescribing guidelines developed by the Opioid Task Force (hereinafter referred to as “the task force”) for the institution, a governing body that is responsible for vetting policies and procedures for all health care providers within the health care system [36]. The task force, established in the fall of 2017, is a multidisciplinary group composed of over 40 physicians, pharmacists, nursing, and trainees who volunteer their time monthly to promote safe prescribing of opioids and promote the evidence-based management of opioid use disorder (OUD). The task force’s thresholds were influenced by recommendations from the Centers for Disease Control and Prevention (CDC) [9,10]. With this insight and guidance, the following metrics were established: (1) total number of opioid prescriptions prescribed, (2) total number of patients to whom opioids were prescribed, (3) total number of prescriptions ordered for nonchronic patients with OUD greater than a 5-day supply, (4) total number of prescriptions ordered for nonchronic patients with OUD with an MEDD greater than 50, and (5) total number of prescriptions ordered for chronic patients with OUD with an MEDD greater than 90.

Defining the Cohort

To generate the cohort of prescribers and prescriptions, the authors first analyzed EHR data stored in the enterprise clinical data warehouse. The analysis was managed through third-party analytics software (Qlik). The authors generated a data set at the granularity of the prescription level based on various pharmaceutical classes of opioid analgesics, opioid antitussives, and their combinations for all prescriptions ordered or filled (or both) at the institution. Variables included prescriber name and unique identifier codes, the authorizing provider (if one existed for prescriptions ordered by advanced practice providers or house staff) and EHR security identifiers, patient name and identifier codes, and characteristics of the prescriptions including medication name and formulation, dosage, quantity dispensed, duration of treatment, and number of refills. The data set was stored server-side for further analysis. Through previously defined methods, the authors also generated the calculated duration of based on the number of tablets and dosage instructions for each prescription [25]. Our organization is divided into 3 divisions. Prescribers were associated with their primary division and clinical department as documented in the EHR. Providers were included in the analysis if they had ordered an opioid prescription in the last 12 months. This analysis automatically occurs quarterly on a 3-month period. The system architecture is described below.

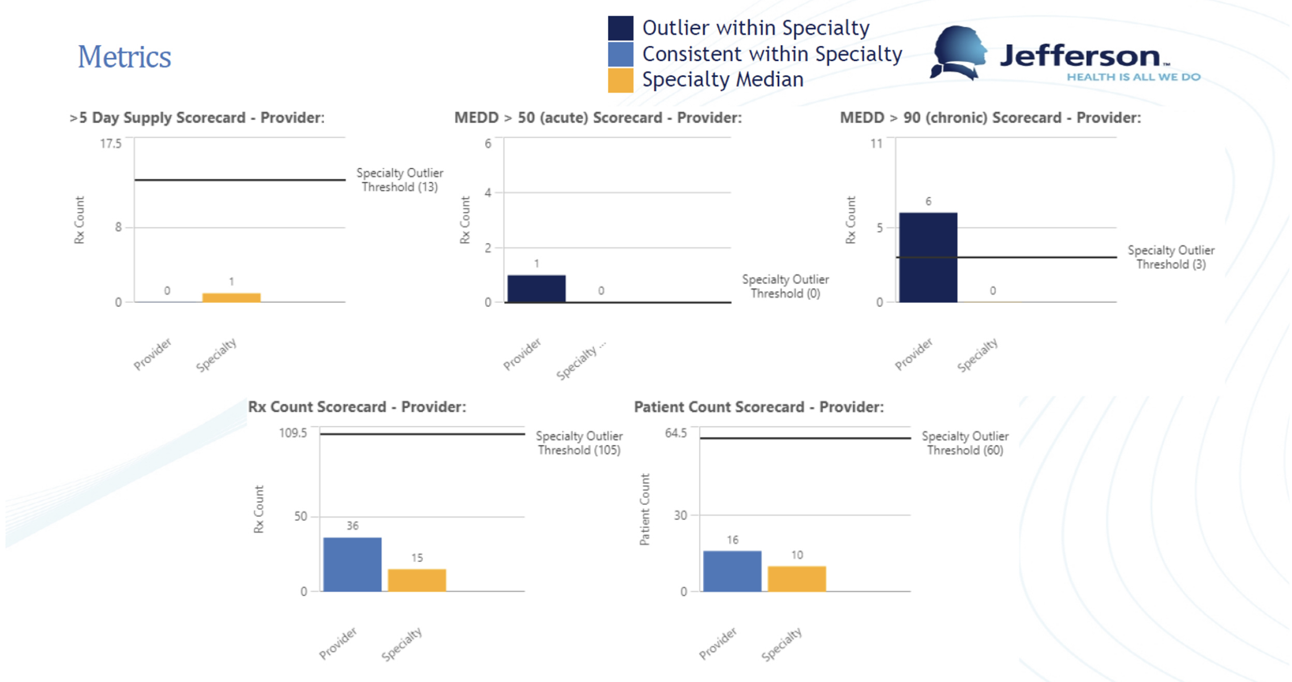
Calculating Thresholds

For each specialty within each division, the authors calculated the median number of opioid prescriptions ordered, the median number of patients to whom prescriptions were ordered for, and the median number of prescriptions for nonchronic patients with OUD greater than 5 days. The authors also calculated the median number of nonchronic patients with OUD per provider who had prescriptions greater than 50 MEDD and the number of chronic patients per provider for prescriptions greater than 90 MEDD. Chronic opioid use was defined as ≥3 opioid prescriptions in the last year with an active opioid prescription, which automatically includes the patient in a chronic-opioid use registry. Using the distributions of these same metrics, the authors calculated the IQRs for each specialty per division. The authors defined outlier thresholds for each specialty within the divisions as 1.5 times the IQR for the previously described metrics, and extreme outliers as 3 times the outlier threshold.

Generating Individual Provider Reports

For individual providers, the authors calculated their values for the metrics described above. These values are compared to the calculated thresholds for outliers for their specialty within their division. If the provider is above the specialty threshold for a given metric, they are labeled an outlier. Provider-level scorecards present each of the 5 metrics as bar charts, with the value in question in the y-axis and separate bars for the provider and specialty median on the x-axis. Additionally, a line is drawn parallel to the x-axis to denote the outlier threshold. Graphical representations are color-coded: navy blue when an outlier and teal blue when consistent with their specialty for metrics with thresholds defined by the task force (greater than a 5-day supply, greater than 50 MEDD for nonchronic patients, and greater than 90 MEDD for chronic patients). The specialty medians are represented in yellow. Each scorecard contains a title page, a list of definitions, and the page containing the graphical representations of the metrics. Figure 1 demonstrates a provider score card.

Figure 1. Example of a provider-level opioid scorecard representing the 5 opioid prescribing metrics, specialty metrics for comparison, and graphical representations of thresholds for that specialty. MEDD: morphine equivalent daily dose.



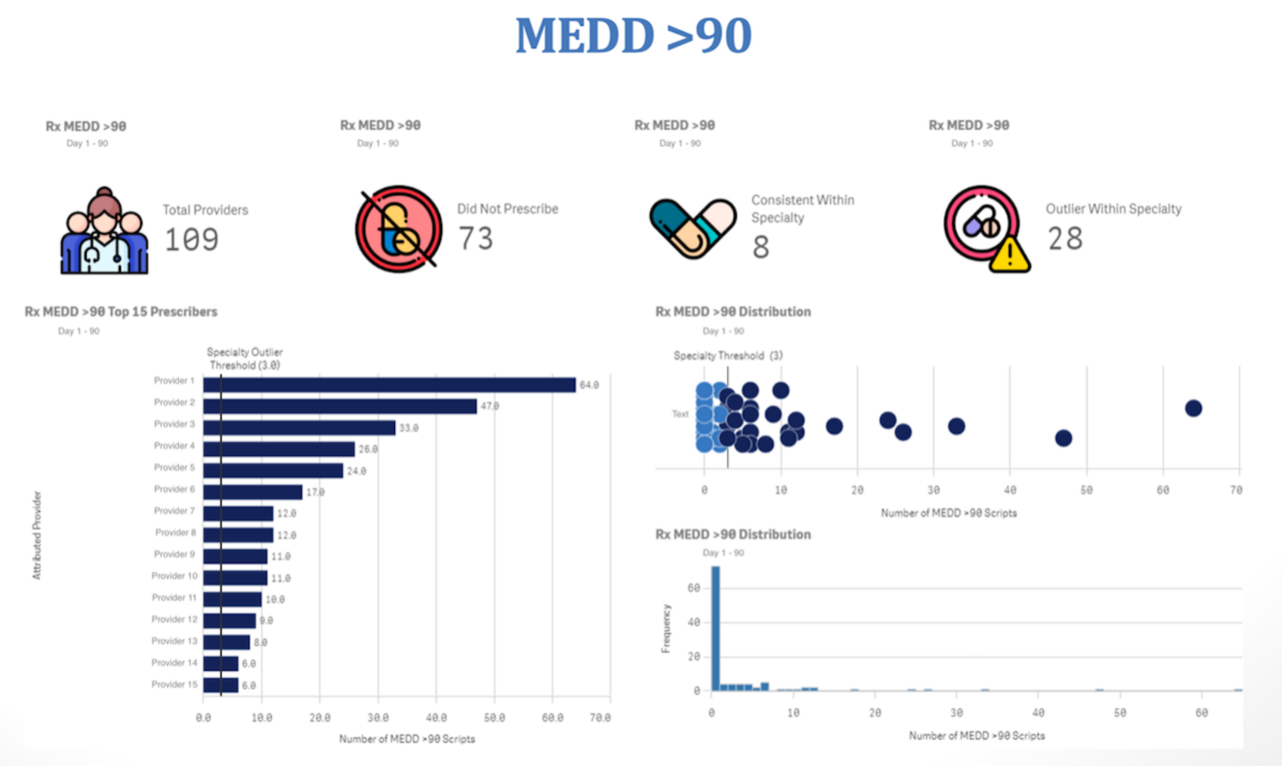
Generating Leadership Scorecards

While it is important for individual prescribers to understand their prescribing habits and perhaps reflect on and potentially modify them, it is also important for individuals in leadership positions to understand the trends in prescribing within their department. To this end, the authors created a scorecard that summarizes the prescribing habits of a department in an easy-to-interpret way with graphical representation for service-line leaders and department chairs. This allows them to discuss prescribing habits with individual prescribers and understand trends within their specialty.

A list of departmental leaders including faculty chairs and service-line leaders was compiled for the enterprise. A crosswalk was generated for each specialty to their service-line leader or chair (or both), with some leaders representing more than 1 specialty.

The same metrics described in the individual provider scorecard are represented for specialty leadership with a graphical representation of the providers in their specialty. Each scorecard includes a cover letter and definitions sheet. Each metric gets a dedicated page including descriptive statistics of the number of providers included according to the criteria, the number of those who did not prescribe any opioids in that quarter, the number of those who prescribed opioids but were within the specialty threshold, and the number of individuals prescribing outside the threshold. Leaders also receive a bar chart with the name of each of the top 15 prescribers with a line on the x-axis for the specialty threshold and color-coding indicating whether each provider is within or outside the threshold. They also receive a scatter plot of all providers color-coded as under or over the threshold to better understand the distribution within their department, and a histogram of the total number of prescriptions used for the calculation of that metric. Figure 2 demonstrates a leadership scorecard example of the greater than 5-day duration metric.

Figure 2. Example of a specialty leadership scorecard with described metrics and a graphical representation of top 15 prescribers by name, a color-coded scatter plot of prescribers above (dark blue) and below (light blue) specialty threshold, and a histogram of prescription frequencies.



Generating Chief Medical Officer Scorecards

Given that the authors' enterprise organizational structure has 4 separate divisions, the authors wanted to create a report that allowed division chief medical officers (CMOs) and the enterprise CMO the opportunity to comprehend and potentially interview those individuals far outside the threshold of their department. While some individuals (such as pain management or hospice care) may be appropriately prescribing outside of what a typical physician would prescribe, others may be unaware of their habits and require intervention. To this end, the authors created an automated report of individual prescribers and their metrics, which includes the previously described extreme outlier threshold set at 3 times the specialty threshold. The report is generated as an Excel (Microsoft Corp) file and distributed via email. This report is automatically shared on a quarterly basis, one for each divisional CMO representing their division and a combined summary for the enterprise CMO. An example of the CMO score card is presented in Figure 3.

Figure 3. Example of the chief medical officer scorecard representing top outliers for the greater than 5 days' supply metric. MEDD: morphine equivalent daily dose.

Top Prescribers for Prescriptions for >5 Days

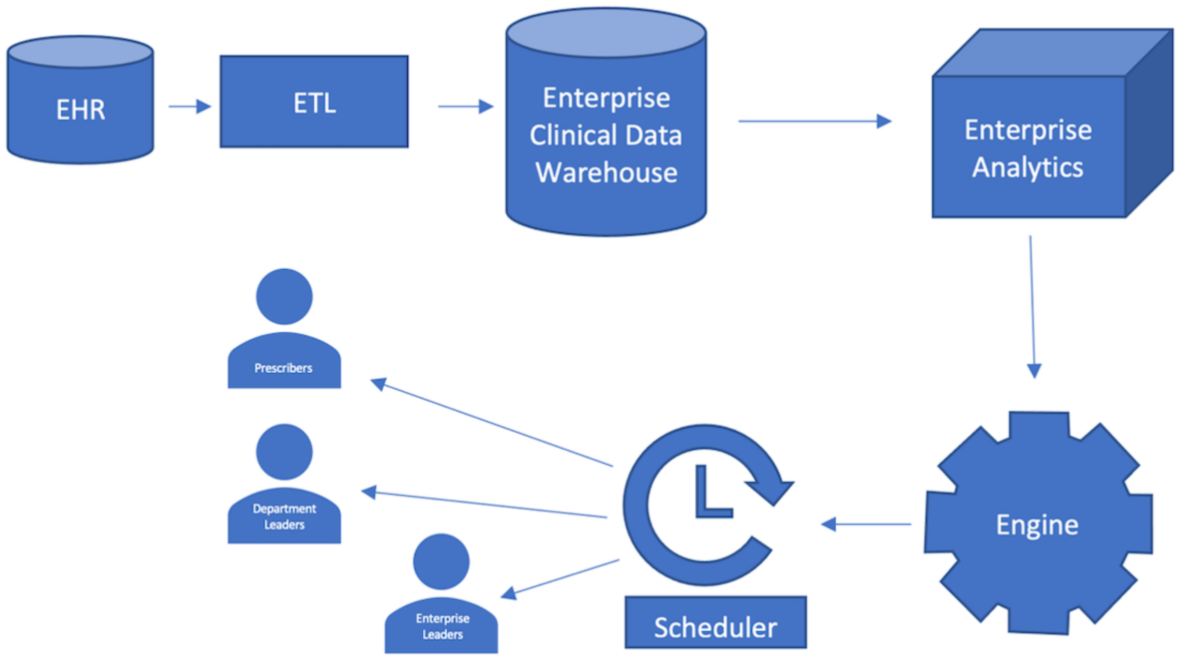
Division	Attributed Provider Specialty	Specialty Threshold for Opioid Rx >5 Days Duration (any value over this is a statistical outlier)	Attributed Provider	Attributed Provider Count of Opioid Rx >5 Day Duration in Last Quarter	% of Opioid Rx with >5 Day Duration out of Opioid Rx by Authorizing Provider	Factor over Specialty Threshold
Division 1	Specialty 1	2.5	Provider 1	152	92.1%	60.8
Division 1	Specialty 2	8.75	Provider 2	44	38.3%	5
Division 1	Specialty 2	8.75	Provider 3	28	25.0%	3.2
Division 1	Specialty 3	5	Provider 4	25	43.9%	5
Division 1	Specialty 4	6.25	Provider 5	257	81.8%	41.1
Division 1	Specialty 4	6.25	Provider 6	136	88.9%	21.8
Division 1	Specialty 4	6.25	Provider 7	106	74.1%	17
Division 1	Specialty 4	6.25	Provider 8	106	44.7%	17
Division 1	Specialty 5	2.5	Provider 9	14	73.7%	5.6

Cover Sheet | Rx >5 Days Outliers | MEDD >90 Outliers | MEDD >50 Outliers | Rx Count Outliers | Patient Count Outliers

System Architecture

The enterprise health system has a robust data analytics system using third-party analytics software. Data from the production EHR are extracted into a database that lives in the enterprise clinical data warehouse (CDW). We used third-party analytics software to extract and manipulate data for analysis and visualization. This third-party software uses its own proprietary language similar to SQL, which is coded by dedicated analysts from the organization's analytics team. Through an iterative process, metrics were generated, and values were validated via the CDW query and review by clinical informatics experts. Scorecard metrics were also compared to other enterprise reports on opioids to confirm validity. Tabular and graphical representations of the data are refreshed daily and are accessible to end users via web browser. Quarterly, a report is created by the Enterprise Analytics team, through the third-party analytics software, generated automatically using the established visualizations, and is disseminated as PDF or XLS files to the appropriate recipients via secure enterprise email. Individual prescriber and leadership reports are generated as PDF files attached to individual emails with hyperlinks to the source analytics software to allow easy access by the individual recipient. There is a web-browser version of the reports that is accessible to those who have been granted access, given the sensitive nature of the reports in question. The CMO scorecards are generated as XLS files per the request of the medical leadership who wanted a quick reference report without graphical representation. Figure 4 demonstrates the system architecture.

Figure 4. System architecture for opioid score card based on enterprise analytics workflows. EHR: electronic health record; ETL: Export, Transform, and Load.



Implementation

The authors used a staged approach to the implementation based on access to clinical data in the CDW. Through the development of these scorecards multiple sites came upon the enterprise EHR providing a natural iterative implementation schedule but also a chance to confirm the enterprise systems could handle increasingly larger calculations and distributions of scorecards. We first piloted in the Center City division, then advancing to the New Jersey and finally the Northern division as each site generated a preimplementation 3-month period of prescribing data in the EHR, to allow for the quarterly nature of the calculations. Phased go-lives also allowed for ongoing learning and iterative improvements in these processes. Providers received an email from their divisions CMO notifying them of pending receive of the first scorecard, and each quarterly scorecard contains a cover letter from the respective CMO, but no additional education was provided outside of referring providers to their clinical leaders with additional questions. CMOs also provided information to clinical leaders about the scorecard via email prior to initial release.

Statistical Analysis

Statistical analysis was performed using R statistical software (The R Foundation, 2021) to calculate proportions and confidence intervals. We also calculated reduction in prescription frequencies via Poisson regression analysis. To compare proportions of prescriptions outside CDC and Hospital guidelines, we performed a chi-square analysis of the first scorecard data (based on the quarter prior when no individual had ever received a scorecard) to the subsequent 4 quarters.

Results

Overview

A pilot version of the first scorecard was distributed in October 2019. The authors performed a staged approach to distribution, starting with the Center City Division, then expanding to New Jersey, then finally the Northern Division. Completed distribution was successful in September 2021.

To date, the authors have distributed 5 iterations of the quarterly score card over 15 months.

A total of 170 scorecards have been released to 81 service line leaders and departmental chairs over 68 departments within the 3 divisions.

Five scorecards have been released to the 4 CMOs across 3 divisions and 1 enterprise CMO.

Prescriber Demographics

Upon the most recent distribution, which was the first to include all divisions, 2034 prescribers (38.9% of the eligible 5231 prescribers) received opioid scorecards from 68 specialties. As described above, this indicates that they have prescribed an opioid in the last 12 months, regardless of outlier status. Of these individuals, 1416 (69.7%) were attending physicians, 331 (16.3%) were nurse practitioners, 168 (8.3%) were physician assistants, and 111 (5.5%) were residents with independent Drug Enforcement Administration licenses.

Outlier Demographics

There were 386 (19.0% of 2034 opioid prescribers, 7.31% of the 5231 licensed prescribers) individuals who were considered outliers in their specialty for at least 1 metric. Of them, 188 (48.7%) were outliers for total number of prescriptions, 162 (42.0%) were outliers for the number of patients prescribed opioids, 125 (32.4%) were outliers for prescriptions ordered for nonchronic patients with prescriptions greater than 5 days, 118 (30.6%) were outliers for prescriptions greater than 50 MEDD for unique nonchronic patients and 113 (29.3%) were outliers for prescriptions greater than 90 MEDD for chronic patients.

Table 1 demonstrates cumulative counts and percentages of outliers from the first 5 scorecards released for each of the 5 metrics from the 5 largest clinical specialties in the organization, deidentified to protect participating physicians. Notably, at least 10% of providers were considered outliers in at least 1 metric during the initial 5 scorecard releases.

Table . Cumulative counts and percentages of outliers from the first 5 scorecards released for each of the 5 metrics from the 5 largest clinical specialties in the organization.

Specialties	Providers, n	Rx count outliers, n (%)	Patient count outliers, n (%)	>5 day supply outliers, n (%)	>50 MEDD[a] outliers, n (%)	>90 MEDD outliers, n (%)
1	212	33 (15.57)	35 (16.51)	42 (19.81)	60 (28.30)	6 (2.83)
2	195	26 (12.26)	28 (13.21)	63 (29.72)	41 (19.34)	39 (18.40)
3	176	19 (8.96)	11 (5.19)	26 (12.26)	51 (24.06)	50 (23.58)
4	101	8 (3.77)	10 (4.72)	42 (19.81)	33 (15.57)	11 (5.19)
5	97	9 (4.25)	7 (3.3)	10 (4.72)	18 (8.49)	15 (7.08)

[a]MEDD: morphine equivalent daily dose.

Prescribing Impact

We examined the iterative scorecard releases to determine if there was an effect on opioid prescribing. Given the phased implementation process, the Center City Division was the only division of the organization with a full 5 quarters of scorecard releases, so this division was isolated for analysis. There were 2,147,710 opioid prescriptions ordered by this division during the study period. From the date of initial release, there was a significant reduction in the frequency of opioid prescriptions at the division, with an associated 1.60% reduction per quarter (P<.001). There was a significant reduction in prescriptions for nonchronic patients longer than 5 days' duration before versus after intervention. There was also a significant reduction in the proportion of prescriptions >50 MEDD for nonchronic patients and an increase in that of prescriptions >90 MEDD for chronic patients (Table 2).

Table . Pre- and postintervention proportions, 95% CIs, and *P* values for >5-day durations, >50 MEDD[a] for nonchronic patients with opioid use, and >90 MEDD for chronic patients with opioid use[b].

Study period	>5-days duration		>50 MEDD for nonchronic patients		>90 MEDD for chronic patients	
	Proportion, %	95% CI	Proportion, %	95% CI	Proportion, %	95% CI
Pre	27.31	27.18 - 27.43	13.06	12.96 - 13.16	6.64	6.56 - 6.71
Post	26.76	26.69 - 26.82	12.14	12.04 - 12.24	6.80	6.76 - 6.84

[a]MEDD: morphine equivalent daily dose.

[b]*P*<.001 for all pre- vs postintervention comparisons.

Discussion

Principal Findings

The objective of this study is to describe an evidence-based peer comparison opioid scorecard. By sharing this process of design and distribution, we hope to inspire similar processes at other organizations. Here we demonstrate the successful distribution of scorecards to 19% (386/2034) of eligible opioid prescribers over a 5-quarter time frame. Initial analysis demonstrates that in all 5 of the largest specialties, at least 10% of providers were outliers in at least 1 metric (Table 1). This is most notable in specialties 1 and 3 where over 20% of providers had prescriptions greater than 50 MEDD for nonchronic patients, and over 19% of providers in specialties 1, 2, and 4 had prescriptions over 5 days' supply for nonchronic patients. These results demonstrate the need for intervention and providing information to individual providers on their prescribing practices at the organization.

Over the study period, the authors demonstrated a quarterly reduction in overall opioid prescriptions of 1.6% per quarter. When comparing the data that were used to generate the initial scorecard (the quarter prior to scorecard release) versus the subsequent 4 quarters, there was a small but significant reduction in the proportion of prescriptions longer than 5 days' duration and greater than 50 MEDD for nonchronic patients (0.6% and 0.3%, respectively). While the proportions were small overall, the median quarterly prescribing rate at this division was over 420,000 prescriptions, yielding a quarterly reduction of over 2500 prescriptions over 5 days and 1200 prescriptions over 50 MEDD. There was also a very small but significant increase in chronic patient prescriptions greater than 90 MEDD (0.16%), resulting in approximate 672 additional patients with this level of MEDD per quarter. Given the nature of this study, it is unclear if these are patients who transitioned from nonchronic to chronic dosage changes, or if this change was due to other external factors that influenced prescribing habits.

Since the early 1990s, opioid prescribing has consistently increased [37] to the point that in 2016, there were 66.5 opioid prescriptions for every 100 persons in the United States [38]. There is a link between long-term use of prescription opioids and a transition to illicit heroin [39] with increasing mortality [40]. Higher doses and longer treatment time of prescription opioids have been linked to an increased risk of chronic opioid use [6-8]. Therefore, a principal method of combating the opioid epidemic is to reduce prescribing by health care providers. To this end, the CDC's 2016 guidelines recommend that caution be exercised when increasing doses to greater than 50 MEDD, and doses greater than 90 MEDD should be avoided. The CDC also states that 3 days or less is often sufficient for acute pain, and more than 7 days is "rarely needed" [9,10]. Our organization has accordingly recommended maintaining prescriptions for acute pain at 5 days and less than 50 MEDD, based on these guidelines. The inclusion of these evidence-based metrics in the scorecard justifies its evidence-based nature.

Peer comparison has been demonstrated to be an effective means of modifying unwanted or unintended behavior among health care providers [14,15,28-31]. Andereck et al [28] and Boyle et al [29] both demonstrated reductions in prescribing rates after peer comparison reports were initiated in their EDs. Michael et al [30] performed a randomized controlled trial of ED physicians and demonstrated a relationship between the perceived rate of prescribing and a reduction when presented with peer comparison data. Suffoletto and Landaou [41] demonstrated that peer comparison was an important contributor to improved prescribing habits compared to audit and feedback alone. While limited to the ED, these research studies justify our approach to peer comparison and confirm that prescriber's own perceptions of their prescribing habits are critical to successful intervention.

Research beyond the ED is limited regarding peer comparison of opioid prescribing. Jacobs et al [31] demonstrated that peer comparison, as well as education and audit feedback, contributed to significant reductions in prescribing by urologists. Weiner et al [42] describe a system-wide initiative that successfully reduced opioid-related morbidity and mortality. They include a mention of "benchmarking reports" with peer comparison elements but do not further detail what was included in these reports or how they were delivered.

Despite evidence of its positive impact on prescribing, a report such as the one presented in this paper has never been produced outside of isolated EDs nor has it been implemented as broadly. The intervention described here is sent to all prescribers, regardless of level of training or clinical specialty, providing them with detailed metrics on their opioid prescribing habits and peer comparison based on clinical specialty. Additionally, most of the previously described department-level interventions focus on opioid prescribing frequency without consideration toward prescribing guidelines such as those recommended by the CDC. The only other system-wide effort described in the literature contains limited information on what was included in their reports, how information was portrayed, and how they were delivered, but rather focused more on other elements of their substantial and impressive system-wide intervention.

The methods described in this paper have the potential to change clinical practice paradigms by using a novel intervention implemented at a larger scale than previously described in the literature. While peer comparison has been used previously in this manner (as described above), little is known about how such a tool could be used on an enterprise hospital level, with multiple divisions encompassing multiple practice specialties across 2 states. As longer and higher doses of opioids are associated with long-term use and potential for OUD, educating prescribers on their opioid prescribing habits can potentially reduce the prescription quantity and duration at the prescriber level, thus potentially mitigating the risk of long-term use at the patient level.

The scorecards received by prescribers contain multiple metrics including the number of prescriptions, number of patients, and MEDD thresholds. Outlier thresholds based off IQRs are calculated quarterly. We appreciate that overtime variability in prescribing patterns may reduce, resulting in lower thresholds that are incapable of driving impact. Currently, all scorecards are delivered with instructions on how the metrics are derived, and they are informational and nonpunitive in nature. Providers are instructed that concerns regarding metrics should be discussed with departmental chairs or service line leaders to determine whether intervention is required. Reinforcement of appropriate prescribing guidelines by leaders was not systematically carried out; however, such systematic enforcement holds promise for further reduction in inappropriate prescribing.

The authors intend to continue to prospectively track the metrics associated with each scorecard release. By measuring descriptive statistics for each specialty to determine whether there are changes in that specialty's prescribing habits before and after intervention, as well as through a time series post intervention, the authors hope to demonstrate the impact that the scorecard and enterprise-level peer comparison can have on opioid prescribing beyond the initial results presented here. The authors will also evaluate frequencies with which individual prescribers are identified as outliers and track any changes in metrics of individual high-prescribing providers. The authors are also tracking quarterly thresholds as a measure of improvement in practice variation and to determine whether prolonged periods of stasis imply that no further reduction in practice pattern variability can be achieved.

Given this effective implementation, positive results could allow for broadening of the scope of the study and for massive scalability. The system described is EHR-agnostic given the use of third-party reporting software, meaning that data sources from other hospital systems could be integrated and similar reports can be generated on a regional, state, or even federal level.

Additionally, while this project is intended to address opioid prescribing for which clear practice guidelines have been established by the CDC, the application of automated peer comparison reports and assessment of their effects on practice patterns can be extended to topics including but not limited to antibiotic prescribing, imaging usage, laboratory test ordering, and specialty consultation. Automated peer comparison could have the potential to reduce cost and improve the quality of care when designed around appropriately established clinical guidelines.

Limitations

There are several limitations that should be addressed. Our study was performed at a single hospital system, and it is unclear if our intervention would have similar effects at institutions dissimilar to ours in different geographical regions. Additionally, we cannot account for all confounders that may have been contributed to reductions in opioid prescribing outside of our intervention. Our organization has historically used multiple interventions to curb the opioid epidemic including but not limited to requiring the review of state prescription drug monitoring programs prior to outpatient prescribing and standardized dosing for opioids in our EHR. However, there were no concurrent internal interventions during the time of the implementation, though external influences such as state and regional educational efforts could have impacted prescribing. As mentioned above, we also cannot account for possible transitions of patients from nonchronic to chronic cohorts during the study period, which may have contributed to the small increase in nonchronic prescribing, though it is also unclear if external factors may have contributed to this finding. Additionally, with regard to prescriptions, we only examined trends in opioid prescribing and did not examine whether there were any contrary changes in other nonopioid medications as an unintended consequence of this intervention. Finally, in late 2022, the CDC updated their opioid prescribing guidelines, which have transitioned from the specific acute and chronic thresholds for MEDD to more general recommendations on prescribing the lowest dose of opioids possible to achieve pain control, with detailed information on the risks of doses above 50 MEDD [43]. The intervention detailed here was developed for the 2016 guidelines, and our organization still supports the prescribing thresholds developed from them. Modifications to this intervention will be developed while pending institutional transitions to the new prescribing guidelines.

Conclusions

Modern EHR technology, advanced analytics, and a robust reporting infrastructure allowed for the generation of an enterprise-level opioid prescribing scorecard that uses peer comparison to provide individual providers with feedback on their prescribing habits. The authors also have successfully developed departmental leadership and CMO reports for oversight and advanced intervention regarding potentially problematic prescribers. Initial results over a 5-quarter period imply reductions in overall opioid prescribing rates as well as improvement in the duration and dose for nonchronic patients, though there was a small increase in prescriptions over 90 MEDD. Future research on the impact of such a scorecard could be pivotal in combating the opioid epidemic, potentially scaling such an intervention to larger geographical regions, and broadening the use of such a tool outside the opioid epidemic. Investment in informatics and analytics holds the potential to have profound impacts on the quality of care and patient safety.

Authors' Contributions

BHS drafted the manuscript with significant contributions to the final version from all authors. The intervention was designed by SH, MM, CM, TB, ML, and BHS with supervision by EAP. SH provided the data with initial analysis by SH, MM, and CM. Further analysis and interpretation were provided by all authors.

Conflicts of Interest

None declared.

References

1. Srivastava AB, Gold MS. Beyond supply: how we must tackle the opioid epidemic. Mayo Clin Proc 2018 Mar;93(3):269-272. [doi: 10.1016/j.mayocp.2018.01.018] [Medline: 29502558]
2. The opioid epidemic: by the numbers. US Department of Health and Human Services. 2016. URL: https://www.hhs.gov/sites/default/files/Factsheet-opioids-061516.pdf [accessed 2024-08-13]
3. Rudd RA, Seth P, David F, Scholl L. Increases in drug and opioid-involved overdose deaths - United States, 2010-2015. MMWR Morb Mortal Wkly Rep 2016 Dec 30;65(50-51):1445-1452. [doi: 10.15585/mmwr.mm655051e1] [Medline: 28033313]
4. What is the U.S. opioid epidemic? US Department of Health and Human Services. 2021. URL: https://www.hhs.gov/opioids/about-the-epidemic/index.html [accessed 2024-08-13]
5. Understanding the epidemic. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/overdose-prevention/about/understanding-the-opioid-overdose-epidemic.html [accessed 2024-08-13]
6. Bohnert ASB, Valenstein M, Bair MJ, et al. Association between opioid prescribing patterns and opioid overdose-related deaths. JAMA 2011 Apr 6;305(13):1315-1321. [doi: 10.1001/jama.2011.370] [Medline: 21467284]
7. Edlund MJ, Martin BC, Russo JE, DeVries A, Braden JB, Sullivan MD. The role of opioid prescription in incident opioid abuse and dependence among individuals with chronic noncancer pain: the role of opioid prescription. Clin J Pain 2014 Jul;30(7):557-564. [doi: 10.1097/AJP.0000000000000021] [Medline: 24281273]
8. Shah A, Hayes CJ, Martin BC. Characteristics of initial prescription episodes and likelihood of long-term opioid use - United States, 2006-2015. MMWR Morb Mortal Wkly Rep 2017 Mar 17;66(10):265-269. [doi: 10.15585/mmwr.mm6610a1] [Medline: 28301454]
9. Dowell D, Haegerich TM, Chou R. CDC guideline for prescribing opioids for chronic pain--United States, 2016. JAMA 2016 Apr 19;315(15):1624-1645. [doi: 10.1001/jama.2016.1464] [Medline: 26977696]
10. Dowell D, Haegerich TM, Chou R. CDC guideline for prescribing opioids for chronic pain - United States, 2016. MMWR Recomm Rep 2016 Mar 18;65(1):1-49. [doi: 10.15585/mmwr.rr6501e1] [Medline: 26987082]
11. Ballantyne JC, Murinova N, Krashin DL. Opioid guidelines are a necessary response to the opioid crisis. Clin Pharmacol Ther 2018 Jun;103(6):946-949. [doi: 10.1002/cpt.1063] [Medline: 29603187]
12. Ivanics T, Nasser H, Kandagatla P, et al. Prescribing habits of providers and risk factors for nonadherence to opioid prescribing guidelines. Am Surg 2021 Jul;87(7):1039-1047. [doi: 10.1177/0003134820956332] [Medline: 33295200]
13. Navathe AS, Emanuel EJ. Physician peer comparisons as a nonfinancial strategy to improve the value of care. JAMA 2016 Nov 1;316(17):1759-1760. [doi: 10.1001/jama.2016.13739] [Medline: 27802553]
14. Meeker D, Linder JA, Fox CR, et al. Effect of behavioral interventions on inappropriate antibiotic prescribing among primary care practices: a randomized clinical trial. JAMA 2016 Feb 9;315(6):562-570. [doi: 10.1001/jama.2016.0275] [Medline: 26864410]
15. Shively NR, Buehrle DJ, Wagener MM, Clancy CJ, Decker BK. Improved antibiotic prescribing within a veterans affairs primary care system through a multifaceted intervention centered on peer comparison of overall antibiotic prescribing rates. Antimicrob Agents Chemother 2019 Dec 20;64(1):e00928-19. [doi: 10.1128/AAC.00928-19] [Medline: 31685466]
16. Percent of hospitals, by type, that possess certified health IT. HealthIT.gov. 2018. URL: https://dashboard.healthit.gov/quickstats/pages/certified-electronichealth-record-technology-in-hospitals.php [accessed 2024-08-13]
17. Parasrampuria S, Blanco M, Barker W. Electronic prescribing of controlled substances among office-based physicians, 2017-2019. : Office of the National Coordinator for Health Information Technology; 2019.
18. Wen H, Hockenberry JM, Jeng PJ, Bao Y. Prescription drug monitoring program mandates: impact on opioid prescribing and related hospital use. Health Aff (Millwood) 2019 Sep;38(9):1550-1556. [doi: 10.1377/hlthaff.2019.00103] [Medline: 31479368]
19. Delgado MK, Shofer FS, Patel MS, et al. Association between electronic medical record implementation of default opioid prescription quantities and prescribing behavior in two emergency departments. J Gen Intern Med 2018 Apr;33(4):409-411. [doi: 10.1007/s11606-017-4286-5] [Medline: 29340937]

20. Zwank MD, Kennedy SM, Stuck LH, Gordon BD. Removing default dispense quantity from opioid prescriptions in the electronic medical record. Am J Emerg Med 2017 Oct;35(10):1567-1569. [doi: 10.1016/j.ajem.2017.04.002] [Medline: 28416266]
21. Montoy JCC, Coralic Z, Herring AA, Clattenburg EJ, Raven MC. Association of default electronic medical record settings with health care professional patterns of opioid prescribing in emergency departments: a randomized quality improvement study. JAMA Intern Med 2020 Apr 1;180(4):487-493. [doi: 10.1001/jamainternmed.2019.6544] [Medline: 31961377]
22. Santistevan JR, Sharp BR, Hamedani AG, Fruhan S, Lee AW, Patterson BW. By default: the effect of prepopulated prescription quantities on opioid prescribing in the emergency department. West J Emerg Med 2018 Mar;19(2):392-397. [doi: 10.5811/westjem.2017.10.33798] [Medline: 29560071]
23. Slovis BH, London KS, Randolph FT, et al. 336 the effect of implementing electronic health record default prescribing preferences on opioid prescriptions written in the emergency department. Ann Emerg Med 2018 Oct;72(4):S133. [doi: 10.1016/j.annemergmed.2018.08.341]
24. Chiu AS, Jean RA, Hoag JR, Freedman-Weiss M, Healy JM, Pei KY. Association of lowering default pill counts in electronic medical record systems with postoperative opioid prescribing. JAMA Surg 2018 Nov 1;153(11):1012-1019. [doi: 10.1001/jamasurg.2018.2083] [Medline: 30027289]
25. Slovis BH, Kairys J, Babula B, et al. Discrepancies in written versus calculated durations in opioid prescriptions: pre-post study. JMIR Med Inform 2020 Mar 31;8(3):e16199. [doi: 10.2196/16199] [Medline: 32229472]
26. Slovis BH, Riggio JM, Girondo M, et al. Reduction in hospital system opioid prescribing for acute pain through default prescription preference settings: pre-post study. J Med Internet Res 2021 Apr 14;23(4):e24360. [doi: 10.2196/24360] [Medline: 33851922]
27. Hays GP, Mycyk MB. Confronting the opioid crisis by taking a long look in the mirror … and at our peers. Acad Emerg Med 2018 May;25(5):594-596. [doi: 10.1111/acem.13397] [Medline: 29498140]
28. Andereck JW, Reuter QR, Allen KC, et al. A quality improvement initiative featuring peer-comparison prescribing feedback reduces emergency department opioid prescribing. Jt Comm J Qual Patient Saf 2019 Oct;45(10):669-679. [doi: 10.1016/j.jcjq.2019.07.008] [Medline: 31488343]
29. Boyle KL, Cary C, Dizitzer Y, Novack V, Jagminas L, Smulowitz PB. Reduction of opioid prescribing through the sharing of individual physician opioid prescribing practices. Am J Emerg Med 2019 Jan;37(1):118-122. [doi: 10.1016/j.ajem.2018.09.052] [Medline: 30343961]
30. Michael SS, Babu KM, Androski C, Reznek MA. Effect of a data-driven intervention on opioid prescribing intensity among emergency department providers: a randomized controlled trial. Acad Emerg Med 2018 May;25(5):482-493. [doi: 10.1111/acem.13400] [Medline: 29498155]
31. Jacobs BL, Rogers D, Yabes JG, et al. Large reduction in opioid prescribing by a multipronged behavioral intervention after major urologic surgery. Cancer 2021 Jan 15;127(2):257-265. [doi: 10.1002/cncr.33200] [Medline: 33002197]
32. Wang TT, Delgado MK, Hersh EV, Panchal N. Peer comparisons to increase responsible opioid prescribing among oral and maxillofacial surgeons. J Oral Maxillofac Surg 2021 May;79(5):955-957. [doi: 10.1016/j.joms.2020.12.013] [Medline: 33428865]
33. Quickpacts: Philadelphia cunty, Pennsylvania. United States Census Bureau. 2022. URL: https://www.census.gov/quickfacts/philadelphiacountypennsylvania2022 [accessed 2024-08-13]
34. Opioid data dashboard. Commonwealth of Pennsylvania. URL: https://data.pa.gov/stories/s/9q45-nckt [accessed 2024-08-13]
35. Percy J. Trapped by the ‘Walmart of Heroin’. The New York Times. 2018. URL: https://www.nytimes.com/2018/10/10/magazine/kensington-heroin-opioid-philadelphia.html [accessed 2024-08-13]
36. Monaghan P. The opioid crisis: confronting america’s other fatal health epidemic. Jefferson Health. 2020. URL: https://thehealthnexus.org/the-opioid-crisis-confronting-americas-other-fatal-health-epidemic [accessed 2024-08-13]
37. Nora D, Volkow MD. America’s addiction to opioids: heroin and presecription drug abuse. National Institute on Drug Abuse. 2014. URL: https://archives.drugabuse.gov/testimonies/2014/americas-addiction-to-opioids-heroin-prescription-drug-abuse [accessed 2024-08-13]
38. Annual surveillance report of drug-related risks and outcomes. Centers for Disease Control and Prevention. 2017. URL: https://www.cdc.gov/overdose-prevention/media/pdfs/pubs/2017-cdc-drug-surveillance-report.pdf [accessed 2024-08-13]
39. Jones CM. Heroin use and heroin use risk behaviors among nonmedical users of prescription opioid pain relievers - United States, 2002-2004 and 2008-2010. Drug Alcohol Depend 2013 Sep 1;132(1-2):95-100. [doi: 10.1016/j.drugalcdep.2013.01.007] [Medline: 23410617]
40. Scholl L, Seth P, Kariisa M, Wilson N, Baldwin G. Drug and opioid-involved overdose deaths — United States, 2013–2017. MMWR Morb Mortal Wkly Rep 2019 Jan 4;67(5152):1419-1427. [doi: 10.15585/mmwr.mm6751521e1]
41. Suffoletto B, Landau A. Nudging emergency care providers to reduce opioid prescribing using peer norm comparison feedback: a pilot randomized trial. Pain Med 2020 Nov 7;21(7):1393-1399. [doi: 10.1093/pm/pnz314] [Medline: 31846029]
42. Weiner SG, Price CN, Atalay AJ, et al. A health system-wide initiative to decrease opioid-related morbidity and mortality. Jt Comm J Qual Patient Saf 2019 Jan;45(1):3-13. [doi: 10.1016/j.jcjq.2018.07.003] [Medline: 30166254]
43. Dowell D, Ragan KR, Jones CM, Baldwin GT, Chou R. CDC clinical practice guideline for prescribing opioids for pain - United States, 2022. MMWR Recomm Rep 2022 Nov 4;71(3):1-95. [doi: 10.15585/mmwr.rr7103a1] [Medline: 36327391]

Abbreviations

CDC: Centers for Disease Control and Prevention
CDW: clinical data warehouse
CMO: chief medical officer
ED: emergency department
EHR: electronic health record
EPCS: electronic prescribing of controlled substances
HITECH: Health Information Technology for Economic and Clinical Health
MEDD: morphine equivalent daily dose
OUD: opioid use disorder

Edited by A Kushniruk; submitted 28.11.22; peer-reviewed by M Keller, P Freda; revised version received 05.05.23; accepted 02.06.24; published 09.09.24.

Please cite as:
Slovis BH, Huang S, McArthur M, Martino C, Beers T, Labella M, Riggio JM, Pribitkin ED
Design and Implementation of an Opioid Scorecard for Hospital System–Wide Peer Comparison of Opioid Prescribing Habits: Observational Study
JMIR Hum Factors 2024;11:e44662
URL: https://humanfactors.jmir.org/2024/1/e44662
doi:10.2196/44662

Original Paper

Usability of an App for Medical History Taking in General Practice From the Patients' Perspective: Cross-Sectional Study

Klara Albrink[1]; Dominik Schröder[1], MSc; Carla Joos[1]; Frank Müller[1], MSc, Dr med; Eva Maria Noack[1], PhD

[1]Department of General Practice, University Medical Center Göttingen, Göttingen, Germany

Corresponding Author:
Eva Maria Noack, PhD
Department of General Practice
University Medical Center Göttingen
Humboldtallee 38
Göttingen, 37073
Germany
Phone: 49 551 39 65745
Email: evamaria.noack@med.uni-goettingen.de

Abstract

Background: A future shortage of physicians, especially in general practice, will result in an increasing workload for health care providers as a whole. Therefore, it is important to optimize patient-encounter processes to increase time efficiency related to visits. Utilizing digital tools to record patients' medical histories prior to a consultation offers great potential to achieve this goal. The collected information can be stored into the practice's electronic medical record, allowing for the general practitioner to review structured information of the patients' complaints and related medical history beforehand, thereby saving time during the encounter. However, the low usability of new digital developments in this setting often hinders implementation.

Objective: The aim of this study was to evaluate the usability of an app designed for medical history taking in general practice to capture the patients' perspective.

Methods: Between November 2021 and January 2022, we recruited 406 patients with acute complaints in one out-of-hour urgent care and seven general practice clinics. These study participants used the app during their waiting time and subsequently assessed its usability by completing the System Usability Scale (SUS), a robust and well-established 10-question survey measuring the perceived usability of products and technologies. Additionally, we collected general participant information, including age, sex, media usage, health literacy, and native language. Descriptive and inferential statistics were applied to identify patient characteristics associated with low or high SUS scores.

Results: We analyzed data from 397 patients (56.7% female, 43.3% male). The mean total SUS score was 77.8 points; 54.4% (216/397) of participants had SUS scores of 80 points or higher, indicating high usability of the app. In a multiple linear regression predicting SUS score, male sex and higher age (65 years or older) were significantly negatively associated with the SUS score. Conversely, a higher health literacy score and German as the native language were significantly positively associated with the SUS score.

Conclusions: Usability testing based on the SUS anticipates successful implementation of the app. However, not all patients will easily adapt to utilizing the app, as exemplified by the participants of older age in this study who reported lower perceived usability. Further research should examine these groups of people, identify the exact problems in operating such an app, and provide targeted solutions.

Trial Registration: German Clinical Trials Register World Health Organization Trial Registration Data Set DRKS00026659; https://trialsearch.who.int/Trial2.aspx?TrialID=DRKS00026659

(JMIR Hum Factors 2024;11:e47755) doi:10.2196/47755

KEYWORDS

digitization; application software; usability; mHealth; history of present illness; medical history taking

Introduction

As in many countries, demographic change is becoming evident in the German health care system, resulting in more complex, multimorbid patients [1] and a shortage of physicians [2]. Moreover, the proportion of older people in the population is rising steadily [3] and people tend to use medical services at a higher rate as they increase in age [4]. In Germany, one group that is particularly affected by this development are general practitioners (GPs) who are the first point of contact for patients requiring medical care and serve as the "gatekeepers" in the German health care system [5]. Approximately 80% of the German population aged 18 years and older are treated by a GP at least once a year [6]. A considerable number of GPs will retire in the upcoming years, resulting in 11,000 GP vacancies expected by 2035. These vacancies will disproportionately impact structurally weak and rural areas [7,8]. Without a sufficient workforce to replace the retired GPs and meet the greater demand for physicians, remaining GP workloads are expected to increase significantly within the next decade [9]. These developments challenge the German health care system at various levels and require attention to address the following key issues: future financing, improving allocation of resources, ensuring access to care, increasing efficiency and effectiveness of health care provision, and strengthened collaboration between providers [10].

To streamline patient care in the upcoming years, it is of importance to optimize patient-encounter processes to increase time efficiency related to visits. In this respect, digital tools offer great potential to support GPs in patient management, documentation workload, and the collection of medical history before consultation.

Digital tools designed to collect patients' medical history can ensure that information is always collected and documented thoroughly in a structured manner and with consistent quality. As many conditions can be diagnosed via a thorough medical history [11,12], these tools can be helpful in maintaining quality of care when time constraints may lead to an otherwise superficial medical history.

As part of our project titled "Digitally assisted information acquisition before medical consultation" (DASI), we developed an app for medical history taking in general practice settings. The app is used by the patient prior to the medical consultation, which could be either in the waiting area or at home. The collected information can be stored in the practice's electronic medical record and eventually be transferred to the individual electronic patient file, which statutory health insurers in Germany have been entitled to use since 2021 [13]. In the electronic patient file, patient data such as medical reports, X-rays, immunization records, and other medical data can be stored and shared among medical providers involved in the care of a particular patient by using the telematics infrastructure [14].

One advantage of the app is that the GP can review structured information of the patients' complaints and related medical history before the encounter. This is particularly helpful for patients that are unknown to the provider, those with many complaints, or those who have a comprehensive medical history. These situations are especially prevalent in out-of-hour urgent care practices. Furthermore, the tool might help patients to reflect on their conditions and enable them to better address their needs when seeing the provider. In this way, we expect that the limited consultation time can be used more efficiently.

Despite Germany's progress in digitalization within the health sector, concerns remain about the limited usability of new digital tools, hindering their full implementation. More than half of German practices see low usability as a strong obstacle to digitalization [15]. The evaluation of a digital tool's perceived usability is of special interest as it is a key determinant of performance for end users. Therefore, the aim of this study was to assess the usability of the app from the patients' perspective and to identify features in need of improvement. The broader aim is to ensure that the app is suitable for implementation in everyday practice, considering that GPs treat a broad range of patients of all ages and various educational and cultural backgrounds [16].

Methods

Study Design and Recruitment

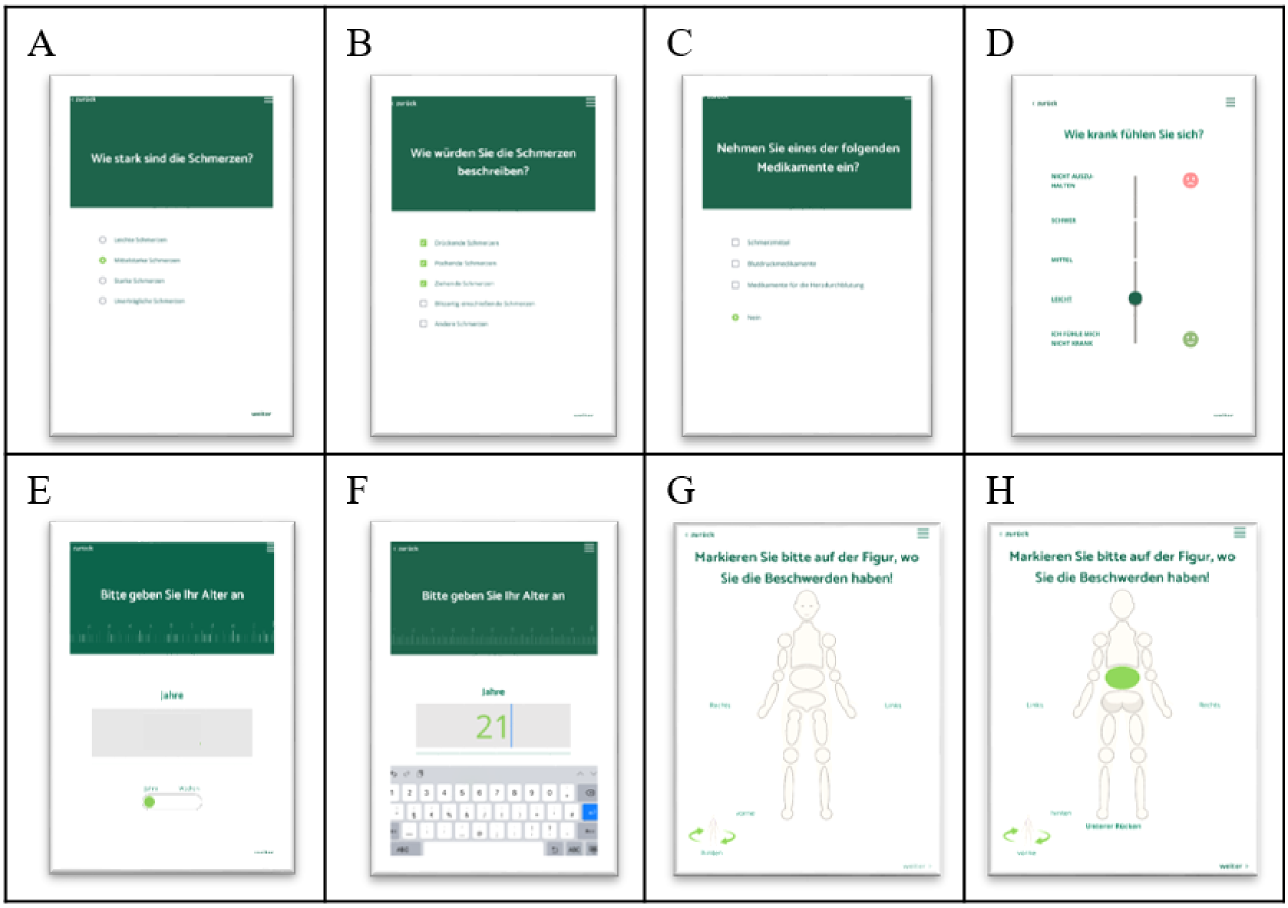
This was a cross-sectional study conducted in Germany in one out-of-hour urgent care practice and seven GP practices to assess the usability of an app designed for medical history taking in general practice settings.

Software and Hardware

The app was developed to take a medical history based on general medical complaints directly from the patients. While there are no international standards for the composition of a standardized patient history, this app was developed based on guidelines and health literature by medical experts from aidminutes GmbH (Hamburg/Buchholz in der Nordheide, Germany). For this study, the content and query structure were further refined for primary care (general practice and out-of-hour practices) by aidminutes GmbH in collaboration with experienced researchers from the Department of General Practice at the University Medical Center Göttingen, Germany. The app was designed to be used by patients in the waiting room before they see the doctor. Patients select one or several complaints and are then guided through a symptom-related questionnaire. In the sense of a branching logic, the app is adaptive to patient responses, which trigger further specific questions about the selected key complaints (eg, how and when a symptom started). Patients are also asked about preexisting conditions, previous treatments and surgeries, current medication, living habits, and chronic conditions in the family history. Information such as biological sex, height, weight, age, as well as the subjectively perceived severity of the complaints are inquired from all patients. More details can be found in the published study protocol [17].

The app was designed to be intuitive for the user such that no prior knowledge or any kind of instruction for its use is necessary. The user interface was designed to be simple to follow and only one question is asked per screen. As the app is operated in the waiting area, sound and video output of an earlier version [18] was omitted due to data protection. The questions are phrased in plain language; medical terminology is avoided or otherwise explained. The questions are substantially comprised by single-choice or multiple-choice questions that can be answered by tapping but also include several data fields (for age, height, and weight) and slider-type questions (Figure 1). The color scheme was designed to ensure reading accessibility for patients who may be color blind. A zoom function can be used for users who may experience visual impairment.

Figure 1. Screenshots of the app for medical history taking in general practice showing different types of questions: (A) single-choice question; (B) multiple-choice question; (C) hybrid question (ie, patients can either select several options or negate all of them); (D) slider for questions including a ranking between items (depicted here as "How sick do you feel?"); (E, F) data entry field (here: "Please enter your age"); and (G, H) selection of a body region on a figure (depicted in figure: "Please mark on the figure where you are suffering from the problems").



As this is a web-based app, it relies on a permanent internet connection. For this study, the app ran on an iPad Mini 4 (Apple Inc, Cupertino, CA, USA) held in an upright position. Tablets were equipped with haptically and visually inconspicuous cases (dark grey polyurethane leather outside and microfiber inside).

Setting

In Germany, GP practices aim at providing preventive, acute, and rehabilitative health care with long-lasting patient-doctor relationships. Out-of-hours urgent care practices provide urgent medical care for acute but not life-threatening cases when other practices are closed. Urgent care practices are often staffed with doctors of various specialties and an established relationship of care between the patients and doctors is not common. These aspects can lead to challenges in efficiently obtaining an accurate medical history and identifying serious health problems. Although the app was designed for general practice, it is also suitable to be used in out-of-hour urgent care practices.

Data Collection

The recruitment of patients was carried out by three study nurses and took place from November 22, 2021, to January 12, 2022. Patients were approached by study nurses in the waiting room of the respective practices before seeing their GP.

Patients meeting the following criteria were eligible to participate in the study: (1) seeking care in a participating practice because of acute somatic and/or psychological complaints, (2) at least 18 years old, and (3) consenting to participate in the study. Patients meeting the following exclusion criteria could not participate in the study: (1) younger than 18 years old (legal minor), (2) patients in an apparent emergency, (3) patients who required immediate medical treatment, and (4) patients who were unable to provide consent.

After the study nurses obtained written informed consent, a tablet on which the app was run on was handed over to the study participants. Participants used the app to report their medical history without an introduction on how to navigate the app. Once finished, they were asked to answer questions on personally perceived usability, media usage, and further sociodemographic data, which were digitally attached to the medical history–taking document. The study nurse in charge was present to observe any problems study participants may have had with using the app and was available to answer questions about the app's content and usability if specifically requested. Data were collected in an anonymized format without any personal information (eg, name or address) linking the results to each study participant. More detailed information on the data collection can be found in the study protocol [17].

Ethical Considerations

The Medical Ethics Committee of the University Medical Center Göttingen approved the study (approval number 26/3/21). A written informed consent form was collected from all patients before their inclusion in the study. Participating in the study was voluntary for patients. Patients could withdraw from participation without giving a reason at any time before they had completed the survey. Subsequently, their data could no longer be deleted because it could not be traced back to the individual.

Measures

The main outcome "usability" was measured using the System Usability Scale (SUS) [19], a commonly used instrument for this purpose [20]. The SUS was developed based on Standard ISO 9241-11 [21], in which usability is measured by the three main attributes of "effectiveness," "efficiency," and "satisfaction" [22,23]. Compared to other instruments, the SUS offers several advantages: (1) it can be analyzed quickly, (2) it is relatively easy to understand by academics from other disciplines [24], (3) it contains only 10 statements for easy completion, and (4) it can be used to evaluate almost any type of user interface [25]. We used the translated and validated German version of the SUS [26] and modified the statements to suit our purpose (see Multimedia Appendix 1).

The SUS consists of 10 statements (Table 1), where statements 1, 3, 5, 7, and 9 are positively connoted and statements 2, 4, 6, 8, and 10 are negatively connoted [19]. The scores for these statements are therefore inverted when calculating the sum. The raters decide on the extent to which they agree or disagree to these statements on a 5-point Likert scale ranging from 0 (strongly disagree) to 4 (strongly agree). The final sum score is multiplied by 2.5, resulting in a score range of 0-100 with higher scores indicating better usability [19].

Lewis and Sauro [27] developed a curved grading scale for SUS scores by comparing more than 200 industrial usability studies and using the percentile ranges, resulting in grades "C" (scores of 62.7-72.5), "B" (scores of 72.6-78.8), and "A" (scores of 78.9-100). As a SUS score of 80 proves an above-average user experience, it has become a common industrial goal. This threshold was therefore used for interpreting our results.

Table 1. Items of the System Usability Scale (SUS) [19].

Items	English version of the statement
SUS 1	I think that I would like to use this system frequently
SUS 2[a]	I found the system unnecessarily complex
SUS 3	I thought the system was easy to use
SUS 4[a]	I think that I would need the support of a technical person to be able to use this system
SUS 5	I found the various functions in this system were well integrated
SUS 6[a]	I thought that there was too much inconsistency in this system
SUS 7	I would imagine that most people would learn to use this system very quickly
SUS 8[a]	I found the system very cumbersome to use
SUS 9	I felt very confident using the system
SUS 10[a]	I needed to learn a lot of things before I could get going with this system

[a]The scores of negatively connoted SUS items were inverted when calculating the sum.

Covariates

Consultations in general practice are attended by patients of different ages and educational as well as cultural backgrounds, who have a different quantities of digital interactions in everyday life. To determine whether these factors have an influence on the personally perceived usability, we surveyed age, sex, media usage, health literacy, and native language. Information about age and sex were part of the app-taken medical history. In addition to the SUS, we asked patients about which digital media tools were available to them in everyday life (possible answers: cell phone/smartphone, computer/laptop/notebook, tablet, television, none, and others; multiple answers were possible) and how many hours a day they used digital media (possible answers: 0-≤1, 1-≤2, 2-≤3, 3-≤4, or 4 or more hours). We asked three questions concerning health literacy as a proxy for education attainment, given that educational achievement is the central determinant of health literacy [28]. Questions covering the three aspects of finding/accessing, evaluating/appraising, and understanding health-related information and content were derived from the European health literacy survey [29,30] adapted for the German language (HLS-GER 2 [31]). The HLS-GER 2 uses a predefined 4-point Likert scale.

Statistical Analysis

Data from the app were saved into a database and subsequently exported to a tab separated format for further analyses. Participants with two or more missing values of the SUS questionnaire were excluded from statistical analysis. In the case of one missing SUS response, we substituted the missing value with a neutral score of 2, as this method has been used with the SUS in previous research [32].

Sociodemographic data are presented as number and percentage of patients for each categorical data point. Mean and SD were utilized for interval or ratio-scaled data, which has become a common industrial goal. Sociodemographic data were compared between participants with SUS scores <80 and ≥80 using the Fisher exact test for 2×2 tables or the Fisher-Freeman-Halton test for categorical variables and the Wilcoxon rank-sum test for continuous variables. A multiple linear regression was conducted using sex, age, native German language, health literacy score, media usage duration per day, sickness level of the participants, and number of stated complaints in the app as independent variables and the SUS score as the dependent variable. Additionally, the individual SUS items were compared according to sex, age (<65 years vs ≥65 years), German native language, and tablet usage with the Wilcoxon rank-sum test. Data are visually presented as boxplots and radar charts. All analyses were carried out using R (4.1.3 under a GNU license) with the packages fmsb [33], psych [34], tidyr [35], dplyr [36], and ggplot2 [37].

Results

Patient Characteristics

We aimed to include approximately 400 patients for this study. This target was set to be able to form subgroups and to ensure that all types of patient complaints were included in our sample, including those selected on a limited basis. In total, individual data from 397 participants were included, with 5 participants having one missing SUS item. Figure 2 shows the flowchart of included patients and Table 2 shows the patients' characteristics.

Figure 2. Flowchart of patient inclusion in the cross-sectional study capturing patients' perceived usability of the app. SUS: System Usability Scale.

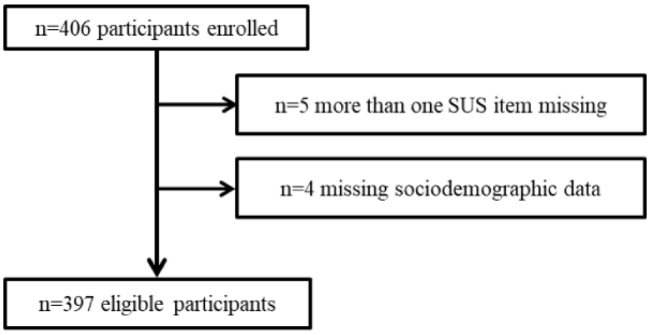


Table 2. Characteristics of the participants of the cross-sectional study capturing patients' perceived usability of the app (N=397).

Characteristics	Value
Sex, n (%)	
Female	225 (56.7)
Male	172 (43.3)
Age (years), median (IQR)	35.0 (25.0)
Age group (years), n (%)	
<30	152 (38.3)
30-65	223 (55.4)
65+	22 (6.3)
Native language German, n (%)	328 (82.6)
Devices used regularly[a], n (%)	
Smartphone	389 (98.0)
Tablet	210 (52.9)
Computer/notebook	310 (78.1)
Television	296 (74.6)
Media usage duration per day (hours), n (%)	
<2	89 (22.4)
2-4	174 (43.8)
>4	134 (33.8)
Self-assessed health literacy, median (IQR)[b]	
Understanding doctor	2.0 (0.0)
Search and understand health information	2.0 (1.0)
Evaluate health information	1.0 (1.0)
"How sick do you feel?"[c], n (%)	
I don't feel sick	32 (8.1)
Just a little	70 (17.6)
Fairly	226 (56.9)
Very	61 (15.4)
Unbearably	6 (1.5)

[a]Multiple selection possible.

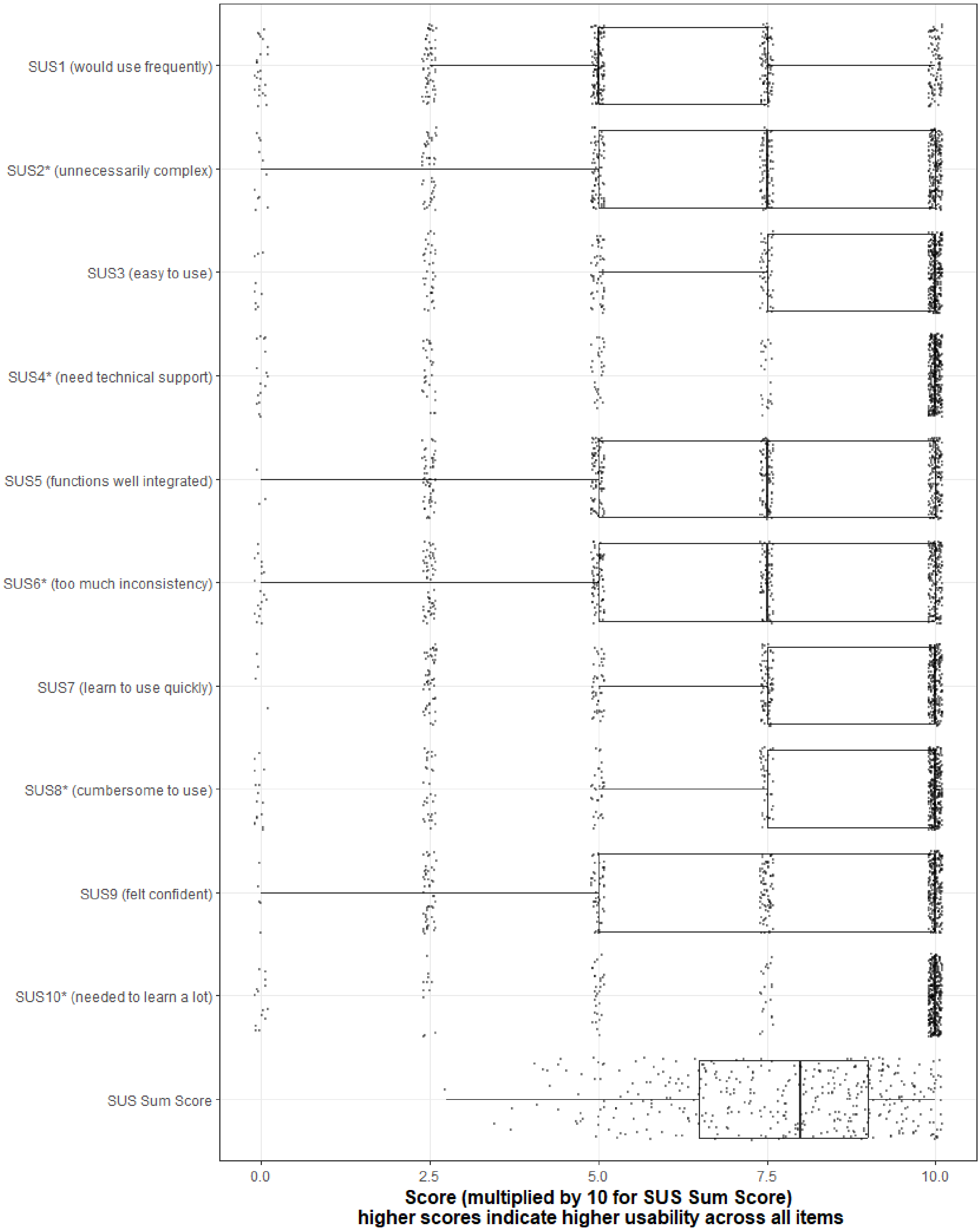
[b]Measured on a 4-point (0-3) Likert-scale (higher scores indicate higher health literacy levels).

[c]Perceived severity of acute complaint.

Usability for All Participants

We found a mean total SUS score of 77.8 points, with 54.4% (216/397) of participants having SUS scores of 80 points or higher, indicating high usability of the app overall. Figure 3 shows boxplots of the individual items in which the scores were calculated for each statement. Irrespective of a positive or negative connotation, a higher score indicates a better result. The maximum score that can be achieved for each item is 10.

Figure 3. SUS items and sum scores for all study participants (N=397). SUS: System Usability Scale. *Scoring was inverted.



Usability Stratified by Sociodemographic Factors

We divided the sample into two groups with the cutoff at a SUS score of 80. Participants with a SUS score of at least 80 were significantly younger, reported higher levels of technology device usage, and higher levels of self-assessed health literacy compared to participants with a SUS score below 80 (Table 3).

Table 3. Sociodemographic variables of study participants stratified by System Usability Scale (SUS) score.

Variable	SUS score<80 (n=181)	SUS score≥80 (n=216)	*P* value
Sex, n (%)			.09[a]
Female	94 (51.9)	131 (60.6)	
Male	87 (48.1)	85 (39.4)	
Age (years), median (IQR)	38.0 (26.0)	32.5 (22.0)	.002[b]
Age group (years), n (%)			.003[c]
<30	55 (30.3)	97 (44.9)	
30-65	111 (60.2)	112 (51.4)	
65+	15 (9.4)	7 (3.7)	
Native language German, n (%)	142 (78.5)	186 (86.1)	.05[a]
Devices used regularly[d], n (%)			
Smartphone	174 (96.1)	215 (99.5)	.03[a]
Tablet	84 (46.4)	126 (58.3)	.02[a]
Computer/notebook	136 (75.1)	174 (80.6)	.22[a]
Television	125 (69.1)	171 (79.2)	.03[a]
Media usage duration per day (hours), n (%)			.08[c]
<2	48 (26.5)	41 (19.0)	
2-4	69 (38.1)	105 (48.6)	
>4	64 (35.4)	70 (32.4)	
Self-assessed health literacy, median (IQR)			
Understanding doctor	2.0 (0.0)	2.0 (1.0)	.16[b]
Search and understand health information	2.0 (0.0)	2.0 (1.0)	.01[b]
Assess confidence of health information	1.0 (1.0)	1.0 (1.0)	.19[b]
"How sick do you feel?"[e], n (%)			.35[c]
I don't feel sick	11 (6.1)	21 (9.7)	
Just a little	30 (16.6)	40 (18.5)	
Fairly	102 (56.4)	124 (57.4)	
Very	34 (18.8)	27 (12.5)	
Unbearably	3 (1.7)	3 (1.4)	

[a]Fisher exact test.
[b]Wilcoxon rank-sum test.
[c]Fisher-Freeman-Halton test.
[d]Multiple selection possible.
[e]Perceived severity of acute complaint.

A multiple linear regression predicting the SUS score was conducted, including sex, age, native German language, health literacy score, media usage duration per day, sickness level of the participants, and number of stated complaints in the app as independent variables (see Table 4). Age, sex, health literacy score, and German native language were significantly associated with SUS score. A higher age (t_{385}=3.30, *P*=.001) and male sex (t_{385}=1.98, *P*=.05) were negatively associated with SUS score, whereby a higher health literacy score (t_{385}=2.83, *P*=.01) and German as a native language (t_{385}=2.51, *P*=.01) were positively associated with SUS score.

Table 4. Multiple linear regression predicting the System Usability Scale sum score.

Variable	β (95% CI)	*P* value
Male sex (reference=female)	–3.21 (–6.40 to –.02)	.05
age (per year)	–.17 (–.27 to –.07)	.001
German not native language (reference=yes)	–5.39 (–9.61 to –1.17)	.01
Does not use tablet (reference=yes)	–1.44 (–4.66 to 1.79)	.38
Average daily media usage (reference=<2 h)		
2-4 h	3.21 (–.95 to 7.37)	.13
>4 h	.01 (–4.46 to 4.48)	.99
Health literacy score (scale 0-9)[a]	1.48 (.45 to 2.51)	.01
How sick do you feel? (score 1-5)[b]	–1.05 (–2.92 to .82)	.27
Number of stated complaints (1-11)	–.98 (–2.11 to .15)	.09

[a]Higher scores indicate a higher level of health literacy.

[b]Perceived severity of acute complaint; higher values indicate a higher level of discomfort.

Differences in Individual Items of the SUS

Stratified according to sex, age, native language, and tablet usage (see Figure 4), significant differences were detected in SUS items 2 (“unnecessarily complex”), 4 (“need technical support”), 7 (“learn to use quickly”), 8 (“cumbersome to use”), and 10 (“needed to learn a lot”).

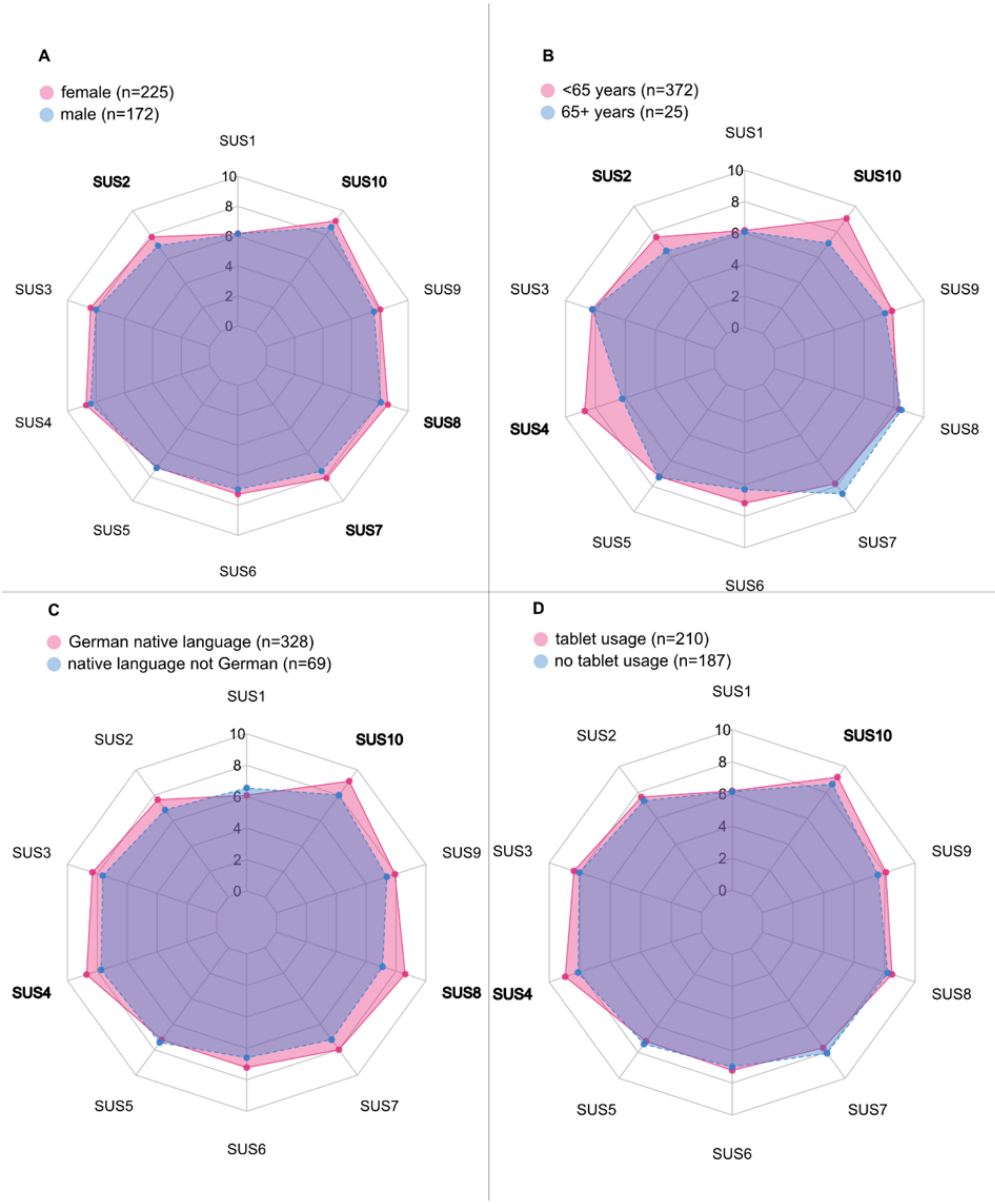
In comparing female and male respondents, all statements were rated more positively by female participants, except for items 1 (“would use frequently”) and 4 (“need technical support”). Female participants also scored significantly higher than male participants on items 2 (“unnecessarily complex”) (mean 7.82 vs 7.11; *P*=.02), 7 (“learn to use quickly”) (mean 8.11 vs 7.53; *P*=.05), 8 (“cumbersome to use”) (mean 8.57 vs 8.08; *P*=.05), and 10 (“needed to learn a lot”) (mean 9.14 vs 8.63; *P*=.03).

Respondents aged 65 years and older scored significantly higher on items 2 (“unnecessarily complex”) (mean 7.58 vs 6.50; *P*=.04), 4 (“need technical support”) (mean 8.72 vs 6.20; *P*<.001), and 10 (“needed to learn a lot”) (mean 9.05 vs 7.; *P*<.001) compared to their counterparts.

German native language speakers scored significantly higher on items 4 (“need technical support”) (mean 8.73 vs 7.75; *P*=.001), 8 (“cumbersome to use”) (mean 8.62 vs 7.10; *P*<.001), and 10 (“needed to learn a lot”) (mean 9.11 vs 8.01; *P*<.001) relative to nonnative speakers.

Lastly, patients who regularly use a tablet had significantly higher SUS scores on items 4 (“need technical support”) (mean 8.95 vs 8.11; *P*<.001) and 10 (“needed to learn a lot”) (mean 9.18 vs 8.64; *P*=.02) in comparison to those of participants who reported reduced levels of tablet use.

Figure 4. Mean System Usability Scale (SUS) score items (see Table 1) stratified according to sex (A), age (B), German as native language (C), and current tablet usage (D). bold: *P*<.05 (Wilcoxon rank-sum test).



Discussion

Principal Findings

In this study, we evaluated the usability of an app in taking medical histories in general practice directly from patients using the SUS [19].

The app achieved a mean SUS score of 77.9, which corresponds to a B+ grade on the curved grading scale [27] and represents a "better" product that does not necessarily need improvement [25]. Other medical devices, even those widely used at home, have lower SUS scores. Kortum and Peres [38] assessed the usability of home health care devices among students, thus representing relatively young, healthy, and well-educated participants. SUS scores for these devices ranged from 65 for an epinephrine injector to 67 for a pregnancy test kit and 81 for a thermometer, even with previous experience using these devices.

To ensure patients can be active participants in the digital medical history–taking process, the app must be easy and intuitive to use without technical introduction or support. This importance is reflected in items 3 (“easy to use”), 4 (“need technical support“), 7 (“learn to use quickly”), and 10 (“needed to learn a lot”). Mean values between 7.8 and 8.9 for these items indicate that intuitive use has been successfully addressed in the development of our app. Item 1, assessing the frequency of app usage, scored the lowest (6.2), which can be explained by the app’s implementation solely in a medical setting and not utilized regularly in leisure time. As such, this finding is the least meaningful for our purpose.

In a pilot study by Melms et al [39], a self-completed tablet-based digital questionnaire designed for collecting medical histories in an emergency department was found to score high with respect to perceived usability. The design and content were similar to those of our app; however, their questions were only based on the SUS, which does not allow direct comparison. Other comparable instruments, although also for emergency departments, have been tested for usability in pilot studies using self-developed satisfaction surveys [40,41], a single question, and researcher or staff documentation of a patient’s need for assistance [42]. In these studies, patients were mostly satisfied with the self-administered medical history–taking tools and reported good ease of use. Taken together, these results give hope that it is possible to design a medical history app that is perceived as user-friendly.

Nonetheless, obstacles to implementing a digital tool in general practice settings can be multifaceted. Surprisingly, we found that sex was significantly associated with usability; female participants had significantly higher SUS scores than male participants. The fact that men scored higher than women for item 4 (“need technical support”) suggests that men felt more confident than women with using the app. Previous studies demonstrated that men tend to report overconfidence in their abilities, especially in fields with a male connotation [43], which computer science certainly represents [44]. Therefore, it is unclear whether men really would have needed less help or whether they overestimated themselves in their technical skills.

Our study suggests that older people are more likely to have difficulties with the handling of such an app. This aligns with a study showing that from the retirement age of 65 years, digital media use among the German population begins to decline dramatically [45] and a positive attitude toward digitalization decreases with increasing age [46]. Older age has a negative impact on the broad usability score given to a user interface [25]. To that end, this study cannot definitively conclude if the older participants of this study actually perceived the app to be of relatively low usability or if their more negative attitude toward the benefit of new technologies prompted them to give lower scores. Due to the small sample size of participants aged 65 years and older, it is not possible to assert how older people in general would cope with the handling of the app. Since GPs are consulted predominantly by older people [47], further research should focus on app testing with older patients to obtain specific feedback, including suggestions for improvement.

Having learned German as a native language was positively associated with a higher SUS score, although only patients with sufficient German language proficiency were included in the study [17]. This could be due to two different reasons: despite the app’s plain language, it is possible that some of the medical history questions or SUS items were not understood properly.

Daily media use was not associated with the SUS score, which suggests that the app is designed to also ensure that people with limited digital experience do not feel overstrained with its operation.

Limitations

Despite our efforts, this study comes with several limitations. The number of older participants (ie, aged 65 years and above) was relatively low in comparison to their constituents in GP practice settings [47]. One potential reason could be a more pronounced skepticism toward digital tools in older generations, leading to an increase in refusal for participation in the study among older patients. However, as no screening lists were maintained, this is mere speculation. A screening list should be obtained in future studies to be able to characterize individuals who declined participation. Another consideration is that people with lower levels of digital media literacy use may not have agreed to participate in the study.

Data collection was performed during the SARS-CoV-2 pandemic, which may have disproportionately impacted study participants as certain patient groups may have avoided seeing a doctor or were more likely to refuse to participate in the study to avoid unnecessary contact. This could have included especially vulnerable groups such as older people or those with multimorbid conditions.

The Likert scale of the SUS questions shown with clickable singular dots was replaced by a slider on December 8, 2021. In the dot-based representation, it was compulsory to make an entry before continuing, whereas the slider was automatically set to the neutral center and could be shifted. This may have led to incorrectly rated items. For example, this may have occurred in instances of the internet faltering or the patient having double-clicked without noticing. Since there were repeated questions about the word “Inkonsistenzen” (SUS item 2), we replaced it by the more common synonym “Unstimmigkeiten” (English translation: discrepancies). Furthermore, hardware as well as the operating system may have influenced the evaluation of the personally perceived usability of a system [48]. For this study, iPad Minis with the iOS operating system were exclusively used. Therefore, possible differences in assessment related to the operating system and hardware are not part of this study.

The SUS is able to classify the usability of a system but is unable to identify specific usability issues nor capture the usability of the system in its entirety. For a more in-depth usability evaluation, different methods could be used (eg, interviews and observations). During data collection, staff were able to observe usability problems. In their observations, multiple-choice, single-choice, and hybrid questions as well as the slider did not appear to cause any difficulties. In contrast, problems concerning the handling of the app arose when participants were required to input free-text entries (eg, age, height, weight). Further, some study participants were unclear on how to open and close the on-screen keyboard. Some participants also did not understand that the figure on which a pain or an injury could be assigned to a body region (see Figure 1E) could be rotated by clicking on an icon at the bottom left of the screen. This means that, for example, back pain may have been falsely reported as abdominal pain. Lastly, an unstable internet connection arose during data collection, which caused the app to be unresponsive intermittently. These factors may have influenced the SUS score.

Conclusion

The app examined in this study for medical history taking passes the usability test based on the SUS and appears to function on par with other digital tools that have become well-integrated in our everyday lives. However, not all people adapted equally well to the app. For successful implementation, all end users, regardless of age, technical affinity, health literacy, or preferred language, must be able to use such a tool. Only if that is attained, providing practical digital solutions can contribute to the efficient and effective delivery of health care services. Therefore, further research should focus on the identification of causes for difficulties of using the app as well as finding appropriate solutions.

Acknowledgments

We would like to thank Julie Ngo, MD for proofreading. We sincerely thank all patients for their participation. We are very grateful to the physicians and their teams who allowed data collection in their practices despite the SARS-CoV-2 pandemic and the extraordinary burden that came along for health care workers. We thank the team at aidminutes GmbH for their commitment and support. This research is funded by the German Innovation Fund (funding number 01VSF19050) of the Federal Joint Committee (G-BA). The funders have not influenced the design of this study and did not play any role during its implementation (eg, data collection and analysis, interpretation, and publication of the results).

Data Availability

The data sets used and analyzed during the study are not publicly available due to the decision of the research ethics board but can be obtained from the authors upon reasonable request within a data sharing agreement.

Authors' Contributions

KA wrote the original draft. KA and EMN were mainly responsible for writing the manuscript. KA, EMN, and DS conceived the study design and analyzed the data. FM and CJ revised the manuscript. All authors have read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Original and German versions of the System Usability Scale (SUS) and the items customized for this study.
[DOCX File , 15 KB - humanfactors_v11i1e47755_app1.docx]

References

1. GBE kompakt 3(2) Demografische Alterung und Folgen für das Gesundheitswesen. Robert Koch Institut. 2012. URL: https://www.rki.de/DE/Content/Gesundheitsmonitoring/Gesundheitsberichterstattung/GBEDownloadsK/2012_2_Demografischer_Wandel_Alterung.pdf?__blob=publicationFile [accessed 2024-01-02]
2. Gesundheit in Deutschland. Gesundheitsberichterstattung des Bundes. 09 Welche Auswirkungen hat der demografische Wandel auf Gesundheit und Gesundheitsversorgung? Robert Koch-Institut. 2015. URL: https://www.rki.de/DE/Content/Gesundheitsmonitoring/Gesundheitsberichterstattung/GBEDownloadsGiD/2015/09_gesundheit_in_deutschland.pdf?__blob=publicationFile [accessed 2024-01-02]
3. Die Bevölkerungsgruppe der älteren Menschen ab 65 Jahren 2021. Statistisches Bundesamt. URL: https://www.destatis.de/DE/Themen/Querschnitt/Demografischer-Wandel/Aeltere-Menschen/bevoelkerung-ab-65-j.html [accessed 2022-05-19]
4. Kürschner N, Weidmann C, Müters S. Who enrolls in a general practitioner model? The behavioral model of health services use and general practitioner-centered care in Germany. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2011 Feb;54(2):221-227. [doi: 10.1007/s00103-010-1205-x] [Medline: 21290277]
5. Schnitzer S, Balke K, Walter A, Litschel A, Kuhlmey A. Do gatekeeping programs increase equality of health care in Germany? A comparison of the health care situation of participants and nonparticipants. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2011 Aug;54(8):942-950. [doi: 10.1007/s00103-011-1317-y] [Medline: 21800242]

6. Prütz F, Rommel A, Thom J, Du Y, Sarganas G, Starker A. Inanspruchnahme ambulanter medizinischer Leistungen in Deutschland – Ergebnisse der Studie GEDA 2019/2020-EHIS. J Health Monitor 2021;6(3):49-71. [doi: 10.25646/8554]
7. GESUNDHEITSDATEN: Niedergelassene Ärztinnen und Ärzte werden immer älter. Kassenärztliche Bundesvereinigung (KBV). 2022. URL: https://gesundheitsdaten.kbv.de/cms/html/16397.php [accessed 2022-05-19]
8. Beerheide R, Haserück A. Versorgungsstrukturen: Hausärztemangel angehen. Deutsches Ärzteblatt 2021;118(38):A-1684 /B-1396.
9. Schulz M, Czihal T, Bätzing-Feigenbaum J, von Stillfried D. Zukünftige relative Beanspruchung von Vertragsärzten – Eine Projektion nach Fachgruppen für den Zeitraum 2020 bis 2035. Versorgungsatlas. 2016. URL: https://www.versorgungsatlas.de/fileadmin/ziva_docs/67/VA-16-02-Bericht-Final_V3.pdf [accessed 2022-05-19]
10. Blümel M, Spranger A, Achstetter K, Maresso A, Busse R. Germany: health system review. Health Syst Transit 2020 Dec;22(6):1-272 [FREE Full text] [Medline: 34232120]
11. Hampton JR, Harrison MJ, Mitchell JR, Prichard JS, Seymour C. Relative contributions of history-taking, physical examination, and laboratory investigation to diagnosis and management of medical outpatients. Br Med J 1975 May 31;2(5969):486-489 [FREE Full text] [doi: 10.1136/bmj.2.5969.486] [Medline: 1148666]
12. Roshan M, Rao AP. A study on relative contributions of the history, physical examination and investigations in making medical diagnosis. J Assoc Physicians India 2000 Aug;48(8):771-775. [Medline: 11273467]
13. Bundesgesetzblatt Jahrgang 2020 Teil I Nr. 46, ausgegeben zu Bonn am 19. Gesetz zum Schutz elektronischer Patientendaten in der Telematikinfrastruktur/Patientendaten-Schutz-Gesetz (PDSG). Bundesrepublik Deutschland. 2020 Oct. URL: https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3_Downloads/Gesetze_und_Verordnungen/GuV/P/PDSG_bgbl.pdf [accessed 2024-01-02]
14. Patientendaten-Schutz-Gesetz. Bundesministerium für Gesundheit. URL: https://www.bundesgesundheitsministerium.de/patientendaten-schutz-gesetz.html [accessed 2024-01-02]
15. Albrecht M, Otten M, Sander M. Praxisbarometer Digitalisierung 2021. KBV. URL: https://www.kbv.de/media/sp/IGES_PraxisBaroDigit2021_Kurzbericht.pdf [accessed 2022-05-19]
16. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Mar;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
17. Albrink K, Joos C, Schröder D, Müller F, Hummers E, Noack EM. Obtaining patients' medical history using a digital device prior to consultation in primary care: study protocol for a usability and validity study. BMC Med Inform Decis Mak 2022 Jul 19;22(1):189 [FREE Full text] [doi: 10.1186/s12911-022-01928-0] [Medline: 35854290]
18. Müller F, Chandra S, Furaijat G, Kruse S, Waligorski A, Simmenroth A, et al. A Digital Communication Assistance Tool (DCAT) to obtain medical history from foreign-language patients: development and pilot testing in a primary health care center for refugees. Int J Environ Res Public Health 2020 Mar 20;17(4):1368 [FREE Full text] [doi: 10.3390/ijerph17041368] [Medline: 32093286]
19. Brooke J. SUS: A 'quick and dirty' usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation In Industry. London: CRC Press; 1995:189-194.
20. Lewis JR. The System Usability Scale: past, present, and future. Int J Hum Comput Interact 2018 Mar 30;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
21. Ergonomics of human-system interaction – Part 11: Usability: Definitions and concepts. International Organization for Standardization (ISO). 2018. URL: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en [accessed 2022-05-19]
22. Abran A, Khelifi A, Suryn W, Seffah A. Usability meanings and interpretations in ISO standards. Softw Qual J 2003;11(4):325-338. [doi: 10.1023/a:1025869312943]
23. Kortum P, Hebl M, Oswald FL. Applying usability measures to assess textbooks. 2014 Presented at: 58th Human Factors and Ergonomics Society Annual Meeting; October 27-31, 2014; Chicago, IL. [doi: 10.1177/1541931214581281]
24. Banger A, Kortum P, Miller J. Determining what individual SUS Scores mean: adding an adjective rating scale. J Usability Stud 2009;4:114-123 [FREE Full text]
25. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
26. Gao M, Kortum P, Oswald FL. Multi-language toolkit for the System Usability Scale. Int J Hum Comput Interact 2020 Aug 19;36(20):1883-1901. [doi: 10.1080/10447318.2020.1801173]
27. Lewis JR, Sauro J. Item benchmarks for the System Usability Scale. J Usability Stud 2018;13(3):158-167 [FREE Full text]
28. Sørensen K, Pelikan JM, Röthlin F, Ganahl K, Slonska Z, Doyle G, HLS-EU Consortium. Health literacy in Europe: comparative results of the European health literacy survey (HLS-EU). Eur J Public Health 2015 Dec;25(6):1053-1058 [FREE Full text] [doi: 10.1093/eurpub/ckv043] [Medline: 25843827]
29. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12:80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
30. HLS-EU CONSORTIUM. Comparative Report of Health Literacy in Eight EU Member States (Second Revised and Extended Version): The European Health Literacy Survey HLS-EU. 2014 Jul 22. URL: https://www.hsph.harvard.edu/wp-content/uploads/sites/135/2015/09/neu_rev_hls-eu_report_2015_05_13_lit.pdf [accessed 2022-05-19]

31. Schaeffer D, Gille S, Berens E, Griese L, Klinger J, Vogt D, et al. Digital health literacy of the population in Germany: results of the HLS-GER 2. Gesundheitswesen 2023 Apr 14;85(4):323-331. [doi: 10.1055/a-1670-7636] [Medline: 34905785]
32. Downey RG, King C. Missing data in Likert ratings: a comparison of replacement methods. J Gen Psychol 1998 Apr;125(2):175-191. [doi: 10.1080/00221309809595542] [Medline: 9935342]
33. Package 'fsmb'. CRAN R project. 2023. URL: https://cran.r-project.org/web/packages/fmsb/fmsb.pdf [accessed 2023-12-29]
34. Revelle WR. psych: procedures for personality and psychological research. Northwestern University Northwestern Scholars. 2022. URL: https://www.scholars.northwestern.edu/en/publications/psych-procedures-for-personality-and-psychological-research [accessed 2022-05-19]
35. tidyr: Tidy Messy Data. URL: https://tidyr.tidyverse.org/ [accessed 2022-05-19]
36. dplyr: A Grammar of Data Manipulation. URL: https://dplyr.tidyverse.org/ [accessed 2023-12-29]
37. Wickham H. ggplot2: Elegant graphics for data analysis, second edition. Cham: Springer; 2016.
38. Kortum P, Peres SC. Evaluation of home health care devices: remote usability assessment. JMIR Hum Factors 2015 Jun 05;2(1):e10 [FREE Full text] [doi: 10.2196/humanfactors.4570] [Medline: 27025664]
39. Melms L, Schaefer JR, Jerrentrup A, Mueller T. A pilot study of patient satisfaction with a self-completed tablet-based digital questionnaire for collecting the patient's medical history in an emergency department. BMC Health Serv Res 2021 Jul 30;21(1):755 [FREE Full text] [doi: 10.1186/s12913-021-06748-y] [Medline: 34330279]
40. Arora S, Goldberg A, Menchine M. Patient impression and satisfaction of a self-administered, automated medical history-taking device in the Emergency Department. West J Emerg Med 2014 Mar;15(1):35-40 [FREE Full text] [doi: 10.5811/westjem.2013.2.11498] [Medline: 24695871]
41. Benaroia M, Elinson R, Zarnke K. Patient-directed intelligent and interactive computer medical history-gathering systems: a utility and feasibility study in the emergency department. Int J Med Inform 2007 Apr;76(4):283-288. [doi: 10.1016/j.ijmedinf.2006.01.006] [Medline: 16473548]
42. Herrick DB, Nakhasi A, Nelson B, Rice S, Abbott PA, Saber Tehrani AS, et al. Usability characteristics of self-administered computer-assisted interviewing in the emergency department: factors affecting ease of use, efficiency, and entry error. Appl Clin Inform 2013;4(2):276-292 [FREE Full text] [doi: 10.4338/ACI-2012-09-RA-0034] [Medline: 23874364]
43. Jakobsson N, Levin M, Kotsadam A. Gender and overconfidence: effects of context, gendered stereotypes, and peer group. Adv Appl Sociol 2013;3(2):137-141. [doi: 10.4236/aasoci.2013.32018]
44. Friedrich JD, Hachmeister CD, Nickel S, Peksen S, Roessler I. Frauen in Informatik: Welchen Einfluss haben inhaltliche Gestaltung, Flexibilisierung und Anwendungsbezug der Studiengänge auf den Frauenanteil? CHE Arbeitspapier Nr 200. Centrum für Hochshulentwicklung (CHE). 2018 Apr. URL: https://www.uni-bremen.de/fileadmin/user_upload/sites/zentrale-frauenbeauftragte/Berichte/CHE_AP_200_Frauen_in_Informatik.pdf [accessed 2022-05-19]
45. Digital skills gap. Initiative D21. 2021. URL: https://initiatived21.de/publikationen/digital-skills-gap [accessed 2022-05-19]
46. Digitale Teilhabe lässt ab 80 Jahren deutlich nach. Initiative D21. 2021. URL: https://initiatived21.de/digitale-teilhabe-laesst-ab-80-jahren-deutlich-nach/ [accessed 2022-05-19]
47. Bergmann E, Kalcklösch M, Tiemann F. Public health care utilisation. Initial results of the Telephone Health Survey 2003. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2005 Dec;48(12):1365-1373. [doi: 10.1007/s00103-005-1167-6] [Medline: 16270187]
48. Calisir F, Cevikcan E, Camgoz AH. Industrial engineering in the big data era. Cham: Springer; 2019.

Abbreviations

DASI: digitally assisted information acquisition before medical consultation
GP: general practitioner
HLS-GER 2: European health literacy survey adapted for the German language
SUS: System Usability Scale

Edited by A Kushniruk; submitted 31.03.23; peer-reviewed by A Simmenroth, R Marshall; comments to author 27.08.23; revised version received 10.10.23; accepted 20.11.23; published 05.01.24.

Please cite as:
Albrink K, Schröder D, Joos C, Müller F, Noack EM
Usability of an App for Medical History Taking in General Practice From the Patients' Perspective: Cross-Sectional Study
JMIR Hum Factors 2024;11:e47755
URL: https://humanfactors.jmir.org/2024/1/e47755
doi:10.2196/47755
PMID:38180798

Original Paper

A Novel Continuous Real-Time Vital Signs Viewer for Intensive Care Units: Design and Evaluation Study

Shiming Yang[1], PhD; Samuel Galvagno[1], MD, PhD; Neeraj Badjatia[2], MD; Deborah Stein[3], MD; William Teeter[4], MD; Thomas Scalea[3], MD; Stacy Shackelford[5], MD; Raymond Fang[6], MD; Catriona Miller[3], PhD; Peter Hu[1], PhD; VS viewer study group[7]

[1]Department of Anesthesiology, University of Maryland School of Medicine, Baltimore, MD, United States
[2]Department of Neurology, University of Maryland School of Medicine, Baltimore, MD, United States
[3]Department of Surgery, University of Maryland School of Medicine, Baltimore, MD, United States
[4]Emergency Medicine, University of Maryland School of Medicine, Baltimore, MD, United States
[5]United States Air Force Academy, Colorado Springs, CO, United States
[6]Department of Surgery, Johns Hopkins University School of Medicine, Baltimore, MD, United States
[7]See Acknowledgments

Corresponding Author:
Shiming Yang, PhD
Department of Anesthesiology
University of Maryland School of Medicine
11 S Paca St. LL01
Baltimore, MD, 21201
United States
Phone: 1 4103284179
Email: syang@som.umaryland.edu

Abstract

Background: Clinicians working in intensive care units (ICUs) are immersed in a cacophony of alarms and a relentless onslaught of data. Within this frenetic environment, clinicians make high-stakes decisions using many data sources and are often oversaturated with information of varying quality. Traditional bedside monitors only depict static vital signs data, and these data are not easily viewable remotely. Clinicians must rely on separate nursing charts—handwritten or electric—to review physiological patterns, including signs of potential clinical deterioration. An automated physiological data viewer has been developed to provide at-a-glance summaries and to assist with prioritizing care for multiple patients who are critically ill.

Objective: This study aims to evaluate a novel vital signs viewer system in a level 1 trauma center by subjectively assessing the viewer's utility in a high-volume ICU setting.

Methods: ICU attendings were surveyed during morning rounds. Physicians were asked to conduct rounds normally, using data reported from nurse charts and briefs from fellows to inform their clinical decisions. After the physician finished their assessment and plan for the patient, they were asked to complete a questionnaire. Following completion of the questionnaire, the viewer was presented to ICU physicians on a tablet personal computer that displayed the patient's physiologic data (ie, shock index, blood pressure, heart rate, temperature, respiratory rate, and pulse oximetry), summarized for up to 72 hours. After examining the viewer, ICU physicians completed a postview questionnaire. In both questionnaires, the physicians were asked questions regarding the patient's stability, status, and need for a higher or lower level of care. A hierarchical clustering analysis was used to group participating ICU physicians and assess their general reception of the viewer.

Results: A total of 908 anonymous surveys were collected from 28 ICU physicians from February 2015 to June 2017. Regarding physicians' perception of whether the viewer enhanced the ability to assess multiple patients in the ICU, 5% (45/908) strongly agreed, 56.6% (514/908) agreed, 35.3% (321/908) were neutral, 2.9% (26/908) disagreed, and 0.2% (2/908) strongly disagreed.

Conclusions: Morning rounds in a trauma center ICU are conducted in a busy environment with many data sources. This study demonstrates that organized physiologic data and visual assessment can improve situation awareness, assist clinicians with recognizing changes in patient status, and prioritize care.

(JMIR Hum Factors 2024;11:e46030) doi:10.2196/46030

KEYWORDS

clinical decision-making; health information technology; intensive care units; patient care prioritization; physiological monitoring; visualization; vital signs

Introduction

Clinicians working in intensive care units (ICUs) must be able to see, understand, and respond quickly to the complex and ever-changing clinical environment of the ICU. They need to be able to collect, analyze, and interpret what is happening and what it means [1]. Situational awareness is essential for ICU clinicians to provide safe and effective care to their patients. When clinicians have good situational awareness, they are better able to identify and respond to changes in their patients' condition and to coordinate care with other members of the health care team. However, clinicians are immersed in a cacophony of alarms and a relentless onslaught of data. Within this frenetic environment, clinicians make high-stakes decisions using multiple data sources and are often oversaturated with information of varying quality. While modern hospitals are equipped with bedside monitors collecting various physiological data in a real-time, continuous, and automated way, these data are not always easily accessible remotely or available to be viewed as a continuous trend [2]. The enormous amount of unprocessed data adds an additional burden on ICU clinicians who work in a dynamic environment with voluminous decision-making requirements. Traditional bedside monitors only show a single patient's instantaneous (static) vital signs (VS) data, limiting the clinician's scope to view a patient's physiological trajectory within a clinically meaningful period of time. Clinicians must rely on separate nursing charts—handwritten or electronic—to review a patient's physiological status. Moreover, auditory alarms often cause "alarm fatigue" instead of increasing situational awareness [3]. Many bedside monitors only display 1 or 2 patients' information; the ability to view an entire unit or ward allows a clinician to prioritize attention to those in most need of critical care support [4]. Improved visualization of patient information may help clinicians cope with information overload in critical care settings by improving situational awareness and supporting clinical decision-making [5]. An automated physiological data-organizing and information-summary system that presents aggregated information from multiple data sources while providing at-a-glance summaries of clinical data can assist ICU clinicians with prioritizing care for multiple patients.

Developed initially for use in aircraft transporting multiple patients who are critically ill, this VS viewer has 2 outcomes of direct and important clinical applicability. First, the VS viewer can provide clinicians with the capability to monitor individual patient trends, improving overall decision-making. Since patients in the ICU require multiple life support treatments to ensure ideal long-term outcomes, improved display of VS patterns could improve patient assessment and clinical decision-making. Second, the VS viewer system allows remote monitoring of groups of patients through a display that provides clinicians with the ability to quickly identify patients in need of rapid intervention. The objective of this work is to evaluate the use of a VS viewer in ICUs at a high-volume level 1 trauma center. We hypothesized that clinicians would subjectively report improved situational awareness and enhanced ability to make clinical decisions with the use of a VS viewer.

Methods

Data and System Design

In the ICUs of the University of Maryland Medical System, GE Marquette Solar 7000/8000 (General Electric) patient VS monitors are networked to provide a collection of real-time patient VS data streams. Each patient monitor collects real-time 240 Hz waveforms and 0.5 Hz trend data, which are transferred through the secure intranet to a dedicated BedMaster server (Excel Medical Electronics) and archived [6]. To increase the system's availability and reliability, a triple-redundant design was used, in which 3 BedMaster servers were used in parallel to collect data from all bed units [7]. Physiological data collected through this system, when they are displayed on the GE Marquette monitor, include electrocardiographic, photoplethysmographic, carbon dioxide, arterial blood pressure, and intracranial pressure (ICP), among others. Trends include heart rate (HR), respiratory rate, temperature, oxygen saturation, end-tidal carbon dioxide, and ICP, among many others. This information provides continuous VS data that relays important physiological information regarding brain perfusion, cardiac stability, overall tissue perfusion, and respiratory status.

During the design of the VS viewer for ICU, our goal was to create a novel physiological data displayer that can reduce ICU clinicians' workload, enhance clinical decision-making, and improve communication in a noisy and confined ICU environment. To achieve the goal, we considered the factors of usability and patient safety, which can be closely related in this application. For usability, current bedside monitors often suffer from insufficient time windows to display physiological trends, a lack of clear indications of patients' physiological status, and a lack of overview of multiple patients for prioritizing [4]. To enhance the clinicians' efficiency while maximizing patients' safety, we adopted the following design strategies: First, the viewer should reduce the information overload for clinicians to access patients' physiological data, current or past, individual or group [8-10]. Second, it should be compatible with the existing patient monitor system so that clinicians can reuse their existing knowledge about the monitor, which may increase the acceptance of the VS viewer [8]. Third, in the user experience design, the viewer should place the user in control [11]. It should use simple colors and graphs to convey efficient information while still providing detailed data for advanced users to access with simple operations [12]. Fourth, the viewer should have reasonable reliability for patients' safety. Redundance was introduced in the design for key components in the system, such as the data collection, database, and web server [7].

The VS viewer adopted a client-server architecture. The server handles 2 types of clients: the bedside monitors and the users. It receives and persists in real-time physiologic data that are transmitted from the bedside monitors. A database records each bedside VS value, bed name, and timestamp. The server also responds to users' requests for viewing data within a given time frame. To continuously present the latest data to the user with low latency, the VS viewer uses the asynchronous Javascript and XML technique to pull the most recent data from the database every minute [13]. Such a method allows the VS viewer to automatically redraw all VS trajectories without refreshing the entire viewing page.

The VS viewer provides a rich interface for data monitoring, exploration, and recording. Data are depicted according to each clinical area of operations, such as the trauma resuscitation unit or emergency department, operating room, computed tomography suite, and individual critical care units. Figure 1 demonstrates the grouping of bed units. On the left panel, a list of all groups can be used as a shortcut to bed units. Selecting a specific unit, a default 24-hour view is displaced for shock index (SI=HR/systolic blood pressure), HR, systolic blood pressure, ICP, cerebral perfusion pressure, brain trauma index, and end-tidal carbon dioxide concentration. If ICP data are not collected, the space is used to plot the next available VS, optimizing the view.

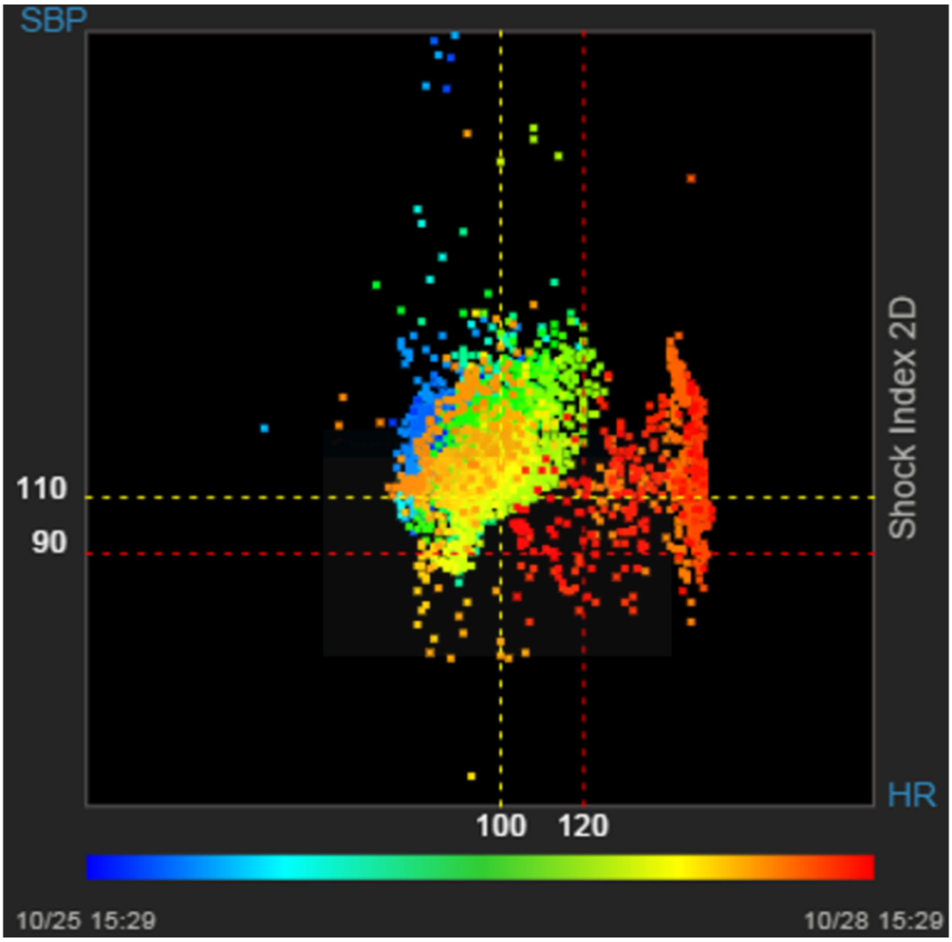
When a bed is selected, a page for this bed (unit view) is displayed. Figure 2 demonstrates the structure of the information. The page is partitioned into multiple areas for navigation, viewing, and tools. Its center is assigned for presenting the selected patient's physiologic data in a time frame (up to 72 hours). VS trajectories are stacked vertically in order of predefined importance. The bottom is reserved for plotting bar segments of all VS that summarize the colored warnings without showing the value changes. This provides a summary of all available VS trends in a condensed space, which could be used to view the physiological stability of the patient over time. To provide an at-a-glance view of other rooms in this group, the left panel lists all the rooms in the current group and updates their VS trajectories in real time. The color-coded warning in the thumbnails enhances situation awareness even when the users are focusing on 1 patient.

The VS viewer has additional diagnostic tools. For example, SI is a commonly used blood transfusion diagnosis tool [14]. The VS viewer adds a 2D SI diagram to show a changing trajectory (Figure 3). To present the temporal information, a heat map is plotted, ranging from blue (cold) to red (warm); blue colors represent past events, whereas red colors represent current data trends. Similarly, the brain trauma index (which is ICP or cerebral perfusion pressure) can also be visualized in the 2D plot [15].

Figure 1. The vital sign viewer in the "group" mode, with a default 24-hour display.

Figure 2. Vital sign (VS) viewer in the “unit” mode, with default 24-hour display. Labeled area 1: navigation menu to other room groups. Area 2: title information for room name, current time, and the next update time. Area 3: user portal. Area 4: list of beds in the same group with their current VS thumbnails. Area 5: the main area to display selected room VS trajectories and the summarization with color-coded patterns. Area 6: diagnostic tools for 2D scatter plots of shock index (=heart rate/systolic blood pressure) and brain trauma index (=intracranial pressure/cerebral perfusion pressure). Area 7: functional buttons for selecting various time ranges for viewing.

Figure 3. An example 2D shock index plot. The colored scatter plot shows the change in shock index (heart rate/systolic blood pressure) from past (blue) to recent (red), thereby depicting a 3-day change in worsening shock index.



Clinical Thresholds

Colored warnings are an effective means to gain a clinician's attention and may be more effective than audible alarms, especially in a noisy, busy, and confined environment [16]. In the VS viewer, VS trajectories with colors may be viewed to highlight the sections where the VS are outside of normal clinical thresholds. For example, too low or too high HR segments are displayed differently from normal HR. Clinical thresholds for VSs were developed after surveying 47 clinicians (24 medical doctors, 18 registered nurses, and 5 respiratory therapists). Among them, 36 clinicians were from the University of Maryland, Baltimore, and 11 from the University of Cincinnati. After the survey was completed, a team of clinicians met to review the results to reach a consensus on the viewer's opinion of their visual appearance. Multimedia Appendix 1 summarizes the optional threshold distributions for some important VS. Based on these threshold values, a consensus set of color-coded cutoffs was determined (Multimedia Appendix 2). These values were set as fixed parameters under consideration of a simplified and consistent user interface.

Survey Design

Clinicians who were scheduled to work in the ICU or on the trauma teams were contacted and trained on how to use the VS viewer. Once trained, ICU and team clinicians were asked to participate in the study. Clinicians were surveyed anonymously from Tuesday to Friday and were asked to conduct rounds normally, using data reported from nurse charts and briefs from fellows to inform their clinical decisions. None of those clinicians participated in the design of the VS viewer. A total of 2 questionnaires were designed to collect clinicians' opinions about a patient's condition and satisfaction with the VS viewer. Clinicians were given a preview survey upon their assessment and formulation of their plan for each patient after traditional rounds and before accessing the viewer. Immediately following the completion of the pre-view survey, the VS viewer was presented to the clinicians on a tablet, displaying the patient's past physiologic data visualized and summarized for up to 72 hours. After reviewing the viewer for up to 1 minute, clinicians completed the postview questionnaire. In both questionnaires, the clinicians answered questions regarding the patient's stability, status, and need for a higher or lower level of care. In the post-view questionnaire, clinicians were also asked if they intentionally planned to implement any of the following interventions after seeing the viewer: (1) changing any current medications, (2) ordering additional medications, (3) ordering additional diagnostic tests, (4) changing ventilation settings, (5) ordering additional labs, (6) physically reexamining this patient, (7) providing fluid bolus, or (8) providing a blood transfusion.

Statistical Methods

A participant's perceiving of the VS viewer's usefulness is represented by a vector consisting of the percentage of the 5 categories (strongly agree, agree, neutral, disagree, and strongly disagree) that he or she assigned to the question "the viewer enhanced my understanding of the patient's condition." We used the Ward method, a hierarchical clustering method, with Manhattan distance to group the participants based on their ratings to the question "the viewer enhanced my understanding of the patient's condition" [17,18]. Between those clusters, we compared the participants' opinion changes on the patients' conditions in 7 questions (Table 1) before and after using the viewer. The chi-square test was used to compare percentage differences.

Table 1. The number of opinion changes for 7 questions (Q1-Q7) before and after seeing the viewer, with respect to the 5 clustered user types.

Questions[a]	Total changes, n (%)	Unique participants, n (%)	C1, n	C2, n	C3, n	C4, n	C5, n	Like (C1 and C2), n (%)	Dislike (C3, C4, and C5), n (%)
Q1	129 (14.2)	16 (66.7)	46	31	10	42	0	77 (59.7)	52 (40.3)
Q2	112 (12.3)	15 (62.5)	38	34	8	32	0	72 (64.3)	40 (35.7)
Q3-6	145 (16)	18 (75)	58	54	3	30	0	112 (77.2)	33 (22.8)
Q7	92 (10.1)	17 (70.8)	20	32	9	31	0	52 (56.5)	40 (43.5)

[a]Please refer to Textbox 1 for the question.

Textbox 1. Questions.

- Q1: Having reviewed the last 24 hours of information during rounds and before and after seeing the 24-hour viewer, do they feel that in the past 24 hours the patient has shown evidence of (a) infection, (b) hemodynamic instability, (c) uncontrolled bleeding, or (d) respiratory deterioration?
- Q2: Over the past 24 hours, has the patient's condition (a) improved significantly, (b) improved slightly, (c) unchanged, (d) deteriorated slightly, or (e) deteriorated significantly?
- Q3: Can the patient be transferred to a lower level of care?
- Q4: Can the patient be transferred to a higher level of care?
- Q5: Does the patient have a traumatic brain injury?
- Q6: Did the patient have intracranial pressure problems in the past 24 hours?
- Q7: Due to the viewer, do they plan for any changes in interventions, including (a) changing any current medications, (b) ordering additional medications, (c) ordering additional diagnostic tests, (d) changing ventilation settings, (e) ordering additional labs, (f) physically reexamining this patient, (g) providing a fluid bolus, or (h) providing a blood transfusion?

Note: These are the questions referenced in Table 1.

Ethical Considerations

The study has been approved by the institutional review board of the University of Maryland School of Medicine (HP-00063086).

Results

Survey Collection

From February 2017 to June 2017, the survey team followed clinicians who agreed to take the surveys. A total of 908 surveys were collected from 24 participants with unbalanced proportions. Among the 908 rounds, 48 (5%) were patients who were newly admitted, and 860 (95%) were not. When asked if the VS viewer enhanced their understanding of the patient's condition, clinicians strongly agreed 45 (5%) times, agreed 514 (56.6%) times, were neutral 321 (35.4%) times, disagreed 26 (2.9%) times, and strongly disagreed 2 (0.2%) times. Figure 4 lists the total surveys each participant contributed and the proportions of ratings on whether the viewer enhanced his or her understanding of the patient's condition during a round.

Results show that physicians' clinical assessments and plans could be influenced by viewing the VS viewer for 1 minute or less, indicated by a "yes" answer to at least 1 of the 8 questions (Q7 in the survey). Of the 908 rounds, a total of 92 (10.1%) rounds had at least 1 "yes" as planning on some changes to the interventions. The most common change was (Q1) changing current medications (36/908, 4%). The next most common changes were (Q6) physically reexamining the patient (31/908, 3.4%), (Q2) ordering additional medications (20/908, 2.2%), and (Q7) providing a fluid bolus (20/908, 2.2%).

We used the Ward method with Manhattan distance to group the participants based on their ratings to the question "the viewer enhanced my understanding of the patient's condition" [17]. For example, 1 participant contributed 62 surveys and rated 2 "strongly agree," 22 "agree," 32 "neutral," 5 "disagree," and 1 "strongly disagree." The vector of percentages (0.03, 0.35, 0.52, 0.08, and 0.02) represents the overall rating that this participant had about the viewer. The 24 participants were clustered into 5 groups, as shown in Figure 5. The 5 groups correspond to the participants who are mostly in favor (C1) of the viewer to those least in favor (C5). There are 6 in C1, 6 in C2, 3 in C3, 7 in C4, and 2 in C5, which shows a very balanced grouping, with half of the participants in the C1 and C2 groups and the other half in the other 3 clusters. This shows that the sampled rounds were done by participants with almost similar proportions of different attitudes toward the viewer. In other words, the survey team sampled the rounds randomly enough so that the collected data were not biased by participants with certain preexisting feelings about the viewer.

Figure 4. Distribution of each participant's rating on if the viewer enhanced his or her understanding of the patient's condition during a round.

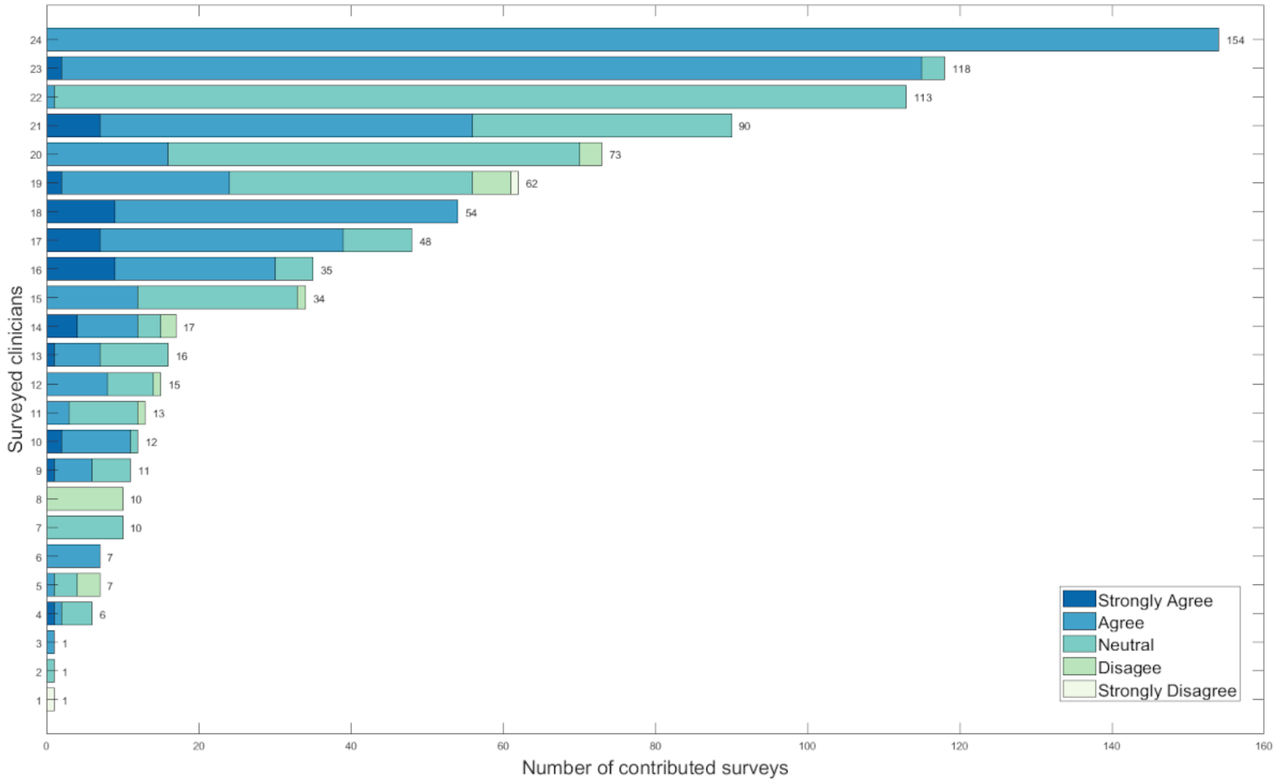
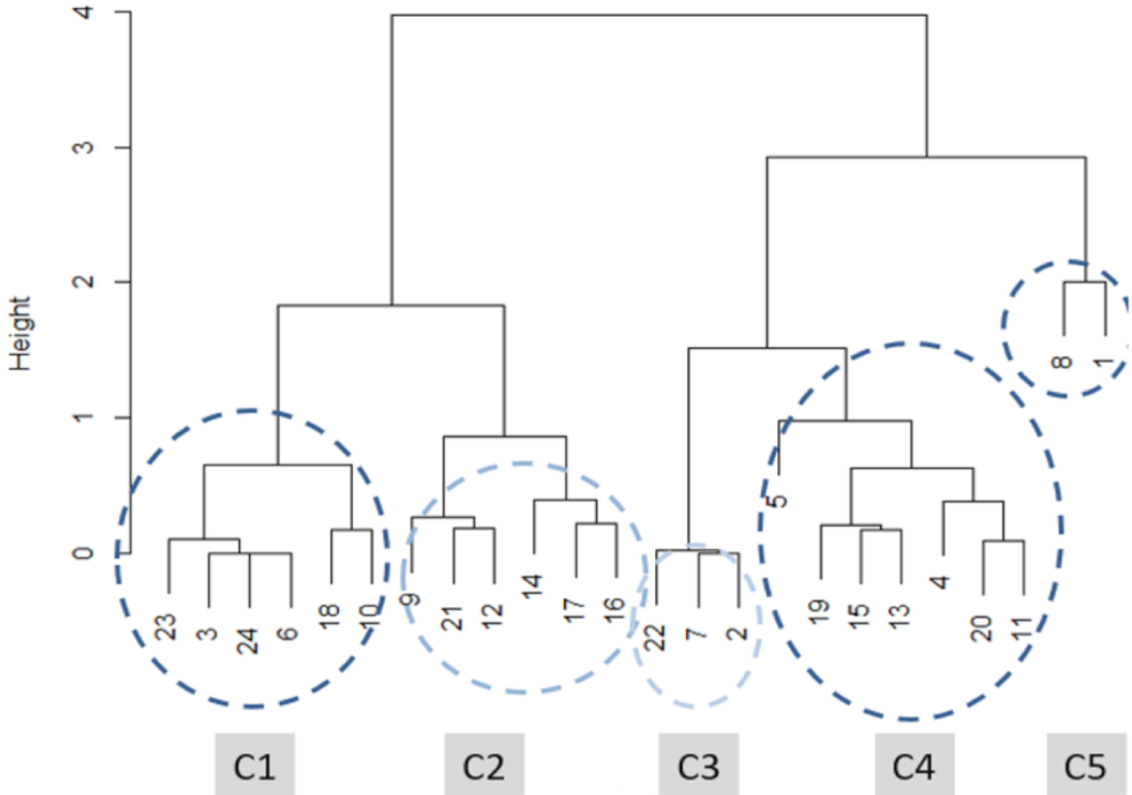


Figure 5. Clusters of participants with similar feelings about the viewer. In total, 24 unique participants are grouped into 5 groups, corresponding to "strongly favor," "favor," "neutral," "dislike," and "strongly dislike.".



Comparisons

We analyzed the opinion changes before and after seeing the viewer, regarding the patient's stability, status, and need for a higher or lower level of care. Instead of summarizing the total changes in opinions, we compared them with respect to the clusters of user types. The participants who were "neutral" (C3) or "strongly dislike" (C5) had low numbers of opinion changes for all 7 questions. Those who were in clusters C1, C2, and C3 had more numbers of opinion changes (Table 1). For simplicity, we can further group the participants into 2 types: those who liked the VS viewer (C1 and C2) and those who disliked it (C3, C4, and C5). The clinicians who liked the VS viewer had a higher rate of changed opinions than those who disliked the VS viewer regarding Q1 to Q6 (Q1: 59.7% vs 40.3%, Q2: 64.3% vs 35.7%, and Q3-6: 77.2% vs 22.8%). When asked if they planned for any changes for interventions (Q7), there was no significant difference between the 2 major groups of clinicians (56.5% vs 43.5%, P=.10).

Discussion

Principal Results

With the development of sensor and computing technologies, vast amounts of high-quality, continuous electronic data, including VS, alarms, and clinical interventions, are collected at the bedside. Those data have the potential to provide an unprecedented view of dynamic physiologic responses to injury, illness, and treatments. Therefore, data gathered from bedsides could assist clinicians in care planning and decision support. However, massive amounts of data that are not well organized or presented still create a barrier for clinicians making full use of them in a busy resuscitation or intensive care environment. Bedside monitors often only display instantaneous readings or a short strip of recent physiologic VS for diagnosis. Clinicians need to rely on separate nursing charts, handwritten or electronic, to review a patient's developing conditions. The VS viewer, which automates physiological data by displaying clear color-coded trends, presents aggregated information from multiple data sources, provides at-a-glance summaries of clinical data, and assists with the prioritization of care for multiple patients.

The use of the VS viewer was subjectively assessed with 908 observations from clinicians working in ICUs at a high-volume level 1 trauma center. Clinicians generally perceived the use of the VS viewer favorably, as evidenced by survey data. The VS viewer was originally developed for the United States Air Force Critical Care Air Transport Teams [19,20]. Critical Care Air Transport Teams transport up to 3 patients who are critically ill in the back of the aircraft, allowing trauma surgeons to perform far-forward damage control surgery, knowing that these patients could be quickly transported rearward with full support. This rapid transport of complex patients with multisystem trauma, shock, burns, and respiratory failure who are in hemodynamic flux requires continual resuscitation, stabilization, advanced care, and life-saving interventions during air transport; however, currently available advanced ICU monitoring systems suitable for the needs of such patients were developed for use in stable, hospital-based settings, not in the crowded, noisy, vibrating, and sometimes frankly jolting environment of air evacuation or long-distance air transport. The noise levels, confined space, limited access to patients, vibration, and overall limited patient visibility make using a VS viewer advantageous in such a setting. Such technology can also be valuable in enhancing emergency medical personnel's decision-making for initial triage. While traditional VS are useful in guiding prehospital care and triage, they represent isolated points in time, and trends and fluctuations in vitals may not be apparent.

In this study, we set the clinical thresholds for colored warnings to be uniform across all beds. This was to make the user interface simplified and more consistent during a survey. Additionally, a set of predefined thresholds from a group of experienced clinicians could be a useful out-of-the-box feature when the VS viewer is deployed in the field. That said, the clinical thresholds could be personalized for each bed. For example, if the bedside monitor allows alarm threshold settings, such settings could be used as the colored warning thresholds in the VS viewer for each bed.

The VS viewer has expanded from ICUs to trauma resuscitation units, operating rooms, neuro ICUs, and pediatric ICUs at the University of Maryland Medical Center. In 2020, during the COVID-19 pandemic, it was deployed to monitor 150 beds in biocontaminated units to reduce the risk of infection and improve efficiency for clinicians in treating their patients.

Innovations

The VS viewer is a multipatient physiological monitor. To the best of our knowledge, we could not find any articles that describe a viewer system with a similar design. In a comprehensive review by Waller et al [5], a total of 17 information displays in ICU settings were designed for specific disease states or body systems, such as cerebral perfusion monitoring for individual patients or monitoring for arterial blood gas trends. The novel user interface presented in this study was designed with the aim of conveying information more efficiently to ICU clinicians in a noisy, confined, and busy environment. It uses color-coded warnings to indicate a patient's status and highlight data that needs attention. The side panel provides a peek at the physiological status of other patients, which can help clinicians keep an eye on other patients even if their attention is focused on a single patient. It uses advanced web front-end techniques to hide large quantities of data behind simple line charts and reveal them when needed.

Clinical Impact

The use of the VS viewer can have several possible influences on clinical assessment and plans. It can help clinicians quickly recognize critical changes in the patient's physiologic status and provide early interventions to prevent further deterioration. The VS viewer can potentially improve patient outcomes by providing clinicians with a concise overview of key information, reducing cognitive load and errors, and improving compliance with evidence-based safety guidelines [12]. It may also help to improve communication efficiency within the ICU team by providing easy access to a shared platform of patient longitudinal data. It can reduce the workload of the ICU team by automating routine tasks such as extracting data from nursing charts.

To prioritize care in high-volume ICUs, intensive care clinicians must be able to rapidly identify physiological events and the need for intervention. The VS viewer can help organize a large amount of data in a busy, noisy ICU environment where close monitoring of patients who are critically ill is essential to detect potentially harmful physiological trends. The presentation of data with temporal, color-coded patterns, and the ability of the

VS viewer to provide at-a-glance data for entire units is advantageous for clinicians working in high-volume ICUs.

The color-coded patterns may reduce the "alarm fatigue" issue in noisy ICUs. The noise burden is common in modern physiologic monitoring systems and has been recognized as a critical patient safety concern in the hospital care setting [21-23]. In noisy environments, such as ICUs, helicopter transportation, or aeromedical evacuation, loud and continuous alarms could reduce their specificity in getting clinicians' responses. Another issue with audible alarms is that they are transient and cannot be replayed once they are gone. While the visual alert patterns could show the longitudinal patterns of physiologic change.

Related Work

The VS viewer with organized and easy-access information could be part of the effort to build the smart ICU or the tele-ICU. The concepts of smart ICU and tele-ICU aim to maximize the use of bedside clinical expertise in assessing and treating patients by providing integrated monitoring and actionable information [24-26]. A survey study of 86 ICU staff in a German university hospital summarized that health providers expect ICU monitoring could be improved by reducing false alarms, using wireless sensors and mobile devices, preparing for the use of AI, and enhancing the digital literacy of ICU staff [27,28]. The VS viewer could be used in both centralized and decentralized architectures of tele-ICU for extending coverage and facilitating patient transfer between hospitals because of its flexible configuration of grouping ICU beds virtually [29]. By making essential clinical information available remotely, the VS viewer allows clinicians to provide care plans when on-site support is infeasible or limited [30,31]. It may potentially reduce exposure to contagious diseases and, hence, increase patient safety.

With continuous physiologic data and other clinical information, the VS viewer has the ability to process real-time data into predictive algorithms, which is also desired for tele-ICU [30]. Beyond being a plain displayer, the VS viewer could embed risk-prediction algorithms that use continuous VS as inputs and may promote more efficient interventions to reduce ICU risk [31]. For example, ICU mortality prediction [32,33], secondary insults after severe brain traumatic injury [34], needs for transfusion [35,36], and neurologic decline in the ICU [37] are reported to have good predictive performances by using variables derived from continuous VS. We have also shown that using risk scores calculated from continuously measured VS, patients requiring endovascular resuscitative interventions can be identified with high accuracy [38]. Moreover, the VS viewer could serve as a platform for predictive model diagnosis by providing clinicians with explainable artificial intelligence [39]. With patient VS data, we can use the Shapley Additive exPlanations algorithm to calculate each variable's contribution to the prediction result [40]. Therefore, the clinicians would know not only the prediction but also the contribution of each variable to the prediction. Such information may help clinicians make more personalized care plans.

Limitations

There are limitations to this work that are worth noting. We collected data from a large number of ICU clinicians compared to trauma team clinicians. Trauma team clinicians are surgeons responsible for the same patient throughout the entire length of stay, regardless of the acuity of the patient. ICU clinicians are intensivists and are only responsible for patients in the ICU. Hence, disparities between both groups of clinicians are inevitable, as each group has different clinical perspectives and patient workloads. As occurs in nearly all survey work, response rates and receptiveness to the surveys varied. Some clinicians were more amenable to being surveyed compared to others. In the collected forms, there were more surveys from some clinicians than from others. To reduce this potential bias, we clustered the participants based on their overall rating on each round, from which we estimated each participant's a priori attitude toward using this viewer. The results show that there was a balanced "favoring" and "non-favoring" of using this viewer.

We only evaluated the viewer based on clinicians' satisfaction and efficiency (potential changes in interventions before and after seeing the viewer). In future studies, randomized controlled trials can be designed to analyze the viewer's impact on patients' outcomes and safety [12].

Conclusions

We designed, implemented, and evaluated an automated physiologic data organizer and visualization platform. It provides at-a-glance summaries and assists with prioritizing care for multiple patients. The VS viewer demonstrates a method to assemble large quantities of data from multiple sources and represents trends in each patient's condition with simple color codes, greatly improving situational awareness. It has the potential to be used in en route care, hospitals with multiple branches, and understaffed hospitals in remote areas. The survey shows that organized physiologic data and visual assessment could assist clinicians in recognizing changes in patient status and prioritizing care.

Acknowledgments

This work was partially funded by USAF 8650-14-2-6D19 and W81XWH-17-C-0034.

The authors thank the VS viewer study group: Colin Mackenzie, MD, Sarah Wade, BS, Lauren Hartsky, BS, Thomas Grissom, MD, Napoleon Roux, MD, Doug Floccare, MD, Ashton Engdahl, BS, Scott Murray, BS, Rosemary Kozar, MD, Sarah Murthi, MD, Laura Buchanan, MD, Neal Reynolds, MD, Gunjan Parikh, MD, Megan Brenner, MD, without whose help the data collection for this study would not have been possible.

Conflicts of Interest

PH, DS, Colin Mackenzie (part of the VS viewer study group), TS, and SY have US Patent Application 17/676,657 filed on February 21, 2022, titled “Method and Apparatus for Monitoring Collection of Physiological Patient Data.”

Multimedia Appendix 1

Surveyed thresholds for heart rate, systolic and diastolic blood pressure, blood oxygen saturation, and temperature.
[DOCX File , 20 KB - humanfactors_v11i1e46030_app1.docx]

Multimedia Appendix 2

VS Viewer color coding threshold values.
[DOCX File , 18 KB - humanfactors_v11i1e46030_app2.docx]

References

1. Endsley MR. Designing for Situation Awareness: An Approach to User-Centered Design, 2nd Edition. United Kingdom: CRC Press; 2016.
2. Yang S, Njoku M, Mackenzie CF. 'Big data' approaches to trauma outcome prediction and autonomous resuscitation. Br J Hosp Med (Lond) 2014 Nov;75(11):637-641. [doi: 10.12968/hmed.2014.75.11.637] [Medline: 25383434]
3. Drew BJ, Harris P, Zègre-Hemsey JK, Mammone T, Schindler D, Salas-Boni R, et al. Insights into the problem of alarm fatigue with physiologic monitor devices: a comprehensive observational study of consecutive intensive care unit patients. PLoS One 2014;9(10):e110274 [FREE Full text] [doi: 10.1371/journal.pone.0110274] [Medline: 25338067]
4. Sebastian K, Sari V, Loy LY, Zhang F, Zhang Z, Feng M. Multi-signal Visualization of Physiology (MVP): A novel visualization dashboard for physiological monitoring of Traumatic Brain Injury patients. : IEEE; 2012 Presented at: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 28 August 2012-01 September 2012; San Diego, CA, USA p. 2000-2003 URL: https://ieeexplore.ieee.org/document/6346349 [doi: 10.1109/embc.2012.6346349]
5. Waller RG, Wright MC, Segall N, Nesbitt P, Reese T, Borbolla D, et al. Novel displays of patient information in critical care settings: a systematic review. J Am Med Inform Assoc 2019 May 01;26(5):479-489 [FREE Full text] [doi: 10.1093/jamia/ocy193] [Medline: 30865769]
6. Excel Medical Electronics, LLC. BedMasterEx Operator's Manual. Jupiter, FL: Excel Medical Electronics, LLC; 2013.
7. Hu PF, Yang S, Li HC, Stansbury LG, Yang F, Hagegeorge G, et al. Reliable collection of real-time patient physiologic data from less reliable networks: a "Monitor of Monitors" system (MoMs). J Med Syst 2017 Jan;41(1):3. [doi: 10.1007/s10916-016-0648-5] [Medline: 27817131]
8. Dziadzko MA, Herasevich V, Sen A, Pickering BW, Knight AMA, Moreno Franco P. User perception and experience of the introduction of a novel critical care patient viewer in the ICU setting. Int J Med Inform 2016 Apr;88:86-91. [doi: 10.1016/j.ijmedinf.2016.01.011] [Medline: 26878767]
9. Pickering BW, Herasevich V, Ahmed A, Gajic O. Novel Representation of Clinical Information in the ICU: Developing User Interfaces which Reduce Information Overload. Appl Clin Inform 2010;1(2):116-131 [FREE Full text] [doi: 10.4338/ACI-2009-12-CR-0027] [Medline: 23616831]
10. Ahmed A, Chandra S, Herasevich V, Gajic O, Pickering BW. The effect of two different electronic health record user interfaces on intensive care provider task load, errors of cognition, and performance. Crit Care Med 2011 Jul;39(7):1626-1634. [doi: 10.1097/CCM.0b013e31821858a0] [Medline: 21478739]
11. Roger SP, Bruce RM. Software Engineering: A Practitioner's Approach. United States: McGraw-Hill Education; 2015.
12. Khairat SS, Dukkipati A, Lauria HA, Bice T, Travers D, Carson SS. The Impact of Visualization Dashboards on Quality of Care and Clinician Satisfaction: Integrative Literature Review. JMIR Hum Factors 2018 May 31;5(2):e22 [FREE Full text] [doi: 10.2196/humanfactors.9328] [Medline: 29853440]
13. Garrett JJ. Ajax: a new approach to web applications. AdaptivePath.com. 2005. URL: https://designftw.mit.edu/lectures/apis/ajax_adaptive_path.pdf [accessed 2022-09-28]
14. King RW, Plewa MC, Buderer NM, Knotts FB. Shock index as a marker for significant injury in trauma patients. Acad Emerg Med 1996 Nov;3(11):1041-1045 [FREE Full text] [doi: 10.1111/j.1553-2712.1996.tb03351.x] [Medline: 8922013]
15. Kahraman S, Hu P, Stein DM, Stansbury LG, Dutton RP, Xiao Y, et al. Dynamic three-dimensional scoring of cerebral perfusion pressure and intracranial pressure provides a brain trauma index that predicts outcome in patients with severe traumatic brain injury. J Trauma 2011 Mar;70(3):547-553. [doi: 10.1097/TA.0b013e31820c768a] [Medline: 21610341]
16. Laughery KR, Wogalter MS. Designing effective warnings. Rev Hum Factors Ergonomics 2016 Nov 03;2(1):241-271. [doi: 10.1177/1557234x0600200109]
17. Murtagh F, Legendre P. Ward’s hierarchical agglomerative clustering method: which algorithms implement Ward’s criterion? J Classif 2014 Oct 18;31(3):274-295. [doi: 10.1007/s00357-014-9161-z]
18. Strauss T, von Maltitz MJ. Generalising Ward's method for use with Manhattan distances. PLoS One 2017;12(1):e0168288 [FREE Full text] [doi: 10.1371/journal.pone.0168288] [Medline: 28085891]

19. Lairet J, King J, Vojta L, Beninati W. Short-term outcomes of US Air Force Critical Care Air Transport Team (CCATT) patients evacuated from a combat setting. Prehosp Emerg Care 2013;17(4):486-490. [doi: 10.3109/10903127.2013.811564] [Medline: 23865806]
20. Ingalls N, Zonies D, Bailey JA, Martin KD, Iddins BO, Carlton PK, et al. A review of the first 10 years of critical care aeromedical transport during operation iraqi freedom and operation enduring freedom: the importance of evacuation timing. JAMA Surg 2014 Aug;149(8):807-813. [doi: 10.1001/jamasurg.2014.621] [Medline: 25074327]
21. Cvach M. Monitor alarm fatigue: an integrative review. Biomed Instrum Technol 2012;46(4):268-277. [doi: 10.2345/0899-8205-46.4.268] [Medline: 22839984]
22. Joint Commission. Medical device alarm safety in hospitals. Sentinel Event Alert 2013 Apr 08(50):1-3. [Medline: 23767076]
23. Oliveira AECD, Machado AB, Santos EDD, Almeida ÉBD. Alarm fatigue and the implications for patient safety. Rev Bras Enferm 2018;71(6):3035-3040 [FREE Full text] [doi: 10.1590/0034-7167-2017-0481] [Medline: 30517409]
24. Halpern NA. Innovative designs for the smart ICU. Chest 2014 Mar;145(3):646-658. [doi: 10.1378/chest.13-0004] [Medline: 27845639]
25. Halpern NA. Innovative designs for the smart ICU: part 3: advanced ICU informatics. Chest 2014 Apr;145(4):903-912. [doi: 10.1378/chest.13-0005] [Medline: 24687712]
26. Avdalovic MV, Marcin JP. When will telemedicine appear in the ICU? J Intensive Care Med 2019 Apr;34(4):271-276. [doi: 10.1177/0885066618775956] [Medline: 29756527]
27. Poncette AS, Spies C, Mosch L, Schieler M, Weber-Carstens S, Krampe H, et al. Clinical requirements of future patient monitoring in the intensive care unit: qualitative study. JMIR Med Inform 2019 Apr 30;7(2):e13064 [FREE Full text] [doi: 10.2196/13064] [Medline: 31038467]
28. Poncette AS, Mosch L, Spies C, Schmieding M, Schiefenhövel F, Krampe H, et al. Improvements in patient monitoring in the intensive care unit: survey study. J Med Internet Res 2020 Jun 19;22(6):e19091 [FREE Full text] [doi: 10.2196/19091] [Medline: 32459655]
29. Guinemer C, Boeker M, Fürstenau D, Poncette AS, Weiss B, Mörgeli R, et al. Telemedicine in intensive care units: scoping review. J Med Internet Res 2021 Nov 03;23(11):e32264 [FREE Full text] [doi: 10.2196/32264] [Medline: 34730547]
30. Hassan E. Tele-ICU and patient safety considerations. Crit Care Nurs Q 2018;41(1):47-59. [doi: 10.1097/CNQ.0000000000000185] [Medline: 29210766]
31. Udeh C, Udeh B, Rahman N, Canfield C, Campbell J, Hata JS. Telemedicine/virtual ICU: where are we and where are we going? Methodist Debakey Cardiovasc J 2018;14(2):126-133 [FREE Full text] [doi: 10.14797/mdcj-14-2-126] [Medline: 29977469]
32. Baker S, Xiang W, Atkinson I. Continuous and automatic mortality risk prediction using vital signs in the intensive care unit: a hybrid neural network approach. Sci Rep 2020 Dec 04;10(1):21282. [doi: 10.1038/s41598-020-78184-7] [Medline: 33277530]
33. Kristinsson Æ, Gu Y, Rasmussen SM, Mølgaard J, Haahr-Raunkjær C, Meyhoff CS, et al. Prediction of serious outcomes based on continuous vital sign monitoring of high-risk patients. Comput Biol Med 2022 Aug;147:105559 [FREE Full text] [doi: 10.1016/j.compbiomed.2022.105559] [Medline: 35635901]
34. Bonds BW, Yang S, Hu PF, Kalpakis K, Stansbury LG, Scalea TM, et al. Predicting secondary insults after severe traumatic brain injury. J Trauma Acute Care Surg 2015 Jul;79(1):85-90; discussion 90. [doi: 10.1097/TA.0000000000000698] [Medline: 26091319]
35. Parimi N, Hu PF, Mackenzie CF, Yang S, Bartlett ST, Scalea TM, et al. Automated continuous vital signs predict use of uncrossed matched blood and massive transfusion following trauma. J Trauma Acute Care Surg 2016 Jun;80(6):897-906. [doi: 10.1097/TA.0000000000001047] [Medline: 27027555]
36. Yang S, Mackenzie CF, Rock P, Lin C, Floccare D, Scalea T, et al. Comparison of massive and emergency transfusion prediction scoring systems after trauma with a new Bleeding Risk Index score applied in-flight. J Trauma Acute Care Surg 2021 Feb 01;90(2):268-273. [doi: 10.1097/TA.0000000000003031] [Medline: 33502145]
37. Podell J, Yang S, Miller S, Felix R, Tripathi H, Parikh G, et al. Rapid prediction of secondary neurologic decline after traumatic brain injury: a data analytic approach. Sci Rep 2023 Jan 09;13(1):403 [FREE Full text] [doi: 10.1038/s41598-022-26318-4] [Medline: 36624110]
38. Zeineddin A, Hu P, Yang S, Floccare D, Lin CY, Scalea TM, et al. Prehospital continuous vital signs predict need for resuscitative endovascular balloon occlusion of the aorta and resuscitative thoracotomy prehospital continuous vital signs predict resuscitative endovascular balloon occlusion of the aorta. J Trauma Acute Care Surg 2021 Nov 01;91(5):798-802. [doi: 10.1097/TA.0000000000003171] [Medline: 33797486]
39. Maul J, Straub J. Assessment of the Use of Patient Vital Sign Data for Preventing Misidentification and Medical Errors. Healthcare (Basel) 2022 Dec 02;10(12):1-87 [FREE Full text] [doi: 10.3390/healthcare10122440] [Medline: 36553964]
40. Lundberg S, Lee SI. A unified approach to interpreting model predictions. ArXiv ;30:1-10 Preprint posted online on 25 Nov 2017. [FREE Full text]

Abbreviations

HR: heart rate
ICP: intracranial pressure
ICU: intensive care unit
SI: shock index
VS: vital signs

Edited by A Kushniruk; submitted 27.02.23; peer-reviewed by AS Poncette, J Walsh, K Gupta; comments to author 14.06.23; revised version received 03.11.23; accepted 20.11.23; published 05.01.24.

Please cite as:
Yang S, Galvagno S, Badjatia N, Stein D, Teeter W, Scalea T, Shackelford S, Fang R, Miller C, Hu P, VS viewer study group
A Novel Continuous Real-Time Vital Signs Viewer for Intensive Care Units: Design and Evaluation Study
JMIR Hum Factors 2024;11:e46030
URL: https://humanfactors.jmir.org/2024/1/e46030
doi:10.2196/46030
PMID:38180791

Original Paper

Characterizing and Comparing Adverse Drug Events Documented in 2 Spontaneous Reporting Systems in the Lower Mainland of British Columbia, Canada: Retrospective Observational Study

Erica Y Lau[1,2*], PhD; Amber Cragg[2], MSc; Serena S Small[2], MA; Katherine Butcher[3], PharmB; Corinne M Hohl[1*], MSc, MD

[1]Department of Emergency Medicine, University of British Columbia, Vancouver, BC, Canada

[2]Centre for Clinical Epidemiology and Evaluation, Vancouver Coastal Health Research Centre, Vancouver, BC, Canada

[3]Pharmaceutical Science, Vancouver General Hospital, Vancouver Coastal Health, Vancouver, BC, Canada

[*]these authors contributed equally

Corresponding Author:
Erica Y Lau, PhD
Department of Emergency Medicine
University of British Columbia
7th Floor, 828 West 10th Avenue Research Pavilion
Vancouver, BC, V5Z 1M9
Canada
Phone: 1 6048754111 ext 68926
Email: erica.lau@ubc.ca

Abstract

Background: Robust adverse drug event (ADE) reporting systems are crucial to monitor and identify drug safety signals, but the quantity and type of ADEs captured may vary by system characteristics.

Objective: We compared ADEs reported in 2 different reporting systems in the same jurisdictions, the Patient Safety and Learning System–Adverse Drug Reaction (PSLS-ADR) and ActionADE, to understand report variation.

Methods: This retrospective observational study analyzed reports entered into PSLS-ADR and ActionADE systems between December 1, 2019, and December 31, 2022. We conducted a comprehensive analysis including all events from both reporting systems to examine coverage and usage and understand the types of events captured in both systems. We calculated descriptive statistics for reporting facility type, patient demographics, serious events, and most reported drugs. We conducted a subanalysis focused on adverse drug reactions to enable direct comparisons between systems in terms of the volume and events reported. We stratified results by reporting system.

Results: We performed the comprehensive analysis on 3248 ADE reports, of which 12.4% (375/3035) were reported in PSLS-ADR and 87.6% (2660/3035) were reported in ActionADE. Distribution of all events and serious events varied slightly between the 2 systems. Iohexol, gadobutrol, and empagliflozin were the most common culprit drugs (173/375, 46.2%) in PSLS-ADR, while hydrochlorothiazide, apixaban, and ramipril (308/2660, 11.6%) were common in ActionADE. We included 2728 reports in the subanalysis of adverse drug reactions, of which 12.9% (353/2728) were reported in PSLS-ADR and 86.4% (2357/2728) were reported in ActionADE. ActionADE captured 4- to 6-fold more comparable events than PSLS-ADR over this study's period.

Conclusions: User-friendly and robust reporting systems are vital for pharmacovigilance and patient safety. This study highlights substantial differences in ADE data that were generated by different reporting systems. Understanding system factors that lead to varying reporting patterns can enhance ADE monitoring and should be taken into account when evaluating drug safety signals.

(JMIR Hum Factors 2024;11:e52495) doi:10.2196/52495

KEYWORDS

adverse drug event reporting systems; side effect; side effects; drug; drugs; pharmacy; pharmacology; pharmacotherapy; pharmaceutic; pharmaceutics; pharmaceuticals; pharmaceutical; medication; medications; patient safety; health information technology; pharmacovigilance; adverse; safety; HIT; information system; information systems; reporting; descriptive statistics; monitoring

Introduction

Over 2 million Canadians visit an emergency department every year because of an adverse drug event (ADE), an unintended and harmful event related to medication use [1,2]. ADEs incur over 700,000 hospital admissions, and cost over CAD $1 billion (USD $7.48 million) in annual health care expenditures across Canada [2,3]. The importance of addressing this issue cannot be overstated: the World Health Organization (WHO) has identified the prevention of ADEs as an urgent global public health priority [4].

In response to this pressing concern, Canada implemented regulations outlined in the Protecting Canadians from Unsafe Drugs Act (Vanessa's Law) which came into full effect on December 16, 2019. This federal legislation mandates prompt reporting of serious adverse drug reactions (ADRs; a subtype of ADEs) and medical device incidents from hospitals to Health Canada within 30 days of documentation [5]. These regulations serve as a safeguard to protect patients and improve drug surveillance.

Postmarketing pharmacovigilance is crucial in the detection, assessment, and prevention of ADEs under real-world conditions [6,7]. Among the various methods used, spontaneous reporting stands out as one of the most widely adopted approaches in pharmacovigilance [8]. When patients or health professionals spontaneously report ADEs, drug safety monitoring agencies evaluate and integrate these reports into databases, enabling ongoing identification of safety signals [7,8]. This method of surveillance captures data from a broad population and allows us to detect drug safety signals that may not have been identified in the randomized trials used for drug licensing and monitor rare ADEs to medications [9].

It is important to recognize, however, that there is considerable variation in ADE reporting systems worldwide in terms of their design, data fields, terminologies [10], and implementations, which may impact the volume and type of ADEs reported [11]. Variation in design also leads to a lack of standardization of reports, which can in turn prohibit interoperability or effective exchange of ADE reports between systems and may prevent comparisons of ADE events, rates, and risk factors across systems [10].

Despite this variation, the diversity of systems may also be a strength. Each system has the potential to complement others, enhancing the overall quantity and quality of ADE data, if variation in design leads to variation in reporting behaviors or the types of reports that can be entered [12]. To leverage this untapped potential, we need to better understand and compare the events collected through diverse reporting systems [13]. Understanding similarities and differences between systems will enable researchers and drug safety monitoring agencies to more effectively use existing data for accurate signal detection, especially for new or rare ADEs, and prioritize the investigation of drug safety signals. This knowledge will also aid stakeholders in optimizing the design and implementation of new reporting systems to enhance ADE data collection and drug safety surveillance and better align systems with their intended purpose [10].

Health Canada, the regulatory authority for postmarketing pharmacovigilance in Canada, oversees the Canada Vigilance Program, collecting reports of suspected ADEs since 1965. Health professionals and consumers can voluntarily submit reports through various channels, including a web-based platform, phone, fax, or mail. Hospitals are required to submit written reports within 30 days, and Health Canada allows them flexibility by permitting the use of existing systems and processes to meet reporting requirements. With Health Canada's approval, hospitals may use a third party, such as a regional health authority or other reporting programs, to submit reports [14].

The province of British Columbia (BC) currently uses 2 approved spontaneous reporting systems that enable hospitals to comply with Vanessa's Law mandates: the BC Patient Safety and Learning System–Adverse Drug Reaction (PSLS-ADR) reporting form and ActionADE. Briefly, PSLS-ADR was developed and implemented as the first province-wide, web-based platform and supports hospitals in meeting the mandatory reporting requirements [14]. ActionADE, implemented later in the timeline, is a research-driven, web-based app that aims to prevent unintentional redispensation of harmful medications by facilitating the sharing of ADE information between providers across health care settings. ADE reporting occurs as a byproduct of enabling safer care provision (Multimedia Appendix 1) [15].

These 2 systems enable a comparison of the quality and quantity of ADE data generated using 2 different designs. Our objective is to describe and compare the ADEs that health care providers documented using PSLS-ADR and ActionADE during the first 3 years following the implementation of Vanessa's Law.

Methods

ADE Reporting Systems

About PSLS-ADR

BC Patient Safety and Learning System (PSLS) is an initiative of the BC Patient Safety Task Force, developed in collaboration with all 6 provincial health authorities and the Health Care Protection Program, which is part of the Risk Management Branch of the Ministry of Finance that insures BC hospitals [16]. BC PSLS is a web-based safety event reporting and management information system designed to support the identification, investigation, and analysis of safety and risk-related events, including safety hazards, near misses, and adverse events [17]. The system underwent a pilot phase in 2007 and was subsequently implemented province-wide in 2008. BC PSLS has been instrumental in promoting patient safety within the health care system in BC [16].

In response to the introduction of Vanessa's Law and in collaboration with Health Canada, BC PSLS launched PSLS-ADR as a new add-on to the existing system in 2014 and released an updated version in 2019 [18,19]. PSLS-ADR is accessible to health care facilities in all health authorities across BC, including acute care hospitals, long-term care facilities, and outpatient clinics. Authorized health care professionals with access to the secure health authority network, including employees, medical staff, paramedics, contractors, students, and volunteers, can submit reports to PSLS-ADR [15]. Once a report is submitted, the system notifies the medication safety officer in the respective health authority to review and respond to the event [20]. The health authorities send eligible reports to Health Canada for Vanessa's Law reporting requirements. Reports are not made available to care providers and not integrated into the electronic medical record. They are only generated for the purposes of pharmacosurveillance (Multimedia Appendix 2).

The PSLS-ADR data fields are based largely on the Canada Vigilance Adverse Reaction Reporting Form, with additional questions enabling medication safety officers, pharmacy representatives, and others to follow-up with reporters or patients, if necessary [20]. The PSLS-ADR reporting form contains 26 required data fields that collect information about the patient, the adverse reaction (eg, seriousness), the suspected health products (types, name, route used, therapy dates, and treatments), and the reporters.

About ActionADE

Previous studies found that 32.5% of ADE cases observed in emergency departments are repeat events [21], often occurring due to the unintentional represcription or redispensation of the same or a same-class medication as one that previously caused harm [22]. This recurrence is attributed to the lack of effective means to communicate and integrate ADE information into clinical workflows. ActionADE, a research-driven initiative, was developed to address this communication gap [23,24].

In collaboration with the Ministry of Health, Vancouver Coastal Health, a technology partner, and health professional organizations and clinicians, our research team developed and piloted ActionADE between 2016 and 2019 using participatory design principles and data standards that were evaluated and subsequently pilot tested to optimize the system's usability [10,11,15,24-28]. In 2020, we began the implementation of ActionADE in 1 hospital (Vancouver General Hospital) and then expanded its use to 6 hospitals operated by Vancouver Coastal Health and Providence Health Care as part of a research initiative. Although providers were encouraged to use ActionADE, they maintain complete autonomy in choosing between the PSLS-ADR and ActionADE systems to meet their needs.

ActionADE is a web-based app that allows providers to document and communicate ADE information, bidirectionally through its integration (or linkage) with BC's central drug database (PharmaNet). ActionADE was accessible to a subset of care providers with an eligible prescriber identification number issued by their respective regulatory college (ie, physicians, pharmacists, and nurse practitioners) [29]. Eligible clinicians submit reports to ActionADE from a designated health authority network, and the data are shared with clinicians within the patient's circle of care via PharmaNet and used to create safety alerts when community pharmacists attempt to redispense culprit or same-class medications. ActionADE complements the PSLS-ADR system by automating ADE reporting to Health Canada (Multimedia Appendix 3).

The ActionADE data fields were developed based on a systematic review of ADE reporting systems worldwide and participatory action research with clinician end users and are compatible with Health Canada's Canada Vigilance Adverse Reaction Reporting Form [10,11,15,27,30]. As ActionADE is integrated with PharmaNet, several fields auto-populate based on the patient's personal health number, including patient's personal and demographic information (ie, name, date of birth, and sex), reporter's information (ie, name, role, and site), and patient's 14-month medication dispensation history. To create a new report, the system auto-populates the patient's information and medication dispensation history, as well as the reporter's information. ActionADE contains 5 required data fields that collect information about the suspect drugs, which is auto generated based on the medication dispensation history or added manually, the ADE type, and details of the event (eg, symptoms or diagnosis, outcome, and certainty; Multimedia Appendix 4).

Study Design

In this retrospective observational study, we analyzed reports documented in PSLS-ADR and ActionADE entered by providers at health care facilities operated by the Vancouver Coastal Health Authority (excluding Providence Health Care, as PSLS-ADR data were unavailable from those facilities) in BC, Canada, between December 1, 2019, to December 31, 2022.

For PSLS-ADR, we included reports documented by authorized health care professionals (eg, employees, medical staff, paramedics, contractors, students, and volunteers) from >120 health care facilities across the province, including hospital, urgent and primary care, long-term care facilities, and community health centers' clinics. For ActionADE, we included reports documented by eligible clinicians from 4 hospitals where ActionADE is implemented: Lions Gate Hospital, Richmond Hospital, UBC (University of British Columbia) Hospital, and Vancouver General Hospital (Multimedia Appendix 5) [31].

We divided this study's period into 4 phases: baseline period, when all hospitals across BC only used PSLS-ADR (December 2019 to February 2020); year 1 (March 2020 to November 2020), when 1 hospital (Vancouver General Hospital) had the option to use ActionADE for piloting purposes while all other sites in BC exclusively used PSLS-ADR; and year 2 (December 2020 to November 2021) and year 3 (December 2021 to December 2022), when the 4 hospitals had the option to use PSLS-ADR or ActionADE and all other sites in BC exclusively used PSLS-ADR (Multimedia Appendix 6).

Data Sources

For this study, we requested ADE reports from PSLS-ADR from the BC PSLS central office and retrieved reports documented in ActionADE during the same period from the ActionADE database. We obtained information about hospital characteristics through the Information Access and Privacy Services at Provincial Health Services Authority, including number of beds, population served, and the number of emergency department visits per year.

Data Extraction

To allow for direct comparisons between the 2 systems, we combined similar variables wherever possible. A clinical pharmacist classified all free-text drug entries from the PSLS-ADR reports into the equivalent generic drug name that would be present if the same report were entered into ActionADE based on the provincial formulary. We translated continuous age from ActionADE into the age categories in PSLS-ADR. We combined information across platforms to produce combined variables for report date, patient demographics (age group and sex), types of ADE (ADRs and nonadverse drug reactions), ADE outcomes (death, emergency visit, hospitalized or hospital extended, life threatening, worsened preexisting condition, permanent disability and fetal defect, other, and unknown) [27], and reporter information (role and facility).

Statistical Analysis

Comprehensive Analysis

First, we performed a comprehensive statistical analysis to provide a global view of coverage, usage, and the types of information captured by both reporting systems. We included all events from both reporting systems, excluding reports related to user errors (eg, duplicate reports), refuted allergies, and reports with incomplete data. We calculated descriptive statistics (eg, means and SDs or frequency and percentages) for the following variables: total number and types of the reporting facilities (hospital vs nonhospital), patient's age group, patient's sex, roles of reporters, proportion of serious events, and the 10 most reported culprit drugs for all events and serious events. We defined serious events based on the Health Canada's definition. This definition includes ADEs that require in-patient hospitalization or prolongation of existing hospitalization, cause congenital malformation, result in persistent or significant disability or incapacity, are life-threatening, or result in death [14].

ADR Analysis

To allow for direct comparisons between the 2 systems, we then conducted a subsample analysis that only included ADR reports (a subtype of ADE) that met Health Canada's definition and that could have been reported in both systems. According to Health Canada, ADRs encompass harmful and unintended responses to a health product, including any undesirable patient effects suspected to be associated with health product use. This definition includes unintended effects, health product abuse, overdoses, interactions (including drug-drug and drug-food interactions), and unusual lack of therapeutic efficacy, all of which are considered reportable adverse reactions [14,32,33]. We included eligible ADR reports from both reporting systems from sites where both systems were available, excluding reports related to user errors (eg, duplicate reports), refuted allergies, and reports with incomplete data. We calculated descriptive statistics for the following variables: patient's age group, patient's sex, proportion of serious events, the 10 most reported culprit drugs, and mean monthly counts of all events and serious events during each phase of this study period, stratified by reporting system. We conducted all analyses using SAS statistical software (version 9.4; SAS Institute).

Ethical Considerations

The UBC (University of British Columbia) clinical research ethics board approved of this research (H18-01332 and H22-00312) and provided a waiver for obtaining informed consent as this study meets the Tri-Council Policy Statement minimal risk criteria.

Results

Comprehensive Analysis

We extracted 3248 reports from both reporting systems. After removing 213 reports related to refuted allergies, erroneous reports, and reports with incomplete data, the analytic cohort for the comprehensive analysis comprised 3035 unique ADEs reported in either system (Figure 1). Of these, 12.4% (375/3035) were entered in PSLS-ADR and 87.6% (2660/3035) were reported in ActionADE.

Approximately 50% of the events occurred in male patients in both PSLS-ADR (178/375) and ActionADE (1285/3035). The highest proportion of events were from patients aged 45-64 years (32.8%, 123/375) in PSLS-ADR and aged 75-84 years (25.3%, 674/2660) in ActionADE. In total, 12 facilities (5 hospitals and 7 nonhospital facilities) entered reports in PSLS-ADR. The primary reporters in PSLS-ADR were medical imaging staff or technicians (170/375, 45.3%) and pharmacists (174/375, 46.4%). Of the 4 hospitals that entered reports in ActionADE, pharmacists were the reporter for 92.1% (2451/3035) of the events. The proportion of serious events was 36% (135/377) in PSLS-ADR and 28.2% (749/3035) in ActionADE (Table 1).

In PSLS, the most common culprit drugs were iohexol, gadobutrol, and empagliflozin, accounting for 46.2% (173/375) of all events. Empagliflozin, ibuprofen, and iohexol represented 11.8% (16/135) of serious events (Tables 2 and 3). Iohexol and gadobutrol are both contrast agents used for diagnostic imaging, whereas empagliflozin is an oral medication primarily prescribed for managing type 2 diabetes mellitus and ibuprofen is an oral, over-the-counter nonsteroidal anti-inflammatory drug used to relieve pain, reduce inflammation, and alleviate fever.

In ActionADE, the most common culprit drugs were hydrochlorothiazide, ramipril, and apixaban, which accounted for 10.5% (356/3391) of all events; hydrochlorothiazide, empagliflozin, and apixaban represented 11.1% (105/951) of serious events (Tables 2 and 3). Hydrochlorothiazide and ramipril are commonly prescribed for hypertension. Apixaban, an oral anticoagulant, is primarily used for stroke prevention in patients with atrial fibrillation and for treatment and prevention of venous thromboembolism.

Figure 1. Flow diagram. ADE: adverse drug event; ADR: adverse drug reaction; PSLS-ADR: Patient Safety and Learning System–Adverse Drug Reaction.

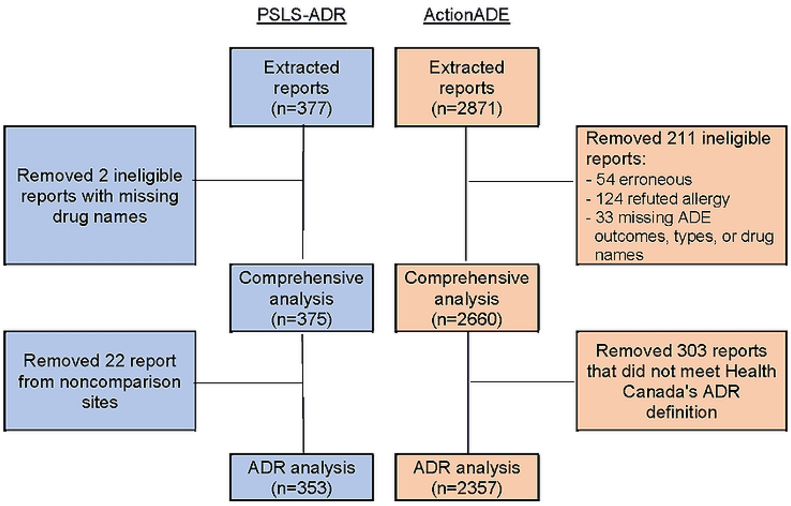


Table 1. Descriptive statistics of all events included in the comprehensive analysis by reporting system[a].

Characteristics	PSLS-ADR[b] (n=375), n (%)	ActionADE[c] (n=2660), n (%)
Type of reporting facilitates		
Hospitals	5 (41.6)	4 (100)
Nonhospitals	7 (58.4)	0 (0)
Patient age group (y)		
<1-19	11 (2.9)	22 (0.8)
20-44	83 (22.1)	299 (11.2)
45-64	123 (32.8)	523 (19.7)
65-74	67 (17.9)	544 (20.5)
75-84	52 (13.9)	674 (25.3)
>84	39 (10.4)	598 (22.5)
Patient sex		
Male	178 (47.5)	1285 (48.3)
Role of reporter		
Physicians	Suppressed[d]	204 (7.7)
Nurses	27 (7.2)	—[e]
Medical imaging staff or technologists	170 (45.3)	—[e]
Nurse practitioners	Suppressed	5 (0.2)
Pharmacists	174 (46.4)	2451 (92.1)
Others	Suppressed	—[e]
Proportion of serious events[f]	135 (36.0)	749 (28.2)

[a]The comprehensive analysis included all events from both reporting systems excluding reports related to errors, refuted allergy, and incomplete data on study variables.

[b]PSLS-ADR: Patient Safety and Learning System–Adverse Drug Reaction.

[c]ADE: adverse drug event.

[d]Cell sizes <5 are suppressed.

[e]These personnel are not eligible to report in ActionADE.

[f]Serious events are those with an outcome of fetal defect, permanent disability, hospitalization, extended hospitalization, life threatening, or death.

Table 2. Most frequently reported culprit drugs for all events in the comprehensive analysis by reporting systems.

System and drug	n (%)
PSLS-ADR[a] (n=375)	
Iohexol	154 (41.1)
Gadobutrol	12 (3.2)
Empagliflozin	7 (1.9)
Rivaroxaban	7 (1.9)
Furosemide	6 (1.6)
Nivolumab	6 (1.6)
Ramipril	6 (1.6)
Unknown generic drug	6 (1.6)
Acetylsalicylic acid	5 (1.3)
Ibuprofen	5 (1.3)
ActionADE[b] (n=2660)	
Hydrochlorothiazide	113 (4.2)
Apixaban	103 (3.9)
Ramipril	92 (3.5)
Acetylsalicylic acid	88 (3.3)
Warfarin	88 (3.3)
Rivaroxaban	79 (3)
Furosemide	77 (2.9)
Empagliflozin	63 (2.4)
Metformin HCL[c]	52 (2)
Spironolactone	50 (1.9)

[a]PSLS-ADR: Patient Safety and Learning System–Adverse Drug Reaction.
[b]ADE: adverse drug event.
[c]HCL: hydrochloride.

Table 3. Most frequently reported culprit drugs for serious events[a] in the comprehensive analysis by reporting systems.

System and drug	n (%)
PSLS-ADR[b] (n=135)	
Empagliflozin	6 (4.4)
Ibuprofen	5 (3.7)
Iohexol	5 (3.7)
Nivolumab	5 (3.7)
Acetylsalicylic acid	Suppressed[c]
Glyburide	Suppressed
Rivaroxaban	Suppressed
Allopurinol	Suppressed
Amlodipine besylate	Suppressed
Apixaban	Suppressed
ActionADE[d] (n=749)	
Hydrochlorothiazide	43 (5.7)
Empagliflozin	26 (3.5)
Apixaban	24 (3.2)
Furosemide	20 (2.7)
Acetylsalicylic acid	18 (2.4)
Rivaroxaban	18 (2.4)
Candesartan cilexetil	16 (2.1)
Ramipril	16 (2.1)
Chlorthalidone	16 (2.1)
Spironolactone	15 (2)

[a]Serious events are those with an outcome of fetal defect, permanent disability, hospitalization, extended hospitalization, life threatening, or death.
[b]PSLS-ADR: Patient Safety and Learning System–Adverse Drug Reaction.
[c]Cell sizes <5 are suppressed.
[d]ADE: adverse drug event.

ADR Analysis

We included a total of 2728 reports that met Health Canada's definition of an ADR from facilities that had the option of using either reporting system during this study's period (Figure 1) [32,33]. Of the included reports, 12.9% (353/2728) were entered in PSLS-ADR, while the majority (2357/2728, 86.4%) were reported in ActionADE.

The distribution of ADR reports by patient sex, age, primary reporters and proportion of serious events for both systems were similar to the comprehensive analysis (Table 4). However, each reporting system revealed distinct patterns of reporting. In PSLS-ADR, iohexol, gadobutrol, and empagliflozin accounted for 44.8% (168/353) of all events, while empagliflozin, ibuprofen, and nivolumab represented 12.1% (16/133) of serious events. In ActionADE, hydrochlorothiazide, ramipril, and apixaban accounted for 12% (284/2357) of all events. Furthermore, hydrochlorothiazide, empagliflozin, and apixaban represented 13.4% (88/671) of serious events (Tables 5 and 6).

A direct comparison in events reportable through both the PSLS-ADR and ActionADE systems revealed an increase in event reporting, including serious events, following the implementation of ActionADE (Figures 2 and 3). Baseline measurements indicate that the mean monthly counts of all events and serious events across sites were 2.9 (95% CI 2.2 to 3.6) and 1.7 (95% CI 0.8 to 2.5), respectively. In period 3, the mean monthly counts of all events and serious events across sites escalated to 27.2 (95% CI 20.4 to 34.0) and 7.0 (95% CI 4.9 to 9.2), respectively, reflecting a 9- and 4-fold increase over time. Furthermore, the mean monthly counts of all events and serious events during this study's period within the ActionADE system were 6- and 4-fold greater than that of PSLS-ADR.

Table 4. Descriptive statistics of events meeting Health Canada's ADR[a] definition[b] across common reporting sites[c] by reporting system.

Characteristics	PSLS-ADR[d] (n=353), n (%)	ActionADE[e] (n=2357), n (%)
Patient age group (y)		
<1-19	11 (3.1)	18 (0.8)
20-44	77 (21.8)	239 (10.1)
45-64	114 (32.3)	450 (19.1)
65-74	64 (18.1)	494 (21)
75-84	49 (13.9)	606 (25.7)
>84	38 (10.8)	550 (23.3)
Patient sex		
Male	173 (49)	1114 (47.3)
Role of reporter		
Physicians	Suppressed[f]	161 (6.8)
Nurses	22 (6.2)	—[g]
Medical imaging staff or technologists[h]	155 (43.9)	—
Nurse practitioners	Suppressed	5 (0.2)
Pharmacists	173 (49)	2190 (92.9)
Others[h]	Suppressed	—
Proportion of serious events[i]	133 (37.7)	671 (28.5)

[a]ADR: adverse drug reactions.

[b]According to Health Canada adverse drug reaction includes unintended effects, health product abuse, overdoses, interactions (including drug-drug and drug-food interactions), and unusual lack of therapeutic efficacy.

[c]Common reporting sites included Vancouver General, University of British Columbia, Lions Gate, and Richmond Hospitals.

[d]PSLS-ADR: Patient Safety and Learning System–Adverse Drug Reaction.

[e]ADE: adverse drug event.

[f]Cell sizes <5 are suppressed.

[g]Not available.

[h]These personnel are not eligible to report in ActionADE.

[i]Serious events are those with an outcome of fetal defect, permanent disability, hospitalization, extended hospitalization, life threatening, or death.

Table 5. Most frequently reported culprit drugs for all events meeting Health Canada's ADR[a] definitions[b] across common reporting sites[c] by reporting system and severity.

System and drug	n (%)
PSLS-ADR[d] (n=353)	
Iohexol	139 (39.4)
Gadobutrol	12 (3.4)
Empagliflozin	7 (2)
Rivaroxaban	7 (2)
Furosemide	6 (1.7)
Nivolumab	6 (1.7)
Ramipril	6 (1.7)
Acetylsalicylic acid	5 (1.4)
Ibuprofen	5 (1.4)
Indapamide	5 (1.4)
ActionADE[e] (n=2357)	
Hydrochlorothiazide	109 (4.6)
Ramipril	88 (3.7)
Apixaban	87 (3.7)
Acetylsalicylic acid	75 (3.2)
Warfarin	74 (3.1)
Rivaroxaban	71 (3)
Empagliflozin	60 (2.5)
Furosemide	56 (2.4)
Metformin HCL[f]	44 (1.9)
Ibuprofen	43 (1.8)

[a]ADR: adverse drug reaction.

[b]According to Health Canada adverse drug reaction includes unintended effects, health product abuse, overdoses, interactions (including drug-drug and drug-food interactions), and unusual lack of therapeutic efficacy.

[c]Common reporting sites included Vancouver General, University of British Columbia, Lions Gate, and Richmond Hospitals.

[d]PSLS-ADR: Patient Safety and Learning System–Adverse Drug Reaction.

[e]ADE: adverse drug event.

[f]HCL: hydrochloride.

Table 6. Most frequently reported culprit drugs for serious events[a] meeting Health Canada's ADR[b] definitions[c] across common reporting sites[d] by reporting system and severity.

System and drug	n (%)
PSLS-ADR[e] (n=133)	
Empagliflozin	6 (4.5)
Ibuprofen	5 (3.8)
Nivolumab	5 (3.8)
Acetylsalicylic acid	Suppressed[f]
Glyburide	Suppressed
Rivaroxaban	Suppressed
Allopurinol	Suppressed
Amlodipine besylate	Suppressed
Apixaban	Suppressed
Clopidogrel bisulfate	Suppressed
ActionADE[g] (n=671)	
Hydrochlorothiazide	43 (6.7)
Empagliflozin	25 (3.7)
Apixaban	20 (3)
Acetylsalicylic acid	19 (2.8)
Chlorthalidone	16 (2.4)
Ramipril	14 (2.1)
Rivaroxaban	14 (2.1)
Ibuprofen	13 (1.9)
Warfarin	13 (1.9)
Candesartan cilexetil	12 (1.8)

[a]Common reporting sites included Vancouver General, University of British Columbia, Lions Gate, and Richmond Hospitals.

[b]ADR: adverse drug reaction.

[c]According to Health Canada adverse drug reaction includes unintended effects, health product abuse, overdoses, interactions (including drug-drug and drug-food interactions), and unusual lack of therapeutic efficacy.

[d]Serious events are those with an outcome of fetal defect, permanent disability, hospitalization, extended hospitalization, life threatening, or death.

[e]PSLS-ADR: Patient Safety and Learning System–Adverse Drug Reaction.

[f]Cell sizes <5 are suppressed.

[g]ADE: adverse drug event.

Figure 2. Mean monthly counts of all events meeting Health Canada's ADR definitions across common reporting sites during this study's period. ADR: adverse drug reaction; PSLS-ADR: Patient Safety and Learning System–Adverse Drug Reaction.

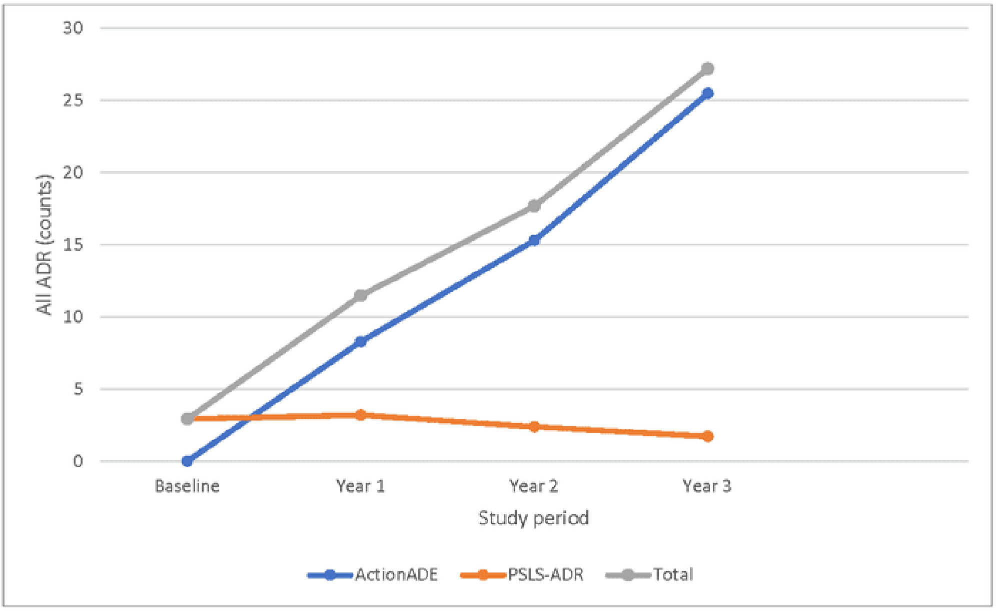
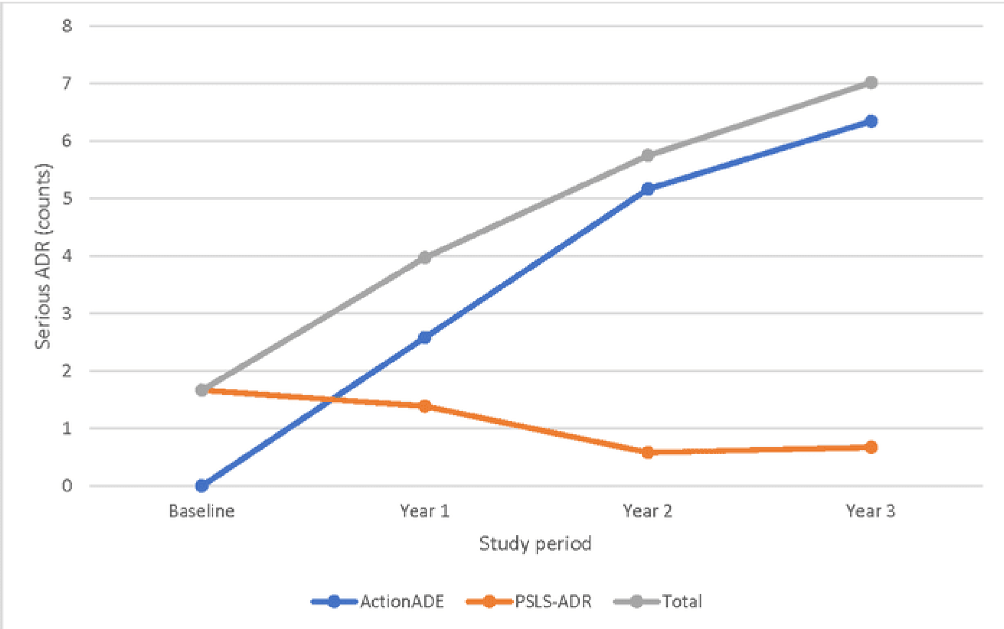


Figure 3. Mean monthly counts of serious events meeting Health Canada's ADR definitions across common reporting sites during this study's period. ADR: adverse drug reaction; PSLS-ADR: Patient Safety and Learning System–Adverse Drug Reaction.



Discussion

Principal Findings

Our study aimed to describe and compare ADEs reported using 2 distinct reporting systems that were developed and implemented in different ways. Both PSLS-ADR and ActionADE are currently in use in BC in the first 3 years following the implementation of Vanessa's Law. We observed differences in reports between the 2 systems regarding their coverage, usage, and the type of ADE data captured.

PSLS-ADR had broader coverage, collecting data from various health care facilities including community health centers, vaccination clinics, and outpatient clinics. Its user base was more diverse including physicians, nurses, medical imaging staff or technologists, nurse practitioners, pharmacists, and other professionals. In contrast, ActionADE coverage was limited to ADEs identified in patients presenting to 4 participating hospitals, with clinical pharmacists as its primary user. The broader coverage of PSLS-ADR can be attributed to its established position as a provincial safety event reporting platform; its accessibility to a broader range of health professions; and a federal mandate for hospitals to stimulate reporting using health authority wide communication efforts including email blasts, information on health authority websites, and presentations to provider groups. Leveraging the insights gained from PSLS-ADR, our research team is actively collaborating with key stakeholders to broaden ActionADE's app. The Vancouver Coastal Health Authority, where ActionADE is presently in use, has recently endorsed it as a standard practice for ADE reporting in new care settings, including long-term care homes, in-patient wards, and community clinics.

Although PSLS-ADR exhibited broader coverage, ActionADE demonstrated higher usage. Our comparative analysis revealed that the average monthly counts of all events and serious events in ActionADE were 6 and 4 times higher, respectively, than in the PSLS-ADR system. Several factors might contribute to these discrepancies in reporting rates. First, PSLS-ADR was designed solely for Vanessa's Law compliance, with reports forwarded to Health Canada for surveillance purposes. ActionADE, on the other hand, serves the dual purpose of functioning as both a clinical communication tool and a means of complying with Vanessa's Law, thus improving patient safety [15]. Reports entered into ActionADE are used to generate preventive alerts in community pharmacies when pharmacists attempt redispensation of a drug that has previously caused the patient harm, which have demonstrated preliminary effectiveness [34]. The potential impact of reporting in ActionADE on patient safety is likely a motivating factor for providers to report ADEs [35]. Furthermore, ActionADE has a proactive implementation support mechanism, which has been shown to be instrumental in enhancing providers' adoption of the reporting platform [35]. Finally, ActionADE used participatory design principles to optimize its design to facilitate use by end users and is integrated with PharmaNet to enable prepopulation of fields to allow reporters to generate reports ≤2 minutes, whereas PSLS-ADR users noted that reports can take 20 minutes to complete [34].

The 2 systems captured adverse events to different culprit drugs. This can be attributed to the more limited accessibility of ActionADE. The most reported drugs in PSLS-ADR were iohexol and gadobutrol, and correspondingly, medical imaging staff or technologists made up a significant proportion of reporters. This suggests that the current workflow for ADR reporting of radiopharmaceuticals is designated to medical imaging staff or technologists. Imaging staff or technologists were unable to use ActionADE at the time of this study due to PharmaNet legislation, which requires that users have prescriber ID restricting use to physicians, pharmacists, and nurse practitioners. This restriction has resulted in fewer radiopharmaceutical ADRs to be reported, as pharmacists generally do not work in radiology departments.

ActionADE frequently captured hydrochlorothiazide-related events, while only a few of such events were captured in PSLS-ADR. Among the ADRs associated with hydrochlorothiazide, electrolyte disturbances, and acute kidney injury were found to be the most common [34], involving multiple additional contributing factors. The specific functionality offered by ActionADE, such as the ability to specify the provider's certainty that the patient's presentation and the option to update or refute events based on new information or alternative diagnoses, likely played a role in encouraging clinicians to report these more complex events [11,15,27,30].

Ibuprofen was the second most commonly reported culprit drug related to serious events in PSLS-ADR, but it barely made the top 10 in ActionADE. This discrepancy may be due to the over-the-counter status of ibuprofen, which means patients can access the medication without a prescription and bypass communication about ADEs from ActionADE that is built into the prescription dispensation process.

While our study primarily focused on comparing these 2 systems, it is crucial to view these findings in the broader context of ADE reporting. Despite these disparities, both systems play vital roles in contributing to patient safety by capturing valuable information on ADEs. PSLS-ADR is an effective means of capturing radiopharmaceutical-related ADEs by imaging staff and technicians who are not trained in taking medication histories or ADE assessments, while ActionADE is more effective for pharmaceutical-related ADEs by clinical pharmacists that are reported and communicated on a patient-level to improve safety. These systems work in a complementary manner, catering to different areas of the health care system and capturing unique data and thus offering a more comprehensive picture of ADEs. For example, a common signal between the 2 systems might indicate a more serious issue for a specific drug irrespective of context (eg, empagliflozin). These findings suggest the need for careful attention to the design and implementation of these systems to ensure they effectively serve their intended users and context of use and ensure data resulting from each system are interpreted correctly by end users. The absence of reporting of one type of event may reflect design, implementation, or user characteristics rather than the absence of these events.

Limitations

To our knowledge, this is the first study to directly compare 2 ADE reporting systems operating within the same jurisdiction. While the results of our study provided valuable insights into the differences between these systems, it is important to acknowledge several limitations that warrant consideration in interpreting the results. First, our study sample was confined to 2 reporting systems, which may not fully encapsulate the diversity of all systems employed across health care settings globally. As a result, the findings may not be generalizable to other reporting systems. Second, our data set was limited to facilities that used PSLS-ADR or ActionADE for reporting. This reduces the generalizability of our findings to the wider array of health care facilities in BC or nationally. It is plausible that unaccounted-for variations in data and reporting practices among facilities not deploying these 2 systems could exist. Third, our study may be susceptible to unmeasured and uncontrolled confounding variables. For example, the level of organizational emphasis on ADE reporting, differences in implementation, available resources, and providers' perceptions could have affected the usage and coverage of the 2 systems under study. This variability might have further influenced the nature of ADE information reported. Fourth, the relatively small number of drugs resulting in ADEs prevented us from conducting a robust quantitative comparison of these events. Furthermore, the data we used were a snapshot in time and may not reflect changes in reporting systems or health care facilities that have occurred since then. Lastly, we consciously chose not to draw comparisons with other studies examining the frequently reported culprit drugs from spontaneous reporting systems in other jurisdictions. This decision stems from the recognition that the diversity in ADE reports—both in terms of numbers and types—is intricately tied to factors such as system design, geography, population characteristics, drug exposures, and the medical system itself. To facilitate meaningful comparisons across studies, a more robust surveillance system is needed.

Conclusions

Understanding the differences between reporting systems can inform future systems design and improvement, including changes to user training and implementation, and inform the use of forthcoming data and procurement decisions for reporting systems. Further research could explore how to integrate the strengths of both systems, potentially leading to more comprehensive safety data to facilitate drug and patient safety and inform pharmacoepidemiologic studies. Continuous evaluation and improvement are essential considering the significant role these systems play to improve our health systems.

Acknowledgments

The ActionADE platform is funded by the Canadian Institutes of Health Research, with support from the Ministry of Health British Columbia, Vancouver Coastal Health, the Lower Mainland Consolidated Pharmacies, and the BC SUPPORT Unit. CMH is supported by the Michael Smith Foundation for Health Research Health Professional-Investigator award. EYL is supported by the Michael Smith Health Research BC Research Trainee Award (GR018656). The funder has no role in the design, conduct, analysis, interpretation of data, or the drafting of scientific papers.

Authors' Contributions

All authors contributed to this study’s conception and design. EYL requested access and analyzed the data and wrote the first draft of this paper. AC, CMH, and SSS contributed to the refinement of the data analysis. All authors contributed to the interpretation of the findings and commented on previous paper versions. All authors read and approved the submitted paper and have agreed to be personally accountable for their contribution.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of (Patient Safety and Learning System–Adverse Drug Reaction) PSLS-ADR and ActionADE.
[DOCX File , 23 KB - humanfactors_v11i1e52495_app1.docx]

Multimedia Appendix 2
Screenshot of Patient Safety and Learning System–Adverse Drug Reaction (PSLS-ADR).
[DOCX File , 282 KB - humanfactors_v11i1e52495_app2.docx]

Multimedia Appendix 3
Screenshot of ActionADE.
[DOCX File , 527 KB - humanfactors_v11i1e52495_app3.docx]

Multimedia Appendix 4
Data fields included in Patient Safety and Learning System–Adverse Drug Reaction (PSLS-ADR) and ActionADE.
[DOCX File , 36 KB - humanfactors_v11i1e52495_app4.docx]

Multimedia Appendix 5
Characteristics of sites that had options to use Patient Safety and Learning System–Adverse Drug Reaction (PSLS-ADR) or ActionADE systems.
[DOCX File , 30 KB - humanfactors_v11i1e52495_app5.docx]

Multimedia Appendix 6
Descriptions of the 4-phase study period.
[DOCX File , 159 KB - humanfactors_v11i1e52495_app6.docx]

References

1. Hohl CM, Dankoff J, Colacone A, Afilalo M. Polypharmacy, adverse drug-related events, and potential adverse drug interactions in elderly patients presenting to an emergency department. Ann Emerg Med 2001;38(6):666-671. [doi: 10.1067/mem.2001.119456] [Medline: 11719747]
2. Zed PJ, Abu-Laban RB, Balen RM, Loewen PS, Hohl CM, Brubacher JR, et al. Incidence, severity and preventability of medication-related visits to the emergency department: a prospective study. CMAJ 2008;178(12):1563-1569 [FREE Full text] [doi: 10.1503/cmaj.071594] [Medline: 18519904]
3. NACRS emergency department visits and lengths of stay. Canadian Institute for Health Information. Ottawa, Ontario: Canadian Institute for Health Information; 2018. URL: https://www.cihi.ca/en/nacrs-emergency-department-visits-and-lengths-of-stay [accessed 2023-12-22]
4. Sheikh A, Dhingra-Kumar N, Kelley E, Kieny MP, Donaldson LJ. The third global patient safety challenge: tackling medication-related harm. Bull World Health Organ 2017;95(8):546-546A [FREE Full text] [doi: 10.2471/BLT.17.198002] [Medline: 28804162]
5. Herder M, Gibson E, Graham J, Lexchin J, Mintzes B. Regulating prescription drugs for patient safety: does Bill C-17 go far enough? CMAJ 2014;186(8):E287-E292 [FREE Full text] [doi: 10.1503/cmaj.131850] [Medline: 24616135]
6. Raj N, Fernandes S, Charyulu NR, Dubey A, Hebbar GS. Postmarket surveillance: a review on key aspects and measures on the effective functioning in the context of the United Kingdom and Canada. Ther Adv Drug Saf 2019;10:2042098619865413 [FREE Full text] [doi: 10.1177/2042098619865413] [Medline: 31384423]
7. The importance of pharmacovigilance: safety monitoring of medical products. World Health Organization. 2002. URL: https://www.who.int/publications/i/item/10665-42493 [accessed 2022-01-23]
8. Alomar M, Tawfiq AM, Hassan N, Palaian S. Post marketing surveillance of suspected adverse drug reactions through spontaneous reporting: current status, challenges and the future. Ther Adv Drug Saf 2020;11:2042098620938595 [FREE Full text] [doi: 10.1177/2042098620938595] [Medline: 32843958]
9. Hazell L, Shakir SAW. Under-reporting of adverse drug reactions : a systematic review. Drug Saf 2006;29(5):385-396 [FREE Full text] [doi: 10.2165/00002018-200629050-00003] [Medline: 16689555]
10. Bailey C, Peddie D, Wickham ME, Badke K, Small SS, Doyle-Waters MM, et al. Adverse drug event reporting systems: a systematic review. Br J Clin Pharmacol 2016;82(1):17-29 [FREE Full text] [doi: 10.1111/bcp.12944] [Medline: 27016266]
11. Hohl CM, Small SS, Peddie D, Badke K, Bailey C, Balka E. Why clinicians don't report adverse drug events: qualitative study. JMIR Public Health Surveill 2018;4(1):e21 [FREE Full text] [doi: 10.2196/publichealth.9282] [Medline: 29487041]
12. Fletcher AP. Spontaneous adverse drug reaction reporting vs event monitoring: a comparison. J R Soc Med 1991;84(6):341-344 [FREE Full text] [doi: 10.1177/014107689108400612] [Medline: 2061900]
13. Lynn RM, Riding K, McIntosh N. The use of electronic reporting to aid surveillance of ADRs in children: a proof of concept study. Arch Dis Child 2010;95(4):262-265 [FREE Full text] [doi: 10.1136/adc.2009.158766] [Medline: 20335236]
14. Mandatory Reporting of Serious Adverse Drug Reactions and Medical Device Incidents by Hospitals: Guidance Document. Ottawa, ON: Health Canada; 2019.
15. Hohl C, Lexchin JR, Balka E. Can reporting of adverse drug reactions create safer systems while improving health data? CMAJ 2015;187(11):789-790 [FREE Full text] [doi: 10.1503/cmaj.150057] [Medline: 25941261]
16. British Columbia patient safety & learning system annual report. BC Patient Safety & Learning System Central Office. 2008. URL: https://bcpslscentral.ca/wp-content/uploads/2014/04/pslsannualreport1011.pdf [accessed 2023-12-22]
17. Cochrane D, Taylor A, Miller G, Hait V, Matsui I, Bharadwaj M, et al. Establishing a provincial patient safety and learning system: pilot project results and lessons learned. Healthc Q 2009;12:147-153 [FREE Full text] [doi: 10.12927/hcq.2009.20717] [Medline: 19667793]
18. Online ADR reporting is coming!. British Columbia Patient Safety & Learning System Central. 2014. URL: https://bcpslscentral.ca/online-adr-reporting-is-coming/#:~:text=Most%20of%20us%20will%20take,taking%20the%20medication%20as%20intended [accessed 2023-12-22]
19. Vanessa's law (ADR & MDI reporting) fact sheet. British Columbia Patient Safety & Learning System Central. URL: https://medicalstaff.islandhealth.ca/sites/default/files/2019-11/ADR%20and%20MDI%20Reporting%20Fact%20Sheet%20and%20FAQ_0.pdf [accessed 2023-12-22]
20. Online ADR reporting is coming!. British Columbia Patient Safety & Learning System. 2014. URL: https://bcpslscentral.ca/online-adr-reporting-is-coming/ [accessed 2023-05-23]
21. Woo SA, Cragg A, Wickham ME, Peddie D, Balka E, Scheuermeyer F, et al. Methods for evaluating adverse drug event preventability in emergency department patients. BMC Med Res Methodol 2018;18(1):160 [FREE Full text] [doi: 10.1186/s12874-018-0617-4] [Medline: 30514232]
22. van der Linden CMJ, Jansen PAF, van Marum RJ, Grouls RJE, Korsten EHM, Egberts ACG. Recurrence of adverse drug reactions following inappropriate re-prescription: better documentation, availability of information and monitoring are needed. Drug Saf 2010;33(7):535-538 [FREE Full text] [doi: 10.2165/11532350-000000000-00000] [Medline: 20553055]
23. Hau JP, Brasher PMA, Cragg A, Small S, Wickham M, Hohl CM. Using ActionADE to create information continuity to reduce re-exposures to harmful medications: study protocol for a randomized controlled trial. Trials 2021;22(1):119 [FREE Full text] [doi: 10.1186/s13063-021-05061-7] [Medline: 33546752]

24. Hohl CM, Woo SA, Cragg A, Wickham ME, Ackerley C, Scheuermeyer F, et al. Repeat adverse drug events associated with outpatient medications: a descriptive analysis of 3 observational studies in British Columbia, Canada. CMAJ Open 2019;7(3):E446-E453 [FREE Full text] [doi: 10.9778/cmajo.20180190] [Medline: 31320328]
25. Chruscicki A, Badke K, Peddie D, Small S, Balka E, Hohl CM. Pilot-testing an adverse drug event reporting form prior to its implementation in an electronic health record. Springerplus 2016;5(1):1764 [FREE Full text] [doi: 10.1186/s40064-016-3382-z] [Medline: 27795906]
26. Hohl CM, Badke K, Zhao A, Wickham ME, Woo SA, Sivilotti MLA, et al. Prospective validation of clinical criteria to identify emergency department patients at high risk for adverse drug events. Acad Emerg Med 2018;25(9):1015-1026 [FREE Full text] [doi: 10.1111/acem.13407] [Medline: 29517818]
27. Peddie D, Small SS, Badke K, Bailey C, Balka E, Hohl CM. Adverse drug event reporting from clinical care: mixed-methods analysis for a minimum required dataset. JMIR Med Inform 2018;6(2):e10248 [FREE Full text] [doi: 10.2196/10248] [Medline: 29954724]
28. Chan E, Small SS, Wickham ME, Cheng V, Balka E, Hohl CM. The utility of different data standards to document adverse drug event symptoms and diagnoses: mixed methods study. J Med Internet Res 2021;23(12):e27188 [FREE Full text] [doi: 10.2196/27188] [Medline: 34890351]
29. PharmaCare policy manual (unified). Appendix C-practitioner ID reference codes. British Columbia Ministry of Health. 2023. URL: https://www2.gov.bc.ca/gov/content/health/practitioner-professional-resources/pharmacare/pharmacare-publications/pharmacare-policy-manual-unified/practitioner-id-reference-codes [accessed 2023-05-22]
30. Peddie D, Small S, Wickham M, Bailey C, Hohl C, Balka E. Designing novel health ICTs to support work, not generate it: five principles. Stud Health Technol Inform 2017;234:262-268. [Medline: 28186052]
31. How's BC doing so far with Vanessa's law? British Columbia Patient Safety & Learning System. 2012. URL: http://bcpslscentral.ca./ [accessed 2023-12-22]
32. Module 1: overview of Vanessa's law and reporting requirements. Health Canada. 2020. URL: https://www.canada.ca/en/health-canada/services/drugs-health-products/medeffect-canada/adverse-reaction-reporting/mandatory-hospital-reporting/education/module-1.html [accessed 2020-01-31]
33. Serious adverse drug reaction reporting form for hospitals. Health Canada. 2020. URL: https://www.canada.ca/content/dam/hc-sc/documents/services/drugs-health-products/medeffect-canada/adverse-reaction-reporting/mandatory-reporting-hospital-drug-eng.pdf [accessed 2023-07-13]
34. Cragg A, Small SS, Lau E, Rowe A, Lau A, Butcher K, et al. Sharing adverse drug event reports between hospitals and community pharmacists to inform re-dispensing: an analysis of reports and process outcomes. Drug Saf 2023;46(11):1161-1172 [FREE Full text] [doi: 10.1007/s40264-023-01348-7] [Medline: 37783974]
35. Lau EY, Small SS, Butcher K, Cragg A, Loh GW, Shalansky S, et al. An external facilitation intervention to increase uptake of an adverse drug event reporting intervention. Front Health Serv 2023;3:1106586 [FREE Full text] [doi: 10.3389/frhs.2023.1106586] [Medline: 37332530]

Abbreviations

ADE: adverse drug event
ADR: adverse drug reaction
BC: British Columbia
PSLS-ADR: Patient Safety and Learning System–Adverse Drug Reaction
PSLS: Patient Safety and Learning System
UBC: University of British Columbia
WHO: World Health Organization

Edited by K Adapa; submitted 05.09.23; peer-reviewed by R Lim, A Austin; comments to author 28.11.23; revised version received 06.12.23; accepted 07.12.23; published 18.01.24.

Please cite as:
Lau EY, Cragg A, Small SS, Butcher K, Hohl CM
Characterizing and Comparing Adverse Drug Events Documented in 2 Spontaneous Reporting Systems in the Lower Mainland of British Columbia, Canada: Retrospective Observational Study
JMIR Hum Factors 2024;11:e52495
URL: https://humanfactors.jmir.org/2024/1/e52495
doi:10.2196/52495
PMID:38236629

Original Paper

Usability and Evaluation of a Health Information System in the Emergency Department: Mixed Methods Study

Christina Østervang[1,2], RN, MScN, PhD; Charlotte Myhre Jensen[2,3], RN, MEd, PhD; Elisabeth Coyne[2,4], RN, PhD; Karin B Dieperink[2,5,6], RN, MCN, PhD; Annmarie Lassen[1,2], DMSci, PhD

[1]Department of Emergency Medicine, Odense University Hospital, Odense, Denmark
[2]Department of Clinical Research, University of Southern Denmark, Odense, Denmark
[3]Department of Orthopedic Surgery and Traumatology, Odense University Hospital, Odense, Denmark
[4]School of Nursing and Midwifery, Griffith University, Brisbane, Australia
[5]Department of Oncology, Odense University Hospital, Odense, Denmark
[6]Family Focused Healthcare Research Center (FACE), University of Southern Denmark, Odense, Denmark

Corresponding Author:
Christina Østervang, RN, MScN, PhD
Department of Emergency Medicine, Odense University Hospital
Kløvervænget 25
Odense, 5000
Denmark
Phone: 45 21757009
Email: christina.oestervang@rsyd.dk

Abstract

Background: A lack of information during an emergency visit leads to the experience of powerlessness for patients and their family members, who may also feel unprepared to cope with acute symptoms. The ever-changing nature and fast-paced workflow in the emergency department (ED) often affect how health care professionals can tailor information and communication to the needs of the patient.

Objective: This study aimed to evaluate the usability and experience of a newly developed information system. The system was developed together with patients and their family members to help provide the information needed in the ED.

Methods: We conducted a mixed methods study consisting of quantitative data obtained from the System Usability Scale questionnaire and qualitative interview data obtained from purposively selected participants included in the quantitative part of the study.

Results: A total of 106 patients and 14 family members (N=120) answered the questionnaire. A total of 10 patients and 3 family members participated in the interviews. Based on the System Usability Scale score, the information system was rated close to excellent, with a mean score of 83.6 (SD 12.8). Most of the participants found the information system easy to use and would like to use it again. The participants reported that the system helped them feel in control, and the information was useful. Simplifications were needed to improve the user experience for the older individuals.

Conclusions: This study demonstrates that the usability of the information system is rated close to excellent. It was perceived to be useful as it enabled understanding and predictability of the patient's trajectory in the ED. Areas for improvement include making the system more usable by older individuals. The study provides an example of how a technological solution can be used to diminish the information gap in an ED context.

(JMIR Hum Factors 2024;11:e48445) doi:10.2196/48445

KEYWORDS

consumer; eHealth; elderly; emergency department; emergency; family members; healthcare professionals; information system; mixed methods research: patients; qualitative interview; questionnaire; technology; usability; usable

Introduction

Background

Clear communication and information are essential to improving care and patient outcomes in the emergency department (ED) [1-5]. A lack of information during ED visits causes patients and their family members to experience a sense of powerlessness and to feel unprepared to cope with acute symptoms [2,3,6]. Due to the hectic nature of the ED and the constant interruptions, communication from health care professionals is often inadequate or not tailored to patients and their families [4,7]. While this problem has been known for many years, it still persists to this date [1].

Health technologies are implemented in many parts of health care systems to promote quality care and treatment [8]. The design and purpose of health technologies range widely from organizational [9] to person-centered intentions [10]. In the ED, technologies may be used as quality dashboards [9] and more personal information systems on patients' own devices to support the delivery of health information [11]. However, the successful use of technology in clinical practice is likely to be ineffective if user needs are not carefully addressed and incorporated before attempting a full-scale implementation [9,12]. Thoroughness in integrating and understanding user perspectives will have a direct impact on how well the technology is suited for clinical practice [13,14].

Based on the current findings, patients in the ED and their family members have unmet information needs [1-4]. Hence, guided by the principles of user-driven activities [15], a health information system was developed [16]. The health information system, which is called "Cetrea Clinical Logistic (CCL) *for patients*," is available for patients in the hospital's emergency room and displays real-time information, including (1) person-centered activities, (2) information videos, (3) a notepad, (4) waiting time, and (5) the nurse and physician responsible for care.

Usability is one of the factors affecting the acceptance of health information systems by users, and it is essential for the effective use of the system [17]. A usability evaluation can identify problems and weaknesses in the design and functionalities in the early development phase [18]. Usability tests allow developers to address and adjust concerns and, thus, avoid implementing technologies that will not be useful in the clinical context.

Therefore, a usability evaluation from an end-user perspective was completed to obtain a nuanced understanding of the sustainable use of the system, specifically from the perspective of patients and their family members.

Objective

The objective of this study is to gain knowledge about the usability and experiences of the newly developed information system, CCL for patients. This study reports on patients' and family members' evaluations of this system.

Participatory Design and Technology

This study is the final phase of a 3-phase participatory design study (Figure 1) [19]. Participatory design is a research methodology based on the epistemological position of genuine involvement and understanding of the needs of future end users. A new technology can be designed to improve a real-life problem [20]. The core principles in participatory design methodology have been the theoretical framework of the overall study. In the initial phase, the author group identified the essential needs of patients in the ED, their family members, and ED clinicians [2,3]. The results from phase 1 informed the second phase, in which an information system, CCL for patients, was developed in a cocreation process [16]. The third phase involved testing and evaluation of the system, which is reported in this study. Reporting the evaluation of participatory-designed health technology is a common part of the research methodology [21,22].

The author group has had no financial interest in the system owners of CCL for patients and has no interest in either marketing or promoting the system.

CCL for patients provides information directly to patients and their family members during their stay in the ED. The information provided relates to treatment and time factors and is adjusted toward the individual patient. CCL is an already existing and implemented system for task management for clinicians' use only [23], whereas CCL for patients is a redesign and further development of the system for patients' use.

The functionalities of the CCL for patients' screen are presented in Figure 2.

Figure 1. Overview of the 3-phase study, highlighting the evaluation phase, which is reported in this study.

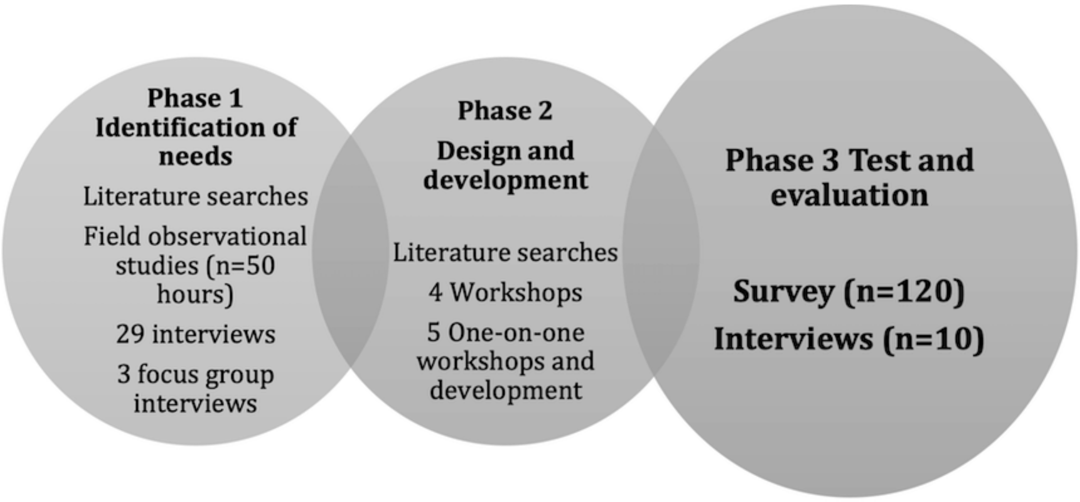
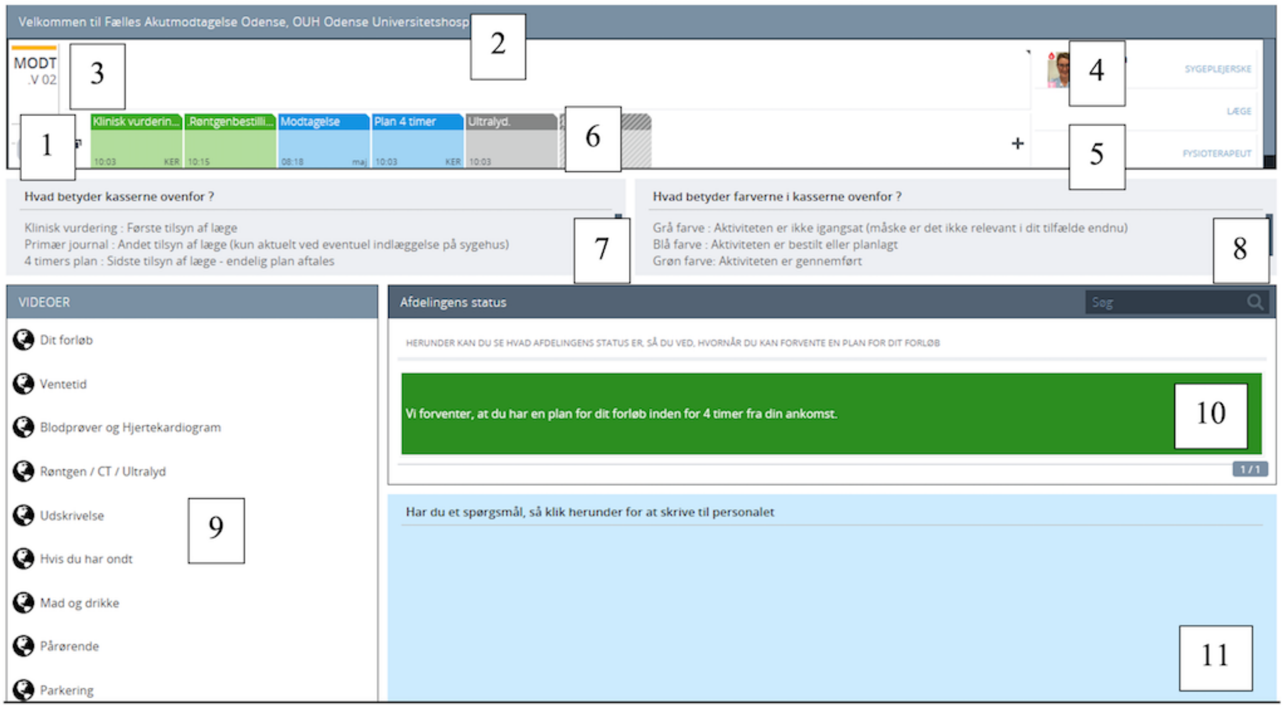


Figure 2. Cetrea Clinical Logistics (CCL) for patients and its functionalities, as displayed to patients and their family members, developed in the second phase of the overall study (the figure has been previously published by Østervang et al [16]). (1) Number of the ED room. (2) Name of the hospital department. (3) The name of the patient (no sensitive information is displayed). (4) The nurse who is responsible for care. (5) The physician who is responsible for treatment. (6) Process line with activities. Displaying nurse assessments, blood samples, electrocardiograms, physician assessments, X-rays, etc. (7) Clarification of the different colors in the process line. Gray: not started; blue: activity scheduled; and green: activity finished. (8) Clarification of special activity names. (9) Link to information videos (eg, information on discharge). (10) Three diverse colors indicate the estimated waiting time: less than 4 hours, equal to 4 hours, or more than 4 hours, respectively. (11) The shared note pad for the patients to write questions to health care professionals or messages from family members.

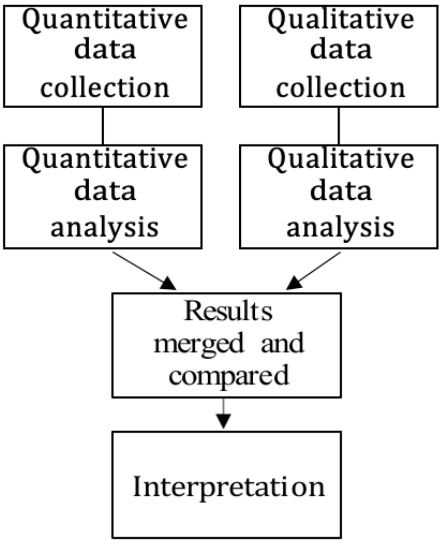


Methods

Research Design

This is a mixed methods study inspired by a convergent parallel design [24]. This design was chosen to obtain nuanced insights into the usability of the system. Further, we adopted this approach to usability testing because quantitative data can identify usability issues and dissatisfaction with program design, while qualitative data can provide detailed information about the causes of the usability issues and point at potential methods for program optimization. As shown in Figure 3 [24], the study contained the following two parts, ending with a merged result: (1) a questionnaire and descriptive characteristics of the participants, and (2) semistructured interviews with patients and their family members.

Figure 3. Diagram for a study using convergent design (Creswell and Clark) [24].



Setting

The data were collected in Odense University Hospital's ED between August 22 and September 29, 2022, on weekdays from 8 AM to 5 PM. The information system was displayed on a laptop personal computer (PC) sitting on the bedside table in the ED room. Four PCs were used during the test phase. They were installed in the specific ED room where the patients participating in the study were admitted.

Inclusion Criteria and Recruitment

All patients admitted to the medical area of the ED without a final plan for treatment and care were eligible for participation. Patients were excluded if they were severely ill or cognitively unable to use the technology. However, patients who were excluded due to a cognitive inability to use the screen but who were still able to give consent for their family members' participation were enrolled if the family member was interested in participating. Patients were recruited by the first author (CØ) or one of 2 research assistants, all of whom have a Master of Nursing Science degree and research experience. Potential participants were identified and discussed with the responsible care nurse before they were approached to reduce the possibility of any concerns.

Quantitative Phase

A survey was conducted to elicit the opinions and experiences of patients and their family members using the information system.

The Questionnaire

The questionnaire, the System Usability Scale (SUS), contained questions regarding the usability of the system. Answers are rated on a 5-point Likert scale from "strongly disagree" to "strongly agree," with 5 representing the highest score (strongly agree) [25]. The participants answered 10 questions from the SUS and 2 questions specific to this study (questions 11 and 12) [25]. These 2 extra questions were added to obtain general information about the participants' experience with CCL for patients (question 11: "I think the system provided a great overview of my stay," and question 12: "I think the information in the system made sense to me"). As SUS has been translated and validated in a Danish hospital context previously (Cronbach α=.87) [26], it was considered suitable for this study.

Sample Size

A total SUS score between 70 and 90 indicates good to excellent usability of the tested system [27]. Based on previous research conducted in Scandinavia using SUS in health care with a reported mean score of 79.81 (SD 14.28), we would gain a 95% CI for a mean score between 77.2 and 82.4 if a total of 120 patients were included [28].

Data Collection

If a patient agreed to participate, the researcher cooperated with the local IT department at the hospital to ensure the patient's access to the system. Initially, the researcher sent the IT department an SMS text message providing information on the PC number and the ED room number. The IT specialist matched the PC and room numbers. Then, the researcher double-checked that the correct information was displayed before handing it to the patient. All participants were given oral guidance on how to use CCL for patients.

The PC with individual information was placed on the bedside table until either the patient left the ED, the patient had used the system for a minimum of 2 hours, or the patient felt ready to perform the evaluation. All of this had to happen no later than 5 PM, when the IT department closed. When returning the PC, the participants were given an iPad to fill out the questionnaire. The data were stored on the logged server OPEN [29], which is part of Odense University Hospital and the University of Southern Denmark.

Qualitative Phase

Interviews were conducted with individual patients or with the patient together with a family member to get a deeper insight into their experiences using the information system.

Interviews

The qualitative part included a subset of the participants from the quantitative part. Before making CCL for patients available to the participants, they were asked whether they were interested in participating in an interview.

All interviews were conducted by the first author (CØ). By taking a phenomenological-hermeneutical stance, CØ was allowed to recognize her perceptions as an experienced emergency nurse within hermeneutic interpretation [30]. To bridle her preconceived ideas, CØ wrote down her preunderstandings of why patients lack information in the ED. This reflection provided an initial focus for both the overall research question and the interview questions.

The interviews were conducted in the hospital room after the participants had completed the questionnaire. Notes and quotes were taken during the interview. A summary of the conversation was generated at the end of the interview in the form of member checking [31]. A semistructured interview guide inspired by Kvale was used [32]. An example of a question is: "What was your experience of using CCL for patients?" The interviews lasted up to 30 minutes. The interviews were conducted until no new themes arose [33].

Sample Size

To obtain maximal variation, a purposive sampling strategy was used [33]. The inclusion criteria were the same as for the quantitative part of the study, but they also ensured representation of differences in age and gender.

Analysis

Analysis of the Questionnaires

Only fully completed questionnaires were analyzed (N=120). There were no missing data, as the questionnaire was only considered complete if all the questions were answered. According to the SUS guidelines, we performed an individual analysis of each participant's SUS score as well as the mean value for the entire population. We separated the 2 self-constructed questions from the original SUS questions in the calculation and interpretation process to ensure that they were accurate and reliable. The final score was between 0 and 100, where a higher score indicates better usability. Odd-numbered questions were positive in tone, and even-numbered questions were negative in tone, so the scale was converted into points ranging from 1 to 5 (1=strongly disagree to 5=strongly agree). The final score was calculated as follows: X = the sum of the points for all odd-numbered questions minus 5. And Y = 25 minus the sum of the points for all even-numbered questions. SUS score = (X + Y) × 2.5 [34]. A system needs a score above 70 to be considered acceptable; better systems will score from the high 70s to the high 80s, and excellent systems will score above 90 [27].

Analysis of the Interviews

The qualitative interviews were analyzed and reported based on Malterud's [35] systematic text condensation. This process consisted of four steps: (1) transcriptions were read several times to get a total impression of the data and to find preliminary themes; (2) we identified and sorted meaning units based on the preliminary themes and arranged them into code groups; (3) the code groups were reviewed, and the content was reduced into condensates; and (4) the meaning and content of the condensates were synthesized and interpreted [35]. The analysis was completed by CØ using NVivo (version 12; QSR International). The trustworthiness and rigor of the qualitative part of the study were evaluated using Guba's [36] definition of quality criteria. As part of steps 2, 3, and 4 in the analysis, the emerging themes and codes were discussed in the author group toward strengthening the credibility and reflexivity of our interpretation of the interviews. Using a systematic approach toward the analysis strategy of all interviews ensured confirmability in the data collection and analysis process.

The SQUIRE 2.0 checklist [37] was used to create transparency and ensure that no important information was missed in the reporting of the study.

Integration of Quantitative and Qualitative Results

To achieve an expanded understanding of the results, the qualitative and quantitative results were compared and integrated as the final step of the analysis using joint display tables [24]. In a joint display table, the 2 results are presented in a way that allows comparison, leading to confirmation, disconfirmation, or expansion of each other [24]. The results from the SUS (quantitative results) are presented on a Likert scale, showing the variation of the grades in the different questions. To elaborate on and verify the answers, supportive qualitative quotes were presented for each question. We divided the grades into low (1-3) and high (4-5) to separate the different perceptions of CCL for patients.

Ethical Considerations

All the participants received verbal and written information about the study in accordance with applicable ethical rules [38] and provided their oral and written consent. The study is registered with the Danish Data Protection Agency, Fortegnelsen (19/22672). Approval of the project was granted by the Regional Committee on Health Research Ethics for Southern Denmark (S-20192000–111).

Results

Quantitative Results

In total, 14 family members and 106 patients agreed to participate. A total of 27 patients declined to participate for three main reasons: (1) no interest, (2) no technical skills, and (3) a lack of mental ability due to the acute situation.

Table 1. Demographic descriptions of the participants.

Demographic description	Patients (n=106)	Family members (n=14)	Total (N=120)
Gender, n (%)			
Female	55 (51.9)	8 (57.1)	63 (52.5)
Male	51 (48.1)	6 (42.9)	57 (47.5)
Age (years), mean (SD)	55.5 (SD 18.7)	66.5 (SD 11.6)	57 (SD 18.3)
Civil status, n (%)			
No partner	39 (36.8)	2 (14.3)	41 (34.2)
In a relationship	67 (63.2)	12 (85.7)	79 (65.8)
Children, n (%)			
Having children	81 (76.4)	14 (100.0)	95 (79.2)
Having children living at home	32 (39.5)	6 (42.9)	38 (40.0)
Technology, n (%)			
Having a smartphone	96 (90.6)	14 (100.0)	110 (91.7)
Using technology on daily basis	102 (96.2)	13 (92.9)	115 (95.8)
Education, n (%)			
Low	21 (19.8)	1 (7.1)	22 (18.3)
Medium	71 (67.0)	9 (64.3)	80 (66.7)
High	14 (13.2)	4 (28.6)	18 (15.0)

The respondents were equally represented by gender, with a mean age of 57 years. The mean age of family members was higher than that of the included patients. Most participants had medium education levels, but low and high educational levels were also represented.

Overall, the participants answered the survey positively. As displayed in Tables 2 and 3, each item could have a score contribution between 1 and 5. All the odd-numbered (positive) questions had a score contribution above 4.27-4.53, and all the even-numbered (negative) questions had a score ranging from 1.52 to 1.99. Question 1 had the most positive answers: 94.2% (113/120) strongly agreed or agreed that they would like to use the system if they were hospitalized again. Question 4 had the highest negative score value, indicating that the participants felt they needed help using the system. Of the participants, 50.8% (61/120) indicated that they were confident using the system, answering “strongly agree” to question 9, and 87.5% (105/120) strongly agreed or agreed that most people would be able to learn to use this system.

Table 2. Results of the System Usability Scale for all participants (N=120) and the System Usability Scale score contribution of individual items.

System Usability Scale analysis item	Value per 5-point Likert scale response, n (%)					Score contribution (1-5), mean (SD)
	1 (strongly disagree)	2 (disagree)	3 (neutral)	4 (agree)	5 (strongly agree)	
1. I think I would like to use this system, if I am admitted again.	0 (0)	0 (0)	7 (5.8)	42 (35)	71 (59.2)	4.53 (SD 0.61)
2. I found the system unnecessarily complex.	61 (50.8)	41 (34.2)	11 (9.2)	5 (4.2)	2 (1.7)	1.72 (SD 0.92)
3. I thought the system was easy to use.	1 (0.8)	1 (0.8)	8 (6.7)	39 (32.5)	71 (59.2)	4.48 (SD 0.73)
4. I think that I would need help from the staff to be able to use this system.	49 (40.8)	44 (36.7)	11 (9.2)	11 (9.2)	5 (4.2)	1.99 (SD 1.12)
5. I found the various functions in the system to be well correlated.	1 (0.8)	1 (0.8)	8 (6.7)	65 (54.2)	45 (37.5)	4.27 (SD 0.69)
6. I thought there was too much inconsistency in this system.	56 (46.7)	52 (43.3)	8 (6.7)	1 (0.8)	3 (2.5)	1.69 (SD 0.84)
7. I would imagine that most people would learn to use this system very quickly.	0 (0)	0 (0)	15 (12.5)	53 (44.2)	52 (43.3)	4.31 (SD 0.68)
8. I found the system very cumbersome to use.	70 (58.3)	42 (35)	5 (4.2)	2 (1.7)	1 (0.8)	1.52 (SD 0.73)
9. I felt very confident using the system.	2 (1.7)	6 (5)	4 (3.3)	47 (39.2)	61 (50.8)	4.33 (SD 0.89)
10. I needed to learn a lot things before I could get going with this system.	67 (55.8)	42 (35)	7 (5.8)	4 (3.3)	0 (0)	1.57 (SD 0.75)

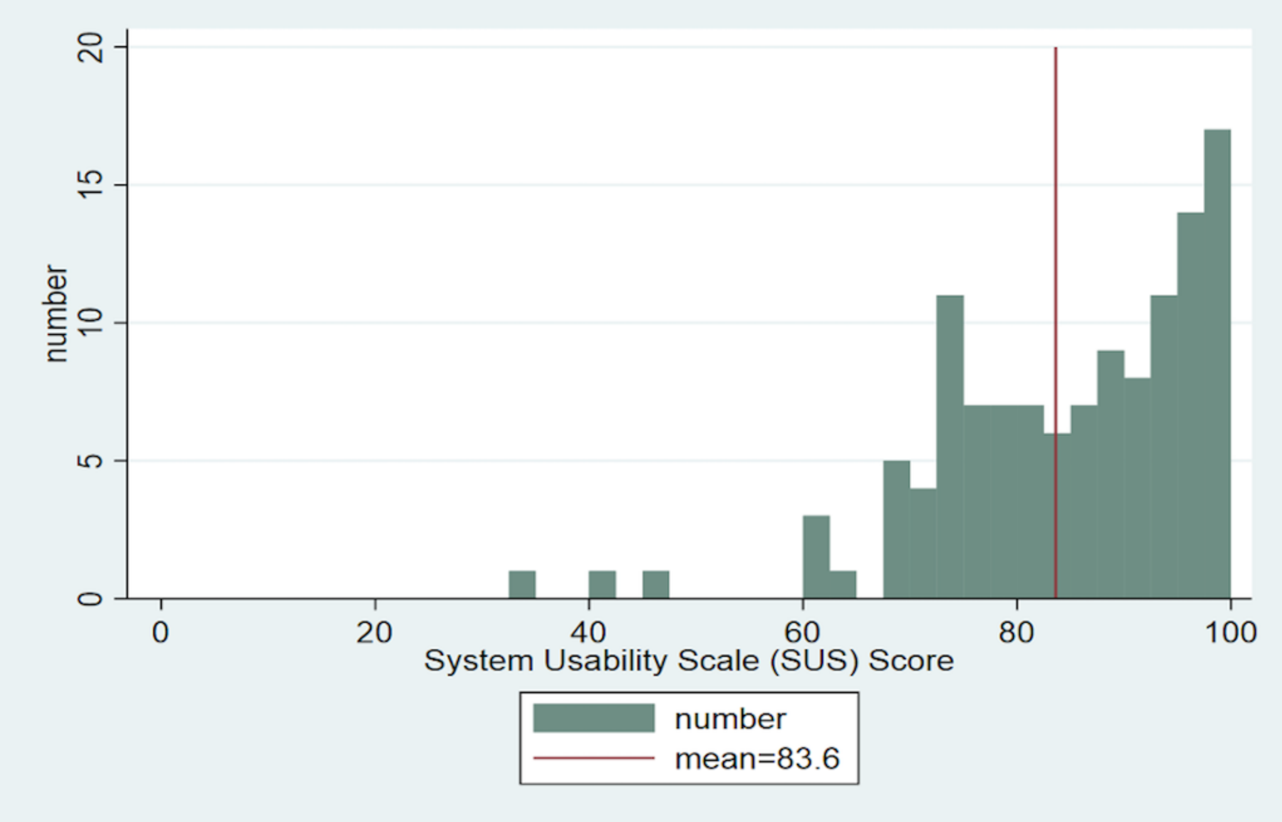
Based on the answers to the 2 self-constructed questions, (Table 3), 57.5% (69/120) of the participants strongly agreed that CCL for patients provided a great overview of their stay, and 87.5% (105/120) agreed or strongly agreed that the information in the system made sense to them.

Table 3. Results of general questions calculated by System Usability Scale principles for all participants (N=120) and the System Usability Scale score contribution of individual items.

System Usability Scale analysis item	Value per 5-point Likert scale response, n (%)					Score contribution (1–5), mean (SD)
	1 (strongly disagree)	2 (disagree)	3 (neutral)	4 (agree)	5 (strongly agree)	
11. I think the system provided a great overview of my stay.	0 (0)	2 (1.7)	10 (8.3)	39 (32.5)	69 (57.5)	4.46 (SD 0.72)
12. I think the information in the system made sense to me.	1 (0.8)	5 (4.2)	9 (7.5)	45 (37.5)	60 (50)	4.32 (SD 0.85)

For all participants, the total mean score for the SUS scale was 83.6 (SD 12.8), indicating that the system had close to excellent usability. The median score was 85, and Figure 4 [27] shows the distribution of the individual answers. The scores covered the entire range from 0 to 20 persons per score, and the majority of individuals scored above 70.

Figure 4. Overview of the System Usability Scale (SUS) rating table with inserted value ranges [27].



Qualitative Results

A total of 10 patients and 3 family members (1 daughter aged 55 years, 1 son aged 65 years, and a husband aged 37 years) were interested in elaborating on their experience of CCL for patients after they had tested the system, and the questionnaire was completed. The patients were aged between 32 and 96 years, with equal representation of men and women, and 3 patients were retired.

The following three main themes emerged from the analysis: (1) future perspectives on usability and design; (2) means toward empowerment; and (3) family implications. These themes will be elaborated on using quotes in the upcoming sections.

Future Perspectives on Usability and Design

The majority of the participants expressed a very positive attitude toward CCL for patients but also offered ideas for the future design of the system. The part of CCL for patients that displayed the estimated waiting time in the ED was found to be intuitive and easy to understand and provided informative insights that prepared the participants for their length of stay. This reduced their frustration with not knowing. However, they expressed concerns about the system's lack of familiarity and that it could be improved if the design was like other systems they used in everyday life, such as email or smartphone apps.

> *The system was not difficult at all but I think it would benefit from more recognizability with others systems, for example, email or iPhone applications.* [Male in his 60s]

Most participants valued the line that displayed the boxes with activities the most. They found this part of the system to be essential, as it was the only part that provided direct, personalized information. While they all expressed that they were able to understand the meaning of the changing colors, they also suggested that the text in the boxes could be provided in plain language or a "help" function with text or video could be used to explain the activity in the box.

> *The line with the boxes could be much larger, as this is the most important part! It would be great if you could choose whether you would like to see only the line or all actions on the screen.* [Joint interview, male patient in his 80s and daughter in her 50s]

The participants all watched more than one video, and there was a consensus that the content in the videos was helpful. A few patients who were placed in the hallway due to crowding found it difficult to listen to the videos not only because of the general noise but also because they were afraid of disturbing others. However, the information provided by using videos instead of text was appreciated.

> *The content in the videos was exactly the information I needed. It was nice to be able to revisit the information in the video.* [Female patient in her 50s]

A participant found CCL for patients to be too general. More personalized information, such as individual test results, should be incorporated. Moreover, patients who were visually impaired found the system difficult to use.

> *I have difficulties with my vision, and I do not think I would have been able to use this without help.* [Female patient in her 60s]

Means Toward Empowerment

All of the participants agreed that the system provided an overview that otherwise would not have been accessible for them. Knowing who their treating nurse and doctor were calmed the participants. They described a feeling of not being forgotten in the hectic environment of the ED. Moreover, they valued being able to follow when activities changed from passive to active. Consistency between actions on the screen and in real life provided them with confidence in health care professionals.

> *When you are here, you can hear people working, but you do not know if anyone is taking care of your situation, or you are forgotten. The system helped us to believe we were not forgotten (…) We loved that when something happens on the screen then it was also reflected in real life. E.g. when the screen said the doctor was on his way- he actually came.* [Joint interview, female patient, and husband in their 30s]

Several of the participants stated that having CCL for patients available made them feel calm, as the system provided predictability. Further, having an overview helped them to remain in control of the course of treatment in the ED. Some of the participants said this system could save the nurses' time, as they felt they were more empowered to handle the situation in the ED since they knew what they were waiting for.

> *The questions that I would have needed a nurse to answer were provided by the system; that was really great.* [Female patient in her 30s]

A few of the participants were worried that CCL for patients would need resources from health care professionals that were already scarce.

> *I am worried that the system takes time away from the patients to support the system.* [Joint interview, male patient in his 80s and daughter in her 50s]

Family Implications

Both patients and family members indicated that giving family members direct access was important. CCL for patients gave the family up-to-date information about the care and treatment-related interventions as soon as they attended the hospital room, and they did not need to wait for a nurse or doctor to get an idea of what was planned.

> *My mother is not able to remember what she is planned for today. I think it was great for me to see she is waiting for X rays.* [Female patient in her 90s and son in his 60s]

Furthermore, the family members reported that the system helped them to support the patient, as they could keep track of the interventions provided by CCL for patients. Family members of older patients felt the system was too complicated for the older individuals to use but appreciated that the system was available for them because it allowed them to talk the patient through the stay in the ED. Moreover, a family member stated that the system made it possible for him to let his wife go to sleep, as they agreed that he would wake her up when he saw that activities were about to happen.

> *It was nice for me to have a system that told us when things were going to happen. My wife fell asleep, and I knew I did not need to wake her up before I could see the box turned into the blue color. It was easy to understand.* [Joint interview, female patient, and her husband in their 30s]

Merged Data

We combined the quantitative and qualitative data in a joint display (Multimedia Appendix 1), providing an assessment of the quantitative and qualitative data together. In this way, the data allow us to expand our understanding of patients and their family members' experiences with CCL for patients. For example, in question 1, the participants were asked whether they would like to use CCL for patients again. The participants who gave a lower score (1-3) to that question were concerned if the system would replace personal appearance from health care professionals, whereas those who gave it a high score (4-5) valued how the systems helped them to keep control.

Furthermore, question 7 regarding people's ability to learn to use the system revealed that the participants who gave a low score (1-3) wanted more simplicity, fearing that the older patients would find the system difficult. Meanwhile, the participants who gave high scores (4-5) felt that the system was easy to use. Regarding question 11, the majority of the patients and their family members stated that CCL for patients provided a great overview of the patient's pathway. They further elaborated on this in the interviews, as they felt that the overview of care in the system helped them to feel less stressed and better understand the treatment pathways.

Discussion

Principal Findings

In this study, we report that the perceived overall usability of the health information system CLL *for patients* is good to excellent, providing information that is needed during the entire emergency process. The participants rated the system highly (a score of 83.6 points) and reported that the system gave them an opportunity to remain in control, as they knew what they were waiting for and who was responsible for care and treatment.

Technology as a Means to Empower Patients and Family Members in the ED

Looking into previous research on testing systems using SUS [28], a mean score of 83.6, as found in this study, would indicate that the tested system was successful. However, while CCL for patients was evaluated positively overall, we also uncovered technical concerns regarding usability limitations, specifically regarding the older individuals. Our results showed a mean patient age of 57 years, which represents a relatively young ED population. However, the mean age of the family members was almost 10 years (9.5 years) older. The older individuals found the system to be complicated to use and felt that it needed simplified functions, such as a zoom function and recognizability (eg, other well-known systems). Echoing these findings, Verma et al [39] investigated the level of eHealth literacy among older adults and caregivers and found that one main barrier to the adoption of eHealth was a lack of familiarity with the tools available. In the development phase of CCL for patients [16], decisions had to be made for the system to work in a clinical setting. One decision was the use of an interface design, which did not allow us to integrate well-known functions, for example, from email or application symbols. Our results highlighted that it might not be possible to design technologies using a one-size-fits-all approach. However, in line with previous research [40], we discovered that the usability testing allowed the developers to adjust and isolate functionalities to provide improved usability outcomes in the future. For example, we found that the participants valued the display with the boxes, which could be promoted in a revised version by the availability of a zoom function.

Furthermore, the participants expressed concerns about whether CCL for patients would influence the health care professionals' available time to provide actual care. Barriers to the adoption of technology systems in clinical settings include the workflow or demand for more human resources [12]. As the information system is a redesigned patient flow system, it would not require changes in workflow or unduly burden professional health care resources. Another consideration was the need for personal test results. They could not be provided in the current form of the system, as it would require a personal log-on to avoid safety issues related to General Data Protection Regulations.

The participants who rated the usability the highest explained that the system made them feel that they were in control of the situation without the fear of being forgotten. The system provided an overview of the care transition and, therefore, offered predictability. This need to be in control has been identified in another study, which described patients' and their relatives' dissatisfaction when visiting the ED [6], as they felt powerless in the ED. Not having knowledge or information available led to such feelings of powerlessness. Nursing rounds were suggested in that study to improve information support [6]. Our results showed that the patients felt more independent because they were able to find the needed information using technology.

Being acutely ill places individuals in a vulnerable situation, and their cognitive capabilities are challenged [2]. Communication from health care professionals and how information is presented have a significant influence on how that information is comprehended [2,41,42]. In this study, we developed information videos related to the journey within the ED, and the participants reported that they were an accessible and usable way to understand information in a stressful situation. Patients and their family members declared that this gave them a feeling of empowerment. Indeed, empowering patients to be in control and involved in their own care is recognized as a core value of high-quality patient-centered care [43]. As Emmamally et al [44-46] noted, improved partnering with family members in the ED is needed. If the family is not included, there is an increased risk of miscommunication and poor understanding of health-related matters [2,44,47-49]. However, creating a closer partnership of care has been described as challenging within the ED due to the high workload, overcrowding, and multitasking [47]. This is echoed in recent findings from studies conducted in a Danish context [2,3], in which family members requested more systematic inclusion in the ED. In this study, the results showed that CCL for patients was perceived as usable and as a useful way to systematically include families during the ED stay.

An update of the Medical Research Council's guidelines for developing and evaluating complex interventions in health care states that appropriate users should be involved in every part of the development, process, and outcome analysis of a complex intervention to ensure sustainable interventions [50]. In line with best practices, the information system has been developed together with representatives of future users of the system, including health care professionals, managers, patients, family members, and IT specialists [16]. For decades, the ED context has been a hectic environment [4,42,51,52]. This creates challenges at both the information and communication levels, affecting whether patients and their families feel in better control during their stay in the ED [1,4,42,51,52]. In this study, we presented and evaluated a simple but unique system that provides timely information to empower individuals without straining health care professionals' resources. The usability test was a crucial and important step to inform changes in functionalities and experiences of using IT in the ED.

Strengths and Limitations

Questionnaires are a common and recognized method for evaluating the usability of health technologies. However, the contextual factors affecting the results are difficult to determine [53]. The SUS did not provide insights on the effectiveness or efficiency of the system, but it is a validated questionnaire and provided an overall understanding of the system [27]. The mixed methods approach [24] enabled the integration of quantitative and qualitative data. This allowed us to obtain an understanding of how the usability was rated and why the results emerged for the specific questions, which is considered a strength of usability testing [40,54].

Additionally, our findings serve as an inspiration to others about how a participatory design process can develop a technology that is aligned with some of the essential needs described by the users of the ED. The findings provide an example of how a technological solution can be used to reduce the information gap in an ED context, as the provision of adequate information to patients and their families is found to be a major challenge in an ED context [2,4,42].

This study also had some limitations. Using a broader evaluation method, for example, a qualitative evaluation questionnaire or an evaluation instrument with more domains, could potentially have provided the study with more nuances [55]. Patients attending the ED outside of the IT department's business hours were not able to use the system. Therefore, we do not know if patients attending the ED in the late evening hours or at night would rate the usability differently. Moreover, no cognitive debriefings or adjustments were made specifically for individuals attending an ED, as these tests were conducted before introducing the questionnaire. Multimedia Appendix 2 [26,56] contains further details about the process as well as final modifications to the questionnaire. In addition, our results are based on a relatively young population (with a mean age of 57 years). Another weakness is that we did not include all users in the evaluation phase, as health care professionals, IT specialists, and managers were only involved in the development phase and not in the usability testing. For the system to be fully useful, it must run on its own or be serviced directly in the ED. These aspects will be considered in the planning of a future implementation process. Moreover, the transferability of the results is limited to countries with comparable access to and understanding of technologies, as in the Danish population and health care system.

Conclusion

Based on the results of this study, the usability of CCL for patients is rated close to excellent by patients and family members. CCL for patients was perceived to be useful, as it enabled understanding of the ED treatment and pathway. The patients indicated that they, from the technology, were able to understand what was going to happen, experienced the feeling of being in control, and found the information to be useful. Areas for improvement include making the system more usable for the older individuals. It is concluded that a technological solution can be used to minimize the information gap in an ED context from the perspective of patients and their family members.

Acknowledgments

The authors would like to thank all those who participated in this study, OPEN [29], and Niels I Agerbek for data management.

The authors received financial support for the research, authorship, and publication of this study from the Region of Southern Denmark, the University of Southern Denmark, and Odense University Hospital.

Authors' Contributions

CØ, CMJ, KBD, EC, and AL wrote the protocol for the study. CØ and AL performed the data analysis. CØ wrote the first draft of the manuscript. All the authors reviewed and edited the manuscript and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Joint display.
[PDF File (Adobe PDF File), 55 KB - humanfactors_v11i1e48445_app1.pdf]

Multimedia Appendix 2
Supplementary file for the Methods section.
[PDF File (Adobe PDF File), 61 KB - humanfactors_v11i1e48445_app2.pdf]

References

1. Graham B, Endacott R, Smith JE, Latour JM. 'They do not care how much you know until they know how much you care': a qualitative meta-synthesis of patient experience in the emergency department. Emerg Med J 2019;36(6):355-363. [doi: 10.1136/emermed-2018-208156] [Medline: 31003992]
2. Østervang C, Lassen AT, Jensen CM, Coyne E, Dieperink KB. What are the needs and preferences of patients and family members discharged from the emergency department within 24 hours? A qualitative study towards a family-centred approach. BMJ Open 2021;11(11):e050694 [FREE Full text] [doi: 10.1136/bmjopen-2021-050694] [Medline: 34753758]
3. Østervang C, Lassen AT, Øelund K, Coyne E, Dieperink KB, Jensen CM. Health professionals' perspectives of patients' and family members' needs in emergency departments and patient pathway improvement: a qualitative study in Denmark. BMJ Open 2022;12(2):e057937 [FREE Full text] [doi: 10.1136/bmjopen-2021-057937] [Medline: 35228292]
4. Sonis JD, Aaronson EL, Lee RY, Philpotts LL, White BA. Emergency department patient experience: a systematic review of the literature. J Patient Exp 2018;5(2):101-106 [FREE Full text] [doi: 10.1177/2374373517731359] [Medline: 29978025]
5. Hermann RM, Long E, Trotta RL. Improving patients' experiences communicating with nurses and providers in the emergency department. J Emerg Nurs 2019;45(5):523-530. [doi: 10.1016/j.jen.2018.12.001] [Medline: 30655007]
6. Forsgärde ES, Attebring MF, Elmqvist C. Powerlessness: dissatisfied patients' and relatives' experience of their emergency department visit. Int Emerg Nurs 2016;25:32-36. [doi: 10.1016/j.ienj.2015.07.004] [Medline: 26254101]
7. Leventhal EL, Schreyer KE. Information management in the emergency department. Emerg Med Clin North Am 2020;38(3):681-691. [doi: 10.1016/j.emc.2020.03.004] [Medline: 32616287]
8. Currie WL, Seddon JJM. A cross-national analysis of eHealth in the European Union: some policy and research directions. Inf Manag 2014;51(6):783-797. [doi: 10.1016/j.im.2014.04.004]
9. Almasi S, Rabiei R, Moghaddasi H, Vahidi-Asl M. Emergency department quality dashboard; a systematic review of performance indicators, functionalities, and challenges. Arch Acad Emerg Med 2021;9(1):e47 [FREE Full text] [doi: 10.22037/aaem.v9i1.1230] [Medline: 34405145]

10. Du Y, Dennis B, Rhodes SL, Sia M, Ko J, Jiwani R, et al. Technology-assisted self-monitoring of lifestyle behaviors and health indicators in diabetes: qualitative study. JMIR Diabetes 2020;5(3):e21183 [FREE Full text] [doi: 10.2196/21183] [Medline: 32857056]
11. Westphal M, Yom-Tov GB, Parush A, Carmeli N, Shaulov A, Shapira C, et al. A patient-centered information system (myED) for emergency care journeys: design, development, and initial adoption. JMIR Form Res 2020;4(2):e16410 [FREE Full text] [doi: 10.2196/16410] [Medline: 32130144]
12. Østervang C, Vestergaard LV, Dieperink KB, Danbjørg DB. Patient rounds with video-consulted relatives: qualitative study on possibilities and barriers from the perspective of healthcare providers. J Med Internet Res 2019;21(3):e12584 [FREE Full text] [doi: 10.2196/12584] [Medline: 30907746]
13. Kushniruk A, Nøhr C. Participatory design, user involvement and health IT evaluation. Stud Health Technol Inform 2016;222:139-151. [Medline: 27198099]
14. Lau AY, Piper K, Bokor D, Martin P, Lau VS, Coiera E. Challenges during implementation of a patient-facing mobile app for surgical rehabilitation: feasibility study. JMIR Hum Factors 2017;4(4):e31 [FREE Full text] [doi: 10.2196/humanfactors.8096] [Medline: 29217504]
15. Simonsen J, Robertson T. Routledge International Handbook Of Participatory Design. London: Routledge; 2012.
16. Østervang C, Lassen A, Schmidt T, Coyne E, Dieperink KB, Jensen CM. Development of a health information system to promote emergency care pathways: a participatory design study. Digit Health 2022;8:20552076221145856 [FREE Full text] [doi: 10.1177/20552076221145856] [Medline: 36601282]
17. Farzandipour M, Nabovati E, Tadayon H, Jabali MS. Usability evaluation of a nursing information system by applying cognitive walkthrough method. Int J Med Inform 2021;152:104459. [doi: 10.1016/j.ijmedinf.2021.104459] [Medline: 34091145]
18. Bligård LO, Osvalder AL. Enhanced cognitive walkthrough: development of the cognitive walkthrough method to better predict, identify, and present usability problems. Adv Hum Comput Interact 2013;2013:1-17 [FREE Full text] [doi: 10.1155/2013/931698]
19. Østervang C, Lassen AT, Jensen CM, Coyne E, Dieperink KB. How to improve emergency care to adults discharged within 24 hours? Acute care planning in emergency departments (The ACE study): a protocol of a participatory design study. BMJ Open 2020;10(12):e041743 [FREE Full text] [doi: 10.1136/bmjopen-2020-041743] [Medline: 33371037]
20. Kensing F. Methods and Practices in Participatory Design. Copenhagen: ITU Press; 2003.
21. Nielsen C, Agerskov H, Bistrup C, Clemensen J. Evaluation of a telehealth solution developed to improve follow-up after kidney transplantation. J Clin Nurs 2020;29(7-8):1053-1063. [doi: 10.1111/jocn.15178] [Medline: 31889357]
22. Jensen CM, Overgaard S, Wiil UK, Clemensen J. Can tele-health support self-care and empowerment? A qualitative study of hip fracture patients' experiences with testing an "app". SAGE Open Nurs 2019;5:2377960819825752 [FREE Full text] [doi: 10.1177/2377960819825752] [Medline: 33415217]
23. Getinge. 2022. URL: http://www.cetrea.com/index.php/da/ [accessed 2024-01-19]
24. Creswell JW, Clark VLP. Designing and Conducting Mixed Methods Research, Third Edition. Thousand Oaks, California: Sage Publications; 2018.
25. Lewis JR. The system usability scale: past, present, and future. Int J Hum-Comput Int 2018;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
26. Hvidt JCS, Christensen LF, Sibbersen C, Helweg-Jørgensen S, Hansen JP, Lichtenstein MB. Translation and validation of the System Usability Scale in a Danish mental health setting using digital technologies in treatment interventions. Int J Hum-Comput Int 2019;36(8):709-716. [doi: 10.1080/10447318.2019.1680922]
27. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum-Comput Int 2008;24(6):574-594. [doi: 10.1080/10447310802205776]
28. Hägglund M, Scandurra I. User evaluation of the Swedish patient accessible electronic health record: system usability scale. JMIR Hum Factors 2021;8(3):e24927 [FREE Full text] [doi: 10.2196/24927] [Medline: 34313596]
29. OPEN. Open patient data explorative network. Clinical Institute University of Southern Denmark. 2022. URL: https://www.sdu.dk/da/forskning/open [accessed 2024-01-19]
30. Green J, Thorogood N. Qualitative Methods for Health Research, Third Edition. Los Angeles: SAGE Publications; 2014.
31. Birt L, Scott S, Cavers D, Campbell C, Walter F. Member checking: a tool to enhance trustworthiness or merely a nod to validation? Qual Health Res 2016;26(13):1802-1811. [doi: 10.1177/1049732316654870] [Medline: 27340178]
32. Kvale S, Brinkmann S. Interview: introduktion til et håndværk, 2. udgave. Kbh: Hans Reitzel; 2009.
33. Moser A, Korstjens I. Series: practical guidance to qualitative research. part 3: sampling, data collection and analysis. Eur J Gen Pract 2018;24(1):9-18 [FREE Full text] [doi: 10.1080/13814788.2017.1375091] [Medline: 29199486]
34. Brooke J. SUS: a 'quick and dirty' usability scale. In: Jordan PW, McClelland IL, Thomas B, Weerdmeester BA, editors. Usability Evaluation in Industry. Boca Raton, FL: CRC Press; 1996:189-194.
35. Malterud K. Systematic text condensation: a strategy for qualitative analysis. Scand J Public Health 2012;40(8):795-805. [doi: 10.1177/1403494812465030] [Medline: 23221918]
36. Guba EG. Criteria for assessing the trustworthiness of naturalistic inquiries. ECTJ 1981;29(2):75-91. [doi: 10.1007/bf02766777]

37. Ogrinc G, Mooney SE, Estrada C, Foster T, Goldmann D, Hall LW, et al. The SQUIRE (Standards for QUality Improvement Reporting Excellence) guidelines for quality improvement reporting: explanation and elaboration. Qual Saf Health Care 2008;17(Suppl 1):i13-i32 [FREE Full text] [doi: 10.1136/qshc.2008.029058] [Medline: 18836062]
38. World Medical Association. World Medical Association declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013;310(20):2191-2194 [FREE Full text] [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
39. Verma R, Saldanha C, Ellis U, Sattar S, Haase KR. eHealth literacy among older adults living with cancer and their caregivers: a scoping review. J Geriatr Oncol 2022;13(5):555-562. [doi: 10.1016/j.jgo.2021.11.008] [Medline: 34810146]
40. Alwashmi MF, Hawboldt J, Davis E, Fetters MD. The iterative convergent design for mobile health usability testing: mixed methods approach. JMIR Mhealth Uhealth 2019;7(4):e11656 [FREE Full text] [doi: 10.2196/11656] [Medline: 31025951]
41. Samuels-Kalow M, Rhodes K, Uspal J, Reyes Smith A, Hardy E, Mollen C. Unmet needs at the time of emergency department discharge. Acad Emerg Med 2016;23(3):279-287 [FREE Full text] [doi: 10.1111/acem.12877] [Medline: 26683867]
42. Vaillancourt S, Seaton MB, Schull MJ, Cheng AH, Beaton DE, Laupacis A, et al. Patients' perspectives on outcomes of care after discharge from the emergency department: a qualitative study. Ann Emerg Med 2017;70(5):648-658.e2. [doi: 10.1016/j.annemergmed.2017.05.034] [Medline: 28712607]
43. Pekonen A, Eloranta S, Stolt M, Virolainen P, Leino-Kilpi H. Measuring patient empowerment—a systematic review. Patient Educ Couns 2020;103(4):777-787. [doi: 10.1016/j.pec.2019.10.019] [Medline: 31767243]
44. Emmamally W, Brysiewicz P. Families' perceptions of support from health care professionals in the three emergency departments in KwaZulu Natal, South Africa. Int J Africa Nurs Sci 2019;10:55-60 [FREE Full text] [doi: 10.1016/j.ijans.2019.01.004]
45. Emmamally W, Erlingsson C, Brysiewicz P. Partnering with families: Is the ED ready? Int Emerg Nurs 2019;42:1. [doi: 10.1016/j.ienj.2019.01.002] [Medline: 30660319]
46. Emmamally W, Erlingsson C, Brysiewicz P. Families' perceptions of relational practice in the emergency department: a qualitative study. Int Emerg Nurs 2020;51:100877. [doi: 10.1016/j.ienj.2020.100877] [Medline: 32479929]
47. Brysiewicz P, Emmamally W. Focusing on families in the emergency department. Int Emerg Nurs 2017;30:1-2. [doi: 10.1016/j.ienj.2016.08.002] [Medline: 27727139]
48. Hoffmann E, Andersen PT, Mogensen CB, Prinds C, Primdahl J. Relatives' negotiation power in relation to older people's acute hospital admission: a qualitative interview study. Scand J Caring Sci 2022;36(4):1016-1026. [doi: 10.1111/scs.13012] [Medline: 34156115]
49. Hoffmann E, Olsen PR. Like an ace up the sleeve: an interview study of nurses' experiences of the contact with relatives in a somatic emergency ward. Scand J Caring Sci 2018;32(3):1207-1214. [doi: 10.1111/scs.12566] [Medline: 29488643]
50. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021;374:n2061 [FREE Full text] [doi: 10.1136/bmj.n2061] [Medline: 34593508]
51. Stevens L, Fry M, Browne M, Barnes A. Fast track patients' satisfaction, compliance and confidence with emergency department discharge planning. Australas Emerg Care 2019;22(2):87-91. [doi: 10.1016/j.auec.2019.01.004] [Medline: 31042527]
52. Nairn S, Whotton E, Marshal C, Roberts M, Swann G. The patient experience in emergency departments: a review of the literature. Accid Emerg Nurs 2004;12(3):159-165. [doi: 10.1016/j.aaen.2004.04.001] [Medline: 15234713]
53. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: a scoping review. Int J Med Inform 2019;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
54. Rochat J, Ehrler F, Siebert JN, Ricci A, Ruiz VG, Lovis C. Usability testing of a patient-centered mobile health app for supporting and guiding the pediatric emergency department patient journey: mixed methods study. JMIR Pediatr Parent 2022;5(1):e25540 [FREE Full text] [doi: 10.2196/25540] [Medline: 35289754]
55. Kidholm K, Ekeland AG, Jensen LK, Rasmussen J, Pedersen CD, Bowes A, et al. A model for assessment of telemedicine applications: mast. Int J Technol Assess Health Care 2012;28(1):44-51 [FREE Full text] [doi: 10.1017/S0266462311000638] [Medline: 22617736]
56. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health 2005;8(2):94-104 [FREE Full text] [doi: 10.1111/j.1524-4733.2005.04054.x] [Medline: 15804318]

Abbreviations

ED: emergency department
PC: personal computer
SUS: System Usability Scale

Edited by P Santana-Mancilla; submitted 26.04.23; peer-reviewed by R Rabiei, R Suomi; comments to author 10.10.23; revised version received 30.11.23; accepted 17.12.23; published 21.02.24.

Please cite as:
Østervang C, Jensen CM, Coyne E, Dieperink KB, Lassen A
Usability and Evaluation of a Health Information System in the Emergency Department: Mixed Methods Study
JMIR Hum Factors 2024;11:e48445
URL: https://humanfactors.jmir.org/2024/1/e48445
doi:10.2196/48445
PMID:38381502

Original Paper

Intention to Use an Electronic Community Health Information System Among Health Extension Workers in Rural Northwest Ethiopia: Cross-Sectional Study Using the Unified Theory of Acceptance and Use of Technology 2 Model

Tesfahun Hailemariam[1], PhD; Asmamaw Atnafu[2], PhD; Lemma Gezie[3], PhD; Jens Kaasbøll[4], Prof Dr; Jorn Klein[5], Prof Dr; Binyam Tilahun[1], PhD

[1]Department of Health Informatics, Institute of Public Health, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia
[2]Department of Health System and Policy, Institute of Public Health, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia
[3]Department of Epidemiology and Biostatistics, Institute of Public Health, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia
[4]Department of Informatics, University of Oslo, Oslo, Norway
[5]Department of Nursing and Health Sciences Campus Porsgrunn, University of South-Eastern Norway, Porsgrunn, Norway

Corresponding Author:
Tesfahun Hailemariam, PhD
Department of Health Informatics, Institute of Public Health
College of Medicine and Health Sciences
University of Gondar
Chechela street
Gondar, 196
Ethiopia
Phone: 251 913305250
Email: tesfahunhailemariam@gmail.com

Abstract

Background: IT has brought remarkable change in bridging the digital gap in resource-constrained regions and advancing the health care system worldwide. Community-based information systems and mobile apps have been extensively developed and deployed to quantify and support health services delivered by community health workers. The success and failure of a digital health information system depends on whether and how it is used. Ethiopia is scaling up its electronic community health information system (eCHIS) to support the work of health extension workers (HEWs). For successful implementation, more evidence was required about the factors that may affect the willingness of HEWs to use the eCHIS.

Objective: This study aimed to assess HEWs' intentions to use the eCHIS for health data management and service provision.

Methods: A cross-sectional study design was conducted among 456 HEWs in 6 pilot districts of the Central Gondar zone, Northwest Ethiopia. A Unified Theory of Acceptance and Use of Technology model was used to investigate HEWs' intention to use the eCHIS. Data were cleaned, entered into Epi-data (version 4.02; EpiData Association), and exported to SPSS (version 26; IBM Corp) for analysis using the AMOS 23 Structural Equation Model. The statistical significance of dependent and independent variables in the model was reported using a 95% CI with a corresponding P value of <.05.

Results: A total of 456 HEWs participated in the study, with a response rate of 99%. The mean age of the study participants was 28 (SD 4.8) years. Our study revealed that about 179 (39.3%; 95% CI 34.7%-43.9%) participants intended to use the eCHIS for community health data generation, use, and service provision. Effort expectancy (β=0.256; P=.007), self-expectancy (β=0.096; P=.04), social influence (β=0.203; P=.02), and hedonic motivation (β=0.217; P=.03) were significantly associated with HEWs' intention to use the eCHIS.

Conclusions: HEWs need to be computer literate and understand their role with the eCHIS. Ensuring that the system is easy and enjoyable for them to use is important for implementation and effective health data management.

(JMIR Hum Factors 2024;11:e47081) doi:10.2196/47081

KEYWORDS

data capturing; data use; eCHIS; electronic community health information system; health extension worker; HEW; intention to use; service provision; Unified Theory of Acceptance and Use of Technology 2; UTAUT2 model

Introduction

Though IT has demonstrated remarkable promise in closing the digital divide in resource-constrained regions and advancing the health care system, there is a global shortage of health workers, which prevents at least half of the world's population from receiving essential health services [1,2]. Training community health workers (CHWs) in low- and middle-income countries has been a recommended approach to closing the global shortage of health workers [2].

Ethiopia has been implementing the Health Extension Program (HEP) since 2003, comprising female health extension workers (HEWs) to improve the health status of its population [3]. Though various strategies were implemented, and substantial progress was observed in enhancing the community health information system (eCHIS), the performance of HEWs has remained low. The possible reasons for the low performance of HEWs are increased workload; lack of motivation; negligence; and skill gaps in health data production, use, and service provision [4-8].

Due to individual, organizational, and interpersonal level impediments, in most resource-constrained countries, particularly in sub-Saharan regions, health care data generated and used for decision-making are incomplete and inaccurate [9,10]. Likewise, quality health data generation and evidence-based decision-making practices are the remaining challenges for the health care system in Ethiopia [8,11,12].

The growing evidence shows that the penetration of mobile technology improves health service delivery and health outcomes across the world [13-19] and is becoming a solution to strengthen health care industries [8,20-22]. Previous studies in Ethiopia [23], Ghana [24], Uganda [25], South Dakota [26], Indonesia [27], Canada [28], Taiwan [29], South Korea [30], and Jordan [31] indicate that data collection using electronic systems may save time over manual data collection [32,33], and there is the potential to improve health care and the productivity of health staff. For example, digital health solutions may enable CHWs to generate quality health data [34], improve health care delivery [35], and help CHWs be more effective in their job at the community level [32,36].

The eCHIS is one of the evidence-based mobile platforms for CHWs in resource-constrained countries [37], which is an easily customizable mobile health (mHealth) platform for health workers to track and support their interactions with clients. It replaces the conventional practice of a CHW manually tracking their work and carrying large client data and documentation [37].

To tackle the challenges that existed with manual health data generation, use, and service provision, the Ethiopian Federal Ministry of Health has taken the initiative to digitize the existing paper-based Community Health Information System through the eCHIS and started piloting it in 6 districts of the Central Gondar zone, Northwest Ethiopia. The ultimate goal of its implementation is to improve the quality of health data production and service delivery at the community level by transforming the culture of information use by using tablet devices.

The first component of the eCHIS is the HEW component, which supports HEWs in family folder management and the provision of reproductive, maternal, newborn, and child health service delivery and follow-up. The second component is the health center referral component, which enables health center workers to confirm referrals and provide referral feedback to HEWs. The focal person component is the third component, which assists focal persons who are designated at the health center level to provide technical and programmatic support to the HEWs. Therefore, it enables HEWs to manage health post–level data and service provision, as it facilitates referral linkage of clients from health posts to health centers and vice versa.

Although using health system technology has expanded worldwide to leverage quality health data production and use, there is a paucity of evidence on users' behavioral intention to use health system technology [38]. The intention to use a new system is how much a health care provider intends, plans, and predicts their future behavioral readiness to use health care technology [39]. Studies show that users' behavioral intention is one of the significant factors of technology acceptance and use.

Hence, it is critical to evaluate the level of users' intention to use IT before implementing it in the health care system [40-42], as it has a significant role in planning and designing effective implementation strategies for health care programs [43]. Moreover, identifying the level of intention to use the eCHIS for community health data production, use, and service provision and its influencing factors could help to be effective in the implementation and strengthening of the program. To the authors' understanding, the level of HEWs' intention to use the eCHIS for community health care data generation, use, and service provision has not been tested using the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) model.

The UTAUT2 model is one of the most mature IT models [44] that has emerged from 8 theoretical models that were primarily developed in psychology and sociology [45]. These include the Technology Acceptance Model, Theory of Planned Behavior, Combined Technology Acceptance Model and Theory of Planned Behavior, Theory of Reasoned Action, Motivational Model, Social Cognitive Theory, Model of PC Utilization, and Innovation Diffusion Theory [45,46].

The UTAUT2 has 3 broad types of integration of concepts. First, the integration was examined in new contexts, new users, and new cultural settings [46]. Second, the addition of new constructs increased the scope of dependent predictors [45]. Third, including independent predictors of the Unified Theory of Acceptance and Use of Technology (UTAUT) variables made comprehension easier [46]. Its extensive replications, applications, and integration extend the theoretical limits of technology adoption. Therefore, the addition of the 3 predictors (hedonic motivation, price value, and habit) to the previously existing 4 constructs in the original UTAUT model (performance expectancy, effort expectancy, social influence, and facilitating conditions) leveraged the adoption and use of technology (eCHIS in this case). This changes the existing relationships of constructs in the original UTAUT and introduces new relationships among constructs known in the UTAUT2.

We used the UTAUT2 constructs to determine HEWs' behavioral intention to use the eCHIS [46], as UTAUT2 perspectives are applicable in the health system and the eCHIS is a form of health system technology. Understanding the intention of HEWs using the UTAUT2 model would give insights to health system leaders on how to digitize community health systems in local settings.

Therefore, this study aimed to investigate HEWs' intention to use the eCHIS and its predicting factors using the UTAUT2 model among HEWs who had received familiarization training on the eCHIS in 6 pilot districts of Northwest Ethiopia.

Since the eCHIS is a form of health system technology, the relationships between UTAUT2 perspectives on accepting and using technology apply to the eCHIS, and the following hypotheses were speculated:

- Hypothesis 1: performance expectancy positively influences HEWs' behavioral intention when using the eCHIS.
- Hypothesis 2: effort expectancy positively influences HEWs' behavioral intention when using the eCHIS.
- Hypothesis 3: social influence positively influences HEWs' behavioral intentions when using the eCHIS.
- Hypothesis 4: facilitating conditions positively influence HEWs' behavioral intentions when using the eCHIS.
- Hypothesis 5: hedonic motivation positively influences HEWs' behavioral intention in using the eCHIS.
- Hypothesis 6: self-efficacy positively influences HEWs' behavioral intention when using the eCHIS.
- Hypothesis 7: habit positively influences HEWs' behavioral intention when using the eCHIS.

In this study, price value was not included in this model because HEWs, the participants in this study, were not directly involved in purchasing the system. Furthermore, the model was not tested on behavioral intention to use the eCHIS among HEWs in Ethiopia.

Methods

Study Design, Period, and Setting

A cross-sectional study design was conducted from January to February 2021 in the Central Gondar zone, Northwest Ethiopia. The Central Gondar zone has 15 districts, of which 6 districts (Wogera, Mirab Dembia, Misrak Dembia, Enfranz, Takusa, and Belesa) were selected as pilot districts for eCHIS implementation in the zone. The estimated total population of the zone was 2,288,440. The zone has a total of 75 health centers and 404 health posts, and there were 897 HEWs (59 urban and 848 rural) during the study period (Central Gondar Zone Health Bureau report, unpublished data, 2020).

Population and Participants of the Study

The source population of the study was HEWs at the primary health care unit level. The study participants were HEWs who were in the pilot districts of the Central Gondar zone and had received initial training for eCHIS implementation. The intervention was skill-oriented training for the implementers of mobile-based community health information system applications based on the training manual prepared by the Ministry of Health, and the training was provided for 1 week by trainers from the regional health bureau and the Ministry of Health. Following the training, each *woreda* (district) led household registration, tablet usage guideline provision, technical support and mentoring, and periodical communications for 1 year.

Provision of Mentorship and Technical Assistance

The University of Gondar assigned three supporting team members who provided technical assistance for implementers with a local mentor every 2 weeks throughout the intervention period. In addition, 1 health information technician (a local mentor) was assigned to provide mentorship and solve eCHIS-related problems during implementation.

Sample Size and Sampling Procedures

The initial sample size was calculated using a single population proportion formula, considering the following assumptions: 50% proportion of intention to use the eCHIS, as there was limited evidence in the area; 95% confidence level for estimations; and 5% margin of error. Using these inputs, the initial sample size was estimated at 385. Considering a 10% nonresponse rate, the final sample size was 422. In the pilot districts, however, the total number of HEWs was 460. Therefore, as the initially determined sample size was closer to the population size, it was planned to include all eligible HEWs in the study.

Study Variables and Measurement

The dependent variable was the intention to use the eCHIS for health data generation and service provision. Based on the UTAUT2, 8 constructs with a 5-point Likert scale were used to assess the intention to use the eCHIS and were considered potential predictors of the study [46].

1. Performance expectancy: the extent to which people believe that using a new technology can improve their job performance [47].
2. Effort expectancy: the degree of ease of use associated with the usage of a new technology [46].
3. Social influence: the degree of importance others recognize in using a new system [45].
4. Facilitating conditions: the degree to which a person perceives that an organization and a technical infrastructure exist to support the intention of people to use technology [45].
5. Hedonic motivation: the motivation to do something due to internal satisfaction [48].

6. Habit: the degree to which users perform the usage of technologies behaviors automatically because of learning [46,49].
7. Self-efficacy: judgment of one's ability to use technology to accomplish a particular job or task [45].
8. Behavioral intention: the degree to which a person has formulated conscious plans to perform or not perform some specified future behaviors [50].

Data Collection Tools and Procedures

Data collection tools were adapted from the source instrument used in the UTAUT2 model [46] in the context of the eCHIS to enhance comprehension by the respondents. The items in the constructs were performance expectancy (4 items), effort expectancy (4 items), social influence (3 items), facilitating condition (4 items), hedonic motivation (3 items), self-expectancy (4 items), habit (4 items), and intention to use (3 items). The source language of the instrument was translated forward into the local language of Amharic, and a backward translation was done to ensure the consistency of the tool. Experts with health management information system backgrounds were invited to review the relevance of each question in the instrument. The experts reviewed the instrument and checked its content and face validity, and the instrument was refined according to the comments given. A pretest was conducted on 5% of the study participants before actual data collection was started, and the tool was refined based on the pretest results. A total of 4 data collectors and 2 supervisors were recruited and trained on the purpose, tools, and procedures of the study. Self-administrated questionnaires were used to collect data from HEWs with the assistance of data collectors and supervisors. The data collection period was from January 28 to February 13, 2021, after 2 weeks of eCHIS familiarization training had been given.

Data Management and Analysis

The data were entered into Epi-data (version 4.02; EpiData Association) and exported to SPSS (version 26; IBM Corp) for descriptive statistics such as frequency, cross-tabulations, and univariate analysis of sociodemographic and model constructs.

Simple and multiple structural equation models were carried out using the AMOS 23 Structural Equation Model in order to test the relationship between observed and latent variables and identify the predicting variables of the intention to use the eCHIS. During analysis, we applied a parceling technique to increase model efficiency [51]. The subset-item-parcel approach was used in order to aggregate items into several parcels and use them as indicators of the target construct [52]. Accordingly, we created 2 parcels for each factor of target latent constructs (such as performance expectancy, effort expectancy, facilitator conditions, self-expectancy, and habit) by aggregating randomly grouped items within each scale [53]. The remaining 3 latent target constructs with 3 indicators per construct, such as social influence, hedonic motivation, and intention to use, remained as they existed in the original UTAUT2 model [46].

The overall model's fitness was measured and assessed using the goodness of fit indices such as chi-square ratio (<3), the goodness of fit index (>0.9), adjusted goodness of fit index (>0.8), normal fit index (>0.9), comparative fit index (>0.9), Tucker-Lewis index (>0.9), and root mean square error of approximation (<0.08). For the structural equation model, standardized path coefficients of the regression weight values were used to estimate the path coefficients of the dependent and independent variables. Standardized coefficients are not dependent on the scales as they vary from –1 to 1, where 0 indicates no relationship, 1 indicates a strong positive relationship, and –1 indicates a strong negative relationship. A critical ratio (regression weight or standard error) was used to evaluate whether the constructs had a significant relationship. The absolute value of a critical ratio greater than 1.96 is an indication of the significance of the path coefficients. In this study, the CI and its *P* value were calculated using bootstrapping, and the statistical significance of dependent and independent variables in the model was reported using a 95% CI with a corresponding *P* value of <.05 (Table 1). The square multiple correlation (R^2) was used to report the proportion of variance so that the intention to use the eCHIS could be explained by the model.

Table 1. Structural equation modeling fitness for intention to use electronic community health information system among health extension workers in Northwest Ethiopia, 2021.

Fit indices	Threshold value	Authors	Results obtained	Conclusion
Chi-square	<3	Bentler [54] (1990)	2.67	Accepted
Goodness-of-fit index	>0.9	Chau [55] (1997)	0.92	Accepted
Adjusted goodness-of-fit index	>0.8	Chau [55] (1997)	0.88	Accepted
Comparative fit index	>0.9	Bentler [54] (1990)	0.97	Accepted
Root mean square error of approximation	<0.05	Browne and Cudeck [56] (1993)	0.08	Accepted
Normed fit index	>0.9	Bentler and Bonett [57] (1980)	0.95	Accepted

Reliability and Validity of the Research

Regarding the reliability and validity of the study, Cronbach α reliability coefficients were computed to determine the internal consistency of the constructs. Cronbach α of .7 or above indicates high reliability; between .5 and .7 indicates moderate reliability; and less than .5 indicates low reliability. We have used 4-item Likert questions to assess the reliability of the constructs. Accordingly, the reliability of the constructs assessed by 3-item questions as follows: performance expectancy (α=.92), effort expectancy (α=.87), facilitating condition (α=.75), self-expectancy (α=.88), habit (α=.84), social influence (α=.78), hedonic motivation (α=.90), and intention to use eCHIS (Table 2). In this study, the magnitude of intention to use the eCHIS was assessed by a 3-item Likert question with a reliability test of Cronbach α=.93.

Table 2. Reliability of the constructs on intention to use the electronic community health information system (eCHIS) among health extension workers in Northwest Ethiopia, 2021.

Constructs	Sample size	Number of items	Cronbach α
Performance expectancy	456	4	.92
Effort expectancy	456	4	.87
Social influence	456	3	.78
Facilitating conditions	456	4	.75
Hedonic motivation	456	3	.90
Self-expectancy	456	4	.88
Habit	456	4	.84
Intention to use the eCHIS system	456	3	.93

Ethical Considerations

Study approval and ethical clearance were obtained from the University of Gondar's ethical review board (R.NO. V/P/RCS/05/2020) and a support letter from the ethical review committee of the Amhara Regional Health Bureau Research and Technology transfer office. Study permission was sought at all levels of governmental administration systems including health offices and health facilities. Written consent was obtained, and participants were informed about the objective, importance of the study, procedure and duration, risk and discomfort, benefits of participating in the study, confidentiality, and the right to refuse or withdraw during data collection. To ensure confidentiality, their names and other personal identifiers were not registered. Participants were not compensated for study participation. We confirm that the provided ethics approval documentation covers the study presented in this manuscript.

Results

Sociodemographic and Other Characteristics of the Study Participants

A total of 456 HEWs participated in the study, with a response rate of 99%. The mean age of the study participants was 28 (SD 4.8) years. More than two-thirds (n=314, 68.9%) of the study participants had work experience of more than 5 years. About half of the participants (n=232, 50.9%) were level 4 (10+4) in their educational status, and the majority of the respondents (n=307, 67.3%) were married. The number of HEWs who had difficulties recharging mobile phones was 307 (67.3%). Our study found that 147 (32.2%) HEWs used Microsoft applications daily , 331 (72.6%) had experience using mobile phones for more than 5 years and above, and 421 (92.3%) had informal mobile phone usage practices or were using personal mobile for health post–related activities (Table 3).

According to the findings of this study, 122 (26.8%), 132 (28.9%), and 162 (35.5%) HEWs strongly agreed to intend, predict, and plan to use the eCHIS, respectively (Table 4).

Table 3. Sociodemographic and informal phone use characteristics of the study participants in Northwest Ethiopia, 2021.

Variables and categories	Values (N=456), n (%)
Age group (years)	
<24	104 (22.8)
25-34	295 (64.7)
≥35	57 (12.5)
Marital status	
Married	307 (67.3)
Single	118 (25.9)
Divorced	26 (5.7)
Widow	5 (1.1)
Work experience (years)	
0-2	55 (12.1)
3-5	87 (19.1)
>5	314 (68.9)
Level of education	
Level I	5 (1.1)
Level II	12 (2.6)
Level III	196 (43)
Level IV	232 (50.9)
Others[a]	11 (2.4)
Difficulty with battery recharging	
Yes	307 (67.3)
No	149 (32.7)
Using Microsoft applications for work and daily life	
Yes	147 (32.2)
No	309 (67.8)
Do you use personal mobile phone for health post–related activities?	
Yes	421 (92.3)
No	35 (7.7)
For how long you have used mobile phone (years)?	
0-5	125 (27.4)
>5	331 (72.6)

[a]Health extension worker with additional diploma, BSc degree, or both.

Table 4. Health extension workers' intention to use the electronic community health information system (eCHIS) in Northwest Ethiopia, 2021 (N=456).

Items	Strongly disagree, n (%)	Disagree, n (%)	Neutral, n (%)	Agree, n (%)	Strongly agree, n (%)
I intend to use the eCHIS system in the future	7 (1.5)	12 (2.6)	12 (2.6)	303 (66.4)	122 (26.8)
I predict I will use the eCHIS system in the future	5 (1.1)	13 (2.9)	9 (2)	297 (65.1)	132 (28.9)
I plan to use the eCHIS system in the future	4 (0.9)	11 (2.4)	15 (3.3)	264 (57.9)	162 (35.5)

Mean Score of All Predictors and Intention to Use the eCHIS Using the UTAUT2 Model

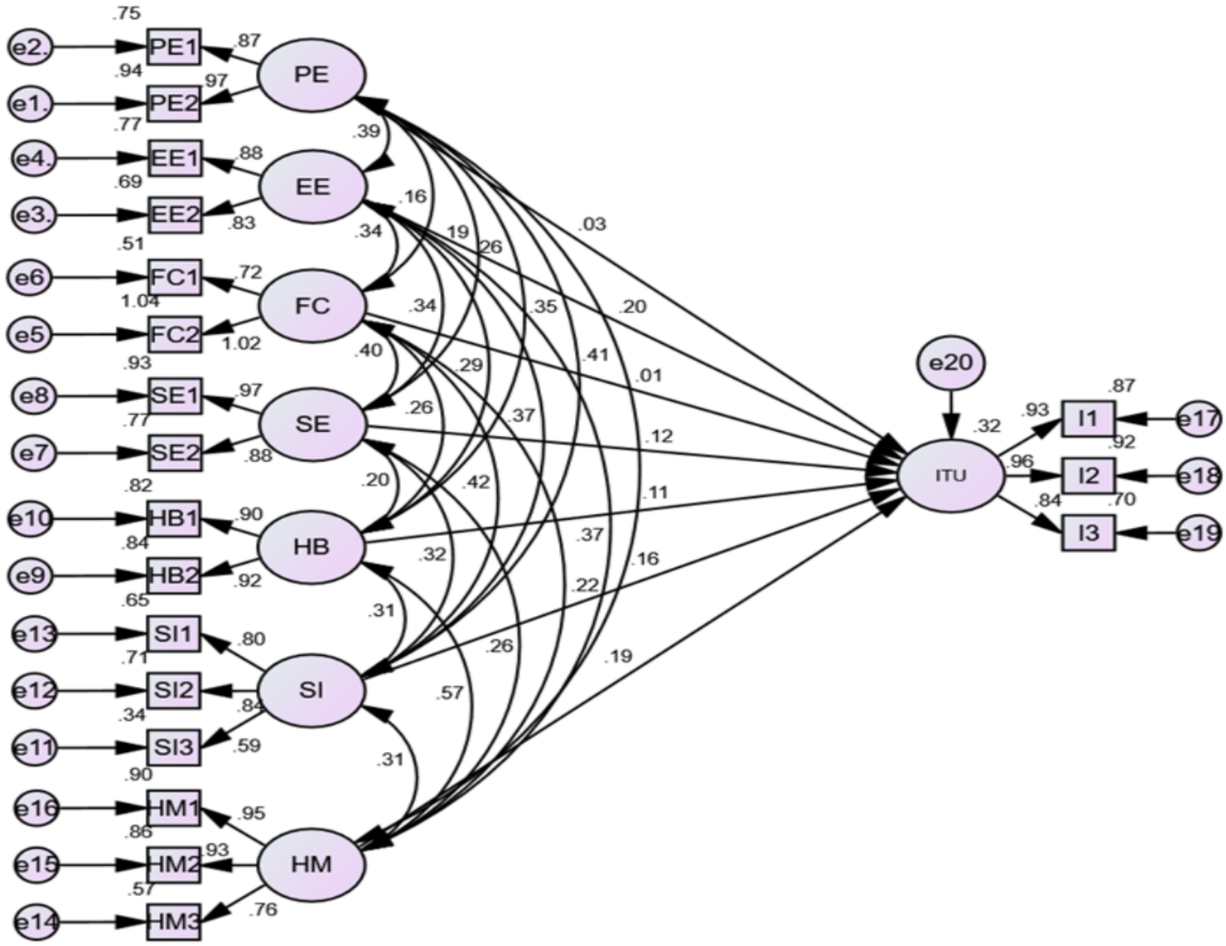
The mean scores of performance expectancy, effort expectancy, facilitating condition, self-expectancy, and habit with 4-item Likert questions were 17.09 (SD 2.58), 16.22 (SD 2.41), 12.07 (SD 3.46), 13.94 (SD 3.62), and 14.75 (SD 3.14), respectively. On the other hand, social influence, hedonic motivation, and intention to use the eCHIS with 3-item Likert questions had a mean score of 11.63 (SD 2.32), 12.23 (SD 2.01), and 12.57 (SD 2.00), respectively.

Our study revealed that 179 (39.3%; 95% CI 34.7%-43.9%) participants who had the intention to use the eCHIS for community health data generation, use, and service provision had scored above the mean. The mean score of the intention to use the eCHIS was 12.57 (SD 2.00). The maximum score of intention to use the eCHIS was 15, while the minimum score was 3.

Simple Structural Equation Model Analysis

The variance in the dependent variable explained by the independent variables was interpreted using square multiple correlation (R^2). The overall R^2 of the intention to use the eCHIS is found to be 32%, the variance that was explained by the independent variables in the model. The bootstrap method with a 95% bias-corrected CI was applied to investigate the significance of path coefficients and factors predicting the model. The predictors with $P<.20$ in the simple structural equation model were considered candidate variables for multiple structural equation model analysis. Due to its undependability to scale, we used a standardized beta coefficient to interpret the influence of predictors on the intention to use the eCHIS. A 95% CI with $P<.05$ was considered to declare an association between dependent and independent variables. The study indicated that effort expectancy has the highest direct effect on HEWs' intention to use the eCHIS, followed by hedonic motivation. The remaining model constructs that have a direct influence on predicting intention to use the eCHIS are social influence and self-expectancy. The structural equation model predicted, with the path coefficients and R^2, is represented in Figure 1, and the path coefficients and *P* value found from the depicted model are presented in the *Results* section. Moreover, the absolute value of the critical ratio of effort expectancy (3.701), self-expectancy (2.468), social influence (2.782), and hedonic motivation (3.311) indicated that predictors had a significant influence on HEWs' intention to use the eCHIS. Overall, 32% of the variance with respect to intention to use the eCHIS was reasonably explained by the predictors in the model.

Figure 1. Predictors and intention to use the electronic community health information system among health extension workers at Central Gonda zone, Northwest Ethiopia, 2021. EE: effort expectancy; FC: facilitating condition; HB: Habit; HM: hedonic motivation; ITU: intention to use; PE: performance expectance; SE: self-expectancy; SI: social influence.



Effort expectancy (the extent to which people believe that using the eCHIS can improve their effort) has a positive influence on HEWs' behavioral intention (β=.256; *P*=.007). Similarly, self-efficacy and social influence had a positive influence on HEWs' behavioral intention (β=.096; *P*=.04), and (β=.203; *P*=.02), respectively. Likewise, hedonic motivation to use eCHIS due to internal satisfaction was found to be (β=0.217; *P*=.03) and had a significant effect on intention to use the eCHIS. Facilitating conditions (β=0.005; *P*=.92), habit (β=0.103; *P*=.07), and performance expectancy (β=0.034; *P*=.61) had no significant influence on intention to use the eCHIS (Table 5).

Table 5. Multiple structural equations modeling association between predictors and intention to use the electronic community health information system among health extension workers in Northwest Ethiopia, 2021.

Hypothesis	Estimate	95% CI	*P* value	Decision
PE[a] ⇒ IU[b]	0.034	–0.91 to 0.190	.61	Not supported
EE[c] ⇒ IU	0.256	0.060 to 0.503	.007	Supported
FC[d] ⇒ IU	0.005	–0.060 to 0.079	.92	Not supported
SE[e] ⇒ IU	0.096	0.004 to 0.191	.04	Supported
HB[f] ⇒ IU	0.103	–0.008 to 0.235	.07	Not supported
SI[g] ⇒ IU	0.203	0.039 to 0.390	.02	Supported
HM[h] ⇒ IU	0.217	0.022 to 0.416	.03	Supported

[a]PE: performance expectance.
[b]IU: intention to use.
[c]EE: effort expectancy.
[d]FC: facilitating condition.
[e]SE: self-expectancy.
[f]HB: habit.
[g]SI: social influence.
[h]HM: hedonic motivation.

Discussion

Principal Findings

In this study, nearly 2 out of 5 HEWs had an intention to use the eCHIS for community health data generation, use, and service provision. Effort expectancy, self-expectancy, social influence, and hedonic motivation were statistically significant predictors of intention to use the eCHIS. The intention to use the eCHIS by HEWs could be associated with the fact that using the eCHIS is not difficult to understand. It saves time and reduces the amount of effort required to complete health-related tasks [58,59]. Furthermore, it simplifies activities and helps them access data easily. The other could be people around them who have the ability to influence their intention to use the system [58]. For example, HEWs' activities should be monitored and evaluated by health system leaders. If they give them more attention, they will be encouraged to use the system. The other could be previous exposure to using informal phone for health system activities, such as reminding clients about their health care appointments and facilitating referral linkage between health centers and health posts, as mHealth enhances communication between health workers and clients [58]. Moreover, using the eCHIS creates a conducive environment for HEWs since their usual data handling approach is exhaustive and takes much time to execute activities at the health post level, and using the eCHIS not only helps them to save their time but also creates motivation to do their job at health post level [58,60].

Regarding factors associated with intention to use the eCHIS, effort expectancy had a positive influence on the intention to use the eCHIS among HEWs. This finding was in accordance with a study conducted in Ethiopia [61], Kenya [61], the United States [62], and Portugal [63] and had a positively significant association with the intention of health care providers to use technology. A possible explanation could be the fact that the less effort the user devotes to using the system, the more likely he or she is to continue to use it. A study in this regard showed that individuals often want to face a system that is easy to use [64]. HEWs might perceive that the eCHIS could help them to do their job aids shortly with less strain and increased work efficiency [65], as using the eCHIS would simplify the tasks they are expected to deliver at health post level. A review in this regard showed that using digital tools simplifies work and helps to access data easily [59]. Furthermore, studies indicate that digital health solutions reduce workload and improve work performance [24,25], reduce errors [34], create motivation and learning opportunities [66], promote health care appointment [67], and are easy to use and improve work efficiency [59]. Using the eCHIS could reduce the workload of HEWs since manual data management practice at health post level is exhaustive and takes much time to collect data and conduct routine activities [59]. Moreover, the referral linkage integrated into the eCHIS, including HEWs, midwives, and focal persons, will harmonize HEWs' activity flow from health posts to health centers and vice versa. Furthermore, using mHealth motivates

CHWs and enables them to perform multiple tasks quickly, reducing efforts and improving performance [60].

The intention to use the eCHIS among HEWs who perceived people around them could influence their behavioral intention was positively associated. The current finding corroborates studies conducted among health care providers using the UTAUT2 model in Ethiopia [61], Morocco [68], Taiwan [62], South Korea [30], and the United States [69], showing that social influence significantly predicted health care providers' intention to use technology. The possible explanation could be that HEWs might perceive peer pressure from health care staff at the *woreda* and facility levels toward using the eCHIS, which could positively influence their intention to use the eCHIS. The other justification could be the fact that HEWs might get trust from the community for the job aids or activities they are expected to deliver. Hence, health system staff need to understand that peer influence has a positive effect on using a new system. Moreover, making people aware of a new system at the *woreda* and facility levels in general and at the kebele leaders, women's development army, and voluntary service providers' levels, in particular, could influence HEWs' behavioral intention to use the eCHIS. A study in this regard showed that the more health workers connected to colleagues, the more they improved the use of digital tools and the quality of care [58].

Our study revealed that the magnitude of intention to use the eCHIS among HEWs who had self-expectancy was positively correlated. The findings of past studies in Ethiopia [61], Malaysia [70], Taiwan [71], and Iran [72] showed that digital literacy was correlated with the intention to use technology in health care industries. The possible reason might be that those who had self-expectancy could not face difficulty in adapting the emerging technology to community-level data management and service provision. The current evidence in the feasibility and effectiveness study on digital health indicated that the level of computer literacy had influenced digital health implementation among CHWs [73]. A possible explanation might be the fact that informal mobile phone usage practices of CHWs for health post–related activities could influence behavioral intention to use the eCHIS. A study indicated that in many different settings, CHWs use their personal phone informally for community-based activities so as to fill the gaps in the health care system [74].

Our study revealed that there is a significant association between intention to use the eCHIS and hedonic motivation or perceived enjoyment from using the eCHIS for community health data generation, use, and service provision. A possible explanation could be the fact that using a new system instead of the usual approach to manage community-level data and service provision may create intrinsic motivation for HEWs to obtain fun or pleasure. A study showed that motivation is an important construct for eHealth users, and it could even be a sufficient reason to adopt newly emerging technology in a contextual environment [75]. In addition, using eHealth technology to deal with community health data generation, use, and service provision may be an enjoyable process and will have a positive influence on the behavioral intentions of the users [72].

HEWs were optimistic about using the eCHIS because it could be related to the production of quality health data, ease of data management, reduced errors and false reports, data protection, and increased accessibility. A study also indicated that using digital tools could enhance the productivity of CHWs [76]. Community health digitization using mobile apps support the services delivered by CHWs [77]. Furthermore, studies show that the digitization of health care data has promising results in improving both health care and health outcomes [13-19] and improving health staff productivity and work efficiency [65]. In our study, the level of users' optimistic perceptions of using the eCHIS could be an advantage in implementing the intervention, as compared to the existing approach [78]. Studies showed that digital health solutions enable CHWs to generate quality health data [34], improve health care delivery [35], and help them to be more effective in their job aids [32,36]. Moreover, digital tools could help them follow the correct order or the required service elements that clients should receive when providing services. It also enables them to communicate with clients in a better way as compared to manual communication since the tool has prespecified data elements that should be asked by CHWs during service delivery [58].

Likewise, it creates enjoyment among users [66] and benefits them by keeping data safe from human and natural factors that could damage the data. In addition, enjoyment could emanate due to the fact that using digital tools can improve data capturing, storing, and reporting of more items that could be more time-consuming during manual data handling and reduce the motivation of health workers to keep data recording [59]. Even though HEWs are optimistic to use the eCHIS, lack of adequate resources for eCHIS implementation at the implementation district could hinder its successful implementation, and therefore resource availability is vital to be effective in community health digitization. Studies show that challenges during digital health solutions implementation, such as the initial and ongoing capacity-building training [73], poor network access and poor access to electricity [58,79], low financial investment [73], and unreliability or absence of infrastructure (eg, electricity and network) [80,81], hinder the implementation. As the skills and knowledge of HEWs vary from one to another, there should be mentoring and supportive supervision during the implementation. Studies showed that the inability to use the system could affect its implementation [26], and intensive training with continuous refreshment could help them realize the digitization of the community health information system [82].

Limitations of the Study

The findings of this study should be interpreted in light of some limitations. Due to the nature of the study, which was cross-sectional, the inability to infer cause-and-effect relationships is present. As the study was focused on HEWs' intention to use the eCHIS in a pilot district in Northwest Ethiopia, the sample size could affect the findings of this study, and covering larger areas at the regional and national levels is possible, including urban HEWs. Finally, the parcel approach used in this study may introduce parameter estimation bias.

Conclusion

In conclusion, 39.3% (179/456) of HEWs scored above the mean of intention to use the eCHIS for community health data management and service provision. Factors associated with the intention to use the eCHIS were effort expectancy, self-expectancy, social influence, and hedonic motivation. The eCHIS has numerous advantages and a promising future in terms of improving data quality, use, and service delivery. Its adoption in the country, however, should focus on identifying all necessary prerequisites for successful implementation and advancing the community health information system. The implementation of the eCHIS should not skip factors that had no significant effect on intention to use the eCHIS, and further studies at the regional and national levels are recommended to investigate their correlation with intention to use the eCHIS. Model explainability was found in the study using factors that existed in the UTAUT2 model; however, it is recommended to examine the moderating effects of CHWs' related variables to examine how the model constructs could influence HEWs' intention to use the eCHIS.

Acknowledgments

We would like to thank the Central Gondar Zone Health Department, electronic community health information system pilot district health offices, and the University of Gondar Comprehensive Specialized Hospital for their provision of necessary information and support during data collection. Our gratitude also goes to the study participants, data collectors, and supervisors who took part in the study. This work would not have been possible without the financial support of the Doris Duke Charitable Foundation (grant 2017187). The mission of the Doris Duke Charitable Foundation is to improve the quality of people's lives through grants supporting the performing arts, environmental conservation, medical research, and child well-being, and through the preservation of the cultural and environmental legacy of Doris Duke's properties.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

TH conducted the study. All authors conceptualized the design of the study, provided a review of the methodology and results, contributed to the reviewing the manuscript, and read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Tamrat T, Kachnowski S. Special delivery: an analysis of mHealth in maternal and newborn health programs and their outcomes around the world. Matern Child Health J 2012;16(5):1092-1101 [FREE Full text] [doi: 10.1007/s10995-011-0836-3] [Medline: 21688111]
2. WHO guidelines on health policy and system support to optimize community health worker programmes. World Health Organization. 2018. URL: https://www.who.int/publications/i/item/9789241550369 [accessed 2024-02-15]
3. Haregu TN, Alemayehu YK, Alemu YA, Medhin G, Woldegiorgis MA, Fentaye FW, et al. Disparities in the implementation of the health extension program in Ethiopia: doing more and better towards universal health coverage. Dialogues Health 2022;1:100047 [FREE Full text] [doi: 10.1016/j.dialog.2022.100047] [Medline: 22489612]
4. Teklehaimanot HD, Teklehaimanot A. Human resource development for a community-based health extension program: a case study from Ethiopia. Hum Resour Health 2013;11(1):39 [FREE Full text] [doi: 10.1186/1478-4491-11-39] [Medline: 23961920]
5. Medhanyie A, Spigt M, Dinant G, Blanco R. Knowledge and performance of the Ethiopian health extension workers on antenatal and delivery care: a cross-sectional study. Hum Resour Health 2012;10(1):44 [FREE Full text] [doi: 10.1186/1478-4491-10-44] [Medline: 23171076]
6. Koblinsky M, Tain F, Gaym A, Karim A, Carnell M, Tesfaye S. Responding to the maternal health care challenge: the Ethiopian health extension program. Ethiop J Health Dev 2010;24(1):105-109 [FREE Full text] [doi: 10.4314/ejhd.v24i1.62951]
7. Teklehaimanot A, Kitaw Y, G/Yohannes A, Girma S, Seyoum S, Ye-Ebiyo Y. Study of the working conditions of health extension workers in Ethiopia. Ethiop J Health Dev 2007;21(3):246-259 [FREE Full text] [doi: 10.4314/ejhd.v21i3.10055]
8. Sebastian MS, Lemma H. Efficiency of the health extension programme in Tigray, Ethiopia: a data envelopment analysis. BMC Int Health Hum Rights 2010;10(1):16 [FREE Full text] [doi: 10.1186/1472-698X-10-16] [Medline: 20546597]
9. Theo L, Rainer S, Bodart C. A framework for designing health information systems. World Health Organization. 2000. URL: https://www.afro.who.int/sites/default/files/2017-06/AHO_Country_H_Infos_Systems_2nd_edition.pdf [accessed 2024-02-15]

10. Tilahun B, Derseh L, Atinafu A, Mamuye A, Mariam TH, Mohammed M, et al. Level and contributing factors of health data quality and information use in two districts in Northwest Ethiopia: social-ecological perspective. BMC Med Inform Decis Mak 2021;21(1):373 [FREE Full text] [doi: 10.1186/s12911-021-01741-1] [Medline: 34972511]
11. Medhanyie A, Spigt M, Kifle Y, Schaay N, Sanders D, Blanco R, et al. The role of health extension workers in improving utilization of maternal health services in rural areas in Ethiopia: a cross sectional study. BMC Health Serv Res 2012;12(1):352 [FREE Full text] [doi: 10.1186/1472-6963-12-352] [Medline: 23043288]
12. Dynes M, Buffington ST, Carpenter M, Handley A, Kelley M, Tadesse L, et al. Strengthening maternal and newborn health in rural Ethiopia: early results from frontline health worker community maternal and newborn health training. Midwifery 2013;29(3):251-259 [FREE Full text] [doi: 10.1016/j.midw.2012.01.006] [Medline: 22342173]
13. Krishna S, Boren SA, Balas EA. Healthcare via cell phones: a systematic review. Telemed J E Health 2009;15(3):231-240 [FREE Full text] [doi: 10.1089/tmj.2008.0099] [Medline: 19382860]
14. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
15. Lim MSC, Hocking JS, Hellard ME, Aitken CK. SMS STI: a review of the uses of mobile phone text messaging in sexual health. Int J STD AIDS 2008;19(5):287-290 [FREE Full text] [doi: 10.1258/ijsa.2007.007264] [Medline: 18482956]
16. Fjeldsoe BS, Marshall AL, Miller YD. Behavior change interventions delivered by mobile telephone short-message service. Am J Prev Med 2009;36(2):165-173 [FREE Full text] [doi: 10.1016/j.amepre.2008.09.040] [Medline: 19135907]
17. Pop-Eleches C, Thirumurthy H, Habyarimana JP, Zivin JG, Goldstein MP, de Walque D, et al. Mobile phone technologies improve adherence to antiretroviral treatment in a resource-limited setting: a randomized controlled trial of text message reminders. AIDS 2011;25(6):825-834 [FREE Full text] [doi: 10.1097/QAD.0b013e32834380c1] [Medline: 21252632]
18. Horvath T, Azman H, Kennedy GE, Rutherford GW. Mobile phone text messaging for promoting adherence to antiretroviral therapy in patients with HIV infection. Cochrane Database Syst Rev 2012;2012(3):CD009756 [FREE Full text] [doi: 10.1002/14651858.CD009756] [Medline: 22419345]
19. Lester RT, Ritvo P, Mills EJ, Kariri A, Karanja S, Chung MH, et al. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. Lancet 2010;376(9755):1838-1845 [FREE Full text] [doi: 10.1016/S0140-6736(10)61997-6] [Medline: 21071074]
20. Vital Wave Consulting. mHealth for development: the opportunity of mobile technology for healthcare in the developing world. United Nations Foundation. Washington DcBerkshire, UK; 2009. URL: https://reliefweb.int/report/world/mhealth-development-opportunity-mobile-technology-healthcare-developing-world [accessed 2024-02-15]
21. MHealth: New Horizons for Health Through Mobile Technologies. Geneva: World Health Organization; 2011.
22. Marshall C, Lewis D, Whittaker M. mHealth Technologies in Developing Countries: a Feasibility Assessment and a Proposed Framework. Herston, Australia: University of Queensland; 2013:1-47.
23. Shiferaw KB, Mehari EA. Modeling predictors of acceptance and use of electronic medical record system in a resource limited setting: using modified UTAUT model. Inform Med Unlocked 2019;17:100182 [FREE Full text] [doi: 10.1016/j.imu.2019.100182]
24. Andreatta P, Debpuur D, Danquah A, Perosky J. Using cell phones to collect postpartum hemorrhage outcome data in rural Ghana. Int J Gynaecol Obstet 2011;113(2):148-151 [FREE Full text] [doi: 10.1016/j.ijgo.2010.11.020] [Medline: 21420086]
25. Chang LW, Kagaayi J, Arem H, Nakigozi G, Ssempijja V, Serwadda D, et al. Impact of a mHealth intervention for peer health workers on AIDS care in rural Uganda: a mixed methods evaluation of a cluster-randomized trial. AIDS Behav 2011;15(8):1776-1784 [FREE Full text] [doi: 10.1007/s10461-011-9995-x] [Medline: 21739286]
26. Wills MJ, El-Gayar OF, Bennett D. Examining healthcare professionals' acceptance of electronic medical records using UTAUT. Issues Inf Syst 2008;9(2):396-401 [FREE Full text]
27. Arman AA, Hartati S. Development of user acceptance model for electronic medical record system. : IEEE; 2015 Presented at: 2015 International Conference on Information Technology Systems and Innovation (ICITSI); November 16-19, 2015; Bandung, Indonesia. [doi: 10.1109/icitsi.2015.7437724]
28. Cocosila M, Archer N. A blended model of electronic medical record system adoption in Canadian medical practices. Commun Assoc Inf Syst 2016;39(1):483-508 [FREE Full text] [doi: 10.17705/1cais.03922]
29. Hung SY, Yu WJ, Tsai MF, Yen DC. What drives physicians’ intention to use electronic medical record system: the roles of perceived service level, computer self-efficacy, and perceived risk. 2013 Presented at: Pacific Asia Conference on Information Systems 2013 Proceedings; June 18, 2013; Virtual p. 194.
30. Venugopal P, Jinka S, Priya SA. User acceptance of electronic health records: cross validation of UTAUT model. Glob Manag Rev 2016;10(3):42-54 [FREE Full text]
31. Tubaishat A. Perceived usefulness and perceived ease of use of electronic health records among nurses: application of technology acceptance model. Inform Health Soc Care 2018;43(4):379-389 [FREE Full text] [doi: 10.1080/17538157.2017.1363761] [Medline: 28920708]
32. Medhanyie AA, Little A, Yebyo H, Spigt M, Tadesse K, Blanco R, et al. Health workers' experiences, barriers, preferences and motivating factors in using mHealth forms in Ethiopia. Hum Resour Health 2015;13(1):2 [FREE Full text] [doi: 10.1186/1478-4491-13-2] [Medline: 25588973]

33. Little A, Medhanyie A, Yebyo H, Spigt M, Dinant GJ, Blanco R. Meeting community health worker needs for maternal health care service delivery using appropriate mobile technologies in Ethiopia. PLoS One 2013;8(10):e77563 [FREE Full text] [doi: 10.1371/journal.pone.0077563] [Medline: 24204872]
34. Florez-Arango JF, Iyengar MS, Dunn K, Zhang J. Performance factors of mobile rich media job aids for community health workers. J Am Med Inform Assoc 2011;18(2):131-137 [FREE Full text] [doi: 10.1136/jamia.2010.010025] [Medline: 21292702]
35. Mushamiri I, Luo C, Iiams-Hauser C, Amor YB. Evaluation of the impact of a mobile health system on adherence to antenatal and postnatal care and prevention of mother-to-child transmission of HIV programs in Kenya. BMC Public Health 2015;15(1):102 [FREE Full text] [doi: 10.1186/s12889-015-1358-5] [Medline: 25886279]
36. The mobile economy: Sub-Saharan Africa 2014. GSMA Intelligence. 2014. URL: https://www.gsma.com/publicpolicy/wp-content/uploads/2015/02/The-Mobile-Economy-Sub-Saharan-Africa-2014-1-pager-FINAL.pdf [accessed 2024-02-15]
37. Mendoza G, Okoko L, Morgan G, Konopka S. mHealth compendium. Volume Two. United States Agency for International Development. 2013. URL: https://msh.org/wp-content/uploads/2013/09/usaid_mhealth_compendium_vol._2_us_letter_final_desktopprint_with_infogr.pdf [accessed 2024-02-15]
38. Klasnja P, Pratt W. Healthcare in the pocket: mapping the space of mobile-phone health interventions. J Biomed Inform 2012;45(1):184-198 [FREE Full text] [doi: 10.1016/j.jbi.2011.08.017] [Medline: 21925288]
39. Ahlan AR, Ahmad BI. User acceptance of health information technology (HIT) in developing countries: a conceptual model. Procedia Technol 2014;16:1287-1296 [FREE Full text] [doi: 10.1016/j.protcy.2014.10.145]
40. Al-Aswad AM, Brownsell S, Palmer R, Nichol JP. A review paper of the current status of electronic health records adoption worldwide: the gap between developed and developing countries. J Health Inform Dev Ctries 2013;7(2):153-164 [FREE Full text]
41. Boonstra A, Broekhuis M. Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions. BMC Health Serv Res 2010;10(1):231 [FREE Full text] [doi: 10.1186/1472-6963-10-231] [Medline: 20691097]
42. Shiells K, Holmerova I, Steffl M, Stepankova O. Electronic patient records as a tool to facilitate care provision in nursing homes: an integrative review. Inform Health Soc Care 2019;44(3):262-277 [FREE Full text] [doi: 10.1080/17538157.2018.1496091] [Medline: 30125140]
43. Gao S, Krogstie J, Siau K. Developing an instrument to measure the adoption of mobile services. Mob Inf Syst 2011;7(1):45-67 [FREE Full text] [doi: 10.1155/2011/831018]
44. Venkatesh V, Davis FD, Morris MG. Dead or alive? the development, trajectory and future of technology adoption research. J Assoc Inf Syst 2007;8(4):267-286 [FREE Full text] [doi: 10.17705/1jais.00120]
45. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425-478 [FREE Full text] [doi: 10.2307/30036540]
46. Venkatesh V, Thong JYL, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Quarterly 2012;36(1):157-178 [FREE Full text] [doi: 10.2307/41410412]
47. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989;13(3):319-340 [FREE Full text] [doi: 10.2307/249008]
48. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
49. Limayem M, Hirt SG, Cheung CMK. How habit limits the predictive power of intention: the case of information systems continuance. MIS Quarterly 2007;31(4):705-737 [FREE Full text] [doi: 10.2307/25148817]
50. Aarts H, Verplanken B, van Knippenberg A. Predicting behavior from actions in the past: repeated decision making or a matter of habit? J Applied Social Pyschol 2006;28(15):1355-1374. [doi: 10.1111/j.1559-1816.1998.tb01681.x]
51. Matsunaga M. Item parceling in structural equation modeling: a primer. Commun Methods Meas 2008;2(4):260-293 [FREE Full text] [doi: 10.1080/19312450802458935]
52. Soliz J, Harwood J. Shared family identity, age salience, and intergroup contact: investigation of the grandparent–grandchild relationship. Commun Monogr 2006;73(1):87-107 [FREE Full text] [doi: 10.1080/03637750500534388]
53. Rogers WM, Schmitt N. Parameter recovery and model fit using multidimensional composites: a comparison of four empirical parceling algorithms. Multivariate Behav Res 2004;39(3):379-412 [FREE Full text] [doi: 10.1207/s15327906mbr3903_1]
54. Bentler PM. Comparative fit indexes in structural models. Psychol Bull 1990;107(2):238-246. [doi: 10.1037/0033-2909.107.2.238] [Medline: 2320703]
55. Chau PYK. Reexamining a model for evaluating information center success using a structural equation modeling approach. Decis Sci 1997;28:309-334 [FREE Full text]
56. Browne MW, Cudeck R. Alternative ways of assessing model fit. In: Bollen KA, Long JS, editors. Testing Structural Equation Models. Newbury Park, CA: Sage Publications; 1993:136-162.
57. Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariance structures. Psychol Bull 1980;88(3):588-606. [doi: 10.1037/0033-2909.88.3.588]

58. Odendaal WA, Watkins JA, Leon N, Goudge J, Griffiths F, Tomlinson M, et al. Health workers' perceptions and experiences of using mHealth technologies to deliver primary healthcare services: a qualitative evidence synthesis. Cochrane Database Syst Rev 2020;3(3):CD011942 [FREE Full text] [doi: 10.1002/14651858.CD011942.pub2] [Medline: 32216074]
59. Cansdale LG, Kelly G, Khashan A, Malata A, Kachale F, Lissauer D, et al. Use of mHealth tools to register birth outcomes in low-income and middle-income countries: a scoping review. BMJ Open 2022;12(10):e063886 [FREE Full text] [doi: 10.1136/bmjopen-2022-063886] [Medline: 36223965]
60. Ballard M, Montgomery P. Systematic review of interventions for improving the performance of community health workers in low-income and middle-income countries. BMJ Open 2017;7(10):e014216 [FREE Full text] [doi: 10.1136/bmjopen-2016-014216] [Medline: 29074507]
61. Ahmed MH, Bogale AD, Tilahun B, Kalayou MH, Klein J, Mengiste SA, et al. Intention to use electronic medical record and its predictors among health care providers at referral hospitals, North-West Ethiopia, 2019: using unified theory of acceptance and use technology 2(UTAUT2) model. BMC Med Inform Decis Mak 2020;20(1):207 [FREE Full text] [doi: 10.1186/s12911-020-01222-x] [Medline: 32883267]
62. Kim S, Lee KH, Hwang H, Yoo S. Analysis of the factors influencing healthcare professionals' adoption of mobile Electronic Medical Record (EMR) using the Unified Theory of Acceptance and Use of Technology (UTAUT) in a tertiary hospital. BMC Med Inform Decis Mak 2016;16(1):12 [FREE Full text] [doi: 10.1186/s12911-016-0249-8] [Medline: 26831123]
63. Ouattara A. Antecedents of Employees' Behavioral Intentions Regarding Information Technology Consumerization. Minneapolis: Walden University; 2017.
64. Garavand A, Mohseni M, Asadi H, Etemadi M, Moradi-Joo M, Moosavi A. Factors influencing the adoption of health information technologies: a systematic review. Electron Physician 2016;8(8):2713-2718 [FREE Full text] [doi: 10.19082/2713] [Medline: 27757179]
65. Kuek A, Hakkennes S. Healthcare staff digital literacy levels and their attitudes towards information systems. Health Informatics J 2020;26(1):592-612 [FREE Full text] [doi: 10.1177/1460458219839613] [Medline: 30983476]
66. Ramachandran D, Canny J, Das PD, Cutrell E. Mobile-izing health workers in rural India. 2010 Presented at: CHI '10: CHI Conference on Human Factors in Computing Systems; April 10-15, 2010; Atlanta Georgia USA. [doi: 10.1145/1753326.1753610]
67. DeRenzi B, Findlater L, Payne J, Birnbaum B, Mangilima J, Parikh T, et al. Improving community health worker performance through automated SMS. 2012 Presented at: ICTD '12: Fifth International Conference on Information and Communication Technologies and Development; March 12-15, 2012; Atlanta Georgia USA. [doi: 10.1145/2160673.2160677]
68. Bennani AE, Oumlil R. IT acceptance by nurses in Morocco: application of a modified unified theory of acceptance and use of technology. IBIMA Bus Rev 2014;2014:849383 [FREE Full text] [doi: 10.5171/2014.849383]
69. Tavares J, Oliveira T. Electronic health record portal adoption: a cross country analysis. BMC Med Inform Decis Mak 2017;17(1):97 [FREE Full text] [doi: 10.1186/s12911-017-0482-9] [Medline: 28679423]
70. Mukred M, Yusof ZM, Alotaibi FM, Mokhtar UA, Fauzi F. The key factors in adopting an Electronic Records Management System (ERMS) in the educational sector: a UTAUT-based framework. IEEE Access 2019;7:35963-35980 [FREE Full text] [doi: 10.1109/access.2019.2904617]
71. Huang WM, Chen T, Hsieh CW. An empirical study on thephysicians' behavioral intentionwith electronic medical recordsystems in Taiwan. 2014 Presented at: Pacific Asia Conference on Information Systems 2014 Proceedings; June 01, 2014; Virtual p. 160.
72. Gagnon MP, Ghandour EK, Talla PK, Simonyan D, Godin G, Labrecque M, et al. Electronic health record acceptance by physicians: testing an integrated theoretical model. J Biomed Inform 2014;48:17-27 [FREE Full text] [doi: 10.1016/j.jbi.2013.10.010] [Medline: 24184678]
73. Agarwal S, Perry HB, Long LA, Labrique AB. Evidence on feasibility and effective use of mHealth strategies by frontline health workers in developing countries: systematic review. Trop Med Int Health 2015;20(8):1003-1014 [FREE Full text] [doi: 10.1111/tmi.12525] [Medline: 25881735]
74. Hampshire K, Mwase-Vuma T, Alemu K, Abane A, Munthali A, Awoke T, et al. Informal mhealth at scale in Africa: opportunities and challenges. World Dev 2021;140:105257 [FREE Full text] [doi: 10.1016/j.worlddev.2020.105257] [Medline: 33814676]
75. Cocosila M, Archer N. Adoption of mobile ICT for health promotion: an empirical investigation. Electron Markets 2010;20(3):241-250 [FREE Full text] [doi: 10.1007/s12525-010-0042-y]
76. WHO Guideline. Recommendations on digital interventions for health system strengthening. World Health Organization. 2019. URL: https://www.who.int/publications/i/item/9789241550505 [accessed 2024-02-15]
77. Braun R, Catalani C, Wimbush J, Israelski D. Community health workers and mobile technology: a systematic review of the literature. PLoS One 2013;8(6):e65772 [FREE Full text] [doi: 10.1371/journal.pone.0065772] [Medline: 23776544]
78. Gustafson DH, Sainfort F, Eichler M, Adams L, Bisognano M, Steudel H. Developing and testing a model to predict outcomes of organizational change. Health Serv Res 2003;38(2):751-776 [FREE Full text] [doi: 10.1111/1475-6773.00143] [Medline: 12785571]

79. Thondoo M, Strachan DL, Nakirunda M, Ndima S, Muiambo A, Källander K, et al. Potential roles of mhealth for community health workers: formative research with end users in Uganda and Mozambique. JMIR Mhealth Uhealth 2015;3(3):e76 [FREE Full text] [doi: 10.2196/mhealth.4208] [Medline: 26206419]
80. White A, Thomas DSK, Ezeanochie N, Bull S. Health worker mHealth utilization: a systematic review. Comput Inform Nurs 2016;34(5):206-213 [FREE Full text] [doi: 10.1097/CIN.0000000000000231] [Medline: 26955009]
81. Russpatrick S, Sæbø J, Romedenne M, Dubent L. The state of community health information systems in West and Central Africa. J Glob Health Rep 2019;3:e2019047 [FREE Full text] [doi: 10.29392//001c.12020]
82. Jeremie N, Kaseje D, Olayo R, Akinyi C. Utilization of community-based health information systems in decision making and health action in Nyalenda, Kisumu County, Kenya. Univers J Med Sci 2014;2(4):37-42 [FREE Full text] [doi: 10.13189/ujmsj.2014.020401]

Abbreviations

CHW: community health worker
eCHIS: electronic community health information system
HEP: Health Extension Program
HEW: health extension worker
mHealth: mobile health
UTAUT: Unified Theory of Acceptance and Use of Technology
UTAUT2: Unified Theory of Acceptance and Use of Technology 2

Edited by A Kushniruk; submitted 07.03.23; peer-reviewed by C Urquhart, D Carvalho, DEE A Jalil; comments to author 22.07.23; revised version received 31.08.23; accepted 05.10.23; published 04.03.24.

Please cite as:
Hailemariam T, Atnafu A, Gezie L, Kaasbøll J, Klein J, Tilahun B
Intention to Use an Electronic Community Health Information System Among Health Extension Workers in Rural Northwest Ethiopia: Cross-Sectional Study Using the Unified Theory of Acceptance and Use of Technology 2 Model
JMIR Hum Factors 2024;11:e47081
URL: https://humanfactors.jmir.org/2024/1/e47081
doi:10.2196/47081
PMID:38437008

Original Paper

Exploring the Use of Persuasive System Design Principles to Enhance Medication Incident Reporting and Learning Systems: Scoping Reviews and Persuasive Design Assessment

Kiemute Oyibo[1], PhD; Paola A Gonzalez[2], PhD; Sarah Ejaz[1], BSc; Tasneem Naheyan[1], MSc; Carla Beaton[3], RPh; Denis O'Donnell[4], PharmD; James R Barker[2†], PhD

[1]Department of Electrical Engineering and Computer Science, Lassonde Research Centre, York University, North York, ON, Canada
[2]Faculty of Management, Dalhousie University, Halifax, NS, Canada
[3]Pharmapod, Think Research Corporation, Toronto, ON, Canada
[4]CareRx, Clinical Excellence, Toronto, ON, Canada
[†]deceased

Corresponding Author:
Kiemute Oyibo, PhD
Department of Electrical Engineering and Computer Science, Lassonde Research Centre
York University
4751 Keele Street
North York, ON, M3J 2N9
Canada
Phone: 1 416 736 5053
Email: kiemute.oyibo@yorku.ca

Abstract

Background: Medication incidents (MIs) causing harm to patients have far-reaching consequences for patients, pharmacists, public health, business practice, and governance policy. Medication Incident Reporting and Learning Systems (MIRLS) have been implemented to mitigate such incidents and promote continuous quality improvement in community pharmacies in Canada. They aim to collect and analyze MIs for the implementation of incident preventive strategies to increase safety in community pharmacy practice. However, this goal remains inhibited owing to the persistent barriers that pharmacies face when using these systems.

Objective: This study aims to investigate the harms caused by medication incidents and technological barriers to reporting and identify opportunities to incorporate persuasive design strategies in MIRLS to motivate reporting.

Methods: We conducted 2 scoping reviews to provide insights on the relationship between medication errors and patient harm and the information system–based barriers militating against reporting. Seven databases were searched in each scoping review, including PubMed, Public Health Database, ProQuest, Scopus, ACM Library, Global Health, and Google Scholar. Next, we analyzed one of the most widely used MIRLS in Canada using the Persuasive System Design (PSD) taxonomy—a framework for analyzing, designing, and evaluating persuasive systems. This framework applies behavioral theories from social psychology in the design of technology-based systems to motivate behavior change. Independent assessors familiar with MIRLS reported the degree of persuasion built into the system using the 4 categories of PSD strategies: primary task, dialogue, social, and credibility support.

Results: Overall, 17 articles were included in the first scoping review, and 1 article was included in the second scoping review. In the first review, significant or serious harm was the most frequent harm (11/17, 65%), followed by death or fatal harm (7/17, 41%). In the second review, the authors found that iterative design could improve the usability of an MIRLS; however, data security and validation of reports remained an issue to be addressed. Regarding the MIRLS that we assessed, participants considered most of the primary task, dialogue, and credibility support strategies in the PSD taxonomy as important and useful; however, they were not comfortable with some of the social strategies such as cooperation. We found that the assessed system supported a number of persuasive strategies from the PSD taxonomy; however, we identified additional strategies such as tunneling, simulation, suggestion, praise, reward, reminder, authority, and verifiability that could further enhance the perceived persuasiveness and value of the system.

Conclusions: MIRLS, equipped with persuasive features, can become powerful motivational tools to promote safer medication practices in community pharmacies. They have the potential to highlight the value of MI reporting and increase the readiness of pharmacists to report incidents. The proposed persuasive design guidelines can help system developers and community pharmacy managers realize more effective MIRLS.

(JMIR Hum Factors 2024;11:e41557) doi:10.2196/41557

KEYWORDS

medication incident; reporting system; persuasive technology; persuasive design; medication; persuasive system design; pharmacy; pharmaceutic; pharmacology; drug reporting; drug event; adverse event; incident management

Introduction

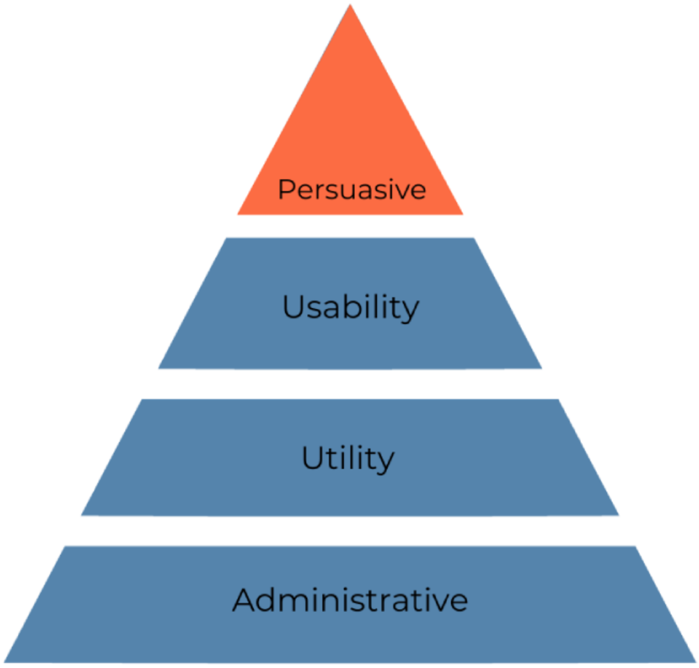
Overview

Medication errors are one of the leading causes of death in many countries worldwide [1,2]. For example, in the United States alone, 7000 to 9000 patients die annually owing to these errors. In Canada, where medical errors (labeled as the third leading cause of death after cancer and heart disease) account for 28,000 deaths annually, every minute and 18 seconds a patient gets harmed because of unintended errors, with medication errors being the most frequent [3]. Wrong medication (eg, because of similar naming, similar packaging, illegible handwriting, and incorrect drug selection) and wrong dose are among the most common medication errors in community pharmacies [3-5]. In particular, advanced drug preparation and administration without double checking [6] and heavy workflow [7] have been identified as key contributing factors to medication errors. However, there may be many more contributing underlying factors that go unreported by pharmacists and other health professionals. For example, a survey on medication administration errors among nurses in South Korea showed that 63.6% of the respondents had been involved in medication errors once or more in the previous month. However, only 28.3% of the participants reported the incidents [6]. Underreporting of medication errors, which is a global issue [8-11], has several implications bordering on shared learning, patient safety, and financial cost. In the United States, for example, psychological or physical pain and distress aside, “the total cost of looking after patients with medication-associated errors exceeds US $40 billion each year, with over 7 million patients affected” [4]. Moreover, underreporting of medication errors and incidents might limit individual and organizational learning from their occurrence [12,13].

The continuous evolution of pharmacotherapy and changing demands on the community pharmacy necessitate constant vigilance to detect new types of medication errors [14]. In a study among hospital pharmacists in South Korea, Hee-Jin et al [15] found that “five or more near misses per month were experienced by 14.8%, 4.3%, and 43.9% of respondents for dispensing, administration, and prescribing errors, respectively.” Moreover, research has shown that medication errors that lead to patient harm are common in medical care including community pharmacy [2,16-19]. Frequent reporting of all medication incidents (MIs) and near-miss events has the potential to improve patient safety through shared learning, which will enable the reduction of recurrence and prevention of MIs in the future [20,21]. Without adequate user reporting, none of the laudable objectives of reporting systems, including identification of gaps and resource development to support patient safety, can be achieved [7]. Medication error reporting is a common metric used to assess the quality of care provided by the health care system [21]. However, research has shown that employees are less motivated to report medication errors [22-25]. Hence, there is a need to find ways to motivate pharmacists and pharmacy technicians to report MIs more often to foster shared learning, prevention of recurrence, and patient safety. The question then is, *How can we motivate pharmacists and pharmacy technicians to report MIs more frequently using persuasive design principles embedded in digital technologies?*

Although some guiding principles have been proposed to alleviate the barriers to MI reporting, these principles, from a user experience (UX) design perspective, are not aimed at motivating pharmacists to report MIs regularly. From our literature search, we identified 4 categories of principles that can guide the design of Medication Incident Reporting and Learning Systems (MIRLS) to improve their adoption and usability. They include administrative principles, usability, utility principles, and persuasive design principles (Figure 1). Administrative principles refer to the organizational processes and policies implemented to enable and encourage employees to report medication errors regularly without fear of consequences. These principles form the basis of MIRLS, upon which the other categories of principles build. Usability and utility principles refer to the UX features that enable a user to report medication errors with ease, effectiveness, efficiency, and satisfaction [26]. Persuasive principles refer to the motivational affordances of a system that facilitate, nudge, and motivate a user to report medication errors. Current MIRLS mainly focus on the administrative, usability and utility-based principles. Typical examples of administrative principles include voluntary use, anonymity, confidentiality, and nonpunitive consequences. Examples of usability and utility-based principles, particularly in the Think Research and Pharmapod system, include ease of use, use of a standard taxonomy, searchability, retrievability, report generation, and root cause analysis [7,14,27].

Figure 1. Four key design principles for Medication Incident Reporting and Learning Systems in community pharmacy.



Apart from the administrative, usability, and utility principles, we argue that persuasive design principles hold potential to increase MI reporting among pharmacists. Persuasive design principles embedded in digital technologies, also known as persuasive technology, can motivate increased reporting of MIs from community pharmacies, as research in other health domains has shown [28]. Hence, this study proposes the use of persuasive design principles, which build on the 3 other categories of principles (Figure 1), to motivate users of MIRLS to report incidents and near misses more often.

Using Think Research, also known as Pharmapod, a cloud-based MIRLS for reporting and reducing incidents in community pharmacies [29], as a case study, this study (1) assesses 1 MIRLS based on the Persuasive System Design (PSD) taxonomy proposed by Oinas-Kukkonen and Harjumaa [30] and (2) proposes persuasive design guidelines to help community pharmacy stakeholders at multiple levels (eg, facility, provincial, and national) integrate persuasive features into their MIRLS. The PSD taxonomy is a widely used framework in the persuasive technology domain for analyzing, designing, implementing, and evaluating persuasive systems. Persuasive strategies from the PSD taxonomy can enhance MIRLS, making them more effective in promoting patient safety and shared learnings among practitioners [31,32]. Moreover, the study presents a summary of the results on the relationship between medication errors and harm and the information system–based barriers to MI reporting to ground the research.

Background and Related Work

In this section, we present an overview of relevant studies on the relationship between medication errors and patient harm and the organizational and information system barriers to reporting.

Medication Errors and Patient Harm

Several studies have been conducted to investigate the prevalence, nature, severity, and effects of MIs. West et al [16] investigated the relationship between medical errors and patient harm in primary care. They found that clinical harm to patients was reported in >10% of the 608 primary care medical error reports, with prescription-related errors most frequently linked to clinical harm. Similarly, Robb et al [17] investigated the relationship between medication and patient harm in hospitals in New Zealand. The authors confirmed the findings of earlier studies that showed that medication-related harms were common in both hospitals and the community, posing a substantial burden for patients and the health care system. In particular, they found that 923 harms were identified among 751 patients, with 28% of them experiencing ≥1 of the medication-related harms. They also found that older and female patients and those who had an increased length of stay were more likely to be harmed. Moreover, 65% of the harms occurred during an inpatient stay and 29% originated from the community and resulted in an admission. Riordan et al [18] investigated discharge prescription errors and their propagation after the discharge of patients. They found that 43% of the patients included in the study experienced postdischarge medication errors, with 86% of them being at risk of moderate harm. Moreover, 88% of the errors were discharge prescription errors that persisted after the discharge.

Most recently, Alqenae et al [2] conducted a systematic review, which they regarded as the first, to explore the prevalence and nature of medication errors and adverse drug events after hospital discharge. The review found that the median rate of medication error was approximately 50% among adult and older patients after hospital discharge, with approximately 20% of the patients in the studies reported to be affected by adverse drug events (such as antibiotics, antidiabetics, analgesics, and cardiovascular drugs) after hospital discharge. Panagioti et al [19] conducted a systematic review and meta-analysis of the prevalence, severity, and nature of preventable patient harm across a range of medical care settings. They found that 5% of the patients were exposed to preventable harm in medical care and 25% of the incidents, which are drug related, accounted for the largest proportion of preventable patient harm, with 12% of the preventable patient harms being severe or leading to death. They asserted that there are limited quality improvement practices specifically targeting incidents that cause preventable harm to patients. They added that designing and implementing evidence-based mitigation strategies specifically targeting preventable patient harm could lead to substantial service quality improvements that are cost effective. This conclusion by Panagioti et al [19], coupled with the prevalence of medication errors in community pharmacy, partly informs this conceptual paper aimed to incorporate persuasive principles in MIRLS to increase medication error reporting and patient safety.

Organizational Barriers to MI Reporting

Researchers have identified several organizational barriers (both administrative and personal) leading to underreporting of medication errors and incidents in community pharmacy [21]. In the long run, these barriers can adversely affect patient safety owing to lack of shared learning among pharmacists within and across organizations because of underreporting [12,13]. Key barriers include fear of consequences such as punitive and disciplinary actions, negative or lack of administrative feedback, poor work climate or culture, inadequate training, and time constraint (Textbox 1) [7,8,21]. For example, Bahadori et al [9] found that the most important reasons for not reporting medication errors were administrative factors including the process of reporting and fear of the consequences of reporting. Research has also shown that personal (ie, sociodemographic) factors can impact medication error reporting. For instance, Aljabari and Kadhim [8] found that younger and lesser experienced professionals and staff with shorter employment periods were less likely to report medication errors. We argue that administrative barriers (such as time constraint and high workload) and perceived low value of the reporting system could be mitigated by using persuasive technologies to facilitate and ensure convenient reporting of MIs and errors. For example, persuasive design features (such as reminders to complete saved draft reports, notifications about the utility and value of reporting, and encouraging messages) may facilitate MI reporting.

Textbox 1. Administrative barriers to reporting medication errors and incidents.

Fear of consequences

- Negative consequences such as blame, shame, professional reputation damage, relationship damage, loss of privileges, medical malpractice lawsuit, relief from certain duties, and loss of job [4,9,33,34].

Lack of feedback

- Lack of useful feedback or negative feedback from administrative teams, such as pharmacy managers, regarding previously reported medication errors [33,34].

Poor work climate or culture

- Blaming staff and not the system or culture, poor support system, poor teamwork, poor organizational leadership, and lack of confidentiality in handling reports [33,35].

Miscommunication

- Poor communication among staff or between staff and patients [36].

Inadequate training of staff

- Difficulty in using the reporting system, poor understanding of the importance or value of reporting, poor understanding of errors, lack of clear definition of incident or near miss, and lack of a well-defined protocol on what events need to be reported [21,35].

Time constraint

- Work pressure and the lack of budgeted time to properly report errors, especially in the midst of a busy work schedule and high workload resulting in lack of enough breaks [7,35,36].

Information System Barriers to Patient Safety

Research has identified technological barriers that hamper patient safety in different health information systems and domains [37-40]. The primary barrier among them is the usability and poor design of health information systems [41]. Ratwani et al [42] found across 3 health care institutions that the usability of electronic health records accounted for more than a third of medication errors in 9000 pediatric patient safety reports. Kushniruk et al [43] evaluated the usability of a handheld prescription writing application. They found various usability problems (most of which relate to interface design) and actual errors in entering prescription data. In particular, they found that certain types of usability problems such as display visibility and ergonomics-related wrong data entry were closely linked to the occurrence of specific types of errors in medication prescription. More recently, Adams et al [37] investigated the medication errors associated with health information technology use and the harm caused to the patient. They found that 55.85% (1508/2700) of the manually reviewed reports described a medication error associated with information technology use and 49.7% (750/1508) of these caused harm to the patient. In particular, they found that 97.35% (1468/1508) of the medication errors associated with information technology were related to usability issues including data entry, workflow support, and alerting. On the basis of these findings, in the current MIRLS domain, we set out to uncover the information technology barriers that border on the usability and utility principles (Figure 1), which may lead to the low perceived value, utility, and use of MIRLS.

PSD of MIRLS

PSD was pioneered by Fogg [44] in the early 2000s in his seminal book, "Persuasive technology: using computers to change what we think and do." This entails the application of behavioral theories from social psychology in the design of technology-based systems to motivate behavior change. Hence, persuasive technology is defined as a motivational tool intentionally designed to change human attitudes and behaviors through persuasive techniques grounded in social psychology [44]. Fogg [44] first proposed a set of 7 persuasive strategies to motivate behavior change. Subsequently, Oinas-Kukkonen and Harjumaa [30] extended the list to 28 persuasive strategies, which are categorized into 4 functional groups (primary task support, dialogue support, social support, and system credibility support), each comprising 7 persuasive strategies. Oyibo [45] extended the primary task support and dialogue support groups with goal setting and verbal persuasion, respectively, increasing the total number of strategies in the PSD taxonomy to 30. The primary task support category, which includes tunneling, tailoring, and self-monitoring, is aimed at helping the user to perform a target behavior easily and effectively. The dialogue support category, which includes praise, reward, and suggestion, is aimed at motivating the user to perform the target behavior through feedback and dialogue with the persuasive system. The social support category, which includes social learning, social comparison, and competition, is aimed at motivating the user through social influence to perform the target behavior. Finally, the system credibility support category, which includes trustworthiness, surface credibility, and authority, is aimed at increasing the user's trust in the system by making the system look professional and credible [46].

Incorporating persuasive features into MIRLS has the potential to improve the rate of error reporting. St-Maurice et al [28] showed that, on average, the percentage of same-day data entries can be increased by 10% for each user by introducing new persuasive design features into a data entry system. On the basis of this prior research finding, we propose guidelines for incorporating persuasive design principles, drawn from the PSD taxonomy, into MIRLS using the Think Research or Pharmapod Incident Management (IM) system as a case study. The PSD taxonomy, which comprises 4 categories of persuasive strategies (primary task support, dialogue support, social support, and system credibility support), is a framework for analyzing, designing, implementing, and evaluating persuasive systems. A systematic review by Win et al [47] showed that primary task support and dialogue support are the most commonly used categories of persuasive strategies in medication management information systems. The review reported that tailoring, self-monitoring, and reminders, which belong to the primary task support category, are more likely to be implemented in medication management information systems than other persuasive strategies. In the case of MIRLS, the proposed persuasive strategy guidelines are aimed at enhancing system utility and facilitating the reporting of near misses and incidents. Research shows that the higher the perceived usefulness of health systems, the higher the number of users who find them more persuasive [48].

Methods

Overview

A total of 2 types of methods were used to address 3 research questions (RQs). They include scoping review and assessment of an existing MIRLS based on administrative, usability, utility, and persuasive features. The RQs are as follows:

1. RQ1. Is there a relationship between medication errors and patient harm?
2. RQ2. What are the information system–based barriers preventing pharmacists and pharmacy technicians from reporting medication errors?
3. RQ3. How can we motivate them to report MIs more frequently using persuasive design principles embedded in digital technologies?

Ethical Considerations

The assessment of our target system was aimed at quality improvement, thus ethical approval was not required [49,50].

Scoping Reviews

To address the first 2 RQs, the authors (KO, SE, and TN) conducted 2 scoping views in August 2023. The first review investigated the relationship between medication errors and patient harm in the pharmacy domain. The second review aimed to uncover usability and utility-related barriers to medication error reporting. We retrieved articles from 6 databases for each study, screened the articles, extracted the relevant data, and presented the results. For the first review, a total of 820 articles were retrieved from PubMed (n=41), Public Health Database (n=89), ProQuest (n=451), Scopus (n=97), ACM (n=42), and Global Health (n=22) using the search string: "(Medic* OR prescri* OR administ* OR drug*) AND (error* OR incident* OR accident* OR nearmiss* OR 'near miss*' OR mistake*) AND patient AND (harm* OR hurt* OR injur* OR wound* OR bruise* OR impairment* OR afflict*) AND pharmac*." A total of 215 duplicates were removed to arrive at 605 unique articles. These articles were screened based on title or abstract to arrive at 91 articles. Next, a full-text review was conducted to arrive at 14 included articles after excluding 77 ineligible articles. Finally, 3 more articles were included to the 14 through Google Scholar search, resulting in 17 articles for the final data analysis. For the second review, a total of 849 articles were retrieved from PubMed (n=268), Public Health Database (n=44), ProQuest (n=90), Scopus (n=448), ACM (n=10), and Global Health (n=45) using the search string: "(Medic* OR prescri* OR administ* OR drug*) AND (error* OR incident* OR accident* OR nearmiss* OR 'near miss*' OR mistake*) AND (report* OR submi* OR log*) AND (system* OR application* OR website* OR tool* OR platform* OR interface* OR technolog*) AND pharmac* AND (barrier* OR hinderance*

OR obstacle* OR drawback* OR setback* OR deterrent* OR limitation* OR shortcoming*." A total of 303 duplicates were removed to arrive at 546 unique articles. These articles were screened based on title or abstract to arrive at 12 articles. Upon the full-text review, we arrived at zero article for data extraction and analysis. Moreover, based on Google Scholar search, we found 1 article [13] that investigated the usability of MIRLS called the Medication Error Reporting App. However, this study did not investigate the relationship between the usability of the app and medication error.

Overview and Initial Assessment of an Existing MIRLS

The authors (KO and PAG) analyzed the Think Research or Pharmapod MIRLS, which is a cloud-based software platform for reporting medication errors (incidents and near misses). As stated on its website, Think Research or Pharmapod describes itself as "the first platform of its kind to pool and share patient safety data across borders, monitoring trends and causes behind medication errors, and empowering healthcare professionals locally to improve their practice" [29]. Our initial review of the system assessed it against the 3 key design principles shown in Figure 1. To assess the administrative and usability and utility principles, the first 2 authors went through the Think Research or Pharmapod system from one interface to another to elicit the supported principles. Next, we used the PSD taxonomy as an assessment framework and 3 assessors (study participants) to identify persuasive strategies fully or partially implemented in the Think Research or Pharmapod IM system. We first assessed the system to identify the existing persuasive strategies and then gathered data from 3 experienced users to propose opportunities for improvement. One of the authors, the vice president of the Quality Improvement and Innovations of Think Research or Pharmapod, arranged for 3 independent and experienced users of the Think Research or Pharmapod IM system from different pharmacies to assess the system against the PSD taxonomy and items. The first assessor was a pharmacist who had 1.5 years of experience using the system. The second assessor was a director in a health care company focused on patient and staff safety, with 1 year of experience working with the system. The third assessor was a senior technology manager at a leading Canadian pharmacy company, with 4 years of experience working with the system.

The authors (KO and PAG) asked the assessors to independently indicate whether each persuasive strategy in the PSD taxonomy is important or useful, present in the system or not, and where it could be found in the system. The implementation of each strategy from the PSD taxonomy was described to the participants in a tabular form. The participants independently responded to the questions and then came together to discuss and confirm their responses and resolve their differences with the first 2 authors. If at least 2 of the 3 assessors indicated or agreed that a given persuasive strategy is important and useful, "yes" is entered into the associated cell in the table, otherwise, "no." Similarly, if at least 2 assessors agreed that the strategy was present in the system (ie, said "yes"), "√" is entered into the cell associated with the status column. However, if ≥2 assessors agreed that the strategy was not present in the system (ie, said "no"), "X" is entered into the associated cell under the status column. Moreover, if at least 1 of the assessors agreed that the strategy was present in the system, but the implementation was limited, "√*" standing for "present but could be improved" is entered into the associated cell under the status column.

Results

In this section, we present the results of the scoping reviews and the initial assessment of the Think Research or Pharmapod IM system.

Scoping Reviews

Medication Errors and Patient Harm

In the first review, 41% (7/17) of the included articles originated from North America (United States [16,51-53], Canada [54,55], and Mexico [56]), 29% (5/17) from Europe (United Kingdom [12], Ireland [18], the Netherlands [57], Sweden [58], and Spain [59]), 23% (4/17) from Asia (Saudi Arabia [60,61], China [62], and Korea [63]), and 6% (1/17) from Oceania (New Zealand [17]). The articles were published between 2001 and 2023, with most of the articles (3/17, 18%) published in 2023. Most of the target populations were from North America (7/17, 41%), followed by Asia (5/17, 29%), Europe (4/17, 23%), and Oceania (1/17, 6%). Of the 17 articles, 1 (6%) each focused on target populations in Africa and South America. Table 1 shows 16 types of harms elicited from the included articles. These were caused by 59 types of medication errors such as wrong drugs, missing or wrong patient weight, prescription errors, dosing error, wrong or unclear dose or strength, wrong patient, and wrong duration, each of which was reported by at least 2 articles. Significant or serious harm was the most frequent harm; it was reported by 65% (11/17) of the articles, followed by death or fatal harm (7/17, 41%) and no harm or potential harm (4/17, 23%).

Table 1. Type or severity of harm caused by medication errors and the number of articles associated with them (N=17).

Type or severity of harm	Articles, n (%)
Significant or serious harm	11 (65)
Death or fatal harm	7 (41)
No harm or potential harm	4 (23)
Inconvenience	3 (18)
Adverse drug events	3 (18)
Mild harm	2 (12)
Moderate harm	2 (12)
Temporary injury or harm	2 (12)
Prolonged hospitalization	2 (12)
Life-threatening harm	2 (12)
Nonlife threatening	1 (6)
Risk to patient or others	1 (6)
Unstable situation	1 (6)
Unknown harm	1 (6)
Permanent harm	1 (6)
Intervention required	1 (6)

Information System Barriers to MI Reporting

One article [13] that investigated the usability of an MIRLS prototype called Medication Error Reporting App found that there was significant improvement in the mean usability score throughout the development process ($P<.001$). However, this mean improvement in usability did not impact the mean time to report medication errors using the app because the mean time was not significantly different between the phases of the development process. Overall, it was found that the testers including pharmacists found the app easy to use, but doctors and nurses were unfamiliar with the medication terms used, especially the medication process in which error occurred and type of error. More importantly, the authors reported that although testers might be willing to adopt the app to make reports in the future, they were apprehensive about data protection issues such as security and abuse of feedback featured in the app [13].

Initial Assessment of Existing MIRLS

In this section, we present the results that emanated from the initial assessment of the Think Research or Pharmapod IM system based on administrative, usability, utility, and persuasive principles.

Administrative and Usability and Utility Principle Support

The assessed system supported at least 75% (6/8) of the administrative guiding principles shown in Textbox 2, including voluntariness, anonymity, confidentiality of information, and nonpunitive measures. It also supported all 7 usability and utility-based principles, including ease of use, searchability and retrievability, standard taxonomy, report generation, and root cause analysis (Table 1).

Textbox 2. Items and questions asked of assessors.

Strategy code

- A codeword representing the persuasive strategy.

System capability

- A description of the persuasive strategy.

Important or useful

- An indication of the importance or usefulness of the strategy (yes or no).

Present in system

- An indication of the presence of the strategy in the system (yes or no).

Interface, tab, or comment

- Provision of the system interface or tab where the persuasive strategy can be found or a comment by the assessor.

Moreover, the system promotes 4 key elements of patient medication safety: report, document, analyze, and share (Figure 2). The analyze and share elements are in addition supported by 6 main continuous quality improvement (CQI) tools. These tools are intended to foster patient safety in community pharmacy within a pharmacy team [20,64]. The tools include event summary, risk matrix, 5-whys template, action plan, learning points, and pharmacy safety self-assessment (Figure 3). An event summary is an incident and a root cause analysis tool. The risk matrix is a color-coded matrix that facilitates the assignment of a risk score based on the probability of recurrence of the incident or near miss at a specific severity level and its impact on a patient if it were to recur. The 5-whys is a tool that facilitates the analysis of an incident or near miss by answering the fundamental question, "Why did the incident occur?" 5 times. The 5-whys is a simple and well-recognized tool for determining the cause and effect of an incident objectively. Action plans is a tool to create and track smart actions of improvement. Learning points organizes identified gaps, for example, in workflows and processes and provides a means to share these learnings. Finally, the pharmacy safety self-assessment is a tool that allows the pharmacy team to proactively identify risks that may compromise patient safety and implement safe medication measures to address them [64].

Figure 2. Four key elements of patient medication safety.

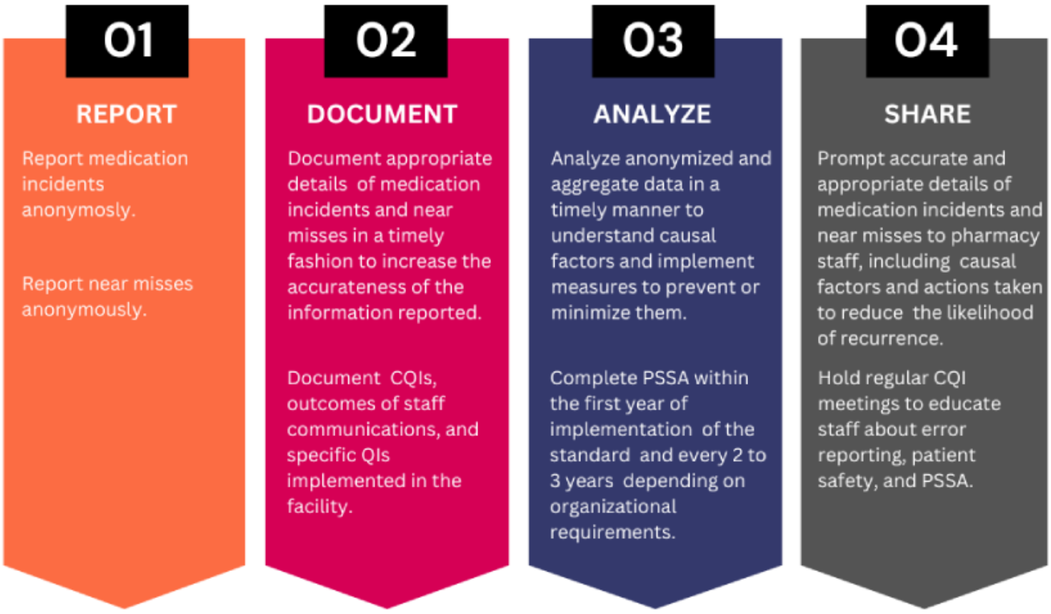
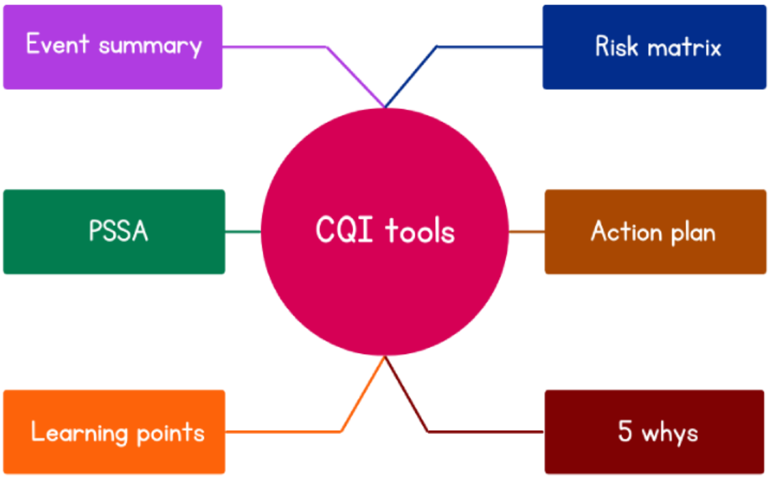


Figure 3. Six continuous quality improvement (CQI) tools in the studied system aimed at fostering fundamental change among pharmacy team members. PSSA: pharmacy safety self-assessment.



Persuasive Principle Support

Tables 2-5 show the results of the assessment of the Think Research or Pharmapod IM system based on the primary, dialogue, social, and credibility support categories of the PSD taxonomy, respectively. The first column of Table 2 captures the coded name of the strategy and its description, the second column describes a yes or no response on the importance and usefulness of the strategy, and the third column describes a yes or no response on the presence of the strategy in the system (ie, status). A fourth column was also provided for the assessors to comment on the assessment of each strategy, for example, the location of the strategy in the system.

Table 2. Guidelines for incorporating the primary task support principles into Medication Incident Reporting and Learning Systems.

Strategy and implementation	I or U[a]	Status
Reduction		
Break down the medication incident and near-miss reporting process into a few simple steps to facilitate reporting [65].	Yes	✓[b]
Tunneling		
Guide the user through the reporting process in a step-by-step fashion, just as a software installation wizard [47].	Yes	X[c]
Goal setting		
Allow the user to set a goal, for example, minimum number of errors or CQI[d] reports to be submitted over a given period such as a week or month.	Yes	X
Self-monitoring		
Allow the user to track their progress after setting a report-based goal or when submitting a report, for example, through the display of a progress bar.	Yes	X
Allow the user to view their number of completed and uncompleted reports and averages per week, month, or year (eg, on their dashboard).	Yes	✓
Allow the user to track the levels of usefulness of their reports (eg, CQI, incident, or near miss) to others, for example, other users or colleagues "like" their anonymous reports as obtainable in YouTube and Facebook.	Yes	X
Tailoring		
Tailor what the user sees (eg, user profile, chart, content, and information) using group-based characteristics such as work experience and designation or role.	Yes	✓*[e]
Personalization		
Personalize the system (eg, information, report, and reminder) based on their interaction, for example, letting the user know where they left off or reminding them about incomplete tasks when they log in [7,66].	Yes	X
Customization		
Allow the user to customize the system (eg, profile, chart, content, information, and reminder) to suit their needs and preferences [66].	Yes	✓*
Simulation		
Show the user a cause-and-effect relationship of the benefit of incident or near miss or CQI reporting, for example, a study chart showing the higher the incidents reported, the lower the number of recurrences.	Yes	X
Rehearsal		
Provide a new user with a simulated environment to rehearse before making an actual report relating to an incident, near miss, or CQI.	Yes	✓*
Provide a new user with video tutorials on how to report a medication incident or near miss.	Yes	✓

[a]I or U: important or useful.

[b]Currently implemented.

[c]Not currently implemented.

[d]CQI: continuous quality improvement.

[e]Partially implemented and could be improved.

Table 3. Guidelines for incorporating dialogue support principles into Medication Incident Reporting and Learning Systems.

Strategy and implementation	I or U[a]	Status
Praise		
As a show of appreciation, praise or congratulate the user for submitting a near-miss or incident or CQI[b] report or for reaching a milestone using textual, visual, or audio-based feedback messages [67].	Yes	X[c]
Reward		
Reward the user with points, badges, etc, when they submit a report (early), achieve a goal or milestone, or others find their report useful (eg, by liking it), etc.	Yes	X
Allow the user over time to grow in the value of their contribution to the community. This can be based on the number, frequency, quality, earliness, and usefulness of their reports (to others), for example, from a silver to a gold valuable contributor of the community.	Yes	X
Reward the user for reporting or sharing action plans that improved safety in the pharmacy.	Yes	X
Reward the user for reporting positive experiences that led to improved safety in the pharmacy, for example, “good news” stories in addition to the negative “error” reports.	Yes	X
Suggestion		
Suggest to the user from time to time based on their profile, role, or interaction with the system new reports that may be interesting and beneficial to their practice [65].	Yes	X
Suggest to the user ways, processes, or methods through which others in the community prevent or address recurrence of certain near misses and incidents.	Yes	X
Suggest to the user standard, process-based solutions (eg, from the user’s pharmacy, province, or professional organization) for addressing certain types of recurring incidents and near misses [65].	Yes	X
Provide the user with a list of “high-alert” medications or types of incidents that occur most often or require extra precautions and suggest best practices to reduce incidents and near misses associated with them [68].	Yes	✓*[d]
Feedback[e]		
Provide the user with summary feedback on their progress toward reaching their monthly, quarterly, or yearly goal (eg, “You have achieved 30% of your goal”).	Yes	X
Provide the user with summary feedback on the usefulness of their reports to others (eg, “5% of the system users in the province [nation] found your report helpful”).	Yes	X
Provide the user monthly, quarterly, or yearly summary feedback highlighting the most recurring types of near misses and incidents (eg, “Poor drug naming caused 5% of the near misses last year”) [65].	Yes	✓[f]
Reminder		
Remind the user from time to time (eg, based on self-set goals) about the need to report near misses and incidents and about the benefits to other users and patient safety.	Yes	X
Remind the user from time to time to complete their CQI action plan that they have started.	Yes	X
Verbal persuasion[e]		
Allow management such as pharmacy managers and supervisors through personally sent messages to encourage users from time to time to report near misses and incidents, for example, “Alice, remember to report your near misses and incidents to improve patient safety. Yes, you can*!*”	N/A[g]	N/A
Emotional appeal[e]		
Use motivational messages to encourage users to report errors, for example, “To err is human, to share is divine” [69].	Yes	X
Liking		
Make the system to be visually attractive, for example, by using visually pleasing or appropriate colors to present charts, content, and important information.	Yes	✓*

[a]I or U: important or useful.

[b]CQI: continuous quality improvement.

[c]Not currently implemented.

[d]Partially implemented and could be improved.

[e]Not originally listed in the Persuasive System Design taxonomy.

[f]Currently implemented.

[g]N/A: not applicable.

Table 4. Guidelines for incorporating social support principles into Medication Incident Reporting and Learning Systems.

Strategy and implementation	I or U[a]	Status
Social learning		
Notify the user by email when other anonymous users submit an incident report (eg, containing the key points) or CQI[b] report that may be of interest to the user, just as in ResearchGate, for example, "John [a pseudonym], here's a new report we think you'll be interested in."	Yes	X[c]
Support chat room and discussion room to foster social support and shared learning [47]. This room can be anonymous.	Yes	X
Support a newsfeed (eg, as in Facebook) to highlight important reports the user may find useful and foster shared learning.	Yes	X
Social comparison		
Allow the user to compare their weekly, monthly, quarterly, or yearly reports with others, maintaining confidentiality (eg, at the city, zone, provincial, or national level).	Yes	✓*[d]
Competition		
Allow the user to see where they are compared with other anonymous (eg, on a leaderboard) at the pharmacy, provincial, or national level based on the total number, frequency, quality, or usefulness of their report to others (eg, over a weekly, monthly, or yearly period).	Yes	X
Cooperation		
Provide users the choice of being paired with another anonymous user, with the goal of motivating one another to achieve individual or collective goals.	No	X
Normative influence		
Inform users about the number of other anonymous users in the pharmacy, province, or nation that are reporting errors in a given period (eg, "10 other people submitted their incident reports today").	Yes	X
Social facilitation		
Make users, who are logged onto the system know that there are other anonymous users elsewhere (eg, in the facility, province, and nation), who are submitting or just submitted a report (eg, "5 other people are currently submitting their incident reports").	Yes	X
Social recognition		
Provide a means for committed user to be publicly recognized for being one of the "most valuable players" of the month, quarter, or year at the pharmacy, provincial, or national level based on certain criteria (eg, number, frequency, quality, or usefulness of their reports to the community).	Yes	X
Allow other users to rate users' reports anonymously based on how useful or helpful it is to them.	No	X

[a]I or U: important or useful.

[b]CQI: continuous quality improvement.

[c]Not currently implemented.

[d]Partially implemented and could be improved.

Table 5. Guidelines for incorporating system credibility support principles into Medication Incident Reporting and Learning Systems.

Strategy	Implementation	I or U[a]	Status
Authority	Present authority-based information and messages (eg, on the value of reporting incidents and near misses and the benefits it can have for the profession, staff, or patient safety) [47].	Yes	X[b]
Third-party endorsement	Demonstrate that the system is approved by authorities such as professional organizations, regulatory bodies, and government, for example, by displaying their corporate logos [65].	Yes	X
Expertise	The visual and functional design of the system should reflect professionalism, expertise, and be up to date to motivate users to use it.	Yes	✓*[c]
Trustworthiness	Build trust into the system, for example, by fostering anonymity, data aggregation, and keeping promises such as it not being used as a punitive tool to hold users accountable [65].	Yes	✓*
Surface credibility	Build surface credibility into the system through its visual design, for example, by reducing advertisements and ensuring users enter accurate information using taxonomy-based pre-defined options, checklists, and drop-downs [47].	Yes	✓[d]
Verifiability	Ensure presented information and messages (eg, on the value of error reporting to the profession, staff, or patient safety) are verifiable, for example, through a link to authority-based websites such as Institute for Safe Medication Practices and World Health Organization.	Yes	X
Real-world feel	The design of the system should mimic the paper-based error reporting forms (eg, [70]) as closely as possible to reduce the cognitive effort required by a new user to make the transition [71].	Yes	✓

[a]I or U: important or useful.

[b]Not currently implemented.

[c]Partially implemented and could be improved.

[d]Currently implemented.

Primary support strategies facilitate the key behaviors promoted by the system, such as reporting. Dialogue support strategies enable users to interact, engage with, and receive feedback from the system through text-, image-, audio-, and video-based dialogue. Social support strategies motivate users through social influence. Finally, credibility support strategies enable users to trust and rely on the system. In summary, based on the assessors' responses, most of the persuasive strategies (29/31, 94%) in the extended PSD taxonomy were considered important or useful, with approximately one-third (14/29, 48%) of them identified as present in the current Think Research or Pharmapod system. Approximately 23% (7/31) and 26% (8/31) of the strategies were considered fully or partially implemented (although they could be improved), respectively. More than 50% (16/31) of the strategies were considered not implemented, with most of them falling under the social support category.

Discussion

We have presented the results of 2 scoping reviews and the initial assessment of the Think Research or Pharmapod system. The following sections discuss the results with a focus on the persuasive design guidelines shown in Tables 2-5, which can inform the persuasive design of future MIRLS.

Summary of Scoping Review Findings

Table 1 shows the types of harm uncovered in the first scoping review. More than half (11/17, 65%) of the included articles reported that medicated errors caused serious harm to patients. In particular, 60% (3/5) of the articles reported serious harm, and 40% (2/5) of the articles reported fatal harm or death caused by medication errors such as wrong dose, drug, patient, and ambulatory pump (eg, [58]). Prescription error [16,18,59,63], wrong drugs [12,58,63], and dosing error [58,59,63] were the most frequent medication errors. For example, in the study by Fyhr and Akselsson [58], most severe medication errors occurred during prescribing and transcribing by physicians. The findings are an indication that medication errors have the potential to cause serious harm to patients, including death; hence, there is a need for interventions aimed to reduce them and increase patient safety (eg, by increasing reporting and shared learning within and across organizations). Moreover, in the second review on the usability of MIRLS, George et al [13] found that an iterative design has the potential to improve the usability of an MIRLS. However, their study suggested that there is a need to address issues surrounding data security and report validation to increase user acceptance and use.

Summary of Administrative and Usability and Utility Assessment

Our assessment shows that the Think Research or Pharmapod system implemented most of the administrative, usability and utility-based principles shown in Textboxes 1, 3 [7,14], and 4 [7,14]. Prior studies advocate most of these principles as essential actions and capabilities aimed at improving incident reporting and shared learning [7,9,14,33-36]. An anonymous reporting, for example, can mitigate the punitive perceptions of incident reporting [20]. However, the system only partially supported persuasive design principles. Persuasive design principles are intended to complement the administrative, usability and utility-based principles by improving the UX and motivating users to see value in reporting MIs and completing the CQI and learning tool reports. Persuasive design may in turn mitigate some of the persistent barriers identified in Textbox 1.

Textbox 3. Administrative guiding principles for designing Medication Incident Reporting and Learning Systems [7,14].

Voluntariness

- Medication reporting will be voluntary.

Inclusiveness

- Professionals and consumers will be encouraged to participate.

Aggregation

- The reporting system will support anonymity and aggregation.

Confidentiality

- The system will provide confidentiality of reported information.

No consequence

- The system will clearly define and support a nonpunitive approach to reporting.

Type of report

- The system will encourage reporting of both potential and actual incidents and near misses.

Feedback

- The system will provide feedback on incident analysis and timely recommendations.

Workflow alignment

- The system should fit with the users' workflow.

Textbox 4. Usability and utility-based guiding principles for designing Medication Incident Reporting and Learning Systems [7,14].

Usability

- The system will be easy to use and time efficient.

Format multiplicity

- The system will support both electronic and paper formats.

Taxonomy

- The system will support standard taxonomy.

Outcome severity

- The system will support levels of severity of outcomes.

Searchability and retrievability

- The system will support searchable and retrievable data.

Report generation

- The system will support report generation.

Root cause analysis

- The system will support root cause analysis.

Summary of Persuasive Design Assessment

In this section, we discuss the results of the system assessment and persuasive design guidelines for designing future MIRLS, taking each category of the PSD taxonomy at a time.

Primary Support Assessment and Guidelines

In the primary support category (Table 2), all the persuasive strategies were considered important or useful, whereas over

55% (5/9) of them were deemed partially or fully implemented by the system.

Reduction

Reduction, which is considered important and present in the system by the assessors, entails breaking down the performance of a complex behavior into a few steps. In the context of MI reporting, this means making the reporting process simple and easy to carry out by users. Reduction is vital to ensuring and facilitating the report of MIs and near misses given the relatively high workload health professionals such as pharmacists handle on a daily basis [65]. In the Think Research or Pharmapod system, for example, to speed up the reporting process, predefined fields and system design widgets such as drop-downs are used to enter information about prescribed drugs, what happened, contributing factors, and harm caused. A critical aspect in realizing the effectiveness of the implementation of this and other PSD guidelines is the fit of the MI reporting task into users' workflow to facilitate regular reporting [7]. However, this examination is beyond the scope of this conceptual study.

Tunneling

Similar to reduction, tunneling (aka guided persuasion) aims at motivating users to report MIs and near misses. The tunneling strategy, which can be likened to the process of installing software on a computer using an installation wizard [72], is used to walk the user through predetermined steps in a structured manner. Two of the assessors agreed that tunneling is important or useful but not present in the studied system, with 1 of them remarking, "the report has four sections, then the CQI has colour coded features but they do not tunnel you in any direction." Once an incident report is completed and saved, the incident analysis interface (an event summary page containing a variety of management tools to prevent the recurrence of similar events in the future) opens automatically. However, the system does not tunnel the user in a specific direction. The third assessor, however, did not find tunneling useful in this context and commented, "No this MIRLS is not like an installation wizard. We like the flexibility provided today." Hence, owing to the mixed responses, a more comprehensive study among a larger target audience is required to understand the perceived usefulness of tunneling.

Goal Setting

Related to the commitment principle proposed by Cialdini [73], goal setting is known as one of the cornerstones of persuasive systems [74]. According to the commitment principle, people are more likely to follow through with a behavior if they make a commitment in written or verbal form to perform the behavior [75]. Studies have shown that people, regardless of culture, are more motivated by the commitment principle than by the other 5 principles of persuasion proposed by Cialdini [76,77]. Goal setting is more likely to be effective if set goals are specific, measurable, achievable, relevant, and timebound (SMART) [77]. The assessors agreed that goal setting is important or useful in MI reporting. One participant thought that the feature was present in the system already. Here, the assessor meant the CQI action planning. In general, both goal setting and action planning are related. However, action planning is concerned with how set goals can be achieved [78]. In the studied system, the CQI actions tool captures both actions (which can be regarded as CQI goals) and action plans (eg, addressing gaps in workflows and processes) [64]. Although we can submit the system-supported CQI goals, it did not support incident reporting goals. Regarding the former, one of the assessors stated that the action plan tab in the system allows a free-form type (such as textboxes that allow the user to type in anything without restrictions). However, it "could be improved by adding prompts for SMART [plans] to guide the user to complete [them] correctly. These action plans are incident specific. They do not allow overall SMART goals around frequency and quality of reporting. [Although], [t]here are dashboards of measurements, they do not include goals or thresholds as a comparison or guide."

Self-Monitoring

Self-monitoring goes hand in hand with goal setting in most implementations [78,79]. In other words, users should be able to visualize their progress toward the realization of their set goals. Self-monitoring is one of the cornerstones of persuasive systems [78] and one of the most requested persuasive features in health apps such as fitness apps [45]. In a systematic review, Matthews et al [80] found that 70% of the included articles evaluated physical activity apps that supported self-monitoring as a persuasive feature to motivate behavior change. Self-monitoring fosters self-reflection and raises users' consciousness of their responsibilities, which culminates in self-regulation and behavior change [78,81,82]. Self-monitoring can be compared with holding a mirror up to the user's face, and if the user does not like what they see, they do something about it. In work environments, employees' engagement in self-monitoring is considered a prerequisite for professional development [82]. In the studied system, self-monitoring is implemented in the form of incident and near-miss reports at the pharmacy, province, or national level. In the data warehouse interface, users can view the number of cases (incidents and near misses); number of events by harm levels, top 5 drugs; and what, why, and when they happened. However, because there is no goal setting for incident report, the system does not support the type of self-monitoring that allows the user to track their progress after setting a report-based goal or when submitting a report, for example, through the display of a progress bar. In addition, the system does not allow the user to track the levels of usefulness of their reports (eg, incident, near miss, and CQI plans) to others. For example, it does not allow other users or colleagues to "like" the user's anonymous reports or to indicate their usefulness.

Tailoring, Personalization, and Customization

All 3 persuasive strategies are related and can be defined as the act of tailoring the user interface elements and content of a system to suit the user's needs, preferences, designation, or role. Tailoring and personalization are carried out by the system, whereas customization is carried out by the user. Although tailoring is enacted by the system based on users' predetermined information (eg, gathered through surveys before using the system), personalization is enacted by the system using information gathered in real time (ie, during user interaction with the system) [81,83]. We observed that tailoring was implemented in the assessed system. This system provides role-based access to certain features. However, the assessors remarked that the tailoring feature can be improved depending on what users need. However, we found that the system does not support personalization. Hence, we recommend MIRLS be personalized based on user interaction, for example, letting the user know where they left off or reminding them about incomplete tasks when they log in [7,66]. In addition, we recommend that users be allowed to customize the system (eg, user profile, chart, content, information, and reminder) to suit their needs and preferences [66].

Simulation

Simulation is a persuasive strategy used to demonstrate the cause and effect of a given behavior. Although the assessors considered it important, it was not currently implemented in the system. Thus, we recommend that MIRLS provide a means for the user to observe a link between the cause and effect of incident and near-miss reporting [80,84]. A typical implementation of the strategy is demonstrating to the user using a graph or chart that the higher the MIs reported using the system, the lower the number of recurrences.

Rehearsal

Rehearsal is a trial performance or practice of a given task so that the user can perform it correctly and easily later. In the assessment, we found that the system already provides a new user with video tutorials (organized in modules) on how to report a MI or near miss. In addition, we recommend that MIRLS provide a new user with a practice environment, in which they can rehearse before using the system to make an actual report.

Dialogue Support Assessment and Guidelines

In the dialogue support category (Table 3), all the persuasive strategies were considered important or useful; however, only 50% (4/8) of them (eg, reminder, feedback, and suggestion) were considered partially implemented by the system.

Praise and Reward

They entail acknowledging, appreciating, and recognizing the user for their effort and time taken to report incidents and near misses for the benefits of other pharmacists and patient safety. As Holden et al [6] noted, “reward and punishment structures may affect individual reporting decisions (e.g. if nurses are rewarded more for productivity than for reporting), as may culture (e.g. blame vs. just culture).” It is yet to be seen how the web-based rewards implemented in a system may influence error reporting. Enacted through well-worded motivational text and well-designed motivational images, symbols, and sounds [67], praise fosters an intimate relationship between the user and the system, making the user feel valued, appreciated, and more open to persuasion [85]. Although considered important by 2 of the assessors, 1 of them had some reservation. The participant stated, “This would emphatically not be wanted. Reward messages coming from an MIRLS technology should not emulate a sports watch. As an advanced user of the system I would find this annoying and a waste of time. If the system helps reduce incidents, a trend report shows proof, that is praise enough.” However, praise and rewards can be targeted to aggregated reports (eg, a pharmacy) on the basis of the number of incidents that reached and did not reach the patient.

Suggestion

This strategy is considered important and partially implemented in the system and can be used as a means of informing users about certain important reports (especially from other anonymous users or generated from the system), which may be useful to them in their practice. A typical suggestion in this context could be a list of “actions to take” for a specific MI or a list of “high-alert” medications that require extra precautions. Other suggestions include new research reports that may be interesting and beneficial to the user or ways, processes, or methods through which other anonymous users in the community prevent or address recurrence of certain medication errors [65]. For example, upon completing a report, the user can be recommended a set of preventive guidelines by the system to mitigate future incidents.

Feedback

Several behavior change theories such as social cognitive theory, goal setting theory [86], and feedback intervention theory consider the provision of feedback as an important ingredient in behavior change [87,88]. An example implementation of the self-monitoring-type of feedback is providing the user with summary feedback on their progress toward reaching their goal (eg, “You have achieved 30% of your goal”). Moreover, feedback entails information about one’s behavior or system-generated figures and statistics. In the context of MIRLS, informational feedback is the information of the user about the impact of their error-reporting behavior on the community or health providers’ medication errors on patient safety. An example of informational feedback is informing pharmacists about the usefulness of their reports to other users in the community (eg, “5% of the system users in the province [nation] found your report helpful”). Another example is providing users with monthly, quarterly, or yearly summary feedback highlighting the most recurring types of errors relevant to their work [65] (eg, “Poor drug naming caused 5% of the near misses last year.” In addition, the solution to this medication error can be included in the feedback message as well; for example, “Poor drug naming caused 5% of the near misses last year; remember to use TALLman lettering when necessary.” The use of uppercase letters in a portion of a drug name helps to draw attention to the dissimilarities between look-alike and sound-alike drug names. Moreover, it helps to alert health care professionals that the name of a given drug can be confused with another drug that has a similar name [89].

Reminder

This refers to an alert on task completion and compliance with certain behavior or expectation [90]. Reminder is closely tied to goal setting in a certain regard. For example, if the user sets a goal (eg, report at least X errors per month), then the user should have the opportunity to set reminders so that they could be reminded at certain preset times to report incidents or near misses if they have any. Reminder has been widely and successfully used in persuasive systems, especially in the health domain, to motivate behavior change [80,91]. In MIRLS, reminders, considered important and partially implemented, can be based on users' self-set goals on medication error reporting as well as CQI-based action plans. For example, based on self-set goals, the system can remind the user at preset times about the need to report near misses and incidents when they occur and about the benefits of the reports to other users in the community and patient safety. For instance, the system can prompt the user at a preset time with a message such as, "Did you have any near misses today or in the last one week? Please report if you did." Moreover, the system can remind the user through this type of message if the user has not logged into it or submitted a report within a certain period. In addition to this reminder-based messages, a direct link to a reporting wizard can be included, allowing users to easily submit a report by simply clicking on the provided link. Persuasive reminders have been widely used in health self-management such as taking one's daily medication and have been effective [92]. Although reminders may be more effective if they are just-in-time [87], in the context of MIRLS, they can be well ahead of time, for example, during the period when a user such as a pharmacist resumes their shift. They can also be at the end of the pharmacist's shift. Therefore, research, in the context of MIRLS, is required to show which of the periods (start or end) is more likely to be effective in motivating reporting of medication errors. In summary, reminders can be general or specific. General reminders are aimed to remind users from time to time to report incidents if they have any. Moreover, specific reminders are aimed to remind users to complete incident report drafts (ie, reports that they started but have not completed). Nevertheless, reminders should be used with caution as they can be overwhelming if overused. As stated by 1 of the assessors, "Reminders can also be annoying to the point of reminder fatigue and disregarded instantly, and overkill for this type of solution." Therefore, users should be allowed to turn them on and off.

Verbal Persuasion

This refers to the act of mentoring and providing encouragement and feedback to help individuals achieve their goals. It is also defined as "the act of telling or convincing a person to perform a task or action to change a behavior or put into action a set of events to achieve an objective" [93]. Research shows that organizational and leadership coaches use verbal persuasion effectively to increase the self-efficacy of their clients and the results they create. The tools for carrying out verbal persuasion include praise (kind words about the user), encouragement (words of affirmation about the user's ability), stories (personal or allegorical stories to help reframe the user's struggle with the task), positive feedback (assessing the user's performance favorably), strengths focus (intentionally linking the task to the user's strengths), and past achievements (acknowledging past wins as an indication of the user's ability to complete the current task) [94]. In the context of MIRLS, praise and encouragement may be used effectively by community pharmacy managers and supervisors to motivate users to report near misses and incidents. However, the use of individual feedback and past achievements may not be possible in MIRLS if, at the pharmacy level, managers and supervisors do not have access to individual users' performance owing to anonymity. In the event that managers had access to individual users' performance, as may be the case in certain pharmacies owing to corporate policy, managers and supervisors could enact verbal persuasion through personal feedback and strengths in addition to praise and encouragement. Although verbal persuasion can be said to be related to the praise and emotional appeal strategies, the main difference is that verbal persuasion is coming directly from a superior (eg, a pharmacy manager) that the user knows rather than the system. A typical message a pharmacy manager can send to an employee to verbally persuade them is, "Alice, remember to report your near misses and incidents to improve patient safety. Yes, you can*!*" Moreover, a typical feedback message from a pharmacy manager is, "Alice, thanks for your constant reporting of near misses—keep it up!" Users (whether reporting frequently or not) may find this type of message motivational. This may motivate users who have not been reporting their errors using the system in recent times to start reporting. Moreover, this type of positive feedback will help address one of the administrative barriers presented in Textbox 1: "Underreporting due to lack of useful feedback or negative feedback from administrative teams such as pharmacy managers" [8].

Emotional Appeal

It is a persuasive strategy designed to elicit an emotional response based on feelings [95]. We argue that motivational messages that appeal to emotion and feeling, such as "To err is human, to share is divine" [69], have the potential to motivate users in the medication error–reporting domain, similar to other domains [81]. In the fitness app domain, for example, Oyibo [96] found that, regardless of gender, health messages that appeal to emotion, such as "Those who do not find time for exercise will have to find time for illness," have the potential to motivate people to start or continue exercising. However, in this study, we found that although a motivational message such as "To err is human, to share is divine" may motivate some pharmacists, as evident in 1 of the assessors' responses ("would love it"), it may demotivate others. One of the assessors commented that the use of emotional appeal is inappropriate in a professional domain such as community pharmacy. The assessor stated, "It is a regulatory requirement to report incidents–no need for motivational messages...like a sports watch or fitbit. It seems unprofessional for a tool such as this to have this. I would NEVER accept this or turn this feature on." The mixed reactions to the use of emotional appeal to motivate incident reporting, similar to praise and reward, require further empirical studies.

Liking

This entails making a system visually attractive and engaging to make it persuasive. This strategy in the PSD taxonomy is drawn from the 6 principles of persuasion proposed by Cialdini [73]. According to Cialdini [73], the more people like someone, the more likely they are to be persuaded by the person. Similarly, in the context of PSD, the more esthetic a system is, the more persuasive users find it and the more likely the users are willing to use it to motivate their behavior change [48,97]. In the context of MIRLS, designers can use visually pleasing user interfaces and appropriate colors to present charts, content, and important information to improve the overall UX.

Social Support Assessment and Guidelines

Overview

In the social support category (Table 3), we only found that social comparison (in the form of benchmarking) was already implemented in the system for a limited number of measurements. However, the user had to filter each time to be able to benchmark the measure of interest (eg, near miss) at one level (eg, in the pharmacy) against another (eg, in the province). The assessors of the system suggested that rather than filtering all the time, it would be better if the benchmarking feature of the system could be enhanced by locking in the error reports—having them appear automatically. Moreover, we recommend guidelines on how to integrate other socially oriented persuasive strategies such as social learning, social facilitation, normative influence, competition, and social recognition. Holden and Karsh [7] found that social influence at the individual, group, organizational, and industry levels has the potential to influence medication error reporting.

Social Learning

This social strategy allows users to observe and imitate the behaviors and achievements of other (anonymous) users of the system [98]. The social learning strategy derives from the social learning theory proposed by Bandura [99]. The social learning theory states that people have the ability to imitate new behavior by coding or storing the ideas about the behavior in their memory, which eventually guide the actual performance of the behavior [100]. In the context of persuasive technology, social learning is simply implemented using the information of the target user about a target behavior performed by other users, for example, through a notification. In the context of MIRLS, a potential approach to implementing social learning is by enabling users to receive notifications (eg, via email) when fellow users in their group submit incident reports. These notifications would contain essential key points from the submitted reports. A typical notification message to this effect is "John [a pseudonym], here's a new report we think you'll be interested in." We believe that messages such as this, which enable one user to learn from others' reports, may motivate the target user to submit their reports given the benefit they derive from them. Given that users may be overwhelmed, they should be given the opportunity to determine the types of messages they wish to receive, the number within a given period such as a week or month, and even opt out completely by turning the feature off. More importantly, owing to privacy concerns, particularly within a facility setting, instead of basing the social learning strategy on key points from reported near misses or incidents, it can be based on the quantity of reports submitted within a specified period (refer to the *Normative Influence* section). According to 1 of the assessors, "I don't think this [first Social Learning implementation] is appropriate if you can see who it is but if it is just numbers it would be useful. [N]otification within a facility could hamper the feeling of safe reporting because anonymity is compromised." A second implementation of social learning is the provision of a news feed that highlights important reports submitted by other anonymous users that the user may find useful. A third implementation is the support of chat rooms or discussion rooms where users can discuss near misses, incidents and lessons learned; share experiences and knowledge; and learn from one another in an anonymous fashion. The chat room and discussion forum feature may be extended and beneficial to nonpharmacists, as evident in 1 of the assessors' comments, "Our users may find this useful. If they have the time, which currently they don't have much of during the pandemic."

Social Comparison

Social comparison allows users to compare their performance with that of others. It is derived from the social comparison theory proposed by Festinger [101], which centers on the belief that individuals have an inner drive to gain accurate self-evaluations through social comparison. It holds that by comparing one's abilities and performances with those of similar others or peers, the individual is able to reduce uncertainty, learn, and improve self. This strategy has been used successfully in persuasive systems [102]. In the assessment of the Think Research or Pharmapod system, we found that social comparison was implemented at the pharmacy and provincial level in the form of benchmarking reports, tables, and dashboards. For example, 1 of the assessors responded thus, "within our own organization we may compare pharmacies with other pharmacies or between provinces of our pharmacies using reports provided." Thus, the implementation of social comparison in the system can be improved. For example, users' error reporting over a particular period, for example, week, month, or year, can be compared anonymously with the average at the pharmacy or provincial level using a bullet chart infographic.

Competition

Similar to social comparison, competition allows users to compare themselves with others, for example, in terms of number of reports, frequency, quality, or usefulness of reports to others. Competition leverages the natural drive of humans to outperform one another [98]. Research on persuasive technology shows that competition, regardless of gender, age, and culture, has the potential to motivate users to perform the target behavior [103]. In the fitness app domain, for example, Oyibo and Vassileva [98] found a significant relationship among social comparison, social learning, and competition, indicating that the more people compare themselves, the more they learn about the performance or achievements of others and the more competitive they become in their behaviors. In the context of MIRLS, users can be allowed to view where they are compared with other anonymous users in small sets (eg, on a leaderboard). The criterion for placement on the leaderboard can include the total number of reports, frequency, quality, or usefulness of the report to others (eg, over a weekly, monthly, or yearly period). The small sets of anonymous users can be drawn from the pool of users at the provincial or national level, which can change from time to time because of the need to foster anonymity. Moreover, the competition feature can be group based, involving anonymous pharmacies, organizations, or provinces. As 1 of the assessors remarked, "Perhaps [my organization] may wish to see how many incidents they are experiencing compared to another organization of the same industry channel and size."

Cooperation

Unlike competition, where users compete to outperform one another, in cooperation, users work together in a collaborative fashion to achieve their individual and collective goals. In the assessment of the Think Research or Pharmapod IM system, we found that providing users the choice of being paired with another (anonymous) user, with the goal of motivating one another to achieve individual or collective goals may not be a good idea. This is based on the premise that the implementation of cooperation in MIRLS may compromise the principle of anonymity of users, upon which MI reporting is founded. Hence, we recommend that cooperation be implemented and used with caution if MIRLS were to support it in a given pharmacy. As commented by 1 of the assessors, "Why would anyone wish to be compared to [cooperate with] another user? Where's the privacy aspect of such a feature?"

Normative Influence

Unlike informational influence, which is conformity to a certain behavior based on the acceptance of evidence about reality provided by others, normative influence is conformity based on an individual's desire to fulfill others' expectations to gain acceptance, fit in, or feel a sense of belonging [104]. In the context of reporting medication errors, the urge for individual users to report near misses and incidents might arise from perceived social pressure rather than actual pressure, considering that the submitted reports are anonymous or deidentified. Thus, a possible way of realizing the normative influence strategy in MIRLS is allowing the user to know about the number of other anonymous users in the facility, province, or nation that are reporting medication errors at a given time. For example, in COVID-19 contact tracing apps, Oyibo and Morita [105] found that socially oriented messages, such as "112 other people reported their COVID-19 diagnosis today," have the potential to motivate app users to report their diagnosis by entering their one-time key into the app. Hence, we recommend that the system informs users at suitable intervals (eg, when they are logged on) about the quantity of other anonymous users within the pharmacy, province, or country who are reporting medication errors within a specific period. A message similar to the message by Oyibo and Morita [105], "10 other people submitted their incident reports today," may be used to normatively influence users to submit their own incident reports as well if they have any pending or have not yet submitted.

Social Facilitation

Social facilitation refers to the improvement in a person's performance as a result of the real, imagined, or implied presence of others. As stated in the study by Mohadis et al [84], "System users are more likely to perform a targeted [behavior] if they discern, via the system, that others are performing the [behavior] along with them." In MIRLS, one way to realize social facilitation is to inform the user when they log on to the system (eg, to make a report) through news feed that they are not alone in their efforts to report an error, as other users elsewhere (eg, in the facility, province, or nation) at the current time are also attempting to making a report or logged on to the system. Motivational messages such as "You are not alone; X others are on the system at the moment submitting a report" could be used to make the user feel the presence of other anonymous users whenever the former is logged into the system. A message such as this may encourage users, who have begun the process of submitting a report, to complete it. This type of message is similar to that which customers get when they are booking a hotel or shopping for a flight ticket on the web (eg, "5 other people are currently shopping for this flight ticket"). Although this type of message is commonly used in the e-commerce domain to create the impression that the user may miss procuring a given flight ticket if they do not act quickly (ie, buy it now), in the domain of medication error reporting, this is not the case. Rather, this type of message is used to let the user know that they are not alone—that there are similar others elsewhere who are trying to do the same task as them (submit a MI report).

Social Recognition

In social psychology, social recognition is the act of recognizing people such as employees for great work, contribution, and achievement by acknowledging them publicly. One possible way of implementing this strategy in an MIRLS is recognizing users for being one of the "most valuable players" of the month, quarter, or year. This can be at the facility, provincial, or national level. The criteria for recognition include the number, frequency, quality, or usefulness of the target user's reports to the community. Although research shows that employees welcome social recognition in the workplace [84], it must be implemented with caution given the anonymity requirement aimed to protect users from punitive measures. We found that users may not welcome the second feature ("allowing other users to rate a user's report anonymously based on how useful or helpful it is to them") as they perceived it as a form of competition. For example, 1 of the assessors commented, "Rating makes this feel like a competition or to call out that can produce negative attitudes. Not helpful. Those entering data into a system may not be the same person who is involved in the incident." Moreover, the user was also concerned about the part of the report being rated as well as privacy and anonymity, "What part of the report is being rated in this scenario?" It is worth noting that we conceived the social recognition rating feature similar to Google Play Store app rating system, in which users can rate an app on a 5-star scale. Although we did not explicitly detail the section of the report being anonymously rated by other users in the study, we intended it to encompass essential elements derived from the report analysis, such as the description of the near miss or incident, the lessons learned by the reporters, and possible recommendations and tips to prevent future recurrence. These key points may have been extracted from a set of similar aggregated reports submitted by different anonymous users at different times and included in the MI analysis report shared with users via the MIRLS by standard bodies such as Assurance and Improvement in Medication Safety (AIMS) [106]. AIMS is a standardized medication safety program that supports CQIs and sets a mandatory consistent standard for medication safety for all pharmacies in Ontario. Its goal is to minimize the risk of harm to patients caused by MIs in the province. Part of its mandate is to aggregate and analyze anonymous MI reports and produce and disseminate the results to stakeholders. This enables practitioners to learn from MIs and have a better understanding of why they occur and how they can be prevented in the future [106]. Although in this study, we did not find the second social recognition feature to be useful to the assessors, there may be a need for a more comprehensive study in future research among a larger audience of community pharmacists to uncover its potential to motivate users to report medication errors more frequently.

System Credibility Support Assessment and Guidelines

Regarding the credibility support category (Table 4), the assessors reported that the system fully or partially supported a number of credibility-related persuasive strategies such as trustworthiness, credibility, expertise, and real-world feel. We discuss all these strategies together with the other 3 strategies in the credibility support category.

Authority

One of the principles of persuasion proposed by Cialdini [73], the authority principle, states that people are more likely to believe and obey those who are in positions of authority. Selassie et al [76] found that frontline staff working with children with autism (supported by a data entry management system) can be persuaded by the authority strategy. Moreover, in the study by Mohadis and Ali [84] on user perception of a physical activity app for older workers, 1 of the participants remarked, "Yeah, incorporating an expert [authority figure's] view is very important so that we become more confident with whatever recommendations that the system offers." In the context of community pharmacy, authority figures and bodies may include researchers, pharmacy managers, and professional bodies such as the Institute for Safe Medication Practices Canada [1]. Thus, we recommend the presentation of authority-based information and messages to users, for example, on the value of reporting medication errors and the benefits it can have for the profession, staff, and patient safety [47].

Third-Party Endorsement

Third-party endorsement is the act of publicly approving or supporting a product, system, or service by a reputable socially influential individual or organization other than the staff or company that owns it. Usually, the third party may have seen, interacted, and used the product, system, or service in question and is satisfied with the results, utility, or experience. In the business world, research has shown that the third-party endorsements have the potential to effectively earn companies the trust and loyalty of customers [21,69]. Moreover, research shows that the expertise and trustworthiness of a third-party organization endorsement have the potential to positively affect the perceived value of a firm, which in turn can positively affect customer loyalty [107]. Hence, to encourage pharmacists to use MIRLS, the designers should demonstrate that the system is approved or endorsed by authoritative bodies such as professional organizations (eg, World Health Organization and Institute for Safe Medication Practices), regulatory bodies, and government. To implement this persuasive strategy in MIRLS, one approach is to incorporate the corporate logos of the endorsing authoritative bodies within the user interface, such as on the system's home page or in the footer, especially if it is a web-based application.

Expertise, Surface Credibility, and Trustworthiness

Research has shown that all 3 strategies are related. For example, Fogg and Tseng [108] postulated that credibility, a perceived quality of a system, comprises 2 key components: trustworthiness and expertise. In other words, a system is perceived to be credible if its perceived trustworthiness and perceived expertise are high. Trustworthiness is a key element in the credibility perception of systems such as websites. It is defined by terms such as well intentioned, truthful, and unbiased [109]. As stated in the study by Fogg et al [109], "the trustworthiness dimension of credibility captures the perceived goodness or morality of the source." Similarly, expertise is a key element in the credibility perception of systems such as websites. It is defined by terms such as knowledgeable, experienced, competent, and professional [109]. As stated in the study by Fogg et al [109], "[t]he expertise dimension of credibility captures the perceived knowledge and skill of the source." In a large-scale website credibility study conducted by Fogg et al [109], the authors found that perceived expertise and perceived trustworthiness have a significant impact on the perceived credibility of websites. In the context of MIRLS, to realize expertise, the visual and functional design of the system should reflect professionalism, expertise, and up-to-dateness to motivate users to use it. Moreover, to implement trustworthiness, the system should foster user anonymity, data deidentification, and data aggregation and live up to promises such as it not being used as a punitive tool to hold users accountable [65]. Finally, perceived credibility can be intentionally built into the system through its visual design, for example, by ensuring users enter accurate information using taxonomy-based option buttons, checklists, and drop-downs and reducing advertisements for a web-based system [47]. In our study, all 3 assessors agreed that perceived expertise is important or useful as well as implemented to a great extent in the system they were currently using. For example, 1 of the assessors commented, "The MIRLS is very easy to use and intuitive, and requires minimal training to get started." However, "there is always room for improvement," remarked another assessor. Failure to foster expertise in the system design may discourage frequent use and completion of tasks, as evident in the assessor's comment, "Performance in speed is always a challenge and [the] latency [experienced in some] areas drive users to drop off or stop using." Regarding trustworthiness, 2 assessors considered it important or useful. However, only 1 assessor considered it to be implemented in the current system. This is partly because of anonymity not being completely fostered in the system. This is evident in 1 of the assessors' comments, "anonymity is fostered outside an organization (eg. when data sent to AIMS) and there is also a choice to report anonymously so the corporate level of an organization does not have visibility. [W]ithin a location the reports are not anonymous." Finally, regarding surface credibility, 2 assessors considered it important or useful and implemented it in their current system. For example, all 3 assessors responded that there were no advertisements in the system and that was very important.

Verifiability

This refers to "the quality or state of being capable of being verified, confirmed, or substantiated" [110]. In the context of

MIRLS, persuasive messages (eg, on the value of error reporting to patient safety) aimed at motivating users should not only be credible but also verifiable. As stated in the study by Jones [111], carefully choosing persuasive messages and supporting materials that are verifiable, specific, and unbiased can be helpful in appealing to logic and increasing users' trust. Verifiability was implemented in WargaFit (a fitness app prototype aimed to encourage simple exercise such as body stretching in an office environment) by the provision of healthy tips accompanied with external links [84]. Similarly, verifiability in MIRLS can be realized through the provision of the source of information or inclusion of the URL in the persuasive message such as "Reporting reduces the number of future errors, diminishing personal suffering and decreasing financial costs" [112]. In our study, 2 assessors considered verifiability useful and not currently implemented in their system. For example, regarding harm levels, 1 of the assessors commented, "There are info points that explain [that] harm level comes from WHO but there is no link to the WHO to verify it."

Real-World Feel

Similar to expertise and trustworthiness, real-world feel is found to positively influence the perceived credibility of websites [109]. Real-world feel is the interaction with and experience of a virtual or electronic product, system, or service as though it is real. This is made possible by the product-, system-, or service-supporting features that mimic and foster real-world interaction and experience. In the case of e-commerce websites, for example, the real-world feel can be fostered by providing contact phone number, contact email address, and a quick response to customer service questions; listing the physical address of the organization behind the website; and showing photos of the members of the organization [109]. In the context of MIRLS, in addition to the aforementioned features, the system should be designed as close as possible to the nonelectronic (paper) version. This has the potential to reduce the cognitive effort required by a new user to make the transition. In the assessment of the Think Research or Pharmapod IM system, assessors stated that it supports real-world feel by mimicking the paper version and allows clients to customize their own forms and notifications or escalations. One way the system designers achieved real-world feel is to allow pharmacies and organizations to customize their MI report forms.

Persuasive System Implementation and Ethical Design Considerations

Our analysis reveals that there is a need to consider and address the ethical implications that may arise from integrating persuasive strategies into the existing MIRLS. These considerations include administrative (eg, anonymity) and choice of persuasive strategies (eg, monetary reward). For example, to ensure that the principle of anonymity is fostered in the implementation of social strategies, user identifications should be limited to pseudonyms, which the users can change from time to time. It is worth noting that a persuasive strategy that may be effective (or welcomed) in one community pharmacy may not be in another. Hence, there may be a need to get the potential users involved in deciding the set of persuasive strategies that will be implemented or effective in a given pharmacy. Thus, the system should offer tailoring capabilities that support the chosen guidelines. Intervention researchers and designers may have to (1) investigate, before implementation, which of the recommended persuasive strategies a given group of pharmacy professionals may be or may not be receptive to and (2) implement only the set of strategies that are likely to be effective, as proven by empirical evidence. For these reasons, MIRLS should be designed in a way that enables pharmacies to turn on and off persuasive strategies that they consider useful and nonuseful, respectively. It is worth noting that some of the persuasive strategies in the PSD taxonomy may have to be combined to realize a holistic and functional persuasive feature that is useful. In other words, some of the persuasive strategies are complementary. For example, praise and feedback strategies must be combined to implement or realize a composite feature that provides immediate feedback of praise to the user upon submitting an incident report. In addition, reminders and verbal persuasion may be combined to realize a composite feature that verbally persuades the user through a reminder. For example, a verbal persuasion message ("Alice, remember to report your near misses and incidents to improve patient safety. Yes, you can*!*") can be sent to the user as a reminder by the pharmacy manager from time to time. Finally, authority, credibility, and verifiability may have to be combined to realize a persuasive message that is not only authoritative and credible but also verifiable.

Contributions

In this study, we have made a number of contributions to knowledge in the domain of community pharmacy and developers of health digital systems. This study is the first to provide guidelines on how to integrate persuasive strategies into MIRLS to increase their utility and motivate users to report MIs and near misses to improve patient safety and promote shared learning. Specifically, we provided MIRLS-specific persuasive design guidelines based on the PSD taxonomy proposed by Oinas-Kukkonen and Harjumaa [30]. Most of the PSD guidelines in the extant literature are concentrated in the domains of healthy eating [113] and physical activity [81,84,114]. Designers of MIRLS can leverage the current set of PSD guidelines in improving future iterations not only in community pharmacy but also in other settings where incident or error reporting is essential and part of the organizational practice. The second contribution is that this study lays the foundation for future empirical research aimed at investigating the effectiveness of persuasive strategies incorporated into MIRLS. Future research efforts should focus on ≥1 of the design guidelines in each of the 4 categories of the PSD taxonomy; implement them; and conduct a field study to examine the perception, acceptance, and adoption of the implemented strategies by the target community pharmacists.

Research Directions

In future work, we look forward to investigating the potential effectiveness of some of the proposed persuasive design guidelines presented in Tables 2-4 and Textbox 4 in field studies. First, we will create prototypes of the persuasive strategies and perform an empirical study to explore which set of strategies might be more effective. In addition, we will analyze the potential influence of demographic variables, such as age, gender, and work experience, on the effectiveness of these strategies. Second, we will select the most persuasive strategies that the target community of pharmacy professionals are most responsive to and implement them in an actual MIRLS (eg, Think Research or Pharmapod). Third, we will conduct a field study (randomized controlled trial) to investigate the effect of the persuasive design on the rate of MI reporting among community pharmacy professionals using different provinces across Canada as case studies. More importantly, owing to the lack of studies on the relationship between system usability and medication error reporting, as our second scoping review shows, we recommend that future work be conducted in this area.

Limitations

Similar to most conceptual papers, our study has limitations owing to its preliminary nature, which stems from the nonmaturity of research on the persuasive design of MIRLS. The first limitation is that the results of the scoping reviews might have been limited one way or the other by the choice of search strings and the subjective assessment, understanding, and interpretations of the extracted data by the researchers that conducted the reviews. Hence, we recommend a more comprehensive review, particularly with regard to the second RQ, in which a formal review led to no included article, other than the article retrieved from Google Scholar search. The second limitation of our study is the convenience sample. In other words, the 3 assessors who assessed the Think Research or Pharmapod system using the PSD taxonomy were not sufficient to be representative of the entire population of community pharmacy professionals using MIRLS across Canada. For example, a persuasive feature that may be important and useful to a group of community pharmacists in one facility may not be useful to another group in another facility. Hence, the findings reported in the last 2 columns of Tables 2-4 and Textbox 4 may not generalize to a larger population sample involving a heterogeneous group of community pharmacists with different roles, working environments, years of working experience, professional qualifications, gender, personality, and economic status, which may influence their responses. In future work, we hope to build on this preliminary study by conducting a formal research (eg, based on storyboards) involving a larger population sample to validate the generalizability of the findings of this study, particularly the effectiveness, acceptability, and adoption of the recommended persuasive strategies presented in Tables 2-4 and Textbox 4.

Conclusions

Although most medical practitioners agree that reporting medication errors improves the quality of care and safety for patients [21], in reality, the rate of reporting remains below expectations [115] owing to lack of motivation and other barriers [22-25]. In this study, we argued that although most current MIRLS have implemented recommended guidelines bordering on favorable administrative measures and utility, they lack motivational affordances that can facilitate or motivate frequent reporting. Hence, using the Think Research or Pharmapod system as a case study, we identified opportunities for incorporating persuasive strategies into MIRLS to make them more effective in motivating behavior change. The proposed persuasive design guidelines can be used by designers and developers in making MIRLS more effective in motivating users to report incidents and near misses more often to reduce risks of recurrence, improve patient safety, and foster shared learning among community pharmacy professionals and stakeholders. However, before the implementation of the recommended persuasive design guidelines in Tables 2-4 and Textbox 4, there is a need for thorough consideration and evaluation of the various ramifications, including administrative, regulatory, and ethical implications. The presented persuasive design guidelines open up new opportunities for persuasive design research in MI reporting. We acknowledge that some of the proposed persuasive strategies may not be suitable or effective in real-life settings. Hence, there is a need for further validation-based research and caution regarding their implementation. In future work, we aim to validate the suitability and effectiveness of the proposed persuasive strategies in motivating behavior change using storyboards, prototypes, and perception and evaluation studies involving community pharmacists across Canada.

Acknowledgments

This project was funded by the Social Sciences and Humanities Research Council Partnership Engage Grant (#1008-2020-0119).

Authors' Contributions

KO and PAG coordinated and supervised the assessment of the Think Research or Pharmapod platform and led the substantive writing of the paper. KO, SE, and TN conducted the scoping reviews including database search, retrieval of articles, screening and selection of included articles, and extraction and tabulation of data. CB facilitated access to the platform, contact with the assessors, and reviewed and edited the paper. DO provided technical assistance in the platform access and assessment and reviewed and edited the paper. KO used the data from the scoping reviews and assessment of the Think Research or Pharmapod platform to write the paper. PAG and JRB contributed to the editing of the paper.

Conflicts of Interest

CB is the vice president of Quality Improvement and Innovations at Think Research or Pharmapod, which is the partner organization for the Partnership Engage Grant that funded this research. All other authors declare no other conflicts of interest.

References

1. Deaths associated with medication incidents: learning from collaborative work with provincial offices of the chief coroner and chief medical examiner. Institute for Safe Medication Practices Canada. 2013 Aug 28. URL: https://tinyurl.com/fycmya97 [accessed 2024-01-11]
2. Alqenae FA, Steinke D, Keers RN. Prevalence and nature of medication errors and medication-related harm following discharge from hospital to community settings: a systematic review. Drug Saf 2020 Jun;43(6):517-537 [FREE Full text] [doi: 10.1007/s40264-020-00918-3] [Medline: 32125666]
3. Desjardins L. Thousands die from medical errors yearly, notes advocacy group. Radio Canada International. 2019 Oct 28. URL: https://tinyurl.com/bdcepsmu [accessed 2023-08-20]
4. Tariq RA, Vashisht R, Sinha A, Scherbak Y. Medication dispensing errors and prevention. In: StatPearls Internet. Treasure Island, FL: StatPearls Publishing; 2013.
5. Mrayyan MT, Shishani K, Al-Faouri I. Rate, causes and reporting of medication errors in Jordan: nurses' perspectives. J Nurs Manag 2007 Sep;15(6):659-670. [doi: 10.1111/j.1365-2834.2007.00724.x] [Medline: 17688572]
6. Kim KS, Kwon SH, Kim JA, Cho S. Nurses' perceptions of medication errors and their contributing factors in South Korea. J Nurs Manag 2011 Apr;19(3):346-353. [doi: 10.1111/j.1365-2834.2011.01249.x] [Medline: 21507105]
7. Holden RJ, Karsh BT. A review of medical error reporting system design considerations and a proposed cross-level systems research framework. Hum Factors 2007 Apr;49(2):257-276. [doi: 10.1518/001872007X312487] [Medline: 17447667]
8. Aljabari S, Kadhim Z. Common barriers to reporting medical errors. ScientificWorldJournal 2021 Jun 10;2021:6494889 [FREE Full text] [doi: 10.1155/2021/6494889] [Medline: 34220366]
9. Bahadori M, Ravangard R, Aghili A, Sadeghifar J, Gharsi Manshadi M, Smaeilnejad J. The factors affecting the refusal of reporting on medication errors from the nurses' viewpoints: a case study in a hospital in Iran. ISRN Nurs 2013 Apr 09;2013:876563 [FREE Full text] [doi: 10.1155/2013/876563] [Medline: 23691354]
10. Yung HP, Yu S, Chu C, Hou IC, Tang FI. Nurses' attitudes and perceived barriers to the reporting of medication administration errors. J Nurs Manag 2016 Jul;24(5):580-588. [doi: 10.1111/jonm.12360] [Medline: 26888342]
11. Higuchi A, Higami Y, Takahama M, Yamakawa M, Makimoto K. Potential underreporting of medication errors in a psychiatric general hospital in Japan. Int J Nurs Pract 2015 May;21 Suppl 2:2-8. [doi: 10.1111/ijn.12169] [Medline: 26125569]
12. Alqenae FA, Steinke D, Carson-Stevens A, Keers RN. Analysis of the nature and contributory factors of medication safety incidents following hospital discharge using national reporting and learning system (NRLS) data from England and Wales: a multi-method study. Ther Adv Drug Saf 2023 Mar 16;14:20420986231154365 [FREE Full text] [doi: 10.1177/20420986231154365] [Medline: 36949766]
13. George D, Hassali MA, Hss AS. Usability testing of a mobile app to report medication errors anonymously: mixed-methods approach. JMIR Hum Factors 2018 Dec 21;5(4):e12232 [FREE Full text] [doi: 10.2196/12232] [Medline: 30578216]
14. David U. Medication error reporting systems: problems and solutions. New Med 2001 Nov;1(2):61-65.
15. Kang HJ, Park H, Oh JM, Lee EK. Perception of reporting medication errors including near-misses among Korean hospital pharmacists. Medicine (Baltimore) 2017 Sep;96(39):e7795 [FREE Full text] [doi: 10.1097/MD.0000000000007795] [Medline: 28953611]
16. West DR, Pace WD, Dickinson LM, Harris DM, Main DS, Westfall JM, et al. Relationship between patient harm and reported medical errors in primary care: a report from the ASIPS collaborative. In: Henriksen K, Battles JB, Keyes MA, Grady ML, editors. Advances in Patient Safety: New Directions and Alternative Approaches (Vol. 1: Assessment). Rockville, MD: Agency for Healthcare Research and Quality; Aug 2008.
17. Robb G, Loe E, Maharaj A, Hamblin R, Seddon ME. Medication-related patient harm in New Zealand hospitals. N Z Med J 2017 Aug 11;130(1460):21-32. [Medline: 28796769]
18. Riordan CO, Delaney T, Grimes T. Exploring discharge prescribing errors and their propagation post-discharge: an observational study. Int J Clin Pharm 2016 Oct;38(5):1172-1181. [doi: 10.1007/s11096-016-0349-7] [Medline: 27473712]
19. Panagioti M, Khan K, Keers RN, Abuzour A, Phipps D, Kontopantelis E, et al. Prevalence, severity, and nature of preventable patient harm across medical care settings: systematic review and meta-analysis. BMJ 2019 Jul 17;366:l4185 [FREE Full text] [doi: 10.1136/bmj.l4185] [Medline: 31315828]
20. Boyle TA, Mahaffey T, Mackinnon NJ, Deal H, Hallstrom LK, Morgan H. Determinants of medication incident reporting, recovery, and learning in community pharmacies: a conceptual model. Res Social Adm Pharm 2011 Mar;7(1):93-107. [doi: 10.1016/j.sapharm.2009.12.001] [Medline: 21397884]
21. Vrbnjak D, Denieffe S, O'Gorman C, Pajnkihar M. Barriers to reporting medication errors and near misses among nurses: a systematic review. Int J Nurs Stud 2016 Nov;63:162-178. [doi: 10.1016/j.ijnurstu.2016.08.019] [Medline: 27637011]
22. Elder NC, Graham D, Brandt E, Hickner J. Barriers and motivators for making error reports from family medicine offices: a report from the American academy of family physicians national research network (AAFP NRN). J Am Board Fam Med 2007;20(2):115-123 [FREE Full text] [doi: 10.3122/jabfm.2007.02.060081] [Medline: 17341747]
23. Soydemir D, Seren Intepeler S, Mert H. Barriers to medical error reporting for physicians and nurses. West J Nurs Res 2017 Oct;39(10):1348-1363. [doi: 10.1177/0193945916671934] [Medline: 27694427]

24. Mirbaha F, Shalviri G, Yazdizadeh B, Gholami K, Majdzadeh R. Perceived barriers to reporting adverse drug events in hospitals: a qualitative study using theoretical domains framework approach. Implement Sci 2015 Aug 07;10:110 [FREE Full text] [doi: 10.1186/s13012-015-0302-5] [Medline: 26250159]
25. Mousavi-Roknabadi RS, Momennasab M, Groot G, Askarian M, Marjadi B. Medical error and under-reporting causes from the viewpoints of nursing managers: a qualitative study. Int J Prev Med 2022 Jul 15;13:103 [FREE Full text] [doi: 10.4103/ijpvm.IJPVM_500_20] [Medline: 36119955]
26. ISO 9241-11:2018 ergonomics of human-system interaction part 11: usability: definitions and concepts. International Organization for Standardization. URL: https://www.iso.org/standard/63500.html [accessed 2023-08-22]
27. Pot JW, de Vries CG. Medical devices in intramural settings and patient safety: availability of incident data and possible solutions to prevent use problems. National Institute for Public Health and the Environment. 2008. URL: https://www.rivm.nl/bibliotheek/rapporten/360020005.pdf [accessed 2024-01-11]
28. St-Maurice J, Burns C, Wolting J. Applying persuasive design techniques to influence data-entry behaviors in primary care: repeated measures evaluation using statistical process control. JMIR Hum Factors 2018 Oct 11;5(4):e28 [FREE Full text] [doi: 10.2196/humanfactors.9029] [Medline: 30309836]
29. Medication safety software for healthcare professionals. Pharmapod. URL: https://pharmapodhq.com/ [accessed 2023-08-21]
30. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009;24 [FREE Full text] [doi: 10.17705/1CAIS.02428]
31. van Gemert-Pijnen L, Kelders SM, Beerlage-de Jong N, Oinas-Kukkonen H. Persuasive health technology. In: eHealth Research, Theory and Development. Milton Park, UK: Routledge; 2018.
32. Widyasari YD, Nugroho LE, Permanasari AE. Persuasive technology for enhanced learning behavior in higher education. Int J Educ Technol High Educ 2019 Apr 30;16:15. [doi: 10.1186/s41239-019-0142-5]
33. Stewart D, Thomas B, MacLure K, Wilbur K, Wilby K, Pallivalapila A, et al. Exploring facilitators and barriers to medication error reporting among healthcare professionals in Qatar using the theoretical domains framework: a mixed-methods approach. PLoS One 2018 Oct 2;13(10):e0204987 [FREE Full text] [doi: 10.1371/journal.pone.0204987] [Medline: 30278077]
34. Bullen K, Hall N, Sherwood J, Wake N, Donovan G. Prescribing error reporting in primary care: a narrative synthesis systematic review. Integr Healthc J 2020 Dec 10;2(1):e000026 [FREE Full text] [doi: 10.1136/ihj-2019-000026] [Medline: 37441317]
35. Holmström AR, Airaksinen M, Weiss M, Wuliji T, Chan XH, Laaksonen R. National and local medication error reporting systems: a survey of practices in 16 countries. J Patient Saf 2012 Dec;8(4):165-176. [doi: 10.1097/PTS.0b013e3182676cf3] [Medline: 23007239]
36. Alsaleh FM, Alsaeed S, Alsairafi ZK, Almandil NB, Naser AY, Bayoud T. Medication errors in secondary care hospitals in Kuwait: the perspectives of healthcare professionals. Front Med (Lausanne) 2021 Dec 20;8:784315 [FREE Full text] [doi: 10.3389/fmed.2021.784315] [Medline: 34988097]
37. Adams KT, Pruitt Z, Kazi S, Hettinger AZ, Howe JL, Fong A, et al. Identifying health information technology usability issues contributing to medication errors across medication process stages. J Patient Saf 2021 Dec 01;17(8):e988-e994. [doi: 10.1097/PTS.0000000000000868] [Medline: 34009868]
38. Cheung KC, van der Veen W, Bouvy ML, Wensing M, van den Bemt PM, de Smet PA. Classification of medication incidents associated with information technology. J Am Med Inform Assoc 2014 Feb;21(e1):e63-e70 [FREE Full text] [doi: 10.1136/amiajnl-2013-001818] [Medline: 24064444]
39. Marcilly R, Schiro J, Beuscart-Zéphir MC, Magrabi F. Building usability knowledge for health information technology: a usability-oriented analysis of incident reports. Appl Clin Inform 2019 May;10(3):395-408 [FREE Full text] [doi: 10.1055/s-0039-1691841] [Medline: 31189203]
40. Hautamäki E, Kinnunen UM, Palojoki S. Health information systems´ usability-related use errors in patient safety incidents. Finn J eHealth eWelfare 2017 Mar 31;9(1):6-17. [doi: 10.23996/fjhw.60763]
41. Kushniruk A, Patel V, Cimino JJ, Barrows RA. Cognitive evaluation of the user interface and vocabulary of an outpatient information system. Proc AMIA Annu Fall Symp 1996:22-26 [FREE Full text] [Medline: 8947620]
42. Ratwani RM, Savage E, Will A, Fong A, Karavite D, Muthu N, et al. Identifying electronic health record usability and safety challenges in pediatric settings. Health Aff (Millwood) 2018 Nov;37(11):1752-1759. [doi: 10.1377/hlthaff.2018.0699] [Medline: 30395517]
43. Kushniruk AW, Triola MM, Borycki EM, Stein B, Kannry JL. Technology induced error and usability: the relationship between usability problems and prescription errors when using a handheld application. Int J Med Inform 2005 Aug;74(7-8):519-526. [doi: 10.1016/j.ijmedinf.2005.01.003] [Medline: 16043081]
44. Fogg BJ. Persuasive Technology: Using Computers to Change What We Think and Do. Amsterdam, The Netherlands: Elsevier Science; 2003.
45. Oyibo K. Investigating the key persuasive features for fitness app design and extending the persuasive system design model: a qualitative approach. Proc Int Symp Hum Factors Ergon Health Care 2021 Jul 22;10(1):47-53 [FREE Full text] [doi: 10.1177/2327857921101022]

46. Bartlett YK, Webb TL, Hawley MS. Using persuasive technology to increase physical activity in people with chronic obstructive pulmonary disease by encouraging regular walking: a mixed-methods study exploring opinions and preference. J Med Internet Res 2017 Apr 20;19(4):e124 [FREE Full text] [doi: 10.2196/jmir.6616] [Medline: 28428155]
47. Win KT, Mullan J, Howard S, Oinas-Kukkonen H. Persuasive systems design features in promoting medication management for consumers. In: Proceedings of the 50th Hawaii International Conference on System Sciences. 2017 Presented at: 50th Hawaii International Conference on System Sciences; January 4-7, 2017; Hilton Waikoloa Village, Hawaii. [doi: 10.24251/hicss.2017.403]
48. Oyibo K, Vassileva J. HOMEX: persuasive technology acceptance model and the moderating effect of culture. Front Comput Sci 2020 Mar;2:1-17 [FREE Full text] [doi: 10.3389/fcomp.2020.00010]
49. TCPS 2 (2022) – Chapter 2: Scope and Approach. Government of Canada. URL: https://tinyurl.com/bdfhatrw [accessed 2023-01-25]
50. Do I need research ethics approval? University of Alberta. URL: https://tinyurl.com/mu8wkue4 [accessed 2024-01-25]
51. Lee BH, Lehmann CU, Jackson EV, Kost-Byerly S, Rothman S, Kozlowski L, et al. Assessing controlled substance prescribing errors in a pediatric teaching hospital: an analysis of the safety of analgesic prescription practice in the transition from the hospital to home. J Pain 2009 Feb;10(2):160-166 [FREE Full text] [doi: 10.1016/j.jpain.2008.08.004] [Medline: 19010736]
52. Calabrese AD, Erstad BL, Brandl K, Barletta JF, Kane SL, Sherman DS. Medication administration errors in adult patients in the ICU. Intensive Care Med 2001 Oct;27(10):1592-1598. [doi: 10.1007/s001340101065] [Medline: 11685299]
53. Cabri A, Barsegyan N, Postelnick M, Schulz L, Nguyen V, Szwak J, et al. Pharmacist intervention on prescribing errors: use of a standardized approach in the inpatient setting. Am J Health Syst Pharm 2021 Nov 23;78(23):2151-2158 [FREE Full text] [doi: 10.1093/ajhp/zxab278] [Medline: 34283219]
54. Boucher A, Ho C, MacKinnon N, Boyle TA, Bishop A, Gonzalez P, et al. Quality-related events reported by community pharmacies in Nova Scotia over a 7-year period: a descriptive analysis. CMAJ Open 2018 Dec 18;6(4):E651-E656 [FREE Full text] [doi: 10.9778/cmajo.20180090] [Medline: 30563921]
55. Stockton KR, Wickham ME, Lai S, Badke K, Villanyi D, Ho V, et al. LO51: incidence of clinically relevant medication errors after implementation of an electronic medication reconciliation process. Can J Emerg Med 2017 May 15;19(S1):S45. [doi: 10.1017/cem.2017.113]
56. Brennan-Bourdon LM, Vázquez-Alvarez AO, Gallegos-Llamas J, Koninckx-Cañada M, Marco-Garbayo JL, Huerta-Olvera SG. A study of medication errors during the prescription stage in the pediatric critical care services of a secondary-tertiary level public hospital. BMC Pediatr 2020 Dec 05;20(1):549 [FREE Full text] [doi: 10.1186/s12887-020-02442-w] [Medline: 33278900]
57. Shanks L, Bil K, Fernhout J. Learning without borders: a review of the implementation of medical error reporting in Médecins Sans Frontières. PLoS One 2015 Sep 18;10(9):e0137158 [FREE Full text] [doi: 10.1371/journal.pone.0137158] [Medline: 26381622]
58. Fyhr A, Akselsson R. Characteristics of medication errors with parenteral cytotoxic drugs. Eur J Cancer Care (Engl) 2012 Sep 16;21(5):606-613 [FREE Full text] [doi: 10.1111/j.1365-2354.2012.01331.x] [Medline: 22335508]
59. Jambrina AM, Santomà A, Rocher A, Rams N, Cereza G, Rius P, et al. Detection and prevention of medication errors by the network of sentinel pharmacies in a southern European region. J Clin Med 2022 Dec 27;12(1):194 [FREE Full text] [doi: 10.3390/jcm12010194] [Medline: 36614995]
60. Aljadhey H, Mahmoud MA, Ahmed Y, Sultana R, Zouein S, Alshanawani S, et al. Incidence of adverse drug events in public and private hospitals in Riyadh, Saudi Arabia: the (ADESA) prospective cohort study. BMJ Open 2016 Jul 12;6(7):e010831 [FREE Full text] [doi: 10.1136/bmjopen-2015-010831] [Medline: 27406640]
61. Al-Harkan A, Al-Harkan N, Al-Najjar A, Al-Hunti A, Al-Rashidi A, Al-Themery A. Investigation of medication errors in a tertiary care hospitals in the Qassim Region, Saudi Arabia. Open Access Maced J Med Sci 2020 Apr 20;8(E):209-212. [doi: 10.3889/oamjms.2020.4330]
62. Qu C, Meng L, Wang N, Chen Y, Yang X, Wang J, et al. Identify and categorize drug-related problems in hospitalized surgical patients in China. Int J Clin Pharm 2019 Feb 4;41(1):13-17. [doi: 10.1007/s11096-018-0777-7] [Medline: 30610549]
63. Han JH, Heo KN, Han J, Lee MS, Kim SJ, Min S, et al. Analysis of medication errors reported by community pharmacists in the Republic of Korea: a cross-sectional study. Medicina (Kaunas) 2023 Jan 12;59(1):151 [FREE Full text] [doi: 10.3390/medicina59010151] [Medline: 36676775]
64. Pharmapod homepage. Pharmapod. URL: https://ca.pharmapodhq.com/ [accessed 2023-08-21]
65. Karsh BT, Escoto KH, Beasley JW, Holden RJ. Toward a theoretical approach to medical error reporting system research and design. Appl Ergon 2006 May;37(3):283-295 [FREE Full text] [doi: 10.1016/j.apergo.2005.07.003] [Medline: 16182233]
66. Orji R, Oyibo K, Tondello GF. A comparison of system-controlled and user-controlled personalization approaches. In: Proceedings of the Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization. 2017 Presented at: UMAP '17; July 9-12, 2017; Bratislava, Slovakia. [doi: 10.1145/3099023.3099116]
67. Salam SN, Yahaya WA, Ali AM. Using persuasive design principles in motivational feeling towards children dental anxiety (CDA). In: Proceedings of the Persuasive Technology, 5th International Conference. 2010 Presented at: Persuasive 2010; June 7-10, 2010; Copenhagen, Denmark. [doi: 10.1007/978-3-642-13226-1_23]

68. Jenkins R, Vaida A. Simple strategies to avoid medication errors. Fam Pract Manag 2007 Feb;14(2):41-47 [FREE Full text] [Medline: 17330716]
69. Koczmara C, Dueck C, Jelincic V. To err is human, to share is divine. Dynamics 2006;17(3):22-25. [Medline: 17009570]
70. Reporting and learning systems for medication errors: the role of pharmacovigilance centres. World Health Organization. 2014 Feb 22. URL: https://www.who.int/publications-detail-redirect/9789241507943 [accessed 2023-08-22]
71. Nielsen J, Molich R. Heuristic evaluation of user interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1990 Presented at: CHI '90; April 1-5, 1990; Seattle, Washington. [doi: 10.1145/97243.97281]
72. Chen P. 7 elements of persuasive design strategy. Association for Talent Development. 2017 Dec 19. URL: https://www.td.org/insights/7-elements-of-persuasive-design-strategy [accessed 2023-08-21]
73. Cialdini RB. Influence, New and Expanded: The Psychology of Persuasion. New York, NY: HarperCollins Publishers; 2021.
74. Orji R, Lomotey R, Oyibo K, Orji F, Blustein J, Shahid S. Tracking feels oppressive and 'punishy': exploring the costs and benefits of self-monitoring for health and wellness. Digit Health 2018;4:2055207618797554 [FREE Full text] [doi: 10.1177/2055207618797554] [Medline: 30202544]
75. Cham S, Algashami A, McAlaney J, Stefanidis A, Phalp K, Ali R. Goal setting for persuasive information systems: five reference checklists. Lect Notes Comput Sci 2019;11433 LNCS:237-253. [doi: 10.1007/978-3-030-17287-9_20]
76. Selassie HH, Oyibo K, Vassileva J. Responsiveness to persuasive strategies at the workplace: a case study. In: Proceedings of the E-Technologies: Embracing the Internet of Things. 2017 Presented at: MCETECH 2017; May 17-19, 2017; Ottawa, ON. [doi: 10.1007/978-3-319-59041-7_17]
77. Oyibo K, Adaji I, Orji R, Olabenjo B, Vassileva J. Susceptibility to persuasive strategies: a comparative analysis of Nigerians vs. Canadians. In: Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization. 2018 Presented at: UMAP '18; July 8-11, 2018; Singapore, Singapore URL: https://dl.acm.org/doi/10.1145/3209219.3209239 [doi: 10.1145/3209219.3209239]
78. Morton K, Sutton S, Hardeman W, Troughton J, Yates T, Griffin S, et al. A text-messaging and pedometer program to promote physical activity in people at high risk of type 2 diabetes: the development of the PROPELS follow-on support program. JMIR Mhealth Uhealth 2015 Dec 15;3(4):e105 [FREE Full text] [doi: 10.2196/mhealth.5026] [Medline: 26678750]
79. Oyibo K, Olagunju AH, Olabenjo B, Adaji I, Deters R, Vassileva J. BEN'FIT: design, implementation and evaluation of a culture-tailored fitness app. In: Proceedings of the Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization. 2019 Presented at: UMAP'19 Adjunct; June 9-12, 2019; Larnaca, Cyprus. [doi: 10.1145/3314183.3323854]
80. Matthews J, Win KT, Oinas-Kukkonen H, Freeman M. Persuasive technology in mobile applications promoting physical activity: a systematic review. J Med Syst 2016 Mar 9;40(3):72. [doi: 10.1007/s10916-015-0425-x] [Medline: 26748792]
81. Oyibo K. Designing culture-tailored persuasive technology to promote physical activity. University of Saskatchewan. 2020. URL: https://harvest.usask.ca/handle/10388/12943 [accessed 2023-08-22]
82. van Loon MH. Self-assessment and self-reflection to measure and improve self-regulated learning in the workplace. In: McGrath S, Mulder M, Papier J, Suart, R, editors. Handbook of Vocational Education and Training. Cham, Switzerland: Springer; 2018.
83. Berkovsky S, Freyne J, Oinas-Kukkonen H. Influencing individually: fusing personalization and persuasion. ACM Trans Interact Intell Syst 2012 Jun;2(2):1-8. [doi: 10.1145/2209310.2209312]
84. Mohadis HM, Ali NM. Designing persuasive application to encourage physical activity at workplace among older workers. In: Proceedings of the Sixth International Conference on Digital Information and Communication Technology and its Applications (DICTAP). 2016 Presented at: Sixth International Conference on Digital Information and Communication Technology and its Applications (DICTAP); July 21-23, 2016; Konya, Turkey. [doi: 10.1109/dictap.2016.7544013]
85. Orji R, Vassileva J, Mandryk RL. Modeling the efficacy of persuasive strategies for different gamer types in serious games for health. User Model User Adapt Inter 2014 Jul 14;24(5):453-498. [doi: 10.1007/s11257-014-9149-8]
86. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation: a 35-year odyssey. Am Psychol 2002 Sep;57(9):705-717. [doi: 10.1037/0003-066x.57.9.705]
87. Schembre SM, Liao Y, Robertson MC, Dunton GF, Kerr J, Haffey ME, et al. Just-in-time feedback in diet and physical activity interventions: systematic review and practical design framework. J Med Internet Res 2018 Mar 22;20(3):e106 [FREE Full text] [doi: 10.2196/jmir.8701] [Medline: 29567638]
88. Kluger AN, DeNisi A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. Psychol Bull 1996 Mar;119(2):254-284. [doi: 10.1037/0033-2909.119.2.254]
89. Special edition: tall man lettering; ISMP updates its list of drug names with tall man letters. Institute for Safe Medication Practices. 2016 Jun 2. URL: https://tinyurl.com/3skfpa2k [accessed 2023-08-22]
90. Langrial S. Exploring the influence of persuasive reminders and virtual rehearsal on the efficacy of health behavior change support systems. University of Oulu. 2014. URL: http://jultika.oulu.fi/files/isbn9789526204765.pdf [accessed 2024-01-11]
91. van Dantzig S, Geleijnse G, van Halteren AT. Toward a persuasive mobile application to reduce sedentary behavior. Pers Ubiquit Comput 2012 Jul 12;17(6):1237-1246. [doi: 10.1007/s00779-012-0588-0]

92. O'Leary K, Liu L, McClure JB, Ralston J, Pratt W. Persuasive reminders for health self-management. AMIA Annu Symp Proc 2016;2016:994-1003 [FREE Full text] [Medline: 28269896]
93. Hollas T, Lane M, Coyne J, Ellis C. What is verbal persuasion. In: Loureiro MJ, Loureiro A, Gerber HR, editors. Advances in Educational Technologies and Instructional Design. Hershey, PA: IGI Global; 2023:58-74.
94. Empowerment and change: the purpose of counselling. Pearson. URL: https://tinyurl.com/t5cxrw5r [accessed 2023-12-21]
95. Kim C, Jeon HG, Lee KC. Discovering the role of emotional and rational appeals and hidden heterogeneity of consumers in advertising copies for sustainable marketing. Sustainability 2020 Jun 25;12(12):5189. [doi: 10.3390/su12125189]
96. Oyibo K. The relationship between perceived health message motivation and social cognitive beliefs in persuasive health communication. Information 2021 Aug 28;12(9):350. [doi: 10.3390/info12090350]
97. Oyibo K, Ali YS, Vassileva J. An empirical analysis of the perception of mobile website interfaces and the influence of culture. In: Proceedings of the Personalization in Persuasive Technology Workshop, Persuasive Technology. 2016 Presented at: Personalization in Persuasive Technology Workshop, Persuasive Technology 2016; April 5, 2016; Salzburg, Austria.
98. Oyibo K, Vassileva J. Investigation of persuasive system design predictors of competitive behavior in fitness application: a mixed-method approach. Digit Health 2019 Nov 01;5:2055207619878601 [FREE Full text] [doi: 10.1177/2055207619878601] [Medline: 31700652]
99. Mcleod S. Albert Bandura’s Social Learning Theory. SimplyPsychology. 2023 Oct 24. URL: https://www.simplypsychology.org/bandura.html [accessed 2023-08-22]
100. Lyons SD, Berge ZL. Social learning theory. In: Seel NM, editor. Encyclopedia of the Sciences of Learning. Boston, MA: Springer; 2012.
101. Festinger L. A theory of social comparison processes. Hum Relat 2016 Apr 22;7(2):117-140. [doi: 10.1177/001872675400700202]
102. Foster D, Linehan C, Kirman B, Lawson S, James G. Motivating physical activity at work: using persuasive social media for competitive step counting. In: Proceedings of the 14th International Academic MindTrek Conference: Envisioning Future Media Environments. 2010 Presented at: MindTrek '10; October 6-8, 2010; Tampere, Finland. [doi: 10.1145/1930488.1930510]
103. Oyibo K, Orji R, Vassileva J. The influence of culture in the effect of age and gender on social influence in persuasive technology. In: Proceedings of the Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization. 2017 Presented at: UMAP '17; July 9-12, 2017; Bratislava, Slovakia. [doi: 10.1145/3099023.3099071]
104. Myers DG. Social Psychology. New York, NY: McGraw-Hill Education; 2009.
105. Oyibo K, Morita PP. The effect of persuasive design on the adoption of exposure notification apps: quantitative study based on COVID alert. JMIR Form Res 2022 Sep 6;6(9):e34212 [FREE Full text] [doi: 10.2196/34212] [Medline: 35580138]
106. AIMS (Assurance and Improvement in Medication Safety). Ontario College of Pharmacists. URL: https://www.ocpinfo.com/regulations-standards/aims-assurance-and-improvement-in-medication-safety/ [accessed 2023-08-22]
107. Yuan CL, Moon H, Kim KH, Wang S, Yu X. Third-party organization endorsement impacts on perceived value and B2B customer loyalty. Ind Mark Manage 2020 Oct;90:221-230. [doi: 10.1016/j.indmarman.2020.07.021]
108. Fogg BJ, Tseng H. The elements of computer credibility. In: Proceedings of the SIGCHI conference on Human Factors in Computing Systems. 1999 Presented at: CHI '99; May 15-20, 1999; Pittsburgh, PA. [doi: 10.1145/302979.303001]
109. Fogg BJ, Marshall J, Laraki O, Osipovich A, Varma C, Fang N, et al. What makes web sites credible?: a report on a large quantitative study. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2001 Presented at: CHI '01; March 31-April 5, 2001; Seattle, WA. [doi: 10.1145/365024.365037]
110. Verifiability. Collins Dictionary. URL: https://www.collinsdictionary.com/dictionary/english/verifiability [accessed 2023-08-22]
111. Jones R. Information and persuasive speaking. In: Communication in the Real World: An Introduction to Communication Studies. Minneapolis, MN: University of Minnesota Libraries Publishing; 2016.
112. Wolf ZR, Hughes RG. Error reporting and disclosure. In: Patient Safety and Quality: An Evidence-Based Handbook for Nurses. Rockville, MD: Agency for Healthcare Research and Quality (US); 2008.
113. Orji R, Mandryk RL. Developing culturally relevant design guidelines for encouraging healthy eating behavior. Int J Hum Comput Stud 2014 Feb;72(2):207-223. [doi: 10.1016/j.ijhcs.2013.08.012]
114. Oyibo K, Orji R, Vassileva J. Developing culturally relevant design guidelines for encouraging physical activity: a social cognitive theory perspective. J Healthc Inform Res 2018 Dec 31;2(4):319-352 [FREE Full text] [doi: 10.1007/s41666-018-0026-9] [Medline: 35415413]
115. Kaldjian LC, Jones EW, Wu BJ, Forman-Hoffman VL, Levi BH, Rosenthal GE. Reporting medical errors to improve patient safety: a survey of physicians in teaching hospitals. Arch Intern Med 2008 Jan 14;168(1):40-46. [doi: 10.1001/archinternmed.2007.12] [Medline: 18195194]

Abbreviations

AIMS: Assurance and Improvement in Medication Safety
CQI: continuous quality improvement

IM: Incident Management
MI: medication incident
MIRLS: Medication Incident Reporting and Learning Systems
PSD: Persuasive System Design
RQ: research question
SMART: specific, measurable, achievable, relevant, and timebound
UX: user experience

Edited by A Kushniruk; submitted 30.07.22; peer-reviewed by C Hohl, R Asbjornsen; comments to author 29.04.23; revised version received 29.08.23; accepted 20.11.23; published 21.03.24.

Please cite as:
Oyibo K, Gonzalez PA, Ejaz S, Naheyan T, Beaton C, O'Donnell D, Barker JR
Exploring the Use of Persuasive System Design Principles to Enhance Medication Incident Reporting and Learning Systems: Scoping Reviews and Persuasive Design Assessment
JMIR Hum Factors 2024;11:e41557
URL: https://humanfactors.jmir.org/2024/1/e41557
doi:10.2196/41557
PMID:38512325

Original Paper

Clinical Decision Support Requirements for Ventricular Tachycardia Diagnosis Within the Frameworks of Knowledge and Practice: Survey Study

Zhao Hu[1*], MM; Min Wang[2*], MS; Si Zheng[2], MS; Xiaowei Xu[2], MS; Zhuxin Zhang[1], MS; Qiaoyue Ge[3], MS; Jiao Li[2], PhD; Yan Yao[1], PhD, MD

[1]Arrhythmia Center, Fuwai Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College/National Center for Cardiovascular Diseases, Beijing, China

[2]Institute of Medical Information, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

[3]West China School of Public Health, West China Fourth Hospital, Sichuan University, Chengdu, China

[*]these authors contributed equally

Corresponding Author:
Yan Yao, PhD, MD
Arrhythmia Center
Fuwai Hospital
Chinese Academy of Medical Sciences & Peking Union Medical College/National Center for Cardiovascular Diseases
Beilishi Road 167
Beijing, 100037
China
Phone: 86 10 88322401
Email: ianyao@263.net.cn

Abstract

Background: Ventricular tachycardia (VT) diagnosis is challenging due to the similarity between VT and some forms of supraventricular tachycardia, complexity of clinical manifestations, heterogeneity of underlying diseases, and potential for life-threatening hemodynamic instability. Clinical decision support systems (CDSSs) have emerged as promising tools to augment the diagnostic capabilities of cardiologists. However, a requirements analysis is acknowledged to be vital for the success of a CDSS, especially for complex clinical tasks such as VT diagnosis.

Objective: The aims of this study were to analyze the requirements for a VT diagnosis CDSS within the frameworks of knowledge and practice and to determine the clinical decision support (CDS) needs.

Methods: Our multidisciplinary team first conducted semistructured interviews with seven cardiologists related to the clinical challenges of VT and expected decision support. A questionnaire was designed by the multidisciplinary team based on the results of interviews. The questionnaire was divided into four sections: demographic information, knowledge assessment, practice assessment, and CDS needs. The practice section consisted of two simulated cases for a total score of 10 marks. Online questionnaires were disseminated to registered cardiologists across China from December 2022 to February 2023. The scores for the practice section were summarized as continuous variables, using the mean, median, and range. The knowledge and CDS needs sections were assessed using a 4-point Likert scale without a neutral option. Kruskal-Wallis tests were performed to investigate the relationship between scores and practice years or specialty.

Results: Of the 687 cardiologists who completed the questionnaire, 567 responses were eligible for further analysis. The results of the knowledge assessment showed that 383 cardiologists (68%) lacked knowledge in diagnostic evaluation. The overall average score of the practice assessment was 6.11 (SD 0.55); the etiological diagnosis section had the highest overall scores (mean 6.74, SD 1.75), whereas the diagnostic evaluation section had the lowest scores (mean 5.78, SD 1.19). A majority of cardiologists (344/567, 60.7%) reported the need for a CDSS. There was a significant difference in practice competency scores between general cardiologists and arrhythmia specialists (P=.02).

Conclusions: There was a notable deficiency in the knowledge and practice of VT among Chinese cardiologists. Specific knowledge and practice support requirements were identified, which provide a foundation for further development and optimization of a CDSS. Moreover, it is important to consider clinicians' specialization levels and years of practice for effective and personalized support.

(JMIR Hum Factors 2024;11:e55802) doi:10.2196/55802

KEYWORDS

clinical decision support system; requirements analysis; ventricular tachycardia; knowledge; clinical practice; questionnaires

Introduction

Sudden cardiac death (SCD) remains a significant public health issue, accounting for 50% of all cardiovascular deaths. The estimated annual incidences of SCD are 60 [1], 40.7 [2,3], and 36.8 [4] per 100,000 people in the United States, China, and Europe, respectively. Ventricular tachycardia (VT) is a major cause or precursor of SCD [5], which can be the initial or sole manifestation of diverse heart diseases [6,7]. VT diagnosis is challenging due to its similarity with some forms of supraventricular tachycardia, the complexity of clinical manifestations, heterogeneity of underlying diseases, and potential for life-threatening hemodynamic instability [6,8]. Diagnostic accuracy and timing are critical for patients with VT, as the stage of diagnosis determines the selection of treatment [9]. However, studies have revealed a substantial prevalence of misdiagnoses of VT [10-13], focusing on differential diagnosis between VT and supraventricular tachycardia. Although diagnostic error has been a challenge along the development of medicine, measuring diagnostic error can be difficult due to detection and reporting biases, with scarce reports indicating error rates of approximately 10%-15% [14]. We could not find additional estimates for the actual diagnostic error of VT; however, it is commonly acknowledged to represent a substantial challenge considering the complexity of the condition [9,15].

Diagnosis represents a complex cognitive process comprising a variety of different problem-solving tasks that are related to the clinical reasoning process, such as taking a medical history, forming a differential diagnosis, ordering examinations, and interpreting clinical findings [16]. The diagnostic process requires not only the retention of knowledge but also the judicious application of that knowledge at opportune moments, namely in clinical practice. A proper diagnosis of VT demands a great volume of knowledge. First, the clinician must be able to identify VT among the spectrum of wide QRS tachycardias by inspecting a list of electrocardiogram (ECG) features and comparing the findings to various diagnostic criteria or algorithms [17,18]. Once VT is identified by ECG interpretation, the next step is to diagnose the underlying diseases from a vast disease spectrum. This is a particularly challenging task, as any disease involving the myocardium can cause VT, such as coronary artery disease (CAD), all types of cardiomyopathies, myocarditis, inherited arrhythmia syndromes, autoimmune or inflammatory diseases, and others [7,9]. Moreover, translating the enormous body of knowledge into proper practice can be difficult [19], which is exacerbated by the fact that VT can cause stress to clinicians due to the probability of hemodynamic instability.

In response to this challenge, the clinical decision support system (CDSS) has emerged as a promising tool to augment the diagnostic capabilities of clinicians. Clinical decision support (CDS) is a process for enhancing health-related decisions with pertinent, organized clinical knowledge and patient information, thus advancing health care delivery [20]. Use of a CDSS can provide clinicians with situation-specific knowledge that aids in making critical clinical decisions such as risk assessment, diagnosis, prognosis, and selection of therapy [21]. A clinical diagnostic decision support system (DDSS) is a computer-based algorithm that assists a clinician with one or more component steps of the diagnostic process [22]. A DDSS is expected to receive relevant patient information and return outputs to assist with the problems the clinician has encountered in the diagnostic process, such as suggesting a likely diagnosis. Some well-known DDSSs such as ISABEL [23] and Dxplain [24] provide a diagnosis list, which can offer a solution to the challenges associated with VT diagnosis. Most CDSSs exhibit efficacy in a laboratory or experimental environment; however, relatively few such systems are being used at present and the rate of use in routine clinical practice is low [20,25-27]. Studies have identified the main barriers to the widespread adoption of CDSSs, including vague requirements, poor integration with the clinical workflow, low user acceptance or trust, and lack of transparency. Among these barriers, comprehensive user requirements engineering should be performed at the very beginning of development, which should be continued iteratively throughout the CDSS design-development-implementation life cycle [25,26,28,29]. To address this gap, several recent studies have aimed at elucidating the clinical requirements for an effective and usable CDSS in the context of specific fields or scenarios [30-34] with a variety of methods, including focus groups [30,35], a workshop [34], expert discussion with a literature review [36,37], semistructured interviews [31,34,35,38], writing user stories [39], and system evaluation [40]. Overall, most studies have adopted a user-centered approach with qualitative analysis.

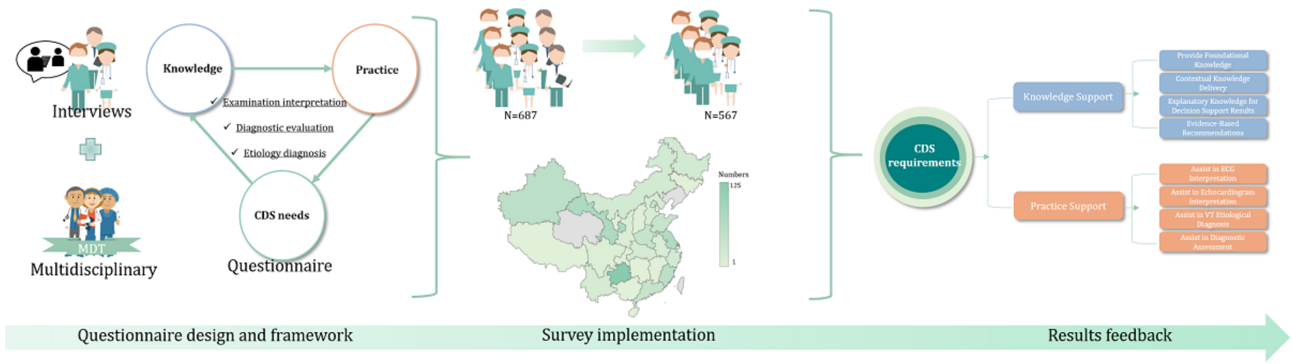
To our best knowledge, although an artificial intelligence model was reported for predicting the in-hospital mortality of VT [8], no CDSS has been developed for VT diagnosis. A recent systemic review of cardiovascular CDSSs found that the complexity of the clinical management of cardiovascular disease itself was a barrier during implementation [27], which emphasizes the need for an authentic clinical requirements analysis. Accordingly, the objective of this study was to analyze the requirements for a VT diagnosis CDSS within the frameworks of knowledge, practice, and CDS needs.

Methods

Study Design and Recruitment Process

Figure 1 shows the overall flow of our study, which consisted of semistructured interviews in the early stages and questionnaires in the later stages. To effectively implement and conduct the questionnaire assessment, we conducted open and explorative interviews about the challenges associated with the management of VT and the expected functions of a CDSS for VT. The interviews were conducted at Fuwai Hospital, the national cardiovascular disease center of China. This hospital actively recruits cardiologists for their fellowships from all regions of China, resulting in a representative sample of interviewees. We sent interview invitations to all 56 cardiologists in the arrythmia center, including cardiologists from the fellowship program or established staff of Fuwai Hospital. Seven cardiologists responded and completed the interview, followed by a brief questionnaire to provide information on demographics and clinical experience (see Multimedia Appendix 1).

Figure 1. Schematic of the overall study workflow and assessment approach. CDS: clinical decision support; ECG: electrocardiogram; MDT: multidisciplinary team; VT: ventricular tachycardia.



A multidisciplinary team was formed to define the purpose of our study and the design of the questionnaire based on the interview results. The multidisciplinary team comprised three arrhythmia specialists, three experts in medical informatics and CDS, and one clinical statistician. The questionnaire was examined by an additional 20 arrhythmia specialists to ensure its clarity and feasibility. We conducted a nationwide cross-sectional survey with an online questionnaire in mainland China from December 31, 2022, to February 15, 2023. We recruited registered cardiologists using a convenience sampling approach from network groups associated with the Asian Heart Rhythm Association (AHRA) on WeChat, the dominant social media app in China. The AHRA is an academic organization focusing on arrhythmias, whose members are all registered cardiologists. Duplicate submissions were prevented through IP address constraints, and only completed responses were included for analysis.

Ethical Considerations

Participants provided online informed consent, which detailed the survey's background, aim, methods, and confidentiality measures. To protect participants' privacy, a signature was not required. Instead, participants clicked the "go on" button at the bottom of the informed consent page if they agreed to participate. According to data privacy protocols, no personal information, including the participants' names or affiliations, was collected. Since patients were not the subject of this study, ethical approval was exempted by the ethics committee of the Institute of Medical Information, Chinese Academy of Medical Sciences/Peking Union Medical College [41]. Each participant received ~US $3 as compensation.

Questionnaire Design

Overview

The questionnaire was divided into four sections (Table 1): demographic information (questions 1-6), knowledge assessment (question 14), practice assessment (questions 7-13), and CDS needs (questions 15-18). A comprehensive version of the questionnaire is provided in Multimedia Appendix 2.

Table 1. Design of the questionnaire.

Section	Content	Related questions
Knowledge	Examination interpretation, etiological diagnosis, diagnostic evaluation, conceptual knowledge	14
Practice	Examination interpretation, etiological diagnosis, diagnostic evaluation	7-13
Clinical decision support needs	Interpretable diagnosis, executable processes, knowledge support	15-18

Knowledge Assessment

Knowledge serves as the theoretical foundation for clinicians to make clinical diagnoses and is thus an essential competency for clinicians. The diagnosis of VT is difficult as it will largely depend on the clinician's familiarity with the vast knowledge of the field. The European Society of Cardiology (ESC) guideline suggests a protocol for VT diagnosis [15]. The multidisciplinary team abstracted the knowledge points from the ESC guideline for collecting information on the participants' self-reported knowledge shortcomings.

Practice Assessment

Areas of Focus

To attain a more accurate gauge of the clinical practice competency, we used simulated cases rather than straightforward questions [42], which can help differentiate practice competency from knowledge. To mitigate the risk of low response rates and careless submissions associated with lengthy surveys [43], we designed two stepwise cases containing seven questions. According to the intention, the questions about clinical practice were divided into three parts: examination interpretation, etiological diagnosis, and diagnostic evaluation. Multiple-choice options were available for all the questions. We standardized the total score for each section to 10 points according to the weighting.

Examination Interpretation

Accurate interpretation of an examination is the basis for a correct etiological diagnosis. ECG is the first-line examination modality for arrhythmias, as nearly all arrhythmia episodes are detected by ECG. Therefore, for this section, we focused on the identification of VT and sites of origin of VT on ECG [15].

Etiological Diagnosis

A correct etiological diagnosis of VT is necessary for appropriate treatment. The main strategy is to identify or exclude structural heart diseases, including CAD, myocarditis, and cardiomyopathies [44]. In this section, we assessed the correctness of a diagnosis of arrhythmogenic right ventricular cardiomyopathy (ARVC) and acute myocarditis as the two cases.

Diagnostic Evaluation

Diagnostic evaluation is a process of collecting clinical information to confirm or exclude a suspected diagnosis. A diagnostic evaluation protocol for VT is recommended in the ESC guideline [15] with the goal of reducing the rate of diagnostic errors. Based on the cases with an etiological diagnosis, we assessed the competency of the participants to arrange further diagnostic evaluations.

CDS Needs

According to the ESC guideline [15] and universal CDSS functionality [25], the multidisciplinary team summarized the results of the interviews to produce a list of functions required for CDS, which could be divided into executable processes, interpretable diagnosis, and knowledge support. We employed this list to poll the functionalities required by the cardiologists for a VT CDSS.

Quality Control of Responses

To ensure the validity and reliability of our survey responses, we used two strategies to filter out potentially low-quality submissions. First, participants who completed the questionnaire in under 2 minutes were excluded. This threshold was determined through a pretest evaluation coupled with multidisciplinary team discussions. Second, responses were considered to be invalid if participants selected all the available options for questions 7, 8, 9, 11, 12, or 13. This exclusion criterion was established based on the consensus opinion of the multidisciplinary team, who deemed such selections to be unreasonable.

Statistical Analysis

We only included valid questionnaire responses in the statistical analysis. All data in the demographic section were categorical. Comparisons were performed using mean, median, range, and percentage. The scores in the practice section are expressed as continuous variables, using the mean, median, and range. The knowledge and CDS sections were phrased as single-choice questions asking clinicians about their subjective views on given statements using a 4-point Likert scale without a neutral option. The internal consistency of the questionnaire was assessed using the Cronbach α value.

In addition, we grouped participants separately by practice years and specialty for further subgroup analyses. The Kruskal-Wallis test was performed to investigate the relationship between practice scores and practice years or specialty. All analyses were conducted in R version 4.0.3 [45]. We analyzed most of the data descriptively using graphics produced by the R package ggplot2.

Results

Sociodemographic Characteristics of Participants

A total of 687 questionnaires were completed. After applying our quality control measures, 567 responses were considered valid, yielding a validity rate of 82.53%. Among the invalid questionnaires, 104 responses were excluded due to a completion time of less than 2 minutes and 16 were excluded for selecting all options in questions 7, 8, 9, 11, 12, or 13. Descriptive statistics regarding the sociodemographic characteristics of participants are presented in Table 2. Of the enrolled participants, 54.50% were men; 93.47% were general cardiologists and the others were cardiac arrhythmia specialists. More than half of the participants were from tertiary A hospitals. Only a small percentage of cardiologists had ever used a CDSS, and the majority reported needing a CDSS to assist them in the management of VT (Table 2).

Table 2. Demographic characteristics of the survey participants (N=567).

Characteristics	Participants, n (%)
Gender	
Woman	258 (45.5)
Man	309 (54.50)
Age (years)	
≤30	89 (15.7)
31-35	152 (26.81)
36-40	129 (22.75)
41-45	92 (16.23)
46-50	60 (10.58)
≥51	45 (7.94)
Department	
Cardiology	530 (93.47)
Cardiac arrhythmia specialty	39 (6.88)
Professional title	
Resident physician	120 (21.16)
Attending	237 (41.8)
Associate chief	145 (25.57)
Chief	65 (11.46)
Years of practice	
<10	247 (43.54)
10-20	213 (37.57)
>20	107 (18.87)
Hospital tier	
Tertiary A	414 (73.02)
Not tertiary A	153 (26.98)
Ever used a CDSS[a]?	
Yes	72 (12.70)
No	495 (87.30)
Is there a need for a CDSS?	
Yes	523 (92.24)
No	44 (7.76)

[a]CDSS: clinical decision support system.

Semistructured Interviews

Textbox 1 summarizes the results of the semistructured interviews, in which we focused on the challenges of VT management and CDSS needs. The responses of the seven cardiologists were focused, with each noting that etiological diagnosis and interpretation of ECG results were their main challenges. The most important demand was the provision of quick and concise recommendations on diagnosis and treatment. The interviewees also expected the CDSS to provide clinical pathways.

Textbox 1. Results of the interviews.

- Challenges in the management of ventricular tachycardia (VT)

1. Etiological diagnosis
2. Wide QRS tachycardia diagnosis on electrocardiogram (ECG)
3. Determination of the location of VT origin on ECG
4. Mechanisms of VT
5. Drug treatment options
6. Options for the treatment of polymorphic VT

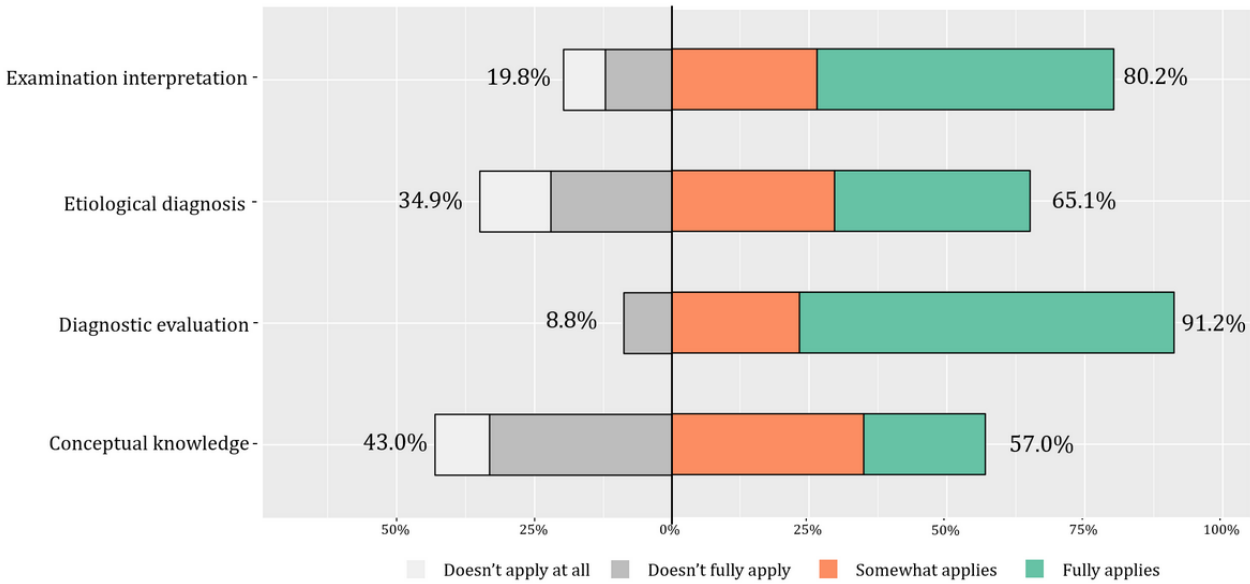
- Clinical decision support system needs

1. Rapid and concise recommendations for diagnosis and treatment
2. Diagnostic and therapeutic pathways for different etiologies
3. Aids in the identification of wide QRS
4. Adjunctive etiological diagnosis
5. Diagnostic supplements for related diseases

Knowledge

Figure 2 shows that there was an overall lack of knowledge with respect to diagnostic evaluation, with 383 of the 567 (68.0%) cardiologists indicating full need of assistant knowledge in diagnostic evaluation. This was followed by examination interpretation, where 305 of the 567 (53.8%) cardiologists were in full need of knowledge regarding the interpretation of ECG, cardiac ultrasound, and other cardiac examinations. The need for conceptual knowledge was relatively lower, even though it still reached nearly 60%.

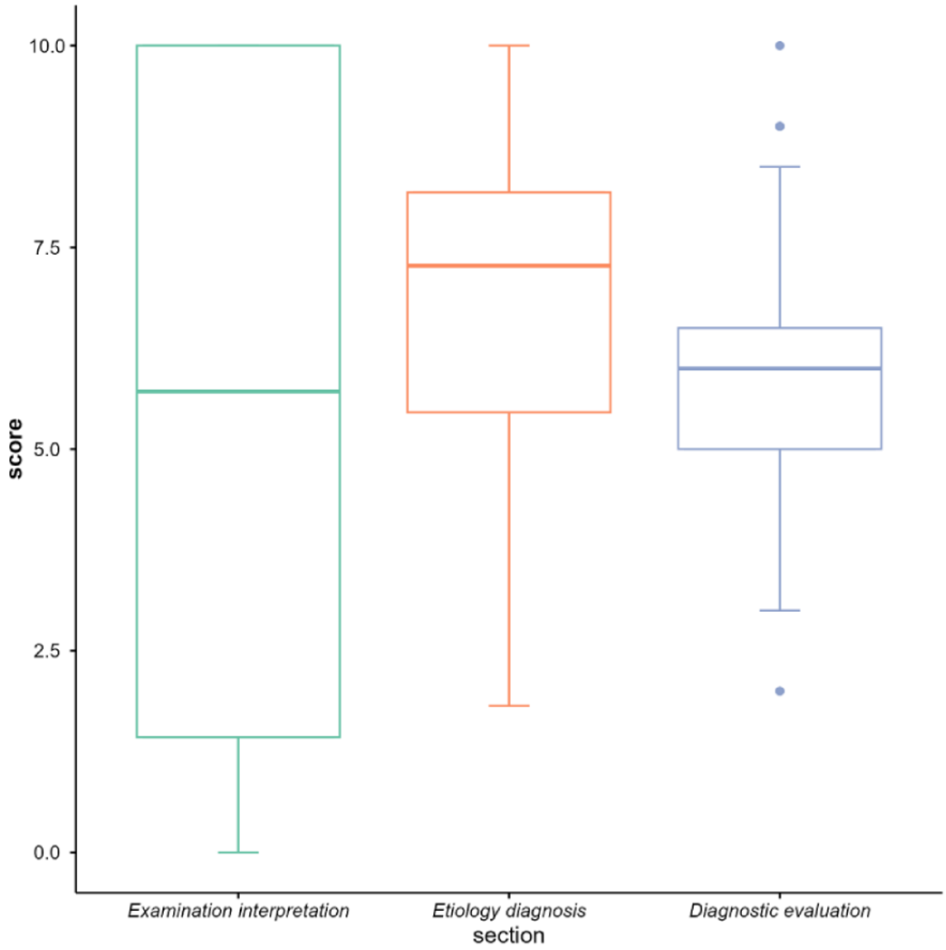
Figure 2. Knowledge assessment.



Practice

The overall average score of the practice questions was 6.11 (SD 0.55), the internal consistency of which was confirmed by a Cronbach α of 0.913. The mean scores of the examination interpretation, etiological diagnosis, and diagnostic evaluation were 6.22 (SD 3.94), 6.74 (SD 1.75), and 5.78 (SD 1.19), respectively. As shown in Figure 3, the etiological diagnosis section was associated with the highest overall score and the distribution of scores was also more concentrated than for the other sections, especially when compared with the distribution of the examination interpretation scores that were more dispersed and polarized.

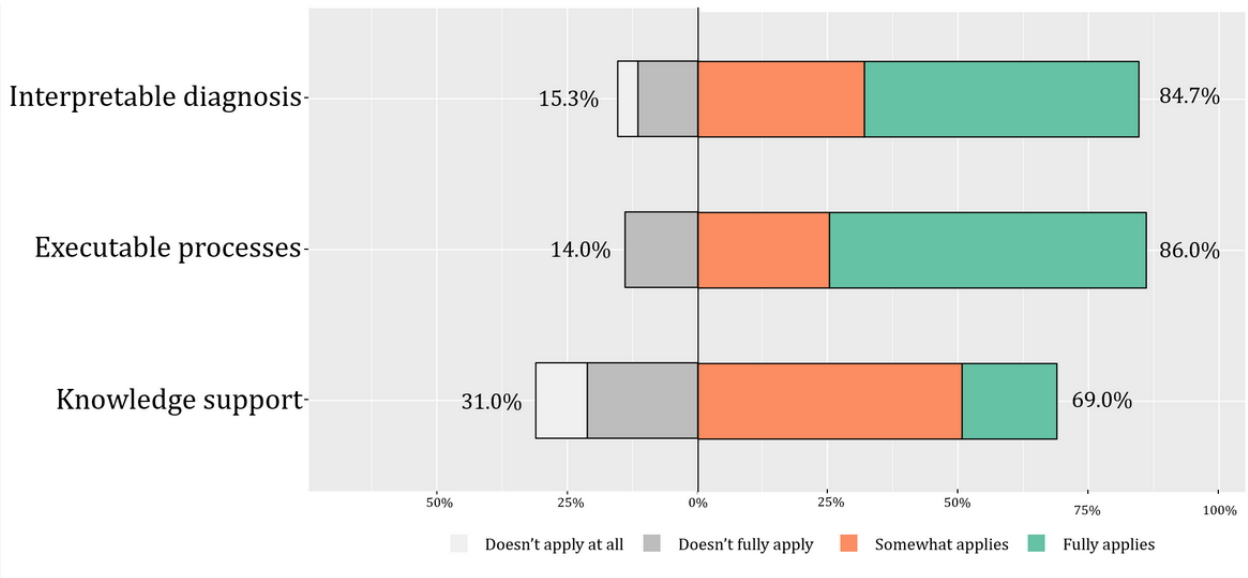
Figure 3. Practice assessment.



CDS Needs

The majority of the surveyed cardiologists reported a positive attitude toward CDS needs (Figure 4). There was relatively higher demand expressed for functions related to executable processes and interpretable diagnosis. In particular, the executable processes function was considered to be an essential requirement of a CDSS by 344 of the 567 cardiologists (60.7%). Knowledge support function received the least support but was still close to 70%.

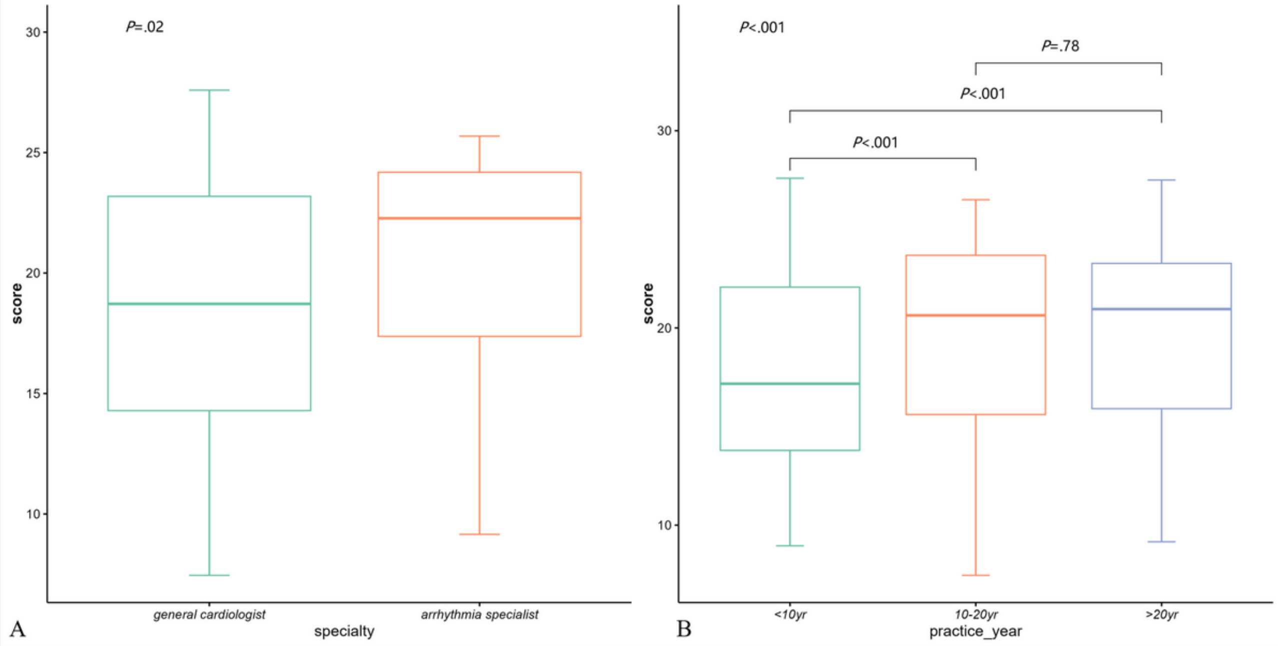
Figure 4. Clinical decision support needs assessment.



Subgroup Analysis

We divided all the cardiologists into subgroups based on specialty (Figure 5A) and practice years (Figure 5B). The Kruskal-Wallis test showed a significant difference in practice competency scores between general cardiologists and arrhythmia specialists (*P*=.02). Subgroup analysis according to years of practice revealed a significant effect of experience on scores. The <10 years group had significantly lower scores compared to those of the 10-20 years and >20 years groups. However, there was no significant difference between those with 10-20 years and >20 years of experience.

Figure 5. Subgroup analyses according to (A) specialty and (B) years of practice.



Discussion

Principal Results

Based on a combination of semistructured interviews and questionnaires, this study conducted a large-scale nationwide survey for cardiologists to understand their knowledge and practice competence about VT diagnosis and their requirements for a related CDSS. The results indicated that knowledge and practice support in examination interpretation, etiological diagnosis, and diagnostic evaluation are considered to be essential for a VT diagnosis CDSS. In addition, the vast majority of the cardiologists gave a positive response with respect to the need for a CDSS.

CDSS Requirements

Previous research on CDSS requirements has primarily relied on methods such as interviews [31,34,35,38,39] and group discussions [30,34,35] to elicit users' subjective needs. Based on recommendations from clinical experts and medical informatics professionals within our research team, it was acknowledged that certain objective requirements might not be articulated by users during interviews. Consequently, a questionnaire was designed to assess and uncover the requirements that might not have been spontaneously expressed during interviews. Previous studies have used questionnaires to investigate the knowledge, attitudes, and practices of health care professionals in various specific tasks [46-53], providing a basis for our questionnaire approach. To objectively reflect cardiologists' knowledge and practice deficiencies, we opted to not directly inquire about specific knowledge points but instead used two case scenarios to simulate authentic VT diagnostic situations, which is proven to be an appropriate method to assess practice competence [54]. The survey results endorsed the advantages of this mixed methods approach. The difficulties in VT diagnosis mentioned by the cardiologists during interviews primarily focused on distinguishing wide QRS tachycardias on ECG and identifying the etiology of VT, with no mention of diagnostic evaluation. However, results from the practice section of the questionnaire indicated poorer competence in diagnostic evaluation compared to etiological diagnosis, suggesting that the interviewees were not consciously aware of their weaknesses in diagnostic evaluation during interviews. Currently, there is no unified systematic method for conducting a CDS requirements analysis. While our method of integrating interviews and questionnaires provides a comprehensive approach, there is still room for improvement. Use of a simulation game has been suggested as a better means for clinical competence assessment [42]. Future research could consider incorporating cognitive analysis [55] and real-world system usability evaluation [56] to further optimize CDSS requirements analysis.

The objective results from case simulations also affirmed the cardiologists' need for decision support (Figure 4). Regarding knowledge requirements, the results from the CDS needs section of the questionnaire indicated that participants had relatively fewer demands for knowledge support compared to direct decision support. Moreover, the cardiologists revealed a preference for automatically prompted relevant knowledge during the diagnostic and therapeutic processes, which can provide more targeted knowledge support (Figure 2). The challenge lies in ensuring that the CDSS accurately identifies the current diagnostic and therapeutic tasks; determines user knowledge gaps; and automatically retrieves, integrates, and presents knowledge support rapidly and accurately [57]. The results of the practice competence highlighted the need for improvement in the interpretation of diagnostic tests, etiological diagnosis, and diagnostic evaluation, suggesting the need for decision support in these three aspects, which were also highlighted as key clinical reasoning [58]. Notably, the accuracy of etiological diagnosis was relatively high, aligning with the lower knowledge demand for an etiological diagnosis (Figure 3). In terms of CDSS needs, the cardiologists favored direct decision support over knowledge support, including explanatory diagnoses and executable evaluation processes, which has also been recognized in recent studies [57,59,60].

Synthesizing the findings of this study, we propose the following recommendations of specific functions of a CDSS for VT diagnosis under a framework of knowledge and practice. With respect to knowledge support, the CDSS needs to (1) provide foundational knowledge by offering fundamental knowledge for each relevant disease that is available for clinicians to retrieve and browse; (2) contextualize knowledge delivery by providing closely related knowledge at decision points, including, but not limited to, the interpretation of diagnostic tests such as ECGs and echocardiograms, wide QRS complex differentiation, etiological diagnosis of VT, and the issuance of diagnostic test orders; (3) explain the knowledge underlying CDSS results; and (4) provide evidence-based recommendations at decision points with available evidence support. With respect to practice support, the CDDS should (1) assist in ECG interpretation, including distinguishing wide QRS complex tachycardias, identifying useful features for etiological diagnosis during sinus rhythm and VT, and recommending diagnostic test orders; (2) assist in echocardiogram interpretation, including the recognition of common etiologies of VT such as old myocardial infarction, ARVC, myocarditis, and the classification of phenotypes of cardiomyopathies; (3) provide suspected etiological diagnoses based on existing information for patients with VT, including acute coronary syndrome, ischemic cardiomyopathy, ARVC, and acute myocarditis, with specific emphasis on alerting clinicians who may not have considered the possibility of acute coronary syndrome; and (4) supplement diagnostic assessments with additional information, including critical medical history, physical examination, laboratory tests, and other examinations. Particularly, using a comprehensive differential diagnosis list is advocated to mitigate premature closure [14], as substantiated by a recent study [61].

Dxplain [24], one of the few DDSSs available for general practice, provides a diagnosis list according to input patient manifestations, which aligns with our proposed structure for VT etiological diagnosis. However, Dxplain lacks knowledge support, examination interpretation, and diagnostic assessment functions, which are highlighted as requirements for a VT CDSS as mentioned above. Another well-known commercial diagnostic support tool, ISABEL, not only serves as a diagnosis reminder but also provides knowledge support (ie, evidence-based knowledge of each disease). However, it does not satisfy the other requirements identified in this study [23,62]. Dr. Mayson [63] is a Chinese commercial CDSS for general practice, which can abstract data from electronic health records to form a diagnosis list as well as provide assistance in diagnostic assessment. Like ISABEL, Dr. Mayson provides a knowledge database for each disease, including clinical practice guidelines. However, the knowledge support is at the disease level rather than the decision level. In addition, this CDSS does not assist with examination interpretation.

Although our study mainly investigated the specific functionalities for VT diagnosis, the results indicated some general CDSS functionalities, including interpretability of decision-making as well as the overall feasibility of the CDSS workflow. Several reviews [64-66] summarized other universal features worthy of consideration, such as integration with the clinical workflow and electronic health record system, reduction of manual input of patient data, execution users' desired action, avoidance of unnecessary alerts, documentation of reasons for rejecting recommendations, as well as the "five rights" of CDS (providing the right information to the right people in the right formats through the right channels at the right time) [67].

We believe that an excellent CDSS should provide tailored assistance for different types of clinicians. Thus, a subgroup analysis was performed according to the clinician characteristics in the practice section (Figure 5). As anticipated, arrhythmia specialists outperformed general cardiologists, which aligns with the findings of previous research [68]. The American College of Cardiology defines different types of cardiovascular specialists that have requirements for different types of support in cardiovascular health care [69]. A CDSS should be tailored to clinicians' specialization levels to assist in diagnostic and therapeutic practices. For highly specialized clinicians facing a narrow spectrum of diseases, CDSS assistance may be limited, while support for foundational diagnostic and therapeutic aspects outside their specialty may be necessary. Conversely, less specialized clinicians facing a broader spectrum of diseases may need support in staying updated with the latest diagnostic and therapeutic advancements. For instance, for less experienced clinicians facing patients with VT, the CDSS should always indicate the possibility of CAD. For experienced clinicians, as they have already cultivated the mindset to exclude CAD, the CDSS might only provide this alert when they miss the diagnosis of CAD. Furthermore, it is expected that the CDSS could continually adapt to individual needs through observing clinician users' behaviors. The impact of years of practice on performance seems to be nonlinear. Clinicians practicing for 10-20 years or more demonstrated better performance than those practicing for less than 10 years. However, there was no significant difference between the 10-20 years and >20 years groups, suggesting that clinical skills may grow in the first 10 years of practice but plateau afterward, thereby challenging the CDSS design to provide targeted support for clinicians with different levels of experience in practice. Additionally, for clinicians entering a bottleneck period in competence growth, the CDSS could facilitate education during practice, thereby supporting lifelong learning. Several studies have been performed in this regard in the areas of pharmaceutical skills [70], imaging interpretation [71], geriatric care [72], and periprocedural antithrombotic use [73].

Most existing CDSSs have been generally designed for health care providers but might not fully consider the diversity of requirements as well as their expertise levels [74]. The genuine needs of health care providers have not been effectively communicated to system developers, resulting in the design of CDSSs that struggle to fulfill their intended role of assistance and workload reduction. Our study centers around the clinical scenario of VT diagnosis, comprehensively exploring support requirements in both knowledge and practice. This investigation can thus provide a foundation for the development of a relevant CDSS. Additionally, we aspire for this study to serve as a reference for clinical needs research, encouraging more health care providers and system developers to scrutinize clinical requirements and establish a groundwork for the development of highly effective CDSSs.

Limitations

Although this study used a combination of structured interviews and questionnaires for assessment, inevitably, some subjective factors from the participants may have biased the results. The questionnaire content of this study was carefully designed based on the results of the interviews as well as the experience of the multidisciplinary team; however, the questionnaire content was unable to cover all aspects of knowledge and practice related to VT diagnosis. Although specific functions for a VT diagnosis CDSS were proposed, they have not been evaluated in a real-world setting. As our team is currently developing a VT CDSS with these functions, more rigorous studies will be conducted to support these findings in our future research.

Conclusions

This comprehensive analysis of VT CDSS requirements using a mixed methods approach identified specific knowledge and practice support requirements. The derived functions provide a foundation for further development and optimization of a CDSS. Moreover, it is important to tailor the CDSS to clinicians' specialization levels and years of practice for effective and personalized support.

Acknowledgments

This study was supported by National High Level Hospital Clinical Research Funding (grant 2022-GSP-GG-25), the Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences (CIFMS) (grant 2021-I2M-1-056), the Key R&D Program of Hunan Province (grant 2021SK2024), the Key Laboratory of Medical Information Intelligent Technology, and the National Key R&D Program of China (grants 2016YFC0901901 and 2017YFC0907503).

Data Availability

The data sets generated and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

ZH, MW, SZ, XX, JL, and YY designed the study. ZH, MW, ZZ, YY, SZ, XX, JL, and QG designed the questionnaire. ZH and ZZ collected the data. MW, ZH, and QG analyzed the data. ZH and MW drafted the manuscript. SZ, XX, JL, and YY critically revised the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Demographic characteristics and clinical experience of the interviewees.

[XLSX File (Microsoft Excel File), 9 KB - humanfactors_v11i1e55802_app1.xlsx]

Multimedia Appendix 2

Complete version of the questionnaire with the participant consent form.

[DOCX File , 979 KB - humanfactors_v11i1e55802_app2.docx]

References

1. Stecker EC, Reinier K, Marijon E, Narayanan K, Teodorescu C, Uy-Evanado A, et al. Public health burden of sudden cardiac death in the United States. Circ Arrhythm Electrophysiol 2014 Apr;7(2):212-217 [FREE Full text] [doi: 10.1161/CIRCEP.113.001034] [Medline: 24610738]
2. Feng X, Hai J, Ma Y, Wang Z, Tse H. Sudden cardiac death in mainland China: a systematic analysis. Circ Arrhythm Electrophysiol 2018 Nov;11(11):e006684. [doi: 10.1161/CIRCEP.118.006684] [Medline: 30571181]
3. Hua W, Zhang L, Wu Y, Liu X, Guo D, Zhou H, et al. Incidence of sudden cardiac death in China: analysis of 4 regional populations. J Am Coll Cardiol 2009 Sep 15;54(12):1110-1118 [FREE Full text] [doi: 10.1016/j.jacc.2009.06.016] [Medline: 19744622]

4. Empana J, Lerner I, Valentin E, Folke F, Böttiger B, Gislason G, ESCAPE-NET Investigators. Incidence of sudden cardiac death in the European Union. J Am Coll Cardiol 2022 May 10;79(18):1818-1827 [FREE Full text] [doi: 10.1016/j.jacc.2022.02.041] [Medline: 35512862]
5. Guandalini GS, Liang JJ, Marchlinski FE. Ventricular tachycardia ablation: past, present, and future perspectives. JACC Clin Electrophysiol 2019 Dec;5(12):1363-1383 [FREE Full text] [doi: 10.1016/j.jacep.2019.09.015] [Medline: 31857035]
6. Libby P, Bonow RO, Mann DL, Tomaselli GF, Bhatt DL, Solomon SD, et al. Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine, 12th edition. Amsterdam: Elsevier; 2021.
7. Roberts-Thomson KC, Lau DH, Sanders P. The diagnosis and management of ventricular arrhythmias. Nat Rev Cardiol 2011 Jun 22;8(6):311-321. [doi: 10.1038/nrcardio.2011.15] [Medline: 21343901]
8. Li L, Ding L, Zhang Z, Zhou L, Zhang Z, Xiong Y, et al. Development and validation of machine learning-based models to predict in-hospital mortality in life-threatening ventricular arrhythmias: retrospective cohort study. J Med Internet Res 2023 Nov 15;25:e47664 [FREE Full text] [doi: 10.2196/47664] [Medline: 37966870]
9. Hayashi M, Shimizu W, Albert CM. The spectrum of epidemiology underlying sudden cardiac death. Circ Res 2015 Jun 05;116(12):1887-1906 [FREE Full text] [doi: 10.1161/CIRCRESAHA.116.304521] [Medline: 26044246]
10. Dancy M, Camm A, Ward D. Misdiagnosis of chronic recurrent ventricular tachycardia. Lancet 1985 Aug 10;2(8450):320-323. [doi: 10.1016/s0140-6736(85)90363-0] [Medline: 2862480]
11. Spice C, Stewart K. Misdiagnosis of ventricular tachycardia. Lancet 1999;354(9196):2165. [doi: 10.1016/S0140-6736(05)77068-9] [Medline: 10609843]
12. Stewart K. Misdiagnosis of ventricular tachycardia. Lancet 1986 Nov 08;2(8515):1103. [doi: 10.1016/s0140-6736(86)90506-4] [Medline: 2877260]
13. Morady F, Baerman JM, DiCarlo LA, DeBuitleir M, Krol RB, Wahr DW. A prevalent misconception regarding wide-complex tachycardias. JAMA 1985 Nov 15;254(19):2790-2792. [Medline: 4057488]
14. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. Arch Intern Med 2005 Jul 11;165(13):1493-1499. [doi: 10.1001/archinte.165.13.1493] [Medline: 16009864]
15. Zeppenfeld K, Tfelt-Hansen J, de Riva M, Winkel BG, Behr ER, Blom NA, ESC Scientific Document Group. 2022 ESC Guidelines for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death. Eur Heart J 2022 Oct 21;43(40):3997-4126. [doi: 10.1093/eurheartj/ehac262] [Medline: 36017572]
16. Denekamp Y, Peleg M. TiMeDDx--a multi-phase anchor-based diagnostic decision-support model. J Biomed Inform 2010 Feb;43(1):111-124 [FREE Full text] [doi: 10.1016/j.jbi.2009.08.001] [Medline: 19665579]
17. Abedin Z. Differential diagnosis of wide QRS tachycardia: a review. J Arrhythm 2021 Oct 09;37(5):1162-1172 [FREE Full text] [doi: 10.1002/joa3.12599] [Medline: 34621415]
18. Katritsis DG, Brugada J. Differential diagnosis of wide QRS tachycardias. Arrhythm Electrophysiol Rev 2020 Nov 05;9(3):155-160 [FREE Full text] [doi: 10.15420/aer.2020.20] [Medline: 33240511]
19. Ray S, Jeremy R, Nishimura R, Simpson IA. Identifying excellence in contemporary cardiology practice: transparency, professionalism and the role of the professional society. Heart 2013 Aug 23;99(16):1144-1145. [doi: 10.1136/heartjnl-2012-303483] [Medline: 23343687]
20. Osheroff J, Teich J, Levick D, Saldana L, Velasco F, Sittig D, et al. Improving Outcomes With Clinical Decision Support: An Implementer's Guide. 2nd ed. Milton Park: Routledge; 2012.
21. Osheroff JA, Teich JM, Middleton B, Steen EB, Wright A, Detmer DE. A roadmap for national action on clinical decision support. J Am Med Inform Assoc 2007;14(2):141-145 [FREE Full text] [doi: 10.1197/jamia.M2334] [Medline: 17213487]
22. Miller RA. Diagnostic decision support systems. In: Berner ES, editor. Clinical Decision Support Systems: Theory and Practice. Cham: Springer International Publishing; 2016:181-208.
23. ISABEL. URL: https://www.isabelhealthcare.com/ [accessed 2024-01-30]
24. DxPlain. mghls. URL: http://www.mghlcs.org/projects/dxplain [accessed 2024-01-30]
25. Greenes RA, Bates DW, Kawamoto K, Middleton B, Osheroff J, Shahar Y. Clinical decision support models and frameworks: seeking to address research issues underlying implementation successes and failures. J Biomed Inform 2018 Feb;78:134-143 [FREE Full text] [doi: 10.1016/j.jbi.2017.12.005] [Medline: 29246790]
26. Khairat S, Marc D, Crosby W, Al Sanousi A. Reasons for physicians not adopting clinical decision support systems: critical analysis. JMIR Med Inform 2018 Apr 18;6(2):e24 [FREE Full text] [doi: 10.2196/medinform.8912] [Medline: 29669706]
27. Lu Y, Melnick ER, Krumholz HM. Clinical decision support in cardiovascular medicine. BMJ 2022 May 25;377:e059818. [doi: 10.1136/bmj-2020-059818] [Medline: 35613721]
28. Timpka T, Johansson M. The need for requirements engineering in the development of clinical decision-support systems: a qualitative study. Methods Inf Med 1994 May;33(2):227-233. [Medline: 8057951]
29. Alvarez R. Discourse analysis of requirements and knowledge elicitation interviews. 2002 Presented at: 35th Annual Hawaii International Conference on System Sciences; January 7-10, 2002; Big Island, HI. [doi: 10.1109/hicss.2002.994388]
30. Bright T. Transforming user needs into functional requirements for an antibiotic clinical decision support system: explicating content analysis for system design. Appl Clin Inform 2013;4(4):618-635 [FREE Full text] [doi: 10.4338/ACI-2013-08-RA-0058] [Medline: 24454586]

31. Schüttler C, Hinderer M, Kraus S, Lang A, Prokosch H, Castellanos I. Requirements analysis for a clinical decision support system aiming at improving the artificial nutrition of critically ill patients. Stud Health Technol Inform 2017;243:137-141. [Medline: 28883187]
32. Miller K, Mosby D, Capan M, Kowalski R, Ratwani R, Noaiseh Y, et al. Interface, information, interaction: a narrative review of design and functional requirements for clinical decision support. J Am Med Inform Assoc 2018 May 01;25(5):585-592 [FREE Full text] [doi: 10.1093/jamia/ocx118] [Medline: 29126196]
33. Jones EK, Banks A, Melton GB, Porta CM, Tignanelli CJ. Barriers to and facilitators for acceptance of comprehensive clinical decision support system-driven care maps for patients with thoracic trauma: interview study among health care providers and nurses. JMIR Hum Factors 2022 Mar 16;9(1):e29019 [FREE Full text] [doi: 10.2196/29019] [Medline: 35293873]
34. Schütze D, Holtz S, Neff MC, Köhler SM, Schaaf J, Frischen LS, et al. Requirements analysis for an AI-based clinical decision support system for general practitioners: a user-centered design process. BMC Med Inform Decis Mak 2023 Jul 31;23(1):144 [FREE Full text] [doi: 10.1186/s12911-023-02245-w] [Medline: 37525175]
35. Kabukye J, Koch S, Cornet R, Orem J, Hagglund M. User requirements for an electronic medical records system for oncology in developing countries: a case study of Uganda. AMIA Annu Symp Proc 2017;2017:1004-1013 [FREE Full text] [Medline: 29854168]
36. Chaffee B. Developing and assessing requirements for clinical decision support. Am J Health Syst Pharm 2003 Sep 15;60(18):1875-1879. [doi: 10.1093/ajhp/60.18.1875] [Medline: 14521039]
37. Shah H, Allard RD, Enberg R, Krishnan G, Williams P, Nadkarni PM. Requirements for guidelines systems: implementation challenges and lessons from existing software-engineering efforts. BMC Med Inform Decis Mak 2012 Mar 09;12(1):16 [FREE Full text] [doi: 10.1186/1472-6947-12-16] [Medline: 22405400]
38. Gulla J, Neri PM, Bates DW, Samal L. User requirements for a chronic kidney disease clinical decision support tool to promote timely referral. Int J Med Inform 2017 May;101:50-57 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.01.018] [Medline: 28347447]
39. Kannan V, Basit MA, Bajaj P, Carrington AR, Donahue IB, Flahaven EL, et al. User stories as lightweight requirements for agile clinical decision support development. J Am Med Inform Assoc 2019 Nov 01;26(11):1344-1354 [FREE Full text] [doi: 10.1093/jamia/ocz123] [Medline: 31512730]
40. Balatsoukas P, Williams R, Davies C, Ainsworth J, Buchan I. User interface requirements for web-based integrated care pathways: evidence from the evaluation of an online care pathway investigation tool. J Med Syst 2015 Nov 7;39(11):183. [doi: 10.1007/s10916-015-0357-5] [Medline: 26446014]
41. Notice on issuing ethical review measures for life sciences and medical research involving humans. National Health Commission, Ministry of Education, Ministry of Science and Technology, Bureau of Traditional Chinese Medicine. URL: https://www.gov.cn/zhengce/zhengceku/2023-02/28/content_5743658.htm [accessed 2024-03-20]
42. Fonteneau T, Billion E, Abdoul C, Le S, Hadchouel A, Drummond D. Simulation game versus multiple choice questionnaire to assess the clinical competence of medical students: prospective sequential trial. J Med Internet Res 2020 Dec 16;22(12):e23254 [FREE Full text] [doi: 10.2196/23254] [Medline: 33325833]
43. Bradburn NM, Sudman S, Wansink B. Asking Questions: The Definitive Guide to Questionnaire Design--For Market Research, Political Polls, and Social and Health Questionnaires. Hoboken, NJ: John Wiley & Sons; 2004.
44. Shivkumar K, Do DH. Catheter ablation of ventricular tachycardia: first, treat the underlying disease. J Am Coll Cardiol 2020 Oct 06;76(14):1657-1659 [FREE Full text] [doi: 10.1016/j.jacc.2020.08.046] [Medline: 33004130]
45. R: A language and environment for statistical computing (Version 4.0.3). URL: https://www.R-project.org/ [accessed 2023-10-21]
46. Zhang C, Shen L, Pan M, Zheng Y, Gu Z, Lin H. Perceptions and knowledge gaps on CHADS-VASc score components: a joint survey of Chinese clinicians and clinical pharmacists. Postgrad Med 2022 Jan 08;134(1):64-77. [doi: 10.1080/00325481.2021.1996815] [Medline: 34694951]
47. Azraii AB, Ramli AS, Ismail Z, Abdul-Razak S, Badlishah-Sham SF, Mohd-Kasim NA, et al. Validity and reliability of an adapted questionnaire measuring knowledge, awareness and practice regarding familial hypercholesterolaemia among primary care physicians in Malaysia. BMC Cardiovasc Disord 2021 Jan 19;21(1):39 [FREE Full text] [doi: 10.1186/s12872-020-01845-y] [Medline: 33468051]
48. Malhotra S, Masood I, Giglio N, Pruetz JD, Pannaraj PS. Current knowledge of Chagas-related heart disease among pediatric cardiologists in the United States. BMC Cardiovasc Disord 2021 Mar 02;21(1):116. [doi: 10.1186/s12872-021-01924-8]
49. Montagni I, Roussel N, Thiébaut R, Tzourio C. Health care students' knowledge of and attitudes, beliefs, and practices toward the French COVID-19 app: cross-sectional questionnaire study. J Med Internet Res 2021 Mar 03;23(3):e26399 [FREE Full text] [doi: 10.2196/26399] [Medline: 33566793]
50. Sturny L, Regard S, Larribau R, Niquille M, Savoldelli GL, Sarasin F, et al. Differences in basic life support knowledge between junior medical students and lay people: web-based questionnaire study. J Med Internet Res 2021 Feb 23;23(2):e25125 [FREE Full text] [doi: 10.2196/25125] [Medline: 33620322]

51. Sadarmin PP, Wong KCK, Rajappan K, Bashir Y, Betts TR. Questionnaire survey of cardiologists' knowledge, attitudes, and guideline application of implantable cardioverter defibrillator therapy. Pacing Clin Electrophysiol 2012 Jun 27;35(6):672-680. [doi: 10.1111/j.1540-8159.2012.03373.x] [Medline: 22452348]
52. Bergman HE, Reeve BB, Moser RP, Scholl S, Klein WM. Development of a comprehensive heart disease knowledge questionnaire. Am J Health Educ 2011 Mar 23;42(2):74-87 [FREE Full text] [doi: 10.1080/19325037.2011.10599175] [Medline: 21720571]
53. Shah S, Davies M, Quinn T. Management of acute coronary syndromes, a questionnaire survey of the clinical practice of cardiologists and other medical physicians belonging to west midland hospitals. Int J Cardiol 2005 Mar 10;99(1):71-75. [doi: 10.1016/j.ijcard.2003.11.043] [Medline: 15721502]
54. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. Lancet 2001 Mar 24;357(9260):945-949. [doi: 10.1016/S0140-6736(00)04221-5] [Medline: 11289364]
55. Patel VL, Arocha JF, Diermeier M, Greenes RA, Shortliffe EH. Methods of cognitive analysis to support the design and evaluation of biomedical systems: the case of clinical practice guidelines. J Biomed Inform 2001 Feb;34(1):52-66 [FREE Full text] [doi: 10.1006/jbin.2001.1002] [Medline: 11376543]
56. Trinkley KE, Kahn MG, Bennett TD, Glasgow RE, Haugen H, Kao DP, et al. Integrating the practical robust implementation and sustainability model with best practices in clinical decision support design: implementation science approach. J Med Internet Res 2020 Oct 29;22(10):e19676 [FREE Full text] [doi: 10.2196/19676] [Medline: 33118943]
57. Mittermaier M, Raza M, Kvedar JC. Collaborative strategies for deploying AI-based physician decision support systems: challenges and deployment approaches. NPJ Digit Med 2023 Aug 05;6(1):137. [doi: 10.1038/s41746-023-00889-6] [Medline: 37543707]
58. Brush JE, Sherbino J, Norman GR. Diagnostic reasoning in cardiovascular medicine. BMJ 2022 Jan 05;376:e064389. [doi: 10.1136/bmj-2021-064389] [Medline: 34987062]
59. Garg AX, Adhikari NKJ, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. JAMA 2005 Mar 09;293(10):1223-1238. [doi: 10.1001/jama.293.10.1223] [Medline: 15755945]
60. Mastrianni A, Sarcevic A, Hu A, Almengor L, Tempel P, Gao S, et al. Transitioning cognitive aids into decision support platforms: requirements and design guidelines. ACM Trans Comput Hum Interact 2023 Jun 10;30(3):1-28. [doi: 10.1145/3582431] [Medline: 37694216]
61. Sibbald M, Monteiro S, Sherbino J, LoGiudice A, Friedman C, Norman G. Should electronic differential diagnosis support be used early or late in the diagnostic process? A multicentre experimental study of Isabel. BMJ Qual Saf 2022 Jun 05;31(6):426-433 [FREE Full text] [doi: 10.1136/bmjqs-2021-013493] [Medline: 34611040]
62. Graber ML, Mathew A. Performance of a web-based clinical diagnosis support system for internists. J Gen Intern Med 2008 Jan 19;23(Suppl 1):37-40 [FREE Full text] [doi: 10.1007/s11606-007-0271-8] [Medline: 18095042]
63. Dr. Mayson. Humei. URL: http://www.huimei.com/specialist [accessed 2024-01-30]
64. Camacho J, Zanoletti-Mannello M, Landis-Lewis Z, Kane-Gill SL, Boyce RD. A conceptual framework to study the implementation of clinical decision support systems (BEAR): literature review and concept mapping. J Med Internet Res 2020 Aug 06;22(8):e18388 [FREE Full text] [doi: 10.2196/18388] [Medline: 32759098]
65. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020 Feb 06;3(1):17. [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
66. Roshanov PS, Fernandes N, Wilczynski JM, Hemens BJ, You JJ, Handler SM, et al. Features of effective computerised clinical decision support systems: meta-regression of 162 randomised trials. BMJ 2013 Feb 14;346(feb14 1):f657-f657 [FREE Full text] [doi: 10.1136/bmj.f657] [Medline: 23412440]
67. Campbell R. The five "rights" of clinical decision support. J AHIMA 2013 Oct;84(10):42-7; quiz 48. [Medline: 24245088]
68. Harrold LR, Field TS, Gurwitz JH. Knowledge, patterns of care, and outcomes of care for generalists and specialists. J Gen Intern Med 1999 Aug;14(8):499-511 [FREE Full text] [doi: 10.1046/j.1525-1497.1999.08168.x] [Medline: 10491236]
69. Fuster V, Hirshfeld JW, Brown AS, Brundage BH, Fye W, Lewis RP, et al. Working group 8: Defining the different types of cardiovascular specialists and developing a new model for training general clinical cardiologists. J Am Coll Cardiol 2004 Jul 21;44(2):267-271 [FREE Full text] [doi: 10.1016/j.jacc.2004.05.027] [Medline: 15261917]
70. Paydar P, Ebrahimpour S, Zehtab Hashemi H, Mohamadi M, Namazi S. Design, development and evaluation of an application based on clinical decision support systems (CDSS) for over-the-counter (OTC) therapy: an educational interventions in community pharmacists. J Adv Med Educ Prof 2023 Apr;11(2):95-104 [FREE Full text] [doi: 10.30476/JAMP.2022.95843.1661] [Medline: 37113680]
71. Shah C, Davtyan K, Nasrallah I, Bryan RN, Mohan S. Artificial intelligence-powered clinical decision support and simulation platform for radiology trainee education. J Digit Imaging 2023 Feb 24;36(1):11-16 [FREE Full text] [doi: 10.1007/s10278-022-00713-9] [Medline: 36279026]
72. Litvin CB, Davis KS, Moran WP, Iverson PJ, Zhao Y, Zapka J. The use of clinical decision-support tools to facilitate geriatric education. J Am Geriatr Soc 2012 Jun;60(6):1145-1149. [doi: 10.1111/j.1532-5415.2012.03960.x] [Medline: 22642270]

73. Strauss AT, James TW, Mathews SC. Fellow education improved through mobile clinical decision support application: a multi-center approach involving peri-procedural antithrombotic use. Gastroenterology 2018 Dec;155(6):2014-2015. [doi: 10.1053/j.gastro.2018.08.034] [Medline: 30165048]
74. Gong Y, Kang H. Usability and clinical decision support. In: Berner ES, editor. Clinical Decision Support Systems: Theory and Practice. Cham: Springer International Publishing; 2016:69-86.

Abbreviations

AHRA: Asian Heart Rhythm Association
ARVC: arrhythmogenic right ventricular cardiomyopathy
CAD: coronary artery disease
CDS: clinical decision support
CDSS: clinical decision support system
DDSS: diagnostic decision support system
ECG: electrocardiogram
ESC: European Society of Cardiology
SCD: sudden cardiac death
VT: ventricular tachycardia

Edited by A Kushniruk; submitted 25.12.23; peer-reviewed by H Younes; comments to author 25.01.24; revised version received 15.02.24; accepted 02.03.24; published 26.03.24.

Please cite as:
Hu Z, Wang M, Zheng S, Xu X, Zhang Z, Ge Q, Li J, Yao Y
Clinical Decision Support Requirements for Ventricular Tachycardia Diagnosis Within the Frameworks of Knowledge and Practice: Survey Study
JMIR Hum Factors 2024;11:e55802
URL: https://humanfactors.jmir.org/2024/1/e55802
doi:10.2196/55802
PMID:38530337

Original Paper

Usability of an Automated System for Real-Time Monitoring of Shared Decision-Making for Surgery: Mixed Methods Evaluation

Christin Hoffmann[1], BA, MSc, PhD; Kerry Avery[1], PhD; Rhiannon Macefield[1], PhD; Tadeáš Dvořák[1], MSci; Val Snelgrove[2], HNC; Jane Blazeby[1], MD; Della Hopkins[3], PhD; Shireen Hickey[4], MBChB; Ben Gibbison[1,5], MD; Leila Rooshenas[1], PhD; Adam Williams[3], MBChB; Jonathan Aning[3], BMedSci; Hilary L Bekker[6,7], PhD; Angus GK McNair[1,3], PhD; The ALPACA Study Team[3]

[1]National Institute for Health and Care Research Bristol Biomedical Research Centre, Bristol Centre for Surgical Research, Bristol Medical School: Population Health Sciences, University of Bristol, Bristol, United Kingdom

[2]Patient representative, Bristol, United Kingdom

[3]North Bristol NHS Trust, Bristol, United Kingdom

[4]Improvement Academy, Bradford Royal Infirmary, Bradford Teaching Hospitals NHS Foundation Trust, Bradford, United Kingdom

[5]University Hospitals Bristol and Weston NHS Foundation Trust, Bristol, United Kingdom

[6]Leeds Unit of Complex Intervention Development (LUCID), Leeds Institute of Health Sciences, School of Medicine, University of Leeds, Leeds, United Kingdom

[7]The Research Centre for Patient Involvement (ResCenPI), Department of Public Health, Aarhus University, Central Denmark Region, Denmark

Corresponding Author:
Christin Hoffmann, BA, MSc, PhD
National Institute for Health and Care Research Bristol Biomedical Research Centre, Bristol Centre for Surgical Research
Bristol Medical School: Population Health Sciences
University of Bristol
Canynge Hall
39 Whatley Road
Bristol, BS8 2PS
United Kingdom
Phone: 44 117 455 5993
Email: c.hoffmann@bristol.ac.uk

Abstract

Background: Improving shared decision-making (SDM) for patients has become a health policy priority in many countries. Achieving high-quality SDM is particularly important for approximately 313 million surgical treatment decisions patients make globally every year. Large-scale monitoring of surgical patients' experience of SDM in real time is needed to identify the failings of SDM before surgery is performed. We developed a novel approach to automating real-time data collection using an electronic measurement system to address this. Examining usability will facilitate its optimization and wider implementation to inform interventions aimed at improving SDM.

Objective: This study examined the usability of an electronic real-time measurement system to monitor surgical patients' experience of SDM. We aimed to evaluate the metrics and indicators relevant to system effectiveness, system efficiency, and user satisfaction.

Methods: We performed a mixed methods usability evaluation using multiple participant cohorts. The measurement system was implemented in a large UK hospital to measure patients' experience of SDM electronically before surgery using 2 validated measures (CollaboRATE and SDM-Q-9). Quantitative data (collected between April 1 and December 31, 2021) provided measurement system metrics to assess system effectiveness and efficiency. We included adult patients booked for urgent and elective surgery across 7 specialties and excluded patients without the capacity to consent for medical procedures, those without access to an internet-enabled device, and those undergoing emergency or endoscopic procedures. Additional groups of service users (group 1: public members who had not engaged with the system; group 2: a subset of patients who completed the measurement system) completed user-testing sessions and semistructured interviews to assess system effectiveness and user satisfaction. We conducted quantitative data analysis using descriptive statistics and calculated the task completion rate and survey response rate (system effectiveness) as well as the task completion time, task efficiency, and relative efficiency (system efficiency). Qualitative thematic analysis identified indicators of and barriers to good usability (user satisfaction).

Results: A total of 2254 completed surveys were returned to the measurement system. A total of 25 service users (group 1: n=9; group 2: n=16) participated in user-testing sessions and interviews. The task completion rate was high (169/171, 98.8%) and the survey response rate was good (2254/5794, 38.9%). The median task completion time was 3 (IQR 2-13) minutes, suggesting good system efficiency and effectiveness. The qualitative findings emphasized good user satisfaction. The identified themes suggested that the measurement system is acceptable, easy to use, and easy to access. Service users identified potential barriers and solutions to acceptability and ease of access.

Conclusions: A mixed methods evaluation of an electronic measurement system for automated, real-time monitoring of patients' experience of SDM showed that usability among patients was high. Future pilot work will optimize the system for wider implementation to ultimately inform intervention development to improve SDM.

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2023-079155

(JMIR Hum Factors 2024;11:e46698) doi:10.2196/46698

KEYWORDS

surgery; shared decision-making; patient participation; mixed methods; surgery; real-time measurement; patient-reported measure; electronic data collection; usability; data collection; patient reported; satisfaction; mobile phone

Introduction

Background

Contemporary health care puts patient-centered care at the heart of its delivery [1-4]. Shared decision-making (SDM) is a form of communication that promotes a dialogue between those involved in making health care choices. Therefore, treatment decisions are based on a shared understanding between patients and health care professionals of the evidence base for treatment and prognosis, patient values, preferences and beliefs, and clinical reasoning to personalize service delivery [5]. SDM is desired by patients and has become a key priority for health care systems globally [6-9]. Ensuring high-quality SDM when discussing and deciding treatments with patients can have many benefits, such as reduced information asymmetry or health service use [10,11]. It has been shown to contribute to good patient outcomes and satisfaction [12-15].

Globally, approximately 310 million operations are performed annually [16]. Surgery is often the only available treatment for a wide variety of minor and major medical conditions, and people increasingly choose surgical treatment (5.3% increase from 2009 to 2014 in the United Kingdom) [17]. Improving surgical patients' experience of SDM before surgery is particularly important because the effects of surgery are immediate and nonreversible. Patients cannot decide to discontinue treatment if the benefits fall short of expectations or side effects become unacceptable. Furthermore, making good surgical decisions may avoid negative impacts on health service costs (eg, through canceled operations) and patient outcomes [18-20].

Strategies aimed at improving SDM in complex health care settings can range from communication skills workshops for health care professionals [21] to educational videos [22] and booklets for patients [23]. However, their effects are mixed [14,15]. Systematic reviews of evidence to improve SDM conclude that achieving long-term change is likely to necessitate interventions that support the implementation of strategies at the organization, clinician, and patient levels [24-26]. However, there is uncertainty about how to realize change on a large scale across health care systems [27-33]. One recommended way to achieve this is through routine monitoring of patients' experience of SDM [34], but robust methods are lacking. Existing approaches to data collection are delayed, potentially affecting patients' accounts of their experience and impacting the ability to respond quickly and effectively before surgical treatments. Advances in technology mean that novel approaches to assessing patients' experiences of SDM can incorporate automated, electronic data capture close to the point of treatment consultations. This offers opportunities for providing information more accurately and in a timely manner, offering an effective way to develop interventions to improve SDM before surgery. Systems routinely collecting electronic patient-reported measure (ePRM) data in other contexts have been shown to improve care and outcomes for patients, including quality of life outcomes in pediatric dermatology [35] or symptom reporting in chronic kidney disease [36]. We developed a novel system to routinely monitor patients' experience of SDM automatically and in real time.

The evaluation of existing ePRM systems highlights the importance of user-friendly processes for their optimal performance [37-40]. Furthermore, the principles of good usability are important because they can be vital to the widespread uptake of ePRM systems by patients and their successful implementation in clinical practice [41-43]. Usability is an outcome defined as the extent to which the system can be used by specified users [44]. Several methods are available to evaluate measures of usability in health care [45-49]. A widely used framework contains standards set by the International Organization for Standardization (ISO) [50,51]. The guidelines recommend evaluating and optimizing the concepts of system effectiveness (the ability of participants to complete the survey), system efficiency (resources required to complete the questionnaire), and user satisfaction (subjective opinions of participants' experience with the measurement system) to achieve good usability.

Aim and Objectives

We aimed to examine the usability of a novel, automated, real-time measurement system to monitor surgical patients' experience of SDM. The specific objectives were to evaluate the measurement system's (1) effectiveness, (2) efficiency, and (3) user satisfaction among a large sample of surgical patients from a wide range of surgical specialties.

Methods

We used quantitative and qualitative methods to examine usability by evaluating the indicators and metrics related to system effectiveness, system efficiency, and user satisfaction. This study adhered to the ISO guideline 9241-11:2018 and followed recommendations for the usability testing of electronic patient-reported outcome measures [49,51].

Context and Setting

This study is part of a wider project to develop, pilot, and evaluate a decision-support intervention that uses real-time monitoring of patients' experiences to improve SDM (the ALPACA Study [52]). The project was initially set up as a quality improvement project at a large acute National Health Service (NHS) Trust in England, United Kingdom, which provides a range of acute and specialized clinical care services in South West England.

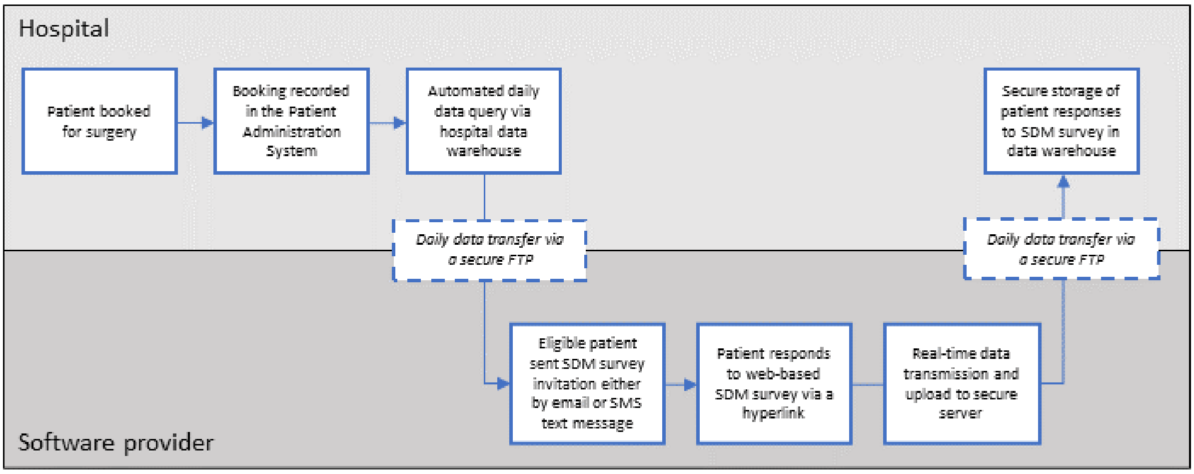
To facilitate automated, real-time data collection of patients' experience of SDM, a customizable off-the-shelf ePRM system (Cemplicity) was procured from a third-party software provider in March 2021. The software provider is an ISO 2001 certified, NHS-authorized ePRM provider, compliant with necessary accessibility and health data governance standards (eg, General Data Protection Regulations and Digital Technology Assessment Criteria). Before deployment and customization, the software provider tested the system development and design. Specifically, the prior rollout of the software across 6 countries and over 3000 health care institutions incorporated feedback from users across different health care settings and patients of diverse age groups, technology literacy, and health confidence. All measurement system interfaces are mobile optimized.

Customization for the purpose of this study was undertaken in collaboration with the software provider and included adapting the following: (1) the system's content and layout to include instruments to assess patients' experience of SDM and (2) data capture mechanisms to implement the system in the NHS Trust.

To assess patients' experience of SDM, 2 validated and widely used patient-reported measures were selected to measure SDM (CollaboRATE and SDM-Q-9). These were chosen by consensus within the study team, which was informed by a systematic review of SDM measurement instruments [53], national guidelines [25], and recommendations and use within the NHS clinical practice [34,54,55]. CollaboRATE is a 3-item instrument measured on a 10-point scale with answer options ranging from 0 ("no effort was made") to 9 ("every effort was made"). SDM-Q-9 consists of 9 items measured on a 6-point scale with answer options ranging from "completely disagree" to "completely agree." The measurement properties of both instruments have been demonstrated to be acceptable [56,57]. The measurement instruments were operationalized into a 12-question electronic survey format, branded to match the NHS Trust guidelines, and integrated into the patient-facing measurement system. Screenshots of the customized content are presented in Multimedia Appendix 1.

To implement the measurement system, secure data exchange processes were established between the software provider and the NHS Trust's information technology system and subsequently widened to various patient cohorts within the surgical departments. Specifically, SQL data queries were developed to identify and extract details of patients booked for surgery from the electronic patient record system that routinely records the patients' demographic and clinical information. The queries were designed to run automatically, securely transferring data from the hospital to the software provider on a daily basis. The 2 SDM measures were administered to patients upon being booked for surgery, with invitations sent either by email or SMS text messaging if no email address was available. Patient responses were received and processed using the measurement system. A reciprocal data feed securely returned response data to the hospital data warehouse for secure storage. A flow diagram of the measurement system process is provided in Figure 1.

Figure 1. Flow diagram of the process of automated real-time shared decision-making (SDM) monitoring through the measurement system. FTP: file transfer protocol.



Study Steering Group

A multidisciplinary study steering group was convened and consisted of a patient and public contributor, health care professionals, methodologists, social scientists, statisticians, and health services researchers. Regular meetings ensured the group's strategic oversight throughout and sought their input into the study design, research activities, and analyzing and interpreting results.

Patient and Public Involvement

We invited a patient and public contributor with lived experience of surgery to the study steering group, which was set up as part of the wider project. The input was sought from the patient and public contributor as appropriate throughout the study (eg, review of patient-facing materials, including survey invitation and instructions, and interim findings from qualitative analyses). In addition, we organized a patient and public advisory meeting which 6 public contributors attended for 1 hour via a Zoom (Zoom Video Communications, Inc) meeting. The aim of the meeting was to obtain patient and public perspectives on the overall project plan and its key challenges. The topics discussed included recruitment, acceptability, and satisfaction with the measurement system, which informed the design aspects of this study.

Usability Concepts

The usability of the measurement system was examined by evaluating metrics and indicators relevant to 3 concepts, including system effectiveness, system efficiency, and user satisfaction. The definitions are summarized in Textbox 1, and the details of their assessment are described subsequently.

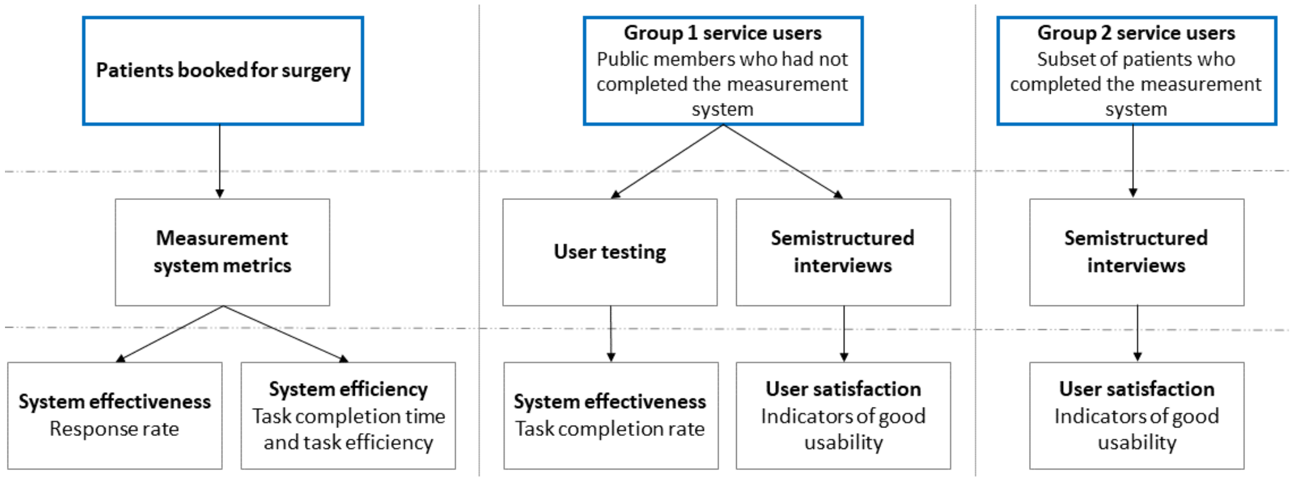
Textbox 1. Definitions of usability concepts.

- System effectiveness: the ability of participants to perform tasks to achieve predetermined goals completely and accurately, without negative consequences (eg, poor layout of the system interface leading to participants missing or accidentally selecting system options) [36,49-51].
- System efficiency: the amount of participant resources required to achieve the prespecified goals [49,58].
- User satisfaction: the subjective opinions of the participants based on their experience of interacting with the system [49]. This includes any subjective reports about likes, dislikes, and recommendations for changes [51].

Participants and Procedures

We used multiple cohorts of participants and procedures for quantitative and qualitative data collection for this study. Figure 2 illustrates the different cohorts of participants and provides an overview of the data collection procedures used to evaluate the usability concepts.

Figure 2. Overview of participant cohorts, data collection procedures, and usability concepts.



Participants and Recruitment

To obtain quantitative measurement system metrics to assess system effectiveness and efficiency (refer to the Quantitative Analysis section for further details), automated, real-time data collection was conducted between April 1 and December 31, 2021, and rolled out across 7 surgical departments: orthopedic, urology, gynecology, neurosurgery, gastrointestinal, vascular and breast. We included adult patients booked for elective surgery in these 7 specialties. Patients aged <18 years, those without the capacity to consent for medical procedures, those undergoing emergency and endoscopic procedures, and those without access to an appropriate internet-enabled device (ie, mobile phone, smartphone, PC, tablet, or similar device) were excluded.

We recruited 2 further groups of service users for user testing and interviews to obtain quantitative and qualitative data to assess system effectiveness and user satisfaction.

Group 1 participants were individuals who had not engaged with the measurement system before user-testing sessions to ensure naive user interactions [59] (refer to the User Testing section for detailed user-testing methods). Service users with experience of surgery were recruited through patient experience panels within 2 NHS Trusts (North Bristol NHS Trust and Bradford Teaching Hospitals NHS Foundation Trust). A panel coordinator identified and approached potential participants via email containing a recruitment advertisement. Sampling was purposive to achieve the maximum possible variation in recognized protected characteristics (eg, sex, disability, and race) and experience of surgery.

Group 2 participants were individuals who had engaged with the measurement system to explore user satisfaction after interacting with the system (refer to the *Semistructured Interviews* section for more details). These were a subset of eligible patients who completed the measurement system. A member of staff with authorized access to the patient administration system and patient response data stored in the data warehouse recruited participants via telephone. We used a purposive sampling strategy to achieve variation in characteristics, including age, ethnicity, sex, type of surgery received, and experience of good or bad SDM (identified through survey responses).

Procedures

Measurement System Metrics

Relevant metrics automatically collected by the measurement system were used to examine usability quantitatively (eg, responses to questionnaire items and timestamps for starting and submitting the survey). Unique entries were recorded for each patient who received the invitation to complete the measurement system. Entries and corresponding data collected between April 1 and December 31, 2021, were available for analysis.

User Testing

Postdeployment user-testing sessions were conducted between June and December 2021 and were performed in a simulated environment.

Group 1 participants were invited to participate in a one-to-one 1-hour videoconference via Zoom with a researcher to complete the measurement system live. Sessions began by reminding participants about the aim and the process of user testing the measurement system. Service users were then sent an SMS text message or email invitation (depending on their preference) that included a test link to the survey. Specific user-testing links were set up to allow simulated completion of the measurement system (ie, responses were not used for live response data). Sessions assessed system effectiveness (including any issues related to system functionality or completion). A concurrent think-aloud technique was applied to vocalize reactions and thinking processes [60-62], supplemented with observational notes of any difficulties encountered [63,64].

User-testing sessions were conducted by 1 member of the study team who had experience in think-aloud methods (AGKM or CH). A topic guide was developed to guide conversations (Multimedia Appendix 2). Sessions were audio recorded and transcribed using unique identifiers to ensure anonymity. Field notes and any problems during the measurement system completion were recorded in a table using Excel (Microsoft Corp).

Semistructured Interviews

We conducted semistructured, in-depth interviews using retrospective probing to explore the service users' views about the indicators of usability of the measurement system [65].

Interviews were conducted with group 1 participants following user testing via the same web-based videoconferencing software. Group 2 participants were invited to take part in an approximately 30- to 45-minute phone or videoconferencing call (according to their preference) during which they reflected on completing the measurement system. The conversations followed a previously tested and refined topic guide that was based on standard usability concepts [51]. An example topic guide can be found in Multimedia Appendix 2.

Interviews were performed by either of the 2 researchers (AGKM or CH), audio recorded, and anonymized during transcription.

Analysis

Quantitative Analyses

All quantitative analyses were performed by 3 researchers (TD, AGKM, and CH) using the statistical software package STATA (version 16.0; StataCorp LLC).

System Effectiveness

We assessed system effectiveness by calculating the user task completion rate based on usability testing sessions and the survey response rate based on measurement system metrics [66].

The user task completion rates were calculated as a percentage of tasks completed by the total number of tasks. A process map was created defining the number and type of tasks (or steps) required to complete the measurement system. Successful completion means that all tasks were completed without user errors. User errors were deviations or problems encountered that interfered with successful task completion. Noncritical errors were defined as those that were successfully addressed by the testers themselves following instructions from the observer. Critical errors were those that required the observer to intervene or take remedial actions.

The survey response rate was calculated as a percentage (number of completed surveys/number of patients invited × 100). Surveys were considered complete when responses to all 3 items of the CollaboRATE measure and at least 7 out of 9 items of the SDM-Q-9 measure were returned.

System Efficiency

We assessed system efficiency by calculating the task completion time and task efficiency based on measurement system metrics [58,66].

The task completion time was defined as the time participants took from the first activity (starting the survey by following the hyperlink) to the last activity (submission of the survey). Task efficiency was defined as the time spent to complete each task (timestamps were recorded in the following format: hh:mm:ss). Analyses were based on those who completed the measurement system for whom typical first and last activity timestamps were available (ie, atypical timestamps were those with no recorded activity time). Extreme outliers were excluded because the system allowed service users to leave and later return to the survey and continue submission (eg, the next day or the following week). These were defined as those with the task completion time >3 times the IQR [67].

Qualitative Analyses of User Satisfaction

User satisfaction was assessed by evaluating service users' self-reported experiences of using the system through user-testing sessions and semistructured interviews [66]. Discussions explored perceptions of usability aspects, including service users' interpretation of the system's ease of use and navigation, their satisfaction with instructions and visual display, and the likelihood of using the system again or recommending it to others.

Transcripts obtained from user-testing sessions and semistructured interviews were uploaded to the qualitative data management software NVivo (version 20.5.1; QRS International) and analyzed using thematic analysis [68]. This involved systematic coding of data to identify commonly mentioned concepts within the data and to develop themes and subthemes. The coding was inductive and iterative and followed predefined steps of data familiarization, generation of initial codes, and searching for themes. Coding was performed by 2 researchers independently who met regularly to review themes. Analysis and interpretation of qualitative data were further supported in 2 ways. First, a report of the interim findings was produced and discussed with the wider multidisciplinary steering group. Second, the presentation of interim findings to the patient and public advisory group sought further input.

Ethical Considerations

This study was part of a project spanning quality improvement and research. Therefore, it was subject to 2 governance processes requiring separate approvals. Monitoring patients' experience of SDM in routine clinical practice was initially approved through a quality improvement proposal at North Bristol NHS Trust (reference: Q80008). This was then incorporated into a larger program of work, where all processes were approved through the appropriate governance framework (Consent and SDM Program Board, reporting to the Clinical Effectiveness and Audit Committee). Ethics approval for conducting interviews with NHS patients was granted by the NHS Health Research Authority North West – Liverpool Central Research Ethics Committee (reference: 21/PR/0345). Participants provided electronic consent through a link to a secure data management platform (version 11.1.18, REDCap [Research Electronic Data Capture]; Vanderbilt University) [69] before any study activity commenced.

Results

Participants and Procedures

A total of 5794 surgical patients received invitations to complete the survey and for whom unique entries were recorded in the measurement system. Of these, 2254 returned the completed surveys (refer to Table 1 for patient characteristics) and provided data for the analysis of measurement metrics.

Table 1. Characteristics of patients who completed the measurement system (N=2254).

Characteristics	Patients, n (%)
Sex	
Female	1243 (55.15)
Male	1011 (44.85)
Age group (y)	
<29	170 (7.54)
30 to 39	213 (9.45)
40 to 49	277 (12.29)
50 to 59	529 (23.47)
60 to 69	555 (24.62)
70 to 79	402 (17.83)
≥80	108 (4.79)
Ethnicity[a]	
Other	104 (4.61)
White British	977 (43.35)
Specialty	
Breast	278 (12.33)
Colorectal	67 (2.97)
General	194 (8.61)
Gynecology	106 (4.7)
Neuro	288 (12.78)
Trauma, orthopedics and spinal	555 (24.62)
Upper gastrointestinal	41 (1.82)
Urology	584 (25.91)
Vascular	141 (6.26)

[a]Missing data: n=1173.

A total of 25 service users (group 1: n=9; group 2: n=16) participated in user-testing sessions and semistructured interviews.

In group 1, a total of 9 service users completed 8 user-testing sessions. Most sessions were completed on a one-to-one basis (7/9, 78%). One session was completed with 2 participants, which included 1 service user with disability and their caregiver who provided additional support. All sessions were held via videoconference and lasted for an average duration of 43 (SD 15.1; range 29-78) minutes. Service users in this group were mostly female participants (6/9, 67%) and self-identified as Asian (1/9, 11%), other White background (1/9, 11%), and White British (7/9, 78%). Details about the surgical experience were known for 4 service users who represented orthopedic (2/4, 50%), upper gastrointestinal (1/4, 25%), and ophthalmic (1/4, 25%) specialties.

In group 2, 16 service users completed semistructured interviews between June and November 2021. Most interviews were conducted via telephone (15/16, 94%), with 1 (6%) interview conducted via videoconference, lasting for an average duration of 36 (SD 9.9; range 21-50) minutes. Most service users in group 2 were female participants (10/16, 62%) and were 51 (SD 15.8) years on average. All participants were from a White British background (16/16, 100%). Efforts were made to recruit participants from a wide range of ethnic minority backgrounds; however, due to a large amount of missing data (Table 1), this was unsuccessful. The characteristics of group 2 participants are presented in Table 2.

Table 2. Characteristics of group 2 service users (n=16).

Characteristics	Service users, n (%)
Age (y), mean (SD; range)	51 (15.8; 23-80)
Sex, n (%)	
Female	10 (62)
Male	6 (38)
Ethnicity, n (%)	
White British	16 (100)
Surgery type, n (%)	
Breast	3 (19)
Colorectal	2 (13)
General	2 (13)
Gynecology	1 (6)
Trauma and orthopedics	2 (13)
Urology	5 (31)
Vascular	1 (6)

Usability Concepts

System Effectiveness

A process map to assess task completion contained 19 tasks (or steps) required to complete the measurement system. Tasks ranged from “Open text message/email” to “Click on ‘Submit’” and are detailed in Multimedia Appendix 3.

A total of 171 tasks across 8 user-testing sessions were submitted by all 9 group 1 participants. One service user reported 2 noncritical errors across 2 tasks when completing the measurement system using a mobile phone. The first error occurred following task 1 “Open text message.” This forced an additional step to resolve a pop-up notification which prompted the service user to select an internet browser to open the survey link. The second error occurred following task 5 “Select response to question 1.” The displayed answer options for CollaboRATE item 1 were cut off at 8, not presenting answer option 9 (every effort was made). Further scrolling was required by the service user to be able to select the answer option 9. Both noncritical errors were managed and resolved without requiring observer input. Consequently, a total completion rate of 98.8% (169/171) was achieved. No critical errors or failures in completing the tasks were reported.

The survey response rate was 38.9% (2254 completed surveys/5794 patients invited × 100).

System Efficiency

Out of the 2254 responses available, 1106 (49.07%) were excluded from analysis. These 1106 responses included 719 (65.01%) responses with an atypical timestamp (ie, no activity time was recorded because the timestamp for the first and last activity was 00:00:00, which was identified as a technical issue and rectified by the software provider) and 387 (34.99%) responses identified as extreme outliers (ie, the task completion time was >12 min). Assessment of the completion time of 1148 (50.93%) of the 2254 responses showed that service users required an average median duration of 3 (IQR 2-4) minutes to complete the measurement system. Calculations of task efficiency showed that the average median time taken per task was 9 (IQR 6-13) seconds.

User Satisfaction

Analysis of qualitative data from user-testing sessions and semistructured interviews with a subset of patients revealed four main themes related to user satisfaction as follows: (1) acceptability, (2) ease of access to the system, (3) ease of use, and (4) satisfaction with the measurement system.

Acceptability

Indicators of Good Acceptability

Service users who were interviewed as part of the qualitative data collection frequently commented on the low burden of completing the measurement system, suggesting good acceptability among the participants. This was mainly because of the low number of questions contributing to the measurement system being considered quick and straightforward to use:

> *Short survey, key thing—not too much of your time.* [PT9, group 1]
>
> *I did it from my phone so yes it was very straightforward.* [PT13, group 2]
>
> *I don’t remember feeling any burden [...], it was quite easy.* [PT19, group 2]
>
> *I don’t think it seemed too long. It was enough. To be honest, if it had been a lot more, I probably wouldn’t bother to do it.* [PT21, group 2]

Furthermore, service users highlighted the common use of web-based surveys to obtain feedback in health care and other general settings. Therefore, they felt a certain level of familiarity with the measurement system, which contributed to the good acceptability:

> *I thought it was, I mean, pretty standard, you know, arial buttons, nought to ten on how much you disagree, agree, disagree to something so yes, familiar with many other surveys that I've seen before.* [PT12, group 2]

Potential Barriers to Acceptability

Some barriers to completing the measurement system were highlighted. For example, participants mentioned that the service users may easily ignore or forget to complete the measurement system as follows:

> *It's easy not to [complete the measurement system], I've had them from places, not about health or anything important like that, but it's easy just [to] think, "Oh, I'll do that later," and then never go back to it.* [PT7, group 1]

Another example included concerns about the number of SMS text messages and surveys received from other sources and the cumulative burden:

> *I mean the good thing about it is it's simple and easy and you just get the nudge, but on the other hand there are lots of other nudges coming through at you.* [PT19, group 2]

This contributed to a small number of service users questioning the credibility of the invitation to complete the measurement system:

> *Something came through via email which to be honest I wasn't sure if it was a genuine thing or if it was something else.* [PT10, group 2]

Solutions to Address Barriers

Service users were asked about the usability of solutions to address these issues and included support for reminder emails:

> *One follow-up is a good idea but not more than one possibly because then people start to feel a bit harassed, but I think a second one is a good idea because of the forgetting thing and they go oh yeah, I'll do it this time.* [PT4, group 1]

Service users thought that the use of email would address this problem for some service users:

> *It is at the top of my email pile again, I'd better do it, so it jogs your memory, texts don't do that, it's a very momentary thing, text messaging.* [PT13, group 2]

Furthermore, service users suggested to increase the personal relevance and awareness of the measurement system:

> *You've got to feel that you're going to benefit, and it's really relevant to you, for you to have the interest to do it.* [PT7, group 1]

> *They [service users] really, really need to know it's coming because I don't know about you but we're very, very careful what we open and if this just appeared with no warning I wouldn't open it.* [PT4, group 1]

Service users mentioned the need to highlight the brevity of the measurement system and the low number of questions:

> *There are people who will fill them in if they're told it's very short, which is why it's important that it says it's short.* [PT9, group 1]

> *I think sometimes if you open one you can see that it's 100 questions you just think I probably won't do that.* [PT14, group 2]

Ease of Access to the System

Indicators of Good Ease of Access to the System

All service users were able to access the measurement system without problems and commented on its ease of access through both methods, email and SMS text message:

> *I think most people nowadays are comfortable with computers and technology.* [PT14, group 2]

Some service users expressed a preference for using either email or their phone to complete the measurement system. However, there was no conclusive evidence to suggest the superiority of either email or SMS text message:

> *I guess that for me making it [come to my phone] makes it more accessible 'cos you don't have to go in your emails. It automatically comes through and you can do it at any time and reply at any time, so you can do it when it's convenient to you and its literally just a text on your phone.* [PT1, group 1]

> *Although I use a smart phone quite a lot, sometimes it's difficult to manipulate it, whilst a laptop I find much more easier to use.* [PT8, group 1]

Furthermore, service users commented on the good comprehensibility and legibility of the content, contributing to good levels of ease of access to the system. For example, comments included that there was a sufficiently large font option for those who required or preferred larger screens:

> *I think the presentation of it on my phone, and I don't have a large phone, I just have a small phone, I could read all that quite easily.* [PT7, group 1]

> *They were really easy to understand[...] The questions were very clear, I thought they were quite well[...] focused and well explained.* [PT24, group 2]

Potential Barriers to Access to the System

Some service users expressed concerns regarding the system's ease of access for certain population groups. Most frequently, concerns were raised in connection with older adults and lack of access to technology. Furthermore, considerations included the ease of access to the measurement system for non–English-speaking service users and those with disabilities:

> *There's also a certain cohort would be using online. [...] I do think people will miss out but if it's just being pinged… whether it's on text or email* [PT3, group 1]

> *People that English isn't their first language, that could be a bit of a consideration.* [PT5, group 1]

Solutions to Address Barriers

The most frequently mentioned solutions were common alternatives to electronic data collection in connection with support measures for questionnaire completion:

> *I mean there's probably still a gap with the older generation who wouldn't be comfortable doing it, and would prefer doing it via communication of phone or in written format.* [PT14, group 2]

Ease of Use

Most often, the simplicity of the system was highlighted in connection with the ease of completing the measurement system. Furthermore, the ease of use was often attributed to the brevity of the measurement system:

> *I actually thought it was quite simple and quite straightforward and easy.* [PT3, group 1]

> *That has been perfectly straightforward, for someone who's not very IT literate, that was all fine.* [PT7, group 1]

> *Yeah, that was very easy, it didn't take very long [...] I remember it did seem simple* [PT23, group 2]

Moreover, most service users commented on the visual display, which was perceived as appealing and very clear. The clear layout of the survey contributed to high comprehensibility among participants:

> *It is very clear and also I quite like the bold type. [...] very clear again and very easy to read.* [PT3, group 1]

> *It's pretty obvious straight off of that where the survey has come from including the logo and almost like the colours of the survey match with the NHS logo [...] I think that part of it makes it really easy.* [PT21, group 2]

> *Yeah, that's laid out really spaced out and easy to read.* [PT6, group 1]

Service users frequently mentioned the ease of navigation and thought it was "basic and straightforward" (PT20, group 2). Others mentioned further details regarding what they liked about the navigation:

> *There is no need to zoom in or zoom out or move around a page or click buttons to find the survey so I think all of that aspect is really easy. [...] It's easy to use and the agree or disagree buttons are really straight to the point.* [PT21, group 2]

One service user also commented on the loading speed of the survey page:

> *I think it's easy to use because it doesn't take long to load which I think is important.* [PT20, group 2]

No service user raised concerns that could be considered barriers to the ease of use of the measurement system.

Overall Satisfaction With the Measurement System

All service users provided positive feedback regarding the abovementioned themes of acceptability, ease of access to the system, and ease of use, which indicated high satisfaction with the measurement system. General supportive comments were made throughout the user-testing sessions and semistructured interviews:

> *Yeah, absolutely brilliant. I'll give that 11 out of 10. [...] Somebody who designed this did a good job.* [PT5, group 1]

All respondents agreed when asked whether they are likely to complete the measurement system again:

> *Yeah, I would definitely respond to it again.* [PT6, group 1]

In addition, there were unprompted comments related to satisfaction with particular features. For example, service users pointed out that they particularly liked the "back buttons" to return to previous questions, the option to pause the measurement system and return at a different time, and the fact that there are contact details of the hospital in case this survey was received in error:

> *You've got the option, you can go back and change something, or if there was something you were worried about that you've done, it's clear that you can go back.* [PT7, group 1]

Discussion

Principal Findings

This study examined the usability of a novel automated and real-time ePRM system to monitor patients' experience of SDM in routine clinical practice. We used a large sample from a diverse range of surgical specialties to evaluate system effectiveness, system efficiency, and user satisfaction.

Overall, the evaluation of the measurement system demonstrated good usability. Metrics relevant to the effectiveness and efficiency showed that the system can be used without problems and completed quickly. The results from qualitative testing sessions and interviews with 25 service users showed that the measurement system has good user satisfaction. It was perceived as acceptable, easy to access, and easy to use. Service users identified potential barriers to acceptability and ease of access to the system, which can inform strategies for the optimization of the measurement system.

Limitations

This study has certain methodological limitations. First, we purposively selected participants to include individuals from a wide socioeconomic background with varying computer literacy skills. While this study exceeded the recommended sample size for usability testing [70-72], service users in our sample were primarily White British (23/25, 92%), English-speaking adults with capacity to consent for medical treatments, and from specific geographic areas of the United Kingdom (West, South West, and North East England). This may limit the generalizability of the study findings. It is uncertain whether the inclusion of more participants from more diverse backgrounds would have elicited different perspectives on the measurement system. Second, only patients who had completed the measurement system were eligible to participate in semistructured interviews. Data protection regulations limited our ability to recruit individuals who had not completed the survey. Therefore, we were unable to explore whether nonengagement with the system was due to reasons related to usability not mentioned by the study participants. Barriers to engagement may align with the themes identified during semistructured interviews, which are partly addressed by ongoing work (refer to the following section). Separately, there is ongoing work which includes conducting follow-up phone calls with patients to explore the reasons for nonengagement. Third, usability may also be evaluated using validated measurement instruments to capture quantitative measures of individuals' perception of usability from a larger, representative sample size [73,74]. This study did not include such measures in addition to the ePRM to avoid distorting usability outcomes. For example, the additional length of the survey may have affected system efficiency and impacted perceptions of ease of use. Instead, we included a range of methods to assess usability to triangulate the data sources [75].

Comparison With Prior Work

Existing research has investigated optimal strategies and methods for collecting ePRMs [40,76-79]. The usability evaluation of electronic platforms is common and has been fundamental in optimizing systems to collect ePRMs across a range of health care settings [80] and also within surgery [81,82]. Less is known about systems that monitor patients' experiences automatically and in real time. We are aware of only 1 recently published protocol describing a similar measurement system [83], but we were unable to identify studies with specific relevance to surgery or SDM. Our study addresses this gap and provides insights into the usability of an automated measurement system that monitors ePRMs for SDM in real time. The measurement system in our study was evaluated for service users undergoing surgical treatment; however, the findings may be applicable to other health care settings.

Evidence of good usability of an automated measurement system that captures surgical patients' experiences in real time supports the measurement systems' potential for scalability. The use of the system is recommended in similar health care settings where policy makers or official bodies wish to audit or monitor patients' experiences of SDM or aim to inform interventions to improve SDM before treatment. System effectiveness and efficiency are central components to service users' successful interaction with any system [51]. The usability concepts evaluated have been shown to be key in other systems rolled out in surgical departments [84] and are likely to play a role in the wider adoption of the measurement system [85]. This study showed that service users were able to successfully complete the measurement system and that they required little time and effort to do so. In addition, good user satisfaction is vital to a system's sustainability and is used as a measure of the success of digital information systems within health care organizations worldwide [86-88]. User satisfaction with ePRM systems and perceived acceptability, in particular, have been shown to be key to their uptake among stakeholders [89,90]. The qualitative evidence obtained from service users in this study demonstrated good acceptability, ease of access to the system, and ease of use, which suggests low concern regarding user satisfaction. Some steps to optimize the system to address identified usability concerns and adapt SDM measurement to other care contexts [91] might be necessary before a wider rollout to other health care settings.

This study highlighted well-known barriers to ease of access to electronic measurement systems [92,93]. Specifically, literacy with electronic systems can be lower in older and frail adults and among individuals without capacity to consent [94-97]. While the measurement system response rate in this study (2254/5794, 38.9%) was notably higher compared to those reported in other studies evaluating measurement systems (eg, 18% in the study by Iversen et al [98], 20% in the study by Bliddal et al [99], or 30% in the study by Arner [82]), it may be indicative of such barriers experienced by surgical patients. The solutions to improve ease of access identified in this study include additional paper-based methods. Furthermore, barriers may be overcome through assisted data collection using a tablet computer at the point of care [100]. Additional resources may be required to ensure full and accurate data capture for adults without capacity to consent to medical treatments completing the measurement system [101]. Similarly, language barriers have been shown to affect service users' ease of access to the system and the quality of responses to ePRM systems [93,102]. Translating content can be key to addressing such language barriers, as demonstrated by widely used quality of life measures [103]. Further work is currently ongoing to address relevant issues to maximize inclusivity (ISRCTN [International Standard Randomised Controlled Trial Number] registry ID: 17951423). Specifically, this line of work seeks to explore the views of underserved groups (eg, limited income, older age, and ethnic minority groups) using qualitative methods to understand how the use of the system and future intervention development can be optimized to maximize inclusivity. This work will consider nondigital materials, translation of study materials, measurement system content, and measurement instruments using appropriate guidance [104] and will include non-English qualitative data collection. Detailed methods will be reported in a separate publication.

High-quality SDM can be a moderator and mediator of health and care quality [105], addressing the challenges of true patient-centered care (eg, reducing asymmetry in medical knowledge between patients and surgeons and addressing issues of individual preferences). To improve patients' experiences of SDM before surgery, additional intervention development work is needed to complement automated, real-time monitoring of SDM experiences. Evidence from other clinical settings suggests that interventions, including real-time feedback, in addition to routine monitoring of ePRMs, can lead to improvements in outcomes or clinical performance [81,106-108]. This study demonstrated the good usability of a measurement system that automatically collects, stores, and retrieves ePRM data and is ready to provide feedback on this information in digital format near to real time. This suggests that the system is ready to provide instantaneous feedback on surgical patients' experience of SDM to clinical teams, which has the potential to improve SDM. Future work will explore the optimal design and feasibility of feedback mechanisms and examine the acceptability of the system. Refinements to optimize the usability and inclusivity of the system are required before evaluating the effectiveness of an intervention to improve SDM. Key to this work will be obtaining wider perspectives from other stakeholders involved in the intervention (eg, health care professionals and stakeholders from the lower-income, ethnic minority, and older age groups). In the long term, strategies to facilitate the implementation of the measurement system in routine clinical care will be investigated and evaluated using evidence-based approaches to intervention design [109].

Conclusions

We examined the usability of a measurement system for automated and real-time ePRM collection to monitor patients' experience of SDM in a large sample using 2 brief, validated instruments. The findings suggest good usability and support scalability of the measurement systems to other secondary health care institutions and will inform its optimization. Complementary work is currently exploring the feasibility and acceptability of monitoring and feedback experience of SDM with patient and professional stakeholders. Future implementation and formal evaluation of the measurement system will be performed to establish whether routine monitoring and feedback of patients' experiences has the potential to improve SDM for surgical patients.

Acknowledgments

The ALPACA Study was funded by the National Institute for Health and Care Research (NIHR) Program Development Grant (NIHR205174). Furthermore, this study was supported by an NIHR Clinician Scientist award to AM (NIHR CS-2017-17-010) and by the NIHR Biomedical Research Centre at the University Hospitals Bristol and Weston National Health Service (NHS) Foundation Trust and the University of Bristol (BRC-1215-20011, NIHR203315). TD was funded by the NIHR Academy through an NIHR Methodology Internship Scheme. The quality improvement project was supported by the North Bristol NHS Trust and the "Next Big Thing Award," funded by the Southmead Hospital Charity. The views expressed in this publication are those of the authors and do not necessarily reflect those of the North Bristol NHS Trust, the UK NHS, or the NIHR. The authors would like to thank Omaid Ahmadyur for supporting the recruitment of the participants.

Authors' Contributions

AGKM and JB developed the idea for this study with input on the conceptualization from HLB, KA, and RM. CH and AGKM prepared the protocol, which was reviewed, discussed, and approved by all coauthors. CH, AGKM, KA, and RM established and formulated the methods for this review, with inputs from JB and VS. Recruitment and data collection were undertaken by CH, DH, SH, and AGKM. Data analysis was performed by CH and TD under the guidance of AGKM and RM. AGKM provided general oversight for this study. AGKM, SH, DH, and JB will take the lead on implementing the findings from this study with the help of CH, VS, and the ALPACA Study Team. The ALPACA Study Team members are: Andy Judge, Andrew Smith, Archana Lingampalli, Barnaby Reeves, Jessica Preshaw, Michael R Whitehouse, Paul Cresswell, Philip Braude, Shelley Potter, Timothy Beckitt, and Timothy Whittlestone.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Example screenshots of the measurement system.
[PDF File (Adobe PDF File), 1154 KB - humanfactors_v11i1e46698_app1.pdf]

Multimedia Appendix 2
Topic guide.
[PDF File (Adobe PDF File), 191 KB - humanfactors_v11i1e46698_app2.pdf]

Multimedia Appendix 3
Process map of tasks.
[PDF File (Adobe PDF File), 89 KB - humanfactors_v11i1e46698_app3.pdf]

References

1. The NHS constitution for England. Department of Health and Social Care. 2021. URL: https://www.gov.uk/government/publications/the-nhs-constitution-for-england/the-nhs-constitution-for-england#introduction-to-the-nhs-constitution [accessed 2022-09-15]
2. The NHS long term plan. National Health Service. 2019. URL: https://www.longtermplan.nhs.uk/ [accessed 2024-03-21]
3. Good surgical practice. Royal College of Surgeons. 2013. URL: https://www.rcseng.ac.uk/standards-and-research/gsp/ [accessed 2024-03-21]

4. Decision making and consent - ethical guidance. General Medical Council. 2020. URL: https://www.gmc-uk.org/professional-standards/professional-standards-for-doctors/decision-making-and-consent [accessed 2024-03-21]
5. Salzburg Global Seminar. Salzburg statement on shared decision making. BMJ 2011 Mar 22;342:d1745. [doi: 10.1136/bmj.d1745] [Medline: 21427038]
6. Elwyn G, Laitner S, Coulter A, Walker E, Watson P, Thomson R. Implementing shared decision making in the NHS. BMJ 2010 Oct 14;341:c5146 [FREE Full text] [doi: 10.1136/bmj.c5146] [Medline: 20947577]
7. Briss P, Rimer B, Reilley B, Coates RC, Lee NC, Mullen P, Task Force on Community Preventive Services. Promoting informed decisions about cancer screening in communities and healthcare systems. Am J Prev Med 2004 Jan;26(1):67-80. [doi: 10.1016/j.amepre.2003.09.012] [Medline: 14700715]
8. Goto Y, Miura H. Challenges in promoting shared decision-making: towards a breakthrough in Japan. Z Evid Fortbild Qual Gesundhwes 2022 Jun;171:84-88. [doi: 10.1016/j.zefq.2022.04.007] [Medline: 35610133]
9. Lu C, Li X, Yang K. Trends in shared decision-making studies from 2009 to 2018: a bibliometric analysis. Front Public Health 2019;7:384 [FREE Full text] [doi: 10.3389/fpubh.2019.00384] [Medline: 31921749]
10. Hughes TM, Merath K, Chen Q, Sun S, Palmer E, Idrees JJ, et al. Association of shared decision-making on patient-reported health outcomes and healthcare utilization. Am J Surg 2018 Jul;216(1):7-12. [doi: 10.1016/j.amjsurg.2018.01.011] [Medline: 29395026]
11. Tai-Seale M, Elwyn G, Wilson CJ, Stults C, Dillon EC, Li M, et al. Enhancing shared decision making through carefully designed interventions that target patient and provider behavior. Health Aff (Millwood) 2016 Apr;35(4):605-612. [doi: 10.1377/hlthaff.2015.1398] [Medline: 27044959]
12. Kashaf MS, McGill E. Does shared decision making in cancer treatment improve quality of life? A systematic literature review. Med Decis Making 2015 Nov;35(8):1037-1048. [doi: 10.1177/0272989X15598529] [Medline: 26246515]
13. Stacey D, Bennett CL, Barry MJ, Col NF, Eden KB, Holmes-Rovner M, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2011 Oct 05(10):CD001431. [doi: 10.1002/14651858.CD001431.pub3] [Medline: 21975733]
14. Durand M, Carpenter L, Dolan H, Bravo P, Mann M, Bunn F, et al. Do interventions designed to support shared decision-making reduce health inequalities? A systematic review and meta-analysis. PLoS One 2014;9(4):e94670 [FREE Full text] [doi: 10.1371/journal.pone.0094670] [Medline: 24736389]
15. Shay LA, Lafata JE. Where is the evidence? A systematic review of shared decision making and patient outcomes. Med Decis Making 2015 Jan;35(1):114-131 [FREE Full text] [doi: 10.1177/0272989X14551638] [Medline: 25351843]
16. Weiser TG, Haynes AB, Molina G, Lipsitz SR, Esquivel MM, Uribe-Leitz T, et al. Estimate of the global volume of surgery in 2012: an assessment supporting improved health outcomes. Lancet 2015 Apr 27;385 Suppl 2:S11. [doi: 10.1016/S0140-6736(15)60806-6] [Medline: 26313057]
17. Abbott TE, Fowler AJ, Dobbs TD, Harrison EM, Gillies MA, Pearse RM. Frequency of surgical treatment and related hospital procedures in the UK: a national ecological study using hospital episode statistics. Br J Anaesth 2017 Aug 01;119(2):249-257 [FREE Full text] [doi: 10.1093/bja/aex137] [Medline: 28854546]
18. Davison BJ, Matthew A, Gardner AM. Prospective comparison of the impact of robotic-assisted laparoscopic radical prostatectomy versus open radical prostatectomy on health-related quality of life and decision regret. Can Urol Assoc J 2014;8(1-2):E68-E72 [FREE Full text] [doi: 10.5489/cuaj.480] [Medline: 24454607]
19. Diefenbach MA, Mohamed NE. Regret of treatment decision and its association with disease-specific quality of life following prostate cancer treatment. Cancer Invest 2007 Sep 11;25(6):449-457. [doi: 10.1080/07357900701359460] [Medline: 17882657]
20. Walsh T, Barr PJ, Thompson R, Ozanne E, O'Neill C, Elwyn G. Undetermined impact of patient decision support interventions on healthcare costs and savings: systematic review. BMJ 2014 Jan 23;348:g188 [FREE Full text] [doi: 10.1136/bmj.g188] [Medline: 24458654]
21. Kissane DW, Bylund CL, Banerjee SC, Bialer PA, Levin TT, Maloney EK, et al. Communication skills training for oncology professionals. J Clin Oncol 2012 Apr 10;30(11):1242-1247 [FREE Full text] [doi: 10.1200/JCO.2011.39.6184] [Medline: 22412145]
22. Ross L, Ashford AD, Bleechington SJ, Dark T, Erwin DO. Applicability of a video intervention to increase informed decision making for prostate-specific antigen testing. J Natl Med Assoc 2010 Mar;102(3):228-236 [FREE Full text] [doi: 10.1016/s0027-9684(15)30529-0] [Medline: 20355352]
23. Smith SK, Trevena L, Simpson JM, Barratt A, Nutbeam D, McCaffery KJ. A decision aid to support informed choices about bowel cancer screening among adults with low education: randomised controlled trial. BMJ 2010 Oct 26;341:c5370 [FREE Full text] [doi: 10.1136/bmj.c5370] [Medline: 20978060]
24. Joseph-Williams N, Lloyd A, Edwards A, Stobbart L, Tomson D, Macphail S, et al. Implementing shared decision making in the NHS: lessons from the MAGIC programme. BMJ 2017 Apr 18;357:j1744 [FREE Full text] [doi: 10.1136/bmj.j1744] [Medline: 28420639]
25. Shard decision making NG197. National Institute for Health and Care Excellence. 2021. URL: https://www.nice.org.uk/guidance/ng197 [accessed 2024-03-21]

26. Légaré F, Adekpedjou R, Stacey D, Turcotte S, Kryworuchko J, Graham ID, et al. Interventions for increasing the use of shared decision making by healthcare professionals. Cochrane Database Syst Rev 2018 Jul 19;7(7):CD006732 [FREE Full text] [doi: 10.1002/14651858.CD006732.pub4] [Medline: 30025154]
27. Ubbink DT, Hageman MG, Legemate DA. Shared decision-making in surgery. Surg Technol Int 2015 May;26:31-36. [Medline: 26054988]
28. Elwyn G, Scholl I, Tietbohl C, Mann M, Edwards AG, Clay C, et al. "Many miles to go …": a systematic review of the implementation of patient decision support interventions into routine clinical practice. BMC Med Inform Decis Mak 2013;13 Suppl 2(Suppl 2):S14 [FREE Full text] [doi: 10.1186/1472-6947-13-S2-S14] [Medline: 24625083]
29. Forcino RC, Meinders MJ, Engel JA, O'Malley AJ, Elwyn G. Routine patient-reported experience measurement of shared decision-making in the USA: a qualitative study of the current state according to frontrunners. BMJ Open 2020 Jun 07;10(6):e037087 [FREE Full text] [doi: 10.1136/bmjopen-2020-037087] [Medline: 32513894]
30. Scholl I, Hahlweg P, Lindig A, Frerichs W, Zill J, Cords H, et al. Evaluation of a program for routine implementation of shared decision-making in cancer care: results of a stepped wedge cluster randomized trial. Implement Sci 2021 Dec 29;16(1):106 [FREE Full text] [doi: 10.1186/s13012-021-01174-4] [Medline: 34965881]
31. Elwyn G, Frosch DL, Kobrin S. Implementing shared decision-making: consider all the consequences. Implement Sci 2016 Aug 08;11:114 [FREE Full text] [doi: 10.1186/s13012-016-0480-9] [Medline: 27502770]
32. Härter M, van der Weijden T, Elwyn G. Policy and practice developments in the implementation of shared decision making: an international perspective. Z Evid Fortbild Qual Gesundhwes 2011;105(4):229-233. [doi: 10.1016/j.zefq.2011.04.018] [Medline: 21620313]
33. Zisman-Ilani Y, Chmielowska M, Dixon LB, Ramon S. NICE shared decision making guidelines and mental health: challenges for research, practice and implementation. BJPsych Open 2021 Sep 02;7(5):e154 [FREE Full text] [doi: 10.1192/bjo.2021.987] [Medline: 34470688]
34. Shared decision-making: summary guide. NHS England Personalised Care. URL: https://www.england.nhs.uk/publication/shared-decision-making-summary-guide/#:~:text=The%20Shared%20Decision%2DMaking%20Summary,services%20that%20informed%20individuals%20want [accessed 2024-03-21]
35. Tyack Z, Simons M, McPhail SM, Harvey G, Zappala T, Ware RS, et al. Improving the patient-centred care of children with life-altering skin conditions using feedback from electronic patient-reported outcome measures: protocol for a hybrid effectiveness-implementation study (PEDS-ePROM). BMJ Open 2021 Apr 09;11(4):e041861 [FREE Full text] [doi: 10.1136/bmjopen-2020-041861] [Medline: 33837095]
36. Kyte D, Anderson N, Auti R, Aiyegbusi OL, Bishop J, Bissell A, et al. Development of an electronic patient-reported outcome measure (ePROM) system to aid the management of patients with advanced chronic kidney disease. J Patient Rep Outcomes 2020 Jul 08;4(1):55 [FREE Full text] [doi: 10.1186/s41687-020-00223-8] [Medline: 32642867]
37. van der Krieke L, Emerencia AC, Aiello M, Sytema S. Usability evaluation of a web-based support system for people with a schizophrenia diagnosis. J Med Internet Res 2012 Feb 06;14(1):e24 [FREE Full text] [doi: 10.2196/jmir.1921] [Medline: 22311883]
38. Richards HS, Blazeby JM, Portal A, Harding R, Reed T, Lander T, et al. A real-time electronic symptom monitoring system for patients after discharge following surgery: a pilot study in cancer-related surgery. BMC Cancer 2020 Jun 10;20(1):543 [FREE Full text] [doi: 10.1186/s12885-020-07027-5] [Medline: 32522163]
39. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. J Clin Oncol 2016 Feb 20;34(6):557-565 [FREE Full text] [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
40. Aiyegbusi OL, Kyte D, Cockwell P, Marshall T, Dutton M, Walmsley-Allen N, et al. Patient and clinician perspectives on electronic patient-reported outcome measures in the management of advanced CKD: a qualitative study. Am J Kidney Dis 2019 Aug;74(2):167-178. [doi: 10.1053/j.ajkd.2019.02.011] [Medline: 31003865]
41. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 05;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
42. O'Cathain A, Croot L, Sworn K, Duncan E, Rousseau N, Turner K, et al. Taxonomy of approaches to developing interventions to improve health: a systematic methods overview. Pilot Feasibility Stud 2019 Mar 12;5(1):41 [FREE Full text] [doi: 10.1186/s40814-019-0425-6] [Medline: 30923626]
43. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
44. Bevan N. Measuring usability as quality of use. Software Qual J 1995 Jun;4(2):115-130. [doi: 10.1007/BF00402715]
45. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: a scoping review. Int J Med Inform 2019 Jun;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
46. Dumas JS, Salzman MC. Usability assessment methods. Rev Hum Factors Ergon 2016 Nov 03;2(1):109-140. [doi: 10.1177/1557234x0600200105]

47. Lin HX, Choong Y, Salvendy G. A proposed index of usability: a method for comparing the relative usability of different software systems. Behav Inf Technol 1997 Jan;16(4-5):267-277. [doi: 10.1080/014492997119833]
48. Yen PY, Bakken S. Review of health information technology usability study methodologies. J Am Med Inform Assoc 2012;19(3):413-422 [FREE Full text] [doi: 10.1136/amiajnl-2010-000020] [Medline: 21828224]
49. Aiyegbusi OL. Key methodological considerations for usability testing of electronic patient-reported outcome (ePRO) systems. Qual Life Res 2020 Feb;29(2):325-333 [FREE Full text] [doi: 10.1007/s11136-019-02329-z] [Medline: 31691202]
50. ISO 9241-11:2018: ergonomics of human-system interaction: part 11: usability: definitions and concepts. International Organization for Standardization. 2018. URL: https://www.iso.org/standard/63500.html#:~:text=ISO%209241%2D11%3A2018%20provides,services%20(including%20technical%20and%20personal [accessed 2024-03-21]
51. Bevan N, Carter J, Earthy J, Geis T, Harker S. New ISO standards for usability, usability reports and usability measures. In: Proceedings of the 18th International Conference on Human-Computer Interaction. Theory, Design, Development and Practice. 2016 Presented at: HCI '16; July 17-22, 2016; Toronto, ON p. 268-278 URL: https://link.springer.com/chapter/10.1007/978-3-319-39510-4_25 [doi: 10.1007/978-3-319-39510-4_25]
52. Hoffmann C, Avery KN, Macefield RC, Snelgrove V, Blazeby JM, Hopkins D, ALPACA Study team. Real-time monitoring and feedback to improve shared decision-making for surgery (the ALPACA Study): protocol for a mixed-methods study to inform co-development of an inclusive intervention. BMJ Open 2024 Jan 18;14(1):e079155 [FREE Full text] [doi: 10.1136/bmjopen-2023-079155] [Medline: 38238045]
53. Gärtner FR, Bomhof-Roordink H, Smith IP, Scholl I, Stiggelbout AM, Pieterse AH. The quality of instruments to assess the process of shared decision making: a systematic review. PLoS One 2018;13(2):e0191747 [FREE Full text] [doi: 10.1371/journal.pone.0191747] [Medline: 29447193]
54. Personalised care shared decision making summary guide. NHS England and NHS Improvement. 2019. URL: https://www.england.nhs.uk/wp-content/uploads/2019/01/shared-decision-making-summary-guide-v1.pdf [accessed 2024-03-21]
55. Barr PJ, Forcino RC, Thompson R, Ozanne EM, Arend R, Castaldo MG, et al. Evaluating CollaboRATE in a clinical setting: analysis of mode effects on scores, response rates and costs of data collection. BMJ Open 2017 Mar 24;7(3):e014681 [FREE Full text] [doi: 10.1136/bmjopen-2016-014681] [Medline: 28341691]
56. Doherr H, Christalle E, Kriston L, Härter M, Scholl I. Use of the 9-item shared decision making questionnaire (SDM-Q-9 and SDM-Q-Doc) in intervention studies-a systematic review. PLoS One 2017;12(3):e0173904 [FREE Full text] [doi: 10.1371/journal.pone.0173904] [Medline: 28358864]
57. Barr PJ, Thompson R, Walsh T, Grande SW, Ozanne EM, Elwyn G. The psychometric properties of CollaboRATE: a fast and frugal patient-reported measure of the shared decision-making process. J Med Internet Res 2014 Jan 03;16(1):e2 [FREE Full text] [doi: 10.2196/jmir.3085] [Medline: 24389354]
58. Barnum CM. Establishing the essentials. In: Barnum CM, editor. Usability Testing Essentials: Ready, Set ...Test!. Burlington, MA: Morgan Kaufmann; 2011:9-23.
59. Hedberg H, Iivari N, Rajanen M, Harjumaa L. Assuring quality and usability in open source software development. In: Proceedings of the 1st International Workshop on Emerging Trends in FLOSS Research and Development. 2007 Presented at: FLOSS '07; May 20-26, 2007; Minneapolis, MN URL: https://ieeexplore.ieee.org/document/4273073 [doi: 10.1109/floss.2007.2]
60. van den Haak MJ, de Jong MD, Schellens PJ. Evaluation of an informational web site: three variants of the think-aloud method compared. Tech Commun 2007;54(1):58-71 [FREE Full text]
61. Alshammari T, Alhadreti O, Mayhew PJ. When to ask participants to think aloud: a comparative study of concurrent and retrospective think-aloud methods. Int J Hum Comput Interact 2015;6(3):48-64 [FREE Full text] [doi: 10.4135/9781412961288.n460]
62. Khajouei R, Farahani F. A combination of two methods for evaluating the usability of a hospital information system. BMC Med Inform Decis Mak 2020 May 04;20(1):84 [FREE Full text] [doi: 10.1186/s12911-020-1083-6] [Medline: 32366248]
63. Farzanfar R, Finkelstein J, Friedman RH. Testing the usability of two automated home-based patient-management systems. J Med Syst 2004 Apr;28(2):143-153. [doi: 10.1023/b:joms.0000023297.50379.3c] [Medline: 15195845]
64. Kaufman DR, Patel VL, Hilliman C, Morin PC, Pevzner J, Weinstock RS, et al. Usability in the real world: assessing medical information technologies in patients' homes. J Biomed Inform 2003;36(1-2):45-60 [FREE Full text] [doi: 10.1016/s1532-0464(03)00056-x] [Medline: 14552846]
65. Birns J, Birns JH, Joffre KA, Leclerc JF, Paulsen CA. Getting the Whole Picture: Collecting Usability Data Using Two Methods-Concurrent Think Aloud and Retrospective Probing. 2002 Presented at: Proceedings of UPA Conference; 2-6 September 2002; London URL: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=ba55a9ff388337e7e561fd1db41b3533051746a9 [doi: 10.1109/ipcc.2003.1245501]
66. Geisen E, Romano Bergstrom J. Developing the usability testing protocol. In: Geisen E, Romano Bergstrom J, editors. Usability Testing for Survey Research. Burlington, MA: Morgan Kaufmann; 2017:111-129.
67. Ghasemi A, Zahediasl S. Normality tests for statistical analysis: a guide for non-statisticians. Int J Endocrinol Metab 2012;10(2):486-489 [FREE Full text] [doi: 10.5812/ijem.3505] [Medline: 23843808]

68. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
69. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
70. Jaspers MW. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. Int J Med Inform 2009 May;78(5):340-353. [doi: 10.1016/j.ijmedinf.2008.10.002] [Medline: 19046928]
71. Virzi RA. Refining the test phase of usability evaluation: how many subjects is enough? Hum Factors 2016 Nov 23;34(4):457-468. [doi: 10.1177/001872089203400407]
72. Nielsen J, Landauer TK. Mathematical model of the finding of usability problems. In: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems. 1993 Presented at: CHI '93; April 24-29, 1993; Amsterdam, The Netherlands p. 206-213 URL: https://dl.acm.org/doi/10.1145/169059.169166 [doi: 10.1145/169059.169166]
73. Brooke J. SUS: a 'quick and dirty' usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation In Industry. London, UK: CRC Press; 1996.
74. Schnall R, Cho H, Liu J. Health information technology usability evaluation scale (Health-ITUES) for usability assessment of mobile health technology: validation study. JMIR Mhealth Uhealth 2018 Jan 05;6(1):e4 [FREE Full text] [doi: 10.2196/mhealth.8851] [Medline: 29305343]
75. Yen PY, Bakken S. Review of health information technology usability study methodologies. J Am Med Inform Assoc 2012;19(3):413-422 [FREE Full text] [doi: 10.1136/amiajnl-2010-000020] [Medline: 21828224]
76. O'Connell S, Palmer R, Withers K, Saha N, Puntoni S, Carolan-Rees G, PROMs, PREMsEffectiveness Programme. Requirements for the collection of electronic PROMS either "in clinic" or "at home" as part of the PROMs, PREMs and Effectiveness Programme (PPEP) in Wales: a feasibility study using a generic PROM tool. Pilot Feasibility Stud 2018;4:90 [FREE Full text] [doi: 10.1186/s40814-018-0282-8] [Medline: 29997899]
77. Palmen LN, Schrier JC, Scholten R, Jansen JH, Koëter S. Is it too early to move to full electronic PROM data collection?: A randomized controlled trial comparing PROM's after hallux valgus captured by e-mail, traditional mail and telephone. Foot Ankle Surg 2016 Mar;22(1):46-49. [doi: 10.1016/j.fas.2015.05.001] [Medline: 26869500]
78. Kyte D, Ives J, Draper H, Calvert M. Current practices in patient-reported outcome (PRO) data collection in clinical trials: a cross-sectional survey of UK trial staff and management. BMJ Open 2016 Oct 03;6(10):e012281 [FREE Full text] [doi: 10.1136/bmjopen-2016-012281] [Medline: 27697875]
79. Liu TC, Ohueri CW, Schryver EM, Bozic KJ, Koenig KM. Patient-identified barriers and facilitators to pre-visit patient-reported outcomes measures completion in patients with hip and knee pain. J Arthroplasty 2018 Mar;33(3):643-9.e1. [doi: 10.1016/j.arth.2017.10.022] [Medline: 29169689]
80. AOA PROMs pilot project final report. Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR). 2020. URL: https://aoanjrr.sahmri.com/documents/10180/681914/AOANJRR+PROMs+Pilot+Final+Report [accessed 2024-03-21]
81. Lungu DA, Pennucci F, De Rosis S, Romano G, Melfi F. Implementing successful systematic patient reported outcome and experience measures (PROMs and PREMs) in robotic oncological surgery-the role of physicians. Int J Health Plann Manage 2020 May;35(3):773-787. [doi: 10.1002/hpm.2959] [Medline: 31793689]
82. Arner M. Developing a national quality registry for hand surgery: challenges and opportunities. EFORT Open Rev 2016 Apr;1(4):100-106 [FREE Full text] [doi: 10.1302/2058-5241.1.000045] [Medline: 28461935]
83. Morris ME, Brusco N, Woods J, Myles PS, Hodge A, Jones C, et al. Protocol for implementation of the 'AusPROM' recommendations for elective surgery patients: a mixed-methods cohort study. BMJ Open 2021 Sep 16;11(9):e049937 [FREE Full text] [doi: 10.1136/bmjopen-2021-049937] [Medline: 34531213]
84. Abbasi F, Khajouei R, Mirzaee M. The efficiency and effectiveness of surgery information systems in Iran. BMC Med Inform Decis Mak 2020 Sep 16;20(1):229 [FREE Full text] [doi: 10.1186/s12911-020-01236-5] [Medline: 32938452]
85. Szajna B. Empirical evaluation of the revised technology acceptance model. Manag Sci 1996 Jan;42(1):85-92. [doi: 10.1287/mnsc.42.1.85]
86. Kalankesh LR, Nasiry Z, Fein RA, Damanabi S. Factors influencing user satisfaction with information systems: a systematic review. Galen Med J 2020;9:e1686. [doi: 10.31661/gmj.v9i0.1686] [Medline: 34466567]
87. Aggelidis VP, Chatzoglou PD. Hospital information systems: measuring end user computing satisfaction (EUCS). J Biomed Inform 2012 Jun;45(3):566-579 [FREE Full text] [doi: 10.1016/j.jbi.2012.02.009] [Medline: 22426283]
88. Mustafa M, Alzubi S. Factors affecting the success of internet of things for enhancing quality and efficiency implementation in hospitals sector in Jordan during the crises of COVID-19. In: Chakraborty C, Banerjee A, Garg L, Rodrigues JJ, editors. Internet of Medical Things for Smart Healthcare: COVID-19 Pandemic. Singapore, Singapore: Springer; 2020:107-140.
89. Handayani PW, Hidayanto AN, Budi I. User acceptance factors of hospital information systems and related technologies: systematic review. Inform Health Soc Care 2018 Dec;43(4):401-426. [doi: 10.1080/17538157.2017.1353999] [Medline: 28829650]
90. Sidani S, Epstein DR, Bootzin RR, Moritz P, Miranda J. Assessment of preferences for treatment: validation of a measure. Res Nurs Health 2009 Aug;32(4):419-431 [FREE Full text] [doi: 10.1002/nur.20329] [Medline: 19434647]

91. Ahmed S, Zidarov D, Eilayyan O, Visca R. Prospective application of implementation science theories and frameworks to inform use of PROMs in routine clinical care within an integrated pain network. Qual Life Res 2021 Nov;30(11):3035-3047 [FREE Full text] [doi: 10.1007/s11136-020-02600-8] [Medline: 32876812]
92. Meirte J, Hellemans N, Anthonissen M, Denteneer L, Maertens K, Moortgat P, et al. Benefits and disadvantages of electronic patient-reported outcome measures: systematic review. JMIR Perioper Med 2020 Apr 03;3(1):e15588 [FREE Full text] [doi: 10.2196/15588] [Medline: 33393920]
93. Schamber EM, Takemoto SK, Chenok KE, Bozic KJ. Barriers to completion of patient reported outcome measures. J Arthroplasty 2013 Oct;28(9):1449-1453. [doi: 10.1016/j.arth.2013.06.025] [Medline: 23890831]
94. Jordan JE, Buchbinder R, Briggs AM, Elsworth GR, Busija L, Batterham R, et al. The health literacy management scale (HeLMS): a measure of an individual's capacity to seek, understand and use health information within the healthcare setting. Patient Educ Couns 2013 May;91(2):228-235. [doi: 10.1016/j.pec.2013.01.013] [Medline: 23419326]
95. Knight J, Ayyash K, Colling K, Dhesi J, Ewan V, Danjoux G, et al. A cohort study investigating the relationship between patient reported outcome measures and pre-operative frailty in patients with operable, non-palliative colorectal cancer. BMC Geriatr 2020 Aug 27;20(1):311 [FREE Full text] [doi: 10.1186/s12877-020-01715-4] [Medline: 32854632]
96. Mason JD, Blencowe NS, McNair AG, Stevens DJ, Avery KN, Pullyblank AM, et al. Investigating the collection and assessment of patient-reported outcome data amongst unplanned surgical hospital admissions: a feasibility study. Pilot Feasibility Stud 2015;1:16 [FREE Full text] [doi: 10.1186/s40814-015-0011-5] [Medline: 27965795]
97. Kwong E, Black N. Feasibility of collecting retrospective patient reported outcome measures (PROMs) in emergency hospital admissions. J Patient Rep Outcomes 2018 Nov 15;2(1):54 [FREE Full text] [doi: 10.1186/s41687-018-0077-y] [Medline: 30467820]
98. Iversen HH, Holmboe O, Bjertnaes O. Patient-reported experiences with general practitioners: a randomised study of mail and web-based approaches following a national survey. BMJ Open 2020 Oct 13;10(10):e036533 [FREE Full text] [doi: 10.1136/bmjopen-2019-036533] [Medline: 33051230]
99. Bliddal S, Banasik K, Pedersen OB, Nissen J, Cantwell L, Schwinn M, et al. Acute and persistent symptoms in non-hospitalized PCR-confirmed COVID-19 patients. Sci Rep 2021 Jun 23;11(1):13153 [FREE Full text] [doi: 10.1038/s41598-021-92045-x] [Medline: 34162913]
100. Wintner LM, Giesinger JM, Zabernigg A, Rumpold G, Sztankay M, Oberguggenberger AS, et al. Evaluation of electronic patient-reported outcome assessment with cancer patients in the hospital and at home. BMC Med Inform Decis Mak 2015 Dec 23;15:110 [FREE Full text] [doi: 10.1186/s12911-015-0230-y] [Medline: 26699708]
101. Aiyegbusi OL, Nair D, Peipert JD, Schick-Makaroff K, Mucsi I. A narrative review of current evidence supporting the implementation of electronic patient-reported outcome measures in the management of chronic diseases. Ther Adv Chronic Dis 2021;12:20406223211015958 [FREE Full text] [doi: 10.1177/20406223211015958] [Medline: 34104376]
102. Hutchings A, Neuburger J, Grosse Frie K, Black N, van der Meulen J. Factors associated with non-response in routine use of patient reported outcome measures after elective surgery in England. Health Qual Life Outcomes 2012 Mar 30;10:34 [FREE Full text] [doi: 10.1186/1477-7525-10-34] [Medline: 22462512]
103. Bevans M, Ross A, Cella D. Patient-reported outcomes measurement information system (PROMIS): efficient, standardized tools to measure self-reported health and quality of life. Nurs Outlook 2014;62(5):339-345 [FREE Full text] [doi: 10.1016/j.outlook.2014.05.009] [Medline: 25015409]
104. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, ISPOR Task Force for TranslationCultural Adaptation. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. Value Health 2005;8(2):94-104 [FREE Full text] [doi: 10.1111/j.1524-4733.2005.04054.x] [Medline: 15804318]
105. Rao JK, Anderson LA, Inui TS, Frankel RM. Communication interventions make a difference in conversations between physicians and patients: a systematic review of the evidence. Med Care 2007 Apr;45(4):340-349. [doi: 10.1097/01.mlr.0000254516.04961.d5] [Medline: 17496718]
106. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback: effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev 2012 Jun 13(6):CD000259. [doi: 10.1002/14651858.CD000259.pub3] [Medline: 22696318]
107. Pattinson RC, Say L, Makin JD, Bastos MH. Critical incident audit and feedback to improve perinatal and maternal mortality and morbidity. Cochrane Database Syst Rev 2005 Oct 19;2005(4):CD002961 [FREE Full text] [doi: 10.1002/14651858.CD002961.pub2] [Medline: 16235307]
108. de Jong RK, Snoek H, Staal WG, Klip H. The effect of patients' feedback on treatment outcome in a child and adolescent psychiatric sample: a randomized controlled trial. Eur Child Adolesc Psychiatry 2019 Jun;28(6):819-834 [FREE Full text] [doi: 10.1007/s00787-018-1247-4] [Medline: 30390148]
109. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021 Sep 30;374:n2061 [FREE Full text] [doi: 10.1136/bmj.n2061] [Medline: 34593508]

Abbreviations

ePRM: electronic patient-reported measure
ISO: International Organization for Standardization
ISRCTN: International Standard Randomised Controlled Trial Number
NHS: National Health Service
REDCap: Research Electronic Data Capture
SDM: shared decision-making

Edited by A Kushniruk; submitted 21.02.23; peer-reviewed by D Lungu, S Scott, S Azadnajafabad; comments to author 04.08.23; revised version received 02.10.23; accepted 02.03.24; published 10.04.24.

Please cite as:
Hoffmann C, Avery K, Macefield R, Dvořák T, Snelgrove V, Blazeby J, Hopkins D, Hickey S, Gibbison B, Rooshenas L, Williams A, Aning J, Bekker HL, McNair AGK, The ALPACA Study Team
Usability of an Automated System for Real-Time Monitoring of Shared Decision-Making for Surgery: Mixed Methods Evaluation
JMIR Hum Factors 2024;11:e46698
URL: https://humanfactors.jmir.org/2024/1/e46698
doi:10.2196/46698
PMID:38598276

Original Paper

Influence of Disease-Related Stigma on Patients' Decisions to Upload Medical Reports to the German Electronic Health Record: Randomized Controlled Trial

Niklas von Kalckreuth[1], MSc; Markus A Feufel[1], MSc, PhD

Division of Ergonomics, Department of Psychology and Ergonomics, Technische Universität Berlin, Berlin, Germany

Corresponding Author:
Niklas von Kalckreuth, MSc
Division of Ergonomics
Department of Psychology and Ergonomics
Technische Universität Berlin
Straße des 17 Juni 135
Berlin, 10623
Germany
Phone: 49 3031470747
Email: niklas.vkalckreuth@tu-berlin.de

Abstract

Background: The rollout of the electronic health record (EHR) represents a central component of the digital transformation of the German health care system. Although the EHR promises more effective, safer, and faster treatment of patients from a systems perspective, the successful implementation of the EHR largely depends on the patient. In a recent survey, 3 out of 4 Germans stated that they intend to use the EHR, whereas other studies show that the intention to use a technology is not a reliable and sufficient predictor of actual use.

Objective: Controlling for patients' intention to use the EHR, we investigated whether disease-specific risk perceptions related to the time course of the disease and disease-related stigma explain the additional variance in patients' decisions to upload medical reports to the EHR.

Methods: In an online user study, 241 German participants were asked to interact with a randomly assigned medical report that varied systematically in terms of disease-related stigma (high vs low) and disease time course (acute vs chronic) and to decide whether to upload it to the EHR.

Results: Disease-related stigma (odds ratio 0.154, $P<.001$) offset the generally positive relationship between intention to use and the upload decision (odds ratio 2.628, $P<.001$), whereas the disease time course showed no effect.

Conclusions: Even if patients generally intend to use the EHR, risk perceptions such as those related to diseases associated with social stigma may deter people from uploading related medical reports to the EHR. To ensure the reliable use of this key technology in a digitalized health care system, transparent and easy-to-comprehend information about the safety standards of the EHR are warranted across the board, even for populations that are generally in favor of using the EHR.

(JMIR Hum Factors 2024;11:e52625) doi:10.2196/52625

KEYWORDS

electronic health record; EHR; technology acceptance; upload behavior; health-related stigma; intention to use; intention-behavior gap; medical reports; stigma; Germany; patient decision; digital transformation; implementation; risk; decision; risk perception; social stigma; safety

Introduction

Background

The digital transformation of health care promises safety and efficiency gains by connecting all players in a health care system [1-3]. One key technology to connect health professionals, insurance providers, and patients is the electronic health record (EHR), which will be implemented nationwide and mandatory for all patients in Germany starting on January 1, 2025. In the EHR, patients' medical data (eg, findings, diagnoses, therapies, vaccinations, discharge reports, emergency data, and medication plans [4,5]) can be digitally documented, exchanged, and viewed [4,6]. Better coordination of health data can ultimately save costs in the health care system [7-9].

In Germany, the Patient Data Protection Act [10] mandates that it is ultimately the patient who controls the type of data that are stored and can be viewed in the EHR. Although a recent survey found that 3 out of 4 Germans state that they intend to use the EHR [11], its success ultimately depends on whether and under what circumstances it is actually *used* to store and share health data. As described below based on the available literature, intention to use is *not* a sufficient and reliable predictor of EHR use. Therefore, in this study, we sought to investigate to what extent intention to use predicts actual use and what additional factors may need to be taken into account to more reliably predict EHR use.

Related Work

The technology acceptance model (TAM) and its extensions such as the unified theory of acceptance and use of technology (UTAUT) assume a positive relationship between intention to use a technology (technology acceptance) and actual use [12-14]. In fact, empirical studies on social networks and online banking show that the greater the intention to use, the more likely the technology will actually be used. However, the same studies also show a statistical discrepancy between intention and behavior, as evidenced by the different variance (R^2) accounted for by the two constructs [15-17]. Questionnaire studies on this so-called "intention-behavior gap" suggest that intention is not a reliable predictor of behavior and consequently that other influencing factors must exist [18,19]. For instance, in the context of social media and electronic commerce, users often have massive privacy concerns to disclose their data and their intentions to use are generally low. Nonetheless, users tend to disclose their data if the benefits they expect from using the applications are sufficiently high [20]; this phenomenon is called the "privacy paradox" and has been confirmed repeatedly [15,20,21]. However, questionnaire studies on digital health technologies show no such paradox and more nuanced patterns. For health technologies, privacy concerns thus far either had no influence [22-24] or have been shown to have a systematic negative impact on intentions and actual technology use [25,26]. In summary, based on the available research, it is unclear to what extent intention to use predicts the actual use of digital health technologies such as the EHR. Theories of technology acceptance infer a direct, positive influence, whereas the results of various questionnaire studies suggest that other factors must play a role given the intention-behavior gap. Although the influence of a few technology-related factors (eg, controllability of data) on the intention to use an EHR have been investigated, a thorough investigation of disease-related factors has not yet been performed.

Methodologically, usage behavior has mostly been investigated using self-report questions about the frequency of use [15,16,27-29], which is associated with several limitations. First, frequency of use is only meaningful if the system is already established and widely used. In the case of new systems such as the EHR in Germany, frequency of use cannot be surveyed. Second, the actual context of use can be difficult to simulate in questionnaire studies, making it difficult to distinguish between intention and behavior [30]. Since the models of technology acceptance described above (ie, TAM and UTAUT) have been evaluated using questionnaires, they may not provide reliable insights into usage behavior in the context of the EHR.

Therefore, to investigate usage behavior regarding the EHR in Germany, we selected a different approach for this study. In terms of uploading behavior, we first identified two possible use cases: (1) users who are living with different acute as well as chronic diseases ("patients with multimorbidity" use case), enabling a direct comparison between different medical findings in terms of risks and benefits of uploading to the EHR; and (2) users who are healthy or have little to no preexisting conditions before they develop a chronic or acute disease ("patients with first contact" use case). To investigate these use cases, we developed and used an interactive prototype of the EHR (ie, a click dummy) to investigate factors influencing the EHR users' decision to upload medical reports. Compared to questionnaire studies, this approach has the advantage that the interaction with the click dummy is closer to a real interaction with the EHR, thereby increasing the ecological validity of behavioral measures [30]. To investigate the first use case, we used a mixed methods design where the experimental intervention was based on an interview study with potential EHR users [31]. The interview study showed that the time course of a disease (chronic vs acute) and disease-related stigma influence people's decisions to upload a medical report to the EHR. The following experiment showed that respondents were more likely to upload a medical report of a chronic disease to the EHR than to upload a report of an acute condition. In contrast, respondents were less likely to upload a report of a disease with high stigma. When a disease with high stigma had a chronic time course, reports were still uploaded. We here report the results of the second use case in which participants interacted with one medical report only.

Methods

Ethical Considerations

This study was approved by the Ethics Committee of the Department of Psychology and Ergonomics (Institut für Psychologie und Arbeitswissenschaft) at Technische Universität (TU) Berlin (tracking number: AWB_KAL_1_230311). Participants volunteered to participate in the survey and informed consent was required. On the first page of the survey, participants were told about the investigator, the study purpose, what data were to be collected during the study, and where and for how long they would be stored. Participants were informed about the duration of the survey (approximately 8 minutes) as well as the compensation for participation. Participants were compensated with €1.60 (US $1.75) for their time and thus according to minimum wage. The participants also had the possibility to download a PDF of the participant information on the first page.

Participants' personal data and responses were kept entirely anonymous and password-protected in the department's data vault. An anonymized data set from the study was made available to other researchers for further analysis with open access. The documentation and availability of the research data collected during the study were managed using the TU repository "DepositOnce," adhering to the regulations for ensuring good scientific practice at TU Berlin, the guidelines of the "DepositOnce" internal research data repository, and data protection regulations. Compliance with these repository guidelines ensures the indexing and findability of the research data by third parties.

Participants

The study was conducted from May 9 to June 10, 2023. Based on an a priori power analysis for a logistic regression with three predictors as well as a false-positive rate α of .05 and a power of $1-\beta=0.80$, we aimed for a sample size of 186 participants. Individuals 18 years and older residing in Germany were eligible to participate in the study. Another prerequisite was that participants had no previous personal experience (own illness) with the diseases mentioned in the medical reports, as affected people deal with disease-related stigma differently than people who are not affected by the disease [32]. Sampling was conducted through Prolific [33], a click worker platform characterized by high data quality [34]. A total of 275 individuals participated in the study. The mean participation time was 9 minutes, 28 seconds (SD 3 minutes, 47 seconds) and the median was 8 minutes, 36 seconds.

Design

We used a 2×2 between-subject study design with the independent variables stigma (high vs low) and time course of illness (chronic vs acute). Stigma was operationalized as the risk that the medical findings could negatively affect the private, professional, or social life of the affected person. For this purpose, the medical reports related to personal lifestyle, as reflected in tests for sexually transmitted diseases [31,32]. The time course is a classification of diseases in terms of their duration. These can be either acute (diseases of short duration that come on quickly) or chronic (diseases that develop slowly or last for a longer time). The dependent variable was the decision to upload the medical report (ie, whether participants were willing to upload the medical findings to the EHR). Furthermore, the intention to use the EHR was included as a covariate.

Materials

The stimuli used in the study were realistic but specially created for the purpose of the study. The medical reports were provided by various hospitals and a medical association. To make the reports appear as realistic as possible, they were edited on the official document heads of these institutions (see Multimedia Appendix 1). In selecting the diseases, both the related stigma and time course were systematically varied. To reflect different disease-related stigma, which covered different risks for professional and social life [35-38], diseases were divided according to their low and high stigmatization potential. To reflect different time courses, diseases were divided according to an acute and chronic time course. Furthermore, diseases were selected to occur regardless of age so that they would be perceived as realistic diseases by an age-diverse sample. Table 1 shows the diseases used as stimuli, categorized by level of stigma potential and time course.

Table 1. Diseases used in the stimuli, categorized by level of stigma potential and time course.

Stigma potential	Acute disease	Chronic disease
Low	Fractured wrist	Type 1 diabetes
High	STD[a] (gonorrhea)	Depression

[a]STD: sexually transmitted disease.

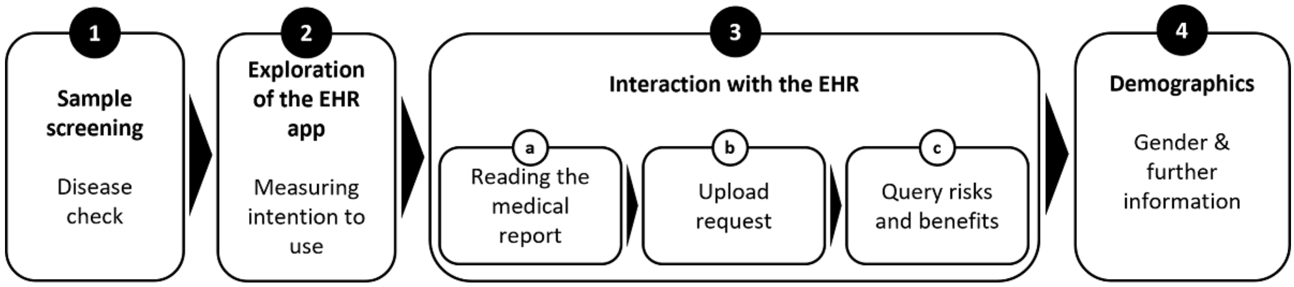
The interface software FIGMA was used to create the click dummy, which was modeled after the mobile EHR app of a German health insurance company (BARMER)—the eCare app—to support a realistic interaction with an EHR. Specifically, the click dummy allowed participants to upload findings, grant or revoke permissions to view medical reports, and create medication plans. Only the "Upload report" function was used in this study.

We used LimeSurvey (version 3.28.3+220315) to create and conduct a 5-page online survey (see Multimedia Appendix 2). The EHR click dummy and the medical reports were incorporated into the survey using iFrame. LimeSurvey software was used to ensure that all questions had to be answered to complete the study and receive the compensation. As in the previous study investigating the first use case [31], in this study, we tested the effect of the independent variables by querying the perceived privacy risk and perceived benefit of uploading findings to the EHR as manipulation checks. Based on the results of this previous study [31], we assumed that high stigma would result in a high perceived privacy risk and a chronic time course would result in a high perceived benefit of uploading the medical report. Perceived privacy risk, perceived benefit, and intention to use were measured using a 7-point Likert scale ranging from 1 ("strongly disagree") to 7 ("strongly agree"). The decision to upload the finding to the EHR was measured using a dichotomous item (yes/no).

Procedure

The study procedure is shown schematically in Figure 1. Before the start of the experiment, participants gave their informed consent. This was followed by screening questions related to disease experience (step 1). Participants who had experience with the diseases in the medical reports were excluded from the study. Subsequently, participants were given 1 minute to interact with the EHR click dummy and were then required to answer questions regarding their intention to use the EHR (step 2). Participants were then asked to interact with the medical report (step 3). In this process, each person was first randomly assigned to one of the four diseases shown in Table 1 and asked to read an easy-to-understand description of the disease of approximately 2-3 sentences (see Multimedia Appendix 3) (step 3a). Participants then decided whether they wanted to upload the report to their EHR (step 3b). Afterward, participants were asked to rate the perceived privacy risks and benefits of uploading the report (step 3c). The survey was completed with the collection of demographic characteristics (age, gender, education level, and experience with EHRs). In this step (step 4), the participants also had the opportunity to declare their responses invalid, while still receiving compensation, in case they did not pay sufficient attention to the instructions provided (eg, due to choosing random answers, inattentively reading questions, or rushing through the survey).

Figure 1. Overview of the study procedure. EHR: electronic health record.



Analysis

We cleaned and analyzed the data using RStudio (version 1.3.1093). Due to lack of variance inhomogeneity or a normal distribution, the analyses regarding perceived privacy risks and benefits were performed using the nonparametric Mann-Whitney *U* test. As mentioned above, we hypothesized that high stigma would result in a high perceived privacy risk and a chronic time course would result in a high perceived benefit of uploading the medical report. The influence of the independent variables (disease-specific stigma and time course) and the covariate "intention to use" on the upload decision were tested using multiple logistic regression with dummy coding. We hypothesized that usage behavior is negatively influenced by disease-specific stigma and positively influenced by time course and intention. To control for demographic and interindividual influences, we used multiple logistic regression with standardized coefficients for better comparability. In doing so, we followed the recommendations for testing control variables [39] and tested the variables that have been shown to be causally related to privacy behavior along with the independent variables. The control variables were age, education level, and experience with the technical system, in this case the EHR [40,41].

Results

Sample Characteristics

A total of 275 observations were collected. Of those, 34 records were excluded, 29 because of participants' previous medical histories, 3 because of incomplete questionnaires, and 2 because they were marked as invalid by participants. Figure 2 shows the flow of participants in the study based on the CONSORT (Consolidated Standards of Reporting Trials) statement [42].

Thus, a sample of 241 observations (146 male, 92 female, 1 diverse, 2 no information provided) was used for further analysis. Table 2 summarizes the demographic characteristics of the sample.

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flow chart. SP: stigma potential; TC: time course.

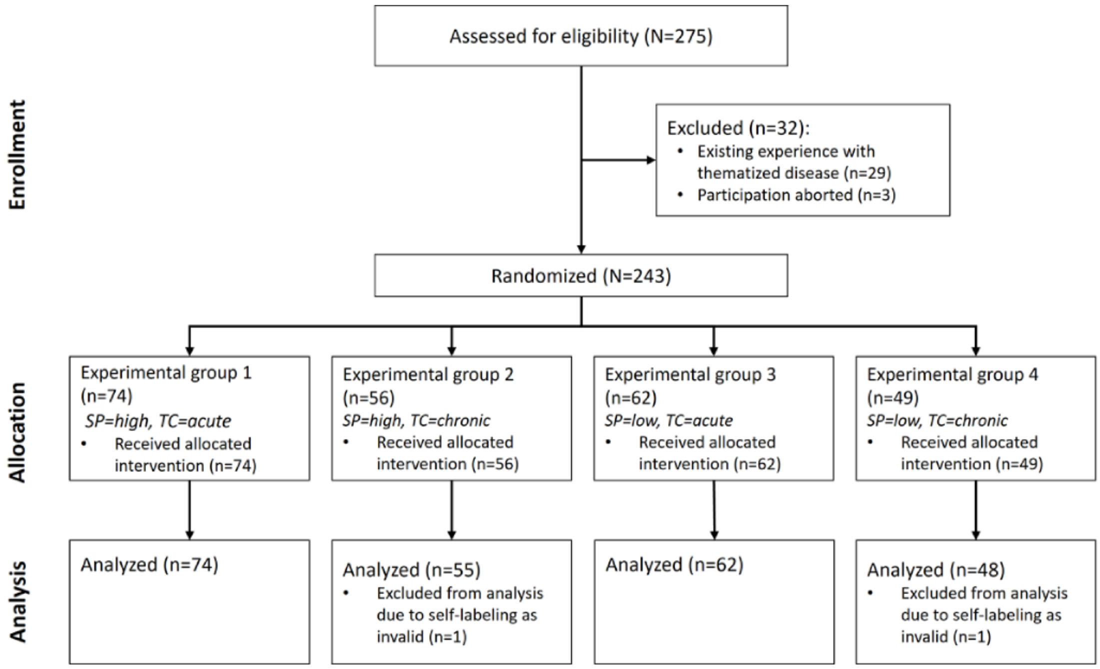


Table 2. Demographic data of the sample (N=241).

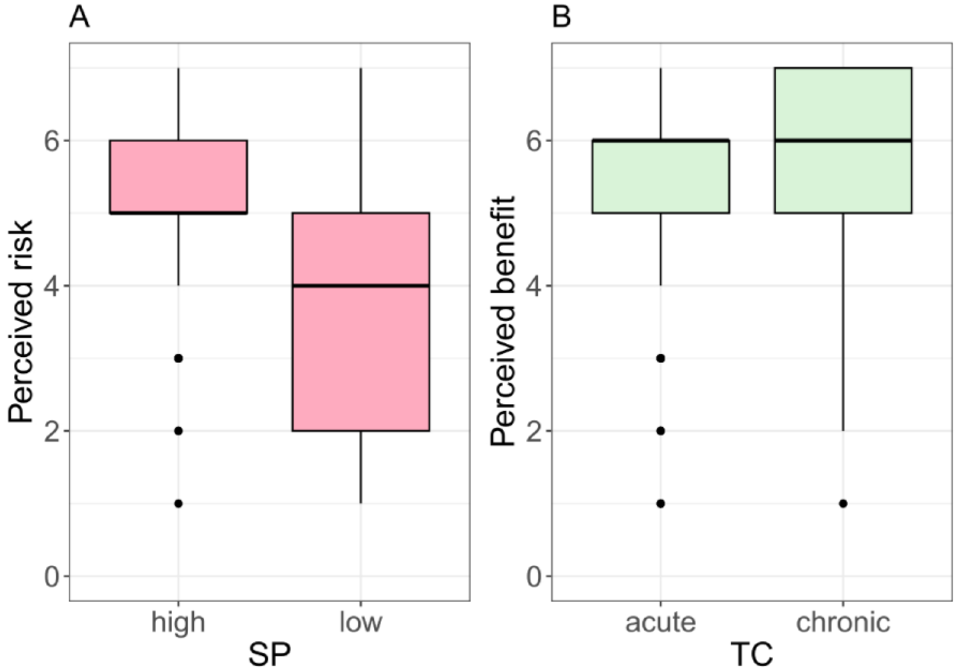
Demographic characteristic	Value
Age (years), mean (SD)	31.31 (9.76)
Gender, n (%)	
Female	92 (38.2)
Male	146 (60.6)
Other	3 (1.2)
Education, n (%)	
No degree	9 (3.7)
School leaving certificate	3 (1.2)
Secondary school certificate	18 (7.5)
General qualification for university entrance	66 (27.4)
Vocational training	33 (13.7)
University degree (bachelor's or master's degree)	112 (46.5)
Experience with the German EHR[a], n (%)	
EHR is unknown	61 (25.3)
EHR is known but not used	164 (68)
Occasional use	14 (5.8)
Regular use	2 (0.8)

[a]EHR: electronic health record.

Risk and Benefit Perception

We first checked whether stigma potential had an effect on privacy risk perception and whether time course had an effect on the benefit perception of uploading (see Figure 3). Mann-Whitney *U* tests showed a significant effect of stigma potential on privacy risk perception (*W*=10,777; *P*<.001), where high stigma was associated with high risk. The effect of the disease time course on benefit perception was not significant (*W*=6379; *P*=.14), with a mean benefit perception of 5.34 (SD 1.39) for acute diseases and of 5.54 (SD 1.43) for chronic diseases. Consequently, in contrast to our study on the first use case with several medical reports [31], there was no relationship found between time course and perceived benefits when there is only one report to upload.

Figure 3. (A) Perceived risk in relation to stigma potential (SP) and (B) perceived benefit in relation to the disease time course (TC). The horizontal line in the box represents the median.



Controls

To investigate the potential association between the decision to upload the medical report and the independent variables disease-specific stigma and time course, we first performed a logistic regression (Hosmer-Lemeshow R^2=0.319, Nagelkerke R^2=0.590, Cox-Snell R^2=0.537; χ^2_{15}=86.973; P<.001) to control for the covariate intention to use and the demographic variables age, sex, education level, and experience with the EHR. The covariate intention to use (odds ratio [OR] 2.497, 95% CI 1.831-3.456; z=5.455; P<.001) showed an association with the decision to upload, whereas none of the control variables had an effect. These variables were consequently removed from the model for further analyses.

Uploading Behavior

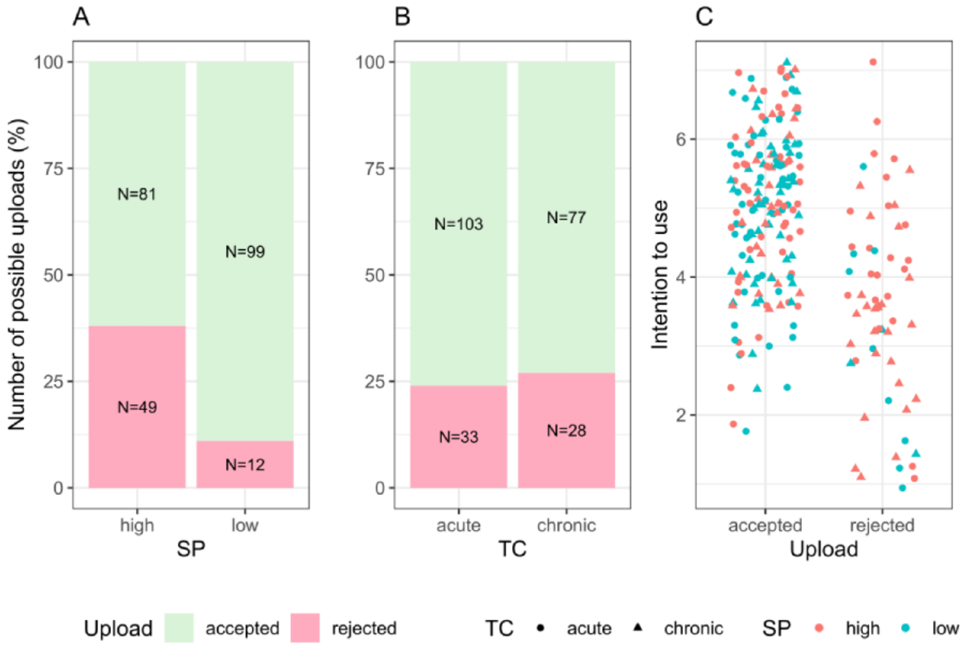
To examine the association between the decision to upload and the independent variables stigma potential and time course, we performed a logistic regression controlling for the covariate intention to use (Hosmer-Lemeshow R^2=0.289, Nagelkerke R^2=0.551, Cox-Snell R^2=0.501; χ^2_3=78.748; P<.001). Intention to use was positively associated with uploading behavior; specifically, as intention to use increased, it was more than twice as likely that the report was uploaded to the EHR. In addition, there was a negative association between stigma and the decision to upload; specifically, when stigma was high, it was six times less likely that the report was uploaded than when stigma was low. Time course of the disease was not associated with the decision to upload a report. The summary of the results of the logistic regression are shown in Table 3.

Table 3. Results of the logistic regression.

Variable	*z* value	*P* value	Odds ratio (95% CI)
Intention to use	6.210	<.001	2.682 (1.971-3.639)
Stigma potential	4.463	<.001	0.154 (0.064-0.336)
Time course	0.244	.81	1.093 (0.537-2.254)

The number of uploads is shown in Figure 4 in relation to the independent variables stigma potential (Figure 4A) and time course (Figure 4B). In addition, we show the relationship between intention to use and the decision to upload a report as a function of the independent variables in Figure 4C.

Figure 4. Number of uploads to the electronic health record with respect to (A) stigma potential (SP) and (B) time course (TC), and (C) the influence of intention to use on the decision to upload as a function of SP and TC.



Discussion

Principal Findings

Our results show that the decision to upload an individual medical report is influenced by people's intention to use the EHR. However, the stigma potential of the diseases mentioned in the reports also influenced this decision. Specifically, uploading diseases with high stigma was associated with higher privacy risk than diseases with low stigma (see Figure 3A). Consequently, stigma potential had a negative influence on the decision to upload records (see Figure 4A), despite generally high intentions to use the EHR.

Thus, intention to use predicts the use of the EHR in part, whereas disease-specific factors such as related stigma can override the general intention. This is particularly evident in Figure 4C where the participants who uploaded reports both with high and low stigma had mostly high intentions to use the EHR (scores>4). However, such a clear distribution of intention to use (scores<4) did not emerge in the case of rejection of uploading. Rather, it is notable that the rejected findings are mainly those with high stigmatization potential (majority of red dots/triangles in Figure 4C). This shows that the effect strength of the stigmatization potential (OR 0.154) is significantly greater than that of the intention to use (OR 2.628). The fact that uploading is rejected due to disease-specific stigma despite high intention to use supports the assumptions of an intention-behavior gap in EHR use [18].

The time course of the diseases had no influence on the decision to upload an individual record. Findings with chronic and acute diseases were uploaded by the majority of participants and with approximately equal frequency (see Figure 3B).

For both use cases, case 1 (patients with multimorbidity) and case 2 (patients after first contact), the results suggest that disease-specific stigma seems to exert an inhibiting influence on the decision to upload. In contrast, the time course only played a role in use case 1, where people interact with multiple reports at a time [31], but not when they interact with only one medical report (use case 2). This difference may be explained by the fact that patients' "health concerns" have a positive influence on their intention to share health data with others [22]. When faced with multiple medical reports, patients may be more aware and concerned about interactions between chronic diseases, because they more strongly affect the patient's health both now and in the future; consequently, the willingness to upload reports about chronic diseases increases. With a single report, interactions between diseases are less present, which means that the time course of a disease may play a reduced role in the decision to upload a record.

Implications

Both the intention to use and the stigma potential of diseases seem to influence whether patients upload an individual medical report to the EHR. Thus, in addition to increasing people's general intention to use the EHR via marketing and information, transparent and easy-to-comprehend information about the safety standards of the EHR (eg, for encrypting data) and the protection of medical records (eg, the control of access rights) are warranted, even for populations that are already in favor of using the EHR. Such combined interventions may help to reduce security concerns and enable realistic risk assessments of a data leak to ultimately ensure reliable use of the EHR as a key technology in any digitalized health care system.

Limitations and Future Directions

We deliberately excluded participants who already had a medical history with the diseases addressed in the stimuli to avoid bias in their responses. Individuals living with a stigmatized disease are more cautious to disclose the information, especially if the disease is not immediately apparent [32,43]. The question arises to what extent the behavior of stigmatized individuals can be simulated under experimental conditions, provided that participants do not exhibit stigmatized characteristics. To further strengthen the validity and generalizability of our results, a follow-up study should examine the perspective of already affected individuals and compare the findings with the results of this study.

Another limitation is that the chronic and acute disease patterns used in the stimuli are not readily comparable. We decided to use the diseases listed in Table 1 as stimuli because they achieved the expected effects in the preliminary study [31]. We could only partially replicate these findings in the present study. For future studies, it would make sense to use diseases that can be more readily compared in terms of their stigma potential and time course (eg, gonorrhea and HIV or a wrist fracture and arthritis) to further strengthen the generalizability of the present findings.

Another limitation is that the distribution of our sample in terms of gender, age, and level of education does not correspond to that of the average German population. In particular, the level of education of our sample was above average. Although we were unable to detect any effects of the control variables age, gender, and level of education in the analysis, the results of this study should be validated with a more representative sample in the future.

Conclusions

In our study, we investigated which disease-specific factors influence whether medical reports are uploaded to the EHR in a German setting. To answer this question, we varied the stigma potential and the time course of diseases in medical reports and controlled for the influence of participants' intention to use the EHR on uploading behavior. We demonstrated that intention to use had a positive effect on the decision to upload a report. In addition, we found that the stigma potential of the disease listed on the medical reports can inhibit uploading behavior. In particular, we found that the intention to use the EHR may be offset by the stigma potential of a specific record.

In summary, despite the fact that 3 out of 4 Germans state that they intend to use the EHR [11], actual use of this technology may depend on disease-specific factors. Consequently, to ensure successful implementation of the EHR, stakeholders in the health system should not only promote the EHR per se but further develop formats and evaluate them with the help of user testing that provide transparent and easy-to-comprehend information about the standards of data security and control in the EHR. Only in this way can users realistically assess the risks associated with individual EHR use and make an informed decision for (or against) EHR use.

Acknowledgments

We acknowledge support from the German Research Foundation and the Open Access Publication Fund of Technische Universität Berlin. We also thank the eV Studienwerk Villigst and the German Federal Ministry of Education and Research, who provided the doctoral scholarship (to NvK) without which this research would not have been possible. We would also like to thank the institutions that provided us with medical findings as well as all those who participated in the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Medical findings used as stimuli.
[PDF File (Adobe PDF File), 377 KB - humanfactors_v11i1e52625_app1.pdf]

Multimedia Appendix 2
Questionnaire.
[DOCX File , 17 KB - humanfactors_v11i1e52625_app2.docx]

Multimedia Appendix 3
Disease descriptions.
[DOCX File , 13 KB - humanfactors_v11i1e52625_app3.docx]

Multimedia Appendix 4
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 26097 KB - humanfactors_v11i1e52625_app4.pdf]

References

1. Haleem A, Javaid M, Pratap Singh R, Suman R. Medical 4.0 technologies for healthcare: features, capabilities, and applications. Internet of Things Cyber-Phys Syst 2022;2:12-30. [doi: 10.1016/j.iotcps.2022.04.001]

2. Hubmann M, Pätzmann-Sietas B, Morbach H. Telemedizin und digitale Akte – Wo stehen wir? Monatsschr Kinderheilkd 2021 Jul 09;169(8):711-716. [doi: 10.1007/s00112-021-01241-6]
3. Kraus S, Schiavone F, Pluzhnikova A, Invernizzi AC. Digital transformation in healthcare: analyzing the current state-of-research. J Bus Res 2021 Feb;123:557-567. [doi: 10.1016/j.jbusres.2020.10.030]
4. Bertram N, Püschner F, Gonçalves A, Binder S, Amelung V. Introduction of an electronic health record in Germany against the background of international experience. In: Klauber J, Geraedts M, Friedrich J, Wasem J, editors. Krankenhaus-Report 2019. Das digitale Krankenhaus. Berlin, Heidelberg: Springer; 2019:3-16.
5. Die elektronische Patientenakte (ePA). Bundesministerium für Gesundheit (BMG). 2021. URL: https://www.bundesgesundheitsministerium.de/elektronische-patientenakte.html [accessed 2023-04-14]
6. Elektronische Patientenakte - ePA. Kassenärztliche Bundesvereinigung (KBV). 2023. URL: https://www.kbv.de/html/epa.php [accessed 2023-04-14]
7. Baron von Osthoff M, Watzlaw-Schmidt U, Lehmann T, Hübner J. Patient group-specific data sovereignty needs and acceptance of electronic health records. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2022 Nov 23;65(11):1197-1203 [FREE Full text] [doi: 10.1007/s00103-022-03589-w] [Medline: 36149472]
8. Cherif E, Mzoughi M. Electronic health record adopters: a typology based on patients' privacy concerns and perceived benefits. Public Health 2022 Jun;207:46-53. [doi: 10.1016/j.puhe.2022.03.010] [Medline: 35486983]
9. Pagliari C, Detmer D, Singleton P. Potential of electronic personal health records. BMJ 2007 Aug 18;335(7615):330-333 [FREE Full text] [doi: 10.1136/bmj.39279.482963.AD] [Medline: 17703042]
10. Patientendaten-Schutz-Gesetz. Bundesministerium für Gesundheit (BMG). 2020. URL: https://www.bundesgesundheitsministerium.de/patientendaten-schutz-gesetz.html [accessed 2023-04-14]
11. Elektronische patientenakte: geplante Wwderspruchslösung trifft auf breite zustimmung. Bertelsmann Stiftung. 2023. URL: https://tinyurl.com/mptd2j95 [accessed 2023-05-17]
12. Bettiga D, Lamberti L, Lettieri E. Individuals' adoption of smart technologies for preventive health care: a structural equation modeling approach. Health Care Manag Sci 2020 Jun 26;23(2):203-214. [doi: 10.1007/s10729-019-09468-2] [Medline: 30684067]
13. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quart 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
14. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quart 2003;27(3):425-478. [doi: 10.2307/30036540]
15. Kim B, Kim D. Understanding the key antecedents of users' disclosing behaviors on social networking sites: the privacy paradox. Sustainability 2020 Jun 24;12(12):5163. [doi: 10.3390/su12125163]
16. Kim B, Min J. The distinct roles of dedication-based and constraint-based mechanisms in social networking sites. Internet Res 2013;25(1):30-51. [doi: 10.1108/intr-11-2013-0253]
17. Oliveira T, Faria M, Thomas MA, Popovič A. Extending the understanding of mobile banking adoption: when UTAUT meets TTF and ITM. Int J Inf Manag 2014 Oct;34(5):689-703. [doi: 10.1016/j.ijinfomgt.2014.06.004]
18. Conner M, Norman P. Understanding the intention-behavior gap: the role of intention strength. Front Psychol 2022 Aug 4;13:923464 [FREE Full text] [doi: 10.3389/fpsyg.2022.923464] [Medline: 35992469]
19. Sheeran P, Webb TL. The intention–behavior gap. Soc Person Psych 2016 Sep 05;10(9):503-518. [doi: 10.1111/spc3.12265]
20. Kokolakis S. Privacy attitudes and privacy behaviour: a review of current research on the privacy paradox phenomenon. Comput Secur 2017 Jan;64:122-134. [doi: 10.1016/j.cose.2015.07.002]
21. Norberg PA, Horne DR, Horne DA. The privacy paradox: personal information disclosure intentions versus behaviors. J Consum Aff 2007 Mar 06;41(1):100-126. [doi: 10.1111/j.1745-6606.2006.00070.x]
22. Cherif E, Bezaz N, Mzoughi M. Do personal health concerns and trust in healthcare providers mitigate privacy concerns? Effects on patients' intention to share personal health data on electronic health records. Soc Sci Med 2021 Aug;283:114146. [doi: 10.1016/j.socscimed.2021.114146] [Medline: 34242891]
23. von Kalckreuth N, Feufel MA. Extending the privacy calculus to the mHealth domain: survey study on the intention to use mHealth apps in Germany. JMIR Hum Factors 2023 Aug 16;10:e45503 [FREE Full text] [doi: 10.2196/45503] [Medline: 37585259]
24. Zhu M, Wu C, Huang S, Zheng K, Young SD, Yan X, et al. Privacy paradox in mHealth applications: an integrated elaboration likelihood model incorporating privacy calculus and privacy fatigue. Telematics Informatics 2021 Aug;61:101601. [doi: 10.1016/j.tele.2021.101601]
25. Dinev T, Albano V, Xu H, D'Atri A, Hart P. Individuals' attitudes towards electronic health records: a privacy calculus perspective. In: Gupta A, Patel VL, Greenes RA, editors. Advances in Healthcare Informatics and Analytics. Cham: Springer International Publishing; 2016:19-50.
26. Li T, Slee T. The effects of information privacy concerns on digitizing personal health records. J Assn Inf Sci Tech 2014 Apr 21;65(8):1541-1554. [doi: 10.1002/asi.23068]
27. Kim SS, Malhotra NK, Narasimhan S. Research Note—Two competing perspectives on automatic use: a theoretical and empirical comparison. Inf Syst Res 2005 Dec;16(4):418-432. [doi: 10.1287/isre.1050.0070]

28. Pavlou PA. Consumer acceptance of electronic commerce: integrating trust and risk with the technology acceptance model. Int J Electron Commer 2014 Dec 23;7(3):101-134. [doi: 10.1080/10864415.2003.11044275]
29. Schretzlmaier P, Hecker A, Ammenwerth E. Extension of the Unified Theory of Acceptance and Use of Technology 2 model for predicting mHealth acceptance using diabetes as an example: a cross-sectional validation study. BMJ Health Care Inform 2022 Nov 15;29(1):e100640 [FREE Full text] [doi: 10.1136/bmjhci-2022-100640] [Medline: 36379608]
30. Distler V, Fassl M, Habib H, Krombholz K, Lenzini G, Lallemand C, et al. Empirical research methods in usable privacy and security. In: Gerber N, Stöver A, Marky K, editors. Human Factors in Privacy Research. Cham: Springer International Publishing; 2023:29-53.
31. von Kalckreuth N, Prümper AM, Feufel MA. The Influence of health data on the use of the electronic health record (EHR) -- a mixed methods approach. 2023 Presented at: AMCIS 2023; August 10-12, 2023; Panama City, Panama URL: https://aisel.aisnet.org/amcis2023/sig_health/sig_health/2
32. Goffman E. Stigma: Notes on the Management of Spoiled Identity. Saddle River, NJ: Prentice Hall; 1963.
33. Prolific. URL: https://www.prolific.co/ [accessed 2023-09-06]
34. Peer E, Brandimarte L, Samat S, Acquisti A. Beyond the Turk: alternative platforms for crowdsourcing behavioral research. J Exp Soc Psychol 2017 May;70:153-163. [doi: 10.1016/j.jesp.2017.01.006]
35. Glozier N. Workplace effects of the stigmatization of depression. J Occup Environ Med 1998 Sep;40(9):793-800. [doi: 10.1097/00043764-199809000-00008] [Medline: 9777563]
36. Strømsvik N, Råheim M, Oyen N, Engebretsen LF, Gjengedal E. Stigmatization and male identity: Norwegian males' experience after identification as BRCA1/2 mutation carriers. J Genet Couns 2010 Aug 20;19(4):360-370. [doi: 10.1007/s10897-010-9293-1] [Medline: 20306122]
37. Bickford J, Barton SE, Mandalia S. Chronic genital herpes and disclosure.... The influence of stigma. Int J STD AIDS 2007 Sep 25;18(9):589-592. [doi: 10.1258/095646207781568484] [Medline: 17784999]
38. Zacks S, Beavers K, Theodore D, Dougherty K, Batey B, Shumaker J, et al. Social stigmatization and hepatitis C virus infection. J Clin Gastroenterol 2006 Mar;40(3):220-224. [doi: 10.1097/00004836-200603000-00009] [Medline: 16633123]
39. Wysocki AC, Lawson KM, Rhemtulla M. Statistical control requires causal justification. Adv Methods Pract Psychol Sci 2022 Jun 28;5(2):251524592210958. [doi: 10.1177/25152459221095823]
40. Schomakers E, Lidynia C, Vervier LS, Calero Valdez A, Ziefle M. Applying an extended UTAUT2 model to explain user acceptance of lifestyle and therapy mobile health apps: survey study. JMIR Mhealth Uhealth 2022 Jan 18;10(1):e27095 [FREE Full text] [doi: 10.2196/27095] [Medline: 35040801]
41. Uncovska M, Freitag B, Meister S, Fehring L. Patient acceptance of prescribed and fully reimbursed mHealth apps in Germany: an UTAUT2-based online survey study. J Med Syst 2023 Jan 27;47(1):14 [FREE Full text] [doi: 10.1007/s10916-023-01910-x] [Medline: 36705853]
42. Cuschieri S. The CONSORT statement. Saudi J Anaesth 2019 Apr;13(Suppl 1):S27-S30 [FREE Full text] [doi: 10.4103/sja.SJA_559_18] [Medline: 30930716]
43. Weiss MG, Ramakrishna J, Somma D. Health-related stigma: rethinking concepts and interventions. Psychol Health Med 2006 Aug;11(3):277-287. [doi: 10.1080/13548500600595053] [Medline: 17130065]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trails
EHR: electronic health record
OR: odds ratio
TAM: technology acceptance model
TU: Technische Universität
UTAUT: unified theory of acceptance and use of technology

Edited by P Santana-Mancilla; submitted 10.09.23; peer-reviewed by JJ Beunza, EM Schomakers; comments to author 07.02.24; revised version received 12.02.24; accepted 21.02.24; published 10.04.24.

Please cite as:
von Kalckreuth N, Feufel MA
Influence of Disease-Related Stigma on Patients' Decisions to Upload Medical Reports to the German Electronic Health Record: Randomized Controlled Trial
JMIR Hum Factors 2024;11:e52625
URL: https://humanfactors.jmir.org/2024/1/e52625
doi:10.2196/52625
PMID:38598271

Original Paper

Academic Detailing as a Health Information Technology Implementation Method: Supporting the Design and Implementation of an Emergency Department–Based Clinical Decision Support Tool to Prevent Future Falls

Hanna J Barton[1], PhD; Apoorva Maru[1], BA, BSc; Margaret A Leaf[2], MSc; Daniel J Hekman[1], MSc; Douglas A Wiegmann[3], PhD; Manish N Shah[1], MD, MPH; Brian W Patterson[1,3], MD, MPH

[1]BerbeeWalsh Department of Emergency Medicine, University of Wisconsin-Madison, Madison, WI, United States

[2]Department of Information Services, UW Health, Madison, WI, United States

[3]Department of Industrial and Systems Engineering, University of Wisconsin-Madison, Madison, WI, United States

Corresponding Author:
Hanna J Barton, PhD
BerbeeWalsh Department of Emergency Medicine
University of Wisconsin-Madison
800 University Bay Dr.
Madison, WI, 53705
United States
Phone: 1 (608) 890 8682
Email: hbarton@wisc.edu

Abstract

Background: Clinical decision support (CDS) tools that incorporate machine learning–derived content have the potential to transform clinical care by augmenting clinicians' expertise. To realize this potential, such tools must be designed to fit the dynamic work systems of the clinicians who use them. We propose the use of academic detailing—personal visits to clinicians by an expert in a specific health IT tool—as a method for both ensuring the correct understanding of that tool and its evidence base and identifying factors influencing the tool's implementation.

Objective: This study aimed to assess academic detailing as a method for simultaneously ensuring the correct understanding of an emergency department–based CDS tool to prevent future falls and identifying factors impacting clinicians' use of the tool through an analysis of the resultant qualitative data.

Methods: Previously, our team designed a CDS tool to identify patients aged 65 years and older who are at the highest risk of future falls and prompt an interruptive alert to clinicians, suggesting the patient be referred to a mobility and falls clinic for an evidence-based preventative intervention. We conducted 10-minute academic detailing interviews (n=16) with resident emergency medicine physicians and advanced practice providers who had encountered our CDS tool in practice. We conducted an inductive, team-based content analysis to identify factors that influenced clinicians' use of the CDS tool.

Results: The following categories of factors that impacted clinicians' use of the CDS were identified: (1) aspects of the CDS tool's design (2) clinicians' understanding (or misunderstanding) of the CDS or referral process, (3) the busy nature of the emergency department environment, (4) clinicians' perceptions of the patient and their associated fall risk, and (5) the opacity of the referral process. Additionally, clinician education was done to address any misconceptions about the CDS tool or referral process, for example, demonstrating how simple it is to place a referral via the CDS and clarifying which clinic the referral goes to.

Conclusions: Our study demonstrates the use of academic detailing for supporting the implementation of health information technologies, allowing us to identify factors that impacted clinicians' use of the CDS while concurrently educating clinicians to ensure the correct understanding of the CDS tool and intervention. Thus, academic detailing can inform both real-time adjustments of a tool's implementation, for example, refinement of the language used to introduce the tool, and larger scale redesign of the CDS tool to better fit the dynamic work environment of clinicians.

(JMIR Hum Factors 2024;11:e52592) doi:10.2196/52592

KEYWORDS

emergency medicine; clinical decision support; health IT; human factors; work systems; SEIPS; Systems Engineering Initiative for Patient Safety; educational outreach; academic detailing; implementation method; department-based; CDS; clinical care; evidence-based; CDS tool; gerontology; geriatric; geriatrics; older adult; older adults; elder; elderly; older person; older people; preventative intervention; team-based analysis; machine learning; high-risk patient; high-risk patients; pharmaceutical; pharmaceutical sales; United States; fall-risk prediction; EHR; electronic health record; interview; ED environment; emergency department

Introduction

Background

New technologies incorporating machine learning–derived content into clinical decision support (CDS) have the potential to bring transformative improvements to clinical care [1-3]. Identifying high-risk patients who merit referral for preventative care services has historically required cumbersome screening, but now can be rapidly completed by risk prediction algorithms that consider the patient's entire electronic health record (EHR) [4-6]. By incorporating machine learning–derived content, clinicians' decision-making can be augmented by insights that may otherwise go unnoticed. Yet the potential benefits of these CDS tools will only be realized when they are designed to fit the clinical contexts in which clinicians work [3,7]. Health information technologies (HITs), including CDS tools, that fail to fit clinicians' decision-making processes and workflows are unlikely to be adopted and even risk increasing clinician burden and burnout [8-10].

However, even technologies that are designed using today's best usability guidance [11] often fail to fit the clinical context upon initial implementation [12,13]. As health systems continue to evolve in response to emergent patient needs and expectations (eg, COVID-19 and its aftermath), regulatory requirements, and staffing challenges, CDS tools are being implemented in increasingly sensitive and complex environments. While implementation science frameworks consider a variety of contextual factors [14-16] and some methods exist for assessing and identifying them [17,18], there is a gap in methods for rapidly identifying contextual factors immediately postimplementation—when it may be easiest to respond to and redesign for emergent barriers to the technology's use, safety, and effectiveness [19].

One method that has the potential to be adapted to rapidly identify contextual factors influencing the implementation of HIT is *academic detailing*. A repurposing of pharmaceutical sales representatives' tactics, academic detailing is defined as a "personal visit by a trained person to health professionals in their own settings" [20]. The goal of these personal visits is to improve care quality and patient outcomes by promoting evidence-based practice through focused clinician education [21]. As an implementation method, academic detailing can be conceptualized as a combination of 3 Expert Recommendations for Implementing Change (ERIC; also known as Evidence-based Recommendations for Implementing Change) strategies: auditing and providing feedback, conducting educational outreach visits, and practice facilitation [14]. The method's attention to the specific contexts in which clinicians make decisions—both by conducting visits in situ and by discussing barriers to and strategies for making evidence-based decisions—may present a unique opportunity to not only promote the use of a newly implemented HIT but also identify contextual factors influencing its initial implementation.

Study Objective

We propose the use of academic detailing as a method for achieving two goals in the implementation of an emergency department (ED)–based CDS tool to prevent future falls: (1) ensuring the correct understanding of the tool and its evidence base and (2) identifying contextual factors influencing the tool's initial implementation. As part of a long-term goal of assessing academic detailing for achieving these 2 aims, the objective of this study was to assess academic detailing through an analysis of the resultant qualitative data.

Methods

Study Context and Setting

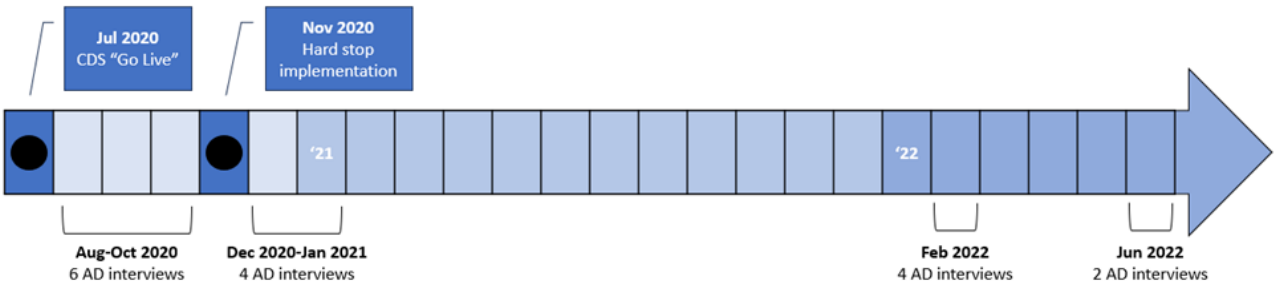
This study was conducted at a large academic medical center located in the Midwestern United States. The associated ED, a level 1 trauma center, treats over 60,000 patients per year. The CDS tool being evaluated is intended to facilitate both screening for outpatient fall risk among older adults presenting to the ED and the referral to a fall prevention clinic for those patients at high risk. Our research team developed an outpatient fall risk prediction algorithm from EHR data and, in concert with our partner health system, designed and implemented a CDS tool to use the algorithm that went live in July 2020 [22,23]. In November 2020, the CDS was updated such that it enforced a "hard stop" in the clinician's workflow and required them to interact with it.

Upon arrival to the ED, all patients aged 65 years and older with an in-system primary care provider are assessed for fall risk algorithmically based on their extant EHR data. For eligible patients who are at high risk for falls, during the discharge process, an interruptive CDS alert is shown to clinicians, which informs them of the patient's risk factors and expedites the placement of a referral order to a mobility and falls clinic, an evidence-based preventive intervention. Patients who are referred are informed both by the nursing staff and in writing and are contacted to schedule an appointment by scheduling staff in the days following their ED visit. This intervention has been described in more detail elsewhere [23,24].

Study Design

To assess academic detailing as a method for simultaneously achieving the goals of ensuring the CDS was understood and identifying contextual factors influencing its implementation, we used a qualitative approach. We conducted 16 semistructured academic detailing interviews with emergency medicine resident physicians (n=10) and advanced practice providers (n=6) who had previously encountered our CDS tool in practice, that is, within the last month. All interviews took place between August 2020 and June 2022, with 6 of the 16 interviews occurring prior to the implementation of the CDS hard stop (Figure 1). We purposively selected a range of participants based on how frequently they responded to the CDS. The academic detailing interviews were led by an intervention expert (AM) who had a comprehensive understanding of the CDS tool and thus was able to identify and correct any misconceptions about the tool and its use—a critical aspect of effective academic detailing [21].

Figure 1. Timeline of clinical decision support implementation and academic detailing interviews. AD: academic detailing; CDS: clinical decision support.



Study Procedure

Interviews were roughly 10 minutes long and took place over the phone or in person while the clinician was on shift. The intervention expert used an interview guide developed using the critical incident technique, which asks the participants to mentally put themselves in the moment they first saw the tool in the EHR [25]. The interview guide (Multimedia Appendix 1) contained questions such as "How and when did you see the tool initially? What was your reaction? How did you make the decision to refer the patient or not?" Additionally, comments made by a participant that suggested an incomplete or inaccurate understanding of the tool were addressed by the intervention expert (eg, "the tool refers patients to the Mobility and Falls clinic, not the Faint and Falls clinic"). Phone interviews were transcribed in real time, while notes were taken during in-person interviews and then written out immediately after.

Ethical Considerations

This study was reviewed and determined to be exempt by the UW-Madison Health Sciences IRB (ID# 2020-1100). Participants were not compensated, and data were deidentified for analysis.

Data Analysis

We conducted an inductive, team-based content analysis [26,27]. Two researchers (AM and MAL) began by independently reviewing and coding 4 interviews, line-by-line, to identify factors that influenced clinicians' decision-making. The researchers then met to compare and refine codes until there was agreement. This process continued iteratively until all interviews were coded; the resultant codebook contained 31 codes and subcodes (eg, patient risk factors and clinician communication). Another researcher (HJB) generated categories of factors that influenced clinicians' referral or nonreferral from the codes through a process of organizing similar and dissimilar codes, periodically incorporating feedback from the research team, until there was agreement.

Results

Overview

We identified five categories of factors that impacted clinicians' use of the CDS: (1) aspects of the CDS tool's design, for example, its features, usability, and how it fits in the clinician's workflow; (2) clinicians' understanding (or misunderstanding) of the CDS or referral process; (3) the busy nature of the ED environment; (4) clinicians' perceptions of the patient and their associated fall risk; and (5) the opacity of the referral process. Table 1 organizes the identified factors by these categories, including a description of the type of clinician education that was done during the academic detailing interviews.

Table 1. Factors impacting clinicians' CDS[a] use identified through academic detailing and clinician education done during academic detailing interviews. Positive and negative factors are indicated by the +, +/–, and – symbols.

Categories of factors impacting clinicians' CDS use through academic detailing	Factors impacting clinicians' CDS use identified through academic detailing	Clinician education done during academic detailing interviews
Aspects of the CDS tool's design	[+] CDS is simple. [+] CDS requires minimal input from the clinician. [+] CDS automatically identifies a high-risk patient and prompts care that the clinician would not otherwise have considered. [+/–] CDS enforces a hard stop in the clinician's workflow. [+/–] CDS alert fires while the clinician is completing discharge in the EHR[b].	• Discussed why the CDS alert fires when it does and the potential benefits and challenges of it firing at a different point in the clinician's workflow.
Clinicians' understanding of the CDS or referral process	[–] Clinician confuses the geriatric mobility and falls clinic with the faint and falls clinic. [–] Clinician believes only patients being seen for a fall are appropriate referrals. [–] Clinician believes referring the patient will be cumbersome, ie, require written justification. [+] Clinician is familiar with the concept of the CDS from an organizational stakeholder's communication.	• Clarified which clinic the referral goes to. • Clarified that referral is appropriate preventative care for patients regardless of their presenting problem. • Demonstrated how simple and quick it is to place a referral via the CDS.
Busy nature of the ED[c] environment	[+/–] A busy ED environment.	
Clinicians' perceptions of the patient and their associated fall risk	[+/–] Clinicians' agreement with the CDS's assessment of the patient's fall risk. [–] Clinicians' perception of the patient's openness to, need for, or benefit from the intervention.	• Demonstrated where in the CDS to find the reasons the patient is being flagged as high risk. • Stressed the potential benefits of a successful referral for both the patient and health system.
Opacity of the referral process	[–] Clinicians lack clarity on where the referral goes once it is sent. [–] Clinicians are uncertain about who should communicate with the patient about the referral, ie, themselves or a nurse. [+/–] Clinicians (do not) have the information necessary for counseling patients on what to expect from the referral and why they are being referred.	• Clarified which clinic the referral goes to. • Clarified the importance of counseling patients on the referral and demonstrated where in the CDS to access information to support counseling patients on the referral. • Demonstrated where in the CDS to find the reasons the patient was flagged as high risk.

[a]CDS: clinical decision support.

[b]EHR: electronic health record.

[c]ED: emergency department.

Aspects of the CDS Tool's Design

The first category of factors that influenced clinicians' use of the CDS was those that related to the design of the CDS tool. Many clinicians described the CDS as user friendly or easy to use, citing the limited number of clicks required and that the CDS did not require the clinician to enter any text. Further, clinicians found that the automatic nature of the CDS tool supported them in providing appropriate care that they otherwise would not have considered:

> *I appreciated how it fired on its own. I wasn't even thinking about falls in the patient because he came in for [condition], not a fall. When it fired, I realized he was a great candidate, but it wasn't something I thought about prior.* [Participant 12]

Further, clinicians described how the CDS integration into their workflow impacted their use of the CDS. For one, the CDS enforced a "hard stop," requiring the clinician to interact with it. While clinicians' feelings on the hard stop varied from being annoyed to finding it valuable, it was only described as impacting CDS use during high-volume times in the ED, which is discussed further in the next section. Clinicians generally appreciated where the CDS fit into their workflow—upon discharging the patient in the EHR—such that it "doesn't seem like an additional step" (participant 15). Some clinicians noted the timing of the CDS as being too late in the care process to have a personal discussion with their patient about the referral, for example, they are discharging their patient in the EHR after they have already done their final visit to the patient's room. Thus, when clinicians saw the CDS alert after their final visit with the patient, they described being less likely to refer the patient because of the additional time necessary to go back to the room to discuss the referral. However, most clinicians thought the CDS alert was well situated in their workflow:

> *It just pops up at discharge. It's just a click, the referral order is already filled out. It's very easy to use. It adds maybe 20 seconds to the discharge process.* [Participant 14]

Clinicians' Understanding of the CDS and Referral Process

The second category of factors that influenced clinicians' use of the CDS was clinicians' understanding, or misunderstanding, of the CDS and the referral process. One such misunderstanding of the CDS was clinicians confusing a separate faint and falls cardiology Clinic with the actual target of the referral, the mobility and falls clinic, which specifically addresses geriatric falls. Consequently, 2 clinicians cited their nonreferral as due to the inappropriateness of the faint and falls clinic for their geriatric patient:

> *I think more of the syncope patients and possible cardiac or peripheral vertigo patients who don't need to be admitted or are younger and are less high risk and are anxious about having syncopal episodes for the first time…. We want to get [them to] an outpatient visit so they'll be more likely to follow up with the clinic and have [care] done. I don't think of it as much who just have a mechanical fall.* [Participant 1]

Further, 6 clinicians described that they would not refer a patient who was being seen for another chief complaint, believing that only patients being seen for a fall are appropriate referrals.

Additionally, 3 clinicians expressed concerns about how cumbersome they believed the referral would be; however, their perception of the tool changed immediately once it was demonstrated that it only required 2 clicks. For example, participant 2 said:

> *The BPA would be less annoying if I knew I didn't need to justify it. If I had known that's all I had to do, I would have clicked [to accept the referral].*

One clinician stated that they expected that the CDS would be cumbersome, that is, that they would have to write out the referral because they thought "[the CDS] would be like the other referrals" that they had come across in the EHR (participant 8).

On the other hand, 1 factor in this category that positively impacted clinicians' use of the CDS was that the CDS was familiar to some clinicians given previous communication from an organizational stakeholder (ie, clinical champion). For example, participant 12 said:

> *When the CDS fired, I knew [clinical champion] sent an email about this. From my interactions with this patient, I thought [they were] at high risk of fall and knew [they would] benefit from it. That's why I tried to place that consult.*

Busy Nature of the ED Environment

The third category influencing CDS use was the busy nature of the ED environment. Five clinicians described the ED environment as a factor impacting their CDS use, whereas at least 1 clinician explicitly said, "the ED environment wouldn't affect whether or not I refer a patient" (participant 4). Clinicians varied in their description of the impact of the busy ED environment, ranging from "I would ignore the [CDS]" (participant 12) to "I just do the referral" (participant 14). Those clinicians who said that the ED environment increased their likelihood of referring the patient cited the CDS's hard stop and a significant amount of text as associated factors that shaped their decision-making. While other clinicians described the amount of text in the CDS as a stressor, it was not otherwise described as influencing clinicians' use of the CDS.

Clinicians' Perceptions of the Patient and Their Associated Fall Risk

The fourth category of factors influencing the use of the CDS was clinicians' perceptions of the patient and their associated fall risk. Overall, most clinicians (10/16) agreed with the CDS's assessment of the patient's fall risk. One clinician said:

> *In general, my reaction has been "oh that kinda makes sense." It was always kind of a surprise in the sense that I hadn't really considered the risk of falls before, but it never seems outlandish that that was a potential concern.* [Participant 3]

A few clinicians described instances of being annoyed by the firing of the CDS when it seemed irrelevant and thus did not use it. Conversely, another clinician was frustrated when the CDS did not fire when they expected to see it. Clinicians described occasionally not referring patients because they appeared to be "independent and functional" (participant 11) or "generally active and stable" (participant 10), or because their fall was "strictly mechanical" (participant 7).

Further, clinicians' perception of the patient's openness to or need for intervention impacted their CDS use. One clinician described factoring their assessment of the patient's openness to going to the mobility and falls clinic into their decision to refer the patient or not:

> *You can kind of get a vibe if someone is going to the doctor. If you tell them there's another doctor you can see; if they don't even want to talk to me, I doubt they're going to go to another doctor. If it's going to be useless, I don't want to waste everyone's time. I like to tell people and if they say I'm not going to that then I won't refer. I think I dictate if I'm going to [refer the patient] based on the conversation.* [Participant 8]

Clinicians also described being more likely to refer patients who they perceived as having a greater need for the intervention. For example, a clinician said, "If the patient seems more anxious or [they] don't have as good of a support system or advocate, I would refer them" (participant 1). Alternatively, clinicians also described assessing how the referral would fit into the patient's care plan, for example, if the patient had an upcoming surgery, the clinician would opt not to refer the patient so as to not "throw an extra thing on top of them" (participant 6).

Opacity of the Referral Process

The final category of factors influencing clinicians' use of the CDS was the opacity of the referral process. One clinician described how they lacked clarity on where the referral goes once it is sent by saying, "it feels like I'm just sending the referral off to the void and I don't know who they're getting referred to" (participant 3). In contrast to a clinician confusing the mobility and falls clinic with the faint and falls clinic, discussed previously, this clinician was specifically pointing to the lack of feedback they received about how the process of referring a patient to the mobility and falls clinic unfolds over time. Consequently, another factor clinicians said impacted their use of the CDS was the ambiguity around who should communicate with the patient about the referral: themselves or a nurse. Clinicians described this factor as being more prominent if they had already spoken to the patient and thus referring the patient would require them to initiate another conversation with the patient themselves or "hope the nurse will tell the patient" (participant 1).

Finally, a few clinicians said they lacked the necessary information to counsel the patient—either about what to expect from the referral and the mobility and falls clinic or about the reasons the patient had been flagged as high risk for falls. One clinician suggested that having guidance on how to counsel the patient might make referring patients an easier choice:

> *I haven't been really having detailed conversation about what this entails and what they should expect. In the moment I hadn't quite seen a link on how to counsel patients on this referral…I do want to refer patients…I just wish I knew what to tell patients.* [Participant 3]

Clinicians also described lacking sufficient information to explain to patients why they were flagged by the CDS as high risk for falls. In particular, clinicians said this information would likely influence their referring of patients being seen for chief complaints other than a fall by making it easier to explain the referral to the patient.

Clinician Education Done During Academic Detailing

During the academic detailing interviews, various misconceptions were addressed directly by the intervention expert through clinician education. First, 1 misconception described previously was the mistaken belief that a referral to the mobility and falls clinic would be inappropriate for people being seen for a chief complaint other than a fall. The intervention expert addressed this by clarifying that the CDS alert fires for any older adult being seen in the ED who is at high risk of falling in the future regardless of their presenting complaint, so barring any contraindications—for example, patient in hospice—it would be appropriate to refer the patient. The intervention expert also clarified, for clinicians who misunderstood, the correct target clinic of the referral (ie, the mobility and falls clinic). Generally, the intervention expert stressed the potential benefits of a successful referral for both the patient (eg, improved quality of life) and the health system (eg, reduced use).

Another misconception that was addressed via academic detailing was the perception that referring a patient would be too cumbersome. By demonstrating that accepting the CDS alert and placing a referral takes only 2 clicks, this misconception was promptly addressed. The intervention expert also demonstrated where in the CDS to access information to support counseling patients on the referral and where in the CDS to find the reasons the patient was flagged as high risk. Finally, for any clinicians who had issues with where the CDS alert fired in their workflow, the intervention expert discussed the reasons for the alert firing when it does and the potential benefits and challenges of it firing at a different time. Oftentimes, after discussion, the clinician had a new appreciation for the complexity of designing the CDS alert.

Discussion

Principal Findings

This study demonstrates the use of academic detailing for supporting the early implementation of HIT, allowing us to identify and begin to address factors that impacted clinicians' use of the CDS while concurrently educating clinicians to ensure the correct understanding of the CDS tool and intervention. By bundling multiple ERIC strategies, academic detailing appears to be a promising method for providing timely feedback to improve HIT implementation.

Addressing Contextual Factors Within Detailing Sessions

A key component of the academic detailing method is its emphasis on clinician education [21] which, in the context of our study, involves correcting clinicians' misconceptions. For example, 1 misconception that we identified and addressed through clinician education was the mistaken belief that a referral to the mobility and falls clinic was only appropriate for people being seen for a fall. Given the nature of this CDS tool, that is, its ability to predict future risk, the impact of this misconception is that the opportunities to intervene in the routine care of high-risk patients being treated for other chief complaints would be missed. As participant 12 articulated, quoted in the "aspects of the tools design" results section, a particular value of the CDS tool is that it runs automatically, that is, does not require clinician initiation; thus, it can prompt the clinician to consider fall risk—and care to address that risk—that they may not have been considering previously. Embedding clinician education into academic detailing thus addressed a high-impact misconception with immediacy. However, it remains to be seen whether and how addressing these misconceptions translates to clinicians' use of the CDS tool. Our future work will explore the impact of these academic detailing sessions on implementation incomes, for example, clinicians' rates of referral and their acceptance of the tool.

Another important misconception to address within the academic detailing interviews was the perception that referring a patient would be too cumbersome. By demonstrating the simplicity of accepting the CDS alert and placing a referral, this misconception was promptly addressed which likely prevented its propagation. However, as described in the Results section, clinicians' perceptions of the CDS tool are situated within the context of the existing EHR and thus are beholden to a broader understanding of how similar tools work (ie, a mental model) [28]. As such, clinicians' responses to CDS alerts can be understood to be habitual, triggered by environmental cues [29]; therefore, solely addressing this misconception at the clinician level is unlikely to sustain CDS use over time. Altering clinicians' mental models of CDS tools and the EHR warrants systems-level redesign.

The content of the clinician education that is included in academic detailing is paramount to its success in increasing the use of an intervention [21]. Previous literature also notes the importance of the relationship between the clinician and the person doing the clinician education [21,30]. For this study, the intervention expert who conducted the academic detailing interviews had extensive experience working with the ED staff and had developed a rapport with them. To carry out academic detailing in another setting, there may be initial relationship- and trust building to do to achieve the detailed results our intervention expert was able to capture. Yet, given their role as a researcher (vs a fellow clinician), there were potentially missed opportunities for educating clinicians on topics that would have been better received from a colleague. For example, the deeply entrenched custom of referring to many older adults' community-based falls as being "mechanical," a catch-all term for falls that does not have an emergent, addressable cause, is known to negatively affect care [31]. This could have potentially been addressed by a colleague; however, in this study, we did not address this clinician perception as it fell outside of the expertise of our intervention expert, that is, outside of the purview of the CDS tool and the referral it recommends.

Addressing Contextual Factors via Redesign

The factors impacting clinicians' use of the CDS point directly to opportunities to intervene in and improve the CDS implementation process (Textbox 1). As discussed previously, clinician education can be done immediately, within the academic detailing interview; however, the clinician education that had to be provided within the interview can inform the redesign of a better rollout (eg, addressing what are likely to be misconceptions up front). Future rounds of academic detailing should thus result in the need for less or different clinician education from the intervention expert.

Textbox 1. Potential approaches for intervening in the health information technology implementation process to improve clinical decision support acceptance and use.

Real time (within academic detailing interview)

- Demonstrating the current capabilities and function of the tool, for example, how easy it is to place a referral, where to access information about why the patient was flagged as high risk, and information to support counseling the patient on the referral.
- Discussing why the clinical decision support tool works the way it does and the potential benefits and challenges of redesigning it.
- Clarifying how the referral works, where it goes, and who is an appropriate candidate for the intervention.
- Addressing problematic or harmful misconceptions, for example, that there is no role the emergency department can play in providing preventive care after "mechanical falls."
- Discussing how successfully using the clinical decision support and placing a referral improves patient outcomes and health system outcomes, for example, by reducing future visits to the emergency department.

Short term (quick fixes)

- Attending regularly scheduled meetings with clinicians to remind them about the clinical decision support and clarify misconceptions about placing the referral.
- Associating the organizational stakeholder's name or image with the clinical decision support.
- Adding the mobility and falls clinic information to the clinical decision support, that is, the phone number and location.

Long term (adaptation and redesign)

- Developing feedback mechanisms for clinicians to hear about successfully referred patients.
- Clarifying roles around patient communication, that is, what is communicated by the clinician versus the nurse, and designing the clinical decision support to support those roles.
- Reviewing clinical decision support tools for potential interaction effects, for example, 2 clinical decision support tools fire on similar populations and are likely to be confused.
- Providing talking points on what the patient can expect after discharge with respect to scheduling and going to an appointment with the mobility and falls clinic.
- Providing talking points that explain why patients being seen for issues other than falls may be referred.
- Personalizing the timing of the clinical decision support alert for clinicians who tend to talk to patients before completing the discharge in the electronic health record, for example, moving the clinical decision support alert earlier in clinicians' workflow.

In the longer term, a variety of approaches could be used to address the factors we identified as impacting the clinicians' use of the CDS. For one, reviewing the CDS tools that are currently implemented in overlapping clinical contexts could identify potential interactions with the newly implemented CDS. To avoid interaction effects, the new CDS could be redesigned to differentiate it from others, for example, to alert a more specific patient population or to have clear visual cues and messaging. Alternatively, a review of the CDS ecosystem may prompt the removal of underused or ineffective CDS tools. Recent research, while limited, suggests that health systems that optimize CDS alerts, that is, reduce unnecessary or less useful alerts, see improved CDS use [32]. Further, those effects are not limited to the optimized CDS but spread to other CDS in the system [32].

Other redesigns that would address factors identified through academic detailing could address the workflow integration of the CDS. For example, for the clinicians who typically talk to a patient before completing the discharge in the EHR, moving the firing of the CDS alert earlier in the clinical workflow may be warranted. Beyond considering the timing of the CDS, to achieve workflow integration as defined by Salwei et al [33,34], the design of the CDS should consider the dimensions of flow, scope of patient journey, and level. An example of such a redesign could be—in the case where the fall risk CDS alert would happen earlier in the clinician's workflow—allowing the clinician to "snooze" the alert until the point at which they have discussed the mobility and falls clinic referral with the patient. This design would increase the chance that the clinician would see the CDS prior to speaking with the patient for the last time, which could promote more meaningful patient counseling on fall risk; however, this design could also have unintended consequences, which should be explored prior to broad implementation.

In designing for CDS use, it is important to remember that increased use does not always equate to increased *appropriate* use (ie, referrals for patients that are a good fit for the mobility and falls clinic intervention). Thus, the findings from academic detailing should also be considered in light of, and be used to design to support, successful teaming between the CDS tool and the clinician. A potential design to promote teamwork between the CDS tool and the clinician could be to include on the CDS a list of exclusion criteria for the mobility and falls clinic that the CDS tool is unable or poorly able to assess (eg, late-stage dementia). The clinician, then, when considering referring the patient would be alerted to where their clinical judgment is especially necessary.

Work Systems Approach to Redesign

Given the breadth of potential redesign options and the challenge of prioritizing efforts to improve not only CDS use, but *appropriate* CDS use, it is pertinent to consider models that can hold and make sense of system complexity. One model that has proven to be valuable across a variety of health care domains and in supporting the design of technologies—the Systems Engineering Initiative for Patient Safety (SEIPS) model—conceptualizes the work of clinicians as happening in a *work system*, which invariably influences care processes and outcomes [35-37]. The SEIPS model, which synthesizes literature on job stress, job design, and health care quality [37,38], provides a theoretical foundation for understanding why the system is achieving certain outcomes and how the system may be redesigned to achieve alternative outcomes.

In a parallel analysis—presented elsewhere [39]—we found that the data we collected using the academic detailing method successfully mapped to the SEIPS model's work system components, for example, the *people* who do the work, the *tasks* they complete, the *tools and technologies* they use, and the *physical* and *organizational environment* they work in. A key aspect of the SEIPS model is the conceptualization of balance—that work system components that negatively influence processes and outcomes (barriers) may be balanced by positive components (facilitators) [37,40]. Thus, through redesign efforts, we can either seek to address the work system barrier or enhance the work system facilitator. Applying a work systems approach to system redesign to address the factors we identified through academic detailing has the potential to result in more sustainable HIT implementation.

Beyond redesigning the CDS itself, as discussed in the previous section, redesigning the work system to clarify the process of referring a patient to the mobility and falls clinic may be essential to promoting the appropriate use of the CDS. This would require creating clarity around who should communicate and about what with the patient (ie, the referring clinician and the nurse). Further, creating transparency around the positive outcomes of past referrals to the mobility and falls clinic (ie, success stories) may promote trust in the referring clinicians that this is an action worth taking.

Limitations

The following limitations of our study should be considered. First, the academic detailing method, as applied here, relies on the clinicians to report what they perceive as influencing their use of the CDS. However, it is possible that clinicians' perceptions differ from what they actually do—a common challenge in understanding people's work is the difference between "work as imagined" versus "work as done" [41]. Second, this study focuses on academic detailing around a specific CDS tool that produces an interruptive alert to which a clinician must respond that they agree or decline to refer the patient. It is possible that there are other considerations for CDS tools and HIT that operate differently from this study (eg, tools that require more in-depth information processing or that must be initiated by the clinician). Further, given this academic detailing method was applied in a live ED setting over nearly 2 years—including multiple waves of COVID-19—a variety of external factors may have contributed to clinicians' use (and perception of their use) of the CDS. Finally, it is yet unclear how many rounds of academic detailing would be required to capture and address the majority of factors impacting the implementation of the HIT. Future research should explore the use of the academic detailing method over a broader range of the implementation process so that the effort and resources required to conduct the interviews are used most effectively.

Conclusions

With HIT developing at rapid speeds, it is essential we develop methods to support its integration into the complex environments in which they will be used. From our initial study, it appears that academic detailing is a promising method for both promoting the correct understanding of a CDS tool and identifying contextual factors influencing its implementation. Thus, academic detailing can inform real-time adjustments of a tool's implementation (eg, refinement of the language used to introduce the tool), and larger scale redesign of the CDS tool to better fit the dynamic work environment of clinicians.

Acknowledgments

The authors would like to acknowledge the clinicians who participated in our academic detailing interviews and shared their insights with us. This work was supported by the Agency for Healthcare Research and Quality (grant R18HS027735; BWP, principal investigator). This work is solely the responsibility of the authors and does not represent the official views of the Agency for Healthcare Research and Quality.

Authors' Contributions

HJB performed data analysis and led the writing and revising of the manuscript. AM performed data collection and analysis and critically reviewed and revised the manuscript. MAL performed data analysis and critically reviewed and revised the manuscript. DJH, DAW, and MNS critically reviewed and revised the manuscript. BWP conceptualized and oversaw the study and critically reviewed and revised the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide for academic detailing interviews.
[DOCX File , 22 KB - humanfactors_v11i1e52592_app1.docx]

References

1. Emanuel EJ, Wachter RM. Artificial intelligence in health care: will the value match the hype? JAMA 2019;321(23):2281-2282. [doi: 10.1001/jama.2019.4914] [Medline: 31107500]
2. Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. N Engl J Med 2016;375(13):1216-1219 [FREE Full text] [doi: 10.1056/NEJMp1606181] [Medline: 27682033]
3. Chen JH, Asch SM. Machine learning and prediction in medicine—beyond the peak of inflated expectations. N Engl J Med 2017;376(26):2507-2509 [FREE Full text] [doi: 10.1056/NEJMp1702071] [Medline: 28657867]
4. Carpenter CR, Griffey RT, Stark S, Coopersmith CM, Gage BF. Physician and nurse acceptance of technicians to screen for geriatric syndromes in the emergency department. West J Emerg Med 2011;12(4):489-495 [FREE Full text] [doi: 10.5811/westjem.2011.1.1962] [Medline: 22224145]
5. Donaldson MG, Khan KM, Davis JC, Salter AE, Buchanan J, McKnight D, et al. Emergency department fall-related presentations do not trigger fall risk assessment: a gap in care of high-risk outpatient fallers. Arch Gerontol Geriatr 2005;41(3):311-317. [doi: 10.1016/j.archger.2005.04.008] [Medline: 15982758]
6. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. J Am Med Inform Assoc 2017;24(1):198-208 [FREE Full text] [doi: 10.1093/jamia/ocw042] [Medline: 27189013]
7. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ 2005;330(7494):765 [FREE Full text] [doi: 10.1136/bmj.38398.500764.8F] [Medline: 15767266]
8. Carayon P, Hoonakker P, Hundt AS, Salwei M, Wiegmann D, Brown RL, et al. Application of human factors to improve usability of clinical decision support for diagnostic decision-making: a scenario-based simulation study. BMJ Qual Saf 2020;29(4):329-340 [FREE Full text] [doi: 10.1136/bmjqs-2019-009857] [Medline: 31776197]
9. Hoonakker PLT, Carayon P, Salwei ME, Hundt AS, Wiegmann D, Kleinschmidt P, et al. The design of PE Dx, a CDS to support pulmonary embolism diagnosis in the ED. Stud Health Technol Inform 2019;265:134-140 [FREE Full text] [doi: 10.3233/SHTI190152] [Medline: 31431589]
10. National Academies of Sciences, Engineering, and Medicine, National Academy of Medicine, Committee on Systems Approaches to Improve Patient Care by Supporting Clinician Well-Being. Taking Action Against Clinician Burnout: A Systems Approach to Professional Well-Being. Washington, DC: National Academies Press; 2019.
11. Zhang J, Johnson TR, Patel VL, Paige DL, Kubose T. Using usability heuristics to evaluate patient safety of medical devices. J Biomed Inform 2003;36(1-2):23-30 [FREE Full text] [doi: 10.1016/s1532-0464(03)00060-1] [Medline: 14552844]
12. Patterson BW, Pulia MS, Ravi S, Hoonakker PLT, Hundt AS, Wiegmann D, et al. Scope and influence of electronic health record-integrated clinical decision support in the emergency department: a systematic review. Ann Emerg Med 2019;74(2):285-296 [FREE Full text] [doi: 10.1016/j.annemergmed.2018.10.034] [Medline: 30611639]
13. Greenes RA, Bates DW, Kawamoto K, Middleton B, Osheroff J, Shahar Y. Clinical decision support models and frameworks: seeking to address research issues underlying implementation successes and failures. J Biomed Inform 2018;78:134-143 [FREE Full text] [doi: 10.1016/j.jbi.2017.12.005] [Medline: 29246790]

14. Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, Matthieu MM, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. Implement Sci 2015;10:21 [FREE Full text] [doi: 10.1186/s13012-015-0209-1] [Medline: 25889199]
15. Damschroder LJ, Reardon CM, Widerquist MAO, Lowery J. The updated consolidated framework for implementation research based on user feedback. Implement Sci 2022;17(1):75 [FREE Full text] [doi: 10.1186/s13012-022-01245-0] [Medline: 36309746]
16. Nilsen P, Bernhardsson S. Context matters in implementation science: a scoping review of determinant frameworks that describe contextual determinants for implementation outcomes. BMC Health Serv Res 2019;19(1):189 [FREE Full text] [doi: 10.1186/s12913-019-4015-3] [Medline: 30909897]
17. McCormack B, McCarthy G, Wright J, Slater P, Coffey A. Development and testing of the Context Assessment Index (CAI). Worldviews Evid Based Nurs 2009;6(1):27-35. [doi: 10.1111/j.1741-6787.2008.00130.x] [Medline: 19207560]
18. Estabrooks CA, Squires JE, Cummings GG, Birdsell JM, Norton PG. Development and assessment of the Alberta context tool. BMC Health Serv Res 2009;9:234 [FREE Full text] [doi: 10.1186/1472-6963-9-234] [Medline: 20003531]
19. Tyre MJ, Orlikowski WJ. Windows of opportunity: temporal patterns of technological adaptation in organizations. Organ Sci 1994;5(1):98-118. [doi: 10.1287/orsc.5.1.98]
20. O'Brien MA, Rogers S, Jamtvedt G, Oxman AD, Odgaard-Jensen J, Kristoffersen DT, et al. Educational outreach visits: effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2007;2007(4):CD000409 [FREE Full text] [doi: 10.1002/14651858.CD000409.pub2] [Medline: 17943742]
21. Yeh JS, Van Hoof TJ, Fischer MA. Key features of academic detailing: development of an expert consensus using the Delphi method. Am Health Drug Benefits 2016;9(1):42-50 [FREE Full text] [Medline: 27066195]
22. Patterson BW, Engstrom CJ, Sah V, Smith MA, Mendonça EA, Pulia M, et al. Training and interpreting machine learning algorithms to evaluate fall risk after emergency department visits. Med Care 2019;57(7):560-566 [FREE Full text] [doi: 10.1097/MLR.0000000000001140] [Medline: 31157707]
23. Jacobsohn GC, Leaf M, Liao F, Maru AP, Engstrom CJ, Salwei ME, et al. Collaborative design and implementation of a clinical decision support system for automated fall-risk identification and referrals in emergency departments. Healthc (Amst) 2022;10(1):100598 [FREE Full text] [doi: 10.1016/j.hjdsi.2021.100598] [Medline: 34923354]
24. Hekman DJ, Cochran AL, Maru AP, Barton HJ, Shah MN, Wiegmann D, et al. Effectiveness of an emergency department-based machine learning clinical decision support tool to prevent outpatient falls among older adults: protocol for a quasi-experimental study. JMIR Res Protoc 2023;12:e48128 [FREE Full text] [doi: 10.2196/48128] [Medline: 37535416]
25. Flanagan JC. The critical incident technique. Psychol Bull 1954;51(4):327-358. [doi: 10.1037/h0061470] [Medline: 13177800]
26. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
27. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
28. Holtrop JS, Scherer LD, Matlock DD, Glasgow RE, Green LA. The importance of mental models in implementation science. Front Public Health 2021;9:680316 [FREE Full text] [doi: 10.3389/fpubh.2021.680316] [Medline: 34295871]
29. Baysari MT, Tariq A, Day RO, Westbrook JI. Alert override as a habitual behavior—a new perspective on a persistent problem. J Am Med Inform Assoc 2017;24(2):409-412 [FREE Full text] [doi: 10.1093/jamia/ocw072] [Medline: 27274015]
30. Van Hoof TJ, Harrison LG, Miller NE, Pappas MS, Fischer MA. Characteristics of academic detailing: results of a literature review. Am Health Drug Benefits 2015;8(8):414-422 [FREE Full text] [Medline: 26702333]
31. Nagaraj G, Hullick C, Arendts G, Burkett E, Hill KD, Carpenter CR. Avoiding anchoring bias by moving beyond 'mechanical falls' in geriatric emergency medicine. Emerg Med Australas 2018;30(6):843-850. [doi: 10.1111/1742-6723.13129] [Medline: 30091183]
32. Chaparro JD, Beus JM, Dziorny AC, Hagedorn PA, Hernandez S, Kandaswamy S, et al. Clinical decision support stewardship: best practices and techniques to monitor and improve interruptive alerts. Appl Clin Inform 2022;13(3):560-568 [FREE Full text] [doi: 10.1055/s-0042-1748856] [Medline: 35613913]
33. Salwei ME, Carayon P, Hoonakker PLT, Hundt AS, Wiegmann D, Pulia M, et al. Workflow integration analysis of a human factors-based clinical decision support in the emergency department. Appl Ergon 2021;97:103498 [FREE Full text] [doi: 10.1016/j.apergo.2021.103498] [Medline: 34182430]
34. Salwei ME. Human-Centered Design of Health Information Technology for Workflow Integration. Madison: The University of Wisconsin; 2020.
35. Carayon P, Hundt AS, Karsh BT, Gurses AP, Alvarado CJ, Smith M, et al. Work system design for patient safety: the SEIPS model. Qual Saf Health Care 2006;15(Suppl 1):i50-i58 [FREE Full text] [doi: 10.1136/qshc.2005.015842] [Medline: 17142610]
36. Holden RJ, Carayon P, Gurses AP, Hoonakker P, Hundt AS, Ozok AA, et al. SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. Ergonomics 2013;56(11):1669-1686 [FREE Full text] [doi: 10.1080/00140139.2013.838643] [Medline: 24088063]

37. Smith MJ, Sainfort PC. A balance theory of job design for stress reduction. Int J Ind Ergon 1989;4(1):67-79. [doi: 10.1016/0169-8141(89)90051-6]
38. Donabedian A. The quality of care. How can it be assessed? JAMA 1988;260(12):1743-1748. [doi: 10.1001/jama.260.12.1743] [Medline: 3045356]
39. Barton HJ, Maru A, Leaf MA, Hekman D, Wiegmann DA, Shah MN. Academic detailing of emergency medicine clinicians to inform post-implementation continuous design of an algorithm-based clinical decision support tool to prevent falls. 2023 Presented at: 14th Organizational Design and Management Conference; 2023; Bordeaux, France URL: https://www.odam2023.org/_files/ugd/2a1c0b_27519dfeb37d45bd82144b59db223bee.pdf
40. Carayon P, Gurses AP, Hundt AS, Ayoub P, Alvarado CJ. Performance obstacles and facilitators of healthcare providers. In: Korunka C, Hoffmann P, editors. Change and Quality in Human Service Work. Munich, Germany: Hampp Publishers; 2005:257-276.
41. Hollnagel E, Clay-Williams R. Work-as-imagined and work-as-done. In: Rapport F, Clay-Williams R, Braithwaite J, editors. Implementation Science: The Key Concepts. London, UK: Routledge; 2022:175-177.

Abbreviations

CDS: clinical decision support
ED: emergency department
EHR: electronic health record
ERIC: Expert Recommendations for Implementing Change, also known as Evidence-based Recommendations for Implementing Change
HFE: human factors engineering
HIT: health information technology
SEIPS: systems engineering initiative for patient safety

Edited by A Kushniruk; submitted 11.09.23; peer-reviewed by K Liu, E Kerns; comments to author 30.12.23; revised version received 08.02.24; accepted 02.03.24; published 18.04.24.

Please cite as:
Barton HJ, Maru A, Leaf MA, Hekman DJ, Wiegmann DA, Shah MN, Patterson BW
Academic Detailing as a Health Information Technology Implementation Method: Supporting the Design and Implementation of an Emergency Department–Based Clinical Decision Support Tool to Prevent Future Falls
JMIR Hum Factors 2024;11:e52592
URL: https://humanfactors.jmir.org/2024/1/e52592
doi:10.2196/52592
PMID:38635318

Original Paper

Electronic Immunization Registry in Rwanda: Qualitative Study of Health Worker Experiences

Thaoussi Uwera[1*], BSc, MSc; Mahima Venkateswaran[2*], MBBS, PhD; Kiran Bhutada[3*], BS, MD; Eleni Papadopoulou[4*], PhD; Enock Rukundo[1*], BSc, MSc; David K Tumusiime[1*], BSc, MSc, PhD; J Frederik Frøen[2*], MD, PhD

[1]Centre of Excellence in Biomedical Engineering and eHealth, University of Rwanda, Kigali, Rwanda

[2]Centre for Intervention Science for Maternal and Child Health (CISMAC), University of Bergen, Bergen, Norway

[3]Global Health Center, Albert Einstein College of Medicine, Bronx, NY, United States

[4]Global Health Cluster, Division for Health Services, Norwegian Institute of Public Health, Oslo, Norway

[*]all authors contributed equally

Corresponding Author:
Thaoussi Uwera, BSc, MSc
Centre of Excellence in Biomedical Engineering and eHealth
University of Rwanda
KG 11 Ave Gasabo
Kigali
Rwanda
Phone: 250 788229321
Email: uwera04@gmail.com

Abstract

Background: Monitoring childhood immunization programs is essential for health systems. Despite the introduction of an electronic immunization registry called e-Tracker in Rwanda, challenges such as lacking population denominators persist, leading to implausible reports of coverage rates of more than 100%.

Objective: This study aimed to assess the extent to which the immunization e-Tracker responds to stakeholders' needs and identify key areas for improvement.

Methods: In-depth interviews were conducted with all levels of e-Tracker users including immunization nurses, data managers, and supervisors from health facilities in 5 districts of Rwanda. We used an interview guide based on the constructs of the Human, Organization, and Technology–Fit (HOT-Fit) framework, and we analyzed and summarized our findings using the framework.

Results: Immunization nurses reported using the e-Tracker as a secondary data entry tool in addition to paper-based forms, which resulted in considerable dissatisfaction among nurses. While users acknowledged the potential of a digital tool compared to paper-based systems, they also reported the need for improvement of functionalities to support their work, such as digital client appointment lists, lists of defaulters, search and register functions, automated monthly reports, and linkages to birth notifications and the national identity system.

Conclusions: Reducing dual documentation for users can improve e-Tracker use and user satisfaction. Our findings can help identify additional digital health interventions to support and strengthen the health information system for the immunization program.

(JMIR Hum Factors 2024;11:e53071) doi:10.2196/53071

KEYWORDS

childhood immunization; electronic immunization registry; digital health interventions

Introduction

In 2021, a reported 18.2 million infants worldwide did not receive basic immunization, and an additional 6.8 million were only partially vaccinated, with associated higher deaths in low- and middle-income countries (LMICs) [1,2]. Health systems worldwide are adopting digital tools to improve immunization service provision as well as monitoring [3]. Digital health interventions (DHIs) have the potential to improve the management and use of health information to enhance health worker performance and provision of care and ultimately improve health outcomes [4,5]. In LMICs, electronic immunization registries (EIRs) were initiated to support improved vaccination coverage among children, primarily through better tracking of children by combining vaccine information from different sources into a single digital record [6,7].

DHIs (in the form of EIRs) are important for immunization programs. For clients, they can help to remind families through SMS text messaging when immunization is due or has been missed. For health workers, they can help ensure that children get the vaccinations they need, improve and simplify the reporting of immunization data, identify high-risk populations for targeted interventions, and allocate resources efficiently and effectively [5-7]. EIRs, enhanced by data-driven DHIs, can help the immunization program achieve its goals of effective immunization coverage and real-time data for decision-making. EIRs can serve their purpose for immunization programs even better if integrated and synergized with DHIs for other programs such as Civil Registration and Vital Statistics (CRVS) and the national identification system. For instance, registration of all newborn babies in EIRs can improve tracking of immunization status and monitoring coverage [7]. EIRs integrated with other programs can strengthen other health services for children by providing a database of newborn babies in the population. Examples include newborn metabolic screening and childhood nutrition programs for the identification and referral of malnourished children [7].

Despite the many opportunities, several challenges hinder the effectiveness of EIRs in LMICs, such as the increased burden of data collection for health workers, which is the result of maintaining paper and digital documentation and reporting systems [7]. The implementation of EIRs, similar to all DHIs, should be aligned with the needs, both in terms of addressing the concerns of the intended users and being relevant to the users [8]. However, there is limited evidence on how to implement digital tools most effectively and sustainably across the full range of health systems [9]. The World Health Organization has highlighted the need for implementation research to identify the crucial factors that affect the implementation of DHIs for health system strengthening [5]. Implementation research can provide a systematic understanding of users' perceptions and experiences and thus enhance the usability and acceptability of DHIs.

In Rwanda, children from 0 to 15 months of age are provided with vaccines against 11 infections according to the Expanded Program on Immunization (EPI), namely, tuberculosis, poliomyelitis, diphtheria, tetanus, measles, pertussis, hepatitis B, *Haemophilus influenzae* type B, rubella, *Streptococcus pneumoniae*, and rotavirus [10]. The latest report from the Global Alliance for Vaccines and Immunization from 2017 identified issues with the immunization health information system such as data quality, population denominators based on projections from census data, and implausible coverage rates of more than 100% [11], similar to other contexts in eastern and southern Africa [12]. Incidents of vaccine dropouts and incomplete immunization, particularly for Pentavalent 3, were also identified. Significant geographic variations in immunization rates were reported, with 1 district in the northern part of the country reporting an overall coverage rate of as low as 88% [11,13].

The introduction of an EIR, known as e-Tracker, was launched in 2019 with the goal of improving overall data quality, data availability for monitoring of immunization defaulters or dropouts, and ultimately increasing immunization coverage [14]. The newly implemented e-Tracker has not yet been subject to research-based evaluations. The aim of this study was to assess the extent to which the immunization e-Tracker responds to stakeholders' needs and identify key areas for improvement in Rwanda's childhood immunization program.

Methods

Study Setting

This study was conducted among immunization nurses and data managers. Supervisors were included at the district hospital level. Health facilities were randomly selected from 5 districts in Rwanda—Gasabo, Rwamagana, Kamonyi, Gicumbi, and Rubavu, 1 from each of the 4 provinces and the City of Kigali of Rwanda. Gicumbi district, which is in the north of Rwanda, has 16 health centers; Kamonyi, in the south, has 13 centers; Rwamagana, in the east, has 15 health centers; Rubavu, in the west, has 13 centers; and Gasabo, in the central city of Kigali, has 16 health centers. In Kamonyi district and Gicumbi district, the routine immunization coverage rates for Pentavalent 3 and measles-rubella 1 in 2018 were 84% and 85%, respectively, while the coverage rate was higher than 89% in the remaining 3 districts. Gicumbi, Gasabo, and Kamonyi were among the districts with the largest percentage of underimmunized children, especially for the third dose of Pentavalent. The e-Tracker was introduced and operationalized in health centers in all districts of Rwanda in 2019.

The study participants were primary users of the immunization e-Tracker, either entering data or using the data: immunization nurses, data managers, and EPI supervisors.

Immunization-related services are organized at different levels of the health system. At the village level, community health workers engage with residents to raise awareness about childhood immunization. All primary health care services, including childhood immunization, are decentralized to the health center level. Immunization is provided at the health centers by immunization nurses or at health posts by the same nurses through community outreach in hard-to-reach areas. There are 499 health centers and 476 health posts in Rwanda [15]. More than 90% of children are immunized at the health center. All immunization sites (centers and posts) have weekly schedules of immunization days.

A health center typically has 2 immunization health workers, a nurse in charge of immunizations, and an assistant to deliver vaccines and keep records of all information pertaining to immunizations.

Figures 1 and 2 show the workflow of immunization at the health facility and the e-Tracker registration process and data visualization, respectively.

Figure 1. Workflow of immunization pertaining all the duties of a health worker (immunization nurse) on a vaccination day.

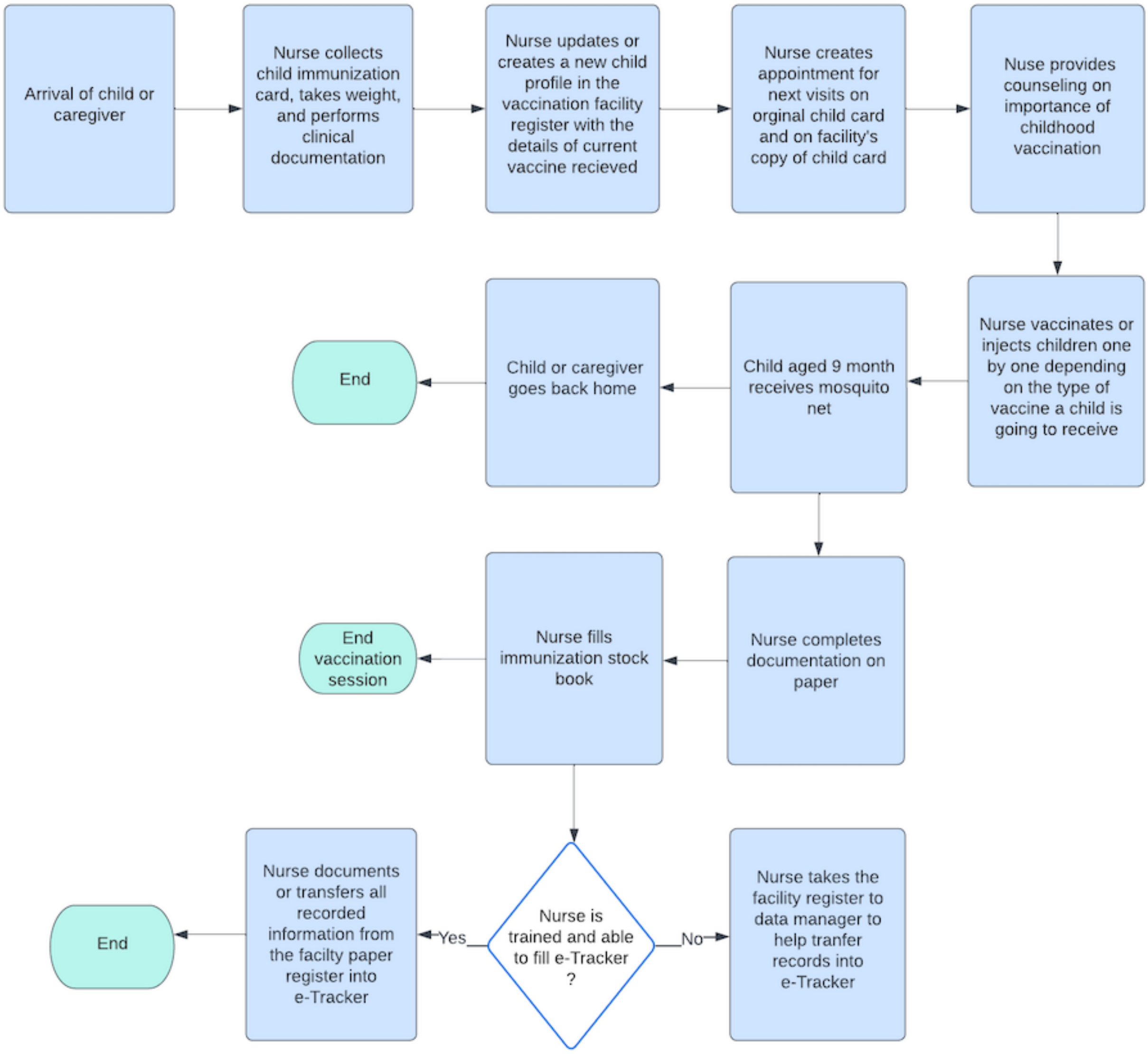
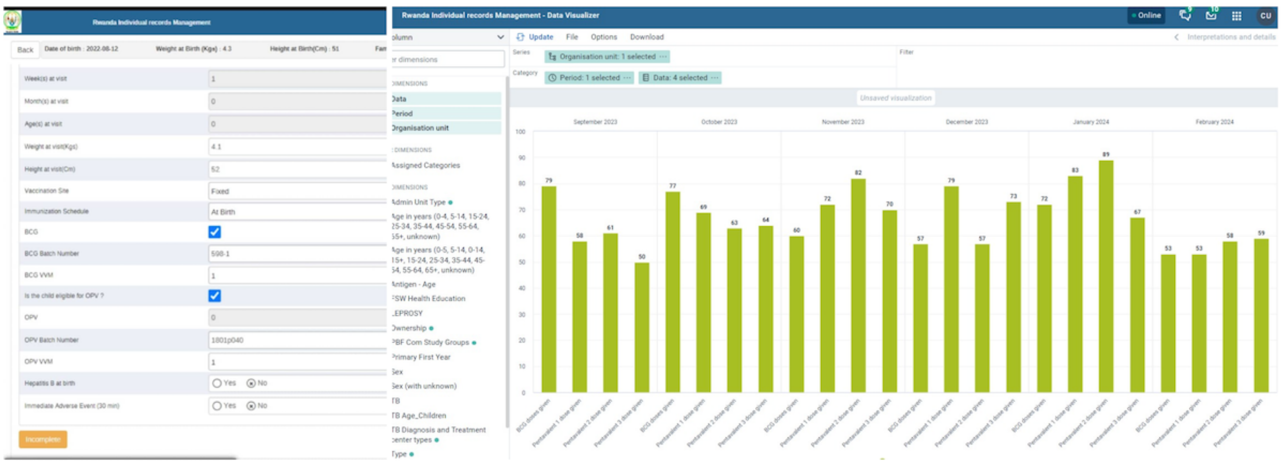


Figure 2. Immunization e-Tracker showing the client registration page and data visualization dashboard.



e-Tracker Implementation and Use

Implementation of the e-Tracker started in 2019 and was operational in all public health facilities. The e-Tracker runs on the District Health Information Software 2 (University of Oslo) platform, one of the most widely used digital health information systems globally [16]. Three cadres of health workers were trained to use the e-Tracker—immunization nurses, data managers, and EPI supervisors. All individual information are first recorded on 2 sets of paper-based forms: the child's immunization cards and the health center's immunization paper registers. The immunization nurse or the data manager then transfers the same information from the paper registers to the e-Tracker. At the end of the month, a set of predefined data is aggregated onto paper reporting forms by immunization nurses and handed over to the data manager, who then enters these data into the aggregate reporting system built in a separate instance of District Health Information Software 2, as a part of the health management information system (HMIS).

EPI supervisors, located at the district hospitals, use the e-Tracker to assess the progress of health facilities by comparing the number of children registered as successfully vaccinated on each indicator against the monthly target provided to the health center (Table 1). The target is an estimate population based on the expected number of births in the area based on census data.

Users who have technical issues with e-Tracker can contact the central help desk. Phone calls or WhatsApp groups are typically used to resolve simple technical issues such as password reset, and for complicated issues, through visits to health centers. Immunization nurses from the health centers have a joint WhatsApp group with their respective EPI supervisors where they communicate issues regarding immunization and e-Tracker–related technical support in their district.

Table 1. Intended use and user roles in the immunization e-Tracker.

User	Intended use and user roles in the e-Tracker
Immunization nurse[a]	• Data entry and registration of new children for immunization • Update and follow up on subsequent immunizations until a child has completed his or her vaccination calendar
Data manager[a]	• Data entry and registration of new children for immunization • Update and follow up on subsequent immunizations until a child has completed his or her vaccination calendar • Generate reports of comparisons of the health center's immunization coverage rate against the target
EPI[b] supervisor	• Review reports from all health centers in the district catchment area and provide recommendations and feedback for improvement based on the data

[a]Data entry tasks could be shared by immunization nurses and data managers.

[b]EPI: Expanded Program on Immunization.

Study Design and Sampling

This study is an implementation research design that used descriptive qualitative methods [17,18] and formative evaluation to assess the extent to which the immunization e-Tracker responds to stakeholders' needs and identify key areas for improvement [19]. This was done through key informant interviews. The Human, Organization, and Technology–Fit (HOT-Fit) evaluation framework guided the data collection and analysis [20]. We chose the HOT-Fit framework because it has the potential to evaluate health information systems; encompasses comprehensive dimensions; and measures the fit between technological, human, and organizational aspects, all of which are critical for system adoption [20].

To select a sample of districts, we first assessed data reports retrieved from the e-Tracker and the national HMIS in the first 3 months of 2020 for all 30 districts in Rwanda. Four immunization indicators—Bacille Calmette-Guérin (BCG) and Pentavalent (penta) first, second, and third doses (penta1, penta2, and penta3)—were reviewed by a program manager together with a researcher (TU) to calculate completeness of data in the e-Tracker (e-Tracker–reported indicator and HMIS-reported indicator). We then selected 5 districts as follows: 1 district among the best performers (Rwamagana district >80%), 1 from the worst performers (Rubavu district <15%), and 3 districts that were in the middle (Gasabo, Gicumbi, and Kamonyi districts: 50%-60%). We randomly included 6 health centers from each of the 5 districts. From this pool of health centers, key informants and participants were purposively sampled among primary users of the e-Tracker. To cover the variation of sites across the districts appropriately, we recruited 1 nurse and 1 data manager from 1 district before moving to the next to diversify the data collected.

Data Collection

This study was carried out in accordance with COREQ (Consolidated Criteria for Reporting Qualitative Research) guidelines (Multimedia Appendix 1) [21].

Based on the 3 constructs of the HOT-Fit framework, we created study-specific definitions for each of the constructs (Table 2) and formulated an interview guide with open-ended questions (Multimedia Appendix 2). Three pilot key informant interviews were conducted with immunization nurses and data managers in 1 district (Gasabo) to validate the tool prior to data collection. We further refined the questions in the guide based on the findings from these interviews (Multimedia Appendix 2).

Table 2. Specific domains of evaluation of the e-Tracker based on the constructs of the HOT-Fit[a] framework.

HOT-Fit constructs and definitions	Study constructs and definitions

Technology: Meets the need of the projected users, is convenient and easy to use, and fits the work patterns of the professionals for whom it is intended and the overall health system	• System quality
	  • Associated with system performance: ease of use, ease of learning, response time, usefulness, system flexibility, and security
	• Information quality
	  • User perspectives and quantitative data: completeness, availability, accuracy, reliability, timeliness, relevance, and consistency
	• Service quality
	  • Service delivered: technical support, quick responsiveness, assurance, empathy, and follow-up service

Human: The person who uses and the use of information output such as reports	• System use
	  • Concerned with the frequency and breadth of health information system inquiries and functions: system users, their levels of use, training, knowledge, belief, expectation, acceptance, or resistance
	• User satisfaction
	  • Evaluation of users' experience in using the system and the potential impact of the system: perceived usefulness, enjoyment, overall satisfaction and satisfaction with specific functions, and decision-making satisfaction

Organization: Nature and factors of a health care institution	• Structure
	  • Nature (type and size), management and communication, clinical process, and workflow process. Leadership, top management support, etc
	• Environment
	  • Financial source, government, politics, and type of population being served

Net benefits	• Quality of care, clinical impact, impact on patient care and communication, and facilitation of information access

[a]HOT-Fit: Human, Organization, and Technology–Fit.

Separate interview guides were used for each category of participants. The interview questions were formulated based on each user's role both in the immunization program and the e-Tracker system. For instance, we asked questions related to user-specific employment and how e-Tracker is related to his or her job. Some e-Tracker technical questions were similar such as whether e-Tracker was easy to use, easy to learn, or about how e-Tracker responds (response time). One author (TU), a current PhD candidate, with experience in IT conducted 14 in-person, in-depth interviews with key informants (e-Tracker end users in primary health care centers and EPI supervisors in their affiliated district hospitals). Interviews were conducted with only the key informant and the interviewer present. The interviews were conducted in Kinyarwanda, took place over approximately 1 hour, and were audio recorded. No notes were taken. The audio was then transcribed in Kinyarwanda and translated into English by a bilingual professional. A group of 2 researchers (TU and ER) reviewed the translations for accuracy. The study team met on a weekly basis to evaluate the data collection process. After 4 interviews per key informant category, the data collector began hearing information repetition. The research team advised undertaking 1 more interview per participant category to ensure that no new findings were discovered. Data saturation was confirmed, and data collection was stopped. No repeat interviews were carried out.

Data Analysis

Translated interview transcripts were uploaded into NVivo 12 (Lumivero). Based on the HOT-Fit framework, a codebook was developed by the team through discussion. Using this agreed-upon codebook, 2 researchers (TU and ER) individually coded the data. A deductive coding style was applied to our data. Discrepancies in coding were discussed and resolved by the team.

The HOT-Fit Framework

After coding was completed by both researchers, the team compiled the relevant data extracts. We performed a framework analysis and worked together to place the extracted data within the HOT-Fit framework [20,22]. We analyzed interview transcripts to find all possible codes from all participants. We identified and summarized codes in accordance with constructs of the HOT-Fit framework and study-specific domains (Table 2). NVivo 12 analysis software was used to manage themes and codes.

Author Reflexivity

Prior to data collection, the interviewer and research team had minimal contact with participants (stakeholder engagement session). The participants were informed that the purpose of the study was to gather their views and experiences on e-Tracker use to assess how the immunization e-Tracker responds to stakeholders' needs and identifies areas for improvement. They were also informed that this was part of a larger project studying the design and implementation of DHIs to improve childhood immunization in Rwanda. Authors entered this study with the belief that an e-Tracker had the potential to positively impact care providers' experiences; however, it took effort to prevent personal bias during data analysis. Due to COVID-19 restrictions, member checking was completed with 1 key informant from each category.

Ethical Considerations

This study was approved by the Rwanda National Ethics Committee (1011/RNEC/2020), the Norwegian (West) Regional Committee for Medical and Health Research Ethics (251925), and the Rwanda Ministry of Health's National Health Research Committee (reference NHRC/2021/PROT/002). All methods were performed in accordance with relevant guidelines and regulations by the World Medical Association Declaration of Helsinki—ethical principles for medical research involving human subjects [23].

The participants were informed about the study objectives, their voluntary participation, and their right to refuse participation at any time. The written informed consent form was obtained from each participant after getting an explanation about the research purpose and confirming their participation in the study. The interviews took place in a safe room with the office door locked at the health facility. The recorded information was transcribed and anonymized. The audio recording device could only be accessed via a security code by the lead author (TU).

Results

Overview

In total, 14 e-Tracker users were interviewed (Table 3), including 5 immunization nurses, 5 data managers, and 4 EPI supervisors (1 EPI supervisor declined being interviewed due to clinical COVID-19 work). Most of the immunization nurses were female (4/5, 80%) and had more than 10 years of work experience (3/5, 60%). In contrast, data managers were mostly male (4/5, 80%), younger, and had work experience of 5 years or less (4/5, 80%). Half of the supervisors (2/4, 50%) were female. The supervisors had varying levels of work experience (Table 3). We present our findings based on the constructs of the HOT-Fit framework.

Table 3. Characteristics of study participants.

Characteristics	Immunization nurses (n=5)	Data managers (n=5)	Supervisors (n=4)
Sex, n (%)			
Male	1 (20)	4 (80)	2 (50)
Female	4 (80)	1 (20)	2 (50)
Age range (years), n (%)			
25-35	1 (20)	2 (40)	1 (25)
36-45	2 (40)	3 (60)	0 (0)
46-55	2 (40)	0 (0)	2 (50)
56 and older	0 (0)	0 (0)	1 (25)
Age (years), mean (SD)	43 (8.29)	35 (5.89)	50 (10.23)
Field of study, n (%)			
Nursing	5 (100)	2 (40)	0 (0)
Laboratory	0 (0)	1 (20)	0 (0)
Computer science	0 (0)	1 (20)	0 (0)
Public health	0 (0)	1 (20)	3 (75)
Midwifery	0 (0)	0 (0)	1 (25)
Working experience (years), n (%)			
≤5	1 (20)	4 (80)	2 (50)
6-10	1 (20)	1 (20)	1 (25)
>10	3 (60)	0 (0)	1 (25)

Technology

System Quality

Data managers and supervisors reported that the e-Tracker was not a complex system. Two (40%) of 5 nurses perceived the e-Tracker as complex due to limited skills of computer literacy (Table 4: section A, construct 1).

Table 4. Summary of main findings in accordance with the HOT-Fit framework and quotes from key informant interviews.

Construct number and main findings	User's quotes
Section A: Technology	
System quality	
1. Ease of learning	"...that system [e-Tracker] is not difficult to use, except that it is not easy for everyone because there are some health centers for example that have immunization nurses who do not know how to use the computer." (EPI[a] supervisor 1)
2. Better data security than paper registers and forms	"e-Tracker is a secure system protected by personal credentials; it is not like paper registers where anyone can access." (Data manager 5)
3. Missing technical functionalities	"...as a person who is in the field and using it [e-Tracker] frequently, I realize that there are some functionalities that the e-Tracker is lacking. For example, it does not show me the next appointment for someone's vaccination or the list of who the nurses should be seeing today." (Data manager 5)
4. Not compatible for community outreach	"...internet connection that is not available, lack of outreach support—all these are challenges with using the e-Tracker." (Immunization nurse 1)
5. Connectivity issues and slow system response	"Things related to e-Tracker are slow, definitely slow. This is a challenge we usually face." (Immunization nurse 2)
Information quality	
6. Incomplete and unreliable data	"There are times when you register a child and when you go back to search him or her, you find that the actual information is not complete, or you find that the e-Tracker contains a duplicate of the child's records." (Immunization nurse 3)
7. Increased documentation workload	"I may fail to get time for instance, and they shift me to provide another health service, but, because there is much information that needs to be entered and I am responsible for that, I go quickly and take like one hour after work, or I come early in the morning to enter them." (Immunization nurse 4)
Service quality	
8. Delays in getting technical support	"It is difficult to get technical assistance because it is from central level and nowhere else...if the problem is simple like the system is off and then back on, those ones are quick and can be done on a phone call or WhatsApp, but bigger technical issues take time." (Data manager 2)
9. Alternative communication lines	"...talking about the other [communication] chain...I just call my superior at the hospital, and he conveys it to the central level technical team...and they gradually communicate with each other, and the information reaches us." (Data manager 4)
Section B: Human	
System use	
10. Does not meet the intended purpose	"What I expected from e-Tracker up to now, I can say that I have not yet seen its results. This may be due to other challenges, but the functionalities required by the nurses to use the e-Tracker well and properly are not yet available." (Supervisor 1)
11. Suboptimal use due to increased documentation workload	"This e-Tracker system is expected to be used by immunization nurses; it has apparently increased their work, which was not easy. That is simply to say, this is beyond their capacity." (Supervisor 1)
User satisfaction	
12. General dissatisfaction with the e-Tracker	"...Discriminating children's cards increases job, in e-Tracker it is simple; just search child and find him easily, but the use of e-Tracker did not stop papers, you complete all existing paper books and forms and then go complete e-Tracker." (Immunization nurse 2)
Section C: Organization	
Structure	
13. Lack of effective training processes	"...data manager who received training has gone, the one who replaced him does not actually know to use e-Tracker, he often called me asking, 'where can I click on?'...you realize that it is slowly by slowly." (Immunization nurse 2)
14. Lack of support for health workers in using technology	"...it happens that you register a child and when you go back to search for him [in the e-Tracker], you miss him simply because you do not know if it is a connection problem, or a low knowledge regarding how to search for him." (Immunization nurse 1)
Environment	
15. Performance-based financing	"We have many duties, and there are so many systems at health center...they come and say we give you PBF after seeing in the system how many children you have entered, and it is understandable that you will not receive any money if you didn't register any child." (Immunization nurse 3)

Construct number and main findings	User's quotes
Section D: Net benefits	
16. Perceived current benefits	"...the e-Tracker has a dashboard for data analysis. Like now, I sometimes say, let me see how many children we have registered this month; for the first, the second and the third dose of Penta, for instance." (Data manager 3)
17. Perceived future benefits for HWs[b]	"e-Tracker has made nothing easier for me. Instead, it has complicated things. Perhaps there is value in the e-Tracker if all these papers and books are removed. Then you may find that the e-Tracker will bring benefits." (Immunization nurse 3)

[a]EPI: Expanded Program on Immunization.

[b]HW: health worker.

Data security in the e-Tracker was generally perceived as satisfactory and better than data security using paper registers (Table 4: section A, construct 2). However, users reported several shortcomings. They cited the lack of several technical functionalities such as client lists, lists of defaulters, unspecific search and register functions, automated routine reports, and linkage to other systems such as birth notification and the national identity system (Table 4: section A, construct 3). Users expressed the need for a more flexible data entry tool that can operate offline, such as handheld tablets instead of desktop computers, to use during community outreach. They also cited poor connectivity and solely relying on health center–purchased internet as one of the most important reasons for the suboptimal use of the e-Tracker (Table 4: section A, construct 4). For system response time, 4 (80%) of 5 health workers and 4 (80%) of 5 data managers reported that the e-Tracker responds slowly. The remaining interviewees, particularly supervisors, located at hospitals with better internet connectivity, reported the opposite that the e-Tracker had a quick response time. Adequate support for network connectivity was lacking (Table 4: section A, construct 5). For example, immunization health workers at health centers were given modems, but they claimed that they were not given financial assistance for continued internet subscriptions.

Information Quality

Data in the e-Tracker were considered incomplete and unreliable and were not actively used by the immunization nurses (Table 4: section A, construct 6). Several underlying issues were identified as contributors to poor information quality. Users were required to document in the e-Tracker in addition to existing paper forms, which created double work. The double entry of data, combined with a mismatch between the data elements in the paper forms and the e-Tracker, results in users skipping some data fields in the e-Tracker. A common response with all users was the lack of time to complete documentations in the e-Tracker due to heavy workloads (Table 4: section A, construct 7). Two (40%) of the 5 interviewed immunization nurses were not trained in e-Tracker use, but even those who were trained and able to use the e-Tracker reported that the time allocated to them to fill the e-Tracker was insufficient. Three (60%) of 5 immunization nurses reported having to work overtime to enter data in the e-Tracker, 1 hour before or after work—a practice that users believed adversely affected data quality.

Service Quality

All interviewed nurses and data managers reported some form of delay in getting technical support (Table 4: section A, construct 8). Users' responses on this issue suggest that they might prefer reporting issues to their supervisor, who could then facilitate communication with the central support team.

Human

System Use

According to all interviewees, the e-Tracker did not meet overall user expectations. Further exploration revealed that users want a system that generates automated monthly reports and reduces documentation workload. The e-Tracker does not automatically generate any reports, and double documentation was identified as an important problem that impacted effective e-Tracker use (Table 4: section B, constructs 10, 11).

User Satisfaction

When asked whether they were satisfied with the e-Tracker, only 2 (40%) of 5 data managers said yes. The lack of technical functionalities and increased documentation workload were the leading causes of dissatisfaction for the data managers and health workers, respectively (Table 4 section B, construct 12).

Organization

Structure

Users described quarterly data quality assessment workshops to encourage e-Tracker use by health workers. Such assessments are usually done by data managers, nurses, and their supervisors by reviewing paper reports and e-Tracker reports and comparing them to HMIS reports for selected vaccination indicators, such as BCG. Health workers reported not receiving enough support in navigating digital systems in general (Table 4 section C, construct 13) and highlighted the need for regular training sessions on how to use the e-Tracker and a plan to deal with staff turnover. The planned training for users in 2021 did not happen due to the COVID-19 pandemic.

Environment

Performance-based financing was provided to the health workers based on the number of newborn babies registered as BCG vaccinated in comparison with their reported number of BCG vaccinations. The interviewees alluded to this as a reason for entering data for this specific indicator into the e-Tracker rather than the indicators for other vaccines. Performance-based financing in this context is based on the number of children registered with BCG vaccination as a way of promoting the registration of newborn babies in the childhood immunization e-Tracker [24] (Table 4: section C, construct 15).

Net Benefits

Participants acknowledged the potential benefits of an e-Tracker provided technical and implementation issues are addressed. For example, all EPI supervisors reported that the tool could be helpful to monitor children's registration and vaccination status without visiting health centers physically. Two (40%) of 5 data managers reported using the e-Tracker for monitoring and evaluation in terms of vaccination coverage for their respective health centers. In contrast, all health workers did not report any net benefits from the current use, although they see that the e-Tracker may contribute positively to their work in the future (Table 4: section D, construct 17).

Key Improvements

Textbox 1 provides a summary of the main recommendations for improvement of the e-Tracker based on our findings.

Textbox 1. Overall recommendations for key improvements highlighted by the users.

Immunization nurses

- Better client search and register function
- Produce lists of expected and missed clients to avoid searching in paper registers
- Facilitate tracking a defaulter or a dropout child and remind parents of the missed appointments
- Improve connectivity
- Offline e-Tracker version that will make it easier to collect data in case of network outage, handheld devices to help immunization outreach in difficult-to-reach areas
- Regular training on e-Tracker use

Data managers

- Generate automatic monthly reports
- Link e-Tracker to other systems such as Civil Registration and Vital Statistics and national identification systems

Expanded Program on Immunization supervisors

- Additional trainings on analysis of e-Tracker data
- Offline e-Tracker version and more devices to support nurses' work at primary health centers

Discussion

Principal Findings

This study explored stakeholders' experiences and perceptions of using the e-Tracker for the Rwandan childhood immunization program. Users of the e-Tracker described several issues that hamper effective data entry as well as data use. Data in e-Tracker were reported to be incomplete and unreliable as result of dual documentation on paper and digitally.

Rwanda is one of the few countries in Africa to implement an EIR at scale. Implementation of the e-Tracker is a top priority for the childhood immunization program. Along with technological resources such as computers and modems, a top-level team and 3 cadres of trained health professionals from each health center across the nation are assigned to support the implementation indicating significant organizational support for change. EIRs allow for real-time monitoring of immunization status and provide data for decision-making, and their evaluations play a key role in identifying strategies to improve their use [25]. Our findings demonstrate the need for technical improvements to fit clinical practice and increase benefits, addressing implementation-related issues such as workflow matching, as well as training and user support. User-informed development of technical functionalities has been shown to be linked to higher adoption of health information systems in a systematic review of 55 studies [26]. Slow response times and delayed IT support adversely affected e-Tracker use in our study, factors also reported in other studies of digital information systems [27,28].

Creating an enabling environment for digital health systems by addressing issues such as training, and capacity strengthening in data entry and use, is equally important to ensure successful implementation [29]. Users cited a general dissatisfaction with the e-Tracker for several reasons including increased workload due to dual documentation and insufficient training. Several studies have reported similar dissatisfaction among users of digital health information systems in many cases as a result of the system's inability to match existing work patterns [26]. On the other hand, users are typically more satisfied when information systems offer good quality data; the higher the quality of the data the higher the satisfaction [27,30]. Users in this study perceived the information in the e-Tracker to be inaccurate and incomplete in comparison with the paper records and registers. None of the entered digital information was used by data managers or nurses for clinical practice.

Immunization nurses are the intended users of the e-Tracker, although the current workflow involves secondary data entry in the e-Tracker by the data manager in several health centers. While data managers and supervisors stated some benefits of the e-Tracker for their work, immunization nurses reported no net benefits of the e-Tracker as it has been implemented in its current version. One of the reasons for this may be that the e-Tracker in its current form is not considered an essential part of the data ecosystem in the immunization health information system, particularly because the monthly reports are still paper based and not generated from the e-Tracker. In a setting such as Rwanda with scarce human resources for health, efficiency and costs are important considerations. Efficiency gains cannot be achieved unless health centers phase out paper immunization records and exclusively use the e-Tracker for data entry [31]. Similarly, a study conducted in Zambia and Tanzania showed that the use of the EIR decreased over time in settings where it was used in parallel with paper-based documentation compared to exclusive use [32]. In most other LMICs, paper-based documentation and reporting consume a significant proportion of health workers' time, which can be alleviated by well-implemented digital tools co-designed with the end user [31,33].

Organizations play a key role in supporting the adoption of digital systems directly and indirectly and sometimes inadvertently skewing priorities [20,30]. For instance, in our study, health workers are provided with performance-based financing based on BCG vaccine coverage rates, which might explain the relatively better completeness of these data in the e-Tracker.

Strengths

This study was conducted in sub-Saharan Africa, where there has been relatively limited research on EIRs and DHIs in general. Our findings are reasonably generalizable to the Rwandan context for two main reasons: (1) we sampled health centers at different stages of e-Tracker use, ranging from low to high, and (2) we included all users of the e-Tracker (immunization nurses, data managers, and supervisors).

Most studies that have applied the HOT-Fit framework have used quantitative methods to evaluate the effectiveness. We chose qualitative methods to gain an in-depth understanding of user-reported barriers and opportunities for e-Tracker use [20]. Our research is aligned with the national health system priorities to improve data use in the immunization program [34]. Key stakeholders, including representatives from the Ministry of Health and the Rwanda Biomedical Center, were involved at every stage of the research. They were consulted and presented with the study plan and results.

Study Limitations

This study has some limitations. The study was conducted in 2021, after a relatively short period of e-Tracker use by the health centers. Since the first introduction of the tool, some improvements have been implemented and these were not captured in our study. For example, nationwide linkages between the CRVS and immunization registry have recently been established and health workers providing immunization can retrieve information about the child from the CRVS. The COVID-19 pandemic and the subsequent restrictions in the years following the implementation of the e-Tracker may have affected training, use, and perceptions. Health workers from the immunization program (immunization nurses, data managers, and EPI supervisors) contributed immensely to the COVID-19 response, which may have affected their attitudes and perceptions toward their general workload and e-Tracker use.

Conclusions

The study findings revealed a low satisfaction level among the users of the immunization e-Tracker in Rwanda due to technical as well as implementation-related factors. Technical functionalities and implementation strategies co-designed with the user can help improve user experience and eventually maximize the benefits of the e-Tracker. Implementation strategies to reduce or remove dual documentation on paper and digital systems and to generate automated digital monthly immunization reports can save valuable time for health workers.

Acknowledgments

The authors thank the Ministry of Health and Rwanda Biomedical Center for their continued collaboration and facilitation of the research. They are also grateful to the participating health centers and the cooperation of study participants. They also acknowledge Michael Mugisha for his invaluable contribution to the conception and design of this research effort.

Conflicts of Interest

None declared.

Multimedia Appendix 1
COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.
[PDF File (Adobe PDF File), 536 KB - humanfactors_v11i1e53071_app1.pdf]

Multimedia Appendix 2
Interview guides.
[PDF File (Adobe PDF File), 125 KB - humanfactors_v11i1e53071_app2.pdf]

References

1. Global immunization coverage 2021. World Health Organization. 2021. URL: https://www.who.int/en/news-room/fact-sheets/detail/immunization-coverage [accessed 2022-10-26]
2. Carter A, Msemburi W, Sim SY, Gaythorpe KAM, Lambach P, Lindstrand A, et al. Modeling the impact of vaccination for the immunization agenda 2030: deaths averted due to vaccination against 14 pathogens in 194 countries from 2021 to 2030. Vaccine 2023;S0264-S0410X:00854-X [FREE Full text] [doi: 10.1016/j.vaccine.2023.07.033] [Medline: 37537094]
3. Tozzi AE, Gesualdo F, D'Ambrosio A, Pandolfi E, Agricola E, Lopalco P. Can digital tools be used for improving immunization programs? Front Public Health 2016;4:36 [FREE Full text] [doi: 10.3389/fpubh.2016.00036] [Medline: 27014673]
4. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
5. World Health Organization. WHO Guideline. Recommendations on Digital Interventions for Health System Strengthening. Geneva: World Health Organization; 2019. URL: https://iris.who.int/bitstream/handle/10665/311941/9789241550505-eng.pdf?sequence=31
6. Closing gaps in vaccine coverage with electronic immunization registries. PATH. URL: https://www.path.org/case-studies/electronic-immunization-registries/ [accessed 2023-01-09]
7. Namageyo-Funa A, Samuel A, Bloland P, Macneil A. Considerations for the development and implementation of electronic immunization registries in Africa. Pan Afr Med J 2018;30:81 [FREE Full text] [doi: 10.11604/pamj.2018.30.81.11951] [Medline: 30344865]
8. Peters DH, Tran NT, Adam T, Alliance for Health Policy and Systems Research. Implementation research in health: a practical guide. World Health Organization. 2013. URL: https://ahpsr.who.int/publications/i/item/9789241506212 [accessed 2024-04-25]
9. Parajuli R, Bohara D, Kc M, Shanmuganathan S, Mistry SK, Yadav UN. Challenges and opportunities for implementing digital health interventions in Nepal: a rapid review. Front Digit Health 2022;4:861019 [FREE Full text] [doi: 10.3389/fdgth.2022.861019] [Medline: 36120714]
10. Comprehensive multi-year plan 2013-2017. Rwanda Ministry of Health. 2012. URL: http://www.nationalplanningcycles.org/sites/default/files/country_docs/Rwanda/attachment_6_revised_cmyp_08.pdf [accessed 2024-04-25]
11. Joint appraisal report 2017 Rwanda. GAVI. 2017. URL: https://www.gavi.org/country/rwanda/documents/ [accessed 2024-04-25]
12. Maïga A, Jiwani SS, Mutua MK, Porth TA, Taylor CM, Asiki G, et al. Generating statistics from health facility data: the state of routine health information systems in Eastern and Southern Africa. BMJ Glob Health 2019;4(5):e001849. [doi: 10.1136/bmjgh-2019-001849] [Medline: 31637032]
13. Nwankwo CM, Orua E. Factors influencing incomplete immunization among under five years old children at CHUK Hospital, Nyarugenge district, Rwanda. Int J Community Med Public Health 2020;7(10):3787-3795 [FREE Full text] [doi: 10.18203/2394-6040.ijcmph20204340]
14. RWANDA - EPI Tracker and Dashboard use case. URL: https://s3-eu-west-1.amazonaws.com/content.dhis2.org/general/EPI+Tracker+%26+Dashboard+Implementation_+Rwanda.pdf [accessed 2024-04-25]
15. Republic of Rwanda, Ministry of Health—Fourth Health Sector Strategic Plan July 2018-June 2024. Rwanda Ministry of Health. 2018. URL: http://npngti.com/wp-content/uploads/2018/06/Rwanda_Nat-Health-Sector-Plan_2018-2024.pdf [accessed 2024-04-25]
16. University of Oslo. The world's largest health information management system — developed through global collaboration led by UiO. Dhis2. URL: https://dhis2.org/ [accessed 2023-02-12]
17. Hamilton AB, Finley EP. Qualitative methods in implementation research: an introduction. Psychiatry Res 2019;280:112516 [FREE Full text] [doi: 10.1016/j.psychres.2019.112516] [Medline: 31437661]
18. Gu Y, Warren J. Chapter 11 Methods for descriptive studies. In: Lau F, Kuziemsky C, editors. Handbook of eHealth Evaluation: An Evidence-based Approach. Victoria, British Columbia Canada V8P 5C2: University of Victoria; 2016.
19. Stetler CB, Legro MW, Wallace CM, Bowman C, Guihan M, Hagedorn H, et al. The role of formative evaluation in implementation research and the QUERI experience. J Gen Intern Med 2006;21(Suppl 2):S1-S8 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00355.x] [Medline: 16637954]
20. Yusof MM, Kuljis J, Papazafeiropoulou A, Stergioulas LK. An evaluation framework for health information systems: Human, Organization and Technology-fit factors (HOT-fit). Int J Med Inform 2008;77(6):386-398. [doi: 10.1016/j.ijmedinf.2007.08.011] [Medline: 17964851]
21. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
22. Meunier PY, Raynaud C, Guimaraes E, Gueyffier F, Letrilliart L. Barriers and facilitators to the use of clinical decision support systems in primary care: a mixed-methods systematic review. Ann Fam Med 2023;21(1):57-69 [FREE Full text] [doi: 10.1370/afm.2908] [Medline: 36690490]

23. World Medical Association. World Medical Association declaration of Helsinki. Ethical principles for medical research involving human subjects. Bull World Health Organ 2001;79(4):373-374 [FREE Full text] [Medline: 11357217]
24. Performance based financing procedures manual for health facilities (Hospitals and Health Centers). Rwanda Ministry of Health. 2021. URL: https://www.scribd.com/document/529709575/1562825127078-PBF-Manual-Hospital-and-Health-Center [accessed 2024-04-25]
25. Deharja A, Hargono A, Santi MW, Nandini N, Damayanti NA. Evaluating the usability of Hospital Information System (HIS) through Human Organization Technology-Fit (Hot-Fit) model. Int Proc 2nd Int Sci Meet Health Inf Manag 2020;5(1171):380-389 [FREE Full text]
26. Yusof MM, Stergioulas L, Zugic J. Health information systems adoption: findings from a systematic review. Stud Health Technol Inform 2007;129(Pt 1):262-266. [Medline: 17911719]
27. Fauzan A, Noviandi N. Evaluation of optima regional health information system with HOT-Fit on technology aspects approach in Johar Baru Health Center Jakarta. J Intell Comput Health Informatics 2020;1(1):9-14 [FREE Full text] [doi: 10.26714/jichi.v1i1.5397]
28. Erlirianto LM, Ali AHN, Herdiyanti A. The implementation of the Human, Organization, and Technology–Fit (HOT–Fit) framework to evaluate the Electronic Medical Record (EMR) system in a hospital. Procedia Comput Sci 2015;72:580-587 [FREE Full text] [doi: 10.1016/j.procs.2015.12.166]
29. Pan American Health Organization. Electronic immunization registry: practical considerations for planning, development, implementation and evaluation. Pan American Health Organization. 2017. URL: https://iris.paho.org/handle/10665.2/34865 [accessed 2024-04-25]
30. Kilsdonk E, Peute LW, Jaspers MWM. Factors influencing implementation success of guideline-based clinical decision support systems: a systematic review and gaps analysis. Int J Med Inform 2017;98:56-64. [doi: 10.1016/j.ijmedinf.2016.12.001] [Medline: 28034413]
31. Venkateswaran M, Nazzal Z, Ghanem B, Khraiwesh R, Abbas E, Khader KA, et al. eRegTime-Time spent on health information management in primary health care clinics using a digital health registry versus paper-based documentation: cluster-randomized controlled trial. JMIR Form Res 2022;6(5):e34021 [FREE Full text] [doi: 10.2196/34021] [Medline: 35559792]
32. Carnahan E, Ferriss E, Beylerian E, Mwansa FD, Bulula N, Lyimo D, et al. Determinants of facility-level use of electronic immunization registries in Tanzania and Zambia: an observational analysis. Glob Health Sci Pract 2020;8(3):488-504 [FREE Full text] [doi: 10.9745/GHSP-D-20-00134] [Medline: 33008860]
33. World Health Organization (WHO). Web Annex M: Cross-cutting acceptability and feasibility issues and implementation considerations for health workers. In: WHO guideline Recommendations on Digital Interventions for Health System Strengthening. Geneva: World Health Organization; 2019:1-5.
34. Mutale W, Chintu N, Amoroso C, Awoonor-Williams K, Phillips J, Baynes C, et al. Improving health information systems for decision making across five sub-Saharan African countries: implementation strategies from the African Health Initiative. BMC Health Serv Res 2013;13(Suppl 2):S9 [FREE Full text] [doi: 10.1186/1472-6963-13-S2-S9] [Medline: 23819699]

Abbreviations

BCG: Bacille Calmette-Guérin
COREQ: Consolidated Criteria for Reporting Qualitative Research
CRVS: Civil Registration and Vital Statistics
DHI: digital health intervention
EIR: electronic immunization registry
EPI: Expanded Program on Immunization
HMIS: health management information system
HOT-Fit: Human, Organization, and Technology–Fit
LMIC: low- and middle-income country

Edited by A Kushniruk; submitted 25.09.23; peer-reviewed by JR Medina, D Chrimes, S Fenton; comments to author 24.01.24; revised version received 20.03.24; accepted 07.04.24; published 28.05.24.

Please cite as:
Uwera T, Venkateswaran M, Bhutada K, Papadopoulou E, Rukundo E, K Tumusiime D, Frøen JF
Electronic Immunization Registry in Rwanda: Qualitative Study of Health Worker Experiences
JMIR Hum Factors 2024;11:e53071
URL: https://humanfactors.jmir.org/2024/1/e53071
doi:10.2196/53071
PMID:38805254

Original Paper

The Solutions in Health Analytics for Rural Equity Across the Northwest (SHARE-NW) Dashboard for Health Equity in Rural Public Health: Usability Evaluation

Elizabeth Heitkemper[1*], RN, PhD; Scott Hulse[2*], MD, MPH, MSc, MSc; Betty Bekemeier[3,4], MPH, PhD; Melinda Schultz[3], MSc; Greg Whitman[3], BA; Anne M Turner[4,5], MD, MPH, MLIS

[1]School of Nursing, The University of Texas at Austin, Austin, TX, United States

[2]School of Medicine, Uniformed Services University of the Health Sciences, Bethesda, MD, United States

[3]School of Nursing, University of Washington, Seattle, WA, United States

[4]School of Public Health, University of Washington, Seattle, WA, United States

[5]School of Medicine, University of Washington, Seattle, WA, United States

[*]these authors contributed equally

Corresponding Author:
Elizabeth Heitkemper, RN, PhD
School of Nursing
The University of Texas at Austin
1710 Red River Street
Austin, TX, 78712
United States
Phone: 1 512 232 4228
Email: e.heit@utexas.edu

Abstract

Background: Given the dearth of resources to support rural public health practice, the solutions in health analytics for rural equity across the northwest dashboard (SHAREdash) was created to support rural county public health departments in northwestern United States with accessible and relevant data to identify and address health disparities in their jurisdictions. To ensure the development of useful dashboards, assessment of usability should occur at multiple stages throughout the system development life cycle. SHAREdash was refined via user-centered design methods, and upon completion, it is critical to evaluate the usability of SHAREdash.

Objective: This study aims to evaluate the usability of SHAREdash based on the system development lifecycle stage 3 evaluation goals of efficiency, satisfaction, and validity.

Methods: Public health professionals from rural health departments from Washington, Idaho, Oregon, and Alaska were enrolled in the usability study from January to April 2022. The web-based evaluation consisted of 2 think-aloud tasks and a semistructured qualitative interview. Think-aloud tasks assessed efficiency and effectiveness, and the interview investigated satisfaction and overall usability. Verbatim transcripts from the tasks and interviews were analyzed using directed content analysis.

Results: Of the 9 participants, all were female and most worked at a local health department (7/9, 78%). A mean of 10.1 (SD 1.4) clicks for task 1 (could be completed in 7 clicks) and 11.4 (SD 2.0) clicks for task 2 (could be completed in 9 clicks) were recorded. For both tasks, most participants required no prompting—89% (n=8) participants for task 1 and 67% (n=6) participants for task 2, respectively. For effectiveness, all participants were able to complete each task accurately and comprehensively. Overall, the participants were highly satisfied with the dashboard with everyone remarking on the utility of using it to support their work, particularly to compare their jurisdiction to others. Finally, half of the participants stated that the ability to share the graphs from the dashboard would be "extremely useful" for their work. The only aspect of the dashboard cited as problematic is the amount of missing data that was present, which was a constraint of the data available about rural jurisdictions.

Conclusions: Think-aloud tasks showed that the SHAREdash allows users to complete tasks efficiently. Overall, participants reported being very satisfied with the dashboard and provided multiple ways they planned to use it to support their work. The main usability issue identified was the lack of available data indicating the importance of addressing the ongoing issues of missing and fragmented public health data, particularly for rural communities.

(JMIR Hum Factors 2024;11:e51666) doi:10.2196/51666

KEYWORDS

data dashboard; rural health; health equity; usability; nursing informatics; dashboard; rural; informatics; satisfaction; think aloud; content analysis; user experience; public health; visualization; information systems

Introduction

Data visualization dashboards developed to address health and equity have become increasingly popular [1,2]. Leveraging the longstanding history of using dashboards to aggregate and analyze data in public health [3] and medicine [4], these new dashboards cover myriad health equity–focused topics and target broad audiences. Recently, Thorpe and Gourevitch [5] identified 15 examples of US-based health dashboards that illustrate this growing trend. Examples range from a COVID-19 dashboard that highlights inequities in cases and deaths by geography to a policy dashboard that aggregates local laws and policies that affect population health [5]. Similar to these dashboards, the solutions in health analytics for rural equity across the northwest (SHARE-NW) dashboard (SHAREdash) was created to address health equity for rural communities.

Delivery and allocation of health services through public health agencies is a key mechanism for achieving health equity in the United States as they provide health prevention and promotion services and care [6]. Nationally, people in rural and frontier jurisdictions have significant health disparities compared with urban populations but are frequently the least well served by their public health agencies—local health departments (LHDs) [7,8]. Exacerbating this is the poor public health data systems, as updating to include information on structural and social factors has not been a top priority in LHDs' activities or spending [8,9]. Research has highlighted the critical need to improve timely and reliable population health data to inform resource allocation and decision-making [10-14]. Consequently, decisions regarding the delivery of public health services and care primarily rely on conventional wisdom. This results in services that frequently do not reflect the needs of the populations they serve resulting in wasteful, harmful, and inequitable inefficiencies that exacerbate existing disparities [15-17]. To address these issues and support LHDs serving rural areas, the goal of SHAREdash is to provide accessible and relevant data that will enable public health professionals to identify, communicate, and address health disparities in their jurisdictions and with their communities. Developed with user-centered participatory design methods and guided by Munzer's Nested Model for Visualization Design and Validation [18], SHAREdash is the first rigorously designed health equity dashboard developed for rural communities that we are aware of [19].

While clear objectives and thoughtful design are critical to ensuring the development of useful dashboards, Thorpe and Gourevitch [5] highlight the importance of evaluating dashboards and the need for a more rigorous assessment of the effectiveness and usefulness of health equity dashboards. Evaluating the performance of a dashboard through end user usability testing is a critical and often missed component of dashboard creation. The International Organization for Standardization defines usability as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specific context of use" [20,21]. Poor usability has been shown to increase errors [22-24], increase the time to complete tasks [25], and reduce user uptake [26-28] and implementation efforts [29].

Proper assessment of technology usability should occur at multiple stages throughout the system development lifecycle (SDLC) and use the methods most appropriate for that respective stage. In a review of usability study methodologies of health information technology by Yen and Bakken [30], the authors outline the importance of conducting multiple usability evaluations that align with the 5 stages of the SDLC. Furthermore, the Yen and Bakken [30] review clarifies the differences in usability evaluation types and goals based on the SDLC stage of the technology (Multimedia Appendix 1). Results from SHAREdash's usability testing for SDLC stages 1 and 2 have been previously published [10,19]. Stage 2 findings were used to make critical changes and inform dashboard completion. Now that SHAREdash is finished and has entered SDLC stage 3, we evaluated its usability by examining all components combined (ie, the finished dashboard). Thus, the aim of this study was to evaluate the SDLC stage 3 evaluation goals of efficiency, satisfaction, and validity for SHAREdash.

Methods

The SHARE-NW Project and Dashboard

SHARE-NW is a partnership research project that was created with the goal of making data available and accessible to rural LHD practitioners, while building their capacity for data use and data-driven decision-making [10]. Partnering with LHDs in Alaska, Idaho, Oregon, and Washington, 7 priority topic areas (obesity, diabetes, tobacco use, mental and behavioral health, violence and injury prevention, oral health, and demographics) were identified during stage 1 of the SDLC for SHAREdash [19]. Data for the dashboard come from 36 unique data sources, including national data from the Centers for Disease Control and Prevention as well as local agencies and health departments (Multimedia Appendix 2). Data were deemed relevant to be included in the dashboard if it (1) addressed 1 of the 7 priority areas and (2) was provided at the county level so that it would be relevant to LHDs. To ensure the dashboard is usable and relevant for users, its features (eg, dynamic filters, pop-up tooltips, and visualizations) were created in collaboration with the staff from partner LHDs during SDLC stage 2. SHARE-NW has also developed a curated repository of web-based trainings and webinars, including new training modules developed in 2021 and launched in 2022 when gaps were found in the related training desired by practitioners. The new training modules developed use problem-based learning to teach audiences how to use and communicate data to promote health equity.

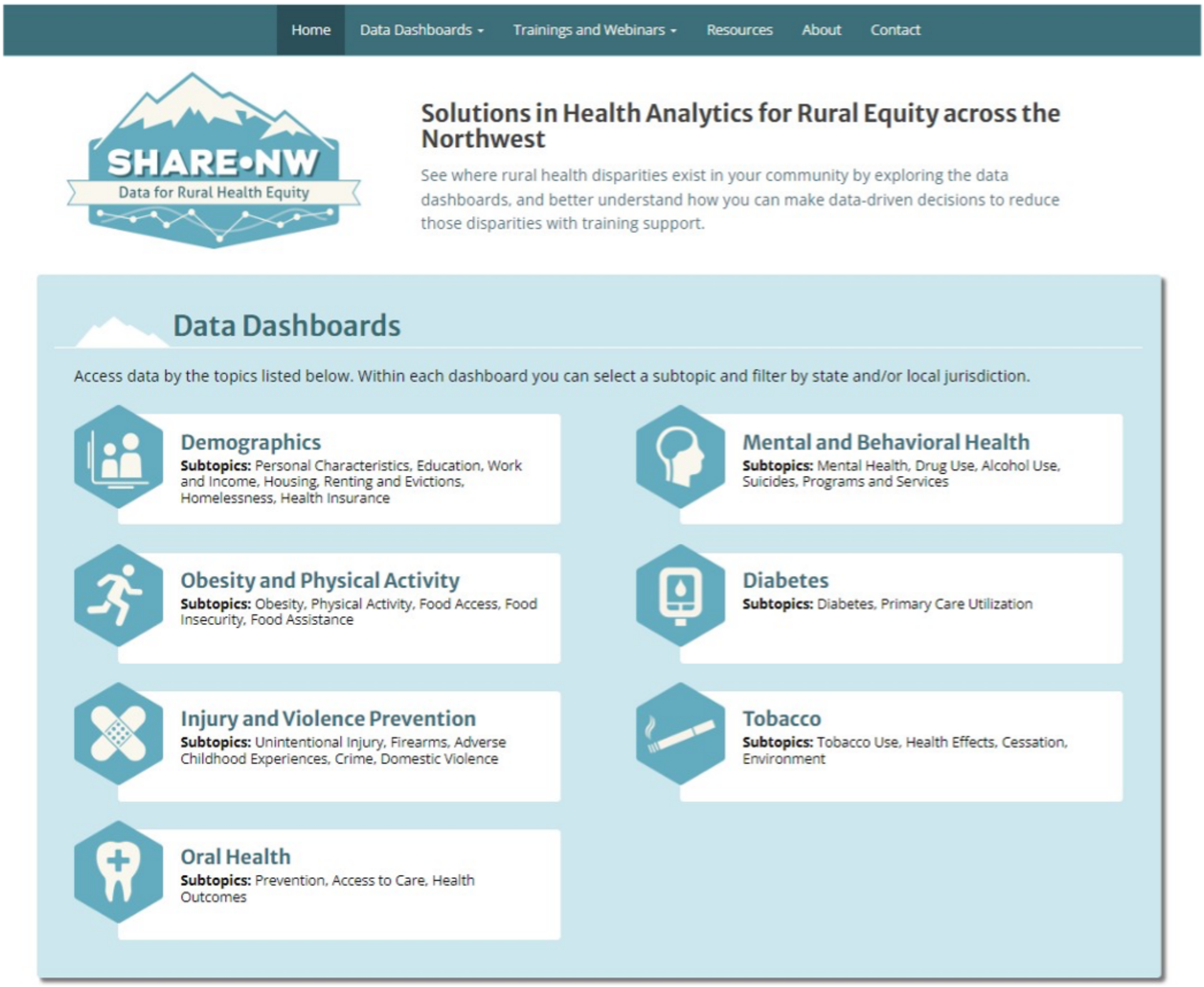
After conducting a needs assessment with rural LHD professionals during SDLC stage 1 [10], members of the SHARE-NW team identified a set of initial design requirements for SHAREdash. These requirements guided design and development decisions that ranged from key decisions, such as the selection of the best software to create the dashboard, to smaller decisions such as which size font to use for a graph label. Together with the findings from the SDLC stage 2 usability study, SHAREdash was completed and launched in August 2021.

SHAREdash Website and Interface

SHAREdash is a Tableau-based dashboard with a header at the top of the main page for users to locate information about the project and team, access resources on relevant topics such as data, communication, and health equity, how to contact a member of the team, and find the dashboards organized by priority topic area. Users can also see relevant trainings and webinars (both via drop-down boxes) on SHAREdash's main page. When users scroll down the main page, they can also find information on the website's purpose and design and see the sources of data powering the website. The largest feature on the main page (Figure 1) links to the 7 dashboards on the priority topic areas mentioned previously [19]. Within each dashboard, users can find state and county-level data organized by relevant subtopics. For example, the topic of "Tobacco" includes subtopics of "Tobacco use," "Health effects," "Cessation," and "Environment."

When users navigate to each of the main topics, they find a header that lists the main topic and each subtopic along the top, such that users can click through them. Within each subtopic, there are several drop-downs that allow users to filter the data. The primary drop-down prompts the user to "Select an Indicator." Some examples of indicators for the topic of "Tobacco" and the subtopic of "Tobacco use" are as follows: "8th graders' current tobacco use include Percent %," "high school students who smoked tobacco in the past month: Percent %," "Adults who currently smoke, Age-adjusted Percent %," and "Current e-cigarette use: Percent %." The remaining drop-downs allow users to filter the results by state, region, rurality (eg, rural or not rural), and jurisdiction types (eg, county). Users can export any of the dashboard views using the 3 options of export image, export to PDF, and share a link that is found along the top of the page. Along the bottom of each dashboard is the clickable link to view the data sources for this dashboard along with a statement explaining which data are listed as unavailable data within the dashboards.

Figure 1. The solutions in health analytics for rural equity across the northwest dashboard (SHAREdash) home page.



Study Setting and Participants

To evaluate this web-based dashboard, our study was conducted from January to April 2022. Participants were recruited from the states upon which SHAREdash was focused—Washington, Idaho, Alaska, and Oregon. All individuals who were rural public health professionals or trainees and had completed at least 1 prior SHARE-NW activity (2017-2022) and agreed to be contacted for future research activities (n=20). Prior SHARE-NW activities included the following: key informant interviews, interviews about the response to COVID-19, web-based surveys, dashboard mock-up testing sessions, dashboard usability testing activities, and group-based dashboard live training sessions. This ensured that all participants met the eligibility criteria of being at least 18 years old, working in public health for at least a year, and were in 1 of the 4 northwest states included in SHAREdash. Given public health differences by state, recruitment efforts ensured that at least 3 of the states were represented. Recruitment of this convenience sample had no additional inclusion or exclusion criteria.

Ethical Considerations

The University of Washington's institutional review board (IRB) approved the study protocol before participant recruitment (STUDY00013451). This study's IRB was approved as a modification to an original approval (MOD00011747; approved on December 13, 2021).

An initial recruitment email briefly summarizing the study purpose was sent to all prior participants in SHARE-NW activities. A follow-up recruitment email was sent 2 weeks after the initial email. The response rate for recruitment was 45% (n=9) with 1 person stating that they did not want to participate and the remaining 10 people not responding. Individuals who expressed interest were sent an email with information about the study, consent to participate, and instructions on scheduling their interview. Dashboard evaluation sessions consisted of 2 think-aloud tasks [31] and open-ended interview questions [32] regarding the participant's occupation and perceptions of the dashboard (Multimedia Appendices 3 and 4). Recruitment stopped when the following metrics were reached: (1) alignment with the published literature on the minimum number of participants needed to identify usability issues [33] and (2) when data saturation was reached. Given how stretched our public health partners were from the COVID-19 pandemic, the study team was cautious to not overburden them with study participation requests.

The think-aloud tasks served 2 purposes—the first was to refamiliarize the participant with SHAREdash and the second was to examine the usability components of effectiveness and efficiency. The first think-aloud task had the participant complete a simple task that consisted of switching between different subtopics, filtering for the participant's county, and changing the time frame being viewed. The second, more complex task included navigating to the right topic, filtering for a specific health outcome type, year, and rate, and identifying the original sources of the data being viewed. During testing, the moderator prompted participants to "think aloud" that is, verbalize their thoughts as they worked through the task. Following the tasks, the participants completed qualitative interviews that asked participants for their perceptions regarding SHAREdash's efficiency, validity, and satisfaction. The semistructured interview guide included questions that asked about the design aesthetics and functionality of SHAREdash, how quickly they are able to perform tasks, and the benefits and issues with using SHAREdash. The evaluation sessions were completed and recorded via a videoconferencing platform since screen sharing was needed for the 2 think-aloud task evaluations. Transcripts were automatically generated by the videoconferencing platform and stored securely in a password-protected cloud-based repository. A member of the research team deidentified and corrected any errors in the verbatim transcripts prior to analysis.

Recruitment and Data Collection

Think-Aloud Task Analysis

Operational definitions of the outcomes align with the International Organization for Standardization definitions of efficiency, satisfaction, and validity [34]. Efficiency and validity were primarily evaluated through the think-aloud tasks. The number of clicks taken to complete the task indicated efficiency and data on the participants' success of task completion were operationalized as "yes" (eg, no assistance needed), "no" (eg, assistance needed), or "partial" (eg, where the moderator confirms participant choices as either correct or incorrect but offers no other assistance) which indicated validity. Task 1 could be completed in a minimum of 7 clicks and task 2 could be done in 9 clicks. Transcripts were automatically generated and edited by a member of the research team for accuracy. Descriptive statistics were used to analyze the data and was completed in Excel. Quotes from the think-aloud tasks were analyzed to evaluate common efficiency issues, examine overall satisfaction, and assess validity. A control arm was not used in this study based on prior work that identified the inability of participants to complete these tasks without SHAREdash [10,19].

Qualitative Analysis

Data analysis of qualitative interview transcripts started with a directed, deductive approach to content analysis that was guided by a codebook comprising the initial codes of efficiency, satisfaction, and validity [35]. From this initial schema, iterative coding categories emerged as themes were developed. Coding was performed in NVivo (Lumivero) to organize data and provide an audit trail. Our interdisciplinary team of researchers met for an initial 90-minute collaborative coding session to talk through coding procedures and develop consensus for initial categories. Subsequent coding was performed independently with researchers meeting for 60-minute coding meetings to discuss categories and resolve discrepancies. Procedures for ensuring credibility, transferability, dependability, and confirmability were incorporated throughout the research process to ensure data trustworthiness. These procedures included taking field notes, team debriefing, reflexive journaling, consideration of negative cases, and maintenance of an audit trail. Data saturation was reached with researchers initially identifying the potential for saturation after the sixth participant interview and later confirming it with the ninth and final participant.

Results

Overview

Interviews lasting an average of 21 (SD 5.4) minutes were conducted between January and April 2022. Of the 9 public health practitioners interviewed, 4 were from Idaho, 3 were from Oregon, and 2 were from Washington (Table 1). Participants all identified as female, and the majority worked for health departments (n=8). Job positions included a director (n=2), managers (n=2), program specialists/coordinators (n=3), an epidemiologist (n=1), and a student/public health intern (n=1). Prior experience with Tableau was minimal with the majority (7/9, 78%) reporting less than 3 months experience to no experience.

Table 1. Demographics of dashboard evaluation participants (N=9).

Characteristic	Values, n (%)
State	
Idaho	4 (44)
Oregon	3 (33)
Washington	2 (22)
Organization type	
Local health department	7 (78)
State health department	1 (11)
Educational institution	1 (11)
Position	
Director	2 (22)
Coordinator	2 (22)
Manager	2 (22)
Other	3 (33)
Tableau experience	
0-3 months	7 (78)
>3 months	1 (11)
Not reported	1 (11)
Sex	
Female	9 (100)
Ethnicity	
Not reported	5 (56)
White	4 (44)

Think-Aloud Task Results

For efficiency, mean clicks were 10.1 (SD 1.4) for task 1 (with a minimum of 7 clicks) and 11.4 (SD 2.0) for task 2 (with a minimum of 9 clicks; Table 2). For task 1, extra clicks occurred when people tried to find the right place to filter for the correct dashboard page. One participant (participant 9) required partial assistance with one of the steps in the first task. They initially thought to navigate to the default subtopic of "Personal Characteristics" instead of the correct subtopic "Homelessness" in the "Demographics" dashboard. Although only 1 participant required assistance with this step, many participants took extra time with it. For task 2, extra clicks resulted from people looking for the sources of the data, which were located at the footer of each dashboard. Most participants were able to find the "View Data Sources" button easily because of the dashboard's instructions or because it was where they expected it based on their prior experience. However, 3 participants noted that they naturally scrolled to the bottom of the dashboard looking for it but confused the "Resources" button with the "View Data Sources" button. This issue not only resulted in extra clicks but also was the point where these participants required confirmation by the moderator to continue. Of the 3 participants, 2 who had this issue stated that they were confused because they expected the data sources to be in a clickable pop-up or an in-text citation, rather than loading onto a separate page.

Table 2. Efficiency task analysis results (N=9).

Task type	Task 1	Task 2
Efficiency		
Clicks, mean (SD)	10.1 (1.4)	11.4 (2.0)
Clicks, median (IQR)	10.5 (9.0-11.0)	11.0 (9.5-12.8)
Validity, n (%)		
Successful	8 (89)	6 (67)
Partial	1 (11)	3 (33)
Not successful	0 (0)	0 (0)

Validity scores for both tasks were high with all participants (n=9) able to complete each task accurately and comprehensively. Most participants received a "yes" indicating that they did not require any prompting—89% (n=8) participants for task 1 and 67% (n=6) participants for task 2. For participants who did not receive a score of "yes," they only required the moderator to either confirm or deny their decisions prior to moving on, resulting in a score of "partial." None of the participants received a "no," indicating they could not finish the tasks.

Qualitative Results

The following 4 themes regarding efficiency were identified: "using the best terms and names to increase efficiency," "drop-down filters reduce efficiency," "minor navigation issues affect efficiency," and "learnability will increase efficiency over time". The primary issue that came up in the think-aloud tasks and the interviews was related to the dashboard labels and names that informed the theme of "using the best terms and names to increase efficiency." Multiple participants brought up the term "jurisdiction" and pointed out that it is less intuitive than the word "county." "I think 'jurisdictions' is obviously not wrong; it just would be a little bit more user-friendly to label it 'county'" (participant 1). Similarly, as identified in the second task analysis, 2 participants found the "Resources" button confusing and suggested renaming it to something more specific to mitigate this confusion.

For the second theme of "drop-down filters reduce efficiency," participants described how the functionality of the filters was difficult to navigate between some options due to the length of the drop-down boxes. For example, filtering to examine a single county requires users to search down a long drop-down list for the exact county they are looking for. Participants described this by saying, "Maybe it would be a nice feature to be able to type in a county versus the drop-down box or having to—well I guess you can 'select all' so you don't have to go through and click them all to select, but just those little things might make it easier" [participant 4]. Another participant described how they expected the interface to be like other software they are used to using such that it allows users to enter free text into a search bar and then, "…when you start typing things it only picks the things that match it" (participant 8).

For the third theme of "minor navigation issues affect efficiency," a few participants had difficulty locating the various subtopics within a dashboard, despite them being listed underneath each dashboard topic. For example, 1 participant looked for the "homelessness" indicator under the wrong subtopic.

> *I was initially thinking, 'Oh 'Homelessness' must be in one of these drop downs, because it was listed as a subtopic,' but then I glanced across the screen, and—you know—I saw 'Homelessness' up in this corner [with the other subtopics].* [participant 2]

Similarly, 3 participants eventually correctly identified that "Demographics" was the dashboard where they would find homelessness data, but they initially looked for the "homelessness" indicator under the "Housing" subtopic instead of the "Homelessness" subtopic. A participant suggested ways that the design of SHAREdash could be updated to more clearly indicate the subtopics.

> *I would think that [the indicator] is definitely going to be in Oral Health. It took a bit when I first looked at [SHAREdash] to realize that there were tabs (e.g., different subtopics). I think the size of the font and the fact that they are the same color as the bar makes it, so they are not standing out.* [participant 1]

These design suggestions were checked with some subsequent participants who agreed that changing the font size and color would help the subtopics stand out.

For the final theme of "learnability will increase efficiency over time," participants spoke about how quickly they were able to figure things out in SHAREdash and reported that with repeated use they thought they would quickly improve over time. Half of the participants stated that first-use learnability was high such that SHAREdash was easy to use the first time they tried. "I would say that there's not a lot of websites out there, where you can pick up on things that quickly. So, I immediately don't have any areas for improvement" [participant 2]. Whereas the remaining half of the participants stated that they felt like they would get progressively better at using SHAREdash over time.

The following 3 themes related to overall satisfaction were identified: "high potential to support work," "enables meaningful comparisons," and "needs more up-to-date data." For the theme of "high potential to support work," participants spoke positively about how much they liked SHAREdash and the myriad ways they could use SHAREdash's various features to support their work. One-third of the participants mentioned how unique and helpful it was to have the ability to export and share graphs.

> *I think it has a lot of features that aren't necessarily easily found [in other dashboards]…blowing it [SHAREdash] up to full screen, downloading it, sharing it—that's not necessarily common with dashboards, so I appreciate that…It could be really useful for like a grant application or demographic reporting for part of a program.* [participant 5]

Two participants mentioned that they might direct others to the dashboard so they could interact with data, and this was described as something that would be "extremely useful" and "super helpful" in their work. Finally, several participants identified specific types of work that SHAREdash would meaningfully support such as completing community health assessments:

> *We would definitely want to look at this in relation to the approach that we took with our CHA [community health assessment]. Most recently, I was trying to mine all of the data sources that are already in existence to inform it and see where some gaps were, and then we did primary data seeking based off those gaps instead of trying to reproduce data that's already in existence. And so, this [SHAREdash] would be a really great one-stop-shop to look at a lot of different ones at one time.* [participant 2]

> *I think it is already something that's on our radar when we talk about this CHA [community health assessment] that I mentioned. So, we're not here to duplicate efforts; let's use what's out there. And so, we'll probably refer to it [SHAREdash] for that.* [participant 5]

For the theme of "enables meaningful comparisons," all but 2 participants reported that they were highly satisfied with SHAREdash and cited the ability to compare their county or region with other neighboring or similar counties in different states as the reason why. Multiple participants stated they wanted to look at counties in nearby states given their close proximity and described how SHAREdash fills this gap since states do not typically share data with one another.

> *Being able to look at data kind of in the same place and say 'Oh, what does your county look like?' You know, which borders us in Oregon, but borders like three of the counties that I oversee. So, what's happening in their county? I can look that up and see if we're seeing similar trends, and the three counties that border that county. So having that originality, I think, is great and is probably a reason that I would go to the website to look at that at some point, or my team would.* [participant 9]

> *It is nice that it includes multiple states, because we are border county in our state, and so a lot of times things that we see are only for Oregon. But we're right next to a couple of Washington counties and it would be great to, you know, compare in that manner as well…It's always really helpful when we can look at, you know, what is our information compared to our neighboring counties, what does our data look like compared to counties of similar size.* [participant 2]

> *I like that you can see a big picture, regionally. So not necessarily just like other counties in Idaho: being able to prepare to other regional and other states and perhaps similar geographic demographic areas that are comparable, but in different states kind of just to see what trends are like there comparatively.* [participant 4]

The third theme of "needs more up-to-date data" described the biggest challenge that participants identified to their overall satisfaction with SHAREdash.

> *I think just what I commented on already is the age of the data that is present. So, it is very difficult to make a decision on data that's extremely outdated. And it's hard to make it relevant to your case. And I know that data can be hard to gather and hard to access, but for those of us who are looking at data to make decisions, that complicates that entire scenario. You want us to use data to make decisions, we need good data to make those decisions. Somebody has to put the data out.* [participant 6]

Several participants acknowledged that none or out-of-date data are typical within public health, particularly for rural areas. "We're used to that, so I think for us that's not a missed expectation to click on it and be like 'Oh, there's not any new data.'…That for us, that's normal" [participant 2]. However, this is a clear barrier to satisfaction and future use of SHAREdash.

There were 2 themes identified on validity called "reputable data sources increases validity" and "impact of missing data decreases validity." Participants spoke about the second theme of "reputable data sources increases validity" by describing their confidence in the data quality and accuracy. A participant described this by stating that, "SHAREdash is a really amazing place to quickly get domestic violence rates across other states. And you can find the source easily. And it is a reputable source too" [participant 1].

Whereas, another participant emphasized more than just the high quality of the data sources, but also the fact that SHAREdash's team provided a second, external check on it:

> *So I think this is a great dashboard and it's so nice because part of my job is to pull [data] from all of these different data sources which I know SHARE has done, and it's been validated and checked and it's a combination of information from various places which is good to have.* [participant 3]

For the "impact of missing data decreases validity" theme, task 1 had participants refined the population of interest to their specific county, which for some participants resulted in SHAREdash indicating that there were no data for their respective county available. Participants described how missing data in the dashboard impacted their ability to completely address the tasks in the think-aloud evaluation and how it would impact their work.

> *I think one of my biggest challenges, and it tends to be a challenge everywhere not just like solely for the dashboards, is that a lot of times when there were things I wanted to look at and there wasn't any data available because our population isn't that big.* [participant 8]

Despite multiple participants acknowledging that problems with data availability for rural areas is a known issue and is not a fault of the dashboard, they still expressed frustration and dissatisfaction about this issue.

> *I was bummed when it didn't have the data that I was looking for. But like I said, it's probably just a result of that data not being available.* [participant 4]

Discussion

Principal Findings

This SDLC stage 3 usability evaluation of SHAREdash, a dashboard designed for rural public health, indicates that overall SHAREdash is an efficient and valid tool that users reported being satisfied with. Task analyses and qualitative findings illustrate how SHAREdash's collaborative co-design process resulted in a tool that is easy to use and supports rural public health professionals' work. Thematic results also identified areas where SHAREdash can be improved to increase its usability such as changing some of the terms and names used and considering alternate ways for users to view and select information that are not just drop-down filters. However, this evaluation also uncovered usability issues related to the lack of public health data that go beyond design aspects and cannot be addressed through modifying SHAREdash's interface or navigation.

Issues related to obtaining quality public health data are well documented in the literature and include the critical problems of a lack of investment in public health data systems and infrastructure [36-39], issues with data quality [40-42] and data fragmentation [43,44], and the sparse data available about rural communities [8,45]. While every effort was made to include as much timely and comprehensive data as possible in SHAREdash, these larger data problems clearly impacted the usability of this tool. Thus, returning to the question posed by Thorpe and Gourevitch [5] regarding whether or not data dashboards for advancing health and equity are fulfilling their promise, findings from our study show that, to fully realize the potential of health equity–focused dashboards substantial investments in public health data need to be made. Unlike health care which benefited from the 2009 Health Information Technology for Economic and Clinical Health Act that supported the adoption of meaningful use of electronic health records [46,47], other systems, such as public health and social services, were not included in these incentives resulting in the numerous data issues seen today. The need for investments in public health data and systems was further clarified and reinforced by the COVID-19 pandemic which magnified many of these ongoing challenges [38,45,48]. To address this urgent problem, supportive policies that fund public health data collection and systems should leverage the successes of the Health Information Technology for Economic and Clinical Health Act and learn from the opportunities to address and alleviate these issues.

Another key finding from this study is also related to data. All the participants emphasized the significance of the trustworthiness of the data in SHAREdash. These results align with prior literature that has articulated the dual importance of dashboards to use data from reputable sources and clearly display or link to original data sources [49]. In a 2020 study by Young and Kitchin [50] that examined user perspectives of 4 different city's dashboards to create design guidelines, the authors stipulate how critical the veracity (eg, accuracy, source, and age) of the included data is. Our findings reinforce this work and indicate the utility of their design guidelines for creating data dashboards of municipal data. Future dashboards of municipal data should use the guidelines provided by Young and Kitchin [50] in the early design and development stages and work with target users to refine them for their specific project needs.

Satisfaction with SHAREdash was high with most participants describing the usefulness of the dashboard in supporting their work. Almost all the participants reported that they would like to make local-level comparisons that cross their respective states and articulated how difficult this currently is. Participants reported how comparisons between counties across different states can be more meaningful than within if they are able to filter for key factors such as population size or number of services available and how helpful it is that SHAREdash facilitates this easily. Enabling such comparisons points to the importance of aggregating large amounts of data across states, particularly for rural health departments that have unique needs and face different challenges than their urban counterparts [51]. It also indicates the importance of continuing to elucidate the unique needs of rural public health. Future research should focus on rural public health so that tailored design guidelines and specialized tools can be developed to support their work in addressing health disparities.

Our SDLC stage 3 usability assessment indicated that SHAREdash is meeting the goal of providing accurate, accessible, and relevant data via a user-centered dashboard to address health equity for rural communities. Next steps for SHAREdash will focus on identifying the elements key to its integration into LHDs using an implementation science approach that is outlined in stage 4 of the SDLC [30]. Planning for this phase is underway and is working closely with future end users to proactively identify and understand barriers to integration as this was a clear lesson learned from a similar study implementing an ICU dashboard [52]. Furthermore, investing in efforts to understand what is needed to support the uptake of health equity–focused dashboards in public health practice is critical to ensuring their impact [5] and aligns with previously identified public health research priorities [53-55] that highlights the importance of using implementation science to translate and assess innovations into public health practice to ensure reach. It is hoped that through the user-centered development and thoughtful translation of informatics, tools such as SHAREdash will address the existing health disparities and improve rural health equity.

Limitations

While our methods were rigorous, this study has limitations. Despite reaching data saturation, the sample size is small, consisting of all female-identifying participants, and limited to the northwest United States. Of note, the public health workforce is 79% women [56], which made diversity by sex difficult to obtain. Future studies would benefit from a larger and broader sample. Additionally, the think-aloud task analysis did not have a control arm where participants completed the tasks without SHAREdash to provide a comparison. Finally, while aligned with the SLC stage 3 usability evaluation components, this study did not examine other aspects of usability that are outlined in the literature, and thus, might have missed certain usability aspects [57].

Conclusions

Evaluating the usability of health equity dashboards is crucial to creating effective and valuable tools. Our findings indicate that SHAREdash, a public health dashboard created to support promoting health equity among rural communities, is an efficient, valid tool that overall users are satisfied with. Results strongly suggest that the utility of dashboards such as SHAREdash would be improved with the availability of more public health data and supportive policies to achieve robust collection of public health data would be beneficial. Future research should continue to focus on building tools that meet the unique needs of professionals working in rural public health to better support and equip them to alleviate rural health disparities.

Acknowledgments

This work was funded by the Department of Health and Human Services, Office of the Assistant Secretary for Health (grant #1 CPIMP171144-01-00). We would like to thank the public health professionals who participated in this study; the SHARE-NW team members who are not authors of this paper for their contributions to this project; and the SHARE-NW Equity Advisory Committee members for their expertise and guidance.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Usability evaluation types and goals based on the system development lifecycle (SDLC) stage (adapted from Yen and Bakken [30], which is published under Creative Commons Attribution-NonCommercial-NoDerivs licence).
[DOCX File , 19 KB - humanfactors_v11i1e51666_app1.docx]

Multimedia Appendix 2
Information on SHAREdash data sources.
[DOCX File , 34 KB - humanfactors_v11i1e51666_app2.docx]

Multimedia Appendix 3
Information on task analyses.
[DOCX File , 18 KB - humanfactors_v11i1e51666_app3.docx]

Multimedia Appendix 4
Semistructured interview guide.
[DOCX File , 20 KB - humanfactors_v11i1e51666_app4.docx]

References

1. Dasgupta N, Kapadia F. The future of the public health data dashboard. Am J Public Health 2022;112(6):886-888 [FREE Full text] [doi: 10.2105/AJPH.2022.306871] [Medline: 35613427]
2. Vahedi A, Moghaddasi H, Asadi F, Hosseini AS, Nazemi E. Applications, features and key indicators for the development of COVID-19 dashboards: a systematic review study. Inform Med Unlocked 2022;30:100910 [FREE Full text] [doi: 10.1016/j.imu.2022.100910] [Medline: 35342788]
3. Carroll LN, Au AP, Detwiler LT, Fu TC, Painter IS, Abernethy NF. Visualization and analytics tools for infectious disease epidemiology: a systematic review. J Biomed Inform 2014;51:287-298 [FREE Full text] [doi: 10.1016/j.jbi.2014.04.006] [Medline: 24747356]
4. Murphy DR, Savoy A, Satterly T, Sittig DF, Singh H. Dashboards for visual display of patient safety data: a systematic review. BMJ Health Care Inform 2021;28(1):e100437 [FREE Full text] [doi: 10.1136/bmjhci-2021-100437] [Medline: 34615664]
5. Thorpe LE, Gourevitch MN. Data dashboards for advancing health and equity: proving their promise? Am J Public Health 2022;112(6):889-892 [FREE Full text] [doi: 10.2105/AJPH.2022.306847] [Medline: 35446603]

6. The foundational public health services. Public Health Accreditation Board. 2022. URL: https://phaboard.org/center-for-innovation/public-health-frameworks/the-foundational-public-health-services [accessed 2024-03-01]
7. Gibson PJ, Shah GH, Streichert LC, Verchick L. Urgent challenges for local public health informatics. J Public Health Manag Pract 2016;22(Suppl 6):S6-S8 [FREE Full text] [doi: 10.1097/PHH.0000000000000479] [Medline: 27684620]
8. Harris JK, Beatty K, Leider JP, Knudson A, Anderson BL, Meit M. The double disparity facing rural local health departments. Annu Rev Public Health 2016;37:167-184 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122755] [Medline: 26735428]
9. Bekemeier B, Marlowe J, Squires LS, Tebaldi J, Park S. Perceived need versus current spending: gaps in providing foundational public health services in communities. J Public Health Manag Pract 2018;24(3):271-280. [doi: 10.1097/PHH.0000000000000612] [Medline: 28832431]
10. Bekemeier B, Park S, Backonja U, Ornelas I, Turner AM. Data, capacity-building, and training needs to address rural health inequities in the Northwest United States: a qualitative study. J Am Med Inform Assoc 2019;26(8-9):825-834 [FREE Full text] [doi: 10.1093/jamia/ocz037] [Medline: 30990561]
11. Bekemeier B, Pantazis A, Yip MPY, Kwan-Gett T. Developing the evidence for public health systems to battle vaccine preventable disease at the local level: data challenges and strategies for advancing research. J Public Health Manag Pract 2017;23(2):131-137. [doi: 10.1097/PHH.0000000000000411] [Medline: 27798522]
12. Burgard SA, Chen PV. Challenges of health measurement in studies of health disparities. Soc Sci Med 2014;106:143-150 [FREE Full text] [doi: 10.1016/j.socscimed.2014.01.045] [Medline: 24561776]
13. Feng W, Martin EG. Fighting obesity at the local level? an analysis of predictors of local health departments' policy involvement. Prev Med 2020;133:106006. [doi: 10.1016/j.ypmed.2020.106006] [Medline: 32007526]
14. Zhang X, Luo H, Gregg EW, Mukhtar Q, Rivera M, Barker L, et al. Obesity prevention and diabetes screening at local health departments. Am J Public Health 2010;100(8):1434-1441 [FREE Full text] [doi: 10.2105/AJPH.2009.168831] [Medline: 20558810]
15. Mays GP, Smith SA, Ingram RC, Racster LJ, Lamberth CD, Lovely ES. Public health delivery systems: evidence, uncertainty, and emerging research needs. Am J Prev Med 2009;36(3):256-265. [doi: 10.1016/j.amepre.2008.11.008] [Medline: 19215851]
16. Mete C, Cioffi JP, Lichtveld MY. Are public health services available where they are most needed? an examination of local health department services. J Public Health Manag Pract 2003;9(3):214-223. [doi: 10.1097/00124784-200305000-00006] [Medline: 12747318]
17. Santerre RE. Jurisdiction size and local public health spending. Health Serv Res 2009;44(6):2148-2166 [FREE Full text] [doi: 10.1111/j.1475-6773.2009.01006.x] [Medline: 19656226]
18. Munzner T. A nested model for visualization design and validation. IEEE Trans Vis Comput Graph 2009;15(6):921-928. [doi: 10.1109/TVCG.2009.111] [Medline: 19834155]
19. Backonja U, Park S, Kurre A, Yudelman H, Heindel S, Schultz M, et al. Supporting rural public health practice to address local-level social determinants of health across Northwest states: development of an interactive visualization dashboard. J Biomed Inform 2022;129:104051 [FREE Full text] [doi: 10.1016/j.jbi.2022.104051] [Medline: 35318149]
20. Abran A, Khelifi A, Suryn W, Seffah A. Usability meanings and interpretations in ISO standards. Softw Qual J 2003;11:325-338 [FREE Full text] [doi: 10.1023/a:1025869312943]
21. Bevan N, Carter J, Earthy J, Geis T, Harker S. New ISO standards for usability, usability reports and usability measures. In: Kurosu M, editor. Human-Computer Interaction: Theory, Design, Development and Practice. HCI 2016. Lecture Notes in Computer Science Vol. 9731. Cham: Springer; 2016:268-278.
22. Kushniruk AW, Triola MM, Borycki EM, Stein B, Kannry JL. Technology induced error and usability: the relationship between usability problems and prescription errors when using a handheld application. Int J Med Inform 2005;74(7-8):519-526. [doi: 10.1016/j.ijmedinf.2005.01.003] [Medline: 16043081]
23. Carayon P, Du S, Brown R, Cartmill R, Johnson M, Wetterneck TB. EHR-related medication errors in two ICUs. J Healthc Risk Manag 2017;36(3):6-15 [FREE Full text] [doi: 10.1002/jhrm.21259] [Medline: 28099789]
24. Brown CL, Mulcaster HL, Triffitt KL, Sittig DF, Ash JS, Reygate K, et al. A systematic review of the types and causes of prescribing errors generated from using computerized provider order entry systems in primary and secondary care. J Am Med Inform Assoc 2017;24(2):432-440 [FREE Full text] [doi: 10.1093/jamia/ocw119] [Medline: 27582471]
25. Gephart S, Carrington JM, Finley B. A systematic review of nurses' experiences with unintended consequences when using the electronic health record. Nurs Adm Q 2015;39(4):345-356. [doi: 10.1097/NAQ.0000000000000119] [Medline: 26340247]
26. Schnall R, Mosley JP, Iribarren SJ, Bakken S, Carballo-Diéguez A, Iii WB. Comparison of a user-centered design, self-management app to existing mHealth apps for persons living with HIV. JMIR Mhealth Uhealth 2015;3(3):e91 [FREE Full text] [doi: 10.2196/mhealth.4882] [Medline: 26385783]
27. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;Suppl:49-56 [FREE Full text] [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
28. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]

29. Kaufman D, Roberts WD, Merrill J, Lai TY, Bakken S. Applying an evaluation framework for health information system design, development, and implementation. Nurs Res 2006;55(Suppl 2):S37-S42. [doi: 10.1097/00006199-200603001-00007] [Medline: 16601633]
30. Yen PY, Bakken S. Review of health information technology usability study methodologies. J Am Med Inform Assoc 2012;19(3):413-422 [FREE Full text] [doi: 10.1136/amiajnl-2010-000020] [Medline: 21828224]
31. Lewis C. Using the "thinking-aloud" method in cognitive interface design. IBM T. J. Watson Research Center. 1982. URL: https://dominoweb.draco.res.ibm.com/2513e349e05372cc852574ec0051eea4.html#:~:text=%22Thinking%2Daloud%22%20is%20a,to%20use%20a%20computer%20system [accessed 2024-05-03]
32. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: a scoping review. Int J Med Inform 2019;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
33. Nielsen J. Usability Engineering. San Francisco, CA: Morgan Kaufmann; 1994.
34. Bevan N, Carter J, Harker S. ISO 9241-11 revised: what have we learnt about usability since 1998? In: Kuroso M, editor. Human-Computer Interaction: Design and Evaluation. HCI 2015. Lecture Notes in Computer Science Vol. 9169. Cham: Springer; 2015:143-151.
35. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
36. Alfonso YN, Leider JP, Resnick B, McCullough JM, Bishai D. US public health neglected: flat or declining spending left states ill equipped to respond to COVID-19. Health Aff (Millwood) 2021;40(4):664-671 [FREE Full text] [doi: 10.1377/hlthaff.2020.01084] [Medline: 33764801]
37. Orr JM, Leider JP, Singh S, Balio CP, Yeager VA, Bekemeier B, et al. Regarding investment in a healthier future: impact of the 2012 Institute of Medicine finance report. J Public Health Manag Pract 2022;28(1):E316-E323. [doi: 10.1097/PHH.0000000000001209] [Medline: 32956294]
38. Degutis LC, Shoaf K, Aragón TJ, Atchison C, Dyjack D, Hites L, et al. The urgent need for public health preparedness funding and support. Am J Public Health 2021;111(3):363-365. [doi: 10.2105/AJPH.2020.306068] [Medline: 33566658]
39. Yeager VA, Balio CP, McCullough JM, Leider JP, Orr J, Singh SR, et al. Funding public health: achievements and challenges in public health financing since the Institute of Medicine's 2012 report. J Public Health Manag Pract 2022;28(1):E244-E255. [doi: 10.1097/PHH.0000000000001283] [Medline: 33605671]
40. Dolley S. Big data's role in precision public health. Front Public Health 2018;6:68 [FREE Full text] [doi: 10.3389/fpubh.2018.00068] [Medline: 29594091]
41. Khoury MJ, Engelgau M, Chambers DA, Mensah GA. Beyond public health genomics: can big data and predictive analytics deliver precision public health? Public Health Genomics 2018;21(5-6):244-250 [FREE Full text] [doi: 10.1159/000501465] [Medline: 31315115]
42. Hoffman S, Podgurski A. Big bad data: law, public health, and biomedical databases. J Law Med Ethics 2013;41(Suppl 1):56-60. [doi: 10.1111/jlme.12040] [Medline: 23590742]
43. Velmovitsky PE, Bevilacqua T, Alencar P, Cowan D, Morita PP. Convergence of precision medicine and public health into precision public health: toward a big data perspective. Front Public Health 2021;9:561873 [FREE Full text] [doi: 10.3389/fpubh.2021.561873] [Medline: 33889555]
44. Iyamu I, Gómez-Ramírez O, Xu AX, Chang HJ, Watt S, Mckee G, et al. Challenges in the development of digital public health interventions and mapped solutions: findings from a scoping review. Digit Health 2022;8:20552076221102255 [FREE Full text] [doi: 10.1177/20552076221102255] [Medline: 35656283]
45. Bekemeier B, Heitkemper E, Backonja U, Whitman G, Schultz M, Jiang Y, et al. Rural public health data challenges during the COVID-19 pandemic: the case for building better systems ahead of a public health crisis. J Public Health Manag Pract 2023;29(4):496-502. [doi: 10.1097/PHH.0000000000001726] [Medline: 36867496]
46. Health IT legislation. HealthIT.gov. 2023. URL: https://www.healthit.gov/topic/laws-regulation-and-policy/health-it-legislation [accessed 2023-08-02]
47. Burde H. Health law the hitech act-an overview. Virtual Mentor 2011;13(3):172-175 [FREE Full text] [doi: 10.1001/virtualmentor.2011.13.3.hlaw1-1103] [Medline: 23127320]
48. Budd J, Miller BS, Manning EM, Lampos V, Zhuang M, Edelstein M, et al. Digital technologies in the public-health response to COVID-19. Nat Med 2020;26(8):1183-1192 [FREE Full text] [doi: 10.1038/s41591-020-1011-4] [Medline: 32770165]
49. Ansari B, Martin EG. Development of a usability checklist for public health dashboards to identify violations of usability principles. J Am Med Inform Assoc 2022;29(11):1847-1858. [doi: 10.1093/jamia/ocac140] [Medline: 35976140]
50. Young GW, Kitchin R. Creating design guidelines for building city dashboards from a user's perspectives. Int J Hum Comput Stud 2020;140:102429 [FREE Full text] [doi: 10.1016/j.ijhcs.2020.102429]
51. Leider JP, Meit M, McCullough JM, Resnick B, Dekker D, Alfonso YN, et al. The state of rural public health: enduring needs in a new decade. Am J Public Health 2020;110(9):1283-1290 [FREE Full text] [doi: 10.2105/AJPH.2020.305728] [Medline: 32673103]

52. Wac M, Craddock I, Chantziara S, Campbell T, Santos-Rodriguez R, Davidson B, et al. Design and evaluation of an intensive care unit dashboard built in response to the COVID-19 pandemic: semistructured interview study. JMIR Hum Factors 2023;10:e49438 [FREE Full text] [doi: 10.2196/49438] [Medline: 37751239]
53. Shelton RC, Cooper BR, Stirman SW. The sustainability of evidence-based interventions and practices in public health and health care. Annu Rev Public Health 2018;39:55-76 [FREE Full text] [doi: 10.1146/annurev-publhealth-040617-014731] [Medline: 29328872]
54. Chambers DA. Commentary: increasing the connectivity between implementation science and public health: advancing methodology, evidence integration, and sustainability. Annu Rev Public Health 2018;39:1-4. [doi: 10.1146/annurev-publhealth-110717-045850] [Medline: 29272164]
55. Estabrooks PA, Brownson RC, Pronk NP. Dissemination and implementation science for public health professionals: an overview and call to action. Prev Chronic Dis 2018;15:E162 [FREE Full text] [doi: 10.5888/pcd15.180525] [Medline: 30576272]
56. Nicolaus T. Public health addressing gender gap in workforce leadership. Nation's Health 2022;52(6):1-6 [FREE Full text]
57. Nielsen J. Usability 101: introduction to usability. NN/g Nielsen Norman Group. 2013. URL: https://www.nngroup.com/articles/usability-101-introduction-to-usability [accessed 2023-08-02]

Abbreviations

IRB: institutional review board
LHD: local health department
SDLC: system development lifecycle
SHAREdash: solutions in health analytics for rural equity across the northwest dashboard
SHARE-NW: solutions in health analytics for rural equity across the northwest

Edited by D Gooch; submitted 07.08.23; peer-reviewed by CH Lai, Anonymous; comments to author 02.01.24; revised version received 24.03.24; accepted 18.04.24; published 05.06.24.

Please cite as:
Heitkemper E, Hulse S, Bekemeier B, Schultz M, Whitman G, Turner AM
The Solutions in Health Analytics for Rural Equity Across the Northwest (SHARE-NW) Dashboard for Health Equity in Rural Public Health: Usability Evaluation
JMIR Hum Factors 2024;11:e51666
URL: https://humanfactors.jmir.org/2024/1/e51666
doi:10.2196/51666
PMID:38837192

Original Paper

Implementing a Hospital Call Center Service for Mental Health in Uganda: User-Centered Design Approach

Johnblack K Kabukye[1,2], MBChB, MSc, PhD; Rosemary Namagembe[3], BSN; Juliet Nakku[4], MBChB, MMed; Vincent Kiberu[5], PhD; Marie Sjölinder[6], PhD; Susanne Nilsson[7], PhD; Caroline Wamala-Larsson[1], PhD

[1]SPIDER - The Swedish Program for ICT in Developing Regions, Department of Computer and Systems Sciences, Stockholm University, Stockholm, Sweden

[2]Uganda Cancer Institute, Kampala, Uganda

[3]Hutchinson Centre Research Institute of Uganda, Uganda Cancer Institute, Kampala, Uganda

[4]Butabika National Referral and Teaching Mental Hospital, Kampala, Uganda

[5]School of Public Health, Makerere University, Kampala, Uganda

[6]RISE - Research Institutes of Sweden, Kista, Sweden

[7]Unit for Integrated Product Development and Design, Department of Machine Design, KTH Royal Institute of Technology, Stockholm, Sweden

Corresponding Author:
Johnblack K Kabukye, MBChB, MSc, PhD
SPIDER - The Swedish Program for ICT in Developing Regions
Department of Computer and Systems Sciences
Stockholm University
Borgarfjordsgatan 12 (NOD Building)
PO Box 7003 Kista
Stockholm, SE-164 07
Sweden
Phone: 46 8162000
Fax: 46 87039025
Email: kabukye@dsv.su.se

Abstract

Background: Mental health conditions are a significant public health problem globally, responsible for >8 million deaths per year. In addition, they lead to lost productivity, exacerbate physical illness, and are associated with stigma and human rights violations. Uganda, like many low- and middle-income countries, faces a massive treatment gap for mental health conditions, and numerous sociocultural challenges exacerbate the burden of mental health conditions.

Objective: This study aims to describe the development and formative evaluation of a digital health intervention for improving access to mental health care in Uganda.

Methods: This qualitative study used user-centered design and design science research principles. Stakeholders, including patients, caregivers, mental health care providers, and implementation experts (N=65), participated in focus group discussions in which we explored participants' experience of mental illness and mental health care, experience with digital interventions, and opinions about a proposed digital mental health service. Data were analyzed using the Consolidated Framework for Implementation Research to derive requirements for the digital solution, which was iteratively cocreated with users and piloted.

Results: Several challenges were identified, including a severe shortage of mental health facilities, unmet mental health information needs, heavy burden of caregiving, financial challenges, stigma, and negative beliefs related to mental health. Participants' enthusiasm about digital solutions as a feasible, acceptable, and convenient method for accessing mental health services was also revealed, along with recommendations to make the service user-friendly, affordable, and available 24×7 and to ensure anonymity. A hospital call center service was developed to provide mental health information and advice in 2 languages through interactive voice response and live calls with health care professionals and peer support workers (recovering patients). In the 4 months after launch, 456 calls, from 236 unique numbers, were made to the system, of which 99 (21.7%) calls went to voicemails (out-of-office hours). Of the remaining 357 calls, 80 (22.4%) calls stopped at the interactive voice response, 231 (64.7%) calls were answered by call agents, and 22 (6.2%) calls were not answered. User feedback was positive, with callers appreciating the inclusion of peer support workers who share their recovery journeys. However, some participant recommendations (eg, adding video call options) or individualized needs (eg, prescriptions) could not be accommodated due to resource limitations or technical feasibility.

Conclusions: This study demonstrates a systematic and theory-driven approach to developing contextually appropriate digital solutions for improving mental health care in Uganda and similar contexts. The positive reception of the implemented service underscores its potential impact. Future research should address the identified limitations and evaluate clinical outcomes of long-term adoption.

(JMIR Hum Factors 2024;11:e53976) doi:10.2196/53976

KEYWORDS

mHealth; mobile health; digital health; digital solution; digital solutions; digital intervention; digital interventions; mental health; awareness; Uganda; Africa; African; user centred; user centered; design; qualitative; focus group; focus groups; call centre; call centres; call center; call centers; mental; experience; experiences; attitude; attitudes; opinion; perception; perceptions; perspective; perspectives; cocreated; cocreation; service; services; mobile phone

Introduction

Mental health conditions are an important public health issue globally, responsible for >8 million deaths per year [1-5]. Three million people die annually from the harmful use of alcohol, and 1 person dies every 40 seconds by suicide [1,2]. An estimated 970.1 million people (12.6% of the global population) experience some form of mental health problem [4]. Mental health conditions account for 5% of the global disability-adjusted life years and 12% to 20% of years lived with disability [4,6]. People with mental health conditions, on average, die 20 years prematurely [4,6] both due to mental as well as physical illnesses because mental health conditions are a risk factor for, or can complicate, physical illnesses, including physical injury and road traffic accidents, HIV or AIDS, cardiovascular diseases, and cancer [5,7]. People with mental health conditions also experience severe human rights violations, stigma, discrimination, abuse, and generally poor socioeconomic status [5,7-9].

Unfortunately, >75% of people with mental health problems do not have access to the care they need [1-3]. This is especially true for Uganda [10,11] and similar low- and middle-income countries (LMICs) where the treatment gap for mental disorders reaches 90% [12-14]. It is estimated that the ratio of mental health workers to population is 200 times smaller in LMICs compared with the high-income countries [3]. In LMICs, mental health is underprioritized in the face of other competing public health challenges such as HIV and AIDS, tuberculosis, malaria, and maternal and child health. Uganda, for example, spends 9.8% of its gross domestic product on health care, but <1% of this goes toward mental health care [10,11]. Consequently, Uganda experiences a shortage of mental health care facilities and professionals and poor and inconsistent access to medication and related mental health services [11]. In addition, most of the health workforce is limited to urban areas, yet >80% of the population lives in rural areas, thus geographically isolated from even the limited care available. Other important challenges facing mental health in Uganda include social norms [15], beliefs (such as witchcraft), lack of awareness of mental health disorders [8,11,16], pervasive stigma, and sociopolitical conflicts [13,17]; these not only result in an increase in the incidence of mental health problems but also lead to many people with mental health problems not seeking care and going undiagnosed.

To address some of the abovementioned challenges and improve access to mental health services in Uganda, we implemented the project *digitalizing mental health care and access in Uganda*. In this project, we followed a user-centered design (UCD) and a cocreation process to set up a hospital call center service to provide mental health information and advice to patients, caregivers, and the general public. This paper aimed to describe the development and formative evaluation of this mental health call center service.

Methods

Ethical Considerations

Ethical approval for the research study was obtained from the Makerere University School of Public Health research ethics committee (#SPH-2021-153) and the Uganda National Council of Science and Technology (#HS1868ES). All participants provided written informed consent before participating in the study activities.

Study Design

We conducted a qualitative case study using the principles of UCD [18-20] and design science research (DSR) [21,22]. UCD focuses on understanding and prioritizing the needs, preferences, and behavior of end users of a product throughout its development life cycle. UCD, therefore, calls for iterative and collaborative engagement of users to ensure high usability and utility of the product. DSR is a structured approach to creating and evaluating innovative solutions or artifacts, where the design process is treated as research that contributes to knowledge for improving the functional performance of artifacts. The steps involved in DSR mirror UCD and include the following: (1) identifying the problem and motivation (understanding user experiences and context of use); (2) defining the objective of the solution (specifying the requirements); (3) designing and development of (often novel) solutions using participatory or cocreation processes; and (4) demonstrating and evaluating the solutions to validate against requirements, assess usability, and long-term adoption. These steps help identify the facilitators and barriers of adoption so that they can be addressed early on in the project life cycle, allow user engagement and facilitate buy-in, ensure that the product fits the context of use and purpose, and has good usability [23,24] and clinical utility [25].

In the following sections, we describe each of the above 4 steps. Note that there was overlap and iterations over the steps as per the UCD best practice. To ease readability, we report the procedure and results from each step. Thereafter, we provide a general discussion and conclusion.

Step 1: Understanding User Experiences and Context to Identify the Problems

Participants and Recruitment

The participants included adults (≥18 years), patients recovering from mental disorders, caregivers of such patients, peer support workers (PSWs), mental health care providers, and persons involved in the implementation of call centers for telecoms or other health care centers. The health care providers, patients, caregivers, and PSWs were recruited from the Butabika National Mental Referral Hospital in Kampala, Uganda, which is also the site of implementation. Sampling was purposive to include different cadres and expertise of providers (informed by the third author, who is the head of Butabika Hospital) and to represent different mental health conditions, levels of education, and socioeconomic status of patients and caregivers to get diversity of experiences and views. The investigators (JKK, JN, and VK) physically approached the health care providers at Butabika Hospital, explained the project's purpose and the research activities involved (including participation in multiple group discussions and workshops), and obtained consent from those interested. These health care providers then reached out to patients under their care, caregivers, and PSWs; provided them with information about the study; and invited those interested for consent by the investigators, who explained the participants' rights and voluntary nature of participation. Participants in the last stakeholder category were recruited through the network of the first author who works in the digital health field in Uganda.

Data Collection

We conducted semistructured focus group discussions (FGDs) in which we explored participants' experience of mental illness and mental health care in Uganda (including unmet information and supportive care needs); experience with call center services from the commercial service sector or other digital health care services; and opinions about a proposed digital mental health service (ie, feasibility, appropriateness, expected benefits, or recommendations for successful implementation). The FGD guide is shown in Textbox 1. There was flexibility in the order of the probes to allow free flow of ideas, with additional probes for clarification added by moderators as issues of interest arose. In addition, certain issues or probes were discussed in detail, paraphrased, or left out as appropriate depending on relevance to the session participant or if such a topic had been sufficiently explored in the prior sessions.

The sessions were conducted in English and Luganda (the lingua franca in Uganda) as appropriate for the participant category. In addition, we held male-only and female-only FGD sessions for patients to reduce the possibility that some participants would overshadow others during the discussion, but other sessions were mixed to ensure rich discussions since diverse viewpoints from different participants inspire others and spur discussion. The first and fourth authors (JKK and VK) were the moderators, while the second author (RN) was a notetaker. The sessions were audio recorded and later transcribed by the second author, who also translated the sessions in Luganda into English for analysis. The FGDs took place in November 2021.

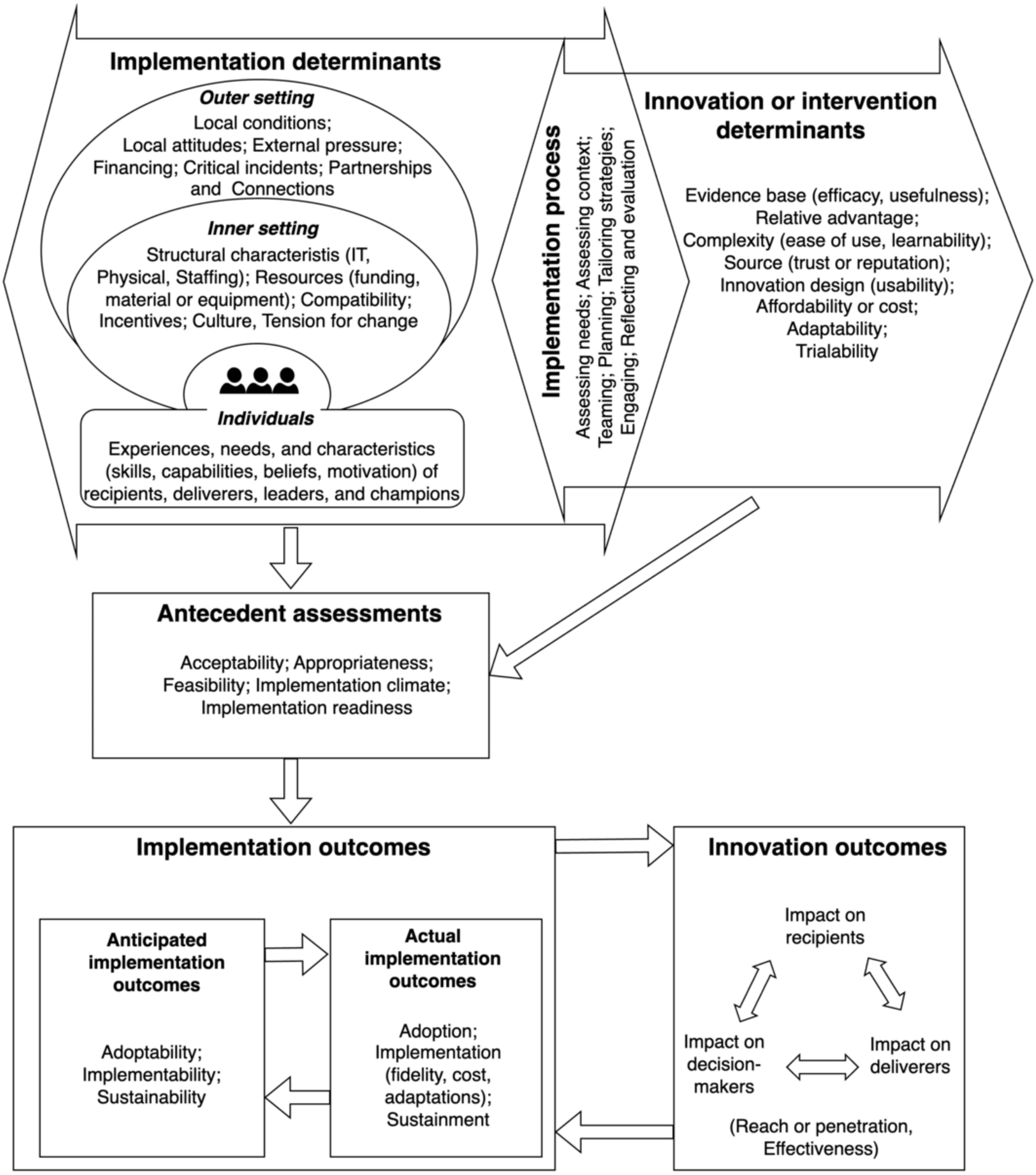
We drew on the Consolidated Framework for Implementation Research (CFIR) [26] to inform data collection (and analysis; see the *Data Analysis* section). The CFIR is a *metatheoretical framework* developed by consolidating several implementation science theories into one comprehensive taxonomy of clearly defined, nonoverlapping constructs related to disseminating and implementing evidence-based interventions. These constructs fall into five domains: (1) the individuals affected or involved in the implementation, (2) the innovation (intervention), (3) the inner setting (organization) where the innovation is implemented, (4) the outer setting (wider societal context), and (5) the process of implementation. The CFIR is one of the most widely used theoretical frameworks to identify implementation barriers and facilitators (ie, a determinant framework) [27-31]. Following a literature review and feedback from researchers who have used the CFIR [29], a recent update, dubbed "CFIR 2.0," has been made, in addition to a CFIR outcomes addendum [32]. These updates have provided further clarification between constructs, including, for example, a distinction between "implementation determinants," which relate to the context, versus "innovation determinants," which relate to the characteristics of the innovation (eg, ease of use, relative advantage, cost, and efficacy of a technology). The implementation and innovation determinants inform the implementation process (needs assessment, user engagement, tailoring to user needs, incentives, marketing, etc) and moderate the anticipated and actual implementation outcomes through the antecedent assessments (tension or readiness for change, feasibility, acceptability, appropriateness, etc). The updates to the CFIR make it also useful for informing the design, implementation, and evaluation of innovations (ie, a process and evaluation framework) [30,31]. Figure 1 shows an adaptation of the CFIR and its recent updates as used in this study [26,29,32].

Textbox 1. Focus group discussion guide.

Topic and questions or probes

- Participants' understanding of mental health and mental problems
  - What is mental health, and what is mental health illness?
  - Do you know any forms of mental illness? What are the signs and symptoms?
  - What do you do with a person who has mental problems? What have you experienced? What is usually done, and what should be the correct thing to do?
  - Where can one get treatment? Probe about alternative healers, witchcraft, religious healers, etc
  - How are mental health or mental illnesses viewed in your community? Probe about stigma, myths, fear, and marginalization
- Mental health information, psychoeducation, and psychosocial support
  - What information about mental health problems or mental health care do you wish you knew early on in your mental illness journey?
  - What issues or topics do you think are the most important to address now? Are there any topics or issues that you still need information about? Give examples.
  - How or where do you get information about mental health and mental illnesses? Which ones are the best or preferred?
  - Tell us any challenges or limitations of these information sources.
  - Are there any services or persons that support you to cope with mental illness or care for your relatives with mental illness? Tell us more about these.
- Telemental health services
  - Tell us about your experience with interactive voice response (IVR) system or call centers: which industry or business? Any challenges and advantages? (moderator to explain IVR if participants do not know and can use the examples of telecoms or bank customer care lines to explain)
  - What are your thoughts on using such IVR systems for mental health information and care (telemental service)? Probes any experience of telemental health services, anticipated benefits, limitations, considerations on how to make it work, concerns about timing, phone ownership and access, privacy, etc. Probe for details and examples.
  - What are the likely barriers or facilitators for such a service?
  - Any thoughts about staffing and the role of peer support workers (PSWs)? Probe about acceptability to patients, benefit to PSWs, any anticipated challenges, and how to mitigate them.
  - Any other thoughts about using technology in mental health care?

Figure 1. The Consolidated Framework for Implementation Research (CFIR) used in this study [26,29,32].



Data Analysis

A directed (deductive) content analysis approach [33] was used. We began with a rapid qualitative analysis [34-36] of the FGDs in order to quickly identify the requirements and other insights needed to inform initial iterations of system development (see *Step 2: Specifying the Requirements of the System* section). Rapid qualitative analysis is aimed at getting actionable and targeted insights in a timely manner and is suitable for studies such as this one, where there is a need to refine and adapt an intervention or program, as opposed to developing new theories. A deductive approach is taken, using existing theories or frameworks (in our case, the CFIR) to summarize the qualitative data into, for example, intervention characteristics or barriers and facilitators. In rapid qualitative research, data collection and analysis occur concurrently and iteratively, with findings from one phase informing the next iteration. The analysis is done not on the transcripts but on the summaries or notes taken during the FGDs or the audio recordings. In addition, multiple data collection methods are used to triangulate findings (eg, FGDs, field observations, debriefing, and reflections by the research team or other stakeholders and literature review). In our study, the rapid analysis was done by the first 4 authors (JKK, RN, JN, and VK) and involved note-taking during FGDs and discussion and summarization of insights after each session. When necessary, recordings were listened to by researchers who, for example, was not present in the session before they contributed to the analysis or for validating the summaries. We summarized the findings into an initial list of mental health information topics to be covered by our system and design considerations (system requirements) based on the experiences and expectations of users, as well as contextual constraints.

Later, a traditional qualitative analysis was done [34-36]. The first and second authors (JKK and RN) independently read 1 of the 7 transcripts and extracted meaningful units or statements and coded them into themes related to the research objectives and the CFIR. They then met to discuss and refine the coding before independently coding the remaining transcripts. Three more meetings were held to compare and refine the coding, after which the findings were shared with all the authors for discussion and interpretation. We focused on saliency [37] rather than frequency of issues and codes, such that even if an issue was mentioned once or by 1 participant category, we coded it as long as it related to the research question and CFIR constructs. As such, we did not count or rank the codes and themes. Basic Office software (Microsoft Corp) was used for coding and summarizing the qualitative data.

Results

Participants and FGD Sessions

We conducted 7 FGDs, each with 8-10 participants, for a total of 65 participants. The participants were fairly balanced by sex (female participants: 35/65, 54%; male participants: 30/65, 46%), and their ages ranged from 21 to 64 years, with a median of 40, IQR 12 years). Each session lasted approximately 1.5 hours. Table 1 shows the details of the FGD sessions.

Table 1. Details of focus group discussion (FGD) sessions (N=65).

FGD session	Stakeholder category	Participants, n (%)	Sex	Language	Notes
1	Patients	10 (15)	Female	Luganda	Sessions in Luganda (the most commonly spoken local language) and separated males and females to ensure participants speak freely and not overshadowed by opposite sex. Diagnoses represented included bipolar affective disorder, schizophrenia, and psychosis.
2	Patients	8 (12)	Male	Luganda	—[a]
3	Caregivers	9 (14)	Mixed	English	Separate sessions in English and in Luganda to get opinions from participants of different education status (English is learned in school in Uganda and is proxy for education and socioeconomic status). Diagnoses represented included bipolar affective disorder, schizophrenia, psychosis, epilepsy, and alcohol and substance use disorder.
4	Caregivers	10 (15)	Mixed	Luganda	—
5	Health care providers	10 (15)	Mixed	English	Staff of Butabika Hospital involved in care for patients and community outreaches, including psychiatrists, psychologists, psychiatric nurses, and psychiatric clinical officers.
6	PSWs[b]	10 (15)	Mixed	English	Volunteers with lived experience of mental illness. They work with Butabika Hospital to share their personal experience and support and educate other patients. They receive small stipends from the hospital, patients they help, or projects and grants to facilitate their work. Diagnoses represented included bipolar affective disorder, schizophrenia, and psychosis.
7	Implementers	8 (12)	Mixed	English	Customer care for telecoms, developers of IVR[c] systems, and implementers of hospital call centers in HIV or AIDS and cancer, private telemedicine company (general care), and mental health NGOs[d].

[a]Not applicable.

[b]PSW: peer support worker.

[c]IVR: interactive voice response.

[d]NGO: nongovernmental organization.

Findings From the FGDs

Multimedia Appendix 1 shows the qualitative findings, including the CFIR domains, constructs, themes, and their explanation. Multimedia Appendix 2 contains illustrative quotes.

Overall, 39 themes emerged across 20 CFIR constructs in all the 5 domains and the antecedent assessments. The themes recurred across the participant groups, supporting their validity. The themes under the *individuals domain* highlighted several challenges that people with mental health conditions in Uganda face, including the limited number of mental health care facilities, long distances to care, lack of mental health information, stigma against patients with mental health problems and their families, financial challenges, and unmet psychosocial needs. The themes also covered contextual issues that explain these challenges. These included issues about the nature of mental illness (chronic and with a high burden of caregiving); organizational issues (*inner setting*), such as understaffing of mental health facilities and frequent medication stockouts; and societal issues (*outer setting*), such as beliefs and cultural norms (eg, belief in witchcraft), which influence how people understand mental health problems and how they seek care. Themes under the domain *innovation determinants* covered participants' perception or expected benefit from the proposed mental health call center service, including affordability; familiarity with similar services and the technology (ubiquitous access to mobile phones); convenience; time and cost saving; and anonymity offered by telephone services, which protect users from the stigma. Finally, themes in the *implementation process* domain encompassed mostly participants' recommendations or strategies for successful implementation, such as linkage with other stakeholders involved in mental health care, marketing of the service (sensitization), training and supervision of staff for quality control, and the need to maintain the human touch rather than attempting to digitalize or automate mental health care delivery. These findings suggested the feasibility, acceptability, and appropriateness of the proposed solution (*antecedent assessments*).

There were also several insights or implicit findings not mentioned by the FGD participants but inferred from observations and the research team's understanding of the context. These are relevant for the implementation and can be mapped to CFIR constructs. For example, there has been an increase in the adoption of telemedicine in Uganda, especially following the COVID-19 pandemic, which has given credibility to such innovations and can explain the general enthusiasm shown by the participants (CFIR construct "evidence base" in *innovation determinants*). In fact, the participants in the implementers' category were themselves involved in implementing call centers for HIV or AIDS, private telemedicine clinics, and mental health NGOs and were aware of the growing scientific evidence globally that supports digital health. The COVID-19 pandemic is also an example of "critical incidents" that can disrupt (or encourage) implementation and delivery of innovations (*outer setting*) according to CFIR 2.0. Other issues included the external project grant (construct: "Financing"), Uganda government's positive digital transformation strategies and policies (construct: "External pressure"), and the position of Butabika Hospital as a national referral that is supposed to be exemplary (construct: "Performance measurement pressure").

Step 2: Specifying the Requirements of the System

Procedure and Team

Requirements were specified based on the understanding of the users' needs, challenges, and contextual constraints from the FGDs. The development team consisted of the first 3 authors (a physician and digital health expert, a research nurse, and a senior consultant psychiatrist, respectively), as well as a psychologist, a psychiatric nurse, and an IT professional specializing in telephone systems. The first 2 authors and the IT professional have previously worked together to set up a similar system at the Uganda Cancer Institute [38] from which they also drew insights. The team held 8 web-based meetings from December 2021 to March 2022 to iteratively discuss the system features, content (mental health information), and setup considerations. We started with the initial list from the rapid qualitative analysis (see the *Data Analysis* section), which we refined to remove conflicting requirements or those that are not feasible due to available resources (eg, video telemedicine). We also agreed on the priority features and mental health information topics.

Results

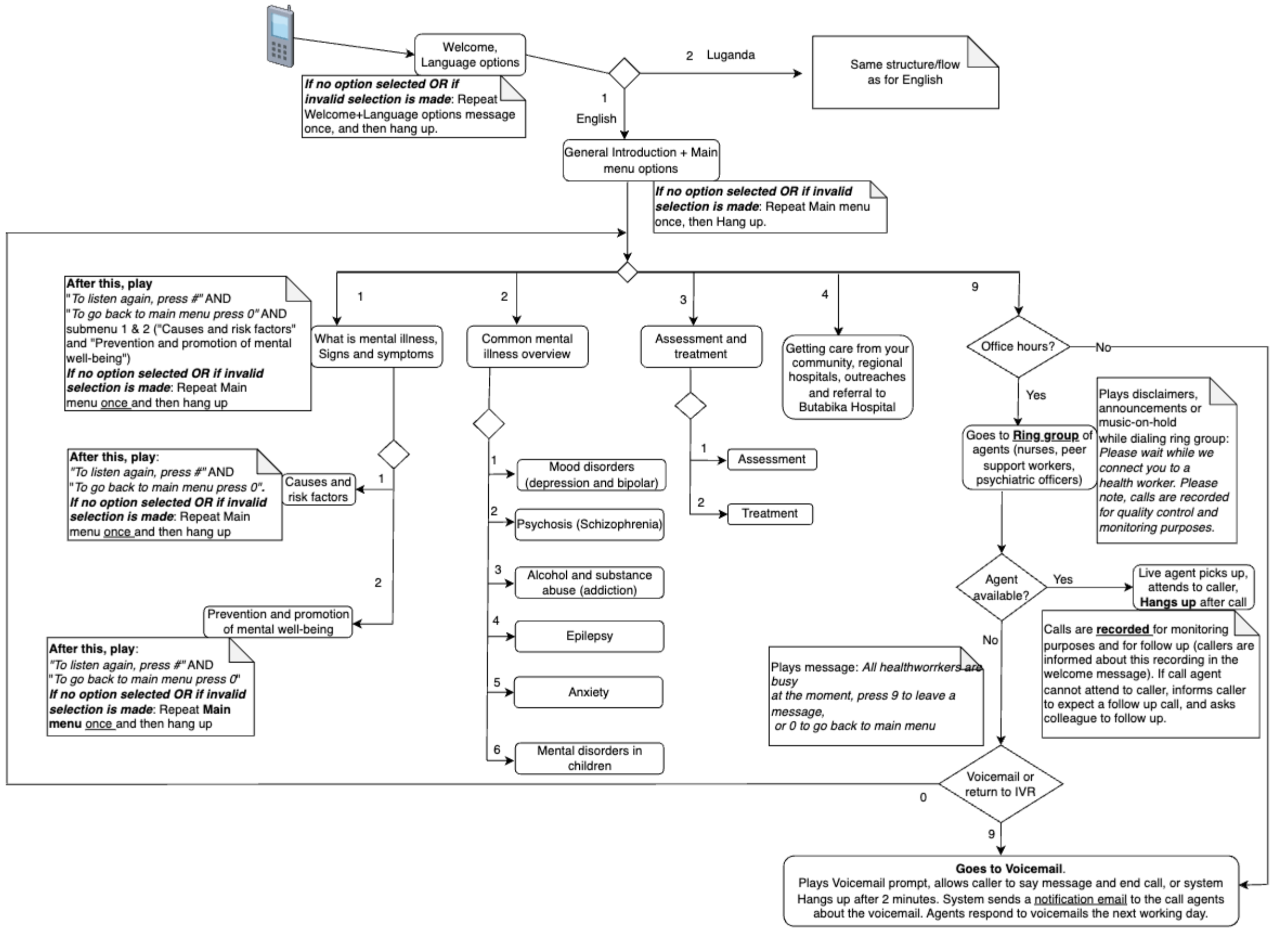
Multimedia Appendix 3 lists the high-level requirements and how they were addressed in the system design and implementation. The key of these requirements is that the system or the intervention provides correct mental health information and psychosocial support in a culturally sensitive and nonstigmatizing manner and in multiple languages. In addition, the system should be easy to use (navigate), accessible 24×7, and affordable (free) to users; there should be no long queues; and it should fit within the workflow of the staff and not increase their workload. Finally, it should ensure privacy and confidentiality to users' information, and risks of harm to users should be minimized through quality control measures, training, and professionalism of staff.

Step 3: Design and Development of the System

Procedure and Team

We designed and developed a telephone system for providing mental health information and advice to callers as per the requirements (Multimedia Appendix 3). The system consists of 3 complimentary components or features: an interactive voice response (IVR), live calls, and voicemails. The IVR is the first component that users interact with, and since it is automated, it is available 24×7. It contains mental health information in audio format in English and Luganda. Callers get navigation instructions and choose from a menu of topics in a self-service manner by pressing the corresponding keys on their phones (eg, *"Thank you for calling Butabika Hospital. Please choose your preferred language. For English, press 1, Bw'oba oyagala kuwuliriza mu Luganda, nyiga 2"*). Figure 2 shows the IVR flow and the topics covered.

Figure 2. Mental health information topics and how they are accessed in the interactive voice response (IVR).



From the IVR, callers can choose to speak directly (live call) with an agent, for example, to seek more clarification on information in the IVR or ask for information that is not covered by the IVR. If it is during office hours, the system connects the caller to the agent. We had a total of 8 agents comprising 2 PSWs, 1 psychologist, 2 psychiatric clinical officers, and 3 psychiatric nurses. The staff do not sit in a physical call center; rather, they are accessible via dedicated mobile phones. All their phones are dialed concurrently ("ring all" strategy), and whoever picks first responds to the caller. Outside working hours, callers are instructed by the system to leave a voicemail, and the call agents returns the calls the next day; this is only possible from a softphone on a computer within the hospital since the caller's number is hidden on the agents' mobile phones for privacy reasons. All agents were encouraged to respond to the calls immediately, and a schedule was created for responding to voicemails. The psychologist provided supervision to the agents and handled any difficult cases, which the agents were encouraged to report or escalate whenever necessary.

Development of the system began by developing the IVR content (mental health messages and navigation instructions), which was done concurrently with the requirements specification process described in *Step 2: Specifying the Requirements of the System* section. The team iteratively wrote the script for mental health messages based on their clinical expertise, reviewed the Luganda translation, and discussed the IVR menu options and caller-system interaction based on the requirements and insights from prior work. We limited the IVR options to a manageable number and organized the information in a logical order, that is, from general information (overview of mental illnesses) to specific information (eg, individual illness such as anxiety or depression). Attention was paid to ease of language (eg, description of concepts or illnesses in addition to naming them and reduction in use of medical jargon); tone (calm, empathic, and nonjudgmental); and cultural appropriateness (eg, acknowledging the role of faith and alternative medicine). The developed content was recorded in a professional audio recording studio and deployed in private branch exchange (PBX) software by the IT professional and the first author.

Results

The telephone system was implemented using Issabel (Issabel LLC), an open-source PBX software based on Asterisk (Sangoma Technologies Corp). It was deployed on a simple server (Intel Core i5 2.6 GHz, 8 GB RAM, 1-TB Hard disk) at Butabika Hospital and connected to a local mobile telecom provider via session initiation protocol with 12 trunks. The calls to the system are reverse billed and therefore are free to the callers.

In sum, we developed a total of 22 messages, 14 (64%) of which were on mental health or other practical information needs elicited from the participants, that is, overview of what mental illnesses are; the causes, signs, and symptoms; the common mental illnesses in Uganda; assessment and management of mental illness; and how to navigate the health care system. The remaining 8 (36%) messages contained navigation instructions or feedback to user (welcome message, language selection, disclaimer, warning in case of emergency, the different menu options, invalid selection, message replay, returning to main menu, and voicemail instructions). The messages were then translated to Luganda for a total of 44 messages. Figure 2 shows the topics addressed by the IVR messages (without some of the navigation messages).

Step 4: Demonstration and Evaluation

Procedure and Participants

Following deployment, we held a 1-day workshop with the PSWs, nurses, and psychiatric clinical officers (n=10) who had participated in the FGDs to test the system, get feedback about the IVR content, and identify and correct any system malfunctions or errors (eg, if there were language mix up or a wrong response for a particular IVR option chosen by the caller).

We held a second workshop to train the call agents on workflows, software system, and phone etiquettes and how to communicate with persons with mental health problems. We also discussed operational issues, for example, definition of office hours when live calls should be allowed, and schedules for returning voicemails and evaluation survey.

The system was advertised via the hospital website and social media channels, posters in the hospital, and personal contacts of the staff and participants. After go-live, we continued to supervise the call agents and held regular review meetings in which we listened to recorded calls and critiqued the conversations, offered support to the call agents (especially the PSW) in case of difficult calls, and collected feedback on usability and user perception of utility of the service.

Results

No major problems were found during the testing workshops, but participants reported that the workshops helped them better understand the service from practically trying it out. They showed enthusiasm for their roles as call agents and became ambassadors who advertised the service to patients and their social networks. Schedules were also drawn for returning calls and office hours defined, which were then programmed into the PBX, sending calls outside these hours to voicemail.

The system went live in August 2021. Detailed results from a survey of the callers and analysis of the use patterns will be reported in a separate study (under preparation), but here we summarize the observations from the first 4 months of operation.

From August to December 2022, a total of 456 calls, from 236 unique numbers (average of 4 calls per day), were made to the system, that is, reaching at least the IVR (automated) component. Of these, 99 (21.7%) calls were made during out-of-office hours for the call agents, so they went to voicemail and were called back within the following days. Of the remaining 357 calls made during office hours, 80 (22.4%) calls stopped at the IVR, while 231 (64.7%) proceeded to speak to a live agent (note that the percentages do not add up to 100% because some callers made multiple calls using the IVR or leaving a voicemail and later called and spoke to a live agent). Furthermore, the 22 (6.2%) calls were never answered by the call agents. On average, live calls were answered within 11 (SD 7) seconds, and their average length was 3.5 (SD 2.8) minutes.

Callers came from all parts of the country (as far as 8 hours by road from Butabika Hospital), although the majority were from the central region (within a 1-hour distance from Butabika Hospital). They included caregivers seeking advice about relatives who were showing symptoms of mental illnesses or those already undergoing care; mental health patients who were relatively stable and were seeking advice about medication or return dates; and others such as clinicians from other health facilities, journalists, and government officials who wanted more information about the call center system or the mental health care services offered at Butabika Hospital. Calls about patients who had “escaped” from the hospital were also common, often made by concerned community members near the hospital who come across a person with mental illness wandering in the community. Generally, the service has been received positively. Callers were especially happy with the PSWs who shared their personal journeys with mental illnesses and recovery, and this encouraged them to overcome the stigma and negativity that they had about mental health care services. The PSWs also reported positive experiences, stating that working as call agents and helping others gave them a sense of purpose and brought order and calmness.

A key challenge was callers who required specific and individualized information that the call agents did not have at hand and could not be prerecorded in the IVR. Such information included requests for prescriptions, questions on stocks of certain medications, availability and cost of certain tests and procedures, or about the condition of a relative who was admitted in the hospital.

Reflexivity

The members of the research team who were involved in data collection and analysis (FGDs, workshops, and analysis meetings) are intimately familiar with the local context and understand participants’ realities (including participants’ access and use of mobile phones and the internet) since they come from the same region of the country and speak the local language (Luganda). This made it easier to communicate with the participants (even for sessions that were held in Luganda) and to understand and relate to the ideas or issues they raised. To reduce potential undue coercion, the clinicians involved in the care of the participants (patients and PSWs) did not participate in the FGDs sessions but participated in data analysis and interpretation. These clinicians were especially important in ensuring that the rest of the research team members were aware of assumptions and potential prejudices, for example, with regard to beliefs in witchcraft as a cause for mental illnesses or in faith healing, common among those with low education status. Clinicians working in mental health care in this context frequently encounter such beliefs and appreciate the importance of respecting them, which was also useful for informing how we crafted the mental health messages in the system. Moreover, 3 of the research team members were from a different high-income country and brought in different perspectives, which helped us question our interpretations and assumptions.

Discussion

Principal Findings

This paper describes the development and implementation of a digital health intervention aimed at improving mental health care in Uganda. Using principles of UCD [18-20] and DSR [21,22], we systematically engaged stakeholders, collected data on target users' experiences of mental health care, their opinions and recommendations about the proposed mental health telephone service, and contextual issues that could influence implementation. We used the CFIR, an established implementation science meta-framework [26-31], to collect and analyze these data and derive system requirements and then iteratively cocreated and tested the system.

We identified several challenges faced by patients with mental health problems and their caregivers in Uganda and peculiarities about the organization and the wider societal context, which supported the proposed innovation. These challenges included the severe shortage of mental health workers and services, lack of awareness, negative beliefs and norms, stigma, huge burden of caregiving, and financial challenges. At the same time, there is a general trend toward digitalization of health care to improve patient experience and efficiency of health care, and participants were enthusiastic about our proposed call center because they were familiar with the technology and considered it as a feasible, affordable, convenient, and efficient way to get mental health services without being stigmatized. The participants also gave several recommendations on how to successfully implement the intervention, for example, by making calls toll free, ensuring 24×7 availability, providing mental health information in multiple languages, using technologies or channels that are appropriate to the context (telephone calls and IVR), sufficient staffing to reducing call waiting times, sensitizing people about the service, and training and supervision of the call agents to ensure quality service. Early evaluation of the intervention shows that clients are very positive about the service, particularly with the use of PSWs (recovering patients) who share their lived experience with others.

Comparison With Prior Work

Prior research has demonstrated the value of mobile health (mHealth) in addressing some of the health care challenges in Uganda and similar contexts elsewhere. Systematic reviews on mHealth in general [39-45] or on specific clinical domains such as HIV or AIDS [45-47] and palliative care [48] have highlighted the improvement of health care coordination and communication between patients and health care providers, patient adherence to treatment and reduction of loss to follow-up, patient engagement and self-care, facilitation of community-based care, and improvement of access to care for rural or geographically isolated populations. Advantages such as ubiquity of mobile technology, affordability and acceptability by patients and health workers, interactivity and personalization, and saving of time and cost of traveling to health facilities have been cited. Examples of prior studies on mHealth in Uganda include use of IVR, SMS text messages, and phone calls to support the management of HIV or AIDS [49,50] and tuberculosis [51]; use of IVR to address barriers to fistula care in Uganda [52]; SMS text messages for stroke rehabilitation [53]; and IVR for provision of cancer awareness and advice [38]. There is also a commercial digital health company that has operated different mHealth services in Uganda for approximately 10 years [54]. Unfortunately, the use of mHealth in mental health in Uganda and Africa in general is limited [8,55,56]. This is likely due to the general underfunding of mental health care services [10-14]. Available research on mHealth in mental health is mostly from developed countries [57-61], with many interventions using the internet and smartphone apps, which might not be accessible or affordable in Uganda or other LMICs. Interventions that use basic phone features such as SMS text messages, IVR, and voice calls are more appropriate in LMICs as they overcome infrastructural limitations. Such interventions are also relevant for low-income and migrant communities in developed countries since these populations face low digital health literacy and language barriers [62-65].

In the previous project led by the first author for the provision of cancer information [38], similar findings in terms of challenges faced by patients, requirements and recommendations for the system, and generally positive reception after implementation were reported. The cancer awareness system mainly used the IVR feature with prerecorded information, with the option to speak to a live agent added as an emergency due to the COVID-19 pandemic. The agents were health care workers (nurses and physicians) who, due to travel restrictions, had been free to handle phone calls. While callers appreciated this feature, it is otherwise not possible given the limited number of health workers. In this study, PSWs helped to address the shortage of health care professionals. A large multinational research study from Uganda and elsewhere has demonstrated the positive benefits of using PSWs, both for their own recovery and for the health care system [66-68]. Our study builds onto this prior work to innovatively and efficiently put this underused resource to use through digital health.

Strengths and Limitations of the Study

A strength of this study is the strong theoretical underpinning. Implementation studies have been faulted in the past for not being theory driven, which undermines the adoption of digital technologies [30]. The UCD and DSR approach used informed a systematic cocreation process of intervention development with user participation, while the CFIR allowed a comprehensive review of user, technological, and contextual issues to inform system requirements. Even so, we could not consider all the requirements or recommendations by the participants when designing the system because of resource limitations or contradictions. For example, some participants recommended adding video calling features to the system to enhance interaction and assessment of affect. Other participants had concerns about continuity of care, which indeed is difficult to achieve with the current call center system that lacks electronic medical records or mechanisms to ensure that callers are directed to agents with whom they have interacted with before. However, adding such features would make the system complex, expensive, and inaccessible to some such as those who mentioned inability to work with smartphones or had connectivity problems. Still, the insights from this comprehensive assessment can inform future incremental iterations of the system during scale-up.

Conclusions

Participants were enthusiastic about the proposed call center because they were familiar with the technology and considered it as a feasible, affordable, convenient, and efficient way to get mental health services without being stigmatized. The system provided mental health information and linkage to health care providers and PSWs. The information in audio format made it accessible even to the people with low literacy, and the automated IVR allowed 24×7 access while reducing the pressure on the health care workforce. Translation to English and Luganda, the 2 most spoken languages in Uganda, increased reach, as did the reverse billing (no cost to the caller) and the use of basic telephone calls as the channel of access since many Ugandans still do not have affordable and reliable internet access.

Recommendations

In this study, people with mental illness, caregivers, and health care providers deemed a telephone-based mental health care service useful and necessary to increase access to mental health information and care and reduce stigma toward people with mental health problems. This positive view needs to be harnessed to scale up the digitalization of mental health care including providing therapy and establishing it in other mental health care settings in line with the current Ugandan digitalization policy and the Third National Development Plan. This method of mental health care may be replicable and scalable in other LMICs with mental health care system and personnel challenges similar to Uganda. Further research is needed to evaluate long-term adoption, patterns of use, and impact on clinical outcomes.

Acknowledgments

This project was funded by Digital Futures, a research collaboration between KTH Royal Institute of Technology, Stockholm University, and RISE Research Institutes of Sweden. Funders had no influence in the conduct of the research study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Summary of qualitative findings coded according to the Consolidated Framework for Implementation Research.

[DOCX File , 17 KB - humanfactors_v11i1e53976_app1.docx]

Multimedia Appendix 2

Consolidated Framework for Implementation Research domains, constructs, themes from focus group discussions, and illustrative quotes.

[XLSX File (Microsoft Excel File), 27 KB - humanfactors_v11i1e53976_app2.xlsx]

Multimedia Appendix 3

Requirements and how they are addressed by the system and its setup.

[DOCX File , 18 KB - humanfactors_v11i1e53976_app3.docx]

References

1. Ingrid D, Brunier A, Sharma H. World mental health day: an opportunity to kick-start a massive scale-up in investment in mental health. World Health Organization. 2020. URL: https://tinyurl.com/2z6bramw [accessed 2024-04-29]
2. Kovacevic R. Mental health: lessons learned in 2020 for 2021 and forward. World Bank Blog. 2021. URL: https://blogs.worldbank.org/health/mental-health-lessons-learned-2020-2021-and-forward [accessed 2024-04-29]
3. Naslund JA, Aschbrenner KA, Araya R, Marsch LA, Unützer J, Patel V, et al. Digital technology for treating and preventing mental disorders in low-income and middle-income countries: a narrative review of the literature. Lancet Psychiatry 2017 Jun;4(6):486-500. [doi: 10.1016/s2215-0366(17)30096-2]
4. GBD 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Psychiatry 2022 Feb;9(2):137-150. [doi: 10.1016/s2215-0366(21)00395-3]
5. The WHO special initiative for mental health (2019-2023): universal health coverage for mental health. World Health Organization. 2019. URL: https://apps.who.int/iris/handle/10665/310981 [accessed 2024-04-29]
6. Mental health. World Health Organization. URL: https://www.who.int/health-topics/mental-health#tab=tab_1 [accessed 2024-04-29]

7. Mental health and the 2030 sustainable development agenda: global inaction on mental health is putting the brakes on development. Mental Health Innovation Network. 2018. URL: https://www.mhinnovation.net/sites/default/files/content/document/Policy%20brief%202030%20SDG.pdf [accessed 2024-04-29]
8. Ben-Zeev D. Mobile health for mental health in west Africa: the case for Ghana. Psychiatr Serv 2018 Jul 01;69(7):741-743. [doi: 10.1176/appi.ps.201700555] [Medline: 29540120]
9. Arias D, Saxena S, Verguet S. Quantifying the global burden of mental disorders and their economic value. EClinicalMedicine 2022 Dec;54:101675 [FREE Full text] [doi: 10.1016/j.eclinm.2022.101675] [Medline: 36193171]
10. Kagaari J. Mental health in Uganda. American Psychological Association. 2021. URL: https://www.apa.org/international/global-insights/uganda-mental-health [accessed 2024-04-29]
11. Molodynski A, Cusack C, Nixon J. Mental healthcare in Uganda: desperate challenges but real opportunities. BJPsych Int 2017 Nov 02;14(4):98-100 [FREE Full text] [doi: 10.1192/s2056474000002129] [Medline: 29093962]
12. Alloh FT, Regmi P, Onche I, van Teijlingen E, Trenoweth S. Mental Health in low-and middle income countries (LMICs): going beyond the need for funding. Health Prospect 2018 Jun 19;17(1):12-17. [doi: 10.3126/hprospect.v17i1.20351]
13. Gberie L. Mental illness: invisible but devastating. Africa Renewal. URL: https://tinyurl.com/mwhvxjzu [accessed 2024-04-29]
14. Patel V, Maj M, Flisher AJ, De Silva MJ, Koschorke M, Prince M, WPA ZonalMember Society Representatives. Reducing the treatment gap for mental disorders: a WPA survey. World Psychiatry 2010 Oct;9(3):169-176 [FREE Full text] [doi: 10.1002/j.2051-5545.2010.tb00305.x] [Medline: 20975864]
15. Ezeugwu CR, Ojedokun O. Masculine norms and mental health of African men: what can psychology do? Heliyon 2020 Dec;6(12):e05650. [doi: 10.1016/j.heliyon.2020.e05650]
16. Esan O, Abdumalik J, Eaton J, Kola L, Fadahunsi W, Gureje O. Mental health care in Anglophone West Africa. Psychiatr Serv 2014 Sep 01;65(9):1084-1087. [doi: 10.1176/appi.ps.201300300] [Medline: 25179185]
17. Musisi S, Kinyanda E. Long-term impact of war, civil war, and persecution in civilian populations-conflict and post-traumatic stress in African communities. Front Psychiatry 2020 Feb 25;11:20 [FREE Full text] [doi: 10.3389/fpsyt.2020.00020] [Medline: 32158407]
18. Abras C, Maloney-Krichmar D, Preece J. User-centered design. In: Bainbridge W, editor. Encyclopedia of Human-Computer Interaction. Thousand Oaks, CA: SAGE Publications; 2004:445-456.
19. De Vito Dabbs A, Myers BA, Mc Curry KR, Dunbar-Jacob J, Hawkins RP, Begey A, et al. User-centered design and interactive health technologies for patients. Comput Inform Nurs 2009;27(3):175-183 [FREE Full text] [doi: 10.1097/NCN.0b013e31819f7c7c] [Medline: 19411947]
20. Mao JY, Vredenburg K, Smith PW, Carey T. The state of user-centered design practice. Commun ACM 2005 Mar;48(3):105-109. [doi: 10.1145/1047671.1047677]
21. Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S. A design science research methodology for information systems research. J Manag Inf Syst 2014 Dec 08;24(3):45-77. [doi: 10.2753/mis0742-1222240302]
22. Iivari J, Venable JR. Action research and design science research - seemingly similar but decisively dissimilar. In: Proceedings of the 17th European Conference on Information Systems. 2009 Presented at: ECIS '09; June 8-10, 2009; Verona, Italy URL: https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1025&context=ecis2009
23. Brooke J. SUS: a 'quick and dirty' usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation In Industry. Boca Raton, FL: CRC Press; 1996:207-212.
24. Brooke J. SUS: a retrospective. J Usability Stud 2013:29-40 [FREE Full text]
25. Smart A. A multi-dimensional model of clinical utility. Int J Qual Health Care 2006 Oct;18(5):377-382. [doi: 10.1093/intqhc/mzl034] [Medline: 16951425]
26. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4(1):50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
27. Kirk MA, Kelley C, Yankey N, Birken SA, Abadie B, Damschroder L. A systematic review of the use of the consolidated framework for implementation research. Implement Sci 2016 May 17;11(1):72 [FREE Full text] [doi: 10.1186/s13012-016-0437-z] [Medline: 27189233]
28. Skolarus TA, Lehmann T, Tabak RG, Harris J, Lecy J, Sales AE. Assessing citation networks for dissemination and implementation research frameworks. Implement Sci 2017 Jul 28;12(1):97 [FREE Full text] [doi: 10.1186/s13012-017-0628-2] [Medline: 28754140]
29. Damschroder LJ, Reardon CM, Widerquist MA, Lowery J. The updated consolidated framework for implementation research based on user feedback. Implement Sci 2022 Oct 29;17(1):75 [FREE Full text] [doi: 10.1186/s13012-022-01245-0] [Medline: 36309746]
30. Nilsen P. Making sense of implementation theories, models and frameworks. Implement Sci 2015 Apr 21;10(1):53 [FREE Full text] [doi: 10.1186/s13012-015-0242-0] [Medline: 25895742]
31. Damschroder LJ. Clarity out of chaos: use of theory in implementation research. Psychiatry Res 2020 Jan;283:112461 [FREE Full text] [doi: 10.1016/j.psychres.2019.06.036] [Medline: 31257020]

32. Damschroder LJ, Reardon CM, Opra Widerquist MA, Lowery J. Conceptualizing outcomes for use with the Consolidated Framework for Implementation Research (CFIR): the CFIR outcomes addendum. Implement Sci 2022 Jan 22;17(1):7 [FREE Full text] [doi: 10.1186/s13012-021-01181-5] [Medline: 35065675]
33. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov 01;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
34. Suchman L, Omoluabi E, Kramer J, Vallin J, Sedlander E, Gitome S, ICAN Research Consortium. Analyzing fast and slow: combining traditional and rapid qualitative analysis to meet multiple objectives of a complex transnational study. Front Sociol 2023 Feb 1;8:961202 [FREE Full text] [doi: 10.3389/fsoc.2023.961202] [Medline: 36818663]
35. Lewinski AA, Crowley MJ, Miller C, Bosworth HB, Jackson GL, Steinhauser K, et al. Applied rapid qualitative analysis to develop a contextually appropriate intervention and increase the likelihood of uptake. Med Care 2021 Jun 01;59(Suppl 3):S242-S251 [FREE Full text] [doi: 10.1097/MLR.0000000000001553] [Medline: 33976073]
36. Gale RC, Wu J, Erhardt T, Bounthavong M, Reardon CM, Damschroder LJ, et al. Comparison of rapid vs in-depth qualitative analytic methods from a process evaluation of academic detailing in the Veterans Health Administration. Implement Sci 2019 Feb 01;14(1):11 [FREE Full text] [doi: 10.1186/s13012-019-0853-y] [Medline: 30709368]
37. Buetow S. Thematic analysis and its reconceptualization as 'saliency analysis'. J Health Serv Res Policy 2010 Apr 01;15(2):123-125. [doi: 10.1258/jhsrp.2009.009081] [Medline: 19762883]
38. Kabukye JK, Ilozumba O, Broerse JE, de Keizer N, Cornet R. Implementation of an interactive voice response system for cancer awareness in Uganda: mixed methods study. JMIR Mhealth Uhealth 2021 Jan 26;9(1):e22061 [FREE Full text] [doi: 10.2196/22061] [Medline: 33496672]
39. McCool J, Dobson R, Whittaker R, Paton C. Mobile health (mHealth) in low- and middle-income countries. Annu Rev Public Health 2022 Apr 05;43(1):525-539 [FREE Full text] [doi: 10.1146/annurev-publhealth-052620-093850] [Medline: 34648368]
40. Hurt K, Walker RJ, Campbell JA, Egede LE. mHealth interventions in low and middle-income countries: a systematic review. Glob J Health Sci 2016 Sep 01;8(9):54429 [FREE Full text] [doi: 10.5539/gjhs.v8n9p183] [Medline: 27157176]
41. Chib A, van Velthoven MH, Car J. mHealth adoption in low-resource environments: a review of the use of mobile healthcare in developing countries. J Health Commun 2015 Mar 27;20(1):4-34. [doi: 10.1080/10810730.2013.864735] [Medline: 24673171]
42. Aranda-Jan CB, Mohutsiwa-Dibe N, Loukanova S. Systematic review on what works, what does not work and why of implementation of mobile health (mHealth) projects in Africa. BMC Public Health 2014 Feb 21;14(1):188 [FREE Full text] [doi: 10.1186/1471-2458-14-188] [Medline: 24555733]
43. Hall CS, Fottrell E, Wilkinson S, Byass P. Assessing the impact of mHealth interventions in low- and middle-income countries--what has been shown to work? Glob Health Action 2014;7:25606 [FREE Full text] [doi: 10.3402/gha.v7.25606] [Medline: 25361730]
44. Abreu FD, Bissaco MA, Silva AP, Boschi SR, Scardovelli TA, Santos MF, et al. The use and impact of mHealth by community health workers in developing and least developed countries: a systematic review. Res Biomed Eng 2021 May 20;37(3):563-582 [FREE Full text] [doi: 10.1007/S42600-021-00154-3]
45. Cooper V, Clatworthy J, Whetham J, Consortium E. mHealth interventions to support self-management in HIV: a systematic review. Open AIDS J 2017 Nov 21;11(1):119-132 [FREE Full text] [doi: 10.2174/1874613601711010119] [Medline: 29290888]
46. Devi BR, Syed-Abdul S, Kumar A, Iqbal U, Nguyen PA, Li YC, et al. mHealth: an updated systematic review with a focus on HIV/AIDS and tuberculosis long term management using mobile phones. Comput Methods Programs Biomed 2015 Nov;122(2):257-265 [FREE Full text] [doi: 10.1016/j.cmpb.2015.08.003] [Medline: 26304621]
47. Catalani C, Philbrick W, Fraser H, Mechael P, Israelski DM. mHealth for HIV treatment and prevention: a systematic review of the literature. Open AIDS J 2013 Sep 20;7(1):17-41 [FREE Full text] [doi: 10.2174/1874613620130812003] [Medline: 24133558]
48. Allsop MJ, Powell RA, Namisango E. The state of mHealth development and use by palliative care services in sub-Saharan Africa: a systematic review of the literature. BMJ Support Palliat Care 2018 Jun 20;8(2):155-163 [FREE Full text] [doi: 10.1136/bmjspcare-2015-001034] [Medline: 27207721]
49. Byonanebye DM, Nabaggala MS, Naggirinya AB, Lamorde M, Oseku E, King R, et al. An interactive voice response software to improve the quality of life of people living with HIV in Uganda: randomized controlled trial. JMIR Mhealth Uhealth 2021 Feb 11;9(2):e22229 [FREE Full text] [doi: 10.2196/22229] [Medline: 33570497]
50. Twimukye A, Bwanika Naggirinya A, Parkes-Ratanshi R, Kasirye R, Kiragga A, Castelnuovo B, et al. Acceptability of a mobile phone support tool (call for life Uganda) for promoting adherence to antiretroviral therapy among young adults in a randomized controlled trial: exploratory qualitative study. JMIR Mhealth Uhealth 2021 Jun 14;9(6):e17418 [FREE Full text] [doi: 10.2196/17418] [Medline: 34121665]
51. Byonanebye DM, Mackline H, Sekaggya-Wiltshire C, Kiragga AN, Lamorde M, Oseku E, et al. Impact of a mobile phone-based interactive voice response software on tuberculosis treatment outcomes in Uganda (CFL-TB): a protocol for a randomized controlled trial. Trials 2021 Jun 13;22(1):391 [FREE Full text] [doi: 10.1186/s13063-021-05352-z] [Medline: 34120649]

52. Tripathi V, Arnoff E, Bellows B, Sripad P. Use of interactive voice response technology to address barriers to fistula care in Nigeria and Uganda. Mhealth 2020 Apr;6:12 [FREE Full text] [doi: 10.21037/mhealth.2019.12.04] [Medline: 32270004]
53. Teriö M, Eriksson G, Kamwesiga JT, Guidetti S. What's in it for me? A process evaluation of the implementation of a mobile phone-supported intervention after stroke in Uganda. BMC Public Health 2019 May 14;19(1):562. [doi: 10.1186/s12889-019-6849-3] [Medline: 31088411]
54. Kamulegey L, Ssebwana J, Abigaba W, Bwanika J, Musinguzi D. Mobile health in Uganda: a case study of the medical concierge group. East Afr Sci 2019 Mar 25;1(1):9-14. [doi: 10.24248/easci.v1i1.12]
55. Pearson AL, Mack E, Namanya J. Mobile phones and mental well-being: initial evidence suggesting the importance of staying connected to family in rural, remote communities in Uganda. PLoS One 2017 Jan 17;12(1):e0169819 [FREE Full text] [doi: 10.1371/journal.pone.0169819] [Medline: 28095427]
56. Adepoju P. Africa turns to telemedicine to close mental health gap. Lancet Digit Health 2020 Nov;2(11):e571-e572. [doi: 10.1016/s2589-7500(20)30252-1]
57. Lattie EG, Stiles-Shields C, Graham AK. An overview of and recommendations for more accessible digital mental health services. Nat Rev Psychol 2022 Feb 26;1(2):87-100 [FREE Full text] [doi: 10.1038/s44159-021-00003-1] [Medline: 38515434]
58. Price M, Yuen EK, Goetter EM, Herbert JD, Forman EM, Acierno R, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. Clin Psychol Psychother 2014 Aug 05;21(5):427-436 [FREE Full text] [doi: 10.1002/cpp.1855] [Medline: 23918764]
59. Wang K, Varma DS, Prosperi M. A systematic review of the effectiveness of mobile apps for monitoring and management of mental health symptoms or disorders. J Psychiatr Res 2018 Dec;107:73-78 [FREE Full text] [doi: 10.1016/j.jpsychires.2018.10.006] [Medline: 30347316]
60. Hilty DM, Ferrer DC, Parish MB, Johnston B, Callahan EJ, Yellowlees PM. The effectiveness of telemental health: a 2013 review. Telemed J E Health 2013 Jun;19(6):444-454 [FREE Full text] [doi: 10.1089/tmj.2013.0075] [Medline: 23697504]
61. Hwang WJ, Ha JS, Kim MJ. Research trends on mobile mental health application for general population: a scoping review. Int J Environ Res Public Health 2021 Mar 02;18(5):2459 [FREE Full text] [doi: 10.3390/ijerph18052459] [Medline: 33801537]
62. Woodward A, Burchert S, Barry AS, Broerse JE, Sondorp E, Bold A, et al. Scalability of digital psychological innovations for refugees: a comparative analysis in Egypt, Germany, and Sweden. SSM Mental Health 2023 Dec;4:100231. [doi: 10.1016/j.ssmmh.2023.100231]
63. Bergman L, Nilsson U, Dahlberg K, Jaensson M, Wångdahl J. Health literacy and e-health literacy among Arabic-speaking migrants in Sweden: a cross-sectional study. BMC Public Health 2021 Nov 25;21(1):2165 [FREE Full text] [doi: 10.1186/s12889-021-12187-5] [Medline: 34823499]
64. Reuland CJ, Godage SK, Wu L, Valenzuela-Araujo D, Cortez JD, Polk S, et al. Information and communication technology access and use among low-income Latino immigrant parents. Matern Child Health J 2021 Dec 23;25(12):1807-1813. [doi: 10.1007/s10995-021-03265-6] [Medline: 34687401]
65. Bender MS, Choi J, Arai S, Paul SM, Gonzalez P, Fukuoka Y. Digital technology ownership, usage, and factors predicting downloading health apps among Caucasian, Filipino, Korean, and Latino Americans: the digital link to health survey. JMIR Mhealth Uhealth 2014 Oct 22;2(4):e43 [FREE Full text] [doi: 10.2196/mhealth.3710] [Medline: 25339246]
66. Puschner B, Repper J, Mahlke C, Nixdorf R, Basangwa D, Nakku J, et al. Using peer support in developing empowering mental health services (UPSIDES): background, rationale and methodology. Ann Glob Health 2019 Apr 05;85(1):53 [FREE Full text] [doi: 10.5334/aogh.2435] [Medline: 30951270]
67. Moran GS, Kalha J, Mueller-Stierlin AS, Kilian R, Krumm S, Slade M, et al. Peer support for people with severe mental illness versus usual care in high-, middle- and low-income countries: study protocol for a pragmatic, multicentre, randomised controlled trial (UPSIDES-RCT). Trials 2020 May 01;21(1):371 [FREE Full text] [doi: 10.1186/s13063-020-4177-7] [Medline: 32357903]
68. Nixdorf R, Nugent L, Aslam R, Barber S, Charles A, Gai Meir L, et al. A multi-national peer support intervention: the UPSIDES pilot phase. Adv Ment Health 2022 Jan 06;20(1):2-14 [FREE Full text] [doi: 10.1080/18387357.2021.2020140]

Abbreviations

CFIR: Consolidated Framework for Implementation Research
DSR: design science research
FGD: focus group discussion
IVR: interactive voice response
LMIC: low- and middle-income country
mHealth: mobile health
PBX: private branch exchange
PSW: peer support worker
UCD: user-centered design

Edited by A Choudhury; submitted 25.10.23; peer-reviewed by A Gangadhara Rao, I Wilson; comments to author 23.02.24; revised version received 01.04.24; accepted 01.05.24; published 06.06.24.

Please cite as:
Kabukye JK, Namagembe R, Nakku J, Kiberu V, Sjölinder M, Nilsson S, Wamala-Larsson C
Implementing a Hospital Call Center Service for Mental Health in Uganda: User-Centered Design Approach
JMIR Hum Factors 2024;11:e53976
URL: https://humanfactors.jmir.org/2024/1/e53976
doi:10.2196/53976
PMID:38843515

Original Paper

Evaluation of a Computer-Aided Clinical Decision Support System for Point-of-Care Use in Low-Resource Primary Care Settings: Acceptability Evaluation Study

Geletaw Sahle Tegenaw[1,2], PhD; Demisew Amenu Sori[3], MD; Girum Ketema Teklemariam[2], PhD; Frank Verbeke[1], MD, PhD; Jan Cornelis[1], PhD, Prof Dr; Bart Jansen[1,4], PhD, Prof Dr

[1]Department of Electronics and Informatics, Vrije Universiteit Brussel, Brussel, Belgium

[2]Faculty of Computing, Jimma Institute of Technology, Jimma University, Jimma, Ethiopia

[3]Department of Obstetrics and Gynecology, College of Health Science, Jimma University, Jimma, Ethiopia

[4]Interuniversitair Micro-Electronica Centrum, Leuven, Belgium

Corresponding Author:
Geletaw Sahle Tegenaw, PhD
Department of Electronics and Informatics
Vrije Universiteit Brussel
Pleinlaan 2
Brussel, 1050
Belgium
Phone: 32 485744961
Email: gtegenaw@vub.be

Abstract

Background: A clinical decision support system (CDSS) based on the logic and philosophy of clinical pathways is critical for managing the quality of health care and for standardizing care processes. Using such a system at a point-of-care setting is becoming more frequent these days. However, in a low-resource setting (LRS), such systems are frequently overlooked.

Objective: The purpose of the study was to evaluate the user acceptance of a CDSS in LRSs.

Methods: The CDSS evaluation was carried out at the Jimma Health Center and the Jimma Higher Two Health Center, Jimma, Ethiopia. The evaluation was based on 22 parameters organized into 6 categories: ease of use, system quality, information quality, decision changes, process changes, and user acceptance. A Mann-Whitney U test was used to investigate whether the difference between the 2 health centers was significant (2-tailed, 95% CI; α=.05). Pearson correlation and partial least squares structural equation modeling (PLS-SEM) was used to identify the relationship and factors influencing the overall acceptance of the CDSS in an LRS.

Results: On the basis of 116 antenatal care, pregnant patient care, and postnatal care cases, 73 CDSS evaluation responses were recorded. We found that the 2 health centers did not differ significantly on 16 evaluation parameters. We did, however, detect a statistically significant difference in 6 parameters (P<.05). PLS-SEM results showed that the coefficient of determination, R^2, of perceived user acceptance was 0.703. More precisely, the perceived ease of use (β=.015, P=.91) and information quality (β=.149, P=.25) had no positive effect on CDSS acceptance but, rather, on the system quality and perceived benefits of the CDSS, with P<.05 and β=.321 and β=.486, respectively. Furthermore, the perceived ease of use was influenced by information quality and system quality, with an R^2 value of 0.479, indicating that the influence of information quality on the ease of use is significant but the influence of system quality on the ease of use is not, with β=.678 (P<.05) and β=.021(P=.89), respectively. Moreover, the influence of decision changes (β=.374, P<.05) and process changes (β=.749, P<.05) both was significant on perceived benefits (R^2=0.983).

Conclusions: This study concludes that users are more likely to accept and use a CDSS at the point of care when it is easy to grasp the perceived benefits and system quality in terms of health care professionals' needs. We believe that the CDSS acceptance model developed in this study reveals specific factors and variables that constitute a step toward the effective adoption and deployment of a CDSS in LRSs.

(JMIR Hum Factors 2024;11:e47631) doi:10.2196/47631

KEYWORDS

low-resource setting; clinical decision support system; point-of-care instrument; evaluation; user acceptance; structural equation modeling; partial least squares structural equation modeling; decision-making; decision making; decision support; caregiver; users; acceptance; system quality

Introduction

The use of health information systems has considerably transformed the health care sector in recent years [1]. Proper and coordinated implementation is beneficial to the enhancement of health care delivery [2,3]. An effective clinical decision support system (CDSS); low-cost, point-of-care diagnostics; effective remote clinics; home-based therapies; and improved communication with patients and across health care facilities are among the benefits [4,5]. Even though the implementation of a CDSS at the point of care has sought to improve treatment quality and resource efficiency, its use in low-resource settings (LRSs) has lagged behind due to a variety of restrictions.

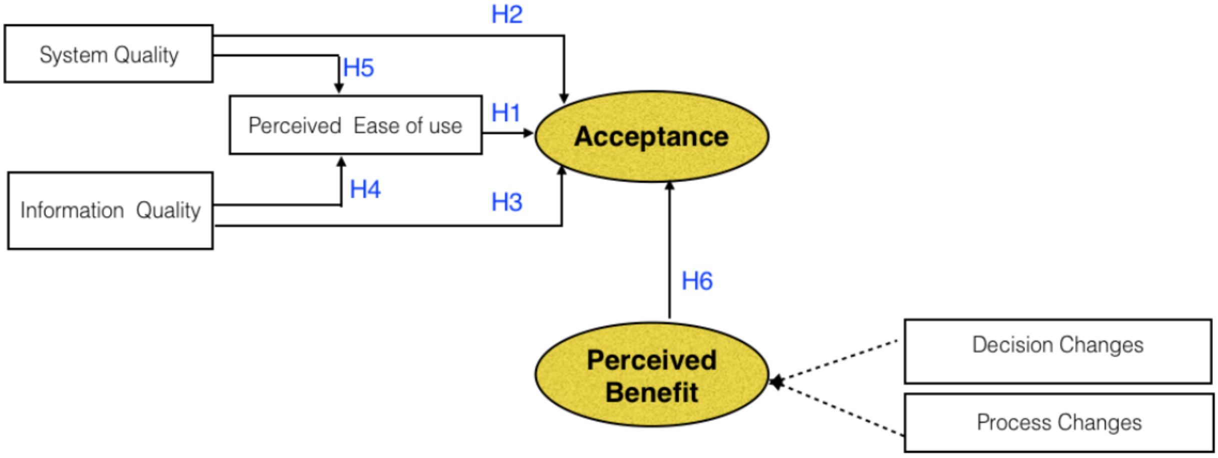
In Ethiopia, the health care system is a 3-tiered system organized into primary, secondary, and tertiary levels of care [6]. Primary health care settings include primary hospitals, health centers, and health posts. Recently, an electronic community health information system and district health information software were implemented in Ethiopian public health centers. These tools are commonly used for routine data management tasks. Frontline workers, however, lacked easy access to decision support systems and other similar point-of-care technologies. Paper-based clinical guidelines (CGs), card sheets, and point-of-care charts were the only available resources, and only limited information is documented on the card sheets [7,8]. Delivering evidence-based services at the point of care by capturing the required clinical data, summarizing and processing them in a consistent manner, and constructing a patient flow sheet to monitor and record the progress of care from the paper-based resources were challenging [7,8]. The Ethiopian national maturity health information assessment survey also revealed that there is a lack of health information infrastructure, a lack of decision support and knowledge management systems, and a lack of parameters and metrics for analyzing the impact of data [9].

Thus, introducing and integrating a CDSS with the existing health information system helps deliver appropriate, consistent, and integrated care. To introduce a CDSS in LRSs, we followed a 3-step approach:

- Step 1: A case study (maternal and childcare health services) needs analysis was conducted in LRSs to assess the available point-of-care evidence of the requirements for a CDSS, such as clinical pathways (CPs) or workflows [7,8].
- Step 2: We conducted a state-of-the-art review to investigate strategies and approaches for designing CDSS instruments for LRSs [10]. The aim was to review existing publications in the LRS context to explore recommended approaches and design considerations for building a CDSS.
- Step 3: A CDSS was developed based on the findings of the needs analysis and a review of the state of the art. The CDSS was designed to reduce delays and support frontline workers. The proposed CP algorithm, in particular, aims to find referrals and locally treatable cases by integrating knowledge-based approaches and historical evidence [11].

The aim of this study was to evaluate the user acceptance of a CDSS in LRSs. Overall, as depicted in Figure 1, this study proposed the following hypotheses to evaluate the user acceptance of the CDSS:

Figure 1. Computer-aided CDSS evaluation model hypotheses. Customized and adopted from Ji et al [20]. CDSS: Clinical decision support system; H: Hypothesis.



- Hypothesis (H)1: The perceived ease of use has a positive effect on the acceptance of a CDSS in LRSs.
- H2: System quality has a positive effect on the acceptance of a CDSS in LRSs.
- H3: Information quality has a positive effect on the acceptance of a CDSS in LRSs.
- H4: Information quality has a positive effect on the perceived ease of use of a CDSS in LRSs.

- H5: System quality has a positive effect on the perceived ease of use of a CDSS in LRSs.
- H6: Perceived benefits have a positive effect on the acceptance of CDSS in LRSs.

Methods

Ethical Considerations

Approval for the research was granted by the Institutional Review Board of the Institute of Health, Jimma University (reference number IHRPGI/467/19).

Study Settings and Participants

This study was conducted in low-resource primary health care centers, with a specific focus on the maternal and childcare health service units at the Jimma Health Center and the Jimma Higher Two Health Center. Both health centers are situated in Jimma Town, in the Oromia region, Southwestern Ethiopia. Each of them serves up to 40,000 people in its geographical area, accepts referrals from community health posts, and refers patients to the nearest hospital, such as the Shanan Gibe General Hospital and the Jimma University Specialized Hospital. The health centers serve and oversee both inpatient and outpatient cases. The number of personnel in the Jimma Higher Two Health Center is 34 and in the Jimma Health Center is 40, whereas in Ethiopia, the health center's maternal and child health service unit employs a much smaller number of health professionals, commonly 5-7 nurses and midwives. There were 5 nurses and midwives at the Jimma Health Center and 4 at the Jimma Higher Two Health Center during our investigation. The maternal and childcare health service unit is expected to serve 2000-2500 antenatal care (ANC), pregnant patient care, and postnatal care (PNC) cases annually.

Participants in the CDSS evaluation were health care professionals, such as midwives and nurses, who worked at the maternal and childcare health service unit at the Jimma Health Center and the Jimma Higher Two Health Center. The inclusion and exclusion criteria were as follows:

- Health care professionals were personnel at the maternal and childcare health service unit and were familiar with the existing clinical workflow, as well as volunteering to evaluate the CDSS.
- The ANC, pregnant patient care, and PNC cases that had been pre-recorded on the evaluation day were suitable for retrospective chart review to evaluate the CDSS.
- Both morning and afternoon evaluations were based on the pre-recorded cases from the respective morning and afternoon visits.

The CDSS evaluation was conducted in the health care professionals' spare time because the number of health care professionals at the maternal and childcare health service unit was limited, and they were so preoccupied and busy with their regular daily activities that it was not feasible to incorporate the evaluation into their routine. The health care professionals completed a questionnaire over the course of a half-day (as a summary of the half-day cases rather than as a case-by-case response), with the morning session taking place from 11:00 to1:00 A.M. and the afternoon session taking place from 5.00 to 18:30 P.M.

The initial evaluation was conducted in August 2022 at the Jimma Health Center. The second round of evaluation took place at the Jimma Health Center and the Jimma Higher Two Health Center from December 20, 2022, until January 15, 2023.

Based on our previous experience [13], obtaining the expected sample size in an LRS was difficult due to a shortage of health care professionals in the maternal and childcare health service unit (usually 4-7).

To determine the optimal strategies, we consulted the existing literature in support of our evaluation study design. Based on the findings of Mburu and Oboko's study [14], we observed that 79 cases were sufficient to assess the use of mobile health (mHealth) interventions in Kenya. Additionally, Mburu and Oboko [14] also reported that 60 subjects were sufficient to detect the small and medium effects of an exogenous latent variable (independent variable) on an endogenous latent variable (dependent variable), according to the findings of Chin and Newsted [15] and Cohen [16], just as using 40 subjects was sufficient for Goodhue et al [17]. The minimum sample size needed to observe an effect with a given power (ie, the probability of observing a statistically significant result at level *P* if a true effect of a certain magnitude is present) is determined by the effect size. The effect size is associated with the path coefficient between a variable that is assumed to describe a cause and a variable that is assumed to be an effect: values<0.02 indicate no effect, values>0.15 indicate a medium effect, and values>0.35 indicate a large effect [17,18]. Moreover, using 70-80 samples was adequate to model functional brain relationship hypotheses in the study by Sideridis et al [19]. However, Sideridis et al [19] also explicitly noted that sample sizes of 50 participants were associated with a root mean square error of approximation of <0.05, suggesting a satisfactory fit.

The study entailed a proof-of-principle CDSS evaluation using a convenience sample of 7 health professionals. Altogether, we reviewed 73 ANC, pregnant patient, and PNC cases.

Procedure and Measurement Instrument

A tutorial and a demonstration were provided to the health care professionals at the 2 health centers prior to using the CDSS. The health care professionals used and assessed the CDSS before completing a questionnaire. They used a retrospective chart review, specifically a half-day of pre-recorded patient card sheet data, to evaluate the CDSS. On the basis of pre-recorded cases, the goal was to evaluate how well the CDSS performed in identifying referrals and locally treatable cases that were actually made. The health care professionals then filled out questionnaires to provide their assessments and feedback on the CDSS. Each evaluation questionnaire was completed based on a half-day of ANC, pregnant patient, and PNC cases, as well as the health care professional's observation of the CDSS reaction to the presented cases. Next, the health care professionals answered a series of 5-point Likert scale items (1=strongly disagree, 2=disagree, 3=neutral, 4=agree, 5=strongly agree) about the CDSS [20]. The measurement instrument consisted of 22 parameters adopted from Ji et al's [12] evaluation framework. The 22 measurement items were classified into 6 factors: system quality, information quality, service quality, perceived ease of use, user acceptability, and perceived benefits. Furthermore, we automated the questionnaire submission, which was accessed via a mobile phone or a laptop. Electronic questionnaire submission was preferred over paper-based alternatives. However, paper-based questionnaire submissions were used in some cases.

The CDSS at the Point of Care

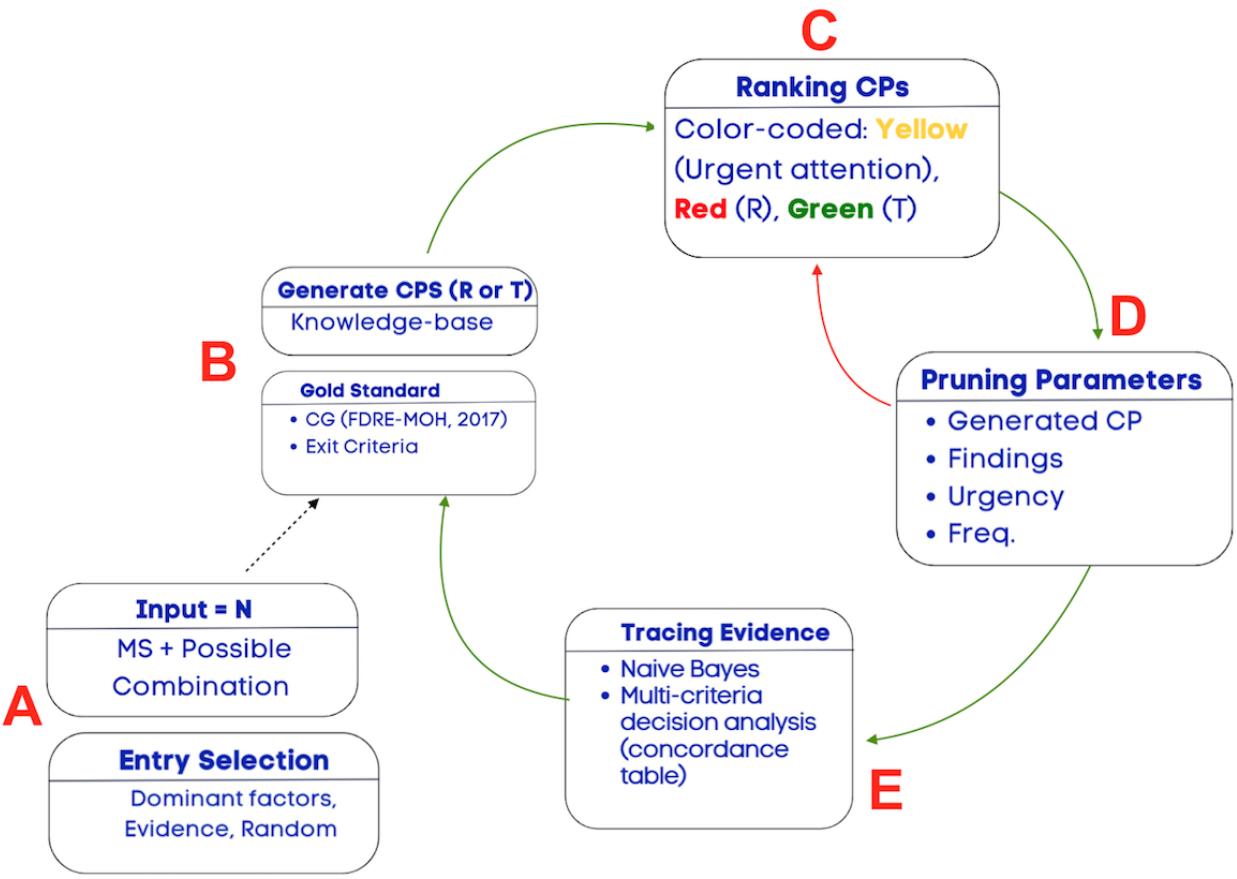
We designed and developed a CDSS to meet the requirements of LRSs. An intelligent clinical wizard, minimum data and data readiness, adaptable features, and low-cost infrastructure are some of the notable requirements and prerequisites of LRSs based on our previous results [7]. Our CDSS incorporates both existing knowledge-based guidelines and data-driven evidence to provide the most relevant information for frontline workers at the time of care delivery [11]. The CDSS provides CPs (or workflows) for point-of-care services. The CP is a critical component of a CDSS for identifying referral and locally treatable cases, which is delivered in the form of a concordance table for multicriteria decision analysis and output [11].

The CDSS has the following major goals:

- Delivering automated CPs and computer-assisted pruning and selection.
- Going beyond existing paper-based evidence that is noninteractive and challenging to grasp, the computerized CDSS was designed to be interactive for ease of use and optimal usage.
- Combining existing CGs and historical evidence (data-driven evidence) to generate an adaptable clinical workflow.

To get the most out of services, the CDSS provides an automated, interactively adaptable CP (or workflow). To reduce arbitrariness in entry point selection, the CDSS provides a range of choices for initiating the CP, such as using evidence from historical records, dominant factors, or randomly initiating the signs and symptoms based on CGs. Figure 2A presents additional information about entry point processing.

Figure 2. CP-processing workflow. CG: Clinical guideline; CP: Clinical pathway; FDRE-MOH: Federal Democratic Republic of Ethiopia Ministry of Health; Freq.: Frequency; MS: Measured symptom.



The process is interactive, and our algorithm uses measured symptoms (MSs) and a combination of MSs to process the CPs:

- First, all CPs based on the first MS are generated, as shown in Figure 2B. CGs are used as the gold standard and criterion for validating the generated CP (also referred to as an exit criterion). If the generated CP is already found on the generated list, the frequency counter is incremented. Otherwise, the generated CP is added (or appended) to the generated list of CPs. Federal Democratic Republic of Ethiopia Ministry of Health version 2017 (FDRE-MOH 2017) is used for CP processing.
- Second, a ranking of CPs is conducted to identify “referral” and “locally treatable” cases. The ranking is color-coded, as shown in Figure 2C, and the ranking criteria are based on CGs.

- Third, the dynamic CP list is pruned, as shown in Figure 2D. CP pruning is based on pruning parameters. If the generated CP list is empty, fall-back and adjustment of the pruning criterion are supported. The pruning process was designed to be interactive, flexible, responsive, and engaging. The user intervention allows for fine-tuning based on domain knowledge and provides trust and understanding for the health care professional. Pruning can also be based on findings if the health care professional requires pruning of specific CP findings. The findings are based on the CGs.
- Fourth, the naive Bayes algorithm and historical records are used to provide data-driven evidence, as shown in Figure 2E. The output is displayed in an easy-to-understand format, using a table to present the evidence. The ranked table provides evidence for assessing various factors, such as symptoms, findings, urgency, CP, CP frequency, accuracy, and prior and posterior probability, to facilitate evidence-based decision-making by the user. Since it provides evidence for analyzing various factors, we refer to it as multicriteria decision analysis. In further detail, the multicriteria output used for decision analysis is displayed in the form of a table, also known as a concordance table. A concordance table is a data (evidence) table used as a cross-reference for integrating evidence from many sources for decision support. In this study, it was used primarily for tracing what evidence was available to support the presented case and identifying the evidence's source (historical records or knowledge-based evidence). A more detailed step-by-step description of the algorithms is found in Figure 2A-E.
- Finally, the preceding steps are repeated for each additional MS.

In the end, the frontline worker must make the final decision based on the suggestions made by the algorithm. For this study and demonstration, the CDSS focused on 3 use cases, namely pregnant patient care, ANC, and PNC services. The sample user interface screenshot for each step is shown in Figures 3-7.

Figure 3. Screenshot of input processing. BP: blood pressure.

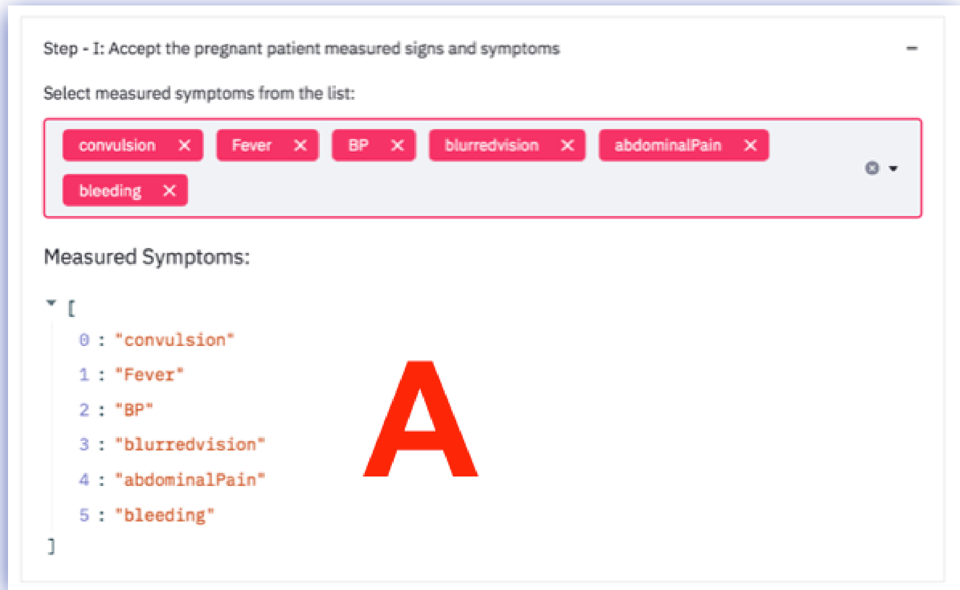


Figure 4. Screenshot of the generated CPs and the gold standard. ANC: antenatal care; CG: clinical guideline; CP: clinical pathway; KB: knowledge base; NC, not classified; PNC: postnatal care; R: referral; T: treatable.

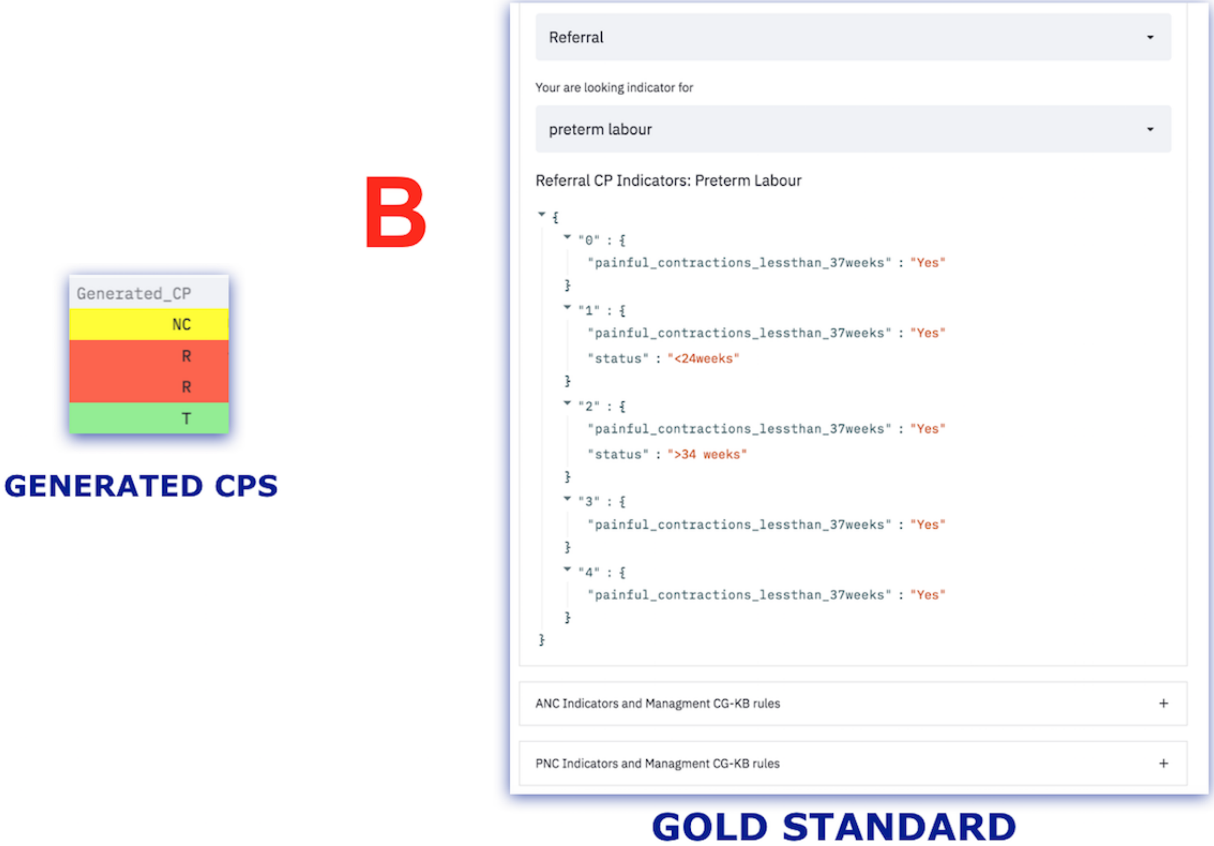


Figure 5. CP ranking. BP: blood pressure; CP: clinical pathway; HC: health center.

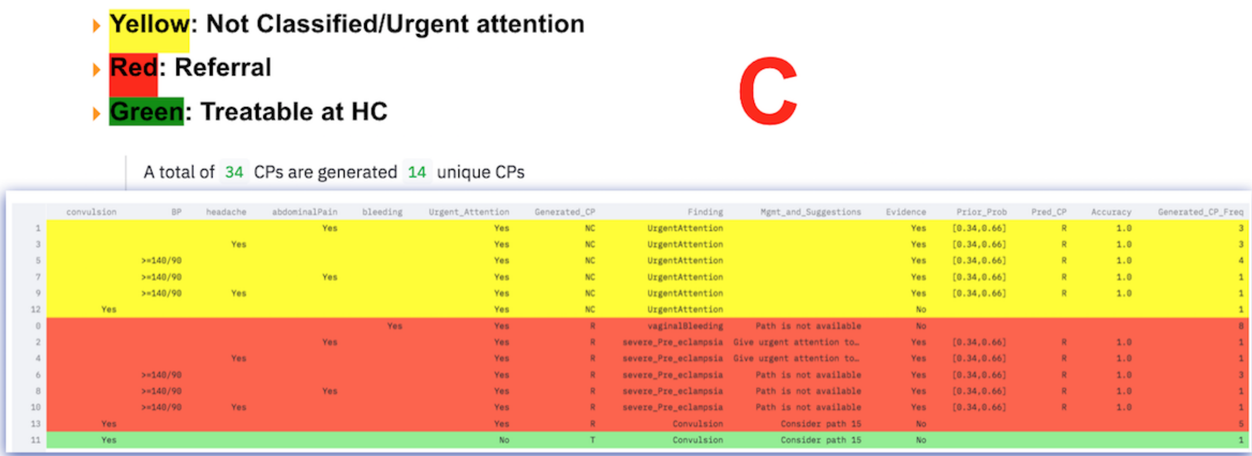


Figure 6. CP pruning. BP: blood pressure; CP: clinical pathway.

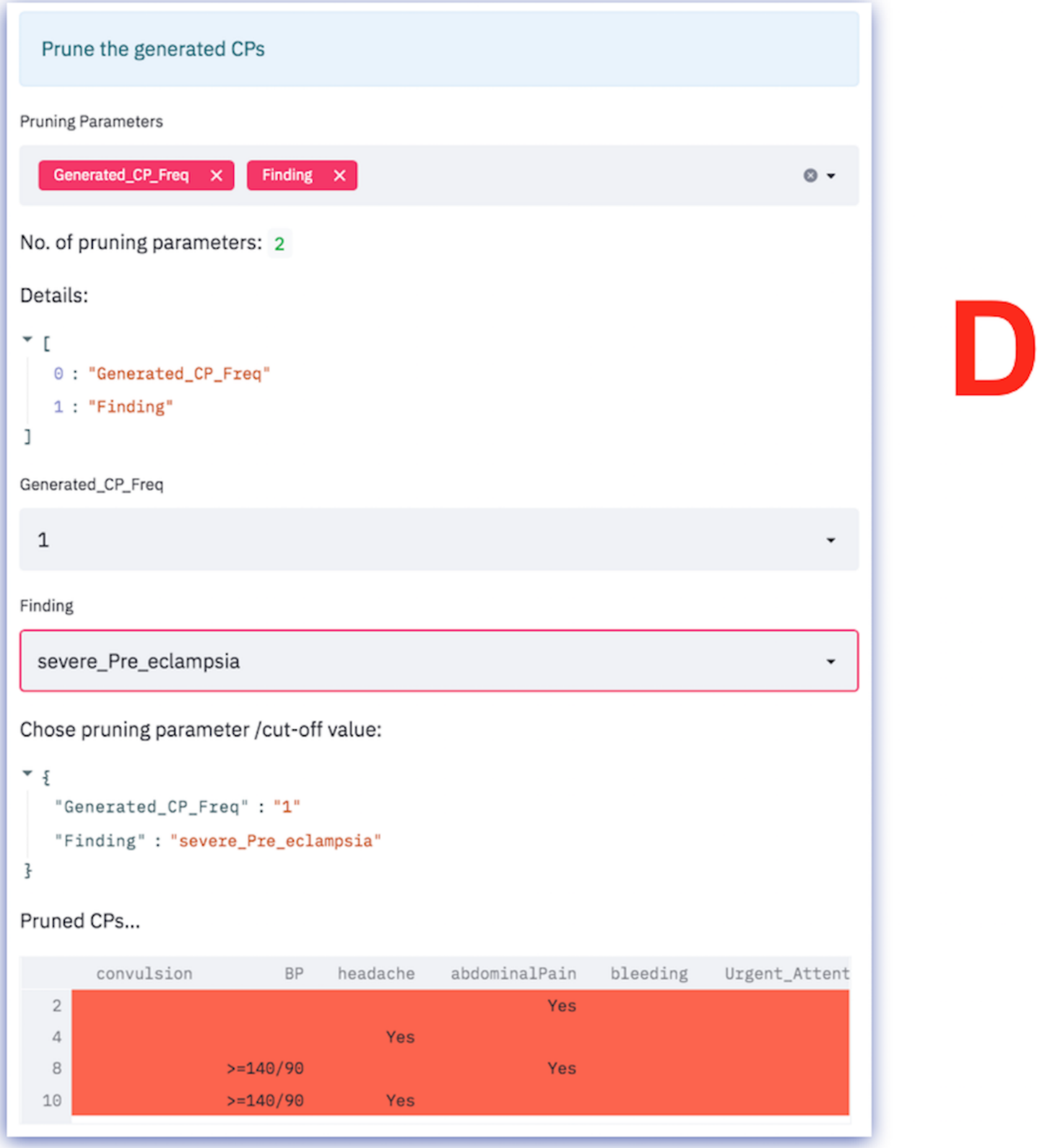
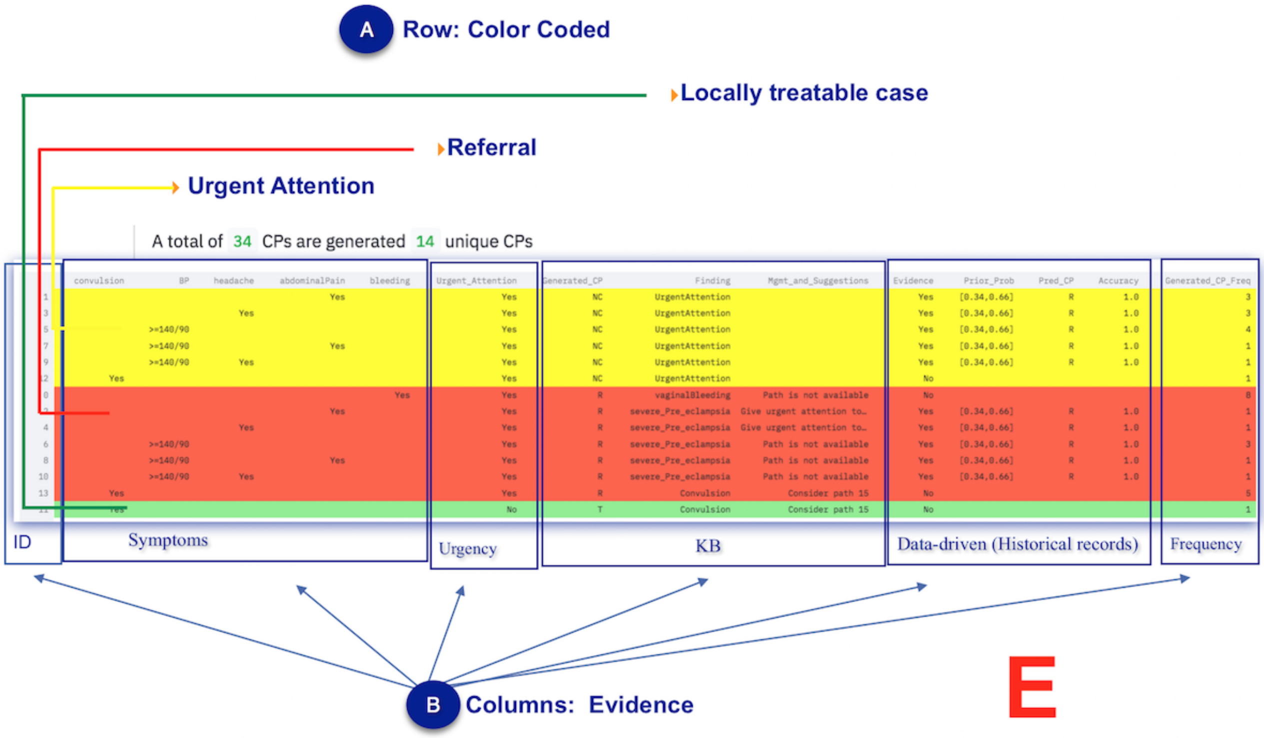


Figure 7. Concordance table. BP: blood pressure; CP: clinical pathway.



Data Analysis

Statistical Package for Social Sciences (SPSS; IBM Corporation) version 26.0 [21], Microsoft Excel [22], Python (version 3.7) [23], and SmartPLS (version 26.0) [24] were used to conduct the analysis and modeling.

We followed the procedures and recommendations of Boone and Boone [25] for the CDSS evaluation based on Likert data analysis. Latent variables were computed by summing the following items:

- The perceived ease of use was a latent variable based on learnability, operability, user interface, data entry, advice to display, and legibility items.
- Response time and stability items were used to assess system quality.
- Information quality was based on security and CP performance items.
- Acceptance included usage, confirmation of expectations, overall quality satisfaction, overall satisfaction, and the intention to use items.
- Perceived benefits were created using decision change (change in order behavior, change in CP) and process change (effectiveness, overall usefulness, adherence to standards, medical quality, and user knowledge and skills) items.

To assess the scale of the CDSS evaluation data set, the validity of the measurement model was checked. Convergent validity was assessed using factor loading and average variance extracted (AVE), with a factor loading threshold of more than 0.70 and an AVE threshold of >0.50 [26,27]. In this study, items with factor loadings of less than 0.70 were candidates for deletion. The internal consistency and reliability of the CDSS evaluation measurement model were assessed using Cronbach α [28] and composite reliability. A recommended value of >.70 for Cronbach α and composite reliability was accepted. We used the heterotrait-monotrait ratio of correlations (HTMT) [26,29] to check the discriminant validity of the measurement model and determined whether the value was less than 0.90 and acceptable. Moreover, perceived benefits are formative second-order construct based on decision changes and process changes. Collinearity was checked to ensure that it did not have a negative impact on the higher-order-construct measurement model, and critical levels of collinearity less than 0.50 were acceptable in this study, as recommended by Hair et al [26].

Following that, item-level and construct-level analyses were performed. On the one hand, an item-level analysis of the CDSS in LRSs between the 2 health centers was conducted. A nonparametric independent-samples statistical test, such as the Mann-Whitney *U* test [30], was used to see whether the 2 health centers were significantly different at the item level. We used the Mann-Whitney *U* test because we could not assume normality in either group and the independent data set observation assumptions were fulfilled, which are preconditions for the use of nonparametric data analysis [31]. Furthermore, there were no significant results from the Shapiro-Wilk test [32] on the normality of our evaluation data set. The significance level used for the inferential statistics was P=.05 and a 95% CI level.

On the other hand, we followed the recommendation of Boone and Boone [25] to use Pearson correlation for construct-level (latent variable) correlation analysis. As a result, Pearson correlation [25,33] was used to examine the factors influencing the acceptance of the CDSS in LRSs and the interrelationships between construct factors. In particular, the relationship between system quality and perceived ease of use, information quality and perceived ease of use, user acceptance and perceived benefits, and user acceptance and information and system quality were explicitly explored.

Finally, structural equation modeling (SEM) is a multivariate statistical analysis technique that is used to analyze structural relationships. It is described in the literature as combined factor analysis and regression analysis for discovering relationships between measured variables and latent constructs [34]. There is a debate on how effective it is to discover causation beyond correlation. In papers dealing with applications of the technique, it is commonly used to express a causal hypothesis in a context where there is semantic information available that supports the validity of the hypothesis or at least does not contradict it [12,14,35]. Our study was a pilot study, not a full, cross-sectional analysis, and it intended to promote the use of partial least squares structural equation modeling (PLS-SEM) [26]. PLS-SEM was used to model the acceptance of the CDSS in LRSs, particularly to model the relationship between the CDSS evaluation measured items and construct variables, as well as between multiple construct variables. We noticed that penalized likelihood estimation algorithms based on regularized structural equation modeling (RegSEM) [36,37] and PLS-SEM [26] were the best candidates for our modeling. We preferred PLS-SEM for the following reasons:

- The SmartPLS [26,38] partial least squares (PLS) algorithm was used to analyze the model's path weight, and it performed well in Mburu and Oboko's [14] study.
- The variation-based structural equation models do not impose a sample size [39] or normality of distribution constraints [26,38].

Overall, to construct the PLS-SEM model for the CDSS in LRSs, first, composite factor analysis was used to examine the validity of the measurement model, including reliability and validity analysis. The relationships in path models with latent variables were then evaluated using PLS-SEM path analysis and coefficients. Finally, the statistical significance of PLS-SEM results, such as path coefficients, outer weights, Cronbach α, and coefficient of determination (R^2) values, was determined using bootstrapping [26]. The bootstrapping settings were percentile bootstrap, 2-tailed test type, and significance level=.05.

Results

Characteristics

The 7 CDSS evaluators were all female (ie, n=4, 57%, from the Jimma Health Center and n=3, 43%, from the Jimma Higher Two Health Center), who worked as health care professionals (eg, midwives and nurses) in the health centers' maternal and childcare health service units. In total, 73 CDSS evaluation responses were recorded based on 116 ANC, pregnant patient care, and PNC cases (n=4, 5%, during the first evaluation period and n=69, 95%, during the second evaluation period). The response was 73 since the evaluation response was based on a summary of half-day cases rather than a case-by-case response. The average time for evaluating the CDSS and completing the questionnaire was 52.35 minutes, with the smallest and longest durations being 31 and 98 minutes, respectively. The Jimma Health Center accounted for 65.5% (76/116) cases, while the Jimma Higher Two Health Center accounted for 34.5% (40/116) cases. Furthermore, we observed that each health center handled 4-6 (3%-5%) cases per day on average. Overall, the first round of evaluation lasted 2 days and included 18 ANC, pregnant patient care, and PNC cases in the Jimma Health Center, which is above average. In round 2, there were 75 ANC cases, 7 pregnant patient care cases, and 16 PNC cases during our evaluation period. The second round of evaluation took place in both health centers, the Jimma Health Center and the Jimma Higher Two Health Center.

The computer-aided CDSS evaluation's mean (SD) score ranged from 4.29 (SD 0.485) to 4.52 (SD 0.503). Table 1 provides more extensive details of each item score.

Table 1. Mean Likert scale scores and reliability analysis for computer-aided CDSS[a] evaluation in LRSs[b].

Construct and items	Value, minimum (maximum)	Score of 73 CDSS evaluation responses, mean (SD)
Perceived ease of use		
Learnability	2 (5)	4.30 (0.545)
Operability	3 (5)	4.29 (0.485)
User interface	3 (5)	4.34 (0.533)
Data entry	3 (5)	4.40 (0.571)
Advice display	3 (5)	4.37 (0.589)
Legibility	1 (5)	4.29 (0.905)
System quality		
Response time	3 (5)	4.38 (0.543)
Stability	2 (5)	4.38 (0.615)
Information quality		
Security	3 (5)	4.32 (0.550)
CP[c] performance	3 (5)	4.37 (0.540)
Decision change		
Change in order behavior	2 (5)	4.08 (0.640)
Change in CP	2 (5)	4.23 (.613)
Process changes		
Effectiveness	3 (5)	4.25 (0.494)
Overall usefulness	3 (5)	4.23 (0.635)
Adherence to standards	3 (5)	4.33 (0.502)
Medical quality	3 (5)	4.29 (0.612)
User knowledge and skills	2 (5)	4.30 (0.570)
Acceptance		
Usage	3 (5)	4.49 (0.580)
Confirmation of expectations	2 (5)	4.34 (0.628)
Satisfaction with overall quality	3 (5)	4.40 (0.571)
Overall satisfaction	—[d]	4.30 (0.570)
Intention to use	4 (5)	4.52 (0.503)

[a]CDSS: clinical decision support system.
[b]LRS: low-resource setting.
[c]CP: clinical pathway.
[d]Not applicable.

CDSS Evaluation Measurement Model

The factor loading of 20 (91%) of 22 items was greater than 0.70. The remaining items, legibility and medical quality, were eliminated since their factor loading value was less than 0.70. All the constructs had Cronbach α values greater than .70, except information quality, for which Cronbach α was .699, which is close to .70. Table 2 provides more information about the measurement model's construct reliability and validity.

To establish discriminant validity, the HTMT on construct factors was used, and the results showed that all constructs passed the test. Table 3 displays the results of the discriminant validity assessment.

Table 2. CDSS[a] measurement model's construct reliability and validity.

Construct and items	Convergent validity		Internal consistency and reliability	
	Factor loading (>0.70)	AVE[b] (>0.50)	Composite reliability (>0.70)	Cronbach α (>.70)
Perceived ease of use		0.588	0.847	.825
Learnability	0.721	—[c]	—	—
Operability	0.738	—	—	—
User interface	0.746	—	—	—
Data entry	0.856	—	—	—
Advise to display	0.836	—	—	—
System quality		0.879	0.869	.863
Response time	0.934	—	—	—
Stability	0.944	—	—	—
Information quality		0.763	0.767	.699
Security	0.825	—	—	—
CP[d] performance	0.930	—	—	—
Decision changes		0.776	0.712	.712
Change in order behavior	0.856	—	—	—
Change in CP	0.856	—	—	—
Process changes		0.650	0.824	.819
Effectiveness	0.773	—	—	—
Overall usefulness	0.813	—	—	—
Adherence to standards	0.896	—	—	—
User knowledge and skills	0.762	—	—	—
Acceptance		0.654	0.871	.867
Usage	0.738	—	—	—
Confirmation of expectations	0.819	—	—	—
Satisfaction with overall quality	0.806	—	—	—
Overall satisfaction	0.846	—	—	—
Intension to use	0.815	—	—	—
Perceived benefits				
Constructed based on decision and process changes	—	0.511	0.848	0.839

[a]CDSS: clinical decision support system.

[b]AVE: average variance extracted.

[c]Not applicable.

[d]CP: clinical pathway.

Table 3. CDSS[a] discriminant validity assessment.

Constructs	Perceived ease of use	Information quality	Perceived benefits	Perceived user acceptance	System quality
Perceived ease of use	—[b]	—	—	—	—
Information quality	0.855	—	—	—	—
Perceived benefits	0.643	0.852	—	—	—
Perceived user acceptance	0.616	0.877	0.877	—	—
System quality	0.545	0.839	0.618	0.779	—

[a]CDSS: clinical decision support system.

[b]Not applicable.

CDSS Evaluation Between the 2 Health Centers

The results of the nonparametric Mann-Whitney *U* test based on the 5-point Likert item evaluation data set collected from the Jimma Health Center and the Jimma Higher Two Health Center revealed that the 2 health centers did not differ significantly in the CDSS item-level evaluation factors, except stability (U=470.5, P=.022), overall usefulness (U=451.0, P=.012), adherence to standards (U=483, P=.024), confirmation of expectations (U=488.5, P=.04), satisfaction with overall quality (U=400.5, P=.001), and overall satisfaction (U=474.5, P=.023). The findings of the CDSS evaluation using the Mann-Whitney *U* test are shown in Table 4.

Table 4. Mann-Whitney U test results (P<.05).

Construct and items	Mean rank		Test statistics[a]	
	Jimma Health Center (n=42)	Jimma Higher Two Health Center (n=31)	Mann-Whitney *U*	Asymptotic significance (2-tailed) *P* value
Perceived ease of use				
Learnability	36.92	37.11	647.5	.962
Operability	38.75	34.63	577.5	.309
User interface	37.48	36.35	631.0	.794
Data entry	37.85	35.85	615.5	.653
Advice display	36.14	38.16	615.0	.650
Legibility	34.76	40.03	557.0	.249
System quality				
Response time	33.48	41.77	503.0	.057
Stability	32.70	42.82	470.5	.022
Information quality				
Security	35.49	39.05	587.5	.409
CP[b] performance	35.65	38.82	594.5	.466
Decision changes				
Change in order behavior	36.48	37.71	629.0	.767
Change in CP	35.51	39.02	588.5	.416
Process changes				
Effectiveness	35.82	38.60	601.5	.489
Overall usefulness	32.24	43.45	451.0	.012
Adherence to standards	33.00	42.42	483.0	.024
Medical quality	34.43	40.48	543.0	.174
User knowledge and skills	34.50	40.39	546.0	.164
Acceptance				
Usage	32.79	42.71	474.0	.024
Confirmation of expectations	33.13	42.24	488.5	.04
Satisfaction with overall quality	31.04	45.08	400.5	.001
Overall satisfaction	32.80	42.69	474.5	.023
Intension to use	35.38	39.19	583.0	.381

[a]Grouping variable: health center.

[b]CP: clinical pathway.

CDSS Evaluation Agreement Score Observation in the Jimma Health Center

Although the total number of observations in the first and second rounds of the CDSS evaluation were not equal, we found a positive mean agreement score increment in the majority of evaluation parameters at the Jimma Health Center, which was calculated using “agree” and “strongly agree” responses. Adherence to the standards agreement score, however, declined from 1.00 to 0.974. The first and second round CDSS evaluation agreement score observations at the Jimma Health Center are shown in Table 5.

Table 5. First and second round CDSS[a] evaluation agreement score observations at the Jimma Health Center.

Construct and items	Round 1: agreement score based on n=4 observations, mean (SD)		Round 2: agreement score based on n=38 observations, mean (SD)	
	Item level	Construct level	Item level	Construct level
Perceived ease of use	—[b]	0.708	—	0.982
Learnability	0.750	—	1.000	—
Operability	0.750	—	1.000	—
User interface	0.750	—	1.000	—
Data entry	0.750	—	1.000	—
Advice to display	0.750	—	1.000	—
Legibility	0.500	—	0.890	—
System quality	—	0.625	—	0.987
Response time	0.750	—	1.000	—
Stability	0.500	—	0.974	—
Information quality	—	0.750	—	0.974
Security	0.750	—	0.947	—
CP[c] performance	0.750	—	1.000	—
Decision changes	—	0.750	—	0.960
Change in order behavior	0.500	—	0.947	—
Change in CP	1.000	—	0.974	—
Process changes	—	0.800	—	0.953
Effectiveness	0.750	—	1.000	—
Overall usefulness	0.750	—	0.842	—
Adherence to standards	1.000	—	0.974	—
Medical quality	0.750	—	0.947	—
User knowledge and skills	0.750	—	0.974	—
User acceptance	—	0.750	—	0.958
Usage	0.750	—	0.947	—
Confirmation of expectations	0.500	—	0.974	—
Satisfaction with overall quality	0.750	—	0.947	—
Overall satisfaction	0.750	—	0.921	—
Intention to use	1.000	—	1.000	—

[a]CDSS: clinical decision support system.

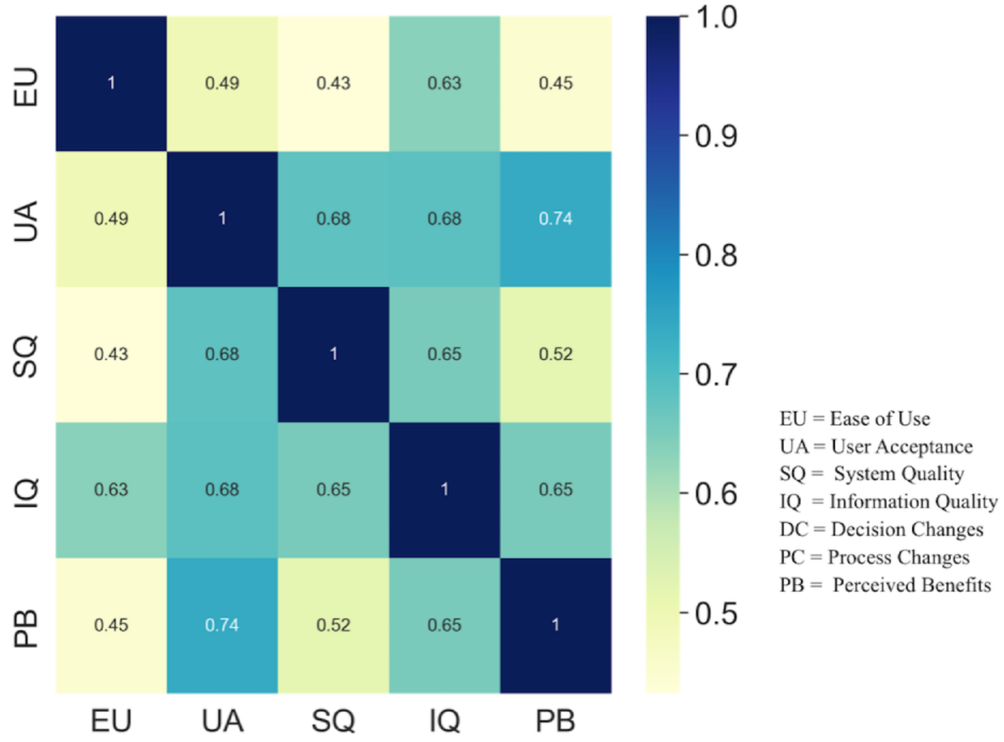
[b]Not applicable.

[c]CP: clinical pathway.

CDSS Evaluation: Construct Factor Interrelationships

We found a significant correlation (r =0.74) between user acceptance and perceived benefits, with perceived benefits as construct factors based on process changes and decision changes. The coefficient of correlation between perceived ease of use and information quality was r=0.63. User acceptance was also correlated with information quality and system quality, with r=0.68. Figure 8 depicts a more detailed Pearson correlation test result.

Figure 8. Pearson correlation (N=73). Correlation values ranging from 0.50 to 0.70 are considered moderate, from 0.70 to 0.90 are considered strong, and from 0.9 to 1.0 are considered very strong [33].



Modeling the Acceptance of the CDSS in LRSs

The perceived user acceptance coefficient of determination (R^2) was 0.703, showing that user acceptance is influenced by system quality, information quality, perceived ease of use, and perceived benefits (Figure 9). More precisely, system quality (β=.321, *P*<.05) and perceived benefits (β=.486, *P*<.05) were shown to have a significant influence. However, the perceived ease of use had no positive effect on CDSS acceptance (β=.015, *P*=.91). Information quality also had no positive effect on CDSS acceptance in this study (β=.149, and *P*=.25).

Furthermore, we found that the perceived ease of use was influenced by system quality, and information quality, with an R^2 value of 0.479. The path coefficient of information quality on the perceived ease of use was β=.021(*P*=.89), and hence, no significant effect was found. The path coefficient of system quality on the perceived ease of use was β=.678 (*P*<.05), that is, a significant influence, whereas the perceived benefits impacted by decision and process changes had an R^2 value of 0.983. The path coefficients of decision changes and process changes were β=.374 and β=.749, respectively, and were significant (*P*<.05). Figure 9 depicts the path weights, *P* values, and coefficient of determination (R^2) for the CDSS evaluation PLS-SEM model developed using the CDSS Jimma Health Center and Jimma Higher Two Health Center evaluation data sets. The results, shown in Figure 9, can be interpreted as perceived ease of use -> perceived user acceptance (β=.015 and *P*=.91), for example. Overall, we found that the perceived ease of use had no positive effect on CDSS acceptance (β=.015, *P*=.91) but, rather, on the system quality (β=.321, *P*<.05) and perceived benefits (β=.486, *P*<.05) of the CDSS. Further information is presented in Table 6.

Figure 9. Computer-aided CDSS evaluation PLS-SEM model generated from the computer-aided CDSS Jimma Health Center and the Jimma Higher Two Health Center evaluation data sets, showing path weights (β), *P* values, and coefficient of determination (R^2). The yellow boxes represent indicators (or parameters). The construct variables are represented by the circle. The path indicates the path weight and *P* value. For example, a 0.321 (.002) value from system quality -> perceived user acceptance shows that β=.321 and *P*=.002. CDSS: clinical decision support system; PLS-SEM: partial least squares structural equation modeling.

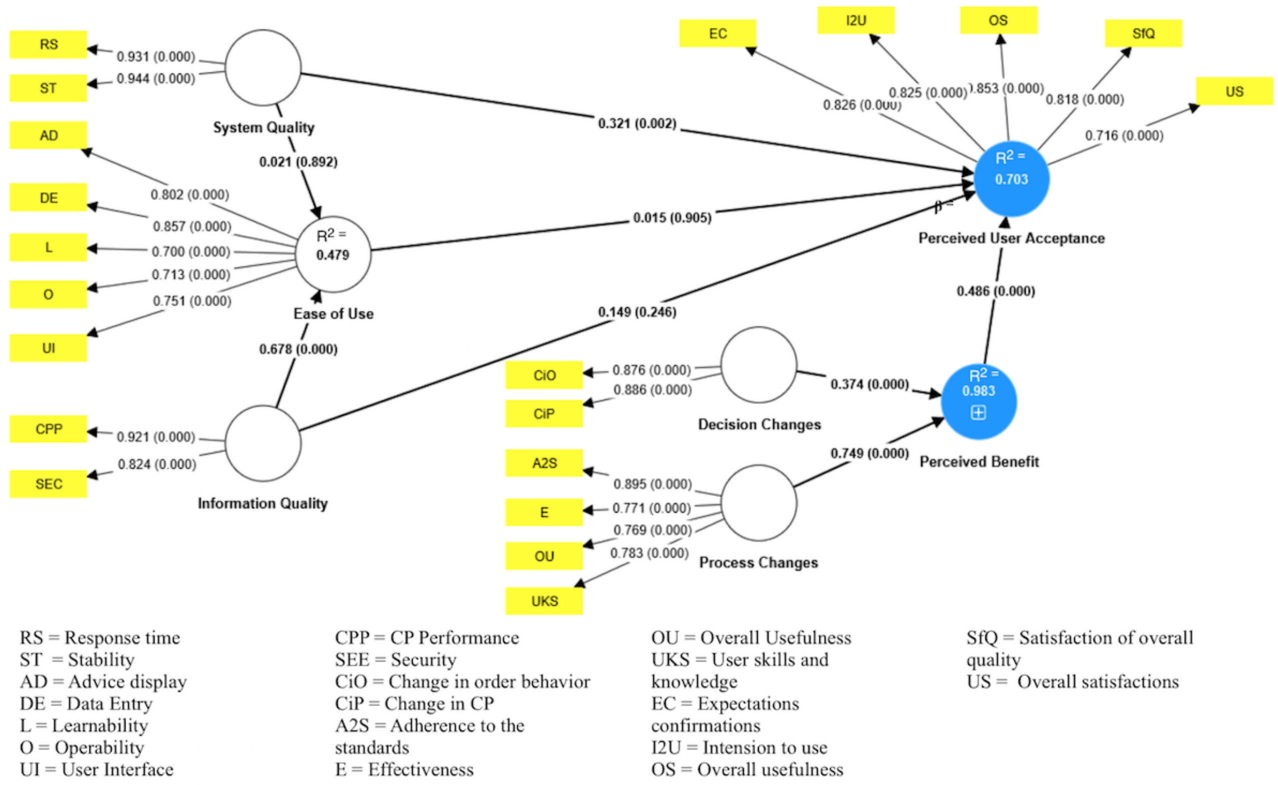


Table 6. Hypotheses conclusion based on PLS-SEM[a] findings (β=.015, *P*=.91).

Hypothesis	Path and relationships	PLS-SEM findings[b]			Conclusion
		β (SD)	*t* Statistics	*P* value	
Hypothesis (H)1: The perceived ease of use has a positive effect on the acceptance of a CDSS[c] in LRSs[d].	Perceived ease of use -> acceptance	.015 (0.123)	0.119	.91	Rejected
H2: System quality has a positive effect on the acceptance of a CDSS in LRSs.	System quality -> acceptance	.321(.102)	3.139	.002	Accepted
H3: Information quality has a positive effect on the acceptance of a CDSS in LRSs.	Information quality -> acceptance	.149 (0.128)	1.162	.25	Rejected
H4: Information quality has a positive effect on the perceived ease of use of a CDSS in LRSs.	Information quality -> perceived ease of use	.678 (0.122)	5.558	<.001	Accepted
H5: System quality has a positive effect on the perceived ease of use of a CDSS in LRSs.	System quality -> perceived ease of use	.021 (0.153)	.135	.89	Rejected
H6: Perceived benefits have a positive effect on the acceptance of a CDSS in LRSs.	Perceived benefits -> acceptance	.486 (0.115)	4.234	<.001	Accepted

[a]PLS-SEM: partial least squares structural equation modeling.

[b]Relationships were significant at *P*<.05.

[a]CDSS: clinical decision support system.

[d]LRS: low-resource setting.

Discussion

Principal Findings

This study aimed to evaluate a CDSS for use at the point of care in primary care LRSs. The health care professionals in this study evaluated user acceptance of the CDSS.

The Cronbach α scale of 22 items appeared to be internally consistent, exceeding the minimum value of .70 required for acceptable reliability [26-28,32]. In this study, the 2 health centers did not differ significantly in terms of the CDSS's perceived ease of use, information quality, and perceived benefits (decision changes and process changes). However, we found a significant difference in system quality, such as stability, and perceived user acceptance, such as overall usefulness, adherence to standards, confirmation of expectations, satisfaction with overall quality, and overall satisfaction. This variation could be attributed to the first round of evaluation, which was based on the Jimma Health Center, or to the fact that more cases were observed in the Jimma Health Center than in the Jimma Higher Two Health Center, but more research and analysis are required. Furthermore, based on the first and second rounds of the CDSS evaluation, we observed a positive agreement score increment at the Jimma Health Center. However, this study was unable to observe a change in the Jimma High Two Health Center, since the first round of evaluation was limited to the Jimma Health Center.

This study highlighted a correlation between construct variables using Pearson correlation. The CDSS's system quality, information quality, and perceived benefits were vital for its acceptance in the LRS. The perceived benefits were based on decision and process changes. In accordance with our results, previous studies have demonstrated that the acceptance and use of mHealth apps in LRSs are influenced by users' perceptions that are aligned with health needs and expectations [14]. However, in this study, the perceived ease of use was moderately correlated with CDSS acceptance, whereas Ji et al [12] suggested that the perceived ease of use has a significant and direct impact on the acceptance of a CDSS. Figure 8 depicts further information about the Pearson correlation between the construct variables of the CDSS. Overall, we observed a low positive Pearson correlation between the perceived ease of use and acceptance, as well as between system quality and the perceived ease of use, when we considered the strength of correlation classification as in Mukaka [33]. System quality and acceptance, information quality and acceptance, and information quality and perceived ease of use all showed a moderately positive correlation. There was a high positive correlation between perceived benefits and acceptance, supporting Pande et al's [40] finding that perceived usefulness is significantly correlated to the intention to use.

This study also used PLS-SEM to evaluate several factors that impact the acceptance of a CDSS in LRSs. The result demonstrated that user acceptance is impacted by system quality, information quality, and perceived benefits, with an R^2 value of 0.703, as shown in Figure 9. The perceived benefits influenced by decision and process changes had an R^2 value of 0.983, whereas the R^2 score for the perceived ease of use as impacted by system and information quality was only 0.479. All retained R^2 values were greater than 0.10, as suggested by Falk and Miller [41]. The R^2 values of the perceived user acceptance and perceived benefits were substantial, as also indicated by the CDSS results of Cohen [18], who reported $R^2>0.26$, and Chin [42], who reported $R^2>0.67$. However, according to the criteria of Hair et al [43], R^2 of perceived benefits is greater than 0.75 and substantial, while R^2 of the perceived ease of use and user acceptance is greater than 0.50 and moderate. However, Mohamed et al [44] showed that the coefficient of determination must be larger than 0.19, the path coefficient between latent variables must be at least 0.1, and the significance level must be at least .05 in order to validate the model. Our CDSS evaluation model meets all these criteria, except the path coefficient from perceived ease of use to perceived user acceptance, which was 0.015. Hair et al [26], however, stated that path coefficients with standardized values greater than 0.20 are typically significant, while in this study, the path coefficient from perceived ease of use to user acceptance was 0.015, which is less than 0.10 and not significant. More information is depicted in Figure 9, which includes the details of the CDSS assessment PLS-SEM model developed from the CDSS Jimma Health Center and Jimma Higher Two Health Center evaluation data sets, including path weights, *P* values, and the coefficient of determination (R^2).

Overall, as shown in Table 6, the PLS-SEM results suggested that the perceived ease of use has no positive effect on CDSS acceptance (β=.015, *P*=.91) but, rather, on system quality (β=.321, *P*=.002) and perceived benefits (β=.486, *P*<.001) of the CDSS. We also observed that information quality had a positive influence on the perceived ease of use (β=.678, *P*<.001). However, system quality had no favorable impact on the perceived ease of use (β=.021, *P*=.89). The detailed conclusions and summary based on PLS-SEM are shown in Table 6.

Limitations

In this study, we evaluated our own proof-of-principle CDSS in LRSs. The small sample size and low number of cases in our study might limit the generalizability of our findings. As a result, difficulties that were not identified during this investigation may be identified during a longitudinal and case-by-case evaluation.

Conclusion

We designed and developed a CDSS based on LRS requirements, which we evaluated in 2 LRSs in Ethiopia: the Jimma Health Center and the Jimma Higher Two Health Center. Our overall result indicates that user acceptance is impacted by system quality, information quality, perceived ease of use, and perceived benefits, with an R^2 value of 0.703. Specifically, system quality and perceived benefits have a direct impact on user acceptance of the CDSS in LRSs. In this study, however, we found that the perceived ease of use and information quality had no positive effect on CDSS acceptability. Overall, the proposed acceptance model includes specific factors and variables, which is an important step toward the successful adoption and implementation of a CDSS in LRSs.

Acknowledgments

We would like to thank the health care professionals at the Jimma Health Center's and the Jimma Higher Two Health Center's maternal and childcare health service units. Sisters Aynalem Wubet, Digiti Abaraya, and Asiya, in particular, deserve special appreciation. Their voluntarism will not be forgotten. The evaluation was possible due to their efforts in administering the clinical decision support system evaluation from beginning to end. We would also like to thank Mr Balew Ayalew (a Jimma University lecturer) for facilitating and assisting with the supervision of the evaluation at the Jimma Higher Two Health Center. Finally, we thank the Network for Advancement of Sustainable Capacity in Education and Research in Ethiopia (NASCERE) program.

Authors' Contributions

GST, DAS, GKT, FV, JC, and BJ conceptualized the research goals and objectives, as well as the methodology. GST conducted data curation, formal analysis, investigation, visualization, and drafting of the manuscript. DAS, GKT, FV, JC, and BJ were involved in the supervision, validation, visualization, and review and editing of the manuscript, as well as the final proofreading. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Haux R. Health information systems - past, present, future. Int J Med Inform 2006 Mar;75(3-4):268-281. [doi: 10.1016/j.ijmedinf.2005.08.002] [Medline: 16169771]
2. Azubuike M, Ehiri J. Health information systems in developing countries: benefits, problems, and prospects. J R Soc Promot Health 1999 Sep 07;119(3):180-184. [doi: 10.1177/146642409911900309] [Medline: 10518358]
3. Pellé KG, Rambaud-Althaus C, D'Acremont V, Moran G, Sampath R, Katz Z, et al. Electronic clinical decision support algorithms incorporating point-of-care diagnostic tests in low-resource settings: a target product profile. BMJ Glob Health 2020 Feb 28;5(2):e002067 [FREE Full text] [doi: 10.1136/bmjgh-2019-002067] [Medline: 32181003]
4. Abbas JJ, Smith B, Poluta M, Velazquez-Berumen A. Improving health-care delivery in low-resource settings with nanotechnology: challenges in multiple dimensions. Nanobiomedicine (Rij) 2017 Mar 29;4:1849543517701158 [FREE Full text] [doi: 10.1177/1849543517701158] [Medline: 29942391]
5. López DM, Rico-Olarte C, Blobel B, Hullin C. Challenges and solutions for transforming health ecosystems in low- and middle-income countries through artificial intelligence. Front Med (Lausanne) 2022 Dec 2;9:958097 [FREE Full text] [doi: 10.3389/fmed.2022.958097] [Medline: 36530888]
6. Alliance for Health Policy and Systems Research. Primary Health Care Systems (PRIMASYS): Case Study from Ethiopia; abridged version. Geneva: World Health Organization; 2017:1-12.
7. Tegenaw G, Amenu D, Ketema G, Verbeke F, Cornelis J, Jansen B. Using clinical guidelines and card sheets for guiding the design of data-driven clinical pathways. J Health Inform Dev Ctries 2021 Nov 12;15(2):1-18.
8. Tegenaw GS, Amenu D, Ketema G, Verbeke F, Cornelis J, Jansen B. Analysis of low resource setting referral pathways to improve coordination and evidence-based services for maternal and child health in Ethiopia. PLoS One 2022 Aug 25;17(8):e0273436 [FREE Full text] [doi: 10.1371/journal.pone.0273436] [Medline: 36007079]
9. Biru A, Birhan D, Melkamu G, Gebeyehu A, Omer AM. Pathways to improve health information systems in Ethiopia: current maturity status and implications. Health Res Policy Syst 2022 Jun 29;20(1):78 [FREE Full text] [doi: 10.1186/s12961-022-00860-z] [Medline: 35768819]
10. Tegenaw G, Amenu D, Ketema G, Verbeke F, Cornelis J, Jansen B. Design approaches for executable clinical pathways at the point of care in limited resource set-tings to support the clinical decision process: review of the state of the art. 2021 Presented at: MobiHealth 2021: International Conference on Wireless Mobile Communication and Healthcare; November 13-14, 2021; Chongqing, PRC. [doi: 10.1007/978-3-031-06368-8_13]
11. Tegenaw G, Amenu D, Ketema G, Verbeke F, Cornelis J, Jansen B. A hybrid approach for designing dynamic and data-driven clinical pathways point of care instruments in low resource settings. 2021 Presented at: MEDINFO 2021: One World, One Health—Global Partnership for Digital Innovation: 18th World Congress on Medical and Health Informatics; October 2-4, 2021; Virtual p. 316. [doi: 10.3233/shti220087]
12. Ji M, Genchev GZ, Huang H, Xu T, Lu H, Yu G. Evaluation framework for successful artificial intelligence–enabled clinical decision support systems: mixed methods study. J Med Internet Res 2021 Jun 02;23(6):e25929 [FREE Full text] [doi: 10.2196/25929] [Medline: 34076581]

13. Tegenaw GS, Amenu D, Ketema G, Verbeke F, Cornelis J, Jansen B. Evaluating a clinical decision support point of care instrument in low resource setting. BMC Med Inform Decis Mak 2023 Mar 30;23(1):51 [FREE Full text] [doi: 10.1186/s12911-023-02144-0] [Medline: 36998074]
14. Mburu S, Oboko R. A model for predicting utilization of mHealth interventions in low-resource settings: case of maternal and newborn care in Kenya. BMC Med Inform Decis Mak 2018 Jul 17;18(1):67. [doi: 10.1186/s12911-018-0649-z] [Medline: 30016943]
15. Chin W, Newsted P. Structural equation modeling analysis with small samples using partial least squares. In: Hoyle RE, editor. Statistical Strategies for Small Sample Research. Thousand Oaks, CA: SAGE Publications; 1999:307-341.
16. Cohen J. A power primer. Psychol Bull 1992 Jul 01;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155] [Medline: 19565683]
17. Goodhue D, Lewis W, Thompson R. PLS, small sample size, and statistical power in MIS research. 2006 Presented at: HICSS'06: 39th Annual Hawaii International Conference on System Sciences; January 4-7, 2006; Kauai, Hawaii p. 202b. [doi: 10.1109/hicss.2006.381]
18. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Abingdon-on-Thames, UK: Routledge; 1988.
19. Sideridis G, Simos P, Papanicolaou A, Fletcher J. Using structural equation modeling to assess functional connectivity in the brain: power and sample size considerations. Educ Psychol Meas 2014 Mar 04;74(5):733-758. [doi: 10.1177/0013164414525397]
20. Bertram D. Likert scales: CPSC 681—topic report. Poincare. 2006. URL: http://poincare.matf.bg.ac.rs/~kristina/topic-dane-likert.pdf [accessed 2024-05-09]
21. SPSS Statistics 26.0.0. IBM Corp. URL: https://www.ibm.com/docs/en/spss-statistics/26.0.0 [accessed 2024-05-09]
22. Microsoft Excel. Microsoft Corporation. 2022. URL: https://office.microsoft.com/excel [accessed 2024-05-09]
23. Van Rossum G, Drake F. Python 3 Reference Manual. Scotts Valley, CA: CreateSpace; 2009.
24. Sarstedt M, Cheah J. Partial least squares structural equation modeling using SmartPLS: a software review. J Market Anal 2019 Jun 27;7(3):196-202. [doi: 10.1057/s41270-019-00058-3]
25. Boone H, Boone D. Analyzing Likert data. JOE 2012 Apr 1;50(2):48. [doi: 10.34068/joe.50.02.48]
26. Hair Jr JF, Hult GT, Ringle CM, Sarstedt M. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). In: A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). Thousand Oaks, CA: SAGE Publications; 2013:211-213.
27. Vinzi V, Chin W, Henseler J, Wang H. Handbook of Partial Least Squares. Berlin: Springer; 2010.
28. Gliem JA, Gliem RR. Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. 2003 Presented at: Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education; 2003; Columbus, OH.
29. Henseler J, Ringle CM, Sarstedt M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. J Acad Mark Sci 2014 Aug 22;43(1):115-135. [doi: 10.1007/s11747-014-0403-8]
30. DePuy V, Berger V, Zhou Y. Wilcoxon-Mann-Whitney test: overview. In: Wiley StatsRef: Statistics Reference Online. Hoboken, NJ: John Wiley & Sons; 2014:14.
31. Grech V, Calleja N. WASP (write a scientific paper): parametric vs. non-parametric tests. Early Hum Dev 2018 Aug;123:48-49. [doi: 10.1016/j.earlhumdev.2018.04.014] [Medline: 29678516]
32. Hendrickson L, Cronbach LJ. Designing evaluations of educational and social programs. In: Contemporary Sociology. San Francisco, CA: Jossey-Bass; 1983:309.
33. Mukaka MM. Statistics corner: a guide to appropriate use of correlation coefficient in medical research. Malawi Med J 2012 Sep;24(3):69-71 [FREE Full text] [Medline: 23638278]
34. Bollen K, Pearl J. Eight myths about causality and structural equation models. In: Handbook of Causal Analysis for Social Research. Dordrecht: Springer Netherlands; 2013:301-328.
35. Jung SY, Hwang H, Lee K, Lee H, Kim E, Kim M, et al. Barriers and facilitators to implementation of medication decision support systems in electronic medical records: mixed methods approach based on structural equation modeling and qualitative analysis. JMIR Med Inform 2020 Jul 22;8(7):e18758 [FREE Full text] [doi: 10.2196/18758] [Medline: 32706717]
36. Huang P, Chen H, Weng L. A penalized likelihood method for structural equation modeling. Psychometrika 2017 Jun 17;82(2):329-354. [doi: 10.1007/s11336-017-9566-9] [Medline: 28417228]
37. Jacobucci R, Grimm KJ, McArdle JJ. Regularized structural equation modeling. Struct Equ Model 2016 Apr 12;23(4):555-566 [FREE Full text] [doi: 10.1080/10705511.2016.1154793] [Medline: 27398019]
38. Hair JF, Risher JJ, Sarstedt M, Ringle CM. When to use and how to report the results of PLS-SEM. Eur Bus Rev 2019 Jan 14;31(1):2-24. [doi: 10.1108/ebr-11-2018-0203]
39. Rosseel Y. Small sample solutions for structural equation modeling. In: Small Sample Size Solutions: A Guide for Applied Researchers and Practitioners. Abingdon-on-Thames, UK: Routledge; 2020:226-238.
40. Pande T, Saravu K, Temesgen Z, Seyoum A, Rai S, Rao R, et al. Evaluating clinicians' user experience and acceptability of LearnTB, a smartphone application for tuberculosis in India. Mhealth 2017 Jul 27;3:30 [FREE Full text] [doi: 10.21037/mhealth.2017.07.01] [Medline: 28828377]
41. Falk R, Miller N. A Primer for Soft Modeling. Akron, OH: University of Akron Press; 1992.

42. Chin WW. The partial least squares approach to structural equation modeling. In: Marcoulides GA, editor. Modern Methods for Business Research. Mahwah, NJ: Lawrence Erlbaum Associates; 1998:295-336.
43. Hair JF, Ringle CM, Sarstedt M. PLS-SEM: indeed a silver bullet. J Mark Theory Pract 2014 Dec 08;19(2):139-152. [doi: 10.2753/mtp1069-6679190202]
44. Mohamed Z, Ubaidullah N, Yusof S. An evaluation of structural model for independent learning through connectivism theory and Web 2.0 towards students' achievement. 2018 Presented at: ICASE 2018: International Conference on Applied Science and Engineering; October 6-7, 2018; Sukoharjo, Indonesia p. 1-5. [doi: 10.2991/icase-18.2018.1]

Abbreviations

ANC: antenatal care
AVE: average variance extracted
CDSS: clinical decision support system
CG: clinical guideline
CP: clinical pathway
FDRE-MOH: Federal Democratic Republic of Ethiopia Ministry of Health
HTMT: heterotrait-monotrait ratio of correlations
LRS: low-resource setting
mHealth: mobile health
MS: measured symptom
PLS: partial least squares
PLS-SEM: partial least squares structural equation modeling
PNC: postnatal care
SEM: structural equation modeling

Edited by A Kushniruk; submitted 28.03.23; peer-reviewed by M Thomson, N Diouf; comments to author 23.07.23; revised version received 11.09.23; accepted 20.01.24; published 11.06.24.

Please cite as:
Tegenaw GS, Sori DA, Teklemariam GK, Verbeke F, Cornelis J, Jansen B
Evaluation of a Computer-Aided Clinical Decision Support System for Point-of-Care Use in Low-Resource Primary Care Settings: Acceptability Evaluation Study
JMIR Hum Factors 2024;11:e47631
URL: https://humanfactors.jmir.org/2024/1/e47631
doi:10.2196/47631
PMID:38861298

Original Paper

Toward Personalized Care and Patient Empowerment and Perspectives on a Personal Health Record in Hemophilia Care: Qualitative Interview Study

Martijn R Brands[1,2], MD; Lotte Haverman[2,3], PhD; Jelmer J Muis[2,3], MSc; Mariëtte H E Driessens[4], PhD; Stephan Meijer[4], MSc; Felix J M van der Meer[5,6], MD, PhD; Marianne de Jong[7], MSc; Johanna G van der Bom[8], MD, PhD, Prof Dr; Marjon H Cnossen[9], MD, PhD; Karin Fijnvandraat[1,2,10], MD, PhD, Prof Dr; Samantha C Gouw[1,2,8], MD, PhD

[1]Department of Pediatric Hematology, Emma Children's Hospital, Amsterdam UMC location University of Amsterdam, Amsterdam, Netherlands
[2]Amsterdam Reproduction & Development, Public Health, Amsterdam UMC location University of Amsterdam, Amsterdam, Netherlands
[3]Child and Adolescent Psychiatry & Psychological Care, Emma Children's Hospital, Amsterdam UMC location University of Amsterdam, Amsterdam, Netherlands
[4]Netherlands Hemophilia Patient Society, Nijkerk, Netherlands
[5]HemoNED Foundation, Leiden, Netherlands
[6]Department of Thrombosis and Hemostasis, Leiden University Medical Center, Leiden University, Leiden, Netherlands
[7]Independent patient representative, Utrecht, Netherlands
[8]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden University, Leiden, Netherlands
[9]Department of Pediatric Hematology and Oncology, Sophia Children's Hospital, Erasmus Medical Center, Rotterdam, Netherlands
[10]Department of Molecular Cellular Hemostasis, Sanquin Research and Landsteiner Laboratory, Amsterdam, Netherlands

Corresponding Author:
Samantha C Gouw, MD, PhD
Department of Pediatric Hematology, Emma Children's Hospital
Amsterdam UMC location University of Amsterdam
Meibergdreef 9
1105 AZ, Amsterdam
Netherlands
Phone: 31 020 566 2727
Email: s.c.gouw@amsterdamumc.nl

Abstract

Background: To enable personalized treatment and shared decision-making in chronic care, relevant health information is collected. However, health information is often fragmented across hospital information systems, digital health apps, and questionnaire portals. This also pertains to hemophilia care, in which scattered information hampers integrated care. We intend to co-design a nationwide digital personal health record (PHR) for patients to help manage their health information. For this, user perspectives are crucial.

Objective: This study aims to assess patients' and health care providers' perspectives regarding the use of a PHR in hemophilia care in the Netherlands, required functionalities, and expectations and concerns.

Methods: In this semistructured interview study, 19 pediatric and adult persons with hemophilia, parents, and women with other inherited bleeding disorders, as well as 18 health care providers working within and outside of hemophilia treatment centers, participated. Perspectives of patients and providers were explored separately. To explore requirements, participants were asked to prioritize functionalities.

Results: Participants expected a PHR would increase the transparency of health information, improve patients' understanding of their illness, and help the coordination of care between health care providers and institutions. Prioritized functionalities included the integration of relevant health information and patient-entered data. Formulated expectations and concerns focused on 4 themes: usability, safety, inclusiveness, and implementation. While patients expressed worries over medicalization (ie, more confrontational reminders of their illness), providers were concerned about an increased workload.

Conclusions: People with hemophilia, their parents, and health care providers welcomed the development of a PHR, as they expected it would result in better coordinated care. Formulated expectations and concerns will contribute to the successful development of a PHR for persons with hemophilia, and ultimately, for all persons with a chronic condition.

(JMIR Hum Factors 2024;11:e48359) doi:10.2196/48359

KEYWORDS

hemophilia; telemedicine; health records; personal; decision-making; shared; patient participation

Introduction

Background

Persons with a chronic health condition, such as the bleeding disorder hemophilia, are encouraged to use eHealth tools to enable personalized treatment and shared decision-making. Hemophilia is an X-linked inherited coagulation factor deficiency of factor VIII (in hemophilia A) or factor IX (in hemophilia B), resulting in an increased bleeding tendency. The severity of hemophilia depends on the residual coagulation factor activity: severe (residual factor activity of <1%), moderate-severe (1%-5%), and mild (6%-40%). Especially in severe hemophilia muscle and joint bleeds are common, leading to painful and disabling joint damage and invalidity. Therefore, lifelong treatment and monitoring is necessary [1,2]. The hallmark of treatment for persons with severe hemophilia consists of replacement therapy with either repetitive prophylactic intravenous infusions of factor concentrate or subcutaneous injections of nonfactor replacement therapy. Preferably, prophylactic treatment is self-administered at home. In case of a bleed, intravenous coagulation factor concentrates are administered, either by people themselves or in the hospital, depending on the severity of a bleed and patients' expertise. Therefore, people are required to closely self-monitor bleeds and treatment efficacy and seek advice from a hemophilia treatment center if necessary.

Successful treatment demands an active role of the patient and well-developed self-management skills. Self-management is the ability to manage the clinical, psychosocial, and societal aspects of illness and its care, and its impact on life [3,4]. eHealth tools can help to facilitate this active role [5-9]. eHealth tools that are used by persons with hemophilia in the Netherlands include a digital treatment diary to log medication use and bleeding episodes [10], a questionnaire portal to complete patient-reported outcomes measures before follow-up visits [11], and patient portals offered by different health care institutions. However, since each tools has a separate log-in procedure, usability is suboptimal. Similarly, individuals who are treated in various care institutions are required to use multiple patient portals. There is no data exchange between these portals, which results in scattered information. Especially for older individuals with comorbidities, as seen in hemophilia, this hampers integrated care [12,13]. The exchange of information between these eHealth tools, patient portals, and different health care institutions is suboptimal. This results in a loss of information and many requests for the same information, which may negatively impact patient safety [14].

Personal health records (PHRs) are being developed to resolve these problems and increase patient empowerment [15,16]. In contrast to a patient portal, a PHR's contents are managed and maintained by individuals, not health care institutions, as illustrated by the definitions presented in Textbox 1. Individuals are able to access and manage their health information and share it with authorized family members or caretakers to help them in managing their care. In addition, people can add self-measurements, such as body weight or data from connected wearables, fill out patient-reported outcomes measures, and complete treatment logs. An example of a PHR is presented in Figure 1. PHRs that are used by people with chronic conditions include PatientsKnowBest (CarePoint) [17] and all certified Dutch PHRs [18]. The Dutch government formulated strict health data exchange guidelines and regulations for PHRs [19]. Therefore, Dutch PHRs will ultimately include integrated medical data from all health care providers involved, regardless of their institution.

Because Dutch PHRs mainly focus on prevalent conditions as hypertension and diabetes, these are not well-equipped in providing for chronic conditions with a high level of self-management, such as hemophilia. Persons with hemophilia may especially benefit from using a PHR, because of the number of health care providers involved in their care, the chronic nature and burden of their condition, and the many self-management skills required from them.

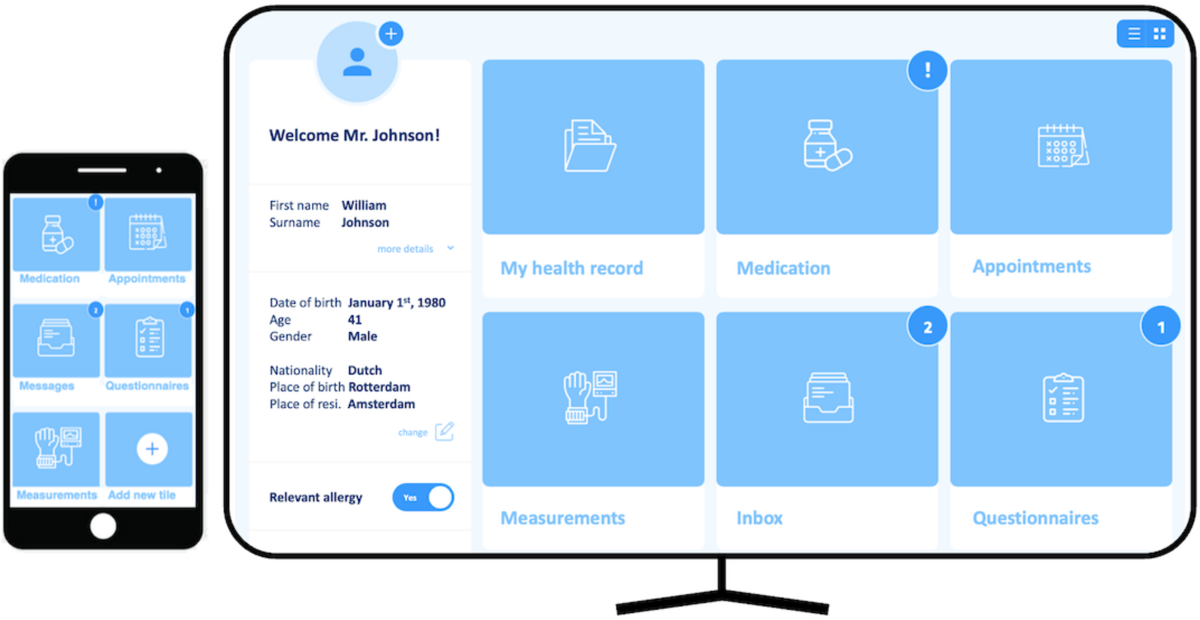
Textbox 1. Definitions.

Contents

- Patient portal: the patient-facing interface of an electronic health record that enables people to view sections of their medical record. This may include access to test results, medication lists, or therapeutic instructions. Health care providers or health care offices determine what health information is accessible for patients. Patient portals often have additional features, such as patient-professional messaging, requesting prescription refills, scheduling appointments, or communicating patient-reported outcomes. By definition, patient portals are "tethered," which refers to a patient portal's connection to an electronic health record.
- Personal health record (PHR): a PHR can have similar features as a patient portal. However, the main difference is that contents are managed and maintained by individuals not health care providers. People can access, manage, and share their health information and that of others for whom they are authorized, such as relatives. Health information from different health care institutions may reside in a single patient-managed PHR. Generally, PHRs are not tethered, with the exception of some, including those currently used in the Netherlands.

Note: The definitions are taken from Brands et al [9].

Figure 1. Example of a personal health record, adapted from MedMij.



Objectives

We intend to co-design a nationwide PHR that meets the needs of persons with hemophilia in the Netherlands, their parents, and health care providers. However, their perceptions on a PHR are not yet known. Therefore, our aim was to assess patients' and health care providers' perspectives regarding the use of a PHR, required functionalities, as well as expectations and concerns. These insights will help to determine whether a PHR is of value in hemophilia care, and will support its further development.

Methods

This is a multicenter, semistructured, qualitative interview study, performed in the Netherlands.

Participants

Study participants were Dutch adolescents and adult persons with hemophilia A and B, parents of young children and adolescents with hemophilia, women with other inherited bleeding disorders, and health care providers working within and outside of hemophilia treatment centers. Adolescents aged 12 to 18 years were interviewed together with a parent. Children <12 years old were represented by the opinions of their parent. Health care providers included: medical specialists, (specialized) hemophilia nurses, allied professionals, and general practitioners (Table 1).

People with hemophilia and parents were recruited through open invitations disseminated by the Dutch Hemophilia Patient Society using their website, email newsletters, and Facebook page. In addition, they were recruited in 2 Dutch hemophilia treatment centers: the Amsterdam UMC and Erasmus MC. We used purposive sampling to include a diverse set of participants. Health care providers were recruited using the direct network of the researchers (MRB and SCG). Inclusion of participants was continued until thematic saturation was achieved, that is, until no new information was introduced in the last 2 interviews. Achieving saturation indicates that additional interviews would not further develop the qualitative theory derived from the data.

Table 1. Participant characteristics.

	Patients[a] (n=19)	Health care providers (n=18)
Sex, n (%)		
Male	14 (74)	2 (13)
Female	5 (26)	16 (87)
Age (y), median (IQR)	39 (18-65)	—[b]
Work experience (y), median (IQR)	—	16 (8-23)
Role or function, n (%)		
Adult	14 (74)	—
Adolescent (aged 12-18 y) interviewed together with a parent	2 (10)	—
Child (aged ≤11 y) represented by the opinion of a parent	3 (16)	—
Pediatric hematologist or hematologist	—	5 (28)
Specialized adult or pediatric hemophilia nurse	—	5 (28)
Pediatric orthopedist	—	1 (6)
Pediatric physiotherapist or physiotherapist	—	2 (11)
Infectiologist	—	1 (6)
Pediatric psychologist	—	1 (6)
Social worker	—	1 (6)
General practitioner	—	1 (6)
Pharmacist	—	1 (6)
Condition, n (%)		
Mild and moderate hemophilia	6 (32)	—
Severe hemophilia	11 (58)	—
Von Willebrand disease type 1	1 (5)	—
Factor VII deficiency	1 (5)	—
Use of prophylaxis, n (%)		
Yes	12 (63)	—
No	7 (37)	—
Comorbidities, n (%)		
Many	6 (32)	—
Some	2 (10)	—
None	11 (58)	—
Digital expertise[c], n (%)		
Proficient	9 (47)	7 (39)
Average	6 (32)	6 (33)
Not proficient	4 (21)	5 (28)
Used teleconsulting, n (%)	5 (26)	10 (56)
Accessed patient portal, n (%)	7 (37)	—
Previous knowledge of PHRs[d], n (%)	7 (37)	9 (50)

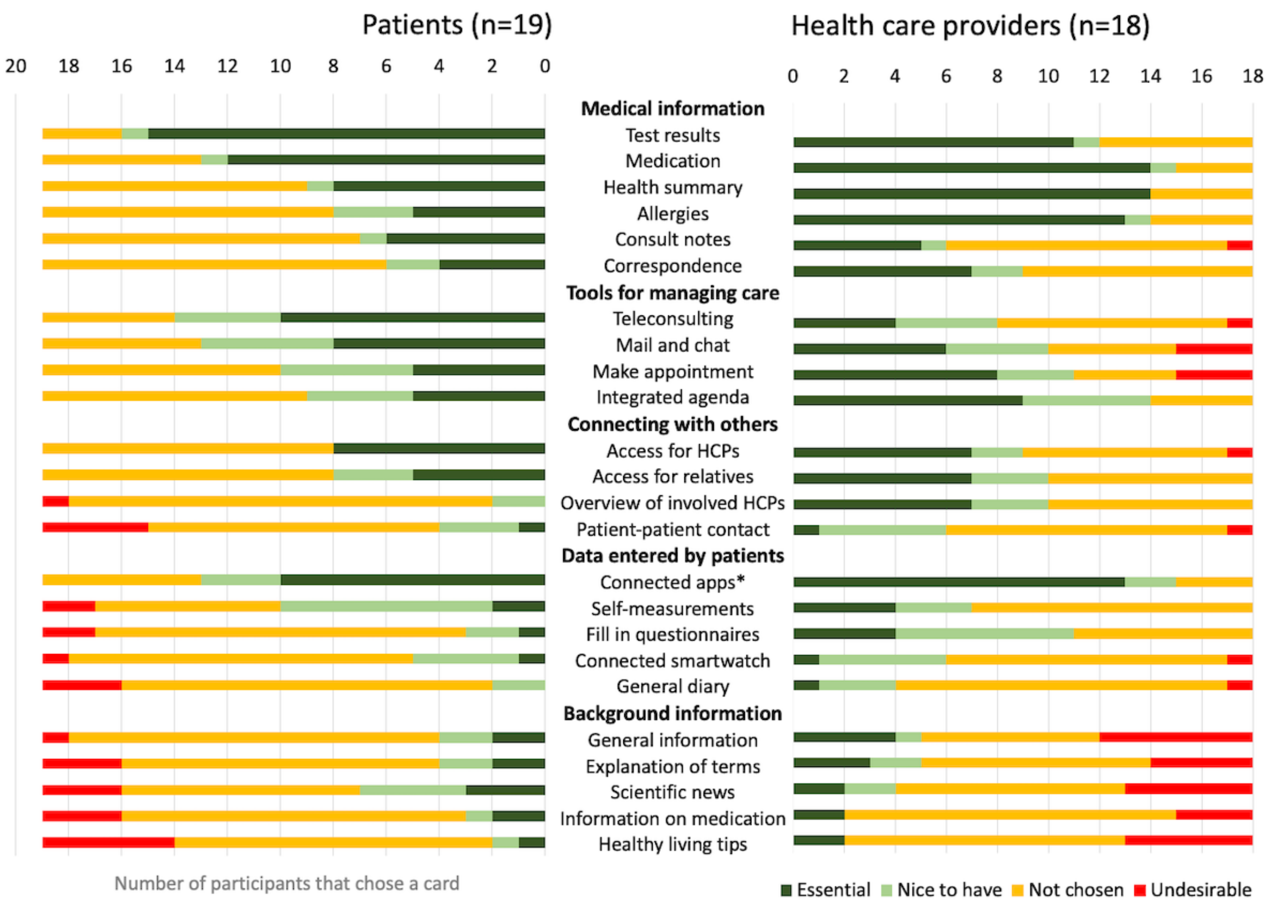
[a]For the 5 parents representing their child, the age and disease characteristics of their child are presented. Among parents, 1 was male.
[b]Not applicable.
[c]A participant's level of digital expertise, as estimated by the interviewers.
[d]PHR: personal health record.

Data Collection

Interviews were conducted by MRB (a physician-researcher) and JJM (a physiotherapist-researcher) between September 1, 2020, and February 1, 2021. Interviews were conducted in person or using video conferencing, the latter due to the COVID-19 pandemic. Duration of interviews ranged from 44 to 96 minutes. All interviews were audiotaped and transcribed verbatim. The item topic list, which was used as an interview guide, was based on clinical expertise and experiences and on previous qualitative research conducted by the authors of this study on the quality of hemophilia care [20]. After discussing relevant personal details and current experiences with patient portals, the concept of a PHR was briefly explained using the image presented in Figure 1. Next, participants were asked about their perspective on the use of a PHR in hemophilia care, required functionalities, as well as expectations and concerns.

Regarding required functionalities, participants first openly discussed their ideas. Second, participants were presented a set of 25 cards, each depicting one functionality. Functionalities were split into 5 categories (Figure 2): medical information, tools for managing care, connecting with others, data entered by patients, and background information. The 25th card stated: “I do not want a PHR.” Cards were based on a qualitative study among 31 older adults with a chronic condition and their caretakers [21]. Cards functioned both as prompts and as an exploration of the most required functionalities. Participants were asked to choose the 5 cards which they considered essential. Third, participants chose the functionalities that they considered nice-to-have and those that they considered undesirable. Finally, participants were asked if they had gained new insights regarding the functionalities.

Figure 2. Functionality cards chosen by participants. *Most participants referred to 2 apps: the digital treatment diary app VastePrik and the patient-reported outcome measure portal KLIK. The 25th card, stating, “I do not want a PHR,” is not depicted; this card was chosen by 1 patient. HCP: health care provider; PHR: personal health record.



Data Analysis

In this qualitative study, a directed form of thematic content analysis was used, because elements of our coding scheme were predetermined by our research question. First, using the qualitative software program MAXQDA (VERBI Software), open coding was used to identify themes in interviews. MRB coded all and JM independently coded a third of the interviews. Next, axial coding was used to find connections, and a coding scheme was drafted by MRB and JJM. Thematic discussions were conducted by MRB and JJM, under the supervision of senior researchers LH and SCG.

Ethical Considerations

The institutional review board of the Amsterdam University Medical Center approved this study (W20_383 # 20.428). All participants signed an informed consent form. Data were deidentified. All participating patients received a gift card worth €20 (US $22) as compensation.

Results

Overview

Thematic saturation was reached at a total of 37 participants: 19 (51%) patients and 18 (49%) health care providers. Characteristics are presented in Table 1. Of the 19 patients, 14 (74%) were male. Patients' median age was 39 (IQR 18-65) years. Furthermore, 11 (58%) had severe hemophilia and 12 (63%) used prophylaxis. In addition, 1 (5%) participant had a factor VII deficiency, and 1 (5%) had von Willebrand disease. Furthermore, 5 (26%) participants were parents of a child with hemophilia. The 18 interviewed health care providers covered all professions involved in the treatment of people with hemophilia. Among them, 7 (39%) worked in adult care, 7 (39%) in pediatric care, and 4 (22%) in both. In total, 16 (89%) were female and median work experience was 16 (IQR 8-23) years.

Perspectives

Overview

In assessing patients' and health care providers' perspectives toward a PHR in hemophilia care, three themes were identified: (1) transparency and ease, (2) understanding and control, and (3) coordination and safety. Statements expressed by patients were either patient-oriented or professional-oriented, that is, related to patients or professionals as the primary beneficiaries of PHRs. The same applied to professionals' statements. Therefore, in exploring themes, and in the matrix in Table S1 in Multimedia Appendix 1, we distinguished patients' and professionals' patient-oriented and professional-oriented views. Quotes are presented in Table S2 in Multimedia Appendix 1.

Transparency and Ease

Virtually all statements on transparency and ease were patient-oriented. Most patients and health care providers expressed a need for improved access by patients to their health information, by integrating health data into one platform. This is illustrated by quote 1 in Table S3 in Multimedia Appendix 1. Most patients and health care providers expected that improved accessibility would also increase transparency in care activities and treatment considerations (quotes 2 and 3). On a different note, some patients and providers questioned what patients would gain from entering more health data in digital tools, especially if they do not experience any symptoms.

Understanding and Control

Patients expressed a desire to gain better, more in-depth understanding of their own health status (quote 4). To achieve this level of understanding, patients said that they require more extensive and detailed health information than is available in the patient portals they were currently using, with additional explanations on what this information means. Such data availability would enable patients to verify if health care providers have understood their symptoms and considerations correctly, and to learn what information is documented (quote 5). In contrast, many patients worried that using a PHR would lead to constant reminders of their illness, and a disturbing confrontation with their illness (quote 6). Several health care providers expressed a similar concern. Finally, many patients expressed a desire to inform health care providers more completely on the impact of illness on their lives, by communicating this through a PHR (quote 7/8).

Health care providers mainly expected that a PHR would improve patients' insight and self-management skills (quote 9/10). However, many questioned whether sufficient patients would use a PHR, since it requires considerable digital expertise, health literacy, and time investment (quote 11). They observed that many patients currently struggle with the interpretation of medical information in patient portals. Eventually, if more health information would become available in a PHR, this could become even more challenging. And would people who use little medication or have few health appointments find it useful? In contrast, nearly all health care providers would strongly welcome a complete overview of medication and other relevant health data (quote 12). Furthermore, more high-quality information entered by patients would help them to better understand the overall impact of illnesses. However, some health care providers perceived PHRs as a burden that increases their workload and responsibility: would they be expected to verify every data entry (quote 13)?

Coordination and Safety

Patients and health care providers elaborated broadly on the final theme: coordination of care and safety. Coordination of care refers to the process of organizing care activities among 2 or more persons involved in a patient's care, often, including the patient. Consequently, patient- and professional-oriented statements overlap. All participants anticipated a PHR would improve coordination of care (quote14/20). Yet, patients stated that a semileading position in the coordination of their care is already expected from them. Three examples include transferring health information from one health care provider to the next, notifying health care providers in case of particularities, and most importantly, instructing those health care providers who are unfamiliar in treating bleeding disorders what to do in case of emergency situations (quote 16). Yet, most patients feel they have insufficient tools and skills to successfully fulfill this role. They expected that accessing and sharing their medical data would enable them to feel more in control (quote 15). Patients hypothesized it would also reduce preventable mistakes related to allergies, contraindications, and medical history. This view was shared by a minority of health care providers (quote 21).

All participants strongly emphasized the need to improve interprofessional collaborations between health care providers working in different organizations and with different backgrounds (quote 17/22). Still, multiple health care providers questioned whether patients should be able to partially control information flows between health care providers. They feared it is likely that patients will not fully understand all information, and do not always feel proficient to determine who should access their data (quote 18). If patients share all medical information with their care providers, it may overwhelm providers with irrelevant details. However, sharing too little could also result in harmful situations if important information is withheld (quote 19). Finally, some health care providers reasoned that a PHR could result in a more rapid diagnosis of bleeding disorders, especially in female patients, because bleeding symptoms are generally presented to multiple health care providers.

Required Functionalities

Before presenting functionality cards, participants frequently discussed the following functionalities: an integrated summary of relevant health data and medication, requesting prescription refills, and a complete overview of care appointments and appointment planning.

An overview of prioritized functionality cards is presented in Figure 2. Nearly all participants expected a PHR to contain a complete summary of relevant health information from multiple institutions, including test results, medication, or allergies. Second, integrated access to communication tools, including teleconsulting and chat was considered either essential or nice-to-have by 76% (28/37) of the participants. For functionalities related to appointment making and viewing, this was 78% (29/37). Yet, many health care providers remarked that patients may not have complete freedom in planning their appointments, to maintain some control as a health care provider. Third, 76% (28/37) of participants considered integrated access to commonly used health apps to be either essential or nice-to-have. Participants mainly referred to 2 health apps: a digital treatment diary and a questionnaire portal. They explained that by choosing the card "connected apps," they implicitly also chose the cards "self-measurements" and "questionnaires." Participants who did not chose the card "connected apps" were all current nonusers of these apps. Finally, some functionalities were considered less beneficial or unwanted: contacting fellow patients, background information, and scientific updates. One patient chose the card "I do not want a PHR," because he did not consider himself to be ill and had low health care use. After discussing functionality cards, 3 additional nice-to-have functionalities were brought up: accessing a PHR from abroad, updates on novel treatment options, and explanations of medical terminology.

Expectations and Concerns

Overview

In analyzing patients' and health care providers' expectations and concerns, 4 themes were identified: (1) usability, (2) safety, (3) inclusiveness and interpretation, and (4) implementation. The theme matrix is shown in Table S3 in Multimedia Appendix 1 and quotes are presented in Table S4 in Multimedia Appendix 1.

Usability

All patients agreed that a PHR must be simple, easy to use, and easily accessible. Many patients stressed on having a low tolerance for crashes and bugs (quote 23). Some expressed usability concerns if PHRs were to have many functionalities. Finally, some said that health care providers working outside of hemophilia treatment centers must also be able to access PHRs, to facilitate shared care.

Health care providers agreed with the necessity of a simple layout (quote 24). They added that an integrated PHR should not only facilitate bleeding disorder care, but all care. Many health care providers stressed that their private work notes must stay private, as is currently the case (quote 25). Finally, health care providers were deeply concerned with an increased workload if they would frequently have to access yet another application with additional health information in the limited time they have (quote 26). Quickly opening PHRs and overseeing all relevant information should meet high usability standards, to not hamper daily practice.

Safety

Safety was often mentioned in 2 distinct manners: safety in terms of privacy and in terms of reliable delivery of care. Related to its first meaning, nearly all patients and health care providers expected safety precautions to prevent data breaches (quote 27). As potential culprits, hackers, "big tech" companies, health care insurers, and "big pharma" companies were suggested (quote 30). The latter 2 terms were mostly mentioned by older patients. In addition, many patients insisted to personally authorize every person that may wish to access their PHR.

The second meaning of safety related to the reliable delivery of care. If a PHR were to become an important source of medical information for both patients and health care providers, many stressed it must always be up-to-date and can never go offline (quote 28). Several participants worried that PHRs could delay the delivery of care if patients must approve all information exchanges. Many health care providers worried that patients will not call their health care providers about emergency problems, but use a PHR's chat function, resulting in a potentially harmful patient and doctor's delay. Finally, several care providers warned that "bad input" results in "bad output" (quote 29/31). If nonvalidated, inaccurate tools are used to upload self-measurements, treatment plans will be based on incorrect information causing safety risks.

Inclusiveness and Interpretation

Many patients and health care providers said that using a PHR should remain optional and may never substitute face-to-face contacts (quote 32). They warned that many patients are likely to experience difficulties using PHRs, including people with low (health) literacy, low digital proficiency, or language barriers (quote 34). Concerns were raised about an increasingly widening "health gap," by focusing on improving care for those patients who already have good access to care services and leaving behind those who do not. Still, even digital-minded, literate patients may struggle with interpreting and overseeing medical data (quote 33/35). Multiple health care providers referred to "data flooding"—patients who are overwhelmed by the large amount of data, making it impossible for them to determine what is relevant (quote 36). Some suggested this could be resolved by only including essential information in a PHR.

Implementation

Many health care providers and some patients mentioned that the complete integration of existing health apps and hospital information systems is essential for the success of PHRs but is challenging (quote 37/39/40). Will vocabulary and abbreviations be exchangeable and understandable for others? Next, many participants stressed it takes “two to tango”—a patient as well as their health care providers need to work with a PHR for it become useful in clinical care. Only if sufficient patients and health care providers use a PHR, it will become part of the routine delivery of care (quote 38).

Discussion

Principal Findings

In this study, we assessed patients’ and health care providers’ perspectives regarding the use of a PHR in hemophilia care in the Netherlands, required functionalities, as well as expectations and concerns. Overall, patients and health care providers welcomed the development of a PHR. Both groups recognized the dual effects of using PHRs, either regarding patients or health care providers as its primary beneficiaries. A PHR was generally thought to increase transparency through access to integrated health data, facilitate patients’ understanding of their illness, and improve coordination of care. Requested functionalities were mostly related to the integration of relevant medical information, medication, appointments, communication tools, and data entered by patients. Concerns were mostly related to usability, privacy, patients’ difficulties in managing information flows, variations in patients’ (digital) health literacy, interoperability, and insufficient uptake of PHRs. While patients more often expressed their worries over medicalization (ie, more confrontational reminders of their illness), providers were concerned about their responsibility for all data entered by patients and a potential increase in workload.

Comparison With Earlier Evidence

The perspectives, required functionalities, and expectations and concerns regarding a PHR or patient portal have not yet been studied among patients with congenital bleeding disorders. More research has been done in other populations.

Perspectives

Perspectives toward PHRs and patient portals were assessed by several studies. Perspectives reflect attitudes that define why people consider (or do not consider) using a PHR or patient portal and reveal their motivation and underlying goals. A Flemish questionnaire study assessed perspectives toward a patient portal among the general population [22]. Similar to our study, they found that increased transparency, understanding, and shared decision-making were anticipated. Unlike our study, PHRs’ warning functions for deteriorating health were often mentioned. This might be because patients with a congenital bleeding disorder are already well-educated in self-monitoring treatment effects. Three other studies mainly identified improved patient understanding as the most anticipated benefit of PHRs: a US mixed methods study among older adult patients [23], a US survey among the general population [24], and a German questionnaire study among persons with psoriasis and their health care providers [25]. The latter study also added interprofessional cooperation as an anticipated benefit [25]. Several reviews on patient portals and PHRs revealed a similar trend, in which transparency [26,27] and understanding and control [26-30], and to a lesser extent, coordination, were often stated as motivations for its use [26,30]. Thus, for the greater part, attitudes overlap, although disease-specific differences do occur.

Study participants expected that a PHR will improve the coordination of care. Multiple health care providers questioned whether patients can (partially) control information flows between health care providers due to their limited comprehension of medical concepts. However, patients stated that they are already expected to adopt a semileading position in the coordination of their care. Therefore, with the appropriate support, a PHR could enable patients to better fulfill this semileading role. However, a PHR is not a panacea for all challenges in the delivery of care. A PHR will only improve physician’s delay if health care providers work together more closely. Similarly, by facilitating the exchange of health data, a PHR will not necessarily change the way that health care providers collaborate. Changes in the organization and delivery of care likely need to be made for PHRs to reach their full potential.

Finally, although the term “patient empowerment” was not expressed by patients nor health care providers, this concept occupies a central role in discussing PHRs’ benefits. Empowerment refers to being able to think critically and make autonomous, informed decisions [31-33]. Contrastingly, the concept “self-management” was often mentioned. It can generally be said that, once empowered, patients perform more self-management behavior [4]. In previous studies, the effects of using PHRs and patient portals on self-management ranged from inconclusive [34-38] to beneficial [39-44]. However, few to none of these PHRs truly enabled patients to collect, manage, and share their health information. Therefore, more research is needed to evaluate PHRs while they maturate.

Required Functionalities

Most required functionalities identified in this study largely matched those identified in previous studies, although 3 were more prominently listed in other studies: requesting prescription refills [22], viewing clinician contact information [23], and proxy-access for a child or parent [24]. This may be because hemophilia treatment center contact information is already well-known among people with a severe congenital bleeding disorder due to long-term treatment relationships and centralized care. Furthermore, adequate alternatives for requesting prescription refills are already available. Finally, proxy-access was considered essential in a specific subgroup of parents and older adult patients in our study.

Expectations and Concerns

In previous studies, the frequently mentioned concerns were usability [16,23,24,26,28,29,34,42-48], privacy and safety [16,23,24,26,27,29,34,42-46,49], inclusiveness [23,27-29,42,43,46,48,50], comprehension of health data [26,29,49], and increased professional workload [16,25,29,48]. Taken together, PHRs are widely expected to be highly user-friendly, understandable, nonobligatory, secure, and inclusive for all patients. Interoperability issues have not previously been reported. However, it is expected that in the rapidly expanding landscape of digital health tools, this concern will become more omnipresent due to the increasing fragmentation of information.

Strengths and Limitations

A strength of this study is that by identifying both patients' and health care providers' perspectives, we were able to assess similarities and differences. Moreover, by using multiple interviewing techniques, such as prompts, participants were stimulated to verbalize their thoughts. Yet, several limitations should be considered. First, by explaining the concept PHR and by using functionality cards, participants may have been influenced. We aimed to limit this risk through the standardization of explanations, and by only presenting cards after we had openly discussed functionalities. Second, by initially including participants through open advertisements and due to the COVID-19 pandemic that forced us to conduct part of the interviews using video conferencing, more digital-minded patients may have been included. Still, by actively including patients from outpatient clinics using purposeful sampling at a later phase, we aimed to minimize such selection. Third, although the gender distribution among the included health care providers represents the current workforce, it is likely that women have different perceptions on a PHR, based on earlier research on gender differences in electronic health record perceptions, although its exact implications remain uncertain [51,52]. Finally, an important limitation is that perspectives and expectations regarding a hypothetical tool are difficult to assess. It may be questioned whether participants truly know what it is that they want and need.

Implications

The greatest challenge in the implementation of PHRs may prove to be the fragmentation of the digital health landscape and its limited interconnectivity. PHRs may not reach their full potential until different hospital information systems and eHealth tools can be adequately connected. Several national and international initiatives have been established that aim to standardize health information technology [53-56]. Until these programs have yielded effects, we must proactively investigate how to promote inclusiveness of PHRs, and safeguard users' privacy. This can only be achieved by involving both patients and health care providers in the development of PHRs to acknowledge their different preferences and to motivate both. In doing so, we would advocate focusing on a PHR intended for a broader population of people with chronic conditions and not only for hemophilia, owing to considerable overlap in expectations and requirements. With hemophilia as a representative use case, these insights may aid the development of a PHR valuable for people with any chronic disease with a high level of self-management.

Conclusions

People with hemophilia, their parents, and health care providers welcomed the development of a PHR in hemophilia care and expressed positive attitudes regarding its use. The benefits of using a PHR on transparency, patients' understanding of their illness, and the improved coordination of care were widely anticipated. However, concerns regarding usability, privacy, inclusiveness, and interoperability need to be taken into account.

Acknowledgments

The authors thank all participants for the interviews. This project was supported by grant NWA.1160.18.038 (principal investigator: MHC) of the Netherlands Organization for Scientific Research (NWO) for the SYMPHONY consortium. The SYMPHONY consortium aims to orchestrate personalized treatment in patients with bleeding disorders. Beneficiaries of the SYMPHONY consortium are Bayer BV, CSL Behring BV, and Swedish Orphan Biovitrum (Belgium) BVBA/SPRL.

Data Availability

The data sets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

MRB conducted 34 interviews and coded all interviews. JJM conducted 3 and coded 12 interviews. Thematic discussions were carried out by MRB, JJM, LH, and SCG. MRB and SCG drafted the first version of the manuscript. All authors provided feedback on the manuscript and approved the final manuscript. No generative artificial intelligence was used in writing this manuscript.

Conflicts of Interest

The institution of MHC has received investigator-initiated research and travel grants as well as speaker fees over the years from Netherlands Organization for Scientific Research (NWO) and Netherlands National research Agenda (NWA), the Netherlands Organization for Health Research and Development (ZonMw), the Dutch Innovatiefonds Zorgverzekeraars, Baxter, Baxalta, Shire, and Takeda, Pfizer, Bayer Schering Pharma, CSL Behring, Sobi Biogen, Novo Nordisk, Novartis, and Nordic Pharma, and for serving as a steering board member for Roche, Bayer, and Novartis, for which fees go to the Erasmus MC as an institution. The institution of KF received unrestricted research grants from CSL Behring and Novo Nordisk and her institution received consultancy fees from SOBI, Grifols, Takeda, Novo Nordisk, and Roche. SCG has received an unrestricted research grant from SOBI. The HemoNED Foundation received a grant or research support from Bayer, CSL Behring, Novo Nordisk, Octapharma, Pfizer, Roche, Sobi, and Takeda. Other authors have no relevant disclosures to report.

Multimedia Appendix 1
Supplementary tables.
[DOCX File , 57 KB - humanfactors_v11i1e48359_app1.docx]

References

1. Leebeek FW, Fischer K. Quality of haemophilia care in The Netherlands: new standards for optimal care. Blood Transfus 2014 Apr;12 Suppl 3(Suppl 3):s501-s504 [FREE Full text] [doi: 10.2450/2014.0041-14s] [Medline: 24922288]
2. Mannucci PM. Hemophilia therapy: the future has begun. Haematologica 2020 Mar;105(3):545-553 [FREE Full text] [doi: 10.3324/haematol.2019.232132] [Medline: 32060150]
3. Schulman-Green D, Jaser S, Martin F, Alonzo A, Grey M, McCorkle R, et al. Processes of self-management in chronic illness. J Nurs Scholarsh 2012 Jun;44(2):136-144 [FREE Full text] [doi: 10.1111/j.1547-5069.2012.01444.x] [Medline: 22551013]
4. Hickmann E, Richter P, Schlieter H. All together now - patient engagement, patient empowerment, and associated terms in personal healthcare. BMC Health Serv Res 2022 Sep 02;22(1):1116 [FREE Full text] [doi: 10.1186/s12913-022-08501-5] [Medline: 36056354]
5. Lancaster K, Abuzour A, Khaira M, Mathers A, Chan A, Bui V, et al. The use and effects of electronic health tools for patient self-monitoring and reporting of outcomes following medication use: systematic review. J Med Internet Res 2018 Dec 18;20(12):e294 [FREE Full text] [doi: 10.2196/jmir.9284] [Medline: 30563822]
6. Risling T, Martinez J, Young J, Thorp-Froslie N. Evaluating patient empowerment in association with eHealth technology: scoping review. J Med Internet Res 2017 Sep 29;19(9):e329 [FREE Full text] [doi: 10.2196/jmir.7809] [Medline: 28963090]
7. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. eHealth for patient engagement: a systematic review. Front Psychol 2016 Jan 08;6:2013 [FREE Full text] [doi: 10.3389/fpsyg.2015.02013] [Medline: 26779108]
8. Morton K, Dennison L, May C, Murray E, Little P, McManus RJ, et al. Using digital interventions for self-management of chronic physical health conditions: a meta-ethnography review of published studies. Patient Educ Couns 2017 Apr;100(4):616-635 [FREE Full text] [doi: 10.1016/j.pec.2016.10.019] [Medline: 28029572]
9. Brands MR, Gouw SC, Beestrum M, Cronin RM, Fijnvandraat K, Badawy SM. Patient-centered digital health records and their effects on health outcomes: systematic review. J Med Internet Res 2022 Dec 22;24(12):e43086. [doi: 10.2196/43086] [Medline: 36548034]
10. VastePrik. Stichting HemoNED. URL: https://hemoned.nl/vasteprik/ [accessed 2024-08-09]
11. Haverman L, Engelen V, van Rossum MA, Heymans HS, Grootenhuis MA. Monitoring health-related quality of life in paediatric practice: development of an innovative web-based application. BMC Pediatr 2011 Jan 12;11:3 [FREE Full text] [doi: 10.1186/1471-2431-11-3] [Medline: 21226913]
12. eHealth Monitor 2019 Report. Nictiz. 2019. URL: https://nictiz.nl/publicaties/ehealth-monitor-2019-rapport/ [accessed 2024-08-09]
13. Smith C. This is the personal website of CEES SMIT. Cees Smith Testimonial. 2015. URL: https://smitvisch.nl/ [accessed 2024-08-12]
14. Inbreng commissiedebat digitale ontwikkelingen in de zorg. Federatie Medisch Specialisten. 2019. URL: https://demedischspecialist.nl/sites/default/files/2023-06/Brief_inbreng_CD_digitale_ontwikkelingen.pdf [accessed 2024-08-14]
15. Roehrs A, da Costa CA, da Rosa Righi R, de Oliveira KS. Personal health records: a systematic literature review. J Med Internet Res 2017 Jan 06;19(1):e13 [FREE Full text] [doi: 10.2196/jmir.5876] [Medline: 28062391]
16. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. J Med Internet Res 2015 Feb 10;17(2):e44 [FREE Full text] [doi: 10.2196/jmir.3171] [Medline: 25669240]
17. Bidmead E, Marshall A. A case study of stakeholder perceptions of patient held records: the Patients Know Best (PKB) solution. Digit Health 2016 Sep 21;2:2055207616668431 [FREE Full text] [doi: 10.1177/2055207616668431] [Medline: 29942567]
18. Choose your personal health record (in Dutch). DigitalCareGuide. 2020. URL: https://www.digitalezorggids.nl/ [accessed 2024-08-19]
19. Wegiz: wet elektronische gegevensuitwisseling in de zorg. Nictiz. URL: https://nictiz.nl/wat-we-doen/programmas/werken-aan-wegiz/#:~:text=Wat%20is%20de%20rol%20van,moeten%20worden%20voor%20het%20wetstraject [accessed 2024-08-09]
20. Brands MR, Haverman L, Muis JJ, Driessens MH, van der Meer FJ, Goedhart G, et al. Patients' and health care providers' perspectives on quality of hemophilia care in the Netherlands: a questionnaire and interview study. Res Pract Thromb Haemost 2023 Apr 23;7(4):100159 [FREE Full text] [doi: 10.1016/j.rpth.2023.100159] [Medline: 37346462]
21. de Kam I, Meijnckens L, Vroonland E. Wat vinden ouderen en mantelzorgers van persoonlijke gezondheidsomgevingen (PGO’s)? ParticipatieKompas. 2020. URL: https://participatiekompas.nl/media/rapportage_proeftuin_inzicht_utrecht-ouderenparticipatie_pgo-_20200604.pdf [accessed 2024-03-05]
22. Van den Bulck SA, Hermens R, Slegers K, Vandenberghe B, Goderis G, Vankrunkelsven P. Designing a patient portal for patient-centered care: cross-sectional survey. J Med Internet Res 2018 Oct 01;20(10):e269 [FREE Full text] [doi: 10.2196/jmir.9497] [Medline: 30287416]

23. Irizarry T, Shoemake J, Nilsen ML, Czaja S, Beach S, DeVito Dabbs A. Patient portals as a tool for health care engagement: a mixed-method study of older adults with varying levels of health literacy and prior patient portal use. J Med Internet Res 2017 Mar 30;19(3):e99 [FREE Full text] [doi: 10.2196/jmir.7099] [Medline: 28360022]
24. Patel VN, Abramson E, Edwards AM, Cheung MA, Dhopeshwarkar RV, Kaushal R. Consumer attitudes toward personal health records in a beacon community. Am J Manag Care 2011 Apr;17(4):e104-e120 [FREE Full text] [Medline: 21774099]
25. Klein TM, Augustin M, Kirsten N, Otten M. Attitudes towards using electronic health records of patients with psoriasis and dermatologists: a cross-sectional study. BMC Med Inform Decis Mak 2020 Dec 30;20(1):344 [FREE Full text] [doi: 10.1186/s12911-020-01302-y] [Medline: 33380329]
26. Benjamins J, Haveman-Nies A, Gunnink M, Goudkuil A, de Vet E. How the use of a patient-accessible health record contributes to patient-centered care: scoping review. J Med Internet Res 2021 Jan 11;23(1):e17655 [FREE Full text] [doi: 10.2196/17655] [Medline: 33427683]
27. Carini E, Villani L, Pezzullo AM, Gentili A, Barbara A, Ricciardi W, et al. The impact of digital patient portals on health outcomes, system efficiency, and patient attitudes: updated systematic literature review. J Med Internet Res 2021 Sep 08;23(9):e26189 [FREE Full text] [doi: 10.2196/26189] [Medline: 34494966]
28. Damen DJ, Schoonman GG, Maat B, Habibović M, Krahmer E, Pauws S. Patients managing their medical data in personal electronic health records: scoping review. J Med Internet Res 2022 Dec 27;24(12):e37783 [FREE Full text] [doi: 10.2196/37783] [Medline: 36574275]
29. Antonio MG, Petrovskaya O, Lau F. The state of evidence in patient portals: umbrella review. J Med Internet Res 2020 Nov 11;22(11):e23851 [FREE Full text] [doi: 10.2196/23851] [Medline: 33174851]
30. Ruhi U, Chugh R. Utility, value, and benefits of contemporary personal health records: integrative review and conceptual synthesis. J Med Internet Res 2021 Apr 29;23(4):e26877 [FREE Full text] [doi: 10.2196/26877] [Medline: 33866308]
31. Mc Namara KP, Versace VL, Marriott JL, Dunbar JA. Patient engagement strategies used for hypertension and their influence on self-management attributes. Fam Pract 2014 Aug;31(4):437-444. [doi: 10.1093/fampra/cmu026] [Medline: 24895137]
32. Greene J, Hibbard JH, Sacks R, Overton V, Parrotta CD. When patient activation levels change, health outcomes and costs change, too. Health Aff (Millwood) 2015 Mar;34(3):431-437. [doi: 10.1377/hlthaff.2014.0452] [Medline: 25732493]
33. Peters AE, Keeley EC. Patient engagement following acute myocardial infarction and its influence on outcomes. Am J Cardiol 2017 Nov 01;120(9):1467-1471. [doi: 10.1016/j.amjcard.2017.07.037] [Medline: 28843395]
34. Dendere R, Slade C, Burton-Jones A, Sullivan C, Staib A, Janda M. Patient portals facilitating engagement with inpatient electronic medical records: a systematic review. J Med Internet Res 2019 Apr 11;21(4):e12779. [doi: 10.2196/12779] [Medline: 30973347]
35. Calvillo J, Román I, Roa LM. How technology is empowering patients? A literature review. Health Expect 2015 Oct;18(5):643-652 [FREE Full text] [doi: 10.1111/hex.12089] [Medline: 23711169]
36. Tuil WS, Verhaak CM, Braat DD, de Vries Robbé PF, Kremer JA. Empowering patients undergoing in vitro fertilization by providing internet access to medical data. Fertil Steril 2007 Aug;88(2):361-368. [doi: 10.1016/j.fertnstert.2006.11.197] [Medline: 17416366]
37. van der Vaart R, Drossaert CH, Taal E, Drossaers-Bakker KW, Vonkeman HE, van de Laar MA. Impact of patient-accessible electronic medical records in rheumatology: use, satisfaction and effects on empowerment among patients. BMC Musculoskelet Disord 2014 Mar 26;15:102 [FREE Full text] [doi: 10.1186/1471-2474-15-102] [Medline: 24673997]
38. Ancker JS, Osorio SN, Cheriff A, Cole CL, Silver M, Kaushal R. Patient activation and use of an electronic patient portal. Inform Health Soc Care 2015;40(3):254-266. [doi: 10.3109/17538157.2014.908200] [Medline: 24786648]
39. Han HR, Gleason KT, Sun CA, Miller HN, Kang SJ, Chow S, et al. Using patient portals to improve patient outcomes: systematic review. JMIR Hum Factors 2019 Dec 19;6(4):e15038 [FREE Full text] [doi: 10.2196/15038] [Medline: 31855187]
40. Price M, Bellwood P, Kitson N, Davies I, Weber J, Lau F. Conditions potentially sensitive to a personal health record (PHR) intervention, a systematic review. BMC Med Inform Decis Mak 2015 Apr 18;15:32 [FREE Full text] [doi: 10.1186/s12911-015-0159-1] [Medline: 25927384]
41. Crouch PC, Rose CD, Johnson M, Janson SL. A pilot study to evaluate the magnitude of association of the use of electronic personal health records with patient activation and empowerment in HIV-infected veterans. PeerJ 2015 Mar 19;3:e852 [FREE Full text] [doi: 10.7717/peerj.852] [Medline: 25802815]
42. Vreugdenhil MM, Ranke S, de Man Y, Haan MM, Kool RB. Patient and health care provider experiences with a recently introduced patient portal in an academic hospital in the netherlands: mixed methods study. J Med Internet Res 2019 Aug 20;21(8):13743 [FREE Full text] [doi: 10.2196/13743] [Medline: 31432782]
43. Hanna L, Gill SD, Newstead L, Hawkins M, Osborne RH. Patient perspectives on a personally controlled electronic health record used in regional Australia. Health Inf Manag 2017 Jan;46(1):42-48. [doi: 10.1177/1833358316661063] [Medline: 27486184]
44. Fylan F, Caveney L, Cartwright A, Fylan B. Making it work for me: beliefs about making a personal health record relevant and useable. BMC Health Serv Res 2018 Jun 14;18(1):445 [FREE Full text] [doi: 10.1186/s12913-018-3254-z] [Medline: 29898730]

45. Abd-Alrazaq AA, Bewick BM, Farragher T, Gardner P. Factors that affect the use of electronic personal health records among patients: a systematic review. Int J Med Inform 2019 Jun;126:164-175 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.03.014] [Medline: 31029258]
46. Entzeridou E, Markopoulou E, Mollaki V. Public and physician's expectations and ethical concerns about electronic health record: benefits outweigh risks except for information security. Int J Med Inform 2018 Feb;110:98-107. [doi: 10.1016/j.ijmedinf.2017.12.004] [Medline: 29331259]
47. de Vries SG, Bekedam MM, Visser BJ, Stijnis C, van Thiel PP, van Vugt M, et al. Travel-related leptospirosis in the Netherlands 2009-2016: an epidemiological report and case series. Travel Med Infect Dis 2018;24:44-50 [FREE Full text] [doi: 10.1016/j.tmaid.2018.05.002] [Medline: 29753855]
48. Thompson MJ, Reilly JD, Valdez RS. Work system barriers to patient, provider, and caregiver use of personal health records: a systematic review. Appl Ergon 2016 May;54:218-242. [doi: 10.1016/j.apergo.2015.10.010] [Medline: 26851482]
49. Zhao JY, Song B, Anand E, Schwartz D, Panesar M, Jackson GP, et al. Barriers, facilitators, and solutions to optimal patient portal and personal health record use: a systematic review of the literature. AMIA Annu Symp Proc 2018 Apr 16;2017:1913-1922 [FREE Full text] [Medline: 29854263]
50. Hemsley B, Rollo M, Georgiou A, Balandin S, Hill S. The health literacy demands of electronic personal health records (e-PHRs): an integrative review to inform future inclusive research. Patient Educ Couns 2018 Jan;101(1):2-15. [doi: 10.1016/j.pec.2017.07.010] [Medline: 28864052]
51. Rotenstein LS, Fong AS, Jeffery MM, Sinsky CA, Goldstein R, Williams B, et al. Gender differences in time spent on documentation and the electronic health record in a large ambulatory network. JAMA Netw Open 2022 Mar 01;5(3):e223935 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.3935] [Medline: 35323954]
52. Rule A, Shafer CM, Micek MA, Baltus JJ, Sinsky CA, Arndt BG. Gender differences in primary care physicians' electronic health record use over time: an observational study. J Gen Intern Med 2023 May;38(6):1570-1572 [FREE Full text] [doi: 10.1007/s11606-022-07837-2] [Medline: 36474003]
53. MedMij homepage. MedMij. URL: https://medmij.nl/ [accessed 2020-08-01]
54. Kay S. IPS update to CEN/TC 251. The International Patient Summary. URL: https://international-patient-summary.net/ips-update-to-cen-tc-251/ [accessed 2024-08-09]
55. Anumula N, Sanelli PC. Meaningful use. AJNR Am J Neuroradiol 2012 Sep;33(8):1455-1457 [FREE Full text] [doi: 10.3174/ajnr.A3247] [Medline: 22790244]
56. European health data space. European Commission. URL: https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en [accessed 2024-08-09]

Abbreviations

PHR: personal health record

Edited by A Kushniruk; submitted 20.04.23; peer-reviewed by Z Mahmoudvand, E van der Velde, C Smith; comments to author 10.10.23; revised version received 07.11.23; accepted 28.06.24; published 03.09.24.

Please cite as:
Brands MR, Haverman L, Muis JJ, Driessens MHE, Meijer S, van der Meer FJM, de Jong M, van der Bom JG, Cnossen MH, Fijnvandraat K, Gouw SC
Toward Personalized Care and Patient Empowerment and Perspectives on a Personal Health Record in Hemophilia Care: Qualitative Interview Study
JMIR Hum Factors 2024;11:e48359
URL: https://humanfactors.jmir.org/2024/1/e48359
doi:10.2196/48359
PMID:

Original Paper

Evaluating the Usability and Quality of a Clinical Mobile App for Assisting Physicians in Head Computed Tomography Scan Ordering: Mixed Methods Study

Zahra Meidani[1], PhD; Aydine Omidvar[2], MD; Hossein Akbari[3], PhD; Fatemeh Asghari[1], MSc; Reza Khajouei[4], PhD; Zahra Nazemi[1], MSc; Ehsan Nabovati[1], PhD; Felix Holl[5], PhD

[1]Health Information Management Research Center, Kashan University of Medical Sciences, Kashan, Iran

[2]Department of Neurosurgery, Kashan University of Medical Sciences, Kashan, Iran

[3]Department of Epidemiology & Biostatistics, Kashan University of Medical Sciences, Kashan, Iran

[4]Department of Health Information Sciences, Faculty of Management and Medical Information Sciences, Kerman University of Medical Sciences, Kerman, Iran

[5]DigiHealth Institute, Neu-Ulm University of Applied Sciences, Neu-Ulm, Germany

Corresponding Author:
Ehsan Nabovati, PhD
Health Information Management Research Center
Kashan University of Medical Sciences
Qotbe Ravandi, Pezeshk Blvd
Kashan, 8715973449
Iran
Phone: 98 55589343
Email: nabovati@kaums.ac.ir

Abstract

Background: Among the numerous factors contributing to health care providers' engagement with mobile apps, including user characteristics (eg, dexterity, anatomy, and attitude) and mobile features (eg, screen and button size), usability and quality of apps have been introduced as the most influential factors.

Objective: This study aims to investigate the usability and quality of the Head Computed Tomography Scan Appropriateness Criteria (HAC) mobile app for physicians' computed tomography scan ordering.

Methods: Our study design was primarily based on methodological triangulation by using mixed methods research involving quantitative and qualitative think-aloud usability testing, quantitative analysis of the Mobile Apps Rating Scale (MARS) for quality assessment, and debriefing across 3 phases. In total, 16 medical interns participated in quality assessment and testing usability characteristics, including efficiency, effectiveness, learnability, errors, and satisfaction with the HAC app.

Results: The efficiency and effectiveness of the HAC app were deemed satisfactory, with ratings of 97.8% and 96.9%, respectively. MARS assessment scale indicated the overall favorable quality score of the HAC app (82 out of 100). Scoring 4 MARS subscales, Information (73.37 out of 100) and Engagement (73.48 out of 100) had the lowest scores, while Aesthetics had the highest score (87.86 out of 100). Analysis of the items in each MARS subscale revealed that in the Engagement subscale, the lowest score of the HAC app was "customization" (63.6 out of 100). In the Functionality subscale, the HAC app's lowest value was "performance" (67.4 out of 100). Qualitative think-aloud usability testing of the HAC app found notable usability issues grouped into 8 main categories: lack of finger-friendly touch targets, poor search capabilities, input problems, inefficient data presentation and information control, unclear control and confirmation, lack of predictive capabilities, poor assistance and support, and unclear navigation logic.

Conclusions: Evaluating the quality and usability of mobile apps using a mixed methods approach provides valuable information about their functionality and disadvantages. It is highly recommended to embrace a more holistic and mixed methods strategy when evaluating mobile apps, because results from a single method imperfectly reflect trustworthy and reliable information regarding the usability and quality of apps.

(JMIR Hum Factors 2024;11:e55790) doi:10.2196/55790

KEYWORDS

mobile apps; user-centered design; user-computer interface; physicians; tomography; x-ray computed; mobile phone

Introduction

Background

Mobile devices and mobile health (mHealth) apps have equipped the health care system with a strategy to improve health through enhanced self-management among patients and access to educational materials for health care professionals [1]. Considering their advantages regarding the fastest and most convenient ways to access health care services, they have been introduced as effective eHealth technology to address health priorities [2]. Recently, a global initiative has been launched to apply mobile technologies to provide health care services and manage various diseases [3]. A 2015 World Health Organization survey revealed that 15,000 mobile apps were available for health care use [4]. However, the continuity in the use of apps is highly challenging, and existing evidence presented poor user engagement and relatively high drop-out rates on apps among patients and health care providers (HCPs) [5]. Earlier research revealed that nearly half of mHealth app users avoid continuous use of them [6]. The drop-out rates in app-based interventions for chronic diseases were reported to be 43% (95% CI 29%-57%) in a meta-analysis by Meyerowitz-Katz et al [7].

Usability has been introduced as a surrogate marker for app quality and user engagement with them to address this challenge [8-10]. Given the significance, assessing the usability and quality of mobile apps occupies a crucial part of app development and users' overall assessment of app quality [6,8]. However, emerging research has debated that mobile apps suffer from usability and quality issues and are limited by their ability to address users' needs [9,11,12]. Physicians use many mobile apps to access a wide range of knowledge and information in educational materials, drug reference guides, x-ray results, laboratory test information, and clinical guidelines [13]. Medical apps were positively perceived, with physicians reporting increased dependency on the apps. The use of apps in the medical setting has steadily grown in recent years [14]. While a considerable number of physicians now use mobile devices and apps for clinical practices globally [7], there are also reports of drop-out rate and short-term engagement among physicians with these mobile apps [1]. Arguably, no clear understanding exists of the physicians' motivations and interests in adopting and long-term use of mobile apps [8]. A variety of factors, from organizational and social factors [15] to users' characteristics (eg, user dexterity and anatomy and positive attitude) [16-19] or mobile features (eg, screen and buttons size, poor resolution, and usability) [6,20-22], would influence the successful adoption of mobile apps among physicians.

Prior Work

Usability and quality issues have been reported as central to user engagement with mobile apps [6,21]. The research team previously developed a mobile app aimed at assisting physicians in prescribing head computed tomography (CT) scans based on appropriateness guidelines, the Head CT Scan Appropriateness Criteria (HAC) mobile app [1]. However, during that study, neurology and neurosurgery residents expressed concerns about usability issues despite their interest in using the app. Therefore, before proceeding with full implementation, it is essential to identify and address usability problems of the app using mixed methods that involve participation from final users.

Goal

In this study, we seek to investigate the usability of HAC app using mixed methods research involving quantitative analysis of the Mobile Apps Rating Scale (MARS) for quality assessment, quantitative and qualitative think-aloud (TA) usability testing, and debriefing across 3 phases.

Methods

Study Setting

This study was conducted as part of a broader effort to develop a mobile app, known as the HAC app, based on clinical guidelines. The development occurred at an academic hospital with Kashan University of Medical Sciences (KAUMS) in Iran, which has 510 beds. This newly developed HAC app allows end users to search for appropriate CT scans based on diseases, signs, symptoms, and modalities, such as CT, CT angiography (CTA), and MRI. Appropriate CT scans refers to imaging studies that are deemed clinically justified and indicated based on established medical criteria, including patient symptoms, signs, and relevant clinical history, in accordance with evidence-based guidelines and best practices in diagnostic radiology.

The study involved 16 medical interns from an academic hospital at KAUMS. For this study, the focus was on assessing the end-user usability through a TA approach [23,24], evaluating the quality of the HAC app using the MARS [25], and conducting informal debriefing sessions to gather insights and opinions for the medical interns regarding the HAC app.

Profile of the HAC App: HAC App Content and Functionality

The HAC app was developed using applied 4-tier architecture, including presentation, data service, business logic, and data access layers. The app was designed using JavaScript in such a way that allows it to be installed and be compatible with the latest version of Android in 2021 (version 12) as well as earlier versions. The HAC app encompasses essential criteria arranged by Care Core guidelines for head CT scans. Care Core provides a list of disease titles, for example, head trauma, which is supplemented by the list of clinical criteria in terms of signs and symptoms of the given disease. Clicking the plus sign (+) provides the detailed clinical criteria. Under each main heading or in front of each condition, the appropriate imaging procedure in MRI, CT, and CTA is provided. A shortlist menu is designed to organize and quickly find frequently used diseases or clinical criteria. It enables users to add common diagnoses to the shortlist menu. Screenshots of the functionalities of the HAC app are presented in Figures 1 and 2.

Figure 1. Head Computed Tomography Scan Appropriateness Criteria app search results for seizure.

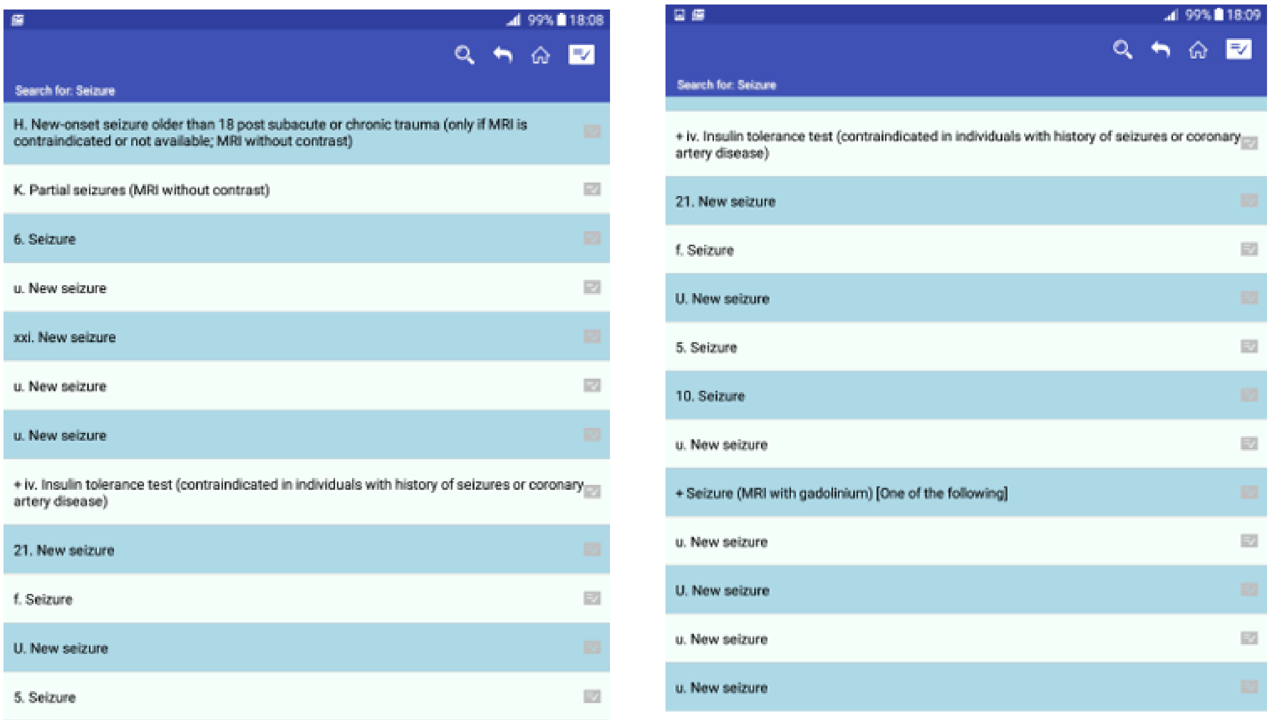
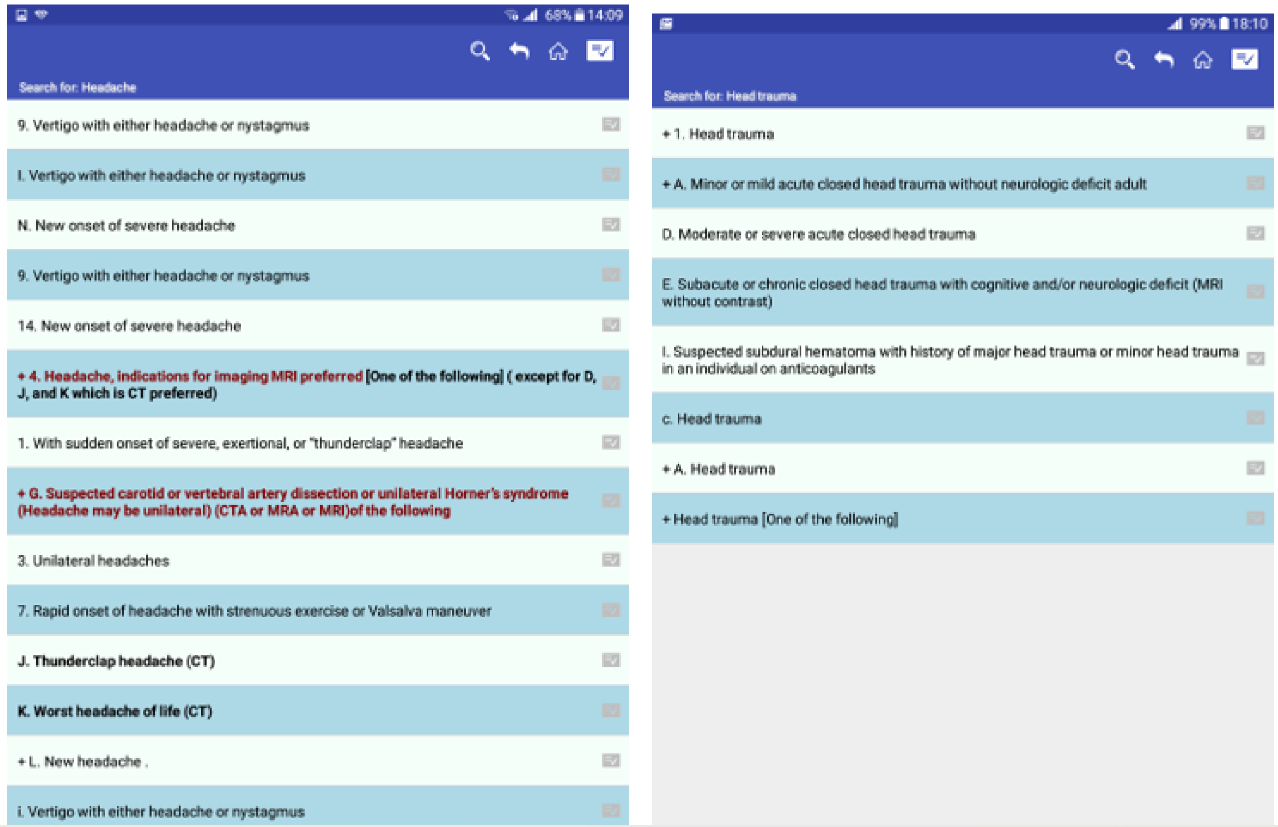


Figure 2. Head Computed Tomography Scan Appropriateness Criteria app search results for head trauma and headache.



Approaches to Conduct the Study

Three approaches have been used to conduct the research and achieve the study objectives.

The TA Usability Testing

In this phase, we tested the HAC app's effectiveness, efficiency, error, and learnability. The study objectives have been determined to ensure the accurate fulfillment of the tasks, the correct selection of the icons and buttons, and end users' use of the mobile app without errors in an efficient way. The TA approach set out to determine the following measures to achieve the objectives:

- The effectiveness of participants' navigation of the app was measured by the accurateness and completeness of the HAC app on CT scan ordering based on diseases, signs, symptoms, and modalities, for example, CT, CTA, and MRI.
- The efficiency of the participants was specified by the number of touch targets on the app screen and the task completion time.
- The simplicity and learnability of the HAC app were measured by the number of tasks that were easily completed and the severity of errors made by the users.
- Errors indicated the number of user mistakes when using the HAC app.

MARS Quality Assessment

To evaluate the HAC app quality in terms of engagement, functionality, Aesthetics, information, and subjective quality, we applied the MARS tool [25], and the following dimensions were addressed:

- the overall quality score of the HAC app and its subscales, including Engagement, Functionality, Aesthetics, Information, and Subjective Quality
- a statistically significant difference between MARS subscales quality score
- a statistically significant difference between 2 sets of pairs of MARS subscales (eg, Engagement and Functionality or Functionality and Aesthetics)
- the correlation between MARS subscales for the HAC app
- the significant relationship between medical interns' characteristics (ie, age, gender, and interest in using mobile apps for learning and clinical practice) with MARS subscales

Debriefing

An informal debrief was conducted to review and digest interns' general ideas about using mobile apps and physicians' expectations for a suitable mobile app. It was also applied to collect underexplored facts for further revision of the HAC app.

Study Design and Data Analysis

Our study design was primarily based on methodological triangulation through the use of mixed methods research, and investigator triangulation to enhance the understanding and interpreting the results [26].

The mixed methods study involved quantitative (MARS quality assessment and TA quantitative usability testing) and qualitative methods (TA qualitative usability testing and debriefing) across 3 phases. By using the technique of investigator triangulation, a variety of researchers, such as medical practitioners, experts in health information technology, and professionals in health information management, were involved in the gathering and analyzing of the data. The details of each phase will be discussed in the subsequent sections.

Phase 1: TA Usability Testing Approach

Design

The study used a TA study design to explore the user's cognition, including feelings, thoughts, and whatever else comes to mind while interacting with a system to perform a task. This standard data collection method for assessing users' cognitive behavior during system interaction helps identify errors and necessary changes [23,24].

There are 2 fundamental usability testing methods: qualitative and quantitative [27]. Qualitative methods primarily aim to explore users' interaction experiences with a product and describe possible issues they encounter [28]. In contrast, the quantitative methods use various metrics, such as task times, completion rates, and errors, to measure and categorize the errors and problems users encounter during usability testing [29]. Both qualitative and quantitative methods were applied in this study to reach the research objectives. Usability evaluation was also conducted in the different stages of a product development life cycle. Formative evaluation was done in the early product development life cycle to shape the design direction. Summative assessment was performed toward the end of the product development (final product) to evaluate its performance against a set of metrics (eg, time on task and success rate) [28]. The participants implemented summative usability testing in this study to evaluate the performance of the HAC app.

Participant Recruitment

Previous evidence confirms that about 5 to 15 participants are sufficient to perform TA to enhance the expected level of problem discovery [30]. We recruited 16 medical interns who participated in 3 phases of the study. We applied social media to attract medical interns to join the study. We posted our research profile on the medical students' academic and social media channels, including the study title, research team, and overall study objectives. We invited those who finished their clinical internship in emergency medicine to participate in this study. Our multidisciplinary research team, including clinicians, significantly streamlined the recruitment process. The research team, consisting of members from diverse disciplines, encouraged their previous students to participate in the research. No rewards or compensations were paid to the participants.

Protocol

The TA usability testing was conducted in multiple sessions of the same activity. Developing a study protocol to ensure the consistency of each activity in each session helped the facilitator give all the necessary information to the participants, which seemed imperative [28]. The study protocol in this study consisted of session introduction, information capture methods (ie, observation and videotaping), task scenarios, user interactions with the product and any identified product problems and difficulties, and measurement criteria, which are discussed in the following sections.

Session Introduction

Interested volunteers were contacted to schedule face-to-face visits. TA sessions were held in the physicians' actual workplace. It is widely believed that evaluations conducted in the field resemble the set-up that matches the user's real work context, providing "ecological validity" to the study and accurately reflecting the users' context [28].

Once the researcher arrived in the field, they gave the participants an overview of the session and the overall goals. They let them know about the presence of any facilitator or observers in the session and the rules for conducting the usability testing.

Ethical Considerations

Overview

This study was approved by the ethics review board at KAUMS (code #IR.KAUMS.MEDNT.REC.1399.075). The participants were informed the same and emphasized the voluntary nature of participation, assuring them of the confidentiality of information. The nonevaluative environment of the TA session was also explained by a trained moderator (researcher). Then, participants who attended the face-to-face meeting consented to participate in the study and TA session run.

Data Collection

The usability data collection protocol was generally implemented via 2 approaches: concurrent TA and retrospective TA protocols [31].

Because concurrent TA is more objective and less dependent on users' memory and prior experience of completed tasks compared with retrospective TA, concurrent TA was adopted as a standardized method to conduct usability testing of the HAC app [28,32].

Considering that most users are uncomfortable installing something on their devices (eg, mobiles or computers) [28] and the importance of using the same tool to capture usability-testing data, interaction with the HAC app was done via an Android mobile phone dedicated only to the research purposes. A portfolio of methods, including video screen-recording software, audio and screen recording, and notetaking, were applied to collect data. Four scenarios containing 4 to 6 tasks were given to the participant to interact with the HAC app. All the activities to accomplish the scenario, including the number of touch targets on the app screen, the task completion time, and the elapsed time, were captured using the free AZ Screen Recorder for Android (AZ Screen Recorder) [33].

Three evaluators facilitated the testing sessions and analyzed the results. The researchers adopted the verbal protocol to collect the data. Although the verbal protocol is the most traditional protocol with limited probing methods compared with active users' participation methods, such as communication - based and coaching protocols, it resembles an authentic context experience by not offering any external assistance to the users [34].

Thus, 1 researcher supervised the evaluation session, but neither user received instruction during the task performance stage. Attention was given to shortening the testing process and keeping a participant on the phone for >10 minutes [28]. Each TA session lasted for about 20 to 30 minutes.

Tasks Scenario

The scenarios (including their goals and actions) were designed to examine different parts and functions of the HAC app and covered the most common tasks that a clinician may use in a typical working application. Usability problems were detected by researchers from analyses of user behavior and expressions during interactions with the system.

Measurement

A coding framework was developed according to 5 usability characteristics and based on the International Organization for Standardization and Nielsen's definitions to recognize the specific user-computer interaction problems in detail to define the measurement criteria [35-37]. Nielsen put forward 5 usability attributes: learnability, efficiency, memorability, errors, and satisfaction [37].

Combining International Organization for Standardization and Nielsen usability attributes yields the following 6 criteria: efficiency, effectiveness, learnability, memorability, errors, and satisfaction. Because the participants only used the HAC app in this study, and there was no need to remember the options for the next session, we omitted memorability in our evaluation. The remaining 5 attributes were integrated into our coding framework [23].

We used the TA method to measure effectiveness, learnability, errors, and efficiency characteristics, and the MARS questionnaire was used to measure satisfaction.

Errors or usability problems were detected based on the analysis of the "critical issues" encountered by the participants during the interactions detected from the video reviews. Critical issues were defined as those that prevented task completion, "severe issues" were defined as those issues that caused significant slowdown or frustration, and "cosmetic issues" were the ones that remained and had minimal effect [38].

Learnability was evaluated by measuring the number of quickly completed tasks.

Data Analysis

Phase 1: TA Usability Testing

TA Quantitative Part

Data analysis and measurements of usability metrics were addressed based on a coding framework mentioned in the study design and protocol section. The usability characteristics and problems and their severity rating are described as follows:

Efficiency was measured by two metrics: (1) the number of touches targeted and (2) the task completion time. The mean time taken for the users to perform each task was based on the following equation:

Efficiency = [(total of full completion of a task (1) or noncompletion (0) / (time spent on a task)] / [(total number of tasks × number of users)] × 100 **(1)**

Effectiveness was measured by the number of completed tasks (ie, task completion rate), indicating the task's success rate. The extent to which the user can fully and accurately achieve their task goals. Effectiveness was measured using the following equation.

Effectiveness = [(number of successfully completed tasks) / (total number of tasks performed)] × 100 **(2)**

The range of effectiveness was taken as "awful" (0%-50%), "bad" (50%-75%), "normal" (75%-90%), and "good" (90%-100%) [24].

Learnability was evaluated by measuring the number of quickly completed tasks.

Errors were identified as the number of user mistakes when performing the tasks.

Satisfaction was measured based on the user's total score on the MARS questionnaire.

TA Qualitative Part

The video reactions of the participants were transcribed verbatim. Usability data, characterized by users' comments, silences, repeated actions, and error messages, were collected through the recordings. Three members of the research team analyzed the obtained content. Transcripts and usability problems were also reviewed to identify the most common concerns. In any case of discrepancy in content analysis, a third-party reviewer was consulted.

These differences were categorized based on the tasks in the scenarios (ie, measurements, zoom and magnifying, and contrast and window level).

Data collected during the TA tasks (phase 1) were analyzed using fundamental inductive content analysis consisting of data reduction, data grouping, and the formation of concepts to answer research questions [39].

The inductive process is a bottom-up process that looks at all the issues as a whole by aggregating similar issues together until all the issues have been sorted into groups. Once all the groups (ie, subcategories) had been sorted, they were labeled to create more significant categories [40,41]. Thus, at the end of this process, we identified significant usability category issues and the specific problems associated with each one.

Phase 2: Evaluation of the Quality of the HAC App Using the MARS Questionnaire

Design

The participants (16 medical interns) were asked to complete the MARS questionnaire immediately after the TA session. MARS is the most popular scale and a highly reliable tool designed to assist researchers, professionals, and clinicians in classifying and assessing the quality of mHealth apps [25].

Data Collection

A validated and reliable Persian language version of MARS was used to collect the HAC app quality data [42].

MARS consists of 23 items in 5 objective quality subscales:

- Engagement encompasses 5 items and mainly focuses on entertainment and interest features of mobile apps.
- Functionality includes 4 items and addresses the ease of use and functional capabilities of mobile apps.
- Aesthetics consist of 3 items and discuss mobile app layout and visual appeal.
- The information includes 7 items and mainly considers quality, quantity, credibility, and visual enhancement of included information.
- The Subjective Quality subscale of MARS focuses on the overall rating of the app, its benefits, and its value.

Data Analysis

Each subscale item was rated a 5-point score from 1 (inadequate) to 5 (excellent). Usually, the mean score and SD were used to rate the quality of apps. Since the number of items in each subscale was different, we also used this formula [(mean of subscale/number of items in subscale)×(100)] to compute the score out of 100 and compare the subscales. To calculate the total HAC app score, the [(total mean of HAC app/total MARS items×(100)].

Friedman test was applied to compare the users' scores in 5 MARS subscales. The Wilcoxon test investigated the mean difference between 2 sets of pairs of MARS subscales. Spearman rank correlation coefficient was used to analyze the positive correlation between MARS subscales. Kruskal-Wallis and 1-way ANOVA tests were used to assess differences between medical interns' characteristics and MARS' subscales. All statistical analyses were performed using SPSS (version 16.0; IBM Inc) at a significance level of .05.

We applied inductive content analysis consisting of data reduction, data grouping, and the formation of concepts to analyze TA qualitative data and transform physicians' ideas into categories in the debriefing phase.

Phase 3: Debrief Participants

A debrief session is an informal conversation to collect users' experiences and [43] any features of the app that they particularly like or dislike, how easy or difficult it is to use, and what they think about the content and design of the app was discussed in the debriefing session. The medical interns' general opinion regarding the effective mobile apps to assist HCPs in education or clinical practice was also investigated in this phase. During the analysis of the recorded videos and voices, it came to our attention that the debrief sessions which were carried out with the participation of clinicians' research team, were the most active and engaging ones. To analyze and present the debriefed data, a narrative analysis method was used [44].

Results

Outline

The findings of each phase will be presented under the same headings in the methods section, including TA quantitative, TA qualitative, MARS quality assessment, and then a debriefing session.

Phase 1: TA Usability Testing

Overview

Table 1 illustrates the scenarios, goals, and actions needed to complete the tasks.

Table 1. Descriptions of the scenarios used in the usability testing.

Scenarios	Goals	Actions
1. A head trauma patient was admitted to the emergency department. Please check if the CT[a] scan indicated a patient with "minor or mild acute closed head trauma without neurologic deficit adult."	According to the guidelines, search for appropriate imaging procedures for a given diagnosis.	1. Selecting the search icon 2. Typing the disease title in the search box 3. Finding the head trauma from the query list 4. Clicking on the plus button 5. Check if the imaging procedure is recommended for the patient.
2. A patient was admitted to the emergency department with "new onset of seizures older than 18 following acute trauma." Please select the appropriate imaging procedure for the case.	To use appropriate imaging procedures for seizures.	1. Opening the search query 2. Typing the seizures into the search box 3. Navigating between items in the search list 4. Selecting the appropriate imaging procedure based on the patient's symptoms
3. Headache and vertigo are common symptoms at the emergency department. Please add headache to the shortlist for forthcoming queries.	To apply a shortlist menu to collect appropriate imaging procedures for common diseases and symptoms.	1. Adding headache to shortlist menu 2. Backing to the first page 3. Opening the shortlist 4. Deselecting the items you are not interested in anymore
4. A patient with proven subarachnoid hemorrhage (negative angiogram) was admitted to the hospital for follow-up. Please check for appropriate imaging procedures.	To use the CT[a] or CTA[b] button to access subarachnoid hemorrhage.	1. Opening list of diseases under the title of CT 2. Finding subarachnoid hemorrhage 3. Moving one step backward 4. Selecting the CTA button 5. Navigating between items in the search list 6. Click on the plus sign to search for detailed information on subarachnoid hemorrhage and its subgroups.

[a]CT: computed tomography.

[b]CTA: computed tomography angiography.

TA Quantitative

Efficiency

On the basis of the equation 1, the HAC app's relative overall efficiency was 97.8%. The average time spent for each scenario was 97.5 seconds, and the number of additional clicks was 0.93. The highest average of performing scenarios belonged to scenario 3 (109.25 seconds), and the lowest average was related to scenario 4 (83.875 seconds). Among the users, the highest total average time for 4 scenarios was related to user number 3 (161.8 seconds), and the lowest time was for user number 11 (58.0 seconds).

Effectiveness

The HAC app's effectiveness in assisting users in performing the scenarios based on the equation 2 was good (97%). Of 16 users, 14 (88%) completed all 4 scenarios, 2 (13%) completed 3 scenarios, and 2 (13%) users had difficulty performing scenario 2, which was focused on searching for "new onset of seizures older than 18 following acute trauma." The characteristics of this scenario that caused usability issues have been discussed under the heading TA qualitative, "inefficient data presentation and information control," and "poor searching capabilities" (Figures 1 and 2).

Learnability

Out of 16 users, 11 (69%) managed to complete 4 scenarios, 4 (25%) users managed to complete 3 scenarios without encountering critical issues, and 2 (13%) users faced critical issues to complete 2 scenarios.

Errors

Out of 16 users, 10 (63%) users did not make any errors while doing the scenarios, and 6 (33%) users were able to do the scenarios with >1 errors (Table 2).

Table 2. Matrix of efficiency and effectiveness of the Head Computed Tomography Scan Appropriateness Criteria (HAC) mobile app[a].

User number	Efficiency, s					Effectiveness
	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Total average	Total scenarios completed
1	98	189	150	210	161.8	3
2	75	101	134	82	98.0	4
3	192	135	129	108	141.0	4
4	92	86	115	57	87.5	4
5	74	125	63	84	86.5	4
6	115	73	129	53	92.5	4
7	115	83	82	52	83.0	4
8	69	93	106	127	98.8	4
9	60	109	89	80	84.5	4
10	70	117	170	82	109.8	3
11	34	57	79	62	58.0	4
12	40	97	125	45	76.8	4
13	109	132	87	64	98.0	4
14	72	83	95	67	79.3	4
15	101	185	95	107	122.0	4
16	110	59	100	62	82.8	4

[a]Average efficiencies: scenario 1=89.125, scenario 2=107.75, scenario 3=109.25, scenario 4=83.87, total average=97.5.

TA Qualitative

The results of the inductive content analysis regarding usability issues were grouped into 8 main categories and discussed below.

Lack of Finger-Friendly Touch Targets

Most participants had difficulty tapping the target buttons, such as the shortlist button, or icons, such as the plus sign (+) on the screen, and it was an intensive task to perform successfully. The participants stated that the given features are inappropriate for finger-touch targets. It might be due to the wrong size of the buttons or the need for more padding between buttons and icons around the edge of the screen. Consequently, it led to selecting the wrong part of the screen and frequent mistapping of the shortlist menu. Most participants often used this statement: “I cannot get the button.” Failure to press the targeted button and retouching the icons multiple times occurred frequently, resulting in a long time on the task and decreased efficiency. Moreover, it caused a failure of task completion by 2 users and reduced the effectiveness of the HAC app.

Poor Search Capabilities

Navigating the diseases and signs and symptoms was case-sensitive to the upper case. It made the searching diagnosis and signs and symptoms keywords awkward. The participant struggled to find diseases and signs and symptoms that had not been typed in upper case. Some participants forgot the “case-sensitive” feature every time they started the new scenario. Thus, participants backed out and jumped over the navigation process or tried to find the given case from a long list of search results. Both situations made it time-consuming and inefficient and caused participant frustration.

Input Problems

The main complaint by the participants was that the font size was inappropriately amplified with the limited mobile size. Participants mentioned that typing on the mobile phone screen was an intensive task. We found some difficulty in typing on the small screen; all the participant’s attention was focused on what they had typed. The “case-sensitive” feature in searching data amplified the problem. The lack of finger-friendly touch targets also made the typing more cognitive load and distracted from their main concerns, interacting with the patients.

Inefficient Data Presentation and Information Control

Another usability issue that caused frustration among users was inefficient data presentation and information control. To apply the HAC app, users entered specific diseases, signs, or symptoms enclosed in the Care Core guideline in the “Index” box. However, the list of clinical criteria under the disease heading was grouped using the plus sign (+) to provide a proper data presentation. A long list of conditions in the form of a dropdown menu enclosing the common signs and symptoms made it confusing for the participant. Since the mobile screen was too small, providing a long list of search results makes it time-consuming and inefficient. The lack of proper information layering and data categorization made it difficult for the participant to scroll the list. The participant commented the following:

> *It requires much attention and is very inconvenient since we need to interact with patients, other colleagues, and clinical settings environment.*

The critical issue was related to bringing cognitive load to the participants.

Unclear Control and Confirmation

Another failure dealt with providing feedback and confirmation. The participants expected the HAC app to inform them about what was happening, using appropriate feedback. For instance, when they were asked to add a given disease to the shortlist, they waited for a dialogue to let them know the conditions were added. The absence of the appropriate feedback resulted in the users being moved to the shortlist and checked on if the command was run. The exact process occurred when they were asked to remove the given disease from the shortlist. The users awaited a confirmation dialogue regarding spoken questions, such as a "yes" or a "no," to remove the disease from the shortlist before executing the removing command. Without a physical response, users did not know the current system status and were not confident about the consequences of their prior actions. They felt confusion and frustration.

Lack of Predictive Capabilities

Some participants expected more predictive capabilities and automation to optimize manual tasks and increase efficiency across various functions. For example, a participant stated the following:

> We prefer the HAC app to automatically move the most visited diseases or signs and symptoms to the shortlist menu.

They believed the sole manually supported feature for making a shortlist menu could be more efficient and less time-consuming.

Poor Assistance and Support

The participants thought some features on the HAC app, such as shortlists highlighted in red or items with the plus sign, were difficult to recall or interpret and caused cognitive load. The participants needed assistance or information to learn more about these features, such as tooltips, which display informative text, such as a description of its function when users hover over, focus on, or tap an icon. They were looking for a help tab and found it unclear because it was at the bottom of the "About us" tab. It caused the HAC app to be less self-descriptive and more dependent on external help, which needed to be clarified and made clearer.

Unclear Navigation Logic

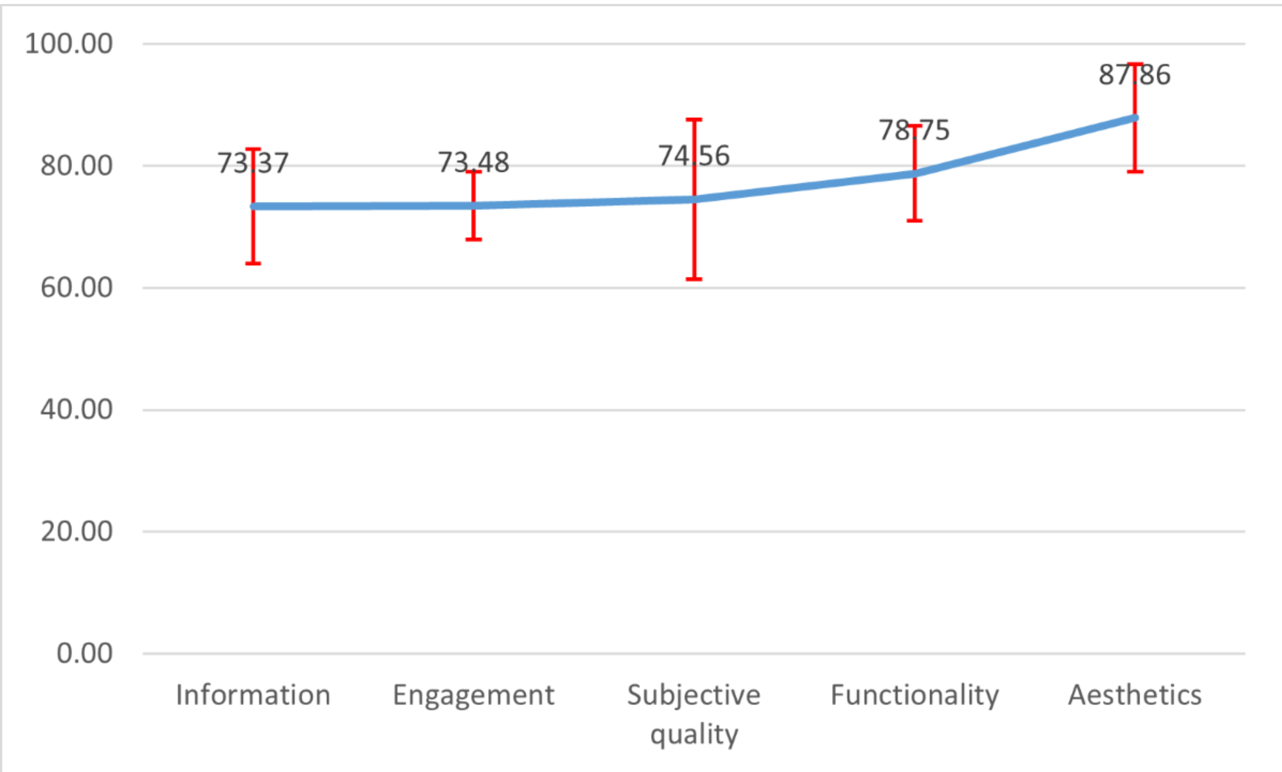
Some fundamental navigation control issues (eg, "back" function) were also reported during usability testing. For example, the participants tended to click the back button to return to the previous page, but it actually led them back to the home page. This drawback can lead to work duplication and frustration in task completion.

Phase 2: the Quality of the HAC App Using MARS

Analysis of Overall Quality Scores of the HAC App

Table 1 indicates that the overall quality score of the HAC app was favorable (82/100). Among the 4 MARS subscales, Information (73.37/100), and Engagement (73.48/100) had the lowest scores while Aesthetics had the highest score (87.86/100; Figure 3).

Figure 3. Overall quality scores of the Head Computed Tomography Scan Appropriateness Criteria app.



Analysis of Significant Differences and Correlation Between MARS Subscales

Using the Friedman test, the users' scores in 5 MARS subscales were compared, and the result revealed a significant difference (*P*<.001).

Wilcoxon test was applied to investigate the mean difference between 2 sets of pairs of MARS subscales. The results indicated a significant relationship between the Aesthetics subscale and Engagement (*P*=.001), Information (*P*=.003), Subjective Quality (*P*=.004), and Functionality (*P*=.02). A significant relationship was also found between the Functionality and Information subscales (*P*=.01; Table 3).

Table 3. The mean differences between 2 sets of pairs of MARS[a] subscale scores.

MARS subscales	Engagement score	Information score	Subjective quality score	Functionality score	Aesthetics score
Information score	0.909	—[b]	—	—	—
Subjective quality score	0.53	0.900	—	—	—
Functionality score	0.057	0.013	0.32	—	—
Aesthetics score	0.001	0.003	0.004	0.02	—

[a]MARS: Mobile Apps Rating Scale.

[b]Not applicable.

Spearman rank correlation coefficient presented a positive correlation between information with functionality subscales, $r_{.588}$, *P*=.02. A positive correlation was also seen between information and satisfaction, $r_{.648}$, *P*=.005. Table 4 indicates, in the subscale Information, the lowest score of the HAC app was "evidence base" (66.2/100), and the highest score was visual information (82/100). In the subscale Engagement, the lowest score of the HAC app was for "customization" (63.6/100), and the highest score was interest (90/100). In the subscale Functionality, the lowest score of the HAC app was for "performance" (67.4/100), and the highest score was "ease of use" (91.2/100). In the subscale Aesthetics, the lowest score of the HAC app was "visual appeal" (83.6/100), and the highest score was "graphics" (91.2/100).

Table 4. Head Computed Tomography Scan Appropriateness Criteria app scoring based on Mobile App Rating Scale 4 subscales.

	Scores, mean (SD)	Score out of 100
Information		
Accuracy: the app contains what is described	3.6 (0.50)	72.4
Goals: specific, measurable, and achievable goals	3.6 (0.50)	72
Quality of information: the app correct, well-written, and relevant content to the goal	3.5 (0.63)	70
Quantity of information: the extent of coverage within the scope of the app	3.4 (0.72)	68
Visual information: visual (eg, charts, images, and videos) to describe concepts	4.1 (0.95)	82
Credibility: legitimate source of app	4.06 (0.25)	81.2
Evidence base: trialed and tested app	3.31 (1.07)	66.2
Engagement		
Entertainment	3.25 (0.44)	65
Interest: fun and entertaining of app	4.5 (0.63)	90
Customization: support all preferences for app features (eg, sound and content)	3.18 (0.54)	63.6
Interactivity: provide feedback, contain reminders, and notifications	3.8 (0.40)	76.3
Target group	3.62 (0.50)	72.4
Functionality		
Performance: accuracy and speed of the app functions and components (buttons and menus)	3.37 (0.80)	67.4
Ease of use: easy to learn how to use the app	4.56 (0.62)	91.2
Navigation: accurate, appropriate, uninterrupted moving between screens	3.8 (0.61)	76
Gestural design: consistency of (taps, swipes, and scrolls) across all components	3.9 (0.57)	78
Aesthetics		
Layout: arrangement and size of buttons, icons, menus, and content on the screen	4.43 (0.72)	88.6
Graphics: the quality and resolution of graphics used for buttons, icons, menus, and content	4.56 (0.51)	91.2
Visual appeal: look of app	4.18 (0.54)	83.6

Medical Interns' Characteristics and MARS Subscales

Of the 16 users participating in the study, none had used the HAC app before, and only 1 (6%) person had used similar applications. Among them, 8 (50%) users believed using mobile apps for learning and clinical practice is helpful and were interested in using them. Figure 4 presents a significant difference between medical interns' interest in using mobile apps for learning and clinical practice (low, medium, high) with the Engagement subscale using the Kruskal-Wallis test (*P*=.03).

Figure 5 also indicates a significant difference between the medical interns' interest in using mobile apps with subjective quality subscales using a 1-way ANOVA test (*P*=.04).

Figure 4. Significant difference between engagement and interest in using the mobile app.

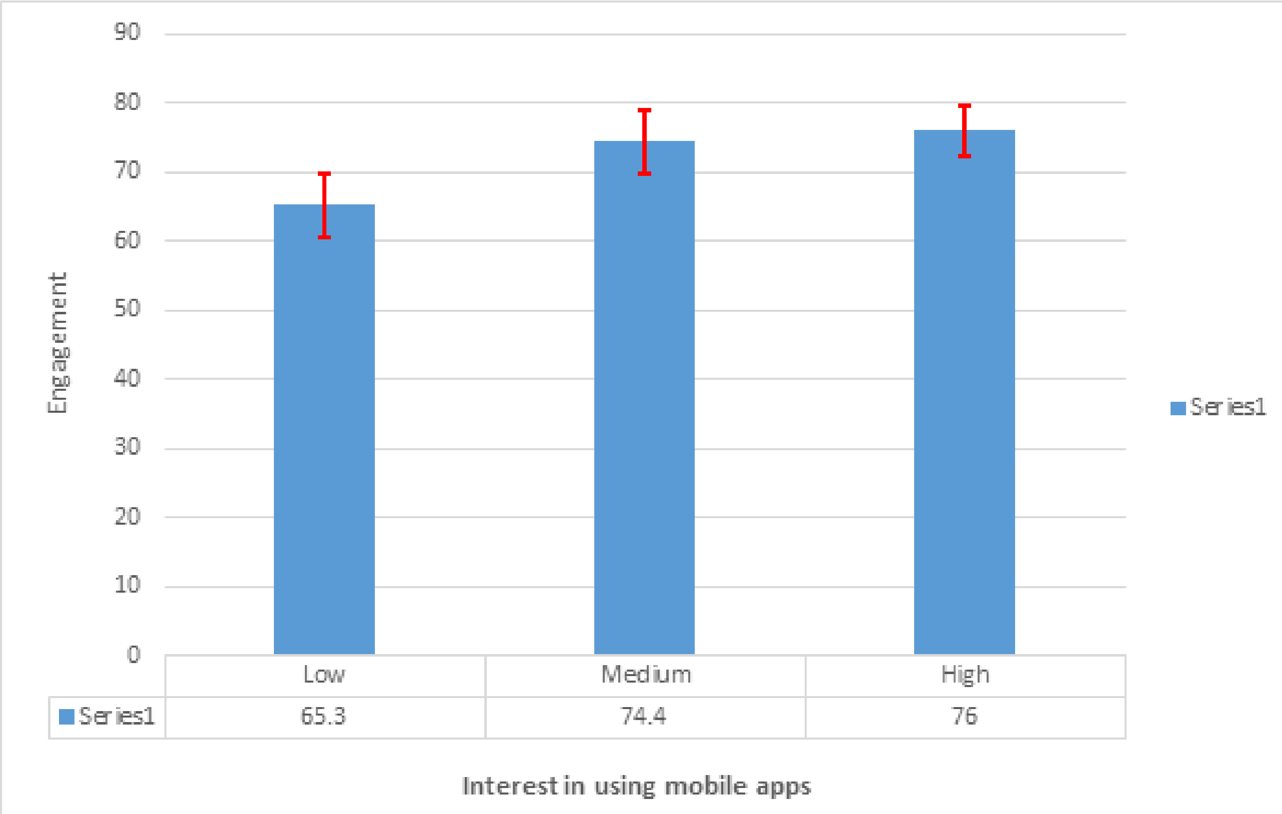
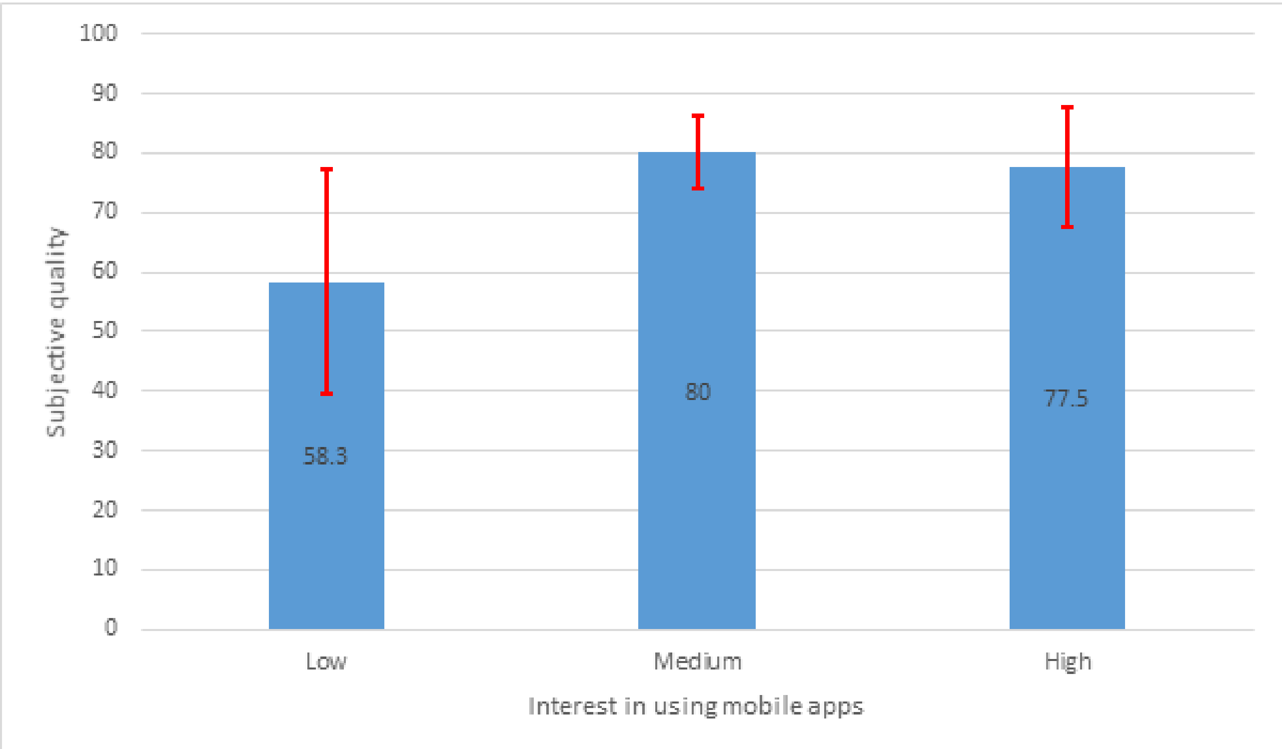


Figure 5. Significant difference between subjective quality and interest in using mobile app.



Phase 3: Debrief

We explored how useful they perceived the app to be, any features they particularly liked or disliked, how easy or difficult it was to use, and what they thought about the content and design of the app, which was discussed in the debriefing session. Although all users appreciated the high simplicity and learnability of the HAC app, they debated that navigation between pages and search capabilities need serious consideration.

One of the participants wanted this tool to be equipped with voice recognition systems:

> *We use this tool while walking or moving in different parts of the hospital, and the possibility of typing or text entry increases the possibility of errors and, as a result, repeating the same action, which will reduce efficiency.* [Participant 2]

Another participant believed this tool should provide access to the app at different times and conditions:

> *I am a doctor, and my hands are bloody; I do not want to touch my mobile too much, and I prefer this app to be able to search for the proper CT scan based on voice.* [Participant 15]

Another participant expected that apps designed for students would pay more attention to the educational needs and learning styles of students:

> *I think this issue is so essential that medical education experts should also be used in the design of apps. Anyway, each of us has a style to learn. If this customization feature is not included in the design, surely some users will not be able to work with this system or at least feel comfortable and useful while working with it.* [Participant 1]

Discussion

Principal Findings

Our findings demonstrated that the HAC app was practical and had acceptable usability in efficiency and effectiveness. It also displayed a positive quality score based on the MARS scale. In contrast, results of the TA usability test revealed that the HAC app has 8 notable usability issues. The results proved that despite the willingness of researchers and the simplicity of quantitative and questionnaire-based approaches to conducting usability testing [6,45], the observational, TA usability testing provided more unbiased, trustworthy, and insightful data in describing mobile app usability.

Nevertheless, through data analysis of the MARS subscales also brought to light the HAC app's usability issues, and its results support the qualitative TA results of this study. This agreement could be explained by the fact that MARS is a scale specifically designed to assess the quality of mobile apps [46]. Typically, the available usability scales and questionnaires are not highly reliable [6]; they are general scales designed primarily for evaluating the usability of computers or websites.

In addition, current usability and quality rating scales focus primarily on developers testing the usability of mobile apps, rather than end users who are patients or HCPs [46].

It is unlikely that usability issues will be thoroughly investigated in sole quantitative and questionnaire-based approaches [47] and need to be complemented by more objective and reliable approaches, such as TA methods.

In this study, HAC effectiveness assessment revealed that most users completed all 4 scenarios, although, 2 users faced problems completing scenario 2, which involved finding an appropriate imaging procedure for the "new onset of seizures" case. This failure may be due to the usability issues we categorized under "poor search capabilities" and "inefficient data presentation and information control." As shown in Figure 1, "poor searching capabilities" and "poor data presentation" caused a long list of seizure conditions, confusing the participant. Since the mobile screen was too small, providing a long list of search results brought more cognitive load to select the correct item, and 2 participants were left to perform this scenario later. However, they never got back to the scenario again. Our results support previous research findings. In the study, Chen et al [48] introduced proper navigation and searching capabilities as significant factors for users' rating of mHealth apps. Schwab and Langell [18] debated that ease of navigation is the foundation of an ideal mobile app since it smooths productivity and increases effectiveness. In the study to explore the usability of the physician-to-physician teleconsultation app in an orthopedic clinic, Choemprayong et al [49] presented mobile app usability issues in terms of data entry errors, presenting large-scale data and difficulty in selecting items from a list, which arise because of limited mobile screen size.

The HAC app also indicated acceptable efficiency and meantime completion for 4 scenarios. However, scenario 3 also showed the highest mean time completion. The problem might arise due to usability issues regarding the "lack of finger-friendly touch targets." The limited screen size of mobile phones results in the inappropriate size of buttons or lack of enough padding between the shortlist button and icons around the edge of the screen. Our results agree with previous studies that tapping the mobile phone buttons correctly is a crucial factor; however, incorrect operations have been reported frequently in previous studies [49-51]. In addition to data presentation, the low resolution of smartphone screens can lead to data input errors [49]. Existing evidence revealed highly significant differences between user effectiveness and efficiency with button sizes. In the study, Conradi et al [22] reported substantial differences in error rate between button sizes (5×5 mm) compared with the other sizes (8×8 mm. 11×8 mm, and 14×14 mm. It has been debated that interaction with mobile devices due to limited screen size and resolution often requires additional considerations and a specially adapted interface. The literature also claimed that key size manipulation should be considered for users' operation posture and activities (eg, standing, sitting, and walking) in mobile phone interactions [22]. However, the wide variation in optimal button size for mobile phones from 2.6 to 41.8 mm represents human–computer interaction in handheld devices. It is still in its infancy and requires more context-awareness to provide assistance based on the knowledge of its environment. Another possible explanation for the highest-time completion for scenario 3 is the usability issue categorized as "unclear control and confirmation" in this study. The participants of this study verbalized a lack of providing feedback on the HAC app when they were asked to add a given disease or sign and symptom to the list. The absence of the confirmation dialogue for successfully adding the given items to the shortlist resulted in the users moving to the shortlist and checking if the command was run. The exact process occurred when they were asked to remove the given disease from the shortlist. This rechecking caused work duplication and led to less efficiency. Work duplication has a significant and negative influence on physicians' performance and has been introduced as physicians' barrier to using mobile apps. In a study, Payne et al [52] found that physicians would use mobile apps to improve care workflow and productivity [38]. In another study, Ely et al [52] found that physicians believed if working with IT-related tools takes more than 2 minutes, they will not be efficient and practical for the point of care (39). Therefore, the effectiveness and efficacy of mobile apps serve as critical factors for physicians' intention to use mobile apps [52,53].

Regarding efficiency measures, our results also indicated significant variation in scenarios' time completion between the users. For example, user number 3 scored the highest total average time, nearly 3 times that of user number 11 (the lowest time), to perform the scenarios. Besides designing an optimal layout, significant variation in scenarios' time completion between the users may be due to the user characteristics. Xiong et al [20] debated that touch accuracy in mobile phones requires proper motor skills and "hand dexterity" in the operating fingers. Schwab and Langell [18] and Ozkan Gokalp-Yavuz [16] also highlighted the importance of user anatomy (eg, average index or thumb fingertip size) and user dexterity (ie, motor skills) in users' efficiency. Cho et al [51] reported usability problems related to the buttons of mobile apps developed using an eye-tracking system and retrospective TA usability evaluation.

The HAC app also showed a favorable quality score based on the MARS scale. However, the HAC app quality suffered from some drawbacks in Engagement and Information, which focus primarily on the effectiveness of apps in terms of interactivity, customizability, sending feedback, alerts, and reminders. Our results support previous results for assessing quality apps used by HCPs. In the study on drug reference apps in Taiwan, Chen et al [48] also reported poor engagement capabilities in terms of lack of entertainment, interactivity, and customization in the studied apps in Taiwan. In the study investigating influential factors in adopting a clinical photo documentation app for clinicians, Jacob discussed some drawbacks in engagement capabilities that need to be added for further revision of a given app [15]. Although few studies exist on using MARS to evaluate clinical apps adopted by HCPs, other relevant evidence supports our findings. In a qualitative study, Pokhrel et al [54] presented that HCPs prefer mobile apps that help them in their clinical practices, including "suggestive diagnosis and treatment after entering." Reports of studies that focused on using other IT toolkits also revealed that the IT tool would be effective among HCPs if it would support interactivity, answer physicians' questions, send feedback, and provide decision reasoning. Sandholzer et al [55] also introduced "prediction capabilities of mobile apps" as the most important preferences of medical students toward specific functionalities of future mobile apps. Despite the HAC app's drawbacks in engagement and information subscales, its quality in aesthetics has shown favorable MARS scoring. In a study of preferences and perceptions of users regarding graphical user interface and user experience, Sandesara et al [56] reported that minimalist design improves user experience and user control to fulfill a task in a specific order and time. The author argued that "simplicity is the ultimate sophistication" [56]. To the best of our knowledge, no study has evaluated and reported the items of each subscale of MARS and research is lacking on the evaluation of adopting and usability testing of a mobile app by HCPs [1]. Lack of related literature to assess the items of each subscales of MARS led to poor in-depth understandings and meaningful perception of apps' quality features in previous evidence. Therefore, it was impossible to compare HAC app quality rating with previous research properly. However, the results of quality assessment using MARS supports TA qualitative findings of this study. HAC app quality scoring in the functionality subscale revealed the minimum score belonged to the performance items, which focuses on the accuracy and speed of the app functions and components such as buttons or menus. Navigation also scored the minimum rating in the given subscale. In the subscale Engagement, the item customization that supports providing all necessary settings for apps features and the item interactivity that allows user input, providing feedback, and containing reminders and notifications also acquired the minimum scoring.

Our findings in the debrief session indicated that physicians with clear awareness and understanding of their clinical context and work processes tend to use other data input methods, such as voice recognition, to interact with the HAC app. The results of physicians' workflow analysis and time and motion studies presented the medical profession as a multitasking job, not only managing patient care but also spending part of their activities on indirect tasks, from doing paper work and documentation to transitioning and traveling within the clinic area, or fetching or bringing something [57,58]. Thus, in designing mobile apps, performance accuracy and time on users' tasks in different positions while walking or standing should be addressed appropriately. It has been argued that interaction with mobile devices while walking influences people's visual acuity and suppresses this ability by nearly 20% compared with visual acuity while standing [22]. Conradi et al [22] debated that walking is prone to a very high number of error occurrences, which is remarkable in smaller buttons. Using mobile apps with text entry methods involves physicians experiencing various interaction issues in terms of difficulty in typing on the small screen, mistapping due to inappropriate size of the buttons or lack of spacing between buttons, poor data presentation, and so on. Any poor mobile interaction is attention-grabbing and makes physicians concentrate solely on interacting with the mobile app to increase their performance accuracy. It would distract them from their main concern, which is interacting with the patients.

Moreover, it results in a long time being on the task and decreases the efficiency and effectiveness of HCPs in clinical settings. Auditory and sonic interfaces occupy less visual attention and make users less engaged in the sole main task. Consequently, users can handle multiple tasks simultaneously [59]. Here, physicians should be equipped with an alternative input method, for example, speech recognition. Evidence revealed that speech recognition has the potential to be a more efficient and effective method to speed up the entry rates while declining the error rates. It was reported that speech recognition supports high entry rates (speaks at a mean entry rate of 13-45 words per minute while walking around) and a low error rate of <2% [60]. Given the requirement that medical interns suggested, they emphasized the importance of "context-awareness" design in mobile apps that focuses on capturing and exploring context-based information to describe any entity (eg, persons, places, objects, and workflows) embedded in the environment to fully understand and characterize users’ tasks [59].

Implications

The evaluation framework used in this study can serve as a guide for the design and improvement of future clinical mobile apps to ensure they meet usability and quality standards for use by HCPs. Identifying usability issues through user feedback and analysis can help developers improve the usability and user satisfaction of clinical mobile apps among HCPs. Moreover, the results of the study can serve as a reference for HCPs and developers in selecting and implementing clinical mobile apps with acceptable usability and quality. It emphasizes the importance of multidisciplinary research, incorporating medical education specialists’ expertise, and considering user characteristics like motor skills and hand dexterity. The mixed methods approach used in the study, including MARS and TA analysis, can be adopted to gather valuable insights into user behavior and inform the design process of future apps for HCPs and developers. The study also suggests context-awareness design as a critical factor in developing meaningful IT–based solutions such as mobile apps.

Limitations

However, our investigation is subject to some limitations. It was conducted using a limited sample a specific target group (medical interns), and attending physicians and residents were not involved in the study. No contributions from IT experts and app developers were included in the evaluation of the HAC app. The study focused on the usability and quality of the HAC app in a specific medical context in Iran, which may limit the applicability of the findings to other health care settings or countries.

Conclusions

A mixed methods approach in evaluating the quality and usability of mobile apps yields valuable insights into the strengths and weaknesses of mobile apps. Adopting a holistic and multifaceted approach in evaluating mobile apps is highly recommended, as exclusively relying on a single methodology does not provide reliable and trustworthy information about the usability and quality of mobile apps. The results also presented that the unique characteristics of mobile devices, such as screen size, the users’ anatomical characteristics, and motor skills, influence users’ interaction and usability with mobile apps. Therefore, considering these characteristics and developing more tailored tools and methods for usability testing of mobile apps can bring potential benefits for developers, decision-makers, and HCPs.

Acknowledgments

The research team express their sincere gratitude to those who participated in the study.

The study was supported financially by the Deputy of Research and Technology of Kashan University of Medical Sciences (KAUMS) in Iran (grant 99190).

Authors' Contributions

ZM, AO, EN, and RK made substantial contributions to the conception and design of the study. FA, ZM, AO, HA, EN, ZN, and FH participated in data collection and performed the statistical analysis. ZM, EN, and FH contributed to manuscript drafting, revision, and approval. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Meidani Z, Atoof F, Mobarak Z, Nabovati E, Daneshvar Kakhki R, Kouchaki E, et al. Development of clinical-guideline-based mobile application and its effect on head CT scan utilization in neurology and neurosurgery departments. BMC Med Inform Decis Mak 2022 Apr 20;22(1):106 [FREE Full text] [doi: 10.1186/s12911-022-01844-3] [Medline: 35443649]
2. Garner SL, Sudia T, Rachaprolu S. Smart phone accessibility and mHealth use in a limited resource setting. Int J Nurs Pract 2018 Feb 20;24(1):10.1111/ijn.12609. [doi: 10.1111/ijn.12609] [Medline: 29159919]
3. Matricardi PM, Dramburg S, Alvarez-Perea A, Antolín-Amérigo D, Apfelbacher C, Atanaskovic-Markovic M, et al. The role of mobile health technologies in allergy care: an EAACI position paper. Allergy 2020 Feb 16;75(2):259-272 [FREE Full text] [doi: 10.1111/all.13953] [Medline: 31230373]
4. Anthes E. Mental health: there's an app for that. Nature 2016 Apr 07;532(7597):20-23. [doi: 10.1038/532020a] [Medline: 27078548]
5. Tarricone R, Petracca F, Cucciniello M, Ciani O. Recommendations for developing a lifecycle, multidimensional assessment framework for mobile medical apps. Health Econ 2022 Sep 06;31 Suppl 1(Suppl 1):73-97 [FREE Full text] [doi: 10.1002/hec.4505] [Medline: 35388585]

6. Zhou L, Bao J, Setiawan IM, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500 [FREE Full text] [doi: 10.2196/11500] [Medline: 30973342]
7. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020 Sep 29;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
8. Adu MD, Malabu UH, Malau-Aduli AE, Malau-Aduli BS. The development of my care hub mobile-phone app to support self-management in Australians with type 1 or type 2 diabetes. Sci Rep 2020 Jan 08;10(1):7 [FREE Full text] [doi: 10.1038/s41598-019-56411-0] [Medline: 31913296]
9. Owen JE, Jaworski BK, Kuhn E, Makin-Byrd KN, Ramsey KM, Hoffman JE. mHealth in the wild: using novel data to examine the reach, use, and impact of PTSD coach. JMIR Ment Health 2015 Mar 25;2(1):e7 [FREE Full text] [doi: 10.2196/mental.3935] [Medline: 26543913]
10. Singh K, Diamantidis CJ, Ramani S, Bhavsar NA, Mara P, Warner J, et al. Patients' and nephrologists' evaluation of patient-facing smartphone apps for CKD. Clin J Am Soc Nephrol 2019 Apr 05;14(4):523-529 [FREE Full text] [doi: 10.2215/CJN.10370818] [Medline: 30898873]
11. Singh K, Drouin K, Newmark LP, Lee J, Faxvaag A, Rozenblum R, et al. Many mobile health apps target high-need, high-cost populations, but gaps remain. Health Aff (Millwood) 2016 Dec 01;35(12):2310-2318. [doi: 10.1377/hlthaff.2016.0578] [Medline: 27920321]
12. Losa-Iglesias ME, Becerro-de-Bengoa Losa K, López-López D, Jimenez-Fernandez R, Corral-Liria I, Vazquez RR, et al. The usability of a heartbeat measuring mobile phone app: an observational study. J Med Syst 2019 Jun 12;43(7):226. [doi: 10.1007/s10916-019-1367-5] [Medline: 31190188]
13. Mickan S, Tilson JK, Atherton H, Roberts NW, Heneghan C. Evidence of effectiveness of health care professionals using handheld computers: a scoping review of systematic reviews. J Med Internet Res 2013 Oct 28;15(10):e212 [FREE Full text] [doi: 10.2196/jmir.2530] [Medline: 24165786]
14. Al-Ghamdi S. Popularity and impact of using smart devices in medicine: experiences in Saudi Arabia. BMC Public Health 2018 Apr 20;18(1):531 [FREE Full text] [doi: 10.1186/s12889-018-5465-y] [Medline: 29678133]
15. Jacob C, Sanchez-Vazquez A, Ivory C. Factors impacting clinicians' adoption of a clinical photo documentation app and its implications for clinical workflows and quality of care: qualitative case study. JMIR Mhealth Uhealth 2020 Sep 23;8(9):e20203 [FREE Full text] [doi: 10.2196/20203] [Medline: 32965232]
16. Ozkan NF, Gokalp-Yavuz F. Effects of dexterity level and hand anthropometric dimensions on smartphone users' satisfaction. Mob Inf Syst 2015;2015:1-9. [doi: 10.1155/2015/649374]
17. Wu P, Zhang R, Luan J, Zhu M. Factors affecting physicians using mobile health applications: an empirical study. BMC Health Serv Res 2022 Jan 04;22(1):24 [FREE Full text] [doi: 10.1186/s12913-021-07339-7] [Medline: 34983501]
18. Schwab T, Langell J. Human factors-based mobile application design for global health. Surg Innov 2018 Nov 22;25(6):557-562. [doi: 10.1177/1553350618814643] [Medline: 30465467]
19. Teferi GH, Tilahun BC, Guadie HA, Amare AT. Smartphone medical app use and associated factors among physicians at referral hospitals in Amhara region, North Ethiopia, in 2019: cross-sectional study. JMIR Mhealth Uhealth 2021 Mar 26;9(3):e19310 [FREE Full text] [doi: 10.2196/19310] [Medline: 33769303]
20. Xiong J, Muraki S, Fukumoto K. The effects of touch button size on touchscreen operability. J Mech Eng Autom 2014 Aug 28;4(8):667-672. [doi: 10.17265/2159-5275/2014.08.006]
21. Islam MN, Karim MM, Inan TT, Islam AK. Investigating usability of mobile health applications in Bangladesh. BMC Med Inform Decis Mak 2020 Feb 03;20(1):19 [FREE Full text] [doi: 10.1186/s12911-020-1033-3] [Medline: 32013965]
22. Conradi J, Busch O, Alexander T. Optimal touch button size for the use of mobile devices while walking. Procedia Manuf 2015;3:387-394. [doi: 10.1016/j.promfg.2015.07.182]
23. Zahiri Esfahani M, Khajouei R, Baneshi MR. Augmentation of the think aloud method with users' perspectives for the selection of a picture archiving and communication system. J Biomed Inform 2018 Apr;80:43-51 [FREE Full text] [doi: 10.1016/j.jbi.2018.02.017] [Medline: 29501909]
24. Farrahi R, Rangraz Jeddi F, Nabovati E, Sadeqi Jabali M, Khajouei R. The relationship between user interface problems of an admission, discharge and transfer module and usability features: a usability testing method. BMC Med Inform Decis Mak 2019 Aug 24;19(1):172 [FREE Full text] [doi: 10.1186/s12911-019-0893-x] [Medline: 31445520]
25. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the mobile application rating scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
26. Thurmond VA. The point of triangulation. J Nurs Scholarsh 2001 Apr 23;33(3):253-258. [doi: 10.1111/j.1547-5069.2001.00253.x] [Medline: 11552552]
27. Pavlíček J, Pavlíčková P. Usability testing methods and usability laboratory management. In: Castro LM, editor. Updates on Software Usability. New York, NY: IntechOpen; 2022.
28. Baxter K, Courage C, Caine K. Understanding Your Users: A Practical Guide to User Research Methods. New York, NY: Morgan Kaufmann; 2015.

29. Sauro J. A Practical Guide to Measuring Usability: 72 Answers to the Most Common Questions about Quantifying the Usability of Websites and Software. New York, NY: Measuring Usability LLC; 2010.
30. Macefield R. How to specify the participant group size for usability studies: a practitioner's guide. J Usability Stud 2009;5(1):34-45.
31. Cho H. Development and usability evaluation of an mHealth application for symptom self-management in underserved persons living with HIV. Columbia University Libraries. 2017. URL: https://academiccommons.columbia.edu/doi/10.7916/D8X06KJW [accessed 2024-04-29]
32. Peute LW, de Keizer NF, Jaspers MW. The value of retrospective and concurrent think aloud in formative usability testing of a physician data query tool. J Biomed Inform 2015 Jun;55:1-10 [FREE Full text] [doi: 10.1016/j.jbi.2015.02.006] [Medline: 25733166]
33. Screen recorder - AZ recorder. Google Play. URL: https://play.google.com/store/apps/details?id=com.hecorat.screenrecorder.free&hl=en&gl=US&pli=1 [accessed 2024-04-29]
34. Olmsted-Hawala EL, Murphy ED, Hawala S, Ashenfelter KT. Think-aloud protocols: a comparison of three think-aloud protocols for use in testing data-dissemination web sites for usability. In: Proceedings of the 2010 SIGCHI Conference on Human Factors in Computing Systems. 2010 Presented at: CHI '10; April 10-15, 2010; Atlanta, GA p. 2381-2390 URL: https://dl.acm.org/doi/10.1145/1753326.1753685 [doi: 10.1145/1753326.1753685]
35. Liu L, Özsu MT. Encyclopedia of Database Systems. Cham, Switzerland: Springer; 2009.
36. Nielsen J. Usability 101: introduction to usability. Nielsen Norman Group. URL: https://www.nngroup.com/articles/usability-101-introduction-to-usability/ [accessed 2016-11-01]
37. Abran A. Usability meanings and interpretations in ISO standards. Softw Qual J 2003;11:325-338 [FREE Full text]
38. Joe J, Chaudhuri S, Le T, Thompson H, Demiris G. The use of think-aloud and instant data analysis in evaluation research: exemplar and lessons learned. J Biomed Inform 2015 Aug;56:284-291 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.001] [Medline: 26071683]
39. Kyngäs H, Mikkonen K, Kääriäinen M. The Application of Content Analysis in Nursing Science Research. Cham, Switzerland: Springer; 2020.
40. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr 18;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
41. Azungah T. Qualitative research: deductive and inductive approaches to data analysis. Qual Res J 2018 Oct 31;18(4):383-400. [doi: 10.1108/qrj-d-18-00035]
42. Raeesi A, Khajouei R, Ahmadian L. Evaluating and rating HIV/AIDS mobile apps using the feature-based application rating method and mobile app rating scale. BMC Med Inform Decis Mak 2022 Oct 30;22(1):281 [FREE Full text] [doi: 10.1186/s12911-022-02029-8] [Medline: 36310157]
43. Henshall C, Davey Z, Jacelon C, Martin C. A usability study to test the effectiveness, efficiency and simplicity of a newly developed Internet-based Exercise-focused Health App for Lung cancer survivors (iEXHALE): protocol paper. Health Informatics J 2020 Jun 21;26(2):1431-1442 [FREE Full text] [doi: 10.1177/1460458219882268] [Medline: 31631739]
44. Smith CP. Content analysis and narrative analysis. In: Reis HT, Judd CM, editors. Handbook of Research Methods in Social and Personality Psychology. Cambridge, MA: Cambridge University Press; 2000:313-335.
45. Ye B, Chu CH, Bayat S, Babineau J, How T, Mihailidis A. Researched apps used in dementia care for people living with dementia and their informal caregivers: systematic review on app features, security, and usability. J Med Internet Res 2023 Oct 12;25:e46188 [FREE Full text] [doi: 10.2196/46188] [Medline: 37824187]
46. Azad-Khaneghah P, Neubauer N, Miguel Cruz A, Liu L. Mobile health app usability and quality rating scales: a systematic review. Disabil Rehabil Assist Technol 2021 Oct 08;16(7):712-721. [doi: 10.1080/17483107.2019.1701103] [Medline: 31910687]
47. Gibson A, McCauley C, Mulvenna MD, Ryan A, Laird L, Curran K, et al. Assessing usability testing for people living with dementia. In: Proceedings of the 4th Workshop on ICTs for improving Patients Rehabilitation Research Techniqu. 2016 Presented at: REHAB '16; October 13-14, 2016; Lisbon, Portugal p. 25-31 URL: https://dl.acm.org/doi/10.1145/3051488.3051492 [doi: 10.1145/3051488.3051492]
48. Chen YC, Liao W, Su M, Lin Y. Personalized and self-management: systematic search and evaluation quality factors and user preference of drug reference apps in Taiwan. J Pers Med 2021 Aug 12;11(8):790 [FREE Full text] [doi: 10.3390/jpm11080790] [Medline: 34442435]
49. Choemprayong S, Charoenlap C, Piromsopa K. Exploring usability issues of a smartphone-based physician-to-physician teleconsultation app in an orthopedic clinic: mixed methods study. JMIR Hum Factors 2021 Dec 20;8(4):e31130 [FREE Full text] [doi: 10.2196/31130] [Medline: 34931991]
50. Ast K. Diagnostic efficacy of handheld devices for emergency radiologic consultation. J Emerg Med 2010 Aug;39(2):273. [doi: 10.1016/j.jemermed.2010.04.044]
51. Cho H, Yen P, Dowding D, Merrill JA, Schnall R. A multi-level usability evaluation of mobile health applications: a case study. J Biomed Inform 2018 Oct;86:79-89 [FREE Full text] [doi: 10.1016/j.jbi.2018.08.012] [Medline: 30145317]

52. Payne KF, Wharrad H, Watts K. Smartphone and medical related app use among medical students and junior doctors in the United Kingdom (UK): a regional survey. BMC Med Inform Decis Mak 2012 Oct 30;12(1):121 [FREE Full text] [doi: 10.1186/1472-6947-12-121] [Medline: 23110712]
53. Ely JW, Osheroff JA, Ebell MH, Bergus GR, Levy BT, Chambliss ML, et al. Analysis of questions asked by family doctors regarding patient care. BMJ 1999 Aug 07;319(7206):358-361 [FREE Full text] [doi: 10.1136/bmj.319.7206.358] [Medline: 10435959]
54. Pokhrel P, Karmacharya R, Taylor Salisbury T, Carswell K, Kohrt BA, Jordans MJ, et al. Perception of healthcare workers on mobile app-based clinical guideline for the detection and treatment of mental health problems in primary care: a qualitative study in Nepal. BMC Med Inform Decis Mak 2021 Jan 19;21(1):21 [FREE Full text] [doi: 10.1186/s12911-021-01386-0] [Medline: 33468120]
55. Sandholzer M, Deutsch T, Frese T, Winter A. Medical students' attitudes and wishes towards extending an educational general practice app to be suitable for practice: a cross-sectional survey from Leipzig, Germany. Eur J Gen Pract 2016 Jun 22;22(2):141-146. [doi: 10.3109/13814788.2016.1144746] [Medline: 27101949]
56. Sandesara M, Bodkhe U, Tanwar S, Alshehri MD, Sharma R, Neagu B, et al. Design and experience of mobile applications: a pilot survey. Mathematics 2022 Jul 06;10(14):2380. [doi: 10.3390/math10142380]
57. Frey SM, Méan M, Garnier A, Castioni J, Wenger N, Egloff M, et al. Inter-hospital comparison of working time allocation among internal medicine residents using time-motion observations: an innovative benchmarking tool. BMJ Open 2020 Feb 16;10(2):e033021 [FREE Full text] [doi: 10.1136/bmjopen-2019-033021] [Medline: 32066604]
58. Tipping MD, Forth VE, Magill DB, Englert K, Williams MV. Systematic review of time studies evaluating physicians in the hospital setting. J Hosp Med 2010 Aug 20;5(6):353-359. [doi: 10.1002/jhm.647] [Medline: 20803675]
59. Lumsden J. Handbook of Research on User Interface Design and Evaluation for Mobile Technology. New York, NY: IGI global; 2008.
60. Kristensson PO, Jameson A. Five challenges for intelligent text entry methods. AI Magazine 2009 Dec;30(4):85-94. [doi: 10.1609/aimag.v30i4.2269]

Abbreviations

CT: computed tomography
CTA: computed tomography angiography
HAC: Head Computed Tomography Scan Appropriateness Criteria
HCP: health care provider
KAUMS: Kashan University of Medical Sciences
MARS: Mobile Apps Rating Scale
mHealth: mobile health
TA: think-aloud

Edited by A Kushniruk, D Kaufman; submitted 01.01.24; peer-reviewed by C Ni, R Spang; comments to author 04.03.24; revised version received 10.05.24; accepted 07.07.24; published 09.09.24.

Please cite as:
Meidani Z, Omidvar A, Akbari H, Asghari F, Khajouei R, Nazemi Z, Nabovati E, Holl F
Evaluating the Usability and Quality of a Clinical Mobile App for Assisting Physicians in Head Computed Tomography Scan Ordering: Mixed Methods Study
JMIR Hum Factors 2024;11:e55790
URL: https://humanfactors.jmir.org/2024/1/e55790
doi:10.2196/55790
PMID:

Original Paper

User-Centered Design and Usability of Voxe as a Pediatric Electronic Patient-Reported Outcome Measure Platform: Mixed Methods Evaluation Study

Samantha J Anthony[1,2,3], MSW, PhD; Sarah J Pol[1], MSc; Enid K Selkirk[1], MSc, PhD; Amarens Matthiesen[1], MA, PhD; Robert J Klaassen[4,5], MD; Dorin Manase[6], BSc; Amanda Silva[6], BSc; Melanie Barwick[1,7,8], MA, PhD; Jennifer N Stinson[1,9], MSc, PhD; Alameen Damer[1], MD; Mowa Ayibiowu[1]; Selina X Dong[4], MD; Stephan Oreskovich[4], MD; Michael Brudno[6,10], PhD

[1]Child Health Evaluative Sciences, Research Institute, Peter Gilgan Centre for Research and Learning, The Hospital for Sick Children, Toronto, ON, Canada

[2]Transplant and Regenerative Medicine Centre, The Hospital for Sick Children, Toronto, ON, Canada

[3]Factor-Inwentash Faculty of Social Work, University of Toronto, Toronto, ON, Canada

[4]Comprehensive Hemophilia Care Clinic, Children's Hospital of Eastern Ontario, Ottawa, ON, Canada

[5]Department of Pediatrics, University of Ottawa, Ottawa, ON, Canada

[6]The Data Aggregation, Translation and Architecture (DATA) Team, University Health Network, Toronto, ON, Canada

[7]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

[8]Department of Psychiatry, University of Toronto, Toronto, ON, Canada

[9]Lawrence S Bloomberg Faculty of Nursing, University of Toronto, Toronto, ON, Canada

[10]Department of Computer Science, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Samantha J Anthony, MSW, PhD
Child Health Evaluative Sciences
Research Institute, Peter Gilgan Centre for Research and Learning
The Hospital for Sick Children
686 Bay Street
Toronto, ON, M5G 0A4
Canada
Phone: 1 416 813 7654 ext 303126
Email: samantha.anthony@sickkids.ca

Abstract

Background: Electronic patient-reported outcome measures (ePROMs) are standardized digital instruments integrated into clinical care to collect subjective data regarding patients' health-related quality of life, functional status, and symptoms. In documenting patient-reported progress, ePROMs can guide treatment decisions and encourage measurement-based care practices. Voxe is a pediatric and user-centered ePROM platform for patients with chronic health conditions.

Objective: We aimed to describe the user-centered design approach involving feedback from end users and usability testing of Voxe's platform features to support implementation in a pediatric health care setting.

Methods: Purposive sampling was used to recruit patients aged 8-17 years from 2 chronic illness populations in 2 pediatric hospitals in Canada. Patients' health care team members were also purposively recruited. One-on-one iterative testing sessions were conducted digitally by research team members with participants to obtain feedback on the appearance and functionalities of the Voxe platform prototype. Patients and health care providers (HCPs) completed Voxe-related task-based activities. International Organization for Standardization key performance indicators were tracked during HCP task-based activities. HCPs also completed the System Usability Scale. To test platform usability, the think-aloud technique was used by participants during the completion of structured tasks. After completing all task-based activities, patient participants selected 5 words from the Microsoft Desirability Toolkit to describe their overall impression and experience with the Voxe platform. Qualitative data about likes, dislikes, and ease of use were collected through semistructured interviews. Feedback testing sessions were conducted with patients and HCPs until Voxe was acceptable to participating end users, with no further refinements identified. Quantitative and qualitative data analysis were completed using descriptive statistics and content analysis.

Results: A total of 49 patients and 38 HCPs were recruited. Patients were positive about Voxe's child-centered design characteristics and notification settings. HCPs rated Voxe as user-friendly and efficient, with the time to complete tasks decreasing over time. HCPs were satisfied with the Voxe platform functionalities and identified the value of Voxe's system notifications, summarized display of ePROM results, and its capacity to integrate with electronic medical records. Patients' and HCPs' high satisfaction rates with the Voxe prototype highlight the importance of being responsive to user suggestions from the inception of eHealth platform developments to ensure their efficient and effective design.

Conclusions: This paper describes the user-centered creation and usability testing of Voxe as an ePROM platform for implementation into clinical care for pediatric patients with chronic health conditions. As a patient-facing platform that can be integrated into electronic medical records, Voxe aligns with measurement-based care practices to foster quality patient-centered approaches to care. End users' positive feedback and evaluation of the platform's user-friendliness and efficiency suggest that Voxe represents a valuable and promising solution to systematically integrate patient-related outcome (PRO) data into complex and dynamic clinical health care settings.

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2021-053119

(JMIR Hum Factors 2024;11:e57984) doi:10.2196/57984

KEYWORDS

eHealth; end user engagement; mobile phone; patient-reported outcome; patient-reported outcome measures; pediatric; user-centered design

Introduction

Patient-Reported Outcome Measures

Patient-reported outcome measures (PROMs) are gaining significant momentum in clinical practice and research to foster a patient-centered approach to health care delivery [1-4]. PROMs are questionnaires used to collect subjective information directly from the patient regarding their health-related quality of life (HRQoL), functional status, and symptoms [5,6]. By directing the foci of clinical encounters, PROMs can facilitate early detection and monitoring of patient symptoms, empower patients to actively participate in their care, enhance health care providers' (HCPs) understanding of patient needs, and influence joint discussions with patients about health outcome priorities [3,4]. In documenting patient-reported progress, PROMs can guide treatment decisions, positively influence patient outcomes [7], and encourage measurement-based care practices [8].

Despite the proposed value of PROMs, low PROM adoption rates have been attributed to factors related to the completion of paper-based PROMs, including limited time and resources among clinicians and low response rates from patients [4]. Digital electronic PROMs (ePROMS) have been designed to overcome cited barriers and to improve PROM data quality and completion time [1,9]. Many benefits have been documented such as greater patient preference and acceptability, higher data quality and response rates, and reduced health care costs [1]. Currently, platforms used for ePROM collection (eg, REDCap [Research Electronic Data Capture]) primarily target the clinical care of adults [9-11], with few designed specifically for the clinical management of children's physical, social, and emotional health [4,10,12-14]. The lack of child-friendly and age-appropriate ePROM platforms needs addressing as children as young as 5 have shown capacity to self-report on their HRQoL [15], and 8 years of age is the recommended age to administer self-reported measures to children [10]. In the context of limited pediatric ePROM platforms, and with considerations around regulatory data privacy and management guidelines, secure servers for data storage and timely, responsive administrative and technical support [16,17], the development of new evidence-based platforms that enhance eHealth solutions for pediatric care should be prioritized.

User-Centered Design

User-centered design (UCD) is a popular design approach for developing eHealth innovations, including ePROM platforms [18]. When applying a UCD approach to optimize usability, compliance, and adoption of ePROM platforms, end users (ie, patients and HCPs) and key stakeholders (eg, decision- and policy-makers) are involved in the platform design processes [9,19,20]. Specifically, UCD outlines that (1) designers should understand end users and user-specific tasks, and (2) design processes are iterative to involve multiple cycles of design, testing, and redesign [21]. With attention to these parameters, using a UCD approach to eHealth platform development can create technologies that are meaningful, manageable, and sustainable for their user and organizational health care systems, potentially impacting the implementation success of eHealth solutions [22,23].

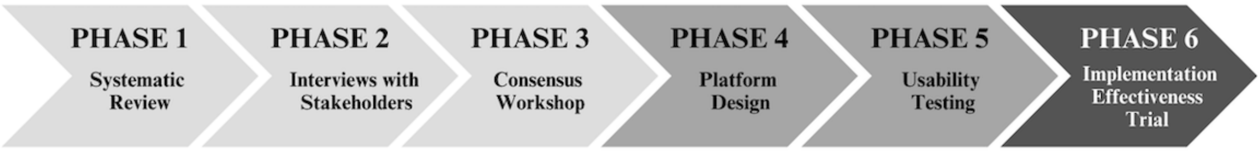
Voxe

Voxe is a pediatric, user-centered, and custom-built ePROM platform that is a progressive web application designed to integrate PROMs into the delivery of clinical care for pediatric patients with chronic health conditions [24]. Despite the profound and multidimensional impact of chronic disease on children's HRQoL, objective outcome metrics (eg, morbidity and mortality) alone are frequently used to determine the success of clinical interventions and care [25,26]. Given the paucity of child-oriented ePROM platforms in health care [27,28], Voxe represents a novel and child-friendly ePROM platform to facilitate the systematic integration of children's subjective evaluations regarding their physical, social, and emotional well-being in the delivery of care [24]. The advantages to having a custom-built progressive web application are (1) the ability to tailor the user experience specifically for pediatric patients, (2) to seamlessly integrate Voxe into the clinical workflow to reduce the barrier to ePROM completion, and (3) to ensure that Voxe is compatible with various electronic devices (eg, mobile phones, tablets, and computers). Additional distinguishing features include Voxe's capacity to incorporate any PROM, its capability to be created in different languages and its potential to integrate with any electronic medical record (EMR). Notably, for the purpose of Phase 5, Voxe was integrated with the EMR Epic.

Objectives

Building on previously completed phases of Voxe's development (Figure 1) [25,29], this paper outlines Voxe's user-centered design approach (Phase 4) involving feedback from end users at 2 pediatric hospitals in Canada and subsequent usability testing (Phase 5) specific to Voxe's platform features. The discussion will highlight our user-centered approach as a strength in prioritizing end user needs prior to Voxe's implementation within a pediatric health care setting.

Figure 1. Overview of key phases involved in Voxe's development.



Methods

Study Participants and Inclusion Criteria

Purposive sampling was used to recruit patients followed by The Hospital for Sick Children (SickKids) Transplant and Regenerative Medicine Centre (TRMC) or Hematology and Oncology program at the Children's Hospital of Eastern Ontario (CHEO) across age, diagnosis, sex, gender, and ethnicity. Members of the patients' interdisciplinary health care teams at SickKids and CHEO were also purposively recruited across professional disciplines, years of practice, sex, gender, and ethnicity.

Patients were eligible if they met the following criteria: (1) heart, kidney, liver, or lung transplant recipients who were a minimum of 3 months posttransplant (SickKids) or followed by the Hematology and Oncology program (CHEO); (2) between 8 and 17 years of age; and (3) able to speak and read in English. Patients with significant cognitive impairments, as determined by a health care team member, were not eligible to participate. Eligible HCPs included any member of the interdisciplinary teams within the Hematology and Oncology program (CHEO; Phase 4) or the TRMC (SickKids; Phases 4 and 5).

Ethical Considerations

Ethical approval to conduct this study was obtained from the Institutional Research Ethics Board at SickKids (1000057043 for Phase 4 and 1000067700 for Phase 5). All participants provided informed consent prior to their involvement in this study, and interview transcripts were deidentified. All participants received a $20 retail store gift card upon completion of study participation.

Phase 4: Generation of the Voxe ePROM Platform

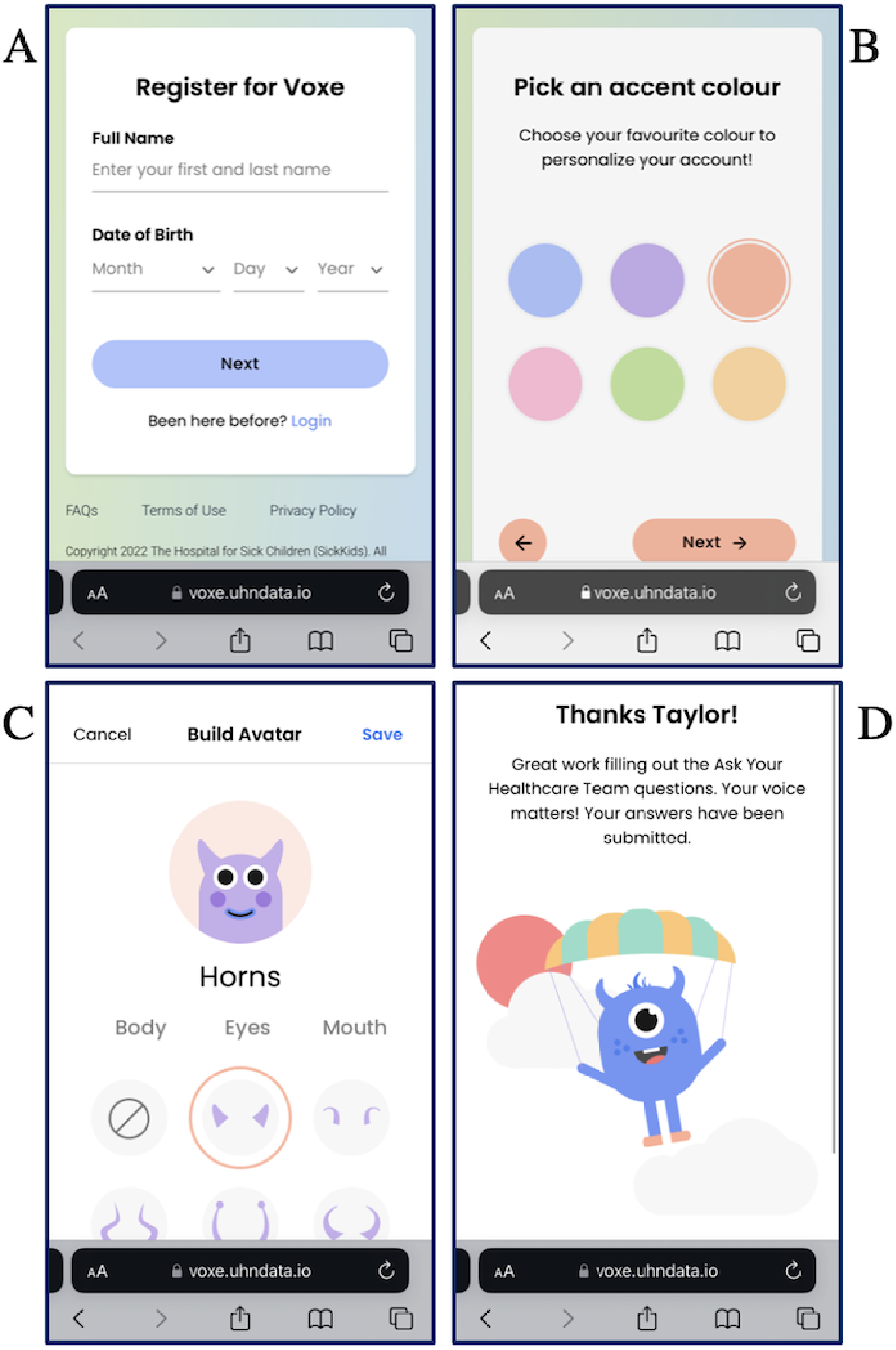
Overview

Previously completed phases of Voxe's development process (Phases 1-3) [30,31] informed the design of preliminary Voxe wireframes. The PedsQL Generic Core Scales [32] were selected as the first ePROM to be designed within Voxe as it is considered the most widely used generic HRQoL pediatric PROM [33]. Voxe wireframes depicting the PedsQL Generic Core Scales adhere to the requirements noted in the e-Booklet for the Electronic Implementation of the PedsQL Generic Core Scales [34] and were reviewed personally by the lead original developer of the PedsQL.

Key stakeholders (eg, decision- and policy makers) were also consulted to identify Voxe users (ie, patients and HCPs) and the possible tasks end users would complete (ie, persona and task inventory development). Following the design of preliminary wireframes, a rapid and iterative testing methodology was used to evaluate and improve Voxe prior to its full development and launch [35]. Through a user-centered approach, one-on-one iterative testing sessions were conducted virtually with patients and HCPs by a member of the research team (SJP, SD) to elicit feedback on Voxe design features.

Patient features identified within Voxe for feedback included: (1) account personalization options, (2) text and email notifications, and (3) the display of Voxe ambassador GIFs. First, after registering for a Voxe account, patients have the option to personalize their account by selecting: (1) 1 of 6 accent colors which populate the header and buttons in the patient's portal and (2) 1 of 6 prebuilt avatars or a custom-built avatar. Figure 2 presents the Voxe Patient Registration (A) and Personalization (B, C) screens. Second, Voxe allows patients to opt for text or email reminder notifications to complete the ePROMs on their preferred device (ie, mobile phone, tablet, or computer). Notifications are delivered by Voxe at 7 and 3 days in advance of clinic appointments. Third, following registration on Voxe and completion of each ePROM, the Voxe ambassador is displayed as a GIF (ie, animated avatar) on the platform. Figure 2 illustrates a still image of one of the Voxe ambassador GIFs (D) on a screen.

Figure 2. Examples of the Voxe Patient Registration (A), Personalization (B, C), and Ambassador GIF (D) Screens.



HCP features-of-use identified within Voxe for feedback included (1) the integration of Voxe with the EMR, (2) system (Epic) notifications of patient Voxe completion, and (3) a presentation view in Voxe to visualize trends in ePROM results over time. Notably, Voxe's presentation screen was designed for HCPs to show patients a high-level summary of their ePROM results through graphs specific to each domain or summary score, depending on the ePROM. Table 1 summarizes patient and HCP features in Voxe.

Table 1. Patient and HCP[a] features in Voxe.

Patient features	HCP features
Account personalization: select accent color, select a prebuilt avatar, and custom-build an avatar	Voxe integration with the EMR[b]
Text or email reminder notifications to complete ePROMs[c]	System notifications of patient Voxe completion
Display of Voxe ambassador GIF (ie, animated avatar) following registration and ePROM completion	Presentation view in Voxe to visualize trends in ePROM results over time

[a]HCP: health care provider.

[b]EMR: electronic medical record.

[c]ePROM: electronic patient-reported outcome measure.

Data Collection

Both patient and HCP participants completed task-based activities pertaining to the Voxe platform. Examples of tasks patients completed include: "This is your first time using Voxe. What would you click on to begin creating your profile?"; "What would you do if you wanted to pick a different colour?"; "You are doing the survey and let's assume you wanted to go back to a previous question. What would you click?" HCPs completed tasks such as "Let's assume that you need to view the patient's PedsQL results from May 24, 2019. What would you click to access their past PROM results?"; "Let's now pretend you want to compare the patient's PedsQL results between January 2019-June 2019 only. Where would you click to do this?"; "Voxe has a patient-friendly presentation feature to show patients their results during clinic and invite better conversation. Click where you would go to show the patient their patient-friendly overview of their results." During HCP task-based activities, International Organization for Standardization key performance indicators (KPIs) were tracked, as KPIs are deemed essential for evaluating the introduction of a novel technology, technique, or process [36,37]. Objective and subjective standards common in user experience design testing [38] were collected to measure (1) effectiveness—accuracy and completeness with which users achieve specific goals, displayed as a percentage of tasks successfully completed by users, and (2) efficiency—resources used in relation to results achieved, represented by the time it takes users to complete standard tasks successfully [39].

Following the completion of task-based activities, HCP participants completed the System Usability Scale (SUS), a 10-item Likert scale questionnaire to assess the KPI satisfaction [40,41]. The SUS is considered a reliable way to evaluate electronic platforms, in which a score of 68 is considered above average [40,41]. Following the completion of task-based activities, patient participants selected 5 words from a list of product reaction words outlined by the Microsoft Desirability Toolkit [42] (eg, creative, easy, and friendly) to describe their overall impression and experience with the Voxe platform.

Qualitative data were collected through semistructured interviews during which participants shared their likes and dislikes of the Voxe platform design and commented on the platform's ease of use and elements of functionality. Interviews were audio-recorded, transcribed verbatim, and deidentified.

Phase 5: Usability Testing of the Voxe ePROM Platform

Overview

To test the usability of the Voxe platform, the think-aloud technique was used in which participants verbalized their thoughts and feelings while interacting with Voxe to complete structured tasks [43,44]. The think-aloud technique is a well-known, formative usability testing approach to identify usability issues in the user interface designs of technologies such as ePROM platforms [45-47]. The think-aloud technique was integral to understanding the end user experience with Voxe and highlighting potential barriers to Voxe adoption that will inform its subsequent implementation [43,44].

Data Collection

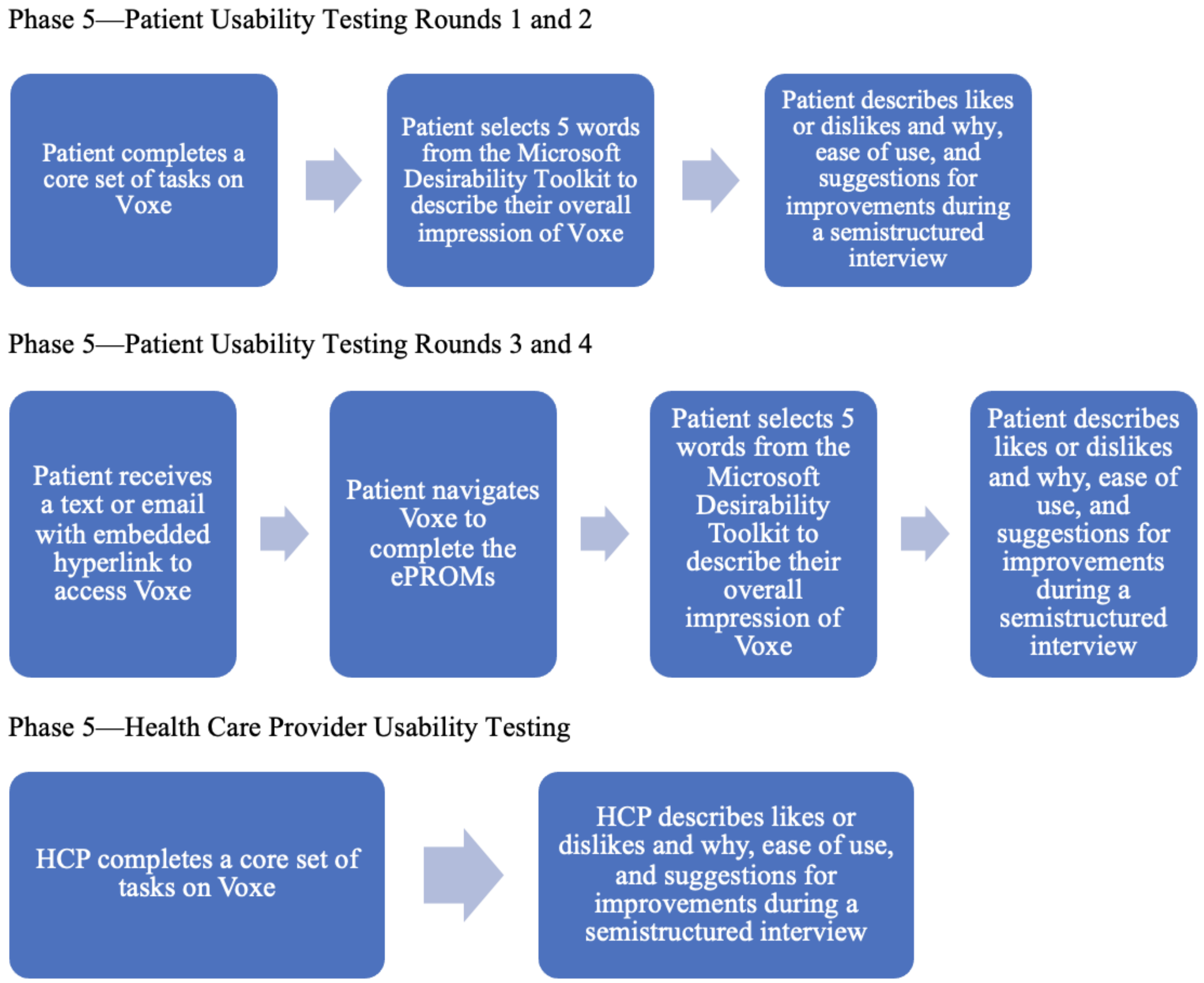
Following the development of interfaces of the Voxe ePROM platform for patients and HCPs, one-on-one iterative testing sessions were conducted virtually by research team members (SJP, AD, MA, SD, and SO) with patient and HCP participants. The purpose of the testing sessions was to obtain patients' and HCPs' feedback on the appearance and functionalities of the Voxe platform prototype.

During the first 2 testing rounds, patient participants were asked to complete a core set of tasks on Voxe, which were presented to them in the form of scenarios that they may encounter while interacting with Voxe. Examples of tasks patients completed include: "This is your first time using Voxe. Please make an account, enter a phone number and log into your account."; "How would you set your profile colour as yellow?"; "A few weeks have passed since your appointment at the hospital. During your appointment your nurse mentioned you can see a summary of your results to the PROMs you answered earlier on Voxe. How would you view your results from surveys you have completed?" The last 2 testing rounds were conducted to simulate "real-world" settings. An automated text message or email with an embedded hyperlink was sent to patients asking them to access Voxe remotely on a smartphone, tablet, or computer. Patients independently logged into Voxe using an anonymous username and password and navigated the platform to complete the ePROMs.

HCP participants accessed Voxe on a computer to complete a core set of tasks which simulate scenarios they may encounter while using Voxe in clinical practice. HCPs completed tasks such as "Click where you would go to view the patient's PedsQL results"; "Let's assume that you need to view the patient's PedsQL results from an earlier date. What would you click to access patient's past PROM results?"; "You are now meeting with the patient in clinic and would like to show them a quick summary of their PedsQL results. Click where you would go to share an overview of their results." Using think-aloud methodology, patient and HCP participants voiced out loud what they were looking at, thinking, doing, and feeling as they navigated the platform [43,48].

After completing task-based activities, patient participants selected 5 words from the Microsoft Desirability Toolkit [42] to describe their overall impression and experience with the Voxe platform. Qualitative data were collected through semistructured interviews to elicit information on what patient and HCP participants liked or disliked and why, the ease of use, elements of functionality in the context of typical practice workflow, and suggestions for improvements. Interviews were audio-recorded, transcribed verbatim, and deidentified. Rounds of iterative feedback testing were conducted with each participant population until Voxe was considered acceptable to participating end users with no further refinements identified [49-51]. Figure 3 presents an overview of usability testing procedures for patients and health care providers.

Figure 3. Overview of Phase 5 usability testing procedures for patients and health care providers. HCP: health care provider.



Data Analysis

Phase 4: Quantitative Data

Quantitative data from the platform creation testing sessions included (1) objective and subjective International Organization for Standardization KPIs, and (2) HCPs' scores on the SUS questionnaire. Descriptive statistics were calculated and summarized as appropriate. The quantitative data were triangulated with qualitative data to provide a comprehensive understanding of end users' experience with Voxe. Further refinements were subsequently made to the Voxe platform design based on the triangulated data.

Phases 4 and 5: Qualitative Data

Research team members experienced in qualitative methods (SJP, AD, MA, SD, and SO) used content analysis to identify and organize meaningful patterns into codes across the data collected from qualitative interviews [52-54]. Codes were primarily developed deductively from key concepts in the interview guides [55], and categories were created to identify areas of similarity by collapsing codes with unifying features [56,57]. Categories were reviewed and refined until a consensus was reached among team members. In the context of the current study, saturation as an end point criterion for completing data collection was determined when Voxe was considered acceptable to end users as measured by no further requests for refinements. NVivo 12 Lumivero was used for qualitative data management [58].

Results

Phases 4 and 5 Patient Participant Results

Patient Participants

A total of 49 patients from SickKids and CHEO participated in iterative testing rounds of Voxe in Phases 4 (platform creation) and 5 (usability testing) between September 2020 and August 2022. Among these, 19 patient participants were boys, 28 were girls, and 2 were nonbinary. Two participants were 8 or 9 years of age, 16 were between 10 and 13 years of age, and 31 were between 14 and 17 years of age. Participants identified as Asian (n=2); Black, Afro-Caribbean, or African American (n=4); Hispanic, Latino, or Spanish (n=1); South Asian (n=9); South Asian and White or Caucasian (n=4); White or Caucasian (n=28); and Other—White and Vietnamese (self-reported) (n=1). At SickKids, 12 and 13 participants participated in Phases 4 and 5, respectively. Of those, 8 participants were involved in both Phases 4 and 5. At CHEO, 12 participants participated in Phases 4 and 5. Eight participants were involved in both Phases 4 and 5. Table 2 reports additional patient participant demographic information.

Table 2. Patient demographics.

Variable	Phase 4 SickKids	Phase 4 CHEO	Phase 5 SickKids	Phase 5 CHEO	Total
Participant sex, n					
Female	7	7	7	7	28
Male	5	5	6	5	21
Participant gender, n					
Boy	4	5	5	5	19
Girl	7	7	7	7	28
Nonbinary	1	0	1	0	2
Participant age in years, n					
8 or 9	0	0	1	1	2
10-13	3	6	4	3	16
14-17	9	6	8	8	31
Transplant type (SickKids only), n					
Kidney	4	—[a]	4	—	8
Liver	4	—	4	—	8
Lung	1	—	2	—	3
Heart	3	—	3	—	6
Time since transplant in years (SickKids only)/diagnosis in years (CHEO[b] only), n					
<6	5	4	4	6	19
6-10	0	3	3	1	7
11-15	7	3	4	3	17
>15	0	1	2	2	5
12 and <1	0	1	0	0	1
Diagnosis (CHEO only), n					
Oncological	—	3	—	3	6
Bleeding disorder	—	6	—	4	10
Red cell disorder	—	1	—	3	4
Thrombotic disorder	—	2	—	2	4
Ethnicity, n					
Asian	1	0	1	0	2
Black, Afro-Caribbean or African American	0	2	0	2	4
Hispanic, Latino, or Spanish	0	0	1	0	1
South Asian	3	1	3	2	9
South Asian and White or Caucasian	1	1	1	1	4
White or Caucasian	7	7	7	7	28
Other—White and Vietnamese (self-reported)	0	1	0	0	1

[a]Not applicable.

[b]CHEO: Children’s Hospital of Eastern Ontario.

Patient Participant Experience

Increased Motivation Through Account Personalization Options

Several patients shared that the variety in options for customizing their personal Voxe account was an attractive feature of the Voxe platform to enhance their motivation to complete the ePROMs. Patients also appreciated that their personalization selections could be amended at any time. The personalization component was described as unique to Voxe, as one participant stated: "My favorite part was building the avatar… Because I think it's what makes it different from other apps…it kind of just felt like you're creating an avatar for a video game, which is pretty fun… [it] like makes doing something for the hospital actually fun" (Usability Testing SickKids #5).

ePROM Completion Facilitated Through Text and Email Notifications

Patients stated a collective preference for Voxe's text notifications on the basis of checking text notifications more frequently than email notifications and being more familiar with texts. Patients also emphasized the convenience of notification alerts appearing on their mobile phone screens rather than being embedded in an email, as one patient stated that they might be "more inclined to miss it [the notification to complete their ePROMs in Voxe] if it was on… email…" (Design Testing CHEO #2). Despite an overarching preference for mobile phone notifications, select patients described the value of offering children a choice of email or text notifications, depending on personal preference. Overall, patients reported that the text and email notification feature of the Voxe platform helped facilitate their ability to complete the ePROMs.

Positive Reinforcement Fostered Through the Voxe Ambassador GIFs

Nearly all patients expressed that the ambassador's presence in the platform reinforced a unique and child-centered sense of delight, engagement, and accomplishment regarding completing ePROMs, as noted by one patient: "If you're feeling down and you saw those animations... it'll probably cheer you up" (Design Testing CHEO #1). Several participants also noted that completing ePROMs in Voxe represents a psychosocial intervention in and of itself to uplift their mood, as one patient shared: "You did what you had to do, and you did it perfectly… it makes you feel better" (Usability Testing CHEO #4).

Phases 4 and 5 Health Care Provider Participant Results

Health Care Provider Participants

A total of 38 HCPs from SickKids and CHEO participated in iterative testing rounds of Voxe (Phases 4 and 5) between April 2020 and January 2023. Of these, 6 HCPs were men and 32 were women. Eleven participants had worked in the SickKids TRMC or at CHEO between 6 months and 5 years, 9 participants had worked there between 6 years and 10 years, 4 participants had worked there between 11 and 15 years, 9 had worked there between 16 and 20 years, and 5 had worked there over 20 years. HCP participants identified as Asian (n=3); Asian and White or Caucasian (n=1); Black, Afro-Caribbean or African American (n=3); South Asian (n=1); White or Caucasian (n=29); and Other—Greek (self-reported) (n=1). Table 3 reports additional HCP participant demographic information. At SickKids, 3 participants were involved in both Phases 4 and 5.

After completion of design testing, 94% (30/32) of HCPs described Voxe as being user-friendly, and 88% (28/32) felt that most people would learn to use Voxe quickly. Task success or effectiveness increased by 11.25% from Round 1 to Round 4. Task completion rate decreased by 7.34 seconds over the 4 rounds of testing. SUS or satisfaction increased from a B (75.94) to an A (83.75) between Round 1 and 4. Table 4 comprises task success, task completion, and system usability score metrics.

Table 3. HCP[a] demographics.

Variable	Phase 4 SickKids	Phase 4 CHEO[b]	Phase 5 SickKids	Total
Participant sex, n				
Female	13	14	5	32
Male	3	2	1	6
Participant gender, n				
Man	3	2	1	6
Woman	13	14	5	32
Clinician type, n				
Dietician	1	0	0	1
Nurse	2	6	2	10
Nurse practitioner	3	1	1	5
Occupational therapist	1	0	0	1
Physician	4	5	2	11
Physician assistant	0	2	0	2
Physiotherapist	1	0	0	1
Psychologist	1	1	0	2
Social worker	2	1	1	4
Other—information coordinator	1	0	0	1
Transplant program or area of work (SickKids only; participants reported all areas of work; total participants: 16 in Phase 4 and 6 in Phase 5), n				
GIFT[c]	1	—[d]	0	1
Heart	5	—	0	5
Kidney	6	—	6	12
Liver	8	—	0	8
Lung	5	—	0	5
Small bowel	1	—	0	1
Other—intestine	1	—	0	1
Department or area of work (CHEO; total: 16 participants), n				
Department of Pediatrics	—	2	—	2
Hematology	—	2	—	2
Hematology and Oncology	—	7	—	7
MDU[e]	—	2	—	2
Oncology	—	3	—	3
Number of years working in the SickKids TRMC[f] (SickKids only)/at CHEO (CHEO only)				
6 months to 5 years	2	7	2	11
6 years to 10 years	4	4	1	9
11 years to 15 years	4	0	0	4
16 years to 20 years	3	3	3	9
More than 20 years	3	2	0	5
Ethnicity, n				
Asian	2	0	1	3
Asian and White or Caucasian	0	1	0	1

Variable	Phase 4 SickKids	Phase 4 CHEO[b]	Phase 5 SickKids	Total
Black, Afro-Caribbean or African American	1	1	1	3
South Asian	1	0	0	1
White or Caucasian	12	13	4	29
Other—Greek (self-reported)	0	1	0	1

[a]HCP: health care provider.
[b]CHEO: Children's Hospital of Eastern Ontario.
[c]GIFT: Group for Improvement of Intestinal Function and Treatment
[d]Not applicable.
[e]MDU: medical day unit.
[f]TRMC: Transplant and Regenerative Centre.

Table 4. Task success, task completion, and system usability score metrics (Phase 4).

Metric	HCP[a] Round 1	HCP Round 2	HCP Round 3	HCP Round 4
Task success (%), mean (SD)	74.17 (12.57)	75.00 (14.32)	79.17 (6.30)	85.42 (11.57)
Task completion (seconds), mean (SD)	23.15 (9.34)	17.05 (7.40)	15.77 (5.20)	15.81 (5.58)
System Usability Score (grade), mean (SD)	75.94 (B); (7.78)	71.88 (C+); (8.81)	73.13 (B-); (12.87)	83.75 (A); (11.73)

[a]HCP: health care provider.

Health Care Provider Participant Perspective

Integrating Voxe With the Electronic Medical Record Is Important

During design testing sessions, HCPs noted the importance of integrating Voxe with the EMR (eg, Epic) as opposed to a standalone ePROM platform to allow HCPs to access PRO data using the existing EMR portal. HCPs associated the value of Voxe's capacity for immediate data transfer between Voxe and the hospital EMR with reducing technological fatigue and optimizing clinical workloads, particularly specific to documentation procedures. For example, 1 HCP described: "I think it would really be very useful if you know that that link could happen between Voxe [and Epic] especially if we're meant to write a plan and to basically get integrated and pull into the note… I think it has to be as seamless as possible" (Design Testing SickKids #5).

System Notifications of Voxe Completion Are Helpful

HCPs discussed the importance of system notifications in EMRs to signify when a patient completes their ePROMs within Voxe. According to HCPs, EMR system notifications would serve as a helpful reminder for HCPs to regularly check patients' ePROM data in the EMR. Respective to notification placement within EMRs, HCPs preferred that the notification be viewed in a prominent area (eg, centered on the page) or in the form of an "In Basket" notification within Epic ("In Basket" is the communication hub in Epic). Several HCPs stated that an "In Basket" notification offered an added advantage of separating PRO data from other clinical data to ensure that ePROMs are reviewed by HCPs. This finding was captured by an HCP, who stated: "…the In Basket… it's a little bit separate from like, say [clinical] results… I feel like if it was mixed in… that could easily get… missed" (Usability Testing SickKids #6).

Using the Presentation View to See Trends Within Results Over Time Is Valuable

HCPs emphasized the value of this feature of Voxe. In particular, HCPs commented on the presentation screen's simplicity and ease of use to identify pertinent topics or issues that may require HCPs' heightened attention during clinical encounters. HCPs also described that the presentation screen's capacity to visualize ePROM results over different time points could represent a tool to facilitate conversations with patients and family members on issues related to HRQoL, including medication management and treatment adherence. For example, an HCP noted: "…if I can see consistently that their pain has affected their physical functioning… that's a tool that I can use to say, 'Well, look how you're rating this, and you still don't want to go on your medication?'… having a longer view is going to be even more powerful in our education and trying to help get buy-in with our plans" (Design Testing CHEO #8).

Discussion

Principal Findings

This paper provides an overview of a UCD approach and usability testing phases of a novel evidence-based pediatric ePROM platform prototype named Voxe. Patients particularly appreciated Voxe's child-friendly options for personalized profiles (eg, color and avatar selection) and the inclusion of Voxe ambassador GIFs (ie, animated avatars) in the platform. HCPs' SUS scores reflected high satisfaction rates with the Voxe platform prototype. Design testing sessions with HCPs also highlighted Voxe's ease of use and unique capacity for integration into the hospital EMR as valuable for streamlining clinical documentation and evaluation processes. Patients' and HCPs' high satisfaction with the Voxe prototype highlight the importance of being responsive to user suggestions from the inception of eHealth platform developments to ensure their efficient and effective design [20,59].

Comparison With Prior Work

Patients' feedback on the Voxe prototype highlights the need to integrate developmentally responsive design considerations (eg, choice of color and avatars) in pediatric ePROM platforms to foster children's sense of engagement and motivation to complete ePROMs. While children and adolescents may be particularly amenable to using eHealth platforms due to their familiarity with technologies such as the internet and mobile phones, eHealth solutions should still respond to children's rapidly shifting development and associated ideas about the novelty or innovative nature of technology [60]. Insufficient consideration of the needs of the intended users in the development of eHealth platforms risks eHealth tools that are not able to fully accomplish their objectives [61], particularly if technologies do not align with end users' daily lives, habits, or rituals [22]. Patients' insights on Voxe's child-friendly characteristics contribute to a limited body of research regarding how to tailor ePROM platforms to the preferences of pediatric patients to optimize ePROM adoption and implementation [14,62,63].

The UCD approach that guided Voxe's development responds to existing calls for user-centered approaches to pediatric eHealth solution developments [64,65]. UCD is recognized as an optimal design approach for creating eHealth platforms [18,66,67] to help overcome eHealth implementation barriers, such as minimal clinical use of eHealth tools and low adoption rates in health care practices [68]. As eHealth technologies are often developed with a marginal level of engagement from end -users [68], Voxe's development phases offer practical steps for facilitating the inclusion of end users' perspectives in creating eHealth platform solutions specific to pediatric health care.

One challenge in pediatric health care practice is that most oral and written communications "with" children occur between adults [69-71]. Voxe's usability testing sessions highlighted Voxe's capacity for motivating pediatric patients to complete ePROMs and the potential to engage them in discussions concerning their care and treatment. For example, Voxe's development directly responded to what children considered most meaningful to them (eg, the incorporation of account personalization and Voxe ambassador GIFs), which may differ from adult-informed ideas about children's needs and preferences. Voxe also offers children time to reflect on pertinent issues they wish to emphasize in upcoming clinic appointments, and HCPs can review and consider patients' responses before clinical visits. Participants highlighted that Voxe's reminder notification system and presentation screen, which display the patient's summarized ePROM results, could be useful tools to facilitate child-provider communication regarding PRO data. The design and functionalities of the Voxe prototype responded to patient needs within the context of clinic visits, encouraging in-advance completion and purposively directing the foci of clinical encounters to patient priorities, offering the potential to improve satisfaction with health care [72-74].

The integration of ePROM data into EMRs presented another contextual variable driving ePROM platform development to address end user needs. Recent estimates suggest that most ePROM platforms are designed as stand-alone platforms, and only 60% (6/10) of ePROM systems offer compatibility for linking with EMRs [75]. In the context of limited pediatric ePROMs, Voxe is one of only a few pediatric-focused [4,76] or combined (ie, children and adults) [10] ePROM platforms to offer the capacity for front-end integration with EMRs. Similar to other cited benefits of ePROM platforms [6,72,74], Voxe can facilitate the integration of PRO data in patient EMRs in a manner that mirrors clinicians' existing workflows relative to documentation and assessment practices. During Voxe's usability evaluations, HCPs emphasized the value of Voxe's capacity for seamless integration into the EMR, which facilitated viewing and sharing mock PRO data in real time, and has the potential to offer time-saving benefits and remove potential burdens associated with logging into a stand-alone platform to review PRO data [74]. Overall, HCPs described how Voxe could help address key barriers pertaining to ePROM uptake in pediatric clinical practices, such as inconsistencies in compatibilities with EMRs, which has been cited among other associated challenges with PRO data management [6].

The Future Implementation of Voxe

The UCD approach and usability testing of Voxe outlined in this paper will inform the full operationalization and implementation of Voxe in pediatric health care settings. Future implementation initiatives will include (1) the delivery of HCP orientation sessions to familiarize HCPs with Voxe and (2) the evaluation (Phase 6) of the Voxe ePROM platform using a hybrid implementation-effectiveness design. Presently, the implementation of ePROMs in health care settings remains sparse and inconsistent [6], and transparent reporting on the use of implementation strategies to guide the future implementation of Voxe [77,78] will contribute to addressing this knowledge gap.

Strengths and Limitations

Voxe was designed and evaluated by an interdisciplinary research team with expertise in pediatric health research, including UCD and mixed methodology. This approach elicited participant perspectives through qualitative methods, garnering insights from a diverse sample of end user participants (ie, patients and HCPs) relative to age and clinician type. Participants completed simple tasks aligned with using Voxe which provided feedback about the changes needed prior to implementation. Evaluating the use of Voxe in more complex, "real-world" situations and to interpret patient ePROM results and trends longitudinally is a future aim of our program of research. We acknowledge that participants represented a small sample recruited from the SickKids TRMC and the Hematology and Oncology program at CHEO which limited eligible chronic health conditions and may have implications on the generalizability of our findings. Future studies should include participation from other pediatric hospitals and clinical programs. Of note, the sample recruited included English-speaking participants only. The absence of the perspectives of non-English speaking individuals to inform the development of Voxe limits our transferability of findings within these populations.

Conclusions

This paper outlines Voxe's UCD approach and the usability testing of Voxe's platform features as an ePROM platform designed for implementation into clinical care delivery for pediatric patients with chronic health conditions. As a patient-facing platform that can be integrated into EMRs, Voxe aligns with measurement-based care practices to foster quality patient-centered approaches to care. End users' positive feedback and evaluation of the platform's user-friendliness and efficiency suggest that Voxe represents a valuable and promising solution to systematically integrate PRO data in complex and dynamic clinical health care settings. Future collection of usage and outcome data will enable cost-benefit analyses to support the long-term integration of eHealth platforms in clinical services [79].

Acknowledgments

The development and usability evaluation of Voxe was funded by SickKids Transplant and Regenerative Medicine Centre—Pilot Grant Competition, Canadian Donation and Transplantation Research Program and Astellas Pharma Canada—Research Innovation Grant, Health Canada—Health Care Policy Contribution Program (agreement 1920-HQ-000105). The authors appreciate all children and health care providers who made time in their schedules to participate in this study. This research would not have been possible without their valuable insights.

Conflicts of Interest

None declared.

References

1. Meirte J, Hellemans N, Anthonissen M, Denteneer L, Maertens K, Moortgat P, et al. Benefits and disadvantages of electronic patient-reported outcome measures: systematic review. JMIR Perioper Med 2020;3(1):e15588 [FREE Full text] [doi: 10.2196/15588] [Medline: 33393920]
2. Grover S, Fitzpatrick A, Azim FT, Ariza-Vega P, Bellwood P, Burns J, et al. Defining and implementing patient-centered care: an umbrella review. Patient Educ Couns 2022;105(7):1679-1688. [doi: 10.1016/j.pec.2021.11.004] [Medline: 34848112]
3. Nguyen H, Butow P, Dhillon H, Sundaresan P. A review of the barriers to using patient-reported outcomes (PROs) and patient-reported outcome measures (PROMs) in routine cancer care. J Med Radiat Sci 2021;68(2):186-195 [FREE Full text] [doi: 10.1002/jmrs.421] [Medline: 32815314]
4. Morris AC, Ibrahim Z, Heslin M, Moghraby OS, Stringaris A, Grant IM, et al. Assessing the feasibility of a web-based outcome measurement system in child and adolescent mental health services - myHealthE a randomised controlled feasibility pilot study. Child Adolesc Ment Health 2023;28(1):128-147 [FREE Full text] [doi: 10.1111/camh.12571] [Medline: 35684987]
5. Krogsgaard MR, Brodersen J, Christensen KB, Siersma V, Kreiner S, Jensen J, et al. What is a PROM and why do we need it? Scand J Med Sci Sports 2021;31(5):967-971. [doi: 10.1111/sms.13892] [Medline: 33249660]
6. Glenwright BG, Simmich J, Cottrell M, O'Leary SP, Sullivan C, Pole JD, et al. Facilitators and barriers to implementing electronic patient-reported outcome and experience measures in a health care setting: a systematic review. J Patient Rep Outcomes 2023;7(1):13 [FREE Full text] [doi: 10.1186/s41687-023-00554-2] [Medline: 36786914]
7. Resnick SG, Oehlert ME, Hoff RA, Kearney LK. Measurement-based care and psychological assessment: using measurement to enhance psychological treatment. Psychol Serv 2020;17(3):233-237. [doi: 10.1037/ser0000491] [Medline: 32881578]
8. Fortney JC, Unützer J, Wrenn G, Pyne JM, Smith GR, Schoenbaum M, et al. A tipping point for measurement-based care. Psychiatr Serv 2017;68(2):179-188. [doi: 10.1176/appi.ps.201500439] [Medline: 27582237]
9. Lee M, Kang D, Kim S, Lim J, Yoon J, Kim Y, et al. Who is more likely to adopt and comply with the electronic patient-reported outcome measure (ePROM) mobile application? A real-world study with cancer patients undergoing active treatment. Support Care Cancer 2022;30(1):659-668. [doi: 10.1007/s00520-021-06473-6] [Medline: 34363495]
10. Haverman L, van Oers HA, Limperg PF, Hijmans CT, Schepers SA, Sint Nicolaas SM, et al. Implementation of electronic patient reported outcomes in pediatric daily clinical practice: the KLIK experience. Tijdschrift voor Kindergeneeskunde 2013;81(S1):48-49. [doi: 10.1007/s12456-013-0048-4]
11. Kyte D, Anderson N, Auti R, Aiyegbusi OL, Bishop J, Bissell A, et al. Development of an electronic patient-reported outcome measure (ePROM) system to aid the management of patients with advanced chronic kidney disease. J Patient Rep Outcomes 2020;4(1):55 [FREE Full text] [doi: 10.1186/s41687-020-00223-8] [Medline: 32642867]
12. Birnie KA, Campbell F, Nguyen C, Lalloo C, Tsimicalis A, Matava C, et al. iCanCope PostOp: user-centered design of a smartphone-based app for self-management of postoperative pain in children and adolescents. JMIR Form Res 2019;3(2):e12028 [FREE Full text] [doi: 10.2196/12028] [Medline: 31008704]
13. Skrabal Ross X, Condon P, Yates P, Walker R, Herbert A, Bradford N. Feasibility of weekly electronic patient- and proxy-reported outcome measures in pediatric oncology. Cancer Nurs 2023. [doi: 10.1097/NCC.0000000000001251] [Medline: 37232526]

14. Meryk A, Kropshofer G, Hetzer B, Riedl D, Lehmann J, Rumpold G, et al. Implementation of daily patient-reported outcome measurements to support children with cancer. Pediatr Blood Cancer 2021;68(11):e29279. [doi: 10.1002/pbc.29279] [Medline: 34383360]
15. Varni JW, Limbers CA, Burwinkle TM. How young can children reliably and validly self-report their health-related quality of life?: an analysis of 8,591 children across age subgroups with the PedsQL 4.0 generic core scales. Health Qual Life Outcomes 2007;5:1 [FREE Full text] [doi: 10.1186/1477-7525-5-1] [Medline: 17201920]
16. Richter J, Kampling C, Schneider M. Electronic patient-reported outcome measures (ePROMs) in rheumatology. Patient reported outcome measures in rheumatic diseases 2016:371-388. [doi: 10.1007/978-3-319-32851-5_15]
17. García Abejas A, Serra Trullás A, Sobral MA, Canelas D, Leite Costa F, Salvador Verges A. Improving the understanding and managing of the quality of life of patients with lung cancer with electronic patient-reported outcome measures: scoping review. J Med Internet Res 2023;25:e46259 [FREE Full text] [doi: 10.2196/46259] [Medline: 37021695]
18. van Velsen L, Ludden G, Grünloh C. The limitations of user-and human-centered design in an eHealth context and how to move beyond them. J Med Internet Res 2022;24(10):e37341 [FREE Full text] [doi: 10.2196/37341] [Medline: 36197718]
19. Brown W, Yen PY, Rojas M, Schnall R. Assessment of the health IT usability evaluation model (health-ITUEM) for evaluating mobile health (mHealth) technology. J Biomed Inform 2013;46(6):1080-1087 [FREE Full text] [doi: 10.1016/j.jbi.2013.08.001] [Medline: 23973872]
20. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
21. Gould JD, Lewis C. Designing for usability: key principles and what designers think. Commun ACM 1985;28(3):300-311. [doi: 10.1145/3166.3170]
22. van Gemert-Pijnen JEC, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
23. Dunlop E, Ferguson A, Mueller T, Baillie K, Laskey J, Clarke J, et al. Involving patients and clinicians in the design of wireframes for cancer medicines electronic patient reported outcome measures in clinical care: mixed methods study. JMIR Form Res 2023;7:e48296 [FREE Full text] [doi: 10.2196/48296] [Medline: 38127422]
24. Anthony SJ, Pol SJ, Lin J, Barwick M, Brudno M, Manase D, et al. Creation of an electronic patient-reported outcome measure platform voxe: a mixed methods study protocol in paediatric solid organ transplantation. BMJ Open 2021;11(10):e053119 [FREE Full text] [doi: 10.1136/bmjopen-2021-053119] [Medline: 34667013]
25. Anthony SJ, Stinson H, Lazor T, Young K, Hundert A, Santana MJ, et al. Patient-reported outcome measures within pediatric solid organ transplantation: a systematic review. Pediatr Transplant 2019;23(6):e13518. [doi: 10.1111/petr.13518] [Medline: 31210404]
26. Didsbury MS, Kim S, Medway MM, Tong A, McTaggart SJ, Walker AM, et al. Socio-economic status and quality of life in children with chronic disease: a systematic review. J Paediatr Child Health 2016;52(12):1062-1069. [doi: 10.1111/jpc.13407] [Medline: 27988995]
27. Bevans KB, Moon JH, Carle AC, Mara CA, Lai J, DiMarco L, et al. Patient reported outcomes as indicators of pediatric health care quality. Acad Pediatr 2014;14(5 Suppl):S90-S96. [doi: 10.1016/j.acap.2014.06.002] [Medline: 25169465]
28. Huang IC, Revicki DA, Schwartz CE. Measuring pediatric patient-reported outcomes: good progress but a long way to go. Qual Life Res 2014;23(3):747-750 [FREE Full text] [doi: 10.1007/s11136-013-0607-2] [Medline: 24362765]
29. Anthony SJ, Young K, Pol SJ, Selkirk EK, Blydt-Hansen T, Boucher S, et al. Patient-reported outcome measures in pediatric solid organ transplantation: exploring stakeholder perspectives on clinical implementation through qualitative description. Qual Life Res 2021;30(5):1355-1364 [FREE Full text] [doi: 10.1007/s11136-020-02743-8] [Medline: 33447959]
30. Anthony SJ, Stinson H, Lazor T, Young K, Hundert A, Santana MJ, et al. Patient-reported outcome measures within pediatric solid organ transplantation: a systematic review. Pediatr Transplant 2019;23(6):e13518. [doi: 10.1111/petr.13518] [Medline: 31210404]
31. Anthony SJ, Young K, Pol SJ, Selkirk EK, Blydt-Hansen T, Boucher S, et al. Patient-reported outcome measures in pediatric solid organ transplantation: exploring stakeholder perspectives on clinical implementation through qualitative description. Qual Life Res 2021;30(5):1355-1364 [FREE Full text] [doi: 10.1007/s11136-020-02743-8] [Medline: 33447959]
32. Varni JW, Seid M, Kurtin PS. PedsQL 4.0: reliability and validity of the pediatric quality of life inventory version 4.0 generic core scales in healthy and patient populations. Med Care 2001;39(8):800-812. [doi: 10.1097/00005650-200108000-00006] [Medline: 11468499]
33. Eiser C, Morse R. Quality-of-life measures in chronic diseases of childhood. Health Technol Assess 2001;5(4):1-157 [FREE Full text] [doi: 10.3310/hta5040] [Medline: 11262421]
34. Devlen J, Crager J, Anfray C, Varni J. e-Booklet for electronic implementation: pediatric quality of life inventory (PedsQL) generic core scales child report (ages 8-12). Mapi Research Trust. 2019. URL: https://eprovide.mapi-trust.org/instruments/pediatric-quality-of-life-inventory [accessed 2024-08-03]
35. Barnum CM. Usability testing essentials: ready, set test!. In: Merken S, editor. Establishing the essentials. Amsterdam: Elsevier; 2020:9-33.

36. ISO 9241-11:2018(en) ergonomics of human-system interaction - Part 11: usability: definitions and concepts. International Organization for Standardization. 2018. URL: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en [accessed 2021-01-20]
37. About us: what we do. International Organization for Standardization. URL: https://www.iso.org/what-we-do.html [accessed 2021-01-20]
38. Barnum CM. Usability testing essentials: ready, set test!. In: Merken S, editor. Analyzing the Findings. Amsterdam: Elsevier; 2020:287-319.
39. ISO 9000(en) Quality management systems - fundamentals and vocabulary. International Organization for Standardization. 2015. URL: https://www.iso.org/obp/ui/#iso:std:iso:9000:ed-4:v1:en [accessed 2021-01-20]
40. Brooke J. SUS—a quick and dirty usability scale. Usability evaluation in industry 1996;189(194):4-7. [doi: 10.1201/9781498710411-35]
41. Haggerty T, Brabson L, Grogg KA, Herschell AD, Giacobbi P, Sedney C, et al. Usability testing of an electronic health application for patient activation on weight management. Mhealth 2021;7:45 [FREE Full text] [doi: 10.21037/mhealth-20-119] [Medline: 34345622]
42. Barnum C. Usability testing essentials: ready, set test!. In: Marken S, editor. Preparing for Usability Testing. Amsterdam: Elsevier; 2020:197-248.
43. Li AC, Kannry JL, Kushniruk A, Chrimes D, McGinn TG, Edonyabo D, et al. Integrating usability testing and think-aloud protocol analysis with "near-live" clinical simulations in evaluating clinical decision support. Int J Med Inform 2012;81(11):761-772. [doi: 10.1016/j.ijmedinf.2012.02.009] [Medline: 22456088]
44. Rose AF, Schnipper JL, Park ER, Poon EG, Li Q, Middleton B. Using qualitative studies to improve the usability of an EMR. J Biomed Inform 2005;38(1):51-60 [FREE Full text] [doi: 10.1016/j.jbi.2004.11.006] [Medline: 15694885]
45. Jaspers MWM, Steen T, van den Bos C, Geenen M. The think aloud method: a guide to user interface design. Int J Med Inform 2004;73(11-12):781-795. [doi: 10.1016/j.ijmedinf.2004.08.003] [Medline: 15491929]
46. Li AC, Kannry JL, Kushniruk A, Chrimes D, McGinn TG, Edonyabo D, et al. Integrating usability testing and think-aloud protocol analysis with "near-live" clinical simulations in evaluating clinical decision support. Int J Med Inform 2012;81(11):761-772. [doi: 10.1016/j.ijmedinf.2012.02.009] [Medline: 22456088]
47. Aiyegbusi OL, Kyte D, Cockwell P, Marshall T, Dutton M, Slade A, et al. Using patient-reported outcome measures (PROMs) to promote quality of care and safety in the management of patients with advanced chronic kidney disease (PRO-trACK project): a mixed-methods project protocol. BMJ Open 2017;7(6):e016687 [FREE Full text] [doi: 10.1136/bmjopen-2017-016687] [Medline: 28667225]
48. Lewis C. Using the "Think Aloud" Method in Cognitive Interface Design. New York: IBM Research; 1982.
49. Stinson JN, Jibb LA, Nguyen C, Nathan PC, Maloney AM, Dupuis LL, et al. Development and testing of a multidimensional iPhone pain assessment application for adolescents with cancer. J Med Internet Res 2013;15(3):e51 [FREE Full text] [doi: 10.2196/jmir.2350] [Medline: 23475457]
50. Macefield R. How to specify the participant group size for usability studies: a practitioner's guide. J Usability Stud 2009;5(1):34-45.
51. Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. J Biomed Inform 2004;37(1):56-76 [FREE Full text] [doi: 10.1016/j.jbi.2004.01.003] [Medline: 15016386]
52. Gibbs G. Analysing qualitative data. London, England: SAGE; 2007.
53. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
54. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: implications for conducting a qualitative descriptive study. Nurs Health Sci 2013;15(3):398-405. [doi: 10.1111/nhs.12048] [Medline: 23480423]
55. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
56. Berends L, Johnston J. Using multiple coders to enhance qualitative analysis: the case of interviews with consumers of drug treatment. Addiction Research & Theory 2009;13(4):373-381. [doi: 10.1080/16066350500102237]
57. Whittemore R, Chase SK, Mandle CL. Validity in qualitative research. Qual Health Res 2001;11(4):522-537. [doi: 10.1177/104973201129119299] [Medline: 11521609]
58. Richards L. Using NVivo in qualitative research. United Kingdom: Sage; 1999.
59. Whitehouse SR, Lam PY, Balka E, McLellan S, Deevska M, Penn D, et al. Co-creation with TickiT: designing and evaluating a clinical eHealth platform for youth. JMIR Res Protoc 2013;2(2):e42 [FREE Full text] [doi: 10.2196/resprot.2865] [Medline: 24140595]
60. Wu YP, Steele RG, Connelly MA, Palermo TM, Ritterband LM. Commentary: pediatric eHealth interventions: common challenges during development, implementation, and dissemination. J Pediatr Psychol 2014;39(6):612-623 [FREE Full text] [doi: 10.1093/jpepsy/jsu022] [Medline: 24816766]
61. Nilsen W, Kumar S, Shar A, Varoquiers C, Wiley T, Riley WT, et al. Advancing the science of mHealth. J Health Commun 2012;17 Suppl 1:5-10. [doi: 10.1080/10810730.2012.677394] [Medline: 22548593]

62. Tyack Z, Simons M, McPhail SM, Harvey G, Zappala T, Ware RS, et al. Improving the patient-centred care of children with life-altering skin conditions using feedback from electronic patient-reported outcome measures: protocol for a hybrid effectiveness-implementation study (PEDS-ePROM). BMJ Open 2021;11(4):e041861 [FREE Full text] [doi: 10.1136/bmjopen-2020-041861] [Medline: 33837095]
63. van Muilekom MM, Teela L, van Oers HA, van Goudoever JB, Grootenhuis MA, Haverman L. Patients' and parents' perspective on the implementation of patient reported outcome measures in pediatric clinical practice using the KLIK PROM portal. Qual Life Res 2022;31(1):241-254 [FREE Full text] [doi: 10.1007/s11136-021-02950-x] [Medline: 34324137]
64. Wiljén A, Chaplin JE, Crine V, Jobe W, Johnson E, Karlsson K, et al. The development of an mHealth tool for children with long-term illness to enable person-centered communication: user-centered design approach. JMIR Pediatr Parent 2022;5(1):e30364 [FREE Full text] [doi: 10.2196/30364] [Medline: 35258466]
65. Badawy SM, Cronin RM, Hankins J, Crosby L, DeBaun M, Thompson AA, et al. Patient-centered eHealth interventions for children, adolescents, and adults with sickle cell disease: systematic review. J Med Internet Res 2018;20(7):e10940 [FREE Full text] [doi: 10.2196/10940] [Medline: 30026178]
66. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;46(Suppl):49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
67. Ten Klooster I, Wentzel J, Sieverink F, Linssen G, Wesselink R, van Gemert-Pijnen L. Personas for better targeted eHealth technologies: user-centered design approach. JMIR Hum Factors 2022;9(1):e24172 [FREE Full text] [doi: 10.2196/24172] [Medline: 35289759]
68. Stewart E, Milton A, Yee HF, Song MJ, Roberts A, Davenport T, et al. eHealth tools that assess and track health and well-being in children and young people: systematic review. J Med Internet Res 2022;24(5):e26015 [FREE Full text] [doi: 10.2196/26015] [Medline: 35550285]
69. Larsson I, Svedberg P, Arvidsson S, Nygren JM, Carlsson I. Parents' experiences of an e-health intervention implemented in pediatric healthcare: a qualitative study. BMC Health Serv Res 2019;19(1):800 [FREE Full text] [doi: 10.1186/s12913-019-4643-7] [Medline: 31690287]
70. Grootens-Wiegers P, Hein I, Staphorst M. Using children's voice to optimize pediatric participation in medical decision making. Am J Bioeth 2018;18(3):14-16. [doi: 10.1080/15265161.2017.1418928] [Medline: 29466122]
71. Baggott C, Baird J, Hinds P, Ruland CM, Miaskowski C. Evaluation of Sisom: a computer-based animated tool to elicit symptoms and psychosocial concerns from children with cancer. Eur J Oncol Nurs 2015;19(4):359-369. [doi: 10.1016/j.ejon.2015.01.006] [Medline: 25691300]
72. Teela L, van Muilekom MM, Kooij LH, Gathier AW, van Goudoever JB, Grootenhuis MA, et al. Clinicians' perspective on the implemented KLIK PROM portal in clinical practice. Qual Life Res 2021;30(11):3267-3277 [FREE Full text] [doi: 10.1007/s11136-020-02522-5] [Medline: 32468402]
73. Engelen V, Detmar S, Koopman H, Maurice-Stam H, Caron H, Hoogerbrugge P, et al. Reporting health-related quality of life scores to physicians during routine follow-up visits of pediatric oncology patients: is it effective? Pediatr Blood Cancer 2012;58(5):766-774. [doi: 10.1002/pbc.23158] [Medline: 21584933]
74. Govindaraj R, Agar M, Currow D, Luckett T. Assessing patient-reported outcomes in routine cancer clinical care using electronic administration and telehealth technologies: realist synthesis of potential mechanisms for improving health outcomes. J Med Internet Res 2023;25:e48483 [FREE Full text] [doi: 10.2196/48483] [Medline: 38015606]
75. Anderson NE, Kyte D, McMullan C, Cockwell P, Aiyegbusi OL, Verdi R, et al. Global use of electronic patient-reported outcome systems in nephrology: a mixed methods study. BMJ Open 2023;13(7):e070927 [FREE Full text] [doi: 10.1136/bmjopen-2022-070927] [Medline: 37438075]
76. Bele S, Paolucci EO, Johnson DW, Quan H, Santana M. Implementing paediatric patient-reported outcome measures in outpatient asthma clinics: a feasibility assessment study. BMJ Open 2023;13(11):e073260 [FREE Full text] [doi: 10.1136/bmjopen-2023-073260] [Medline: 37945296]
77. IJzerman RV, van der Vaart R, Breeman LD, Arkenbout K, Keesman M, Kraaijenhagen RA, et al. An iterative approach to developing a multifaceted implementation strategy for a complex eHealth intervention within clinical practice. BMC Health Serv Res 2023;23(1):1455 [FREE Full text] [doi: 10.1186/s12913-023-10439-1] [Medline: 38129824]
78. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
79. Bergmo TS. How to measure costs and benefits of eHealth interventions: an overview of methods and frameworks. J Med Internet Res 2015;17(11):e254 [FREE Full text] [doi: 10.2196/jmir.4521] [Medline: 26552360]

Abbreviations

CHEO: Children’s Hospital of Eastern Ontario
ePROM: electronic patient-reported outcome measure
EMR: electronic medical record
HCP: health care provider

HRQoL: health-related quality of life
KPI: key performance indicator
PRO: patient-reported outcome
PROM: patient-reported outcome measure
REDCap: Research Electronic Data Capture
SUS: System Usability Scale
TRMC: Transplant and Regenerative Medicine Centre
UCD: user-centered design

Edited by A Kushniruk; submitted 05.03.24; peer-reviewed by M Gorges, E McCabe; comments to author 22.04.24; revised version received 06.06.24; accepted 12.07.24; published 19.09.24.

Please cite as:
Anthony SJ, Pol SJ, Selkirk EK, Matthiesen A, Klaassen RJ, Manase D, Silva A, Barwick M, Stinson JN, Damer A, Ayibiowu M, Dong SX, Oreskovich S, Brudno M
User-Centered Design and Usability of Voxe as a Pediatric Electronic Patient-Reported Outcome Measure Platform: Mixed Methods Evaluation Study
JMIR Hum Factors 2024;11:e57984
URL: https://humanfactors.jmir.org/2024/1/e57984
doi:10.2196/57984
PMID:

Original Paper

Triage Accuracy and the Safety of User-Initiated Symptom Assessment With an Electronic Symptom Checker in a Real-Life Setting: Instrument Validation Study

Ville Liu[1], DMD, PhD Program; Minna Kaila[2], MD, PhD; Tuomas Koskela[3,4], MD, PhD

[1]Faculty of Medicine, University of Helsinki, Helsinki, Finland

[2]Public Health Medicine, Faculty of Medicine, University of Helsinki, Helsinki, Finland

[3]Department of General Practice, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland

[4]The Wellbeing Services County of Pirkanmaa, Tampere, Finland

Corresponding Author:
Ville Liu, DMD, PhD Program
Faculty of Medicine
University of Helsinki
Ruusulankatu 21 B 32
Helsinki, 00250
Finland
Phone: 358 400642104
Email: villeliu@hotmail.com

Abstract

Background: Previous studies have evaluated the accuracy of the diagnostics of electronic symptom checkers (ESCs) and triage using clinical case vignettes. National Omaolo digital services (Omaolo) in Finland consist of an ESC for various symptoms. Omaolo is a medical device with a Conformité Européenne marking (risk class: IIa), based on Duodecim Clinical Decision Support, EBMEDS.

Objective: This study investigates how well triage performed by the ESC nurse triage within the chief symptom list available in Omaolo (anal region symptoms, cough, diarrhea, discharge from the eye or watery or reddish eye, headache, heartburn, knee symptom or injury, lower back pain or injury, oral health, painful or blocked ear, respiratory tract infection, sexually transmitted disease, shoulder pain or stiffness or injury, sore throat or throat symptom, and urinary tract infection). In addition, the accuracy, specificity, sensitivity, and safety of the Omaolo ESC were assessed.

Methods: This is a clinical validation study in a real-life setting performed at multiple primary health care (PHC) centers across Finland. The included units were of the walk-in model of primary care, where no previous phone call or contact was required. Upon arriving at the PHC center, users (patients) answered the ESC questions and received a triage recommendation; a nurse then assessed their triage. Findings on 877 patients were analyzed by matching the ESC recommendations with triage by the triage nurse.

Results: Safe assessments by the ESC accounted for 97.6% (856/877; 95% CI 95.6%-98.0%) of all assessments made. The mean of the exact match for all symptom assessments was 53.7% (471/877; 95% CI 49.2%-55.9%). The mean value of the exact match or overly conservative but suitable for all (ESC's assessment was 1 triage level higher than the nurse's triage) symptom assessments was 66.6% (584/877; 95% CI 63.4%-69.7%). When the nurse concluded that urgent treatment was needed, the ESC's exactly matched accuracy was 70.9% (244/344; 95% CI 65.8%-75.7%). Sensitivity for the Omaolo ESC was 62.6% and specificity of 69.2%. A total of 21 critical assessments were identified for further analysis: there was no indication of compromised patient safety.

Conclusions: The primary objectives of this study were to evaluate the safety and to explore the accuracy, specificity, and sensitivity of the Omaolo ESC. The results indicate that the ESC is safe in a real-life setting when appraised with assessments conducted by triage nurses. Furthermore, the Omaolo ESC exhibits the potential to guide patients to appropriate triage destinations effectively, helping them to receive timely and suitable care.

International Registered Report Identifier (IRRID): RR2-10.2196/41423

(JMIR Hum Factors 2024;11:e55099) doi:10.2196/55099

KEYWORDS

nurse triage; emergency department triage; triage; symptom assessment; health services accessibility; telemedicine; eHealth; remote consultation; eHealth; primary health care; primary care; urgent care; health services research; health services

Introduction

Background

Seeking information online regarding medical symptoms is a common and well-known phenomenon among people and patients worldwide [1-4]. Usually, the general public searches online for symptoms associated with their medical condition before receiving a medical diagnosis. This includes websites of support groups, written blogs by patients, websites created by editors of popular media, governmental sites, and artificial intelligence (AI) interfaces. However, self-diagnostic web-based sources may be of varying quality in terms of reliability, with misleading information and possibly false advertising [5,6].

To address these problems, digital health care applications are spreading online, including self-diagnosis tools and electronic symptom checkers (ESCs) [7-10]. These are meant to provide solutions and information to the user seeking to learn more about symptoms or a condition they have or think they might have. In cases where access to human health care experts may be limited, telehealth services have tremendous promise for transforming the provision of health care services [11]. Conversely, studies find that healthier users use digital services more often than others and are also more likely to be younger, female, and more highly educated, and to have higher income levels [10,12,13].

Based on the user's input, ESCs use algorithms to make diagnostic suggestions, offer advice on what action to take, and help in identifying the relevant condition. This is medical triage, and it involves directing patients to the most suitable location within an appropriate time frame. In clinical practice, triage assessment and guidance are usually done by health care professionals either over the phone or face to face, for example, at a health care center [14]. Triage takes up a lot of professionals' time and its quality varies. Therefore, even the partial digitalization of triage in health care organizations could increase service uniformity, enhance efficiency, and free up working hours [15,16]. This inherently requires that health organizations and teams reorganize their workflows and work distributions to support clinical processes [17,18].

The Omaolo ESC questionnaires and algorithms in use are based on research evidence, probabilities, and expert opinions as to whether the condition described is mild and self-limiting. In terms of urgency, an assessment is made on how soon the condition would worsen without treatment or whether it requires the intervention of a health care professional. However, as with clinical decision-making in general, making an accurate diagnosis requires user-provided information, clinical examinations, various diagnostic tests, and potential consultations with other health care professionals [19,20].

Previous studies have evaluated the accuracy of the diagnostics of ESCs and triage using clinical case vignettes [21-27]. Variation exists between different ESCs, and the conditions being assessed, including the triage capabilities [8,27-29]. In some studies, the diagnostic accuracy of clinicians has been shown to be superior in both primary and specialized health care when compared with ESC tools [22,23]. These studies have shown that users may be referred to as self-care even if they need professional help, and users for whom self-care would suffice are referred to unnecessary counseling. There are risks and the potential for error in ESC-based triage [8,21-29]. In particular, self-care guidance should be limited to cases where it is safe and appropriate. There is currently a limited amount of evidence available on the impact of ESCs on seeking treatment with real-life users [12,30]. However, respondents were satisfied with the ESC services they use [13,31,32].

A study comparing the accuracy of physicians' and computer diagnostics found that physicians listed the correct diagnosis first more often across all study vignettes compared with ESCs (79.1%-65.3% vs 40.5%-24.3%; P <.001) as well as in the top 3 diagnoses listed (84.3% vs 51.2%; P <.001) [24]. There is limited evidence of live clinical patient safety hazards associated with the use of ESCs, as safety has mainly been evaluated with the use of clinical vignettes [7,8,21,23,24,26]. When comparing AI and human doctors concerning triage and diagnosis, some AI systems were able to provide triage and diagnostic information on a level of clinical accuracy and safety comparable to human doctors [8,23,33]. However, ESCs on average make the user's triage more sensitive to the need for more urgent care than the user would need [22,34].

The seamless integration of ESCs into the broader health care triage process is crucial for achieving their intended goals, such as preventing emergency departments overcrowding and providing more accurate symptom assessment and triage for citizens. ESCs can offer citizens a preliminary triage level for their symptoms before contacting health care services [35]. In addition, ESCs and eHealth applications can serve an educational purpose by providing users with structured, research-based disease and treatment information that is easily accessible [36,37]. From a clinical perspective, the ability to accurately identify cases where self-care suffices is paramount in assessing an app's utility in preventing overcrowding and the "unnecessary use of healthcare services" [38].

Description of the Omaolo Electronic Symptom Checkers

Omaolo is a national web-based service for health care and social welfare. The purpose of Omaolo is to promote the health and well-being of citizens. Omaolo supports self-care and helps people to contact public health care professionals, if necessary. Omaolo is a medical device with a Conformité Européenne marking, manufactured by government-owned DigiFinland and Duodecim Publishing Company Ltd. Omaolo was granted a CE certificate in accordance with the requirements of European Union Regulation 2017/745 (Medical Device Regulation) in May 2022 [39]. The aim of the Omaolo ESC is to identify, based on an assessment of alarm symptoms and other pre-existing conditions, situations that require immediate or urgent assessment and to conduct follow-up examinations and treatment without delay in situations where conservative treatment may lead to complications. The questionnaires and the algorithms the ESC uses are based on evidence and legal requirements [40].

The ESC operates as [41] the user initially receives reliable information about the symptom (articles from the Health Library Duodecim) and a short summary. If unable to decide on the course of action needed, the user can answer the ESC's questions. The ESC will then suggest the estimated needed treatment and its urgency. The results of the completed survey made by identified users are saved and prompted to be sent to a regional health care professional through Omaolo. The ESC algorithm initially seeks to identify alarming symptoms and then prompt the user to contact the nearest emergency department immediately. The idea is to identify situations where a professional assessment is necessary and to determine the urgency of the assessment. The user is encouraged to consider whether they may have symptoms that are not covered by the information or survey provided. As the questionnaires might not cover all possible situations that could be due to other illnesses, treatments, or other causes that the user may have, the following help text is displayed to the user at the end of the query: “If you have symptoms that have not been covered in the survey or other illnesses or medications that you think affect your need for treatment, contact your PHC provider or, in an emergency, the nearest emergency department.” [41]. Description of the Omaolo service is described in detail in the study protocol [40].

Objectives

We hypothesize that the Omaolo ESC assessments are safe to use compared with the assessment of a triage nurse.

The aim is to study the clinical validity of the Omaolo ESC and to evaluate its exact triage accuracy, specificity, sensitivity, and safety. These parameters can be used to determine if Omaolo ESC can direct the right patient user to the right place at the right time. The main research question was: how well does triage by the ESC match the triage of a nurse [40] (Multimedia Appendix 1)?

Methods

Study Design and Setting

The study setting was primary health care (PHC), and the data were collected at 18 PHC centers nationwide in Finland. The data were collected between June 1, 2018, and December 31, 2020. The study used the version of the Omaolo ESC that was in use in 2018. Each participating organization provided a study space where it is practically possible for the users (patients) to complete the ESC questionnaire and a nurse with at least 2 years' experience of triage nurse work in primary care to perform triage (Textbox 1).

Textbox 1. Nurse's form questions.

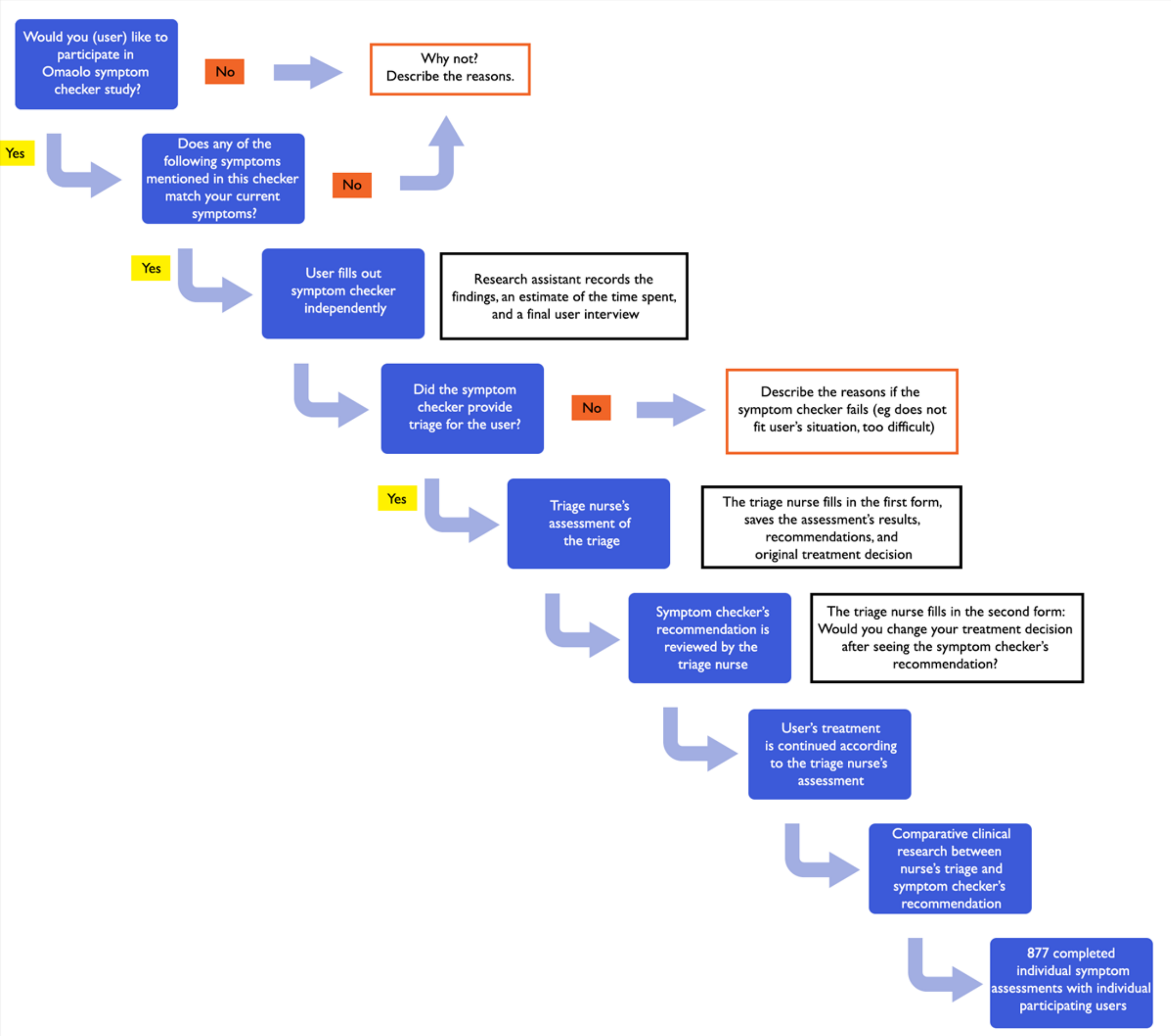
- How did the user arrive at the reception (walk-in or via telephone contact)?
- Did you consult a doctor to assess the user's need for treatment (triage)?
- What was the most significant thing (observation, symptom, or discovery) that influenced your decision-making?
- Do you feel the need to change your assessment of the need for treatment (triage) after seeing the responses and recommendations of the electronic symptom checker? (yes or no)
- If yes, why did you change the assessment of the need for treatment (triage) after seeing the responses and recommendation of the electronic symptom checker?
- If you feel it is necessary to change your triage assessment, reselect where the user should be referred to according to the classification terms of the electronic symptom checker recommendation.

A total of 119 individual nurses took part in this study. A nurse's average age was 40 (SD 10.5) years (median 37, IQR 33-47 years), and their average amount of work experience in triage was 9.9 (SD 8.5) years (median 7, IQR 4-14 years).

In total, 3 in every 10 recruited patient users arrived through walk-in at the PHC centers, and the rest first contacted their center through telephone. Upon physically arriving at the PHC centers, the patients were asked if they were willing to participate in the study.

The patients answered the ESC when they arrived at the PHC center (on arrival at the center, they also filled out a consent form). Filling out the questions of the ESC was done in a separate quiet space without the research assistant interfering. Next, a triage nurse made a triage assessment of the same patient and filled out the study questionnaire related to the assessment of the patient's triage. The triage nurses did not get to know the result of the ESC until they had assessed the patient's condition themselves. After completing their questionnaire, the triage nurses finally got to see the results of the ESC triage concerning the same patient. Based on that result, the triage nurses filled out another questionnaire inquiring whether they felt it necessary to change their own assessment-based action recommendation after seeing the action recommendation of the ESC. The organization also ensured that the patient population of the study remained unscreened. No user-identifying age or gender data were collected for this study [40] (Figure 1).

Figure 1. Flowchart showing the collection of the symptom checker and nurse triage data.



Assessment of Electronic Symptom Checker Coverage and Triage Suggestion Levels

The results of the ESC triage and the assessments of nurses were analyzed from the completed study forms. Each assessment was first analyzed individually, and the results concerning particular symptom assessments were combined.

The Ministry of Social Affairs and Health (MSAH) has provided a practical classification for levels of emergency and the Finnish Institute for Health and Welfare (THL) referral class classification with codes [42] (Figures 2-4).

Figure 2. An example showing how the triage comparison chart (confusion matrices) in Figure 3 was constructed with color-coded differences in triage levels (overly conservative, overly conservative but suitable, exactly matched, safe but under conservative, and potentially unsafe [resolved]). Matching rows (triage nurse) to their respective columns (symptom assessment) results in a safety assessment. P0-P4=classification of emergency care criteria. L2-4=referral urgency classes. The columns show how often, by ESC, the symptom assessment was overestimated, underestimated, or accurately matched compared with the assessment made by the triage nurse and the decision made on further referral. ESC: electronic symptom checkers; FN: false negative; FP: false positive; TN true negative; TP: true positive.

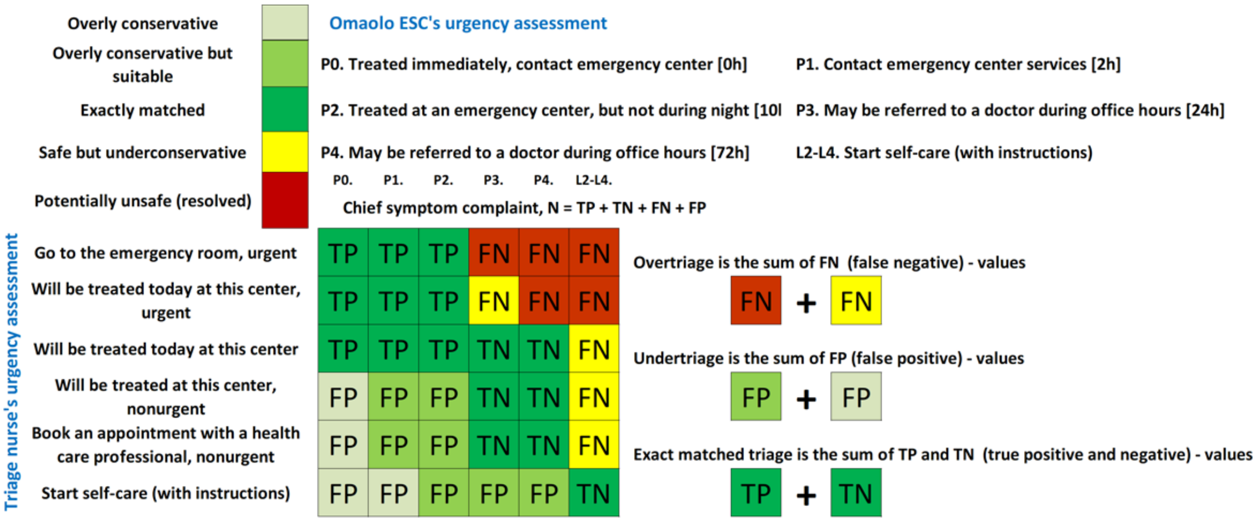


Figure 3. Triage comparison chart (confusion matrices) with color-coded differences in triage levels. ESC: electronic symptom checker.

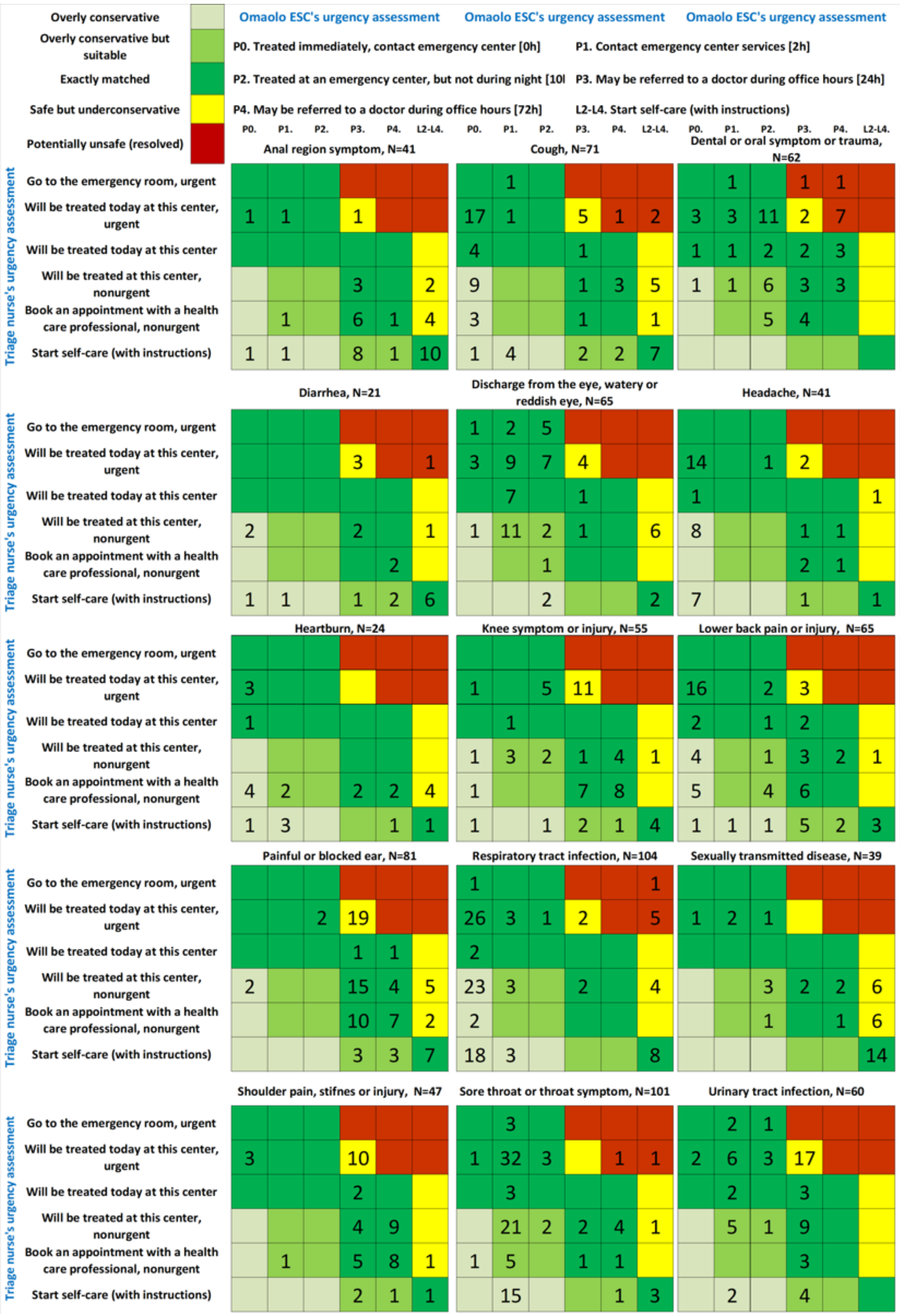
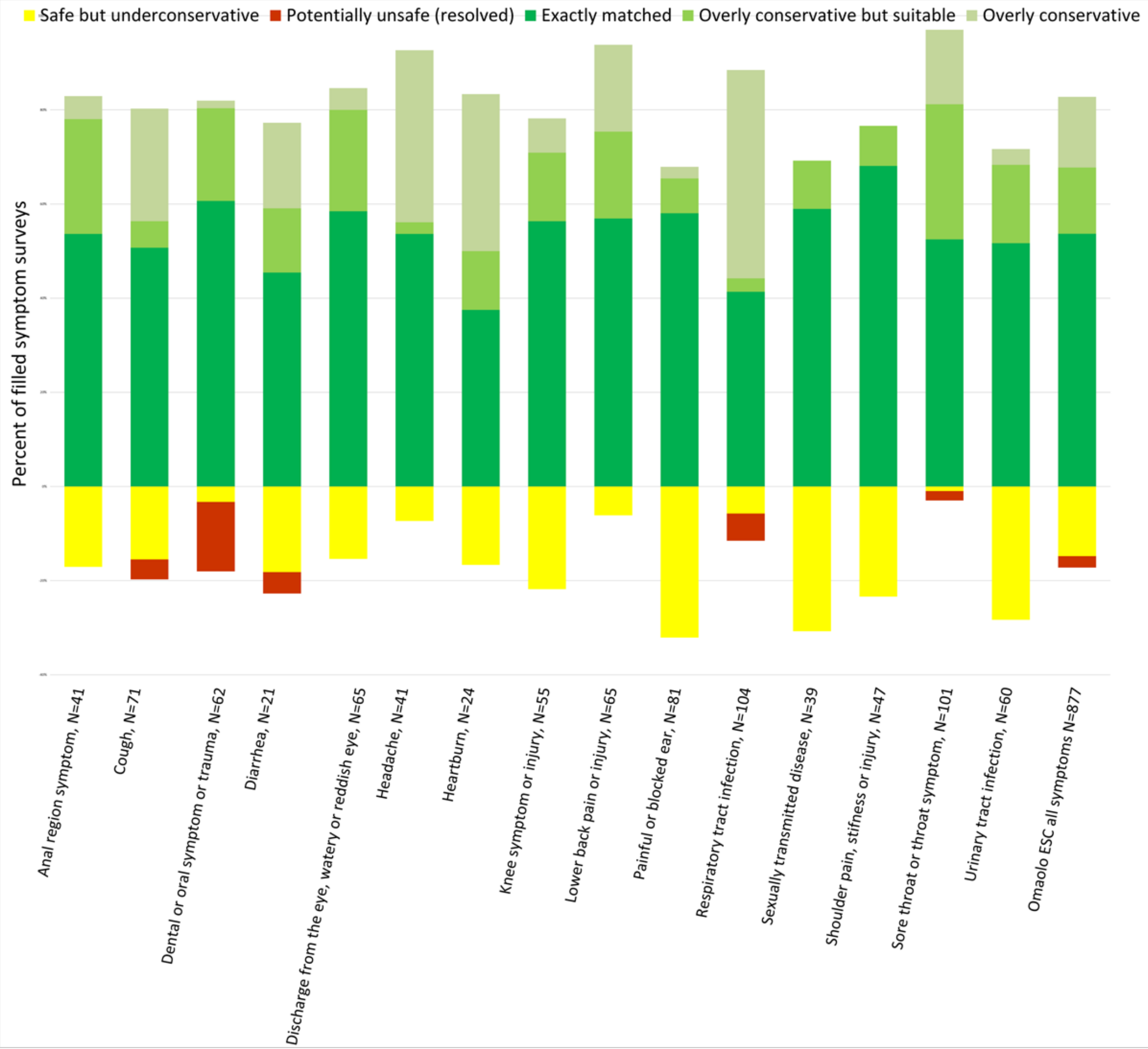


Figure 4. The individual bar heights (y-axis, %) reflect the proportional correspondence and accuracy levels of the electronic symptom checker triage with the nurse's triage (overly conservative, overly conservative but suitable, exactly matched, safe but under conservative, and potentially unsafe [resolved]). Individual electronic symptom checkers are depicted on the x-axis.



These classification levels are as follows ranging from P0 to P4 and L2 to L4:

- P0: Treated immediately, contact emergency center (within 0 h)
- P1: Contact emergency center services (within 2 h)
- P2: Treated at an emergency center, but not during the night (within 10 h)
- P3: May be referred to a doctor during office hours (within 24 h)
- P4: May be referred to a doctor during office hours (within 72 h)
- L2-L4: Start self-care (with instructions)

Making use of the MSAH emergency levels, the Omaolo ESC and nurse triage were matched as, (1) exactly matched (ESC's and triage nurse's triage were the same), (2) overly conservative but suitable (ESC's assessment was 1 triage level higher than the nurse's triage), (3) safe but under conservative (ESC's triage level was 1 triage level lower than the nurse's triage), (4) overly conservative (ESC's triage level was 2 levels higher than the nurse's triage), and (5) potentially unsafe (triage nurse assessed as urgent or an on-call duty but ESC assessed as nonurgent or self-care).

Assigning Accuracy Gold Standards to the Triage Nurse and the Electronic Symptom Checker

The nurse triage assessment is the gold standard, to which the recommendations of the ESC were compared. An assessment was also defined as potentially unsafe if a case was assessed by the nurse as urgent or an on-call duty but assessed by the ESC as nonurgent or self-care. These assessments were further investigated to ensure patient safety.

An external expert group was founded to analyze possible critical assessments and user safety. The group included one Omaolo developer nurse, 3 physicians with specialist degrees in general practice, 1 physician with a specialist degree in public health, 3 personnel working in the ESC development department, and 1 coder responsible for the Omaolo ESC production.

Statistical Methods and Metrics for Assessing Triage Accuracy

Triage accuracy was calculated as the percentages of matches for each individual ESC and 95% CI. The calculations were based on the initial nurses' triage assessments. The hypothesis is that the Omaolo ESC is safe to use [40].

We set the safe symptom assessment based on a previous study, as at most one level of urgency less urgent than the assessment of a nurse, and the performance of the ESC as 97% safe accuracy across all completed assessments [23]. We also assumed that misdiagnosis by physicians occurs in approximately 5% of outpatients [43]. We estimated the required sample size by assuming the given range of safe advice at 97% and using a 95% CI, and we computed the CI estimate for the true proportion of safe ESC assessments [23].

Triage comparison charts (confusion matrices) are created in which true positive (TP) represents the outcome where the model correctly predicts the positive class (condition is detected when present). True negative (TN) is the outcome where the model correctly predicts the negative class (condition is not detected when absent). False positive (FP) represents the outcome where the model incorrectly predicts the positive class (condition is detected when absent). False negative (FN) is the outcome where the model incorrectly predicts the negative class (condition is not detected when present). These values are crucial components in calculating the positive predictive values (PPVs), negative predictive values (NPVs), and Matthews correlation coefficient (MCC). The PPV, NPV, and MCC formulas are given as:

These values are typically extracted from a confusion matrix, as illustrated in Figure 3 and presented in Table 1 of the study.

Table 1. Results of matching the nurse triage (gold standard) with the Omaolo electronic symptom checker recommendation in 15 different symptoms[a].

Absolute magnitude of the Matthews correlation coefficient	Interpretation
±0.00-0.10	Negligible correlation
±0.10-0.39	Weak correlation
±0.40-0.69	Moderate correlation
±0.70-0.89	Strong correlation
±0.90-1.00	Very strong correlation

[a]The Omaolo electronic symptom checker recommendation is defined as safe if the critical condition was not assessed as urgent or on-call duty by the nurse but assessed by the electronic symptom checker as nonurgent or self-care. The CI range in the column reporting the number of completed assessments is assumed at safe advice of 97% and using a 95% CI. The table also shows positive predictive values and negative predictive values following the Matthews correlation coefficient.

The MCC ranges from 1 to +1, where ±1 indicates perfect agreement or disagreement, and 0 indicates no relationship.

Ethical Considerations

This study was reviewed by the Pirkanmaa hospital district's ethics committee (ETL-Code: R18126), and regional permission was additionally granted by each participating organization, all according to the regulations of the University of Tampere.

The most significant ethical issue related to the research setting is that the user's participation in the research does not affect their chances of receiving timely treatment. All users who fill out the symptom checker will be forwarded to an appointment with an experienced nurse. Denial of treatment for users who refused to participate in the study was strictly prohibited.

When users (patients) were recruited for this study, the research assistant informed the user about the study, distributed the study information sheet, and then asked if the user was willing to participate in the study. If, after being informed, the user was willing to participate in the study, they were asked to sign a consent form in which the user acknowledged that they had received sufficient information about the study and agreed to participate in it. The user was given an information sheet about the study with contact information in case the user wanted to ask more about the study. The user was paid no amount of compensation.

The patient user's consent form was disconnected from the response form with a personal identification code, that is, the users were completely anonymized. No medical record data were collected or combined with research forms. The users could withdraw their consent to the study at any time. However, the completed forms cannot be destroyed after the data collection because the consent form containing the personal data does not have the identification code to identify the appropriate study forms to be destroyed.

Results

Overview

A total of 877 patient user assessment cases were successfully collected. No patient user identifying age or gender data were collected for this study.

Quantifying and Comparing the Levels of Urgency and Triage Accuracy

The ESC's and nurse's triage were exactly matched in 53.7% (471/877; 95% CI 49.2%-55.9%) of the cases in all symptom assessments. Considering ESC's individual main symptom triage suggestions, the most exactly matched assessments were found for shoulder pain, stiffness, or injury (32/47, 68%), dental or oral symptom or trauma (37/61, 61%), and sexually transmitted disease (23/39, 59%).

The mean value for exactly matched or overly conservative but suitable for all symptom assessments was 66.6% (584/877; 95% CI 63.4%-69.7%). Safe assessments of ESCs accounted for 97.6% (856/877; 95% CI 95.6%-98.0%) of all assessments made (Table 1, Figures 2-4). Concerning acute cases in which the nurse evaluated that a user needed to be treated urgently the ESC's exactly matched accuracy was 70.9% (244/344; 95% CI 65.8%-75.7%), and in cases whether medical care should be sought or self-care is sufficient, matches were found in 65.9% of cases (351/533; 95% CI 61.7%-70.0%). Sensitivity for the Omaolo ESC was 62.6% and specificity was 69.2%. The proportions of evaluations occurred at a ratio of 100 suitable to 25 overtriage to 22 undertriage. The overly conservative triage (overtriage) suggestions made by the ESC occurred most often for respiratory tract infection in 44% (46/104), heartburn in 33% (8/24), and headache in 37% (15/41) of cases (Figure 4).

The question "Do you feel the need to change your assessment of the need for treatment (triage) after seeing the responses and recommendation from the ESC? (Yes/No)" was answered with the "Yes" option in 19 out of 877 assessments. In answering the follow-up question "If yes, why did you change the assessment of the need for treatment (triage) after seeing the responses and recommendation by the ESC?", the nurses frequently stated that the symptoms they were told by the user did not match with the ones the user had stated while filling the ESC (Table 2 and Multimedia Appendix 2). In the last section, in the case of an affirmative answer, the path of changing the triage ended with "If you feel it is necessary to change your triage assessment, reselect where the user should be referred to according to the classification terms of the ESC recommendation." In these 19 assessments, the nurses were found to have chosen a triage suggestion closer to that of the ESC assessment (Multimedia Appendix 2). In 80 cases across all completed 877 assessments, the triage nurse consulted a doctor in assessing the triage.

Table 2. Triage nurse form questions results, the y-axis represents the number of observations.

Symptom	Completed matched assessments, n	Percentage of safely matched advice symptom assessments, % (95% CI)	Exactly matched symptom assessments, % (95% CI)	Positive predictive value (precision), %	Negative predictive value, %	Matthews correlation coefficient
Anal region symptom	41	100 (91.4-100)	53.7 (37.4-69.3)	14.3	74.1	–0.133
Cough	71	95.8 (88.1-99.1)	50.7 (38.6-62.8)	52.2	48.1	0.004
Dental or oral symptoms or trauma	62	85 (74.2-93.1)	59.7 (46.5-72.0)	62.9	57.7	0.204
Diarrhea	21	95.2 (76.2-99.9)	45.5 (24.4-67.8)	0	66.7	–0.37
Discharge from the eye, watery or reddish eye	65	100 (90.0-99.6)	58.5 (45.6-70.6)	66.7	28.6	–0.036
Headache	41	100 (91.4-100)	53.7 (37.4-69.3)	50.0	66.7	0.138
Heartburn	24	100 (85.8-100)	37.5 (18.8-59.4)	26.7	55.6	–0.233
Knee symptom or injury	55	100 (93.5-100)	56.4 (42.3-69.7)	36.8	66.7	0.035
Lower back pain or injury	65	100 (94.5-100)	56.9 (44.0-69.2)	46.7	80.0	0.253
Painful or blocked ear	81	100 (95.6-100)	58.0 (46.5-68.9)	20.0	49.0	–0.115
Respiratory tract infection	104	94.2 (88.0-97.9)	41.3 (31.8-51.4)	40.2	45.4	–0.118
Sexually transmitted disease	39	100 (91.0-100)	59.0 (42.1-74.4)	50.0	61.3	0.093
Shoulder pain, stiffness, or injury	47	100 (92.5-100)	68.1 (52.9-80.9)	42.9	72.5	0.12
Sore throat or throat symptom	101	98 (93.0-99.8)	54.5 (42.3-62.5)	48.3	78.6	0.19
Urinary tract infection	60	100 (94.0-100)	51.7 (38.4-64.8)	57.1	46.9	0.04
In total	877	97.6 (96.4-98.5)	53.7 (50.3-57.1)	49.4	59.5	0.089

The Analysis of Critical Assessments

In total, 21 critical assessments were identified for further analysis. Details are given in Multimedia Appendix 3. Further analysis showed that there were no indications that patient safety was endangered. In Figures 3 and 4, these patient user cases are marked as "Potentially unsafe" (resolved).

Discussion

Principal Findings

The findings suggest that while exact matches of the Omaolo ESC recommendations with the gold standard (nurse triage) occurred in just over half of the cases, nearly all cases were evaluated as safe, with urgency levels being at most 1 level less urgent compared with the nurses' triage. Concerning acute assessments, an exact match was found in nearly 3 out of 4 cases. This study assessed for the first time the safety of the Omaolo ESC within the Finnish PHC context.

Comparison to Previous Work

In a systematic review, the diagnostic accuracy and the triage-making abilities of ESC services such as Ada, Babylon, Buoy, K Health, Mediktor, Symptomate, and Your.MD were compared with those of general practitioners' assessment using clinical case vignettes. The average safe operating recommendations ranged from 90.1% (SD 7.4%) [23]. By contrast, the general practitioner's percentage of safe advice stood at 97.0% (SD 2.5%) [23]. Comparatively, the proportion of safe Omaolo ESC assessments across all investigated cases was 97.6% (856/877) using similar methods for safety assessment. These findings underscore the safety of Omaolo ESC compared with assessments by experienced nurses, particularly notable given the real-life setting of our study. By comparing Omaolo ESC's accuracy to internationally reported results, we can gauge its overall performance and capabilities [8,22-24,44].

A study examining Ada's use by 378 "walk-in" patients in urgent care compared its triage accuracy with that of a triage nurse using the Manchester Triage System, conducted under similar circumstances as the Omaolo ESC's triage accuracy study. Ada exhibited an undertriage rate of 8.9% (34/378) and an overtriage rate of 57.1% (216/378). Out of 378 participants, 344 (91%) were triaged identically or more conservatively, while 34 (8.9%) were undertriaged by the app. The assessment was deemed safe in 94.7% (358/378) of patients when compared with the Manchester Triage System assessment. In the Omaolo ESC study involving 877 users, 726 (83%) were triaged identically or more conservatively, with 151 (17%) being undertriaged. Notably, Omaolo ESC exhibited a 29% (255/877) overtriage rate, with evaluations occurring at a ratio of 100 suitable evaluations to 25 overtriages to 22 undertriages. Compared with Ada, in the Omaolo study, overtriage rates were lower.

A recent systematic review concluded that the median accuracy of studied apps in determining the necessity of emergency care was 80% (IQR 74.6%-86.8%) [20]. For Omaolo ESC's triage of acute cases where a nurse assessed urgent treatment as necessary, exact matches occurred in 244 out of 344 cases, representing 70.9% (95% CI 65.8%-75.7%) of cases, while matches indicating whether medical care should be sought or if self-care is sufficient occurred in 351 out of 533 cases, totaling 65.9% (95% CI 61.7%-70.0%) of cases. This is in line with the international figure of 73.3% (IQR 70.5%-82.3%) [8]. In addition, the median app sensitivity was 51.9% (IQR 40%-78.2%), and the median specificity was 93.3% (IQR 87.4%-96.4%) [8]. Omaolo ESC exhibited a sensitivity of 62.6% and a specificity of 69.2%.

ESC's capability to detect individuals in need of immediate treatment is vital for user safety. In addition to that, an ESC that holds promise for safely assisting in self-triage and that helps prevent overcrowding of emergency departments could bring added value to health care. Notably, concerning Omaolo, the least overtriage was observed in chief symptoms of sexually transmitted diseases (4/39, 10.3%), shoulder pain stiffness or injury (4/47, 8.5%), and painful or blocked ears (8/81, 9.9%). Conversely, more sensitive and risk-averse chief symptom assessments such as headaches (16/41, 39%), heartburn (11/24, 45.8%), and respiratory tract infections (49/104, 47.1%) exhibited higher rates of false positives, raising concerns about overcrowding and possible unnecessary health care service use. However, due to potentially serious conditions, these ESCs are set to be sensitive in order to rule out alarming symptoms.

A relatively high PPV was found in cough, dental or oral symptoms or trauma, discharge from the eye or watery or reddish eye, and urinary tract infection assessments. This indicates reduced false positives, beneficial when false positives have high costs or the condition is not severe. High PPV minimizes overtreatment and unnecessary costs. Conversely, a moderate PPV found from other assessments is acceptable if follow-up tests are inexpensive and harmless precautionary measures are taken. For the symptom assessments of anal region symptoms, diarrhea, heartburn, headache, and sexually transmitted diseases a low number of cases has to be taken into consideration when evaluating these values.

A relatively high NPV was found in assessments of anal region symptoms, dental or oral symptoms or trauma, diarrhea, headache, heartburn, knee symptoms or injury, lower back pain or injury, sexually transmitted diseases, shoulder pain and stiffness or injury, and sore throat or symptom. High NPV is crucial for serious or contagious conditions needing early intervention, minimizing false negatives. Moderate NPV observed in this study for cough, discharge from the eye, watery or reddish eye, painful or blocked ear, respiratory tract infection, and urinary tract infection is acceptable if the condition is not severe or resolves on its own.

The MCC shows weak positive relationships for dental or oral symptoms or trauma and lower back pain or injury and weak negative relationships for heartburn and diarrhea. Other assessments have negligible relationships, suggesting the symptom checker's predictions are slightly better than random guessing.

Strengths and Limitations of the Study

The real-life setting presents both strengths and potential weaknesses for this study. There are notable concerns regarding potential selection bias. First, users completing the ESC in PHC center waiting rooms may experience different symptoms compared with those using the ESC outside of such environments. Moreover, while the Omaolo ESC is designed for users over the age of 15 years, only individuals over the legal age of 18 years were recruited for this study, potentially limiting the generalizability of findings. Furthermore, users who were unable to independently complete the ESC due to technological limitations (such as difficulty using a computer mouse or tablet devices) were excluded from the study, introducing another potential limitation.

In addition, the rarity of users with serious acute symptoms in this setting may skew the study results, as their symptoms may not be severe enough to interfere with ESC questionnaire completion. However, it is worth noting that the Omaolo ESC prompts users to urgently contact health care services if they are unable to complete the questionnaire due to the severity of their symptoms.

Furthermore, potential users with mild, self-treatable symptoms may have been excluded as such cases may remain unreported in this study context. Nonetheless, the study focused on completed ESC triage assessment accuracy, specificity, and sensitivity.

In some instances, the triage classification made by nurses may have been influenced by user-reported needs for sick leave certificates, potentially biasing the calculated ESC accuracy.

The data accumulation process was hindered by the COVID-19 pandemic, which slowed down research activities [45,46]. However, the exponential growth in individual Omaolo users during the pandemic, particularly with the use of the coronavirus ESC, was noted [47]. To address the scarcity of completed assessments for some ESCs and urgent cases, future research will supplement data with electronic patient cases, known as case vignettes. This approach will also allow for the assessment of ESC performance across rare symptoms and conditions in a primary care setting.

Finally, it is important to acknowledge that, for this study, the nurse's triage assessment was considered the gold standard. An alternative approach could have been to use outcomes during follow-up, such as revisits or hospitalizations after 30 days, to determine the success of the triage.

Implications and Future Research

In this study, heartburn and diarrhea were relatively infrequent as chief symptoms for safety and accuracy evaluations. However, it is crucial to monitor and supplement the safety assessments of these symptoms in future case vignette studies. Despite their limited representation in this study, the present results suggest that heartburn and diarrhea are likely safe to assess using the ESC. To address more common and less urgent situations, vignette studies will be instrumental. These vignettes, sourced from various contexts, will undergo thorough testing to ensure that individual vignettes' difficulty and correlation with overall assessment are carefully considered [48].

Moving forward, there should be an evolution toward more standardized methodologies and studies tailored to specific settings. Regulation and standardization of evaluations are vital for ensuring the transferability of findings across different contexts [49]. In addition, adopting a patient-centered approach is essential for evaluating ESCs effectively. A standardized process with clear specifications for vignette-based clinical evaluation is necessary to guide developers and facilitate objective comparisons among ESCs. This approach will enhance the reliability and validity of ESC assessments and promote their widespread adoption in clinical practice [43,50].

The data suggests that the Omaolo ESC is reliable for preliminary symptom assessment and triage, demonstrating a high level of safety in aligning with triage recommendations from experienced nurses, especially in acute cases. Omaolo uses structured questions and fixed algorithms designed by professionals to provide medically qualified recommendations, though it does not use AI. The question of whether AI-based ESC systems would be a desirable advancement or introduce additional risk and uncertainty for patient safety is complex.

AI-based ESC systems have the potential to enhance efficiency and accuracy by continuously learning from vast datasets and adapting to evolving medical knowledge. They can rapidly process information and offer consistent assessments across different users and contexts, potentially covering a broader range of symptoms and conditions. However, there are risks associated with AI-based systems, including reliance on data quality and a potential lack of nuanced clinical judgment compared with human triage nurses.

While AI-based ESC systems can complement triage processes, human supervision, and oversight are essential to ensure patient safety. Human triage nurses provide contextual understanding, empathy, and critical thinking that AI systems may lack, intervening when AI-generated recommendations are uncertain or potentially harmful. Therefore, AI-based systems should be viewed as tools to augment rather than replace human triage nurses.

In conclusion, ESCs should augment traditional triage rather than substitute for it, potentially leading to benefits such as reduced phone calls and increased accessibility to health services. Omaolo ESC, with its acceptable specificity and accuracy, holds promise for preventing unnecessary use of primary health care services. In addition, well-structured ESC assessments systematically collect user medical history and symptom information, evidenced by triage nurses' decisions to adjust triage based on additional user-provided information.

Impacts of the Study on the Omaolo Electronic Symptom Checker Service

The results highlighting the safety of Omaolo have been crucial for the continuation of the ESC service in Finnish health care.

Conclusions

The primary objectives of this study were to evaluate the safety and provide essential insights into the accuracy, specificity, and sensitivity of the Omaolo ESC. The results indicate that the ESC is safe for use compared with assessments conducted by triage nurses. Furthermore, the Omaolo ESC exhibits the potential to guide patients to appropriate triage destinations aptly, ensuring they receive timely and suitable care.

Acknowledgments

This research was carried out in addition to the work of primary health care centers, municipalities, health care district–based health care organizations nationwide, Omaolo project employees, and Finnish Medical Society Duodecim employees.

Data Availability

The electronic datasets generated during and analyzed in this study are available from the corresponding author upon reasonable request. Paper forms used to create these datasets are stored in a locked cabinet. The consent form containing the user's name and signature is kept in a separate locked cabinet. All paper forms are stored in the Tampere University. The forms are disposed of with a shredder 10 years after the end of the study.

Conflicts of Interest

MK is the president of the Finnish Medical Society Duodecim. TK worked as a salaried part-time editor for Duodecim Medical publications until 2018. VL works in the MD PhD program of the Faculty of Medicine at the University of Helsinki.

Multimedia Appendix 1
Quantitative clinical study.
[DOCX File , 15 KB - humanfactors_v11i1e55099_app1.docx]

Multimedia Appendix 2
Details of affirmative answers to "do you feel the need to change your assessment of need for treatment (triage) after seeing the responses and recommendation from the electronic symptom checker?".
[DOCX File , 18 KB - humanfactors_v11i1e55099_app2.docx]

Multimedia Appendix 3
Identified conflict situations (21 critical cases). The findings of the inspection, the identified conflict, the final result, and the declared final result (assessed by the evaluation team).
[DOCX File , 20 KB - humanfactors_v11i1e55099_app3.docx]

References

1. Hochberg I, Allon R, Yom-Tov E. Assessment of the frequency of online searches for symptoms before diagnosis: analysis of archival data. J Med Internet Res 2020 Mar 06;22(3):e15065 [FREE Full text] [doi: 10.2196/15065] [Medline: 32141835]
2. Sebastian C, Ahmed M, Guido Z, Bevan K. Search engines vs. symptom checkers: a comparison of their effectiveness for online health advice. 2021 Presented at: Proceedings of the Web Conference 2021; June 3, 2021; Ljubljana, Slovenia p. 206-216.
3. EPatient survey. Health&Care Management. 2020. URL: https://www.hcm-magazin.de/epatient-survey-2020-digital-health-studies/150/10992/407743 [accessed 2021-03-06]
4. Aboueid S, Meyer S, Wallace JR, Mahajan S, Chaurasia A. Young adults' perspectives on the use of symptom checkers for self-triage and self-diagnosis: qualitative study. JMIR Public Health Surveill 2021 Jan 06;7(1):e22637 [FREE Full text] [doi: 10.2196/22637] [Medline: 33404515]
5. Tasnim S, Hossain MM, Mazumder H. Impact of rumors and misinformation on COVID-19 in social media. J Prev Med Public Health 2020;53(3):171-174 [FREE Full text] [doi: 10.3961/jpmph.20.094] [Medline: 32498140]
6. McLellan J, Heneghan C, Roberts N, Pluddemann A. Accuracy of self-diagnosis in conditions commonly managed in primary care: diagnostic accuracy systematic review and meta-analysis. BMJ Open 2023;13(1):e065748 [FREE Full text] [doi: 10.1136/bmjopen-2022-065748] [Medline: 36627158]
7. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020;3:4. [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]
8. Schmieding ML, Kopka M, Schmidt K, Schulz-Niethammer S, Balzer F, Feufel MA. Triage Accuracy of Symptom Checker Apps: 5-Year Follow-up Evaluation. J Med Internet Res 2022 May 10;24(5):e31810 [FREE Full text] [doi: 10.2196/31810] [Medline: 35536633]
9. Battineni G, Sagaro GG, Chinatalapudi N, Amenta F. Applications of machine learning predictive models in the chronic disease diagnosis. J Pers Med 2020;10(2):21 [FREE Full text] [doi: 10.3390/jpm10020021] [Medline: 32244292]
10. Morse KE, Ostberg NP, Jones VG, Chan AS. Use characteristics and triage acuity of a digital symptom checker in a large integrated health system: population-based descriptive study. J Med Internet Res 2020;22(11):e20549 [FREE Full text] [doi: 10.2196/20549] [Medline: 33170799]

11. van Veen T, Binz S, Muminovic M, Chaudhry K, Rose K, Calo S, et al. Potential of mobile health technology to reduce health disparities in underserved communities. West J Emerg Med 2019;20(5):799-802 [FREE Full text] [doi: 10.5811/westjem.2019.6.41911] [Medline: 31539337]
12. Rodgers M, Raine G, Thomas S, Harden M, Eastwood A. Informing NHS policy in 'digital-first primary care': a rapid evidence synthesis. Southampton (UK): NIHR Journals Library 2019. [Medline: 31869020]
13. Prepared by ipsos MORI and york health economics consortium with prof. chris salisbury for NHS hammersmith and fulham CCG and NHS England. Evaluation of Babylon GP at hand Final evaluation report. 2019. URL: http://allcatsrgrey.org.uk/wp/download/informatics/Evaluation-of-Babylon-GP-at-Hand-Final-Report.pdf [accessed 2024-08-17]
14. NHS 111 minimum data set, England, March 2021. NHS England. URL: https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2021/04/NHS-111-MDS-Apr-2021-Statistical-Note.pdf [accessed 2024-08-17]
15. Aboueid S, Meyer SB, Wallace JR, Mahajan S, Nur T, Chaurasia A. Use of symptom checkers for COVID-19-related symptoms among university students: a qualitative study. BMJ Innov 2021;7(2):253-260. [doi: 10.1136/bmjinnov-2020-000498] [Medline: 34192014]
16. Judson TJ, Odisho AY, Neinstein AB, Chao J, Williams A, Miller C, et al. Rapid design and implementation of an integrated patient self-triage and self-scheduling tool for COVID-19. J Am Med Inform Assoc 2020;27(6):860-866 [FREE Full text] [doi: 10.1093/jamia/ocaa051] [Medline: 32267928]
17. Laukka E, Kujala S, Gluschkoff K, Kanste O, Hörhammer I, Heponiemi T. Leaders' support for using online symptom checkers in Finnish primary care: Survey study. Health Informatics J 2021;27(4):14604582211052259 [FREE Full text] [doi: 10.1177/14604582211052259] [Medline: 34821152]
18. Kujala S, Hörhammer I. Health care professionals' experiences of web-based symptom checkers for triage: cross-sectional survey study. J Med Internet Res 2022;24(5):e33505 [FREE Full text] [doi: 10.2196/33505] [Medline: 35511254]
19. Thompson C, Dowding D. My library My History Essential Decision Making and Clinical Judgement for Nurses. Netherlands: Elsevier Health Sciences; 2009.
20. National Academies of Sciences, Engineering, and Medicine. In: Balogh EP, Miller BT, Ball JR, editors. Improving Diagnosis in Health Care. Washington, DC: The National Academies Press; 2015.
21. Ceney A, Tolond S, Glowinski A, Marks B, Swift S, Palser T. Accuracy of online symptom checkers and the potential impact on service utilisation. PLoS One 2021;16(7):e0254088 [FREE Full text] [doi: 10.1371/journal.pone.0254088] [Medline: 34265845]
22. Semigran HL, Linder J, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ 2015;351:h3480 [FREE Full text] [doi: 10.1136/bmj.h3480] [Medline: 26157077]
23. Gilbert S, Mehl A, Baluch A, Cawley C, Challiner J, Fraser H, et al. How accurate are digital symptom assessment apps for suggesting conditions and urgency advice? A clinical vignettes comparison to GPs. BMJ Open 2020;10(12):e040269 [FREE Full text] [doi: 10.1136/bmjopen-2020-040269] [Medline: 33328258]
24. Semigran HL, Levine DM, Nundy S, Mehrotra A. Comparison of physician and computer diagnostic accuracy. JAMA Intern Med 2016;176(12):1860-1861. [doi: 10.1001/jamainternmed.2016.6001] [Medline: 27723877]
25. Shen C, Nguyen M, Gregor A, Isaza G, Beattie A. Accuracy of a popular online symptom checker for ophthalmic diagnoses. JAMA Ophthalmol 2019;137(6):690-692 [FREE Full text] [doi: 10.1001/jamaophthalmol.2019.0571] [Medline: 30973602]
26. Munsch N, Martin A, Gruarin S, Nateqi J, Abdarahmane I, Weingartner-Ortner R, et al. Diagnostic accuracy of web-based COVID-19 symptom checkers: comparison study. J Med Internet Res 2020;22(10):e21299 [FREE Full text] [doi: 10.2196/21299] [Medline: 33001828]
27. Wallace W, Chan C, Chidambaram S, Hanna L, Iqbal FM, Acharya A, et al. The diagnostic and triage accuracy of digital and online symptom checker tools: a systematic review. NPJ Digit Med 2022;5(1):118 [FREE Full text] [doi: 10.1038/s41746-022-00667-w] [Medline: 35977992]
28. Yoshida Y, Thomas Clark G. Accuracy of online symptom checkers for diagnosis of orofacial pain and oral medicine disease. J Prosthodont Res 2021;65(2):186-190 [FREE Full text] [doi: 10.2186/jpr.JPOR_2019_499] [Medline: 32938875]
29. Yu SWY, Ma A, Tsang VHM, Chung LSW, Leung S, Leung L. Triage accuracy of online symptom checkers for accident and emergency department patients. Hong Kong J Emerg Med 2020;27(4):217-222. [doi: 10.1177/1024907919842486]
30. El-Osta A, Webber I, Alaa A, Bagkeris E, Mian S, Taghavi Azar Sharabiani M, et al. What is the suitability of clinical vignettes in benchmarking the performance of online symptom checkers? An audit study. BMJ Open 2022;12(4):e053566 [FREE Full text] [doi: 10.1136/bmjopen-2021-053566] [Medline: 35477872]
31. Meyer AND, Giardina TD, Spitzmueller C, Shahid U, Scott TMT, Singh H. Patient perspectives on the usefulness of an artificial intelligence-assisted symptom checker: cross-sectional survey study. J Med Internet Res 2020;22(1):e14679 [FREE Full text] [doi: 10.2196/14679] [Medline: 32012052]
32. Liu VD, Sellgren L, Kaila M, Koskela T. Usability of online symptom checkers. Finnish Journal of eHealth and eWelfare 2021;13(2):100-112 [FREE Full text]
33. Baker A, Perov Y, Middleton K, Baxter J, Mullarkey D, Sangar D, et al. A comparison of artificial intelligence and human doctors for the purpose of triage and diagnosis. Front Artif Intell 2020;3:543405 [FREE Full text] [doi: 10.3389/frai.2020.543405] [Medline: 33733203]

34. Hill M, Sim M, Mills B. The quality of diagnosis and triage advice provided by free online symptom checkers and apps in Australia. Med J Aust 2020;212(11):514-519. [doi: 10.5694/mja2.50600] [Medline: 32391611]
35. Ommundsen Y, Roberts GC, Kavussanu M. Perceived motivational climate and cognitive and affective correlates among Norwegian athletes. J Sports Sci 1998;16(2):153-164. [doi: 10.1080/026404198366867] [Medline: 9531004]
36. Arellano Carmona K, Chittamuru D, Kravitz RL, Ramondt S, Ramírez AS. Health information seeking from an intelligent web-based symptom checker: cross-sectional questionnaire study. J Med Internet Res 2022;24(8):e36322 [FREE Full text] [doi: 10.2196/36322] [Medline: 35984690]
37. Painter A, Hayhoe B, Riboli-Sasco E, El-Osta A. Online symptom checkers: recommendations for a vignette-based clinical evaluation standard. J Med Internet Res 2022;24(10):e37408 [FREE Full text] [doi: 10.2196/37408] [Medline: 36287594]
38. Schmieding ML, Mörgeli R, Schmieding MAL, Feufel MA, Balzer F. Benchmarking triage capability of symptom checkers against that of medical laypersons: survey study. J Med Internet Res 2021;23(3):e24475 [FREE Full text] [doi: 10.2196/24475] [Medline: 33688845]
39. DigiFinland Oy. URL: https://digifinland.fi/wp-content/uploads/2023/07/FI22_871876_Iss4_DigiFinland_Oy_EU-QMS_2023-05-31.pdf [accessed 2024-08-17]
40. Liu V, Koskela TH, Kaila M. User-initiated symptom assessment with an electronic symptom checker: protocol for a mixed methods validation study. JMIR Res Protoc 2023;12:e41423 [FREE Full text] [doi: 10.2196/41423] [Medline: 37467041]
41. Omaolo instructions for use. Omaolo. URL: https://www.omaolo.fi/kayttoohjeet/omaolo-instructions.pdf [accessed 2024-08-17]
42. Omaolo symptom checker triage classifications. DigiFinland. URL: https://digifinland.fi/omaolohelp/verkkokoulutus/2-oirearviot/2-1-oirearvio/prioriteetti-eli-kiireellisyys/ [accessed 2024-08-17]
43. Singh H, Meyer AND, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. BMJ Qual Saf 2014;23(9):727-731 [FREE Full text] [doi: 10.1136/bmjqs-2013-002627] [Medline: 24742777]
44. Cotte F, Mueller T, Gilbert S, Blümke B, Multmeier J, Hirsch MC, et al. Safety of triage self-assessment using a symptom assessment app for walk-in patients in the emergency care setting: observational prospective cross-sectional study. JMIR Mhealth Uhealth 2022;10(3):e32340 [FREE Full text] [doi: 10.2196/32340] [Medline: 35343909]
45. Valmorri L, Vertogen B, Zingaretti C, Miserocchi A, Volpi R, Clemente A, et al. Clinical research activities during COVID-19: the point of view of a promoter of academic clinical trials. BMC Med Res Methodol 2021;21(1):91 [FREE Full text] [doi: 10.1186/s12874-021-01291-0] [Medline: 33931012]
46. Nguyen HL, Tran OT, Ha DA, Phan VH, Nguyen CT, Nguyen GH, et al. Impact of the COVID-19 pandemic on clinical research activities: survey of study participants and health care workers participating in a hypertension trial in Vietnam. PLoS One 2021;16(7):e0253664 [FREE Full text] [doi: 10.1371/journal.pone.0253664] [Medline: 34264973]
47. Omaolo-palvelu laajenee tämän vuoden aikana kattamaan lähes koko Suomen. URL: https://digifinland.fi/omaolo-palvelu-laajenee-taman-vuoden-aikana-kattamaan-lahes-koko-suomen/ [accessed 2024-08-17]
48. Kopka M, Feufel MA, Berner ES, Schmieding ML. How suitable are clinical vignettes for the evaluation of symptom checker apps? A test theoretical perspective. Digit Health 2023;9:20552076231194929 [FREE Full text] [doi: 10.1177/20552076231194929] [Medline: 37614591]
49. Pairon A, Philips H, Verhoeven V. A scoping review on the use and usefulness of online symptom checkers and triage systems: how to proceed? Front Med (Lausanne) 2022;9:1040926 [FREE Full text] [doi: 10.3389/fmed.2022.1040926] [Medline: 36687416]
50. Riboli-Sasco E, El-Osta A, Alaa A, Webber I, Karki M, El Asmar ML, et al. Triage and diagnostic accuracy of online symptom checkers: systematic review. J Med Internet Res 2023;25:e43803 [FREE Full text] [doi: 10.2196/43803] [Medline: 37266983]

Abbreviations

AI: artificial intelligence
ESC: electronic symptom checker
FIHW: Finnish Institute for Health and Welfare
FN: false negative
FP: false positive
GP: general practitioner
MSAH: Finnish Ministry of Social Affairs
NPV: negative predictive values
PHC: primary health care
PPV: positive predictive values
THL: Finnish Institute for Health and Welfare
TN: true negative
TP: true positive

Edited by A Parush; submitted 02.12.23; peer-reviewed by B Mueller, H Sievanen; comments to author 20.03.24; revised version received 13.05.24; accepted 16.07.24; published 26.09.24.

Please cite as:
Liu V, Kaila M, Koskela T
Triage Accuracy and the Safety of User-Initiated Symptom Assessment With an Electronic Symptom Checker in a Real-Life Setting: Instrument Validation Study
JMIR Hum Factors 2024;11:e55099
URL: https://humanfactors.jmir.org/2024/1/e55099
doi:10.2196/55099
PMID:

Original Paper

User Requirements for an Electronic Patient Recruitment System: Semistructured Interview Analysis After First Implementation in 3 German University Hospitals

Alexandra Stein[1*], Dipl Jur; Romina Blasini[2*], MSc; Cosima Strantz[3], MSc; Kai Fitzer[4], MSc; Christian Gulden[3], Dr; Torsten Leddig[1], Dr rer nat; Wolfgang Hoffmann[1], Prof Dr Med

[1]Institute for Community Medicine, Section Epidemiology of Health Care and Community Health, University Medicine Greifswald, Greifswald, Germany

[2]Institute of Medical Informatics, Justus Liebig University, Giessen, Germany

[3]Medical Informatics, Institute for Medical Informatics, Biometrics and Epidemiology, Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany

[4]Core Unit Data Integration Center, University Medicine Greifswald, Greifswald, Germany

[*]these authors contributed equally

Corresponding Author:
Alexandra Stein, Dipl Jur
Institute for Community Medicine
Section Epidemiology of Health Care and Community Health
University Medicine Greifswald
Ellernholzstr 1-2
Greifswald, 17487
Germany
Phone: 49 383486 ext 22708
Email: alexandra.stein@med.uni-greifswald.de

Abstract

Background: Clinical trials are essential for medical research and medical progress. Nevertheless, trials often fail to reach their recruitment goals. Patient recruitment systems aim to support clinical trials by providing an automated search for eligible patients in the databases of health care institutions like university hospitals. To integrate patient recruitment systems into existing workflows, previous works have assessed user requirements for these tools. In this study, we tested patient recruitment systems KAS+ and recruIT as part of the MIRACUM (Medical Informatics in Research and Care in University Medicine) project.

Objective: Our goal was to investigate whether and to what extent the 2 different evaluated tools can meet the requirements resulting from the first requirements analysis, which was performed in 2018-2019. A user survey was conducted to determine whether the tools are usable in practice and helpful for the trial staff. Furthermore, we investigated whether the test phase revealed further requirements for recruitment tools that were not considered in the first place.

Methods: We performed semistructured interviews with 10 participants in 3 German university hospitals who used the patient recruitment tools KAS+ or recruIT for at least 1 month with currently recruiting trials. Thereafter, the interviews were transcribed and analyzed by Meyring method. The identified statements of the interviewees were categorized into 5 groups of requirements and sorted by their frequency.

Results: The evaluated recruIT and KAS+ tools fulfilled 7 and 11 requirements of the 12 previously identified requirements, respectively. The interviewed participants mentioned the need for different notification schedules, integration into their workflow, different patient characteristics, and pseudonymized screening lists. This resulted in a list of new requirements for the implementation or enhancement of patient recruitment systems.

Conclusions: Trial staff report a huge need of support for the identification of eligible trial participants. Moreover, the workflows in patient recruitment differ across trials. For better suitability of the recruitment systems in the workflow of different kinds of trials, we recommend the implementation of an adjustable notification schedule for screening lists, a detailed workflow analysis, broad patient filtering options, and the display of all information needed to identify the persons on the list. Despite criticisms, all participants confirmed to use the patient recruitment systems again.

(JMIR Hum Factors 2024;11:e56872) doi:10.2196/56872

KEYWORDS

patient recruitment system; clinical trial recruitment support system; clinical trials; recruit; recruitment; recruiting; participant; participants; research; digital health; usability; interview; interviews; qualitative; experience; experiences; attitude; attitudes; opinion; perception; perceptions; perspective; perspectives; database; databases; information system; information systems; search; searches; searching; retrieval

Introduction

Background

Clinical trials are the gold standard of evidence-based medicine and are indispensable for medical progress. New diagnostics, therapies, and medications usually need to be evaluated in a randomized clinical trial. Despite the importance of clinical trials, it is often difficult for trial staff to identify a sufficient number of patients who meet the specific eligibility criteria of clinical trials and who are willing to participate. Therefore, many trials fail to include enough patients, thereby leading to statistical and financial as well as ethical problems in medical research [1-3]. One reason for this is the lack of time capacity of the trial staff [2].

Electronic systems can help to identify potential trial participants in hospitals or other health care institutions by generating a screening list of all patients who fulfill the eligibility criteria [4-6]. For example, in 2015, McCowan et al [7] published a report on stakeholders from various countries in Europe for the project EHR4CR (electronic health records systems for clinical research), which aimed to enhance the utilization of electronic health records for clinical research. Their findings indicated that a significant proportion of stakeholders perceived that a platform could facilitate the implementation of clinical trials [7].

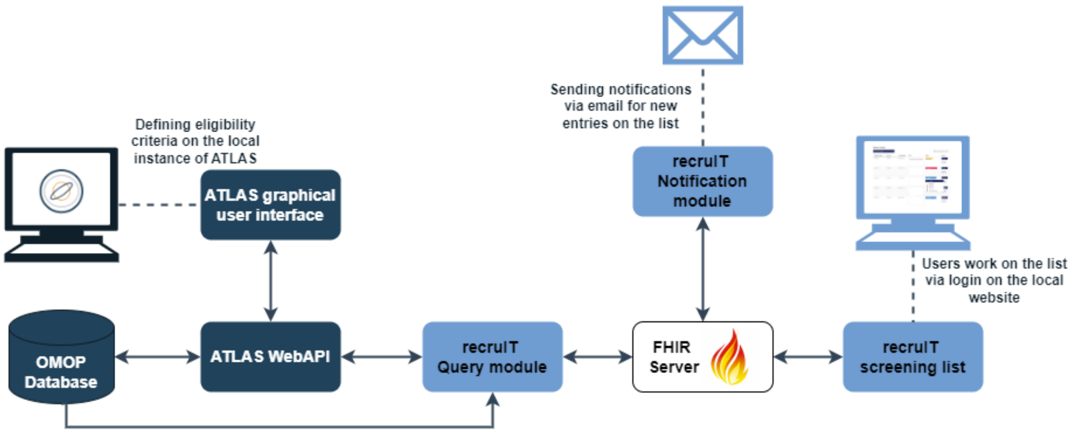
Most of the described patient recruitment systems (PRSs) were implemented for a specific site or trial. The PRS approach is time-consuming and costly and therefore not scalable for other trials [8,9]. Few systems have been built with a generic approach, independent of specific use cases to support a wide range of experiments [5].

Medical Informatics in Research and Care in University Medicine

Data integration centers were established at university hospitals as part of the MIRACUM (Medical Informatics in Research and Care in University Medicine) project, a large-scale initiative in German medical informatics focusing on research and care in university medicine. One part of MIRACUM was the so-called Use Case 1 (alerting in care), which aimed to develop and evaluate a hospital-wide PRS in a multicentric study across all participating sites. The implemented systems, namely, recruIT and KAS+, were evaluated in this feedback analysis. Both systems are briefly presented in the following paragraphs.

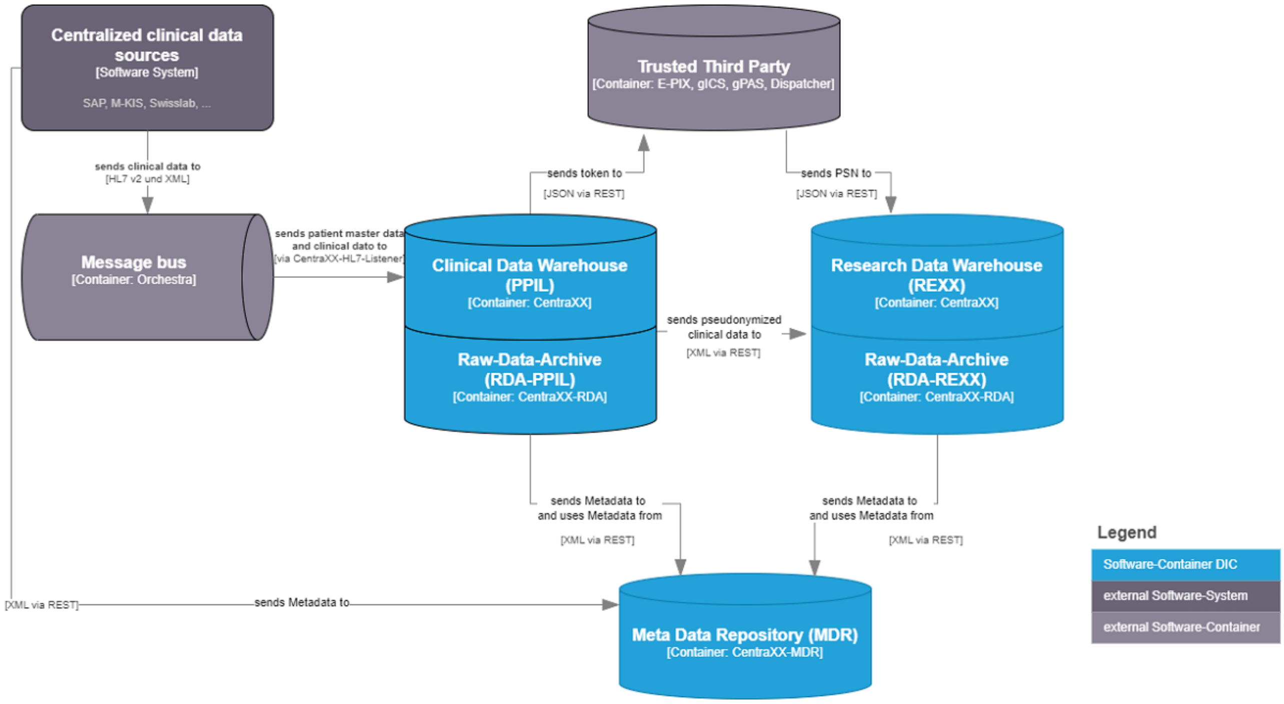
As part of the MIRACUM project, a recruitment system has been in place at several sites to support a wide range of trials [10]. Based on previously identified system requirements [11], the software recruIT (MIRACUM project) was developed. The system is shown in Figure 1 and described in detail in [11]. This system relies on the Observational Medical Outcomes Partnership (OMOP) common data model, which is a software tool of the Observational Health Data Sciences and Informatics (OHDSI) [12]. The eligibility criteria of the trials can be formulated using the ATLAS (OHDSI) software. RecruIT generates a list of potentially eligible patients, which can be accessed through an internal website that shows all the basic information such as patient number, age, and gender of all entries [5].

Figure 1. Architecture of the recruIT system. recruIT components are displayed in light blue, and Observational Health Data Sciences and Informatics (OHDSI) components are displayed in dark blue. The eligibility criteria are portrayed with the ATLAS graphical user interface. The query module triggers the search for new patients and writes all the results in the central Fast Healthcare Interoperability Resources (FHIR) store. The graphical user interface of recruIT (screening list) displays the results as a website. Users are informed about new results via email. API: application programming interface; FHIR: Fast Healthcare Interoperability Resources; OMOP: Observational Medical Outcomes Partnership.



Within the KAS+ infrastructure (Figure 2), all clinical systems transmit the patient data via HL7v2 and XML to the communication server orchestra. This distributes the data between the clinical systems and immediately transfers the data to the research platform. This consists of 2 CentraXX instances and 2 CentraXX raw-data-archives. The clinical data are read into Privacy Protection and Interface Layer, and if informed consent has been given, the data are pseudonymized using the trusted third party tools and transferred to REXX. Within the research platform, the trials are administered and the inclusion and exclusion criteria are defined. If it is configured for a defined study, CentraXX immediately checks any new patient's data to determine whether a patient may be eligible for a trial and sends the proposal to the hospital information system (HIS).

Figure 2. Architecture of the KAS+ system. Data integration center components are shown in light blue, and external components are shown in gray. Eligibility criteria are managed in the CentraXX instance called Privacy Protection and Interface Layer. The CentraXX query module initiates the search for new patients, writes all the results to its internal database, and sends proposals to the hospital information system. PPIL: Privacy Protection and Interface Layer.



Requirements of PRS

In order for the system to be useful to the trial staff and clinicians, it needs to be fully integrated into their workflow [13]. Research has been conducted on the topic of implementing and evaluating PRS and on data elements needed for that purpose [11,14,15]. For example, Schreiweis and Bergh [14] performed unstructured interviews and identified PRS requirements of different health care actors. Although Schreiweis and Bergh [14] described the fundamental prerequisites, the specific desires of researchers for a PRS with integration in diverse workflows remain largely unidentified. Aside from the capacity to search for eligible individuals with the assistance of software, there is a paucity of information regarding the specific requirements researchers have for a PRS [14]. In a previous work [11], a number of people involved in patient recruitment were interviewed to assess how the recruitment process currently works, which data sources are useful, and which features they need from a PRS in general. With this information, a list of requirements was developed that a PRS should fulfill in order to meet the requirements of the trial staff.

Objective

The goal of our work is to investigate whether and to what extent the tools of the MIRACUM project can fulfill the requirements resulting from the initial requirements analysis. Feedback should come from the real-world environment of patient recruitment. Therefore, a test phase is needed in which the trial staff will use the tool in their day-to-day work. A user survey will be conducted to determine whether the tools are usable in practice and helpful for the trial staff. The survey can also show whether additional requirements might arise from the test phase that were not considered in the initial requirements analysis. To avoid misunderstandings, we will refer to studies using the PRS as "trials" and to the investigation described here as our "study."

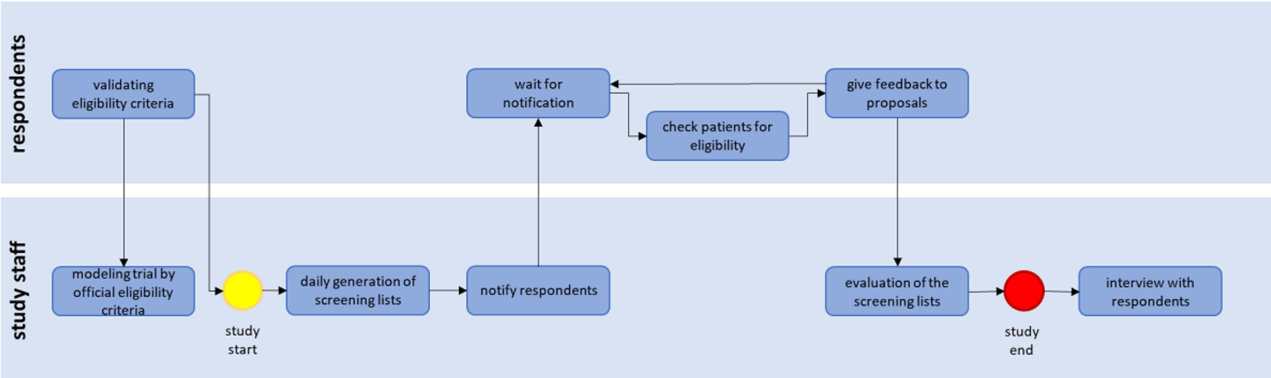
Methods

General Procedure

We conducted semistructured interviews with users of recruIT and KAS+ and derived requirements and feedback on the systems from these. Users (trial staff) had access to PRS instruments for at least 1 month. After this testing period, interviews were conducted according to the interview guide.

The test phase was part of an evaluation study to review the effectiveness of the software tools at 7 university hospitals. More detailed information regarding this study can be found in [16]. Figure 3 shows the tasks the study staff and the respondents had during the study.

Figure 3. Processes of the study involving users and study staff.



All respondents supervised at least 1 trial during the testing phase. For this study, 3 university hospitals from the evaluation study were included with all the respondents who gave interviews for analysis, either recorded and transcribed or stenographed. The other 4 university hospitals could not provide recorded or transcribed interviews, which is why they could not be included in this evaluation.

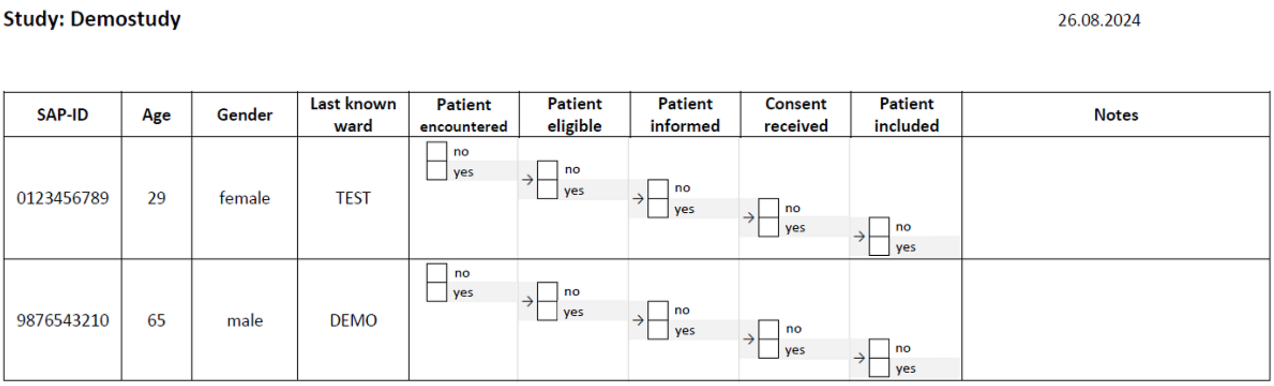
A few eligibility criteria were given in the study itself, such as the exclusion of trials with focus on psychological diagnosis. Other criteria were established by the sites to consider the local particularities: the exclusion of trials regarding children or cancer diagnosis, as the size of that site did not make such a tool necessary because the staff know the suitable patients. Another criterion for the trial selection was the expected recruitment of at least 4 patients over the course of 1 year to generate analyzable data.

The interview partners were selected at the respective study locations by the primary investigators. This approach was designed to leverage the domain knowledge and the local networks of the investigators to recruit test individuals for the study in an optimal manner. Potential individuals were invited to participate in this study, and if they consented, a time was arranged for a face-to-face interview.

PRS

From KAS+, only part of the PRS was used during the study to meet the requirements of the ethics committee and generate the feedback necessary for the evaluation. We used the search engine in CentraXX, and the parameters used were defined together with the trial personnel. From the search results generated by CentraXX, we created the so-called screening lists with an SQL query. These results were copied to a template with feedback options. An example list is shown in Figure 4. Each morning, participants received an email with the screening list if any potentially eligible patients were identified or a notice that no suggestions had been generated. These lists were then used by the trial staff according to their usual recruitment workflows. The tabular format provides information on age, gender, and the last ward stored in the system for each patient ID. The adjacent checkboxes are used to record the recruitment status. They are also required to record the feedback on the proposals necessary for the study.

Figure 4. A mock screening list of the CentraXX system, which is provided to trial personnel in the form of a PDF file.

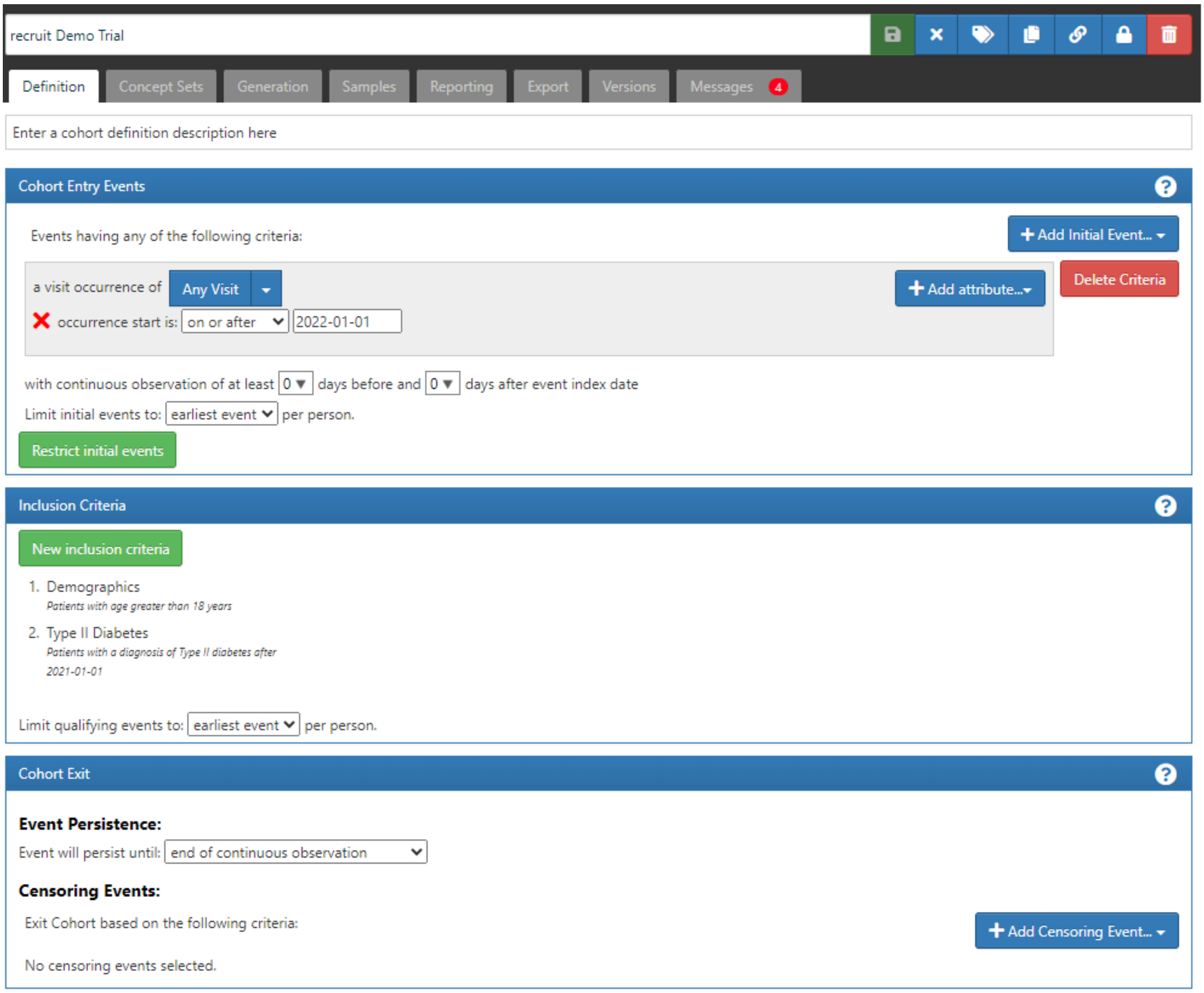


Study: Demostudy 26.08.2024

SAP-ID	Age	Gender	Last known ward	Patient encountered	Patient eligible	Patient informed	Consent received	Patient included	Notes
0123456789	29	female	TEST	no / yes	no / yes	no / yes	no / yes	no / yes	
9876543210	65	male	DEMO	no / yes	no / yes	no / yes	no / yes	no / yes	

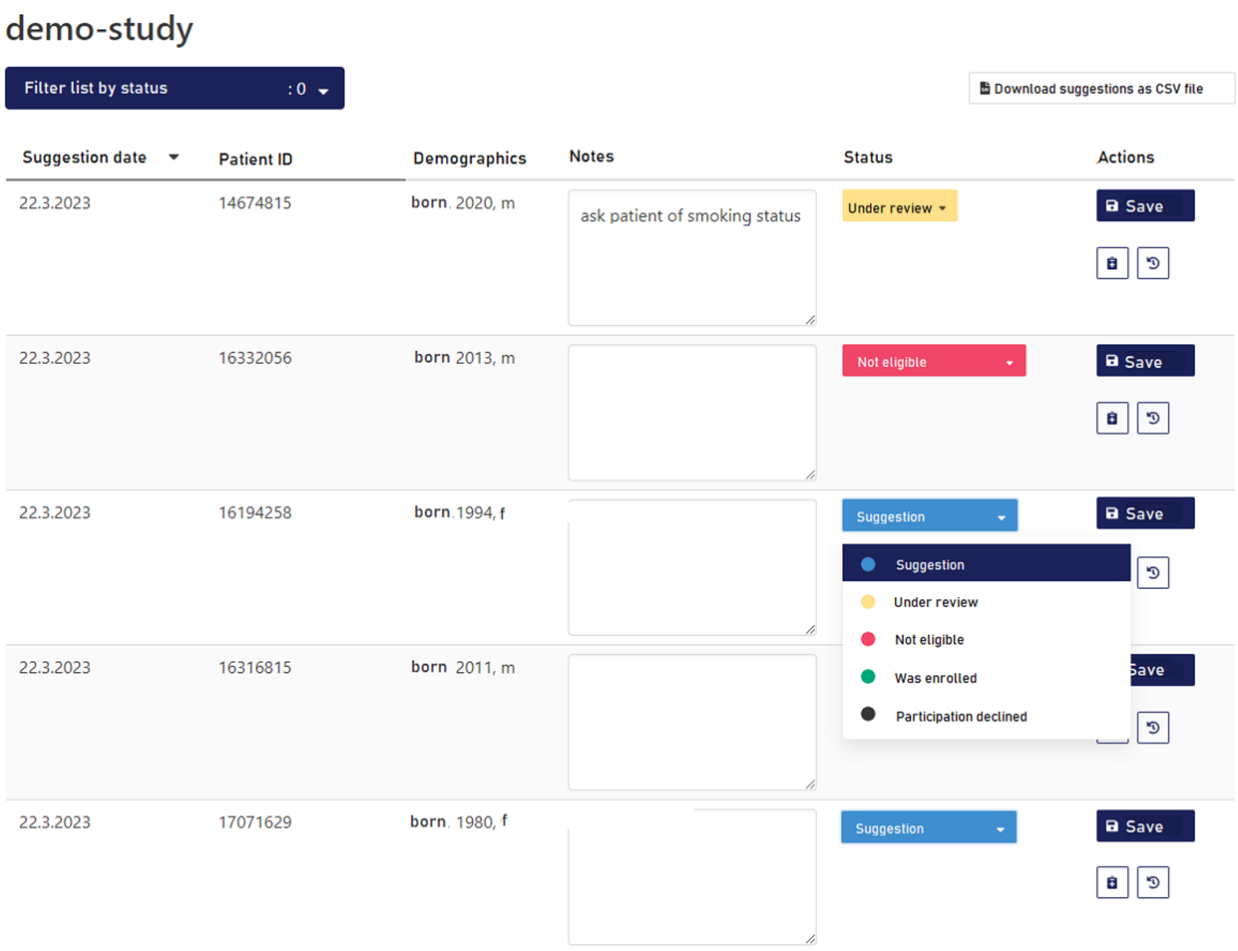
Of the 3 sites included, 2 utilized the recruIT system to generate screening lists. The initial step in utilizing the system was to translate the eligibility criteria of all the participating trials to ATLAS cohorts. In the OMOP common data model, all information is represented by a medical terminology system. Consequently, we also identified the corresponding codes and units of the aforementioned terminology systems for each eligibility criterion, prior to their portrayal in ATLAS. This procedure is also described in [17]. Figure 5 illustrates a cohort definition in ATLAS for 1 trial.

Figure 5. Sample trial as represented in the ATLAS software. All eligibility criteria are defined under inclusion criteria. In this example, the trial is looking for people who have been hospitalized since 2022, have type 2 diabetes, and are older than 50 years.



Both sites used individual configurations in accordance with local ethics committee recommendations and data protection regulations. This leads to different information shown on the web-based screening list, which is shown in Figure 6. In both sites, a patient identification number was displayed as well as the date of the first suggestion of the patient and the recruitment status. The latter can be updated by the trial staff, and a text box is provided for each entry to store additional free text regarding the proposal. Additionally, the list shows when a patient is not eligible anymore, for example, when he/she is discharged from the hospital or in case that the patient has been enrolled in another trial. For 1 site, some more information about the patients was shown on the list. This included gender, birth year, and information about the last visit and ward. The systems were updated on a daily basis, and notifications were configured either daily, weekly, or several times a week, in accordance with the user's wishes.

Figure 6. Exemplar representation of the screening list of the recruIT system. The original screenshot was overwritten with English translation.



Fulfillment of Requirements

Identifying the requirements met by the tools is the first step. For this purpose, the results of Fitzer et al [11] were used, and each requirement was compared with the functional scope. These requirements were extracted and compiled into a table. Subsequently, it was indicated for both tools whether they completely fulfill, partially fulfill, or do not fulfill these requirements at all. For the KAS+ site, both the versions used in the study context and HIS integration were assessed.

Interviews

The authors used semistructured interviews. Most of the questions were open-ended. These questions asked interviewees to describe the process of identifying eligible patients with and without PRS. Additional questions were included, that is, if there were problems with usability and whether the system could be integrated into their workflow, with room for additional statements about their experiences. Moreover, we added 2 questions that required only a yes or no answer: whether the system could be integrated into their workflow and whether they would use it again. In addition, we asked for some demographic data, which include age and experience with patient recruiting. The full list of questions and their order is shown in Multimedia Appendix 1. Although some of the questions required a "yes" or "no" answer, participants were given the opportunity to provide more detailed responses in full text, and if they did, we included those responses as well in our analysis.

For organizational reasons, 1 site asked additional questions as described in the last part of the table in Multimedia Appendix 1. Once all the interviews were transcribed, they were independently coded by 2 authors (RB and AS) according to Meyring method [18,19]. In this approach, the text to be analyzed was first examined for its key statements, and these were then summarized. These statements were then generalized into codes, and codes with the same meaning were summarized. The generalization was performed on the basis of a previously defined category system, which was then checked again against the source material [18,19]. After categorizing the codes, we structured and sorted them by using categories. Furthermore, any statements that contained a requirement for recruitment tools were marked. Afterward, the responses from the interviews were compared with the requirements already identified in [11] and checked to see if they were the same or if new ones had been mentioned.

Ethics Approval

This study received institutional review board approval from the ethics committees of Friedrich-Alexander University Erlangen-Nuremberg (approval 89_20B) as well as of Justus Liebig University Giessen (approval AZ 193/20) and Greifswald University Medicine (approval BB 084/20). Within this study, no identifying personal data were centrally collected and analyzed. No compensation was offered.

Results

Study Participants

This study consists of a total of 11 participants, comprising 7 clinical trial investigators, 2 research assistants, and 2 physicians. In 1 instance, the interview was conducted with 2 individuals simultaneously. The ages of the interviewees ranged from 25 to 34 years (2 participants), 35 to 44 years (4 participants), and 45 to 54 years (4 participants). The average number of years of professional experience in patient recruitment was 10.4 years (range 1-21 years). Four participants worked in the field of neurology, 2 in cardiology, and 1 person each in the fields of dermatology, internal medicine, neurosurgery, and rheumatology. The participants used the screening list for 1-3 trials each.

Degree of Compliance With Requirements

In [11], the following 6 categories of requirements are described: notifications, overview of patients, overview of trials, search, patient data, and user management and interface. We omitted the category “overview of trials” in this study since it is implemented as part of another tool at all participating sites. The category “patient data” contains data elements that can be used for searching; all other categories are shown in Table 1. Both systems fulfill the main requirement of (1) generating a list of eligible patients and (2) notifying users. In addition, it is possible to tag participants, make notes, and track the recruitment status. Both investigated systems lacked integration with existing HISs. Comparison with clinical trial eligibility criteria is possible with diagnoses, demographics, laboratory results, and vital signs in both tools. The treatment data mentioned in [11] can only be partially queried by the tools.

Table 1. List of requirements defined by Fitzer et al [11] and implementation in the recruIT and KAS+ systems. Since a KAS+ test environment with different properties was used for this study, this is also indicated.

Requirements	recruIT (n=12)	KAS+ evaluation environment (n=12)	KAS+ (n=12)
Notifications			
Users are instantly notified if new suggestions are available	Yes[a]	Yes	Yes
Notifications are adjustable to individual preferences by the user	Yes	No[b]	No
Overview of patients			
Supports a list of all patient suggestions	Yes	Yes	Yes
Possibility to check suggestions by themselves	Yes	No	Yes
The list with suggestions is integrated into existing systems	No	No	Yes
Option to mark participants	Yes	Yes	Yes
Option to make notes	Yes	Yes	Yes
Option to track the recruitment status	Yes	No	Yes
Edit recruitment list by manually adding patients to list	No	No	Yes
Edit recruitment list by removing patients from list	Yes	No	
Integrating patient summaries into the patient recruitment system	Yes	No	Yes
Search			
Offers sophisticated search options	Partially[c]	Yes	Yes
User management and interface			
Contain a sophisticated rights concept to account for the various roles in the trial and at the clinical center	Partially	No	Yes
Requirements fulfilled	8	5	11

[a]Yes: implemented.

[b]No: not implemented.

[c]Partially: partially implemented.

Interview Results

As the participants had no restrictions on how to integrate the tool into their workflow, the kind of integration varied. For 3 participating trials, the tool was a permanent part of the workflow; for others, it was used when the staff had spare time (n=2). Two interviewees made it the preferred source for patient recruitment. When analyzing the transcribed interviews, we found 54 different codes. To put the codes into context with each other, we defined 5 groups: system integration, parameter, precision, system evaluation, and user interface.

System Integration

Statements about the update frequency differed between the requirement to enable real-time recruitment, daily, and weekly updates. The requirement to flexibly adjust the update and notification interval per trial was also mentioned by 3 participants. All other statements required a more flexible integration into the daily workflow of the trial staff. The lists should be processed flexibly, when there is time in the clinical daily work, and the list should be integrated into the local HIS. Generally, there should be no system discontinuities.

Parameter

Three respondents indicated that not all relevant criteria were available in the system, and 9 respondents mentioned specific data elements, namely, medications, pulmonary parameters, lung transplant list, laboratory results, cardiac echocardiography findings, general findings, admission letters, and alcohol abuse. In addition, 6 respondents expressed the necessity of filtering the list by ward, while 1 respondent proposed that this should encompass the entire patient journey. One respondent cited poor data quality and inadequate utilization of the International Statistical Classification of Diseases (ICD) and Related Health Problems for all criteria related to diagnosis (n=3).

User Interface

The majority of the statements in this group pertain to patient identification. Patient numbers should be fully displayed, and pseudonymous lists are considered impractical to use. In 1 interview, the full name of each patient was also requested, while in another one, this was mentioned as not relevant. The new features cited included more options for patient recruitment status (n=1), integration of better categorization and tagging options in the list (n=1), and the ability to sort suggestions by ward (n=1). In addition, 1 person commented that it would be nice to add a third category of soft exclusion criteria that would result in a warning on the generated list. This would be especially helpful in the case of discretionary decisions, for example, if patients are excluded due to a certain diagnosis, all patients on the list who had this diagnosis in the past should be flagged on the list so that the trial staff know that they need to check whether this diagnosis is still valid.

Precision

It was mentioned that the list contains too many suggestions that are not eligible for the trial (n=5) as well as too many eligible persons that are not on the list (n=2). The list contained no or few false positives (n=3) or false negatives (n=1). In addition, 1 respondent mentioned that subsequent adjustments to the filter criteria resulted in more accurate screening lists.

System Evaluation

The results regarding the feasibility of integrating the system into the workflow of trial staff were mixed. Overall, most of the interviewees (n=7) reported satisfaction with the system and expressed their desire to use it again in the future (n=8). However, some interviewees mentioned that they would only use the system for specific trials (n=2). Furthermore, 2 individuals highlighted that using the system resulted in labor savings during the recruitment process (n=2) and positively impacted recruitment numbers (n=1). The system was capable of reaching different groups of people compared to traditional recruitment methods, which, as a result, broadens the pool of potential patients (n=2).

Evaluated Requirements

Requirements were derived from the statements of interviewees and are shown by frequency and category. We identified 4 requirements that were stated by 4 persons in the interviews. These were that whole patient numbers should be shown on the list to identify the patients properly. Further, they require a possibility to filter patients on the list after hospital wards. Two less concrete requirements were the better integration of the application in the clinical workflow and less false-positive suggestions on the list. All other requirements are shown in Multimedia Appendix 2. Multimedia Appendix 3 includes the code assignment for the requirements.

Five requirements were identified upon analyzing the interviews, which are highly similar to the ones in [11] (Multimedia Appendix 4). Three people mentioned the previously unimplemented requirements of integration into the local HIS and having flexible access to the lists. Although highlighting patients on the list is already possible, 1 interviewee proposed the ability to categorize and mark suggestions. This implies that the current implementation does not fully satisfy the users. One interviewee mentioned the need for more status options and an iterative patient search.

Discussion

Principal Findings

A comparison of the PRSs in question revealed that 8 (recruIT) and 11 (KAS+) of the 12 requirements identified in the previous analysis by Fitzer et al [11] can be fulfilled. Ten interviews were conducted with individuals involved in the recruitment of individuals for clinical trials at 3 distinct sites. Additionally, further requirements of the participants were identified. These requirements could be classified into different categories, and it was determined that integration into existing workflows is of particular importance for our interviewees. Many of the identified requirements are directly related to this.

Degree of Compliance With Requirements

Although certain requirements outlined in [11] were not implemented in the systems under evaluation, none of the interviewees mentioned any of them. Based on the results of this study, it is assumed that both manually adding or removing patients from the list and implementing a sophisticated role and rights concept do not have a high priority for the interviewees. However, it cannot be determined whether these requirements would be useful in the context of the PRS. Given that this study is confined to a limited number of trial centers, it is possible that these requirements will only become relevant when more people are involved and multiple trials are supported.

Interview Results

Certain interviewees mentioned new filtering options such as filtering for wards, despite the rarity or absence of these criteria in trial protocols. This indicates that for PRS implementation, official eligibility criteria alone might not be sufficient; additional filtering criteria that are specific to recruitment workflows may also be relevant. Further investigation may prove valuable in identifying other criteria that could enhance patient filtering.

Many of the mentioned parameters lead to diagnostic examinations which, taken together, occur often [20,21]. Diagnostic examinations can vary widely, and the resulting data that need to be queried by a PRS can vary as well. This can create challenges in collecting data from the local HIS. Access to high quality data from different clinical systems and electronic health records, which is an important part of a PRS, remains an unresolved issue and is the subject of ongoing research and development [22-24]. This finding was also reported by McCowan et al [7], who conducted stakeholder interviews for the project EHR4CR in 2015. Over half of the interviewees expressed the opinion that problems could arise from the lack of functionality in their HISs and the absence of crucial data items in the primary care systems [7]. Problems with filtering can arise from data that are documented unstructured, incorrect, or too late. As described in a data completeness analysis in 2022 [25], some data elements are found in less than 50% of electronic health records in German hospitals. Presumably for this reason, 1 participant mentioned that the data quality was not good enough.

Accuracy of suggestions is an area with several influencing factors such as the type of trial, general accessibility of the data, and data quality. One reason for false positives can be that not all of the important criteria are accessible, leading to suggestions that are technically correct, although the patient is still not eligible for the trial. The same result is achieved when there are fuzzy criteria, which need to be judged by trial staff. This is a problem also identified by Li et al [26] in 2021. They described that different scopes of research can lead to different definitions. In order to address this problem, we included the trial personnel in the definition of filter criteria. However, we learned that some criteria have to be checked manually, such as the cause of a disease or life expectancy [26]. Penberthy et al [27] also identified a high rate of false-positive suggestions in the evaluation of their PRS. They concluded that this was due to incomplete information about the patients, which prevented the exclusion criteria from being fully checked. However, some people have mentioned that few false positives are possible. Nevertheless, it is unclear whether this observation is based on concrete numbers or on the expectations of the participants.

The population examined in clinical studies is often criticized as not being representative. Older adults, women, and ethnic minorities in particular are less frequently included in clinical trials than they are represented in the general population [28-30]. Especially with the possibility to access a broader pool of patients, these tools could be used to face the underrepresentation of different groups. The ability of research staff to identify additional patients from diverse hospital wards is a phenomenon that Penberthy et al [27] also observed in their PRS evaluation. Additionally, including persons from other wards can happen more often to reach patients who are primarily treated for a different disease or health issue than that addressed in the trial.

Half of the participants were able to integrate the PRS in their daily routine, while others stated that this would not be fully possible. On closer examination of all statements of these persons, we could identify potential reasons for the missing integration and could see that 2 of these persons also criticized that pseudonymized lists are not practical. One of them stated that the lists should be generated earlier in the morning, and another demanded that the full patient names be included in the list. It is possible that integration could be easier if these issues are worked on. Despite the lack of comprehensive investigation into the PRS requirements, the integration of a PRS into existing systems, such as the official HIS, is a topic that is frequently discussed in various academic publications. In addition to the findings of Fitzer et al [11], Dugas et al [31] were able to derive this conclusion from a case study, while Schreiweis and Bergh [14] reached the same conclusion through stakeholder interviews.

Features Already Implemented in KAS+

As mentioned above, the KAS+ system was not used with its full capabilities due to the study requirements. For each proposal, feedback was necessary, especially if it was marked as false positive—this was not possible with HIS integration. Therefore, this integration was not used for this study, which also disabled the connected features.

This is why some mentioned features are already implemented but have not been used, like the integration into HIS. Trial staff can access a screening list, which is constantly updated. Various filters can be applied to this screening list and electronic health records can be accessed directly from this list, provided that the user has sufficient rights. All suggestions are shown with a consent status that indicates, for example, whether they have signed an informed consent for this study, rejected, or withdrawn it.

Implementation of Requirements

We could identify 32 requirements from the analyzed interviews. It is not possible to say which of the requirements are specific to a trial center or medical discipline and which are valid for a broader field of users. We assume that requirements mentioned by more than one person are at least not specific to one process. From all requirements, 12 were expressed in at least 2 interviews and are listed in Multimedia Appendix 4.

Adjustable Update Interval

Three of the participants addressed the notification or update interval of the tool. By implementing adjustable intervals, all these requirements could be met. At least the features mentioned above should be available: weekly, daily, and real-time updates. It would be even better, especially with changing shift schedules, if the intervals could be chosen completely freely, that is, users could also specify certain weeks, days, or times when they want to receive notifications.

Integration in Workflow

Users want to be able to adapt the system to suit their needs, which correlates with the demand for a flexible PRS to be integrated into the daily workflow. This is particularly related to the demand for integration into the local HIS, which would also reduce system discontinuities in the solutions. The lack of integration sometimes leads to time-consuming workarounds, mainly caused by typing information from one system into the other. Reducing this work would therefore mean that the use of the screening list would take less time.

The tools should be embedded in the daily work of the trial centers. Trial staff work in a variety of workflows involving different groups of people, departments, and information sources [11,32]. In order to implement a working integration into existing processes, it is necessary to know them in detail. To the best of our knowledge, there is as yet no publicly described preliminary work on which to build [32]. For this reason, it makes sense to perform a complete workflow analysis before developing and implementing a PRS in order to avoid system discontinuities and other application issues. Furthermore, it is advisable to integrate the PRS directly into existing information systems when feasible.

Broad Filtering Options

The filtering options when generating the list have to cover criteria, which are relevant for the identification of eligible persons. Several studies have been conducted to find out which data elements are necessary to check all the eligibility criteria. Therefore, the criteria were bundled into data element groups. The studies that examined this issue list a broad range of data elements and their frequency, which can be used as a guide for the first implementation of a PRS [20,21]. Additionally, we could show that the filtering for wards and the multiple presence of parameters is necessary in the eyes of our participants. The PRS can only consider those data elements that are present in the clinical systems. However, there are data elements that are not routinely collected or are not of sufficient quality. As a primary requirement for the implementation of a PRS, it is therefore necessary that the system has access to a data pool that is as complete and up-to-date as possible. Nevertheless, eligibility criteria can be highly specialized. Thus, a more flexible approach where data elements can be extended continuously would be a way to face these issues.

Pseudonymization

The screening list should always show enough information to find the persons easily in HIS in order to keep the effort in locating the patients as low as possible. We consider this as the reason for full patient numbers or patient names to be shown on the list. Patient data should be pseudonymized for authorized users before being displayed on the screening list. A similar observation was made by Butte et al [33].

Limitations

The main limitation of our study was the small number of participating trials: 10 trials at 3 different locations. Participants used a different PRS as already described above, and all locations had dissimilar local conditions, which might have had an impact on the results. The investigated trials varied in type, design, and the duration for which the trial staff used the PRS. Also, the way the interviews were documented varied slightly; while 2 of the included sites transcribed the interviews, 1 filled the form stenographically. Therefore, there is no additional information for the bounded questions for 1 trial. As we did only a qualitative analysis and used only our participants' opinions regarding the false-positive and false-negative rates, we have no evidence that they always correlate with the quantitative numbers. Since the KAS+ test environment worked with daily generated PDF lists, many patients had already left the clinic when the trial staff checked the lists. Moreover, the KAS+ system would remove no longer suitable patients from the list, while this was not the case within the PDFs. This may have led to a higher false-positive rate.

Conclusion

The trial staff had a high workload with the recruitment of patients. Especially in retrospective recruitment, where often hundreds of files of a ward have to be searched manually, the time required can be enormous and files of other wards are not even included. If a filter system such as recruIT or KAS+ succeeds in generating a list in which this number can be reduced, time can be saved, even if there are false-positive entries in this list. Although the evaluated PRS does not actually yet meet all requirements, all participants would use the system again, at least for certain trials, which shows the need of any kind of support.

Participants stated that, even with more accurate suggestions, a manual control is crucial, as there will always be discretionary criteria or other aspects that need a human judgment, which cannot be done by a PRS. The recruitment efficacy of the system can vary across different trials. Nonetheless, it remains to be seen. In any case, participants do not want a support system for each and every trial, in particular, if there are already well-established processes in place or if the identification of a test individual depends heavily on the doctor's subjective assessment.

Our results are in line with test runs of comparable recruitment tools but also show that study personnel must be closely involved in the development to meet their needs like the filtering option for current wards or scheduled notifications. The next steps should be the exploration of the most needed parameters to increase the quality of the suggestions, the integration into HIS, and the implementation of an adjustable update and notification interval, as these are the most important aspects shown in this evaluation.

The future enhancement of the tools should be done in cooperation with the study personnel to create a tool that can easily be integrated into the workflow. To ensure this, future evaluations with a larger group of participants and a wider array of trials are necessary for a comprehensive analysis.

Acknowledgments

This study was created as part of the MIRACUM (Medical Informatics in Research and Care in University Medicine) project, which is funded by the German Ministry of Education and Research (funding: FKZ 01ZZ1801A/D/M).

Conflicts of Interest

None declared.

Multimedia Appendix 1

The full list of questions used as an interview guideline.

[DOCX File , 17 KB - humanfactors_v11i1e56872_app1.docx]

Multimedia Appendix 2

List of all the requirements derived from the qualitative analysis of interviews, ordered by the frequency of nomination.

[PNG File , 53 KB - humanfactors_v11i1e56872_app2.png]

Multimedia Appendix 3

List of all the requirements derived from qualitative analysis of interviews, ordered by category and shown with the code assignment for the requirements.

[DOCX File , 18 KB - humanfactors_v11i1e56872_app3.docx]

Multimedia Appendix 4

List of requirements defined by Fitzer et al [11] and implementation in the recruIT and KAS+ systems. Since a KAS+ test environment with different properties was used for this study, this is also indicated. In addition, the list shows how often the requirements were mentioned by the participants.

[DOCX File , 24 KB - humanfactors_v11i1e56872_app4.docx]

References

1. Carlisle B, Kimmelman J, Ramsay T, MacKinnon N. Unsuccessful trial accrual and human subjects protections: an empirical analysis of recently closed trials. Clin Trials 2015 Feb;12(1):77-83 [FREE Full text] [doi: 10.1177/1740774514558307] [Medline: 25475878]
2. Desai M. Recruitment and retention of participants in clinical studies: Critical issues and challenges. Perspect Clin Res 2020;11(2):51. [doi: 10.4103/picr.picr_6_20]
3. Halpern SD, Karlawish JHT, Berlin JA. The continuing unethical conduct of underpowered clinical trials. JAMA 2002 Jul 17;288(3):358-362. [doi: 10.1001/jama.288.3.358] [Medline: 12117401]
4. Embi PJ, Jain A, Clark J, Bizjack S, Hornung R, Harris CM. Effect of a clinical trial alert system on physician participation in trial recruitment. Arch Intern Med 2005 Oct 24;165(19):2272-2277 [FREE Full text] [doi: 10.1001/archinte.165.19.2272] [Medline: 16246994]
5. Reinecke I, Gulden C, Kümmel M, Nassirian A, Blasini R, Sedlmayr M. Design for a modular clinical trial recruitment support system based on FHIR and OMOP. Stud Health Technol Inform 2020 Jun 16;270:158-162. [doi: 10.3233/SHTI200142] [Medline: 32570366]
6. Thadani SR, Weng C, Bigger JT, Ennever JF, Wajngurt D. Electronic screening improves efficiency in clinical trial recruitment. J Am Med Inform Assoc 2009;16(6):869-873 [FREE Full text] [doi: 10.1197/jamia.M3119] [Medline: 19717797]
7. McCowan C, Thomson E, Szmigielski CA, Kalra D, Sullivan FM, Prokosch H, et al. Using electronic health records to support clinical trials: a report on stakeholder engagement for EHR4CR. Biomed Res Int 2015;2015:707891-707898 [FREE Full text] [doi: 10.1155/2015/707891] [Medline: 26539523]
8. Ni Y, Bermudez M, Kennebeck S, Liddy-Hicks S, Dexheimer J. A real-time automated patient screening system for clinical trials eligibility in an emergency department: design and evaluation. JMIR Med Inform 2019 Jul 24;7(3):e14185 [FREE Full text] [doi: 10.2196/14185] [Medline: 31342909]
9. Campillo-Gimenez B, Buscail C, Zekri O, Laguerre B, Le Prisé E, De Crevoisier R, et al. Improving the pre-screening of eligible patients in order to increase enrollment in cancer clinical trials. Trials 2015 Jan 16;16:15 [FREE Full text] [doi: 10.1186/s13063-014-0535-7] [Medline: 25592642]
10. Prokosch H, Acker T, Bernarding J, Binder H, Boeker M, Boerries M, et al. MIRACUM: Medical Informatics in Research and Care in University Medicine. Methods Inf Med 2018 Jul 17;57(S 01):e82-e91. [doi: 10.3414/me17-02-0025]
11. Fitzer K, Haeuslschmid R, Blasini R, Altun FB, Hampf C, Freiesleben S, et al. Patient recruitment system for clinical trials: mixed methods study about requirements at ten university hospitals. JMIR Med Inform 2022 Apr 20;10(4):e28696 [FREE Full text] [doi: 10.2196/28696] [Medline: 35442203]
12. Reich C, Ostropolets A. Standardized vocabularies. The Book of OHDSI. 2019. URL: https://ohdsi.github.io/TheBookOfOhdsi/ [accessed 2022-11-07]
13. Cuggia M, Besana P, Glasspool D. Comparing semi-automatic systems for recruitment of patients to clinical trials. Int J Med Inform 2011 Jun;80(6):371-388. [doi: 10.1016/j.ijmedinf.2011.02.003] [Medline: 21459664]

14. Schreiweis B, Bergh B. Requirements for a patient recruitment system. Stud Health Technol Inform 2015;210:521-525. [Medline: 25991202]
15. Lai YS, Afseth JD. A review of the impact of utilising electronic medical records for clinical research recruitment. Clin Trials 2019 Apr;16(2):194-203. [doi: 10.1177/1740774519829709] [Medline: 30764659]
16. Boeker M, Zöller D, Blasini R, Macho P, Helfer S, Behrens M, et al. Effectiveness of IT-supported patient recruitment: study protocol for an interrupted time series study at ten German university hospitals. Trials 2024 Feb 16;25(1):125 [FREE Full text] [doi: 10.1186/s13063-024-07918-z] [Medline: 38365848]
17. Blasini R, Buchowicz KM, Schneider H, Samans B, Sohrabi K. Implementation of inclusion and exclusion criteria in clinical studies in OHDSI ATLAS software. Sci Rep 2023 Dec 18;13(1):22457 [FREE Full text] [doi: 10.1038/s41598-023-49560-w] [Medline: 38105303]
18. Mayring P, Fenzl T. Qualitative inhaltsanalyse. In: Handbuch Methoden der Empirischen Sozialforschung. Wiesbaden: Springer VS; 2019:648.
19. Kuckartz U. Qualitative inhaltsanalyse: methoden, praxis, computerunterstützung. In: Grundlagentexte Methoden. Weinheim Basel: Beltz Juventa; 2018.
20. Gulden C, Landerer I, Nassirian A, Altun FB, Andrae J. Extraction and prevalence of structured data elements in free-text clinical trial eligibility criteria. Stud Health Technol Inform 2019;258:226-230. [Medline: 30942751]
21. Ateya MB, Delaney BC, Speedie SM. The value of structured data elements from electronic health records for identifying subjects for primary care clinical trials. BMC Med Inform Decis Mak 2016 Jan 11;16:1 [FREE Full text] [doi: 10.1186/s12911-016-0239-x] [Medline: 26754574]
22. Eichler H, Bloechl-Daum B, Broich K, Kyrle PA, Oderkirk J, Rasi G, et al. Data rich, information poor: can we use electronic health records to create a learning healthcare system for pharmaceuticals? Clin Pharmacol Ther 2019 Apr;105(4):912-922 [FREE Full text] [doi: 10.1002/cpt.1226] [Medline: 30178490]
23. Dhayne H, Haque R, Kilany R, Taher Y. In search of big medical data integration solutions - a comprehensive survey. IEEE Access 2019;7:91265-91290. [doi: 10.1109/access.2019.2927491]
24. Aguduri LS, Merzweiler A, Yüksekogul N, Meyer N, Brandner A, Heinze O. Modeling clinical data transformation for a medical data integration center: An openEHR approach. 2019 Sep Presented at: Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e. V. (GMDS); September 8; Dortmund. [doi: 10.3205/19gmds161]
25. Vass A, Reinecke I, Boeker M, Prokosch HU, Gulden C. Availability of structured data elements in electronic health records for supporting patient recruitment in clinical trials. Stud Health Technol Inform 2022 Jun 06;290:130-134. [doi: 10.3233/SHTI220046] [Medline: 35672985]
26. Li M, Cai H, Nan S, Li J, Lu X, Duan H. A patient-screening tool for clinical research based on electronic health records using OpenEHR: development study. JMIR Med Inform 2021 Oct 21;9(10):e33192 [FREE Full text] [doi: 10.2196/33192] [Medline: 34673526]
27. Penberthy L, Brown R, Puma F, Dahman B. Automated matching software for clinical trials eligibility: measuring efficiency and flexibility. Contemp Clin Trials 2010 May;31(3):207-217 [FREE Full text] [doi: 10.1016/j.cct.2010.03.005] [Medline: 20230913]
28. Flores LE, Frontera WR, Andrasik MP, Del Rio C, Mondríguez-González A, Price SA, et al. Assessment of the inclusion of racial/ethnic minority, female, and older individuals in vaccine clinical trials. JAMA Netw Open 2021 Feb 01;4(2):e2037640 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.37640] [Medline: 33606033]
29. Alessy SA, Davies EA, Rawlinson J, Baker M, Lüchtenborg M. How representative are colorectal, lung, breast and prostate cancer patients responding to the National Cancer Patient Experience Survey (CPES) of the cancer registry population in England? A population-based case control study. BMJ Open 2019 Dec 16;9(12):e034344 [FREE Full text] [doi: 10.1136/bmjopen-2019-034344] [Medline: 31848175]
30. Etti M, Fofie H, Razai M, Crawshaw AF, Hargreaves S, Goldsmith LP. Ethnic minority and migrant underrepresentation in COVID-19 research: Causes and solutions. EClinicalMedicine 2021 Jun;36:100903 [FREE Full text] [doi: 10.1016/j.eclinm.2021.100903] [Medline: 34095791]
31. Dugas M, Lange M, Berdel WE, Müller-Tidow C. Workflow to improve patient recruitment for clinical trials within hospital information systems - a case-study. Trials 2008 Jan 11;9:2 [FREE Full text] [doi: 10.1186/1745-6215-9-2] [Medline: 18186949]
32. Trinczek B, Schulte B, Breil B, Dugas M. Patient recruitment workflow with and without a patient recruitment system. Stud Health Technol Inform 2013;192:1124. [Medline: 23920898]
33. Butte AJ, Weinstein DA, Kohane IS. Enrolling patients into clinical trials faster using RealTime Recuiting. Proc AMIA Symp 2000:111-115 [FREE Full text] [Medline: 11079855]

Abbreviations

EHR4CR: electronic health records systems for clinical research
HIS: hospital information system
ICD: International Statistical Classification of Diseases

MIRACUM: Medical Informatics in Research and Care in University Medicine
OHDSI: Observational Health Data Sciences and Informatics
OMOP: Observational Medical Outcomes Partnership
PRS: patient recruitment system

Edited by A Choudhury; submitted 30.01.24; peer-reviewed by C Gundler, A Hassan, S Jimenez-Jorge; comments to author 12.04.24; revised version received 03.07.24; accepted 11.07.24; published 27.09.24.

Please cite as:
Stein A, Blasini R, Strantz C, Fitzer K, Gulden C, Leddig T, Hoffmann W
User Requirements for an Electronic Patient Recruitment System: Semistructured Interview Analysis After First Implementation in 3 German University Hospitals
JMIR Hum Factors 2024;11:e56872
URL: https://humanfactors.jmir.org/2024/1/e56872
doi:10.2196/56872
PMID:

Original Paper

Exploring Patient, Proxy, and Clinician Perspectives on the Value and Impact of an Inpatient Portal: A Reflexive Thematic Analysis

Simone Schmidt[1,2], BA, MA, PhD; Adam Boulton[1], BSN; Benita Butler[1], BA, BSN; Timothy Fazio[1,3], MBBS, MIS

[1]EMR Team, Royal Melbourne Hospital, Melbourne, Australia

[2]School of Computing and Information Systems, University of Melbourne, Melbourne, Australia

[3]Department of Medicine, Melbourne Medical School, University of Melbourne, Melbourne, Australia

Corresponding Author:
Simone Schmidt, BA, MA, PhD
School of Computing and Information Systems
University of Melbourne
Melbourne Connect
700 Swanston St
Melbourne, 3053
Australia
Phone: 61 481247167
Email: simone.schmidt@unimelb.edu.au

Abstract

Background: Research exploring perspectives on inpatient portals reports that patients desire the information affordances of inpatient portals, and clinicians recognize their value for improving patient experience but also express caution regarding sharing aspects of the medical record. This study contributed to the existing literature on inpatient portals by considering the psychosocial dimension of clinician resistance to information sharing with inpatients and the power dynamic associated with clinician-patient information asymmetry. Along with the information affordances commonly discussed in this area, this study explored perspectives on the novel option to audio record consultations via an inpatient portal.

Objective: This study aims to understand patient, proxy, and clinician perspectives on the value and impact of an inpatient portal within the Australian context. It explores clinician resistance and receptivity to sharing aspects of the medical record with patients and the power dynamic that characterizes the relationship between clinician and patient. It considers how an inpatient portal might assist in the transformation of this relationship such that this relationship could be characterized by greater information symmetry.

Methods: Interviews were conducted with patients (n=20), proxies (n=4), and clinicians (n=21) recruited from 3 areas within the Royal Melbourne Hospital, where the portal would later be implemented. A largely inductive reflexive thematic analysis was conducted.

Results: Patient and proxy participants reported that they wanted to understand what is happening in their care for peace of mind and that an inpatient portal could support this understanding. Clinician participants reflected on how they might transform their information-sharing practice to provide greater transparency in their relationship with patients. Participants considered the types of information that could be shared and how this information could be shared via an inpatient portal. Four key themes were generated: (1) affording the patient and proxy awareness, control, and reassurance through sharing accessible and meaningful information; (2) protecting the clinician and safeguarding quality health care in information sharing; (3) flexibly deploying the functions depending upon clinician, patient, proxy, and context; and (4) moving toward person-centered care: empowerment and equity via an inpatient portal.

Conclusions: An inpatient portal provides an opportunity to reconceptualize the medical record and how this information might be shared with patients while they are admitted to the hospital, such that they have more understanding as to what is happening in their care, which ultimately supports their well-being. The transition to a more transparent information-sharing culture in the Australian hospital context will take time. An inpatient portal is a critical step in facilitating this transition and creating more informational symmetry in the clinician-patient relationship.

(JMIR Hum Factors 2024;11:e52703) doi:10.2196/52703

KEYWORDS

inpatient portal; patient perspective; clinician perspective; information sharing; clinician-patient relationship; person-centered care; reflexive thematic analysis; qualitative research; mobile phone

Introduction

Background

Over the last decade, initiatives have been implemented across various countries to provide patients access to the medical record. Of these initiatives, Open Notes [1] is perhaps the most notable. Initiated in the United States in 2010, Open Notes is a call to action to provide patients access to their notes so that they can have more knowledge of, become more involved in, and have improved experience and outcomes in their health care. This initiative began as a pilot project to evaluate patient access to primary care notes [2]. As of 2021, it is now a US federal rule to provide patients access to their medical records [3]. Another key initiative is Planetree [4]. Founded in 1978, this international organization is dedicated to improving patient care from the patient perspective and has published numerous articles on the benefit of sharing the record with inpatients (refer to the study by Frampton et al [5]). Sweden's Journalen is also another noteworthy initiative. Implemented in 2012, this system offers anyone from the age of 16 years access to their outpatient records from hospitals, primary care, mental health care, and dental care [6]. Alongside these initiatives, studies from the last decade have focused on and have been motivated by the incentive to share medical records with inpatients via portals for the reasons outlined earlier, that is, to enable the patient, as the patient-centered or person-centered discourse puts it, to become more "empowered" in their care [7-16]. These studies all underlined the inpatient portal's beneficial impact on patient experience.

Studies (noted earlier) on the value and impact of sharing medical records with inpatients from both the patient and clinician perspectives are concentrated in the United States. These studies are generated from patient and clinician participant perspectives on hypothetical patient access to components of the medical record or from the actual clinician and patient experience of the sharing of and access to components of the record. These studies reported that clinicians are generally accepting of providing inpatients access to their records due to the benefits they provide patients. Also in these studies, clinician participants convey innovative ways the record could be shared with inpatients. However, clinician participants generally express caution that patient access to their records during their hospital stay may result in patient misunderstanding and anxiety as well as potentially compromise the quality of care. Patients generally report wanting access to their records and appreciate the potential transparency and inclusion in their care that this information sharing could provide. Significantly, studies involving the patient experience of the inpatient record report a decrease or no increase in patient anxiety [9,10,13,14] and therefore suggest a misunderstanding from the clinician perspective of the patient experience and a possible reduction of this experience to a limiting stereotype [17].

Objectives of This Study

The motivation for our study was to explore patient, proxy (a nominated representative of the patient and in this study a partner involved in their care), and clinician perspectives of the MyChart Bedside inpatient portal within the Australian context before its future implementation. MyChart Bedside is an app by Epic Systems Corporation that extends the electronic medical record, providing patients access to aspects of their record (including notes, test results, and vital signs) while admitted to the hospital, via their smartphone or a hospital-provided tablet (Figures 1 and 2 show screenshots of the inpatient portal). To our knowledge, there are no studies of this kind exploring patient and clinician perspectives of patient access to their inpatient record in Australia.

Figure 1. Screenshot of inpatient portal with menu bar, notes, and recording features as displayed on a tablet (MyChart is a registered trademark of Epic Systems Corporation).

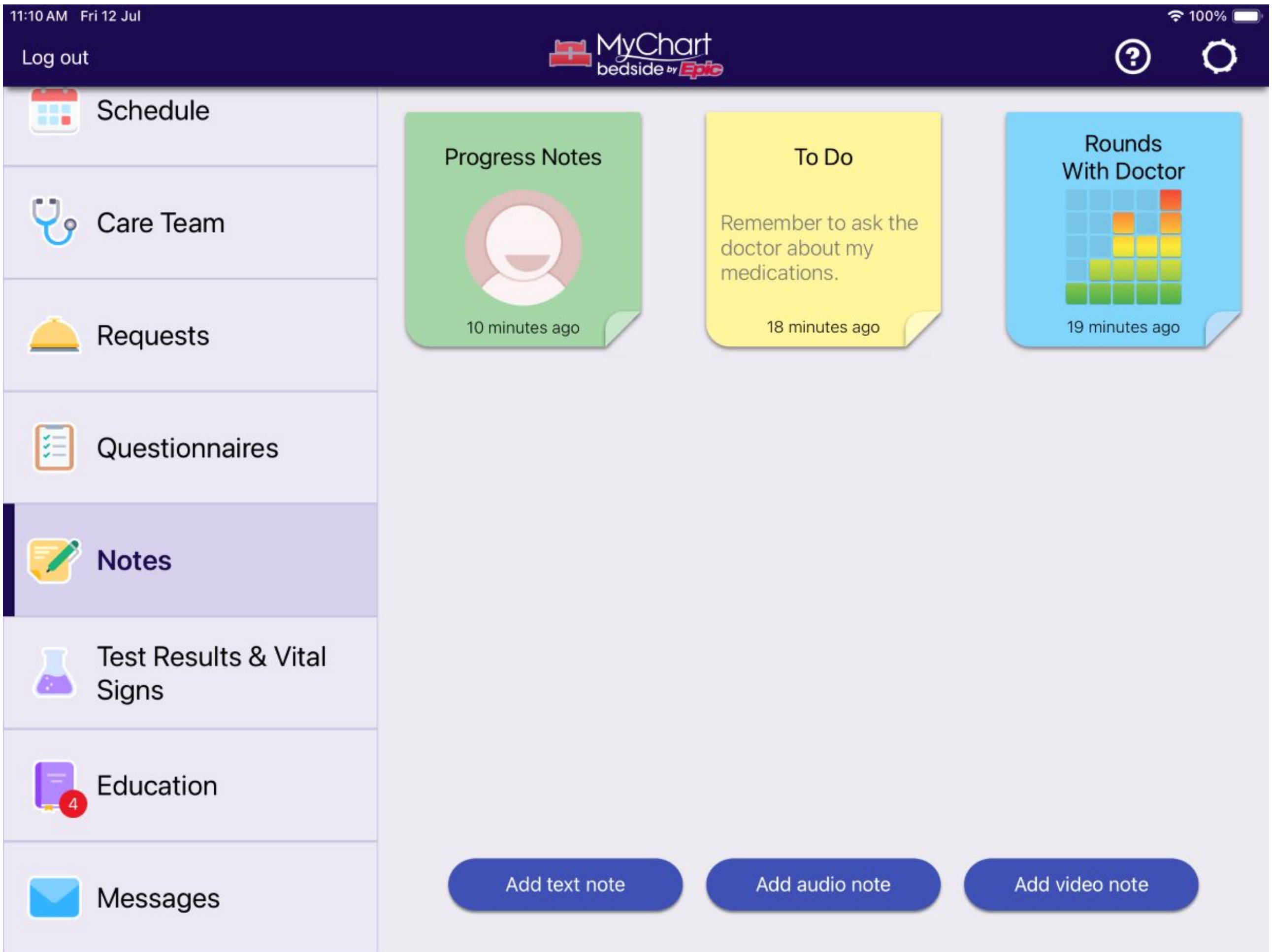
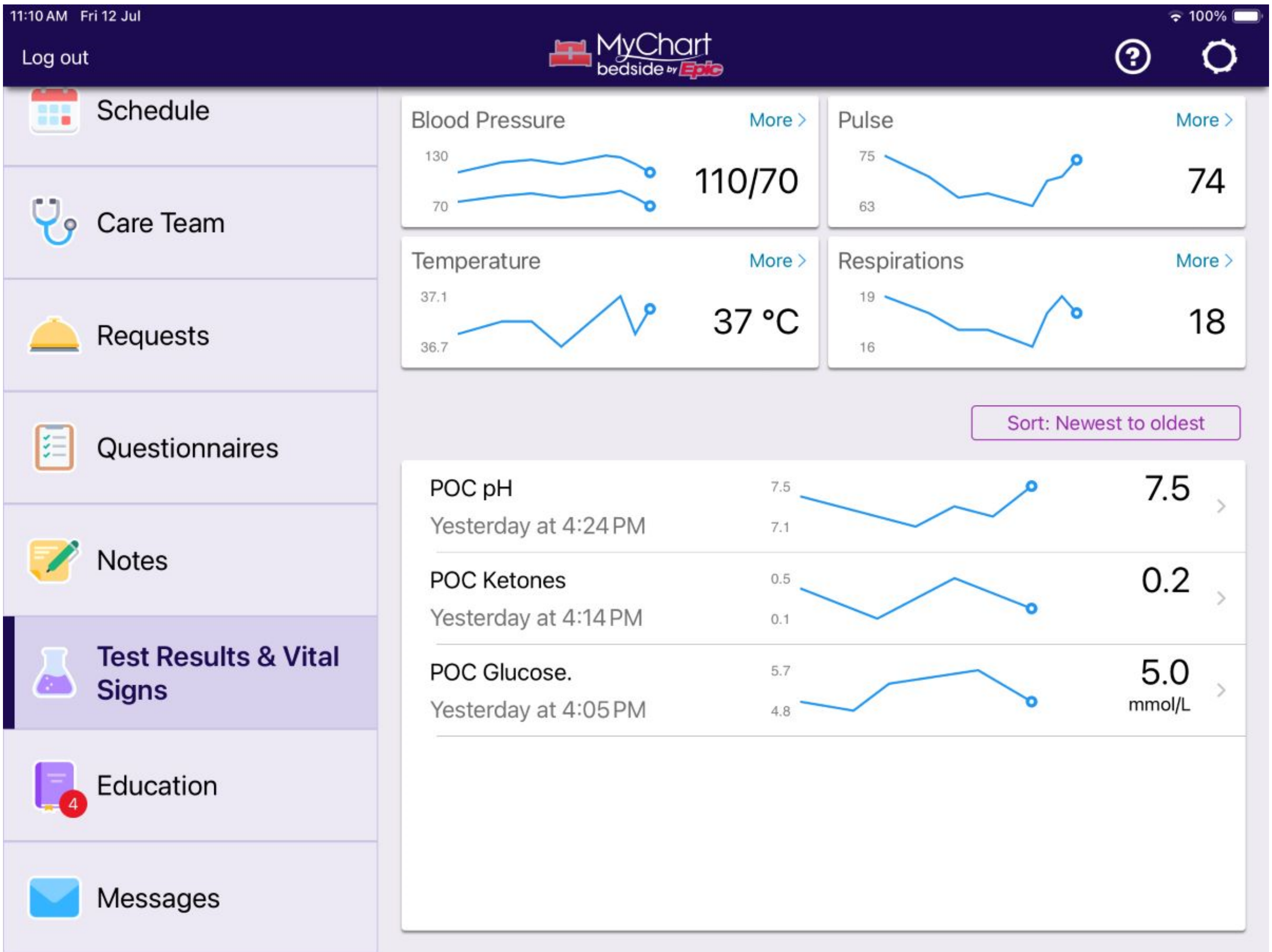


Figure 2. Screenshot of inpatient portal with menu bar and test result feature as displayed on a tablet (MyChart is a registered trademark of Epic Systems Corporation).



In this qualitative study, we used the term *person-centered*, as person-centered discourse acknowledges the patient as more than solely a patient with a set of symptoms to be diagnosed and acted upon, and it acknowledges the carer as not only a carer but as a person with unique preferences and needs [18,19]. Initiatives that provide patients access to their records and studies focused on sharing information with inpatients (cited earlier) are aligned with the ethos of person-centered care. This ethos calls for health care to move away from the traditional paternalistic model of medicine characterized by an information asymmetry, where the clinician withholds information from the patient and proxy and positions them as a passive recipient in health care, to move toward a more balanced relationship characterized by greater transparency in information sharing between clinicians and patients and proxies, supporting greater patient and proxy collaboration with the clinician in their health care.

A significant contribution of this study to the studies in this area is that it approaches not only the patient and proxy but also the clinician through a person-centered lens. In this approach, the clinician is understood not solely in terms of their clinical function but also as a person who, just like the patient and the proxy, although for different reasons, may be vulnerable in health care. This study pointed to how recognition of clinician vulnerability may provide insights into their resistance to information sharing, which is the common result in studies exploring perspectives on patient access to their record. This study also shares with the person-based approach [20] the acknowledgment of the significance of the psychosocial dimension of its participants when exploring their perspectives. The person-based approach builds on user-centered and human-centered design and the principles of usability and acceptability and deploys a psychosocial lens to design and evaluate digital health interventions for behavior change and to enhance well-being. This study that has a different but related focus—an inpatient portal is also intended to enhance patient well-being—explored, where relevant, the psychosocial dimension of its participants when considering their perspectives and acknowledging their preferences about the acceptance of and vision of the sociotechnical system that the portal would generate.

This study was conducted within the context of the Royal Melbourne Hospital (RMH), a metropolitan, quaternary, and adult teaching hospital. Since August 2020, RMH has been offering its outpatients and their nominated proxies access to portions of their records, including select notes, test results, and medication lists via an outpatient portal (accessible via their smartphone or PC) referred to as "Health Hub." The next logical step envisioned by RMH is to offer inpatients access to information pertaining to their hospital stay via an inpatient portal, MyChart Bedside. MyChart Bedside provides information affordances, including those that are the focus of this study—access to notes, test results, and the ability to message the care team and audio record consultations during the patient's hospital stay. The primary incentive of our study was to determine the levels of receptivity and resistance among the clinician, patient, and proxy population to these affordances before the portal's future implementation throughout the hospital. The primary research question that guided our study was "What is the foreseen value and impact of an inpatient portal from the patient, proxy, and clinician perspective?" Patients, proxies, and clinicians were interviewed to understand their perspectives on the portal's functions, and this paper presents a reflexive thematic analysis of their perspectives. The exploration of patient, proxy, and clinician perspectives on the prospect of patients recording their consultations provides a novel contribution to the existing literature on patient portals that does not discuss this function.

This study builds on the knowledge generated in previous studies with particular attention to the psychosocial and cultural dimension of information sharing within the context of a hospital and the power dynamics involved between the patient and proxy and clinician. More generally, it points to why, despite the patient's and proxy's desire for more information on their health care and the clinician's understanding of the value of sharing information with the patient and proxy, the transition to a more transparent, symmetrical relationship through information sharing between clinicians and patients and proxies is complex. This study describes this complexity while suggesting ways it can be addressed such that health care contexts, such as that of Australia, can move toward more person-centered care.

Methods

Recruitment

Clinician, patient, and proxy participants were recruited for interviews from areas within the hospital where the portal would later be implemented. These areas included ward 7B (hereafter referred to as the leukemia ward), hospital in the home (HITH) subacute, and HITH acute. These areas care for patients with blood cancers, notably leukemia, and extend inpatient care to a patient's home for patients recovering from such conditions as having a stroke (HITH subacute) or breast cancer surgery (HITH acute). In total, 5 patients and 1 proxy were recruited from each of the HITH areas; 10 patients and 2 proxies were recruited from the leukemia ward. Recruitment of patients and proxies was facilitated by the nursing unit manager from the leukemia ward, a nurse educator from HITH acute, and a clinical coordinator from HITH subacute. Potential participants were selected via convenience sampling, that is, based on their mental and physical ability to participate in an interview and provide informed consent, their fluency in English (interviews were conducted in English as the hospital operates within an English-speaking environment), and their willingness to be interviewed. Author SS phoned the selected potential participants, and those interested in being interviewed nominated a date and time for the interview. More details on patient and proxy demographic characteristics (including context recruited from, age range, the highest level of education, employment status, comfort with technology, and understanding of their health condition) can be found in the Results section.

Nursing unit managers from the 3 areas and the chief medical information officer nominated the clinicians to be recruited. Their purposeful selection included a range of clinician types (including physicians, nurses, and allied health staff) and levels of seniority. A total of 9 (43%) of the 21 clinicians were recruited from the leukemia ward, 7 (33%) clinicians were recruited from HITH subacute, and 5 (24%) clinicians were recruited from HITH acute (refer to the Results section for clinician demographic characteristics, including context recruited from, clinician type, age range, level of education, and comfort with technology).

Generation of the Dataset

Interviews

Participants were interviewed for 30 to 60 minutes via a videoconference platform. A semistructured interview guide was used. The interview explored the portal's following functions: (1) access to the notes, (2) access to the test results, (3) ability to message clinicians, and (4) ability to record consultations.

Participants were invited to consider whether and how these information affordances were useful and how they would impact care. Interviews were recorded and transcribed.

Analysis

Reflexive thematic analysis was used in the analysis of the dataset. In this qualitative method, researchers acknowledge their epistemological position, that they coproduce knowledge with participants, and that knowledge is not "found" but generated [21]. This study took a contextualist epistemological position, which located knowledge as generated by people's experiences and perspectives from a particular context at a particular point in time. Author SS took a largely inductive approach in the analysis of the dataset and conducted semantic coding of the transcripts in NVivo (Lumivero), describing the dataset and staying close to the participants' voices. The codes and the subsequent themes were outputs generated from the dataset rather than preconceived through a theoretical framework. However, in the generation of a dataset, its analysis, and reporting, there is always a level of interpretation involved. The disciplinary background of author SS in the humanities and particular interest in ethics of sociotechnical systems informed the interpretation of the dataset. The disciplinary backgrounds of authors AB (nursing and informatics), BB (nursing), and TF (medicine and informatics) informed their review of the analysis and its finalization.

Ethical Considerations

This study was approved as a quality assurance project by the RMH Office for Research in August 2022 (QA2022087) and required only verbal consent from participants. Participants received an information sheet detailing that their information would be deidentified in published material as well as in data storage. Patient and proxy participants were remunerated with gift vouchers.

Results

Participant Demographic Characteristics

The participant demographic characteristics are presented in Table 1 and Textbox 1. Patient and proxy participants were aged between 21 and 80 years, and clinician participants were aged between 21 and 60 years. Most of our patient and proxy participants had sound digital and health literacy: all patient and proxy participants reported a good understanding or a very good understanding of their health condition, and most reported being comfortable or very comfortable with technology. The education levels of the patient and proxy participants varied. Textbox 1 presents the clinician participant types recruited, grouped by each specific health care context. As can be seen, nurses were our most represented clinician participant type, followed by physicians.

Table 1. Demographic characteristics of participants.

Demographic characteristics	Patients and proxies (n=24), n (%)	Clinicians (n=21), n (%)
Age range (y)		
21-30	2 (8)	6 (29)
31-40	6 (25)	7 (33)
41-50	3 (12)	6 (29)
51-60	6 (25)	2 (10)
61-70	6 (25)	—[a]
71-80	1 (4)	—
Highest level of education		
Secondary	9 (38)	—
Technical and further education	3 (13)	—
Bachelor	6 (25)	12 (57)
Postgraduate	6 (25)	9 (43)
Employment status		
Employed	16 (67)	—
Unemployed	4 (17)	—
Retired	4 (17)	—
Hospital context		
Leukemia ward	12 (50)	9 (43)
HITH[b] acute	6 (25)	5 (24)
HITH subacute	6 (25)	7 (33)
Comfort with technology		
Not comfortable	2 (8)	—
Comfortable	12 (50)	10 (48)
Very comfortable	10 (42)	11 (52)
Understanding of health		
Good	16 (67)	—
Very good	8 (33)	—

[a]Not available.

[b]HITH: hospital in the home.

Textbox 1. Types of clinicians recruited from each of the areas of future implementation.

Leukemia ward clinicians

- Nursing unit manager
- Graduate nurse
- Assistant nursing unit manager
- Social worker
- Dietician
- Pharmacist
- Consultant
- 2 physicians

Hospital in the home subacute clinicians

- Clinical coordinator
- Home visit nurse
- Occupational therapist
- 2 speech therapists
- Physiotherapist
- Consultant

Hospital in the home acute clinicians

- Nurse educator
- 2 home visit nurses
- Pharmacist
- Consultant

Thematic Overview

In total, 4 themes were generated from the dataset as conveyed in Figure 3: focused on the patient's voice, theme 1 formed a major part of the analysis. It was concerned with how the inpatient portal's functions can serve the patient and proxy, affording them awareness, control, and reassurance in their health care through accessible and meaningful information. After understanding what information the patient and proxy wants and the reason for it, we then need to understand how patient and proxy access to this information might impact the clinician's practice. Theme 2 was focused on clinicians' concerns about how the portal's functions might impact them and their provision of care. Theme 3 considered how these functions can be flexibly deployed to serve the patient and proxy, protect the clinician, and safeguard quality care. Theme 4 presented an overall reflection on the value and impact of an inpatient portal from the patient, proxy, and clinician perspective, exploring the portal's perceived strengths and limitations in relation to key concepts in person-centered care: *empowerment* and *equity*.

Figure 3. Thematic overview.

Theme 1: Affording the Patient and Proxy Awareness, Control, and Reassurance Through Sharing Accessible and Meaningful Information

Overview

The most overwhelming factor generated in the patient and proxy interviews was that patients and proxies experience anxiety while admitted to the hospital due to the uncertainty about their situation. Patient and proxy participants explained that their anxiety is not so much a result of their health condition but their lack of awareness of what is happening in their health care. This theme was concerned with how functions that share meaningful information with the patient and proxy could bring them greater awareness of their situation and, by extension, a greater sense of control and peace of mind in their health care. Throughout this paper, *meaningful information* for the patient and proxy is understood as selective information regarding their health care that is tailored, relevant, and makes sense to them. Patients and clinicians conveyed the meaningfulness of information shared by the inpatient portal would depend upon the patient's understanding of this information and, if necessary, whether it can be made meaningful by the clinician in conversation with the patient. Participants also considered the meaningfulness of this information could depend on added explicatory summaries; patient-friendly language; the selection of relevant information; and whether the patient wants this information generally speaking or at specific moments during their health care. Significant to this theme is the idea that the portal affords patients the ability to access information when they are ready because, as patient participants expressed, they experience periods of being unable to take information in due to the emotional and physical impact of their condition. This theme has been generated primarily from the patient and proxy perspective of functions that provide the option to share test results and notes and to record consultations. The clinician's perspective on these functions contributes to this theme in a secondary capacity in terms of its degrees of alignment to or divergence from the patient and proxy perspective. The themes mentioned subsequently are labeled with the function they address and a patient or proxy quote or paraphrased perspective conveying a common patient and proxy perspective on the benefit of the function.

Theme 1.1 Test Results: It's Good to Know Where You Sit Whether It's Good or Bad

Patient participants from the leukemia ward considered it would be valuable if the portal could afford them more timely access to their results, rather than them having to wait for their clinician to share this information:

> *As a leukaemia patient, the one thing you're always wanting to know is...your blood levels...those counts mean that...we're stuck in hospital or we can...become day patients, which is what we all aim for.... First thing in the morning...pathology has already got your results, but you don't know what they are.... If it's on the app [portal...] you can...get the blood counts from...the night before.* [Participant 38]

Patient participants more generally explained that access to their results helps them understand their health, and, as many expressed, this understanding reassures them. One patient expressed, "It's good to know where you sit; whether it's good or bad." Another stated that having access to these results would help them prepare for their consultations and feel more in control. Another explained that having access to their results could function as a kind of diary whereby they could track their health progress as they receive treatment. Accessibility was a key reason why patient participants were interested in having their results on the portal—having access to them whenever they needed them and being able to share this information with their care network (general physicians, family, and carers) was important to them. Some patient participants expressed the need for an explicatory summary or additional information to help them contextualize and understand their results:

> *It would be great to have...something that would explain...these results...It's valuable in the sense that you can see things out of range, but you can't put the context around it....You're like, "I'm not quite sure what they're looking for with that."* [Participant 43]

Some clinician participants echoed the aforementioned perspective, stating that it would benefit patients to receive their results with an explanatory statement. A clinical coordinator stated the following:

> *I am on the health portal and love seeing my results, but I also know how to interpret them....If [they] came with explanations and general advice, I think that would be good.... Like what happens if my potassium is low? What do I do about it? My blood sugar is high. What do I do about it?* [Participant 4]

A more common clinician perspective was that results should not be shared with patients before being discussed with their clinicians. A pharmacist explained the following:

> *If you don't have the skills to interpret it you shouldn't be given information without explanation.*

However, there were clinician participants who had no reservations regarding sharing results with patients and made no mention of the importance of the clinician first discussing these results with their patient or providing an explanatory statement. Echoing a patient perspective cited earlier, an assistant nursing unit manager from the leukemia ward stated the following:

> *A lot of them actually wake up at about three o'clock...in the morning because they know that that's when results come back and they ask.... It would be good for them to just log on themselves and actually have a look, "Oh, I'm going to get some platelets today. Oh, my white cells are—I don't think I can go home today," because they know. They're very well-informed. They know exactly what's happening.* [Participant 12]

Theme 1.2 Notes: Knowing What Track You're on Gives You Peace of Mind

Patient and proxy participants were interested in access to their progress notes for a greater understanding of and to keep track of their health care and to have something to refer to, particularly at times when it is difficult to take in or retain information from consultations:

> *You know that this is the track you're on...you can...see if something's changed.... If there was something major coming up and then I'd forgotten about it or missed something, I could go back into the notes and go, "Okay, no, we've decided that this is what we're going to do".... It'll give you...peace of mind and definitely something to refer back to, because it can be so overwhelming. Sometimes you don't retain all that information.* [Participant 22]

> *We would be able to look at the notes and if there was something that we just needed to reconfirm...or if...we didn't quite understand what the doctor said.... I need to know exactly how she's going, not only for her wellbeing but for mine...so I can do what I have to do as a carer looking after her.* [Participant 11]

A proxy participant, who is also a nurse and so understands what the progress notes involve, explained why access to their partner's notes would have been helpful:

> *We couldn't really visit him very much in hospital because visiting hours are restricted, so it would really help us to have access to check that everything was okay.... I guess it would just be extra reassurance.* [Participant 1]

However, some patient participants were aware that they would not be able to understand notes in their current form. A patient stated the following:

> *I was exposed to the hospital notes.... The language used is mainly for the medical department, so mainly it's a jargon...a bunch of numbers and blood counts that...I couldn't translate to tangible information for myself, so—yeah, it hasn't been useful, but I'd like some more transparency there.* [Participant 22]

Some patient participants suggested a patient-friendly, summary record alternative to the clinician-facing notes, as they assumed the latter would be difficult to understand and therefore not useful.

Clinician participants were comfortable sharing admission and discharge notes with patients and proxies, but most were reluctant to share progress notes. Those open to sharing progress notes echoed the patient suggestion of a patient-facing note that could function as a summary record, filtering out aspects of the notes that could offend, confuse, or distress the patient. Some clinician participants considered how certain sections of the notes, such as those from the ward rounds and assessment notes, could be of value to the patient. A physician stated the following:

> *If I say, "...the next couple of days, the plan is to do a CT...to give you some transfusions—" this and that...I don't think much of that would sink in.... I almost...want to give them a...a list to...say, "Okay, this is what's happening today. This is what's happening tomorrow."...That's what's captured in a ward round note.... If they have access to the ward round note, it'll be nice for them to see what's on the agenda...to understand what's happening for their illness....* [Participant 26]

A speech therapist stated the following:

> *If I did a swallowing assessment...and had a report, I would write it in a patient-friendly manner and...share it, whereas if it was just a progress note with no home therapy program...then I would maybe just keep that for the other clinicians.* [Participant 17]

There was a minority (2/21, 10%) of clinician participants who were open to sharing notes and wrote them as if a patient could access them. A physiotherapist explained that this was how they were trained to write notes. A social worker explained that they write their notes as if a patient could read them but simultaneously expressed discomfort at the prospect of a patient reading their notes.

Theme 1.3 Recording Consultations: I'd Feel More Informed and Relaxed About My Care Because I Wouldn't Need to Remember Everything

Most patient and proxy participants were interested in recording their consultations due to similar reasons for wanting access to their notes—the view that it is difficult to take information in when you are unwell and that having this record of information would bring awareness, control, and reassurance in their health care. A patient stated the following:

> *I'd feel more informed...and more relaxed about my care, because I wouldn't need to remember everything. I've felt anxious some of the time because I figured I was being told something very important, but especially when I'd just come out of ICU...I could barely relate what had happened to me in terms of what the doctor had said.... I didn't feel confident...talking to the doctor. I needed someone else...there to take in what they were saying.* [Participant 39]

This perspective conveys how the recording function could be beneficial for patients who do not have someone with them who can help them process information communicated to them. Another common reason for patient interest in recording consultations was to be able to share this information with their care network. Patient participants expressed that sometimes when they try to relay consultation information to their family, they are not always able to do so accurately, if at all. The following quote is from the proxy, who is also a nurse:

> *Because [he...] was unable to speak...he was unable to tell me what happened in those consultations.... That was really hard. He had a whole team of people come and see him, but he was unable to...relay that information. So it would be great to have them recorded.* [Participant 1]

One patient participant stated that they would not need to record their consultations because they have "pretty good recall." A proxy participant explained that they would not want this because they trust their partner's clinicians. Formerly a police officer, this participant associated recording with surveillance.

Despite having reservations, many clinician participants could see the value in providing patients with the option to record their consultations. They considered it could function as a record for patients to return to or share with their care network if they were not proficient in English. They also considered it could be valuable as a form of education and echoing the patient participants, a record for when patients cannot take their consultations in, for when their recall is limited, and to share with families absent from consultations. A graduate nurse stated the following:

> *When the doctors come in...it can be quick.... They don't process the information a hundred percent....To have that...conversation recorded that they can go back and listen to...might be helpful because sometimes you need to hear information a few times to process it.* [Participant 3]

There was a minority (2/21, 10%) of clinician participants who, due to their ethos of transparency, had no problem with patients recording their consultations. There were others—also a minority (3/21, 14%)—who were opposed to it, due to the vulnerability they felt in relation to it. Most (16/21, 76%) clinician participants had "mixed feelings" regarding this prospect.

Theme 2: Protecting the Clinician and Safeguarding Quality Health Care in Information Sharing

Overview

It was immediately apparent in the interviews that just as patients and proxies experience anxiety due to lack of awareness and therefore control in their health care, clinicians experience vulnerability in sharing information with patients. This vulnerability could be a result of the clinician's level of seniority, expertise, training, experience, work culture, the medicolegal implications of their field, or a combination of all these factors. This theme focuses on the clinicians' concern regarding patients recording consultations and accessing progress notes. Clinician participants considered how patient access to these information affordances might result in scrutiny of their practice, which could negatively impact them as well as compromise the quality of care provided. This theme is developed solely from the clinician's perspective. The following subthemes are labeled with the function they address and a common clinician perspective.

Theme 2.1 Recording Consultations: This Would Make a Lot of Clinicians Uncomfortable

Clinician participants recognized that enabling patients to record their consultations would benefit patients but felt uneasy about this. They explained that they did not want their words taken out of context to be misinterpreted, misappropriated, and vulnerable to any medicolegal implications. A speech therapist considered that if consultations were recorded, they (the therapist) might overwhelm the patient with information to safeguard themselves rather than giving the patient the information that is most meaningful for them at that moment in their health care. They stated the following:

> *We should be very aware of what we're saying, and...comfortable enough with what we're saying that we're happy...to be recorded....Some of the things that I probably said when I was working three or four years ago are no longer actually accurate because of new evidence so I think there would need to be some safeguards....What can happen with those recordings, where can they go? Is it just accessible through the app or is it going to be uploaded by the patient somewhere else?* [Participant 31]

A physician considered the following:

> *It'll be helpful to play back those recordings and try and understand...things a bit better...and ask questions. So, I can see that that's pretty valuable. It does make me...nervous when patients say, "Can I record things?"...That is something that would make a lot of clinicians uncomfortable.... Perhaps, that is more of a reflection of us, and what we're worried about with our practice than what is best for the patient.... A lot of concern...would be the medico-legal implications of recorded information, and information, perhaps, being taken out of context....* [Participant 26]

Several clinician participants stated that they would support the recording of consultations if there was consent from the clinician and regulations to ensure information would not be misinterpreted and misappropriated. There was also the view that this function should only be deployed in certain situations where it was needed.

Theme 2.2 Sharing Progress Notes: I'd be Worried About the Self-Filtering That Might Happen and What Impact That Might Have on Clinician Communication and Patient Care

In their consideration of sharing progress notes, a speech therapist reflected as follows:

> *It's hard because it's like a culture in the hospital, you wouldn't expect for the notes to be shared, so it's a whole change of thinking.* [Participant 17]

This perspective that sharing notes with patients would require a cultural shift and an evolution of thinking perhaps explains why many clinician participants were resistant to the idea of sharing progress notes. Importantly, several clinician participants understood why patients may want access to the notes and stated they would want access if they were patients; however, they also felt reservations about sharing these notes with patients when considering the clinician's perspective.

Clinician participants were generally concerned about how patient and proxy access to progress notes might negatively impact clinical practice and patients and proxies. Many clinician participants explained that these notes are a mode of communication between clinicians and include differential diagnoses, queries, social issues, and observations on mood and behavior that may confuse, offend, or distress the patient or proxy. They explained that if they were to share these notes, they would need to change the way they write, and this may impede communication with their colleagues and degrade the quality of care. A nurse explained the following:

> *I'd be worried about people not potentially fully documenting things [...because they are] worried what the patient might see or how they might react to what they've read.... That's certainly an area that's going to take...more exploration as the program unfolds.* [Participant 2]

A dietician expressed the following:

> *I'd just be worried at the level of self-filtering that might in fact happen and what impact that might actually have on communication between teams and...patient care....There's perhaps a question about the purpose of why they might need access to everything when it comes to their notes....* [Participant 40]

As with the clinicians' perspective on recording consultations, there was the idea that it should be optional to share specific notes depending upon the patient, the clinician, and the situation, rather than a default sharing of all notes.

Theme 3: Flexibly Deploying the Functions Depending on Clinician, Patient, Proxy, and Context

Overview

Throughout the interviews, particularly when interviewing clinicians, it became clear that for the functions to be of value to and to have a beneficial impact on the clinician, patient and proxy, and health care context, they would need to be flexibly used. The flexible use of functions meant that they would not be used in the same way in every situation, but their use would depend upon the clinician, patient, proxy, and context. Rather than a one-size-fits-all approach, it was conveyed by clinicians that certain notes and results (rather than others) may be beneficial to share, and in some cases, recording consultations would be appropriate. The flexible use of functions would enable the clinician to adapt the information-sharing system according to their shifting levels of comfort, the shifting needs of the patient, and the particularities of the context.

This theme introduces a new function—the messaging function—and draws from both patient and clinician perspectives. The interviews conveyed that if the messaging function were to be implemented, it would need to be done flexibly, depending on the clinician, patient, and context. Furthermore, regulations would need to be put in place for its purpose and to set patient expectations regarding clinician response. Many clinician participants, particularly physicians, were opposed to this function due to the extra work it would entail in responding to messages. Patient participants also expressed concern that this function would place too much burden on physicians. However, there were clinicians who could see how messaging could be an effective means of communication with patients in their respective roles as clinical coordinators, speech therapists, and occupational therapists from the HITH subacute context. There was unanimous agreement among patient and nurse participants from the leukemia ward that messaging would be an effective means of patient-nurse communication and would support quality care. The following subtheme is labeled with the function it addresses and a key patient perspective from the leukemia ward that was echoed by nurse participants from this same health care context.

Theme 3.1 Messaging: It Would Be So Good for Nurses and for Us as Patients Because Nurses Could Go to Where They Need to Be

Nurse participants stated after a patient buzzes them, they must go into their room to determine what they need, whereas messaging would immediately convey to them what patients need, and they would be able to determine who to attend to first depending upon the urgency of the situation. They stated messaging would be particularly useful in situations where they would otherwise need to put on protective wear to go into the patient's room, which takes time. One nurse explained the following:

> *On our busy days, our buzzers are just going off the hook [which...] adds...stress to the nursing staff. Having something like that to [...know] the things that are urgent and the things that aren't...would help.* [Participant 12]

Another nurse stated the following:

> *If you could get a message on your rover [nurses' communication device] telling you exactly what they need...it would save time....* [Participant 3]

Patient participants shared the nurses' perspectives, adding the insight that messaging would be beneficial if they needed to let their nurse know that they were in a critical condition:

> *My temperature can be raging at 41 and I can be close to death, and I push the button, or I need a bottle of cordial, and I push the button. And the nurse...doesn't know whether I'm nearly dead or whether I want the bottle of cordial....I think the app would be brilliant.... It would be so good for the nurse, and so good for us as patients because then the nurses could go to where they really need to be....* [Participant 41]

The interviews made clear how messaging could improve patient-nurse communication in the leukemia ward, enabling nurses to more effectively manage their patients' needs and potentially reduce nurse and patient stress.

Theme 4: Toward Person-Centered Care: Empowerment and Equity Via an Inpatient Portal

Overview

This theme explores the perspectives of the patient, proxy, and clinician participants on the value and impact of an inpatient portal through 2 critical and related concepts in person-centered care: *empowerment* and *equity*. The adage *knowledge is power* was voiced throughout the interviews when considering the potential value of the portal's information affordances. The terms *empowerment* and *control* were often expressed—and were often expressed together—by participants when they considered the portal's value. A significant insight generated by the interviews was that if a person shifts from the position of the clinician to that of patient or proxy, they become disempowered in their access health care information. Within the discourse on person-centered care in digital health, as concerns the context of this study, the proposition is that an inpatient portal could facilitate greater equity for the patient and proxy in their relationship with their clinician and in their health care through providing patients and proxies with accessible and meaningful information. This theme explores this proposition and highlights the potential strengths and limitations of the portal concerning the concepts of empowerment and equity. The following subthemes are partially labeled with participant quotes.

Theme 4.1 The Portal Empowers Patients to Be More in Control Like an Equal Person in Their Care

Reflecting on the value of the portal, a nurse stated the following:

> *That's the thing that stood out to me the most...it would be really good for them.... It empowers them to be more in control and like an equal person in their care.*

The term *empowerment* became especially poignant when clinicians considered when they had been patients and had experienced an immediate disempowerment in their role reversal. A physician explained the following:

> *When they change roles...from a healthcare professional to a patient, even though...they've done so many years of study and training, immediately...being a patient, just kind of disempowers them in a way that's very confronting....As the treating team, you have all the information...you just feel like, "Yep, we know what's going on." But as a patient, you are relying on someone to tell you what the results are and where you're at with your treatment.... To just feel like you're in control...that's what the app [portal] might help with.* [Participant 26]

This perspective conveys the dramatic shift in access to information and, by extension, power that occurs when a clinician becomes a patient. The clinician knows what's happening (to a degree), and the patient depends on the clinician for this information. As the nurse who was the proxy participant in this study expressed, she would have wanted access to her husband's progress notes and the ability to record his consultations, so she could have a better understanding of what was happening to him. This would have reassured her. This participant nurse did not hesitate in expressing her desire for this information, despite also being a clinician and therefore familiar with the Australian hospital information-sharing culture, which restricts patient access to such information. As the aforementioned quote conveys, perhaps the reason why there was no reservation in accessing this information, as is normally the case from the clinician's perspective, was because, in the position of patient or proxy, one is disempowered. Reflecting upon the value of the inpatient portal, a patient participant stated the following:

> *You'd just feel a little bit in control. At the time, we're totally out of control.* [Participant 9]

Another explained the following:

> *I think it's empowering to know more about your condition. If you know more, you understand more, you can make better decisions.... When you know more, you're just more in control.* [Participant 27]

Theme 4.2 The Portal Highlights and Marginalizes the Less Equitable Population

Considering the value of the portal, the proxy participant, who is also a nurse, reflected on caring for her husband, who is a physician and had just had a stroke:

> *I'm very aware that for us, as awful as it's been, it's been fairly seamless because we understand what's going on. I can't imagine how terrifying it must be for people who don't understand.... If people are...informed and they feel like they can ask questions, I think it [the portal] could only be a good thing. Knowledge is power and people feel better equipped to support their loved ones if they've got all the information because it's very intimidating, this whole talk about MRIs and CTs, and clot retrieval.... We have to embrace the technology.... It's easy for English speakers and people who don't have a pre-existing disability.... It's not that easy for everyone.* [Participant 1]

This perspective raises the issue of equity of access for people who are not proficient in English or have a disability that impedes their intake of information. The fact that the portal would only be accessible in English concerned some clinician participants. Clinician, patient, and proxy participants were also concerned that the portal would not benefit older adults or those who were not technology proficient. A consultant considered the following:

> *Is it empowering? Yeah, if it gives the information in an accurate way.... But...there's still that bias towards the tech-savvy.... Older [people from] non-English speaking backgrounds [...this] population who would most benefit—one, would not be able to use it, and two, the materials we provide are not in their language of choice. So [it...] highlights and actually marginalizes the less equitable population.* [Participant 10]

This perspective shared by several clinician and patient participants points to the limitations of the portal's affordances.

Discussion

Overview

This study's patient and proxy participants wanted health care information accessible via an inpatient portal to increase their control and, by extension, decrease their anxiety during their hospital stay. Clinician participants generally wished to support inpatient access to information via the portal but also expressed resistance and uncertainty in this area. Clinicians will need more experience in information sharing via inpatient portals in health care contexts like that of Australia to understand how it might impact care. A flexible deployment of the inpatient portal's functions could support exploration in this area. Clinician and patient participants shared ideas on how to reconceptualize the record such that it is accessible and meaningful to the patient and proxy. This study highlighted how an inpatient portal can help address the information asymmetry and power imbalance that characterizes the patient-clinician relationship.

All patients and proxies interviewed were highly interested in access to notes, and test results via an inpatient portal. All but one participant was interested the ability to record their consultations. Patient participants explained that, when unwell, it is difficult to process and retain information and having the option of accessing meaningful information—information that is relevant and makes sense to them—regarding their health care when they are ready to process it would be valuable. When considering information sharing via the inpatient portal, 3 key clinician perspectives were generated. The first 2 perspectives were held by a minority (5/21, 24%) of clinician participants. They involved either (1) an openness without reservation to sharing information with patients and proxies or (2) a complete resistance to sharing information with patients and proxies. The third perspective, held by most (16/21, 76%) clinician participants, involved an openness to information sharing with patients with reservations as to what, how, between whom, and in what context information should be shared. This perspective generated theme 3 focused on the flexible deployment of the functions depending upon the clinician, patient, proxy, and health care context. The importance of flexible deployment of the portal's information affordances, particularly in relation to high-risk situations, is noted elsewhere [22]. Furthermore, as another study has stated, when sharing health information, there should not be a one-size-fits-all approach, as different people need different information at different times and in different contexts [7].

This study contributes to existing studies on sharing medical records via an inpatient portal by acknowledging that how this sociotechnical system is configured will be determined not only by the individual patient's need for awareness, control, and reassurance in their health care but also by the clinician's need to protect their practice. Some clinician participants self-reflexively considered their attitude toward sharing inpatient notes or enabling the patient to record consultations. These participants, when reflecting on their attitudes toward information sharing, considered that these attitudes could be a result of their education, their work culture, and the broader medicolegal implications of their field. The significance here is that just as patient and proxy anxiety is a critical component to consider in the case of the value of information sharing through an inpatient portal, clinician vulnerability and indeed, anxiety, in relation to this process must also be acknowledged. As 1 clinician participant conveyed sharing inpatient notes is not part of the existing culture of the hospital and requires a new way of thinking, and as another clinician participant expressed, it will need exploration.

Shifting hospital culture to accommodate new attitudes and exploring new approaches to information sharing with inpatients will take time. Indeed, in the American context, OpenNotes charts the evolution of information-sharing culture in health care over half a century [1]. In the Australian context, where the prospect of sharing the record with inpatients is still in its infancy, a flexible sociotechnical system is needed while clinicians adjust to the changing information-sharing culture and become more comfortable with becoming increasingly transparent with patients and proxies. However, clinicians will need a high level of awareness when making the decision not to share information with patients and proxies. They will need to know if this decision is motivated by a limited attitude toward the patient that disempowers them and excludes them from their health care, or whether this decision is motivated by concern for the patient's well-being or because of the complexity of the treatment that requires time before meaningful information can be conveyed to the patient [17,22]. It has been noted that policies are needed to guide clinicians in the sharing of sensitive information in this transforming information-sharing culture [12].

The evolution of health care information sharing will require clinicians to continually reflect upon their attitudes. As raised in the Introduction section, a common clinician perspective is that if inpatients have access to their notes and tests, they will become anxious. However, as noted previously in this paper, studies exploring the inpatient experience of such information have shown that patient anxiety does not increase and indeed, in some cases, decreases [9,12,14]. This common clinician misunderstanding that access to information will increase patient anxiety has been argued to be the result of reducing the patient to a stereotype, which in turn produces a *hermeneutical injustice*: where a person's right to understand their situation is impeded [17,23]. What is overwhelmingly reported in this study is that patient anxiety is the result of not knowing what is happening more so than knowing what is happening. As 1 patient participant so clearly stated, "It's good to know where you sit; whether it's good or bad."

This study has generated significant insights into how clinicians are inquiring into their current practice and reflecting upon why they may be resistant to certain forms of information sharing, despite understanding the value it would have for the patient and proxy. As a physician reflecting on clinician resistance to recording considered, "Perhaps, that is more of a reflection of us, and what we're worried about with our practice than what is best for the patient." This reflection highlights the tension between professional expectations and wanting to provide patients with greater access to information. This clinician self-reflexivity is a critical component in the movement toward greater information symmetry and in addressing the power imbalance in the patient- and proxy-clinician relationship. As insights from clinician participants conveyed, the impetus for developing a more symmetrical relationship is most clear when the clinician places themselves in the position of the patient or proxy and understands the disempowerment the patient and proxy experiences due to limited access to their health care information. Indeed, compelling insights came from clinicians who drew from their direct experience of the reversal of roles and their recognition that the patient and proxy is disempowered in their relation to the clinician.

A perspective was raised in this study that the portal would highlight and marginalize less equitable populations, including older people and people who do not speak English. However, a study conducted in the United States has noted that less equitable populations benefit from patient portals the most: Black people, people who have low education levels, and those who do not speak English at home reported access to the record as more important than White people, people with high levels of education and native English speakers [24]. However, this idea is contradicted in a more recent study in the inpatient context that reported that African American patients used the inpatient portal less than White patients [25]. For our study, many of our patient and proxy participants had only a secondary school education, which indicates that a desire for health care information is not dependent on being highly educated. It is noteworthy that a patient who has difficulty speaking English or has low digital literacy may share their portal's information affordances with their care network, such as their family. This idea of the value of a digital platform being accessible by multiple parties involved in a patient's care is emphasized elsewhere [26].

A key insight generated from this study was that patients and proxies want access to information that is *meaningful*—information that is tailored, relevant, and makes sense to them. This perspective led to participants' conceptualization of how the record could be shared. Clinician participants considered what aspects of the notes would be useful to the patient, such as the ward round notes and assessments. Patient and clinician participants considered how the notes, if they were shared, would need to be in patient-friendly language and a summary form. This idea of the summary note has been raised in another study that reported that information sharing enabled by patient portals may call for a redesign of notes [7]. This idea of reconceptualizing the record has been highlighted in another study, which states the importance of clinicians "writing notes that patients will read" [22]. A patient participant from an earlier study exploring patient experiences on access to their records explained, "I read until I found technical stuff, and then I would jump over it...The parts I read, and I understood, were interesting and beneficial" [13]. This perspective is echoed by patient perspectives generated in this study that conveyed that unless patients understand the information, the information will not be useful. Likewise, clinician participants considered the importance of sharing useful information with patients; 1 clinician participant stated that information is not necessarily a good thing and could result in information overload. As noted in a previous study, selecting information that is meaningful for the patient and that does not overwhelm them will require an exploration of how best to present this information to the patient via the portal's interface [12].

Another significant contribution of this study to the literature on inpatient portals is the exploration of clinician and patient perspectives on enabling patients to record consultations. Although these perspectives have been explored in studies focused solely on the topic of recording [27], the option to record consultations has not been explored as a function of the inpatient portal. Clinician participants considered how it would be valuable for the patient to record consultations if they were not proficient in English so they could return to the consultation or share it with their care network. Clinician and patient participants considered the recording function useful to capture the details of an intervention or education, to share a consultation with the patient's care network, and to give the patient time to process the information because when unwell it is difficult to take information in, particularly when one has, as several patient participants put it, "chemo brain." Most clinician participants expressed resistance to enabling this function, cautioning that before such a function could be implemented, there should be safeguards in place to protect the clinician from any medicolegal consequences that could result from such a practice. Research in health care recording in the Australian context details patient rights and legal responsibilities [28]. In Victoria, if the consultation is in person, the patient has a legal right to record it and does not require the consent of the health care provider; consent is advised to protect the patient-clinician relationship. However, for the patient to share the recording with others, clinician consent is legally required. To record and share telehealth consultations in Victoria, the patient is legally required to have the consent of the clinician [28]. Neither patient nor clinician participants in this study were aware of this legal framework. Perhaps if they were, the prospect of recording consultations would be normalized, rather than being a subject of desire from the patient's perspective and a subject of fear from the clinician's perspective.

This study enabled patient, proxy, and clinician participants to reflect on how they want information shared in the inpatient context and the type of person-centered sociotechnical system they want to create to support the well-being of everyone involved in the health care context but most importantly the patient. Within the Australian context, the transition to a more transparent patient-proxy-clinician information-sharing relationship is still very much in process. An inpatient portal is a valuable step in this process.

Principal Results

The participants' consideration of an inpatient portal involved a reconceptualization of the medical record and information-sharing practice within the hospital. Participants considered how the portal's potential influence on and redesign of information sharing could result in enhanced patient-clinician communication and more transparency in the provision of care, which could contribute to patient and proxy well-being. This study highlighted how the clinician-patient information asymmetry correlates with the power imbalance involved in the clinician-patient relationship. Alongside the influence of an inpatient portal that initiates a rethinking of the medical record and how it can be shared with the inpatient, clinician self-reflexivity is a valuable step in addressing this power imbalance and information asymmetry. This study also underlined that clinician's vulnerability needs to be acknowledged in the evolving information-sharing culture and that the portal's functions could be flexibly deployed with the intention to protect the clinician's practice. Finally, this study reported the value of recording consultations from the patient and proxy perspective and the messaging function from the patient and nurse perspective within particular contexts.

Limitations

Study participants were recruited from only 3 contexts within the hospital; participants from other contexts may have provided different perspectives. Only patients and proxies fluent in English were included in the sample due to the need to conduct the interviews in English. Furthermore, all patient and proxy participants had a good understanding of their health, and most were comfortable using technology. This could be because the types of patients and proxies who agree to an interview focused on health care technology are usually people who have reasonable health and technology literacy. The study sample could therefore reflect how the less equitable patient and proxy population is excluded from the health care context.

Comparison With Prior Work

This study echoed results reported in previous studies: patients desire the information affordances of an inpatient portal to support their well-being, and clinicians understand the value of these information affordances but express caution regarding certain information sharing. This study has contributed to the knowledge of patient and clinician perspectives on inpatient portals through its exploration of the power dynamic that characterizes the patient-clinician relationship as well as drawing attention to clinician vulnerability in information sharing with patients. It considered how an inpatient portal might assist in transforming the hospital's information-sharing culture to support person-centered care. It provided a novel contribution to the literature on inpatient portals by exploring perspectives on the recording function. Finally, it explored perspectives on an inpatient portal from within the Australian context, pointing to how, in this context, health care is still in the process of transitioning to person-centered care.

Conclusions and Future Work

Participants explored how information could be shared via an inpatient portal. Patient and proxy participants expressed their desire for the portal's information affordances to support their well-being. Clinician participants reflected on their resistance and receptivity to information sharing with patients and proxies. Clinician participants expressed that there needs to be further exploration of information sharing via an inpatient portal to understand its value and impact. This study is the first stage of a 2-part research project; our next step is to explore the experiences of patients using the inpatient portal in a recent implementation in the leukemia ward. It is important to note that this recent implementation of the inpatient portal in the leukemia ward has not been informed by the results of this study but was realized as an independent project, and this implementation does not include access to progress notes nor the recording or messaging functions. However, once the 2 stages of this study are published, their combined results will inform future implementations of the inpatient portal throughout the hospital and potentially lead to patient access to more comprehensive information affordances within the inpatient context.

Acknowledgments

The authors would like to thank Cameron Grant, Annie Robson Hammond, Janelle Walters, Bronwyn West, and Julie Wong for their assistance in participant recruitment. The authors thank Jessica Rixon for sending the patient and proxy participants gift vouchers and Julie Louie for sourcing inpatient portal images. The authors would like to thank Sarah Boltuk for facilitating Epic Systems Corporation's copyright permission for the portal images and Dr Simon D'Alfonso for reviewing this paper. The authors would also like to thank the study participants for their insights.

Data Availability

The datasets generated during and analyzed during this study are not publicly available to protect participant privacy.

Authors' Contributions

SS conducted the interviews and analysis and wrote the manuscript. All authors reviewed and developed the manuscript. All authors were involved in research design and recruitment.

Conflicts of Interest

None declared.

References

1. OpenNotes. URL: https://www.opennotes.org/ [accessed 2023-04-24]
2. Delbanco T, Walker J, Darer JD, Elmore JG, Feldman HJ, Leveille SG, et al. Open notes: doctors and patients signing on. Ann Intern Med 2010 Jul 20;153(2):121-125 [FREE Full text] [doi: 10.7326/0003-4819-153-2-201007200-00008] [Medline: 20643992]
3. Salmi L, Blease C, Hägglund M, Walker J, DesRoches CM. US policy requires immediate release of records to patients. BMJ 2021 Feb 18;372(426):1-2. [doi: 10.1136/bmj.n426] [Medline: 33602667]
4. Planetree International. URL: https://www.planetree.org/ [accessed 2024-06-19]
5. Frampton SB, Horowitz S, Stumpo BJ. Open medical records. Am J Nurs 2009 Aug;109(8):59-63. [doi: 10.1097/01.NAJ.0000358504.52667.71] [Medline: 19641416]
6. Armstrong S. Patient access to health records: striving for the Swedish ideal. BMJ 2017 May 02;357:j2069. [doi: 10.1136/bmj.j2069] [Medline: 28465316]
7. Fritz Z, Griffiths FE, Slowther A. Custodians of information: patient and physician views on sharing medical records in the acute care setting. Health Commun 2021 Dec 19;36(14):1879-1888 [FREE Full text] [doi: 10.1080/10410236.2020.1803553] [Medline: 32814466]
8. McAlearney AS, Sieck CJ, Gaughan A, Fareed N, Volney J, Huerta TR. Patients' perceptions of portal use across care settings: qualitative study. J Med Internet Res 2019 Jun 06;21(6):e13126 [FREE Full text] [doi: 10.2196/13126] [Medline: 31172960]
9. McAlearney AS, Fareed N, Gaughan A, MacEwan S, Volney J, Sieck C. Empowering patients during hospitalization: perspectives on inpatient portal use. Appl Clin Inform 2019 Jan 13;10(1):103-112 [FREE Full text] [doi: 10.1055/s-0039-1677722] [Medline: 30759491]
10. Grossman LV, Creber RM, Restaino S, Vawdrey DK. Sharing clinical notes with hospitalized patients via an acute care portal. AMIA Annu Symp Proc 2017;2017:800-809 [FREE Full text] [Medline: 29854146]
11. O'Leary KJ, Lohman ME, Culver E, Killarney A, Randy Smith G, Liebovitz DM. The effect of tablet computers with a mobile patient portal application on hospitalized patients' knowledge and activation. J Am Med Inform Assoc 2016 Jan;23(1):159-165 [FREE Full text] [doi: 10.1093/jamia/ocv058] [Medline: 26078412]
12. Woollen J, Prey J, Wilcox L, Sackeim A, Restaino S, Raza S, et al. Patient experiences using an inpatient personal health record. Appl Clin Inform 2016;7(2):446-460 [FREE Full text] [doi: 10.4338/ACI-2015-10-RA-0130] [Medline: 27437053]
13. Prey JE, Restaino S, Vawdrey DK. Providing hospital patients with access to their medical records. AMIA Annu Symp Proc 2014;2014:1884-1893 [FREE Full text] [Medline: 25954461]
14. Wilcox L, Morris D, Tan D, Gatewood J. Designing patient-centric information displays for hospitals. Proc SIGCHI Conf Hum Factor Comput Syst 2010 Apr 10;2010:2123-2132 [FREE Full text] [doi: 10.1145/1753326.1753650] [Medline: 27990502]

15. Smith CA, Kelly MM. In anticipation of sharing pediatric inpatient notes: focus group study with stakeholders. J Particip Med 2022 May 30;14(1):e37759 [FREE Full text] [doi: 10.2196/37759] [Medline: 35635743]
16. McAlearney AS, Gaughan A, MacEwan SR, Fareed N, Huerta TR. Improving acceptance of inpatient portals: patients' and care team members' perspectives. Telemed J E Health 2020 Mar 01;26(3):310-326 [FREE Full text] [doi: 10.1089/tmj.2019.0026] [Medline: 31081723]
17. Blease C, Salmi L, Rexhepi H, Hägglund M, DesRoches CM. Patients, clinicians and open notes: information blocking as a case of epistemic injustice. J Med Ethics 2021 May 14;48(10):785-793 [FREE Full text] [doi: 10.1136/medethics-2021-107275] [Medline: 33990427]
18. Santana MJ, Manalili K, Jolley RJ, Zelinsky S, Quan H, Lu M. How to practice person-centred care: a conceptual framework. Health Expect 2018 Apr 19;21(2):429-440 [FREE Full text] [doi: 10.1111/hex.12640] [Medline: 29151269]
19. Ekman I, Swedberg K, Taft C, Lindseth A, Norberg A, Brink E, et al. Person-centered care--ready for prime time. Eur J Cardiovasc Nurs 2011 Dec 01;10(4):248-251. [doi: 10.1016/j.ejcnurse.2011.06.008] [Medline: 21764386]
20. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
21. Braun V, Clarke V. Thematic Analysis: A Practical Guide. Thousand Oaks, CA: Sage Publications; 2022.
22. Hägglund M, McMillan B, Whittaker R, Blease C. Patient empowerment through online access to health records. BMJ 2022 Sep 29;378:e071531 [FREE Full text] [doi: 10.1136/bmj-2022-071531] [Medline: 36175012]
23. Fricker M. Epistemic Injustice: Power and the Ethics of Knowing. Oxfordshire, UK: Oxford University Press; 2007.
24. Walker J, Leveille S, Bell S, Chimowitz H, Dong Z, Elmore JG, et al. OpenNotes after 7 years: patient experiences with ongoing access to their clinicians' outpatient visit notes. J Med Internet Res 2019 May 06;21(5):e13876 [FREE Full text] [doi: 10.2196/13876] [Medline: 31066717]
25. Walker DM, Hefner JL, Fareed N, Huerta TR, McAlearney AS. Exploring the digital divide: age and race disparities in use of an inpatient portal. Telemed J E Health 2020 May 01;26(5):603-613 [FREE Full text] [doi: 10.1089/tmj.2019.0065] [Medline: 31313977]
26. Morley J, Floridi L. Enabling digital health companionship is better than empowerment. Lancet Digit Health 2019 Aug;1(4):e155-e156 [FREE Full text] [doi: 10.1016/S2589-7500(19)30079-2] [Medline: 33323183]
27. Moloczij N, Krishnasamy M, Butow P, Hack TF, Stafford L, Jefford M, et al. Barriers and facilitators to the implementation of audio-recordings and question prompt lists in cancer care consultations: a qualitative study. Patient Educ Couns 2017 Jun;100(6):1083-1091 [FREE Full text] [doi: 10.1016/j.pec.2017.01.005] [Medline: 28117193]
28. Prictor M, Johnston C, Hyatt A. Overt and covert recordings of health care consultations in Australia: some legal considerations. Med J Aust 2021 Feb;214(3):119-23.e1. [doi: 10.5694/mja2.50838] [Medline: 33131072]

Abbreviations

HITH: hospital in the home
RMH: Royal Melbourne Hospital

Edited by E Borycki; submitted 04.10.23; peer-reviewed by E Ray Chaudhuri, J Jih, E Nilsson; comments to author 30.05.24; revised version received 19.07.24; accepted 23.09.24; published 20.11.24.

Please cite as:
Schmidt S, Boulton A, Butler B, Fazio T
Exploring Patient, Proxy, and Clinician Perspectives on the Value and Impact of an Inpatient Portal: A Reflexive Thematic Analysis
JMIR Hum Factors 2024;11:e52703
URL: https://humanfactors.jmir.org/2024/1/e52703
doi:10.2196/52703
PMID:

Optimization of a Web-Based Self-Assessment Tool for Preconception Health in People of Reproductive Age in Australia: User Feedback and User-Experience Testing Study

Edwina Dorney[1], BAPSC, MBBS (Hons), MPH; Karin Hammarberg[2], RN, PhD; Raymond Rodgers[3], BAgrSc, MAgrSc, PhD; Kirsten I Black[1], MBBS, MM, PhD

[1]Faculty of Medicine and Health, The University of Sydney, Susan Wakil Health Building, Western Avenue, Camperdown, Australia
[2]Global and Women's Health, School of Public Health, and Preventative Medicine, Monash University, Melbourne, Australia
[3]Robinson Research Institute School of Biomedicine, University of Adelaide, Adelaide, Australia

Corresponding Author:
Edwina Dorney, BAPSC, MBBS (Hons), MPH
Faculty of Medicine and Health, The University of Sydney, Susan Wakil Health Building, Western Avenue, Camperdown, Australia

Abstract

Background: Good preconception health reduces the incidence of preventable morbidity and mortality for women, their babies, and future generations. In Australia, there is a need to increase health literacy and awareness about the importance of good preconception health. Digital health tools are a possible enabler to increase this awareness at a population level. The Healthy Conception Tool (HCT) is an existing web-based, preconception health self-assessment tool, that has been developed by academics and clinicians.

Objective: This study aims to optimize the HCT and to seek user feedback to increase the engagement and impact of the tool.

Methods: In-depth interviews were held with women and men aged 18 - 41 years, who spoke and read English and were residing in Australia. Interview transcripts were analyzed, and findings were used to inform an enhanced HCT prototype. This prototype underwent user-experience testing and feedback from users to inform a final round of design changes to the tool.

Results: A total of 20 women and 5 men were interviewed; all wanted a tool that was quick and easy to use with personalized results. Almost all participants were unfamiliar with the term "preconception care" and stated they would not have found this tool on the internet with its current title. User-experience testing with 6 women and 5 men identified 11 usability issues. These informed further changes to the tool's title, the information on how to use the tool, and the presentation of results.

Conclusions: Web-based self-assessment tools need to be easy to find and should communicate health messages effectively. End users' feedback informed changes to improve the tool's acceptability, engagement, and impact. We expect that the revised tool will have greater reach and prompt more people to prepare well for pregnancy.

(JMIR Hum Factors 2024;11:e63334) doi:10.2196/63334

KEYWORDS
technology; internet; eHealth; user experience; patient engagement; self-assessment tool; preconception; health communication

Introduction

Overview

Preconception care aims to optimize the physical and psychological health of women and their partners prior to conception [1]. Good preconception health reduces the incidence of preventable morbidity and mortality for women and their babies [1,2]. In Australia, there is a need to increase awareness about the importance of good preconception health and access to preconception care [3]. This is evidenced by the high proportion of women entering pregnancy above a healthy weight, the low proportion of women taking folic acid before conception, and the low rate of good prepregnancy glycemic control in women with diabetes [4-7].

Barriers to the delivery of preconception care include inadequate knowledge in the community about the importance of optimal preconception health and a lack of presentation to health care providers for preconception assessment [8,9]. Additional barriers include women from low socioeconomic backgrounds having lower levels of functional health literacy [10]. This is associated with lower rates of consulting a health professional for preconception care and worse preconception health behavior [10,11]. Women also identify lack of time and living in rural and remote communities as prohibitive factors to seeking preconception care [12].

Evidence suggests that people of reproductive age are keen to learn about preconception health and adopt positive health behavior change before a pregnancy. Most women of reproductive age report a preference for internet-based information sources [13] and use technology to find information on pregnancy health [13,14]. Digital health tools, such as web-based self-assessment tools, are a promising medium to increase knowledge and awareness about preconception health among people of reproductive age across geographical locations [15,16]. This includes increased understanding of the benefits of good preconception health, the risks of poor preconception health, and the opportunities to improve health before conception [17,18]. Impacts of web-based preconception health self-assessment tools include reduced rates of preconception alcohol consumption, improved uptake of folic acid supplementation [19], and also act as a catalyst for clients to initiate discussions with health care providers [17].

The Healthy Conception Tool

The Healthy Conception Tool (HCT) was an existing web-based preconception health self-assessment tool, developed by YourFertility, a Commonwealth-funded fertility health promotion program in Australia. The HCT asks questions about a person's general and reproductive health from which people then get a personalized output of results. They are then encouraged to take the results to their doctor to discuss what they can do to optimize their health. The HCT contains a section for both women and men. The HCT was developed in 2017 in collaboration with academics, clinicians, and researchers at the Robinson Research Institute. Some feedback from users was used in the initial development of the HCT; however, formal usability testing was not performed. Exploring usability can identify issues with how people engage and interact with a tool, and the findings can be used to improve experience to maximize the effectiveness and efficiency of the tool [20].

As the HCT is a potential enabler to promote the importance of preconception health, particularly for people in rural and remote areas who face additional challenges in accessing health care, optimization and usability testing can enhance a tool for these populations. This paper describes the process taken to engage with communities using a person-centered approach to assess usability and optimize an existing web-based preconception health self-assessment tool for people in Australia.

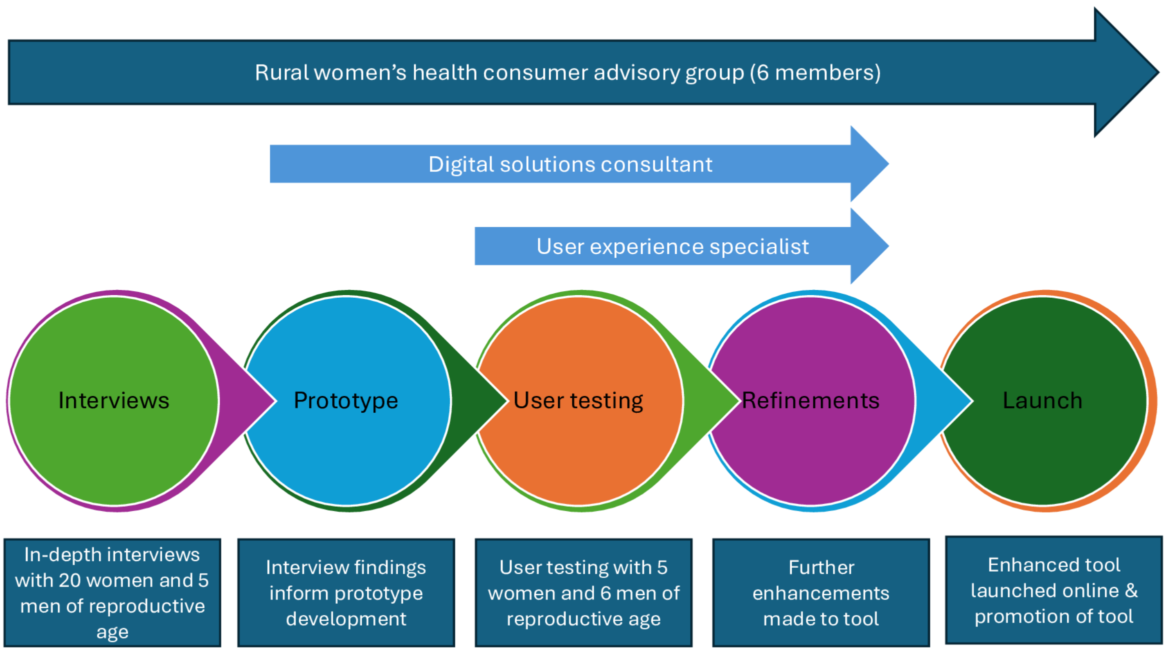
Methods

Overview

In-depth user interviews and usability testing were used to optimize the existing HCT (Figure 1).

A rural women's health consumer advisory group (RWH-CAG) was established to oversee this body of work and ensure person-centered design in all aspects. A total of 6 women aged 19 - 30 years old, who resided in rural and remote locations in Australia, were selected from a pool of 17 applicants. Selected applicants had lived experience and demonstrated knowledge of issues around the access of sexual and reproductive services to rural Australian women, particularly in the areas of family planning and preconception care. The role of the RWH-CAG was to inform the study design and recruitment methods (Figure 1).

Figure 1. Process of optimizing a web-based self-assessment tool for preconception care.



Recruitment

People of reproductive age (18 - 41 years old), residing in Australia, who could speak and read English, were eligible to participate in both the interviews and user-experience testing. People of all relationship statuses were eligible to take part.

Recruitment for interviews was via social media with advertisements targeting regional and remote community parent groups and noticeboards. Recruitment for user-testing was coordinated through a user-experience agency (Nomat Australia) with an Australia-wide registered pool of user-experience participants. The participants meeting the recruitment specifications listed above were sent an invitation to take part.

Process

In-Depth Interviews

In-depth interviews to understand participants' experiences of using the HCT were performed from February to September 2021. In particular, what features of the HCT were most likely to increase their knowledge and influence behavior change, and to explore ways in which the tool could be improved. Given COVID-19 quarantine restrictions, interviews were conducted via telephone or internet-based videoconferencing at the participants' request. Before the interview, participants were required to complete the HCT. Interviews were recorded and transcribed, and data were analyzed by two researchers who have reproductive and public health expertise, using an inductive thematic approach [21,22]. Data were coded, manually without assistance of software programs, to identify consumer likes and dislikes across content and design features of the self-assessment tool. These findings were further discussed among all members of the research team, and agreement was reached on key domains and features for optimization

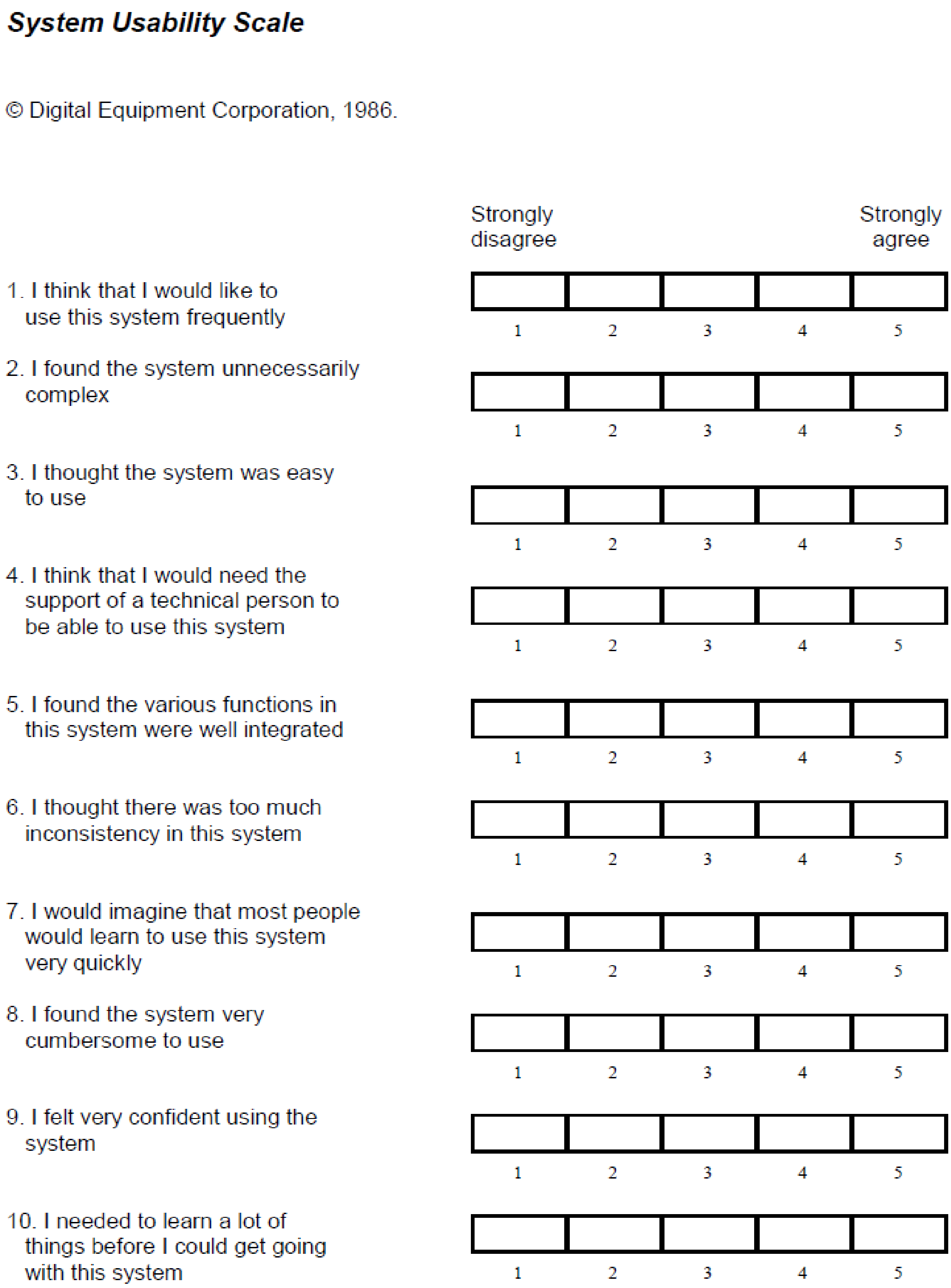
Prototype Development

A series of strategy, planning, and design workshops were held with a digital solutions consultant (Sentius Australia). These workshops aimed to develop creative and digital solutions for the new tool design, informed by the interview findings. A prototype was developed by Sentius and the research team, with a list of features identified for user testing.

Optimization of Prototype With User Testing

The prototype underwent user-experience testing, conducted by a user-experience specialist to identify any usability issues and validate design changes that had been made from May to June 2022. Again, given the COVID-19 pandemic restrictions, the participants took part in a 1-hour videoconferencing call for user-experience testing. In this call, the participants completed the HCT in their chosen setting, to reflect their typical context. The participants shared their smart device screen so that their tool navigation could be directly observed. The participants completed the System Usability Scale (SUS) [23] and had the option to provide additional feedback on their experience. The SUS is a 10-item questionnaire to enable a standardized assessment of usability of an interface (Figure 2) and is an accepted tool to benchmark the usability of Digital Health Applications [24]. Usability problems were categorized as minor, moderate, or critical relating to the impact they had on tool engagement and completion (Table 1).

Figure 2. System Usability Scale [23].



System Usability Scale

© Digital Equipment Corporation, 1986.

	Strongly disagree 1	2	3	4	Strongly agree 5
1. I think that I would like to use this system frequently					
2. I found the system unnecessarily complex					
3. I thought the system was easy to use					
4. I think that I would need the support of a technical person to be able to use this system					
5. I found the various functions in this system were well integrated					
6. I thought there was too much inconsistency in this system					
7. I would imagine that most people would learn to use this system very quickly					
8. I found the system very cumbersome to use					
9. I felt very confident using the system					
10. I needed to learn a lot of things before I could get going with this system					

Table . Category definitions for identified usability issues.

Classification	Definition
Minor	Causes some hesitation or irritation
Moderate	Causes task failure for some users
Critical	Leads to task failure. Issue inhibits users from completing core tasks or may impact core business objectives

Prototype Refinement and Enhancements

User-experience testing results informed additional changes to the tool. Another round of planning workshops was held to decide on the final changes to the tool. These were tested amongst the research group for quality assurance.

Launch

When all changes were made, the enhanced tool was launched on the internet.

Ethical Considerations

This study was approved the University of Sydney Human Research Ethics Committee (2020/430; June 22, 2020). Verbal consent was obtained prior to the interview and user testing. Recruitment for both processes ended when no new data were obtained, and previously identified themes were repeating. Participants were given a Aus $50 (US $37.50) VISA gift card in recognition of their time.

Results

Overview

In-depth interviews with 20 women and 5 men, and user-experience testing with 6 women and 5 men of reproductive age were performed. The participants included people planning and not planning a pregnancy in the next 12 months, from metropolitan, regional, and remote locations in Australia. The participant demographics are shown in Table 2.

Table . Participant demographics.

Demographics	Interview participants (n=25), n	User testing participants (n= 11), n
Gender		
Female	20	6
Male	5	5
Pregnancy intention		
Planner	10	6
Nonplanner	15	5
Place of residence		
Metropolitan	11	9
Regional and rural	14	2

Interview Results

All participants completed the HCT prior to their interviews, using either a mobile device or laptop, with some completing the tool on both a mobile and laptop device. Attributes of the self-assessment tool, as identified from the interviews, are presented below.

General Features

Overall, participants were very positive about the HCT. Most participants liked that it was quick to use, taking on average 5 minutes to complete, and that it was a single source for multiple topics of pregnancy preparation.

Source Credibility

The participants expressed that it was important to know the tool was from a credible source. This was attributed to endorsement from reputable professional organizations such as academic institutions.

> *I look for does it come from a trusted source, is it from the Royal Children's Hospital website or is it from Breastfeeding Victoria's website, not just a mummy blog or a website that has obvious ads or links to creams that they want to sell you.* [P19, female, regional, nonplanner]

Target Audience

Participants liked that the tool was inclusive of partners with information for both men and women.

> *I liked that it had a section for men and women, I think that's really good, that men are included in it as well.* [P6, female, regional, nonplanner]

Finding the Tool

The participants identified that a good tool is one that is intuitive to find on a simple web-based search. Almost all participants stated that they would not have found this tool on the internet with its current title.

Information Gained

The participants liked the links to further information, particularly the multiple and varied topics that were collated together.

> *And I think the link that it had when you clicked on the little info icon, they took you to places, and there was like so much information in those places.* [P6, female, regional, nonplanner]

Presentation of Results

The participants stated the presentation of their results was too general and that they wanted more personalized results. Consistently, the participants conveyed that the results must be meaningful to the individual to encourage behavior change.

User Experience

A total of 4 key domains for exploration in the user-experience testing were identified from the interviews:

1. User interface
2. Tool navigation
3. User experience
4. Results

A summary of the features for testing within these 4 domains is shown in Table 3. More details on the interview findings relating to these domains and features are available in Multimedia Appendix 1.

Table . Domains and features of a web-based self-assessment tool for preconception care to be assessed in usability testing.

Domain	Features to test
User interface	
Color	• Color scheme
Images	• Real life images • Icons
Tool navigation	
Answer mechanisms	• Free text • Radio buttons • Scroll bar • Specified range
User experience	
Explanatory text	• Explain how to answer the questions and what to do on completion
Sequence of questions	• Sequence of questions for logical flow
Language	• Title and text to avoid medical terminology • Inclusive language and answer options • Tone • Appropriate health literacy
Accessibility of additional information	• Display of additional information
Results	
Timing of results display within the tool	• Give result with each individual question answered • Give all results at the end of the tool
Visual display of results	• Visual display of results with color coding system
Prioritized ordering and personalization of results	• Prioritized display of results
Mechanism to receive and keep results	• Email • Print

User-Experience Testing

All participants completed all questions in the prototype on their first attempt and responded well to its simplicity. The amount of information provided was considered appropriate, the supplied links were determined to be valuable, and information was perceived to be presented in a way that was easy to digest and use. The average SUS score was 91.82 (SD 4.62), which is a high score compared with an industry average in Australia of 68 (SD 12.5). The participants responded well to the given color scheme.

A total of 11 usability problems were identified, 4 moderate and 7 minor (Table 4).

Additional observations of significance were the impact of the organizational logos at the bottom of the tool. Without prompting, several participants commented on the organization logos and indicated that seeing this added credibility to the tool. New features to increase reach of the tool were also noted and included the option to share the tool with the user's contacts.

Additional refinements were made to the tool following user-experience testing to address the problems identified. Some key changes are shown in Figure 3.

Table . The findings on user-experience testing for a web-based self-assessment tool for preconception care.

Issue	Observation	Recommendation	Severity
Results			
Presentation of results—traffic light system colours	Consistently, the color coding on the results page was not obvious to participants. Particularly, the use of orange instead of red was misunderstood. "*I don't understand why they've done that...maybe change the choice of colours to red, yellow and green? I would be fine with red*" (P36, male, planner, metropolitan)	Update to green, orange, red and introduce descriptive headings within the results page.	Moderate
Grouping preference	Consistently, participants indicated that they preferred the layout of the second version of the results page (order as per traffic lights). Specifically, participants suggested that grouping results led to increased understanding of the content grouping. "*Now it makes more sense to see the colours categorised in groups.*" (P28, female, non-planner, metropolitan)	Implement grouping of results as the standard.	Moderate
User experience			
Help icon: visibility	The help icon was deemed to be of value to participants, however it was not immediately obvious to all participants. The visual hierarchy needs to be increased, in order to prompt a higher use rate. "*No, I don't recall seeing these help icons, that would have been good to know for this question [fertile window].*" (P26, female, non-planner, metropolitan)	Consider making the help icon 15% larger and changing the color to HEX #14c797 to increase prominence.	Moderate

Issue	Observation	Recommendation	Severity
Answer clarification: smoking and alcohol	For both the smoking and alcohol use questions, some participants answered ”no” despite describing infrequent use. This indicated that just having 2 possible answers causes misreporting in some instances. “*It’s a tad confusing, the smoking and drinking. It’s a yes and a no answer but there’s a lot of gradations in there...someone might have two glasses a week.*” (P34, male, planner, metropolitan)	Include help icon to explain how to answer the question.	Moderate
Question clarification			
Chemicals	Participants often displayed confusion as to what kind of chemicals this question referred to. “*I don’t know much about chemical exposure...maybe a Tefal pan? I’m not sure. That would have been a question I could use some extra information for.*” (P28, female, nonplanner, metropolitan)	Include help icon text to expand on the types of household chemicals that could impact fertility.	Minor
STI check	On occasion, participants expressed surprise at the inclusion of the STI check question. Specifically, participants did not feel it was relevant to them if they were in a long-term relationship. *“Why is this question relevant to pregnancy?”* (P30, male, planner, metropolitan)	Include help icon text to explain the relationship between STI checks and future fertility. Reiterate STIs can be asymptomatic	Minor
Prescription meds	Participants indicated confusion around the types of prescription medication that would be included here. *“I’m on the mini [contraceptive] pill, but I wouldn’t tick yes here.*” (P27, female, planner, metropolitan)	Add help icon text to explain prescription medications	Minor

Issue	Observation	Recommendation	Severity
Folic acid	Folic acid prompted some additional discussion from participants, with questions raised about what it was, the dose required, and the function it served. "*I would answer no to taking folic acid...I haven't even heard that mentioned. Now I wonder if that is important or not...I am not sure what it is.*" (P33, female, planner, regional)	Add help icon text to give a brief explanation of the importance of folic acid. Mention that folic acid is included in pre-natal supplements.	Minor
Results page: BMI information	Participants indicated that the BMI information required additional context to be more effective. "*The BMI information is good, but you just want a quantified benefit. You always hear about eating healthy and being healthy...it just becomes noise. Unless there's a real benefit, what's the point?*" (P30, male, planner, metropolitan)	Expand results section to include specific statistics connecting BMI with fertility. Include more targeted exercise suggestions.	Minor
Language: male version	Throughout the text on the male pages, some of the wording occasionally confused male participants. Some male participants did not feel that this was relevant to them and questioned whether the results were actually targeted towards the male user.	Tailor copy for participant gender.	Minor
User interface			
Image: sperm regeneration	Without being prompted, male participants consistently drew attention to the sperm regeneration fact at the top of the results page. Participants indicated that this statistic was new to them, and that they valued having this information included.	On the results page, incorporate this information within the main content.	Minor

Figure 3. Enhancements to the self-assessment tool questions and results.

Discussion

Principal Findings

This work explored what features of a web-based self-assessment tool for preconception care are important to people of reproductive age. In particular, what features will increase engagement and completion of the tool, and what are the best ways to present the results and information so that a user will act upon them.

Our findings showed that participants value a tool that is intuitive to find in a web-based search and is quick and simple to use. This is consistent with findings of other eHealth modalities, where ease of use and simplicity is a determinant in user engagement [25,26]. Clear information about what each question is asking was important to maintain engagement in the tool. For questions that were not intuitive, or required explanation, an information icon was placed adjacent to the question text, in a different color to emphasize its presence. The information was visible by hovering over the icon, as participants indicated they did to want to be directed to an additional page, as this interfered with user experience. Having information presented in a way that is easily understood by the user has been identified as a key quality indicator of web-based health sites [27].

A key finding in both the interviews and user-experience testing was the need to have information clearly presented and easy to digest at an appropriate health literacy level. The International Reproductive Health Education Collaboration recently devised recommendations for developing and implementing tools to improve fertility literacy [28]. This included the recommendation to understand the target population and include user perspectives when developing education tools. In both the interviews and user-experience testing. there was a balance of planners and nonplanners (those not planning a pregnancy in the next 12 mo) to ensure the information presented was accessible and relevant to all people of reproductive age regardless of pregnancy intention. This also led to some key changes including to the title of the tool. Almost all of the interview participants would not have found the tool on the internet, and as such the tool was renamed with a plain English title as informed by participants of "Healthy you, Healthy baby."

Another key feature for users was the personalized and ordered presentation of tool results. Results were grouped into categories that required the user's attention and action. These were color-coded to convey visual priority and accompanied by explanatory text, in a positive tone to complement the color scheme.

The participants expressed the importance of knowing that a source is credible, and this is acknowledged by the inclusion of logos from trusted organisations. Trust in eHealth sites is a recognized determinant of user engagement with web-based health information sources [29]. Accreditation or endorsement with recognized logos from reputable institutions can increase trust in a platform [30].

eHealth platforms have been shown to be effective in improving user health knowledge, behavior change, and health outcomes [31,32]. Several eHealth platforms have been shown to be effective for the communication of preconception health information [33-35]. A web-based app for people with sickle cell disease and sickle cell trait providing information about prepregnancy health was found to be acceptable and usable and increased consumer knowledge [33]. An eHealth lifestyle coaching program for women prior to pregnancy has been shown to increase healthy eating behaviours [34,35].

Studies have suggested improvement in eHealth intervention designs to increase their effectiveness [32]. This includes adopting a holistic approach to promote user engagement [36]. Our approach, informed by a rural women's health consumer advisory group, of in-depth interviews followed by user-experience testing enabled a detailed understanding of our target audience needs and expectations. The opportunity to test the consumer, informed prototype, and validated design has delivered an enhanced tool for people of reproductive age in Australia.

Strengths

The use of both interviews and user-experience testing techniques are a strength of this study and provided additional iterations to enhance the self-assessment tool. The involvement of the RWH-CAG from conception to completion of this project also ensure a person-centred approach.

Limitations

The tool was only explored by people who can speak and read English, and therefore does not capture the preferences of people from culturally and linguistically diverse backgrounds. As these populations can face challenges with access to care, this is a priority area for future research. This study used the SUS as an instrument within the usability testing. Further enhancements may be achieved by using additional tools such as Neilsen's guidelines in combination with the SUS [37].

Conclusion

As a web-based tool, "Healthy You, Healthy Baby" has the potential to improve knowledge among people of reproductive age about the importance of optimal preconception health, including those who experience health inequities, such as women and men in rural and remote areas. The tool can be adapted to other priority populations, including people from culturally and linguistically diverse backgrounds to further improve the delivery of preconception care.

Acknowledgments

The authors acknowledge the input of the rural women's health consumer advisory group members for the invaluable contribution to the design and development of recruitment materials for this study. We also acknowledge the contribution of Renee da Silva and Joanna Anagnostou who were the Health Promotion Coordinators at YourFertility during the time of this research. The authors also acknowledge the collaboration with Kim Morrison and Mel King at Sentius for digital design, and with Chris Gray and Eliza Crisp at Nomat for user experience testing. This research was funded by the National Women's Health Strategy Health Promotion Grants Scheme GO2745.

Authors' Contributions

ED contributed to the conceptualization, methodology, investigation, data curation, formal analysis, writing of the original draft, review and edits, visualization, project administration, and funding acquisition. KH performed the conceptualization; formal analysis; and the writing, reviewing and editing of the manuscript. RR contributed to the conceptualization and the writing, reviewing, and editing of the manuscript. KIB performed the conceptualization; data curation; formal analysis; writing, reviewing, and editing of the manuscript; supervision; and funding acquisition.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview findings outlining the domains and features for testing of the online self-assessment tool for preconception care.
[DOCX File, 25 KB - humanfactors_v11i1e63334_app1.docx]

References

1. Preconception care: maximizing the gains for maternal and child health - policy brief. World Health Organization. 2013. URL: https://www.who.int/publications/i/item/WHO-FWC-MCA-13.02 [accessed 2024-12-06]
2. Stephenson J, Heslehurst N, Hall J, et al. Before the beginning: nutrition and lifestyle in the preconception period and its importance for future health. Lancet 2018 May;391(10132):1830-1841. [doi: 10.1016/S0140-6736(18)30311-8]
3. Boyle JA, Black K, Dorney E, et al. Setting preconception care priorities in Australia using a Delphi technique. Semin Reprod Med 2022 Jul;40(3-04):214-226. [doi: 10.1055/s-0042-1749683] [Medline: 35760312]
4. Lang AY, Harrison CL, Barrett G, Hall JA, Moran LJ, Boyle JA. Opportunities for enhancing pregnancy planning and preconception health behaviours of Australian women. Women Birth 2021 Mar;34(2):e153-e161. [doi: 10.1016/j.wombi.2020.02.022] [Medline: 32312651]
5. Malek L, Umberger W, Makrides M, Zhou SJ. Poor adherence to folic acid and iodine supplement recommendations in preconception and pregnancy: a cross-sectional analysis. Aust N Z J Public Health 2016 Oct;40(5):424-429. [doi: 10.1111/1753-6405.12552] [Medline: 27523027]
6. Australia's mothers and babies. Australian Institute of Health and Welfare (AIHW). 2023. URL: https://www.aihw.gov.au/reports/mothers-babies/australias-mothers-babies/contents/antenatal-period/maternal-body-mass-index [accessed 2024-11-10]
7. Abell SK, Boyle JA, de Courten B, et al. Impact of type 2 diabetes, obesity and glycaemic control on pregnancy outcomes. Aust N Z J Obstet Gynaecol 2017 Jun;57(3):308-314. [doi: 10.1111/ajo.12521] [Medline: 27593528]
8. Mazza D, Chapman A, Michie S. Barriers to the implementation of preconception care guidelines as perceived by general practitioners: a qualitative study. BMC Health Serv Res 2013 Jan 31;13(1):36. [doi: 10.1186/1472-6963-13-36] [Medline: 23368720]
9. Goossens J, De Roose M, Van Hecke A, Goemaes R, Verhaeghe S, Beeckman D. Barriers and facilitators to the provision of preconception care by healthcare providers: a systematic review. Int J Nurs Stud 2018 Nov;87:113-130. [doi: 10.1016/j.ijnurstu.2018.06.009] [Medline: 30096578]
10. Endres LK, Sharp LK, Haney E, Dooley SL. Health literacy and pregnancy preparedness in pregestational diabetes. Diabetes Care 2004 Feb;27(2):331-334. [doi: 10.2337/diacare.27.2.331] [Medline: 14747209]
11. Fransen MP, Hopman ME, Murugesu L, Rosman AN, Smith SK. Preconception counselling for low health literate women: an exploration of determinants in the Netherlands. Reprod Health 2018 Nov 23;15(1):192. [doi: 10.1186/s12978-018-0617-1] [Medline: 30470239]
12. Poels M, Koster MPH, Boeije HR, Franx A, van Stel HF. Why do women not use preconception care? A systematic review on barriers and facilitators. Obstet Gynecol Surv 2016 Oct;71(10):603-612. [doi: 10.1097/OGX.0000000000000360] [Medline: 27770130]
13. Khan NN, Boyle JA, Lang AY, Harrison CL. Preconception health attitudes and behaviours of women: a qualitative investigation. Nutrients 2019 Jun 29;11(7):1490. [doi: 10.3390/nu11071490] [Medline: 31261954]
14. Musgrave L, Cheney K, Dorney E, Homer CSE, Gordon A. Addressing preconception behavior change through mobile phone apps: systematic review and meta-analysis. J Med Internet Res 2023 Apr 19;25:e41900. [doi: 10.2196/41900] [Medline: 37074767]
15. Van Dijk MR, Koster MP, Rosman AN, Steegers-Theunissen RP. Opportunities of mHealth in preconception care: preferences and experiences of patients and health care providers and other involved professionals. JMIR Mhealth Uhealth 2017 Aug 17;5(8):e123. [doi: 10.2196/mhealth.7834] [Medline: 28818820]
16. Gardiner P, Hempstead MB, Ring L, et al. Reaching women through health information technology: the Gabby preconception care system. Am J Health Promot 2013;27(3 Suppl):eS11-eS20. [doi: 10.4278/ajhp.1200113-QUAN-18] [Medline: 23286652]
17. Bello JK, Adkins K, Stulberg DB, Rao G. Perceptions of a reproductive health self-assessment tool (RH-SAT) in an urban community health center. Patient Educ Couns 2013 Dec;93(3):655-663. [doi: 10.1016/j.pec.2013.09.004] [Medline: 24094839]
18. Levis DM, Westbrook K. A content analysis of preconception health education materials: characteristics, strategies, and clinical-behavioral components. Am J Health Promot 2013;27(3 Suppl):S36-S42. [doi: 10.4278/ajhp.120113-QUAL-19] [Medline: 23286661]

19. Agricola E, Pandolfi E, Gonfiantini MV, et al. A cohort study of a tailored web intervention for preconception care. BMC Med Inform Decis Mak 2014 Apr 15;14:33. [doi: 10.1186/1472-6947-14-33] [Medline: 24731520]
20. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Feb;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
21. Braun V, Clarke V, Hayfield N, Terry G. In: Liamputtong P, editor. Handbook of Research Methods in Health Social Sciences: Springer; 2019:843-896. [doi: 10.1007/978-981-10-5251-4_103]
22. Braun V, Clarke V. What can "thematic analysis" offer health and wellbeing researchers? Int J Qual Stud Health Well-being 2014;9:26152. [doi: 10.3402/qhw.v9.26152] [Medline: 25326092]
23. Brooke J. SUS—A Quick and Dirty Usability Scale. Usability Eval in Ind 1996;189:4-7 [FREE Full text]
24. Hyzy M, Bond R, Mulvenna M, et al. System usability scale benchmarking for digital health apps: meta-analysis. JMIR Mhealth Uhealth 2022 Aug 18;10(8):e37290. [doi: 10.2196/37290] [Medline: 35980732]
25. Lazard AJ, Watkins I, Mackert MS, Xie B, Stephens KK, Shalev H. Design simplicity influences patient portal use: the role of aesthetic evaluations for technology acceptance. J Am Med Inform Assoc 2016 Apr;23(e1):e157-e161. [doi: 10.1093/jamia/ocv174] [Medline: 26635314]
26. Wang H, Zhang J, Luximon Y, Qin M, Geng P, Tao D. The determinants of user acceptance of mobile medical platforms: an investigation integrating the TPB, TAM, and patient-centered factors. Int J Environ Res Public Health 2022 Aug 29;19(17):10758. [doi: 10.3390/ijerph191710758] [Medline: 36078473]
27. Tao D, LeRouge C, Smith KJ, De Leo G. Defining information quality into health websites: a conceptual framework of health website information quality for educated young adults. JMIR Hum Factors 2017 Oct 6;4(4):e25. [doi: 10.2196/humanfactors.6455] [Medline: 28986336]
28. Martins MV, Koert E, Sylvest R, et al. Fertility education: recommendations for developing and implementing tools to improve fertility literacy†. Hum Reprod 2024 Feb 1;39(2):293-302. [doi: 10.1093/humrep/dead253] [Medline: 38088127]
29. Sbaffi L, Rowley J. Trust and credibility in web-based health information: a review and agenda for future research. J Med Internet Res 2017 Jun 19;19(6):e218. [doi: 10.2196/jmir.7579] [Medline: 28630033]
30. Rowley J, Johnson F, Sbaffi L. Students' trust judgements in online health information seeking. Health Informatics J 2015 Dec;21(4):316-327. [doi: 10.1177/1460458214546772] [Medline: 25193449]
31. Sawesi S, Rashrash M, Phalakornkule K, Carpenter JS, Jones JF. The impact of information technology on patient engagement and health behavior change: a systematic review of the literature. JMIR Med Inform 2016 Jan 21;4(1):e1. [doi: 10.2196/medinform.4514] [Medline: 26795082]
32. Duan Y, Shang B, Liang W, Du G, Yang M, Rhodes RE. Effects of eHealth-based multiple health behavior change interventions on physical activity, healthy diet, and weight in people with noncommunicable diseases: systematic review and meta-analysis. J Med Internet Res 2021 Feb 22;23(2):e23786. [doi: 10.2196/23786] [Medline: 33616534]
33. Wilkie DJ, Telisnor G, Powell-Roach K, et al. CHOICES for sickle cell reproductive health: a protocol of a randomized preconception intervention model for a single gene disorder. PLoS ONE 2023;18(12):e0294907. [doi: 10.1371/journal.pone.0294907] [Medline: 38060589]
34. Hojeij B, Schoenmakers S, Willemsen S, et al. The effect of an eHealth coaching program (smarter pregnancy) on attitudes and practices toward periconception lifestyle behaviors in women attempting pregnancy: prospective study. J Med Internet Res 2023 Jan 31;25:e39321. [doi: 10.2196/39321] [Medline: 36719733]
35. van Dijk MR, Koster MPH, Oostingh EC, Willemsen SP, Steegers EAP, Steegers-Theunissen RPM. A mobile app lifestyle intervention to improve healthy nutrition in women before and during early pregnancy: single-center randomized controlled trial. J Med Internet Res 2020 May 15;22(5):e15773. [doi: 10.2196/15773] [Medline: 32412417]
36. Barello S, Triberti S, Graffigna G, et al. eHealth for patient engagement: a systematic review. Front Psychol 2015;6:2013. [doi: 10.3389/fpsyg.2015.02013] [Medline: 26779108]
37. Khajouei R, Farahani F. A combination of two methods for evaluating the usability of a hospital information system. BMC Med Inform Decis Mak 2020 May 4;20(1):84. [doi: 10.1186/s12911-020-1083-6] [Medline: 32366248]

Abbreviations

HCT: Healthy Conception Tool
RWH-CAG: rural women's health consumer advisory group
SUS: System Usability Scale

Edited by A Kushniruk; submitted 20.06.24; peer-reviewed by C Pizzuti, R Marshall; revised version received 22.09.24; accepted 14.10.24; published 24.12.24.

Please cite as:
Dorney E, Hammarberg K, Rodgers R, Black KI
Optimization of a Web-Based Self-Assessment Tool for Preconception Health in People of Reproductive Age in Australia: User Feedback and User-Experience Testing Study
JMIR Hum Factors 2024;11:e63334
URL: https://humanfactors.jmir.org/2024/1/e63334
doi:10.2196/63334

Nutrition Management Miniprograms in WeChat: Evaluation of Functionality and Quality

Hui Sun[1], BSN; Yanping Wu[1], MSN; Jia Sun[2], BSN; Wu Zhou[1], BSN; Qian Xu[1], BSN; Dandan Hu[1], BSN

[1]School of Medicine, Southeast University, No.87 Dingjiaqiao, Hunan Road, Naning, China

[2]The Affiliated Lianyungang Hospital of Xuzhou Medical University, Lianyungang, China

Corresponding Author:
Yanping Wu, MSN
School of Medicine, Southeast University, , No.87 Dingjiaqiao, Hunan Road, Naning, , China

Abstract

Background: With the rise in people's living standards and aging populations, a heightened emphasis has been placed in the field of medical and health care. In recent years, there has been a drastic increase in nutrition management in domestic research circles. The mobile nutritional health management platform based on WeChat miniprograms has been widely used to promote health and self-management and to monitor individual nutritional health status in China. Nevertheless, there has been a lack of comprehensive scientific evaluation regarding the functionality and quality of the diverse range of nutritional miniprograms that have surfaced in the market.

Objective: This study aimed to evaluate the functionality and quality of China's WeChat nutrition management miniprogram by using the User Version of the Mobile Application Rating Scale (uMARS).

Methods: This observational study involves quantitative methods. A keyword search for "nutrition," "diet," "food," and "meal" in Chinese or English was conducted on WeChat, and all miniprograms pertaining to these keywords were thoroughly analyzed. Then, basic information including name, registration date, update date, service type, user scores, and functional scores was extracted from January 2017 to November 2023. Rating scores were provided by users based on their experience and satisfaction with the use of the WeChat miniprogram, and functional scores were integrated and summarized for the primary functions of each miniprogram. Moreover, the quality of nutrition management applets was evaluated by 3 researchers independently using the uMARS.

Results: Initially, 27 of 891 miniprograms identified were relevant to nutrition management. Among them, 85.2% (23/27) of them offered features for diet management, facilitating recording of daily dietary intake to evaluate nutritional status; 70.4% (19/27) provided resources for nutrition education and classroom instruction; 59.3% (16/27) included functionalities for exercise management, allowing users to record daily physical activity; and only 44.4% (12/27) featured components for weight management. The total quality score on the uMARS ranged 2.85-3.88 (median 3.38, IQR 3.14-3.57). Engagement scores on the uMARS varied from 2.00 to 4.33 (median 3.00, IQR 2.67-3.67). Functional dimension scores ranged from 3.00 to 4.00 (median 3.33, IQR 3.33-3.67), with a lower score of 2.67 and a higher score of 4.33 outside the reference range. Aesthetic dimension scores ranged from 2.33 to 4.67 (median 3.67, IQR 3.33-4.00). Informational dimension scores ranged from 2.33 to 4.67 (median 3.33, IQR 2.67-3.67).

Conclusions: Our findings from the uMARS highlight a predominant emphasis on health aspects over nutritional specifications in the app supporting WeChat miniprograms related to nutrition management. The quality of these miniprograms is currently at an average level, with considerable room for functional improvements in the future.

(JMIR Hum Factors 2024;11:e56486) doi:10.2196/56486

KEYWORDS

nutrition management; WeChat mini-program; User Version of the Mobile Application Rating Scale; uMARS; function and quality evaluation

Introduction

With the intensification of the population's aging tendency, people pay increasing attention to the health of older adults. Specifically, nutrition management is of paramount importance for older adults, in that nutritional needs not only impact their quality of life but also significantly influence overall health status [1]. Aging, a normal process, is accompanied by physiological changes such as a loss of muscle mass, reduction in bone density, and decline in metabolic rate; Therefore, it is of vital importance to adjust nutritional intake among older individuals, especially those with weak digestive systems [2].

Malnutrition among older adults is a significant universal concern. Senile malnutrition, characterized by inadequate and imbalanced nutrition, arises from various factors, such as inappropriate dietary choices, insufficient intake, absorption disorders, etc [3]. The aging demographic shift contributes to a rise in the population of older adults, and the health status of the older adults directly impacts societal sustainability [4]. Consequently, there is an urgent need for effective nutrition policies and intervening measures for older adults. Among them, nutrition education, promotion of a balanced diet, optimization of medical and health care systems are essential strategies to prevent and improve senile malnutrition [5].

Thus far, the government has issued a series of policies, including the "Healthy China 2030 Plan Outline," which emphasizes the promotion of self-disciplined health behaviors and encourages balanced diets. Concurrently, the National Nutrition Plan (2017-2030) advocates for the integration of "Internet+nutrition and health" [6], endorsing the use of technology to manage public health and nutrition. Nowadays, this shift toward digital health management and intelligent nutrition support system for older adults reassure the growing emphasis on leveraging science and technology in health care. Through internet platforms and mobile apps, the dietary habits of older adults can be more effectively monitored, and personalized dietary recommendations can be provided to identify and address possible malnutrition issues promptly [7]. WeChat, an instant messaging software for smart terminals, was launched by Tencent on January 21, 2011, and it emerged in 2017 with its distinct advantages, including user-friendly miniprograms that do not require downloading, occupy minimal mobile phone memory, and include payment capabilities, circles of friends, public platforms, WeChat miniprograms, and other functionalities. All these features have contributed greatly to its widespread adoption among users [8].

Despite the growing number of mobile apps for nutrition guidance, there is remarkable variability among them due to a lack of specificity. Studies in China have predominantly focused on exploring the effects of interventions and the significance of nutrition research; yet, there remains a notable absence of scientific evaluation regarding the functionality and quality of commercially available nutrition miniprograms [9,10]. Accordingly, this study aimed to carry out a comprehensive search and assessment of relevant miniprograms using the User Version of the Mobile Application Rating Scale (uMARS) and develop WeChat-based applets for nutrition management.

Methods

Search Strategy

A keyword search for "nutrition," "diet," "food," and "meal" in Chinese or English was independently conducted on WeChat by 2 researchers, and all miniprograms pertaining to these keywords were thoroughly analyzed. They personally experienced the relevant miniprograms registered on WeChat between January 2017 and November 2023. Screening for miniprograms strictly adhered to predefined inclusion and exclusion criteria.

The inclusion criteria were as follows: (1) the miniprogram's functional content pertained to diet and nutrition, (2) it was available for free use, (3) its content was written in Chinese or English, and (4) it was compatible with mobile phones or tablets. The exclusion criteria were as follows: (1) it was never updated or maintained, (2) it was designed for commercial ordering and canteen services, (3) it was specifically for purchasing food, (4) it solely records data. During the screening process, the 2 researchers resorted to another researcher, if necessary, to resolve any discrepancies.

Sample Size

In this study, an observational research design was adopted to systematically search and quantitatively evaluate the function and quality of nutrition management miniprograms on WeChat. Using keywords such as "nutrition," "diet," "food," and "meal," a total of 891 miniprograms were initially identified. These miniprograms were then screened based on predefined inclusion and exclusion criteria. After the screening process, 27 eligible miniprograms were selected for detailed analysis.

Quality Assessment of WeChat Miniprograms Related to Nutrition Management

As described in a previous study [11], the uMARS was originally developed as an end user evaluation tool based on the Mobile Application Rating Scale (MARS), and it has been widely used for evaluating diverse categories of mobile health apps, which include those focusing on weight loss and nutrition, as well as the management of conditions such as rheumatism and ankylosing spondylitis [12].

As a comprehensive tool for evaluating the user experience of mobile apps, the uMARS was commonly used to gauge usability, user satisfaction, functionality, and other pertinent factors [13]. The 5 dimensions of the uMARS are engagement, functionality, aesthetics, information, and subjective quality, each dimension encompassing 3 - 5 questions and each question scored on a scale of 0-5 points (20 items in total) [14,15]. In the engagement dimension, the evaluator can evaluate the miniprogram's entertainment value, level of interest, customization, interactivity, and target audience appeal. The functional dimension facilitates the evaluation of performance, ease of use, navigation, and gesture design. The aesthetics section focuses on layout, graphics quality, and visual appeal, and the information section assesses information quality, quantity visual information, and the credibility of the information source [16-18]. Notably, to ensure evaluation consistency, the subjective scale is not included in the assessment process due to the highly subjective nature of evaluators' personal opinions and preferences.

Participants

Evaluation of the uMARS was conducted independently by 2 nutrition experts and 1 experiencer. Before the evaluation, each evaluator was required to read and further familiarize themselves with dimensions and items of the scale. All evaluators must comply with the consensus of the scoring criteria reached by the group discussion and evaluate each applet independently.

User Scores

User scores, a built-in feature of WeChat applets, are intended to assess users' overall satisfaction and experience, and a 5-point satisfaction scale is commonly used to measure people's satisfaction levels with the applets for research and surveys. Developers use these ratings to gauge the satisfaction levels of existing users and identify any areas for enhancement. If a feature of a WeChat miniprogram is not properly activated or if no users have participated in the rating, the default rating is set to 0.

Functional Scores

From among the 27 miniprograms selected, 16 functions were identified on the basis of registration and usage. Each miniprogram's functionality was evaluated by assigning 1 point for each aspect assessed, contributing to the cumulative total score. Feature scores ranged from 0 to 16, illustrating the comprehensiveness of the miniprograms' features.

Data Collection

Basic information regarding the 27 selected miniprograms was collected, encompassing metrics such as name, registration date, update date, service type, user scores, and functional scores. Subsequently, the functionality and quality of these miniprograms were quantitatively assessed using the uMARS.

Statistical Analyses

The functional score of each miniprogram was collected and expressed as quantity and percentage values. uMARS scores were described as mean and SD or IQR values. All data were analyzed using SPSS (version 26.0; IBM Corp).

Ethical Considerations

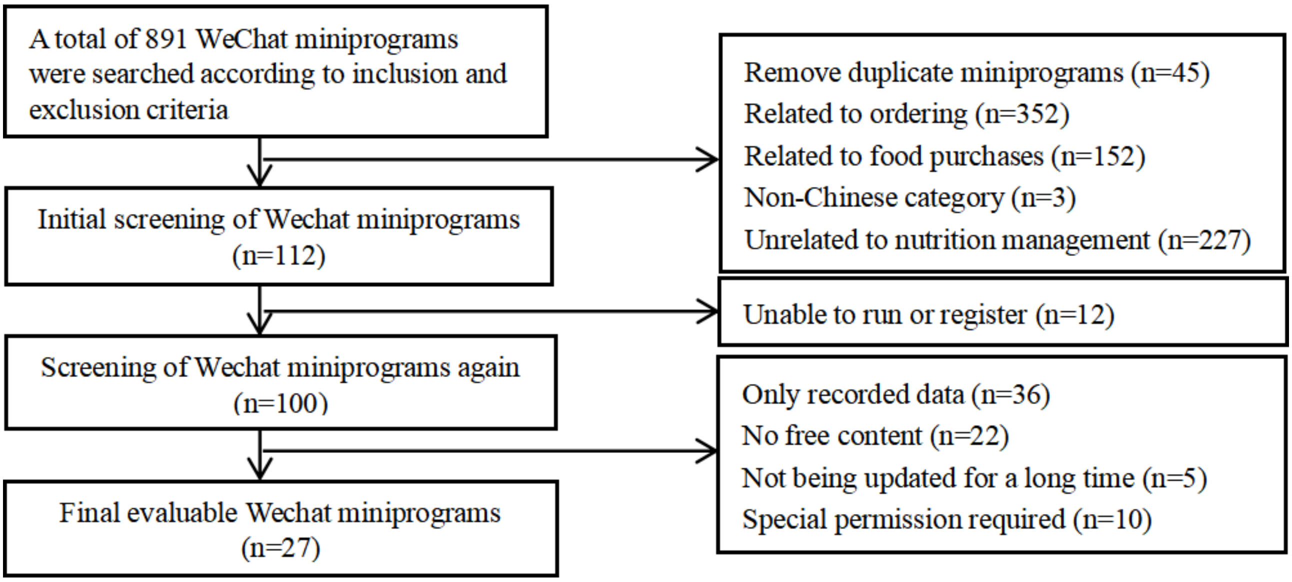
This study was approved by Zhongda Hospital affiliated to Southeast University (2024ZDSYLL194-P01).

Results

Screening of Miniprograms for Nutrition Management

We identified 891 miniprograms based on search terms (Figure 1). After removing the duplicate and unrelated miniprograms, 112 miniprograms were filtered. Ultimately, 27 miniprograms met the inclusion and exclusion criteria were selected for further study.

Figure 1. Flowchart for screening of nutrition management miniprograms.



Characteristics of Nutrition Management Miniprograms

We conducted a thorough search for miniprograms related to nutrition management through WeChat and used them after registration. Interestingly, these miniprograms shared several common characteristics, including name, registration date, update date, service type, user scores, and functional scores (Table 1).

Table . Characteristics of the nutrition- related WeChat mini-programs[a].

Name	Registration date	Update date	Service type	User scores	Functional scores
Peppermint nutritionist	Nov 15, 2018	Feb 10, 2023	Health management, medical information, food and beverage, and health products	—[b]	10
YOU nutrition	Apr 17, 2020	Nov 3, 2023	Health management, health care products, food and beverage, drug information display, and medical equipment sales platform	5	9
The more accurate and nutritious the diet	Oct 12, 2021	Nov 2, 2023	Food and beverage and health products	4.6	6
Long light nutrition diet therapy	Jul 19, 2021	Nov 2, 2023	Web-based education, educational information services, and medical information	—	7
Nutrition pagoda	Sep 14, 2022	Mar 10, 2023	Health management	3.6	7
Little Ann dietitian	Jan 15, 2021	Nov 4, 2023	Health management and equipment management	4.6	6
Abbott Medical Nutrition Care	Sep 18, 2019	Oct 28, 2023	Health management	4.5	5
Peppermint nutrition Pro	Dec 10, 2018	Jan 10, 2023	Health management	3.7	4
Nutritionist world	Aug 12, 2022	Nov 2, 2023	Health care products and food and beverage	3.5	4
Nutritional meal companion	Jul 9, 2022	Oct 28, 2023	Catering information service	—	6
Better One Nutritional fat reduction	Jan 19, 2022	Oct 28, 2023	Beauty service	—	6
Nutrition weight loss service platform	Aug 28, 2023	Nov 4, 2023	Health management and web-based fitness	—	8
Carkaka |Meal control card assistant	Feb 17, 2023	Oct 26, 2023	Recipe drinks, community/forum, and health management	4.1	6
High uric acid diet	Oct 27, 2017	Nov 6, 2023	Information inquiry, health management, and community/forum	4.3	6
Food diary	Mar 8, 2020	Nov 1, 2023	Catering information service and medical information	4.7	3
AI Dietary dietitian	Dec 5, 2022	Nov 6, 2023	Information query, video customer service, and health management	—	6
Little Doctor's diet diary	Nov 10, 2022	Jul 10, 2023	Health data statistics	—	5
High potassium diet	Dec 22, 2017	Nov 5, 2023	Information inquiry, health management, and community/forum	4.6	6
Chestnut food diary	Nov 27, 2018	Oct 28, 2023	Health management and medical information	4.3	8

Name	Registration date	Update date	Service type	User scores	Functional scores
Dietary calories	Oct 4, 2021	Sep 12, 2023	Information inquiry, food and beverage information service, and health management	4.2	5
Low carb diet assistant	Jun 21, 2019	Nov 12, 2022	Health management	4.4	4
Diet evaluation	Feb 28, 2023	Jun 12, 2023	Information, health management, and drug information display	—	2
Food notes	Feb 1, 2023	Sep 12, 2023	Health management	4.1	4
Sannuo Health	Sep 20, 2019	Nov 6, 2023	Medical information, community/forum, drug information display, and medical device manufacturer	—	9
Pick fruit health	May 27, 2020	Oct 13, 2023	Food and beverage, equipment management, medical equipment, health products, and health management	4.0	6
Mint Health	Apr 26, 2017	Oct 13, 2023	Catering information service and information inquiry	4.6	7
Peak Health Butler	Oct 11, 2022	Jul 13, 2023	Health management and information inquiry	4.2	5

[a](1) Health management: managing health through lifestyle changes; (2) medical information: information about diseases and health; (3) food and beverage: information about food nutrients; (4) health products: goods for maintaining or improving well-being; (5) drug information display: platform for medication details and usage guidance; (6) medical equipment sales platform: marketplace for health care equipment transactions; (7) educational information services: to provide information resources for health purposes; (8) medical information: information related to health care and treatments; (9) equipment management: access basic medical equipment data on the web; (10) catering information service: information hub for dining options and nutrition; (11) beauty service: offerings for cosmetic and aesthetic treatments; (12) web-based fitness: exercise and wellness programs accessible via the internet; (13) recipe drinks: formulations for beverages with health benefits; (14) community/forum: platform for discussions and interactions among users; (15) video customer service: support assistance provided through video communication; (16) health data statistics: analysis and presentation of health-related data; and (17) medical device manufacturer: producer of health care equipment.

[b]Not applicable.

WeChat Miniprogram for Nutrition Management

Following registration and use, the functional evaluation mainly focused on the nutrition management module of the miniprogram. This module encompassed 4 primary functions (diet management, weight management, exercise management, and nutrition education through class, video, or popular science content) and 12 auxiliary functions (Table 2). These auxiliary functions encompassed specific features such as comparisons, analysis, and recommendations, along with capabilities for managing blood sugar, blood pressure, and sleep. Furthermore, it included features for nutrition assessment, questionnaire survey, dietitian consultation, access to a nutrition marketplace, generation of a nutrition report, monitoring of biochemical indicators, participation in nutrition-related social circles, and health assessment [19-21].

Table . WeChat miniprogram for nutrition management.

WeChat mini-program function		Miniprograms, n (%)
Main function		
	Food record/management/analysis/clock in	23 (85.2)
	Exercise recording/management/analysis/clocking	16 (59.3)
	Weight recording/management/analysis/clocking	12 (44.4)
	Nutrition class/video/popular science	19 (70.4)
Auxiliary function		
	Food list	4 (14.8)
	Food comparison	3 (11.1)
	Food inquiry/analysis	8 (29.7)
	Food/recipe recommendations	13 (48.1)
	Blood pressure/blood sugar/sleep management	7 (25.9)
	Nutrition assessment/questionnaire	8 (29.7)
	Nutrition expert consultation	13 (48.1)
	Nutrition mall	10 (37.0)
	Nutrition report	7 (25.9)
	Biochemical index	5 (18.5)
	Nutrition sharing/friend circle/community	10 (37.0)
	Health assessment	8 (29.6)

As depicted in Table 2, the average functional score across all miniprograms was 6 points, ranging from 2 points in diet assessment to 10 points in peppermint nutrition food. Among the analyzed miniprograms, 85.2% (23/27) of them offered diet management features, facilitating the recording of daily dietary intake to assess nutritional status; 70.4% (19/27) of them provided nutrition knowledge and classroom teaching functionalities; 59.3% (16/27) of them offered exercise management capabilities, enabling users to record their daily physical activities; and only 44.4% (12/27) of them incorporated weight management functionalities.

In the functional evaluation, we conducted a comprehensive evaluation of the 4 main functions of each miniprogram. Simultaneously, we scrutinized and evaluated the auxiliary features to ensure that users could access comprehensive nutrition management services. Through this comprehensive assessment, we aimed to furnish users with detailed feedback, assisting them in selecting a high-quality nutrition management miniprogram, which was tailored to their requirements. Furthermore, this endeavor aimed to enhance users' health management practices and improve their overall and quality of life [22].

uMARS Quality Rating

An overview of the engagement, functionality, aesthetics, and information scores for the top 5 and bottom 5 miniprograms was presented in Table 3. Among them, the highest-ranking miniprogram was the WeChat applet mint health, with a uMARS score of 3.88 (SD 0.73), followed by AI (artificial intelligence) dietary dietitian, chestnut food diary, peppermint nutritionist, and sannuo health in turn. Notably, the user score approached 4.6 and the functional scores were 7.0. In contrast, the lowest ranking was WeChat applet-nutritionist world, with a uMARS score of 2.85 (SD 1.09), followed by diet evaluation, peppermint nutrition pro, nutritional meal companion, and long light nutrition diet therapy. In addition, the participation, function, aesthetics, and information scores of miniprograms are summarized in Table 3.

The uMARS total quality median score, calculated on a scale of 5, was 3.38 (IQR 3.14-3.57), with an overall range from 2.85 to 3.88, indicating that most miniprograms achieved scores above 3 points. The uMARS score of engagement ranged from 2.00 to 4.33, with a median score of 3.00 (IQR 2.67-3.67). The functional dimension varied from 3.00 to 4.00, and the median score was 3.33 (IQR 3.33-3.67), with a lower score of 2.67 and a higher score of 4.33 outside the reference range. The aesthetic dimension spanned from 2.33 to 4.67, with a median score of 3.67 (IQR 3.33-4.00), and the informational dimension ranged from 2.33 to 4.67, with a median score of 3.33 (IQR 2.67-3.67; Figure 2). Detailed uMARS scores for all 27 miniprograms are provided in Multimedia Appendix 1.

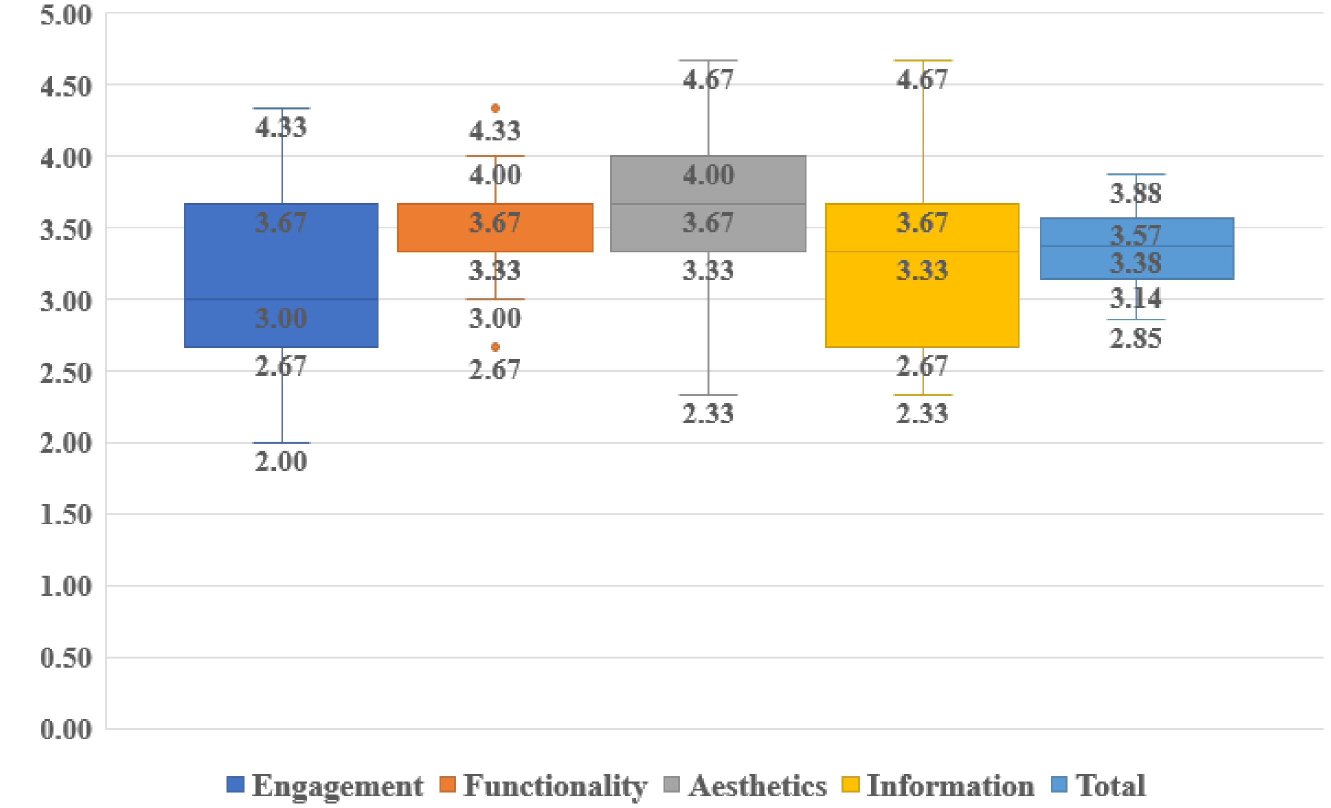
Table . The 5 highest- and lowest-scoring miniprograms (N=27) based on the User Version of the Mobile Application Rating Scale (uMARS).

WeChat applet	Engagement score, mean (SD)	Functionality score, mean (SD)	Aesthetics score, mean (SD)	Information score, mean (SD)	uMARS score, mean (SD)
Five highest-scoring miniprograms [a]					
	3.73 (0.96)	3.83 (0.72)	4.11 (0.33)	3.92 (0.67)	3.88 (0.73)
	3.73 (0.88)	3.67 (0.49)	4.33 (0.50)	3.83 (0.58)	3.85 (0.68)
	3.53 (0.92)	3.67 (0.49)	4.11 (0.60)	4.08 (0.79)	3.81 (0.76)
	3.47 (1.13)	3.67 (0.49)	4.00 (0.50)	4.00 (0.60)	3.75 (0.79)
	3.60 (1.12)	3.67 (0.65)	3.67 (0.50)	4.00 (0.74)	3.73 (0.82)
Five lowest-scoring miniprograms [b]					
	2.80 (0.86)	3.17 (0.58)	3.44 (1.01)	2.91 (0.79)	3.04 (0.82)
	2.73 (1.03)	3.08 (0.79)	3.33 (0.50)	3.17 (0.94)	3.04 (0.87)
	2.93 (1.22)	3.33 (0.49)	3.11 (0.93)	2.58 (1.24)	2.98 (1.04)
	2.93 (1.16)	3.33 (0.49)	3.11 (1.05)	2.50 (1.17)	2.96 (1.03)
	2.53 (1.89)	3.08 (0.90)	3.22 (1.09)	2.75 (1.14)	2.85 (1.09)

[a]Five highest-scoring miniprograms: (1) mint health, (2) AI (artificial intelligence) dietary dietitian, (3) chestnut food diary, (4) peppermint nutritionist, and (5) sannuo health.

[b]Five lowest-scoring miniprograms: (1) long light nutrition diet therapy, (2) nutritional meal companion, (3) peppermint nutrition pro, (4) diet evaluation, and (5) nutritionist world.

Figure 2. The User Version of the Mobile Application Rating Scale overall and section-specific scores of the nutrition management miniprograms (N=27).



Discussion

Principal Findings

A single user rating may not accurately gauge the quality of the miniprograms, and some existing evaluation systems for nutrition management miniprograms lack a scientifically grounded approach to promoting human nutrition. At present, the field of nutrition management miniprograms in China is still in its nascent stage and requires continual refinement of technique [18]. Analysis of functional scores revealed that only 18.5% of the miniprograms achieved scores above 8 points, indicating a need for improvement in their nutrition-related functionalities and health care. Among the identified issues, concerns were raised regarding the accuracy and comprehensiveness of the food database, potentially resulting in access to inaccurate nutrition. Furthermore, most miniprograms lack personalized services, consequently failing to offer tailored nutrition advice based on users' specific requirements and health conditions [23].

This study differs from prior research in several key aspects. First, many nutrition miniprograms now target specific disease types, tailoring diet management to the unique needs of patients. Second, some programs incorporate social circles to enhance user engagement, consequently fostering greater usage among patients [24]. Third, a subset of programs collaborates with the medical industry, engaging professional medical teams during development and establishing expert consultation platforms accessible via the WeChat miniprogram, tablet computer app, and computer software. Fourth, certain programs leverage data analysis and AI to deliver personalized nutrition advice and services, catering to individuals' specific needs [25]. Fifth, the seamless integration of these programs within WeChat enables direct access without the need for installation or downloads, ensuring faster and more convenient processes. Finally, developers continuously refine and expand nutrition management programs in response to evolving technological advancements and user preferences, thus integrating new functionalities and enhancing user experience to elevate program quality and competitiveness [26]. Notably, our evaluation further revealed that the quality of nutrition management miniprograms varies significantly, with many exhibiting incomplete content, imperfect functionality, and limited individualization and intelligence [27].

Limitations

Although this study reports some interesting and significant findings, there are several limitations associated with the use of WeChat miniprograms. First, nutrition miniprograms on WeChat are still undeveloped and nascent compared to nutrition apps with high usage rates. Second, due to the inability to download WeChat miniprograms, relevant data such as download counts and software size cannot be fully obtained, limiting our analysis of the data displayed on the platform [28]. Additionally, our research focuses solely on analyzing the functionality and quality of miniprograms within the WeChat platform, resulting in a relatively narrow scope. Accordingly, future studies could expand their scope by conducting questionnaire surveys among users of the existing miniprograms or by integrating user feedback more effectively [29]. Simultaneously, encouraging active involvement from health professionals in the development of mobile health apps is crucial for ensuring their effectiveness and relevance in promoting better health outcomes [30]. Moving forward, the trajectory of development should prioritize the enhancement of quality, introduction of innovative features, and fostering of active participation of professionals, consequently providing more scientifically grounded and practical personalized dietary guidance and enriching research on domestic nutrition management miniprograms [31-33].

Conclusions

Our findings from the uMARS highlight a predominant emphasis on health aspects over nutritional specifications in the application of WeChat miniprograms related to nutrition management. The quality of these small programs is currently at an average level, with considerable room for functional improvements in the future.

Acknowledgments

This study is supported by the Jiangsu Social Science Fund in 2022: Research on the path and countermeasures of digital Technology to promote the construction of the Jiangsu Health Community (22GLB010).

Authors' Contributions

This paper is a collaborative effort of all authors. HS and YPW developed and designed the study. HS, JS, WZ, QX, and DDH conducted the search for the miniprograms, data analysis, and scale evaluation. HS and YPW drafted the manuscript, and all authors participated in the review and editing of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
uMARS (User Version of the Mobile Application Rating Scale) scale score.
[XLSX File, 20 KB - humanfactors_v11i1e56486_app1.xlsx]

References

1. Hua N, Zhang Y, Tan X, et al. Nutritional status and sarcopenia in nursing home residents: a cross-sectional study. Int J Environ Res Public Health 2022 Dec 18;19(24):17013. [doi: 10.3390/ijerph192417013] [Medline: 36554892]
2. Rus GE, Porter J, Brunton A, et al. Nutrition interventions implemented in hospital to lower risk of sarcopenia in older adults: a systematic review of randomised controlled trials. Nutr Diet 2020 Feb;77(1):90-102. [doi: 10.1111/1747-0080.12608] [Medline: 32022999]
3. Seemer J, Kiesswetter E, Fleckenstein-Sußmann D, et al. Effects of an individualised nutritional intervention to tackle malnutrition in nursing homes: a pre-post study. Eur Geriatr Med 2022 Jun;13(3):741-752. [doi: 10.1007/s41999-021-00597-y] [Medline: 34854062]
4. Lindner-Rabl S, Wagner V, Matijevic A, et al. Clinical interventions to improve nutritional care in older adults and patients in primary healthcare - a scoping review of current practices of health care practitioners. Clin Interv Aging 2022;17:1-13. [doi: 10.2147/CIA.S343307] [Medline: 35023909]
5. Mikkelsen S, Geisler L, Holst M. Healthcare professionals' experiences with practice for managing disease-related malnutrition in general practice and proposals for improvement: a qualitative study. Scand J Caring Sci 2022 Sep;36(3):717-729. [doi: 10.1111/scs.13033] [Medline: 34541700]
6. Ding Y, Lu X, Xie Z, Jiang T, Song C, Wang Z. Evaluation of a novel WeChat applet for image-based dietary assessment among pregnant women in China. Nutrients 2021 Sep 10;13(9):3158. [doi: 10.3390/nu13093158] [Medline: 34579035]
7. Wu Y, Wang X, Gao F, Liao J, Zeng J, Fan L. Mobile nutrition and health management platform for perioperative recovery: an interdisciplinary research achievement using WeChat applet. Front Med (Lausanne) 2023 May;10:1201866. [doi: 10.3389/fmed.2023.1201866] [Medline: 37293309]
8. Huo J, Wu X, Gu C, et al. Using a WeChat mini-program-based lactation consultant intervention to increase the consumption of mother's own milk by preterm infants in the neonatal intensive care unit: a study protocol for a cluster randomized controlled trial. Trials 2021 Nov 24;22(1):834. [doi: 10.1186/s13063-021-05731-6] [Medline: 34819117]
9. LeBeau K, Huey LG, Hart M. Assessing the quality of mobile apps used by occupational therapists: evaluation using the user version of the Mobile Application Rating Scale. JMIR Mhealth Uhealth 2019 May 1;7(5):e13019. [doi: 10.2196/13019] [Medline: 31066712]
10. Li Y, Ding J, Wang Y, Tang C, Zhang P. Nutrition-related mobile apps in the China app store: assessment of functionality and quality. JMIR Mhealth Uhealth 2019 Jul 30;7(7):e13261. [doi: 10.2196/13261] [Medline: 31364606]
11. Adam A, Hellig JC, Perera M, Bolton D, Lawrentschuk N. 'Prostate Cancer Risk Calculator' mobile applications (Apps): a systematic review and scoring using the validated user version of the Mobile Application Rating Scale (uMARS). World J Urol 2018 Apr;36(4):565-573. [doi: 10.1007/s00345-017-2150-1] [Medline: 29222595]
12. Fijačko N, Gosak L, Cilar L, et al. The effects of gamification and oral self-care on oral hygiene in children: systematic search in app stores and evaluation of apps. JMIR Mhealth Uhealth 2020 Jul 8;8(7):e16365. [doi: 10.2196/16365] [Medline: 32673235]
13. Shinohara Y, Yamamoto K, Ito M, et al. Development and validation of the Japanese version of the uMARS (user version of the mobile app rating system). Int J Med Inform 2022 Sep;165:104809. [doi: 10.1016/j.ijmedinf.2022.104809] [Medline: 35728358]
14. Martin-Payo R, Carrasco-Santos S, Cuesta M, Stoyan S, Gonzalez-Mendez X, Fernandez-Alvarez MDM. Spanish adaptation and validation of the User Version of the Mobile Application Rating Scale (uMARS). J Am Med Inform Assoc 2021 Nov 25;28(12):2681-2686. [doi: 10.1093/jamia/ocab216] [Medline: 34613400]
15. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72. [doi: 10.2196/mhealth.5849] [Medline: 27287964]
16. Ko S, Lee J, An D, Woo H. Menstrual tracking mobile app review by consumers and health care providers: quality evaluations study. JMIR Mhealth Uhealth 2023 Mar 1;11:e40921. [doi: 10.2196/40921] [Medline: 36857125]
17. Shetty VB, Soon WHK, Roberts AG, et al. A novel mobile health app to educate and empower young people with type 1 diabetes to exercise safely: prospective single-arm mixed methods pilot study. JMIR Diabetes 2021 Oct 14;6(4):e29739. [doi: 10.2196/29739] [Medline: 34647896]
18. Yu H, Tan L, Zhu T, Deng X. A WeChat applet-based national remote emergency system for malignant hyperthermia in China: a usability study. BMC Med Inform Decis Mak 2023 Sep 5;23(1):175. [doi: 10.1186/s12911-023-02275-4] [Medline: 37670310]
19. Chen D, Shao J, Zhang H, et al. Development of an individualized WeChat mini program-based intervention to increase adherence to dietary recommendations applying the behaviour change wheel among individuals with metabolic syndrome. Ann Med 2023;55(2):2267587. [doi: 10.1080/07853890.2023.2267587] [Medline: 37898907]
20. Feng Y, Zhao Y, Mao L, et al. The effectiveness of an eHealth family-based intervention program in patients with uncontrolled type 2 diabetes mellitus (T2DM) in the community via WeChat: randomized controlled trial. JMIR Mhealth Uhealth 2023 Mar 20;11:e40420. [doi: 10.2196/40420] [Medline: 36939825]
21. Liu H, Feng J, Shi Z, et al. Effects of a novel applet-based personalized dietary intervention on dietary intakes: a randomized controlled trial in a real-world scenario. Nutrients 2024 Feb 19;16(4):565. [doi: 10.3390/nu16040565] [Medline: 38398889]

22. Haas R, Aşan H, Doğan O, Michalek CR, Karaca Akkan Ö, Bulut ZA. Designing and implementing the MySusCof app-a mobile app to support food waste reduction. Foods 2022 Jul 26;11(15):2222. [doi: 10.3390/foods11152222] [Medline: 35892807]
23. Tu W, Yan S, Yin T, et al. Mobile-based program improves healthy eating of ulcerative colitis patients: a pilot study. Dig Health 2023;9:20552076231205741. [doi: 10.1177/20552076231205741] [Medline: 37829613]
24. Lambrecht A, Vuillerme N, Raab C, et al. Quality of a supporting mobile app for rheumatic patients: patient-based assessment using the User Version of the Mobile Application Scale (uMARS). Front Med (Lausanne) 2021;8:715345. [doi: 10.3389/fmed.2021.715345] [Medline: 34368202]
25. Larson D, Henning J, Burgermaster M. Smartphone applications (apps) for nutrition education: a qualitative analysis of outpatient dietitian perspectives. J Nutr Educ Behav 2023 Aug;55(8):596-603. [doi: 10.1016/j.jneb.2023.05.247] [Medline: 37354197]
26. Daniels K, Rathore FA, Bonnechère B. Mobile health: is your next rehabilitation's specialist in your pocket? J Pak Med Assoc 2024 Mar;74(3):599-601. [doi: 10.47391/JPMA.24-21] [Medline: 38591311]
27. Wang X, Zeng H, Li L, et al. Personalized nutrition intervention improves nutritional status and quality of life of colorectal cancer survivors in the community: a randomized controlled trial. Nutrition 2022;103-104:111835. [doi: 10.1016/j.nut.2022.111835] [Medline: 36183482]
28. Ding B, Gou B, Guan H, Wang J, Bi Y, Hong Z. WeChat-assisted dietary and exercise intervention for prevention of gestational diabetes mellitus in overweight/obese pregnant women: a two-arm randomized clinical trial. Arch Gynecol Obstet 2021 Sep;304(3):609-618. [doi: 10.1007/s00404-021-05984-1] [Medline: 33570656]
29. Lull C, von Ahnen JA, Gross G, et al. German mobile apps for patients with psoriasis: systematic search and evaluation. JMIR Mhealth Uhealth 2022 May 26;10(5):e34017. [doi: 10.2196/34017] [Medline: 35617014]
30. Bardus M, Ali A, Demachkieh F, Hamadeh G. Assessing the quality of mobile phone apps for weight management: user-centered study with employees from a Lebanese university. JMIR Mhealth Uhealth 2019 Jan 23;7(1):e9836. [doi: 10.2196/mhealth.9836] [Medline: 30672742]
31. Salas-Groves E, Galyean S, Alcorn M, Childress A. Behavior change effectiveness using nutrition apps in people with chronic diseases: scoping review. JMIR Mhealth Uhealth 2023 Jan 13;11:e41235. [doi: 10.2196/41235] [Medline: 36637888]
32. Schaafsma HN, Jantzi HA, Seabrook JA, et al. The impact of smartphone app-based interventions on adolescents' dietary intake: a systematic review and evaluation of equity factor reporting in intervention studies. Nutr Rev 2024 Mar 11;82(4):467-486. [doi: 10.1093/nutrit/nuad058] [Medline: 37330675]
33. Vietzke J, Schenk L, Baer NR. Middle-aged and older adults' acceptance of mobile nutrition and fitness tools: a qualitative typology. Digit Health 2023;9:20552076231163788. [doi: 10.1177/20552076231163788] [Medline: 36937695]

Abbreviations

AI: artificial intelligence
MARS: Mobile Application Rating Scale
uMARS: User Version of the Mobile Application Rating Scale

Edited by A Kushniruk, E Borycki; submitted 18.01.24; peer-reviewed by A Hassan, B Lawford, F Carrouel; revised version received 17.06.24; accepted 03.07.24; published 12.09.24.

Please cite as:
Sun H, Wu Y, Sun J, Zhou W, Xu Q, Hu D
Nutrition Management Miniprograms in WeChat: Evaluation of Functionality and Quality
JMIR Hum Factors 2024;11:e56486
URL: https://humanfactors.jmir.org/2024/1/e56486
doi:10.2196/56486

Original Paper

Assessing the Usability and Feasibility of Digital Assistant Tools for Direct Support Professionals: Participatory Design and Pilot-Testing

Patrice D Tremoulet[1], PhD; Andrea F Lobo[2], PhD; Christina A Simmons[1], PhD; Ganesh Baliga[2], PhD; Matthew Brady[2]

[1]Department of Psychology, Rowan University, Glassboro, NJ, United States

[2]Department of Computer Science, Rowan University, Glassboro, NJ, United States

Corresponding Author:
Patrice D Tremoulet, PhD
Department of Psychology
Rowan University
201 Mullica Hill Rd
Robinson Hall Room 115K
Glassboro, NJ, 08028
United States
Phone: 1 8562564500 ext 53777
Email: tremoulet@rowan.edu

Abstract

Background: The United States is experiencing a direct support professional (DSP) crisis, with demand far exceeding supply. Although generating documentation is a critical responsibility, it is one of the most wearisome aspects of DSPs' jobs. Technology that enables DSPs to log informal time-stamped notes throughout their shift could help reduce the burden of end-of-shift documentation and increase job satisfaction, which in turn could improve the quality of life of the individuals with intellectual and developmental disabilities (IDDs) whom DSPs support. However, DSPs, with varied ages, levels of education, and comfort using technology, are not likely to adopt tools that detract from caregiving responsibilities or increase workload; therefore, technological tools for them must be relatively simple, extremely intuitive, and provide highly valued capabilities.

Objective: This paper describes the development and pilot-testing of a digital assistant tool (DAT) that enables DSPs to create informal notes throughout their shifts and use these notes to facilitate end-of-shift documentation. The purpose of the pilot study was to assess the usability and feasibility of the DAT.

Methods: The research team applied an established user-centered participatory design process to design, develop, and test the DAT prototypes between May 2020 and April 2023. Pilot-testing entailed having 14 DSPs who support adults with IDDs use the first full implementation of the DAT prototypes during 2 or 3 successive work shifts and fill out demographic and usability questionnaires.

Results: Participants used the DAT prototypes to create notes and help generate end-of-shift reports. The System Usability Scale score of 81.79 indicates that they found the prototypes easy to use. Survey responses imply that using the DAT made it easier for participants to produce required documentation and suggest that they would adopt the DAT if this tool were available for daily use.

Conclusions: Simple technologies such as the DAT prototypes, which enable DSPs to use mobile devices to log time-stamped notes throughout their shift with minimal effort and use the notes to help write reports, have the potential to both reduce the burden associated with producing documentation and enhance the quality (level of detail and accuracy) of this documentation. This could help to increase job satisfaction and reduce turnover in DSPs, both of which would help improve the quality of life of the individuals with IDDs whom they support. The pilot test results indicate that DSPs found the DAT easy to use. Next steps include (1) producing more robust versions of the DAT with additional capabilities, such as storing data locally on mobile devices when Wi-Fi is not available; and (2) eliciting input from agency directors, families, and others who use data about adults with IDDs to help care for them to ensure that data produced by DSPs are relevant and useful.

(JMIR Hum Factors 2024;11:e51612) doi:10.2196/51612

KEYWORDS

technology prototype; data collection; documentation; direct support professionals; intellectual and developmental disabilities; pilot test; mobile phone

Introduction

Background

In 2019, more than 2 million adults with intellectual and developmental disabilities (IDDs) were living in the United States [1]. Many rely upon direct support professionals (DSPs) for assistance with activities of daily living, such as hygiene, dressing, taking medications properly, eating, accessing and navigating stores, learning vocational skills, participating in therapeutic activities, and socializing [2]. It is widely understood that the quality of life of adults with IDDs is significantly impacted by the quality of the support they receive from DSPs [3-11]. Unfortunately, there has been a shortage in the DSP workforce for more than a decade [7,12]. This shortage and a DSP turnover rate of 44.8% in 2017 led the President's Committee for People with Intellectual Disabilities to declare a crisis in the direct support workforce [13]. The DSP shortage and high turnover rate, each of which is associated with reduced quality of life for those served by DSPs [3,7], have both significantly increased since the COVID-19 pandemic [14]. Therefore, it is not surprising that the rate of COVID-19 infection among adults with IDDs was disproportionally high and that their quality of life decreased between 2019 and 2020 [15].

DSPs have indicated that generating required documentation is one of the least rewarding and most onerous aspects of their jobs [16]. Several factors make this task challenging, including difficulty recalling details of work performed many hours ago; fatigue; being rushed because employers require them to check out on a time clock at specific times to avoid overtime pay; fear of being accused of copying and pasting content from previous shifts; frequent interruptions by clients or other staff members; and, for some, challenges in writing in a nonnative language [16]. High turnover increases the communication demands on DSPs, including the need to generate detailed documentation to help bring new DSPs up to speed on their clients' needs. However, overworked DSPs who prioritize clients' medical and behavioral health needs may struggle to find time to document and share all relevant data during shift changes, unintentionally leaving their clients vulnerable to medical errors and inadequate support [14].

Technology enabling in-the-moment data collection can increase the efficiency of direct support staff and raise the quality of the data they produce, both of which can help improve care for their clients [17-19]. However, most data collection and reporting tools used to capture data about individuals with IDDs are targeted toward providers who work one-on-one with children [16] (eg, paraprofessionals or behavior technicians). These tools are not appropriate for DSPs who support multiple adults with IDDs during their shifts. In addition, the responsibilities of DSPs are much wider in scope than those of paraprofessionals and behavior technicians, who are typically not responsible for tasks such as administering medication and helping prepare meals. DSPs' responsibilities are closer to those of home health workers who provide older adults or other adults with medical assistance and help with activities of daily living. Many high-income countries are already facing shortages of in-home health workers, while demand for them, as well as for DSPs, is expected to grow in many countries [20-22].

Recognizing that DSPs and the adults with IDDs who depend upon them could both potentially benefit from improved data collection and documentation, the research team used an established user-centered design methodology called Interaction Design and Engineering for Advanced Systems (IDEAS) [23] to design and pilot-test a suite of technology components called a digital assistant tool (DATs) to support data collection and reporting. This methodology relies upon frequent input and feedback from the target users, in this case, DSPs, to ensure that novel technology solutions will be useful, usable, and accepted by the DSPs [23].

Objectives

The primary purpose of the pilot study described here was to assess the usability of the DAT prototypes. The main objectives of the pilot study were as follows:

1. Identify design inconsistencies and usability problem areas of the DAT.
2. Observe representative users interacting with the DAT prototypes to help assess whether this technology could be effective, efficient, and well received by DSPs.
3. Establish baseline performance and user satisfaction levels in anticipation of more widespread testing of improved versions of the DAT.

Methods

Study Design

Overview

This project applied the IDEAS methodology, a user-centered participatory design approach that relies upon frequent input and feedback from target users to ensure that novel technology solutions will be useful, usable, and accepted by the users [23]. There are 6 steps in the IDEAS process: needs analysis, requirements generation, design and engineering, interface review, implementation, and evaluation. The first phase of this project included the first 2 steps; the second phase included the third, fourth, and fifth steps; and the third phase, which is the focus of this paper, comprises the last step.

Phase 1: Needs Analysis and Requirements Generation

To conceptualize the potential use of technology by DSPs, the research team sought to understand their perspectives on current data collection and documentation techniques and their ideas on how digital technology could be applied to support their work. The results of this exploratory descriptive research, which included focus groups, ethnographic observations, and a survey, are described elsewhere [19]. Using the findings of this formative research, the team developed a list of design principles for our first set of prototypes, shown in Textbox 1.

Textbox 1. Summary of formative research results and the corresponding design principles for the initial digital assistant tool (DAT) prototypes.

- Direct support professionals (DSPs) need to track and remember large amounts of information about multiple clients.
  - The DAT should automatically store notes, it should allow users to associate notes with specific clients, and it should provide access to notes on demand.
- DSPs must continuously monitor clients for safety while tracking behaviors.
  - The DAT should enable users to quickly and easily create notes while attending to clients, it should run on mobile and wearable devices, and it should not require a lengthy authentication process.
- DSPs must not allow clients to recognize when DSPs are creating notes about them.
  - The DAT should enable DSPs to create notes unobtrusively.
- DSPs are comfortable using smartphones but desire simple, easy-to-use data logging capabilities.
  - The DAT should prioritize usability—its user interfaces should be very simple, DAT log-ins should not time out during a shift, and the DAT should provide readable language.
- DSPs do not want employers to be able to access their notes.
  - The DAT should feature high levels of security and privacy and allow only DSPs to see their notes.
- DSPs must provide either chronological or categorically organized reports.
  - The DAT should time-stamp notes, it should allow users to associate notes with a topic or category, and it should allow users to sort and filter notes.
- DSPs must not copy text from prior days' reports into the current report.
  - The DAT should allow users to copy time-stamped notes into a clipboard or Word document to use to help write reports.

Phase 2: Design and Engineering, Interface Review, and Implementation

Once the research team had decided on the initial set of capabilities for the DAT, they designed a suite of technology prototypes that (1) enable quick and easy in-the-moment data collection; and (2) allow DSPs to access a private, secure web portal to review, sort, filter, and organize their notes to facilitate end-of-shift documentation. This suite includes 4 components: a mobile app that currently runs on Android smartphones; a private, secure web application that allows DSPs to access, review, and organize notes that they created with the mobile application; a cloud-based center that houses the data; and an administrative website for creating and managing user accounts. As these components were being architected, researchers shared user interface design concepts for the mobile app and the web portal with DSPs to obtain their feedback. The initial wireframes for the mobile app included multiple screens that would allow users not only to create notes but also to review and edit them (Multimedia Appendix 1). After 4 iterations, the team settled upon a very simple single-screen note creation design for the mobile app, also called the Note Creation App; and a spreadsheet-like view of saved notes for the web application, also referred to as the Note Review App (both of which were implemented; refer to Figures 1-3).

Figure 1. The mobile app, used to create informal notes. This screen shows the touchscreen keyboard used to type new notes.

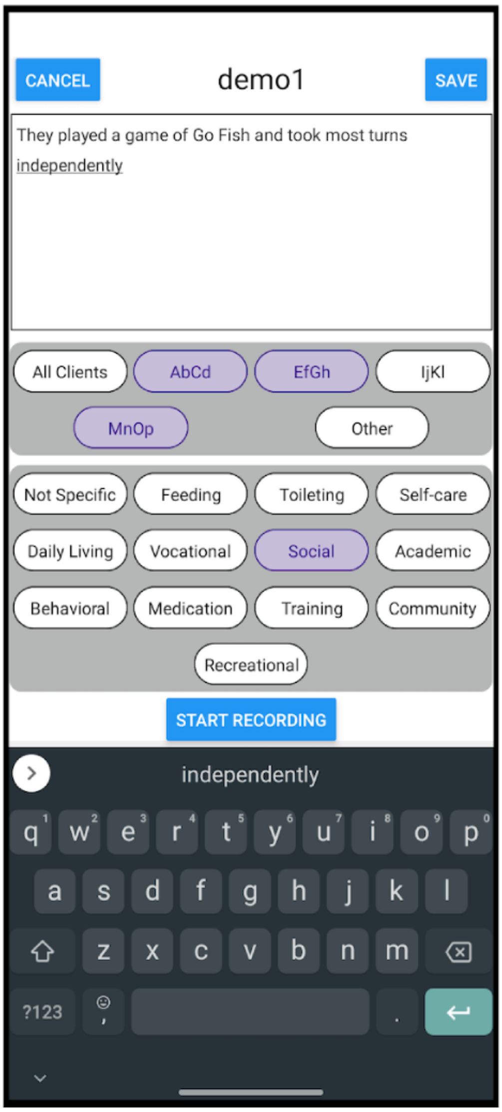
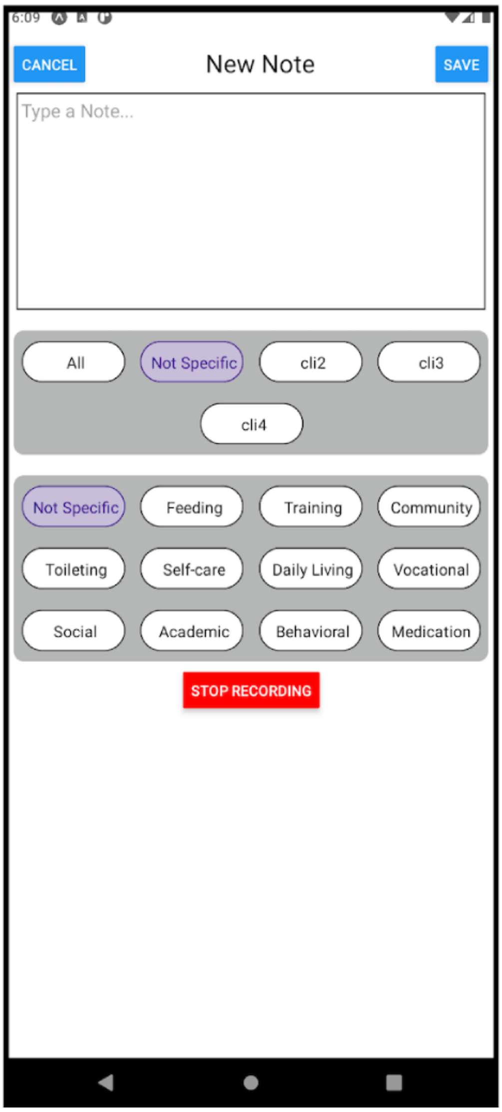


Figure 2. The mobile app, used to create informal notes. This screen shows the screen used for voice-based note creation.



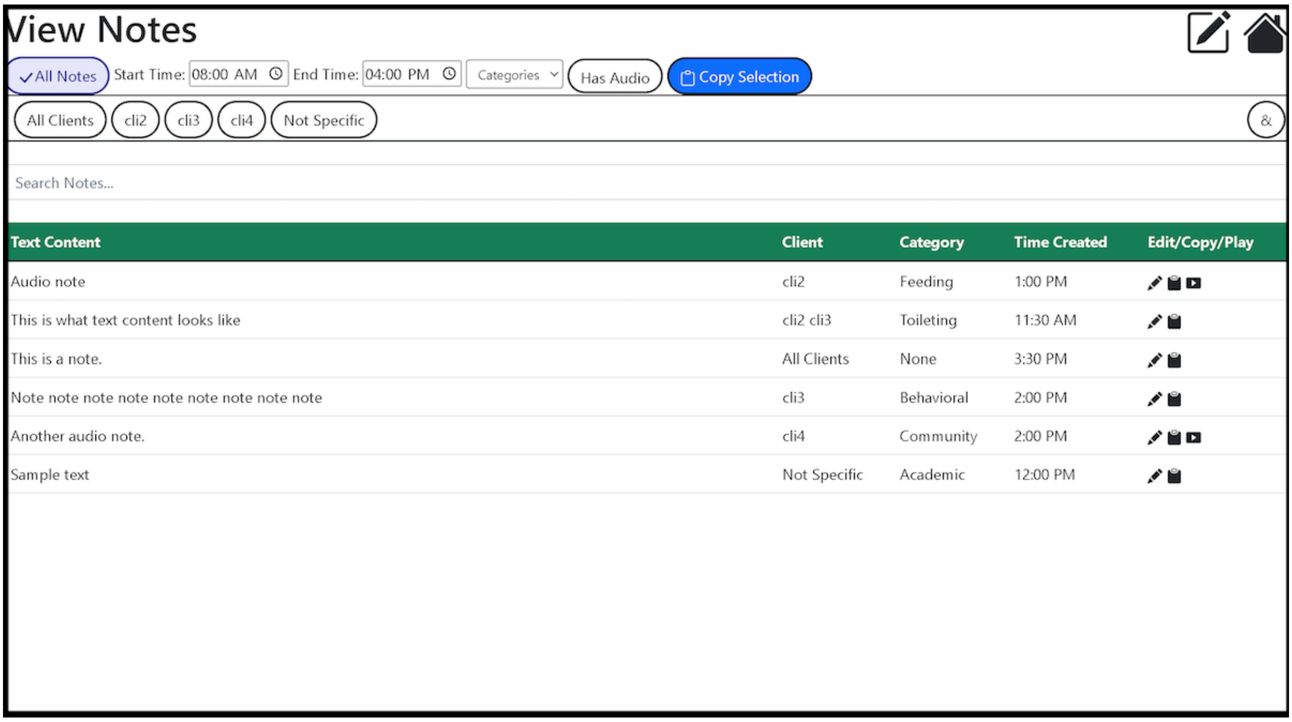


Figure 3. The web application, used to organize notes and copy data into report generation software. cli2: client 2; cli3: client 3; cli4: client 4.

Phase 3: Evaluation

Once the initial prototypes were implemented, engineering and psychology undergraduate students tested them in laboratory settings; the engineering students focused on performance testing, and the psychology students focused on providing usability feedback. After all known bugs had been fixed, and usability improvements had been implemented, the team began scheduling pilot tests.

Participants and Recruitment

Overview

The pilot study was intended to determine *proof of concept*; therefore, the number of participants was not determined based on traditional power analysis calculations. Guidance on how many participants to include in pilot studies varies, with some recommending between 10 and 30 [24,25] and others suggesting 12 [26,27]. Meanwhile, the System Usability Scale (SUS), the instrument used to measure baseline usability, requires at least 8 to 10 participants to produce reliable results [28]. The research team worked with 3 DSP service provider partners to recruit 16 participants with the goal of having at least 12 (75%) of them test the prototypes during multiple shifts.

The service providers' senior management members notified their staff via email and during staff meetings that members of the research team would be bringing technology tools designed to be used by DSPs on specific dates. All staff who were working to support adults with IDDs in the designated program for 3 consecutive shifts during the research team visits were eligible to serve as pilot test participants. Due to variability in testing sites, including staff characteristics, the level of support needed by clients served by the staff, and the timing of the staff's work shifts, demographic data and results are reported separately for each test site.

Site 1 (Day Program)

The first pilot-testing site was a day program for adults with IDDS where DSPs and clients were each assigned to 1 of 5 different rooms. Of the 10 DSPs who worked in this setting on the first day of testing, 6 (60%) were selected to serve as pilot testers; 2 (20%) were excluded because they were not scheduled to work 3 shifts in a row in the same setting, and 2 (20%) were willing but unable to participate because there were only 6 smartphones with the DAT mobile app installed available; thus, the first 6 DSPs who volunteered were enrolled. All 6 participants were native English speakers; 5 (83%) were women. Of the 6 participants, 2 (33%) were aged between 18 and 24 years, 1 (17%) was aged between 25 and 34 years, 1 (17%) was aged between 35 and 44 years, and 2 (33%) were aged between 45 and 54 years. All identified as Black or African American. Of the 6 participants, 1 (17%) had a high school degree, while the remaining 5 (83%) had taken some college classes but did not have college degrees. Experience working with adults with IDDs ranged from 1.5 to >20 years, with an average of 8 years, 9.6 months. The amount of time they had held their positions ranged from 4 months to 3 years. Due to last-minute schedule changes, of the 6 participants, 2 (33%) used the DAT prototypes for 2 consecutive shifts, whereas the other 4 (67%) used them for 3 consecutive shifts.

Site 2 (Private Home)

The second test site was a private home that housed a single adult with a high level of support needs. Of the 12 different staff members who worked in this home, 4 (33%) were eligible to participate, based on their schedules, and all volunteered to serve as pilot testers. All 4 participants from this site were women and native English speakers; 1 (25%) was aged between 18 and 24 years, and the remaining 3 (75%) were aged between 25 and 34 years. Of the 4 participants, 1 (25%) identified as Black or African American, 1 (25%) as Hispanic, and 2 (50%)

as White. Of the 4 participants, 1 (25%) had a high school degree, 2 (50%) had some college but no degree, and 1 (25%) had a college degree. Their experience working with individuals with IDDs ranged from 7 months to 7 years, with an average of 3 years, 11.4 months. The amount of time they had been working in the private home ranged from 3 months to 1.5 years.

Site 3 (Provider-Managed Group Homes)

The third test site included 3 group homes, all located in the same county, that housed 3 to 4 adults with IDDs. Each of the 2 staff members in each of the 3 homes volunteered to serve as pilot testers. Of the 6 volunteers, 2 (33%) had work schedule changes that prevented them from using the DAT prototypes after their first shift and were removed from the study; the remaining 4 (67%) volunteers were included in the study. All 4 study participants were women and native English speakers. Of the 4 participants, 3 (75%) identified as Black Hispanic, and 1 as *other*; 1 (25%) had a college degree, and 3 (75%) had high school degrees and no college experience. Of the 4 participants, 1 (25%) was aged between 18 and 24 years, 1 (25%) was aged between 25 and 34 years, and 2 (50%) were aged between 45 and 54 years. The amount of experience they had working with adults with IDDs ranged from 1 to 22 years, with an average of 11 years, 9 months. The amount of time they had held their positions ranged from 1 to 22 years, with an average of 7 years, 4 months. Due to last-minute schedule changes, of the 4 participants, 2 (50%) used the DAT prototypes for 2 consecutive shifts, whereas 2 (50%) used them for 3 consecutive shifts.

Materials

Each participant was provided with an Android smartphone (Motorola Moto G Power 2021 running Android version 11 RZBS31.Q2-14327-25) that had the DAT mobile app preinstalled. For testing at sites 2 and 3, each smartphone also had shortcuts to all study surveys (demographic survey, post–shift 1 survey, and post–final shift survey) located on its main screen. Shortcuts were not provided on the smartphones during testing at site 1 because the research team was present to provide links in person at that site. All participants used employer-provided laptops to access the DAT web application at the end of their shifts.

Procedure

On the first day of testing, the principal investigator (PI) obtained informed consent from each participant. She explained that they would be awarded gift cards at the end of the multiday test trial that would be credited with US $75 per shift for their first 2 shifts and US $100 for their third shift. Next, she asked participants to use a computer to complete a short demographic questionnaire, administered through Qualtrics (Qualtrics International Inc; Multimedia Appendix 2).

While participants were filling out the demographic survey, the research team used the administrative website to create user accounts and location-based shifts, which included the initials of the clients whom each participant would be supporting and relevant data categories for the DSPs who worked in this setting (eg, feeding, toileting, behavior, medication, and social skills). These categories were identified by asking staff supervisors to select from a list of possible categories; they were also invited to add items not on the list.

After the participant finished the demographic survey, the researcher showed them the web application, which was populated with sample data previously entered by the research team. The researcher explained to the participant that this application would be used at the end of their shift to support writing the required reports. Subsequently, the researcher provided each participant with an Android smartphone, demonstrated how to use the mobile app, and invited the participants to try creating an audio note as well as a text-based note.

Participants were also invited to review, sort, and filter their sample notes using the web application during the training session. Once participants indicated that they knew how to use both the mobile app and the web application, they were given an opportunity to ask questions and informed that they could request help or address questions to the PI at any point during their shifts. They were then directed to use the smartphone that the research team had provided to log audio and text-based notes about their work while otherwise performing their job as usual throughout their shift.

Site 1 Testing

During pilot-testing in the day program, 2 to 3 members of the research team stayed on site during all 3 pilot-testing shifts. One research team member monitored the use of the mobile app through the administrative website, and 2 other members intermittently visited the rooms where the DSPs who were testing the DAT were assigned to work to observe or answer questions. In 1 room, the ratio of DSPs to clients was 1:3 on day 1 and 1:4 on days 2 and 3. In the second room, the ratio of DSPs to clients was 2:5 on day 1 and 2:4 on days 2 and 3. In the third room, the ratio was 2:3 on days 1 and 2 and 3:3 on day 3 (the third DSP in this room was in training and did not have access to the DAT). Finally, the fourth room, which was only included on day 3, had a ratio of 3:8, and only 1 DSP in this room used the DAT.

Throughout all their shifts, the participants used the mobile app to create short notes. They had the option to use voice or a touchscreen keyboard to enter text, and they could also use touchscreen buttons to associate notes with ≥1 clients or to indicate a specific category (eg, feeding, self-care, and medication).

Toward the end of the first day of pilot-testing, when participants were ready to start working on their required reports, a researcher helped them log in to the web application so that they could review and organize the notes that they had created during the shift. If needed, the researcher reviewed functionality, such as sorting, filtering, and copying 1 or multiple notes to a clipboard. The team then observed how the DSPs used the web application to help them create end-of-shift reports. Most of the DSPs (5/6, 83%) decided to paste their notes into either the clipboard or a text file and use 1 of these to draft the text portion of their end-of-shift report, rather than copying note content directly into the application that they use to submit their reports. Once the participants had finished their reports, they were directed to use either the laptop used to create reports or the smartphone to fill out a web-based post–shift 1 survey (Multimedia Appendix 3). Participants were asked to rate 7 items, using a Likert scale ranging from 1 to 5, to indicate their level of agreement. Of the 7 items, 3 (43%) described their experience using the mobile app to log notes, their experience using the web application to review and organize the notes, and their opinions about the quality of the behavioral data and end-of-shift reports that they had provided that day. The items in this survey, which were also included in the final survey, were developed by the authors for this study and were not tested for validity or reliability.

At the end of days 2 and 3, the research team members were available to assist the participants, but all of them were able to access the web application and use it to help them with their reports independently. At the end of day 3, the participants were directed to fill out the final survey (Multimedia Appendix 4). The survey began with 17 items that were rated on a Likert scale ranging from 1 to 5 showing level of agreement. The first 7 items, which were about data collected and reports created, were the same as the first 7 statements rated after the first shift. The last 10 items were those included on the SUS survey, a well-established survey for assessing ease of use and user satisfaction [28,29]. Eleven open-ended questions followed the 17 statements. They asked participants to share their overall impressions of the mobile app and the web application, what they thought about the layout of the mobile app and the web application, what concerns they had about using the DAT, what it was like writing end-of-shift reports while using the DAT, and what additional feedback they wanted to share. These questions were designed to obtain feedback on existing functionality and elicit suggestions for changes or additions.

Once the final survey was submitted, the research team collected the smartphones, and the PI provided gift cards.

Site 2 Testing

As the private home housed a single adult, testing at this site was sequential, with only 1 (25%) of the 4 participants using the DAT prototypes on any given shift. In addition, only the PI visited this site when the DSPs who worked there participated in the pilot test. The PI met with each participant in the home approximately 1 hour before their shift started and obtained informed consent, provided the participant with a smartphone running the DAT mobile app, and taught them how to use the DAT mobile app and the DAT web application. Once training was completed, the PI placed shortcuts to the post–shift 1 survey as well as the final survey on the main screen of the smartphone and emailed the participant a private, secure link to the web application site where they could review and organize their notes. Finally, the PI answered any questions that the participant had, made sure the participant had a power cord and knew how to charge the smartphone with the mobile app, provided a phone number and email address to use in case questions or technical difficulties arose during their shifts, and departed. Although the research team members were not on site, they were able to monitor the use of the DAT mobile app and the DAT web application using the DAT administrative portal. All participants used the smartphone to create notes and accessed the web portal at the end of their shifts without support from the research team. Toward the end of each participant's first shift, the PI sent a reminder email that contained a link to the post–shift 1 survey (which was also accessible via a link on the smartphone). Toward the end of the participants' third shift, the PI sent an email to confirm a meeting time to deliver gift cards and pick up the smartphones; the email also contained a link to the final survey.

Site 3 Testing

Two members of the research team visited each of the 3 group homes during the first day of testing at site 3 to obtain informed consent, hand out the smartphones with the DAT mobile app, and conduct training. The research team members had placed shortcuts to the demographic, post–shift 1, and final surveys on the home screens of the smartphones in advance. After training was complete, the research team members helped participants to create a bookmark to conveniently access the web application on the computers that the participants used to produce their end-of-shift documentation. The PI also emailed links to the surveys and a secure link to access the web application so that the participants could easily access the surveys and their private website for reviewing and organizing their notes from their computers. Next, the researchers gave participants an email address and a phone number that could be used to reach the PI and encouraged them to use these if they had questions or concerns while using the DAT prototypes. The researchers then departed. During the rest of this shift, the research team members monitored the use of the DAT mobile app and the DAT web application via the administrative website. Toward the end of the first shift, the PI emailed participants reminders to fill out the post–shift 1 survey; and toward the end of the third shift, the PI emailed reminders to fill out the final survey. After the third shift had ended for all participants, the PI visited each of the group homes to pick up the smartphones and chargers and distribute gift cards.

Data Preparation and Analysis

Survey data were extracted from Qualtrics. For each site, descriptive statistics for demographic data were summarized; and the frequencies of the Likert scale responses, except for the 10 SUS items, were computed. The responses to the SUS items from all 14 participants were used to compute a single SUS score, using established computations [29]. The research team members also analyzed open-ended responses from all participants thematically, using inductive open coding [30]. Following the 6-step method formulated by Braun and Clarke [30] as explained by Maguire and Delahunt [31], 2 members of the research team categorized different response types for each question and used these to establish codes; next, these researchers and the PI collaboratively identified high-level themes that ran across multiple questions. Finally, the research assistants independently coded responses to all open-ended questions using the final code set.

In addition, in keeping with the exploratory nature of this work, the research team members analyzed the notes that were collected using the mobile app to understand how DSPs used this technology. In particular, the team reviewed the number of text-based and audio notes that each DSP produced during each of their shifts, the number of words in each type of note, and the number of notes that were started but canceled (not saved) by each DSP during each shift.

Ethical Considerations

This research complies with the American Psychological Association code of ethics and was approved by the Rowan University Institutional Review Board (PRO-2020-001085). Informed consent was obtained from each participant on the first day of testing. Regarding compensation, at the end of the multiday test trial, participants were awarded gift cards that were credited with US $75 per shift for their first 2 shifts and US $100 for their third shift.

Results

Note Logs

Data logs from 34 DSP shifts were imported into Excel (Microsoft Corp) for analysis. Across these shifts, a total of 373 notes were saved. Another 41 notes were started, but they were canceled (not saved). Table 1 shows, for each site, the medians and IQRs of all saved notes per shift, audio notes saved per shift, text-based notes saved per shift, canceled notes per shift, notes created during a DSP's first pilot-testing shift, notes created during a DSP's second pilot-testing shift, notes created during a DSP's third pilot-testing shift, words per audio note, words per text note, and words per note.

Table 1. Types of direct support professional notes produced at each site during the pilot-testing.

Computation	Site 1, median (IQR)	Site 2, median (IQR)	Site 3, median (IQR)
Audio notes per shift	2 (1-9)	0 (0)	3 (2-5)
Text-based notes per shift	8 (4-14)	10 (7.5-14)	7 (4-12)
Total notes per shift	10 (5-12)	10 (7.5-14)	10 (2-16)
Canceled notes per shift	0 (0-1.5)	0 (0-3.5)	0 (0-1.5)
Notes created during first shift of pilot-testing	6 (4-12)	7.5 (6-9)	8.5 (5-12)
Notes created during second shift of pilot-testing	10 (5-12)	10 (5.5-14.5)	6 (2-10)
Notes created during third shift of pilot-testing	6 (8.5-12.5)	12.5 (6-16)	9.5 (7-12)
Words per audio note	9 (6-17)	N/A[a]	8.5 (8-14)
Words per text note	12 (6-22.5)	6.5 (3-9)	11 (8-15)
Words per note	10 (6-22)	6.5 (3-9)	10.5 (8-15)

[a]N/A: not applicable.

Post–Shift 1 Survey Ratings Questions

The post–shift 1 survey began with 7 statements that were rated on a scale ranging from 1 to 5 to show level of agreement (1=strongly disagree, 2=disagree, 3=neutral, 4=agree, and 5=strongly agree). Table 2 shows these statements as well as the frequencies of each of the ratings given to each statement.

Table 2. Frequencies of the post–shift 1 survey Likert scale responses for each test site. Response scale ranged from 1=strongly disagree to 5=strongly agree.

Statement	Site 1 (n=6), response; n (%)	Site 2 (n=4), response; n (%)	Site 3 (n=4), response; n (%)
I am confident that today's data sheets are accurate	• 4; 1 (17) • 5; 5 (83)	• 5; 4 (100)	• 3; 2 (50) • 4; 1 (25) • 5; 1 (25)
I found it easy to record behavior data for all clients today	• 4; 1 (17) • 5; 5 (83)	• 5; 4 (100)	• 3; 2 (50) • 4; 1 (25) • 5; 1 (25)
I believe today's behavior data will be valuable to others	• 4; 1 (17) • 5; 5 (83)	• 5; 4 (100)	• 3; 3 (75) • 4; 1 (25)
I found it easy to write session notes today	• 5; 6 (100)	• 3; 1 (25) • 5; 3 (75)	• 3; 2 (50) • 5; 2 (50)
I am confident today's session notes contain all necessary information	• 3; 1 (17) • 5; 5 (83)	• 5; 4 (100)	• 2; 1 (25) • 4; 1 (25) • 5; 2 (50)
I am confident today's session notes contain only relevant information	• 4; 1 (17) • 5; 5 (83)	• 5; 4 (100)	• 2; 1 (25) • 4; 2 (50) • 5; 1 (25)
I believe today's session notes will be valuable to others (parents, supervisors, behavior analyst)	• 4; 2 (33) • 5; 4 (67)	• 5; 4 (100)	• 2; 1 (25) • 3; 1 (25) • 4; 1 (25) • 5; 1 (25)

Post–Final Shift Survey Ratings and SUS Ratings

All 14 participants who used the DAT during multiple shifts completed the post–final shift survey. Table 3 shows the frequencies of responses to the first 7 statements.

The SUS score for the 14 participants was 81.79, which is quite high for an initial prototype; this score corresponds to *excellent* usability [32]. This score is also in the range where a product is likely to be recommended by users to other potential users [33].

The final survey also contained 11 open-ended questions: 4 (36%) were about the mobile app, 6 (55%) about the web application, and 1 (9%) asked if any initial concerns that participants had about using the DAT prototypes had been alleviated after using them during multiple shifts.

Table 3. Frequencies of responses to post–final shift survey non–System Usability Scale Likert scale items.

Statement	Site 1 (n=6), response; n (%)	Site 2 (n=4), response; n (%)	Site 3 (n=4), response; n (%)
I am confident that today's data sheets are accurate	• 4; 1 (17) • 5; 5 (83)	• 5; 4 (100)	• 3; 1 (25) • 4; 1 (25) • 5; 2 (50)
I found it easy to record behavior data for all clients today	• 5; 5 (100)	• 5; 4 (100)	• 3; 2 (50) • 4; 1 (25) • 5; 1 (25)
I believe today's behavior data will be valuable to others	• 4; 1 (17) • 5; 5 (83)	• 5; 4 (100)	• 3; 1 (25) • 4; 2 (50) • 5; 1 (25)
I found it easy to write session notes today	• 5; 5 (100)	• 5; 4 (100)	• 3; 2 (50) • 4; 1 (25) • 5; 1 (25)
I am confident today's session notes contain all necessary information	• 4; 1 (17) • 5; 5 (83)	• 5; 4 (100)	• 3; 1 (25) • 4; 1 (25) • 5; 2 (50)
I am confident today's session notes contain only relevant information	• 4; 2 (33) • 5; 4 (67)	• 5; 4 (100)	• 3; 2 (50) • 5; 2 (50)
I believe today's session notes will be valuable to others (parents, supervisors, behavior analyst)	• 4; 1 (17) • 5; 5 (83)	• 5; 4 (100)	• 3; 1 (25) • 4; 2 (50) • 5; 1 (25)

Thematic Analysis of Open-Ended Question Responses

Overview

One-third of the question responses (5/14, 36%) were coded by each of the 2 coders. Intercoder reliability was 92.56%. After coding was complete, the research assistants worked with the PI and identified the following 5 overarching themes in the survey response data.

Theme 1: Using the DAT Was a Positive Experience

Participants reported that the mobile app was easy to use and that it was helpful to be able to create notes in real time. Comments after using it for a single shift included "it's amazing" (Participant 4) and "I would use it on a daily basis" (Participant 12). The participants also provided predominantly positive feedback on the web application, such as "it's great," (Participant 2) "simple and easy," (Participant 7) and "organized and helpful" (Participant 6).

Theme 2: Using the DAT Made It Easier to Create End-of-Shift Reports

Survey responses also revealed that participants found that using the DAT facilitated writing end-of-shift reports. Their comments included "[The DAT] made it easy to keep track of everything throughout the day" (Participant 14) and "made [writing reports] a lot easier for me" (Participant 1). A participant wrote, "I had a great experience [writing reports] and I like [the DAT] very much" (Participant 7).

Theme 3: The DAT Helped Increase the Accuracy of End-of-Shift Reports

Participants indicated that using the DAT enabled them to create more accurate reports. One wrote, "It help [sic] to maintain accurate notes as the day goes along" (Participant 7); another reported that "It gives you the opportunity to keep the notes accurate" (Participant 9); and a third simply commented, "increases accuracy" (Participant 2).

Theme 4: Additional and Improved Features Are Desired

Although participants provided positive responses when asked about their experience using the DAT, they also noted that the tools could provide even greater benefit if they were enhanced. Suggestions included allowing users to access the web application to review and edit notes on their smartphone, improving the transcription accuracy; allowing users to store notes locally when smartphones are not connected to Wi-Fi; offering additional or more specific categories; and allowing users to copy time stamps when copying note text from the web application.

Theme 5: More or Better Training Would Help

The survey responses by 3 (21%) of the 14 participants suggested that they would have liked additional training. One of them noted that "it looks easy but when I'm by myself it's a different story" (Participant 12); the second wrote, "I wish I could get a better understanding" (Participant 11); and the third reported that "it would be easy if I get the hang of it" (Participant 12).

Discussion

Principal Findings

All 14 participants were able to use the mobile app without assistance during all shifts, but their ability to use the web application independently at the end of their first shift varied. All participants were able to use both the mobile app and the web application without help from the research team members during their second and third shifts. This result is consistent with an average SUS score of >80, indicating that the DAT prototypes are very easy to use. The hardest part about using the web application for most of the participants was typing in the URLs to access the private, secure notes review app. In future testing, aliases (shorter active links that take users to the same place as the longer URLs) should be set up to help avoid this problem. In addition, research team members should always help participants create a bookmark during training to make it easy to access the web application, a strategy that was not adopted until testing was conducted at site 3. All participants indicated that they found it easier to write reports at the end of their shift after using the DAT.

At site 1, all 6 participants were able to use the DAT prototypes for at least 2 successive shifts. These participants, whose ages were fairly well distributed across the range of 18 to 54 years, agreed or strongly agreed with 6 (86%) of the 7 survey statements after day 1. For the statement "I am confident today's session notes contain all necessary information," of the 6 participants, 1 (17%) selected *neither agree nor disagree*, and the remaining 5 (83%) all agreed or strongly agreed (Table 2). When given the same set of questions in the final survey, all 6 participants agreed or strongly agreed with all 7 statements (Table 3). This is consistent with the largely positive responses that these participants provided to the open-ended questions in the final survey. All indicated that they enjoyed using the DAT prototypes and would be interested in using future versions.

The site 2 participants were younger, with ages ranging from 18 to 34 years; therefore, they had less experience working with adults with IDDs than site 1 participants. Their responses to the Likert statements also indicated a high level of agreement; after day 1, all 4 participants strongly agreed with 6 (86%) of the 7 statements; 1 (25%) participant selected *neither agree nor disagree* for the statement "I found it easy to write session notes today," and the remaining 3 (75%) participants all strongly agreed with this statement. All 4 site 2 participants responded *strongly agree* to all 7 statements in the final survey. These participants also provided responses to the open-ended questions indicating that they enjoyed using the DAT prototypes and would be interested in using future versions. These 4 participants provided more suggestions for additional features than the participants from the other 2 sites.

Site 3 participants fell into 2 different age ranges. Of the 4 participants, 2 (50%) were aged between 18 and 34 years, and 2 (50%) were aged between 45 and 54 years. Compared to site 1 and site 2 participants, site 3 participants provided a wider range of responses to the 2 surveys. As can be seen in Table 2, after the first shift using the DAT prototypes, a participant disagreed with the statements "I am confident today's session notes contain all necessary information," "I am confident today's session notes contain only relevant information," and "I believe today's session notes will be valuable to others (parents, supervisors, behavior analyst)." Responses to the other 4 statements ranged from *neutral* to *strongly agree*. Responses to the final survey ranged from *neutral* to *strongly agree* for all 7 statements, indicating more variable sentiments among site 3 participants at the end of the pilot test than among site 1 and site 2 participants.

When we grouped the site 3 participants' responses based on their ages, we found that those in the age category of 45 to 54 years (2/4, 50%) generally reported a less positive experience using the DAT prototypes than the other participants (2/4, 50%). In fact, a review of the data logs from these 2 participants revealed that they only used the mobile app at the end of their shifts. It seemed that they used the DAT mostly as a transcription service. As the DAT prototypes used a freely available web-based transcription service which was far from perfect, and time stamps are not helpful when all notes are created at the end of shifts, these DSPs would have been better off using a transcription app on the computers they used to create their end-of-shift reports.

While most of the test participants (12/14, 86%) indicated that they valued the ability to collect data about their work in the moment, several of them (6/14, 43%) noted that their ability to benefit from this capability was reduced because the mobile app prototype requires Wi-Fi to save notes. Even with this limitation, the DSPs were very enthusiastic about the DAT prototypes; 1 participant asked whether she could keep using them after the pilot test, and another asked when they would be available for daily use. Future work includes enabling the mobile app to store data locally so that it can be used to collect data while DSPs and clients are out in the community, which is a key part of some DSPs' client interactions.

The average numbers of notes created per shift and canceled notes per shift was roughly the same for each site. The number of words per note averaged 6.5 at site 2 compared to approximately 10 at sites 1 and 3. To our surprise, audio notes had roughly the same number of words as text-based notes. While the number of words seemed low at first, this is consistent with the intent that note creation be quick, easy, and informal, helping jog memories during report creation. Meanwhile, the average number of notes that DSPs saved during their shifts increased for most DSPs across their shifts. We anticipated that the number of DSPs who created few notes initially would increase in subsequent shifts, and those who created many would create fewer as they learned which types of notes were most helpful. Although 1 DSP's notes fit this pattern, it seems that the other DSPs created more notes as they became more accustomed to using the mobile app during their shift. In a few cases, participants did not include any content; they just created a note with a client and category selected, indicating that just having a time associated with a client and category would be useful when writing reports. In addition, participants indicated that being able to sort notes by category can directly help with writing reports that have specific requirements, such as detailing all instances of medication administration or all food intake during a shift.

Overall, the results from the pilot test are promising, suggesting that DSPs would be willing to use mobile devices to enable in-the-moment data collection, provided that the collected data facilitate efficient generation of required end-of-shift documentation. Feedback from the test participants suggests that technology such as the DAT could help to increase efficiency, effectiveness, and job satisfaction among DSPs.

Comparison With Prior Literature

Many sources, some more than a decade old, have warned about, or described, a shortage of DSPs in the United States [7,34-39]. Factors such as heavy workload, onerous documentation requirements, and burnout contribute to a high turnover rate, which exacerbates stress on DSPs [22,40-49]. Some of this literature points out that technology could potentially help to reduce the time DSPs must spend on their other responsibilities, such as documentation, so that they can spend more time on direct support [7,38].

Technology has been successfully applied to increase documentation efficiency and decrease workload in special education and clinic- and home-based therapy programs for children with autism spectrum disorder [17,18,50,51]. By enabling users to quickly and easily record in-the-moment data, these technologies help improve the quality of documentation they generate [52]. However, there is relatively little work that explores how technology could be applied to support direct support workers, such as DSPs and home health aides (HHAs), who provide care to adults [53,54]. Most of the research that addresses having direct support workers leverage technology is qualitative and focuses on workers who support patients without IDDs [55-57]. One 2022 review paper identified only 1 study that created technology expressly intended to support the work of direct support workers [53]. This study, which also entailed creating a smartphone app to use to help create reports, was conducted in Japan nearly 2 decades ago. The authors who described this effort concluded that enabling direct support workers to use smartphones to create reports saves time and reduces costs [58]. Another 2022 review, which surveyed "the technological landscape" of direct support workers, noted that there is a paucity of evidence about how information and communication technologies can be used by these workers [57]. The authors of the review went on to assert that none of the existing technology-based interventions that could be used to facilitate home care were specifically designed to support the workflows of HHAs and concluded that "there is an urgent need for research that centers on the needs and perspectives of HHAs and using human-centered methods to engage HHAs in the design of technologies that truly support their essential caregiving work" [53]. The human-centered research reported here was specifically focused on developing technology for DSPs, but many of the design principles identified in Textbox 1 are also relevant to creating technologies to support the work of direct support workers who care for individuals without IDDs, such as HHAs.

However, inserting technology that is intended to facilitate sharing data among many care providers does not always increase documentation efficiency or decrease workload. Consider the migration from paper to electronic health records (EHRs), which significantly increased clinicians' workload [59-62]. EHRs have also been associated with a decrease in the amount of direct patient support time [63-65]. In contrast to EHRs, many of which grew out of billing systems [66], the DAT prototypes were developed by following an established user-centered design methodology [23] that ensured that our team considered DSPs' needs, goals, and constraints and obtained their feedback throughout our effort. As in the case of EHRs, a strong motivation behind creating the DAT prototypes helped to facilitate care for the individuals whose data are being logged electronically [67-70].

Multiple studies have pointed out that clinician burnout, which has been connected to the advent of EHRs, is correlated with decreased patient safety [71-74]. Similarly, DSP burnout and turnover are associated with poor health and quality of life for their clients [3,5]. Hence, technology such as the DAT prototypes, which are designed to reduce the burden of required documentation, improve job satisfaction, and reduce turnover among DSPs, can also help to improve the quality of life of adults with IDDs whom they support.

Limitations

Several limitations warrant mention. Our pilot test was limited to 14 DSPs working in 1 geographic region. However, different work settings, ranging from a private home with a single resident and group homes with 3 to 4 residents to a day program at a large agency, as well as the range of time that the DSPs who served as pilot testers had worked with adults with IDDs (<1 year to 20 years) increase the generalizability of our results. The pilot test length was also quite short, and we did not measure the amount of time spent generating documentation before the pilot test. These factors prevented us from assessing the impact of the DAT prototypes on documentation efficiency and quality. Nevertheless, the pilot test was long enough to establish feasibility.

Another limitation is that only 10 (71%) of the 14 pilot test volunteers were able to use the DAT prototypes for 3 successive shifts. Scheduling issues, often driven by staffing shortages that caused a staff member to be moved to a different location at the last minute, prevented several participants (4/14, 29%) from using the DAT as planned. In addition, we did not collect data about participants' familiarity and comfort with technology, although our observations while on site during the first shift provided insights about these factors. The fact that the DSPs were paid to participate is another limitation, although responses to the open-ended questions suggest that at least some of the participants were interested in continuing to use the tools when they would not be paid.

In addition, while the qualitative data from the test users was generally very positive, the study did not use any instruments that assess the likelihood of user adoption, such as the technology acceptance model [75]. Future work on technologies to support DSPs should be informed by the framework developed by Venkatesh et al [76] for understanding user acceptance of IT and might also consider the strategies outlined by Sebastian et al [77] for increasing user acceptance of novel technologies.

Implications

The results of the pilot test are promising, suggesting that upgraded versions of the DAT prototypes or similar technologies have the potential to reduce the burden of completing end-of-shift reports, while improving the quality of data produced by DSPs. Making it quick and simple to produce time-stamped in-the-moment notes facilitates logging more accurate and more detailed client data. Better data about adults with IDDs would enable family members, health providers, therapists, health providers, and behavior analysts to better support these adults. Long-term benefits of using the DAT could include (1) reducing DSP workload; (2) increasing the time DSPs spend interacting with adults with IDDs; (3) enabling DSPs to provide more consistent and appropriate support; (4) increasing DSP job satisfaction; (5) improving medical and behavioral support for adults with IDDs; and (6) providing a foundation for technology use that increases independence in adults with IDDs, thereby improving their quality of life. Digital documentation could also facilitate timely access to information about adults with IDDs for the diverse stakeholders who help support them.

However, there are risks associated with adopting technology-based data collection tools, including loss of data or an inability to log new data during technology failures. Moreover, to achieve the goal of capturing higher-quality data during work shifts, it will be necessary to allow users to customize the mobile app based on specific clients' support plans. This will add complexity, which could negatively impact user acceptance because many DSPs do not have a great deal of experience with technology. However, it is possible that supervisors or other employees at agencies that employ DSPs could learn to use an administrative portal to customize the mobile app based on clients' needs for support.

In any case, it will be important to obtain input from other stakeholders, such as behavioral supervisors, agency directors, and families, in future efforts, particularly as selection of suitable technology is likely the responsibility of the workspace and employer, and communication of data extends beyond DSPs. In the long term, technologies that support data collection will need to be integrated into existing report generation tools. Eventually, these technologies could even leverage artificial intelligence to create first drafts of DSPs' end-of-shift reports. If this sort of capability is developed, it will be very important to identify strategies for ensuring privacy and security and to consider the ethical implications of using artificial intelligence technologies [78,79].

Finally, while the DAT development effort focused on providing technology that supports DSPs, it is likely that many other types of direct workers could benefit from a similar platform (eg, HHAs, care workers, personal care assistants, certified nursing assistants, and nursing home assistants) [57,80-83]. This is significant because the United States is now facing a severe shortage across the entire direct support workforce [84,85] due in large part to the COVID-19 pandemic. As in the case with DSPs, the shortage in the direct support workforce is harmful to both these workers and those who rely upon them for support [86,87]. A report from the Centers for Disease Control and Prevention revealed that in 2016 a total of 61 million adults (approximately 1 in 4) living in the United States had a disability that impacts major life activities [88]. Many of these adults do, or will eventually, depend upon direct support workers for assistance with activities of daily living.

In summary, future efforts should not only increase the capabilities and robustness of the initial DAT prototypes and consider the needs of family members, medical providers, behavior analysts, and others who would benefit from timely accurate data about adults with IDDs but also explore how these tools could be adapted to meet the needs of other types of direct support workers by eliciting information and feedback from these workers.

Conclusions

DSPs play a critical role in the care of adults with IDDs. Technology can help mitigate the high turnover rates, poor job satisfaction, and the burden of necessary data collection and documentation that negatively impact DSPs' ability to care for these adults. The user-centered research effort reviewed here produced proof-of-concept prototypes of tools intended to improve the effectiveness and job satisfaction of DSPs. The results of the pilot test indicate that these tools are likely to provide the intended benefits to DSPs and thus have the potential to help improve the quality of life of clients served by DSPs.

Future research should include testing more robust feature-enhanced versions of the DAT over longer periods in even more diverse settings where DSPs provide support to adults with IDDs. Additional work should also include identifying or developing an instrument to reliably assess report quality and time-motion studies of DSPs before and during longer trials to help quantify how much time DSPs spend generating required documentation. One time-motion study of physicians working on hospital wards found that the physicians believed that they were spending more time on documentation and other administrative tasks than they actually were [89].

In any case, the work reported here, despite its limitations, provides valuable insights into how technology could benefit DSPs and the people they support. Feedback from DSPs indicates that the highest-priority feature enhancement for the DAT prototypes is enabling the mobile app to store data locally to support in-the-moment data collection without Wi-Fi connectivity. Several other enhancements, such as shared task lists, were identified as part of the initial user needs analysis activities performed at the start of this effort [16]. In addition to adding some of these capabilities, future work should identify the needs and constraints of the service providers who employ DSPs to identify barriers to adopting data collection and documentation technologies, such as costs, adaptability in small operations, the need to protect confidentiality, minimizing potential technology damage, and preventing data loss. Such work could help enable future versions of the DAT to supply all caregivers and service providers with the information necessary for better overall service, outcomes, and quality of life for adults with IDDs.

Finally, this line of research needs to be expanded because it could have a profound impact on the health and welfare of several other adults beyond those with IDDs who are supported by direct support workers: older adults, individuals with physical disabilities, and individuals with severe mental illness. Furthermore, the direct support workers themselves could also benefit from technologies such as DATs that enable quick and easy in-the-moment data collection and facilitate end-of-shift reporting.

Acknowledgments

Funding for this work was provided by a grant from the New Jersey Department of Health Governor's Council for Medical Research and Treatment of Autism (CAUT2020APL016). The authors would like to thank the direct support professionals who participated in these studies for sharing their valuable experiences and perspectives, the leadership of the 3 organizations where we conducted the pilot tests, and Adam Jonas and Morgan Murphy for helping with the thematic analysis of the responses to the open-ended survey questions.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Initial designs for the digital assistant tool mobile app.
[PDF File (Adobe PDF File), 938 KB - humanfactors_v11i1e51612_app1.pdf]

Multimedia Appendix 2
Demographic survey filled out by all pilot test participants.
[DOCX File , 21 KB - humanfactors_v11i1e51612_app2.docx]

Multimedia Appendix 3
Post–shift 1 survey filled out after using the digital assistant tool prototypes for a single shift.
[DOCX File , 16 KB - humanfactors_v11i1e51612_app3.docx]

Multimedia Appendix 4
Final survey filled out after participants had used the digital assistant tool prototypes for 2 or 3 shifts.
[DOCX File , 16 KB - humanfactors_v11i1e51612_app4.docx]

References

1. Larson SA, Neidorf J, Pettingell SL, Sowers M. In-home and residential long-term supports and services for persons with intellectual or developmental disabilities: status and trends through 2019. Minneapolis: University of Minnesota, Institute on Community Integration. 2023. URL: https://ici.umn.edu/products/AiMjLBqCS8mSwjLmIPPvKg [accessed 2024-04-03]
2. Mlinac ME, Feng MC. Assessment of activities of daily living, self-care, and independence. Arch Clin Neuropsychol 2016 Sep 29;31(6):506-516. [doi: 10.1093/arclin/acw049] [Medline: 27475282]
3. Friedman C. Direct support professionals and quality of life of people with intellectual and developmental disabilities. Intellect Dev Disabil 2018 Aug;56(4):234-250. [doi: 10.1352/1934-9556-56.5.234] [Medline: 30024847]
4. Friedman C. The impact of continuity and security on quality of life. Intellect Dev Disabil 2022 Apr 01;60(2):101-112. [doi: 10.1352/1934-9556-60.2.101] [Medline: 35297986]
5. Friedman C. The impact of direct support professional turnover on the health and safety of people with intellectual and developmental disabilities. Inclusion. 2021 2021;9(1):63-73 [FREE Full text] [doi: 10.1352/2326-6988-9.1.63]
6. Personal outcome measures: measuring personal quality of life. The Council on Quality and Leadership. 2017. URL: https://www.c-q-l.org/wp-content/uploads/2020/03/2017-CQL-POM-Manual-Adults.pdf [accessed 2024-04-03]
7. Macbeth JM. Commitment, capacity and culture: solutions for the direct support workforce crisis. The National Alliance for Direct Support Professionals. 2011. URL: https://tinyurl.com/7w8v9k35 [accessed 2024-04-03]
8. Claes C, Van Hove G, Vandevelde S, van Loon J, Schalock R. The influence of supports strategies, environmental factors, and client characteristics on quality of life-related personal outcomes. Res Dev Disabil 2012 Jan;33(1):96-103. [doi: 10.1016/j.ridd.2011.08.024] [Medline: 22093653]
9. Felce D, Lowe K, Jones E. Staff activity in supported housing services. Res Intellect Disabil 2002 Dec 11;15(4):388-403. [doi: 10.1046/j.1468-3148.2002.00130.x]
10. Felce D, Perry J. Living with support in the community. In: Odom SL, Horner RH, Snell ME, Blacher JB, editors. Handbook of Developmental Disabilities. Oxfordshire, United Kingdom: Routledge; 2007:410-428.

11. McConkey R, Abbott S, Walsh PN, Linehan C, Emerson E. Variations in the social inclusion of people with intellectual disabilities in supported living schemes and residential settings. J Intellect Disabil Res 2007 Mar 18;51(Pt 3):207-217. [doi: 10.1111/j.1365-2788.2006.00858.x] [Medline: 17300416]
12. Li S. Direct care personnel recruitment, retention, and orientation. Community-University Institute for Social Research. 2004. URL: https://tinyurl.com/4a7awczz [accessed 2024-04-03]
13. Brandt J. Report to the president 2017: America's direct support workforce crisis. President's Committee for People with Intellectual Disabilities (PCPID). 2017. URL: https://tinyurl.com/mry7afc8 [accessed 2024-04-03]
14. Hall SA, Anderson LL, Pettingell SL, Zhang A, Bershadsky J, Hewitt A, et al. Direct support professional and frontline supervisor perspectives on work-life in a pandemic. Inclusion (Wash) 2022;10(4):314-326 [FREE Full text] [doi: 10.1352/2326-6988-10.4.314] [Medline: 38528916]
15. Friedman C. The COVID-19 pandemic and quality of life outcomes of people with intellectual and developmental disabilities. Disabil Health J 2021 Oct;14(4):101117 [FREE Full text] [doi: 10.1016/j.dhjo.2021.101117] [Medline: 34144897]
16. Simmons CA, Moretti AE, Lobo AF, Tremoulet PD. Direct support professionals' perspectives on using technology to help support adults with autism spectrum disorder: mixed methods study. JMIR Form Res 2023 Apr 25;7:e40722 [FREE Full text] [doi: 10.2196/40722] [Medline: 37097738]
17. Hayes GR, Gardere LM, Abowd GD, Truong KN. CareLog: a selective archiving tool for behavior management in schools. In: Proceedings of the 2008 SIGCHI Conference on Human Factors in Computing Systems. 2008 Presented at: CHI '08; April 5-10, 2008; Florence, Italy p. 685-694 URL: https://dl.acm.org/doi/10.1145/1357054.1357164 [doi: 10.1145/1357054.1357164]
18. Bangerter A, Manyakov NV, Lewin D, Boice M, Skalkin A, Jagannatha S, et al. Caregiver daily reporting of symptoms in autism spectrum disorder: observational study using web and mobile apps. JMIR Ment Health 2019 Mar 26;6(3):e11365 [FREE Full text] [doi: 10.2196/11365] [Medline: 30912762]
19. Sleeper JD, LeBlanc LA, Mueller J, Valentino AL, Fazzio D, Raetz PB. The effects of electronic data collection on the percentage of current clinician graphs and organizational return on investment. J Organ Behav Manage 2017 Jan 10;37(1):83-95. [doi: 10.1080/01608061.2016.1267065]
20. Kroezen M, Schäfer W, Sermeus W, Hansen J, Batenburg R. Healthcare assistants in EU member states: an overview. Health Policy 2018 Oct;122(10):1109-1117. [doi: 10.1016/j.healthpol.2018.07.004] [Medline: 30033205]
21. Desa U. World population prospects 2019: highlights. United Nations Department for Economic and Social Affairs. 2019. URL: https://population.un.org/wpp/Publications/Files/WPP2019_10KeyFindings.pdf [accessed 2024-04-03]
22. Lakin KC, Burke MM. Looking forward: research to respond to a rapidly aging population. Res Pract Persons Severe Disabl 2019 Nov 01;44(4):280-292. [doi: 10.1177/1540796919882356]
23. Craven PL, Tremoulet PD, Regli SH. Incorporating human systems engineering in advanced military technology development. In: McNeese MD, Forster PK, editors. Cognitive Systems Engineering: An Integrative Living Laboratory Framework. Boca Raton, FL: CRC Press; 2017:341-359.
24. Isaac S, Michael WB. Handbook in Research and Evaluation: A Collection of Principles, Methods, and Strategies Useful in the Planning, Design, and Evaluation of Studies in Education and the Behavioral Sciences. 3rd edition. New York, NY: Edits Publications; 1995.
25. Hill R. What sample size is "enough" in internet survey research. Interpers Compute Technol 1998;6(3-4):1-12 [FREE Full text]
26. Julious SA. Sample size of 12 per group rule of thumb for a pilot study. Pharm Stat 2005 Nov 24;4(4):287-291. [doi: 10.1002/pst.185]
27. van Belle G. Statistical Rules of Thumb. Hoboken, NJ: John Wiley & Sons; 2011.
28. Brooke J. SUS: a retrospective. J Usability Stud 2013;8(2):29-40 [FREE Full text]
29. Brooke J. SUS: a 'quick and dirty' usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation In Industry. Boca Raton, FL: CRC Press; 1996:4-7.
30. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
31. Maguire M, Delahunt B. Doing a thematic analysis: a practical, step-by-step guide for learning and teaching scholars. Ir J High Educ 2017;9(3) [FREE Full text]
32. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123 [FREE Full text]
33. Sauro J. Does better usability increase customer loyalty? MeasuringU. 2010. URL: https://measuringu.com/usability-loyalty/ [accessed 2024-04-03]
34. Swanson-Aprill L, Luz C, Travis A, Hunt J, Wamsley S. Policy brief direct care workforce shortage. Michigan Aging and Adult Services Agency. 2019. URL: https://tinyurl.com/2s3n949n [accessed 2024-04-02]
35. Taylor SJ. The direct support workforce crisis: can unions help resolve this? Center on Human Policy. 2008. URL: https://eric.ed.gov/?id=ED503912 [accessed 2024-04-02]
36. Bradley VJ. How COVID-19 may change the world of services to people with intellectual and developmental disabilities. Intellect Dev Disabil 2020 Oct 01;58(5):355-360. [doi: 10.1352/1934-9556-58.5.355] [Medline: 33032314]

37. Embregts PJ, Tournier T, Frielink N. Experiences and needs of direct support staff working with people with intellectual disabilities during the COVID-19 pandemic: a thematic analysis. J Appl Res Intellect Disabil 2021 Mar 21;34(2):480-490 [FREE Full text] [doi: 10.1111/jar.12812] [Medline: 32954585]
38. Position statement on the direct support professional workforce. New Jersey Council on Developmental Disabilities. 2019. URL: https://njcdd.org/wp-content/uploads/DSP-NJCDD-Position-Statement.pdf [accessed 2024-04-03]
39. Laws CB. The direct support professional workforce crisis: challenges, state approaches, and opportunities for Georgia. The University of Georgia. 2019. URL: https://tinyurl.com/3pt3c3s8 [accessed 2024-04-03]
40. Pettingell SL, Houseworth J, Tichá R, Kramme JE, Hewitt AS. Incentives, wages, and retention among direct support professionals: national core indicators staff stability survey. Intellect Dev Disabil 2022 Apr 01;60(2):113-127 [FREE Full text] [doi: 10.1352/1934-9556-60.2.113] [Medline: 35297985]
41. Thornton CR. Direct support staff retention and turnover in the field of applied behavior analysis: a national survey. Temple University Graduate Board. 2018. URL: https://tinyurl.com/4zbdkpp6 [accessed 2024-04-05]
42. Kazemi E, Shapiro M, Kavner A. Predictors of intention to turnover in behavior technicians working with individuals with autism spectrum disorder. Res Autism Spectr Disord 2015 Sep;17:106-115. [doi: 10.1016/j.rasd.2015.06.012]
43. Bogenschutz M, Nord D, Hewitt A. Competency-based training and worker turnover in community supports for people with IDD: results from a group randomized controlled study. Intellect Dev Disabil 2015 Jun;53(3):182-195. [doi: 10.1352/1934-9556-53.3.182] [Medline: 26107852]
44. Bossink LW, van der Putten AA, Vlaskamp C. Understanding low levels of physical activity in people with intellectual disabilities: a systematic review to identify barriers and facilitators. Res Dev Disabil 2017 Sep;68:95-110. [doi: 10.1016/j.ridd.2017.06.008] [Medline: 28750208]
45. Gray-Stanley JA, Muramatsu N, Heller T, Hughes S, Johnson TP, Ramirez-Valles J. Work stress and depression among direct support professionals: the role of work support and locus of control. J Intellect Disabil Res 2010 Aug 15;54(8):749-761 [FREE Full text] [doi: 10.1111/j.1365-2788.2010.01303.x] [Medline: 20633203]
46. Couderc S, Cousson-Gélie F, Pernon E, Porro B, Miot S, Baghdadli A. Burnout among direct support workers of adults with autism spectrum disorder and intellectual disability. Scand J Caring Sci 2023 Mar;37(1):131-140. [doi: 10.1111/scs.13008] [Medline: 34075606]
47. Gray-Stanley JA, Muramatsu N. Work stress, burnout, and social and personal resources among direct care workers. Res Dev Disabil 2011 May;32(3):1065-1074 [FREE Full text] [doi: 10.1016/j.ridd.2011.01.025] [Medline: 21316918]
48. Mutkins E, Brown R, Thorsteinsson E. Stress, depression, workplace and social supports and burnout in intellectual disability support staff. J Intellect Disabil Res 2011 May;55(5):500-510. [doi: 10.1111/j.1365-2788.2011.01406.x] [Medline: 21418365]
49. Ryan C, Bergin M, Wells JS. Work-related stress and well-being of direct care workers in intellectual disability services: a scoping review of the literature. Int J Dev Disabil 2019 Mar 27;67(1):1-22 [FREE Full text] [doi: 10.1080/20473869.2019.1582907] [Medline: 34141395]
50. Siyam N. Factors impacting special education teachers' acceptance and actual use of technology. Educ Inf Technol 2019 Jan 8;24(3):2035-2057. [doi: 10.1007/s10639-018-09859-y]
51. Bak MY, Plavnick JB, Dueñas AD, Brodhead MT, Avendaño SM, Wawrzonek AJ, et al. The use of automated data collection in applied behavior analytic research: a systematic review. Behav Anal Res Pract 2021 Nov;21(4):376-405. [doi: 10.1037/bar0000228]
52. Nadkar RS. Increasing accuracy of data collection by direct support professionals with task clarification and performance feedback. Temple University Libraries. 2018. URL: https://scholarshare.temple.edu/handle/20.500.12613/1990 [accessed 2024-04-03]
53. Kuo EF, Cho J, Olaye I, Delgado D, Dell N, Sterling MR. Understanding the technological landscape of home health aides: scoping literature review and a landscape analysis of existing mHealth apps. J Med Internet Res 2022 Nov 11;24(11):e39997 [FREE Full text] [doi: 10.2196/39997] [Medline: 36176033]
54. Bartle V, Albright L, Dell N. "This machine is for the aides": tailoring voice assistant design to home health care work. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. 2023 Presented at: CHI '23; April 23-28, 2023; Hamburg, Germany p. 1-9 URL: https://dl.acm.org/doi/abs/10.1145/3544548.3581118 [doi: 10.1145/3544548.3581118]
55. Franzosa E, Judon KM, Gottesman EM, Koufacos NS, Runels T, Augustine M, et al. Home health aides' increased role in supporting older veterans and primary healthcare teams during COVID-19: a qualitative analysis. J Gen Intern Med 2022 Jun 22;37(8):1830-1837 [FREE Full text] [doi: 10.1007/s11606-021-07271-w] [Medline: 35319082]
56. Islam SM, Halooq A, Dening J, Uddin R, Laranjo L, Chow CK, et al. Healthcare providers' perspectives on using smart home systems to improve self-management and care in people with heart failure: a qualitative study. Int J Med Inform 2022 Nov;167:104837. [doi: 10.1016/j.ijmedinf.2022.104837] [Medline: 36126353]
57. Perez H, Neubauer N, Marshall S, Philip S, Miguel-Cruz A, Liu L. Barriers and benefits of information communication technologies used by health care aides. Appl Clin Inform 2022 Jan 09;13(1):270-286 [FREE Full text] [doi: 10.1055/s-0042-1743238] [Medline: 35263800]
58. Ogawa H, Yonezawa Y, Maki H, Sato H, Hahn AW, Caldwell WM. A Java mobile phone-based "home helper" care report creation support system. Biomed Sci Instrum 2004;40:76-79. [Medline: 15133938]

59. Lou SS, Lew D, Harford DR, Lu C, Evanoff BA, Duncan JG, et al. Temporal associations between EHR-derived workload, burnout, and errors: a prospective cohort study. J Gen Intern Med 2022 Jul 16;37(9):2165-2172 [FREE Full text] [doi: 10.1007/s11606-022-07620-3] [Medline: 35710654]
60. Shan Y, Shang J, Yan Y, Ye X. Workflow interruption and nurses' mental workload in electronic health record tasks: an observational study. BMC Nurs 2023 Mar 09;22(1):63 [FREE Full text] [doi: 10.1186/s12912-023-01209-9] [Medline: 36890555]
61. Arndt BG, Beasley JW, Watkinson MD, Temte JL, Tuan WJ, Sinsky CA, et al. Tethered to the EHR: primary care physician workload assessment using EHR event log data and time-motion observations. Ann Fam Med 2017 Sep 11;15(5):419-426 [FREE Full text] [doi: 10.1370/afm.2121] [Medline: 28893811]
62. Gregory ME, Russo E, Singh H. Electronic health record alert-related workload as a predictor of burnout in primary care providers. Appl Clin Inform 2017 Dec 20;08(03):686-697. [doi: 10.4338/aci-2017-01-ra-0003]
63. Baumann LA, Baker J, Elshaug AG. The impact of electronic health record systems on clinical documentation times: a systematic review. Health Policy 2018 Aug;122(8):827-836. [doi: 10.1016/j.healthpol.2018.05.014] [Medline: 29895467]
64. Joukes E, Abu-Hanna A, Cornet R, de Keizer N. Time spent on dedicated patient care and documentation tasks before and after the introduction of a structured and standardized electronic health record. Appl Clin Inform 2018 Jan 17;9(1):46-53 [FREE Full text] [doi: 10.1055/s-0037-1615747] [Medline: 29342479]
65. Asan O, D Smith P, Montague E. More screen time, less face time - implications for EHR design. J Eval Clin Pract 2014 Dec 19;20(6):896-901 [FREE Full text] [doi: 10.1111/jep.12182] [Medline: 24835678]
66. Kim E, Rubinstein SM, Nead KT, Wojcieszynski AP, Gabriel PE, Warner JL. The evolving use of electronic health records (EHR) for research. Semin Radiat Oncol 2019 Oct;29(4):354-361. [doi: 10.1016/j.semradonc.2019.05.010] [Medline: 31472738]
67. Transforming health care through information technology. President’s Information Technology Advisory Committee Panel on Transforming Health Care. 2006. URL: https://www.nitrd.gov/pubs/pitac/pitac-hc-9feb01.pdf [accessed 2024-04-03]
68. Evans RS. Electronic health records: then, now, and in the future. Yearb Med Inform 2018 Mar 06;25(S 01):S48-S61. [doi: 10.15265/iys-2016-s006]
69. Kohli R, Tan SS. Electronic health records: how can IS researchers contribute to transforming healthcare? MIS Q 2016;40(3):553-574 [FREE Full text]
70. Burde H. The HITECH act: an overview. AMA J Ethics 2011;13(3):172-175 [FREE Full text]
71. Nijor S, Rallis G, Lad N, Gokcen E. Patient safety issues from information overload in electronic medical records. J Patient Saf 2022 Sep 01;18(6):e999-1003 [FREE Full text] [doi: 10.1097/PTS.0000000000001002] [Medline: 35985047]
72. Aldosari B. Patients' safety in the era of EMR/EHR automation. Inform Med Unlocked 2017;9:230-233. [doi: 10.1016/j.imu.2017.10.001]
73. Middleton B, Bloomrosen M, Dente MA, Hashmat B, Koppel R, Overhage JM, American Medical Informatics Association. Enhancing patient safety and quality of care by improving the usability of electronic health record systems: recommendations from AMIA. J Am Med Inform Assoc 2013 Jun 01;20(e1):e2-e8 [FREE Full text] [doi: 10.1136/amiajnl-2012-001458] [Medline: 23355463]
74. Meeks DW, Smith MW, Taylor L, Sittig DF, Scott JM, Singh H. An analysis of electronic health record-related patient safety concerns. J Am Med Inform Assoc 2014 Nov 01;21(6):1053-1059 [FREE Full text] [doi: 10.1136/amiajnl-2013-002578] [Medline: 24951796]
75. Davis FD. Technology acceptance model: TAM. In: Al-Aufi AS, Al-Suqri MN, editors. Information Seeking Behavior and Technology Adoption: Theories and Trends. Hershey, PA: IGI Global; 1989:205-219.
76. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]
77. Sebastian G, George A, Jackson Jr G. Persuading patients using rhetoric to improve artificial intelligence adoption: experimental study. J Med Internet Res 2023 Mar 13;25:e41430 [FREE Full text] [doi: 10.2196/41430] [Medline: 36912869]
78. Sebastian G. Privacy and data protection in ChatGPT and other AI chatbots: strategies for securing user information. SSRN J 2023 [FREE Full text] [doi: 10.2139/ssrn.4454761]
79. Sebastian G. Exploring ethical implications of ChatGPT and other AI chatbots and regulation of disinformation propagation. SSRN J 2023. [doi: 10.2139/ssrn.4461801]
80. Ming J, Kuo E, Go K, Tseng E, Kallas J, Vashistha A, et al. "I go beyond and beyond" examining the invisible work of home health aides. Proc ACM Hum Comput Interact 2023 Apr 16;7(CSCW1):1-21. [doi: 10.1145/3579492]
81. Saari M, Patterson E, Killackey T, Raffaghello J, Rowe A, Tourangeau AE. Home-based care: barriers and facilitators to expanded personal support worker roles in Ontario, Canada. Home Health Care Serv Q 2017 Nov 08;36(3-4):127-144. [doi: 10.1080/01621424.2017.1393482] [Medline: 29048246]
82. King S, Liu L, Stroulia E, Nikolaidis I. Using simulations to integrate technology into health care aides' workflow. Int J Adv Corp Learn 2013 Jul 23;6(2):28. [doi: 10.3991/ijac.v6i2.2982]
83. Sterling MR, Dell N, Tseng E, Okeke F, Cho J, Piantella B, et al. Home care workers caring for adults with heart failure need better access to training and technology: a role for implementation science. J Clin Transl Sci 2020 Apr 06;4(3):224-228 [FREE Full text] [doi: 10.1017/cts.2020.36] [Medline: 32695493]

84. Kreider AR, Werner RM. The home care workforce has not kept pace with growth in home and community-based services. Health Aff (Millwood) 2023 May 01;42(5):650-657. [doi: 10.1377/hlthaff.2022.01351] [Medline: 37075251]
85. Jumabhoy S, Jung HY, Yu J. Characterizing the direct care health workforce in the United States, 2010-2019. J Am Geriatr Soc 2022 Feb 22;70(2):512-521. [doi: 10.1111/jgs.17519] [Medline: 34687042]
86. Smallwood N, Harrex W, Rees M, Willis K, Bennett CM. COVID-19 infection and the broader impacts of the pandemic on healthcare workers. Respirology 2022 Jun 19;27(6):411-426 [FREE Full text] [doi: 10.1111/resp.14208] [Medline: 35048469]
87. Harootunian L, Buffett A, O'Gara B, Perry K, Serafini MW, Hoagland GW. A multipronged approach to alleviating the direct care workforce shortage. Health Aff forefront 2023 [FREE Full text] [doi: 10.1377/forefront.20230608.815571]
88. Okoro CA, Hollis ND, Cyrus AC, Griffin-Blake S. Prevalence of disabilities and health care access by disability status and type among adults - United States, 2016. MMWR Morb Mortal Wkly Rep 2018 Aug 17;67(32):882-887 [FREE Full text] [doi: 10.15585/mmwr.mm6732a3] [Medline: 30114005]
89. Westbrook JI, Ampt A, Kearney L, Rob MI. All in a day's work: an observational study to quantify how and with whom doctors on hospital wards spend their time. Med J Aust 2008 May 05;188(9):506-509. [doi: 10.5694/j.1326-5377.2008.tb01762.x] [Medline: 18459920]

Abbreviations

DAT: digital assistant tool
DSP: direct support professional
EHR: electronic health record
HHA: home health aide
IDD: intellectual and developmental disability
IDEAS: Interaction Design and Engineering for Advanced Systems
PI: principal investigator
SUS: System Usability Scale

Edited by A Parush; submitted 05.08.23; peer-reviewed by G Sebastian, A AL-Asadi; comments to author 14.01.24; revised version received 07.03.24; accepted 16.03.24; published 25.04.24.

Please cite as:
Tremoulet PD, Lobo AF, Simmons CA, Baliga G, Brady M
Assessing the Usability and Feasibility of Digital Assistant Tools for Direct Support Professionals: Participatory Design and Pilot-Testing
JMIR Hum Factors 2024;11:e51612
URL: https://humanfactors.jmir.org/2024/1/e51612
doi:10.2196/51612
PMID:38662420

Original Paper

Assessing the Feasibility and Preliminary Effects of a Web-Based Self-Management Program for Chronic Noncancer Pain: Mixed Methods Study

Pascale Marier-Deschenes[1,2], PhD; Anne Marie Pinard[1,2,3], MEd, MD; Laura Jalbert[4], BA, MSc; Annie LeBlanc[1,4], PhD

[1]Laval University, Medicine Faculty, Québec, QC, Canada

[2]CIRRIS, Centre interdisciplinaire de recherche en réadaptation et intégration sociale, Québec, QC, Canada

[3]CHU de Québec, Université Laval, Québec, QC, Canada

[4]VITAM, Centre de recherche en santé durable, Québec, QC, Canada

Corresponding Author:
Pascale Marier-Deschenes, PhD
CIRRIS, Centre interdisciplinaire de recherche en réadaptation et intégration sociale
525, boul. Wilfrid-Hamel Québec
Québec, QC, G1M 2S8
Canada
Phone: 1 418 649 3735
Email: pascale.marier-deschenes.1@ulaval.ca

Abstract

Background: In Canada, adults with chronic noncancer pain face a persistent insufficiency of publicly funded resources, with the gold standard multidisciplinary pain treatment facilities unable to meet the high clinical demand. Web-based self-management programs cost-effectively increase access to pain management and can improve several aspects of physical and emotional functioning. Aiming to meet the demand for accessible, fully automated resources for individuals with chronic noncancer pain, we developed a French web- and evidence-based self-management program, *Agir pour moi* (APM). This program includes pain education and strategies to reduce stress, practice mindfulness, apply pacing, engage in physical activity, identify and manage thinking traps, sleep better, adapt diet, and sustain behavior change.

Objective: This study aims to assess the APM self-management program's feasibility, acceptability, and preliminary effects in adults awaiting specialized services from a center of expertise in chronic pain management.

Methods: We conducted a mixed methods study with an explanatory sequential design, including a web-based 1-arm trial and qualitative semistructured interviews. We present the results from both phases through integrative tables called joint displays.

Results: Response rates were 70% (44/63) at postintervention and 56% (35/63) at 3-month follow-up among the 63 consenting participants who provided self-assessed information at baseline. In total, 46% (29/63) of the participants completed the program. We interviewed 24% (15/63) of the participants. The interview's first theme revolved around the overall acceptance, user-friendliness, and engaging nature of the program. The second theme emphasized the differentiation between microlevel and macrolevel engagements. The third theme delved into the diverse effects observed, potentially influenced by the macrolevel engagements. Participants highlighted the features that impacted their self-efficacy and the adoption of self-management strategies. We observed indications of improvement in self-efficacy, pain intensity, pain interference, depression, and catastrophizing. Interviewees described these and various other effects as potentially influenced by macrolevel engagement through behavioral change.

Conclusions: These findings provided preliminary evidence that the APM self-management program and research methods are feasible. However, some participants expressed the need for at least phone reminders and minimal support from a professional available to answer questions over the first few weeks of the program to engage. Recruitment strategies of a future randomized controlled trial should focus on attracting a broader representation of individuals with chronic pain in terms of gender and ethnicity.

Trial Registration: ClinicalTrials.gov NCT05319652; https://clinicaltrials.gov/study/NCT05319652

(JMIR Hum Factors 2024;11:e50747) doi:10.2196/50747

KEYWORDS

persistent pain; eHealth; self-paced intervention; web-based program; evidence based; web based; self-management; pain; chronic pain; mixed methods study; pain treatment; pain education

Introduction

Background

The prevalence of chronic pain, including chronic cancer pain in adults, is estimated to be between 18% and 21%, with severe repercussions for all aspects of the lives of those affected, their families, and society [1-3]. Chronic pain affects patient-perceived health status and psychological functioning; decreases energy levels; and hinders engagement with physical, emotional, cognitive, and social activities [3-6]. These impacts can strain familial and social relationships and affect work performance [7]. Living with chronic pain often involves increased medical expenditures and detrimentally affects one's financial well-being [3,6]. In addition, the wait for services is not without added consequences to these repercussions, with long wait time (12-30 months) being associated with further deterioration in pain-related interference, psychological distress, and pain acceptance [3,8-10].

In Canada, adults with chronic noncancer pain face a persistent insufficiency of publicly funded resources, with the gold standard multidisciplinary pain treatment facilities being unable to meet the high clinical demand [1,2,11]. Since 2019, the Canadian Task Force has reaffirmed the necessity to implement equitable and innovative ways to deliver health interventions in a timely manner in the public network [12,13]. Web-based self-management programs that include exercise, sleep hygiene, pacing, and a healthy lifestyle are endorsed as part of the therapeutic considerations and recommendations for chronic noncancer pain management [14]. These programs have shown an impact on patients' pain intensity, pain interference [15,16], anxiety [15,17], depression [17,18], stress [18], catastrophizing, and self-efficacy [19].

The lack of accessible and reliable unguided web-based self-management programs tailored to French-speaking individuals with chronic noncancer pain is a significant yet solvable health services gap. Over the years, individuals with lived experience, organizations, and researchers have stressed the relevance and importance of actively involving patient partners in the health intervention development process [20-25]. Therefore, a novel French web-based self-management program for chronic noncancer pain developed in collaboration with individuals with lived experience could meet the specific needs of French-speaking individuals [26].

Objectives

This study aims to (1) assess the feasibility and acceptability of the *Agir pour moi* (APM) self-management program and trial procedures and (2) explore preliminary outcomes in individuals living with chronic noncancer pain.

Methods

Study Design

We conducted a mixed methods sequential explanatory study consisting of a single-arm, pre- and postintervention trial, followed by qualitative, semistructured interviews with adults experiencing chronic noncancer pain and awaiting services from a center of expertise in chronic pain management [27-29]. We registered the trial at ClinicalTrials.gov (NCT05319652) and followed the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) and guidelines for reporting nonrandomised pilot and feasibility studies [30,31].

Ethical Considerations

The University Hospital Centre (CHU) de Québec-Université Laval Research Ethics Board approved the study (#2023-6312).

Knowledge Users' Involvement

A total of 7 individuals with lived experience of chronic pain, 5 health care professionals with experience and expertise in chronic pain management, 3 medical students, and 1 graphic designer contributed to the program's codevelopment. We engaged in web-based, phone, or email conversations over a period of 1.5 years. All knowledge users could contribute to various aspects of program development, including its identity (eg, colors, logo, and name), structure (eg, lesson sequence, content organization, and navigation), content (eg, self-management strategies, theoretical content, and testimonials), and learning modalities and behavior change techniques (eg, personal plans, reflective activities, and interactive scenarios). These knowledge users were not further involved across the duration of the trial, but individuals with lived experience initially guided the team toward reducing the questionnaire burden to a minimum for the participants.

Study Setting, Participants, and Recruitment

We recruited participants from the center of expertise in chronic pain management waitlist at the CHU de Québec-Université Laval. This center provides superspecialized services intended for complex chronic pain cases requiring a technical platform and multidisciplinary team. Most individuals referred to such centers experience significant impairments, including high pain levels interfering with their daily life, moderate to extremely severe depression, and pain-related sleep disturbance. Most of them take prescription analgesic medication and have already consulted different types of health care professionals [2].

Using the center's assigned priority level, between June and August 2022, we sent 500 invitation letters to adults (aged >18 years) with chronic noncancer pain (for >3 months) unlikely to receive services within the next 6 months. Interested individuals were to email us to set an eligibility interview, confirming that they understood French, had access to a computer and high-speed internet, had not started a new treatment for pain within the last 1 month and agreed to notify us before starting a new one, were available for the duration of the study, and were able to provide informed consent. We excluded individuals who participated in a chronic pain self-management program within the last year or those who were scheduled for surgical treatment within 6 months. Following the assessment for eligibility, a research team member explained the study procedures and recorded verbal informed consent.

Intervention

The codevelopment of the APM self-management program (thereafter APM or the program) is detailed [32] and available in the study by Marier-Deschenes et al [33]. Briefly, we designed a cognitive behavioral therapy (CBT)–centered, web-based self-management program that would enable participants to develop their self-management skills autonomously (eg, goal setting), practice suggested self-management strategies (eg, pacing), and sustain new behaviors (eg, respect of limits). Despite the diverse nature of our targeted population, the proposed self-management strategies to explore and develop are mostly universal, spanning areas such as managing thoughts and emotions, gradually resuming physical activity, practicing pacing, and adopting good sleep hygiene. The program is self-guided (ie, unguided) in that it provides the same information as face-to-face programs offered in tertiary pain clinics but without therapeutic support from health care professionals [34]. It is structured around weekly lessons over an 8-week period with content that encompasses 26 different behavior change techniques [35] and a downloadable personal plan (Table 1).

Table 1. Summary of Agir pour moi (APM) program's topics and self-management strategies with associated content.

Week	Topics and strategies	Lesson headers
Foreword	What does APM offer?	• What is self-management? • Who is this program for? • Will you have less pain? • Key attitudes to adopt • How to navigate the program?
Week 1	Introduction	• What is chronic pain? • Are you ready for self-management? • How to set specific, measurable, appealing, realistic, and time-bound (SMART) objectives?
Week 2	Engage in well-being activities	• Reduce stress • Experience mindfulness
Week 3	Practice pacing	• Follow-up on last week's objective • Evaluate your energy expenditures
Week 4	Practice pacing, continued	• Follow-up on last week's objective • Planning your weeks
Week 5	Engage in physical activity	• Follow-up on last week's objective • Stretching exercises • Engage in physical activity that you enjoy
Week 6	Take care of your thoughts	• Follow-up on last week's objective • Identify thinking traps • Perceive the positive
Week 7	Revise your lifestyle habits	• Follow-up on last week's objective • Promote sleep • Adapt your diet
Week 8	Plan for the future	• Reflect on previous objectives and further goals • Sustain the change

The program incorporates a variety of media, including photos, infographics, interactive scenarios, tables, audio recordings, and videos. All content is fully narrated, with short audio clips accompanying written information in each lesson, ensuring accessibility in both formats. Interactive exercises such as quizzes, drag-and-drop questions, and real-life scenarios enhance understanding.

Participants were encouraged to set and track their own weekly objectives related to the topic in their personal plan, which serves as our learner's workbook. This plan features reflective, observational, monitoring, problem-solving, and action-planning activities.

Participants were advised to allocate 60 to 90 minutes weekly for program activities. They had the flexibility to divide the lessons into multiple short sessions; completing each lesson in 1 sitting was not necessary. While participants were encouraged to follow the program sequentially, all 8-week content was readily accessible.

Following the program poses minimal health risks. The program incorporates low-intensity activities, such as stretching exercises, which might cause temporary discomfort when resumed. However, the risks associated with physical inactivity, including the development or worsening of chronic illnesses, outweigh those of gradually resuming physical activity.

Quantitative Data Collection and Outcomes

Overview

Participants were assigned a log-in user ID and password for the program's web-based platform. They completed self-reported questionnaires on the web at 3 time points: preintervention, postintervention, and 3 months after completing the program. We sent an email reminder to those who did not log in at least once a week or complete questionnaires at the appropriate time. Participants who completed all questionnaires were eligible for a random computerized drawing of 5 CAD $75 (US $55.89) gift cards. We provided participants facing technical difficulties with phone support. We used REDCap (Research Electronic Data Capture; Vanderbilt University), a secure web application, for creating and managing surveys and databases.

As this was a feasibility study, we did not perform a power calculation on measures of effect but rather aimed at estimating the number of eligible participants and the potential recruitment rate from the center of expertise in chronic pain management waitlist. Therefore, this study is not appropriately powered to assess APM's efficacy [36].

Feasibility and Acceptability Outcomes

We considered the following outcomes in assessing the feasibility of the intervention and research methods and the acceptability of the program: (1) feasibility of recruitment (number of referred adults who responded to the invitation and consented to participate in the study and number of interested adults excluded based on inclusion and exclusion criteria), (2) feasibility of data collection (rate of response to and completion of the questionnaires at each time point), (3) acceptability for those who engaged with the program (mean score to the Acceptability eScale, which includes dimensions of usability and satisfaction) [37,38], and (4) engagement (number of lessons completed). Participants completing at least 6 (75%) out of the 8 weekly lessons were defined as program completers.

Effects Measures

We opted for the French versions of the following self-reported measures based on the recommendations of the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials [39] (Table 2).

Table 2. Self-reported measures.

Measures	Items, n	Constructs	Score range	High score meaning
Pain Self-Efficacy Questionnaire [40,41]	10	Individual's confidence to perform activities while experiencing pain	0-60	Better self-efficacy
Pain intensity subscale of the Brief Pain Inventory [42,43]	4	Worst, least, average, and current pain intensity	0-10	Worse pain intensity
Pain interference subscale of the Brief Pain Inventory	7	Impact of pain on general activity, mood, walking ability, normal work, sleep, relationships, and enjoyment of life	0-10	Worse pain interference
Anxiety subscale of the Hospital Anxiety and Depression Scale [44]	7	State of anxiety	0-21	Worst anxiety symptoms
Depression subscale of the Hospital Anxiety and Depression Scale	7	State of depression	0-21	Worst depressive symptoms
Pain Catastrophizing Scale [45]	13	Catastrophic thinking and maladaptive responses to pain	0-52	Worst catastrophizing
Patient Global Impression of Change Scale	1	Patient's rating of overall improvement	1-7	Greater change

Statistical Analyses

We performed descriptive statistics using means (SD) for continuous outcomes and frequencies (%) for categorical outcomes. We compared pre-, post-, and follow-up intervention scores for effect measures using repeated-measures linear models. All statistical analyses were conducted using R software (version 4.3.0; R Foundation for Statistical Computing) in RStudio (version 2023.06.0; Posit PBC) [46,47]. Models were fit using lme4 (version 1.1-33; R Foundation for Statistical Computing) [48]. Model fit evaluations and assumption checks were done through visualizations using performance (version 0.10.4) [49]. Effects were considered significant when the 95% CI for the estimates did not include 0.

Qualitative Data Collection and Outcomes

Overview

We formed a heterogeneous group of 20 potential participants using their Acceptability eScale and pre- and posteffect measure scores. We invited participants with positive and negative impressions of the intervention and those for whom we could observe the effects on functioning or not. From a practical perspective, we made the decision to not interview individuals who did not engage with the program at all, as they would not have been able to provide valuable insights into the program's acceptability and feasibility. However, we conducted interviews with participants who did not complete the program; they were just not specifically selected based on this criterion. We conducted semistructured, audio-recorded, 40-minute phone interviews 5 to 7 months after the intervention. We achieved data saturation with 15 interviews (12/15, 80% women and 3/15, 20% men) and did not deem it necessary to conduct further interviews [50].

Data Analysis

We analyzed the transcriptions using inductive and deductive thematic analysis based on the motivational model for pain self-management [51]. The lead author read the interviews multiple times to obtain a detailed understanding, then coded them according to the research questions and with consideration for the model's components. According to the model by Jensen et al [51], the willingness to embrace pain self-management behaviors is influenced by 2 primary factors. First, it is molded by beliefs concerning the perceived importance of these behaviors, encompassing considerations of cost/benefit ratio, learning history, and current contingencies. Second, self-efficacy, denoting personal beliefs about one's abilities to accomplish a specific task, also plays a pivotal role in shaping the inclination toward behavior change. Furthermore, to ensure the validity of the analysis, a research associate coded and discussed 3 interviews. Then, the lead author further identified meaningful units and assembled them into descriptive categories. She analyzed, interpreted, and summarized categories into 3 explanatory themes that were then discussed among all coauthors [52].

Integration

We addressed our study's main objectives by integrating the quantitative and qualitative results, drawing on all relevant data. We opted for 2 approaches: a weaving approach through the narrative and 2 joint displays presenting categories and associated quotes explaining the quantitative data [53,54].

Results

Feasibility

From the 500 invitations sent, 74 (15%) individuals expressed interest, of which 65 (13%) were confirmed eligible. Of these 65 eligible individuals, 63 (97%) consented to participate (Figure 1). A total of 9 (12%) of the 74 participants were ineligible due to unavailability during summer, current services from a pain clinic or rehabilitation center, lack of belief in program helpfulness, absence of computer access, low literacy level, and being unreachable. Response rates were 70% (44/63) at postintervention and 56% (35/63) at 3-month follow-up. A total of 15 (24%) of the 63 participants exited the study for reasons mostly unrelated to the intervention. Missing data were attributed to connectivity problems with the REDCap platform.

Participants were almost exclusively White (61/63, 97%) and female (44/63, 70%), with a mean age of 54 (range 24-75) years. On average, participants experienced chronic pain symptoms for 12 (SD 13.6) years, and most (34/63, 54%) had chronic musculoskeletal pain (Table 3).

Figure 1. Flow diagram of the participants' recruitment, enrollment, and engagement.

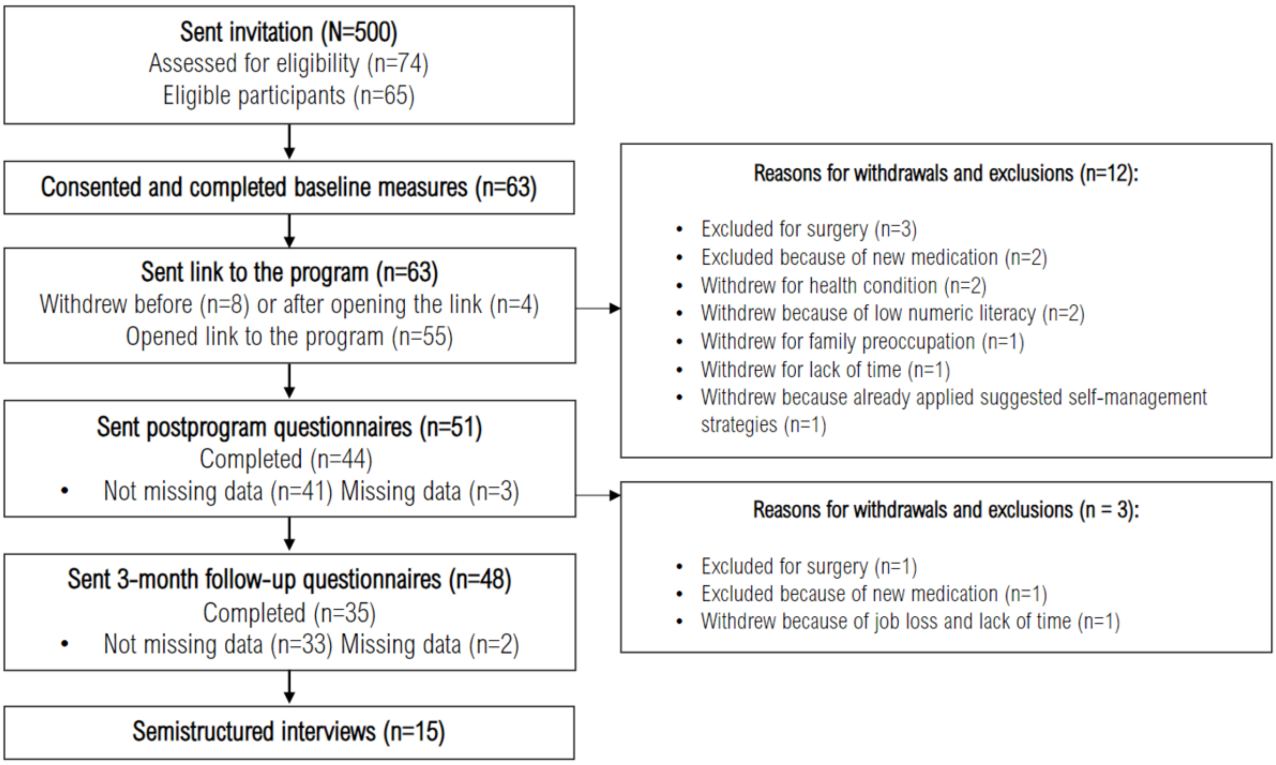


Table 3. Participants' sociodemographic and clinical characteristics (N=63).

Sociodemographic characteristics	Values, n (%)
Age (years)	
20 to 29	2 (3)
30 to 39	6 (10)
40 to 49	13 (21)
50 to 59	17 (27)
60 to 69	19 (30)
70 to 79	6 (10)
Gender	
Women	44 (70)
Men	17 (27)
Prefer not to answer	2 (3)
Race	
White	61 (97)
People of color	2 (3)
Marital status	
Married	26 (41)
Living common law	20 (32)
Single	10 (16)
Widowed	3 (5)
Divorced or separated	4 (6)
Education level	
No certificate, diploma, or degree	3 (5)
High-school diploma or equivalency certificate	17 (27)
Apprenticeship, trades certificate, or diploma	6 (10)
College, CEGEP, or other nonuniversity certificate or diploma	20 (32)
University certificate or diploma below bachelor level	5 (8)
University diploma or degree at bachelor level or above	12 (19)
Employment	
Full-time	18 (29)
Part-time	6 (10)
Off work for a short or undetermined term	8 (13)
Off work for a long term or disabled	12 (19)
Unemployed	2 (3)
Retired	16 (25)
Other	1 (2)
Household income	
CAD 0-$49,999 (US $0-$37,257)	19 (30)
CAD $50,000-$99,999 (US $37,258-$74,515)	24 (38)
CAD ≥$100,000 (US ≥$74,516)	12 (19)
Do not know	2 (3)
Prefer not to answer	6 (10)
Duration of chronic pain (years)	

Sociodemographic characteristics	Values, n (%)
1 to 5	26 (41)
6 to 10	15 (24)
11 to 15	6 (10)
16 to 20	5 (8)
≥21	11 (17)
Type of chronic pain	
Chronic widespread pain (includes fibromyalgia syndrome)	24 (38)
Complex regional pain syndrome	8 (13)
Chronic musculoskeletal pain (eg, cervical pain and low back pain)	34 (54)
Chronic headache and orofacial pain (eg, migraine)	10 (16)
Chronic visceral pain	5 (8)
Chronic neuropathic pain	9 (14)
Inflammatory arthritis (eg, rheumatoid arthritis and psoriatic arthritis)	8 (13)
Osteoarthritis	22 (35)
Other	6 (10)

[a]CEGEP : College of General and Professional Teaching.

Acceptability

Among the 63 participants, 38 (60%) completed the Acceptability eScale. The average total score for these participants was 25.2 out of 30 (acceptability threshold=24). In Table 4, we have presented the integrated results of the quantitative and qualitative phases. The mean score for each of the 6 items from the Acceptability eScale is listed in the first column. Each item score ranges from 1 to 5, with higher scores reflecting higher acceptability. The interviews revealed a consistent theme: a program that is globally acceptable, easy to use, and engaging. Interviewees felt they could opt for what worked for them once they had experienced it completely. They would keep most of the program as is, with some potential minor improvements.

Table 4. Joint display of participants' perceptions of the Agir pour moi program's acceptability.

Quantitative results	Qualitative findings	Integrated analysis

Mean score of the ease-of-use item: 4.0/5 (SD 0.9); 5 (8%) of the 63 participants needed support at least once to log in	• Ease of navigation
	  ◦ "Everything is very easy. We can go forward, we can go back, we can resume, we can close, come back. No, everything is perfect." [INT[a] 7]
	• User-friendly, except for the platform hosting the web-based program
	  ◦ "As I was saying, when you're a participant, and you're logged in, I think it is super user-friendly ... It's more when you get to the big link page, I believe that's a little less user-friendly." [INT 13]
		The program was user-friendly once logged in, but connecting to the hosting platform could have been more intuitive.

Mean score of the strategies' helpfulness item: 4.1/5 (SD 0.9)	• Picking the tool you need
	  ◦ "That's it, because if you don't have that resource, what do you do? On a day when things aren't going well, you brood, you grumble all day long; you're in pain, you're angry, and you lose your patience, so if you think a little bit after having followed the program, you go back and find the tool you needed in it. Because as far as I'm concerned, the program wasn't one singular tool; it was a toolkit, and then you take what you need." [INT 10]
	  ◦ "I was happy to see all this in the program, because it helped me a lot personally." [INT 8]
	• Adaptable to one's situation
	  ◦ "It's good to have all the suggestions in there because, ultimately, we keep what suits us and works well for us. So there's something for everyone, for everyone's tastes, and it's not the same stuff that works for everyone." [INT 12]
	• Developing further the strategies that suit you best
	  ◦ "It's been like a springboard to the rest of my journey, to do some reading... ... it sure helps with the perception of managing what we're capable to manage on our own." [INT 6]
		Some strategies benefited many participants, while others only reached a minority. Interviewees latched on to at least 1 lesson that triggered something in them. They acquired something useful out of it, although not all appreciated the same strategy or strategies.

Mean score of the required time item: 4.2/5 (SD 0.7)	• Objective-dependent efforts
	  ◦ "I didn't think it was very demanding, and after that, the rest is up to you... then you do it at your own pace, so I thought it was very appropriate." [INT 13]
	• Motivating and minimal effort required
	  ◦ "Oh, for me, it wasn't that much effort, no. Perhaps the first week's lessons were a little longer than the others that followed. But no, it's really not...it didn't take me any effort, no. It was motivating." [INT 7]
	• A little demanding yet tangible
	  ◦ "I think that's what makes it interesting. Because just reading is okay, but afterward, when you do the exercise, you push yourself a bit on the spot; sometimes, you don't really feel like doing it, but it makes it more tangible." [INT 11]
		The time and effort required to follow the program and apply self-management strategies were adequate, but interviewees highlighted the necessity to reach a certain degree of readiness to change because taking action required some investment.

Quantitative results	Qualitative findings	Integrated analysis

Mean score of the use appreciation item: 4.3/5 (SD 0.7)	• Adapting use to your schedule
	  • "As I said, we could follow it whenever we wanted. I chose the time of the day when I was in better shape. That was something I thought was very nice, you know, not having to connect at a fixed time." [INT 15]
	• Staying engaged
	  • "It went really well. Then, you know, you'd scroll, then you'd click, scroll, click, so that, you know, it kept you present; you had to be there... It's not like a video you start and then lose focus. That's very interesting too." [INT 11]
	• Fun program
	  • "it's super fun." [INT 13]
		Participants appreciated navigating the program at their convenient time. It was fun to use, and the need to scroll through the content kept them engaged.

Mean score of the comprehensibility item: 4.5/5 (SD 0.6)	• Very simple explanation • "It's very...it's simple. It's very well explained." [INT 7]	The program presented well-explained, easy to understand information.

Mean score of the satisfaction item: 4.1/5 (SD 0.7)	• Participants would recommend the program
	• "I really enjoyed the program. I liked that there were videos, it made it more dynamic, I thought they were well done, well constructed ... I'd definitely recommend it." [INT 12]
	• Down-to-earth expectations, no promises
	  • "I thought the program was interesting... Knowing what's in it, I'd do it again today... In the beginning, you don't make any promises. In the program, it says 'learning to live with chronic pain', but there's no promise; it doesn't say: 'Hey, when you get to the eighth week, you're pain-free'." [INT 10]
	  • "But I think the program doesn't apply to me... You know, it's like, geez, I thought I would discover something amazing. It's been four years now, I've seen three internists, lots of doctors, and I went to a maxillofacial specialist back in September, and we're trying to figure it out, but nobody knows what it is. I'm still waiting to find out." [INT 2]
		All interviewees mentioned they would recommend the program to someone in a similar situation, reflecting satisfaction. Participants with high expectations for very specific problems might have been less satisfied.

Total mean score of the Acceptability eScale[b]: 25.2/30 (SD 3.0)	• No change required
	  • "Well, it went well. I liked it a lot, I really liked the way it was put together, the way it was presented, gradually if you like. The program is super well done, I'm going to revise it, but I wouldn't change a thing if you asked me if there was anything to change." [INT 1]
	• Use of multiple media
	  • "I loved it because there were testimonials. It wasn't just reading. I put on my headphones, and I didn't have to read. I listened. I like to listen, and then as I filled out my sheet, I'd make notes on the paper as I listened. Sometimes I'd take breaks, put it on pause and come back." [INT 9]
	• Promoting access for everyone
	  • "God, yes, it's acceptable, and everyone should have access to it." [INT 7]
		Overall, the program was well developed for our target users and proposed an appreciated gradual approach to the application of different strategies.

Of the 15 interviewees, 10 (67%) were still accessing the program or using the personal plan on an occasional to regular basis for >5 months after they finished it for the first time.		While the web-based format did not suit everyone, participants appreciated using it whenever they wanted and having the possibility to go back to previous lessons for a refresher. Not traveling to learn self-management strategies was a significant advantage. Had the program been offered through weekly classes at a specific time and place, some participants would have been unable to drive and attend.

Quantitative results	Qualitative findings	Integrated analysis
	• Web-based content accessibility • "It helps a lot that it's online, so you can do it when it suits you. Because for me, if I'd had to travel, you know, appointments and things like that, I'd have had a lot more trouble because it's hard for me to go on the road, whereas here, it was much more accessible, which is really great." [INT 12] • Possible punctual use once the program is done • "Then again, it's not a program that once it's done, it's done, and you can't go back to it; you can go back and look. So you can continue to use it." [INT 10]	

[a]INT: interviewee.

[b]The Acceptability eScale has 6 items with a total score ranging from 6 to 30. Higher scores represent a high level of acceptability.

Engagement

Of the 63 participants who consented to the study, 46 (73%) started the program, 29 (46%) finished at least 6 weeks' lessons, and 26 (41%) completed all the lessons (Table 5).

Participants generally followed the lessons in order. Among the 19 noncompleters who did not withdraw from the study, 10 (53%) wrote back to us after receiving the email reminder. These participants provided personal reasons for not following the program according to schedule, including sickness or mental health issues among close relatives, the death of a parent, estate management, and increased symptoms. Including participants who withdrew from the study, 21 (33%) out of 63 participants were not connecting weekly or stopped at some point for reasons external to the study or the intervention. Reasons related to the program included having already applied the proposed strategies and having trouble connecting. Of the 63 participants, 5 (8%) needed help logging in independently (n=3, 60% only required information by email and n=2, 40% received phone support and succeeded with a step-by-step explanation but did not connect afterward).

Table 5. Participants' weekly lessons' level of completion (N=63).

Weekly lessons	Completed the lesson, n (%)	Partially completed the lesson, n (%)	Never initiated the lesson, n (%)
Lesson 1	42 (67)	4 (6)	17 (27)
Lesson 2	37 (59)	6 (10)	20 (32)
Lesson 3	33 (52)	6 (10)	24 (38)
Lesson 4	32 (51)	3 (5)	28 (44)
Lesson 5	30 (48)	5 (8)	28 (44)
Lesson 6	29 (46)	3 (5)	31 (49)
Lesson 7	28 (44)	4 (6)	31 (49)
Lesson 8	26 (41)	4 (6)	33 (52)

Interviewees mentioned other factors potentially undermining the engagement in completing the program. Interviewees highlighted that timing and pain acceptance played an essential role in their perseverance or lack thereof. For example, understanding that no specific cause for her pain issue might ever be found was a turning point in a participant's proactivity:

> *When I was still at the stage where I thought we would find a cause, it is something I really wouldn't have been ready to do, the program... Doing the program, really, it sort of happened right at the time I got to the stage of realizing: 'Okay, now I'm going to stay that way. What am I going to do with this?' I felt a lot of psychological distress. Then, looking more into this (psychological) aspect through the program, it was as if I needed to do this. I'm not done, but it's come a long way. And also, time matters; getting used to accepting.* [INT 11]

For some interviewees, occasional or regular coaching would have been necessary to sustain motivation, answer questions, and revise their personal plan. Of the 63 participants, 2 (3%) mentioned their need for support and feedback:

> *There were ups and downs. There was no one to answer my questions. I found that very, very hard, especially the first few weeks.* [INT 5]
>
> *I really need someone who'll say, "Go, we'll do this, we'll do that." In writing, I'll read, and I'll say to myself, "Oh my god, that's wonderful," but I'll do it 2 or 3 times, and then I'll give up... But I've always been like that; I've always needed someone to push me, not in everything, but especially since I've been ill.* [INT 9]

There are participants who related less to parts of the program because their activity level was not adequately represented:

> *We always assume that people aren't physically active when in pain. Or if they are, clearly, it can't be too much. But you know, on my part, I will get injured before I stop [laughs].* [INT 11]

Furthermore, interviewees described factors mostly contributing to their engagement in completing the program. They expressed a connection to the program's content and felt more hopeful, supported, and less alone listening to the multiple integrated testimonials of individuals with lived experiences applying the strategies:

> *For me, it was seeing lots of testimonials from many people ... that's what really attracted me. To understand it better and to see that that's how it is: everyone's gone down exactly the same path I went through, and we all arrive at the same point, not being able to get out, not seeing anyone. It seemed like it was just me who was going through this. So, it gave me a boost. It also gave me a bit of confidence... I had the impression of being accompanied.* [INT 14]

Moreover, a participant living far from a major urban area highlighted the appeal of these short videos:

> *Here, in my neighborhood, there's no program like that. I know there are places, there are meetings for people with chronic pain to chat, have a coffee, and things like that. We don't have that here. I found it fun to listen to them and know we're not alone in this.* [INT 9]

On the one hand, interviewees thought the weekly connection and review of personal objectives supported their motivation and helped them stay focused:

> *I guess I needed to be held by the hand for a while, and to be guided through it... There was some kind of follow-up. So you weren't left to your own devices as much.* [INT 12]

On the other hand, it might have been too much for some individuals:

> *There were too many things. Every week there was something to do, you know. You didn't have time to swallow the information and were already moving on to other things... Maybe it would have taken a week or two between each chapter. What made it easier was to stop for a few weeks, then think about all the information.* [INT 5]

There is a fine line between what feels like a comprehensive program and what feels like an overwhelming task for unsupported participants. While most interviewees showed interest in all the presented topics and perceived the lessons' sequence as logical, gradual, and positive, specific strategies within these lessons might have lacked appeal to some participants. A suggestion was made that participants could start with their most appealing lessons after covering the pain education section. While we invited everyone to complete the tasks in order, the content was freely available and one could decide to skip parts of the program if desired.

Through the interviewee discourse, we could distinguish microlevel engagement, including the number of lessons they have followed, from macrolevel engagement, referring to the depth of involvement with the behavior change process, such as applying strategies consistently in a real-world setting [55,56]. Behavior change could be challenging, with participants facing expectations from others and their own. However, the program provided them with a better understanding of why it is beneficial to do so, and some participants made it a priority. Getting into the habit of doing something sometimes required broader or prior changes such as setting boundaries with the extended family. The notion of beginning with the easiest tasks was mentioned to underscore the initial challenge of implementing a strategy. Interviewees described the integration of certain habits into their routine as gradual, sometimes evolving over several months, without the participants realizing it. Over 5 months after finishing the program, interviewees described what they continued to apply and how some strategies became part of their routine:

> *Over the weeks, I really selected what had a positive impact on me. Then I do it regularly, almost every day. I'm into meditation, cardiac coherence, managing energy, stretching, and the gratitude journal. The other things, I can't think of anything else I could do more.* [INT 6]

> *I still do so to this day, which is unusual for me. So, I thought it was really, really good... I'm more active now too, I do my exercises almost daily. I used to say to myself, "It's no use, I won't be able to do it. I don't have the energy. I'm in pain. I'm out of shape." Now I say, "Look, do it, even if it's just two minutes today, it'll at least be two minutes, you'll have done it."[INT 12]*

Preliminary Effects Outcomes

We observed an indication of improvements over time in self-efficacy, pain interference, depression, and pain catastrophizing between baseline and postintervention (Table 6). The results of the linear model corrected for the individual (treatment effects) are presented in Table 7. In addition, 24 (56%) out of 43 participants reported some level of improvement on the Patient Global Impression of Change Scale after the intervention.

From a qualitative perspective, 14 (93%) out of 15 interviewees reported a different set of effects 5 to 7 months later. These effects included, among others, shorter and less frequent pain attacks, better management of pain, higher sense of control, less comparison with life before pain, improved psychological state, more patience, less frustration and irritability, forgiveness toward oneself, higher activity level, more self-care, and increased social life.

In Table 7, we have presented the integrated results of the quantitative and qualitative phases, with the overarching theme that the various effects observed were potentially influenced by macrolevel engagements.

The effects of the lifelong task of self-management might become noticeable over time. A participant mentioned recognizing these effects several months after completing the program:

> *While I was going through (the program), well, you know, it was okay. In my case, it's really like it slowly permeated me, but in a positive way, I mean.* [INT 11]

Therefore, conducting the interviews 5 to 7 months after program completion was deemed appropriate.

Table 6. Mean scores for outcome measures at baseline, postintervention, and 3-month follow-up.

Measure	Subscale or construct	Baseline (N=63), mean (SD)	Postintervention (n=43), mean (SD)	3-month follow-up (n=34), mean (SD)
Pain Self-Efficacy Questionnaire[a]	Self-efficacy	26.9 (13.8)	32.4 (13.3)	32.5 (12.4)
Brief Pain Inventory	Pain intensity	5.5 (1.6)	4.8 (1.5)	5.0 (1.6)
Brief Pain Inventory	Interference	5.7 (2.0)	4.4 (2.0)	4.7 (1.9)
Hospital Anxiety and Depression Scale	Anxiety	9.0 (4.4)	8.4 (4.2)	8.2 (4.0)
Hospital Anxiety and Depression Scale	Depression	8.5 (4.0)	7.2 (3.9)	7.5 (3.8)
Pain Catastrophizing Scale	Catastrophizing	25.3 (10.9)	20.1 (12.2)	20.5 (13.2)

[a]A higher score in the Pain Self-Efficacy Questionnaire indicates a high level of self-efficacy.

Table 7. Joint display of participants' perceptions of the Agir pour moi (APM) program's effects.

Quantitative results	Qualitative findings	Integrated analysis

• Pain Self-Efficacy Questionnaire: the scores increased by 3.06 points on average after intervention (95% CI 1.26 to 4.85), *P*=.002.
• No effects were detected when examining postintervention and follow-up metrics.
	• Better self-efficacy
	  • "I'm less inclined to compare myself with my previous life. I see more of what I'm capable of doing today with the abilities I have. That's the main thing, I think." [INT[a] 12]
	  • "Even though I've seen specialists who gave me medication, in the end, I think I got better by doing this on my own and saying to myself, Okay, I'm basically taking charge... I feel like I can control my pain peaks a bit more, and I know why I will have them." [INT 4]
	  • "Let's just say I've relearned how to gain confidence in myself and then say you're capable, go ahead, go take a walk, you can do it." [INT 8]
		Participants' belief in their capacity to do certain things to achieve their goal increased.

• BPI[b] pain intensity subscale scores decreased by 0.32 points on average after intervention (95% CI –0.55 to –0.10), *P*=.007.
• No effects were detected when examining postintervention and follow-up metrics.
	• No trend in pain intensity changes
	  • "I can say that my pain is less present and that it's not all I think about anymore. So yes, for my pain, it helped a lot." [INT 5]
	  • "It's just that instead of being in pain for nine hours at a time, well, not only have I hardly had any attacks for six, seven months, but when it happens, well the two times it happened, it lasted two, three hours or so, then it's stopped really suddenly, instead of lasting a whole night. So I tell you, it's not so bad." [INT 14]
	  • "But the pain remains the same. Sometimes when I'm doing your stuff, I've managed to get away for half an hour or an hour, but it comes right back." [INT 3]
	  • "I've done things, like mediation and all that, but it doesn't work." [INT 2]
		There were improvements for some interviewees and no improvements for others. While not measured with the BPI subscale, the frequency of pain crises decreased in some participants. However, as mentioned straightforwardly at the beginning of the APM, pain reduction is not the main objective of a self-management program.

• BPI pain interference subscale: scores decreased by 0.78 points on average after intervention (95% CI –1.19 to –0.37), *P*=.001.
• No effects were detected when examining postintervention and follow-up metrics.
	• Less interference
	  • "I realize that the energy is coming back a little. It's coming back. It's not a complete crash like it used to be for me, and then I'd take a week to recover because I'd gone over the edge. Now, I don't do that anymore ... And in the end, I do a lot more than I used to do that way." [INT 12]
	  • "You know, I was more or less able to do some movements, some I couldn't do at all anymore. There are some that I've gradually managed to recover a little, it's not to the maximum here, but ... like putting on my shoes." [INT 15]
	  • "Well, I fall asleep faster when I do these exercises. Because I always try to go to bed at the same time, and sometimes sleep doesn't come. So I tell myself I'm going to bed anyway and do some breathing exercises. Then I fall asleep, which doesn't take long." [INT 15]
		Reducing pain interference translates into a gain in energy, a more stable ability to perform tasks, a new capacity to do movements, increased social activities, and better sleep.

Quantitative results	Qualitative findings	Integrated analysis

• HADS[c] depression subscale: scores decreased by 0.73 points after intervention (95% CI –1.21 to –0.25), *P*=.005.
• No effects were detected when examining postintervention and follow-up metrics.
	• Positive change in depressive state
	  • "It had been seven years since I'd stopped putting any effort into it and let myself fall, so it...no, no, it whipped me, and then I seemed to become a bit like myself again. I was letting myself go, then it was like: okay, go, I've been sinking for seven years, and now it's time to get back on. It gave me a good boost. ... My mood has changed. I seem to be less, sorry about my French, but I'm less (swear) angry all day long." [INT 14]
	• Better mood
	  • "how can I put it, patience, my patience came back, better than when I couldn't do my things. Yes, yes, that, I've made some gains." [INT 15]
	  • "It's also psychological, you know, like being less on edge in my head, I was a lot like (swears), I can't do this anymore, I can't. So I'm in a better mood with the kids. ... It also has a lot to do with irritability, because you know when you're in pain, you're always irritable, so if I'm in less pain, I'm less irritable, and I'm more likely to want to go outside with them." [INT 11]
		Most interviewees did not explicitly talk about depression but many reported being in a better mood, being less frustrated, and being less irritable.

• HADS anxiety subscale: nonsignificant score decrease of 0.37 points after intervention (95% CI –0.85 to 0.12), *P*=.14. • No effects were detected when examining postintervention and follow-up metrics.	• Anxiety • "it's really taken my anxiety level about it down a notch." [INT 11]	Following the program does not appear to impact the anxiety state. Only 1 interviewee briefly mentioned a decrease.
• Pain Catastrophizing Scale: scores decreased by 2.83 points after intervention (95% CI –4.35 to –1.30), *P*=.001. • No effects were detected when examining postintervention and follow-up metrics.	• Less panic • "For me, what was a real game changer ... was also realizing that most people in this condition tend to have the same thoughts; thinking, for example, that there's something perhaps serious hidden behind it, thinking that it will never stop, that you could die from it, and all that. It was good for me, because these are patterns I really have. Then, I thought of myself more like a normal person. ... The panic I used to feel about my pain has almost completely disappeared, and I realize how much I can change my situation myself." [INT 11]	The testimonials and theoretical content helped normalize some participants' catastrophic thoughts and guide them in confronting those and adopting more adapted views. Following the program can reduce catastrophizing.

[a]INT: interviewee.

[b]BPI: Brief Pain Inventory.

[c]HADS: Hospital Anxiety and Depression Scale.

Discussion

Principal Findings

This paper described a pilot, mixed methods study assessing feasibility and acceptability and exploring the preliminary outcomes of a new self-directed, web-based program for chronic pain among adults awaiting superspecialized services. Collaborating with patient partners experiencing diverse types of pain was a key strategy enabling us to align the program closely with the varied needs and expectations of our wide-ranging audience. We opted for a simple yet attractive layout, providing clear instructions and features to aid those with attention and concentration challenges.

Feasibility, Acceptability, and Engagement

Chronic noncancer pain affects individuals in different ways and to different degrees. Those awaiting tertiary services in Canada experience severe impairments and present with a poor biopsychosocial profile [2]. To specifically recruit these individuals who are not yet patients at the center of expertise in chronic pain management, we could not directly reach out to them. The hospital's archives had to send them invitation letters. Estimating the response rate in such a specialized context proved challenging because we lacked a benchmark for our expectations. However, a study by Thiblin et al [57], which involved a comparable internet-administered, CBT-based self-help intervention, achieved an 11% enrollment rate by sending out 509 invitation letters. Considering the documented high dropout rates in similar trials [19,58], we anticipated that 500 invitations would suffice to ensure a minimum of 30 participants responding to questionnaires at all 3 time points. Our enrollment rate for a 3-month recruitment period was similar to or better than those of studies recruiting in tertiary pain treatment facilities [59,60]. The consent rate among potentially eligible adults was acceptable. However, because 4 (6%) out of 63 participants exited the study for surgery purposes and 3 (5%) others for significant changes in their medication, we might need to reconsider how we address these eligibility criteria during the phone interview. Furthermore, considering there are between 1000 and 1500 individuals awaiting services from the center of expertise and that they all would not meet our eligibility criteria, our response rate, although similar to those observed in other studies, would require conducting a future full trial across multiple centers.

Our study yielded comparable data collection results to other web-based intervention pilot trials with approximately 50% (35/63) of response at 3-month follow-up [61,62]. While studies with higher financial incentives (US $25-$80 per assessment) during visit assessments or initial motivational interviews had better response rates [60,63-66], we purposefully chose to stay as close as possible to real life where no incentives are offered. We received no negative feedback on the number or length of questionnaires. However, sending email reminders to participants who did not log in or complete the questionnaires at the appropriate time was suboptimal. Making phone calls in addition to email reminders might have provided us with reasons for disengagement and ensured participants received the reminders.

Some interviewees highlighted that the log-in process was not intuitive, leading us to consider modifying this aspect before conducting a full trial. The log-in was essential to the research project but is not part of APM itself and will not apply in real life. Once logged in, accessing APM at their most convenient time and place was a significant asset for our participants, consistent with patients' preferences [21].

APM's codevelopment with health care professionals and people with lived experience of chronic pain allowed a tailored approach to this population's needs and preferences in a web-based self-management program [32]. The interviewees' description of APM aligned with the acceptability score as a globally acceptable, easy-to-use, and engaging program. We based our 24 (80%) out of 30 threshold score for the Acceptability eScale based on the study by Tariman et al [38], suggesting that 80% of the highest possible summary score indicates good program acceptability. However, a score <24 would not automatically deem the program unacceptable. We must examine individual item scores to assess specific program weaknesses. All items scored ≥4 (≥80%) out of 5, as show in Table 4, indicating no significant flaws requiring major modifications. Minor improvements we could make include adding testimonials from highly active individuals with chronic pain who learned to pace themselves and mentioning that lessons are preferably followed in order but can also be explored based on personal preferences after completing week 1. APM effectively promoted behavioral change, guided participants in taking action, and served as a reference in the longer term.

While participants' weekly lessons' level of completion was lower than anticipated, it was consistent with what had been observed in other feasibility and pilot studies [62,67-69] and could be explained mainly by reasons external to the study. Our qualitative results alleviated concerns about potential flaws and did not point toward questioning the participant's appeal to the program. Overall, we are confident that the program and trial procedures are both feasible and acceptable.

Preliminary Effects

We explored pre- and postintervention effects as preliminary indications of potential changes in self-efficacy, pain intensity and interference, anxiety, depression, and catastrophizing. Findings yielded relevant results, but these should be interpreted cautiously.

Nevertheless, the qualitative interviews pointed in the same direction as our preliminary quantitative findings. Furthermore, these aligned with the results of a meta-analysis suggesting that following internet-delivered, CBT-centered interventions for chronic pain can lead to small significant improvements in pain interference and intensity, depression, anxiety, self-efficacy, and catastrophizing, with greater treatment effects in anxiety, pain interference, and intensity in guided compared to unguided interventions [34]. Therefore, depending on their perceived importance of change and self-efficacy, individuals with chronic pain may require additional support in reaching readiness to make sustainable changes.

Because we still did not know precisely what clinical, intervention, and study characteristics positively impacted the effects of unguided CBT-based self-management programs for chronic pain, APM offered several self-management strategies [70]. However, we did not expect participants to implement all of them once they completed the program. Participants used this program as a toolbox, as mentioned by an interviewee.

No adverse events were reported throughout the course of this study.

Limitations

This study has some limitations. First, we cannot make definitive statements regarding APM's effects without an appropriate control group, randomization process, and sample size. Indeed, we neither designed nor appropriately powered this feasibility trial to test a specific hypothesis [36]. Furthermore, while the participants presented various pain conditions, most of the female participants were White, as this is the case in similar trials [18,71,72], and were all attached to a single center. Recruitment strategies of a future randomized controlled trial should focus on attracting a broader representation of individuals with chronic pain in terms of gender, ethnicity, and health care institutions. In Quebec City, where our study was conducted, <10% of the population identified as members of a visible minority group in 2021. Expanding our research to cities with greater ethnic diversity could enhance our sociodemographic data and improve the relevance of our findings. Shifting to web-based recruitment methods might allow us to create tailored invitation messages for specific demographic groups, using a casual design and images, instead of overwhelming potential participants with excessive written information. In our feasibility study, we had to use standardized letters to recruit participants from the waitlist of the center of expertise in chronic pain management. We acknowledge the need to adopt more flexible parameters for future large-scale studies. Furthermore, we might adjust our eligibility criterion, not limiting participation to those awaiting specialized services. This broader criterion could yield a more diverse sample, aligning with our aim to reach a wider demographic. However, we are mindful of the potential impacts on adherence and user satisfaction this broader criterion might pose. Nevertheless, these adjustments reflect our dedication to conducting a comprehensive and inclusive trial, ultimately contributing to a more nuanced understanding of chronic pain management. This may lead us to unforeseen modifications in our program.

Despite widespread internet access in Canada, disparities in internet speed, affordability, and digital literacy persist. APM, being exclusively web-based, poses a limitation in reaching individuals from remote regions and Indigenous communities as well as those in low-income households, older adults, and individuals with disabilities. These groups are disproportionately affected by the digital divide in Canada, making it challenging for them to access our program.

Future Direction

This study provides an initial understanding of APM's potential benefits for this group of individuals with chronic pain awaiting specialized services. Through the interviews, we acknowledged we had not captured the effects on the temporal aspect of pain, such as shorter and less frequent pain attacks, which were crucial for some participants. Therefore, we could consider adding measures capturing these aspects in a future trial.

Developing a web- and evidence-based, patient-centered, free-of-charge, user-friendly, and French self-management program for chronic noncancer pain represents a potential response to the clearly expressed needs of individuals with this condition. Although the literature increasingly emphasizes the importance of personalization in eHealth, our limited financial resources hindered us from incorporating advanced features. We deliberately chose to focus on fundamental aspects and prioritize what we could offer and support in the long term, establishing the groundwork for a web-based program that could potentially evolve. As a result, the current version did not include personalized features, but it was still perceived as usable and useful. It will be essential to document how the program's implementation makes it possible to respond quickly and more equitably to some of the needs of patients waiting for services or who live far from large centers. APM is currently being used without restrictions in other French-speaking regions and countries. Anyone can use it freely, but a potential hurdle faced when using it abroad pertains to adapting to the accents in testimonial videos and Quebec-specific expressions.

Conclusions

The study findings provided preliminary evidence that the APM program and research methods were both feasible, as suggested by perceived acceptability and engagement. Furthermore, it provided preliminary indications of potential improvements in self-efficacy, pain intensity, interference, depression, and catastrophizing. The study yielded essential results to undertake a future complete trial.

Acknowledgments

The authors would like to thank Sergio Cortez Ghio, a biostatistician who performed the quantitative analysis. The authors would also like to extend their thanks to all those who contributed to the codevelopment of APM. PMD acknowledges support from a Frederick Banting and Charles Best Canada Graduate Scholarship Doctoral program from the Canadian Institutes of Health Research, a Leadership and Sustainable Development scholarship from Laval University, and a scholarship from the Strategic Research Development Fund from VITAM—Center for Sustainable Health Research. AMP holds the Chair in Educational Leadership on Chronic Pain—Medisca, which contributed to the development of the program. Additional funding for this study came from discretionary funds. The funders did not play any role in the design, management, analysis, interpretation, or writing of the manuscript.

Data Availability

The data sets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

PMD, AMP, and AL contributed to the study conception and design. PMD and AMP supported the details of recruitment and data collection. PMD, AL, and LJ analyzed the results. PMD and AL drafted the manuscript. All authors commented on and revised its previous versions before reading and approving the manuscript to be published.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1181 KB - humanfactors_v11i1e50747_app1.pdf]

References

1. Ndegwa S, Dolcine B, Frey N, Quay T. Access to and availability of non-pharmacological treatments for chronic non-cancer pain in Canada: an environmental scan. Canadian Agency for Drugs and Technologies in Health (CADTH). 2018 Nov. URL: https://www.cadth.ca/sites/default/files/pdf/es0313_access-to-non-pharmacological-treatment-of-pain.pdf [accessed 2024-04-18]
2. Choinière M, Dion D, Peng P, Banner R, Barton PM, Boulanger A, et al. The Canadian STOP-PAIN project - part 1: who are the patients on the waitlists of multidisciplinary pain treatment facilities? Can J Anaesth 2010 Jun;57(6):539-548. [doi: 10.1007/s12630-010-9305-5] [Medline: 20393821]
3. Dueñas M, Ojeda B, Salazar A, Mico JA, Failde I. A review of chronic pain impact on patients, their social environment and the health care system. J Pain Res 2016;9:457-467 [FREE Full text] [doi: 10.2147/JPR.S105892] [Medline: 27418853]
4. Burke AL, Mathias JL, Denson LA. Psychological functioning of people living with chronic pain: a meta-analytic review. Br J Clin Psychol 2015 Sep;54(3):345-360. [doi: 10.1111/bjc.12078] [Medline: 25772553]
5. Moriarty O, McGuire BE, Finn DP. The effect of pain on cognitive function: a review of clinical and preclinical research. Prog Neurobiol 2011 Mar;93(3):385-404. [doi: 10.1016/j.pneurobio.2011.01.002] [Medline: 21216272]
6. Reid KJ, Harker J, Bala MM, Truyers C, Kellen E, Bekkering GE, et al. Epidemiology of chronic non-cancer pain in Europe: narrative review of prevalence, pain treatments and pain impact. Curr Med Res Opin 2011 Feb;27(2):449-462. [doi: 10.1185/03007995.2010.545813] [Medline: 21194394]
7. Ashton-James CE, Anderson SR, Mackey SC, Darnall BD. Beyond pain, distress, and disability: the importance of social outcomes in pain management research and practice. Pain 2022 Mar 01;163(3):e426-e431 [FREE Full text] [doi: 10.1097/j.pain.0000000000002404] [Medline: 34252908]
8. Deslauriers S, Roy JS, Bernatsky S, Blanchard N, Feldman DE, Pinard AM, et al. The burden of waiting to access pain clinic services: perceptions and experiences of patients with rheumatic conditions. BMC Health Serv Res 2021 Feb 18;21(1):160 [FREE Full text] [doi: 10.1186/s12913-021-06114-y] [Medline: 33602224]
9. Blanchard N, Deslauriers S, Gervais-Hupé J, Hudon A, Roy JS, Bernatsky S, et al. "It feels like an endless fight": a qualitative study exploring healthcare utilization of persons with rheumatic conditions waiting for pain clinic admission. BMC Musculoskelet Disord 2022 Sep 22;23(1):878 [FREE Full text] [doi: 10.1186/s12891-022-05808-6] [Medline: 36131335]
10. Burke AL, Mathias JL, Denson LA. Waiting for multidisciplinary chronic pain services: a prospective study over 2.5 years. J Health Psychol 2020 Aug;25(9):1198-1212. [doi: 10.1177/1359105317752828] [Medline: 29322830]
11. Choinière M, Peng P, Gilron I, Buckley N, Williamson O, Janelle-Montcalm A, et al. Accessing care in multidisciplinary pain treatment facilities continues to be a challenge in Canada. Reg Anesth Pain Med 2020 Dec;45(12):943-948. [doi: 10.1136/rapm-2020-101935] [Medline: 33024007]
12. Campbell F, Hudspith M, Choinière M, El-Gabalawy H, Laliberté J, Sangster M. Working together to better understand, prevent, and manage chronic pain : what we heard / a report by the Canadian pain task force. Health Canada. 2020. URL: https://publications.gc.ca/site/eng/9.892793/publication.html [accessed 2024-04-05]
13. Campbell F, Hudspith M, Anderson M, Choinière M, El-Gabalawy H, Laliberté J, et al. Chronic pain in Canada: laying a foundation for action: a report of the Canadian pain task force. Health Canada. 2019. URL: https://www.canada.ca/content/dam/hc-sc/documents/corporate/about-health-canada/public-engagement/external-advisory-bodies/canadian-pain-task-force/report-2019/canadian-pain-task-force-June-2019-report-en.pdf [accessed 2024-04-05]
14. Devan H, Perry MA, van Hattem A, Thurlow G, Shepherd S, Muchemwa C, et al. Do pain management websites foster self-management support for people with persistent pain? A scoping review. Patient Educ Couns 2019 Sep;102(9):1590-1601. [doi: 10.1016/j.pec.2019.04.009] [Medline: 30981410]
15. Buhrman M, Syk M, Burvall O, Hartig T, Gordh T, Andersson G. Individualized guided internet-delivered cognitive-behavior therapy for chronic pain patients with comorbid depression and anxiety: a randomized controlled trial. Clin J Pain 2015 Jun;31(6):504-516. [doi: 10.1097/AJP.0000000000000176] [Medline: 25380222]
16. Irvine AB, Russell H, Manocchia M, Mino DE, Cox Glassen TC, Morgan R, et al. Mobile-Web app to self-manage low back pain: randomized controlled trial. J Med Internet Res 2015 Jan 02;17(1):e1 [FREE Full text] [doi: 10.2196/jmir.3130] [Medline: 25565416]
17. Dear BF, Gandy M, Karin E, Staples LG, Johnston L, Fogliati VJ, et al. The pain course: a randomised controlled trial examining an internet-delivered pain management program when provided with different levels of clinician support. Pain 2015 Oct;156(10):1920-1935 [FREE Full text] [doi: 10.1097/j.pain.0000000000000251] [Medline: 26039902]
18. Bromberg J, Wood ME, Black RA, Surette DA, Zacharoff KL, Chiauzzi EJ. A randomized trial of a web-based intervention to improve migraine self-management and coping. Headache 2012 Feb;52(2):244-261 [FREE Full text] [doi: 10.1111/j.1526-4610.2011.02031.x] [Medline: 22413151]
19. Moman RN, Dvorkin J, Pollard EM, Wanderman R, Murad MH, Warner DO, et al. A systematic review and meta-analysis of unguided electronic and mobile health technologies for chronic pain-is it time to start prescribing electronic health applications? Pain Med 2019 Nov 01;20(11):2238-2255. [doi: 10.1093/pm/pnz164] [Medline: 31386151]

20. Moore SK, Guarino H, Acosta MC, Aronson ID, Marsch LA, Rosenblum A, et al. Patients as collaborators: using focus groups and feedback sessions to develop an interactive, web-based self-management intervention for chronic pain. Pain Med 2013 Nov;14(11):1730-1740 [FREE Full text] [doi: 10.1111/pme.12200] [Medline: 23859438]
21. Ledel Solem IK, Varsi C, Eide H, Kristjansdottir OB, Mirkovic J, Børøsund E, et al. Patients' needs and requirements for eHealth pain management interventions: qualitative study. J Med Internet Res 2019 Apr 01;21(4):e13205 [FREE Full text] [doi: 10.2196/13205] [Medline: 30877780]
22. Banner D, Bains M, Carroll S, Kandola DK, Rolfe DE, Wong C, et al. Patient and public engagement in integrated knowledge translation research: are we there yet? Res Involv Engagem 2019;5:8 [FREE Full text] [doi: 10.1186/s40900-019-0139-1] [Medline: 30805202]
23. Karazivan P, Dumez V, Flora L, Pomey MP, Del Grande C, Ghadiri DP, et al. The patient-as-partner approach in health care: a conceptual framework for a necessary transition. Acad Med 2015 Apr;90(4):437-441. [doi: 10.1097/ACM.0000000000000603] [Medline: 25607943]
24. Slattery P, Saeri AK, Bragge P. Research co-design in health: a rapid overview of reviews. Health Res Policy Syst 2020 Feb 11;18(1):17 [FREE Full text] [doi: 10.1186/s12961-020-0528-9] [Medline: 32046728]
25. Kirwan JR, de Wit M, Frank L, Haywood KL, Salek S, Brace-McDonnell S, et al. Emerging guidelines for patient engagement in research. Value Health 2017 Mar;20(3):481-486 [FREE Full text] [doi: 10.1016/j.jval.2016.10.003] [Medline: 28292494]
26. Eccleston C, Blyth FM, Dear BF, Fisher EA, Keefe FJ, Lynch ME, et al. Managing patients with chronic pain during the COVID-19 outbreak: considerations for the rapid introduction of remotely supported (eHealth) pain management services. Pain 2020 May;161(5):889-893 [FREE Full text] [doi: 10.1097/j.pain.0000000000001885] [Medline: 32251203]
27. Ivankova NV, Creswell JW, Stick SL. Using mixed-methods sequential explanatory design: from theory to practice. Field Methods 2016 Jul 21;18(1):3-20. [doi: 10.1177/1525822X05282260]
28. Creswell JW, Clark VL. Designing and Conducting Mixed Methods Research. Thousand Oaks, CA: SAGE Publications; 2007.
29. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
30. Lancaster GA, Thabane L. Guidelines for reporting non-randomised pilot and feasibility studies. Pilot Feasibility Stud 2019;5:114 [FREE Full text] [doi: 10.1186/s40814-019-0499-1] [Medline: 31608150]
31. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
32. Marier-Deschenes P, Pinard AM, LeBlanc A. Codéveloppement d'un programme d'autogestion de la douleur chronique en ligne: un projet de recherche basé sur la conception et axé sur l'engagement des patients. Can J Pain 2024;8(1):2302487 [FREE Full text] [doi: 10.1080/24740527.2024.2302487] [Medline: 38482505]
33. Marier-Deschênes P, Pinard AM, Soucy P, Le Capitaine C, Fauteux C, Gagnon-Hamel E. Agir pour moi. Réseau québecois de recherche sur la douleur. URL: https://www.gerermadouleur.ca/agir-pour-moi/ [accessed 2024-04-05]
34. Gandy M, Pang ST, Scott AJ, Heriseanu AI, Bisby MA, Dudeney J, et al. Internet-delivered cognitive and behavioural based interventions for adults with chronic pain: a systematic review and meta-analysis of randomized controlled trials. Pain 2022 Oct 01;163(10):e1041-e1053. [doi: 10.1097/j.pain.0000000000002606] [Medline: 35121696]
35. Michie S, Wood CE, Johnston M, Abraham C, Francis JJ, Hardeman W. Behaviour change techniques: the development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative data). Health Technol Assess 2015 Nov;19(99):1-188 [FREE Full text] [doi: 10.3310/hta19990] [Medline: 26616119]
36. Arain M, Campbell MJ, Cooper CL, Lancaster GA. What is a pilot or feasibility study? A review of current practice and editorial policy. BMC Med Res Methodol 2010 Jul 16;10:67 [FREE Full text] [doi: 10.1186/1471-2288-10-67] [Medline: 20637084]
37. Micoulaud-Franchi J, Sauteraud A, Olive J, Sagaspe P, Bioulac S, Philip P. Validation of the French version of the acceptability E-scale (AES) for mental E-health systems. Psychiatry Res 2016 Mar 30;237:196-200. [doi: 10.1016/j.psychres.2016.01.043] [Medline: 26809367]
38. Tariman JD, Berry DL, Halpenny B, Wolpin S, Schepp K. Validation and testing of the acceptability e-scale for web-based patient-reported outcomes in cancer care. Appl Nurs Res 2011 Feb;24(1):53-58 [FREE Full text] [doi: 10.1016/j.apnr.2009.04.003] [Medline: 20974066]
39. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, IMMPACT. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005 Jan;113(1-2):9-19. [doi: 10.1016/j.pain.2004.09.012] [Medline: 15621359]
40. Tonkin L. The pain self-efficacy questionnaire. Aust J Physiother 2008;54(1):77. [doi: 10.1016/s0004-9514(08)70073-4] [Medline: 18567187]
41. Nicholas MK. The pain self-efficacy questionnaire: taking pain into account. Eur J Pain 2007 Feb;11(2):153-163. [doi: 10.1016/j.ejpain.2005.12.008] [Medline: 16446108]

42. Poundja J, Fikretoglu D, Guay S, Brunet A. Validation of the French version of the brief pain inventory in Canadian veterans suffering from traumatic stress. J Pain Symptom Manage 2007 Jun;33(6):720-726 [FREE Full text] [doi: 10.1016/j.jpainsymman.2006.09.031] [Medline: 17531912]
43. Tan G, Jensen MP, Thornby JI, Shanti BF. Validation of the brief pain inventory for chronic nonmalignant pain. J Pain 2004 Mar;5(2):133-137 [FREE Full text] [doi: 10.1016/j.jpain.2003.12.005] [Medline: 15042521]
44. Roberge P, Doré I, Menear M, Chartrand E, Ciampi A, Duhoux A, et al. A psychometric evaluation of the French Canadian version of the hospital anxiety and depression scale in a large primary care population. J Affect Disord 2013 May;147(1-3):171-179. [doi: 10.1016/j.jad.2012.10.029] [Medline: 23218249]
45. French DJ, Noël M, Vigneau F, French JA, Cyr CP, Evans RT. L'Échelle de dramatisation face à la douleur PCS-CF: adaptation canadienne en langue française de l'échelle «pain catastrophizing scale». Can J Behav Sci 2005 Jul;37(3):181-192. [doi: 10.1037/h0087255]
46. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. URL: https://www.r-project.org/ [accessed 2024-04-05]
47. RStudio: integrated development environment for R. Posit Software. URL: https://posit.co/download/rstudio-desktop/ [accessed 2024-04-05]
48. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
49. Lüdecke D, Ben-Shachar M, Patil I, Waggoner P, Makowski D. performance: an R package for assessment, comparison and testing of statistical models. J Open Source Softw 2021 Apr;6(60):3139. [doi: 10.21105/joss.03139]
50. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
51. Jensen MP, Nielson WR, Kerns RD. Toward the development of a motivational model of pain self-management. J Pain 2003 Nov;4(9):477-492 [FREE Full text] [doi: 10.1016/s1526-5900(03)00779-x] [Medline: 14636816]
52. Creswell JW, Plano Clark VL. Designing and Conducting Mixed Methods Research 3rd edition. Thousand Oaks, CA: SAGE Publications; 2018.
53. Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs-principles and practices. Health Serv Res 2013 Dec;48(6 Pt 2):2134-2156 [FREE Full text] [doi: 10.1111/1475-6773.12117] [Medline: 24279835]
54. Guetterman TC, Fetters MD, Creswell JW. Integrating quantitative and qualitative results in health science mixed methods research through joint displays. Ann Fam Med 2015 Nov;13(6):554-561 [FREE Full text] [doi: 10.1370/afm.1865] [Medline: 26553895]
55. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
56. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018 Nov 16;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]
57. Thiblin E, Woodford J, Öhman M, von Essen L. The effect of personalised versus non-personalised study invitations on recruitment within the ENGAGE feasibility trial: an embedded randomised controlled recruitment trial. BMC Med Res Methodol 2022 Mar 06;22(1):65 [FREE Full text] [doi: 10.1186/s12874-022-01553-5] [Medline: 35249543]
58. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020 Sep 29;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
59. Schlicker S, Baumeister H, Buntrock C, Sander L, Paganini S, Lin J, et al. A web- and mobile-based intervention for comorbid, recurrent depression in patients with chronic back pain on sick leave (Get.Back): pilot randomized controlled trial on feasibility, user satisfaction, and effectiveness. JMIR Ment Health 2020 Apr 15;7(4):e16398 [FREE Full text] [doi: 10.2196/16398] [Medline: 32293577]
60. Guarino H, Fong C, Marsch LA, Acosta MC, Syckes C, Moore SK, et al. Web-based cognitive behavior therapy for chronic pain patients with aberrant drug-related behavior: outcomes from a randomized controlled trial. Pain Med 2018 Dec 01;19(12):2423-2437 [FREE Full text] [doi: 10.1093/pm/pnx334] [Medline: 29346579]
61. Perry J, VanDenKerkhof EG, Wilson R, Tripp DA. Guided internet-based psycho-educational intervention using cognitive behavioral therapy and self-management for individuals with chronic pain: a feasibility study. Pain Manag Nurs 2017 Jun;18(3):179-189. [doi: 10.1016/j.pmn.2016.12.003] [Medline: 28433488]
62. Schultz R, Smith J, Newby JM, Gardner T, Shiner CT, Andrews G, et al. Pilot trial of the reboot online program: an internet-delivered, multidisciplinary pain management program for chronic pain. Pain Res Manag 2018;2018:9634727 [FREE Full text] [doi: 10.1155/2018/9634727] [Medline: 30254712]
63. Carpenter KM, Stoner SA, Mundt JM, Stoelb B. An online self-help CBT intervention for chronic lower back pain. Clin J Pain 2012 Jan;28(1):14-22 [FREE Full text] [doi: 10.1097/AJP.0b013e31822363db] [Medline: 21681084]

64. Rini SC, Porter LS, Somers TJ, McKee DC, DeVellis RF, Smith M, et al. Automated internet-based pain coping skills training to manage osteoarthritis pain: a randomized controlled trial. Pain 2015 May;156(5):837-848 [FREE Full text] [doi: 10.1097/j.pain.0000000000000121] [Medline: 25734997]
65. Chiauzzi E, Pujol LA, Wood M, Bond K, Black R, Yiu E, et al. painACTION-back pain: a self-management website for people with chronic back pain. Pain Med 2010 Jul;11(7):1044-1058. [doi: 10.1111/j.1526-4637.2010.00879.x] [Medline: 20545873]
66. Menga G, Ing S, Khan O, Dupre B, Dornelles AC, Alarakhia A, et al. Fibromyalgia: can online cognitive behavioral therapy help? Ochsner J 2014;14(3):343-349 [FREE Full text] [Medline: 25249800]
67. Mehlsen M, Heegaard L, Frostholm L. A prospective evaluation of the chronic pain self-management programme in a Danish population of chronic pain patients. Patient Educ Couns 2015 May;98(5):677-680. [doi: 10.1016/j.pec.2015.01.008] [Medline: 25662621]
68. Perry J, VanDenKerkhof EG, Wilson R, Tripp DA. Development of a guided internet-based psycho-education intervention using cognitive behavioral therapy and self-management for individuals with chronic pain. Pain Manag Nurs 2017 Apr;18(2):90-101. [doi: 10.1016/j.pmn.2016.12.004] [Medline: 28262481]
69. Ventura J, Sobczak J, Chung J. The chronic knee pain program: a self-management model. Int J Orthop Trauma Nurs 2018 May;29:10-15. [doi: 10.1016/j.ijotn.2018.01.001] [Medline: 29409721]
70. Nicholl BI, Sandal LF, Stochkendahl MJ, McCallum M, Suresh N, Vasseljen O, et al. Digital support interventions for the self-management of low back pain: a systematic review. J Med Internet Res 2017 May 21;19(5):e179 [FREE Full text] [doi: 10.2196/jmir.7290] [Medline: 28550009]
71. Williams DA, Kuper D, Segar M, Mohan N, Sheth M, Clauw DJ. Internet-enhanced management of fibromyalgia: a randomized controlled trial. Pain 2010 Dec;151(3):694-702 [FREE Full text] [doi: 10.1016/j.pain.2010.08.034] [Medline: 20855168]
72. Trudeau KJ, Pujol LA, DasMahapatra P, Wall R, Black RA, Zacharoff K. A randomized controlled trial of an online self-management program for adults with arthritis pain. J Behav Med 2015 Jun;38(3):483-496 [FREE Full text] [doi: 10.1007/s10865-015-9622-9] [Medline: 25686538]

Abbreviations

APM: Agir pour moi
CBT: cognitive behavioral therapy
CHU: University Hospital Centre
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
REDCap: Research Electronic Data Capture

Edited by A Kushniruk; submitted 26.07.23; peer-reviewed by P Gazerani, J Pearson, J Dudeney; comments to author 11.10.23; revised version received 23.10.23; accepted 20.03.24; published 03.05.24.

Please cite as:
Marier-Deschenes P, Pinard AM, Jalbert L, LeBlanc A
Assessing the Feasibility and Preliminary Effects of a Web-Based Self-Management Program for Chronic Noncancer Pain: Mixed Methods Study
JMIR Hum Factors 2024;11:e50747
URL: https://humanfactors.jmir.org/2024/1/e50747
doi:10.2196/50747
PMID:38701440

Original Paper

A Web-Based Peer Support Network to Help Care Partners of People With Serious Illness: Co-Design Study

Elizabeth A O'Donnell[1,2*], MPH; Aricca D Van Citters[2*], MS; Inas S Khayal[2,3], PhD; Matthew M Wilson[4,5], MD; David Gustafson[6], PhD; Amber E Barnato[2,4,5], MD, MPH, MS; Andrea C Buccellato[3], BFA; Colleen Young[7], BA; Megan M Holthoff[2], MSHS; Eugene Korsunskiy[8], MFA; Stephanie C Tomlin[9], MS; Amelia M Cullinan[4,5], MD; Alexandra C Steinbaugh[5], MSW; Jennifer J Hinson[9], BHCA; Kristen R Johnson[5], MS; Andrew Williams[9], BS; Ruth M Thomson[4,5], DO; Janet M Haines[9], BA; Anne B Holmes[9], BA; Ann D Bradley[9], RN, BSN, MEd; Eugene C Nelson[2], DSC, MPH; Kathryn B Kirkland[2,4,5], MD

[1]Communications, Marketing and Community Health, Alice Peck Day Memorial Hospital, Lebanon, NH, United States
[2]The Dartmouth Institute for Health Policy & Clinical Practice, Geisel School of Medicine at Dartmouth, Lebanon, NH, United States
[3]Dartmouth Cancer Center, Geisel School of Medicine, Dartmouth, Lebanon, NH, United States
[4]Palliative Medicine, Geisel School of Medicine at Dartmouth, Lebanon, NH, United States
[5]Section of Palliative Medicine, Dartmouth Health, Lebanon, NH, United States
[6]College of Engineering, University of Wisconsin, Madison, WI, United States
[7]Mayo Clinic Connect, Mayo Clinic, Rochester, MN, United States
[8]Thayer School of Engineering, Dartmouth College, Hanover, NH, United States
[9]Patient and Family Advisors, Dartmouth Health, Lebanon, NH, United States
[*]these authors contributed equally

Corresponding Author:
Aricca D Van Citters, MS
The Dartmouth Institute for Health Policy & Clinical Practice
Geisel School of Medicine at Dartmouth
One Medical Center Drive, WTRB Level 5
Lebanon, NH, 03756
United States
Phone: 1 603 646 5668
Email: aricca.d.van.citters@dartmouth.edu

Abstract

Background: Care partners of people with serious illness experience significant challenges and unmet needs during the patient's treatment period and after their death. Learning from others with shared experiences can be valuable, but opportunities are not consistently available.

Objective: This study aims to design and prototype a regional, facilitated, and web-based peer support network to help active and bereaved care partners of persons with serious illness be better prepared to cope with the surprises that arise during serious illness and in bereavement.

Methods: An 18-member co-design team included active care partners and those in bereavement, people who had experienced serious illness, regional health care and support partners, and clinicians. It was guided by facilitators and peer network subject-matter experts. We conducted design exercises to identify the functions and specifications of a peer support network. Co-design members independently prioritized network specifications, which were incorporated into an early iteration of the web-based network.

Results: The team prioritized two functions: (1) connecting care partners to information and (2) facilitating emotional support. The design process generated 24 potential network specifications to support these functions. The highest priorities included providing a supportive and respectful community; connecting people to trusted resources; reducing barriers to asking for help; and providing frequently asked questions and responses. The network platform had to be simple and intuitive, provide technical support for users, protect member privacy, provide publicly available information and a private discussion forum, and be easily accessible. It was feasible to enroll members in the ConnectShareCare web-based network over a 3-month period.

Conclusions: A co-design process supported the identification of critical features of a peer support network for care partners of people with serious illnesses in a rural setting, as well as initial testing and use. Further testing is underway to assess the long-term viability and impact of the network.

(JMIR Hum Factors 2024;11:e53194) doi:10.2196/53194

KEYWORDS

human-centered design; caregivers; care partners; serious illness; peer support; online support network; virtual network; online network; caregiver; unmet need; unmet needs; active care; bereaved care; bereavement; clinician; clinicians; function; functions; specification; information; emotional support; technical support; privacy protection; rural; viability; impact; engineering design; care provider; care providers; mortality; quality of life; tertiary care; caregiving

Introduction

Care partners of people with serious illnesses are often overlooked and poorly understood by health care professionals, lack support and educational resources, and are likely to experience significant challenges and unmet needs [1,2]. For many, the work of caring for a person with a serious illness can bring deep satisfaction and can also be challenging [3]. Care partners experience burdens in every area of their lives—emotional, physical, social, spiritual, and financial [2,4,5]. The death of someone with a serious illness (as well as the events leading up to it) also brings hardship, including stress related to loneliness, grief, trauma, role recognition, and self-identity [6-9]. Social isolation and grief are strongly correlated with subsequent depression and related symptoms in bereaved spouses, including sadness, appetite loss, and lower quality of life [5,10]. Bereaved care partners may also face challenges navigating household, financial, social planning, and legal affairs, as well as reintegrating into their local community [11] and accessing available resources. Addressing care partner needs has become a pressing health, economic, and social imperative [12].

Increasingly, care partners are joining online peer support networks to obtain emotional support, access information, and connect and share with others in similar circumstances [13-17]. While people often prefer connecting with health professionals for medical information, they prefer connections with peers over professionals for accessing emotional support or practical advice [18]. In the case of serious illness, when patients may not be well enough to use online peer support networks themselves, care partners are more likely to participate [19].

Previous research has demonstrated a number of variables that contribute to an online peer support network's success or failure [20-22]. Communities that have a clear, defined purpose; foster a strong sense of community; and have a high level of activity are more likely to be successful [20]. Additionally, sustained organizational and financial support for maintaining an online community from inception to maturity is essential, including support for a community manager who sets the tone for the community, creates content, conducts outreach, and fosters a sense of community [20,23]. Successful online networks also harness the interests and abilities of their users to strengthen the community. Networks with more active users are generally more successful [21] because they maintain a critical mass to allow for diversity in experiences and individual attributes, allowing for the natural formation of relationships and answering questions.

Evidence on the impact of online peer support networks for care partners is promising [14,24-26]. For care partners of people with cancer, studies show evidence of decreased care partner emotional distress [27], negative mood [28,29], and sense of burden [29], as well as increases in quality of life and self-efficacy [27]. For care partners of people with dementia, online networks can lead to improvement in self-efficacy [30], decision-making confidence [31], and care partner and patient relationship quality [32]. Care partners also benefit from being able to freely express their sentiments and provide mutual support in a dedicated digital space apart from their loved ones [33,34]. Even people who observe network activity without participating report that reading about the experiences of others is empowering and informative [35,36]. Online networks offer certain advantages: 24/7 home access, flexibility (communication is often asynchronous), anonymity, and a wide range of expertise and experience not limited by geography [37-39].

Online peer support networks for care partners often target specific health conditions (eg, breast cancer and Alzheimer dementia) or stages of caregiving (treatment vs bereavement), but infrequently support care partners of people with diverse conditions or the transition between stages of caregiving. They may also fail to provide active facilitation and moderation; identify and vet regional resources and support from local peers; or provide the possibility of meeting in person. A co-design process can elicit those factors that matter most to the people for whom the network is intended to serve and ensure the successful adoption of the proposed solution [40].

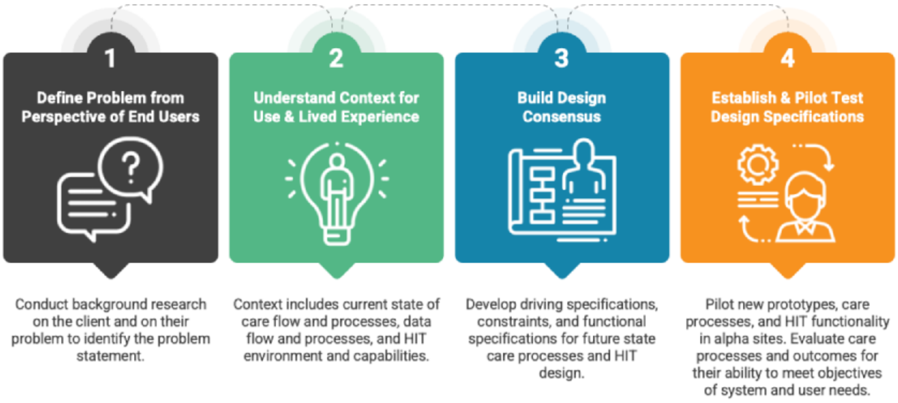
The objective of this paper is to describe a co-design process and the resulting key functions and specifications for a regional, facilitated, and web-based peer support network that can meet the needs of active and bereaved care partners of persons with serious illnesses.

Methods

Overview

We applied a co-design framework (Figure 1 [41,42]), which combines human-centered design [43] and engineering design [44,45] processes, to create the specifications for a regional, facilitated, and web-based peer support network. The framework includes 4 stages: defining the problem, understanding the context for use, developing and building consensus around functions and specifications that fulfill identified needs, and establishing and pilot-testing design specifications.

Figure 1. Co-design framework (reproduced from The Dartmouth Institute for Health Policy & Clinical Practice, which is published under Creative Commons Attribution 4.0 International License). HIT: health information technology.



Ethical Considerations

The study was approved by the Dartmouth College institutional review board (#2000907).

A waiver of written informed consent was used for the surveys, as the only link to the survey respondent would have been the written informed consent document. No identifiable information was collected and individuals were not paid for participating in human subjects research.

Target Population

The target population, hereafter referred to as end users, was defined as care partners (ie, informal caregivers or family members) supporting or providing care to adults (aged 18 years or older) with a serious illness and those who have experienced the loss of someone to a serious illness. The term care partner was chosen by the co-design team to reflect their role and relationship in partnering with a person with a serious illness. Serious illness has been defined as one that carries a high risk of mortality and either negatively impacts a person's daily function or quality of life, or excessively strains their care partners [46]. Care partners include relatives, spouses or partners, friends, neighbors, or others who have a significant personal relationship with, and who provide a broad range of assistance to, a person with a serious illness. We focused on care partners living in New Hampshire and Vermont, the catchment area for Dartmouth-Hitchcock Medical Center.

Participants

We formed an 18-member team to co-design the network. The team included 2 active and bereaved care partners of people with serious illness, 4 adults with serious illness, 6 interdisciplinary palliative care clinical team members, and 6 support service staff. Care partners and people with serious illnesses were recruited by our clinical team partners. Clinical team members were affiliated with Dartmouth-Hitchcock Medical Center, a rural tertiary care academic medical center in New Hampshire. Facilitation of the co-design process was led by researchers with expertise in co-design, evaluation, and quality improvement (EAO and ADVC). The design process was informed by consultation with an expert in human-centered design (EK) and a systems engineer (ISK). To ensure our design aligned with best practices, we consulted with external advisors with expertise in facilitated support networks (DG and CY), met with regional health care and support partners, and obtained input from an external advisory committee with expertise in scaling innovations, business, and serious illness.

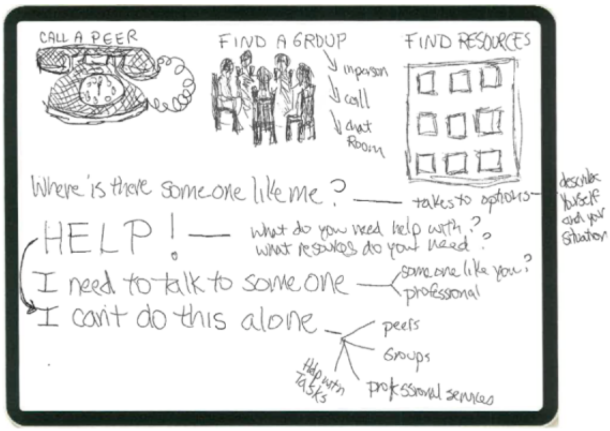
The co-design team met twice a month for 8 months (April to November 2019) to identify and prioritize the functions and specifications of the network and met monthly for 6 months (December 2019 to May 2020) to test prototypes.

Defining the Problem and Understanding Context for Use

We conducted human-centered design exercises [43] to elicit community needs and assets, define the problem, and understand the context for use. We drew upon stories of serious illnesses shared by care partners to identify needs arising from lived experiences. Design facilitators shadowed [43] outpatient palliative care visits and attended interdisciplinary palliative care team meetings to further understand the context of use; services provided by the care team; and the daily lives of people with serious illness, their care partners, and clinicians [47]. We developed empathy maps [43] to reflect and articulate what end users hear, see, say, do, think, and feel, and to identify points of pain and gain. We used a visual thinking exercise [48] to sketch ideas for an ideal support network (example, Figure 2). Design exercises were reviewed during design sessions to discuss critical functions and important features that fulfill and support these functions.

We supplemented design activities with surveys of potential end users. Between December 2018 and April 2019, we collected 28 surveys from a convenience sample of active care partners presenting at the Dartmouth-Hitchcock Medical Center outpatient palliative care clinic and 21 surveys from a convenience sample of bereaved care partners affiliated with the clinic or regional health organizations (eg, hospice) to elicit information on the challenges, needs, and desire for peer connection among active and bereaved care partners (surveys are provided in Multimedia Appendix 1). We used descriptive statistics to summarize categorical data, and thematic analyses to identify themes from open-ended questions.

Figure 2. Sketch of a serious illness community network, created by a co-design team member.



Building Design Consensus

We collated, organized, and systematically described the identified needs into a hierarchical list. We converted the highest-level needs into functions for the network to achieve. At the highest level, for example, “care partner need for information” was converted to “connecting care partners to information” and “care partner need for emotional support” was converted to “facilitating emotional support.” We identified potential specifications associated with each function of the network.

Co-design team members independently rated the importance of each specification using a five-item prioritization scale: (5) must have, (4) high importance (feasible without), (3) should have (very important), (2) could have (consideration), and (1) desirable (will not have at this time) [49]. Team members were provided an opportunity to describe their understanding and thought process for preference identification as a group. Scores were weighted and averaged by respondent type to ensure that patients and care partners, clinicians, and support service staff had equal participation weight (eg, scores from 6 clinicians were averaged to create 1 clinician average).

Pilot-Testing Design Specifications

We tested a series of prototypes to identify the importance of different network components. These included storytelling exercises, face-to-face group conversations, one-to-one matching of care partners, group videoconferencing, an online discussion forum, and a “Caregiver Day” event at the health center. Summaries are provided in Multimedia Appendix 2.

After finalizing design functions and specifications, we identified potential vendors to host the network by conducting an environmental scan of web-based networks focused on people with serious illnesses and care partners and mapping desired design specifications to vendor capabilities. Four vendors provided demonstrations of their platforms between April and May 2020. Vendor selection was driven by the vendor’s ability to provide the prioritized functions and specifications, the cost to build and maintain the network, and being a US-based company. Following vendor selection, the web-based network platform was customized by the design team to deliver upon design functions and specifications.

User acceptance testing was conducted between April and July 2021. Four care partners and 2 people who previously had a serious illness were invited to register as founding members of the network in April. These founding members were encouraged to invite care partners they knew into the web-based community to test the feasibility of enrolling members. Three clinical champions (physicians and social workers), a chaplain, and a staff member who manages complementary care programs referred care partners to the network. The research team met weekly with the vendor during the user acceptance testing period to resolve issues. The web-based platform remained available for registered members to use while issues were addressed. The network moderator met monthly with the design team to plan future improvements using a quality improvement framework [50].

Results

Problem Definition

The co-design process led to clarity around the objective of the network: to help care partners cope with the surprises that arise during serious illness and bereavement. The network, named “ConnectShareCare,” was intended to supplement existing services, to be provided outside of clinical encounters with the health care system or regional professional support and service organizations, and to tap into the wisdom of those with lived experiences.

If successful, the design team anticipated that the network would benefit 4 groups, as outlined in Table 1.

Table 1. Anticipated impact of ConnectShareCare.

Audience	Anticipated objectives
Care partners	• Improve access to information that can guide decision-making • Improve sense of empowerment in making decisions and providing support • Decrease sense of distress and social isolation
Community partners	• Improve understanding of needs and gaps in service • Provide a system to share assets or resources
Clinicians or health care system	• Address gaps in services that are not currently met • Improve availability to see patients who seek services • Improve efficiency of health care encounters
Quality improvement leaders and researchers	• Improve understanding of the needs of care partners • Align services with care partners and community needs • Demonstrate a positive impact of the network over time

Context for Use and Lived Experiences

Our co-design process identified that active and bereaved care partners have different needs but have common interests in sharing information and providing or receiving support. Active care partners who completed an assessment survey were most challenged by emotional difficulties (eg, worry, uncertainty, or lack of control; 12/28, 43%), providing care and emotional support (7/28, 25%), and practical matters (6/28, 21%). Bereaved care partners were most challenged by loneliness (10/21, 48%), managing grief and emotional difficulties (6/21, 29%), and managing practical matters (5/21, 24%). Active care partners were most helped by support from friends, family, or other social connections (12/28, 43%), as well as by medical professionals (9/28, 32%), while bereaved care partners were most helped by support from friends, family, or other social connections (16/21, 76%) and by developing self-care strategies that led to personal resilience and growth (12/21, 57%).

Most active (18/27, 67%) and bereaved (18/21, 86%) care partners were interested in 1 or more forms of connecting with other people who have shared a similar care experience. Both active and bereaved care partners anticipated that a network could provide support, knowledge, and resources but anticipated challenges associated with time to participate and with forming personal connections.

The series of human-centered design exercises and interviews led to additional insights. First, active and bereaved care partners may benefit from connecting and sharing information with each other. Peer-generated information from care partners who have shared a similar experience feels more authentic, detailed, and actionable. Second, care partners wished to belong to a local support community that was connected through geographic proximity and could provide recommendations for local resources. Third, a web-based network enables care partners to access information at any place or time, allows anonymity, improves access for people who are home-bound or grieving, and may reach an increased number of care partners. Fourth, trained staff who can moderate, promote, and manage the web-based community and volunteers who can recruit and engage users are important. Paid or volunteer moderators can play an important role in listening, making connections, and highlighting information, services, and programs. Fifth, a web-based network would benefit from supplemental opportunities for the community to meet face-to-face or through digital programming.

The design process also identified several potential risks and possible mitigation strategies. First, there was a risk of causing harm to vulnerable end users if the design failed to provide a safe and supportive environment, protect the privacy of sensitive information, or enact acceptable data ownership guidelines. Second, there was a risk associated with the usability of the network among end users who were less facile with web-based services. Third, there was a risk associated with the inability to form a personal connection with peers through a web-based network. Other potential risks included those associated with competition from other networks, inaccurate content or information, and care partners having minimal time to participate in a web-based support network due to other responsibilities.

Build Design Consensus

Two primary functions emerged from the design activities: to support care partners in (1) providing each other with emotional support and (2) exchanging helpful information and resources. We developed specifications related to these functions (Table 2), as well as the form of the network, including the user interface, data and security, and other considerations (Table 3).

Table 2. Prioritization ranking of design specifications associated with network functions (weighted average scores across patients and families [n=5]; clinicians [n=4]; clinical support service providers [n=5]; and network advisors [n=2]).

Network functions and prioritization ranking of associated design specifications	Ranking[a]
Function 1: provide each other with emotional support	
Must have	
• Provide a supportive and respectful space • Ability for established "Guidelines and Ground Rules" to be clearly visible to users • Ability to protect individual identity (opt-out options for sharing personal information; opportunity to keep geographic location private) • Allows moderator functionality for policing interactions and blocking users if necessary	5.00
Should have (very important)	
• Incorporate and help facilitate, one-to-one connections • Ability to locate "true peer" (similar users) through the platform via matching on similar life circumstances (through backend algorithm or user profile details: type of loss, disease, time caregiving, or time since loss) • Allow for private one-to-one messaging to facilitate a more personal connection, not monitored by an external entity	3.27
• Includes opportunity for storytelling based on personal user content or experience	3.00
• Provides an opportunity to share solutions	3.00
Could have (consideration)	
• Differentiate between whether people want to feel heard or want to hear solutions • Ability for users to designate whether they are looking to hear solutions or feedback or simply share	2.88
Function 2: exchange helpful information and resources	
Must have	
• Provide connections to trusted and curated local, national, and international resources • Ability to host webinars in order to share educational content • Ability for newsfeed or wall that features newly published content • Provide document or resource repository related to user needs • Ability to highlight and share local events via calendar, bulletin board, or other • Robust search function available to find targeted resources within the platform	4.56
High importance (feasible without)	
• Reduce the difficulty of asking for help by normalizing needing help • Ability to add a button in various locations that asks "Having a hard time asking for help?" and that opens a new page that contains tips or guidelines on how to ask for help and what to expect when asking for help	3.88
• Provide frequently asked questions list and answers • Site provides a list, or ability to create a list, of the most popular or frequently asked questions and answers (eg, "How do I cope with stress?")	3.76
Should have (very important)	
• Ability to identify most common needs • Ability to organize conversations around themes (or topics) and make it easy for someone with a specific question, topic, or theme to locate information pertaining to it • Ability to "like," (showing interest, support) posts, topics, or comments so that users can see which posts are popular and most useful • Ability to follow a discussion thread, topic, etc. Once a user has "followed" something or someone, they can receive a notification when there is new content posted • Ability to bookmark posts (to save content) that users would like to revisit	3.47
• Ability to be supported by local or regional expert moderator (community manager)	3.38

Network functions and prioritization ranking of associated design specifications	Ranking[a]
• Include the ability to publish videos related to the content of the network • Allow users to record and post videos instantly (personal and other)	3.06

[a]Prioritization ranking: (5) must have, (4) high priority (feasible without), (3) very important (should have), (2) consideration (could have), (1) desirable (will not have at this time).

Table 3. Prioritization ranking of the form of the network: user interface, data and security, and other considerations (weighted average scores across patients and families [n=5]; clinicians [n=4]; clinical support service providers [n=5]; and network advisors [n=2]).

Form of the network and its prioritization ranking	Ranking[a]
User interface	
Must have	
• Simple or intuitive interface • Provide support, guidance, and assistance with how to navigate and use platform (ideal: offer video tutorials) • Passes Web Content Accessibility Guidelines (WCAG). Example: large font • Ability to easily identify new content (since user's last login)	5.00
• Provides IT technical support (for members)	5.00
• Easy access to support user engagement • Smooth and simple login process • Optimized for mobile device • Real-time information and comments available and accessible (not prescreened by community moderator)	4.53
High importance (feasible without)	
• Aesthetically refined • Pleasing to the eye, organized, and appropriate imagery • Symmetrical and aligned (looks modern)	4.00
• Appropriate use of pop-ups and other interactive elements	4.00
• Does not allow advertising • Does not have advertisements on the platform itself • Does not send any unsolicited promotional emails related to the platform or other	3.76
Data and security	
Must have	
• Secure platform or user privacy protected (Health Insurance Portability and Accountability Act [HIPAA] compliant)	5.00
• Public forum for information and resources, but opportunities for private discussion forums	5.00
Should have (very important)	
• Data are owned by the co-design team's institution (not the vendor) • Establish terms and conditions for how information or data will be accessed, stored, and used • No selling of data to for-profit or not-for-profit entities (pharmaceuticals or other) for financial gains	3.47
• Data access and analysis • Provides actionable metrics related to user activity and engagement (including IP address) • Data analysis capabilities available within local database • Create an extract of selected data • Ability to survey users	3.00
Other considerations	
High importance (feasible without)	
• Health-focused support network • Network software is targeted to health-focused communities, has features and functions relevant to health, self-management, etc (eg, health needs assessment) and has experience working in peer-to-peer health care	4.00
• Sustainability (retaining users) • Provides facilitation and network growth support (through designated pump primers, marketing, etc) • Opportunity to offer member incentives (through badges, quality improvement initiatives: creating educational material, etc)	4.00

Form of the network and its prioritization ranking	Ranking[a]
• Scalability • Interactive and responsive; ability to customize and add measures and functionality over time	4.00

[a]Prioritization ranking: (5) must have, (4) high priority (feasible without), (3) very important (should have), (2) consideration (could have), (1) desirable (will not have at this time).

The most highly prioritized specifications to support each function (Table 2) included providing a supportive and respectful space; providing connections to trusted and curated local, national, and international resources; reducing the difficulty of asking for help by normalizing needing help; and providing curated resources to address the most common concerns (eg, easy access to frequently asked questions and answers).

The user interface (Table 3) must be simple and intuitive, provide technical support for users, and be easy to access. It was highly important for the user interface to be aesthetically refined, include appropriate use of interactive elements, and not allow external advertising. The platform must be secure and protect user privacy. It must be available as a public forum for information but also allow participants to communicate via discussions not visible to others outside of the network. Other highly important considerations included hosting by a vendor with experience in providing health-focused networks and providing features that support sustainability and scalability (such as member incentives and the ability to customize or add functionality over time).

Pilot Test Design Specifications

Following the vendor selection process, we worked with CareHubs, an online health network vendor, to build the ConnectShareCare network. The network included (1) a single support community for active and bereaved care partners; (2) a short list of curated resources based on needs identified during the design (planning ahead, practical issues, emotional issues, communication issues, and family resources); (3) a calendar of online and in-person regional support programming; (4) a story from an end user about their experience as a care partner; (5) a roster with member profiles; (6) a help center; and (7) community guidelines. Screenshots are included in Multimedia Appendix 3.

The initial feasibility of the network was demonstrated through the active participation of founding members, beginning in April 2021, with expansion to include 12 new care partner members in May 2021 and 15 new members in June 2021. During this period, the network had an average of 135 posts per month. A total of 16 members posted a message to the network, with an average of 11 members posting per month.

Discussion

Principal Findings

A co-design process generated a useful and feasible regional, facilitated, and web-based peer support network for care partners of people with serious illnesses. The co-design process ensured that all voices were heard, especially among people who typically may not work together. Design decisions were made collectively and systemically, which allowed network functions and specifications to be identified and prioritized. By doing so, the co-design process ensured that the most critical decisions were responsive to regional needs and preferences. The resulting network connects active and bereaved care partners with peers to facilitate emotional support and exchange information related to caregiving for people with serious illness.

Comparison With Prior Work

Similar to other networks [15,33,51-62], ConnectShareCare has a clear purpose, includes mechanisms to foster a strong sense of community and support among regional care partners, and provides value to a variety of groups. The network builds on the resources, wisdom, and experiences of care partners. The inclusion of a moderator helps ensure a safe environment that is protected from misinformation, trolling, or cyberbullying. The moderator sets the tone and etiquette with members, modeling behavior and other preventative measures, as well as moderating posts, facilitating connections, and providing feedback to adjust member behaviors [20,23].

Our design process had several strengths. First**,** our process engaged people who would be end users of the network in making critical design decisions [40]. In contrast to asking end users about single decisions, our process allowed end users to make decisions in the context of all other design requirements and options. This led to a more systemic approach to engagement that ensures that decisions are optimized to fit together. Second, our process brought together people with different expertise (in being a person with serious illness or a care partner, in medicine, in health network design, and in community management and moderation) that may typically not work together to create services, and each had different needs to maintain the value of the network. The process also engaged people with engineering design [44,45] and user-centered design [43] expertise to ensure that the process was rigorous enough to produce a network tailored to the needs of end users. Third, our design was responsive to regional needs by addressing gaps in available services and drawing upon local assets. We intentionally worked with multiple health care and support organizations in the region to broaden our network and reach care partners most in need of support, regardless of where formal health care services were received. Care partners are a vulnerable population who often receive minimal structured support from the health system, yet they have significant knowledge to contribute on how to navigate health care systems, health and social resources, and losses at every stage (eg, relationships, identity, and freedom). This knowledge is often actionable by peers. Finally, our process may have supported growth in network participation due to health care and regional support partners feeling heard and included in the design process.

Limitations

This project has certain limitations that should be considered by those wishing to adapt the methods and findings to different contexts. First, the project was built around a recognition that people who live in north-east rural areas can be particularly isolated and lack access to sources of support. It is unclear whether the network will meet the needs of people in other regions of the country. Second, similar to local demographic characteristics, our design team had limited racial or ethnic diversity. We do not know how greater racial, ethnic, or sociodemographic diversity would enhance or create barriers to its success. Third, the network is built around an asynchronous model. This can be very important because it respects the different schedules that people are on; however, it limits the opportunities for people to hear and see each other in real time. It is unknown whether a network that combines synchronous and asynchronous components would be useful in our context. Finally, while the design process requires extensive back and forth among participants and may not be feasible in other situations, it also represents a strength in creating a network that more closely reflects community needs. In our situation, the decision to create a new network was a result of the recognition that other solutions are not likely to fulfill the needs of our end users.

Conclusions

Care partners of people with serious illness often lack support and are likely to experience significant challenges and unmet needs. We followed a structured co-design process to collaboratively identify and prioritize the functions and specifications of a regional web-based facilitated peer support network to help care partners cope with the surprises that arrive during serious illness and bereavement. The network was designed to provide emotional support and exchange information related to serious illness caregiving. The coproduction of accessible peer-led information, resources, and support may extend the scope of services offered by a health system to support lay care partners—becoming part of a sustainable, person-centered value-creation system [63,64]. Opportunities exist to evaluate the feasibility of actively engaging community members and moderators in the network [65,66] and will be reported on in a publication under development. Moreover, there is a need to understand effective mechanisms to recruit and retain participants and provide a safe environment to people who are in vulnerable situations; to monitor the network life cycle through metrics related to activity and growth [20,22]; and to consider the creation of network subgroups to support care partners of people with particular illnesses (eg, cancer, dementia, and Parkinson disease) or people from specific minority populations. Finally, there is an opportunity to understand the impact of a regional support network on care partner quality of life, self-confidence, loneliness, and isolation [67,68]; and on health system reputation, use, and visibility.

Acknowledgments

The authors wish to thank the many other people who contributed to the design and implementation of the ConnectShareCare network and the many active and bereaved care partners who participated in and provided feedback on the network. Special thanks are given to Deadra Ashton, James Bowling, Mary Elizabeth Byrnes, Nancy Duhaine, Ellen Flaherty, Melissa Garland, Malavika Govindan, Amanda Hoggard, Marv Klassen-Landis, Jane Masters, Catherine Reed, and Donna Soltura for their feedback and contributions to the design of the network. In addition, the authors wish to thank the Upper Valley Elder Forum; Upper Valley Bereavement Partnership Group; Dartmouth-Hitchcock: Volunteers, Palliative Care Program, Arts and Humanities Program, and Aging Resource Center; Dartmouth Cancer Center; Visiting Nurse and Hospice of Vermont and New Hampshire; Lake Sunapee Visiting Nurses Association and Hospice Care; Bayada Home Care; CommunityCare of Lyme; Community Nurse Connection; Shire Digital; and CareHubs for their support of the ConnectShareCare network. This work was supported by the Gordon and Betty Moore Foundation (grant number 7485) and the Couch Family Fund at The Dartmouth Institute for Health Policy and Clinical Practice.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Surveys administered to active and bereaved caregivers.
[DOCX File , 479 KB - humanfactors_v11i1e53194_app1.docx]

Multimedia Appendix 2
Prototype and testing results.
[DOCX File , 23 KB - humanfactors_v11i1e53194_app2.docx]

Multimedia Appendix 3
ConnectShareCare screenshots.
[DOCX File , 1751 KB - humanfactors_v11i1e53194_app3.docx]

References

1. Northouse LL, Katapodi MC, Song L, Zhang L, Mood DW. Interventions with family caregivers of cancer patients: meta-analysis of randomized trials. CA Cancer J Clin 2010;60(5):317-339 [FREE Full text] [doi: 10.3322/caac.20081] [Medline: 20709946]
2. Choi S, Seo J. Analysis of caregiver burden in palliative care: an integrated review. Nurs Forum 2019;54(2):280-290. [doi: 10.1111/nuf.12328] [Medline: 30737798]
3. Lewis C, Abrams MK, Schneider EC, Shah T. Coping with serious illness in America: relying on family and friends. The Commonwealth Fund. 2019. URL: https://www.commonwealthfund.org/publications/2019/feb/coping-serious-illness-america [accessed 2024-04-10]
4. McDonald J, Swami N, Pope A, Hales S, Nissim R, Rodin G, et al. Caregiver quality of life in advanced cancer: qualitative results from a trial of early palliative care. Palliat Med 2018;32(1):69-78. [doi: 10.1177/0269216317739806] [Medline: 29130418]
5. Götze H, Brähler E, Gansera L, Schnabel A, Gottschalk-Fleischer A, Köhler N. Anxiety, depression and quality of life in family caregivers of palliative cancer patients during home care and after the patient's death. Eur J Cancer Care (Engl) 2018;27(2):e12606. [doi: 10.1111/ecc.12606] [Medline: 27859889]
6. Utz RL, Swenson KL, Caserta M, Lund D, deVries B. Feeling lonely versus being alone: loneliness and social support among recently bereaved persons. J Gerontol B Psychol Sci Soc Sci 2014;69(1):85-94 [FREE Full text] [doi: 10.1093/geronb/gbt075] [Medline: 24056690]
7. U.S. Surgeon General. Our epidemic of loneliness and isolation: the U.S. surgeon general's advisory on the healing effects of social connection and community. U.S. Department of Health and Human Services. 2023. URL: https://www.hhs.gov/sites/default/files/surgeon-general-social-connection-advisory.pdf [accessed 2024-04-10]
8. Ugalde A, Krishnasamy M, Schofield P. Role recognition and changes to self-identity in family caregivers of people with advanced cancer: a qualitative study. Support Care Cancer 2012;20(6):1175-1181. [doi: 10.1007/s00520-011-1194-9] [Medline: 21611864]
9. Duffy M, Wild J. Living with loss: a cognitive approach to prolonged grief disorder—incorporating complicated, enduring and traumatic grief. Behav Cogn Psychother 2023;51(6):645-658 [FREE Full text] [doi: 10.1017/S1352465822000674] [Medline: 37170765]
10. Fried EI, Bockting C, Arjadi R, Borsboom D, Amshoff M, Cramer AOJ, et al. From loss to loneliness: the relationship between bereavement and depressive symptoms. J Abnorm Psychol 2015;124(2):256-265. [doi: 10.1037/abn0000028] [Medline: 25730514]
11. Kim Y, Carver CS, Cannady RS. Bereaved family cancer caregivers' unmet needs: measure development and validation. Ann Behav Med 2020;54(3):164-175 [FREE Full text] [doi: 10.1093/abm/kaz036] [Medline: 31602463]
12. Reinhard SC, Caldera S, Houser A, Choula RB. Valuing the invaluable 2023 update: strengthening supports for family caregivers. AARP Public Policy Institute. 2023. URL: https://www.aarp.org/ppi/info-2015/valuing-the-invaluable-2015-update.html [accessed 2024-04-10]
13. Fox S, Duggan M, Purcell K. Family caregivers are wired for health. Pew Research Center. Washington, D.C; 2013. URL: https://www.pewresearch.org/internet/2013/06/20/family-caregivers-are-wired-for-health/ [accessed 2024-04-10]
14. Shin JY, Kang TI, Noll RB, Choi SW. Supporting caregivers of patients with cancer: a summary of technology-mediated interventions and future directions. Am Soc Clin Oncol Educ Book 2018;38:838-849 [FREE Full text] [doi: 10.1200/EDBK_201397] [Medline: 30231412]
15. Gage-Bouchard EA, LaValley S, Mollica M, Beaupin LK. Cancer communication on social media: examining how cancer caregivers use facebook for cancer-related communication. Cancer Nurs 2017;40(4):332-338. [doi: 10.1097/NCC.0000000000000418] [Medline: 27442210]
16. Ahmad A, Premanandan S, Langegård U, Cajander Å, Johansson B, Carlsson M, et al. Positive design framework for carer eSupport: qualitative study to support informal caregivers of patients with head and neck cancer in Sweden. JMIR Cancer 2023;9:e45748 [FREE Full text] [doi: 10.2196/45748] [Medline: 37252765]
17. Dale J, Nanton V, Day T, Apenteng P, Bernstein CJ, Grason Smith G, et al. Uptake and use of care companion, a web-based information resource for supporting informal carers of older people: mixed methods study. JMIR Aging 2023;6:e41185 [FREE Full text] [doi: 10.2196/41185] [Medline: 37733406]
18. Fox S. Peer to peer healthcare: many people—especially those living with chronic or rare diseases—use online connections to supplement professional medical advice. Pew Research Center. Washington DC; 2011. URL: https://www.pewresearch.org/internet/2011/02/28/peer-to-peer-health-care-2/ [accessed 2024-04-10]
19. Kinnane NA, Milne DJ. The role of the internet in supporting and informing carers of people with cancer: a literature review. Support Care Cancer 2010;18(9):1123-1136. [doi: 10.1007/s00520-010-0863-4] [Medline: 20336326]
20. Young C. Community management that works: how to build and sustain a thriving online health community. J Med Internet Res 2013;15(6):e119 [FREE Full text] [doi: 10.2196/jmir.2501] [Medline: 23759312]
21. Iriberri A, Leroy G. A life-cycle perspective on online community success. ACM Comput Surv 2009;41(2):1-29. [doi: 10.1145/1459352.1459356]

22. Preece J. Sociability and usability in online communities: determining and measuring success. Behav Inform Technol 2010;20(5):347-356. [doi: 10.1080/01449290110084683]
23. Skousen T, Safadi H, Young C, Karahanna E, Safadi S, Chebib F. Successful moderation in online patient communities: inductive case study. J Med Internet Res 2020;22(3):e15983 [FREE Full text] [doi: 10.2196/15983] [Medline: 32181743]
24. Oliver DP, Patil S, Benson JJ, Gage A, Washington K, Kruse RL, et al. The effect of internet group support for caregivers on social support, self-efficacy, and caregiver burden: a meta-analysis. Telemed J E Health 2017;23(8):621-629. [doi: 10.1089/tmj.2016.0183] [Medline: 28328392]
25. Ploeg J, Ali MU, Markle-Reid M, Valaitis R, Bartholomew A, Fitzpatrick-Lewis D, et al. Caregiver-focused, web-based interventions: systematic review and meta-analysis (part 2). J Med Internet Res 2018;20(10):e11247 [FREE Full text] [doi: 10.2196/11247] [Medline: 30368439]
26. Dam AEH, de Vugt ME, Klinkenberg IPM, Verhey FRJ, van Boxtel MPJ. A systematic review of social support interventions for caregivers of people with dementia: are they doing what they promise? Maturitas 2016;85:117-130. [doi: 10.1016/j.maturitas.2015.12.008] [Medline: 26857890]
27. Northouse L, Schafenacker A, Barr KLC, Katapodi M, Yoon H, Brittain K, et al. A tailored web-based psychoeducational intervention for cancer patients and their family caregivers. Cancer Nurs 2014;37(5):321-330 [FREE Full text] [doi: 10.1097/NCC.0000000000000159] [Medline: 24945270]
28. DuBenske LL, Gustafson DH, Namkoong K, Hawkins RP, Atwood AK, Brown RL, et al. CHESS improves cancer caregivers' burden and mood: results of an eHealth RCT. Health Psychol 2014;33(10):1261-1272 [FREE Full text] [doi: 10.1037/a0034216] [Medline: 24245838]
29. Chih MY, DuBenske LL, Hawkins RP, Brown RL, Dinauer SK, Cleary JF, et al. Communicating advanced cancer patients' symptoms via the internet: a pooled analysis of two randomized trials examining caregiver preparedness, physical burden, and negative mood. Palliat Med 2013;27(6):533-543 [FREE Full text] [doi: 10.1177/0269216312457213] [Medline: 22988042]
30. Marziali E, Garcia LJ. Dementia caregivers' responses to 2 internet-based intervention programs. Am J Alzheimers Dis Other Demen 2011;26(1):36-43 [FREE Full text] [doi: 10.1177/1533317510387586] [Medline: 21282276]
31. Bass DM, McClendon MJ, Brennan PF, McCarthy C. The buffering effect of a computer support network on caregiver strain. J Aging Health 1998;10(1):20-43. [doi: 10.1177/089826439801000102] [Medline: 10182416]
32. McKechnie V, Barker C, Stott J. The effectiveness of an internet support forum for carers of people with dementia: a pre-post cohort study. J Med Internet Res 2014;16(2):e68 [FREE Full text] [doi: 10.2196/jmir.3166] [Medline: 24583789]
33. Oliver DP, Washington K, Wittenberg-Lyles E, Gage A, Mooney M, Demiris G. Lessons learned from a secret Facebook support group. Health Soc Work 2015;40(2):125-133. [doi: 10.1093/hsw/hlv007] [Medline: 26027421]
34. Demiris G, Washington K, Ulrich CM, Popescu M, Oliver DP. Innovative tools to support family caregivers of persons with cancer: the role of information technology. Semin Oncol Nurs 2019;35(4):384-388 [FREE Full text] [doi: 10.1016/j.soncn.2019.06.013] [Medline: 31229338]
35. van Uden-Kraan CF, Drossaert CHC, Taal E, Shaw BR, Seydel ER, van de Laar MAFJ. Empowering processes and outcomes of participation in online support groups for patients with breast cancer, arthritis, or fibromyalgia. Qual Health Res 2008;18(3):405-517. [doi: 10.1177/1049732307313429] [Medline: 18235163]
36. Han JY, Hou J, Kim E, Gustafson DH. Lurking as an active participation process: a longitudinal investigation of engagement with an online cancer support group. Health Commun 2014;29(9):911-923 [FREE Full text] [doi: 10.1080/10410236.2013.816911] [Medline: 24345206]
37. Colvin J, Chenoweth L, Bold M, Harding C. Caregivers of older adults: advantages and disadvantages of internet-based social support*. Fam Relat 2003;53(1):49-57. [doi: 10.1111/j.1741-3729.2004.00008.x]
38. Hopwood J, Walker N, McDonagh L, Rait G, Walters K, Iliffe S, et al. Internet-based interventions aimed at supporting family caregivers of people with dementia: systematic review. J Med Internet Res 2018;20(6):e216 [FREE Full text] [doi: 10.2196/jmir.9548] [Medline: 29895512]
39. Lee E. Do technology-based support groups reduce care burden among dementia caregivers? a review. J Evid Inf Soc Work 2015;12(5):474-487. [doi: 10.1080/15433714.2014.930362] [Medline: 25794367]
40. Sanders EBN, Stappers PJ. Co-creation and the new landscapes of design. CoDesign 2008;4(1):5-18 [FREE Full text] [doi: 10.1080/15710880701875068]
41. Van Citters AD. CC BY 4.0 DEED attribution 4.0 international. Creative Commons. 2023. URL: https://creativecommons.org/licenses/by/4.0/ [accessed 2024-04-10]
42. Tips and tools. The Dartmouth Institute for Health Policy & Clinical Practice. URL: https://sites.dartmouth.edu/coproduction/knowledge-center/tools/ [accessed 2024-04-26]
43. Stickdorn M, Schneider J. This is Service Design Thinking: Basics, Tools, Cases. Amsterdam: BIS Publishers; 2011.
44. Buede DM, Miller WD. The Engineering Design of Systems: Models and Methods. Hoboken, NJ: John Wiley & Sons, Inc; 2016.
45. Khayal IS. Designing technology and healthcare delivery systems to support clinician and patient care experiences: a multi-stakeholder systems engineering co-design methodology. : IEEE; 2019 Presented at: 2019 IEEE International

Symposium on Technology and Society (ISTAS); November 15-16, 2019; Medford, MA, USA p. 1-6. [doi: 10.1109/istas48451.2019.8937932]
46. Kelley AS, Bollens-Lund E. Identifying the population with serious illness: the "denominator" challenge. J Palliat Med 2018;21(S2):S7-S16 [FREE Full text] [doi: 10.1089/jpm.2017.0548] [Medline: 29125784]
47. Khayal IS, Barnato AE. What is in the palliative care 'syringe'? a systems perspective. Soc Sci Med 2022;305:115069 [FREE Full text] [doi: 10.1016/j.socscimed.2022.115069] [Medline: 35691210]
48. Rohde M. Sketching: the visual thinking power tool. A List Apart. 2011. URL: https://alistapart.com/article/sketching-the-visual-thinking-power-tool/ [accessed 2024-04-10]
49. Agile Business Consortium. Chapter 10: MoSCoW prioritisation. DSDM Project Framework. 2014. URL: https://www.agilebusiness.org/dsdm-project-framework/moscow-prioririsation.html [accessed 2024-04-10]
50. Langley GJ, Nolan KM, Nolan TW, Norman CL, Provost LP, Moen RD. The Improvement Guide: A Practical Approach to Enhancing Organizational Performance, 2nd Edition. San Francisco: Jossey-Bass Publishers; 2009.
51. Hwang J, Toma CL, Chen J, Shah DV, Gustafson D, Mares ML. Effects of web-based social connectedness on older adults' depressive symptoms: a two-wave cross-lagged panel study. J Med Internet Res 2021;23(1):e21275 [FREE Full text] [doi: 10.2196/21275] [Medline: 33439143]
52. Gustafson DH, Kornfield R, Mares ML, Johnston DC, Cody OJ, Yang EF, et al. Effect of an eHealth intervention on older adults' quality of life and health-related outcomes: a randomized clinical trial. J Gen Intern Med 2022;37(3):521-530 [FREE Full text] [doi: 10.1007/s11606-021-06888-1] [Medline: 34100234]
53. Kotsenas AL, Aase L, Arce M, Timimi FK, Dacy M, Young C, et al. The social media DNA of Mayo Clinic-and health care. J Am Coll Radiol 2018;15(1 Pt B):162-166 [FREE Full text] [doi: 10.1016/j.jacr.2017.09.026] [Medline: 29128500]
54. Kotsenas AL, Arce M, Aase L, Timimi FK, Young C, Wald JT. The strategic imperative for the use of social media in health care. J Am Coll Radiol 2018;15(1 Pt B):155-161 [FREE Full text] [doi: 10.1016/j.jacr.2017.09.027] [Medline: 29128501]
55. Baumel A. Online emotional support delivered by trained volunteers: users' satisfaction and their perception of the service compared to psychotherapy. J Ment Health 2015;24(5):313-320. [doi: 10.3109/09638237.2015.1079308] [Medline: 26485198]
56. Baumel A, Correll CU, Birnbaum M. Adaptation of a peer based online emotional support program as an adjunct to treatment for people with schizophrenia-spectrum disorders. Internet Interv 2016;4:35-42 [FREE Full text] [doi: 10.1016/j.invent.2016.03.003] [Medline: 30135788]
57. Baumel A, Schueller SM. Adjusting an available online peer support platform in a program to supplement the treatment of perinatal depression and anxiety. JMIR Ment Health 2016;3(1):e11 [FREE Full text] [doi: 10.2196/mental.5335] [Medline: 27001373]
58. Wittenberg-Lyles E, Washington K, Oliver DP, Shaunfield S, Gage LA, Mooney M, et al. "It is the 'starting over' part that is so hard": using an online group to support hospice bereavement. Pall Supp Care 2014;13(2):351-357. [doi: 10.1017/s1478951513001235]
59. Bateman DR, Brady E, Wilkerson D, Yi E, Karanam Y, Callahan CM. Comparing crowdsourcing and friendsourcing: a social media-based feasibility study to support Alzheimer disease caregivers. JMIR Res Protoc 2017;6(4):e56 [FREE Full text] [doi: 10.2196/resprot.6904] [Medline: 28396304]
60. Costello RE, Anand A, Evans MJ, Dixon WG. Associations between engagement with an online health community and changes in patient activation and health care utilization: longitudinal web-based survey. J Med Internet Res 2019;21(8):e13477 [FREE Full text] [doi: 10.2196/13477] [Medline: 31469082]
61. Patel R, Smeraldi F, Abdollahyan M, Irving J, Bessant C. Analysis of mental and physical disorders associated with COVID-19 in online health forums: a natural language processing study. BMJ Open 2021;11(11):e056601 [FREE Full text] [doi: 10.1136/bmjopen-2021-056601] [Medline: 34740937]
62. Sik D, Németh R, Katona E. Topic modelling online depression forums: beyond narratives of self-objectification and self-blaming. J Ment Health 2023;32(2):386-395. [doi: 10.1080/09638237.2021.1979493] [Medline: 34582309]
63. Oliver BJ, Forcino RC, Batalden PB. Initial development of a self-assessment approach for coproduction value creation by an international community of practice. Int J Qual Health Care 2021;33(Supplement_2):ii48-ii54 [FREE Full text] [doi: 10.1093/intqhc/mzab077] [Medline: 34849960]
64. Fjeldstad Ø, Johnson JK, Margolis PA, Seid M, Höglund P, Batalden PB. Networked health care: rethinking value creation in learning health care systems. Learn Health Syst 2020;4(2):e10212 [FREE Full text] [doi: 10.1002/lrh2.10212] [Medline: 32313837]
65. O'Donnell EA, Buccelleto AC, Van Citters AD, Nelson EC, King J, Holthoff M, et al. ConnectShareCare builds connections among care partners of people with serious illness. Alzheimers Dement 2023;19(S19):e073484. [doi: 10.1002/alz.073484]
66. O'Donnell E, Van Citters A, Buccellato A, Haines J, Cullinan A, Wilson M, et al. ConnectShareCare: a peer-to-peer online facilitated support network for care partners of people living with serious illness. Northern New England Clinical Oncology Society (NNECOS). URL: https://www.nnecos.org/22-abstracts [accessed 2024-04-10]
67. Gustafson DH, McTavish F, Hawkins R, Pingree S, Arora N, Mendenhall J, et al. Computer support for elderly women with breast cancer. JAMA 1998;280(15):1305. [doi: 10.1001/jama.280.15.1305] [Medline: 9794300]

68. Gustafson DH, DuBenske LL, Namkoong K, Hawkins R, Chih M, Atwood AK, et al. An eHealth system supporting palliative care for patients with non-small cell lung cancer: a randomized trial. Cancer 2013;119(9):1744-1751 [FREE Full text] [doi: 10.1002/cncr.27939] [Medline: 23355273]

Edited by A Kushniruk; submitted 17.10.23; peer-reviewed by N Kaufman, E Kent; comments to author 08.02.24; revised version received 20.03.24; accepted 20.03.24; published 08.05.24.

Please cite as:
O'Donnell EA, Van Citters AD, Khayal IS, Wilson MM, Gustafson D, Barnato AE, Buccellato AC, Young C, Holthoff MM, Korsunskiy E, Tomlin SC, Cullinan AM, Steinbaugh AC, Hinson JJ, Johnson KR, Williams A, Thomson RM, Haines JM, Holmes AB, Bradley AD, Nelson EC, Kirkland KB
A Web-Based Peer Support Network to Help Care Partners of People With Serious Illness: Co-Design Study
JMIR Hum Factors 2024;11:e53194
URL: https://humanfactors.jmir.org/2024/1/e53194
doi:10.2196/53194
PMID:38717809

Original Paper

Gamification Approach to Provide Support About the Deferral Experience in Blood Donation: Design and Feasibility Study

Roberto Espinoza Chamorro[1], MS; Luciano H O Santos[2], PhD; Yukiko Mori[3], MD, PhD; Chang Liu[3], PhD; Goshiro Yamamoto[3], PhD; Tomohiro Kuroda[3], PhD

[1]Graduate School of Informatics, Kyoto University, Kyoto, Japan

[2]Fitting Cloud Inc., Kyoto, Japan

[3]Kyoto University Hospital, Kyoto, Japan

Corresponding Author:
Roberto Espinoza Chamorro, MS
Graduate School of Informatics
Kyoto University
36-1 Yoshida-Honmachi, Sakyo-ku
Kyoto, 606-8501
Japan
Phone: 81 075 753 7531
Email: rech91@outlook.com

Abstract

Background: Multiple studies have examined the impact of deferral on the motivation of prospective blood donors, proposing various policies and strategies to support individuals who undergo this experience. However, existing information and communications technology systems focused on blood donation have not yet integrated these ideas or provided options to assist with the deferral experience.

Objective: This study aims to propose an initial gamified design aimed at mitigating the impact of the deferral experience by addressing the drivers of awareness and knowledge, interaction and validation, and motivation. Additionally, the study explores the feasibility of implementing such a system for potential users.

Methods: We conducted a literature review focusing on the dynamics of motivation and intention related to blood donation, as well as the deferral situation and its impact on citizens. Through this review, we identified weak donor identity, lack of knowledge, and reduced motivation as key factors requiring support from appropriate interventions. These factors were then defined as our key drivers. Taking these into account, we proposed a gamification approach that incorporates concepts from the MDA framework. The aim is to stimulate the aforementioned drivers and expand the concept of contribution and identity in blood donation. For a preliminary evaluation, we designed a prototype to collect feedback on usability, usefulness, and interest regarding a potential implementation of our proposed gamification approach.

Results: Among the participants, a total of 11 citizens interacted with the app and provided feedback through our survey. They indicated that interacting with the app was relatively easy, with an average score of 4.13 out of 5 when considering the 11 tasks of interaction. The SUS results yielded a final average score of 70.91 from the participants' answers. Positive responses were received when participants were asked about liking the concept of the app (3.82), being likely to download it (3.55), and being likely to recommend it to others (3.64). Participants expressed positivity about the implementation of the design but also highlighted current shortcomings and suggested possible improvements in both functionality and usability.

Conclusions: Although deferral is a common issue in blood donation, there is a missed opportunity in existing ICT services regarding how to effectively handle such experiences. Our proposed design and implementation seem to have captured the interest of prospective users due to its perceived positive usefulness and potential. However, further confirmation is needed. Improving the design of activities that currently rely heavily on extrinsic motivation elements and integrating more social components to create an enhanced activity loop for intrinsic motivation could further increase the value of the proposed project. Future research could involve conducting a more specialized and longitudinal design evaluation with a larger sample size.

(JMIR Hum Factors 2024;11:e50086) doi:10.2196/50086

KEYWORDS

blood donation; deferral experience; Theory of Planned Behavior; Self-Determination Theory; gamification; ICT design; motivation; patient education; prototype; feasibility

Introduction

Blood Donation and Deferral Experience

Blood donation is an altruistic, socially responsible, and even a self-care activity. However, the commitment required to participate can be deterred by uncomfortable experiences, negatively affecting motivation to donate [1]. Preventing such experiences and reducing their impact could preserve the intention and willingness of citizens to continue donating blood.

One significant deterrent is the deferral experience [2], which means being disqualified from donating blood based on eligibility criteria. It is especially impactful in young and first-time donors, often resulting in abandonment [3]. Negative emotions are commonly reported during the deferral process [4]. Such negative experiences can deter potential donors when shared with their social circles [5]. Similarly, studies consistently identify negative interactions with staff, feelings of rejection, and confusion about deferral reasons as primary factors reducing return intention and motivation [6]. Some deferred donors misunderstand their deferral conditions, erroneously believing they cannot donate for longer periods or even permanently [7]. Communication and information gaps contribute to these misconceptions [8].

Challenges and Potential Solutions

To counter deferral effects, strategies such as enhanced communication, clear deferral information, and targeted recruitment show promise [4]. However, these solutions require substantial planning and resources, often unavailable to many blood centers. In that regard, information and communications technology (ICT) platforms, inspired by successful implementations in health promotion and telemedicine [9,10], could facilitate such support. As existing apps focus mostly on the donation itself and on supporting citizens for their next donation [11,12], there is an opportunity to offer unique value by tailoring systems to address the deferral experience. However, to that end, understanding the psychological responses and specific needs of deferred donors is crucial. Temporary deferrals necessitate motivation, health improvement, and eligibility [13], whereas permanently deferred donors (those unable to donate anymore) could still contribute indirectly through activities such as promoting donation.

Regarding the motivation topic, an approach that has gained notoriety is gamification, which involves the use of game-design elements in nongaming contexts [14]. Some blood donation centers worldwide use gamification, rewarding donors with badges, gifts, and certificates [15]. Furthermore, government blood donation apps in countries such as the United States and Canada have integrated gamified elements into their ICT services, aiming to boost donor motivation [16,17]. Although the impact on blood donation is yet to be studied, gamification has proven effective in therapy commitment and health self-monitoring [18,19]. Considering this, gamification holds promise for increasing motivation in the blood donation context.

However, motivation is not the only factor to consider for a possible proposal. Previous studies have identified various deterrents, stemming from deferrals, health conditions, or environmental factors, which influence citizens' future intentions and behaviors [4]. In this study, we reviewed previous findings, as well as the results of a preliminary survey, to identify pertinent topics and form the foundation of our proposed design for an ICT system that aims to support (also) deferred donors.

Theory of Planned Behavior and Extensions in Blood Donation

The Theory of Planned Behavior (TPB) [20], an extension of the Theory of Reasoned Action, asserts that specific behavior is determined by intention, influenced by attitude, subjective norm, and perceived behavioral control. In the context of blood donation, the TPB proposed that positive evaluation of the act (attitude), social expectations (subjective norm), and belief in individual control over donation (perceived behavioral control) dictate the decision to donate.

Previous studies have found that the TPB explains between 32% and 50% of the variance in intention and 27% and 36% of the variance in behavior [21,22]. To enhance predictive power, the framework was extended due to the inconsistent link of the subjective norm [23]. In blood donation, additional constructs were incorporated based on psychological differences among nondonors, novices, and repeat donors. Moral norms, descriptive norms, past behavior, and self-identity were included as predictors [20,24,25].

Systematic reviews have shown that self-efficacy, donor identity, and anticipated regret have medium positive effects on both intention and behavior. Conversely, deferral has a medium negative impact, leading to a decrease in subsequent donations among experienced donors [26,27]. Past behavior or habit explains 19% additional variance in blood donation behavior for those donating 5 times or more [28]. Habit, suggested to be context bound, is viewed as an external motivator, whereas self-identity, which pertains to one's role in society as a blood donor, is defined as an internal motivator [29-31]. Both habit and self-identity significantly influence repeat blood donation behavior, with past behavior likely forming identity [23].

Self-Determination Theory and Motivation and Gamification

As the TPB applies mostly to situational-level intentions [32], blood donation studies primarily rooted in the TPB have expanded their scope to incorporate Self-Determination Theory (SDT) in the last decade [33,34]. SDT, a theory of human behavior and personality development, emphasizes social-contextual factors supporting human growth through satisfying basic psychological needs for *competence* (effectiveness of my actions in my current environment), *relatedness* (social involvement and relation with others), and *autonomy* (internal need to be responsible for your own meaningful choices) [35]. It proposes that internally motivated behaviors persist, while external motivations can become internalized under appropriate socioenvironmental conditions [36].

SDT categorizes behavior on a continuum from amotivation (nonregulated behavior) to extrinsic motivation (external to integrated regulation) to intrinsic motivation (intrinsic regulation). Extrinsic motivation refers to acting in a certain way or doing a specific action because it leads to a separable outcome or reward. By contrast, intrinsic motivation refers to acting in a certain way or doing a specific action because the act itself is inherently satisfying. Integrating the TPB and SDT, studies have revealed that SDT's motivational orientations explained an additional 14% of the variance in blood donation intention compared with TPB-only models [37]. Amotivation had a negative direct effect on intention, while external motivation had no overall effect on intention but a positive effect on amotivation [38]. By contrast, introjected regulation had positive direct and indirect effects on intention, and autonomous motivation predicted intention directly and via attitudes, subjective norms, and perceived behavioral control [33,38].

Gamification Concepts and Frameworks

As SDT discusses the impact of motivation on behavior, it was considered the foundation for implementing gamification, as it does not aim to directly affect an outcome, but to change a target behavior (by affecting psychological factors) that can lead to that outcome [39,40]. To achieve this, the system can utilize its various design components, as outlined by the Mechanics, Dynamics, and Aesthetics Framework (MDA) [41], which served as the primary reference for our study. The framework comprises mechanics, which encompasses specific game components such as data representation and algorithms; dynamics, which refers to the interactions between these mechanics and player inputs over time; and aesthetics, which aims to elicit desirable emotional responses from players when they engage with the game system. These components are integrated to drive either extrinsic or intrinsic motivation, considering the targeted changes in human behavior [42-44].

Extrinsic motivation can drive behavior but may fade without external rewards, while intrinsic motivation leads to long-term positive effects on intention and behavior [45,46]. Thus, most gamified approaches recommend prioritizing intrinsic motivation in the design process. In that regard, users can be categorized according to the recognized characteristics and that drives them in the gamified implementations [47]: socializers (motivated by *relatedness*), free spirits (motivated by *autonomy*), achievers (motivated by *competence*), philanthropists (motivated by *purpose and meaning*), players (motivated by *rewards*), and disruptors (motivated by *change*). The players and disruptors categories can be further divided according to their behavior.

Considering the previous concepts and relationships of gamification and SDT, DiTommaso and Taylor [39] defined a framework in which they propose the following steps for design: discover the reason to gamify, identify players' profiles and motivational drivers, set up goals and objectives, describe skills and desired outcomes, and playtest among others. Another design framework with similar foundations is the Six Steps to Gamification [48], which also takes influence from the MDA. It proposes the following steps: definition of business objectives, target expected behavior, description of players, design of activity loops, do not forget the fun, and deploy appropriate tools. Although not domain specific, these adaptable frameworks can guide gamification projects and were also used for reference in our study.

Deferral Experience and Effects in Return Rate

From the literature review, we chose to focus on recurrent and impactful issues related to the deferral experience, especially the ones that aligned with the constructs from the TPB and SDT. For example, the construct of self-identity (blood donor identity in this case) from the extended TPB can be associated with the negative feelings from a deferral. More specifically, a deferral, which can generate a feeling of rejection in the unsuccessful participant [13], can threaten the citizen's self-perception as a capable blood donor (identity), as the inability to participate diminishes their possibility of building experiences and forming a habit (especially in the cases of new and young donors). Similarly, confusion and misunderstandings in deferral make a successful blood donation seem more complex and difficult than it is, affecting citizens' perceived behavioral control (TPB construct). As indicated by Gemelli et al [1] and Hillgrove et al [13] negative experiences can reduce the motivation for future involvement, particularly for long-term or permanently deferred donors, eroding their sense of self-efficacy.

To further explore the relationship between the deferral experience and intention, we also took into account the findings of a preliminary survey involving Japanese citizens [49], in which a total of 208 participants were recruited. In the survey, the dependent variable was "Intention to donate again after deferral" (a 6-point Likert scale question with the values 1=not anymore, 2=not for a while, 3=I don't know, 4=maybe, after a while, 5=yes, unless rejected again, and 6=yes, I would). Citizens were asked whether they heard or knew about the concept of deferral and whether they had experienced a deferral case, as well as their future intention in a possible deferral scenario. The results implied a possible relation between deferral and reduced intention to donate (following previous studies). However, the data also suggested a positive relation between preventive awareness of the deferral experience and intention to donate. Donors and nondonors who had knowledge about the deferral concept indicated higher intention of future participation even after a possible deferral scenario.

Objectives

Considering the literature review, we focused on recurrent issues that could be addressed with a gamification approach, taking into account the connections between the deferral experience, their issues, and motivation. The topics we chose were as follows:

- Lack of knowledge about deferral: Some of the negative feelings appear because citizens are not knowledgeable of the topic, are not retaining the information, or have misunderstood it.

- Weak donor identity: citizens feeling rejected and lacking validation.
- Reduced motivation: citizens losing interest in addressing the deferral reason or losing interest in contributing to the future.

Additionally, considering the evaluated strategies to mitigate the negative impact of deferral from the analyzed literature [4,50], we defined the main drivers for our approach. First, to provide awareness and knowledge about deferral by making learning interesting to the citizens (*awareness and knowledge*). Next, to increase the scenarios of interaction and validation for deferred donors to nurture their identity (*interaction and validation*). Lastly, to provide motivational drivers for deferred donors to regularly engage in activities related to blood donation (*motivation*).

After that, we worked on the design of activities that could be implemented with the gamification framework while targeting the drivers selected regarding the deferral experience. After completing the initial design, we implemented a prototype with basic features and integration for a feasibility study, collecting feedback about the usability and receptivity of potential users to discuss the future value of the idea of offering a service regarding the deferral experience, our proposed design, and its implementation.

For this study, we explored the following research questions (RQs):

- RQ1: Will our gamified design that focuses on the previously mentioned drivers with regard to the deferral experience in blood donation have a positive reception from potential users?
- RQ2: Will our initial prototype implementation of the design be considered usable and useful in its current iteration?

Methods

Conceptualizing a Gamification Approach for the Deferral Experience

Overview

In this study, we are adopting an approach similar to the gamification frameworks mentioned previously [40,41,48], while also taking into account the unique requirements of individuals in blood donation. We have adapted the steps and elements of these frameworks to provide support specifically addressing the deferral experience and focusing on the main drivers mentioned.

Definition of Approach Objectives

We redefined our target users to include not only deferred donors but also regular donors and potential donors who might face deferral in the future. Our focus broadened to cater to anyone interested in the topic, aligning with our objective of providing deferral support. We concentrated on 3 main issues: lack of knowledge about deferral, weak donor identity, and reduced motivation, translating these into drivers for our gamification approach: *awarenessand knowledge*, *interaction and validation*, and *motivation*.

Target Expected Behavior

The next step was to define the citizens’ expected behavior when interacting with our proposed gamification implementation. For our approach, we wanted the design to nurture the drivers, and as a consequence, possibly affect future intentions.

For *awareness and knowledge*, we expected users to engage in educational activities that both teach them about and test their understanding of the deferral experience and strategies for improvement. For *interaction and validation*, we expected users to get involved in discussions, in sharing experiences, and in supporting one another, improving the sense of community. In terms of *motivation*, our goal was to encourage users to access the system regularly, ideally once or twice per week, considering the prolonged pace between blood donations.

Description of Users

For our target group, while we initially expected to focus on the deferred donors, the results from the preliminary survey guided us to design the service as a preemptive one (including regular donors and nondonors), to nurture the identity of the users and prepare them against a deferral scenario. Designed primarily for young citizens (20-30 years) yet accessible to older individuals, the approach incorporated specific design elements reflecting the regional context (Japan). However, the core of the approach was intended to be adaptable, considering possible future adaptations for other regions.

In the context of the gamification approach, considering that the potential users (citizens) would not have the same goals or motivations (following the connection with SDT), for this study, we focused on targeting the players, the socializers, the free spirits, the achievers, and the philanthropists.

Design of Activity Loops

Macrolevel Progression Loops

The gamification approach aims to motivate citizens, particularly deferred blood donors, to stay engaged with blood donation–related activities. Although encouraging future donations is the ultimate goal, maintaining interest in the topic and promoting contributions to other related areas are also crucial. The design focuses on creating macrolevel progression loops for the drivers of *awareness and knowledge*, as well as *interaction and validation.*

Initial Outline

User progression is represented through levels. Levels increased based on experience points earned from various activities. Points earned could be exchanged for basic title characters. Special characters are unlocked as users progress, with higher levels requiring more points for unlocking. Higher user levels unlock additional activity options, which yield more points.

For the microlevel, we first defined some basic loops for the foundation of the design. For example, one of the initial hurdles considered was that, independent of any learning or social activity that could be designed, their value would not be achieved if the users were not motivated to access the ICT system. In that regard, we considered a simple loop of providing a reward to initially push the user to use the system: if the user logs-in to the system, they receive a message about their current streak and earn some points. Users will earn more points according to how often they connect to the app and how high is their level. With regard to *awareness and knowledge*, we aimed to make the users both learn about deferral and review their current knowledge. For this purpose, the initial idea for this loop was that as users learn more, they can face harder challenges. And the more successful they are, the more complex information they will be taught. With regard to *interaction and validation*, we aimed to provide some activities in which users could interact with other users, and the more interactions and levels the user has, the more options of interaction would be available to them.

However, we needed to solidify the ideas for the microlevel. To achieve this, we opted to elaborate on the design with greater detail. We chose to do this by following the MDA framework, first from the user perspective, then transitioning to the designer perspective to finalize the activities' design.

Definition of MDA Aesthetics

For the driver *awareness and knowledge*, we aimed to nurture a habit in the users of learning about deferral. For *interaction and validation*, the expected behavior was to generate regular engagement in the users. To that end, specifically for the players, we first selected *submission*, which means the design would allow users to interact with the system as a pastime. Our goal was to present a variety of activities offering rewards and collectible items to enhance user enjoyment. However, this approach may heavily rely on extrinsic motivations, potentially overshadowing the altruistic aspect of blood donation. Thus, we needed to be cautious in its implementation to avoid solely focusing on rewards. To address this, we selected *fellowship* as a social framework to appeal to users who value social experiences.

We considered possible ways to make users interact with others, possibly in cooperation or competition. *Challenge* (experience as obstacle course), *discovery* (experience as uncharted territory), and *expression* (experience as self-discovery) were also chosen as they are more related to intrinsic motivations, which we wanted to favor over the extrinsic motivation, which was aimed to be used only as the trigger for the conduct of the users.

Definition of MDA Dynamics

We initially drafted dynamics outlines to connect the drivers and the aesthetics. For instance, in terms of *awareness and knowledge,* users could opt to heighten the difficulty of their learning process, introducing an element of risk that could generate a *challenge.* Additionally, we explored the possibility of randomizing the information users received, with variations based on their actions within the environment, thus fostering a sense of *discovery.*

By contrast, for *interaction and validation*, we aimed for users to be able to choose the type of recognition they would get, allowing for *expression*. They should also decide what they could share with others and try to encourage them to perform certain actions, creating *fellowship*. From these initial ideas, we expanded into more detailed dynamics in the designer perspective iteration of the MDA.

Definition of MDA Mechanics

Generic mechanics are introduced, incorporating points, levels, and characters for onboarding. Points served as rewards for participating in different activities (the amount was adjusted per result), to create a sense of progression (the historical record was tracked to calculate the current level of the user), and to be used as a currency in the system. Levels were also used for progression. They increased according to the number of participations, providing recognition and incremental rewards. They were used as a certain multiplier in the activity rewards and to unlock new and special characters in the exchange store.

Characters were chosen as part of the representation and recognition of the users, being the main extrinsic reward of the gamification approach. However, they were integrated to appeal to both extrinsically and intrinsically motivated players, aiming to reduce the dependency on the extrinsic component. For example, with customization, they would target free spirits; if they were collectible, they would target players and achievers. As general rules, every registered user was provided with the same starting character; they could acquire more in the shop by exchanging the points they collected through the activities; they could also upgrade (defined as "evolve" inside the app) them by exchanging multiples of the same one. One character at a time could be selected to use it as their icon in their social activities, and characters would change their appearance if the user stood inactive in the system for more than a week.

Some social interaction components were included, such as a comment section and a simple feed wall for users' posts. Both of them had an upvote or downvote mechanism for users to indicate their relevance or popularity. A certain degree of user anonymity was incorporated to reduce possible social burdens of participants when creating content. However, for regular comments, the app showed their current character (and title) and their username. The main posts were put on hold until approved by an administrator, to reduce possibly harmful or misleading content; however, regular comments did not have this restriction. These mechanics aimed to engage the socializer, the free spirit, and the philanthropist types of users.

After this first iteration, we started with the designer perspective, in which we focused on linking all the previous concepts together, defining the more specific activities available in the system for the users.

MDA-Based Features and Feedback Loops

Finally, we describe the design of our proposed features for the gamification approach, integrating all the previous considerations and concepts.

The first feature we defined was the "Login Reward." Usage of blood donation apps tends to be low because of the timed nature of donating blood. However, to handle learning and engagement, as part of the onboarding, we chose to encourage users to interact with the system more often. To that end, we rewarded points if users log-in to the app regularly with up to

5 rewards per week, increasing the amount per consecutive access. We linked the reward to the level mechanic, providing additional points according to the level. Regarding the *awareness and knowledge* driver, we also included a message of advice and information regarding blood donation deferral. As dynamics, users could choose to access the app as usual or connect more times to increase their multiplier. Besides, as the level was linked to the rewards, users could choose to increase their level through other activities to receive more points. However, as users were not forced to read the advice message, we connected it with other activities to create the intended aesthetics and more complex activity loops.

The next features we defined, *quizzes* and *social poll*, were aimed to be connected with the broader activity loop and the *awareness and knowledge* driver. Quizzes have been implemented in other blood donation apps, so we included additional mechanics to make it less extrinsic, create new dynamics, and reach the intended aesthetics. We incorporated a life mechanic that resets daily, along with a difficulty level that becomes unlockable as users progress through levels, giving less incentive to guess the answer while also providing a higher risk-higher reward choice to more expert users. Additionally, feedback was provided according to the result, either congratulating the user for their right answer or guiding the user on the mistake. Furthermore, we connected the questions with the content shared through other features, so invested users will feel rewarded for learning on their own. Similar ideas were considered for the content of the *social poll*, but some mechanics that could allow for social interaction were included. Once per week, users could vote between different facts related to blood donation deferral, according to what they felt was the most interesting one. At the end of the week, users were notified of the most popular choice, and the ones who chose it were able to claim reward points. If desired, users could either discuss outside or through the app to try to get information about other users' preferences or to coordinate a specific choice for benefit. Additionally, previous results and facts were accessible, so users could review the content and discuss it for self-learning or connection with users.

The *news sharing* feature was also connected to the previous features and the *awareness and learning* driver, as content shared on the former would be used in the latter ones. Users could like their favorite entries and could comment about them. Comments or replies from administrative users had a special identifier while regular users had the default. Administrators would try to reply to important questions, but the content of the comments or discussions was up to the users, giving them freedom for communication.

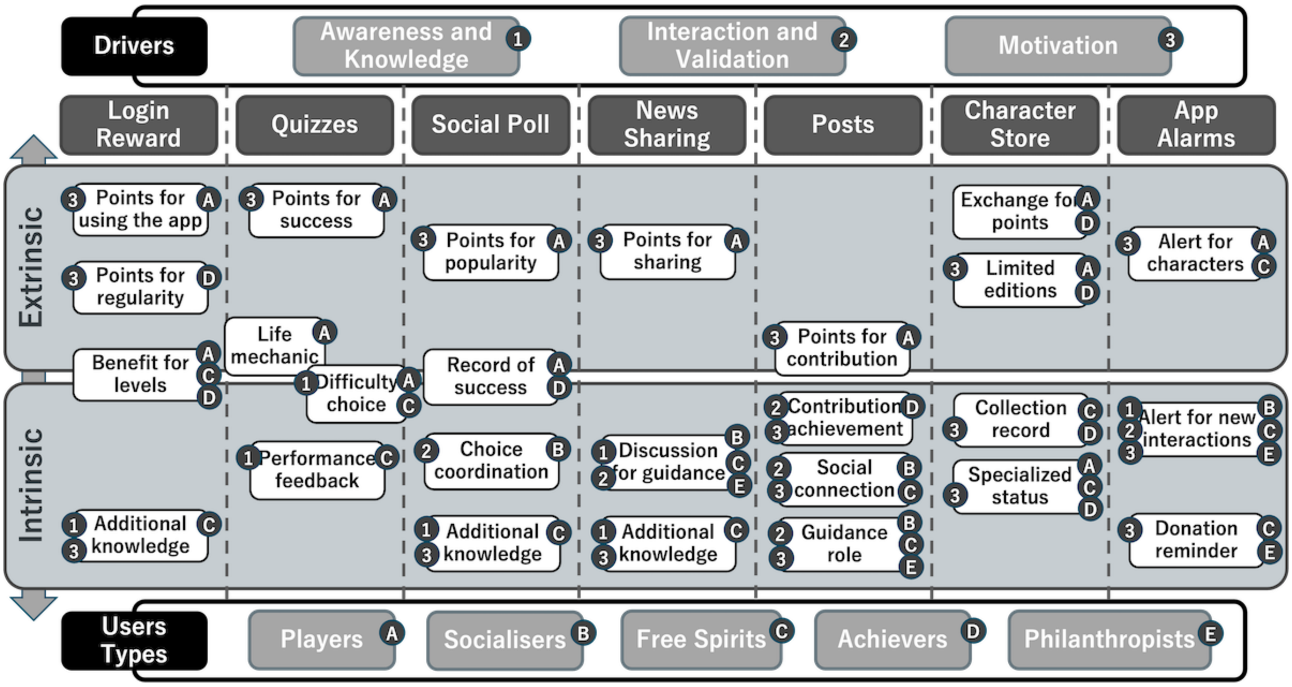
Similarly, to provide more options for the *interaction and engagement* driver, we defined the *posts* feature. Users could create posts for discussion (questions, anecdotes, suggestions, among others). If approved, the posts were shown in the app anonymously, displaying their current relevancy score. Every week, users who created new posts with high relevance would be rewarded points. While posts were defined to require approval by the administrator users, that restriction was not included in the evaluation. Posts would show in the user's feed, by order or relevancy and created date. We aimed to reward users for meaningful content, which in itself could motivate the participation of other users in the discussions. Besides, as the more relevant ones would be highlighted in the app feed, it could provide a sense of self-worth by knowing that one's content received a good reception from the community or that it provided value to the community, eventually motivating them to participate again in the discussions.

The next feature was "Application Alarms," aiming to provide users with some mechanics that could support their preferences. Users had the option to enable up to 3 types of notifications: notifications when new characters were implemented, notifications for news and discussions, and a reminder of the calculated end of the deferral period. The aim was for users to voluntarily choose to get informed about their topics of interest within the app.

Finally, for this initial scope of the design, we included features that, while not creating a proper loop by themselves, were required to connect the previous features and their loops. The first one is the "Character Store," in which users can exchange their points for available characters. The list of characters was updated in a regular schedule, with new characters being highlighted, while locked characters had a gray background. The store showed the required level, price, and current amount collected for each character. Characters being collectible were used as an extrinsic way to motivate the users to keep getting points through the other activities. By contrast, the upgrade option was linked to the title achieved by the user, which meant a special title for their effort. Users had the option to concentrate their points on either one objective or the other or to participate as much as possible to pursue both simultaneously. The other feature was the "Profile," in which users had access to their stats (level/points), their character collection (including the upgrades and selection), and additional settings for the account.

Some of the mentioned intended connections between the drivers, the users, the features, and their gamification elements can be seen in Figure 1, which provides a more general outline of what we aimed to integrate as part of the activity loops.

Figure 1. Outline of connection between drivers, users, features, and gamification elements.

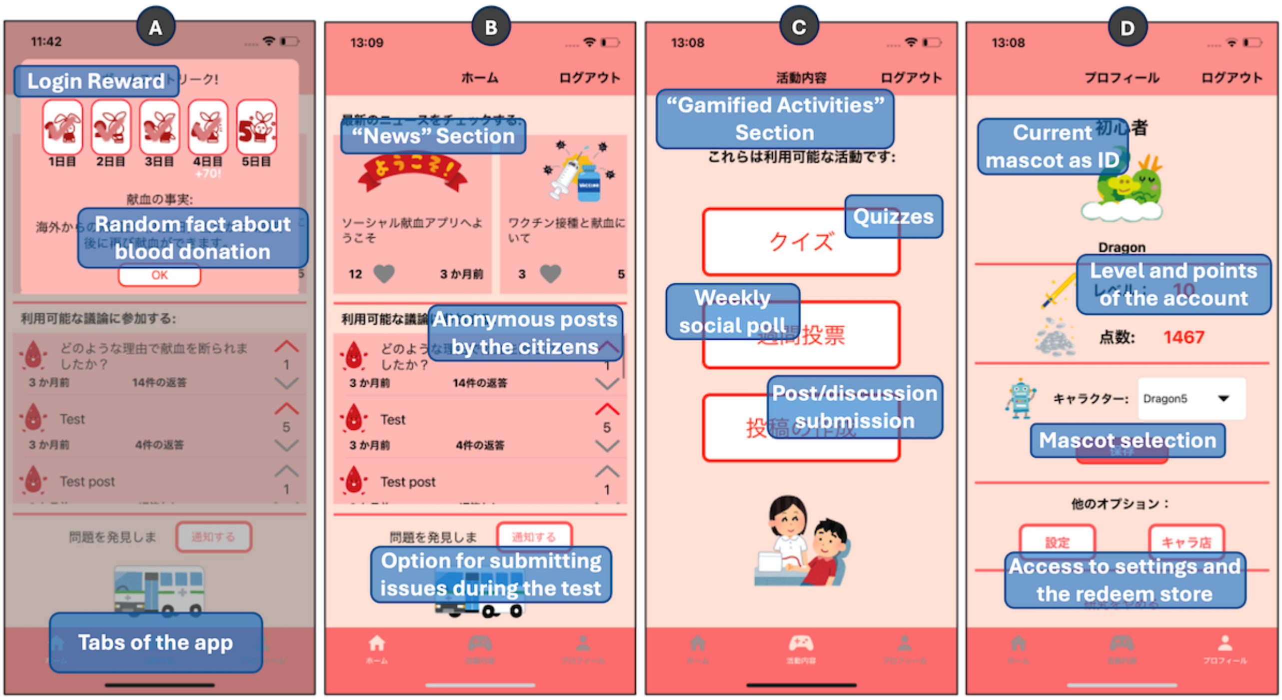


Deploy of the Concept

The proposed design was implemented in a basic mobile prototype, named as “Social Blood” app, to encapsulate the idea of a more interactive role from the citizens in blood donation.

For the icon and the other illustrations of the prototype, public domain images were selected from the Japanese web page Irasutoya [51] for the test deployment. The main screens of the app are displayed in Figure 2.

Figure 2. Screenshots of the main sections (Japanese version) of the prototype: (A) Home tab with the "Login Reward" as a pop-up; (B) Home tab screen; (C) Activity tab screen; and (D) Profile tab screen.



A welcome screen was created for user registration with either an email or Facebook account. An additional functionality (In-app Survey) not related to the design was included for data collection. The app would check for surveys requested by the researcher or the staff and ask the user to answer them. Once the pending surveys were completed, the user was redirected to the main part of the app. If the user was accessing the app for the first time in the day, the “Login Reward” feature was shown to them.

A “Home” tab was created as the main interface available to the user. This section included features related to both learning and interaction, such as “News” and “Posts.” The user could check the number of likes and votes of any entry on this screen. An additional option not related to the design, "Error Report," was also considered in this screen, to let users notify the administrators if any issue was found during the use of the app. An "Activity" tab was created to include the features that support learning. Users can access the "Quiz" and "Social Poll" features on this screen. Users could create a discussion post on this screen. Finally, a "Profile" tab was also created to show to the user information regarding the gamified elements of the app. This section of the app connects to the "Character Store" feature and to the "Settings Screen" screen, which includes the "Application Alarms" feature.

Recruitment of Participants

To collect initial feedback regarding the prototype for its usability and acceptance by prospective users, a survey was performed with volunteers recruited on social networks. A digital flyer was posted with details of contact for the interested parties (Multimedia Appendix 1). Prospective participants were required to have an iPhone (Apple Inc.), be between 20 and 50 years old, and live in Japan for at least the last two months. Participants were recruited from June 17 to July 2, 2021. No incentives were used for the recruitment. Interested citizens received a Google Form (Google LLC/Alphabet Inc.) with the informed consent details and registration (Multimedia Appendix 2). If they signed up for participation, they later received an email with the following: a link to download the app, the user manual of the app (Multimedia Appendix 3), a list of main tasks to complete inside the app (Multimedia Appendix 4), and another Google Form link that contained an anonymous survey (Multimedia Appendix 5). Participants were asked to first download the TestFlight app from the Apple Store, and from there, install and use the approved version of the research prototype for a few days. They could then follow the tasks and complete the anonymous survey either through the app or through the Google Form once they deemed their test as completed. They could test the app and submit their answers to the survey until July 11, 2021.

Ethical Considerations

The study focused on collecting preliminary feedback (usability and acceptance), so no sensible information was stored, and no risk nor effect was involved for the participants. With those points in consideration, considering the guidelines of the Kyoto University Graduate School and Faculty of Medicine Ethics Committee, it was not required to apply for ethical approval.

Evaluation Details and Data Collected

Participants were instructed to attempt to complete the list of primary tasks outlined in the prototype app (Multimedia Appendix 4). Hereafter, these tasks are referred to as follows: Register and log-in (1. Log-in), fill out the survey (2. Survey), interact with news posts (3. News), interact in the discussion posts (4. Discussion), participate in the quiz activity (5. Quiz), participate in the weekly poll activity (6. Poll), submit a simple post (7. Post), acquire a new character (8. Buy), upgrade a new character in the Character Store (9. Evolve), select a new character for your profile (10. Select), and finalize their session (11. Log out).

In the Google Form, participants were asked to answer the following sections: demographic questions (age and gender), difficulty of task completion (questions about the previously mentioned list of tasks), System Usability Scale (SUS), and follow-up questions divided into acceptance questions (a Likert scale of 5 items) and opinion questions (free-text answers). Further details of these questions are provided in Multimedia Appendix 5. Additionally, participants were given a contact email to ask for support in case they had issues during the testing.

Data Analysis

Statistics of mean, SD (σ), and standard error of the mean (σ_M) were calculated for the average of task difficulty and the final SUS results. The SUS score per participant was calculated according to the standard assignment of points per type of question, which included positive and negative values [52,53].

Qualitative answers were grouped and summarized (if possible) following a simple semantic approach: we grouped the answers for each question and summarized the main ideas according to positive or negative feedback regarding the topic of the question. For the questions regarding the status of the app, we used the labels interface, functionality, and gamification to group the answers. The same categorization was followed for the questions regarding suggestions and improvements. Additional comments were not segregated but were individually considered and described, provided they were not redundant.

The analysis of the questions was carried out after the submission deadline for participant results had elapsed. Only submissions that were completed and received before the deadline were taken into account for the final analysis.

Results

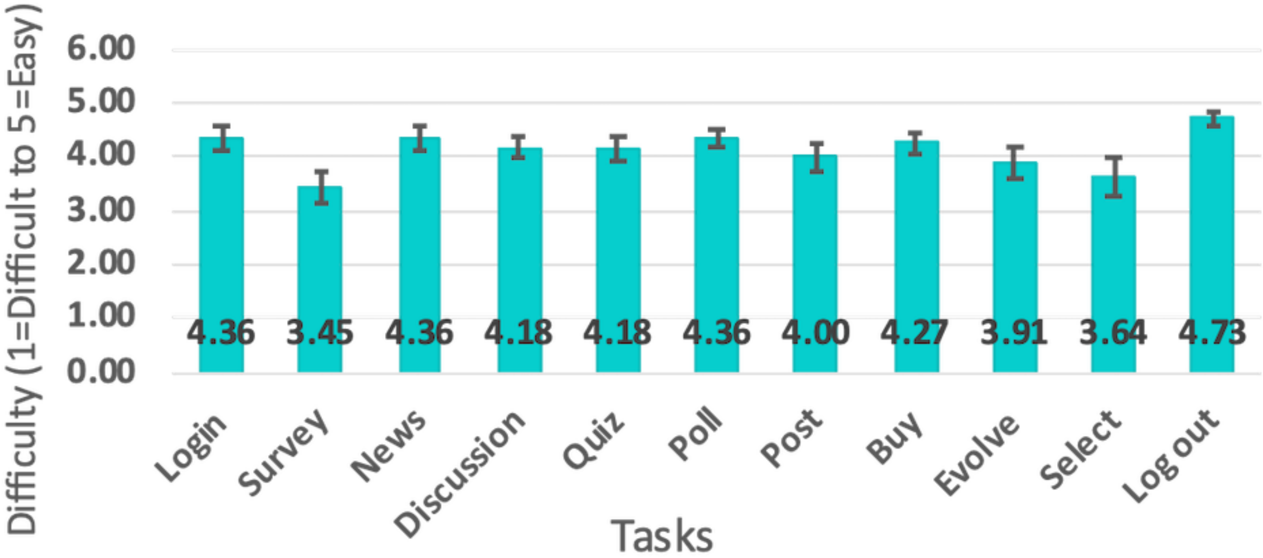
Overview

From May 17 to July 2, 15 participants were recruited for the preliminary evaluation. A total of 13 participants created an account for the app and 11 participants submitted the final form. The full content of the answers is available in Multimedia Appendix 6. Participants ranged in age from 20 to 34 years. There were 10 male and 1 female participants. Descriptive statistics were used regarding the demographic variables of the participants.

Difficulty of Task Completion

Only 4 participants asked for support during the period of the evaluation. Questions were related to tasks 2, 7, 8, 9, and 10. Only tasks 9, 10, and 11 had 1 case each of not being completed. Regarding the difficulty level of each task, the results indicated that all of them were relatively easy to complete (4.13 average). The tasks that were considered the most difficult were task 2 (Completing the in-app surveys) and task 10 (Selecting a new character), as shown in Figure 3.

Figure 3. Subjective results of difficulty of each task.



System Usability Score

The SUS evaluation of our proposed app, as seen in Multimedia Appendix 7, showed a final score of 70.91 (scale 0-100, with 100 being the best usability), slightly above the average SUS score of 68 (C grade, percentile range of 41-59). The highest SUS score received by participants was 95, while the lowest score was 30.

Regarding the score per question, item 1 (“I think that I would like to use this application frequently”) showed the average lowest score from all the lists, with a value of 3.18. The highest score was for items 3 (“I thought the application was easy to use”) and 7 (“I would imagine that most people would learn to use this application very quickly”), with a value of 4.00.

Follow-Up Questions

Acceptance and Qualitative Questions

Regarding the acceptance questions, participants responded positively to the app, expressing interest in its concept (3.82), likelihood to download it (3.55), and likelihood to recommend it to others (3.64).

For the qualitative questions, we summarized the answers for the main topics of the survey.

Goals of the App

Most participants considered a blood donation app concept useful or helpful. From them, 2 participants highlighted the possible impact of the deferral experience. The other 2 participants focused on the service being an app as a core value of the project.

Factors That Could Motivate Usage

Three participants emphasized that being aware of how they can contribute can help maintain their motivation; 3 participants mentioned interaction with others and popularity of the app as their motivation; 3 participants focused on the gamification aspects as one of their factors; 2 participants highlighted the social components as their drivers; and 2 participants indicated possible personal benefits for motivation.

Preferences About the App

Three participants liked the interactive possibilities of the app; 2 participants indicated the Quiz as their preferred feature; 3 indicated sharing and discussing as their favorite activities; 2 mentioned liking the activities involving characters; 1 participant indicated to like the interaction in general; and 1 participant indicated that they liked the aesthetic of the app the most.

Weaknesses of the App

Some participants recommended support of more languages so more citizens could benefit from the app. One participant indicated that the Quiz activity required improvement but did not specify reasons. Two participants indicated that the character functionality could be improved. One participant complained about the compulsory survey in the app because of its duration. One participant felt that not all the gamification features were connecting well with the goal of the project.

Current Status of the App

Only 1 participant mentioned that the current features might not be sufficient to support the goals of the app. They mentioned that while the app can be used to support deferred donors, it might not motivate them to promote blood donation. The other participants provided feedback regarding adjustments or fixes for the current version of the app (Textbox 1).

Textbox 1. Feedback regarding adjustments or fixes for the current version of the app.

1. Regarding the User Interface

   Possibly change the color palette of the app or allow theme selection, as the red color might create discomfort. Implement support for the "Dark" mode, as it created issues with the color of text in the News and Post features.

2. Regarding Functionality

   Reduce the length of the In-App Survey. If possible, implement support for more languages, as it could help international students who want to donate.

3. Regarding Gamification

   Adjust the point requirement for characters, as it was too high in the test. Move the character selection option to a grid, so users can look at their whole collection when choosing. Add feedback messages in the Quiz about the points acquired.

Improvements and Suggestions

Participants were also asked about what they wanted to see for implementation in the future (Textbox 2).

Textbox 2. Participants' expectations for implementation.

1. Regarding Functionality

   Consider the inclusion of a feature to find locations for blood donations. Consider the inclusion of features to share the news and discussions on social networks. Allow to link or upload videos in the comments.

2. Regarding Gamification

   Consider adding the creation of groups or friend requests. Consider adding a "Gacha" option to acquire exclusive characters. Consider adding a ranking or certificate, similar to what is implemented in "Duolingo" [54].

Additional Comments

Some concerns about the information allowed in the Post feature were mentioned, as it could be nonrelated or harmful to the users. The usefulness of the app would be higher if medical institutions could provide information within it. It was suggested to highlight to the users the core goal of the app during the registration. It was also suggested to allow donors to know when their blood is used, as it could help to motivate them to continue to donate blood.

Discussion

Principal Findings

Current ICT services in blood donation aim to improve the citizens' experience but do not focus on the deferral experience and its effects on prospective donors. This paper contributes to the field by debating the viability of implementing a system focusing on deferral and proposing a novel design to expand the concept of contribution and identity in blood donation. Our study indicates a missed opportunity in current services related to deferral. Potential users seem interested in an app supporting them in this area, and social gamification could make the role of a blood donor more approachable. However, our results, although slightly positive, require further validation due to the limitations, leaving room for discussion regarding the gamified design and the implemented prototype.

Proposed Gamified Approach Reception and Shortcomings

- RQ1. Will our gamified design that focuses on the previously mentioned drivers with regard to the deferral experience in blood donation have a positive reception from potential users?

Participants' favorable responses (average Likert scale score of 3.82) and positive opinions about the proposed functionality gave us an initial indication that the proposed project could be beneficial for the community. These results seem to align with ideas and concepts previously discussed in other studies. Previous studies discussed the relationship between knowledge of blood donation and intention to donate blood. However, only a couple of reviewed studies evaluated the ratio of knowledge regarding deferral. From our preliminary survey in Japan [49], 33% of nondonors did not know about the concept of deferral, with an additional 11% also unaware of the concept. Similar results were shown in [55], in which 90% of the participants never heard about the "donor deferral" term. This unawareness regarding deferral could be related to the positive response from the participants in our project, as either it introduced them to a new but relevant concept or it emerged as a service that could be valuable because of the low level of current support, which can be considered from their answers in the open questions. Furthermore, as participants expressed their positive intention to download the app (3.55) and to recommend it to others (3.64), the results suggest that there could not only be an interest but also an emerged necessity that has not appeared before because of the lack of awareness.

However, the current data are insufficient to reach a proper conclusion about the project acceptance, not only because of the small sample but also because of the scope of the participants, as it is not a proper representation of the target population. Additionally, the positive reception from the users could have been influenced by the Hawthorne effect [56], as the participants were aware of being part of an experiment, and the topic was related to a social contribution project. In this regard, a higher-scale study is required for further validation and analysis to decrease the effects of noise in the data and allow for more significant results.

Regarding the gamification aspect and its value, while the mentioned results were positive, their approval could have been related more to the goal of providing support. We delved into the comments of the participants about the design itself for possible conclusions. When asked about motivation to use the app and its best feature, some participants did indicate that the gamified aspects caught their interest and could even be driving motivators, highlighting the characters as part of it. These answers seem to suggest that the gamified components can play a role in, at least, capturing the interest of potential users. However, more detailed data are required to determine how beneficial is the integration of the gamified concepts in our proposed project. For example, asking participants for specific reasons why a gamification feature seems motivating to them or why it might feel discouraging. Besides, an additional evaluation regarding the impact is considered, as the value of the proposal can be confirmed if a positive effect can be determined. Comparing which features have more or less effect could also be important, as it could allow for the identification of factors to consider for future ICT-gamified implementations in blood donation.

Currently, while we discussed the importance of integrating the type of users, their motivations, and the MDA elements to nurture the drivers of interest, we have no specific data to indicate if our design has the desired effect or not. The data limitation becomes important with our goal of nurturing the intrinsic motivation of the citizens (prospective users), as we cannot recognize if the potential interest is related to the components that nurture the intrinsic motivation or the ones that do so for the extrinsic motivation. Analyzing some of the comments, most of the positive focus was on the *quiz* and the *characters*. These features, although designed considering an activity loop that could nurture intrinsic motivation, might not reach that goal in their current state. This weakness appears to be echoed in the feedback from 1 participant, who expressed dissatisfaction with the current state of the app, feeling that it falls short of achieving our design goals and lacks sufficient integration of features. As some participants showed interest in the social activities of the design (which are more related to intrinsic motivation), it might be worth it to redesign the current gamified activities to incorporate and integrate social components as part of the progress of the users.

We previously mentioned that some restrictions should be considered in a system related to blood donation, as some interactions could clash with the altruistic nature behind the donation act. To address that complexity, having a deeper understanding of game design itself is required. Learning from different and successful implementations of player interactions in game environments can lead us to a design that can properly nurture prospective blood donors' social motivation. From the case studies in *The Gamification of Learning and Instruction Fieldbook* [57], an interesting idea is the implementation of specific types of leaderboards that encourage various forms of participation, thereby creating a stronger activity loop. Building on that concept, although we aim to steer clear of incentivizing competition in donation participation, we could adapt similar interactive mechanisms to enhance engagement in learning activities. For instance, in the Quiz activity, introducing a monthly leaderboard alongside corresponding achievements could offer users more personalized motivation compared with simply rewarding points. Another intriguing option could involve allowing users to accumulate questions they have answered correctly, which they could then use in a soft-competition interaction. In this scenario, users could anonymously challenge others using their question collections until their opponent provides an incorrect answer. With this revised structure, points serve as the initial incentive to engage with the Quiz feature. However, the interaction with others serves as an intrinsic motivator, encouraging users to strive for higher-difficulty questions to challenge others. Additionally, users may be motivated to continue learning or recalling information to avoid losing in these interactions.

Applying a similar rework infused with deeper game design insights could greatly enhance the experience for prospective users. However, before this step, gathering additional data on the project's reception and soliciting input from more citizens would be invaluable. This information will help define the direction for implementing gamification strategies to encourage blood donation participation.

Prototype Implementation Usability and Usefulness

- RQ2. Will our initial prototype implementation of the design be considered usable and useful in its current iteration?

Based on the initial average SUS score of 70.91, it seems that our proposed implementation is progressing in the right direction in terms of usability. Additionally, participants did not report significant issues regarding how to use the main options of the app, as they rated the difficulty level closer to "Somewhat Easy." However, similar to the reception, we cannot draw definitive conclusions due to the small sample size and the potential influence of the Hawthorne effect. Moreover, the scores may have been positively biased due to the presence of an instruction manual and the support provided. Taking these factors into account, we directed our attention to the individual responses for more in-depth discussion.

Regarding difficulty, the activities with lower scores were those related to managing the characters (*evolving and selecting*), as well as the added survey functionality. From the comments, it appears that the functionality for upgrading the characters to their additional forms is not intuitive. The issue may stem from the fact that the options are spread across different screens, making it challenging to locate and connect them. Consolidating all the actions related to character management onto a single screen, separate from the character acquisition process, could potentially make the interface easier to use. In the case of the survey, the only complaint received was regarding its length, with participants finding it too long to complete. However, it received the lowest rating among the activities, suggesting that other participants may have also encountered issues with it. We can hypothesize that, aside from the length of the activity itself, participants may have been dissatisfied with its mandatory status rather than being optional. We could enhance the data collection process within the app by integrating it with gamification concepts, offering initial extrinsic rewards to users interested in participating. Ideally, we should also establish a loop that fosters participation through intrinsic motivation. This could involve designing activities or incentives that align with users' intrinsic interests, values, or desires for personal growth or contribution. Furthermore, from specific results of the SUS score, the participant who gave the lowest score (30) cited issues with the user interface. Taking this into consideration, future implementations of the proposed design should allocate adequate time for interface functionality and compatibility tests.

Another point for analysis from the SUS results is the average score assigned to item 1 ("I think that I would like to use this application frequently") of the survey. Although participants expressed positive sentiments regarding downloading the app and recommending it to others, the responses to item 1 indicated a nearly neutral position regarding the desire to use the app frequently. Indeed, the variation in results could stem from differing perspectives among user types. Nondonors might not use the app as frequently, even if they appreciate its concept. Similarly, donors with no deferral experience might use it for reference purposes, but perhaps not as frequently as deferred donors. Another possible and simple reason could be that participants might have had different interpretations of the term "frequently." Besides the inclusion of the "user category" variable, future survey evaluations could use a support question to help identify the regularity (if either daily, weekly, or monthly, as examples) of usage of our proposed implementation.

Further data collection is still required to obtain more detailed feedback about the current implementation, as there might be additional issues or shortcomings from the usability or the usefulness that were not captured because of the small number of participants.

Limitations

The study has multiple limitations that affect the reliability and generalization of the results. The small sample size of only 11 participants from Japan limits our ability to capture the true opinions of the various groups within the target population (prospective blood donors). Nondonors, donors, and deferred donors could have different perspectives and specific improvements regarding the design. Besides, although the recruitment was performed with Japanese material, we cannot confirm that only Japanese citizens participated in the evaluation. We have to consider that, although blood donation is seen as an altruistic and social activity in general, there can be differences in how individual values contribute to society according to one's cultural background.

Recruiting participants through social networks and including ownership of an iPhone as part of the criteria may have biased the sample toward individuals with higher levels of technological literacy, potentially influencing the results of the SUS score. However, to mitigate this bias, consistent guidance materials and tasks were provided to ensure a similar starting point for all participants.

The final version of the prototype for evaluation was created within a limited timeframe and programmed solely by 1 (REC) researcher. This constraint impacted the resources available for implementing content and graphical user interface options. The workforce constraints also impacted the choice of the target system for development, leading to the selection of iOS for release due to the developer's familiarity with it. Additionally, while the introduction of elements and activities involving user donation was considered, the acquisition and integration of these data into the current iteration of the project proved infeasible due to limitations related to permissions, partnerships, and time constraints.

Not collecting quantitative results regarding the value of the gamification aspect of the proposed design represents a significant weakness in the evaluation. Furthermore, the anonymous nature of the responses prevented the possibility of soliciting more detailed explanations regarding certain qualitative answers or comments from participants about the gamified components of the app. Indeed, it is crucial to address these shortcomings in future evaluations. Planning for a recruitment process that ensures a sufficient number of participants and obtaining ethical approval are essential steps for conducting a more comprehensive evaluation.

Conclusions

ICT systems have gained significant recognition and reliability across various fields, including within the realm of blood donation. We sought to explore previous work related to deferred donors and identify areas for further improvement. In addition to providing automated services, certain ICT projects have prioritized enhancing user motivation by incorporating gamification into their design. However, upon reviewing the current literature, it became apparent that only a few, if any, of the existing systems have specifically addressed the experience of deferral or its implications. In this research, we introduced an innovative ICT gamified design and implementation aimed at addressing this overlooked issue. Additionally, we offered an initial assessment of the project's potential reception, usability, and usefulness. Further enhancements can be made to the design of activities, which currently rely primarily on extrinsic motivation elements, to incorporate more social interaction. This would create an enriched activity loop that fosters intrinsic motivation. Further research could involve a more specialized and longitudinal design evaluation with a larger sample size. Understanding which specific features or gamification elements influence citizens' intentions or behaviors regarding their role in blood donation could be crucial for future design endeavors. Moreover, it could serve as a reference point for official ICT implementations in blood donation services.

Acknowledgments

The authors thank the citizens of Kyoto for their participation and feedback regarding the prototype of the study. REC holds a Japanese Government (Monbukagakusho: MEXT) Scholarship for the Doctoral Program. This work was supported by JST (CREST grant JPMJCR21M1), Japan.

Authors' Contributions

REC performed the literature research, design of the study, development of the app, and data analysis. LHOS and YM reviewed and validated the steps of the study design, especially the preliminary evaluation of the concept. All authors contributed points for discussion and implications based on the results. Additionally, all authors participated in drafting and revising the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study recruitment flyer—Japanese version.
[PNG File , 767 KB - humanfactors_v11i1e50086_app1.png]

Multimedia Appendix 2
Informed consent—Google Form sections for recruitment.
[DOCX File , 45 KB - humanfactors_v11i1e50086_app2.docx]

Multimedia Appendix 3
App user manual—English version (from right to left).
[PDF File (Adobe PDF File), 2982 KB - humanfactors_v11i1e50086_app3.pdf]

Multimedia Appendix 4
List of tasks for the app evaluation.
[DOCX File , 20 KB - humanfactors_v11i1e50086_app4.docx]

Multimedia Appendix 5
Study main survey—acceptance, System Usability Scale (SUS), and feedback.
[DOCX File , 22 KB - humanfactors_v11i1e50086_app5.docx]

Multimedia Appendix 6
Study survey data—formatted, anonymized, and translated to English.
[XLSX File (Microsoft Excel File), 17 KB - humanfactors_v11i1e50086_app6.xlsx]

Multimedia Appendix 7
System Usability Scale (SUS) results from each user with average and SD.
[PNG File , 150 KB - humanfactors_v11i1e50086_app7.png]

References

1. Gemelli CN, Thijsen A, Van Dyke N, Masser BM, Davison TE. Emotions experienced when receiving a temporary deferral: perspectives from staff and donors. ISBT Science Series 2018 Oct 14;13(4):394-404. [doi: 10.1111/voxs.12463]
2. van Dongen A. Easy come, easy go. Retention of blood donors. Transfus Med 2015 Aug 07;25(4):227-233. [doi: 10.1111/tme.12249] [Medline: 26399971]
3. Spekman ML, van Tilburg TG, Merz E. Do deferred donors continue their donations? A large-scale register study on whole blood donor return in the Netherlands. Transfusion 2019 Dec 17;59(12):3657-3665 [FREE Full text] [doi: 10.1111/trf.15551] [Medline: 31621923]
4. Davison TE, Masser BM, Gemelli CN. Deferred and deterred: a review of literature on the impact of deferrals on blood donors. ISBT Science Series 2019 Aug 28;15(1):3-10. [doi: 10.1111/voxs.12520]
5. Tynell E, Norda R, Ekermo B, Sanner M, Andersson S, Björkman A. False-reactive microbiologic screening test results in Swedish blood donors-how big is the problem? A survey among blood centers and deferred donors. Transfusion 2007 Jan;47(1):80-89. [doi: 10.1111/j.1537-2995.2007.01067.x] [Medline: 17207234]

6. Mathew S, King M, Glynn S, Dietz S, Caswell S, Schreiber G. Opinions about donating blood among those who never gave and those who stopped: a focus group assessment. Transfusion 2007 Apr;47(4):729-735. [doi: 10.1111/j.1537-2995.2007.01177.x] [Medline: 17381633]
7. Eder A, Goldman M, Rossmann S, Waxman D, Bianco C. Selection criteria to protect the blood donor in North America and Europe: past (dogma), present (evidence), and future (hemovigilance). Transfus Med Rev 2009 Jul;23(3):205-220. [doi: 10.1016/j.tmrv.2009.03.003] [Medline: 19539875]
8. Clement M, Shehu E, Chandler T. The impact of temporary deferrals on future blood donation behavior across the donor life cycle. Transfusion 2021 Jun 29;61(6):1799-1808. [doi: 10.1111/trf.16387] [Medline: 33780022]
9. Rahimi B, Nadri H, Lotfnezhad Afshar H, Timpka T. A Systematic Review of the Technology Acceptance Model in Health Informatics. Appl Clin Inform 2018 Jul 15;9(3):604-634 [FREE Full text] [doi: 10.1055/s-0038-1668091] [Medline: 30112741]
10. Morales J, Inupakutika D, Kaghyan S, Akopian D, Yin Z, Evans M, et al. Technology-based health promotion: Current state and perspectives in emerging gig economy. Biocybern Biomed Eng 2019 Jul;39(3):825-842 [FREE Full text] [doi: 10.1016/j.bbe.2019.07.006] [Medline: 32313347]
11. Sundarde C, Jain S, Shaikh E. Advancement of Blood Donation Application. Academia. 2015 Nov. URL: https://www.academia.edu/19419947/Advancement_of_Blood_Donation_Application [accessed 2022-01-21]
12. Ouhbi S, Fernández-Alemán JL, Toval A, Idri A, Pozo JR. Free blood donation mobile applications. J Med Syst 2015 May 3;39(5):52. [doi: 10.1007/s10916-015-0228-0] [Medline: 25732077]
13. Hillgrove TL, Doherty KV, Moore VM. Understanding non-return after a temporary deferral from giving blood: a qualitative study. BMC Public Health 2012 Dec 10;12(1):1063 [FREE Full text] [doi: 10.1186/1471-2458-12-1063] [Medline: 23227817]
14. Deterding S, Sicart M, Nacke L, O'Hara K, Dixon D. Gamification. using game-design elements in non-gaming contexts. In: CHI EA '11: CHI '11 Extended Abstracts on Human Factors in Computing Systems. New York, United States: Association for Computing Machinery; 2011 May 07 Presented at: CHI '11: CHI Conference on Human Factors in Computing Systems; May 7 - 12, 2011; Vancouver, Canada p. 2425-2428 URL: https://dl.acm.org/doi/10.1145/1979742.1979575 [doi: 10.1145/1979742.1979575]
15. Huotari K, Hamari J. Defining gamification: a service marketing perspective. In: MindTrek '12: Proceeding of the 16th International Academic MindTrek Conference. New York, United States: Association for Computing Machinery; 2012 Oct 03 Presented at: AcademicMindTrek '12: International Conference on Media of the Future; 2012; Tampere Finland p. 17-22 URL: https://dl.acm.org/doi/10.1145/2393132.2393137 [doi: 10.1145/2393132.2393137]
16. Domingos DCL, Lima LFSG, Messias TF, Feijo JVL, Diniz AAR, Soares HB. Blood hero: An application for encouraging the blood donation by applying gamification. : IEEE; 2016 Aug Presented at: 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'16); August 16 - 20, 2016; Orlando, FL, USA p. 5624-5627. [doi: 10.1109/EMBC.2016.7592002]
17. Sabani AC, Manuaba IBK, Adi E. Gamification: Blood Donor Apps for IOS Devices. Journal of Games, Game Art, and Gamification 2021 Oct 19;1(1):14-26. [doi: 10.21512/jggag.v1i1.7247]
18. Xu L, Shi H, Shen M, Ni Y, Zhang X, Pang Y, et al. The Effects of mHealth-Based Gamification Interventions on Participation in Physical Activity: Systematic Review. JMIR Mhealth Uhealth 2022 Feb 03;10(2):e27794 [FREE Full text] [doi: 10.2196/27794] [Medline: 35113034]
19. Ossenbrink L, Haase T, Timpel P, Schoffer O, Scheibe M, Schmitt J, et al. Effectiveness of Digital Health Interventions Containing Game Components for the Self-management of Type 2 Diabetes: Systematic Review. JMIR Serious Games 2023 Jun 01;11:e44132 [FREE Full text] [doi: 10.2196/44132] [Medline: 37261900]
20. Ferguson E, France CR, Abraham C, Ditto B, Sheeran P. Improving blood donor recruitment and retention: integrating theoretical advances from social and behavioral science research agendas. Transfusion 2007 Nov;47(11):1999-2010. [doi: 10.1111/j.1537-2995.2007.01423.x] [Medline: 17958528]
21. Giles M, Cairns E. Blood donation and Ajzen's theory of planned behaviour: an examination of perceived behavioural control. Br J Soc Psychol 1995 Jun 06;34 (Pt 2)(2):173-188. [doi: 10.1111/j.2044-8309.1995.tb01056.x] [Medline: 7620844]
22. Khani Jeihooni A, Jormand H, Saadat N, Hatami M, Abdul Manaf R, Afzali Harsini P. The application of the theory of planned behavior to nutritional behaviors related to cardiovascular disease among the women. BMC Cardiovasc Disord 2021 Dec 07;21(1):589 [FREE Full text] [doi: 10.1186/s12872-021-02399-3] [Medline: 34876014]
23. Conner M, Armitage CJ. Extending the Theory of Planned Behavior: A Review and Avenues for Further Research. J Applied Social Pyschol 2006 Jul 31;28(15):1429-1464. [doi: 10.1111/j.1559-1816.1998.tb01685.x]
24. Robinson NG, Masser BM, White KM, Hyde MK, Terry DJ. Predicting intentions to donate blood among nondonors in Australia: an extended theory of planned behavior. Transfusion 2008 Dec 05;48(12):2559-2567. [doi: 10.1111/j.1537-2995.2008.01904.x] [Medline: 18717776]
25. Masser BM, Bednall TC, White KM, Terry D. Predicting the retention of first-time donors using an extended Theory of Planned Behavior. Transfusion 2012 Jun 18;52(6):1303-1310 [FREE Full text] [doi: 10.1111/j.1537-2995.2011.03479.x] [Medline: 22257135]
26. Masser BM, White KM, Hyde MK, Terry DJ. The psychology of blood donation: current research and future directions. Transfus Med Rev 2008 Jul;22(3):215-233. [doi: 10.1016/j.tmrv.2008.02.005] [Medline: 18572097]

27. Bagot KL, Murray AL, Masser BM. How Can We Improve Retention of the First-Time Donor? A Systematic Review of the Current Evidence. Transfus Med Rev 2016 Apr;30(2):81-91. [doi: 10.1016/j.tmrv.2016.02.002] [Medline: 26971186]
28. Bednall TC, Bove LL, Cheetham A, Murray AL. A systematic review and meta-analysis of antecedents of blood donation behavior and intentions. Soc Sci Med 2013 Nov;96:86-94. [doi: 10.1016/j.socscimed.2013.07.022] [Medline: 24034955]
29. Aarts H, Verplanken B, van Knippenberg A. Predicting Behavior From Actions in the Past: Repeated Decision Making or a Matter of Habit? J Applied Social Pyschol 2006 Jul 31;28(15):1355-1374. [doi: 10.1111/j.1559-1816.1998.tb01681.x]
30. Callero PL. Role-Identity Salience. Social Psychology Quarterly 1985 Sep;48(3):203. [doi: 10.2307/3033681]
31. Charng H, Piliavin JA, Callero PL. Role Identity and Reasoned Action in the Prediction of Repeated Behavior. Social Psychology Quarterly 1988 Dec;51(4):303. [doi: 10.2307/2786758]
32. Hagger MS, Chatzisarantis NLD. Integrating the theory of planned behaviour and self - determination theory in health behaviour: A meta - analysis. British J Health Psychol 2010 Dec 24;14(2):275-302. [doi: 10.1348/135910708x373959]
33. Williams LA, Sun J, Masser B. Integrating self - determination theory and the theory of planned behaviour to predict intention to donate blood. Transfusion Med 2018 Nov 11;29(S1):59-64. [doi: 10.1111/tme.12566]
34. Pentecost R, Arli D, Thiele S. It’s my choice! Investigating barriers to pro-social blood donating behaviour. MIP 2017 Apr 03;35(2):243-258. [doi: 10.1108/mip-03-2016-0055]
35. Deci EL. Self-determination theory. In: Ryan RM, editor. Handbook of Theories of Social Psychology. London: Sage Publications Ltd; 2011:416-436.
36. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist 2000;55(1):68-78. [doi: 10.1037/0003-066X.55.1.68]
37. France CR, France JL, Carlson BW, Frye V, Duffy L, Kessler DA, et al. Applying self-determination theory to the blood donation context: The blood donor competence, autonomy, and relatedness enhancement (Blood Donor CARE) trial. Contemp Clin Trials 2017 Feb;53:44-51 [FREE Full text] [doi: 10.1016/j.cct.2016.12.010] [Medline: 27979752]
38. M’Sallem W. Role of motivation in the return of blood donors: mediating roles of the socio-cognitive variables of the theory of planned behavior. Int Rev Public Nonprofit Mark 2021 Jun 23;19(1):153-166. [doi: 10.1007/s12208-021-00295-2]
39. DiTommaso D. Beyond gamification: Architecting engagement through game design thinking. Slideshare, a Scribd company. Nov 11. URL: https://www.slideshare.net/slideshow/beyond-gamification-architecting-engagement-through-game-design-thinking/10125282 [accessed 2024-02-13]
40. DiTommaso D, Taylor C. Beyond gamification: designing behavior change games. In: CHI PLAY '14: Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play. New York, United States: Association for Computing Machinery; 2014 Oct 19 Presented at: CHI PLAY '14: The annual symposium on Computer-Human Interaction in Play; October 19 - 21, 2014; Toronto Ontario Canada p. 475 URL: https://dl.acm.org/doi/10.1145/2658537.2662410 [doi: 10.1145/2658537.2662410]
41. Hunicke R, LeBlanc M, Zubek R. MDA: A Formal Approach to Game Design and Game Research. In: AAAI Workshop Papers 2004. 2004 Presented at: Nineteenth National Conference on Artificial Intelligence (AAAI-04); July 25 - 29, 2004; San Jose, California p. A.
42. Ryan, Deci. Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. Contemp Educ Psychol 2000 Jan;25(1):54-67. [doi: 10.1006/ceps.1999.1020] [Medline: 10620381]
43. Hassan L. Governments Should Play Games. Simulation & Gaming 2016 Dec 19;48(2):249-267. [doi: 10.1177/1046878116683581]
44. Koivisto J, Hamari J. The rise of motivational information systems: A review of gamification research. International Journal of Information Management 2019 Apr;45:191-210. [doi: 10.1016/j.ijinfomgt.2018.10.013]
45. Vallerand RJ, Blssonnette R. Intrinsic, Extrinsic, and Amotivational Styles as Predictors of Behavior: A Prospective Study. J Personality 1992 Sep 28;60(3):599-620. [doi: 10.1111/j.1467-6494.1992.tb00922.x]
46. Ryan RM, Rigby CS, Przybylski A. The Motivational Pull of Video Games: A Self-Determination Theory Approach. Motiv Emot 2006 Nov 29;30(4):344-360. [doi: 10.1007/s11031-006-9051-8]
47. Marczewski A. Introduction to gamification part 4: motivation (R.A.M.P, Maslow, SDT and more). Gamified UK. 2019. URL: https://www.gamified.uk/2019/01/30/introduction-to-gamification-part-4-motivation-r-a-m-p-maslow-sdt-and-more/ [accessed 2022-03-15]
48. Werbach K, Hunter D. For the Win: How Game Thinking Can Revolutionize Your Business. Philadelphia, PA: Wharton School Press; Oct 30, 2012.
49. Espinoza R, Hiragi S, Santos L, Morris K, Ueshima H, Yamamoto G, et al. Blood donation and deferral awareness in Japan: issues and opportunities for information and communications technology interventions. In: Proceedings of the Annual Conference of the Institute of Systems, Control and Information Engineers. 2021 May Presented at: The 65th Annual Conference of the Institute of Systems, Control and Information Engineers (SCI'21); May 26-28, 2021; Okayama University (Online) p. 1112-1116 URL: https://cir.nii.ac.jp/crid/1520853834864279168
50. Gemelli CN, Kruse SP, Thijsen A, Van Dyke N, Davison TE. Educating donors about deferrals through the provision of tailored materials: Findings from a survey of deferred donors. Transfusion 2020 Dec 07;61(3):822-829. [doi: 10.1111/trf.16212]
51. Takashi M. Irasutoya. URL: https://www.irasutoya.com [accessed 2021-03-13]

52. Sauro J, Lewis JR. Chapter 8: Standardized usability questionnaires. In: Quantifying the User experience: Practical Statistics for User Research. Burlington, MA: Morgan Kaufmann; 2012:185-240.
53. Brooke J. SUS: A 'Quick and Dirty' Usability Scale. In: Usability Evaluation In Industry. London, England: CRC Press; Nov 1996.
54. Teske K. Duolingo. CALICO 2017 Aug 24;34(3):393-401 [FREE Full text] [doi: 10.1558/cj.32509]
55. Atherley AE, Taylor CG, Whittington A, Jonker C. Knowledge, attitudes and practices towards blood donation in Barbados. Transfus Med 2016 Dec 16;26(6):415-421. [doi: 10.1111/tme.12359] [Medline: 27634655]
56. Adair JG. The Hawthorne effect: A reconsideration of the methodological artifact. Journal of Applied Psychology 1984;69(2):334-345. [doi: 10.1037//0021-9010.69.2.334]
57. Kapp KM. The Gamification of Learning and Instruction Fieldbook: Ideas into Practice. Hoboken, NJ: Wiley; 2013.

Abbreviations

ICT: information and communications technology
RQ: research question
SDT: Self-Determination Theory
SUS: System Usability Scale
TPB: Theory of Planned Behavior

Edited by A Kushniruk; submitted 01.07.23; peer-reviewed by F Gallego Duran, T Barros Pontes e Silva; comments to author 22.08.23; revised version received 31.10.23; accepted 07.04.24; published 14.06.24.

Please cite as:
Espinoza Chamorro R, Santos LHO, Mori Y, Liu C, Yamamoto G, Kuroda T
Gamification Approach to Provide Support About the Deferral Experience in Blood Donation: Design and Feasibility Study
JMIR Hum Factors 2024;11:e50086
URL: https://humanfactors.jmir.org/2024/1/e50086
doi:10.2196/50086
PMID:38875005

Original Paper

Application of an Adapted Health Action Process Approach Model to Predict Engagement With a Digital Mental Health Website: Cross-Sectional Study

Julien Rouvere[1], MA; Brittany E Blanchard[1], PhD; Morgan Johnson[2], MS; Isabell Griffith Fillipo[1], BA; Brittany Mosser[1], MSW, LICSW; Meghan Romanelli[3], PhD, LCSW; Theresa Nguyen[4], LCSW; Kevin Rushton[4], BS; John Marion[4], MA; Tim Althoff[5], PhD; Patricia A Areán[6], PhD; Michael D Pullmann[1], PhD

[1]Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, WA, United States

[2]Merck, Upper Gwynedd, PA, United States

[3]School of Social Work, University of Washington, Seattle, WA, United States

[4]Mental Health America, Alexandria, VA, United States

[5]Allen School of Computer Science & Engineering, University of Washington, Seattle, WA, United States

[6]NIMH, Bethesda, MD, United States

Corresponding Author:
Julien Rouvere, MA
Department of Psychiatry and Behavioral Sciences
University of Washington School of Medicine
1959 NE Pacific Street
Box 356560
Seattle, WA, 98195-6560
United States
Phone: 1 206 221 5498
Email: rouvere@uw.edu

Abstract

Background: Digital Mental Health (DMH) tools are an effective, readily accessible, and affordable form of mental health support. However, sustained engagement with DMH is suboptimal, with limited research on DMH engagement. The Health Action Process Approach (HAPA) is an empirically supported theory of health behavior adoption and maintenance. Whether this model also explains DMH tool engagement remains unknown.

Objective: This study examined whether an adapted HAPA model predicted engagement with DMH via a self-guided website.

Methods: Visitors to the Mental Health America (MHA) website were invited to complete a brief survey measuring HAPA constructs. This cross-sectional study tested the adapted HAPA model with data collected using voluntary response sampling from 16,078 sessions (15,619 unique IP addresses from United States residents) on the MHA website from October 2021 through February 2022. Model fit was examined via structural equation modeling in predicting two engagement outcomes: (1) choice to engage with DMH (ie, spending 3 or more seconds on an MHA page, excluding screening pages) and (2) level of engagement (ie, time spent on MHA pages and number of pages visited, both excluding screening pages).

Results: Participants chose to engage with the MHA website in 94.3% (15,161/16,078) of the sessions. Perceived need (β=.66; *P*<.001), outcome expectancies (β=.49; *P*<.001), self-efficacy (β=.44; *P*<.001), and perceived risk (β=.17-.18; *P*<.001) significantly predicted intention, and intention (β=.77; *P*<.001) significantly predicted planning. Planning was not significantly associated with choice to engage (β=.03; *P*=.18). Within participants who chose to engage, the association between planning with level of engagement was statistically significant (β=.12; *P*<.001). Model fit indices for both engagement outcomes were poor, with the adapted HAPA model accounting for only 0.1% and 1.4% of the variance in choice to engage and level of engagement, respectively.

Conclusions: Our data suggest that the HAPA model did not predict engagement with DMH via a self-guided website. More research is needed to identify appropriate theoretical frameworks and practical strategies (eg, digital design) to optimize DMH tool engagement.

(JMIR Hum Factors 2024;11:e57082) doi:10.2196/57082

KEYWORDS

Health Action Process Approach (HAPA); digital health; health behavior; Mental Health America (MHA); digital mental health engagement; mental health website

Introduction

Background

Access to mental health support is poor, with barriers including mental health stigma and limited time, availability, and accessibility [1]. According to the National Institute of Mental Health, it was estimated that 57.8 million adults in the United States lived with a mental illness in 2021 [2]. Digital Mental Health (DMH) is seen as the next generation of mental health care, in that it is readily accessible to most, affords anonymity when appropriate, and can be used when a consumer needs it, rather than when a provider is available. It serves a particularly valuable role where services are scarce, with studies finding disproportionate use of online mental health screening tools in rural areas [3].

In general, engagement with DMH is poor [4-7]. When using mobile apps, consumers in 1 study disengaged in as little as 2 weeks [8], and another study found a median 15-day retention rate of 3.9% [9]. However, there is limited research on engagement with mental health focused websites, which likely have different usage patterns than other approaches to DMH, such as mobile apps or digitally mediated therapy. Mental health websites are typically free, usually used for short-term information gathering, and may include a wide variety of self-guided resources such as psychoeducation articles, self-administered screening tools for common mental health problems, links to find therapists, downloadable self-monitoring tools (eg, medication management charts), ways to connect with peer support groups, and interactive self-help tools (eg, cognitive reframing activities using artificial intelligence). One website developed and maintained by Mental Health America (MHA) [10] features all of these DMH resources contained within approximately 500 pages, is free and openly accessible, and is visited by 9-10 million people a year, with 6.5 million using self-administered screening tools and 3.5 million accessing other resources. However, resources other than screening tools are underutilized. In particular, those who self-administer mental health screening tools are unlikely to engage with the other valuable resources on the website. For instance, in 2021, only 70% of those who used the website to self-administer a screening tool visited a nonscreening-related page on MHA. Only 15% subsequently read any articles despite the valuable psychoeducation and other materials available that could help them improve their emotions and behavior (Mental Health America, unpublished data, January 2024).

We partnered with MHA to conduct collaborative research with the goal of increasing optimal engagement with DMH resources on their website. We defined optimal engagement with DMH as engaging with resources that are aligned with users' needs and interests and motivate users toward taking positive mental health actions. The Health Action Process Approach (HAPA) is a theoretical framework aligned with this goal. The HAPA model has been described in detail elsewhere [11-13], but here we describe it as applied to DMH tool engagement. The HAPA model is separated into 2 phases: *motivational* (when *intentions* are developed) and *volitional* (when *health actions* occur). During the motivational phase, preintention factors shape behavioral intention. Several mechanisms operate to support or hinder progression through the HAPA stages of health behavior change. Movement from preintention to intention is influenced by *perceived need* to engage in a health action [14], *task self-efficacy*, the degree to which consumers believe that they can engage in behavior change, *perceived risks* of not acting, and *outcome expectancies*, the positive and negative anticipated consequences of engaging in a behavior. If consumers have a high perceived need for change, high task self-efficacy, positive outcome expectancies of engaging with a DMH tool, and high perceived risks of not acting, they are more likely to move through the 2 stages of the motivational phase: *preintention* to *intention*. If task self-efficacy, perceived needs and risks, or outcome expectancies are low, then the consumer will be unlikely to set intentions. The volitional phase describes the process by which a consumer moves through the action stages: *behavioral intention* to *planning* to *acting* (ie, implementing a health behavior; in this case, engaging with DMH). Consumers with high intentions are more likely to implement a health behavior if they engage in planning. Other mechanisms in the traditional HAPA model include *maintenance* and *recovery self-efficacy* (the belief in the ability to continue a health behavior in the face of challenges, or restart if stopped) and *coping planning* (plans for how to continue an action in the face of challenges). These mechanisms are less relevant for the current research because they are focused on ongoing maintenance of behavioral actions. The HAPA model has robust empirical support in predicting health behavior change, such as dietary behaviors and medication adherence [15,16]. For example, 1 study found that medication adherence was significantly predicted by intention, task self-efficacy, coping self-efficacy, and coping planning in a sample of patients with type 2 diabetes [16].

We aimed to assess the applicability of the HAPA model to DMH engagement on a mental health website. We conducted the current research to assess whether the HAPA model fit our data and to examine the strength with which HAPA model constructs (eg, self-efficacy, perceived need, and perceived risks) were related to DMH engagement. This information would be used to develop and test tailored engagement strategies that match stage (eg, preintention, intention) and mechanism (eg, self-efficacy, perceived need) to the individual to improve DMH engagement. Tailored nudges (ie, strategies using a cue, such as a message, to influence user behavior) using this language could be developed and tested to help consumers move through the stages of behavior change (ie, from preintention or intention into the volitional stage of acting) to act on health-promoting behaviors (eg, through self-help tools, therapy, and building social connections). For instance, upon completing a screening tool, an MHA website visitor who endorses low self-efficacy could be shown language intended to increase self-efficacy to thereby increase their motivation to transition from self-administering screenings to initiating self-help or professional treatment.

There is precedent for applying the HAPA model to develop stage-matched interventions. Lippke et al [17] found that for intenders, but not preintenders, a physical activity planning intervention produced changes in intention and planning which were subsequently associated with increased physical activity. Self-efficacy, which was not targeted by the intervention, was not impacted by the intervention and not associated with physical activity. The authors concluded that interventions will be more successful when tailored to the participant's stage of change, with intenders being more receptive to volition-based interventions as predicted by the HAPA model. Similar findings have been found when developing interventions tailored to stage of change and HAPA mechanisms. A study on hand hygiene behavior in hospital units found that interventions tailored to empirically assessed HAPA components (eg, providing information on the risks of not acting for units that scored low on risk perceptions) were associated with a decrease in infections, whereas units that received untailored interventions did not experience a decrease in infections [18].

Previous work has described how the HAPA model can be adapted for ongoing digitally delivered psychotherapy [19], but the use of DMH self-guided websites is likely to have different engagement patterns and associations. Website usage is likely to be more sporadic, short-term, and used by people in the preintention and intention stages, as they are determining what kinds of actions they want to take. In addition, the operationalization of a health action on a DMH website includes actions such as clicking on a navigation page, reading an article about depression, or using a web-based single-session intervention. Planning a health action in this context may be a seamless and instantaneous process that is difficult to separate from intention. Intention to act, planning, and action (clicking a link) may occur within a time span ranging from seconds to minutes, rather than days to weeks as in usual applications of the HAPA model [12,20-25]. Furthermore, it is unclear whether or how principles of marketing (strong call to actions) or user design (color and placement) determine user behavior as compared with the HAPA model.

A meta-analysis of 95 studies using the HAPA model [13] provides a summary of empirical literature applying the HAPA model to a wide variety of health behavior outcomes, such as dietary behavior and physical activity [26]. Self-efficacy and outcome expectancies were moderately correlated with intention, and risk perception effect sizes were small but statistically significant. In the meta-analytic path analyses, self-efficacy, intention, and action planning significantly predicted outcomes. Due to the small and nonsignificant role of risk perception, the authors suggested that this mechanism may play only a minor role in predicting health behaviors.

Only 3 (3.2%) of the studies included in the meta-analysis used an observed behavioral measure. The remaining studies used self-report measures of outcomes, and none of the studies examined DMH. Because using self-report measures for all HAPA constructs and outcomes is likely to inflate associations due to method bias, observable behavioral outcomes are necessary to reduce these biases [27]. In addition, most studies had a greater than 4-week time lag between the measurement of HAPA constructs and behavior. Whether the HAPA model is relevant for behaviors that occur with lags of seconds or minutes rather than weeks or months remains unknown.

An additional study not included in the meta-analysis examined the HAPA model in a DMH context and included both self-report and observed measures of engagement [14]. This study of engagement with a web-based intervention for trauma had mixed findings: self-reported engagement was associated with HAPA motivational constructs but observed engagement was not. These findings support the idea that self-reported engagement is more strongly associated with self-reported HAPA constructs than objective measures and that planning may be most effective for those with low self-efficacy.

In sum, several gaps exist in the HAPA model literature. Studies have relied primarily on self-report measures, with so few including observed or objective measures that a moderator comparison of self-report and objective measures could not be conducted in the meta-analysis. In addition, no studies applying the HAPA model to DMH were included in the meta-analysis because so few exist. We are aware of only 1 previous study that applied the HAPA model to DMH, which found attenuated effects when using an objective measure for engagement behavior [14]. Therefore, more research is needed to determine the strength of the HAPA model in a DMH context and which uses objective measures of behavioral outcomes.

Aim of This Study

The purpose of this cross-sectional study was to test whether the HAPA model predicted DMH engagement and to determine which HAPA model constructs were most strongly associated with DMH engagement. In doing so, this manuscript contributes to the broader literature on the HAPA model by focusing on DMH engagement, employing an objective measure of health behavior, and measuring outcomes within seconds or minutes between the motivational and volitional phases of the HAPA model. In addition to contributing novel insight toward understanding engagement with DMH via the HAPA model, these findings could be applied to strategies to increase engagement by focusing on HAPA constructs in website design.

Methods

Participants

Data were collected via voluntary response sampling. Participants were a naturalistic sample of people who self-administered a mental health screener on the MHA website between October 2021 and February 2022. The inclusion criteria were clicking on the "Next Steps" survey and completing it for their own needs (as opposed to reporting that it was completed for others, such as their children). The exclusion criteria were completing less than 70% of the Next Steps survey or taking the "parent test" or "youth screener" (Pediatric Symptom Checklist [28,29]); there was no age restriction. The final analytic sample included data from 16,078 sessions (15,619 unique IP addresses from US residents) on MHA.

Procedure

Users of the MHA website can self-administer any of several evidence-based mental health screening measures for depression, postpartum depression for new and expecting parents, anxiety, attention-deficit/hyperactivity disorder, psychosis or schizophrenia-spectrum disorders, posttraumatic stress disorder, eating disorders, substance or behavioral addictions, and youth mental health (separate versions for youth or caregivers to complete). Participants could click on any of these tests and take the online screener; for example, clicking "Depression Test" showed users the Patient Health Questionnaire-9 [30]. This study focuses on data collected from the Next Steps survey, which was an option participants could choose among these screeners and was also provided via a link on the results pages for screening measures. Participants could complete as many screeners as desired. After completion of each screener, including the Next Steps survey, the participants were shown an optional demographics survey. Data submissions were chunked into sessions, which were defined as all website interactions via a unique IP address bounded by a gap of at least 30 minutes.

Ethical Considerations

All materials and procedures for this study were approved by the University of Washington Institutional Review Board (STUDY00010958). All data were protected under the University of Washington Institutional Review Board. Data sharing between MHA and the University of Washington followed a Data Security Protocol, which required data to be deidentified via random identifiers and stored securely on an encrypted institutional server accessible only by authorized research team members. The study's external Data Safety and Monitoring Board oversaw study progress and participant safety. The Institutional Review Board and Data Safety and Monitoring Board determined that there were no anticipated risks for participants and no further application for ethical approval was required.

Measures

Demographic Survey

The demographic survey included items about age range, race, ethnicity, and gender identity. A checkbox was provided for participants to indicate whether they identify as transgender. If participants completed more than 1 demographic measure during a session and these values differed, which was rare (<3%), we report "more than 1 answer given," which should be distinguished from "more than 1 of the above." For instance, an IP address that completed only 1 demographics measure and endorsed "Asian" and "Black" was categorized as "more than 1 of the above," whereas an IP address that reported "Asian" on 1 demographics measure and "Black" on another screener was categorized as "more than 1 answer given." It is possible that multiple people (eg, in the same household) may have completed the surveys from the same IP address.

Next Steps Survey

The Next Steps Survey was developed by the research team at the University of Washington collaboratively with MHA (Multimedia Appendix 1). Because this was a naturalistic study and participants were not incentivized to participate, the team aimed to use measures that were pragmatic, brief, tailored to the website activities, and consistent with MHA's values and approach. No existing measures were suitable, as studies using the HAPA model often apply individualized items due to the study-specific nature of the constructs. Therefore, items were developed by the study team by adapting existing items developed and tested in other studies of the association between HAPA constructs and behavior [13,14,20,22,31-34]. The resulting 22-item survey designed to assess HAPA model constructs in the context of DMH consisted of four 5-item subscales (outcome expectancies, intention, self-efficacy, and planning) and 2 single items for perceived risk and perceived need. The 5-item subscales comprised questions that asked about the five types of resources available on the MHA website: (1) learning more about mental health (LMH), (2) connecting with others (CO) who have mental health conditions, (3) learning about treatment options (LTO), (4) receiving mental health treatment (RMH), and (5) using online self-help tools (ONL).

Outcome expectancies were measured by participant ratings on their belief that the 5 types of resources would improve their mental health on a 4-point scale ranging from 0 (*Definitely won't help*) to 3 (*Definitely will help*). Intention items assessed whether participants intended to act on the 5 resources on a scale of 0 (*I definitely will not do this*) to 3 (*I definitely will do this*). Self-efficacy items asked participants to rate their confidence in their ability to act on the 5 items on a scale of 0 (*Not at all confident*) to 3 (*Very confident*). Planning items were yes/no questions that asked participants whether they had a specific plan to take any of those actions.

Perceived need was measured on a scale of 0 (*No*), 1 (*I don't know*), and 2 (*Yes*). Perceived risk was measured on a 4-point scale from 0 (*No, it will get better on its own*) to 3 (*Yes, it will get much worse*). If participants completed more than 1 Next Steps survey during a single session, we report on data provided during the first completion within that session.

Engagement With Web-Based Content

Choice to engage was a dichotomous outcome indicating that participants visited any web pages, excluding screening pages. To correct for extremely brief web page visits that did not represent true engagement, based on literature that shows that users evaluate website design within seconds [35], we defined engagement as requiring at least 3 seconds on the MHA website excluding time spent on screening pages.

We operationalized level of engagement with 2 indicators: number of pages visited, or engaged pages (EP), and total engaged minutes (EM). EP was computed by summing the total number of web pages visited on the MHA website excluding screening pages. EM was computed by summing the total amount of time spent on MHA's website excluding time spent on screening pages. Because time spent per page was computed from when a page on MHA was loaded to the time participants clicked on a link to another page on MHA, time spent on the final page viewed (eg, before closing a browser or visiting another website) could not be included in EM.

Analyses

The goals of these analyses were to examine whether the HAPA model predicted engagement with web-based resources and which HAPA constructs were significant predictors of website engagement. We did this by testing the model fit of the adapted HAPA model to our data using structural equation modeling (SEM).

Data were collected from 21,329 sessions in which participants took the Next Steps survey. Of these, 75.9% (16,179/21,329) completed the Next Steps survey for their own needs and did not take the Pediatric Symptom Checklist. Analyses included data from only those participants who completed at least 70% of the 22 Next Steps survey items (N=16,078). Because participants completed the Next Steps survey during each session and may have exhibited different engagement patterns after ≥30 minutes of inactivity on the website, IP addresses with multiple sessions (n=605) were retained in the analysis, with the number of sessions for any 1 IP address ranging from 1 to 10. Response categories with less than 10% endorsement were combined with the next adjacent response category for accurate parameter estimation. Following collapsing of response categories, proportions of endorsement for each of the Next Steps survey items were no greater than 58.9% for any 1 response option.

Prior to SEM, we examined correlations between the Next Steps survey items, EM, and EP. We used the weighted least squares means and variance–adjusted (WLSMV) estimator, which computes a tetrachoric or polychoric correlation matrix using pairwise complete observations. Next, measurement models using unidimensional confirmatory factor analysis (CFA) were assessed for each reflective latent construct (self-efficacy, outcome expectancies, intention, and planning) using model fit statistics, with factor loadings as the parameters of interest. The measurement model for level of engagement was underidentified and model fit indices could not be estimated. The adapted HAPA model was tested by incorporating the HAPA constructs and their measured variables to form the theorized model (Figure 1). This model excludes maintenance self-efficacy, recovery self-efficacy, and coping planning, as these constructs are not relevant to individuals engaging with a website during a single session. Maintenance self-efficacy refers to one's belief that they can overcome barriers to taking a health action, and coping planning refers to identifying methods to cope with those barriers. These constructs were excluded because there were no foreseeable barriers to the health action (ie, clicking a link) in the context of this study. Recovery self-efficacy refers to restarting a health behavior once it ends, which in our study would be returning to the website after usage. This is less relevant to our study as our research question is focused on an immediate behavioral outcome.

The first model predicted choice to engage, and the second analysis predicted level of engagement for those who did engage. All parameters were modeled and allowed to freely vary, and models were assessed by examining model fit indices and path coefficients. EM was modeled as a censored variable to account for the unrecorded time spent on the final page visited. Assumptions of multivariate normality and linearity were assessed in MPlus. While SEM is robust to nonnormality, recommended values of acceptable skewness range from –2 to +2, and acceptable values of kurtosis range from –7 to +7 [36,37]. Skewness and kurtosis for both EM and EP were well out of the acceptable range (EM skewness=5.19, kurtosis=40.58; EP skewness=5.74, kurtosis=92.56), indicating nonnormal distributions. The highest proportion of values was at 0 (5.6%) for EM and 2 for EP (13.6%; Table 1). The WLSMV estimator is used to analyze ordinal response data, does not require the observed variables to be normally distributed, and assumes that the underlying latent response variable is normally distributed [38-41].

Figure 1. Hypothesized structural equation model based on the adapted Health Action Process Approach model.

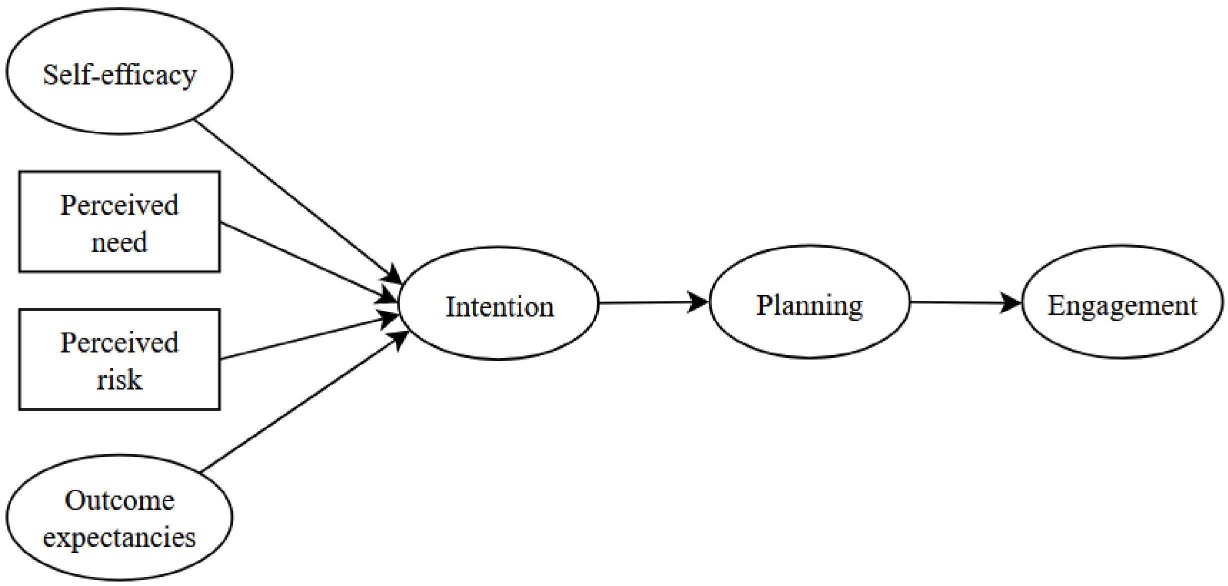


Table 1. Ranges, means, variances, and percentage of zeros for engaged minutes and engaged pages.

	Range	Mean	Variance	Percentage of zeros
Engaged minutes				
Full sample[a]	0-167.72	4.30	114.53	5.55[b]
≥3 seconds[c]	0.05-167.72	4.56	120.27	—[d]
Engaged pages				
Full sample	0-202	6.92	80.33	4.60[e]
≥3 seconds	1-202	7.32	82.34	—

[a]N=16,078.
[b]892/16,078.
[c]n=15,161.
[d]Not applicable.
[e]740/16,078.

The WLSMV estimator uses pairwise deletion for missing data; the first and second models used data from 99.5% (16,003/16,078) and 93.8% (15,086/16,078) of the full analytic sample, respectively. The analytic samples exceeded the recommended sample size of 300 for CFA [42]. While there are no absolute standards for an adequate sample size for SEM, the sample exceeded recommendations of a sample size of 10-20 per estimated parameter or a range from 200 to 1000 observations for WLSMV estimation depending on model complexity [43-47].

There is evidence that traditional cutoffs for model fit based on the maximum likelihood estimator (eg, standardized root mean squared residual [SRMR] <0.08, comparative fit index [CFI] ≥0.95, and root mean square error of approximation [RMSEA] <0.06) may not be appropriate for models with censored, ordinal, and dichotomous variables, and alternative indices of fit are being explored in the SEM literature [48-52]. However, to the best of our knowledge, there is no well-accepted method for interpreting model fit for the WLSMV estimator. For the purposes of this analysis, we reported traditional model fit indices and used traditional interpretations.

Theoretically consistent model modifications informed by Mplus MODINDICES outputs were conducted as exploratory analyses to determine whether model fit could be improved. Planning was removed from the full models such that intention predicted engagement. Next, intention and planning were removed such that the motivational HAPA constructs (ie, self-efficacy, perceived need, perceived risk, and outcome expectancies) predicted engagement. Descriptive data were analyzed in R (version 4.3.2; R Core Team, 2023) [53], and SEMs were conducted using Mplus (version 8.4; Muthén & Muthén) [54].

Results

Descriptive Statistics

The analytic samples included data from 16,078 sessions and a subsample of sessions in which participants spent at least 3 seconds on MHA outside of screening pages (n=15,161). See Table 2 for demographic characteristics. Correlations between the 22 Next Steps survey items, EM, and EP are shown in Multimedia Appendix 2.

Table 2. Demographic characteristics.

	Total sample (N=16,078)	<3 Seconds engagement (n=917)	≥3 Seconds engagement (n=15,161)
Age (years), n (%)			
11-17	2592 (16.1)	150 (16.4)	2442 (16.1)
18-24	4323 (26.9)	129 (14.1)	4194 (27.7)
25-34	2582 (16.1)	91 (9.9)	2491 (16.4)
35-44	1099 (6.8)	41 (4.5)	1058 (7)
45-54	515 (3.2)	21 (2.3)	494 (3.3)
55-64	262 (1.6)	8 (0.9)	254 (1.7)
≥65	90 (0.6)	6 (0.7)	84 (0.6)
More than 1 answer given[a]	125 (0.8)	1 (0.1)	124 (0.8)
Missing	4490 (27.9)	470 (51.3)	4020 (26.5)
Race and ethnicity, n (%)			
American Indian or Alaska Native	133 (0.8)	7 (0.8)	126 (0.8)
Asian	2088 (13)	85 (9.3)	2003 (13.2)
Black or African American (non-Hispanic)	772 (4.8)	39 (4.3)	733 (4.8)
Hispanic or Latino	962 (6)	28 (3.1)	934 (6.2)
Middle Eastern or North African	378 (2.4)	20 (2.2)	358 (2.4)
More than 1 of the above	427 (2.7)	11 (1.2)	416 (2.7)
Native Hawaiian or other Pacific Islander	41 (0.3)	3 (0.3)	38 (0.3)
Other	574 (3.6)	27 (2.9)	547 (3.6)
White (non-Hispanic)	5631 (35)	208 (22.7)	5423 (35.8)
More than 1 answer given	174 (1.1)	3 (0.3)	171 (1.1)
Missing	4898 (30.5)	486 (53)	4412 (29.1)
Gender identity[b], n (%)			
Another gender identity	502 (3.1)	24 (2.6)	478 (3.2)
Woman	8652 (53.8)	324 (35.3)	8328 (54.9)
Man	2444 (15.2)	104 (11.3)	2340 (15.4)
More than 1 answer given	76 (0.5)	2 (0.2)	74 (0.5)
Nonbinary	2 (0.01)	0 (0)	2 (0.01)
Missing	4402 (27.4)	463 (50.5)	3939 (26)
Transgender, n (%)			
No	11102 (69.1)	430 (46.9)	10672 (70.4)
Yes	176 (1.1)	11 (1.2)	165 (1.1)
More than 1 answer given	461 (2.9)	13 (1.4)	448 (3)
Missing	4339 (27)	463 (50.5)	3876 (25.6)
Depression (PHQ-9)[c], n (%)			
Minimal	139 (0.9)	1 (0.1)	138 (0.9)
Mild	513 (3.2)	4 (0.4)	509 (3.4)
Moderate	1041 (6.5)	15 (1.6)	1026 (6.8)
Moderately severe	1666 (10.4)	11 (1.2)	1655 (10.9)
Severe	2367 (14.7)	23 (2.5)	2344 (15.5)
Missing	10,352 (64.4)	863 (94.1)	9489 (62.6)

	Total sample (N=16,078)	<3 Seconds engagement (n=917)	≥3 Seconds engagement (n=15,161)
Anxiety (GAD-7)[d], n (%)			
Minimal	157 (1)	1 (0.1)	156 (1)
Mild	718 (4.5)	6 (0.7)	712 (4.7)
Moderate	1298 (8.1)	10 (1.1)	1288 (8.5)
Severe	2667 (16.6)	17 (1.9)	2650 (17.5)
Missing	11,238 (69.9)	883 (96.3)	10,355 (68.3)

[a]"More than 1 answer given" includes participants who provided different responses in multiple sessions on Mental Health America.

[b]The original response options for gender identity included "female" and "male." We report these as "woman" and "man", respectively, to distinguish from sex assigned at birth.

[c]PHQ-9: Patient Health Questionnaire-9 [30].

[d]GAD-7: General Anxiety Disorder-7 [55].

Measurement Models

Self-efficacy, outcome expectancies, intention, and planning originally had 5 categorical items corresponding to the 5 types of resources: LMH, CO, LTO, RMH, and ONL. The CO items had the lowest loadings on self-efficacy, outcome expectancies, and intention (β=.49-.55; *P*<.001). The CO item ("Do you have a specific plan to connect with others who have mental health conditions?") was the third-lowest loading item for planning (β=.61; *P*<.001). To maintain consistency among the measurement models, all CO items were removed, and CFAs were conducted using the remaining 4 items (LMH, LTO, RMH, and ONL). Standardized coefficients and model fit indices for the final confirmatory 1-factor models are shown in Table 3. All standardized factor loadings of each latent construct were significant (*P*<.001).

Table 3. Standardized coefficients and model fit indices for measurement models.

Construct, items	Standardized λ[a] (SE)	Chi-square (*df*)	*P* value	CFI[b]	RMSEA[c] (90% CI)	SRMR[d]
Self-efficacy		613.4 (2)	<.001	0.99	0.14 (0.13-0.15)	0.02
LMH[e]	0.78 (0.004)					
LTO[f]	0.93 (0.003)					
RMH[g]	0.74 (0.01)					
ONL[h]	0.65 (0.01)					
Outcome expectancies		694.6 (2)	<.001	0.98	0.15 (0.14-0.16)	0.03
LMH	0.69 (0.01)					
LTO	0.92 (0.01)					
RMH	0.77 (0.01)					
ONL	0.55 (0.01)					
Intention		759.3 (2)	<.001	0.98	0.15 (0.14-0.16)	0.03
LMH	0.68 (0.01)					
LTO	0.93 (0.004)					
RMH	0.76 (0.01)					
ONL	0.53 (0.01)					
Planning		863.4 (2)	<.001	0.88	0.16 (0.16-0.17)	0.07
LMH	0.53 (0.01)					
LTO	0.995 (0.02)					
RMH	0.47 (0.01)					
ONL	0.45 (0.01)					

[a]λ: standardized regression coefficient; all coefficients were significant (*P*<.001).
[b]CFI: comparative fit index.
[c]RMSEA: root mean square error of approximation.
[d]SRMR: standardized root mean squared residual.
[e]LMH: learning more about mental health.
[f]LTO: learning about treatment options.
[g]RMH: receiving mental health treatment.
[h]ONL: using online self-help tools.

Structural Models

Predicting Choice to Engage

The first model predicted whether participants chose to engage with the MHA website. The model comprised the HAPA constructs and a dichotomous outcome variable, choice to engage (Figure 2). Model fit was poor: χ^2_{147}=40,022.3, *P*<.001, CFI=0.81, RMSEA=0.13, 90% CI 0.129-0.131, and SRMR=0.14. Of note, the RMSEA is often inflated in models with small *df* and should be interpreted with caution [56]. The strongest predictor of intention was perceived need (β=.66, SE=0.02, *P*<.001), and intention significantly predicted planning (β=.77, SE=0.01, *P*<.001). However, choice to engage was not significantly predicted by planning (β=.03, SE=0.02, *P*=.18). The model accounted for 83% of the variance in intention, 58.7% of the variance in planning, and 0.1% of the variance in choice to engage, although these estimates should be interpreted with caution due to poor model fit. See Table 4 for latent variable correlations.

Figure 2. Structural equation model predicting choice to engage. Chi-square(147)=40,022.3, *P*<.001, comparative fit index=0.81, root mean square error of approximation=0.13, 90% CI 0.129-0.131, and standardized root mean square residual=0.14. *P<.001. INT: intention; LMH: learning more about mental health; LTO: learning about treatment options; OE: outcome expectancies; ONL: using online self-help tools; PL: planning; RMH: receiving mental health treatment; SE: self-efficacy.

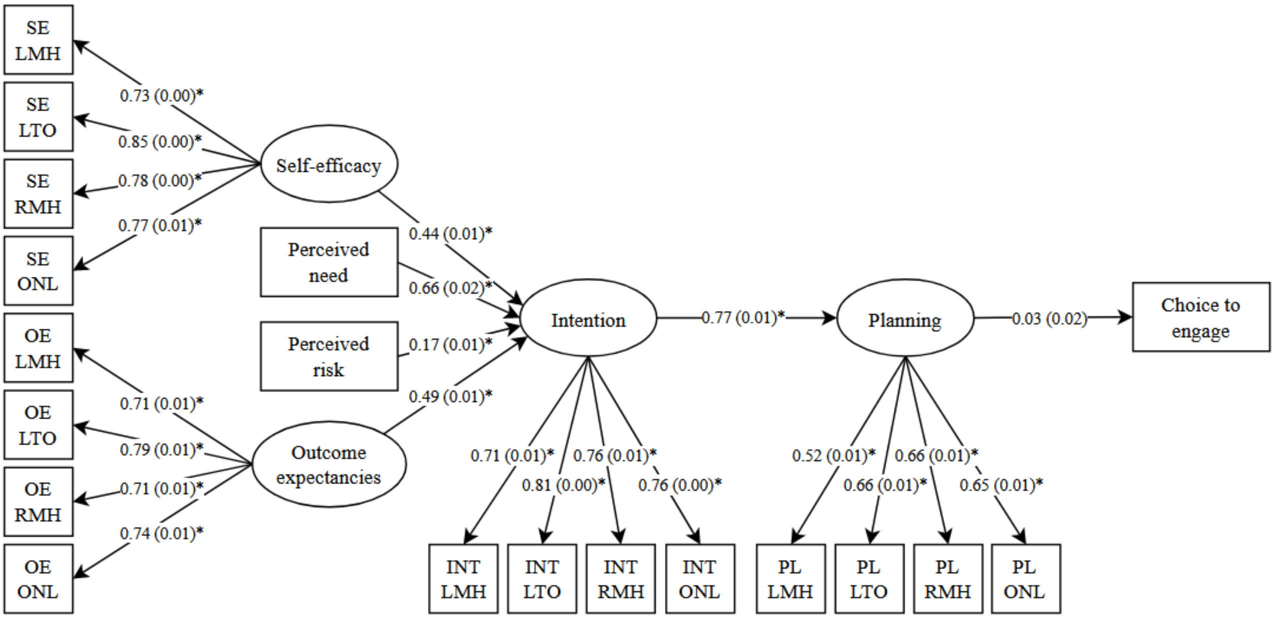


Table 4. Latent variable correlations.

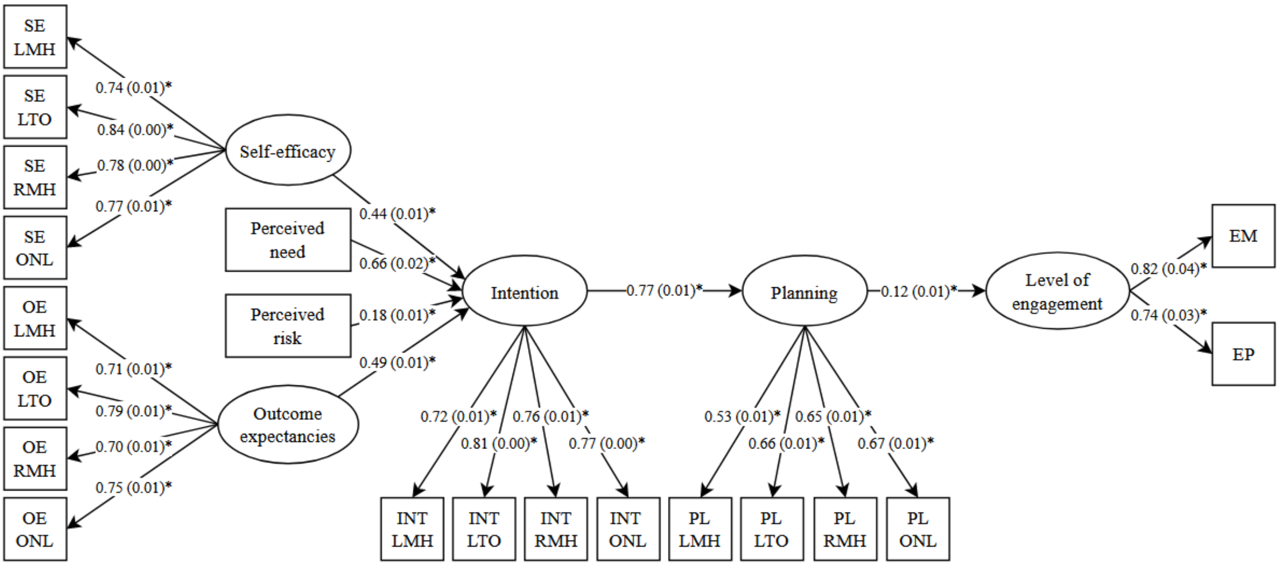
	Intention	Self-efficacy	Outcome expectancies	Planning
Model 1: Predicting choice to engage, *r* (*P* value)				
Self-efficacy	0.77 (<.001)	—[a]	—	—
Outcome expectancies	0.79 (<.001)	0.68 (<.001)	—	—
Planning	0.77 (<.001)	0.59 (<.001)	0.61 (<.001)	—
Choice to engage	0.02 (.18)	0.02 (.18)	0.02 (.18)	0.03 (.18)
Model 2: Predicting level of engagement, *r* (*P* value)				
Self-efficacy	0.77 (<.001)	—	—	—
Outcome expectancies	0.79 (<.001)	0.67 (<.001)	—	—
Planning	0.77 (<.001)	0.59 (<.001)	0.60 (<.001)	—
Level of engagement	0.09 (<.001)	0.07 (<.001)	0.07 (<.001)	0.12 (<.001)

[a]Correlations on the principal diagonal of the correlation matrix and redundant correlations are not shown.

Predicting the Level of Website Engagement

Using the same structural model, we predicted level of engagement within the subsample of participants who engaged with MHA for at least 3 seconds (Figure 3). Similar to the first model, model fit was poor: χ^2_{164}=36,925.1, *P*<.001, CFI=0.84, RMSEA=0.12, 90% CI 0.121-0.123, and SRMR=0.14. The second model accounted for 83% of the variance in intention and 58.6% of the variance in planning. While the association between planning and level of engagement was statistically significant (β=.12, SE=0.01, *P*<.001), the model was able to account for only 1.4% of the variance in level of engagement. See Table 4 for latent variable correlations.

Figure 3. Structural equation model predicting level of engagement. Chi-square(164)=36,925.1, *P*<.001, comparative fit index=0.84, root mean square error of approximation=0.12, 90% CI 0.121-0.123, and standardized root mean square residual=0.14. **P*<.001. EM: engaged minutes; EP: engaged pages; INT: intention; LMH: learning more about mental health; LTO: learning about treatment options; OE: outcome expectancies; ONL: using online self-help tools; PL: planning; RMH: receiving mental health treatment; SE: self-efficacy.



Sensitivity Analyses

The first sensitivity analysis was conducted by computing correlations between time spent on pages of different content types (navigation, condition, treatment, provider, connect, and DIY tool), the number of pages visited per content type, and the Next Step survey items. Time spent and number of pages visited by content type exhibited low correlations with the Next Steps survey items (*r*=–0.08 to 0.26). A second sensitivity analysis examined the second SEM predicting level of engagement with EM and EP winsorized at the 95th percentile. Model fit remained poor: χ^2_{154}=39,373.3, *P*<.001, CFI=0.82, RMSEA=0.12, 90% CI 0.121-0.123, and SRMR=0.14.

Exploratory Analyses

Because planning had the worst model fit among the measurement models, it was removed from each structural model. The first exploratory analysis examined the model predicting choice to engage without planning (ie, intention predicted choice to engage). This model accounted for 88% of the variation in intention, but intention was not significantly associated with choice to engage (β=.01, SE=0.02, *P*=.41), χ^2_{86}=25,070.8, *P*<.001, CFI=0.86, RMSEA=0.14, 90% CI 0.133-0.136, and SRMR=0.13. In the second model, planning was removed such that intention predicted level of engagement. Intention significantly predicted level of engagement (β=.09, SE=0.01, *P*<.001), and the model accounted for 88% of the variance in intention and 0.8% of the variance in level of engagement, χ^2_{99}=23,252.1, *P*<.001, CFI=0.89, RMSEA=0.13, 90% CI 0.123-0.126, and SRMR=0.13.

Next, self-efficacy, perceived need, perceived risk, and outcome expectancies were modeled to directly predict the engagement outcomes. In the first model, perceived need (β=.11, SE=0.04, *P*=.01) and perceived risk (β=.10, SE=0.02, *P*<.001) were significant predictors of choice to engage, χ^2_{41}=8,015.8, *P*<.001, CFI=0.92, RMSEA=0.11, 90% CI 0.108-0.112, and SRMR=0.14. The path coefficients for self-efficacy (β=.04, SE=0.03, *P*=.20) and outcome expectancies (β=–.03, SE=0.03, *P*=.26) were not statistically significant. The model accounted for 0.9% of the variance in choice to engage. In the second model, level of engagement was significantly associated with perceived need (β=.18, SE=0.03, *P*<.001), perceived risk (β=.09, SE=0.01, *P*<.001), and self-efficacy (β=.08, SE=0.02, *P*<.001) but not outcome expectancies (β=–.01, SE=0.02, *P*=.62). Only 1.7% of the variance in level of engagement was accounted for by the model, χ^2_{50}=8,060.1, *P*<.001, CFI=0.94, RMSEA=0.10, 90% CI 0.101-0.105, and SRMR=0.13.

Discussion

Principal Findings

This study found limited support for the HAPA model in predicting engagement with web-based mental health resources, although the adapted HAPA model was useful for understanding and predicting self-reported motivations, intentions, and planning. We found support for the HAPA model in the motivational phase, with a large percentage of explained variance in health behavior intention (83%) and planning (59%). Self-efficacy, perceived need, perceived risk, and outcome expectancies significantly predicted intention to engage. This is consistent with a meta-analysis, which found strong support for the motivational phase of the HAPA model [13]. The results also highlight the importance of perceived need, the strongest predictor of intention, when investigating engagement with DMH. Although referenced in prior studies of the HAPA model, perceived need is not often included as a construct in HAPA studies, including the meta-analysis. Our results indicate that perceived need may be an important construct to regularly measure in future HAPA research.

Consistent with previous work [57-59], we found evidence of the intention-behavior gap in predicting DMH website engagement. Neither intention nor planning was a strong predictor of the engagement outcomes. Choice to engage was not associated with self-efficacy, outcome expectancies, intention, or planning. Latent variable correlations between level of engagement and self-efficacy (*r*=0.07), outcome expectancies (*r*=0.07), intention (*r*=0.09), and planning (*r*=0.12) were statistically significant, but effect sizes were trivial and should be viewed with caution because of the poor fit of the overall model and other limitations, such as the use of a self-report measure and measuring engagement using MHA's available metadata (number of pages visited and time spent on MHA, excluding the final page viewed). However, these coefficients are aligned with meta-analytic results, which found coefficients less than 0.18 for all HAPA constructs in predicting behavioral outcomes [13]. A number of studies examining the HAPA framework have also found evidence supporting the preintention-intention-planning phases but limited or no support for the association between planning and behavior [60-62]. For example, previous work demonstrated that HAPA constructs predicted intention and planning but did not find evidence for planning predicting behavior (physical activity) across samples of middle-aged, physically inactive women, adults with obesity, and walking duration in older adults [60-62]. One study of a 2-week web-based intervention similarly found greater support for the motivational phase than the volitional phase, as well as stronger associations between self-reported motivational HAPA constructs and self-reported behavior but weaker associations with self-reported motivational constructs and objective measures of engagement [14].

Assuming that our estimates in predicting DMH engagement are correct, we suspect that 1 reason our study had weaker estimates than those found in the meta-analysis may have been our use of objective measures of behavior rather than self-report outcomes. The stronger associations at the motivational phase of the model in our study and in the meta-analysis may be inflated by method bias, as all measures were self-reported except for engagement, though the context of DMH may also have played a role. Only 3 studies in the meta-analysis included objective measures of behavior, none focused on DMH, and none focused on extremely short time lags for measuring behavior [13]. While our findings align with those in another study that used an objective measure of engagement [14], the relative lack of objective measures in existing HAPA studies limits our ability to contextualize our findings within the larger HAPA literature. Due to this, it is unknown whether the lack of association with engagement in our study as compared with other studies was the result of the specific context of a mental health resource website, the use of an objective measure of behavior, or the short time lag of measurement.

These findings have important implications for researching and enhancing user engagement with self-guided mental health websites. Studies on various DMH forms, such as web-based interventions or digitally delivered psychotherapy, consistently show that real-world engagement is poorer than in controlled lab experiments. Future studies could focus on understanding the factors influencing small-scale engagement actions, such as split-second decisions to click on website resources. These actions may collectively contribute to more substantial, long-term health behaviors. We suspect that website design, specifically strategies ensuring easy access and appeal of resources aligned with user needs, interests, and identity [63], may be a better predictor of health behavior in this context than the HAPA model. The sheer volume of materials on websites such as MHA can be overwhelming, making the quality, relevance, and visual appeal of resource links potentially more influential on DMH website engagement than constructs such as self-efficacy and outcome expectancies [64]. Tailoring links based on information from screening and demographic surveys is being explored in our ongoing research through a randomized trial of personalized website design.

Strengths and Limitations

Strengths of the study include a large real-world sample of participants across the United States seeking information on a broad spectrum of mental health conditions. Another strength was the application of SEM, a robust statistical technique, to a real-life DMH setting with a well-established theoretical basis for the hypothesized model. There has been little research on the mechanisms driving engagement with DMH websites, and to our knowledge, this study is the first to fill this gap in the literature by examining the HAPA model in the context of DMH website engagement.

This study has several limitations. The Next Steps survey, used as a self-report measure, is a proprietary tool that lacks psychometric evaluation and assesses simplified HAPA model constructs. Other studies of the HAPA model have distinguished, for example, between task self-efficacy, coping self-efficacy, and recovery self-efficacy, and between action planning and coping planning [16]. Nevertheless, this approach enabled the assessment of HAPA constructs in relation to the specific tools provided by the MHA website and aligns with the methodology of many other published HAPA studies, which also use study-specific measures without psychometric evaluation. In addition, this study did not examine these constructs across multiple time periods. The stability of these constructs over time in this context remains unexplored, leaving open the possibility that factors such as self-efficacy may fluctuate. However, our study assessed these constructs as closely as possible to the engagement behavior. Our measurement of engagement was limited to MHA's metadata, focusing on pages visited and time spent, excluding time spent on the final page viewed. While this approach may not capture the full spectrum of engagement behaviors, it was the only feasible approach for gathering a very large number of responses in a real-world setting. This study relied on a convenience sample, as participants self-selected into the survey. This could introduce bias, limiting the generalizability of our findings. Moreover, the exploratory models were iteratively honed and are likely overfit to the data. However, the modified models also resulted in poor model fit, suggesting that the adapted HAPA model did not fit the data.

Future Directions

Future research could address these limitations by using psychometrically sound measures, exploring temporal dynamics of HAPA constructs, broadening engagement assessments, and using more diverse participant recruitment strategies. The study design could be modified such that the HAPA constructs are measured over multiple time points to examine change or consistency in appraisal-motivational factors over time. Researchers may also explore other frameworks that might be more appropriate for understanding and enhancing engagement with DMH (eg, Social Norms Theory [65], Health Belief Model [66], and I-Change Model [67]). We encourage future researchers to continue to fill these gaps, especially by including objective measurements of behavioral outcomes and by conducting additional research across the wide array of DMH approaches beyond website usage.

Conclusions

As online mental health services become increasingly popular, there is a growing need for standardized measures of online engagement with all DMH, including mental health resource websites such as MHA. Despite the recent proliferation of DMH tools, they are underutilized by the communities they intend to serve, and few studies have examined factors impacting DMH engagement and how it can be optimized. This study aimed to identify whether and which HAPA constructs (self-efficacy, perceived need, perceived risk, outcome expectancies, intention, and planning) were significant predictors of DMH engagement. Despite some evidence supporting the motivational phase of the HAPA model in this context, the key finding was that there was insufficient support for the HAPA model in predicting engagement with DMH. The literature has yet to identify predictors of real-world DMH engagement, and doing so is an emerging and crucial research priority. It is possible that poor model fit was due in part to the study limitations. Researchers are encouraged to explore the HAPA model and other frameworks of health behaviors to identify predictors of DMH engagement. Understanding factors that optimize DMH engagement is urgently needed to inform evidence-based approaches toward tailoring DMH websites to better serve people seeking online mental health support.

Acknowledgments

All authors contributed to this manuscript. This research was supported by a grant from the National Institute of Mental Health (R01MH125179). BEB is supported by the University of Washington’s Institute of Translational Health KL2 Program through the National Center for Advancing Translational Sciences of the National Institutes of Health (KL2TR002317) and the National Institute of Drug Abuse Loan Repayment Program (L30DA056956). TA is supported in part by a National Science Foundation award (NSF CAREER IIS-2142794). MR is supported by the National Institute of Mental Health of the National Institutes of Health (award number K01MH131795).

Conflicts of Interest

TN, KR, and JM are employees of Mental Health America. All other authors have no conflicts of interest to disclose.

Multimedia Appendix 1
Next Steps survey.
[DOC File , 45 KB - humanfactors_v11i1e57082_app1.doc]

Multimedia Appendix 2
Item correlations.
[DOC File , 81 KB - humanfactors_v11i1e57082_app2.doc]

References

1. Bridging science and service: a report by the national advisory mental health council's clinical treatment and services research workgroup. NIMH. 1998. URL: https://www.nimh.nih.gov/about/advisory-boards-and-groups/namhc/reports/bridging-science-and-service-a-report-by-the-national-advisory-mental-health-councils-clinical-treatment-and-services-research-workgroup [accessed 2024-02-01]
2. NIMH. Mental illness. 2023. URL: https://www.nimh.nih.gov/health/statistics/mental-illness [accessed 2024-02-01]
3. Yom-Tov E, Lekkas D, Heinz MV, Nguyen T, Barr PJ, Jacobson NC. Digitally filling the access gap in mental health care: an investigation of the association between rurality and online engagement with validated self-report screens across the United States. J Psychiatr Res 2023;157:112-118 [FREE Full text] [doi: 10.1016/j.jpsychires.2022.11.024] [Medline: 36462251]
4. Hatch A, Hoffman JE, Ross R, Docherty JP. Expert consensus survey on digital health tools for patients with serious mental illness: optimizing for user characteristics and user support. JMIR Ment Health 2018;5(2):e46 [FREE Full text] [doi: 10.2196/mental.9777] [Medline: 29895514]
5. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. J Med Internet Res 2019;21(7):e12869 [FREE Full text] [doi: 10.2196/12869] [Medline: 31333198]
6. Ng MM, Firth J, Minen M, Torous J. User engagement in mental health apps: a review of measurement, reporting, and validity. Psychiatr Serv 2019;70(7):538-544 [FREE Full text] [doi: 10.1176/appi.ps.201800519] [Medline: 30914003]

7. Yeager CM, Benight CC. If we build it, will they come? Issues of engagement with digital health interventions for trauma recovery. Mhealth 2018;4:37 [FREE Full text] [doi: 10.21037/mhealth.2018.08.04] [Medline: 30363749]
8. Flett JAM, Hayne H, Riordan BC, Thompson LM, Conner TS. Mobile mindfulness meditation: a randomised controlled trial of the effect of two popular apps on mental health. Mindfulness 2018;10(5):863-876. [doi: 10.1007/s12671-018-1050-9]
9. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
10. Take a mental health test. Mental Health America. URL: https://screening.mhanational.org/screening-tools [accessed 2024-02-01]
11. Schwarzer R. Modeling health behavior change: how to predict and modify the adoption and maintenance of health behaviors. Appl Psychol 2008;57(1):1-29. [doi: 10.1111/j.1464-0597.2007.00325.x]
12. Schwarzer R, Lippke S, Luszczynska A. Mechanisms of health behavior change in persons with chronic illness or disability: The Health Action Process Approach (HAPA). Rehabil Psychol 2011;56(3):161-170. [doi: 10.1037/a0024509] [Medline: 21767036]
13. Zhang CQ, Zhang R, Schwarzer R, Hagger MS. A meta-analysis of the health action process approach. Health Psychol 2019;38(7):623-637. [doi: 10.1037/hea0000728] [Medline: 30973747]
14. Yeager CM, Shoji K, Luszczynska A, Benight CC. Engagement with a trauma recovery internet intervention explained with the health action process approach (HAPA): longitudinal study. JMIR Ment Health 2018;5(2):e29 [FREE Full text] [doi: 10.2196/mental.9449] [Medline: 29636323]
15. Steca P, Pancani L, Greco A, D'Addario M, Magrin ME, Miglioretti M, et al. Changes in dietary behavior among coronary and hypertensive patients: a longitudinal investigation using the health action process approach. Appl Psychol Health Well Being 2015;7(3):316-339. [doi: 10.1111/aphw.12050] [Medline: 26218746]
16. Ranjbaran S, Shojaeizadeh D, Dehdari T, Yaseri M, Shakibazadeh E. Determinants of medication adherence among Iranian patients with type 2 diabetes: an application of health action process approach. Heliyon 2020;6(7):e04442 [FREE Full text] [doi: 10.1016/j.heliyon.2020.e04442] [Medline: 32695914]
17. Lippke S, Schwarzer R, Ziegelmann JP, Scholz U, Schüz B. Testing stage-specific effects of a stage-matched intervention: a randomized controlled trial targeting physical exercise and its predictors. Health Educ Behav 2010;37(4):533-546. [doi: 10.1177/1090198109359386] [Medline: 20547760]
18. Gaube S, Fischer P, Lermer E. Hand(y) hygiene insights: applying three theoretical models to investigate hospital patients' and visitors' hand hygiene behavior. PLoS One 2021;16(1):e0245543 [FREE Full text] [doi: 10.1371/journal.pone.0245543] [Medline: 33444410]
19. Zech JM, Johnson M, Pullmann MD, Hull TD, Althoff T, Munson SA, et al. An integrative engagement model of digital psychotherapy: exploratory focus group findings. JMIR Form Res 2023;7:e41428 [FREE Full text] [doi: 10.2196/41428] [Medline: 37099363]
20. Chiu CY, Lynch RT, Chan F, Berven NL. The health action process approach as a motivational model for physical activity self-management for people with multiple sclerosis: a path analysis. Rehabil Psychol 2011;56(3):171-181. [doi: 10.1037/a0024583] [Medline: 21767037]
21. Chiu CY, Fitzgerald SD, Strand DM, Muller V, Brooks J, Chan F. Motivational and volitional variables associated with stages of change for exercise in multiple sclerosis. Rehab Couns Bull 2012;56(1):23-33. [doi: 10.1177/0034355212439898]
22. Duan YP, Wienert J, Hu C, Si GY, Lippke S. Web-based intervention for physical activity and fruit and vegetable intake among Chinese university students: a randomized controlled trial. J Med Internet Res 2017;19(4):e106 [FREE Full text] [doi: 10.2196/jmir.7152] [Medline: 28396306]
23. Joveini H, Rohban A, Ardebili HE, Dehdari T, Maheri M, Hashemian M. The effects of an education program on hookah smoking cessation in university students: an application of the health action process approach (HAPA). J Subst Use 2019;25(1):62-69. [doi: 10.1080/14659891.2019.1664655]
24. Teng Y, Mak WW. The role of planning and self-efficacy in condom use among men who have sex with men: an application of the Health Action Process Approach model. Health Psychol 2011;30(1):119-128. [doi: 10.1037/a0022023] [Medline: 21299300]
25. Zarski AC, Berking M, Reis D, Lehr D, Buntrock C, Schwarzer R, et al. Turning good intentions into actions by using the health action process approach to predict adherence to internet-based depression prevention: secondary analysis of a randomized controlled trial. J Med Internet Res 2018;20(1):e9 [FREE Full text] [doi: 10.2196/jmir.8814] [Medline: 29326097]
26. Schwarzer R, Schuz B, Ziegelmann JP, Lippke S, Luszczynska A, Scholz U. Adoption and maintenance of four health behaviors: theory-guided longitudinal studies on dental flossing, seat belt use, dietary behavior, and physical activity. Ann Behav Med 2007;33(2):156-166. [doi: 10.1007/BF02879897] [Medline: 17447868]
27. Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol 2003;88(5):879-903. [doi: 10.1037/0021-9010.88.5.879] [Medline: 14516251]

28. Jellinek MS, Murphy JM, Robinson J, Feins A, Lamb S, Fenton T. Pediatric symptom checklist: screening school-age children for psychosocial dysfunction. J Pediatr 1988;112(2):201-209. [doi: 10.1016/s0022-3476(88)80056-8] [Medline: 3339501]
29. Gardner W, Murphy M, Childs G, Kelleher K, Pagano M, Jellinek M, et al. The PSC-17: a brief pediatric symptom checklist with psychosocial problem subscales. A report from PROS and ASPN. Ambul Child Health 1999;5(3):225-236.
30. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
31. Larson M, Cook CR, Brewer SK, Pullmann MD, Hamlin C, Merle JL, et al. Examining the effects of a brief, group-based motivational implementation strategy on mechanisms of teacher behavior change. Prev Sci 2021;22(6):722-736 [FREE Full text] [doi: 10.1007/s11121-020-01191-7] [Medline: 33226575]
32. Lyon AR, Pullmann MD, Dorsey S, Levin C, Gaias LM, Brewer SK, et al. Protocol for a hybrid type 2 cluster randomized trial of trauma-focused cognitive behavioral therapy and a pragmatic individual-level implementation strategy. Implement Sci 2021;16(1):3 [FREE Full text] [doi: 10.1186/s13012-020-01064-1] [Medline: 33413511]
33. Merle JL, Cook CR, Pullmann MD, Larson MF, Hamlin CM, Hugh ML, et al. Longitudinal effects of a motivationally focused strategy to increase the yield of training and consultation on teachers' adoption and fidelity of a universal program. School Ment Health 2023;15(1):105-122 [FREE Full text] [doi: 10.1007/s12310-022-09536-z] [Medline: 35936515]
34. Tang C, Raat H, Yan M, Zhang Q, Li K, Jiang M, et al. Application of the health action process approach model for reducing excessive internet use behaviors among rural adolescents in China: a school-based intervention pilot study. BMC Public Health 2021;21(1):986 [FREE Full text] [doi: 10.1186/s12889-021-10999-z] [Medline: 34039318]
35. Pengnate S, Sarathy R, Lee J. The engagement of website initial aesthetic impressions: an experimental investigation. Int J Hum Comput Interact 2018;35(16):1517-1531. [doi: 10.1080/10447318.2018.1554319]
36. Byrne BM. Structural Equation Modeling With AMOS: Basic Concepts, Applications, and Programming, 2nd ed. New York, NY: Routledge/Taylor & Francis Group; 2010.
37. Hair JF, Black W, Babin BJ, Anderson R. Multivariate Data Analysis: Pearson Educational International, 7th ed. Upper Saddle River, NJ: Prentice Hall; 2010.
38. Asparouhov T, Muthén BO. Computationally efficient estimation of multilevel high-dimensional latent variable models. Alexandria: American Statistical Association; 2007 Presented at: Proceedings of the 2007 JSM meeting. Salt Lake City, UT, Section on Statistics in Epidemiology; 2007; Alexandria URL: http://statmodel.com/download/JSM2007000746.pdf
39. Wang J, Wang X. Structural Equation Modeling Applications Using Mplus. Hoboken, NJ: John Wiley & Sons Ltd; 2012.
40. Muthén B, Muthén L, Asparouhov T. Estimator choices with categorical outcomes. 2015. URL: https://www.statmodel.com/download/EstimatorChoices.pdf [accessed 2024-02-01]
41. Brown T. Confirmatory Factor Analysis for Applied Research (Methodology in the Social Sciences Series) Second Edition. New York, NY: The Guilford Press; 2015.
42. Tabachnick BG, Fidell LS. Using Multivariate Statistics. 4th ed. Boston, MA: Allyn and Bacon; 2001.
43. Flora DB, Curran PJ. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. Psychol Methods 2004;9(4):466-491 [FREE Full text] [doi: 10.1037/1082-989X.9.4.466] [Medline: 15598100]
44. Nunnally JC. Psychometric Theory. 2nd ed. New York, NY: McGraw-Hill; 1978.
45. Nussbeck FW, Eid M, Lischetzke T. Analysing multitrait-multimethod data with structural equation models for ordinal variables applying the WLSMV estimator: what sample size is needed for valid results? Br J Math Stat Psychol 2006;59(Pt 1):195-213. [doi: 10.1348/000711005X67490] [Medline: 16709286]
46. Rhemtulla M, Brosseau-Liard P, Savalei V. When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. Psychol Methods 2012;17(3):354-373. [doi: 10.1037/a0029315] [Medline: 22799625]
47. Schreiber JB. Core reporting practices in structural equation modeling. Res Social Adm Pharm 2008;4(2):83-97. [doi: 10.1016/j.sapharm.2007.04.003] [Medline: 18555963]
48. Hipp JR, Bollen KA. 9. Model fit in structural equation models with censored, ordinal, and dichotomous variables: testing vanishing tetrads. Sociol Methodol 2016;33(1):267-305. [doi: 10.1111/j.0081-1750.2003.00133.x]
49. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equation Model Multidisciplinary J 1999;6(1):1-55. [doi: 10.1080/10705519909540118]
50. Nye CD, Drasgow F. Assessing goodness of fit: simple rules of thumb simply do not work. Organ Res Methods 2010;14(3):548-570. [doi: 10.1177/1094428110368562]
51. Xia Y, Yang Y. RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: the story they tell depends on the estimation methods. Behav Res Methods 2019;51(1):409-428. [doi: 10.3758/s13428-018-1055-2] [Medline: 29869222]
52. Yu CY. Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes [doctoral dissertation]. Los Angeles, CA: University of California; 2002:11-17.
53. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2023. URL: https://www.R-project.org/ [accessed 2024-01-29]

54. Muthén LK, Muthén BO. Mplus User's Guide. 1998. URL: https://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf [accessed 2024-02-01]
55. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
56. Kenny DA, Kaniskan B, McCoach DB. The performance of RMSEA in models with small degrees of freedom. Sociol Methods Res 2014;44(3):486-507. [doi: 10.1177/0049124114543236]
57. Sheeran P. Intention—Behavior Relations: A Conceptual and Empirical Review. Eur Rev Soc Psychol 2002;12(1):1-36. [doi: 10.1080/14792772143000003]
58. Sheeran P, Webb TL. The intention–behavior gap. Soc Pers Psychol Compass 2016;10(9):503-518. [doi: 10.1111/spc3.12265]
59. Sniehotta FF, Scholz U, Schwarzer R. Bridging the intention–behaviour gap: planning, self-efficacy, and action control in the adoption and maintenance of physical exercise. Psychol Health 2005;20(2):143-160. [doi: 10.1080/08870440512331317670]
60. Parschau L, Barz M, Richert J, Knoll N, Lippke S, Schwarzer R. Physical activity among adults with obesity: testing the health action process approach. Rehabil Psychol 2014;59(1):42-49. [doi: 10.1037/a0035290] [Medline: 24446673]
61. Barg CJ, Latimer AE, Pomery EA, Rivers SE, Rench TA, Prapavessis H, et al. Examining predictors of physical activity among inactive middle-aged women: an application of the health action process approach. Psychol Health 2012;27(7):829-845 [FREE Full text] [doi: 10.1080/08870446.2011.609595] [Medline: 21867395]
62. Labudek S, Fleig L, Jansen C, Kramer-Gmeiner F, Nerz C, Becker C, et al. Applying social cognition models to explain walking duration in older adults: the role of intrinsic motivation. J Aging Phys Act 2021;29(5):744-752. [doi: 10.1123/japa.2020-0296] [Medline: 33567407]
63. Sharma A, Rushton K, Lin IW, Nguyen T, Althoff T. Facilitating self-guided mental health interventions through human-language model interaction: a case study of cognitive restructuring. New York, NY, USA: Association for Computing Machinery; 2024 Presented at: Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24); May 11-16, 2024; Honolulu, HI p. 1-29. [doi: 10.1145/3613904.3642761]
64. Garett R, Chiu J, Zhang L, Young SD. A literature review: website design and user engagement. Online J Commun Media Technol 2016;6(3):1-14 [FREE Full text] [Medline: 27499833]
65. Berkowitz AD. The social norms approach: theory, research, and annotated bibliography. 2004. URL: http://www.alanberkowitz.com/articles/social_norms.pdf [accessed 2024-05-15]
66. Janz NK, Becker MH. The health belief model: a decade later. Health Educ Q 1984;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]
67. Kasten S, van Osch L, Candel M, de Vries H. The influence of pre-motivational factors on behavior via motivational factors: a test of the I-Change model. BMC Psychol 2019;7(1):7 [FREE Full text] [doi: 10.1186/s40359-019-0283-2] [Medline: 30786921]

Abbreviations

CFA: confirmatory factor analysis
CFI: comparative fit index
CO: connecting with others
DMH: Digital Mental Health
EM: engagement minutes
EP: engaged pages
HAPA: Health Action Process Approach
LMH: learning more about mental health
LTO: learning about treatment options
MHA: Mental Health America
ONL: using online self-help tools
RMH: receiving mental health treatment
RMSEA: root mean square error of approximation
SEM: structural equation modeling
SRMR: standardized root mean squared residual
WLSMV: weighted least squares means and variance—adjusted

Edited by D Gooch; submitted 07.02.24; peer-reviewed by R Gore, S Ranjbaran; comments to author 18.04.24; revised version received 04.06.24; accepted 17.06.24; published 07.08.24.

Please cite as:
Rouvere J, Blanchard BE, Johnson M, Griffith Fillipo I, Mosser B, Romanelli M, Nguyen T, Rushton K, Marion J, Althoff T, Areán PA, Pullmann MD
Application of an Adapted Health Action Process Approach Model to Predict Engagement With a Digital Mental Health Website: Cross-Sectional Study
JMIR Hum Factors 2024;11:e57082
URL: https://humanfactors.jmir.org/2024/1/e57082
doi:10.2196/57082
PMID:

Original Paper

Lessons Learned From Developing Dashboards to Support Decision-Making for Community Opioid Response by Community Stakeholders: Mixed Methods and Multisite Study

Naleef Fareed[1], MBA, PhD; Ramona G Olvera[2], PhD; Yiting Wang[3], MFA; Michael Hayes[4], BA; Elizabeth Liz Larimore[5], MSc; Peter Balvanz[6], MPH; Ronald Langley[5], PhD; Corinna A Noel[7], PhD; Peter Rock[5], MPH; Daniel Redmond[8], BA; Jessica Neufeld[9], MPH; Sarah Kosakowski[6], MPH; Daniel Harris[10], PhD; Marc LaRochelle[6], MPH, MD; Timothy R Huerta[2,3], PhD; LaShawn Glasgow[4], DrPH; Emmanuel Oga[3], MPH, MD; Jennifer Villani[11], MPH, PhD; Elwin Wu[9], PhD

[1]Department of Biomedical Informatics, College of Medicine, The Ohio State University, Columbus, OH, United States
[2]Center for the Advancement of Team Science, Analytics, and Systems Thinking, College of Medicine, The Ohio State University, Columbus, OH, United States
[3]Department of Research Information Technology, College of Medicine, The Ohio State University, Columbus, OH, United States
[4]Research Triangle Institute, Research Triangle Park, NC, United States
[5]Center for Drug and Alcohol Research, University of Kentucky, Lexington, KY, United States
[6]Clinical Addiction Research and Evaluation Unit, Section of General Internal Medicine, Boston Medical Center, Boston, MA, United States
[7]Department of Public and Ecosystem Health, Cornell University, Ithaca, NY, United States
[8]Institute for Biomedical Informatics, University of Kentucky, Kentucky, KY, United States
[9]Social Intervention Group, School of Social Work, Columbia University, New York, NY, United States
[10]Institute for Pharmaceutical Outcomes and Policy, University of Kentucky, Lexington, KY, United States
[11]National Institute on Drug Abuse, Bethesda, MD, United States

Corresponding Author:
Naleef Fareed, MBA, PhD
Department of Biomedical Informatics
College of Medicine
The Ohio State University
370 W. 9th Avenue
Columbus, OH, 43210
United States
Phone: 1 614 685 3251
Email: naleef.fareed@osumc.edu

Abstract

Background: Data dashboards are published tools that present visualizations; they are increasingly used to display data about behavioral health, social determinants of health, and chronic and infectious disease risks to inform or support public health endeavors. Dashboards can be an evidence-based approach used by communities to influence decision-making in health care for specific populations. Despite widespread use, evidence on how to best design and use dashboards in the public health realm is limited. There is also a notable dearth of studies that examine and document the complexity and heterogeneity of dashboards in community settings.

Objective: Community stakeholders engaged in the community response to the opioid overdose crisis could benefit from the use of data dashboards for decision-making. As part of the Communities That HEAL (CTH) intervention, community data dashboards were created for stakeholders to support decision-making. We assessed stakeholders' perceptions of the usability and use of the CTH dashboards for decision-making.

Methods: We conducted a mixed methods assessment between June and July 2021 on the use of CTH dashboards. We administered the System Usability Scale (SUS) and conducted semistructured group interviews with users in 33 communities across 4 states of the United States. The SUS comprises 10 five-point Likert-scale questions measuring usability, each scored from 0 to 4. The interview guides were informed by the technology adoption model (TAM) and focused on perceived usefulness, perceived ease of use, intention to use, and contextual factors.

Results: Overall, 62 users of the CTH dashboards completed the SUS and interviews. SUS scores (grand mean 73, SD 4.6) indicated that CTH dashboards were within the acceptable range for usability. From the qualitative interview data, we inductively created subthemes within the 4 dimensions of the TAM to contextualize stakeholders' perceptions of the dashboard's usefulness and ease of use, their intention to use, and contextual factors. These data also highlighted gaps in knowledge, design, and use, which could help focus efforts to improve the use and comprehension of dashboards by stakeholders.

Conclusions: We present a set of prioritized gaps identified by our national group and list a set of lessons learned for improved data dashboard design and use for community stakeholders. Findings from our novel application of both the SUS and TAM provide insights and highlight important gaps and lessons learned to inform the design of data dashboards for use by decision-making community stakeholders.

Trial Registration: ClinicalTrials.gov NCT04111939; https://clinicaltrials.gov/study/NCT04111939

(JMIR Hum Factors 2024;11:e51525) doi:10.2196/51525

KEYWORDS

data visualizations; dashboards; public health; overdose epidemic; human-centered design

Introduction

Background

Data dashboards are tools, often published digitally on websites or dedicated apps, that present visualizations; they are increasingly used to display data about behavioral health, social determinants of health, chronic and infectious disease risks, and environmental risks to inform or support public health endeavors [1-4]. Dashboards can be an evidence-based approach used by communities to influence public awareness and decision-making and to focus the provision of resources and interventions in health care toward specific populations [2,3,5-8]. For example, local public health agencies across the nation have communicated health data about the COVID-19 pandemic through dashboards, using these as tools to generate awareness and motivate behavior change (eg, adherence to public health guidelines) [9].

Despite widespread use, evidence on how to best design and use dashboards in the public health realm is limited [1,9-12]. There is also a notable dearth of studies that examine and document the complexity and heterogeneity of dashboards in community settings. Experiences from the Healing Communities Study (HCS) provided an opportunity to empirically learn how dashboards and health data visualizations can be assessed to inform design and use, especially among community end users.

The HCS is implementing the Communities That HEAL (CTH) intervention aimed at reducing overdose deaths by working with community coalitions in selected counties and cities highly affected by opioid deaths in each state [13]. In January 2020, as part of the community engagement component of the CTH, 4 research sites followed a common protocol to develop community dashboards to support community coalitions in selecting evidence-based practices (EBPs) to reduce opioid overdose deaths in their respective communities. The protocol stipulated which key metrics to present to each community regarding opioid overdose deaths and associated factors that may contribute to their prevalence [14]. The CTH intervention protocol involved dashboard cocreation by HCS researchers and community stakeholders from each community that incorporated the principles of user-centered design [15].

We define community stakeholders for this analysis as individuals in the community who are engaged in fighting the overdose crisis, including coalition members (eg, county public health officials and behavioral health practitioners) and community research staff. The CTH intervention envisioned community-tailored dashboards as a tool that coalition members would use to discuss and understand baseline conditions and trends, to inform EBP selections, and monitor research outcomes of interest to the community. Community research staff, often with preexisting ties to the community, were to support and lead coalitions through the use of data dashboards for decision-making and ongoing monitoring. These stakeholders were the anticipated dashboard end users who were not expected to have expertise in the use of dashboards, even though community research staff had additional training in community engagement and overall study protocols. Our research team oversaw the design, implementation, maintenance, and evolution of the dashboards used by these stakeholders.

The CTH dashboard cocreation involved, at a minimum, iterative show-and-tell sessions in which feedback from community stakeholders was provided on wireframes with the goal of refining the dashboard and its ability to align with specific objectives (eg, to address local challenges in fighting the opioid crisis and to highlight key performance indicators). Although the CTH intervention required the same specific core components of the dashboards (eg, use of predetermined metrics, annotations for metrics, and granularity of the data presented), communities could incorporate unique components based on preferences and resources, such as displaying local data (eg, county opioid overdose death rates) acquired by the HCS. Thus, dashboards varied in layout, interface, and content across the 4 sites.

Objectives

To elucidate lessons learned from providing dashboards to the coalitions and community stakeholders, we investigated the following questions on the CTH dashboards: Are the CTH dashboards usable and useful for community decision-making?

Are the dashboards easy to use and understand? Will the dashboards be used in the long term, and, if yes, for what purposes will they be used? Our study is novel because of four specific areas: (1) the use of a qualitative approach (instead of a quantitative approach) to expound on the technology adoption model (TAM) constructs, informing existing perspectives on dashboard usability; (2) the investigation of dashboard usability across 67 diverse communities; (3) the generation of themes and subthemes on the usability of a dashboard in the substance abuse domain; and (4) the identification of common themes on dashboard usability from a TAM and System Usability Scale (SUS) perspective [16,17] by comparing feedback from 4 unique applications of a dashboard that was tailored to the needs of end users. Our findings will be used to establish current perceptions of the dashboards and prioritize gaps in knowledge, design, and use that can be considered for future dashboard applications.

Methods

Ethical Considerations

Advarra Inc, the HCS's single institutional review board, approved the study protocol (Pro00038088). The HCS study is a registered trial (ClinicalTrials.gov NCT04111939).

Research Setting

Data collection and analyses were conducted as part of the HCS, a 4-site, waitlisted community-level cluster-randomized trial seeking to significantly reduce opioid overdose deaths by implementing the CTH intervention in 67 communities across Kentucky, Massachusetts, New York, and Ohio [13,18]. For the first wave of the HCS, each site created interactive dashboards with community stakeholder input to support local data-driven decision-making using community-level metrics for 33 communities randomized to receive the intervention (ie, wave 1 communities) from January 2020 through June 2022 [14]. Our analysis involved all wave 1 communities: 8 communities from Kentucky, 8 communities from Massachusetts, 8 communities from New York, and 9 communities from Ohio.

The CTH Dashboards

Common Components Across the Dashboards

More details about the cocreation of the dashboards and required core components can be found elsewhere [14]. Briefly, each site was responsible for developing a secure portal for each community receiving the CTH intervention. The portals contained downloadable intervention materials (eg, information on EBPs, community profiles, and community landscape data). Each portal had a CTH dashboard with data visualizations (eg, bar and line graphs and tables) displaying metrics related to the HCS's primary outcome of reducing opioid deaths (eg, opioid overdose rates) and secondary outcomes connected to EBPs (eg, number of naloxone kits distributed and number of buprenorphine prescriptions filled).

Unique Components of the CTH Dashboards

Each site developed its CTH dashboard using different software, including Power BI (Microsoft Corp), Tableau (Salesforce Inc), SharePoint (Microsoft Corp), D3.js (Mike Bostock and Observable, Inc; Data-Driven Documents), Drupal (Drupal Community), and different types of visualizations, such as bar graphs, line plots, and tables with directional markers [14]. The 4 distinct dashboards displayed community-specific data that were accessible only to that community's stakeholders. Data acquisition and display were limited by site-specific data use agreements, which informed the timing of the data (lags), data suppression, and granularity (eg, aggregation by state vs county), as well as the requirement of user and password-protected access in some cases.

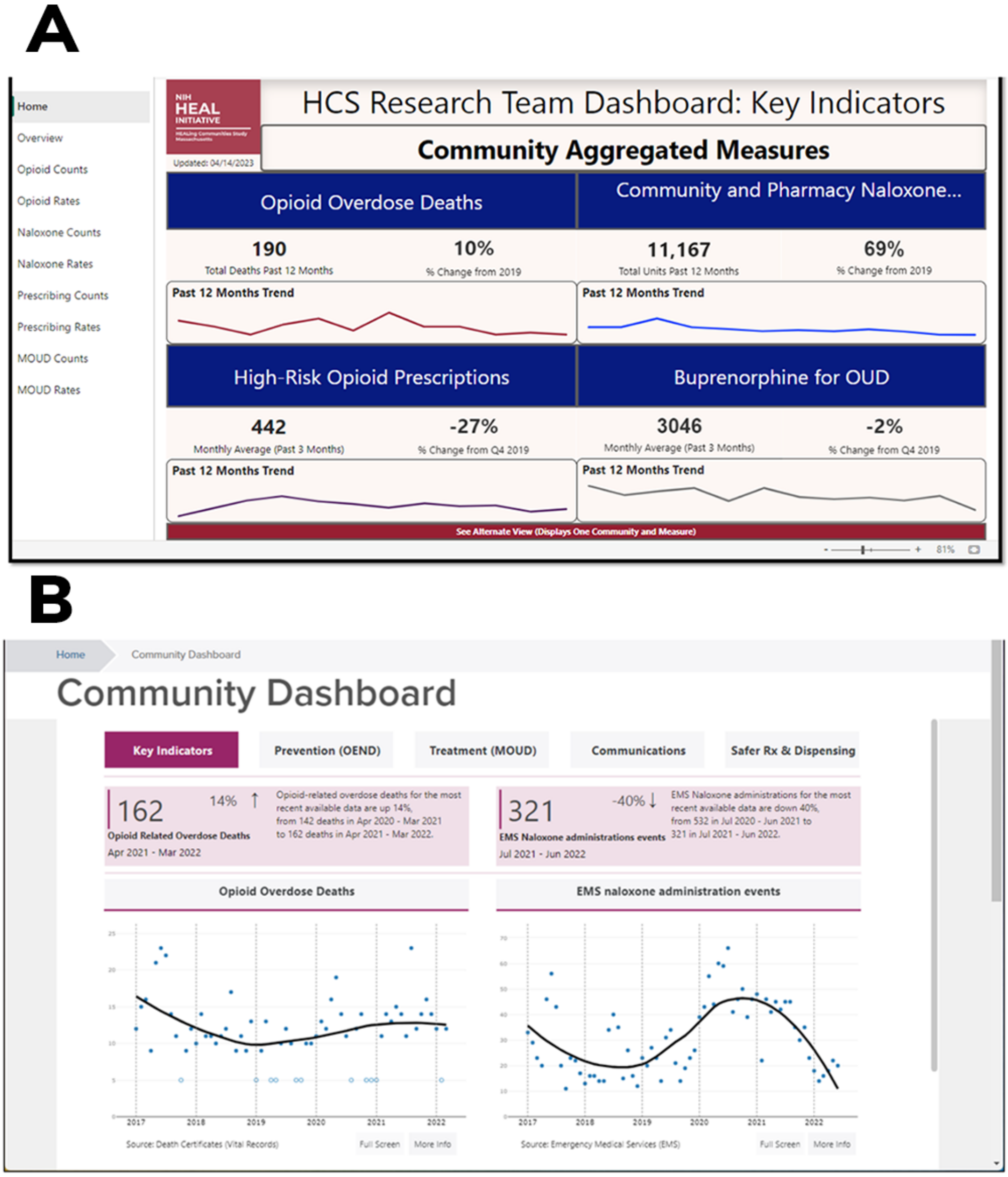
Study Design, Sampling, and Recruitment

Between June and July 2021, we conducted a mixed methods assessment of the use of the CTH dashboards. This period reflects the postimplementation phase of the second version of the cocreated dashboards (Figures 1 and 2) and was a stable time during which no fundamental revisions were made to the dashboards across the HCS. We administered the SUS to and conducted TAM-informed semistructured interviews with community stakeholders involved in the HCS. We collected sociodemographic data from participants to help characterize respondents. The interview data and SUS scores were concurrently examined to identify factors that influenced dashboard perceived use and usability.

Each site sampled and recruited participants. The sample was drawn from community stakeholders in the HCS communities who had a community portal account (ie, active users). Researchers from each site generated a roster of eligible participants from site server audit log files. Our research team worked with field staff to achieve a diverse participant pool of community stakeholder active users within each community. The sample included coalition members, including stakeholders in the HCS communities responsible for championing the use of a specific CTH intervention component, and HCS community staff who had the role of coordinating the use of data for decision-making. These active users were invited via email to participate in a group interview via Zoom (Zoom Communications, Inc) with up to 6 participants, although a few interviews were conducted individually due to scheduling challenges.

Figure 1. Mockups of the Communities That HEAL (CTH) dashboards: (A) CTH dashboard for Kentucky and (B) CTH dashboard for Ohio. Please note that any names of communities present are not real, and only synthetic data are used in the images. DAWN: deaths avoided with naloxone; EMS: emergency medical service; HCS: Healing Communities Study; MOUD: medication for opioid use disorder; OD: overdose; OUD: opioid use disorder; TBD: to be determined.

Figure 2. Mockups of the Communities That HEAL (CTH) dashboards: (A) CTH dashboard for Massachusetts and (B) CTH dashboard for New York. Please note that any names of communities present are not real, and only synthetic data are used in the images. EMS: emergency medical service; HCS: Healing Communities Study; HEAL: Helping to End Addiction Long-term; MOUD: medication treatment for opioid use disorder; NIH: National Institutes of Health; OUD: opioid use disorder.



Data Collection

Mixed methods evaluation data were collected by each site, which were then shared with the HCS data coordinating center (DCC) for analysis. Interviewers at each site first received common training on the research protocol, interview guide, and interviewing techniques. During the review, participants verbally provided consent and then were asked to individually fill out a short survey on REDCap [19]. The survey included sociodemographic questions and the SUS. After participants completed the survey, the research team displayed the site-specific CTH dashboard with synthetic community data to reorient participants to the dashboard used in their community. Next, the research team asked participants semistructured, open-ended questions based on the TAM dimensions of perceived usefulness, perceived ease of use, intention to use, contextual factors (Multimedia Appendix 1). Table 1 provides operational definitions for each of the TAM constructs used to develop our interview guide questions. We focused on 4 constructs to explore how the dashboards were used for community decision-making, their ease of use, the intent for future use, and the context around dashboard use. Our interviews lasted approximately 1 hour and were recorded and transcribed.

Table 1. Technology adoption model themes and operationalized definitions[a].

Construct	Operationalized definition
Perceived usefulness	
Description	Alignment of the dashboard and its functions with the community stakeholder's expectations for goals and tasks
Benefits	How the dashboard positively influences the work and decision-making expectations of the community stakeholder
Drawbacks	How the dashboard did not meet the work and decision-making expectations of the community stakeholder
Perceived ease of use	
Description	Alignment of the technical functionality of the dashboard with the community stakeholder's workflow (needs and desires)
Barriers	Specific challenges faced with using the portal
Facilitators	Specific resources needed to support the use of the portal
Intention to use	
Description	The willingness of the community stakeholder to use the dashboard in the future, even if modified slightly
Acceptance	The approachability of the dashboard as a technological tool to accomplish work
Preference	Specific improvements, changes, or recommendations to the dashboard
Contextual factors	
Description	Circumstances (eg, social, cultural, and historical circumstances) that influence a community stakeholder's use of the dashboard
Community data orientation	Collective perceptions about how the community connects and works with data, including community data tools and approaches for decision-making

[a]Definitions derived from Davis [16].

Data Analysis

The validated SUS consisted of 10 five-point Likert scale questions. Each SUS item was scored (scale of 0 to 4), and their scores were summed and multiplied by 2.5 to assign a total score from 0 to 100 (lowest to highest usability). For the qualitative data, 1 research site constructed the codebook using an iterative constant comparative method [20]. Initially, the team created general codes using concepts from the TAM from the interview guide. Two researchers (RGO and YW) each coded an interview independently, compared coding, and discussed agreement and differences with a senior researcher (NF), creating a refined codebook with consensus codes and clarified definitions. The remaining interviews from this site were coded by these researchers, with review by the senior researcher. One researcher (RGO) working with the team inductively created subthemes for each TAM dimension, revised the codebook, and coded all the interviews from the site. The established codebook and coded interviews were then shared with the DCC. Two new coders from the DCC were trained by the experienced coders; each independently coded 6 test interviews and compared results in meetings with the experienced coders. Then, the DCC coders coded all remaining interviews (n=14) with weekly meetings to discuss differences and ensure alignment. The DCC coders then reviewed the original 7 interviews coded by the research site to confirm alignment across all data. After coding, the DCC conducted an exploratory analysis to better understand the emergent themes across different sites. Interview transcriptions were analyzed using NVivo (version 12; QSR International) [21].

The reporting of our qualitative methods and results adheres to the COREQ (Consolidated Criteria for Reporting Qualitative Research) guidelines (Multimedia Appendix 2; Table 1) [22]. Altogether, scientific rigor was supported by the use of participant IDs and labels to ensure data were appropriately associated with participants across communities; encouraging consistency in data collection and the fidelity of the interview guide through mentorship and weekly meetings to ensure agreement and the reliability of codes and results; discussions with various experts in the data visualization field during the development of themes; and triangulation (data from multiple HCS communities and diverse roles) and parallel data collection (SUS and TAM) to achieve theoretical sufficiency for themes and diverse representation across sites [23,24].

Results

Descriptive Summary of Survey Respondents and Interview Participants

A total of 159 individuals from across all HCS sites were invited to participate in our study. Of them, 62 (39%) individuals enrolled and participated in interviews. We conducted a total of 17 group interviews (with an average of 3 participants per group) and 10 individual interviews (with an average of 7 interviews per site). Community coalition members represented 56% (35/62) of interviewees, with the remaining 44% (27/62) comprising community staff hired by the HCS to work with their specific community coalition (Multimedia Appendix 2; Table 2). At least 1 person was interviewed from every community. Participants from all sites typically had a master's degree, and a majority of participants from Ohio and Massachusetts were younger. Participants also typically identified as non-Hispanic White and female across sites (Multimedia Appendix 2; Table 3).

Table 2. Subthemes under perceived usefulness.

	Illustrative quotes
Benefits	
Knowledge dissemination	• "I think it's a terrific dashboard. I love the way that the information is on there. I think it really captures every aspect of what we're seeing on the street" (Ohio). • "Our subcommittee particularly likes to see in the slides [from dashboard data] what's going on month by month. We usually give them an overall of how many [naloxone kits] we've distributed in total, but they also like to see how we're trending. So that's basically been our main use of it" (Kentucky). • "When you dive into the data dashboard, I would say that the biggest use for it has been just displaying any recent data we have pretty much on a quarterly basis, and that mainly happens at subgroup meetings" (Massachusetts).
Decision-making tool	• "[Deidentified person] can speak more to how data impacted our strategies. She was really helpful in the sense that like, she would give us the data, we would identify the issue, and then I would take that and talk to our partners to figure out how to move forward, but I wasn't the one really digesting it, pulling the stories from it" (New York). • "Yes, I do believe that it definitely helped when it came to those decisions because it helped the entire coalition know certain areas that needed extra help and what kind of help that they needed. It also helped us form...helped us do the problem solving in the original start of this coalition" (Massachusetts). • "Because it helped us choose...There's 10 strategies that you could have chose, but by looking at the data that's available on the dashboard, then you say, 'All right, these are the ones that we need to tackle here in our community, because it's right here in front of our face. These are the important ones'" (Kentucky).
Access to more data	• "So, we don't see the practical application of this portal, but the information is on there. How to access it is, in my opinion, absolutely suited for any organization that would need that data that they don't have access to normally" (Ohio). • "But the Medicaid data, we never access that as the health department. I'm not entirely sure why but it's a really helpful data source for us to be able to use because we serve a lot of, you know, medically underserved population. So, being able to have access to those numbers does help" (Ohio). • "I got that from the dashboard, because I think that our sense was that there was more naloxone available than there probably was because there's so many other sources now. I mean, it's not just coming from harm reduction...And then we were trying to identify what was behind it because we were seeing that naloxone was a huge factor in deaths being down. I mean, that was something we had to really make sure was occurring. And if there was a distribution site, as it were, that was sort of flailing a little, we wanted to do something about that. So, that information came from the dashboard. I did not have a sense of that from the community" (Massachusetts). • "[L]ike residents receiving buprenorphine. I don't know where you're getting that information, prescription drug monitoring programs. I wish we could get that like firsthand, but this is great to look at because I don't get that data. And it's current as of 2021" (New York).
Drawbacks	
Time constraints	• "[G]iven COVID and everything that happened we just didn't have [data] available, and the timeline really just didn't align with decision-making, and I think that is just a fundamental issue with the study, like we talked about, but I do think that that makes it challenging" (Massachusetts). • "Particularly since we've been working out of the office, there's greater demands of people's time and lower bandwidth available" (Kentucky). • "But this has also just been a unique for us Wave 1 folks, a unique time, given the pandemic and everything. We have so many other things too...that are not typical. That some of the HEALing [Communities] Study...has just been a little bit extra so it's just been kind of a hard time..." (Ohio).
Misalignment of the dashboard as a tool	• "So, I would say that in [deidentifed community], a lot of it is redundant. They already have access to this data. It's already in a location where they feel comfortable going and know how to go to. So, they don't want to use something new. If they don't have to" (Ohio). • "I think, you know, had we had it earlier, I think it would have been easier to incorporate it into the coalitions. I think now it's just trickier" (New York). • "We did not use the data from the dashboard largely because, I think, everyone around our virtual table already knew all the data from our community dashboards and understood. Everyone is there because they work in this space, and so they understand exactly where the trends are, where the issues are, and so forth" (Kentucky).
Disutility of data	• "I think the factor is that the data is just so lagged, and the information that we need to work off of, we just need something way more current than what's available...It's just -- the data is just not very useful for us right now, unfortunately" (New York). • "The problem is much bigger than what the numbers from [deidentified agency] can show us and so when we're looking at it that way, it's better to get that real time data from each other rather than rely on HCS to get old data that we've kind of already gone through" (Ohio). • "Well, we've looked at the data, but I think we haven't really used it, and the reason is two-fold. One, it's too soon to be able to see much of an impact on the practices. We're still rolling some of them out. And secondly, COVID, that has screwed up all the data, and we don't, at the moment, have a good way to separate the impacts of the pandemic isolation and so forth from changes in practice in our community. Everyone right now is just kind of, with what's going on, doing the best we can. We can't read the data to detect impact of anything that we're doing right now" (Kentucky).

Table 3. Subthemes under perceived ease of use.

	Illustrative quotes
Barriers	
Access to dashboard	• "This might be a little petty, but just the fact that I have to use a password to access it. I think especially people on our coalition, they may not write down their passwords that they use for [the HCS[a]] because it's not their full-time job. So. if they forget the password, they're less likely to go through the steps to retrieve it and get in there, so they may not use it as much if it were just open access" (Kentucky). • "I remember my login info and its only because my computer remembers it. I think my username was given to me and it's not what I normally use for things so I think if my computer didn't hold my username and password, I would have had to fiddle with it every time I try to open it" (New York). • "Another thing too, and this is my last thought is that it's kind of odd who can have access and who cannot have access. It feels very gate kept in a way. And that's going to be a barrier just anyway, so you know, we have our voting is set up for one vote per agency, which has led to just one representative per agency so that one representative for that agency is the only one who has access but their partner or a contact of theirs within their agency is the one emailing me for the information. They don't have access to the dashboard. So that kind of does create a barrier as well and it's definitely something that I don't like about it. I don't like there's a password. I don't like that you have to log in" (Ohio).
Data manageability	• "[Describing barriers and challenges] first of all it's not being able to download the data but also not being able to compare two different data sets" (Ohio). • "[I]nitially, they wanted us to have a couple slides and go over overdoses every month, and all of us were like, 'What? There's so many different things, and it looks insane. I don't know how to explain that.' Once I felt more comfortable with filtering things out to make it digestible, it's been more useful. But initially, it was kind of crazy to look at" (Kentucky). • "I might add that, as a CDM [community data manager] in a cluster community, these communities don't often combine their data, so on the data dashboard for...those aggregate counts, or that aggregated together from each community, so it's really hard to see what's working, although we want to break silos. They don't identify together, so I think that's been a little bit hard, not to be able to see that separate breakdown" (Massachusetts).
Unable to locate usable data	• "I would say, a major barrier in [deidentifed community], as well, are the suppression rules, you know, because we have such a small population" (Ohio). • "[I]t's a rural community or communities, and so often our data is suppressed for the variables that require suppression. So, I think that just based on it's great to remain confidentiality and is hard to understand if there's an increase from one to four, et cetera, or zero to four. So, the suppression in rural communities is that point" (Massachusetts). • "And so, we were anticipating and we were being told over and over again, that this data is coming, wait to see the dashboard, and then behind the scenes, we would get contacted and say, you guys got to push this dashboard, you guys got to keep pushing it, and so I'll go on the dashboard and there's nothing to push" (New York).
Presentation of numbers and labels	• "So, I'm assuming if any numerator is a one through 10 or a denominator's one through 10, the data gets suppressed, but I guess I'm not 100% sure, and then...And then I noticed like on the overdose deaths for May of '21. There's a zero value there. I don't know if that's like you know when there's a zero on the dashboard. Are those true zeros or is it like a placeholder? Like do we really not have any overdose deaths in May at all. Which seems kind of, like it, you know not right, I guess" (Ohio). • "Sometimes I wonder if, you know, the line across whatever the imputed data value [ie, values masked due to suppression rules] that is chosen for a particular measure, now that there's the trend line, I wonder if just leaving off the imputed value all together, so it doesn't show on the plot. Because I feel like if someone looks at it fast, they're not looking at it close enough to be like, 'Oh, why is March missing?' They wouldn't even notice that and that might clear up a bunch of confusion" (New York). • "[On their naloxone administration event data] But yeah, it's like on a five, it's like five. And I remember somebody asking me like, 'Why is it stuck at five instead of zero?' I don't know why. And I couldn't explain that. So, like, if I can't explain that, then like I don't -- like then they look at me, and, you know, whatever, we're all supposed to be experts and we all have degrees in-in, you know, in bio stats or data or whatever, we're experienced and if we can't explain that, then-then they question that" (New York). • "My thing is super minor, and it's a visual thing...On our specific dashboard, there's the cover page, and the first one in the upper hand corner, it still says opioid overdose deaths, but if you click in it's more like, I would say it'd be more accurate to describe it as opioid overdose events or related events, because it seems like, if someone were to click on it, they'd only be able to access the death data...When I present it, people were like, 'Oh, I didn't realize there was EMS data there too,' so it's totally a minor label thing" (Massachusetts).
Facilitators	

	Illustrative quotes
Additional support	• "I think, just like before maybe HEALing Communities leaves our county...It might be nice to like have a refresher to some of our stakeholders about the dashboard and what information is out there. You know, maybe like a you know presentation, or something and how to get access to it and that kind of thing. I know that was done a while back, because I feel like...I'm even a data person and I, you know still kind of struggle with the dashboard of HEALing Communities so and there's a lot on there obviously that I didn't even know about" (Ohio). • "It may be helpful just for me to have some kind of training on how to put the data together though. Because I can go in there and just mess with everything. But looking at the community profile, the dashboards, and how to build those specific charts and things, I'm not sure exactly the functionality of it. So maybe just a video or just being able to ask questions or something like that would be pretty helpful" (Kentucky). • "I was going to say for troubleshooting too, I think we've run into that challenge. Of course, coalition members will come to us saying they've run into an error or whatever the case may be, and it's just challenging to triage some of that, to be like, 'Hold on. Let me get in touch, try to figure out what's going on, and then I'll try to get back to you as soon as possible.' And I think everyone in [Organization] has been really responsive" (Massachusetts).
Data accessibility	• "So, I do like that up at the top of the dashboard it has like where you are compared to 2017, it compares the various years to 2017 so that you see...the -- if it's increasing or decreasing compared to that, like really steady fixed baseline. I also like that you can look at the data by month, by quarter, and by year, because I think that is also really helpful. Especially like with overdose deaths, where there can be some seasonality to that, it's helpful to see that" (Ohio). • "I like that you're able to add more than one indicator to kind of look at how the trends have changed over the last couple of years and being able to put in more than one thing to get a graph and look at. Because it's very helpful when it's all in one place. Because I did a lot of data for a grant for that [deidentified] grant that I wrote a couple of years ago. And it was so hard to get all that data because you had to look at multiple different programs and now it's just all on one program. So that's really helpful" (Kentucky). • "It's quick, it's like a bottom line, you know. If you, if you had to throw together a newsletter or a speech or a letter to the editor, you could, you know, you could glance over it and get some up to date, real time numbers, you know, for that area, pretty reliably" (New York).
Navigability	• "I think the dashboards, I think they're beautiful, I think you guys did a really great job...I think it looks great, it just -- the data is just not very useful for us right now, unfortunately" (New York). • "It's very easy to use and I'm kind of a very visual person. So, for me, being able to look at graphs makes it a lot easier. And so, I use a lot of graphs...having those graphs provides me with a very quick and easy picture of ups and downs is really valuable" (Massachusetts). • "I like the charts. I think that as someone who...I mean not only do I write grants and those charts are really helpful to stick in there for the reviewer, especially when you have something that you want to really emphasize. But also doing presentations and reports and things like that for other people, it really helps them to be able to see the trend and understand you know that. So, I just like the fact that it makes really nice, simple, pretty charts" (Ohio). • "One of the things that I liked about it is being able to go in there and see the trends. It's like it's talking about pharmacy dispensing, and naloxone dispensing in the community, whether that's trending up and down. I thought that was pretty helpful. It's a few months behind, but an accurate view of what's going on in the community right now" (Kentucky).

[a]HCS: Healing Communities Study.

Perceived Usability and Acceptance of the CTH Dashboards

Of the 62 participants, 58 (94%; Kentucky: n=13, 21%; Massachusetts: n=14, 23%; New York: n=15, 24%; Ohio: n=16, 26%) completed the SUS survey. The grand mean for overall SUS scores across all HCS sites was 73 (SD 4.6), indicating that the CTH dashboards were within the 70th and 80th percentile of SUS responses (Multimedia Appendix 2; Table 4). This score is recognized in the literature as being within the acceptable range [25,26]. The overall SUS score for the New York dashboard was the highest (mean 77, SD 11.4) and akin to the SUS scores of Microsoft Word and Amazon. The overall SUS score for the Ohio dashboard was the lowest (mean 67, SD 12.3), which was similar to the usability of a GPS system. The overall SUS scores for the Kentucky and Massachusetts dashboards were 75.8 (SD 10.4) and 70.2 (SD 16.3), respectively.

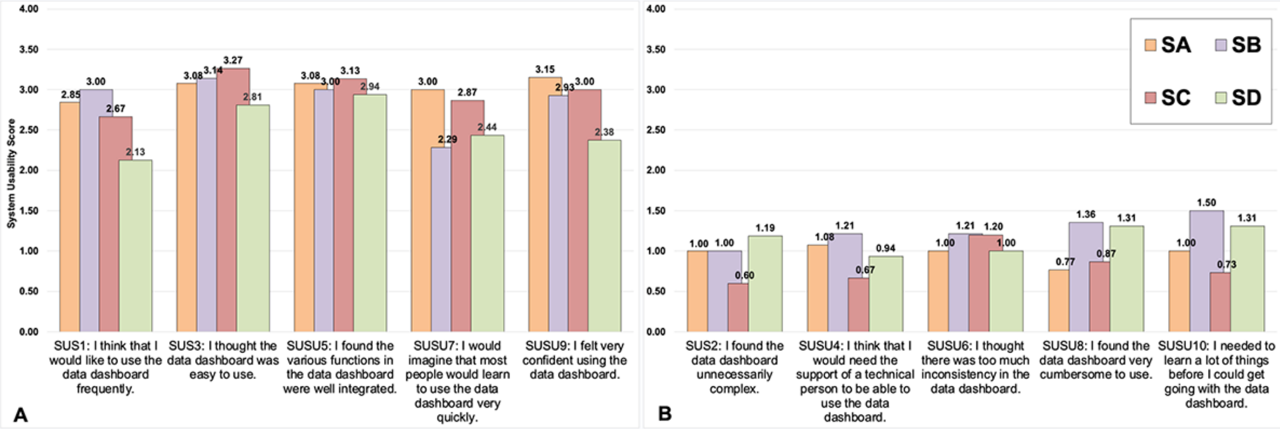
Figure 3 illustrates the breakdown of each SUS domain and contrasts the average scores by participants from each HCS site. The New York dashboard scored the highest across most of the SUS domains. Although Ohio participants did not score the dashboard as favorably as participants from the other sites across many of the domains, it should be noted that the responses were generally positive. The discussion that follows expounds on these SUS scores, with additional support from the data from our TAM-informed interviews. We link specific SUS dimensions around usability with closely associated TAM themes on usefulness. This novel approach highlights potential drivers of SUS scores based on design considerations elucidated by a specific set of TAM themes and subthemes.

Table 4. Subthemes under intention to use.

	Illustrative quotes
Acceptance	
Alternate data uses	• "We are always, always, always pushing data driven decision-making with the coalitions and more often than not, the coalitions have a lack of data that they can specify down to a county or even a municipality. And so, having access to something like this...can potentially tap into the [deidentified community] data and be able to use that data to justify a program need or a grant..." (New York). • "[W]e have opportunities...from new funding sources that we have never had before. So they're coming towards us and kind of questioning where our gaps are and where we think that we could grow. So, I do think there's things on the dashboard that we will probably be utilizing in the future" (Ohio). • "I'm doing a study for my dissertation actually on...overdose deaths and things of that nature, and trauma, retention and treatment. All sorts of things. So, I'll definitely be looking at the numbers that you guys have because they're recent numbers. I know it's valid data and things of that nature" (Kentucky).
Future strategy monitoring and modification	• "Maybe the dashboard will become more relevant as time goes on because it'll be a longer period of time that we'll be looking at and so we'll have more time to reflect on and we'll have a bigger window of time that we're looking at" (Massachusetts). • "So I think it'll be even more important later on in the study after the community has been established, and the goals or your strategies have been in place for a while, and you can start seeing the effect" (Kentucky). • "Maybe as things change in terms of opioid response, there might be a need to start looking at some other forms of data, but the data supplied now are okay, thank you" (New York).
Use conditional on data changes	• [When asked whether they would use the dashboard next year] "So I feel, I feel like if the data is, if I can track the same data on the dashboard that my DC [data coordinator] is giving to us, then yes" (Ohio). • "[T]he reality is that the workload is very, very high and we're frequently, and this won't surprise you, we're frequently coordinators and CEFs [community engagement facilitators] in a position of hurry up and do this today. And so it doesn't allow for a lot of time for reflection and kind of that kind of thoughtful consideration...maybe it should be built into something a little bit more so that it is part of something" (Massachusetts). • "I think it would be good for a community like the HEALing [Communities Study] to have that presence long term and having it be a problem-solving person means it could be one person serving multiple sites and not just ongoing doing something, but really targeted to needs. I think would be very helpful" (New York).
Preference	
Content	• "If we could break it down like [deidentied participant] said earlier by zip code that would be phenomenal because you know we wouldn't have to wait for the meeting, we'd be able to see it anytime" (Ohio). • "Well, I would love to see the information for the rest of the study in the community personally. I mean, I can understand why they limited it...I don't mind comparing, you know, this one township to different counties and seeing how people are doing, you know...if the numbers showed something drastic, I would call over there and say, 'What are you doing that we're not doing?'" (New York) • "One is that it feels a bit, I don't know, not really, doesn't get that accurate. Maybe it just can't. Maybe there's no way for it possibly to give an accurate illustration of the communities that it is supposed to represent, and maybe there's just no way to do that. Maybe it's just because I know the community so well, and maybe if somebody didn't know the communities at all, they would look at that and say, 'Wow, this is really helpful'" (Massachusetts).
Function	• "I just wish it was publicly available, because if I could send a link to somebody on my coalition there's much better chance they'd be able to see it and access it" (Ohio). • "It's not accessible to general community or other groups that are not inside the [HCS[a]] structure. So, I guess there's that piece, so thinking about that. Is there a purpose and a reason to make it public?" (Kentucky) • "Something that has been frustrating for me, from some of the HCS data that we get is we get a lot of plots, but we never get the underlying data. And always what I want to do is be able to download it, download it into like visualization software that I use, add those like timeline components, and then present it in a way that I know my community will understand" (New York).
Aesthetics	• "My thing now is how can I make it into a Facebook post-type of thing to grab people's attention; that we can start the conversation and engage them in the conversation. Having tools or in a format that we can say, 'Let's share this part and this part.' Perhaps get some people engaged in talking about this issue or finding a champion about this issue. How can we make it usable like that?" (Kentucky) • "I think it could be structured a little bit differently. You know specific to Healing Communities Study, you know, if we, if we really intended EBP [evidence-based practice] selection and monitoring for these tools, then I think they would have been they should be structured a little bit differently. The dashboard is getting there. I know we have like an MOUD view and a safer prescribing view, which I think is really helpful and you know if that was more like the primary way of looking at it. I think that would really increase its use for what we hope it would be used for" (Ohio).

[a]HCS: Healing Communities Study.

Figure 3. System Usability Scale (SUS) results by domain and study site (n=58). Raw SUS scores are presented. (A) SUS questions that were framed positively. (B) SUS questions that were framed negatively. Each SUS item is scored on a scale of 0 (strongly disagree) to 4 (strongly agree). A total of 4 participants did not respond to the SUS. SA: Kentucky; SB: Massachusetts; SC; New York; SD: Ohio.



Perceived Usefulness

Sites were generally positive in their reported perceptions about inconsistencies with the dashboards (SUS6); however, there was variability in the scores across states when it came to confidence in using the dashboards (SUS9). For example, participants from Ohio reported the least confidence, whereas participants from Kentucky scored the confidence domain the highest.

The alignment of the dashboards and community stakeholders' expectations may have influenced perceptions of confidence in using the dashboards and consistency in using the dashboards. On the basis of the TAM-informed interviews, 3 subthemes emerged regarding the benefits of using the CTH dashboards: *knowledge dissemination* (the dashboard was used to increase the awareness of activities in a community [eg, sharing data and validating existing assumptions]); *decision-making tools* (the dashboard was used for choosing EBPs in a community and evaluating EBPs strategies); and *access to more data* (broader and granular dashboard data about a community were available). In contrast, 3 subthemes emerged regarding the drawbacks of using the CTH dashboards: *time constraints* (the use of the dashboard felt time intensive, and other priorities distracted from its use); *misalignment of dashboard as a tool* (the use of the dashboard was not aligned with the HCS's goals and workflow); and *disutility of data* (dashboard data were not translatable to anything actionable). Table 2 provides illustrative quotes from the interviews for each of the primary themes under perceived usefulness.

Perceived Ease of Use

Around integration (SUS5), perceptions among all participants were generally positive. Participants from Ohio did not respond as positively about ease of use and complexity of use (SUS2 and SUS3). Participants from Massachusetts, New York, and Kentucky scored the dashboards higher in these areas. Participants from Ohio and Massachusetts also scored training, support, and the cumbersomeness of the system (SUS4, SUS7, and SUS8) less favorably.

The alignment of the technical functionality of the dashboards with users' workflows may be linked to the users' perceptions of the dashboards' complexity and ease of use and the need for training and support. Our TAM-informed interviews revealed 4 subthemes around barriers to use: *access to dashboard* (participants faced challenges with accessing the dashboard); *data manageability* (accessing the data users needed was cumbersome); *unable to locate usable data* (data in the dashboard were not easy to find because of the lack of availability [due to lags in reporting or suppression rules] or because of navigation issues on the dashboard); and *presentation of numbers and labels* (the data were displayed in ways that made their use difficult for participants). However, we discovered 3 subthemes that facilitated the use of the dashboards: *additional support* (training or information that informed the use of the dashboard or data); *data accessibility* (the dashboard was useful because of aspects of the technology or the way it is displayed); and *navigability* (different ways in which the dashboard was easy to use and navigate and aspects of the dashboard that could be changed to make it better). Table 3 provides illustrative quotes for each of the primary themes under perceived ease of use.

Intention to Use

Scores of the SUS items on the frequency of planned use and the knowledge needed to use the dashboard (SUS1 and SUS10) were equivocal and lower across all the sites. Participants from Massachusetts reported that they wanted to use the dashboard frequently, but they needed to learn more about how to use it. Participants from Ohio provided the lowest score for the SUS item on the frequency of planned use of the dashboard and, similar to their Massachusetts counterparts, indicated the need to learn more about how to use the dashboard.

The willingness of community stakeholders to use the dashboard in the future may influence their desire to use the dashboard frequently and learn how to use it. During the TAM-informed interviews, participants provided several recommendations for improving the dashboard that could promote future use. A total of 3 subthemes emerged regarding the acceptance of the dashboards in the future: *alternate data uses* (data on the dashboard will be useful but for purposes not initially part of the focus of the HCS); *future strategy monitoring and modification* (the dashboard will be used as intended for HCS purposes, including for EBP strategy monitoring and data-driven decision-making about these strategies); and *use conditional on data changes* (the dashboard will be useful if some conditions are met [to help with its use]). In addition, 3 subthemes emerged regarding preferences with desirable use over time: *content* (participants desired changes to the data on the dashboard, including the provision of different data, to improve the usability of the data and their interpretation); *function* (participants desired additional tools or different ways to navigate the dashboard and use data); and *esthetics* (participants described ways in which the appearance of the dashboard could be improved to effectively communicate information). Table 4 provides illustrative quotes on recommendations for each of the primary themes under intention to use.

Contextual Factors

Collective circumstances may influence the dashboard's perceived usability and use. A total of 2 subthemes emerged regarding how the community connects and works with data (community data orientation): *comfort with data* (collective perceptions of the comfort with using tools such as dashboards that contain data to make decisions); *established tools and approaches* (perceptions that ranged from those of human-centered data sources [eg, existing relationship with a coroner's office] to technology-based sources that were community preestablished substitutes to the HCS dashboard). Table 5 provides exemplary quotes for each of the primary themes under the community context that illustrate how dashboards and their use may have been perceived when implemented.

Table 5. Subthemes under contextual factors.

Community data orientation	Illustrative quotes
Comfort with data	• "[T]he treatment providers, they're not necessarily data people, but there's some of us that are looking at it and using it. I guess that's all that matters in the long run. That's something that we're always going to be dealing with the people in the coalition that we have" (Massachusetts). • "I think it goes back to who your audience is...If you're looking to speak to people who, you know, I kind of consider myself a middle person for that aspect, right, so if you're trying to speak to community members who have no data background and things like that, they're -- first of all, they're not even going to notice that [a date] is missing..." (New York). • "[S]o we are planning on either in our August or September drug coalition meeting kind of...showing people how it works, showing people the data that's in there, and hopefully getting people a little bit more oriented towards...looking at that dashboard to see if what we're doing is making a difference, because I don't think that we are, as a coalition I don't think we are data oriented and enough really if that makes sense" (Ohio).
Established tools and approaches	• "I mean, I think the primary source or the primary...way that data is distributed in [the community] is really just through personal connections...I mean you know I think every coalition meeting we've had our health commissioner comes and has the...harm reduction clinic numbers, you know written down on the back of an envelope or a napkin or something and that was you know the you know the organizations are generally very good about sharing that individual level data...You know, to help, to help people" (Ohio). • "So, they've got their ODMAP [Overdose Detection Mapping Application Program] data that comes in through HIDTA [high intensity drug trafficking areas] and the police departments, and then they have the [Department of Health] data that shows how many Narcan saves happened and how many overdoses occurred. And then on top of that, you have things like coroner's reports and coroner responses to actual deaths" (New York). • "I would say there's a couple of places we'll look depending on what it is. We use the KIPRC site a lot, Kentucky Injury Prevention. We use that depending on what they have. We also keep a data dashboard from the Health Department's perspective on substance use and opioid disorder. So, we have a dashboard on ours that looks at things. We do a daily dashboard on overdose visits to the ED and EMS runs and things like that...Periodically, we might get things from the Office of Drug Control Policy, the state. We also have a local Office of Drug Control Policy, which pulls some local data for all of our counties as well. So, I would say those are probably our main sources of substance use data" (Kentucky).

Discussion

Principal Findings

Our assessment indicated that the community stakeholders in the HCS found the CTH dashboards to be usable, as measured by the SUS, and easy to use and understand, as indicated by the themes identified through our TAM-informed interviews. Some respondents indicated the usefulness of the dashboards, with many indicating areas for improvement. From these findings, we have synthesized and prioritized the following gaps and lessons learned for future consideration, which are generalizable to community stakeholders engaged in dashboard use. In the spirit of Chen and Floridi [27], our lessons learned are specifically about different visualization pathways for use in dashboards among community stakeholders. Prior research has identified similar practices as supportive of the effective use (eg, greater engagement, cognitive alignment between end users, and decision aids) of dashboards. The lessons learned that we describe subsequently may help researchers design higher-fidelity dashboards that future scientific studies should consider when developing similar interventions and more general tools that integrate data visualizations for use among community stakeholders in community-oriented studies.

Prioritized Gaps in and Lessons Learned About Designing Dashboards for Community Stakeholders

Prioritized Gap: Cognitive Dissonance With the Dashboards

Many of the community stakeholders already had frontline knowledge of their community opioid crisis and its complexities. Hence, some community stakeholders may have seen the dashboards as tools simply providing hard data, which were secondary to their lived experiences within their communities and knowledge about contextual nuances with these communities. There may have been, arguably, a cognitive disconnect between what the community stakeholder expected from a dashboard and what the dashboard provided. This disconnect may have been exacerbated by factors such as adherence to the main HCS research protocol and purpose (eg, a brief timeline for development and programming before deployment and the prohibition of data downloads from visualization), use of suppression rules that masked values and contributed to the spareness of data, lags in metrics due to reporting, and limited connections of the data with local resources for community stakeholders to act upon. Other challenges included balancing the need for "real-time" data with the validity of these data due to retrospective corrections.

Lesson 1: Use Storytelling via Dashboards

Cognitive and information science theories suggest the importance of aligning information representation formats provided by decision aids with mental representations required for tasks and cognitive styles of individuals [28-30]. We propose storytelling via dashboards as an effective, historically validated [31] approach to achieving this alignment. At a minimum, this involves providing basic data summaries in plain English, as done by the New York HCS site, and mixing and matching measures in graphs to allow the illustration of specific points, as adopted by the Ohio and Kentucky sites. Richer storytelling may involve multiple iterations or cocreations with community stakeholders that help craft the right set of qualitative contexts and information that situates quantitative data to evoke understandings that fit with internal mental models of the task and lived experiences [32,33]. Powerful stories could provide community stakeholders with structured templates to analyze, justify, and communicate data. Our findings suggested that there were indeed community stakeholders who were not comfortable with using data; individuals with such predispositions could find data presented as stories more meaningful, especially during decision-making processes.

Lesson 2: Link Actionable Insights to Useful Resources

Community stakeholders may require additional support with using insights garnered from dashboards. Dashboards can be seen as a mediator between data and a call to action to address specific issues. Moreover, dashboards can present structured sets of actions a community stakeholder can undertake, also known as an actionable impetus [34], as guidance on addressing a problem identified with the data [35-37]. For example, some community stakeholders in the HCS would have appreciated a set of recommendations for EBPs or resources within a local neighborhood that a coalition could have acted upon, given an outbreak of opioid deaths. Design considerations could have made affordances (eg, alerts in the user interface) to create an impetus for action, which has been described as an "actionable data dashboard" [34]. Such actionable dashboards could provide important links to specific issues with the opioid crisis, such as the linkage of distribution centers for naloxone kits to community hot spots where there is unequal access to these kits based on factors such as race and ethnicity [38].

Prioritized Gap: Gatekeeping of the Data and Dashboards

Access to the dashboards was a noteworthy barrier under perceived ease of use. Some study participants indicated that this barrier hindered the use of data to facilitate decision-making among community members. Dashboard requirements limiting access to certain stakeholders was due to study protocol requirements to prevent communities from benchmarking data across communities. However, the user and password verification requirements were seen as inconvenient, and access control was seen as a form of gatekeeping that limited information sharing with key stakeholders who did not have dashboard access and were critical to decision-making in communities.

Lesson 3: Use Processes and Tools That Promote Access and Sharing

Arguments to enhance the use of data included approaches that facilitate human-human interactions in the sensemaking of data within dashboards and visualizations [3,39,40]. In addition to storytelling, other practices include improving how study insights are shared with communities and how communities are encouraged to share insights; permissible sharing includes dashboard export features or providing embedded codes so that data or narratives can be used in media, reports, presentations, websites, and social media content [41].

Prioritized Gap: Low Engagement With the Dashboards

Several themes (eg, data manageability, additional support, use conditional on data changes, and aesthetics) highlighted additional challenges to perceived ease of use and intention to use that could be represented by a broader concept used in the technological literature known as digital engagement [42-44]. The HCS missed opportunities to promote community stakeholder engagement with the CTH dashboards through which enhancements could have been made in areas such as dashboard training and clearer protocols around how community stakeholders used the dashboards during coalition meetings.

Lesson 4: Use a Multisector and Interdisciplinary Approach to Understand and Improve Engagement With Data Visualizations

The use of audit log file data or eye tracking to monitor and characterize use and user preferences (eg, line chart vs bar graphs) can inform design decisions and tailored interventions to foster behavioral changes that support the effective use of data among community stakeholders [44-46]. Our study, while recognizing early on that log files can support engagement, faced challenges with implementing a standard data model to track visual preferences and visit frequencies among community stakeholders, a decision that could have supported systematic transformations to the dashboards. In addition, our findings demonstrated that the combined use of SUS scores and TAM-informed interview data was beneficial in obtaining a comprehensive perspective on how engagement, via usability, can be transformed. For example, prior research has indicated that barriers in usability may have been overlooked if only measured by a single approach, such as exclusively using the SUS [47].

The use of innovative training (eg, dashboard navigators and indexed videos) and workflow changes (eg, starting meetings with referencing insights from a dashboard) are 2 other practice-based examples that can foster engagement among community stakeholders by supporting the realignment of existing mental models and the building of new ones with visual representations to achieve specific tasks [48,49]. For example, the Massachusetts HCS site adopted "community data walks," in which expert dashboard users provided active demonstrations to community stakeholders on how to use the CTH dashboards.

Our findings suggest that the fidelity of the cocreation process to develop the CTH dashboards had gaps. The HCS is one of the first large-scale applications of the cocreation process to design dashboards within the community context. The pragmatic, multisetting nature of the study reflects challenges to balancing a research agenda with the expectations of diverse communities. Interventions and strategies (eg, multiple plan-do-study-act cycles during the cocreation process) that can address gaps in fidelity in using cocreation could be adopted to ensure that the final outputs are indeed relevant to end users' work and promotes higher engagement. There were no "checklists" to support our implementation at this level of complexity, and we hope that the lessons we learned establish the foundations for such an approach for future endeavors.

Limitations

Our definition of community stakeholders may be idiosyncratic to our study and parent study research protocols. However, the 4-site design provides a diverse sample of individuals who represent different professional roles and social and demographic groups; the samples were not homogeneous across sites, including in terms of technology literacy, which was not assessed yet could shape end users' experience with the dashboards [48]. The COVID-19 pandemic may have influenced perceptions of use of and intentions to use the CTH dashboards, given pandemic-driven competing demands, technology-mediated workflow, etc, experienced by the respondents. This may have resulted in conservative perceptions of dashboard usability; however, the public health domain is generally faced with competing demands and emerging situations, which might mitigate such notions. Our research study was relatively large in scope; hence, ensuring the fidelity of our assessment in general and the interviews specifically was challenging. However, as noted in the Methods section, we worked closely across sites to maintain rigor from study inception. We acknowledge that although the SUS is a validated tool, it may have lacked specificity for assessing CTH dashboards. Other TAM constructs (eg, external environment) and more updated versions of the TAM exist [50,51]. However, one of our original goals was to build a framework for dashboard use among community stakeholders; therefore, we chose to use a fundamental version of the TAM that had both the critical constructs for our focus and did not have other constructs that may have imposed a priori assumptions on dashboard usability from other fields and studies.

Conclusions

Data dashboards can be an evidence-based approach to support community-based public health decision-making. These technological interventions to visualize and interact with data can support transformations in community health, including during public health crises, such as the COVID-19 pandemic and the opioid epidemic. Our study is a novel assessment using both the SUS and TAM to examine the usability of dashboards among community stakeholders. Participants provided multifaceted perspectives on the usability of the CTH dashboards and their intention to use these dashboards. On the basis of our findings, we presented important gaps that motivated the consolidation of lessons learned regarding dashboard use among community stakeholders.

Acknowledgments

This research was supported by the National Institutes of Health (NIH) and the Substance Abuse and Mental Health Services Administration through the NIH HEAL (Helping to End Addiction Long-term) Initiative under the award numbers UM1DA049394, UM1DA049406, UM1DA049412, UM1DA049415, and UM1DA049417 (ClinicalTrials.gov Identifier: NCT04111939).

The authors wish to acknowledge the participation of the Healing Communities Study communities, community coalitions, and community advisory boards and state government officials who partnered with us on this study. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH, the Substance Abuse and Mental Health Services Administration, or the NIH HEAL Initiative.

The authors are grateful to many individuals who were instrumental in the development of the Healing Communities Study. In particular, the authors acknowledge the work of those who supported coordination and the collection of data for this study.

Authors' Contributions

All authors contributed to the design, writing, and review of the manuscript. NF led in the writing of the manuscript. RGO, JV, LG, EW, and MH contributed to the original draft. Formal analysis was conducted by NF, RGO, LG, YW, and MH. All authors reviewed and edited subsequent versions of the manuscript.

Conflicts of Interest

JV reports affiliation with the National Institute on Drug Abuse.

Multimedia Appendix 1
Semistructured group interview guide on dashboards.
[DOCX File , 16 KB - humanfactors_v11i1e51525_app1.docx]

Multimedia Appendix 2
Supplementary tables with additional data.
[DOCX File , 39 KB - humanfactors_v11i1e51525_app2.docx]

References

1. Lange SJ, Moore LV, Galuska DA. Data for decision-making: exploring the division of nutrition, physical activity, and obesity's data, trends, and maps. Prev Chronic Dis 2019 Sep 26;16:E131 [FREE Full text] [doi: 10.5888/pcd16.190043] [Medline: 31560645]
2. Wu DT, Chen AT, Manning JD, Levy-Fix G, Backonja U, Borland D, et al. Evaluating visual analytics for health informatics applications: a systematic review from the American Medical Informatics Association Visual Analytics Working Group Task Force on Evaluation. J Am Med Inform Assoc 2019 Apr 01;26(4):314-323 [FREE Full text] [doi: 10.1093/jamia/ocy190] [Medline: 30840080]
3. Streeb D, El-Assady M, Keim DA, Chen M. Why visualize? Untangling a large network of arguments. IEEE Trans Vis Comput Graph 2021 Mar;27(3):2220-2236. [doi: 10.1109/TVCG.2019.2940026] [Medline: 31514139]
4. Wahi MM, Dukach N. Visualizing infection surveillance data for policymaking using open source dashboarding. Appl Clin Inform 2019 May;10(3):534-542 [FREE Full text] [doi: 10.1055/s-0039-1693649] [Medline: 31340399]
5. Fareed N, Swoboda CM, Lawrence J, Griesenbrock T, Huerta T. Co-establishing an infrastructure for routine data collection to address disparities in infant mortality: planning and implementation. BMC Health Serv Res 2022 Jan 02;22:4 [FREE Full text] [doi: 10.1186/s12913-021-07393-1]
6. Fareed N, Swoboda CM, Lawrence J, Griesenbrock T, Huerta T. Co-establishing an infrastructure for routine data collection to address disparities in infant mortality: planning and implementation. BMC Health Serv Res 2022 Jan 02;22(1):4 [FREE Full text] [doi: 10.1186/s12913-021-07393-1] [Medline: 34974826]
7. Hester G, Lang T, Madsen L, Tambyraja R, Zenker P. Timely data for targeted quality improvement interventions: use of a visual analytics dashboard for bronchiolitis. Appl Clin Inform 2019 Jan;10(1):168-174 [FREE Full text] [doi: 10.1055/s-0039-1679868] [Medline: 30841007]
8. Kummer BR, Willey JZ, Zelenetz MJ, Hu Y, Sengupta S, Elkind MS, et al. Neurological dashboards and consultation turnaround time at an academic medical center. Appl Clin Inform 2019 Oct;10(5):849-858 [FREE Full text] [doi: 10.1055/s-0039-1698465] [Medline: 31694054]
9. Fareed N, Swoboda CM, Chen S, Potter E, Wu DT, Sieck CJ. U.S. COVID-19 state government public dashboards: an expert review. Appl Clin Inform 2021 Mar;12(2):208-221 [FREE Full text] [doi: 10.1055/s-0041-1723989] [Medline: 33853140]
10. Everts J. The dashboard pandemic. Dialogue Hum Geogr 2020 Jun 17;10(2):260-264. [doi: 10.1177/2043820620935355]
11. Vahedi A, Moghaddasi H, Asadi F, Hosseini AS, Nazemi E. Applications, features and key indicators for the development of Covid-19 dashboards: a systematic review study. Inform Med Unlocked 2022;30:100910 [FREE Full text] [doi: 10.1016/j.imu.2022.100910] [Medline: 35342788]
12. Dixon BE, Grannis SJ, McAndrews C, Broyles AA, Mikels-Carrasco W, Wiensch A, et al. Leveraging data visualization and a statewide health information exchange to support COVID-19 surveillance and response: application of public health informatics. J Am Med Inform Assoc 2021 Jul 14;28(7):1363-1373 [FREE Full text] [doi: 10.1093/jamia/ocab004] [Medline: 33480419]
13. Chandler RK, Villani J, Clarke T, McCance-Katz EF, Volkow ND. Addressing opioid overdose deaths: the vision for the HEALing communities study. Drug Alcohol Depend 2020 Dec 01;217:108329 [FREE Full text] [doi: 10.1016/j.drugalcdep.2020.108329] [Medline: 33075691]
14. Wu E, Villani J, Davis A, Fareed N, Harris DR, Huerta TR, et al. Community dashboards to support data-informed decision-making in the HEALing communities study. Drug Alcohol Depend 2020 Dec 01;217:108331 [FREE Full text] [doi: 10.1016/j.drugalcdep.2020.108331] [Medline: 33070058]
15. Mao AY, Chen C, Magana C, Caballero Barajas K, Olayiwola JN. A mobile phone-based health coaching intervention for weight loss and blood pressure reduction in a national payer population: a retrospective study. JMIR Mhealth Uhealth 2017 Jun 08;5(6):e80 [FREE Full text] [doi: 10.2196/mhealth.7591] [Medline: 28596147]
16. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
17. Brooke J. SUS: a 'quick and dirty' usability scale. In: Usability Evaluation In Industry. Boca Raton, FL: CRC Press; 1996.

18. HEALing Communities Study Consortium. The HEALing (Helping to End Addiction Long-term SM) communities study: protocol for a cluster randomized trial at the community level to reduce opioid overdose deaths through implementation of an integrated set of evidence-based practices. Drug Alcohol Depend 2020 Dec 01;217:108335 [FREE Full text] [doi: 10.1016/j.drugalcdep.2020.108335] [Medline: 33248391]
19. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
20. Fram SM. The constant comparative analysis method outside of grounded theory. Qual Rep 2013;18(1):1-25. [doi: 10.46743/2160-3715/2013.1569]
21. NVivo (version 12). Lumivero. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home [accessed 2022-12-15]
22. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
23. Gorbenko K, Mohammed A, Ezenwafor EI, Phlegar S, Healy P, Solly T, et al. Innovating in a crisis: a qualitative evaluation of a hospital and Google partnership to implement a COVID-19 inpatient video monitoring program. J Am Med Inform Assoc 2022 Aug 16;29(9):1618-1630 [FREE Full text] [doi: 10.1093/jamia/ocac081] [Medline: 35595236]
24. Colorafi KJ, Evans B. Qualitative descriptive methods in health science research. HERD 2016 Jul;9(4):16-25 [FREE Full text] [doi: 10.1177/1937586715614171] [Medline: 26791375]
25. Kortum PT, Bangor A. Usability ratings for everyday products measured with the system usability scale. Int J Hum Comput Interact 2013 Jan 03;29(2):67-76. [doi: 10.1080/10447318.2012.681221]
26. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
27. Chen M, Floridi L. An analysis of information visualisation. Synthese 2012 Sep 26;190:3421-3438. [doi: 10.1007/s11229-012-0183-y]
28. Luo W. User choice of interactive data visualization format: the effects of cognitive style and spatial ability. Decis Support Syst 2019 Jul;122:113061. [doi: 10.1016/j.dss.2019.05.001]
29. Vessey I. Cognitive fit: a theory-based analysis of the graphs versus tables literature. Decis Sci 1991 Mar;22(2):219-240. [doi: 10.1111/j.1540-5915.1991.tb00344.x]
30. Engin A, Vetschera R. Information representation in decision making: the impact of cognitive style and depletion effects. Decis Support Syst 2017 Nov;103:94-103. [doi: 10.1016/j.dss.2017.09.007]
31. Kosara R, Mackinlay J. Storytelling: the next step for visualization. Computer 2013 May;46(5):44-50. [doi: 10.1109/mc.2013.36]
32. Schuster R, Koesten L, Gregory K, Möller T. "Being simple on complex issues" -- accounts on visual data communication about climate change. arXiv Preprint posted online November 18, 2022. [doi: 10.1109/TVCG.2024.3352282]
33. Anderson KL. Analytical engagement - constructing new knowledge generated through storymaking within co-design workshops. In: Anderson KL, Cairns G, editors. Participatory Practice in Space, Place, and Service Design. Wilmington, DE: Vernon Press; 2022.
34. Verhulsdonck G, Shah V. Making actionable metrics “actionable”: the role of affordances and behavioral design in data dashboards. J Bus Tech Commun 2021 Oct 29;36(1):114-119. [doi: 10.1177/10506519211044502]
35. Cheng CK, Ip DK, Cowling BJ, Ho LM, Leung GM, Lau EH. Digital dashboard design using multiple data streams for disease surveillance with influenza surveillance as an example. J Med Internet Res 2011 Oct 14;13(4):e85 [FREE Full text] [doi: 10.2196/jmir.1658] [Medline: 22001082]
36. Fanzo J, Haddad L, McLaren R, Marshall Q, Davis C, Herforth A, et al. The food systems dashboard is a new tool to inform better food policy. Nat Food 2020 May 13;1:243-246. [doi: 10.1038/s43016-020-0077-y]
37. Eckerson WW. Performance Dashboards: Measuring, Monitoring, and Managing Your Business. Hoboken, NJ: John Wiley & Sons; 2005.
38. Nolen S, Zang X, Chatterjee A, Behrends CN, Green TC, Linas BP, et al. Evaluating equity in community-based naloxone access among racial/ethnic groups in Massachusetts. Drug Alcohol Depend 2022 Dec 01;241:109668 [FREE Full text] [doi: 10.1016/j.drugalcdep.2022.109668] [Medline: 36309001]
39. van Wijk JJ. The value of visualization. In: Proceedings of the VIS 05. IEEE Visualization. 2005 Presented at: VISUAL 2005; October 23-28, 2005; Minneapolis, MN. [doi: 10.1109/visual.2005.1532781]
40. Bruner J. Acts of Meaning: Four Lectures on Mind and Culture. Cambridge, MA: Harvard University Press; 1992.
41. Nelson LA, Spieker AJ, Mayberry LS, McNaughton C, Greevy RA. Estimating the impact of engagement with digital health interventions on patient outcomes in randomized trials. J Am Med Inform Assoc 2021 Dec 28;29(1):128-136 [FREE Full text] [doi: 10.1093/jamia/ocab254] [Medline: 34963143]
42. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018 Nov 16;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]

43. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
44. Fareed N, Walker D, Sieck CJ, Taylor R, Scarborough S, Huerta TR, et al. Inpatient portal clusters: identifying user groups based on portal features. J Am Med Inform Assoc 2019 Jan 01;26(1):28-36 [FREE Full text] [doi: 10.1093/jamia/ocy147] [Medline: 30476122]
45. Morgan E, Schnell P, Singh P, Fareed N. Outpatient portal use among pregnant individuals: cross-sectional, temporal, and cluster analysis of use. Digit Health 2022 Jul 06;8:20552076221109553 [FREE Full text] [doi: 10.1177/20552076221109553] [Medline: 35837662]
46. Nadj M, Maedche A, Schieder C. The effect of interactive analytical dashboard features on situation awareness and task performance. Decis Support Syst 2020 Aug;135:113322 [FREE Full text] [doi: 10.1016/j.dss.2020.113322] [Medline: 32834262]
47. Tieu L, Schillinger D, Sarkar U, Hoskote M, Hahn KJ, Ratanawongsa N, et al. Online patient websites for electronic health record access among vulnerable populations: portals to nowhere? J Am Med Inform Assoc 2017 Apr 01;24(e1):e47-e54 [FREE Full text] [doi: 10.1093/jamia/ocw098] [Medline: 27402138]
48. Chrysantina A, Sæbø JI. Assessing user-designed dashboards: a case for developing data visualization competency. In: Proceedings of the 15th IFIP WG 9.4 International Conference on Social Implications of Computers in Developing Countries. 2019 Presented at: ICT4D 2019; May 1-3, 2019; Dar es Salaam, Tanzania. [doi: 10.1007/978-3-030-18400-1_37]
49. McAlearney AS, Walker DM, Sieck CJ, Fareed N, MacEwan SR, Hefner JL, et al. Effect of in-person vs video training and access to all functions vs a limited subset of functions on portal use among inpatients: a randomized clinical trial. JAMA Netw Open 2022 Sep 01;5(9):e2231321 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.31321] [Medline: 36098967]
50. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manag Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
51. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003 Sep;27(3):425-478. [doi: 10.2307/30036540]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
CTH: Communities That HEAL
DCC: data coordinating center
EBP: evidence-based practice
HCS: Healing Communities Study
SUS: System Usability Scale
TAM: technology adoption model

Edited by A Kushniruk; submitted 02.08.23; peer-reviewed by TN Nguyen, Z Shan; comments to author 07.10.23; revised version received 08.11.23; accepted 05.05.24; published 09.09.24.

Please cite as:
Fareed N, Olvera RG, Wang Y, Hayes M, Larimore EL, Balvanz P, Langley R, Noel CA, Rock P, Redmond D, Neufeld J, Kosakowski S, Harris D, LaRochelle M, Huerta TR, Glasgow L, Oga E, Villani J, Wu E
Lessons Learned From Developing Dashboards to Support Decision-Making for Community Opioid Response by Community Stakeholders: Mixed Methods and Multisite Study
JMIR Hum Factors 2024;11:e51525
URL: https://humanfactors.jmir.org/2024/1/e51525
doi:10.2196/51525
PMID:39250216

Original Paper

A Chatbot-Based Version of the World Health Organization–Validated Self-Help Plus Intervention for Stress Management: Co-Design and Usability Testing

Valentina Fietta[1,2], MSc; Silvia Rizzi[1], MSc; Chiara De Luca[3], MSc; Lorenzo Gios[1], PhD; Maria Chiara Pavesi[3], MSc; Silvia Gabrielli[1], PhD; Merylin Monaro[2], PhD; Stefano Forti[1], MSc

[1]Fondazione Bruno Kessler, Trento, Italy
[2]Department of General Psychology, University of Padova, Padova, Italy
[3]Istituto Pavoniano Artigianelli, Trento, Italy

Corresponding Author:
Valentina Fietta, MSc
Fondazione Bruno Kessler
Via Sommarive, 18, Povo TN
Trento, 38123
Italy
Phone: 39 049 827 6628
Email: valentina.fietta.1@phd.unipd.it

Abstract

Background: Advancements in technology offer new opportunities to support vulnerable populations, such as pregnant women and women diagnosed with breast cancer, during physiologically and psychologically stressful periods.

Objective: This study aims to adapt and co-design the World Health Organization's Self-Help Plus intervention into a mobile health intervention for these target groups.

Methods: On the basis of the Obesity-Related Behavioral Intervention Trials and Center for eHealth Research and Disease Management models, low-fidelity and high-fidelity prototypes were developed. Prototypes were evaluated by 13 domain experts from diverse sectors and 15 participants from the target groups to assess usability, attractiveness, and functionality through semantic differential scales, the User Version of the Mobile Application Rating Scale questionnaire, and semistructured interviews.

Results: Feedback from participants indicated positive perceptions of the mobile health intervention, highlighting its ease of use, appropriate language, and attractive multimedia content. Areas identified for improvement included enhancing user engagement through reminders, monitoring features, and increased personalization. The quality of the content and adherence to initial protocols were positively evaluated.

Conclusions: This research provides valuable insights for future studies aiming to enhance the usability, efficacy, and effectiveness of the app, suggesting the potential role of a chatbot-delivered Self-Help Plus intervention as a supportive tool for pregnant women and women with a breast cancer diagnosis.

(JMIR Hum Factors 2024;11:e64614) doi:10.2196/64614

KEYWORDS

acceptance and commitment therapy; ACT; well-being; pregnancy; breast cancer; eHealth; mobile health; mHealth; development; usability; user-centered design

Introduction

Background

The growing awareness of the profound significance of mental health for individuals and society has spurred an expanding body of research to scrutinize global population trends and the strategies used to address this issue. Empirical evidence consistently reveals an enduring surge in requests for psychological support, yet this burgeoning demand remains largely unmet due to scarce available resources and services [1]. Consequently, people's needs remain unmet. The factors contributing to the challenge of accessing mental health services are multifaceted. In general, these impediments encompass issues such as suboptimal service quality, inadequate levels of mental health literacy, pervasive stigma, and formidable cost barriers [2]. Within this context, developing and implementing strategies to fortify and enhance the health care system becomes increasingly imperative, rendering it more accessible to the population. Significantly, particular attention is being devoted to the prospective role of digital technologies, which can enhance the sustainability of the health care system by providing 24/7 support to patients and optimizing health care provider interventions [3]. These innovations aim to surmount the aforementioned impediments, advancing digital health as an integral and foundational strategy to foster equitable, affordable, and universally accessible mental health care [4]. A flourishing body of literature corroborates the potential of emergent technologies, encompassing telemedicine, mobile health (mHealth) initiatives, and digital therapies, in facilitating a seamless continuum of care, extending from clinical settings to patients' homes while embracing a staged care approach [3].

Furthermore, digital health may be particularly suitable for low-intensity mental health interventions [5]. This terminology refers to specific programs wherein the active engagement of health care professionals and specialists is not necessarily required. These interventions are grounded in empirically validated [6], evidence-based psychological practices seamlessly integrated into a self-help paradigm, whether guided or unguided. This intervention genre is conventionally designed to be transdiagnostic, offers facile adaptability across diverse contexts, and is readily implementable by nonprofessional operators. Given their structural attributes and overarching mission, low-intensity interventions represent a valuable conduit for augmenting access to pragmatic, evidence-based psychological treatments, catering to a broad spectrum of recipients. These encompass from individuals among the general population to those presenting with limited or mild symptomatic manifestations associated with distress and mental illness [6]. In summary, low-intensity interventions are positioned as a pivotal resource for addressing situations characterized by mild distress, in which failure to intervene effectively could potentially precipitate the escalation of these conditions into pathological states [7].

Psychological distress, encompassing stress, anxiety, and depression, frequently co-occurs with physical illnesses such as breast cancer [8]. Cultivating a more optimistic outlook has been demonstrated to play a role in disease management and recovery, underscoring the significance of holistically addressing physical conditions and mental health issues [9,10].

Even a completely different health condition from those mentioned previously, such as pregnancy, can expose women to similar psychological symptoms. Pregnancy is characterized by major transformations that significantly impact the woman physically, mentally, and socially. How the woman adapts to these changes determines the quality of her life and her levels of well-being [11]. Where adaptation is not functional, symptoms of psychological distress may occur; the most common conditions are anxiety, stress, and depression [11-14]. To date, psychoeducational interventions that promote women's psychological well-being during pregnancy are scarce and tend to focus mainly on samples of women with psychiatric symptomatology (eg, perinatal depression disorder) [15]. For this reason, our study is part of the digital health framework to support health prevention strategies in the first 1000 days of life [16].

Evidence of the effectiveness of psychological interventions targeting pregnant women or women with breast cancer is increasing [17,18], but access to care services still presents several challenges. Many women face geographic barriers, with specialized centers often far from their homes. In addition, a shortage of qualified personnel, such as psychologists, further limits access to specialized care [19]. The stigma associated with mental health problems can also prevent women from seeking psychological support [20]. Therefore, in numerous instances, these target groups are excluded from accessing the requisite services. However, low-intensity interventions emerge as pivotal in addressing this lacuna, and the strategic combination with mHealth methodologies can yield an effective, sustainable, and inclusive framework for augmenting the scalability of mental health interventions.

This model is poised to cater to a comprehensive user base, encompassing prevention for individuals at potential risk of mental distress and intervention for patients grappling with mild to moderate mental distress [7].

Self-Help Plus

Self-Help Plus (SH+) is a low-intensity group intervention for stress management initially developed to target populations that are numerous or hard to reach by health care professionals under the principle of improving and facilitating access to health care interventions [21]. The SH+ package has been incorporated into the expanding array of low-intensity psychological interventions endorsed by the World Health Organization (WHO) [21]. By design, SH+ is a transdiagnostic intervention that is applicable, meaningful, and safe for people with and without mental disorders. SH+ is based on acceptance and commitment therapy (ACT) [22,23], a form of cognitive behavioral therapy [24,25].

The SH+ intervention package has 3 main components: a prerecorded audio course, a facilitator manual, and a self‐help booklet for participants. This material has been translated into multiple languages and can be easily accessed on the web at the WHO website [26]. The audio material imparts key information about stress management and guides participants through individual exercises and small group discussions. The intervention is structured into five sessions focused on acceptance- and mindfulness-based techniques for stress management: (1) *grounding* (mindfulness), (2) *unhooking* (cognitive distancing), (3) *acting on your values* (value-based behavioral activation), (4) *being kind* (gratitude), and (5) *making room* (acceptance).

Preliminary studies report positive effects of SH+ with a potential impact on mental well-being; it had a significant long-term efficacy in a target population of refugees and asylum seekers exposed to stressful situations [27]. Other studies show a still debatable effect of SH+ when applied to health care professionals during the COVID-19 pandemic, whereas there emerges a need for further examining the potential role of the confounding effects of nonspecific factors [28].

This Research

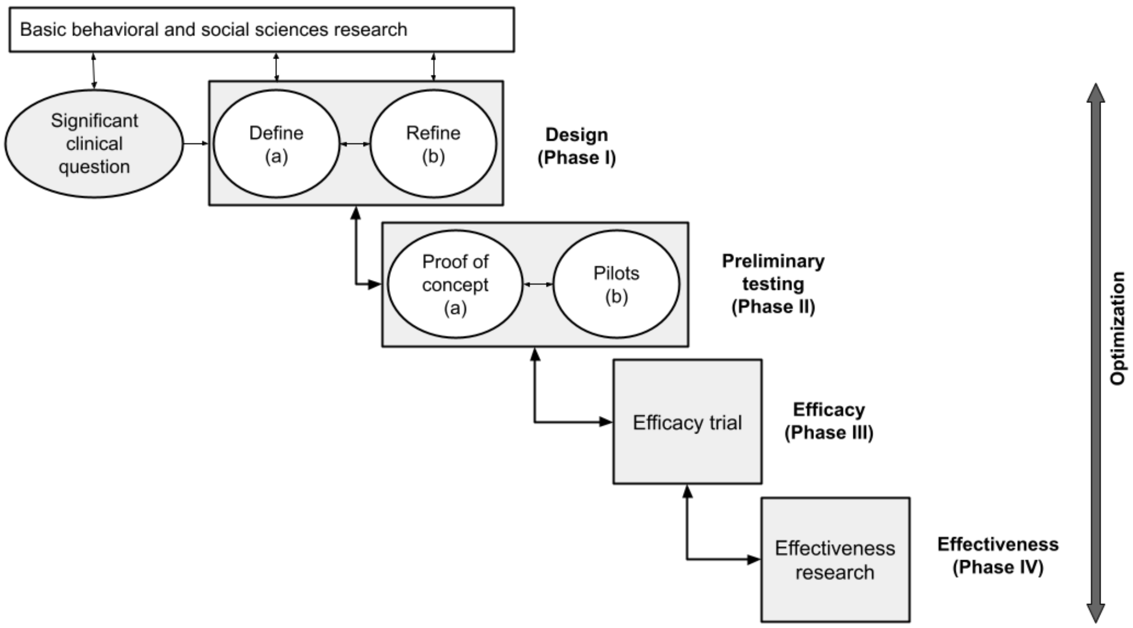
This research fits into the landscape of WHO strategies by adapting the stress management intervention developed by the WHO itself, SH+, with two main goals: (1) to assess the viability of this intervention when targeting specific subgroups (women with breast cancer and pregnant women) and (2) to validate the applicability of the intervention as a chatbot-delivered and preventive action. This SH+ intervention, which has already been validated and tested on some specific vulnerable populations (eg, asylum seekers) [29], will be fully available to users through digital tools. In particular, it will be delivered through a mobile app and guided by a virtual assistant, ALBA. This research aims to assess the prototype of the ALBA app to gather feedback and needs from key stakeholders to further refine the app from a qualitative perspective of usability, accessibility, and acceptability of the intervention delivered via a chatbot.

Methods

Overview

The stress management intervention was developed iteratively following the Obesity-Related Behavioral Intervention Trials (ORBIT) model [30], as illustrated in Figure 1 [30], which shows the pathway followed to translate a human-guided intervention into a possible digital therapy. In particular, the design and development process encompassed a multidisciplinary approach and continuous, systematic evaluation throughout, as the Center for eHealth Research and Disease Management (CeHRes) comprehensive road map approach recommended to improve the uptake and impact of eHealth technologies [31].

Figure 1. The Obesity-Related Behavioral Intervention Trials model.



The multidisciplinary project team, consisting of psychology, eHealth research, and communication experts, had biweekly meetings during the design and development phase. User-centered design methodologies ensured user involvement throughout the design and development process. Patient representatives, health care providers, and security experts were consulted throughout. The stress management intervention was developed in iterative processes through a combination of (1) intervention content development, identified and adjusted from the evidence-based cognitive behavioral stress management concept; and (2) iterative software development (phase 1—low-fidelity prototypes and phase 2—high-fidelity prototypes) and formative evaluation.

Intervention Content Development

A primary goal of this study was to adapt a validated stress management intervention, SH+ [21], into a new technology-based stress management intervention for pregnant women and women with breast cancer. To do so, the development involved a wide range of expertise encompassing psychology, eHealth IT, interaction design, and specialized knowledge in pregnancy and oncology. A multidisciplinary team was assembled to address diverse user needs, ensure psychological coherence, and integrate technological requirements and user-centered design principles. The focus was on achieving adaptability of the proposed tools in the digital domain [31] guided by 2 pivotal methodologies: user-centered design [32,33] and service design [34].

The development process comprised the following stages: (1) literature review of WHO protocols and papers on digital mental health and the specific psychological needs of target populations; (2) exploration gathering insights from user representatives (eg, patients with breast cancer and pregnant women), health care providers, and eHealth experts (including designers and developers); and (3) content adaptation customizing SH+ intervention manual content addressing software development and potential privacy and security issues. The intervention content was adapted and tailored by the entire research team through iterative processes to fit a 5 module–based intervention in electronic format. Adjustments were made to ensure easy language, short sentences, and focus on clear content for small screens.

This holistic approach, merging diverse expertise and user-centric methodologies, underscores the dedication to crafting a robust and efficient chatbot-driven intervention tailored for women dealing with breast cancer and pregnancy.

Phase 1: Iterative Development and Low-Fidelity Prototypes

The adaptation of the SH+ intervention through the implementation of the ALBA chatbot was based on a novel approach to delivering psychological support. Users can engage with a comprehensive and effective intervention through gamification, personalized sessions, reminders, and feedback. Therefore, a key aspect is represented by a multilevel structure to consider the different sections of the app and, simultaneously, to guarantee proper levels of user engagement, adherence, and overall impact on users' well-being. In the first iteration in 2023, a total of 3 psychologists and 2 communication experts tested and gave feedback on the prototype to ensure that the intervention program was logically built and would meet the stakeholder requirements.

On this basis, the following methodology was applied. The group of experts was divided into pairs, in which one person assumed the role of the chatbot whereas the other adopted the user's perspective, reading their respective segments of the dialogue aloud.

The pairs were reorganized for each of the 5 distinct modules to ensure a diverse spectrum of interactions and exhaustive coverage of potential dialogue scenarios and to avoid biases. This method facilitated the exploration of varied interaction dynamics and the collection of data on multiple communication styles. The oral recitation of dialogues served as a mechanism to evaluate several aspects of the chatbot's effectiveness, such as dialogue realism, rhythmicity and repetition of the texts, and communication fluency [35]. The verbalization process aids in gauging how seamlessly the chatbot replicates a humanlike conversation, identifying any inconsistencies or unnatural responses. Thus, listening to the dialogue's progression allows for the assessment of the conversation's smoothness, which encompasses the coherence and pertinence of the chatbot's replies. Another function of this method was to identify any ambiguities or misinterpretations that might emerge during interactions with the chatbot [35]. Simulating the conversation enabled the experts to offer immediate critiques on every facet of the dialogue, contributing to rapid and focused content refinement.

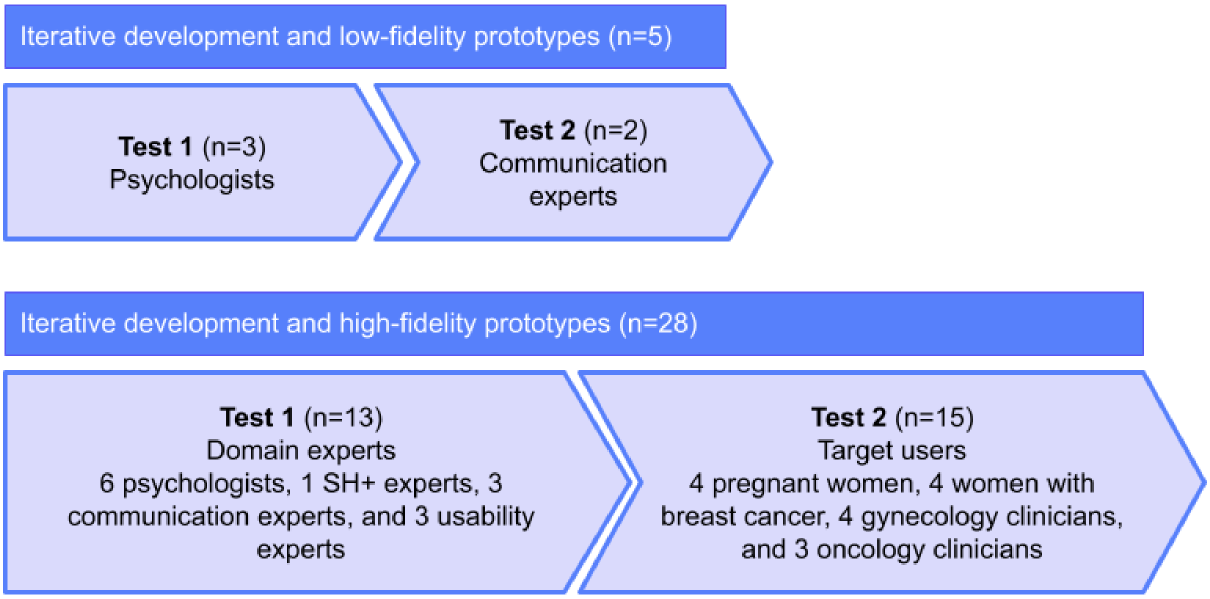
In a nutshell, this procedure aimed to analyze the chatbot's capability to engage in realistic, empathetic, and psychologically suitable conversations, leveraging direct feedback and expert psychologists' insights as key metrics for evaluation.

Phase 2: Iterative Development and High-Fidelity Prototypes

After minor adjustments, the paper prototype was implemented into an electronic format using the Landbot tool (HELLO UMI S.L.) [36]. Landbot is a chatbot generator that allows one to create, test, and deploy conversational chatbots via WhatsApp and other chat channels.

During the period spanning the end of 2023 and the beginning of 2024, a 2-phase evaluation of the ALBA prototype was conducted involving an overall sample of 28 participants, as elaborated on in Figure 2. In particular, of the 28 participants involved, 13 (46%) were domain experts (n=6, 21% psychologists), 1 (4%) was an SH+ expert, 3 (11%) were communication experts, 3 (11%) were usability experts, and 15 (54%) were users—8 (29%) target users (n=4, 50% women who were currently pregnant or had given birth within the previous year, 4/28, 14% of the total sample; and n=4, 50% women who were in current breast cancer disease status or follow-up from it, 4/28, 14% of the total sample) and 7 (25%) target clinicians (n=4, 57% gynecology clinicians [eg, obstetricians and gynecologists; 4/28, 14% of the total sample] and n=3, 43% oncology clinicians [eg, oncologists and case managers; 3/28, 11% of the total sample]). The recruitment of participants was based on personal contacts of researchers selected based on representativeness of the key target user groups addressed by the ALBA solution.

Figure 2. Software development and formative evaluation (N=37). SH+: Self-Help Plus.



Variable Identification

In an attempt to gather the necessary information, key variables were identified for investigation: communication, session structure, materials, engagement, functionality, esthetics, information, subjective and perceived impact, interaction, communication mode, involvement consistency, general comments, amelioration of technical implementation, and content adherence. The first 4 variables were assessed through the semantic differential tool [37], the following 5 were assessed through the User Version of the Mobile Application Rating Scale (uMARS) [38], and the latter 6 were assessed through an ad hoc semistructured interview.

The semantic differential tool is an instrument consisting of a series of scales, each of which is composed of a pair of bipolar adjectives between which a rating scale (5 positions) is placed. Given the study's variables, a list of subvariables was chosen to create ad hoc items for the research. Table 1 shows the chosen variables and their respective subvariables. On the basis of the target subject domain, each person was asked to evaluate and determine variables, as reported in Table 1 (single items are reported in Multimedia Appendix 1).

Table 1. List of variables investigated through the semantic differential tool and the people involved in evaluating the individual variables.

Variable and subvariable	Psychologists	SH+[a] expert	Communication experts	Usability experts	Target group	Clinicians
Communication						
Empathy and listening	✓[b]	✓	✓		✓	✓
Smoothness and fluidity	✓	✓	✓	✓	✓	✓
Chatbot interaction	✓	✓	✓	✓	✓	✓
Lexicon	✓	✓	✓	✓	✓	✓
Session structure						
Interaction length	✓	✓	✓	✓	✓	✓
Materials						
Audio tracks	✓	✓	✓		✓	✓
Infographics and videos	✓	✓	✓		✓	✓

[a]SH+: Self-Help Plus.

[b]Involved.

The uMARS questionnaire, on the other hand, evaluates mobile apps by covering 4 objective dimensions (engagement, functionality, esthetics, and information) and 1 subjective dimension. Briefly, the questionnaire consists of 20 items covering the inquired variables as follows—engagement (n=5, 25%), functionality (n=4, 20%), esthetics (n=3, 15%), and information (n=4, 20%)—and 4 items belonging to the subjective quality domain. A section on perceived impact (6 items) also assesses users' perceptions of the app's usefulness. Each answer is rated on a 5-point scale (1=inadequate, 2=poor, 3=acceptable, 4=good, and 5=excellent) measuring the usability of mHealth apps.

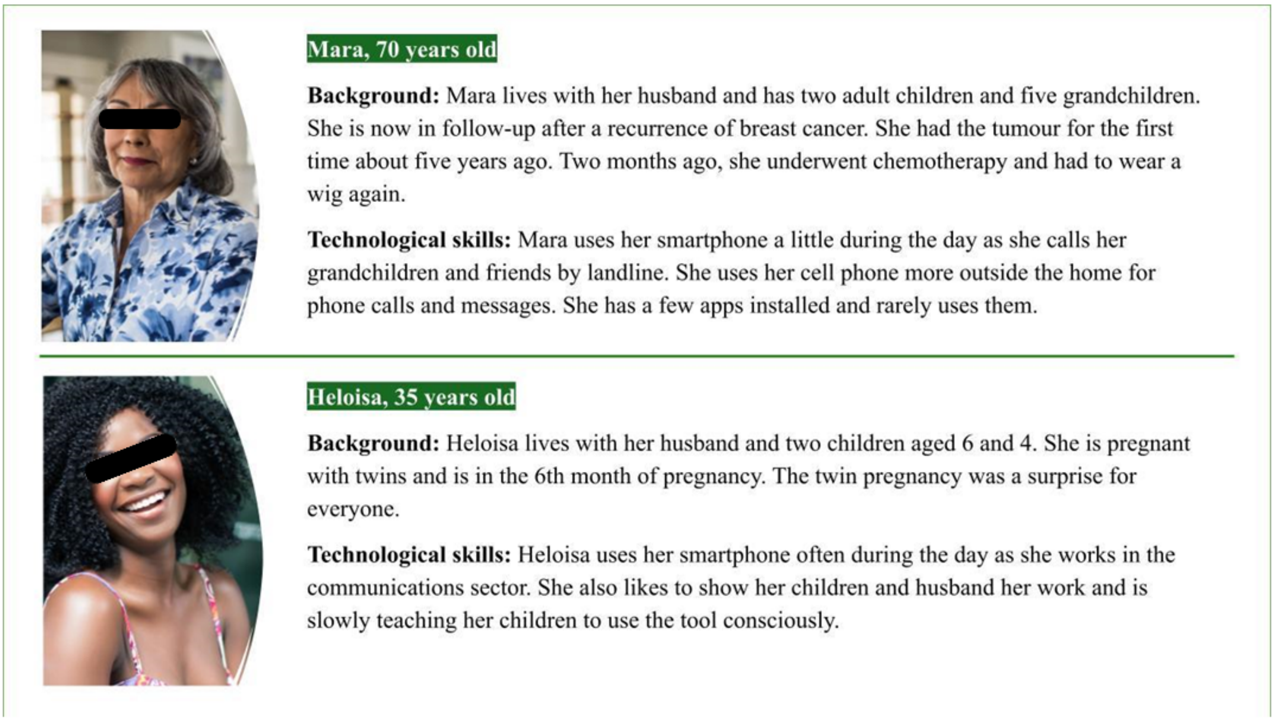
Interviews, instead, are suitable for a more in-depth investigation of users' attitudes and preferences regarding new technological solutions as open-ended discussions with users can help researchers better understand the issues and concerns related to the possible future adoption of these solutions [39]. Multimedia Appendix 2 shows the list of topics and questions posed during the interviews.

Context-Specific Methodologies

Personas

Personas are a user-centered and service design method used to create and visualize fictional representations of the target group [40]. Personas are an effective method for all project team members to better understand the target group for which the app is built. Personas in this study contained information about the pregnancy background, challenges, and technology use. Figure 3 provides illustrated examples of study personas. Psychologists had to take the perspective of one of these personas before the reading.

Figure 3. Examples of personas.



Focus on Experience and Expertise

This study aimed to involve experts from various fields and did not require all participants to respond to every specific ad hoc item of the semantic differential tool and questions of the semistructured interview. Usability experts were not asked to evaluate specific intervention content, whereas the SH+ expert was queried about the fidelity of the content to the original intervention during the interview. Clinicians, psychologists, and participants from the pregnancy context were asked questions relevant to that context, whereas those from the oncology field responded to questions specific to their area of work or direct life experience.

Procedure

All data were collected confidentially with participants' informed consent. Recruitment was conducted through word of mouth and direct acquaintance, clearly stating the study's objectives. Only volunteers aged ≥18 years were included. Before the study, participants received a privacy notice and consent form via Google Forms, allowing them to consent to participation and data processing.

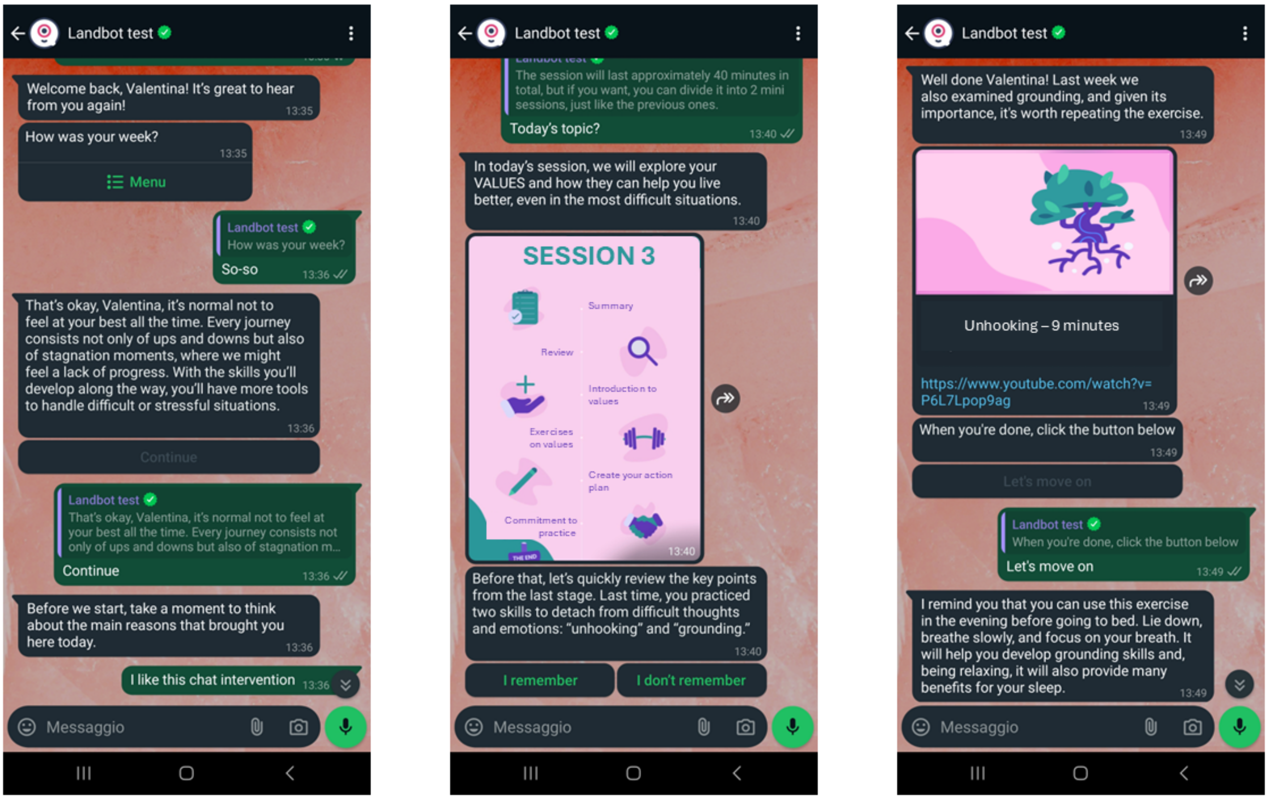
Operationally, the experimental procedure was structured as follows:

1. The information notice and informed consent were displayed.
2. Participants who decided to take part in the study were asked to fill out a questionnaire that collected some generic biographical information (age, schooling, gender, and employment status) and some information related to their knowledge and use of mobile apps.
3. Next, participants interacted with the ALBA app prototype for 6 weeks. Each session lasted approximately 40 minutes. During the study, the participants were asked to test the ALBA app on their phones; in particular, they had the opportunity to explore the interface and different sections of the app via mock-ups and read, listen to, and interact with the chatbot via WhatsApp during the dialogue session (see Figure 4 for ALBA chat mock-ups). The session

consisted of structured dialogues with predefined buttons for response options or free text. The text was not analyzed in this prototype phase as it was not the main purpose of this study. The dialogue was constructed to be as standardized as possible to maintain the already tested validity of the intervention. Participants were also informed about the future implementation of reminders and feedback for the activities proposed by the chatbot.

4. Questionnaires were then presented to the participants regarding their overall experience with the system. In the final stages, participants were asked to evaluate the experience they had with the ALBA app by answering 2 questionnaires. Specifically, the semantic differential tool [37] and the Italian version of the uMARS [41] were used for the evaluation through questionnaires.
5. Finally, a brief semistructured interview was conducted. After the usability assessment, participants were invited to join a web-based interview to further report on their expectations, preferences, and concerns regarding the ALBA solution tested. The interview questions were specifically designed ad hoc for our study, focusing on topics relevant to the evaluation of the app, such as interaction, communication mode, user involvement and consistency, technical implementation, and content adherence. The inclusion criteria for participating in the interview were having completed both the prototype test and the questionnaires as well as providing consent to participate in this additional phase. The exclusion criteria, on the other hand, were having dropped out during the test, not completing the questionnaires, or not providing consent for the interview. A total of 28 interviews were conducted by a researcher and audio recorded to enable a more detailed analysis of participants' responses. They were then analyzed and processed using qualitative tools. The interviewer initially provided a brief introduction to the interview objectives. Then, participants were asked to answer a series of semistructured questions regarding their expectations and preferences for using the app.

Figure 4. Examples of WhatsApp chat mock-ups. ALBA delivers content and monitors users' experiences by assigning tasks and providing motivational feedback. Through ALBA, users can interact, receive clear instructions, and access exercises and multimedia content.



To sum up, this study was divided into 2 main phases, each with distinct timelines and activities. Phase 1 focused on the iterative development of low-fidelity prototypes and spanned 6 months. Phase 2, dedicated to the iterative development of high-fidelity prototypes, lasted 4 months. Each step of phase 2 was carefully structured with specific time frames to ensure thorough data collection and analysis—10 minutes for informed consent completion, 15 minutes for pretest questionnaire completion, 6 weeks for app testing by participants, 20 minutes for posttest questionnaires, and 20 minutes for semistructured interviews.

Ethical Considerations

All data were collected in Italian and pseudonymized (deidentified) with participants' informed consent. Confidential audio recordings of semistructured interviews were used for data analysis, and participants were identified only by numeric codes. At the study's conclusion, participants can request the research outcomes from the research manager. Participants did not receive any compensation. This study was approved by the University of Padua Ethics Committee of Psychological Research on August 1, 2023 (reference 238-b).

Data Analysis

Data analysis for the quantitative results of the semantic differential tool and uMARS questionnaires was conducted using JASP and R (R Foundation for Statistical Computing) [42,43]. Due to the small sample size, nonparametric tests were used [44]. When applicable, the Wilcoxon signed rank test (*W*) was used as a nonparametric alternative to the 1-sample 2-tailed *t* test [44-46], and the rank-biserial correlation (*r*) was reported to indicate the strength of association along with its corresponding 95% CI [46]. All analysis results were considered significant with a critical *P* value set at .05.

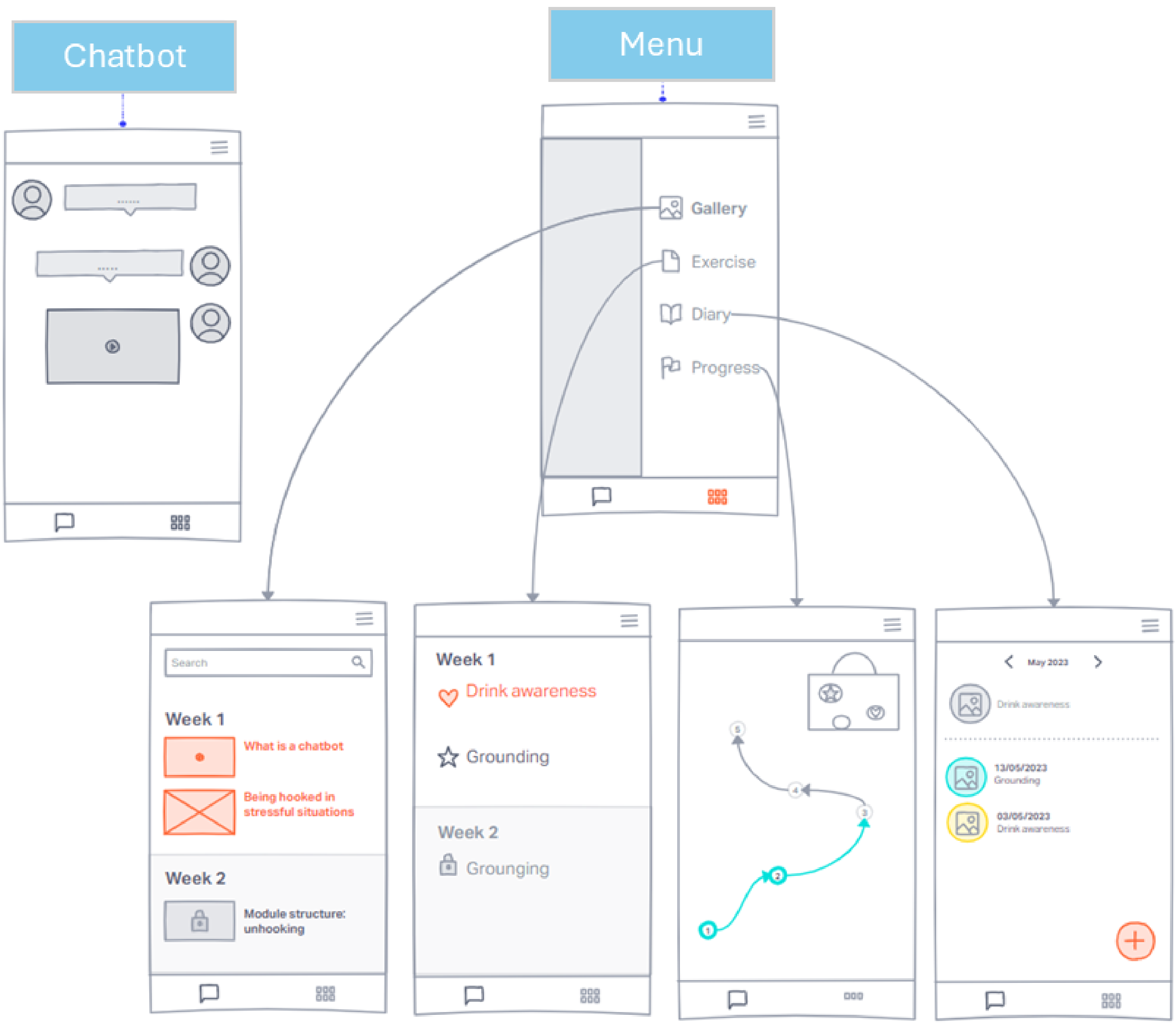
The data collected during the interviews were analyzed using a qualitative method. A thematic analysis was conducted [47], organizing the themes into tables based on different contexts and participant types (eg, psychologists and clinicians). The responses were analyzed by grouping the most prominent themes emerging from the ad hoc initial topics (around which the questions were formulated) into subvariables to address thematic redundancy within the participant sample. For this analysis, the interviews were first recorded and conducted by one author, whereas transcription and analysis were conducted by 2 other authors. Consensus on thematic relevance and redundancy was reached when both authors agreed; in case of disagreement, a third author was consulted to achieve a two-thirds majority. Finally, a comprehensive report was created highlighting the main findings with references to the specific groups where applicable.

Results

Phase 1: Iterative Development and Low-Fidelity Prototypes

Specifically, the app was designed to include 5 sections: *Chatbot*, *Exercises*, *Diary*, *Gallery*, and *Progress*. The first low-fidelity prototype version of the app was developed (Figure 5).

Figure 5. Low-fidelity prototype version of the ALBA app.



Functions of App Sections

ALBA assigns homework during the session delivery that the users should carry out during the following days, so the app *Diary* section supports self-monitoring of exercise progress and completeness of sessions, reinforcing users' empowerment. The completion of exercises is monitored in the evenings before the next session. All the exercises, divided week by week, are shown in the *Exercise* section. The *Gallery* page features all the multimedia material delivered by ALBA, such as educational videos, images, and intervention introductions. In addition, the app features a section called *Progress*, which provides users with an overview of their "journey" toward increased well-being. Using the journey metaphor, users can see the sticker badges earned by practicing exercises between sessions appearing on their suitcase illustration. These badges serve as long-term positive reinforcement and gratification for their efforts. Users receive badges for completion of exercises, and this is another way in which they can track their progress. Again, this gamification approach was selected to further reinforce empowerment and the sense of self-efficacy of the user.

Results of the Interaction With the Low-Fidelity Prototype

Regarding the interaction with the low-fidelity prototype in the co-design phase, 2 main themes emerged.

Communication

First, it was possible to review the communication style through role-play. To make the protocol more realistic, some parts of the dialogue were revised. The changes were made from a grammatical and syntactic point of view to make the text more fluid when reading. The changes also took empathy into account. This attention creates a feeling of trust in the relationship with ALBA. Through empathic communication, the person can feel in a safe space within which they are reassured and protected. For example, after these changes, ALBA provides the option to skip a particularly sensitive answer. It also lets the user find a calm place before doing audio exercises, and moreover, it gives the possibility to review the concepts of the previous session.

Duration

A second critical result that emerged from the specialists who tested the low-fidelity prototype was allowing the user to divide the session into 2 mini sessions. Given that the time needed to complete a single session is approximately 40 minutes, which is generally considered an excessive time to dedicate to interaction with an app. For this reason, appropriate changes were made to the dialogue to provide a partial and optional closure after 20 minutes followed by a gradual resumption of the session on the subsequent day. Thus, to support adherence, the session can be divided into 2 mini sessions, preventing user fatigue and improving the usability of both the chatbot and the app itself.

Other Improvements

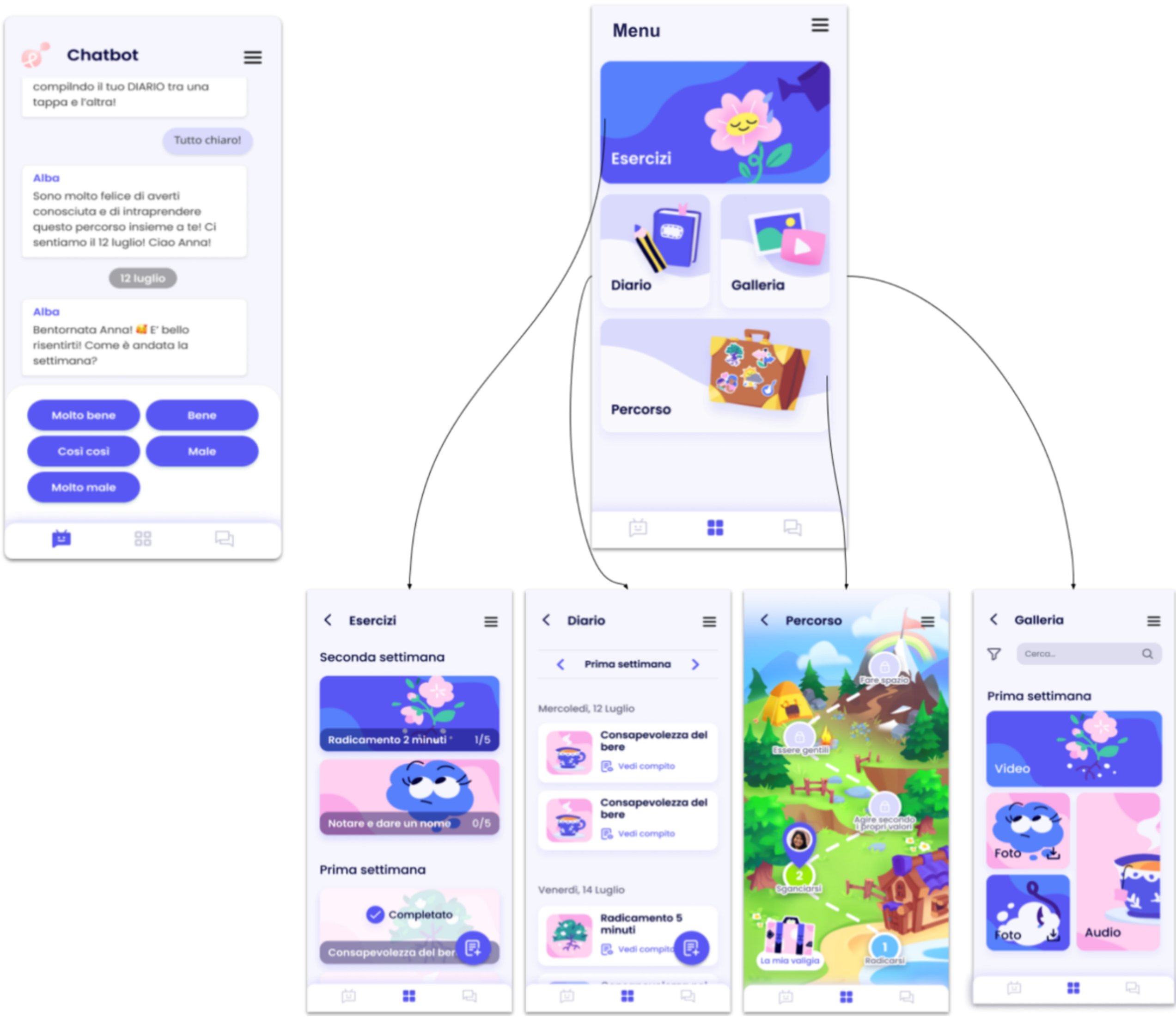
At the same time, writing errors were corrected, and the remaining parts of the original protocols designed for groups were adapted to the individual intervention. Mainly, attention was placed on the translation from group to individual gratitude exercises, the final exercise of each session.

Phase 2: Iterative Development and High-Fidelity Prototypes

Overview

To better understand the app and the entire intervention, more realistic mock-ups were created (Figure 6) after minor adjustments on the basis of the information gathered. This allowed users to better evaluate the app in its final appearance.

Figure 6. More realistic mock-ups of the ALBA app with its different sections.



In total, 28 individuals (n=4, 14% men and n=24, 86% women) participated in the pilot evaluation; the average age was 41.39 (SD 11.66; range 28-72) years. The average number of schooling years of the sample was 18.98 (SD 2.88; range 13-23).

Quantitative Results

Semantic Differential Tool

In this case, using a semantic differential–based questionnaire also facilitated the observation of the respondents' average positioning with respect to the 3 macrovariables under investigation. In particular, concerning the subvariables, several significant results emerged from the Wilcoxon analysis, as reported in Table 2. The graphical representation of mean values derived from the semantic differential tool for each item is reported in Multimedia Appendix 3. As seen in this figure, a tendency toward the positive semantic pole (right pole) emerged in the feedback of this group of participants.

Table 2. Results of the semantic differential tool (N=28)[a].

	Participants, n (%)	Values, mean (SD)	Values, median (range)	*W*	*P* value	*r* (95% CI)
Empathy and listening	25 (89)	3.55 (0.67)	3.60 (2.40-5.00)	213.50	<.001	0.85 (0.64 to 0.94)
Smoothness and fluidity	28 (100)	3.89 (0.69)	4.00 (2.00-5.00)	266.00	<.001	0.93 (0.83 to 0.97)
Chatbot interaction	28 (100)	3.72 (0.55)	3.75 (2.50-4.50)	271.50	<.001	0.97 (0.92 to 0.99)
Lexicon	28 (100)	4.41 (0.61)	4.50 (2.50-5.00)	404.00	<.001	0.99 (0.98 to 1.00)
Session structure	28 (100)	2.95 (0.63)	3.00 (1.00-4.00)	64.50	.87	−0.05 (−0.54 to 0.47)
Audio tracks	25 (89)	3.97 (0.44)	4.00 (3.00-5.00)	300.00	<.001	1.00 (1.00 to 1.00)
Infographics and videos	25 (89)	4.08 (0.50)	4.00 (3.00-5.00)	300.00	<.001	1.00 (1.00 to 1.00)

[a]For the Wilcoxon test, the effect size is given by the matched rank-biserial correlation.

The results indicate that participants generally responded positively across various subvariables. Significant positive responses were observed for empathy and listening, smoothness and fluidity, chatbot interaction, lexicon, audio tracks, and infographics and videos. The session structure subvariable did not show a significant positive trend, indicating a neutral or mixed response. The effect sizes ranged from moderate to large, suggesting varying degrees of impact for the different subvariables under investigation.

uMARS Results

The uMARS evaluated the respondents' average positioning in 4 key dimensions. Detailed results of the Wilcoxon tests are summarized in Table 3.

Table 3. Results of the User Version of the Mobile Application Rating Scale (N=28)[a].

	Participants, n (%)	Values, mean (SD)	Values, median (range)	*W*	*P* value	*r* (95% CI)
Engagement	28 (100)	3.55 (0.43)	3.60 (3.00-4.60)	300.00	<.001	1.00 (1.00-1.00)
Functionality	28 (100)	4.16 (0.51)	4.25 (3.00-5-00)	378.00	<.001	1.00 (1.00-1.00)
Esthetics	28 (100)	3.86 (0.49)	4.00 (3.00-4.67)	300.00	<.001	1.00 (1.00-1.00)
Information	28 (100)	4.20 (0.52)	4.25 (2.50-5.00)	405.00	<.001	1.00 (0.99-1.00)
Subjective items	28 (100)	3.20 (0.50)	3.25 (2.25-4.00)	233.50	.06	0.44 (0.02-0.72)
Perceived impact	28 (100)	3.66 (0.60)	3.83 (2.67-4.50)	264.00	<.001	0.91 (0.79-0.97)

[a]For the Wilcoxon test, the effect size is given by the matched rank-biserial correlation.

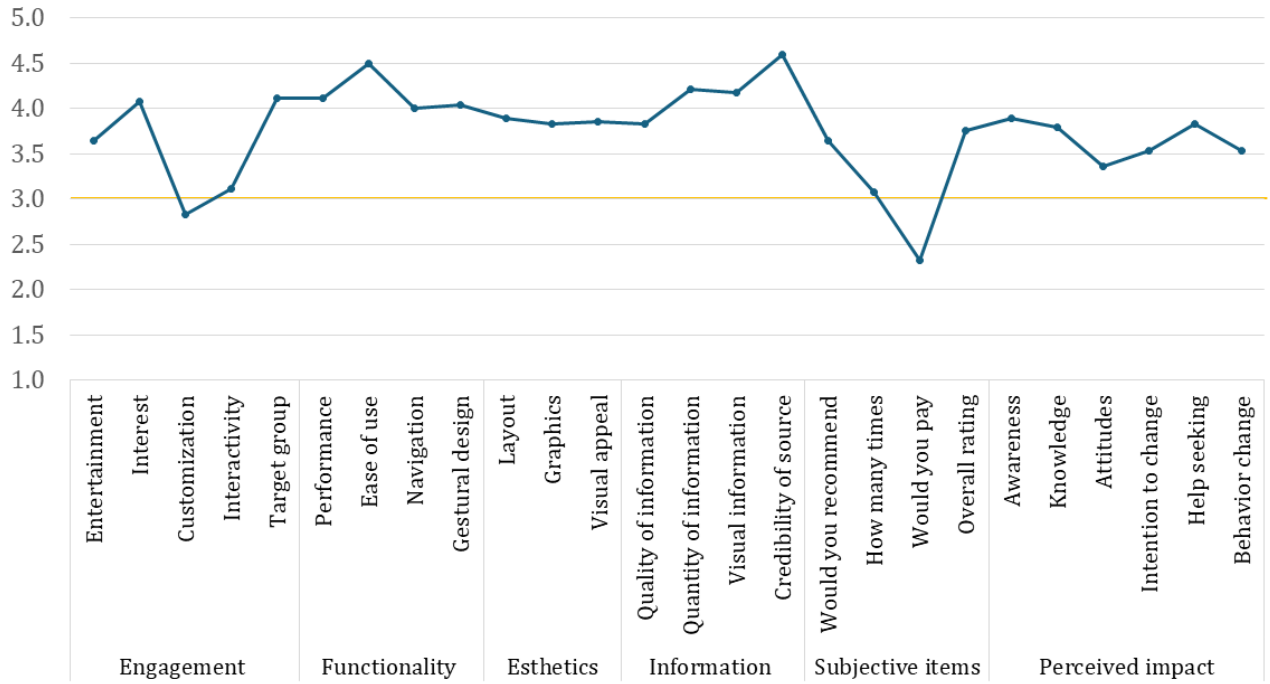
The results indicate consistently high ratings across all dimensions of the uMARS. Significant positive responses were noted for *Engagement*, *Functionality*, *Esthetics*, *Information*, and *Perceived Impact*. Effect sizes were uniformly high, particularly for the Wilcoxon tests, indicating a strong positive skew in user perceptions for each dimension evaluated. However, items such as "Customization" and "Interactivity" did not show significant positive trends in the *Engagement* scale, with "Customization" even indicating a negative effect size, suggesting variability or mixed responses from users.

Considering the *Subjective Items* scale singularly (which did not show a significant trend), the items "Would you recommend" (median 4.00; SD 0.62; *P*<.001; effect size=1.00) and "Overall rating" (median 4.00; SD 0.52; *P*<.001; effect size=1.00) received a strong positive response. This suggests that most users would recommend the app to others, with a robust positive effect size indicating widespread satisfaction with the app's performance, utility, and perceived value. However, for the item "How many times" (median 3.00; SD 0.77; *P*=.64; effect size=0.13), responses were mixed regarding the frequency of app use, with no significant trend emerging. This indicates that users were likely to use the app on average 3 to 10 times in the following year. Moreover, for the item "Would you pay" (median 2.00; SD 0.72; *P*<.001; effect size=−1.00), there was a significant negative response, indicating that users were generally unwilling to pay for the app. The strong negative effect size underscores a consistent reluctance to incur costs for app use.

Regarding the relevant *Perceived Impact* scale of the uMARS, items such as "Awareness" (median 4.00; SD 0.74; *P*<.001; effect size=0.24), "Knowledge" (median 4.00; SD 0.79; *P*<.001; effect size=0.26), "Attitudes" (median 3.50; SD 0.73; *P*=.02; effect size=0.26), "Intention to Change" (median 4.00; SD 0.69; *P*=.001; effect size=0.26), and "Help Seeking" (median 4.00; SD 1.09; *P*=.003; effect size=0.23) were rated positively. This indicates that the app positively impacted users, although the effect sizes were lower compared with those of other items, suggesting a moderate consensus and some variability in responses. Conversely, the "Behavior Change" item result (median 4.00; SD 0.64; *P*<.001; effect size=0.88) highlights that the app had a significant positive impact on users, with a high effect size indicating strong agreement among users on the app's effectiveness in promoting behavioral modifications.

The mean and statistical significance for every item are reported in Multimedia Appendix 4. Figure 7 shows the graphical distribution of item scores.

Figure 7. Graphical distribution of User Version of the Mobile Application Rating Scale item scores.



Qualitative Results: Semistructured Interviews

One psychologist conducted the qualitative interview to gather additional information. Interviews were conducted and analyzed according to the identified thematic variables.

Set 1: Interaction

The interaction with the chatbot was generally well received. The clinicians appreciated the good overall interaction, pointing out the effectiveness of the videos and images as well as the presence of multiple response options. However, they found it difficult to go back to previous answers, noting repetitiveness and lack of novelty in the content. The communication experts praised the fluidity of the interaction and particularly appreciated the personality of the app but criticized the excessively fragmented design of the messages.

The SH+ expert highlighted the confidence given by the chatbot, which calls the user by name and provides relevant answers, but it was not clear when user responses could be expressed in free text (chatting and not pressing the default buttons) and expressed a preference for the human figure in the videos. The usability experts appreciated the guidance and psychoeducational support offered by the chatbot but pointed out the 1-way interaction and the excessive number of messages in a row.

Among the features preferred by those who tested the app were videos with exercises, the personality of the app, heterogeneous content, and images with goals and values. However, among the elements to be improved were some aspects of the videos. In particular, the psychologists pointed out that professional voices could greatly improve the effectiveness of the videos, which were nevertheless appreciated for their content. The SH+ expert, on the other hand, found the videos and audio a little too slow and monotonous but greatly appreciated the chatbot's confidence and relevant answers. Finally, the usability experts found the videos and graphics useful but criticized the excessive length of the messages.

There was some discord regarding the possibility of correcting oneself and going back. In fact, the SH+ expert and psychologists considered it very useful and suggested that it should be included, whereas the communication experts did not find it particularly useful to go back. The usability experts did not find frequent errors, and the psychologists considered the consequences of errors to be not serious. On the other hand, the target group appreciated the practical examples provided during the sessions, which were considered important and highly functional to the objectives of the intervention. The variety of response options was also positively evaluated. However, it was suggested to reduce the length of the videos and messages as shorter and more focused content was deemed to improve the overall experience.

Set 2: Communication Mode

Concerning clarity and thoroughness, all categories of experts appreciated the possibility to better investigate concepts through examples and videos. In particular, the communication experts emphasized the clarity of the content, although they suggested using more concise answers. The same suggestion was made by the SH+ expert, who found some dialogues too complex and articulate. The clinicians disagreed, instead pointing out the simple and accessible language.

The usability experts also noted that the repetition of information was positive, especially for weekly use, but found the audio and video a little too slow and monotonous. Similarly, the psychologists appreciated the repetition of videos to better explain concepts and suggested alternating text with images or videos to lighten the information load.

The language used by the chatbot was generally well received. The clinicians appreciated the simple language and short sentences, whereas the communication and usability experts praised the clear, appropriate, and friendly language. The psychologists and SH+ expert also found the language appropriate to the content and easy to understand. However, the SH+ expert considered the tone to be a little slow. It was suggested to keep the language simple but to consider more complexity and variety in the content using a more active and engaging tone.

Finally, the target group generally appreciated the appropriateness of the sentence length and the terminology used as the language was perceived as accessible and aligned with the needs of future users. In addition, the inclusion of emojis was seen as effective and realistic in conveying emotions. Users found these visual elements to be a helpful and engaging way to express their feelings. All categories of experts appreciated and found useful the visual mode of emojis, which was also considered effective by the usability experts. However, it was suggested to offer both emojis and words as a response option to accommodate different preferences. However, one area for improvement pointed out by the target group was the length of the written messages. Some users reported that, sometimes, the messages were too long and could become heavy to read.

Set 3: Involvement and Consistency

User involvement and consistency were strengths for some groups, whereas others experienced difficulties. The clinicians found the tone positive and encouraging and the communication fast. The SH+ expert found the communication engaging, whereas the psychologists found good engagement in the videos. However, the clinicians found it difficult to feel engaged without a physical person, whereas the usability experts found the interaction similar to reading a book, needing more interactivity and exchange. For these reasons, it was suggested that the level of personalization and interaction be improved to increase engagement.

The customization of the chatbot was considered good but with room for improvement. The clinicians appreciated the personalization through the use of the user's name and the space for personal choices but noted standard answers for all and the lack of specific personalization elements. The SH+ expert gave a score of 8/10 to personalization, whereas the psychologists noted that the answers were standard. It was suggested to offer the possibility to add photos and avatars and improve the customization of answers. The target group found the content engaging and motivating to continue, with good personalization using names. However, they noted that it became less engaging in the long term, with repetition feeling lengthy and tedious and the exercises sometimes being monotonous.

Set 4: General Questions

The communication experts noted that the timing should be personalized according to individual circumstances, emphasizing that some users might benefit more from reminders and progress tracking. Indeed, target users indicated that the chatbot could be helpful during times of high stress and change, such as during chemotherapy or radiation and after childbirth, but less useful immediately after diagnosis or in early pregnancy. In contrast, the clinicians suggested that the ideal time for chatbot use in oncology is during and after chemotherapy or radiation treatment, whereas during pregnancy, it is in the first and second trimesters. In addition, the SH+ expert expressed that the chatbot was generally useful but highlighted the need for it to be adaptable across different stages of treatment and pregnancy. The psychologists found consistency with the ACT model and good care in the messages but reported that negative responses were not always considered.

Set 5: Technical Implementation

Reminders and notifications were considered useful for maintaining consistency. However, it was emphasized that reminders should not be too intrusive. The communication experts emphasized the importance of sending reminders in a nondisruptive way, whereas the SH+ expert stressed the need for reminders to maintain consistency. The usability experts agreed with the clinicians but suggested integrating reminders and immediate feedback to enhance user engagement. In particular, the target group appreciated the overall coherence of the content with the topic addressed and the usefulness of reminders to maintain consistency in use. However, they also reported usability issues, especially when too many messages were received in sequence, which made them difficult to manage and slowed down interaction with the chatbot.

Concerning gamification aspects, the preference for gradual coloring of the stickers was expressed by most groups. The clinicians suggested that gradual coloring gives the idea of progress, whereas the communication experts preferred immediate gratification. The SH+ expert indicated a preference for receiving the sticker at the end, whereas the usability experts suggested giving the reward immediately to maintain engagement. The target group preferred a gradual color progression of the stickers as they completed the exercises. Indeed, this approach helps visualize progress, providing them satisfaction and motivation. At the same time, a lot of users also expressed the desire for immediate rewards and feedback to maintain engagement. Indeed, the use of pop-up positive feedback provides instant gratification and motivation.

Set 6: Content Adherence

The SH+ expert reported that, on a scale from 1 to 5, our intervention adhered at a level of 4 to the original intervention. In terms of content, the app takes them up and is coherent, yet it is very innovative in terms of the ways it is delivered.

Table 4 briefly highlights key positive and negative aspects derived from the user interviews.

Table 4. Main positive and negative aspects that emerged from the user interviews.

Variable and subvariable	Positive aspects	Negative aspects (and suggestions)
Set 1: interaction		
General+alternatives for answers	Positive interaction (n=5)[a]; multiple options for answers (n=10)	Limited alternatives in some cases (n=5); fatigue in going back to previous answers (n=2)
Best and worst features	Confident and personalized chatbot; interactive videos and images (n=9)	Long and sometimes monotonous videos (n=4)
Going back for mistakes	—[b]	Need for option to correct mistakes (n=13)
Set 2: communication mode		
Clarity	Clear language (n=10); suitable for all users (n=2)	Sometimes overly simplistic and repetitive (n=1)
Emojis	Visual and easy to express emotions (n=8)	Offer both emojis and words for responses to cater to different preferences (n=6)
Length of and terms used in messages	Terms appropriate to the content (n=13)	Messages can be too lengthy (n=5)
Set 3: involvement and consistency		
Engagement	Positive and encouraging tone (n=1)	Improve personalization (n=7)
Personalization	Use of user's name (n=4); space for personal input (n=3)	Lacks deeper personalization (n=6)
Set 4: general questions		
Concerns and criticism	Reflective questions after exercises (n=1)	Even if the user can read the message, they can still decide not to proceed (n=1)
Ideal time	Oncology: during (n=11) and after (n=8) treatment; pregnancy: second trimester (n=10) and after (n=12) childbirth	Not ideal immediately after diagnosis (n=5) or in early pregnancy (n=4)
Set 5: technical implementation		
Reminders	Helpful for maintaining consistency (n=22)	Ensure that reminders are supportive and not overwhelming (n=5)
Pop-up positive feedback	Provides instant gratification and motivation (n=17)	Offer immediate feedback but consider a weekly summary for sustained engagement (n=1)
Progress in stickers	Gradual coloring indicates progress and provides satisfaction (n=19)	—
Set 6: content adherence		
Adherence to SH+[c] protocol	Generally aligns well with SH+ protocol (n=1)	Repetitive content (n=3)

[a]Numbers in parentheses indicate the response frequency.

[b]No statements.

[c]SH+: Self-Help Plus.

Discussion

Principal Findings

This study evaluated the adaptation of the WHO SH+ intervention for stress management. The first step involved developing the SH+ protocol, which was implemented through a mobile app with the support of the interactive ALBA chatbot. ALBA guides users through the 5-week program, corresponding to the 5 SH+ sessions, and facilitates navigation through various app sections. Notably, this adaptation introduced several innovations: the intervention was designed to further reinforce interaction and feedback to foster empowerment and self-efficacy, and it was tailored to female users, explicitly targeting 2 populations of interest in our study—pregnant women and women diagnosed with breast cancer.

The ORBIT and the CeHRes comprehensive road map approaches were adopted to evaluate the app, starting with a preliminary phase focused on refining the dialogues and low-fidelity mock-ups of the app's sections. This initial phase was crucial for developing coherent, accurate, and engaging dialogues and ensuring a reliable adaptation of the original SH+ content. Given that the protocol had already been validated, the chatbot's structured and standardized dialogues allowed for effective transmission of the proven content without the risk of artificial hallucinations, undertaken risks, or biases, which can occur with more advanced chatbot models based on large language models [48]. Giving value to the methodology adopted, regarding content adherence, the SH+ expert rated the intervention's adherence to the original at 4 out of 5, appreciating the innovative delivery methods while the core content was maintained. However, an explicit limitation of our chatbot emerged—the lack of flexibility in personalizing responses and interactions, as highlighted by participants in phase 2. Indeed, it is worth noting that psychologists found consistency with the ACT model but noted that negative responses were not always considered by ALBA. This rigidity presents a double-edged sword—while it ensures the psychological rigor of the initial intervention, it also reduces the flexibility of personalized response options [49].

Additional relevant findings from the low-fidelity prototype review pertain to the app's organization, featuring a well-defined structure in sections evaluated during the second phase regarding usability. The integration of gamification and feedback aspects based on user progress monitoring to maintain engagement and immediate or delayed reinforcement from classic behaviorism to sustain motivation was also significant. The app provided customization in session management by proposing interruptions and clarification moments for the presented content, which users could accept or decline.

From the second phase of this study, further essential results emerged that will guide the iterative development of the app. Valuable insights were gathered by involving various stakeholder groups identified from both the pregnancy and oncology contexts. Both expert groups and target users, the final app users, provided quantitative feedback through the semantic differential tool and uMARS questionnaires and qualitative feedback through semistructured interviews. The evaluation of dialogues and mock-ups, following modifications from phase 1, confirmed a generally positive assessment of the app and the ALBA chatbot.

In particular, the semantic differential tool results indicated that ALBA's communication was empathetic and fluid, and the interaction with users was deemed appropriate and acceptable. In addition, the interview reports highlighted that the interaction with the chatbot was generally well received, although criticisms included difficulties in navigating back to previous answers, repetitiveness, lack of novelty, and fragmented message design. Specific feedback highlighted the chatbot's confidence, relevant responses, and psychoeducational support while suggesting improvements such as professional voices for videos and reducing video and message length. Overall, both the target group and experts acknowledged the strengths of the chatbot's interaction, particularly its engaging multimedia elements and responsive nature. However, their perspectives diverged significantly on several key aspects. The target group expressed a desire for more innovative content and a reduction in repetitiveness, emphasizing that fresh and varied interactions would enhance their experience. In contrast, the experts, particularly the communication and usability specialists, focused on the fragmented design of the messages and the overwhelming volume of content, suggesting that a more cohesive and streamlined message structure would improve interactive usability. Indeed, user satisfaction significantly improves when chatbots provide quick, relevant, and friendly responses [50]. This capability reduces wait times and enhances the perception of service efficiency. Moreover, positive interactions with chatbots not only increase short-term satisfaction but also contribute to long-term user loyalty as users with positive experiences are more likely to return and use the service again, thereby strengthening their relationship with the app [51].

Regarding the communication modality, from the semantic differential tool emerged that the language was clear and understandable. Experts valued the clarity and thoroughness of using examples and videos to investigate concepts, suggesting more concise answers and alternating text with images or videos. Thus, multimedia content, such as images, videos, and audio provided by the chatbot, was also positively recognized both in the semantic differential tool and interview outputs. While the chatbot's language was also praised for simplicity and clarity in the interviews, some found the tone slow and recommended a more active tone, and the use of emojis was well received, with a recommendation to offer both emoji and word response options; the target group found the terms satisfactory but suggested reducing message length. In comparing the perspectives of experts and the target group on communication mode, notable differences emerged. The experts emphasized the importance of clarity and the effective use of various communication modalities, highlighting the need for dynamic and engaging delivery. In contrast, the target group appreciated the overall clarity and accessibility of the language used but expressed concerns about the length of certain messages. While the experts focused on enhancing different communication modalities, the target group prioritized brevity and engaging content. This divergence in focus highlights specific areas for improvement, indicating that refining the chatbot's communication strategies could enhance user satisfaction and overall effectiveness.

Overall, the session structure was considered appropriately lengthy and moderately light in content according to the semantic differential tool results. The importance of using inclusive and comprehensible language in chatbots is increasingly recognized in academic literature, emphasizing how this can enhance user experience and engagement. Studies show that chatbots that use such language can significantly improve user satisfaction and accessibility, making interactions more effective and welcoming for a diverse audience [52,53].

Regarding the uMARS standardized questionnaire, equally encouraging results were obtained for the engagement, functionality, esthetics, and information variables. However, improvements are needed for customization and interactivity, which could have received a more clearly positive rating. In accordance with this, in the interviews, the usability experts appreciated the guidance and psychoeducational support from the chatbot but noted a 1-way interaction and too many consecutive messages. Indeed, the experts, including the clinicians, SH+ specialist, and psychologists, noted in the interviews the difficulties related to the lack of physical presence and the static nature of the interaction, which felt akin to reading a book. They emphasized the need for enhanced personalization and interaction to foster deeper engagement, with suggestions to include features such as photos and avatars. Additional insights on engagement from interview responses were mixed. Although personalization through the use of the user's name was appreciated by the target group, there was a need for more specific customization—while the content was initially engaging and motivating for the target group, it became less engaging over time, leading to suggestions for reducing repetition and monotonous exercises. This divergence highlights the experts' focus on improving interaction dynamics and customization, whereas the target group prioritized sustained engagement and the need for variety to maintain interest over time.

Enhancements are expected in these areas by better integrating the *Diary* and *Exercise* sections with the chatbot's messages regarding weekly exercise management once the app is fully implemented. In addition, features such as reminders and feedback, which were not available to testers, are expected to improve the perception of app personalization. The literature emphasizes the effectiveness of engagement strategies and reminders in improving user interaction with chatbots. Recent research shows that chatbots that use personalized engagement techniques and timely reminders can significantly enhance user commitment over time and adherence to recommended actions, leading to better outcomes and increased user satisfaction [54].

Additional information from the interviews provides insights about future technical implementation. Both the experts and the target group recognized the utility of reminders and notifications but mentioned that they should not be intrusive, with preferences for nondisruptive reminders and maintaining consistency. Moreover, most groups preferred a gradual coloring of stickers to indicate progress, providing satisfaction and motivation, although the target group also expressed a desire for immediate rewards, indicating a nuanced understanding of how to balance engagement and consistency. The app's ease of use and the high perceived credibility of the source were 2 very positive aspects highlighted. Both results will be further investigated in future feasibility studies using standardized tools to measure app usability [55] and user trustworthiness [56].

A good overall rating emerged from the uMARS items regarding subjective impact and perception, with behavior-change and help-seeking initiatives aligning with the intervention principles. These results are promising for the app's effective use but should be considered with the potential bias from psychological experts who favor such interventions. In addition, participants indicated a reluctance to pay for the app hypothetically and suggested that they would use it 3 to 10 times per year. While the first point may seem moderate, the app will be provided for free by the health care system, and some willingness to pay adds value. The second point warrants further investigation as the app is designed for continuous use over 5 weeks and it is unclear whether further use throughout the year implies single accesses or restarting of the intervention. The interviews indicated that the app could be proposed to target women at various stages of breast cancer care or before or after childbirth in other contexts of interest. The clinicians suggested using the chatbot during and after chemotherapy or radiation for oncology patients and during the first and second trimesters for pregnant women. Similar ideas emerged from the target group. They indicated that the chatbot could be helpful during chemotherapy or radiation and after childbirth and, in contrast, less useful immediately after diagnosis or in early pregnancy. This flexibility is due to the different triggers and timings of stress-related issues in both contexts [57,58].

Strengths, Limitations, and Future Directions

This study highlights several strengths, limitations, and future directions for the app. Among the strengths, the chatbot's structured and standardized dialogues ensured an effective delivery of the validated content, and the empathetic dialogues and integration of gamification and feedback mechanisms were positively received by participants. In addition, the app's ease of use and high perceived credibility of the contents coming from a validated WHO protocol were crucial for user adoption. However, notable limitations include the chatbot's rigidity in personalizing responses and interactions and some repetitive content.

There was also a clear need for more specific customization options and improving answer personalization. While initially engaging, the content became less motivating over time, prompting suggestions to reduce repetition and monotonous exercises.

Future directions involve enhancing the chatbot's flexibility and personalization by incorporating user-specific elements. Implementing reminders and feedback mechanisms is also expected to improve personalization perception. Furthermore, conducting future feasibility studies using standardized tools to measure app usability and user trustworthiness will be essential. Exploring the frequency and context of app use over a year will help better understand user needs and improve continuous engagement effectiveness in pregnancy and oncological contexts.

Conclusions

This research evaluated the adaptation of the SH+ intervention for stress management through a mobile app guided by the ALBA chatbot. The implementation of the protocol tailored for pregnant women and women with a breast cancer diagnosis showcased several innovations, including interactive elements, gamification, and personalized feedback mechanisms. Using the ORBIT and CeHRes methodologies, this study validated the structured dialogues of the chatbot for effective content transmission while acknowledging limitations such as rigidity in personalization. Despite these challenges, the app's organization, user-friendly interface, and perceived credibility were notable strengths identified through participant feedback. Moving forward, addressing customization shortcomings, enhancing engagement strategies, and conducting further usability studies will be critical to refining the app's effectiveness and user satisfaction across diverse health care contexts.

Acknowledgments

The authors extend their gratitude to all study participants who dedicated their time, shared their knowledge, and contributed with their invaluable personal and professional experiences to the research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of variables, subvariables, and adjectives investigated using the semantic differential tool.
[DOCX File , 8 KB - humanfactors_v11i1e64614_app1.docx]

Multimedia Appendix 2
Questions posed to participants attending the interview evaluation.
[DOCX File , 14 KB - humanfactors_v11i1e64614_app2.docx]

Multimedia Appendix 3
Graphical representation of mean values derived from the semantic differential tool.
[PNG File , 93 KB - humanfactors_v11i1e64614_app3.png]

Multimedia Appendix 4
Wilcoxon signed ranked test analysis of the User Version of the Mobile Application Rating Scale items.
[DOCX File , 18 KB - humanfactors_v11i1e64614_app4.docx]

References

1. World mental health report: transforming mental health for all. World Health Organization (WHO). URL: https://www.who.int/publications/i/item/9789240049338 [accessed 2024-04-29]
2. Comprehensive mental health action plan 2013-2030. World Health Organization (WHO). URL: https://www.who.int/publications/i/item/9789240031029 [accessed 2024-04-29]
3. Cripps M, Scarbrough H. Making digital health “solutions” sustainable in healthcare systems: a practitioner perspective. Front Digit Health 2022;4:727421 [FREE Full text] [doi: 10.3389/fdgth.2022.727421] [Medline: 35434699]
4. Park SY, Nicksic Sigmon C, Boeldt D. A framework for the implementation of digital mental health interventions: the importance of feasibility and acceptability research. Cureus 2022 Sep;14(9):e29329 [FREE Full text] [doi: 10.7759/cureus.29329] [Medline: 36277565]
5. Mediavilla R, McGreevy KR, Felez-Nobrega M, Monistrol-Mula A, Bravo-Ortiz MF, Bayón C, et al. Effectiveness of a stepped-care programme of internet-based psychological interventions for healthcare workers with psychological distress: study protocol for the RESPOND healthcare workers randomised controlled trial. Digit Health 2022;8:20552076221129084 [FREE Full text] [doi: 10.1177/20552076221129084] [Medline: 36211795]
6. Chen C, Wang X, Xu H, Li Y. Effectiveness of digital psychological interventions in reducing perinatal depression: a systematic review of meta-analyses. Arch Womens Ment Health 2023 Aug;26(4):423-439 [FREE Full text] [doi: 10.1007/s00737-023-01327-y] [Medline: 37231116]
7. Scalable psychological interventions for people in communities affected by adversity: a new area of mental health and psychosocial work at WHO. World Health Organization (WHO). URL: https://iris.who.int/bitstream/handle/10665/254581/WHO-MSD-MER-17.1-eng.pdf [accessed 2024-04-29]
8. Guarino A, Polini C, Forte G, Favieri F, Boncompagni I, Casagrande M. The effectiveness of psychological treatments in women with breast cancer: a systematic review and meta-analysis. J Clin Med 2020 Jan 12;9(1):209 [FREE Full text] [doi: 10.3390/jcm9010209] [Medline: 31940942]
9. Fasano J, Shao T, Huang H, Kessler A, Kolodka O, Shapiro C. Optimism and coping: do they influence health outcomes in women with breast cancer? A systemic review and meta-analysis. Breast Cancer Res Treat 2020 Oct;183(3):495-501 [FREE Full text] [doi: 10.1007/s10549-020-05800-5] [Medline: 32691379]
10. Stewart DE, Cheung AM, Duff S, Wong F, McQuestion M, Cheng T, et al. Attributions of cause and recurrence in long-term breast cancer survivors. Psychooncology 2001;10(2):179-183 [FREE Full text] [doi: 10.1002/pon.497] [Medline: 11268144]
11. Goodman JH, Guarino A, Chenausky K, Klein L, Prager J, Petersen R, et al. CALM Pregnancy: results of a pilot study of mindfulness-based cognitive therapy for perinatal anxiety. Arch Womens Ment Health 2014 Oct;17(5):373-387 [FREE Full text] [doi: 10.1007/s00737-013-0402-7] [Medline: 24449191]
12. Woods SM, Melville JL, Guo Y, Fan MY, Gavin A. Psychosocial stress during pregnancy. Am J Obstet Gynecol 2010 Jan;202(1):61.e1-61.e7 [FREE Full text] [doi: 10.1016/j.ajog.2009.07.041] [Medline: 19766975]
13. Fonseca A, Gorayeb R, Canavarro MC. Women s help-seeking behaviours for depressive symptoms during the perinatal period: socio-demographic and clinical correlates and perceived barriers to seeking professional help. Midwifery 2015 Dec;31(12):1177-1185 [FREE Full text] [doi: 10.1016/j.midw.2015.09.002] [Medline: 26433622]

14. Li H, Bowen A, Bowen R, Muhajarine N, Balbuena L. Mood instability, depression, and anxiety in pregnancy and adverse neonatal outcomes. BMC Pregnancy Childbirth 2021 Aug 25;21(1):583 [FREE Full text] [doi: 10.1186/s12884-021-04021-y] [Medline: 34429072]
15. Fenwick J, Toohill J, Gamble J, Creedy DK, Buist A, Turkstra E, et al. Effects of a midwife psycho-education intervention to reduce childbirth fear on women's birth outcomes and postpartum psychological wellbeing. BMC Pregnancy Childbirth 2015 Oct 30;15:284 [FREE Full text] [doi: 10.1186/s12884-015-0721-y] [Medline: 26518597]
16. Gabrielli S, Mayora Ibarra O, Forti S. A holistic digital health framework to support health prevention strategies in the first 1000 days. JMIR. Preprint posted online December 23, 2023 2023 [FREE Full text] [doi: 10.2196/preprints.55235]
17. Zhang Q, Zhao H, Zheng Y. Effectiveness of mindfulness-based stress reduction (MBSR) on symptom variables and health-related quality of life in breast cancer patients-a systematic review and meta-analysis. Support Care Cancer 2019 Mar;27(3):771-781. [doi: 10.1007/s00520-018-4570-x] [Medline: 30488223]
18. Ghorbani V, Zanjani Z, Omidi A, Sarvizadeh M. Efficacy of acceptance and commitment therapy (ACT) on depression, pain acceptance, and psychological flexibility in married women with breast cancer: a pre- and post-test clinical trial. Trends Psychiatry Psychother 2021;43(2):126-133 [FREE Full text] [doi: 10.47626/2237-6089-2020-0022] [Medline: 34392667]
19. Dilworth S, Higgins I, Parker V, Kelly B, Turner J. Patient and health professional's perceived barriers to the delivery of psychosocial care to adults with cancer: a systematic review. Psychooncology 2014 Jun;23(6):601-612 [FREE Full text] [doi: 10.1002/pon.3474] [Medline: 24519814]
20. Solo il 17% delle donne con tumore al seno usufruisce del supporto psico-oncologico. Europa Donna. URL: https://europadonna.it/2021/04/13/risultati-fortemente-2-supporto-psicologico-tumore-seno/ [accessed 2021-04-13]
21. Epping-Jordan JE, Harris R, Brown FL, Carswell K, Foley C, García-Moreno C, et al. Self-Help Plus (SH+): a new WHO stress management package. World Psychiatry 2016 Oct;15(3):295-296 [FREE Full text] [doi: 10.1002/wps.20355] [Medline: 27717271]
22. Hayes SC, Luoma JB, Bond FW, Masuda A, Lillis J. Acceptance and commitment therapy: model, processes and outcomes. Behav Res Ther 2006 Jan;44(1):1-25. [doi: 10.1016/j.brat.2005.06.006] [Medline: 16300724]
23. Hayes SC, Strosahl KD, Wilson KG. Acceptance and Commitment Therapy: The Process and Practice of Mindful Change. 2nd edition. New York, NY: The Guilford Press; 2012.
24. Hayes SC, Hofmann SG. The third wave of cognitive behavioral therapy and the rise of process-based care. World Psychiatry 2017 Oct;16(3):245-246 [FREE Full text] [doi: 10.1002/wps.20442] [Medline: 28941087]
25. Hofmann SG, Asnaani A, Vonk IJ, Sawyer AT, Fang A. The efficacy of cognitive behavioral therapy: a review of meta-analyses. Cogn Ther Res 2012 Jul 31;36(5):427-440 [FREE Full text] [doi: 10.1007/S10608-012-9476-1]
26. Self-Help Plus (SH+): a group-based stress management course for adults. World Health Organization (WHO). URL: https://www.who.int/publications/i/item/9789240035119 [accessed 2024-04-29]
27. Turrini G, Purgato M, Tedeschi F, Acartürk C, Anttila M, Au T, et al. Long-term effectiveness of self-help plus in refugees and asylum seekers resettled in western Europe: 12-month outcomes of a randomised controlled trial. Epidemiol Psychiatr Sci 2022 Jun 08;31:12 [FREE Full text] [doi: 10.1017/s2045796022000269]
28. Riello M, Purgato M, Bove C, Tedeschi F, MacTaggart D, Barbui C, et al. Effectiveness of self-help plus (SH+) in reducing anxiety and post-traumatic symptomatology among care home workers during the COVID-19 pandemic: a randomized controlled trial. R Soc Open Sci 2021 Nov;8(11):210219 [FREE Full text] [doi: 10.1098/rsos.210219] [Medline: 34849238]
29. Karyotaki E, Sijbrandij M, Purgato M, Acarturk C, Lakin D, Bailey D, et al. Self-Help Plus for refugees and asylum seekers: an individual participant data meta-analysis. BMJ Ment Health 2023 Jul;26(1):e300672 [FREE Full text] [doi: 10.1136/bmjment-2023-300672] [Medline: 37524517]
30. Czajkowski SM, Powell LH, Adler N, Naar-King S, Reynolds KD, Hunter CM, et al. From ideas to efficacy: the ORBIT model for developing behavioral treatments for chronic diseases. Health Psychol 2015 Oct;34(10):971-982 [FREE Full text] [doi: 10.1037/hea0000161] [Medline: 25642841]
31. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 05;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
32. Abras C, Maloney-Krichmea D, Preece J. User-centered design. In: Bainbridge WS, editor. Berkshire Encyclopedia of Human-Computer Interaction. Great Barrington, MA: Berkshire Publishing Group; 2004:757-768.
33. de Vito Dabbs A, Myers BA, Mc Curry KR, Dunbar-Jacob J, Hawkins RP, Begey A, et al. User-centered design and interactive health technologies for patients. Comput Inform Nurs 2009;27(3):175-183 [FREE Full text] [doi: 10.1097/NCN.0b013e31819f7c7c] [Medline: 19411947]
34. Stickdorn M, Schneider J. This Is Service Design Thinking: Basics, Tools, Cases. New York, NY: Bis Publishers; 2010.
35. See A, Roller S, Douwe K, Weston J. What makes a good conversation? How controllable attributes affect human judgments. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019 Presented at: NAACL '19; June 2-7, 2019; Minneapolis, MN p. 1702-1723 URL: https://doi.org/10.18653/v1/n19-1170 [doi: 10.18653/v1/n19-1170]
36. Intuitive conversational chatbot builder. Landbot. URL: https://landbot.io/ [accessed 2024-04-29]
37. Osgood CE, Suci GJ, Tannenbaum PH. The Measurement of Meaning. Champaign, IL: University Of Illinois Press; 1957.

38. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the mobile application rating scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
39. Morgan DL. Focus groups. In: Michalos AC, editor. Encyclopedia of Quality of Life and Well-Being Research. Thousand Oaks, CA: Sage Publications; 2005.
40. Adlin T, Pruitt J. The Essential Persona Lifecycle: Your Guide to Building and Using Personas. Burlington, MA: Morgan Kaufmann; 2010.
41. Morselli S, Sebastianelli A, Domnich A, Bucchi C, Spatafora P, Liaci A, et al. Translation and validation of the Italian version of the user version of the Mobile Application Rating Scale (uMARS). J Prev Med Hyg 2021 Mar;62(1):E243-E248. [doi: 10.15167/2421-4248/jpmh2021.62.1.1894] [Medline: 34322643]
42. JASP (Version 0.18.2). JASP. URL: https://jasp-stats.org/previous-versions/ [accessed 2024-04-29]
43. Open source and professional software for data science teams. RStudio. URL: http://www.rstudio.com/ [accessed 2024-04-29]
44. Field AP, Miles J, Field Z. Discovering Statistics Using R. Thousand Oaks, CA: Sage Publications; 2012.
45. King AP, Eckersley RJ. Statistics for Biomedical Engineers and Scientists: How to Visualize and Analyze Data. Boston, MA: Academic Press; 2019.
46. Tomczak M, Tomczak E. The need to report effect size estimates revisited. An overview of some recommended measures of effect size. Trends Sport Sci 2014;21(1):19.
47. Terry G, Hayfield N, Clarke V, Braun V. Thematic analysis. In: Willig C, Rogers WS, editors. The SAGE Handbook of Qualitative Research in Psychology. Thousand Oaks, CA: Sage Publications; 2017:17-36.
48. Ke L, Tong S, Cheng P, Peng K. Exploring the frontiers of LLMs in psychological applications: a comprehensive review. arXiv. Preprint posted online January 3, 2024 2024 [FREE Full text]
49. Pryss R, Kraft R, Baumeister H, Winkler J, Probst T, Reichert M, et al. Using chatbots to support medical and psychological treatment procedures: challenges, opportunities, technologies, reference architecture. In: Baumeister H, Montag C, editors. Digital Phenotyping and Mobile Sensing: New Developments in Psychoinformatics. Cham, Switzerland: Springer; 2019:249-260.
50. Go E, Sundar SS. Humanizing chatbots: the effects of visual, identity and conversational cues on humanness perceptions. Comput Human Behav 2019 Aug;97:304-316 [FREE Full text] [doi: 10.1016/j.chb.2019.01.020]
51. Jenneboer L, Herrando C, Constantinides E. The impact of chatbots on customer loyalty: a systematic literature review. J Theor Appl Electron Commer Res 2022 Jan 21;17(1):212-229 [FREE Full text] [doi: 10.3390/jtaer17010011]
52. Hill J, Randolph Ford W, Farreras IG. Real conversations with artificial intelligence: a comparison between human–human online conversations and human–chatbot conversations. Comput Human Behav 2015 Aug;49:245-250 [FREE Full text] [doi: 10.1016/j.chb.2015.02.026]
53. Chaves AP, Egbert J, Hocking T, Doerry E, Gerosa M. Chatbots language design: the influence of language variation on user experience with tourist assistant chatbots. ACM Trans Comput Hum Interact 2022 Jan 16;29(2):1-38 [FREE Full text] [doi: 10.1145/3487193]
54. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied conversational agents in clinical psychology: a scoping review. J Med Internet Res 2017 May 09;19(5):e151 [FREE Full text] [doi: 10.2196/jmir.6553] [Medline: 28487267]
55. Lewis JR. The system usability scale: past, present, and future. Int J Hum Comput Interact 2018 Mar 30;34(7):577-590 [FREE Full text] [doi: 10.1080/10447318.2018.1455307]
56. Gulati S, Sousa S, Lamas D. Design, development and evaluation of a human-computer trust scale. Behav Inform Technol 2019 Aug 31;38(10):1004-1015 [FREE Full text] [doi: 10.1080/0144929x.2019.1656779]
57. Chaix B, Bibault JE, Pienkowski A, Delamon G, Guillemassé A, Nectoux P, et al. When chatbots meet patients: one-year prospective study of conversations between patients with breast cancer and a chatbot. JMIR Cancer 2019 May 02;5(1):e12856 [FREE Full text] [doi: 10.2196/12856] [Medline: 31045505]
58. Chua JY, Choolani M, Chee CY, Chan YH, Lalor JG, Chong YS, et al. Insights of parents and parents-to-be in using chatbots to improve their preconception, pregnancy, and postpartum health: a mixed studies review. J Midwifery Womens Health 2023;68(4):480-489. [doi: 10.1111/jmwh.13472] [Medline: 36734375]

Abbreviations

ACT: acceptance and commitment therapy
CeHRes: Center for eHealth Research and Disease Management
mHealth: mobile health
ORBIT: Obesity-Related Behavioral Intervention Trials
SH+: Self-Help Plus
uMARS: User Version of the Mobile Application Rating Scale
WHO: World Health Organization

Edited by A Kushniruk, E Borycki; submitted 22.07.24; peer-reviewed by P Delir Haghighi; comments to author 13.09.24; revised version received 24.09.24; accepted 30.09.24; published 18.10.24.

Please cite as:
Fietta V, Rizzi S, De Luca C, Gios L, Pavesi MC, Gabrielli S, Monaro M, Forti S
A Chatbot-Based Version of the World Health Organization–Validated Self-Help Plus Intervention for Stress Management: Co-Design and Usability Testing
JMIR Hum Factors 2024;11:e64614
URL: https://humanfactors.jmir.org/2024/1/e64614
doi:10.2196/64614
PMID:39355954

Original Paper

Expanding a Health Technology Solution to Address Therapist Challenges in Implementing Homework With Adult Clients: Mixed Methods Study

Brian E Bunnell[1], PhD; Kaitlyn R Schuler[1,2], PhD; Julia Ivanova[3], PhD; Lea Flynn[1]; Janelle F Barrera[1], MPH; Jasmine Niazi[1]; Dylan Turner[3]; Brandon M Welch[3,4], PhD

[1]Department of Psychiatry and Behavioral Neurosciences, Morsani College of Medicine, University of South Florida, Tampa, FL, United States
[2]Department of Psychology, University of North Carolina Wilmington, Wilmington, NC, United States
[3]Doxy.me Research, Doxy.me Inc, Rochester, NY, United States
[4]Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, United States

Corresponding Author:
Brian E Bunnell, PhD
Department of Psychiatry and Behavioral Neurosciences
Morsani College of Medicine
University of South Florida
3515 E Fletcher Ave
Tampa, FL, 33613-4706
United States
Phone: 1 8139748607
Email: bbunnell@usf.edu

Abstract

Background: Homework is implemented with variable effectiveness in real-world therapy settings, indicating a need for innovative solutions to homework challenges. We developed Adhere.ly, a user-friendly, Health Insurance Portability and Accountability Act–compliant web-based platform to help therapists implement homework with youth clients and their caregivers. The initial version had limited functionality, was designed for youth clients and their caregivers, and required expanding available features and exercises to suit adult clients.

Objective: The purpose of this study was to better understand barriers and potential solutions to homework implementation experienced by therapists seeing adult clients and obtain their input on new features and exercises that would enable Adhere.ly to better meet their needs when working with this population.

Methods: This study used an exploratory, sequential mixed methods design that included 13 semistructured focus groups with mental health therapists and clinic leaders and a survey administered to 100 therapists. Analyses were performed using the NVivo qualitative analysis software and SPSS.

Results: The findings revealed common barriers, such as clients and therapists being busy, forgetting to complete homework, managing multiple platforms and homework materials, and clients lacking motivation. Adhere.ly was perceived as a potential solution, particularly its user-friendly interface and SMS text-message based reminders. Therapists suggested integrating Adhere.ly with telemedicine and electronic health record platforms and adding more exercises to support manualized therapy protocols and therapy guides.

Conclusions: This study highlights the importance of technology-based solutions in addressing barriers to homework implementation in mental health treatment with adult clients. Adhere.ly shows promise in addressing these challenges and has the potential to improve therapy efficiency and homework completion rates. The input from therapists informed the development of Adhere.ly, guiding the expansion of features and exercises to better meet the needs of therapists working with adult clients.

(JMIR Hum Factors 2024;11:e56567) doi:10.2196/56567

KEYWORDS

mental health; mental illness; mental disease; mental disorder; homework; homework challenge; therapy; therapist; barriers; adult client; adult; technology-based solution; health technology; digital health; digital technology; digital intervention; mobile phone

Introduction

Background

Mental health disorders affect approximately 20% of individuals in the United States and are associated with costly physical, social, and occupational problems [1-3]. Most evidence-based mental health treatments (eg, cognitive behavioral therapy [CBT]) involve introducing and practicing therapeutic mental health exercises during therapy sessions, assigning patients to practice those therapeutic exercises between therapy sessions, and reviewing that practice during the next therapy session [4,5]. This between-session practice, or homework, facilitates treatment processes, strengthens learning, and enables the skills learned in therapy to generalize to the client's everyday environment, leading to improved maintenance of treatment gains [6-8].

Despite its many benefits, homework is implemented with variable effectiveness in real-world clinical settings. Only 68% of general mental health care providers and approximately 55% of family care providers report using homework "often" to "almost always" [9,10]. Therapists report using homework in an average of 57% of sessions, and only 25% of therapists report using expert-recommended systematic procedures for implementing homework (ie, specifying frequency, duration, and location and writing down homework assignments for patients) [11]. A national survey revealed that 93% of mental health therapists estimate rates of client adherence to homework to be low to moderate [10], and studies generally report low to moderate rates of client homework adherence [12,13].

Previous research has outlined numerous barriers to the successful implementation of homework during mental health treatment [6,7,9,14-21]. In our prior work, we found that youth therapists often have difficulty consistently and effectively designing, introducing, practicing, assigning, and assessing homework exercises, especially given the time constraints and workload of everyday clinical practice. We also found that youth and caregiver clients often have difficulty remembering to complete assignments and understanding the rationale and practical implementation (ie, when, where, and how) of homework exercises [14]. As part of this prior work, we also explored potential health technology solutions to these barriers. Therapists and youth and caregiver clients suggested digitized therapeutic exercises with instructions for clients, automated SMS text message–based reminders to complete those therapeutic exercises for homework, behavior and symptom tracking, reports, and activity summaries. Therapists and clients also expressed beliefs that a health technology solution with these features would likely increase therapist use of and client adherence to homework and have positive effects on the therapeutic relationship, treatment efficiency, and treatment effectiveness [14].

Informed by these results and using an iterative, user-centered design approach, we developed Adhere.ly, a user-friendly, Health Insurance Portability and Accountability Act–compliant web-based platform to help youth therapists implement homework with youth clients and their caregivers [22]. Upon initial development, the major components of Adhere.ly were designed to help youth therapists (1) *practice* digitized therapeutic mental health exercises (eg, relaxation, affect and emotion, cognitive, exposure, and parent-child activities) with their youth and caregiver clients in the session, (2) *remind* clients to practice homework between therapy sessions via automated SMS text message or email messages with links to those exercises to be opened and completed in the clients' smartphone or computer browser, and (3) *review* clients' homework completion and relevant data for certain exercises (eg, self-monitoring ratings and anxiety ratings during exposure exercises). Adhere.ly is accessed via computer, tablet, and smartphone browser and does not require any downloads by clients or therapists. Clients do not create accounts or need to log in to complete exercises assigned by therapists.

Objectives

The initial version of Adhere.ly included limited functionality and exercises designed solely for youth clients and their caregivers. Although therapeutic approaches with children and adults share many theoretical foundations, such as CBT, there are numerous developmental considerations that must be taken into account when providing therapy to children and adolescents [23,24]. Some of these include the focus and goals of the therapy (eg, individual vs systemic), the involvement of others (eg, spouses vs caregivers), and the way in which therapeutic concepts and techniques are communicated (eg, didactic vs gamified) [25]. As we sought to expand Adhere.ly to support therapists in treating adult mental health clients, we felt that it would be critical to obtain the input of these key stakeholders. The first step toward accomplishing this goal was to better understand barriers and potential solutions to homework implementation experienced by therapists seeing adult clients. Next, we sought to obtain their input on new features and exercises that would enable Adhere.ly to better meet their needs when working with this population.

Methods

Overall Design

We used an exploratory, sequential (qualitative then quantitative) mixed methods design to understand (qualitative) and prioritize (quantitative) (1) the challenges and goals of stakeholders relative to the use of health technology solutions such as Adhere.ly and (2) new features and exercises to support therapists delivering evidence-based therapy to adult clients. The qualitative component of this design included remote (ie, videoconference or telephone), ≤90-minute screen- and audio-recorded semistructured focus groups with mental health therapists and clinic leaders. The quantitative component included a survey among mental health therapists.

Ethical Considerations

This study was approved by the University of South Florida Institutional Review Board (STUDY002555). For the focus groups, we scheduled interested individuals and obtained individual verbal consent and demographic information before starting the group interview. For the survey, after following the link and completing a screener, participants clicked a button to indicate their electronic informed consent to participate. All focus group and survey data were deidentified before analysis. Focus group participants received a US $50 gift card for their time, and we compensated participants with a US $50 gift card for their time upon completion of the survey.

Focus Groups

Participants and Recruitment

We conducted a total of 13 focus groups with mental health therapists (n=18) and clinic leaders (n=16) between September 1, 2021, and October 25, 2021. A range of 2 to 6 mental health therapists and exactly 2 clinic leaders were included in each of the 8 clinic leader focus groups, and therapists and clinic leaders were interviewed separately. We did not meet our initial recruitment goal of 24 mental health therapists and 24 clinic leaders due to recruitment challenges during the COVID-19 pandemic.

All focus group participants were English-speaking adults (aged ≥18 years). We included therapists who had obtained at least a master's degree in social work, counseling, clinical psychology, or a related field and carried active adult mental health treatment caseloads. We included clinic leaders who served in a supervisory, managerial, or ownership role for an organization focused on providing mental health treatment services. Research staff emailed mental health therapists and clinic leaders registered with Doxy.me, a commercial telemedicine platform, with information about the study and an invitation to participate.

Data Collection and Instruments

In total, 2 master's-level and 1 bachelor's-level research staff members trained in social and behavioral sciences research facilitated the focus groups, and all 3 are authors on this manuscript. The focus group facilitators were university employees and not employed by Adhere.ly, LLC. Facilitators were trained with a detailed semistructured interview guide and protocol, including a series of role-play exercises and strategies to minimize potential bias. The focus group semistructured interview guide was developed following the Consolidated Framework for Implementation Research (CFIR) codebook [26] and the Integrated Technology Implementation Model (ITIM) [27]. All the interview questions are included in Multimedia Appendix 1. All focus groups were recorded and transcribed using the Dovetail transcription software (Dovetail Research Pty Ltd). Transcriptions were reviewed and edited to ensure accuracy while listening to the recordings.

The semistructured focus groups began with questions about challenges that therapists experience while trying to engage adult clients in homework during therapy and strategies that therapists have used to try to overcome those challenges. Next, participants viewed a 5-minute video tutorial on Adhere.ly and its current exercises and features. We then asked participants to share their perspectives on how Adhere.ly might address challenges to engaging clients in homework and for their input on changes that might be made to Adhere.ly to meet the needs of therapists treating adult clients, including the addition of new features and therapeutic exercises.

Data Analysis

The qualitative analysis was completed by a team of 4 research staff members, 3 of whom are authors on this manuscript. All research staff members had training in sociobehavioral sciences, including 2 with doctoral degrees in the social sciences (ie, clinical psychology and health and human performance), 1 with a master's degree in psychology, and another with a bachelor's degree in psychology. In total, 2 members of the analysis team also facilitated the focus groups. The researchers coded the focus group recordings using NVivo (version 11; QSR International), a qualitative analytical platform. Procedures followed the COREQ (Consolidated Criteria for Reporting Qualitative Research) [26]. The team read through the transcripts line by line to identify sensitizing categories reflective of the data. Researchers used a hybrid deductive-inductive process using thematic analysis [27,28]. The researchers coded emerging categories based on prior work that informed the focus group questions [14,29-31]. The unit of analysis was meaningful phrases [27,28].

Specifically, the team completed an initial round of deductive coding using categories identified through (1) prior work evaluating barriers to implementing homework through telemedicine [14,25] and (2) the CFIR and ITIM frameworks [26,27]. Categories related to homework challenges and requested exercises and features were created using evidence from previous studies on the platform [14,25], including but not limited to motivational and internet access–related issues and care provider attitudes toward homework. Categories following the CFIR and ITIM frameworks included but were not limited to the adaptability, compatibility, complexity, efficiencies, and relative advantages of Adhere.ly.

Results of this initial round were used to develop survey questions that asked respondents to priority rank the therapist- and client-related barriers to homework and the requested features and exercises identified in the first round of coding.

During the second round of coding, researchers refined the codebook over 4 iterations of inductive coding to identify categories, and disagreements were resolved through consensus with the research team. The research team coded a third of the focus groups collaboratively until the Cohen κ was at least 0.61, indicating substantial agreement, and the remaining focus groups were coded independently. Focused coding was used to refine the coding and ensure that the data were coded completely with minimal redundancy [32]. A full list of categories and the number of associated references are provided in Multimedia Appendix 1.

Survey

Participants and Recruitment

We administered the 15-minute, web-based survey among 100 English-speaking adult (aged ≥18 years) mental health therapists who had obtained at least a master's degree. Similar to the recruitment of focus group participants, research staff emailed mental health therapists registered with Doxy.me, including those who had participated in the focus groups, with information about the study, an invitation to participate, and a link to the survey screener. Survey responses were collected between October 21, 2021, and December 21, 2021.

Data Collection and Instruments

We developed a 59-question survey using insights obtained from the focus groups and the literature on homework and process-based CBT [33]. The survey began with screening questions about age, specialty (eg, mental health counselor vs social worker vs psychologist), and education (ie, master's degree vs doctoral degree vs neither), followed by questions about personal (ie, sex, race, and ethnicity) and professional (eg, ages and disorders treated, practice type, primary reimbursement, and treatment paradigms followed) demographics. The survey then asked therapists questions related to their use of homework with their adult clients. These included questions about how often they asked clients to complete homework exercises using a 4-point Likert scale (ie, 1=*never*; 4=*most of the time*), how important they thought homework was to improving their clients' outcomes on a 5-point scale (ie, 1=*not at all*; 5=*extremely*), and how important it was for them to know the results of their clients' homework assignments (eg, whether they completed them and how it went) on the same 5-point scale (ie, 1=*not at all*; 5=*extremely*). The survey also included questions about the extent to which certain factors were a barrier to them using homework with their clients (eg, not knowing what or how to assign homework, forgetting, and being too busy) and their clients completing homework assignments (eg, not knowing why or how to do the homework assignment or what to do, forgetting, being too busy, and avoiding homework due to distress) using a 4-point scale (ie, 1=*not a barrier*; 4=*significant barrier*). Therapists then viewed a 5-minute video tutorial on Adhere.ly and its current exercises and features. Afterward, the survey asked them to rank a list of exercises (eg, relaxation, interpersonal, self-monitoring, and behavioral activation) and a separate list of features (eg, assessments, integration with telemedicine platforms, integration with electronic health record [EHR] software, and therapy guides) in the order in which they should be prioritized when added to Adhere.ly to meet therapists' needs when seeing adult clients. All survey questions are included in Multimedia Appendix 2.

Data Analysis

We cleaned the survey data and used SPSS (version 28; IBM Corp) to calculate descriptive statistics. We averaged the rank order of the features and exercises and used mean rankings to determine relative priority.

Results

Focus Groups

Participants

The demographic characteristics of the focus group participants are shown in Table 1. On average, participants were aged 47.97 (SD 13.21) years and were largely female (25/34, 74%), White individuals (23/34, 68%), and non-Hispanic or Latino individuals (31/34, 91%). As shown in Table 1, demographic characteristics were similar for the therapist and clinic leader groups.

Table 1. Focus group participant demographics (N=34).

Demographics	Overall sample	Therapists (n=18)	Clinic leaders (n=16)
Age (y), mean (SD)	47.97 (13.21)	49.7 (15.80)	46.1 (9.60)
Sex (female), n (%)	25 (74)	12 (67)	13 (81)
Race, n (%)			
Asian	1 (3)	0 (0)	1 (6)
Black	8 (24)	5 (28)	3 (19)
Multiracial	2 (6)	2 (11)	0 (0)
White	23 (68)	11 (61)	12 (75)
Ethnicity (non-Hispanic or Latino), n (%)	31 (91)	16 (89)	15 (94)

The therapist and clinic leader focus groups were analyzed together, and these data were reported in a unified coding scheme comprising categories of responses. The number of therapists and clinic leaders that made statements relevant to each category is reported in Multimedia Appendix 1. In addition, illustrative quotes are provided for each category. It should be noted that only therapists and clinic leaders were participants, so references made to client-related barriers are therapist and clinic leader perceptions of those client barriers.

Challenges to Engaging Clients in Homework

Participants identified 7 major categories related to challenges in engaging clients in homework (44 total references). Overall, most of the focus group participants reported challenges related to therapists being too busy to complete, assign, or review homework (9/44, 20% of references) or perceiving that their clients were too busy to complete homework (8/44, 18% of references). However, the number of participants who made statements within both categories differed substantially between the clinic leader and therapist focus groups.

One participant commented the following:

> *We have not had much luck with getting anything returned to us, whether it's homework or releases of information...*

Another participant stated the following:

> *...as an individual provider or even in a clinic, you don't have the time to be constantly checking in every day with all the clients to get them to practice what you're telling them to practice.*

Most of the references associated with the category of therapists being too busy to assign homework and perceptions that clients were too busy to complete homework were from the clinic leader focus groups (7/9, 78% of references) as opposed to the therapist focus groups (2/9, 22% of references). Slightly more clinic leaders (5/8, 62% of references) than therapists (3/8, 38% of references) endorsed perceiving that clients forget to complete homework. While these were the 2 most common categories of responses in the clinic leader focus groups, neither was the most frequent in the therapist focus groups.

In addition, many focus group participants reported struggling to find the right platform or medium for homework (7/44, 16% of references). The number of relevant responses did not differ substantially between the therapist (3/7, 43% of references) and clinic leader (4/7, 51% of references) focus groups. Participants reported difficulty sending homework materials over telemedicine platforms as well as having to manage multiple web-based platforms and applications to meet their clients' needs.

One participant commented the following:

> *When I know that I'm going to be assigning homework assignments, I'll have the worksheet or paper in front of me so that I can give it to them. We use Doxy.me for our tele-health and it's a little hard, I can sometimes screen-share, but sometimes our tele-health is over the phone and I'm like, okay, just try and Google this on your own. Good luck. It makes it a little challenging.*

The next 2 most frequently reported challenges included perceptions that clients lack motivation or discipline to complete homework consistently (7/44, 16% of references) and perceptions that clients receive insufficient instruction or rationale for completing homework (6/44, 14% of references). A comparable number of therapists and clinic leaders made statements that fell under these 2 categories.

One participant commented the following:

> *I think when it comes to giving a client an assignment, it's always a challenge as to whether they're going to follow through with it and utilize it. And also, I mean, at least they'll try it once and whether they're going to try it again, will depend on if it worked or not, or if it was so empowering to them that they're willing to keep using it.*

Few participants mentioned challenges related to perceptions of clients avoiding homework due to anticipated distress (4/44, 9% of references) and clients experiencing external (ie, internet or phone access), educational, or cognitive functioning barriers to completing homework (3/44, 7% of references).

For example, one participant stated the following:

> *So, like you're in therapy and you're in the moment...and then in between sessions, you just sort of like, don't want to go there. So, you avoid.*

Another participant stated the following:

> *I would also add that we have some resource barriers here sometimes. So, I'm in a fairly rural part of the country. And as a result of that, the ability to obtain a consistent internet or phone signal sometimes is really challenging. So, there's some technological barriers in that regard.*

Related to educational or cognitive barriers to completing homework, another participant stated the following:

> *And then I also think the education and cognitive abilities of the person sometimes plays a role in making sure that the materials we're providing are appropriate for the person. So, if I try giving them really complicated interventions and I send them home with a sheet and their reading isn't strong enough to be able to comprehend what the sheet says that can also be a barrier in some situations.*

While the category related to client avoidance was evident in both the therapist (1/4, 25% of references) and clinic leader (3/4, 75% of references) focus groups, the category related to socioeconomic status, rural versus urban location, educational level, or cognitive ability challenges was evident only in the clinic leader focus groups (3/44, 7% of references).

Overall, most clinic leaders emphasized therapists' and clients' busy schedules and perceptions that clients forget to complete their homework. Most therapists emphasized perceptions that clients lack motivation or discipline to complete homework consistently; however, the number of participants whose responses fell within these categories was more evenly distributed across categories in the therapist focus groups than in the clinic leader focus groups.

Current Strategies for Overcoming Challenges

Participants' responses fell into 5 categories of strategies or responses that therapists used to overcome challenges in engaging clients in between-session exercises (57 total references). Most participants referenced using motivational and reinforcement strategies to encourage homework completion during the sessions (23/57, 40% of references); however, nearly all the participants who mentioned this strategy were clinic leaders (19/23, 83% of references) as opposed to therapists (4/23, 17% of references):

> *We have to really just plan out how important the assignments are to their treatment, and just for them to know that if they want to get better and achieve the goals that we've set in therapy, then the homework assignments are a very important part of that, and they need to take it seriously.*

Many participants also mentioned using multiple applications and web platforms to send and track homework (15/57, 26% of references). Similar to the previous category, most participants who mentioned this strategy were clinic leaders (13/15, 87% of references) rather than therapists (2/15, 13% of references):

> *...some of us have gotten creative. We realized that the paper homework doesn't work...or having the client record their thoughts on their phone or on a paper does not work. So, we get written permission from the client to share Google Drive, which is probably not the most HIPAA compliant activity to be doing. We have to go through the whole hassle of getting the client to sign that they're okay with this and that they're sharing that Google document. Knowing that the therapist is able to review it and see it in real time tends to have a little bit of an effect on compliance.*

The third most frequently referenced strategy was adapting the treatment plan to "meet clients where they are at," which largely entailed changing homework assignments or completing them in the session rather than continuing to give the same homework assignments that the clients had not previously completed (14/57, 25% of references). This category was equally referenced by clinic leaders (7/14, 50% of references) and therapists (7/14, 50% of references):

> *I don't penalize and you know, I don't shame or anything like that when people don't do their homework. However, like a good teacher, if you don't do your homework, you're doing it in class. And so, we just break out the pen and paper, I get silent, turn the lights down, if I need to, or the lights up, whatever is going to be conducive. But sometimes people, this is the one time they get a chance to really just be still and really think, cause they're on autopilot so much.*

The last 2 strategies were referenced by only a few focus group participants and included having clients set their own reminders either electronically or in written form (3/57, 5% of references), which only emerged in the clinic leader focus groups, and providing accountability via email, SMS text message, or phone (2/57, 4% of references), which was mentioned by only 1 participant in the clinic leader and therapist focus groups each. For example, one participant commented the following:

> *...we've definitely had clients who forget things that we've had them put it in their calendar in their phone.*

How Adhere.ly Might Address Challenges to Engaging Clients in Homework

Participants' responses to how Adhere.ly might address challenges to engaging clients in homework were divided into 11 categories (143 total references). Overall, the most common responses were that the utility of Adhere.ly would vary by client demographics and clinical needs (28/143, 19.6% of references), which was emphasized almost equally by therapists (18/28, 64% of references) and clinic leaders (10/28, 36% of references), and that Adhere.ly is easy to use and would appeal to a wide range of therapists and clients (27/143, 18.9% of references), which was emphasized by more therapists (12/27, 44% of references) than clinic leaders (15/27, 56% of references).

Regarding the utility of Adhere.ly varying by client demographics and clinical needs, participants mentioned demographics such as age, which they thought might influence attitudes toward technology, and smartphone use. Many participants also mentioned generally taking the approach of seeing how specific clients responded to new interventions before they drew conclusions about effectiveness and being unable to comment on Adhere.ly until they had piloted it with their clients. A smaller subset of these participants mentioned that Adhere.ly may not be helpful for specific clinical populations, including those in crisis or with serious substance abuse problems:

> *I know a lot of my older people...my older ladies, high anxiety, depression, they're going to have difficulty understanding it. I mean, it's something that I think could be very useful to a certain group of folks.*
>
> *I think this also speaks to a particular demographic. I see a fair percentage of Medicare patients, some of whom don't have smartphones, and some who are very smart about using their smartphone. So, you know, defaulting to that, it's enough of a challenge to get people, some of them, to do telemedicine. So, we'd have to look at whether this is a demographic that is comfortable doing things on the phone.*

Regarding ease of use, participants mentioned that Adhere.ly seemed more accessible and efficient than other applications and platforms, which was mentioned by nearly the same number of therapists (12/143, 8.4% of references) and clinic leaders (15/143, 10.5% of references):

> *I definitely agree that even compared to some other apps...the user interface seems a lot simpler. I kind of poked around before and created an account and it actually moves pretty quickly and doesn't lag and it's just easier to navigate.*

Many participants reported perceptions that Adhere.ly would increase the overall efficiency of therapy (23/143, 16.1% of references), which was mentioned by most participants across the therapist (11/23, 48% of references) and clinic leader (12/23, 52% of references) focus groups. Several specifically mentioned that Adhere.ly's reminder feature would improve homework completion rates (16/143, 11.2% of references), although this was primarily emphasized by clinic leaders (15/16, 94% of references) rather than therapists (1/16, 6% of references). One participant commented the following:

> *I think it reminds me of Fitbit. It's kind of the same concept of getting people to get up and move or exercise and, it's just, again, reinforcing a different habit.*

Another participant stated the following:

> *...being able to know that there's some real practical, useful tools and automated support reinforces the fact that folks are getting what they need in between sessions. Cause it's absolutely difficult to have to start over again week by week because you feel like you lost what happened between the sessions.*

Many participants also stated that the utility of Adhere.ly would vary by therapist demographics and theoretical orientation (12/143, 8.4% of references), although more therapists (18/143, 12.6% of references) emphasized this than clinic leaders (10/143, 7% of references). Some of these comments referenced the association between age and attitudes toward technology or attitudes toward technology alone, as illustrated by the following quotes:

> *I have some [therapists] that are really tech savvy. And so, this is going to seem like a breeze, and it would be pretty easy for them to incorporate it. And I have others that wish that we didn't have electronic health records and wish we could still do pencil and paper for everything. And so, I would imagine this would probably align with comfort level in part, but as long as it was fairly user-friendly, I would imagine probably most of my clinicians would at least try it out.*

> *I have staff under me who are in their fifties. I have staff in their thirties. So, I think it just varies. I can't speak to everybody's level of comfortability. That's the best answer I can give you.*

Therapists also mentioned that Adhere.ly might be especially useful when implementing evidence-based CBTs over telehealth but might be less useful if using other interventions or doing therapy in person:

> *It's not as much a need for in-person services as for integrated CBT interventions on doxy.me. That feels like a need. Like, to me, that's the missing piece. I think I mentioned this in our first meeting, even as a clinical supervisor, if I have a very beginning therapist and I'm trying to teach her how to teach the cognitive triangle or how to teach a thought record, and I'm doing it over zoom, which I did for so long.*

Some participants also mentioned that Adhere.ly would easily fit into therapists' in-session workflows (11/143, 7.7% of references), which was mentioned by slightly more clinic leaders (7/11, 64% of references) than therapists (4/11, 36% of references). Participants explained that therapists could easily set up Adhere.ly exercises with their clients at the end of the session, which would minimize the time they would need to spend doing homework outside the session:

> *Many times, we have patients back-to-back and so the ability to do it in between appointments isn't always there. I think if the provider saw the value in this, I would imagine they would incorporate it into the session on the back end as kind of a wrap up between that and the following session. I think if it was like an add-on that had to be done in between sessions, that could create some challenges only because of the learning curve and time commitments.*

Other less commonly referenced categories included Adhere.ly's ability to capture treatment data (9/143, 6.3% of references) and the use of text-based communication (6/143, 4.2% of references) coming from their therapist as opposed to an external platform (5/143, 3.5% of references) to increase client engagement and homework completion. For example, one participant commented the following:

> *I like the tracking. I like the little charts that show you how they're doing.*

Another participant said the following:

> *I like that it goes straight to their email or their texts without a log-on...if there's no login, I find that clients are much more likely to click on a link and then follow through.*

Primarily clinic leaders emphasized the benefits of capturing treatment data (8/9, 89% of references) and the use of text-based communication (5/6, 83% of references), whereas more therapists emphasized the benefits of clients receiving direct communication from their therapist through Adhere.ly (4/5, 80% of references).

Suggested Features to Add to Adhere.ly

When asked which features should be added to Adhere.ly (63 references in total), a total of 7 categories emerged, although most of the references included in this account were from the clinic leader focus groups (52/63, 83% of references). The most common response was to add a demonstration video or other web-based walk-through for clients and therapists (14/63, 22% of references), and 86% (12/14) of these references were from the clinic leader focus groups. One participant stated the following:

> *...make an electronic training that actually sends the links to the clinicians so that they see what it's like, and they're sort of exposed to it from the client side.*

This was followed by integration of Adhere.ly with EHR platforms and Doxy.me (12/63, 19% of references), which was only mentioned by clinic leaders. One participant stated the following:

> *I think the toughest part would be if people have to log into a separate platform...the more integrated it is into something you're already using or the easier it is to get in and out of, the more helpful it would be.*

Other suggestions included enhancement of audiovisual features (ie, gamification and more color and visuals; 11/63, 17% of references), which was primarily mentioned by clinic leaders (8/11, 73% of references) rather than therapists (3/11, 27% of references):

> *There's got to be some type of draw that keeps somebody continuing to engage with electronic interventions...like gamification. For example, can they achieve ribbons or trophies or can go toward something that shows them that they're making progress? I know at least with some of the programs that I utilize now, that's what I really like about it.*

Several participants also suggested expanding the reminder feature (ie, recurring notifications, customizations, and parent reminders; 10/63, 16% of references), which was equally mentioned by therapists (5/10, 50% of references) and clinic leaders (5/10, 50% of references). For example, a participant suggested adding "a way to set the reminder up for an open-ended period of time until it's changed, rather than every week having to get them there and change the reminder." Other suggestions mentioned by several participants included adding more options for registration (ie, self-registration and customized communication preferences, including more client identifiers and a feature for group, couple, and family registration; 8/63, 13% of references) and adding an e-consent form (5/63, 8% of references). One participant emphasized the following:

> *Share-ability! Where can I share the information? Is it with the client? Is it with their parent? Is it with a family member? Clinician share-ability I think goes a long way to making something useful.*

Both of these features were only mentioned by clinic leaders.

There were several features that did not fit neatly into the other categories, all of which were referenced by only 1 participant, such as options for clients to provide feedback on exercises, a feature allowing PDF export of client results, and ensuring sixth- and seventh-grade reading level across the content. These features were also only requested by clinic leaders.

Suggested Therapeutic Exercises to Add to Adhere.ly

When asked which therapeutic exercises should be added to Adhere.ly, participants responded within 7 major categories (81 references in total), most of which were identified by therapists (47/81, 58% of references) These categories included the following: homework from evidence-based therapy protocols (40/81, 49% of references), homework for specific skills or strategies (17/81, 21% of references), homework for specific client demographics (6/81, 7% of references), homework for specific mental health conditions (5/81, 6% of references), personalized or custom exercises (5/81, 6% of references), and brief or single-item assessments (3/81, 4% of references).

Regarding homework for evidence-based therapy protocols, most therapists and clinic leaders identified the following interventions as useful additions: mindfulness-based approaches (11/40, 28% of references), trauma-focused therapies (ie, eye movement desensitization and reprocessing, cognitive processing therapy, and prolonged exposure; 8/40, 20% of references), dialectical behavior therapy (5/40, 12% of references), behavioral activation (4/40, 10% of references), safety planning (3/40, 8% of references), CBT (3/40, 8% of references), and acceptance and commitment therapy (2/40, 5% of references). One participant explained why they emphasized trauma-focused interventions:

> *Much of my focus is trauma. And so many of the approaches I use include things like EMDR and somatic experiencing. So...if I was going to assign anything to anyone, it would be related to developing self-regulation skills and somatic awareness skills to help them be able to cope with trauma symptoms. So that would be one way that I can see providing homework to people is to do something like that.*

While mindfulness and stress reduction interventions were mentioned by an equal number of therapists (6/11, 55% of references) and clinic leaders (5/11, 45% of references), most of the remaining interventions were mentioned by more clinic leaders than therapists, although the overall number of suggested interventions was greater for therapists than clinic leaders.

With respect to homework for specific skills and strategies, therapists most frequently requested journaling (7/17, 41% of references), creative activities (3/17, 18% of references), recovery plans (2/17, 12% of references), and strength-based activities and positive psychology (2/17, 12% of references). These distributions can be viewed in greater detail in Multimedia Appendix 1.

One participant suggested the following:

> *I think those basic needs things like, are you sleeping? Are you eating? Have you had water today? For that kind of stuff, they need reminders because when they're really depressed, as we know, it's hard to remember those basic things.*

Regarding homework for specific client demographics (5/81, 6% of references), some therapists referenced emotion regulation games that are engaging for children (3/5, 60% of references) despite the prompt to suggest interventions for adult users, as well as couples exercises (2/5, 40% of references). For example, one participant suggested the following:

> *...for clinicians working with couples...there are lots of lots of things like scanning for opportunities to express appreciation and respect...prompt them to initiate a conversation to show appreciation.*

No clinic leaders suggested interventions for specific demographics or mental health conditions. The remaining categories, including homework for specific mental health conditions and personalized or custom exercises, did not have a pattern of minor categories and were only mentioned by a few therapists.

Survey

Participants

The demographic characteristics of the survey respondents are shown in Table 2. On average, therapists were aged 45.6 (SD 14.23) years and were largely female (69/100, 69%), White individuals (79/100, 79%), and non-Hispanic or Latino individuals (85/100, 85%). Therapists were mostly mental health counselors (39/100, 39%), psychologists (28/100, 28%), and social workers (22/100, 22%) who worked in individual practice (55/100, 55%) and small clinic (33/100, 33%) settings and were primarily reimbursed via private (53/100, 53%) or public (32/100, 32%) insurance. Most therapists primarily saw adults (93/100, 93%) and adolescents (51/100, 51%) with anxiety (99/100, 99%), trauma- and stressor-related (90/100, 90%), and mood (85/100, 85%) disorders following the cognitive behavioral paradigm (69/100, 69%).

Table 2. Demographic and professional characteristics of the survey respondents (N=100).

Demographics	Values
Age (y), mean (SD)	45.64 (14.23)
Sex (female), n (%)	69 (69)
Race, n (%)	
American Indian or Alaska Native	1 (1)
Asian	3 (3)
Black	13 (13)
Multiracial	11 (11)
White	79 (79)
Ethnicity (non-Hispanic or Latino), n (%)	85 (85)
Highest degree, n (%)	
Master's	67 (67)
Doctoral	33 (33)
Specialty, n (%)	
Mental health counselor (eg, LMHC[a] or LPC[b])	39 (39)
Psychologist (eg, PhD or PsyD)	28 (28)
Social worker (eg, LCSW[c])	22 (22)
Marriage and family therapist (eg, LMFT[d])	9 (9)
Other mental health therapist	1 (1)
Type of clinic or organization, n (%)	
Individual practice	55 (55)
Network of health care providers or small clinic	33 (33)
Hospital or large clinic	10 (10)
Educational setting	2 (2)
Primary method of reimbursement, n (%)	
Private insurance	53 (53)
Public insurance (Medicare or Medicaid)	32 (32)
Out of pocket by the client	15 (15)
Age groups treated, n (%)	
Adults (aged 18-64 y)	93 (93)
Adolescents (aged 11-17 y)	51 (51)
Older adults (aged ≥65 y)	32 (32)
Children (aged 0-10 y)	26 (26)
Mental health disorders treated, n (%)	
Anxiety disorders	99 (99)
Trauma- and stressor-related disorders	90 (90)
Mood disorders	85 (85)
Disruptive, impulse control, and conduct disorders	44 (44)
Personality disorders	42 (42)
Somatic symptom and related disorders	33 (33)
Substance-related and addictive disorders	23 (23)
Primary treatment paradigm, n (%)	

Demographics	Values
Cognitive behavioral	69 (69)
Interpersonal	10 (10)
Existential or humanistic	7 (7)
Family systems	6 (6)
Psychodynamic or analytic	8 (8)

[a]LMHC: licensed mental health counselor.
[b]LPC: licensed professional counselor.
[c]LCSW: licensed clinical social worker.
[d]LMFT: licensed marriage and family therapist.

Perceptions on Homework

Most therapists (94/100, 94%) reported that they generally asked their clients to practice therapeutic skills and exercises for homework *some* to *most of the time* (mean ranking 3.51, SD 0.61). Most therapists (73/100, 73%) felt that homework is *moderately* to *extremely* important to improving their clients' outcomes (mean ranking 4.08, SD 0.93). Most therapists (69/100, 69%) also believed that it is *moderately* to *extremely* important for them to know the results of their clients' homework assignments (eg, whether they completed them and how it went).

Barriers to Therapists Using Homework With Clients

Therapists perceived having difficulty getting clients to complete assignments (mean ranking 2.66, SD 0.79), not wanting to overwhelm or distress clients (mean ranking 2.05, SD 0.93), and being too busy or not having time (mean ranking 2.01, SD 0.90) as *minor* to *moderate* barriers to them using homework with their clients. Therapists perceived forgetting to assign homework (mean ranking 1.84, SD 0.90), not knowing what to assign (mean ranking 1.72, SD 0.71), not knowing how to address clients not completing assignments (mean ranking 1.69, SD 0.84), not knowing how to assign homework (mean ranking 1.34, SD 0.59), and not being trained to assign homework (mean ranking 1.29, SD 0.59) as *minor* barriers to nonbarriers to them using homework with their clients.

Barriers to Clients Completing Homework Assignments

Therapists perceived clients having a busy or chaotic home life as a *moderate* barrier to clients completing homework assignments (mean ranking 3.00, SD 0.80). Therapists also perceived clients forgetting about homework (mean ranking 2.86, SD 0.94), avoiding completing assignments due to distress or symptoms (mean ranking 2.66, SD 0.71), receiving little reward or reinforcement for completing assignments (mean ranking 2.10, SD 0.88), and viewing assignments as boring (mean ranking 2.06, SD 0.81) as *minor* to *moderate* barriers to clients completing homework. Furthermore, therapists perceived clients not knowing how (mean ranking 1.91, SD 0.77), what (mean ranking 1.88, SD 0.81), or why (mean ranking 1.73, SD 0.75) assignments should be completed as *minor* barriers to nonbarriers to their clients completing homework assignments.

New Adhere.ly Features to Prioritize for Therapists Seeing Adult Clients

With respect to adding new features to Adhere.ly, therapists primarily suggested adding self-report questionnaires or assessments (mean ranking 6.76, SD 1.46), integration with telemedicine platforms (mean ranking 5.18, SD 2.26), and therapy guides or treatment protocols (mean ranking 4.64, SD 2.31). These were followed by the ability to edit text language in SMS text message or email reminders (mean ranking 4.47, SD 2.08), customized exercises for their practice (mean ranking 3.97, SD 2.92), integration with EHR software (mean ranking 3.94, SD 2.38), incentives or rewards for clients (mean ranking 2.69, SD 1.82), integration with wearables (mean ranking 2.68, SD 1.89), and integration with other treatment software (mean ranking 1.67, SD 1.97).

New Adhere.ly Exercises to Prioritize for Therapists Seeing Adult Clients

Regarding adding new therapeutic exercises to Adhere.ly, therapists suggested prioritizing relaxation (eg, breathing, muscle relaxation, mindfulness, and grounding; mean ranking 7.38, SD 2.20) and self-monitoring (eg, journaling, emotions, thoughts, and behaviors; mean ranking 6.57, SD 2.08). These were followed by coping and emotion regulation (eg, cognitive reappraisal and acceptance; mean ranking 5.49, SD 2.33), behavioral activation (eg, scheduling pleasant and important activities; mean ranking 5.04, SD 2.18), interpersonal (eg, social skill training; mean ranking 4.88, SD 2.45), and cognitive (eg, restructuring, flexibility and reappraisal, and modifying core beliefs; mean ranking 4.64, SD 2.71) exercises. Lower-priority exercises included problem-solving (eg, simplification, visualization, and planful problem-solving; mean 3.37, SD 2.24), behavioral (eg, contingency management, stimulus control, and shaping; mean ranking 2.87, SD 2.19), exposure (eg, building imaginal and in vivo exposure hierarchies; mean ranking 2.59, SD 2.32), and couple (eg, communication, compromising, and problem-solving; mean ranking 2.17, SD 2.85) exercises.

Discussion

Principal Findings

This mixed methods study explored common barriers experienced by mental health therapists when implementing homework with their adult clients. This study also explored potential solutions to those barriers, including health technology such as Adhere.ly, a web-based platform designed to help therapists engage adult clients in homework. Therapists in this study provided their input on new features and exercises that would enable Adhere.ly to better meet their needs when working with this population.

Consistent with previous research, therapists reported facing substantial challenges when implementing homework with adult clients [6,7,9,14-21]. While most therapists rated homework as moderately to extremely important, only a small percentage reported consistent and successful implementation. These results align with previous findings [9-11]. The most used evidence-based therapy protocols, particularly cognitive behavioral interventions, rely on homework exercises, such as worksheets and coping skill practice [4,5], yet there are few solutions available to support therapists in carrying out this critical therapy component [14]. The significant barriers faced by mental health care providers and the lack of available tools to resolve these barriers have significant implications for the successful implementation of evidence-based therapies in real-world practice settings.

Common Barriers to Homework

Primary barriers to therapists using homework with their adult clients differed significantly between the focus groups and survey in terms of barrier rankings.

While therapists and clients being too busy to assign or complete homework, respectively, was the top-ranked barrier in the focus groups, difficulty getting clients to complete assignments was the top-ranked therapist barrier in the surveys. Therapists being too busy to assign or review homework was the third priority-ranked barrier in the surveys. The emphasis on busyness and competing demands aligns with previous research [14,15]. Even when clients are motivated, therapists and clients juggle numerous responsibilities, commitments, and in-session priorities, which are frequently cited barriers to completing, assigning, and reviewing homework [15]. In the focus groups, therapist and client barriers were merged due to statements often referencing both therapist and client busyness, which is the mostly likely reason for the different rankings in the survey and focus groups. Furthermore, primarily clinic leaders identified therapist busyness as a barrier in the focus groups, and the surveys did not include clinic leaders.

Similarly, therapist perceptions that clients forget to complete their homework was ranked second in the focus groups and in the survey. Previous research [14,15] has identified clients forgetting as a primary barrier and emphasized the need for solutions that remind and encourage adult clients to complete homework given their other responsibilities. In contrast with child and adolescent clients, adults have no external reinforcement to stay focused on their therapy goals aside from their therapist, whom they may see at most once a week. Low rates of homework completion might be expected given that these barriers are coupled with clients finding assignments boring and receiving little reward or reinforcement for completing them, which was the third-ranked client barrier in the survey. Given that adult clients also struggle to remember what to do for homework and how to do it, solutions that remind clients to complete homework while providing easy-to-use and easy-to-access exercises with clear instructions have tremendous potential to overcome these common barriers to homework adherence.

The remaining barriers identified in the focus groups and the survey differed significantly. The third-ranked barrier in the focus groups was therapists struggling to find the right medium or tools with which to assign homework, followed by clients lacking the motivation to complete homework, clients receiving insufficient reinforcement for completing homework, clients avoiding homework due to distress, and other client barriers to homework (eg, lack of internet or smartphone or cognitive challenges). In contrast, the remaining survey priority-ranked categories were not wanting to overwhelm or distress clients and not knowing what to assign or how.

We attribute these differences to developing survey items using a combination of previous research [14,15] and a preliminary content analysis of the focus group results, using different samples for the focus groups and the survey, and the inclusion of clinic leaders in the focus groups who may have significantly different concerns from therapists (eg, emphasis on structural or administrative challenges). Furthermore, there is limited previous research in this area on which to base survey items, and prior work may not adequately capture barriers to implementing therapy homework. Indeed, the survey priority rankings were overall low for the client and therapist barrier question, indicating that the items selected for the survey were not representative of therapists' concerns. These discrepancies emphasize the importance of this mixed methods study given the critical role of homework in evidence-based therapy protocols. The qualitative findings will be used to develop future survey items that are more reflective of therapists' and clinic leaders' experiences in real-world practice settings.

Focus group participants reported that therapists' general therapeutic approach to addressing these challenges includes using motivational and reinforcement strategies to "meet clients where they are," including completing homework in the session and adapting homework assignments so that they are less time-consuming to complete. Therapists reported using multiple web platforms, mobile apps, and other materials (eg, paper homework) to create, assign, and remind clients of homework. For example, therapists often use phone reminders, paper materials, and various mobile apps to follow homework assignments prescribed according to specific therapy protocols. However, this approach results in therapists having to use several different platforms and mediums to implement homework, which can be inefficient and time-consuming for both the therapist and client, leading to frustration and lower homework adherence. Participants reported high perceived value of an all-in-one solution to support them in this process. This perceived value highlights the importance of adapting Adhere.ly for use with adult clients.

Adhere.ly as a Solution

Overview

Overall, participants viewed Adhere.ly as a potential solution to these challenges by offering a user-friendly, all-in-one solution. They felt that that Adhere.ly would likely reduce inefficiencies, increase homework completion rates, and be easy to use. Participants most frequently showed interest in Adhere.ly's SMS text message or email reminder feature. This feature allows clients to simply click the link from their SMS text messages or email and complete brief therapy exercises.

Most participants stated that Adhere.ly would be helpful for a wide range of therapists and clients while also believing that its utility would depend on therapist and client demographics. The most frequently mentioned concern in this domain was that some populations might not have access to the internet or cell phones with data plans and some clients or therapists might need to be technologically savvy to use the platform. As a result, Adhere.ly may resolve barriers to homework for some but not all therapist and client populations. Some therapists suggested that generations who grew up around these technologies might find SMS text message–based reminders and exercises especially motivating and helpful. Therapists and clients who did not grow up around these technologies might find learning a new application burdensome or less engaging. Additional training and support resources, while alleviating some barriers, would not address underlying attitudes toward or existing patterns of technology use [34].

Suggested Features and Exercises

Although the survey and focus groups identified similar features and exercises to add to make Adhere.ly compatible with adult clients, there were significant differences in the extent to which these features and exercises were emphasized by both samples. Regarding features, the survey respondents ranked enhanced assessment, integration with telemedicine platforms (eg, Doxy.me), therapy guides, customizable reminders, and EHR integration as the top 5 most important features to add. Focus group participants emphasized a demonstration video to orient clients and therapists regarding the application, EHR integration, gamification, and custom reminders as the most important features to add. We suspect that these differences are largely due to the focus group sample including clinic leaders because all the focus group participants who emphasized the demonstration video and EHR integration and most of the participants who mentioned gamification were clinic leaders. In future studies exploring the utility of health technology to enhance the implementation of evidence-based therapies, we will include clinic leaders in both the surveys and focus groups and ensure that extensive demographic and professional information is collected in both samples.

Regarding requested therapy exercises, evidence-based therapy protocol homework assignments (eg, mindfulness-based therapies and trauma-focused interventions) were most frequently requested, followed by specific coping skill assignments (eg, journaling). Eye movement desensitization and reprocessing, cognitive processing therapy, and dialectical behavior therapy were the most frequently referenced interventions, reflecting a high need for trauma-focused homework solutions. The survey options were worded differently than the focus group exercise categories, but the results were fairly similar across the samples. Survey respondents also emphasized mindfulness and relaxation-based exercises, self-monitoring exercises, coping and emotion regulation, behavioral activation, interpersonal skills, and cognitive exercises as the top-ranked exercises to add. Given that 69% (69/100) of the survey respondents identified as cognitive behavioral therapists, it makes sense that the most requested exercises aligned with second- or third-wave cognitive behavioral interventions. Focus group participants were more likely to discuss the overall need for including evidence-based therapy protocol homework with an emphasis on mindfulness-based and trauma-focused interventions.

Limitations

While this study offers valuable insights into common barriers faced by therapists to implementing homework with their adult clients and how health technology solutions such as Adhere.ly might help, there are some limitations to consider. This study included a small sample of therapists and clinic leaders for the focus groups and a separate sample of therapists for the survey. Recruitment was impacted by the COVID-19 pandemic, which led to smaller-than-recommended focus groups; however, this study was conducted during a period of unprecedented need for telehealth solutions, which we believe is a strength.

While efforts were made to ensure diversity in this sample, the findings may not reflect all mental health care providers, and true integration of the qualitative and quantitative findings would require at least some overlap in the survey and focus group samples. Furthermore, this study did not adequately assess therapist and clinic leader demographics to match the demographics collected in the survey, which will be corrected in future studies. Lack of inclusion of this information impacted our ability to assess how participants' contexts may have influenced their responses. Specifically, the therapists and clinic leaders who consented to participate may have greater comfort with or interest in health technology solutions than the average mental health clinician.

In addition, these data were obtained using focus groups and self-report assessments, which are subject to inherent biases and limitations. While we aimed to reduce the potential for researcher bias, as described in the Methods section, all researcher contributions are also subject to inherent biases and limitations. This study aimed to assess therapist and clinic leader perceptions on Adhere.ly's utility for adult-serving clinicians given that its utility for therapists serving children and families has been previously assessed. As a result, the findings cannot be generalized to therapists working across all populations and settings.

Conclusions

This study identified primary challenges that therapists face while implementing homework in adult-serving mental health outpatient treatment. The findings indicate that, with the additional features and therapeutic exercises requested by respondents, Adhere.ly has strong potential to address some of these challenges and increase overall homework completion rates, which would theoretically support improved clinical outcomes.

The results of this study were used to inform the user-centered development of the requested features and exercises, which can be accessed freely through the Adhere.ly platform. However, the broader implications of this study are that it was conducted at a time of unprecedented need for telehealth solutions and, at present, evidence-based therapies are inadequately implemented in real-world practice settings due to the aforementioned challenges. There is incredibly limited research on evidence-based therapy homework implementation despite its clear implications for the implementation science literature.

Ongoing research is exploring the feasibility and effectiveness of using Adhere.ly in mental health therapy practice settings. Specifically, we are evaluating its ability to contribute to improved homework implementation and adherence and, in turn, clinical and functional outcomes. Future studies will also address the limitations of this study through continued refinement of Adhere.ly and study methodology. Overall, this study contributes to a growing but still small body of literature on therapy homework implementation and highlights the importance of innovative technology solutions that address common challenges experienced by therapists.

Acknowledgments

The research reported in this publication was supported by the National Institute of Mental Health of the National Institutes of Health under award R41MH126734. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

BEB, BMW, and DT are shareholders of Adhere.ly LLC, which licenses the Adhere.ly software. BMW and DT are shareholders of Doxy.me Inc, a commercial telemedicine company. BEB, JI, and JFB are employees of Doxy.me Inc. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Adhere.ly feature categories.
[DOCX File , 32 KB - humanfactors_v11i1e56567_app1.docx]

Multimedia Appendix 2
Health care provider survey.
[DOCX File , 24 KB - humanfactors_v11i1e56567_app2.docx]

References

1. Whitney DG, Peterson MD. US national and state-level prevalence of mental health disorders and disparities of mental health care use in children. JAMA Pediatr 2019 Apr 01;173(4):389-391 [FREE Full text] [doi: 10.1001/jamapediatrics.2018.5399] [Medline: 30742204]
2. Key substance use and mental health indicators in the United States: results from the 2020 national survey on drug use and health. Substance Abuse and Mental Health Services Administration. URL: https://www.samhsa.gov/data/sites/default/files/reports/rpt35325/NSDUHFFRPDFWHTMLFiles2020/2020NSDUHFFR1PDFW102121.pdf [accessed 2022-09-22]
3. Hu TW. Perspectives: an international review of the national cost estimates of mental illness, 1990-2003. J Ment Health Policy Econ 2006 Mar;9(1):3-13. [Medline: 16733267]
4. Kazantzis N, Whittington C, Dattilio F. Meta-analysis of homework effects in cognitive and behavioral therapy: a replication and extension. Clin Psychol Sci Pract 2010;17(2):144-156. [doi: 10.1111/j.1468-2850.2010.01204.x]
5. Kazantzis N, Whittington C, Zelencich L, Kyrios M, Norton PJ, Hofmann SG. Quantity and quality of homework compliance: a meta-analysis of relations with outcome in cognitive behavior therapy. Behav Ther 2016 Sep;47(5):755-772. [doi: 10.1016/j.beth.2016.05.002] [Medline: 27816086]
6. Kazantzis N, Deane FP, Ronan KR, L'Abate L. Using Homework Assignments in Cognitive Behavior Therapy. Milton Park, UK: Routledge; Dec 2005.
7. Hudson JL, Kendall PC. Showing you can do it: homework in therapy for children and adolescents with anxiety disorders. J Clin Psychol 2002 May;58(5):525-534. [doi: 10.1002/jclp.10030] [Medline: 11967878]
8. Scheel MJ, Hanson WE, Razzhavaikina TI. The process of recommending homework in psychotherapy: a review of therapist delivery methods, client acceptability, and factors that affect compliance. Psychother Theory Res Pract Train 2004;41(1):38-55 [FREE Full text] [doi: 10.1037/0033-3204.41.1.38]
9. Dattilio FM, Kazantzis N, Shinkfield G, Carr AG. A survey of homework use, experience of barriers to homework, and attitudes about the barriers to homework among couples and family therapists. J Marital Fam Ther 2011 Apr;37(2):121-136. [doi: 10.1111/j.1752-0606.2011.00223.x] [Medline: 21457279]
10. Kazantzis N, Lampropoulos GK, Deane FP. A national survey of practicing psychologists' use and attitudes toward homework in psychotherapy. J Consult Clin Psychol 2005 Aug;73(4):742-748. [doi: 10.1037/0022-006X.73.4.742] [Medline: 16173863]
11. Kazantzis N, Deane FP. Psychologists' use of homework assignments in clinical practice. Prof Psychol Res Pract 1999;30(6):581-585. [doi: 10.1037//0735-7028.30.6.581]

12. Helbig S, Fehm L. Problems with homework in CBT: rare exception or rather frequent? Behav Cogn Psychother 2004 Jun 14;32(3):291-301. [doi: 10.1017/S1352465804001365]
13. Gaynor ST, Lawrence PS, Nelson-Gray RO. Measuring homework compliance in cognitive-behavioral therapy for adolescent depression: review, preliminary findings, and implications for theory and practice. Behav Modif 2006 Sep;30(5):647-672. [doi: 10.1177/0145445504272979] [Medline: 16894234]
14. Bunnell BE, Nemeth LS, Lenert LA, Kazantzis N, Deblinger E, Higgins KA, et al. Barriers associated with the implementation of homework in youth mental health treatment and potential mobile health solutions. Cognit Ther Res 2021 Apr;45(2):272-286 [FREE Full text] [doi: 10.1007/s10608-020-10090-8] [Medline: 34108776]
15. Kazantzis N, Shinkfield G. Conceptualizing patient barriers to nonadherence with homework assignments. Cogn Behav Pract 2007 Aug;14(3):317-324. [doi: 10.1016/j.cbpra.2006.08.003]
16. Callan JA, Dunbar-Jacob J, Sereika SM, Stone C, Fasiczka A, Jarrett RB, et al. "Barriers to cognitive behavioral therapy homework completion scale- depression version": development and psychometric evaluation. Int J Cogn Ther 2012;5(2):219-235 [FREE Full text] [doi: 10.1521/ijct.2012.5.2.219] [Medline: 24049556]
17. Garland A, Scott J. Using homework in therapy for depression. J Clin Psychol 2002 May;58(5):489-498. [doi: 10.1002/jclp.10027] [Medline: 11967875]
18. Bru L, Solholm R, Idsoe T. Participants' experiences of an early cognitive behavioral intervention for adolescents with symptoms of depression. Emot Behav Difficulties 2013;18(1):24-43. [doi: 10.1080/13632752.2012.675138]
19. Leahy RL. Improving homework compliance in the treatment of generalized anxiety disorder. J Clin Psychol 2002 May;58(5):499-511. [doi: 10.1002/jclp.10028] [Medline: 11967876]
20. Simons AD, Marti CN, Rohde P, Lewis CC, Curry J, March J. Does homework "matter" in cognitive behavioral therapy for adolescent depression? J Cogn Psychother 2012 Jan;26(4):390-404. [doi: 10.1891/0889-8391.26.4.390]
21. Coon DW, Rabinowitz YG, Thompson LW, Gallagher-Thompson D. Older adults. In: Kazantzis N, Deane FP, Ronan KR, L'Abate L, editors. Using Homework Assignments in Cognitive Behavioral Therapy. Milton Park, UK: Routledge; 2005.
22. Adhere.ly. URL: https://www.adhere.ly/ [accessed 2022-10-18]
23. Kingery JN, Roblek TL, Suveg C, Grover RL, Sherrill JT, Bergman RL. They're not just "little adults": developmental considerations for implementing cognitive-behavioral therapy with anxious youth. J Cogn Psychother 2006 Sep;20(3):263-273. [doi: 10.1891/jcop.20.3.263]
24. Sauter FM, Heyne D, Michiel Westenberg P. Cognitive behavior therapy for anxious adolescents: developmental influences on treatment design and delivery. Clin Child Fam Psychol Rev 2009 Dec;12(4):310-335 [FREE Full text] [doi: 10.1007/s10567-009-0058-z] [Medline: 19568935]
25. Carr A. What Works with Children, Adolescents, and Adults?: A Review of Research on the Effectiveness of Psychotherapy. Milton Park, UK: Taylor & Francis; 2008.
26. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
27. Bernard HR. Research Methods in Anthropology: Qualitative and Quantitative Approaches. Walnut Creek, CA: AltaMira Press; 2006.
28. Clarke V, Braun V. Thematic analysis. J Posit Psychol 2016 Dec 09;12(3):297-298. [doi: 10.1080/17439760.2016.1262613]
29. Bunnell BE, Kazantzis N, Paige SR, Barrera J, Thakkar RN, Turner D, et al. Provision of care by "real world" telemental health providers. Front Psychol 2021 May 07;12:653652 [FREE Full text] [doi: 10.3389/fpsyg.2021.653652] [Medline: 34025519]
30. Damschroder LJ. The consolidated framework for implementation research website. Consolidated Framework for Implementation Research. URL: https://cfirguide.org/ [accessed 2020-08-25]
31. Schoville R, Titler MG. Integrated technology implementation model: examination and enhancements. Comput Inform Nurs 2020 Nov;38(11):579-589. [doi: 10.1097/CIN.0000000000000632] [Medline: 32520784]
32. Miles MB, Huberman AM. Qualitative Data Analysis: An Expanded Sourcebook. Thousand Oaks, CA: SAGE Publications; 1994.
33. Hofmann S, Hayes SC. Process-Based CBT: The Science and Core Clinical Competencies of Cognitive Behavioral Therapy. Oakland, CA: New Harbinger Publications; 2018.
34. Heaslip V, Richer S, Simkhada B, Dogan H, Green S. Use of technology to promote health and wellbeing of people who are homeless: a systematic review. Int J Environ Res Public Health 2021 Jun 25;18(13):6845 [FREE Full text] [doi: 10.3390/ijerph18136845] [Medline: 34202242]

Abbreviations

CBT: cognitive behavioral therapy
CFIR: Consolidated Framework for Implementation Research
COREQ: Consolidated Criteria for Reporting Qualitative Research
EHR: electronic health record

ITIM: Integrated Technology Implementation Model

Edited by A Kushniruk; submitted 19.01.24; peer-reviewed by KC Stryczek, JJ Kim; comments to author 09.05.24; revised version received 19.09.24; accepted 14.10.24; published 12.12.24.

Please cite as:
Bunnell BE, Schuler KR, Ivanova J, Flynn L, Barrera JF, Niazi J, Turner D, Welch BM
Expanding a Health Technology Solution to Address Therapist Challenges in Implementing Homework With Adult Clients: Mixed Methods Study
JMIR Hum Factors 2024;11:e56567
URL: https://humanfactors.jmir.org/2024/1/e56567
doi:10.2196/56567
PMID:

Original Paper

Interest in mHealth Among Patients With Low Back Pain: Cross-Sectional Study

Jonas Ammundsen Ipsen[1,2], MSc; Louise Fleng Sandal[3], PhD; Natalie Hong Siu Chang[1,4], MSc; Berit Schiøttz-Christensen[5], MD, PhD; Karen Søgaard[3,6], PhD; Anders Hansen[1,4], PhD

[1]Department of Regional Health Research, University of Southern Denmark, Odense, Denmark
[2]Department of Physical Therapy and Occupational Therapy, Lillebaelt Hospital, University Hospital of Southern Denmark, Kolding, Denmark
[3]Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark
[4]Medical Research, Spine Centre of Southern Denmark, University Hospital of Southern Denmark, Middelfart, Denmark
[5]Research Unit of General Practice, University of Southern Denmark, Odense, Denmark
[6]Department of Clinical Research, University of Southern Denmark, Odense, Denmark

Corresponding Author:
Jonas Ammundsen Ipsen, MSc
Department of Physical Therapy and Occupational Therapy
Lillebaelt Hospital
University Hospital of Southern Denmark
Sygehusvej 24
Kolding, 6000
Denmark
Phone: 45 76362832
Email: jonas.ammundsen.ipsen@rsyd.dk

Abstract

Background: Digitally supported self-management tailored to an individual's need, in addition to usual care, may reduce pain-related disability compared to usual care alone, and patients with low back pain (LBP) using mobile health (mHealth) solutions express positive experiences. Hence, implementing mHealth solutions designed to support self-management is desirable from a clinical and patient perspective. Easily accessible mHealth solutions that can support the self-management of patients with LBP are available, but interest may be subgroup specific. Understanding the characteristics and preferences of patients with LBP labeled as interested may help to reach relevant LBP patient groups and inform the development and implementation of effective interventions with mHealth for patients with LBP.

Objective: This study aims to explore the proportion of patients with LBP labeled as interested in testing an mHealth solution designed to support self-management in addition to usual care and to assess how these patients differ from those who were labeled as not interested.

Methods: This exploratory cross-sectional study analyzed demographic and patient-reported outcomes from the SpineData registry, a Danish registry of patients with LBP in an outpatient setting. Between February and December 2019, the SpineData registry was used to assess the preliminary eligibility of patients for a clinical trial (selfBACK). Patients were labeled as interested or uninterested depending on if they responded to an invitation to be tested for eligibility for the trial Outcomes were selected from the International Classification of Functioning core set of LBP using a clinical approach. Associations were assessed in a backward selection process, and the proportion of variance explained was assessed with pseudo-R^2 statistic.

Results: This study included 843 patients, with 181 (21%) individuals labeled as interested in participating in the selfBACK trial. Notably, the cohort labeled as interested differed from their uninterested counterparts in two key aspects: age (36-65 years: 116/181, 64.1% vs 347/662, 52.4%; *P*=.003) and smoking status (smokers: 22/181, 12.5% vs 174/662, 26.6%; *P*<.001). Those aged 36-65 years had higher odds of being labeled as interested compared to individuals aged 18-35 years (odds ratio [OR] 0.43, 95% CI 0.26-0.71) and those 65 years or older (OR 0.77, 95% CI 0.53-1.15). Nevertheless, age accounted for only a modest proportion of variance (R^2=0.014). Smokers demonstrated lower odds of being labeled as interested (OR 0.39, 95% CI 0.24-0.64), with smoking status explaining a similarly small proportion of variance (R^2=0.019). Collectively, age and smoking status accounted for 3.3% of the variance.

Conclusions: Our investigation revealed that 181 (21%) individuals with LBP invited to participate in the mHealth solution trial for self-management expressed interest. Generally, the characteristics of those labeled as interested and uninterested were comparable. Of note, patients aged 36-65 years had a higher frequency of being labeled as interested compared to their younger and older counterparts.

(JMIR Hum Factors 2024;11:e48729) doi:10.2196/48729

KEYWORDS

low back pain; mHealth solutions; mobile health; characteristics; patient interest; transferability; representativeness

Introduction

Digital health interventions delivered with smartphones (mobile health [mHealth] solutions) are accessible to most patients across age, geography, and socioeconomic status. Thus, clinicians' expectations of mHealth solutions are significant, and the availability of new solutions on the commercial market every day also indicates a strong general interest in using mHealth solutions [1,2]. Nevertheless, many mHealth solutions have limited download rates, and if downloaded, the use can be scarce [3,4]. This discrepancy may indicate a need for a better understanding of potential users and their characteristics.

For patients with low back pain (LBP; not attributable to a recognizable, known specific pathology such as infections, fractures, or structural deformity), self-management support is recommended as the first line of treatment [5-8]. This may involve empowering patients to know when to consult for diagnostic assessment, symptom relief, or advice [9]. Digitally supported self-management may be delivered through smartphone apps or digital platforms to facilitate and enhance such self-management practices. Research indicates that the integration of such digitally supported self-management strategies, when combined with standard care, can lead to a reduction in pain-related disability [10]. Further, evidence supports that mHealth solutions designed to support self-management are accepted by patients with chronic LBP [9]. Therefore, there is a growing interest in implementing mHealth solutions designed to support self-management from both clinical and patient perspectives. However, despite the potential benefits, the level of acceptance and use of these interventions remains an area that requires further investigation.

However, studies on other patient groups using mHealth solutions report that lower age, higher education, higher income, higher BMI, and higher self-perceived health are associated with increased use [4,11]. In contrast, the cost of using these apps is a significant barrier [11].

Individuals with LBP who use mHealth solutions to self-manage may thus represent a specific subset within the general population. Therefore, this study aimed to investigate the percentage who expressed interest in participating in a trial evaluating an mHealth solution designed for self-management alongside standard care, as well as to evaluate potential distinctions between those who were labeled as expressing interest and those who were not.

Methods

Study Design

This exploratory cross-sectional study used demographic and patient-reported outcomes (PROs) from an internet-based multiuser clinical registry (SpineData) [12]. Reporting follows the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines [13].

Setting

Data were collected at the Spine Centre of Southern Denmark, an outpatient hospital that performs clinical spine evaluations [12]. General practitioners or chiropractors typically refer patients to the Spine Centre, which performs a multidisciplinary assessment of its patients, with more than 10,000 new cases yearly.

Before patients are evaluated at the Spine Centre, they provide information in the local SpineData registry [14]. The registry is designed based on the biopsychosocial model of health, and information is collected across the health domains of pain, activity limitation, work participation, psychological factors, physical impairment, and contextual factors [12]. To mitigate nonresponse and missing information, SpineData uses a "waterfall" model (eg, patients in employment are not asked to respond to causes for unemployment). SpineData has an overall completion rate of 80% and approximately 60% of patients agreed to their responses being used for research [14]. The use of this registry allows for the comprehensive assessment of patients consulted at the Spine Centre and provides a rich source of data for research studies, such as the one presented in this paper.

Participants

Between February and December 2019, SpineData was used to identify eligible patients based on the following criteria: consenting to be contacted for research projects, proficiency in Danish, and experiencing LBP in the past 14 days that exceeded their leg pain in severity. Patients with previous back surgery, who were actively filing for a pension, or who were younger than 18 years were not invited. All patients matching the eligibility criteria were sent a letter of invitation to hear more about the selfBACK trial. One reminder was sent. The patients who did not respond to either invitation or reminder were labeled uninterested. The selfBACK trial investigated the effectiveness of the selfBACK digital decision support system that provided patients with LBP individually tailored digital support in an app format using three content domains: (1) physical activity, (2) education, and (3) exercise programs. The trial investigated the additive effect of the selfBACK system in addition to usual care. Participants in this trial were recruited from primary health care such as chiropractors, physiotherapists, and general practitioners in addition to the Spine Center of Southern Denmark. Recruitment was performed in Denmark and Norway. The population within this study concerns the pool of patients seen at the Spine Center, who would have received an invitation to eligibility screening to the selfBACK trial based on their answers given in the SpineData clinical registry. In this study, all patients who matched the preliminary eligibility criteria for the selfBACK trial were included [15].

Outcomes

The variables of interest were selected from the SpineData registry, based on the International Classification of Function core set for LBP and clinical reasoning [16]. The demographics and clinical characteristics comprised the domains of pain, activity limitation, work participation, and psychological and contextual factors (Textbox 1).

Textbox 1. Detailed description of the content and handling of included outcomes.

Sex

- Male or female

Age

- Patients were categorized into age groups ≤35, 36-65, and >65 years.

BMI

- The anthropometric variables of height and weight were used to calculate BMI (kg/m^2).

Smoker status

- Categorical variable that was dichotomized to smoker and nonsmoker strata. If a patient indicated cigarette use of any kind, they were categorized as a smoker.

Alcohol consumption

- Categorical variable that was stratified into two groups based on the consumption of more than 14 alcoholic beverages a week. The threshold was based on the recommendation of the Danish Ministry of Health [17].

Comorbidities

- This variable was based on four dichotomous variables: allergies, including medication; cancers; heart disease; and lung disease. If a patient replied yes to one of these variables, they were categorized as having comorbidities.

Current work status

- The work status variables consisted of different ways of participating in the labor market: working full- or part-time, flex job, in education, job training due to inability to maintain habitual job function, unemployed, early retirement, pensions, stay at home, and other. The variable was dichotomized to working or not by grouping patients indicating working part- or full-time, flex, and students in one group and the remainder in another group.

Multiple pain sites

- SpineData contains a freehanded pain drawing. The pain drawing was post defined into 46 anatomical regions. In this study, the regions were grouped into 9 areas: neck, shoulders, upper back, elbows, lower back, wrists/hands, hips/thighs, knees, and ankles/feet, inspired by Øverås et al [18]. Patients with two or more pain sites were considered as having multiple pain sites.

Average back pain

- The average back pain in the last 14 days was measured on a 0-10 numeric rating scale, with 10 indicating the worst imaginable pain.

STarT BaCK screening tool [19]

- The STarT Back scores categorize patients into three strata based on their risk of developing chronicity: low risk, moderate risk, and high risk of chronicity:
  - Low risk: <3
  - Moderate risk: ≥4 and subscore ≤3
  - High risk: ≥4 and subscore ≥4

EQ-5D-5L-VAS [20]

- Numeric rating scale score spanning from 0 to 100, with 100 representing the best possible health state

Oswestry Disability Index (ODI) [21]

- The ODI is a questionnaire containing 10 items that are scored from 0 to 5. The maximum score is 50 points, which indicates that the patient is bedbound. The ODI has been found valid for patients with low back pain [22].
- To estimate the patients' functional level, the ODI Stata package was used. The ODI package allows for the imputation of data for one missing value. The missing values in one section were replaced with the average score for all sections.

Anxiety [23]

- Numeric score rating from 0 to 10, with 0 indicating no anxiety and 10 a high degree of anxiety

Social isolation [23]

- Numeric score rating from 0 to 10, with 0 indicating no loneliness and 10 a high degree of loneliness

Catastrophization (terrible pain that will never improve) [23]

- Numeric score rating from 0 to 10, with 0 indicating no catastrophization and 10 a high degree of catastrophization

Catastrophization (the pain is overwhelming) [23]

- Numeric score rating from 0 to 10, with 0 indicating no catastrophization and 10 a high degree of catastrophization

Risk of persisting pain [23]

- Numeric score rating from 0 to 10, with 0 indicating no risk of persisting pain and 10 a high risk of persisting pain

Feelings of sadness, depression, or hopelessness [23]

- Numeric score rating from 0 to 10, with 0 indicating no feelings of depression and 10 a constant presence of depression

Loss of interest or joy [23]

- Numeric score rating from 0 to 10, with 0 indicating no loss of interest or joy and 10 never feeling interest or joy

Fearing activity will damage the back [23]

- Numeric score rating from 0 to 10, with 0 indicating no fear that physical activity will damage the back and 10 completely agreeing that physical activity will damage the back

Fearing activity will increase the pain [23]

- Numeric score rating from 0 to 10, with 0 indicating completely disagreeing to avoid physical activity and 10 completely agreeing to avoid physical activity

Exposure

Patients were allocated into two groups based on their response to being invited for eligibility screening for the selfBACK trial. Those who responded positively to the invitation to be screened were labeled as interested in using the mHealth solution, whereas those who did not respond were labeled as uninterested.

Statistical Methods

The demographics and baseline characteristics of patients who were or were not labeled as interested in the digital mHealth intervention were assessed using the χ^2 test for categorical variables and 2-tailed Student *t* test for continuous variables. Baseline characteristics are reported as the proportion and percentage or mean and SD.

To assess the strength of associations between PROs and patients labeled as interested in mHealth or not, we used univariate and multivariate logistic regression analysis with an odds ratio (OR) and 95% CI. The associations were assessed in a backward selection process, and the proportion of variance explained was assessed with McFadden pseudo-R^2 statistic. Statistical analyses were performed with Stata statistical software (Release 17; StataCorp LLC). Missing information was handled using pairwise deletion. The ODI Stata package allows for data imputation for one missing value. The missing values in one section were replaced with the average score for all sections. To avoid overparameterizing the model, we aimed for a 1:10 patient-to-variable ratio.

Ethical Considerations

The Region of Southern Denmark was the data controller for this project, which is included in its records on personal data processing activities (file 21/13433). Data processing in the project was regulated by the Danish Act on Research Ethics Review of Health Research Projects section 14, subsection 2, which states that health research based solely on questionnaire surveys and registry data is exempt from the obligation to notify the committees. Following the Danish Health Care Act, we obtained approval for using hospital record data for scientific purposes from the council of the Region of Southern Denmark (file 21/25588). After merging, analyses were run on pseudonymized data, and the results presented in this manuscript do not enable the identification of single data participants. Hence, following national laws, no additional informed consent was collected and no remuneration was offered to patients.

Results

Overview

From February to the end of December 2019, 5796 patients (~80% of those invited) completed the SpineData registry before their diagnostic assessment at the Spine Centre. Of the total sample, 843 (15%) were invited to the selfBACK trial. The mean age of the cohort was 52 (SD 16.2) years, with an even distribution of sexes (male: n=429, 50.1%), and a mean BMI of 27.5 kg/m^2.

Of the 843 patients invited to the eligibility screen for the trial, 181 (21%) accepted the invitation and were stratified into the group who were labeled as interested in the mHealth solution.

Of the 21 included variables, 8 had complete responses, and none of the remaining 13 variables had more than 2.5% missing responses.

Comparison of Patients Who Were Labeled as Interested and Uninterested in an mHealth Solution

Patients labeled as interested in using the mHealth solution were aged 36-65 years (*P*=.003) and had a lower proportion of smokers (*P*<.001) compared to the patients labeled as uninterested. The remaining variables were not different between the patients labeled as interested and uninterested (Table 1).

Table 1. Baseline characteristics of patients labeled as interested in the mobile health solution compared to the uninterested patients.

Baseline characteristic[a]	Interested (n=181)	Uninterested (n=662)	*P* value
Female, n (%)	93 (51.3)	321 (48.5)	.49
Age (years), n (%)			.003
18-35	21 (11.6)	146 (22.1)	
36-65	116 (64.1)	347 (52.4)	
>65	44 (24.3)	169 (25.5)	
BMI (kg/m^2), mean (SD)	28.1 (5.8)	27.2 (5.0)	.05
Smokers, n (%)	22 (12.5)	174 (26.6)	<.001
<14 alcohol consumption per week, n (%)	173 (95.5)	636 (96.0)	.76
Has comorbidities, n (%)	84 (46.7)	208 (42.3)	.29
Working, n (%)	95 (52.4)	367 (55.4)	.49
Has multiple pain sites, n (%)	136 (75.1)	462 (69.7)	.24
Average back pain (score range: 0-10), mean (SD)	6.3 (2.0)	6.3 (1.9)	.91
STarT Back tool, n (%)			.34
Low risk	54 (29.8)	164 (24.7)	
Moderate risk	46 (25.5)	169 (25.5)	
High risk	81 (44.7)	329 (49.7)	
EQ-5D-5L-VAS (score range: 0-100), mean (SD)	59.0 (22.1)	55.2 (23.0)	.05
Oswestry Disability Index (score range: 0-50), mean (SD)	30.3 (15.6)	31.1 (14.9)	.50
Anxiety (score range: 0-10), mean (SD)	3.8 (3.0)	3.8 (3.1)	.99
Loneliness (score range: 0-10), mean (SD)	1.4 (2.4)	1.3 (2.2)	.67
Catastrophization (terrible pain that will never improve; score range: 0-10), mean (SD)	4.8 (2.9)	5.0 (3.0)	.48
Catastrophization (the pain is overwhelming; score range: 0-10), mean (SD)	3.7 (3.1)	4.1 (3.1)	.24
Risk of persisting pain (score range: 0-10), mean (SD)	6.8 (2.6)	6.8 (2.6)	.91
Sadness (score range: 0-10), mean (SD)	3.5 (3.1)	3.6 (3.1)	.71
Loss of interest or joy (score range: 0-10), mean (SD)	4.3 (3.3)	4.3 (3.2)	.88
Fearing activity will damage the back (score range: 0-10), mean (SD)	3.4 (2.9)	3.8 (3.2)	.08
Fearing activity will increase the pain (score range: 0-10), mean (SD)	4.8 (3.2)	4.4 (3.3)	.16

[a]Missing: 8 of the 21 variables had complete responses, and none of the remaining 13 variables had more than 2.5% missing responses.

Our results suggest that patients aged 36-65 years were more likely to be labeled as interested in mHealth solutions compared to patients between 18-35 years (OR 0.43, 95% CI 0.026-0.711) and 65 years or older (OR 0.77, 95% CI 0.525-1.153) and explained a limited proportion of variance (R^2=0.014). Smoker (OR 0.39, 95% CI 0.244-0.636) and the association explained a limited proportion of variance (R^2=0.019). Combined, the associations of age and smoking explained 3.3% of the proportion of variance.

These findings were supported by univariate regression analysis and a comparison of patients who were labeled as expressing interest in the mHealth solution to those who did not. The proportion of variance explained in the group of patients labeled as interested in mHealth solutions across the 21 selected variables was 0.059, with age and smoking status accounting for 0.033 of the variance (Table 2).

BMI (*P*=.05), overall perception of health measured using the EQ-5D-5L-VAS score (*P*=.05), and fear that activity will damage the back (*P*=.08) were borderline significant.

Table 2. Associations to being labeled as interested and proportion of variance explained.

	Odds ratio (95% CI)	SE	*Z*	People invited, N	*P* value (*P*>|z|)	R^2
Age (years)				830		0.014
36-65 (reference)	—	—	—		—	
18-35	0.43 (0.02-0.71)	0.114	–3.17		.002	
>65	0.77 (0.52-1.15)	0.151	–1.48		.14	
Smoking				830		0.019
Nonsmoker (reference)	—	—	—	—	—	—
Smoker	0.39 (0.24-0.64)	0.098	–3.72		<.001	

Discussion

Principal Results

This study aimed to explore the proportion of patients with LBP who were labeled as interested in using an mHealth solution designed to support self-management in addition to usual care and assess how these patients differed from those who were labeled as not interested. We found that 21% of the eligible patients were labeled as interested in using the mHealth solution. The groups had no statistically significant differences except that patients labeled as interested were more frequently within the 36-65 years age range and were nonsmokers.

Comparison With Prior Work

Previous evidence of the characteristics and associations of patients with LBP and their interest in mHealth solutions is limited. Contrary to Krebs and Duncan [4], we found a nonsignificant association between BMI and no association between being younger and labeled as interested in mHealth solutions. The key differences between Krebs and Duncan [4] and this study are the target populations (general population) and the type of mHealth solutions included (fitness apps or calorie trackers). Similar to our results, Philip et al [24] identified an association between higher age and increased use of mHealth solutions among patients with chronic pain. We suggest that the differences in results between Krebs and Duncan [4], Philip et al [24], and this study were due to differences between participants from the general population and patients with LBP or chronic pain. Three recent studies have assessed the characteristics and associations of users and nonusers of different mHealth apps, all using participants from the general population, but still lacking consensus. Walrave et al [25] identified no sociodemographic differences between users and nonusers of contact tracking alert apps, including the Belgian Corona alert app. A study of the general US population identified strong associations of age, gender, and education level with the use of fitness apps and calorie counters [26]. Lim et al [27] identified that female patients with higher education were more prevalent users of mHealth apps. Although this lack of consensus regarding patient interest could indicate a call for more research, it could also reflect that the interest in mHealth solutions may be characterized by patients' preferences and perspectives on the relevance of mHealth solutions.

Strengths and Limitations

This study benefitted from several strengths. First, we had access to comprehensive information on the patients participating through the SpineData registry. Further, we benefitted from the fact that SpineData has been in routine use for several years and is frequently updated per clinician and evidence demand [14]. Thus, the PROs were collected using validated questionnaires or questions designed for the LBP population and International Classification of Function core set [14,16,28,29]. The included patients were identified using a computer algorithm, and patients were sent one invitation and one reminder invitation to be screened for eligibility. Thus, the risk of unconscious bias in the recruitment was eliminated. However, using a single data source (SpineData) also limited the variables available to investigate in the univariate model. Low education and economic status have been associated with limited use and adoption of mHealth solutions [26,30], but this information was unavailable in SpineData. Smoking is reportedly more prevalent among patients with a lower socioeconomic or sociodemographic status [31,32]. Further, the use of one registry meant we only had access to PROs, which may be affected by recall bias. The statistically significant difference between being labeled as interested in mHealth solutions by smoking status could reflect a difference in education level. Thus, education level is a parameter that could differentiate the patients labeled as interested and those labeled as uninterested in the mHealth solution, although this hypothesis remains unanswered. Patients referred to the Spine Centre usually have pain for extended periods and at a higher intensity than patients in the primary sector [33]. Thus, these patients potentially have more complex LBP issues than those with LBP who were not referred, which means that our study population may be a subgroup of the general LBP population. The terms "interested" and "uninterested" pose a challenge due to their vague nature. We recognize the distinction between demonstrating a "cursory" interest and moving toward actual participation. After extensive discussions among authors, we chose the terms "interested" and "uninterested." Despite their less-than-optimal nature, we believe these terms best suit the context where we categorize patients based on their response to an invitation, progressing from screening to eligibility for participation in a trial evaluating an mHealth solution supporting self-management in patients with LBP. Further, some patients might be interested in testing an mHealth solution but uninterested in participating in a trial or vice versa. Further, those labeled as uninterested in the mHealth solution in this study might see advantages in mHealth solutions that they found more relevant like how to stop smoking or lose weight [34]. This study only addresses patient characteristics; however, investigating clinicians' perspectives on the use and adoption of mHealth solutions in LBP self-management will similarly inform on barriers to and facilitators of increased mHealth adoption in clinical practice. However, as the SpineData clinical registry only entails patient data, this perspective was not possible in this study. Thus, the results of this study should be interpreted with caution regarding generalizability, and future qualitative or mixed methods studies could explore patients' preferences and perceptions of the relevance of mHealth solutions. Another important area of research can be clinicians' acceptability of mHealth solutions and the need for rigorous demonstrations of safety and efficacy to alleviate any reservations or hesitance among clinicians.

Conclusion

This study aimed to explore the characteristics of patients labeled as interested or uninterested in participating in a trial testing an mHealth solution designed to improve self-management. Our study identified that 21% (n=181) of eligible patients with LBP were labeled as interested in participating in the trial testing an mHealth solution to support self-management. Overall, the patients labeled as interested and uninterested, except for age and smoking status were similar. Interestingly, patients aged 36-65 years were more frequently labeled as interested in the mHealth solution. Thus, patients aged 36-65 years may be more interested in adopting mHealth solutions. How to increase interest in mHealth solutions among younger and older patients with LBP is an important consideration for future research and developers, especially as the findings of the selfBACK trial indicate an increased effect for older patients.

Authors' Contributions

AH, BSC, and KS conceptualized the study. AH, LFS, and NHSC acquired the data and data permissions. JAI performed the data analysis and drafted the manuscript, with support from NHSC, AH, and LFS. All authors helped draft and critically revised the manuscript for important intellectual content and approved the final version.

Conflicts of Interest

None declared.

References

1. Bates DW, Landman A, Levine DM. Health Apps and Health Policy: What Is Needed? JAMA 2018 Nov 20;320(19):1975-1976. [doi: 10.1001/jama.2018.14378] [Medline: 30326025]
2. Aitken M. Digital health trends 2021: innovation, evidence, regulation, and adoption. IQVIA. 2021 Jul 22. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports-and-publications/reports/digital-health-trends-2021 [accessed 2023-11-15]
3. mHealth economics 2017/2018 – connectivity in digital health. Research2Guidance. 2017. URL: https://research2guidance.com/product/connectivity-in-digital-health/ [accessed 2022-12-25]
4. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
5. Hutting N, Johnston V, Staal JB, Heerkens YF. Promoting the use of self-management strategies for people with persistent musculoskeletal disorders: the role of physical therapists. J Orthop Sports Phys Ther 2019 Apr;49(4):212-215. [doi: 10.2519/jospt.2019.0605] [Medline: 30931733]
6. Caneiro JP, Roos EM, Barton CJ, O'Sullivan K, Kent P, Lin I, et al. It is time to move beyond 'body region silos' to manage musculoskeletal pain: five actions to change clinical practice. Br J Sports Med 2020 Apr;54(8):438-439. [doi: 10.1136/bjsports-2018-100488] [Medline: 31604698]
7. Corp N, Mansell G, Stynes S, Wynne-Jones G, Morsø L, Hill JC, et al. Evidence-based treatment recommendations for neck and low back pain across Europe: A systematic review of guidelines. Eur J Pain 2021 Feb;25(2):275-295 [FREE Full text] [doi: 10.1002/ejp.1679] [Medline: 33064878]
8. Qaseem A, Wilt TJ, McLean RM, Forciea MA, Denberg TD, Barry MJ, et al. Noninvasive treatments for acute, subacute, and chronic low back pain: a clinical practice guideline from the American College of Physicians. Ann Intern Med 2017 Apr 04;166(7):514-530 [FREE Full text] [doi: 10.7326/M16-2367] [Medline: 28192789]
9. Kongsted A, Ris I, Kjaer P, Hartvigsen J. Self-management at the core of back pain care: 10 key points for clinicians. Braz J Phys Ther 2021;25(4):396-406 [FREE Full text] [doi: 10.1016/j.bjpt.2021.05.002] [Medline: 34116904]
10. Sandal LF, Bach K, Øverås CK, Svendsen MJ, Dalager T, Stejnicher Drongstrup Jensen J, et al. Effectiveness of app-delivered, tailored self-management support for adults with lower back pain-related disability: a selfBACK randomized clinical trial. JAMA Intern Med 2021 Oct 01;181(10):1288-1296 [FREE Full text] [doi: 10.1001/jamainternmed.2021.4097] [Medline: 34338710]

11. Rai A, Chen L, Pye J, Baird A. Understanding determinants of consumer mobile health usage intentions, assimilation, and channel preferences. J Med Internet Res 2013 Aug 02;15(8):e149 [FREE Full text] [doi: 10.2196/jmir.2635] [Medline: 23912839]
12. Kent P, Kongsted A, Jensen TS, Albert HB, Schiøttz-Christensen B, Manniche C. SpineData - a Danish clinical registry of people with chronic back pain. Clin Epidemiol 2015;7:369-380 [FREE Full text] [doi: 10.2147/CLEP.S83830] [Medline: 26316820]
13. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007 Oct 20;370(9596):1453-1457 [FREE Full text] [doi: 10.1016/S0140-6736(07)61602-X] [Medline: 18064739]
14. Chang N. Rygmedicinsk Forskning - SpineData og MiRD databaserne. Sygehus Lillebælts. 2019. URL: http://tinyurl.com/48vu9h2m [accessed 2020-07-05]
15. Sandal LF, Stochkendahl MJ, Svendsen MJ, Wood K, Øverås CK, Nordstoga AL, et al. An App-Delivered Self-Management Program for People With Low Back Pain: Protocol for the selfBACK Randomized Controlled Trial. JMIR Res Protoc 2019 Dec 03;8(12):e14720 [FREE Full text] [doi: 10.2196/14720] [Medline: 31793897]
16. Kirschneck M, Kirchberger I, Amann E, Cieza A. Validation of the comprehensive ICF core set for low back pain: the perspective of physical therapists. Man Ther 2011 Aug;16(4):364-372. [doi: 10.1016/j.math.2010.12.011] [Medline: 21269870]
17. Notat: Sundhedsstyrelsens udmeldinger om indtag af alkohol. Sundhedsstyrelsen. 2023 Jul 19. URL: https://www.sst.dk/da/Viden/Alkohol/Alkoholforebyggelse/Sundhedsstyrelsens-udmeldinger-om-alkohol [accessed 2020-03-25]
18. Øverås CK, Nilsen TIL, Nicholl BI, Rughani G, Wood K, Søgaard K, et al. Multimorbidity and co-occurring musculoskeletal pain do not modify the effect of the SELFBACK app on low back pain-related disability. BMC Med 2022 Feb 08;20(1):53 [FREE Full text] [doi: 10.1186/s12916-022-02237-z] [Medline: 35130898]
19. STarTBack online. STarTBack. URL: https://startback.hfac.keele.ac.uk/training/resources/startback-online/ [accessed 2024-02-01]
20. Soer R, Reneman MF, Speijer BLGN, Coppes MH, Vroomen PCAJ. Clinimetric properties of the EuroQol-5D in patients with chronic low back pain. Spine J 2012 Nov;12(11):1035-1039. [doi: 10.1016/j.spinee.2012.10.030] [Medline: 23199409]
21. Fairbank JC, Pynsent PB. The Oswestry Disability Index. Spine (Phila Pa 1976) 2000 Nov 15;25(22):2940-52; discussion 2952. [doi: 10.1097/00007632-200011150-00017] [Medline: 11074683]
22. Schwind J, Learman K, O'Halloran B, Showalter C, Cook C. Different minimally important clinical difference (MCID) scores lead to different clinical prediction rules for the Oswestry disability index for the same sample of patients. J Man Manip Ther 2013 May;21(2):71-78 [FREE Full text] [doi: 10.1179/2042618613Y.0000000028] [Medline: 24421616]
23. Kent P, Mirkhil S, Keating J, Buchbinder R, Manniche C, Albert HB. The concurrent validity of brief screening questions for anxiety, depression, social isolation, catastrophization, and fear of movement in people with low back pain. Clin J Pain 2014 Jun;30(6):479-489. [doi: 10.1097/AJP.0000000000000010] [Medline: 24281277]
24. Philip L, Roberts A, Currie M, Mort A. Technology for older adults: maximising personal and social interaction: exploring opportunities for eHealth to support the older rural population with chronic pain. Scott Geogr J 2015 Jan 15;131(3-4):181-193. [doi: 10.1080/14702541.2014.978806]
25. Walrave M, Waeterloos C, Ponnet K. Reasons for nonuse, discontinuation of use, and acceptance of additional functionalities of a COVID-19 contact tracing app: cross-sectional survey study. JMIR Public Health Surveill 2022 Jan 14;8(1):e22113 [FREE Full text] [doi: 10.2196/22113] [Medline: 34794117]
26. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017 Apr 19;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]
27. Lim D, Norman R, Robinson S. Consumer preference to utilise a mobile health app: A stated preference experiment. PLoS One 2020;15(2):e0229546 [FREE Full text] [doi: 10.1371/journal.pone.0229546] [Medline: 32084250]
28. Morsø L, Albert H, Kent P, Manniche C, Hill J. Translation and discriminative validation of the STarT Back Screening Tool into Danish. Eur Spine J 2011 Dec;20(12):2166-2173 [FREE Full text] [doi: 10.1007/s00586-011-1911-6] [Medline: 21769444]
29. Johnsen LG, Hellum C, Nygaard OP, Storheim K, Brox JI, Rossvoll I, et al. Comparison of the SF6D, the EQ5D, and the oswestry disability index in patients with chronic low back pain and degenerative disc disease. BMC Musculoskelet Disord 2013 Apr 26;14:148 [FREE Full text] [doi: 10.1186/1471-2474-14-148] [Medline: 23622053]
30. Faber JS, Al-Dhahir I, Reijnders T, Chavannes NH, Evers AWM, Kraal JJ, et al. Attitudes toward health, healthcare, and eHealth of people with a low socioeconomic status: a community-based participatory approach. Front Digit Health 2021;3:690182 [FREE Full text] [doi: 10.3389/fdgth.2021.690182] [Medline: 34713165]
31. Hitchman SC, Fong GT, Zanna MP, Thrasher JF, Chung-Hall J, Siahpush M. Socioeconomic status and smokers' number of smoking friends: findings from the International Tobacco Control (ITC) Four Country Survey. Drug Alcohol Depend 2014 Oct 01;143:158-166 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.07.019] [Medline: 25156228]

32. Tomioka K, Kurumatani N, Saeki K. The association between education and smoking prevalence, independent of occupation: a nationally representative survey in Japan. J Epidemiol 2020 Mar 05;30(3):136-142 [FREE Full text] [doi: 10.2188/jea.JE20180195] [Medline: 30828035]
33. Hansen A, Morsø L, Stochkendahl MJ, Andersen MKK, Schiøttz-Christensen B, Madsen SD, et al. Demographic and clinical characteristics of patients with low back pain in primary and secondary care settings in Southern Denmark. Scand J Prim Health Care 2023 Jun;41(2):152-159 [FREE Full text] [doi: 10.1080/02813432.2023.2196548] [Medline: 37154804]
34. Pan J, Dong H, Bryan-Kinns N. Perception and initial adoption of mobile health services of older adults in London: mixed methods investigation. JMIR Aging 2021 Nov 19;4(4):e30420 [FREE Full text] [doi: 10.2196/30420] [Medline: 34807836]

Abbreviations

LBP: low back pain
mHealth: mobile health
OR: odds ratio
PRO: patient-reported outcome
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by A Kushniruk; submitted 04.05.23; peer-reviewed by P Poulin, S Kheirinejad, C Baxter; comments to author 29.10.23; revised version received 12.12.23; accepted 20.01.24; published 12.02.24.

Please cite as:
Ipsen JA, Sandal LF, Chang NHS, Schiøttz-Christensen B, Søgaard K, Hansen A
Interest in mHealth Among Patients With Low Back Pain: Cross-Sectional Study
JMIR Hum Factors 2024;11:e48729
URL: https://humanfactors.jmir.org/2024/1/e48729
doi:10.2196/48729
PMID:38345837

Original Paper

A Mobile Health App to Support Home-Based Aerobic Exercise in Neuromuscular Diseases: Usability Study

Tim Veneman[1,2], MSc; Fieke Sophia Koopman[1,2], MD, PhD; Sander Oorschot[1,2], MSc; Pien G Koomen[1,2], PT; Frans Nollet[1,2], MD, PhD; Eric L Voorn[1,2], PhD

[1]Amsterdam University Medical Center location University of Amsterdam, Rehabilitation Medicine, Amsterdam, Netherlands

[2]Amsterdam Movement Sciences, Rehabilitation & Development, Amsterdam, Netherlands

Corresponding Author:
Eric L Voorn, PhD
Amsterdam University Medical Center location University of Amsterdam
Rehabilitation Medicine
Meibergdreef 9
Amsterdam, 1105 AZ
Netherlands
Phone: 31 20 56 64421
Email: e.l.voorn@amsterdamumc.nl

Abstract

Background: Home-based aerobic exercise in people with neuromuscular diseases (NMDs) has benefits compared to exercise in the hospital or a rehabilitation center because traveling is often cumbersome due to mobility limitations, and societal costs are lower. Barriers to home-based aerobic exercise include reduced possibilities for monitoring and lack of motivation. To overcome these and other barriers, we developed a mobile health app: Keep on training with ReVi (hereafter referred to as ReVi).

Objective: We aimed to determine the usability of the ReVi app.

Methods: Patients followed a 4-month, polarized, home-based aerobic exercise program on a cycle or rowing ergometer, with 2 low-intensity sessions and 1 high-intensity session per week supported by the ReVi app. The app collected training data, including heart rate and ratings of perceived exertion, provided real-time feedback on reaching target intensity zones, and enabled monitoring via an online dashboard. Physiotherapists instructed patients on how to use the ReVi app and supervised them during their training program. Patients and physiotherapists separately evaluated usability with self-developed questionnaires, including 9 questions on a 5-point Likert scale, covering the usability elements efficiency, effectiveness, and satisfaction.

Results: Twenty-nine ambulatory adult patients (n=19 women; mean age 50.4, SD 14.2 years) with 11 different slowly progressive NMDs participated. Both patients and physiotherapists (n=10) reported that the app, in terms of its efficiency, was easy to use and had a rapid learning curve. Sixteen patients (55%) experienced 1 or more technical issue(s) during the course of the exercise program. In the context of effectiveness, 23 patients (81%) indicated that the app motivated them to complete the program and that it helped them to exercise within the target intensity zones. Most patients (n=19, 70%) and physiotherapists (n=6, 60%) were satisfied with the use of the app. The median attendance rate was 88% (IQR 63%-98%), with 76% (IQR 69%-82%) of time spent within the target intensity zones. Four adverse events were reported, 3 of which were resolved without discontinuation of the exercise program.

Conclusions: The usability of the ReVi app was high, despite the technical issues that occurred. Further development of the app to resolve these issues is warranted before broader implementation into clinical practice.

(JMIR Hum Factors 2024;11:e49808) doi:10.2196/49808

KEYWORDS

neuromuscular disorders; endurance training; home-based exercise; eHealth; tele-rehabilitation; app; exercise; aerobic exercise; mhealth; mobile app; neuromuscular disease; usability

Introduction

Physical fitness is an important health marker [1,2] and is strongly associated with daily life functioning [3] and independent living [4] at older age. People with neuromuscular disease (NMD) often have reduced physical fitness caused not only by the underlying disease but also by an inactive lifestyle [5-7]. Aerobic exercise is an important aspect of rehabilitation treatment for NMD, as it contributes to improved physical fitness [8]. The integration of exercise programs into everyday life was recently identified as one of the major research priorities for individuals with NMD [9].

People with NMD usually perform their aerobic exercise program in a hospital, rehabilitation center, or physiotherapy practice under the direct supervision of a physiotherapist. However, center-based exercise may be cumbersome for individuals with NMD, who are often limited in their mobility. Moreover, center-based exercise requires the availability of physiotherapy staff, whose number is often limited as many countries are reducing health care services [10,11]. This amplifies the need for alternative modes of exercise intervention delivery that maintain high quality and effectiveness [12,13].

Transferring aerobic exercise from the hospital environment to the home or community may be a beneficial way to reduce travel time and societal costs. A recently developed training guide called B-FIT is an example of a home-based aerobic exercise program specifically developed for NMD [14]. Feasibility of the B-FIT exercise program has been demonstrated for different types of NMD, and patients and physiotherapists were satisfied with its use [14]. A barrier to use B-FIT was that some patients experienced the program as insufficiently challenging. This requires attention because poor motivation has been reported as a major barrier to exercise in people with NMD [7,15]. Furthermore, physiotherapists perceived initiation of the program as time-consuming; most of the worksheets, including exercise testing results and the training schedules, needed to be filled out by hand. A more general concern regarding exercise in the home environment is the reduced possibility to monitor exercise sessions. This is particularly important for the vulnerable population of people with NMD, as it may put them at risk for under- and overtraining.

To overcome these barriers to home-based aerobic exercise for people with NMD, we developed a mobile health (mHealth) app called Keep on training with ReVi (hereafter referred to as the ReVi app). The ReVi app aims to improve patients' adherence to the B-FIT exercise program by (1) offering a structured exercise program, (2) providing insight into training progression, and (3) improving motivation through auditory encouragement. For physiotherapists, the ReVi app aims to improve their opportunities for supervision by enabling them to monitor progress and provide feedback from a distance and also to reduce the time investment to initiate the exercise program.

The primary aim of this study was to assess the usability of the ReVi app for assisting and monitoring home-based aerobic exercise according to the B-FIT training guide in people with NMD. We also evaluated the attendance rate, the time spent within target intensity zones, and the occurrence of adverse effects.

Methods

Design

A multicenter prospective pilot study was conducted at the outpatient departments of rehabilitation medicine of 2 university hospitals and in 3 rehabilitation centers in the Netherlands. All centers were specialized in treatment of NMD. This study included 2 different cohorts; in one cohort, the ReVi app was applied as part of usual care at the Department of Rehabilitation Medicine at the Amsterdam UMC, location Amsterdam Medical Center (AMC). The other cohort consisted of patients using the ReVi app in the intervention group of an ongoing multicenter randomized controlled trial on the efficacy of a physical activity program, which combines the B-FIT aerobic exercise program and motivational interviewing coaching to improve physical fitness in people with NMD [16].

Ethical Considerations

The medical ethics review committee of the AMC waived the need for medical ethical approval for the usual care cohort, and approved the study protocol of the randomized controlled trial (NL62104.018.17). All patients provided informed consent.

Participants

The inclusion criteria applied to both cohorts were (1) diagnosis of a slowly progressive NMD, (2) age ≥18 years, and (3) possession of a smartphone or tablet. Exclusion criteria were (1) contraindication for being physically active, (2) inability to follow verbal or written instructions, and (3) insufficient competence in the Dutch language. In addition, patients in the randomized controlled trial had to be motivated to improve their reduced physical fitness and were excluded if they had participated in an exercise program for a period longer than 4 weeks in the past 6 months. For the purpose of this study, we included only data of patients who completed at least 12 of the 48 possible training sessions, to ensure sufficient experience with the use of the ReVi app to evaluate its usability. We aimed to include a total of 30 patients in this study.

Physiotherapists were included in the study if they supervised at least 1 patient. Physiotherapists that were already exposed to the B-FIT training guide followed a half-day training course to refresh their knowledge on the use of the B-FIT training program and learn the use of the ReVi app. Physiotherapists that were not exposed to the B-FIT training guide followed a full-day training course to learn both the B-FIT training program and the use of the app. Furthermore, they received an instruction manual with a step-by-step guide on the use of the app.

ReVi App

The ReVi app (Amsterdam UMC) was built by a company (everywhereIM BV) specialized in the development of medical apps. The app was available for iOS and Android and it was developed in the Dutch language. An expert group consisting of physiotherapists, rehabilitation physicians, exercise physiologists, patients with different types of NMD, and representatives of the Dutch Society of Muscle Diseases and of the app builder actively participated in the development of the ReVi app. Expert group meetings were organized to discuss the aims of the app, to identify essential functionalities, and to provide feedback on so-called functional designs (on paper). The primary objective during this initial developmental phase was to create an app to assist a 16-week aerobic exercise regimen. If the study yields favorable results, the next developmental stage will be initiated to enhance the app's functionality and further explore the possibilities for offering longer-term support to home-based aerobic exercise. The data protection officer of Amsterdam UMC (location AMC) was also involved in the app's development process to ensure that personal data processing was organized in accordance with the General Data Protection Regulation (GDPR).

B-FIT Aerobic Exercise Program

The ReVi app was programmed with the B-FIT aerobic exercise program. This 16-week, polarized, home-based exercise program consisted of 2 low-intensity sessions below the anaerobic threshold (AT) and 1 high-intensity session above the AT per week. Patients visited the study center prior to the start, midway through, and after completion of the exercise program for a face-to-face meeting with their supervising physiotherapist. During each visit, an exercise test was executed. During the visits midway through and after completion of the exercise program, patients received feedback on training progress based on exercise testing results and based on data in the ReVi dashboard (see section App Description).

In the usual care cohort, target intensity zones were based on indirect assessment of the AT using ratings of perceived exertion (RPEs) during a submaximal exercise test [17]. In the randomized controlled trial cohort, target intensity zones were based on direct assessment of the AT during an exercise test through visual inspection of the gas exchange plots using the V-slope method [18]. If training based on heart rate was not feasible, for instance in patients using β-blocking agents, training was based on RPEs using the 6-20 Borg scale. Each training session consisted of several exercise intervals interspersed with recovery periods. Training sessions were performed in the home environment (eg, at home, in the gym, or at a physiotherapy practice) on a bicycle or rowing ergometer. A more detailed description of the B-FIT aerobic exercise program can be found in Multimedia Appendix 1 [14].

App Description

Physiotherapists created a personal account for a web-based dashboard that was used to create and manage ReVi app accounts of patients they supervised. The dashboard could be accessed using a desktop or laptop computer. Two-way verification using Google Authenticator (Google Inc) was required to sign in. The physiotherapists created patient accounts by sending a link to the patients' email addresses. Via this link, a password was created. Patients used the ReVi app on a mobile phone or tablet. Logging in to the ReVi app required their personal email address and password.

After signing in to the ReVi app, the home menu opened, from which 2 menus could be chosen: the Settings menu and the Training menu, which provided an overview of the program (Figure 1). Through the Settings menu the type of training could be chosen: training based on heart rate or based on Borg scale. For training based on heart rate, a Bluetooth connection with a heart rate monitor was established (in these cohorts, the device was the Polar H10; Polar Electro) and could be tested. Additionally, contact details of the physiotherapist were entered to enable patients to contact their therapists via the ReVi app.

The Training menu provided an overview of the training sessions (Figure 2). By selecting training sessions, the training protocol, including exercise intervals and recovery periods, was shown (Figure 3). During training sessions, the ReVi app guided users by illustrating their target intensity zones. In case of heart rate–based training, a heart rate chest strap was provided to the patient. The app was Bluetooth connected to the heart rate chest strap to continuously monitor heart rate (Figure 4). Patients rated their perceived exertion every final minute of the exercise interval or recovery period using the 6-20 Borg scale (Figure 5). During RPE-based training, patients rated their perceived exertion every minute. The ReVi app provided auditory feedback during training sessions. When patients trained within the target intensity zone, they were encouraged to continue. If the heart rate or Borg scale was not within the target intensity zone for at least 20 seconds, the ReVi app provided auditory instructions to increase or decrease the resistance. Directly after completion of the exercise session, an overview of the exercise results was shown (Figure 6).

Heart rate and Borg score data were saved by the ReVi app and sent to the web-based dashboard. Physiotherapists could access the training data of the patients they supervised; patients only had access to their own exercise data. The dashboard included, for each training session, a table with the percentage of time spent within the target intensity zones, the average heart rate for each exercise interval and recovery period, and the accompanying RPE (Figure 7). Additionally, a graph illustrated the actual heart rate or RPE with reference to the target intensity zones.

Figure 1. Screenshot of the ReVi app home screen.

Figure 2. Screenshot of the exercise program overview.

Figure 3. Screenshot of the exercise session protocol; this is an overview of the intensity and duration of each exercise or recovery bout.

Figure 4. Screenshot of the exercise session live screen; the actual achieved intensity (heart rate or rating of perceived exertion) and the target intensity zone during the exercise session are shown.

Figure 5. Screenshot of the Borg scale.

Figure 6. Screenshot of the exercise session results; the graph shows heart rate progression over time and the percentage of time spent within the target intensity zones.

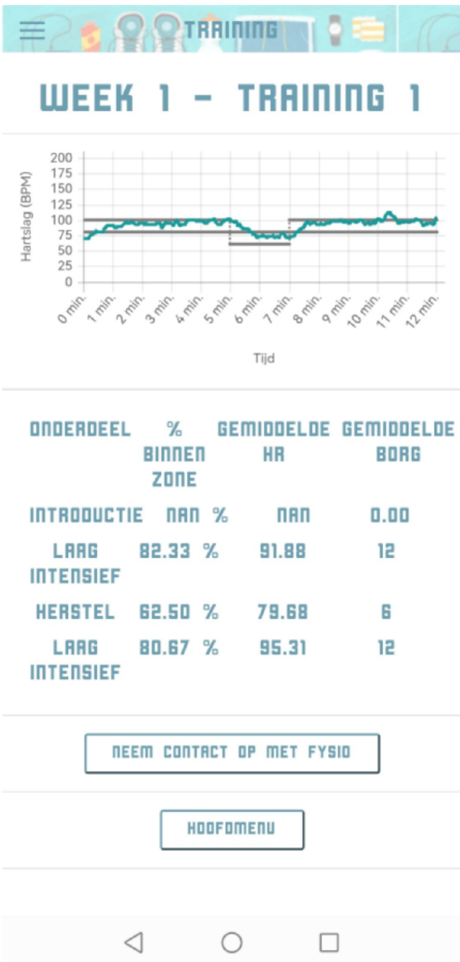
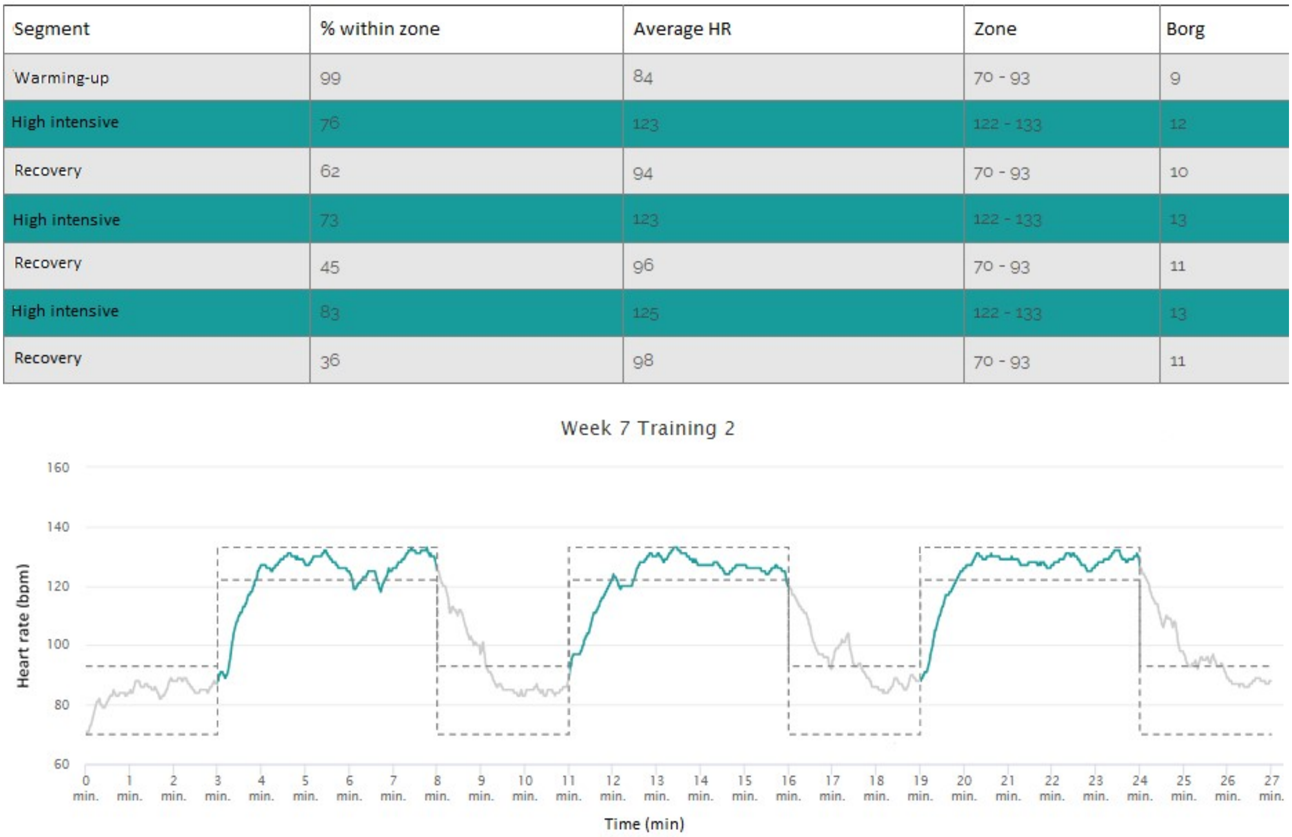


Figure 7. Screenshot of the ReVi app web-based dashboard. The percentage of time spent within the target intensity zones is presented in the table. The graph shows the heart rate during the training session (solid line), as well as the target intensity zones (grey dashed lines). The ReVi app dashboard is only available in Dutch, but for this paper the screenshot was translated to English. HR: heart rate.



Segment	% within zone	Average HR	Zone	Borg
Warming-up	99	84	70 - 93	9
High intensive	76	123	122 - 133	12
Recovery	62	94	70 - 93	10
High intensive	73	123	122 - 133	13
Recovery	45	96	70 - 93	11
High intensive	83	125	122 - 133	13
Recovery	36	98	70 - 93	11

Outcomes

ReVi App Usability

The primary outcome was the usability of the ReVi app, defined according to the International Organization for Standardization (ISO) as follows: “Usability is the extent to which a product can be used by specified users to achieve specified goals with efficiency, effectiveness and satisfaction in a specified context of use” [19]. Efficiency refers to the resources expended in relation to the accuracy and completeness with which users achieve goals (eg, ease of use, learning time, and additional effort of using the ReVi app during training sessions). Effectiveness refers to the extent to which the ReVi app has completed its goals to motivate patients and support patients to train within the targeted heart rate zones. Satisfaction assesses positive or negative attitudes toward the use of the ReVi app [20].

Self-developed questionnaires were used to assess the usability of the ReVi app among patients and physiotherapists. The questionnaires were developed by the study team, which consisted of researchers, rehabilitation physicians, and a physiotherapist. The questionnaires were reviewed by 2 patients and another physiotherapist before the final version was developed. The questionnaires contained questions pertaining to the 3 major aspects of usability: efficiency, effectiveness, and satisfaction. The usability questionnaires for patients and physiotherapists included 12 and 13 questions, respectively, of which 2 were open questions (Multimedia Appendices 2 and 3). Nine of the closed questions were scored on a 5-point Likert scale (1=strongly disagree; 2=disagree; 3=neither agree nor disagree; 4=agree; 5=strongly agree). Patients filled in the questionnaire after their last completed training session; physiotherapists did so after completion by the last patient they supervised.

Attendance Rate and Time Within Target Intensity Zones

For assessing attendance rates and the time spent within target intensity zones, we used data collected in the ReVi app dashboard. The attendance rate was defined as the percentage of followed training sessions. From the followed training sessions, we determined the percentage of time spent within target intensity zones for low- and high-intensity exercise intervals combined and separately.

Adverse Events

Adverse events related to the exercise program, such as severe muscle fatigue, joint pain, or muscle pain, were recorded. Patients were instructed to contact the physiotherapist to report adverse events. In addition, physiotherapists checked for adverse events during each patient visit.

Data Analysis

Descriptive statistics are used to present patient and physiotherapist characteristics. The data from the questions that were scored on a 5-point Likert scale were reduced by combining “agree” and “strongly agree” responses to form an “agree” category, and response options of “strongly disagree” and "disagree" were combined to form "disagree." Frequencies were calculated on the basis of the total number of responses to each question on the usability questionnaire and expressed as percentages. Data analysis was performed using SPSS (version 28.0; IBM Inc).

Results

Study Group

Between January 2020 and November 2021, 23 patients started their exercise program as part of the usual care cohort, of which 20 patients were included in the study. Three patients were excluded because they executed less than 12 exercise sessions. Reasons included technical problems with the ReVi app (n=1), medical issues (n=1), and a lack of motivation (n=1). Nine other patients participating in the ongoing randomized controlled trial were also included and started between July 2021 and December 2021.

Patient characteristics are shown in Table 1. Twenty-three patients were treated at the outpatient clinic of the Department of Rehabilitation of the Amsterdam UMC (location AMC), supervised by 6 physiotherapists. The other 6 patients were treated by 4 physiotherapists at Rehabilitation Center Klimmendaal (Arnhem; n=2), Basalt Rehabilitation Center (Leiden; n=2), University Medical Center Utrecht (Utrecht; n=1), and Sint Maartenskliniek (Nijmegen; n=1). Twenty-eight patients trained based on heart rate and 1 patient based on the Borg scale. Twenty-seven patients performed the exercise program using a bicycle ergometer and 2 patients used a rowing ergometer.

Table 1. Respondent profile.

Characteristics	Respondents
Patients (n=29)	
Age (years), mean (SD)	50.4 (14.2)
Female, n (%)	19 (66)
Sum score of manual muscle testing for the legs[a], median (range)	75 (60-80)
Peak workload baseline submaximal exercise test (watts), median (range)	100 (50-210)
Types of neuromuscular disorder, n	
Charcot-Marie-Tooth disease	7
Myotonic dystrophy	4
Nonspecific myopathy	4
Congenital dystrophy	3
Limb girdle muscular dystrophy	3
Mitochondrial myopathy	2
Inclusion body myositis	2
Becker muscular dystrophy	1
Postpolio syndrome	1
Dermatomyositis	1
Chronic inflammatory demyelinating polyradiculoneuropathy	1
Physiotherapists (n=10)	
Female, n (%)	7 (70)
Patients supervised in study (n), median (range)	2 (1-9)
Prior experience with use of the ReVi app, n (%)	4 (40)

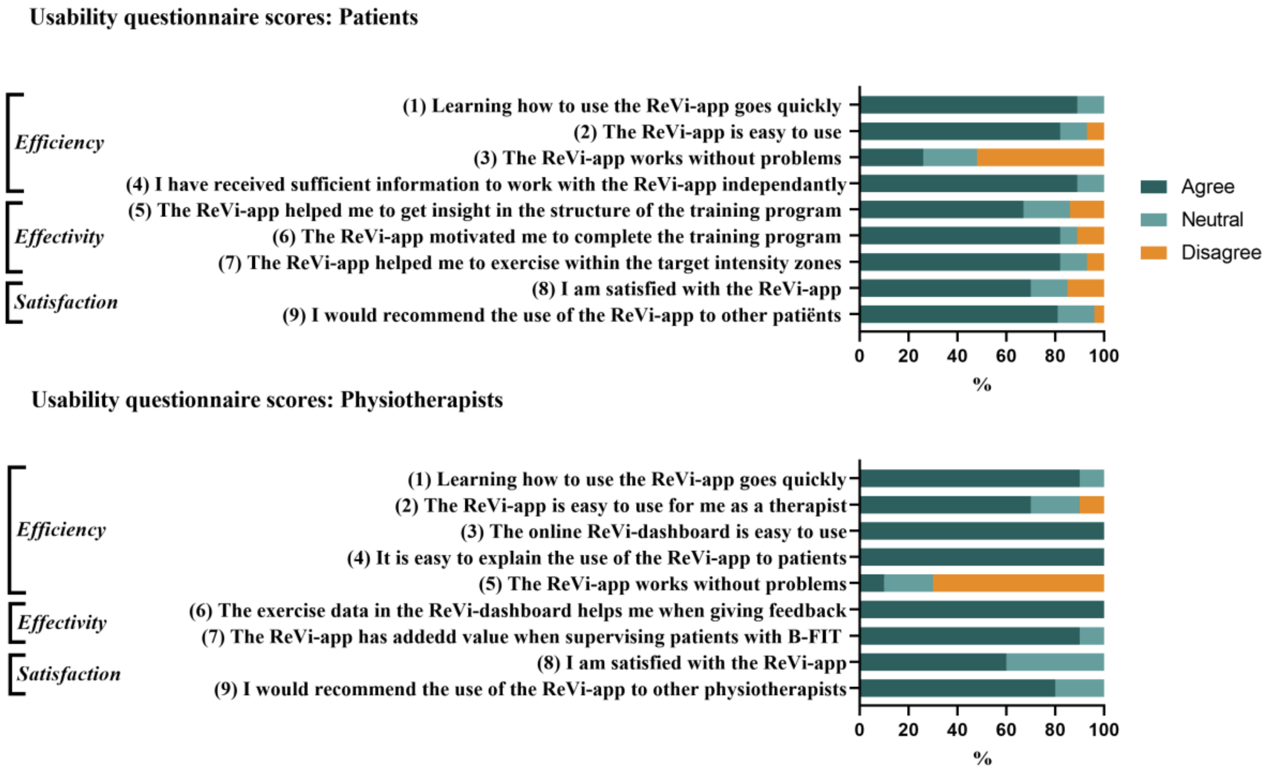
[a]Sum score for muscle strength of the legs was calculated by adding 16 muscle groups. Each muscle group had a score between 0 and 5, and the sum score ranged from 0 to 80 [21].

Primary Outcome

Usability

Twenty-seven patients and all 10 physiotherapists filled in and returned the usability questionnaire. Two patients did not return the usability questionnaire despite multiple requests. Questionnaire scores of patients and physiotherapists are presented in Figure 8.

Figure 8. Statements and outcomes for the patient usability questionnaire (n=27) and the physiotherapist usability questionnaire (n=10). Scores are given on a Likert scale, ranging from 1-5 (1 = strongly disagree, 5 = strongly agree). Frequency data were reduced by combining “agree” and “ strongly agree” responses to form an “agree” category, and response options of “strongly disagree” and “disagree” were combined to form “disagree”.



Efficiency

Twenty-four patients (89%) reported that learning how to use the ReVi app went quickly and 22 patients (81%) found that the ReVi app was easy to use. Seven patients (26%) agreed with the statement “the ReVi app works without problems.” In 16 of the total of 29 patients (55%), 1 or more technical issues occurred during the course of the ReVi app training program. The most-reported technical issues were connection problems with the heart rate monitor and a bug in the app that hindered saving of exercise data in week 11 of the exercise program.

The majority of therapists reported that the ReVi app was easy to use (n=7, 70%) and all therapists found the use of the app easy to explain to patients. Nine therapists (90%) experienced technical issues using the ReVi app.

Effectiveness

Eighteen patients (67%) reported that the ReVi app provided insight into the structure of the exercise program. Twenty-two patients (81%) agreed that the app motivated them to complete the program and that it helped them to maintain exercise within the target intensity zones.

All therapists reported that the web-based dashboard helped them to provide feedback to patients and that the ReVi app had added value for supervision. The most important benefit reported by the physiotherapists was that the ReVi app allowed insight into the number of sessions that were followed and the exercise intensity that was achieved during training sessions.

Satisfaction

Nineteen patients (70%) were satisfied with the use of the ReVi app and 22 patients (81%) would recommend the use of the ReVi app to other patients with NMD. The most important reasons to recommend its use to others were that the app provided structure, helped them to train within the target intensity zones, and motivated them to complete their training sessions. The most-reported reason for patients not to recommend the ReVi app to others was the occurrence of technical issues.

Six therapists (60%) reported that they were satisfied with the use of the ReVi app and 8 therapists (80%) would recommend the use of the app to other physiotherapists. Reasons to recommend its use to others were that it was easy to use, enabled monitoring from a distance, and provided data that could be used to give tailored feedback to patients. The most-reported reason to not recommend the ReVi app to others was the technical issues that occasionally occurred when using the app.

Secondary Outcomes

Attendance and Time Within Target Intensity Zones

Twenty of the 29 patients (69%) completed the exercise program. Reasons for discontinuation among the other 9 patients were technical problems with the ReVi app (n=4), medical issues (n=2), closing of the local gym due to COVID-19 measures (n=2), and a lack of motivation (n=1).

Figure 9 shows the attendance rate for each patient, as well as the time spent within the target intensity zones. The median attendance rate was 88% (IQR 63%-98%). During the attended training sessions, patients spent a median of 76% (IQR 69%-82%) of the time within their target intensity zones (Figure 10). The median percentage of time spent within the low intensity zones was 85% (IQR 81%-92%), and in the high intensity zones it was 59% (IQR 45%-70%).

Figure 9. Attendance rates for individual patients ordered from most to least sessions, and the percentage of time spent within the target intensity zones during corresponding sessions. * patient trained based on Borg scale.

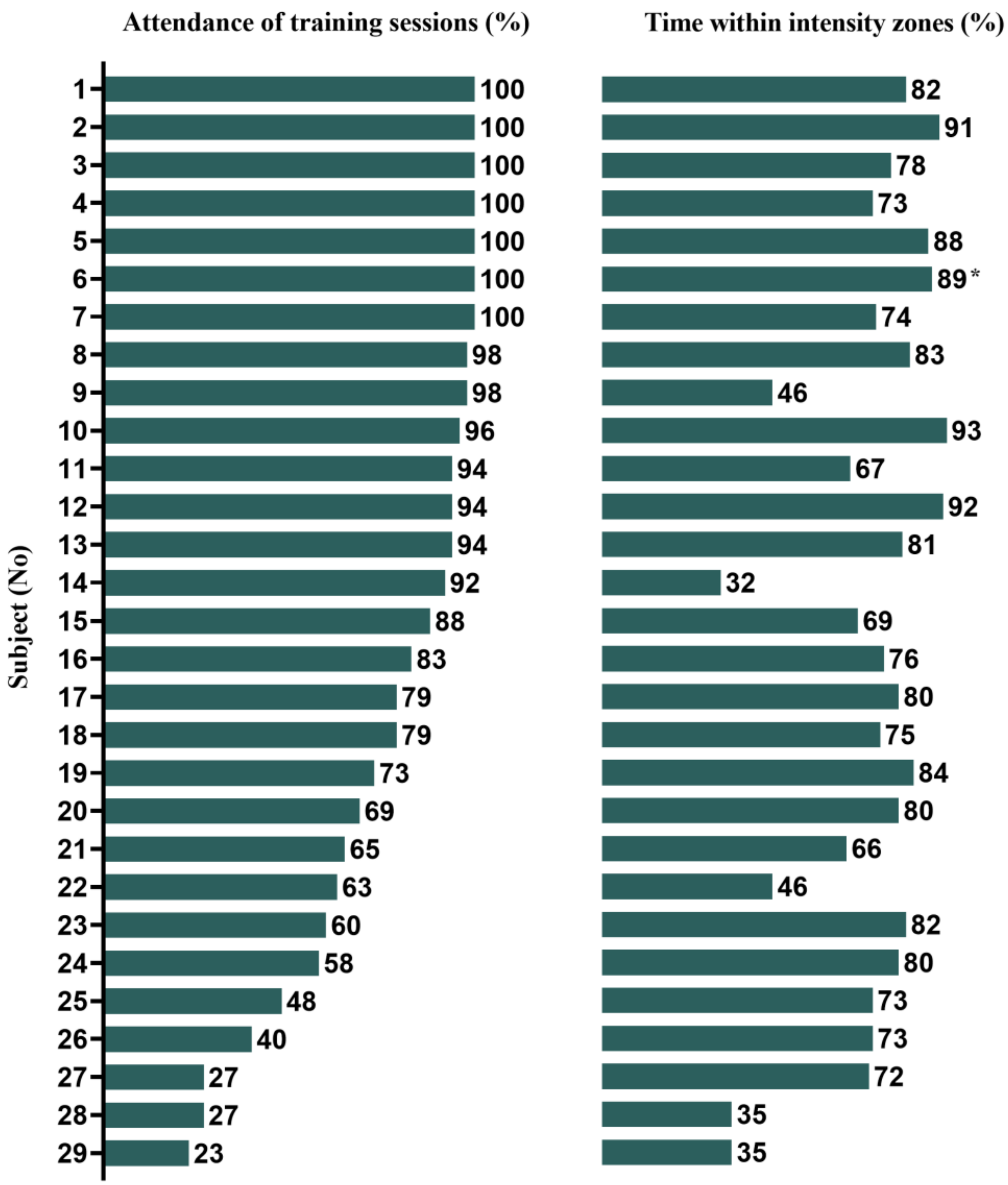
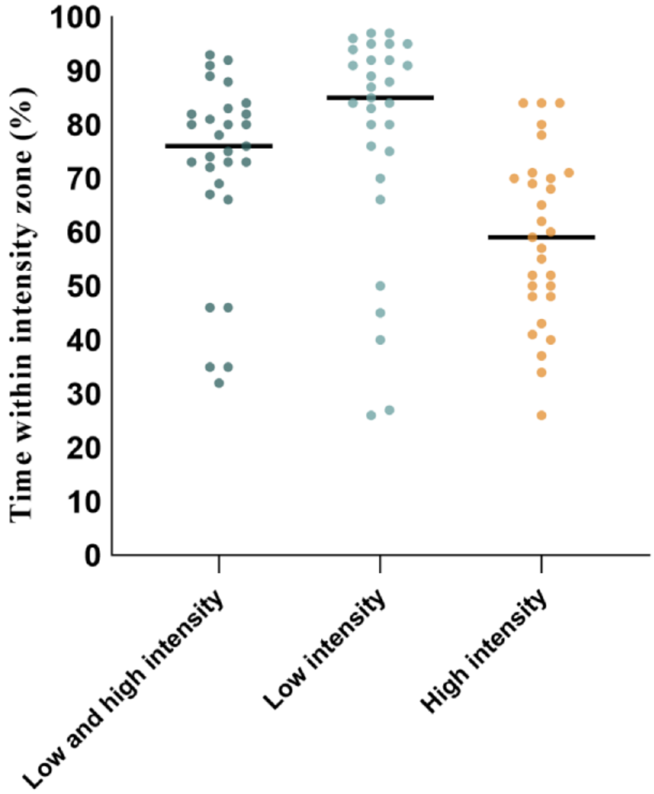


Figure 10. The percentage of time spent within the target intensity zones during A) both high and low intensity exercise intervals, B) low intensity exercise intervals and C) high intensity exercise intervals. Black lines indicate the median. Each dot represents a single patient.



Adverse Events

Four adverse events were reported: fatigue (n=2), knee joint pain (n=1), and high blood pressure during training (n=1). In the patient with high blood pressure during training, the rehabilitation physician and physiotherapist decided to terminate the exercise program. The other 3 adverse events were resolved without discontinuation of the exercise program.

Discussion

Principal Findings

This study provides insight into the usability of the ReVi app among people with NMD to support home-based aerobic exercise according to the B-FIT training program. The different components of usability, including efficiency, effectiveness, and satisfaction, were all judged as good by physiotherapists and patients, despite the occurrence of technical issues.

Patients were generally positive about the efficiency of the ReVi app due to its rapid learning curve and ease of use. Patients could independently work with the app based on the instructions that they received from their treating physiotherapist. Adequate instructions are known to be a key facilitator of patient engagement with mHealth apps [22]. With regards to its effectiveness, patients reported that the most important goals of the ReVi app were achieved: its use motivated them to complete the exercise program and helped them to exercise within their target intensity zones. These outcomes were supported by the findings that patients attended the majority of training sessions and spent most time within the target intensity zones. Patients were mostly satisfied with the use of the app, which concurs with other studies on apps supporting home-based physical exercise programs in amyotrophic lateral sclerosis, which is a rapidly progressive type of NMD [23], and a variety of other patient populations [24-26].

Physiotherapists were positive about the efficiency of the ReVi app. This was mainly due to the rapid learning curve and its ease of use; a half- or full-day training course was required for physiotherapists to learn how to work with the app and the B-FIT training guide, depending on prior experience with B-FIT. In terms of effectiveness, physiotherapists reported that the most important goals of the ReVi app were achieved. They found the app helpful when monitoring patients during their home-based program, mainly because it enabled them to provide feedback based on exercise data. They were generally satisfied with the use of the ReVi app and would recommend the use of the app to other physiotherapists.

While efficiency, effectiveness, and user satisfaction were overall judged as positive, one of the efficiency items was clearly judged as insufficient: 55% (n=16) of the patients and 90% (n=9) of the physiotherapists experienced technical issues. Most of these issues were solved, but in all cases, this required the help of a physiotherapist, researcher, or software developer. Technical issues are known to negatively impact usability and decrease adherence and engagement with mHealth tools [22]. They often cause patients to stop their mHealth interventions, leading to high dropout rates, and they are reported as a main barrier to further implementation of mHealth or eHealth apps [22,27]. This is consistent with our finding that the most important reason for discontinuation of the exercise program was when the ReVi app did not function well. Therefore, resolving technical issues is an important concern for further implementation of the ReVi app in clinical rehabilitation practice on a broader scale. This also underlines the importance of offering technical support when using mHealth tools such as the ReVi app [28]. Physiotherapists could play an important role in this, but that would require sufficient proficiency with mHealth. Moreover, considering the limited availability of physiotherapy personnel, it is essential for successful implementation of mHealth tools like the ReVi app to minimize technical issues and provide access to additional technical support for more complex problems.

The attendance rate, time within target intensity zones, and adverse events found in this study suggest that training in the home environment with the help of the ReVi app is a good alternative to center-based training. The attendance rate of 88% and time within target intensity zones of 76% are in line with adherence rates found in other studies evaluating aerobic exercise programs for NMD [29-37] that were mostly conducted in a hospital or rehabilitation center. Comparison of the attendance rate and time within target intensity zones between this study and past studies on exercise for NMD is hampered by incomplete or absent descriptions of adherence assessment methods in most other studies. In some studies, it is unclear if reported values are for attendance rates, the time spent within exercise zones, or training time. Moreover, some studies excluded patients who dropped out, leading to overestimated adherence. In this light, the attendance rate in our study may have been impacted by excluding patients who performed less than 12 exercise sessions and by the finding that some patients performed several training sessions without using the ReVi app. Despite these uncertainties, the attendance rate and time within target intensity zones found in our study seem to be in line with values reported in other aerobic exercise studies. The limited number of adverse events reported in this study also concurs with other studies on center-based aerobic exercise programs for NMD [32,38,39]. This further strengthens the notion that home-based aerobic exercise supported by the ReVi app may be considered a safe and feasible alternative for center-based exercise programs, which is in line with earlier research in telemonitoring of home-based exercise for amyotrophic lateral sclerosis [23].

Limitations

Patients with a positive attitude toward the use of mHealth may have been more inclined to participate in this study, causing selection bias and limiting generalizability to people with NMD and less affinity for mHealth. Also, most patients trained under supervision of a physiotherapist experienced in treating patients with NMD, which limits generalizability of our results to other health care settings, such as primary care physiotherapy practices.

Future Studies

Implementation of mHealth, such as with the ReVi app, in rehabilitation care presents some major challenges, such as the comfort of patients and therapists with the use of technology, legal and ethical considerations regarding patient monitoring and the protection of privacy rights, and integration of mHealth tools into current working protocols [40,41]. Additionally, specific application design requirements have to be considered for NMD patients who experience reduced hand functionality due to muscle weakness. These requirements may include sufficiently large buttons and input fields. As a consequence of these challenges, the scientific literature on telerehabilitation in NMD patients is still limited [42,43]. To enable the broader implementation of mHealth in clinical practice, research is warranted into other facilitators of and barriers to the implementation of mHealth specific to neuromuscular rehabilitation.

Conclusions

The usability of the ReVi app in terms of perceived efficiency, effectiveness, and user satisfaction is high, despite the occurrence of technical issues. Combined with the high attendance rate and time spent within target intensity zones and low number of adverse events, the ReVi app can be considered a promising tool to support home-based aerobic exercise in rehabilitation practice for NMD. Further development of the ReVi app to resolve technical issues is warranted before broader implementation into clinical rehabilitation practice.

Acknowledgments

We thank all participating patients and physiotherapists. This research project was funded by IPSEN Pharmaceutics BV.

Conflicts of Interest

None declared.

Multimedia Appendix 1
B-FIT exercise program.
[DOCX File , 110 KB - humanfactors_v11i1e49808_app1.docx]

Multimedia Appendix 2
Patient usability questionnaire.
[DOCX File , 36 KB - humanfactors_v11i1e49808_app2.docx]

Multimedia Appendix 3
Therapist usability questionnaire.
[DOCX File , 35 KB - humanfactors_v11i1e49808_app3.docx]

References

1. Eriksen L, Curtis T, Grønbæk M, Helge JW, Tolstrup JS. The association between physical activity, cardiorespiratory fitness and self-rated health. Prev Med 2013 Dec;57(6):900-902. [doi: 10.1016/j.ypmed.2013.09.024] [Medline: 24099878]
2. Ortega FB, Ruiz JR, Castillo MJ, Sjöström M. Physical fitness in childhood and adolescence: a powerful marker of health. Int J Obes (Lond) 2008 Jan;32(1):1-11. [doi: 10.1038/sj.ijo.0803774] [Medline: 18043605]
3. Arnett SW, Laity JH, Agrawal SK, Cress ME. Aerobic reserve and physical functional performance in older adults. Age Ageing 2008 Jul;37(4):384-389. [doi: 10.1093/ageing/afn022] [Medline: 18287178]
4. Cress ME, Meyer M. Maximal voluntary and functional performance levels needed for independence in adults aged 65 to 97 years. Phys Ther 2003 Jan;83(1):37-48. [Medline: 12495411]
5. Aitkens S, Kilmer DD, Wright NC, McCrory MA. Metabolic syndrome in neuromuscular disease. Arch Phys Med Rehabil 2005 May;86(5):1030-1036. [doi: 10.1016/j.apmr.2004.09.012] [Medline: 15895353]
6. McDonald CM. Physical activity, health impairments, and disability in neuromuscular disease. Am J Phys Med Rehabil 2002 Nov;81(11 Suppl):S108-S120. [doi: 10.1097/00002060-200211001-00012] [Medline: 12409816]
7. Phillips M, Flemming N, Tsintzas K. An exploratory study of physical activity and perceived barriers to exercise in ambulant people with neuromuscular disease compared with unaffected controls. Clin Rehabil 2009 Aug;23(8):746-755. [doi: 10.1177/0269215509334838] [Medline: 19506002]
8. Cup EH, Pieterse AJ, Ten Broek-Pastoor JM, Munneke M, van Engelen BG, Hendricks HT, et al. Exercise therapy and other types of physical therapy for patients with neuromuscular diseases: a systematic review. Arch Phys Med Rehabil 2007 Nov;88(11):1452-1464. [doi: 10.1016/j.apmr.2007.07.024] [Medline: 17964887]
9. Nierse CJ, Abma TA, Horemans AMC, van Engelen BGM. Research priorities of patients with neuromuscular disease. Disabil Rehabil 2013 Mar;35(5):405-412. [doi: 10.3109/09638288.2012.694964] [Medline: 22747255]
10. Ratcliffe J, Laver K, Couzner L, Cameron ID, Gray L, Crotty M. Not just about costs: the role of health economics in facilitating decision making in aged care. Age Ageing 2010 Jul;39(4):426-429. [doi: 10.1093/ageing/afq041] [Medline: 20444804]
11. World report on ageing and health. World Health Organization. 2015. URL: https://www.who.int/publications/i/item/9789241565042 [accessed 2024-02-27]
12. Lai B, Young HJ, Bickel CS, Motl RW, Rimmer JH. Current trends in exercise intervention research, technology, and behavioral change strategies for people with disabilities: A scoping review. Am J Phys Med Rehabil 2017 Oct;96(10):748-761. [doi: 10.1097/PHM.0000000000000743] [Medline: 28398967]
13. Lai B, Kim Y, Wilroy J, Bickel CS, Rimmer JH, Motl RW. Sustainability of exercise intervention outcomes among people with disabilities: a secondary review. Disabil Rehabil 2019 Jun;41(13):1584-1595. [doi: 10.1080/09638288.2018.1432704] [Medline: 29409367]
14. Voorn EL, Koopman FS, Nollet F, Brehm MA. Individualized aerobic exercise in neuromuscular diseases: a pilot study on the feasibility and preliminary effectiveness to improve physical fitness. Phys Ther 2021 Mar 03;101(3):pzaa213 [FREE Full text] [doi: 10.1093/ptj/pzaa213] [Medline: 33332538]
15. Essery R, Geraghty AWA, Kirby S, Yardley L. Predictors of adherence to home-based physical therapies: a systematic review. Disabil Rehabil 2017 Mar;39(6):519-534. [doi: 10.3109/09638288.2016.1153160] [Medline: 27097761]
16. Oorschot S, Brehm MA, van Groenestijn AC, Koopman FS, Verhamme C, Eftimov F, et al. Efficacy of a physical activity programme combining individualized aerobic exercise and coaching to improve physical fitness in neuromuscular diseases (I'M FINE): study protocol of a randomized controlled trial. BMC Neurol 2020 May 13;20(1):184 [FREE Full text] [doi: 10.1186/s12883-020-01725-0] [Medline: 32404133]
17. Voorn EL, Gerrits KH, Koopman FS, Nollet F, Beelen A. Determining the anaerobic threshold in postpolio syndrome: comparison with current guidelines for training intensity prescription. Arch Phys Med Rehabil 2014 May;95(5):935-940. [doi: 10.1016/j.apmr.2014.01.015] [Medline: 24491465]
18. Wasserman K, Hansen J, Sue D, Stringer W, Whipp B. Principles of Exercise Testing and Interpretation, Including Pathophysiology and Clinical Applications. Philadelphia, PA: Lippincott Williams & Wilkins; 2005.
19. Ergonomic requirements for office work with visual display terminals (VDTs) - Visual display requirements. British Standards Institution. 1998. URL: https://knowledge.bsigroup.com/products/ergonomic-requirements-for-office-work-with-visual-display-terminals-vdts-visual-display-requirements?version=standard&tab=preview [accessed 2024-02-05]
20. Harrison R, Flood D, Duce D. Usability of mobile applications: literature review and rationale for a new usability model. J Interact Sci 2013;1(1):1 [FREE Full text] [doi: 10.1186/2194-0827-1-1]
21. Secer H, Daneyemez M, Gonul E, Izci Y. Medical research council: aids to examination of the peripheral nervous system. UK Research and Innovation. 1976. URL: https://www.ukri.org/wp-content/uploads/2021/12/MRC-011221-AidsToTheExaminationOfThePeripheralNervousSystem.pdf [accessed 2024-02-27]

22. Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrão J, et al. Barriers to and facilitators of engagement with remote measurement technology for managing health: systematic review and content analysis of findings. J Med Internet Res 2018 Jul 12;20(7):e10480 [FREE Full text] [doi: 10.2196/10480] [Medline: 30001997]
23. Braga AC, Pinto A, Pinto S, de Carvalho M. Tele-monitoring of a home-based exercise program in amyotrophic lateral sclerosis: a feasibility study. Eur J Phys Rehabil Med 2018 Jun;54(3):501-503 [FREE Full text] [doi: 10.23736/S1973-9087.18.05129-8] [Medline: 29363923]
24. Arensman R, Kloek C, Pisters M, Koppenaal T, Ostelo R, Veenhof C. Patient perspectives on using a smartphone app to support home-based exercise during physical therapy treatment: qualitative study. JMIR Hum Factors 2022 Sep 13;9(3):e35316 [FREE Full text] [doi: 10.2196/35316] [Medline: 36098993]
25. Bizzarini E, Chittaro L, Frezza M, Polo M, Malisan C, Menosso R, et al. A mobile app for home-based exercise in spinal cord injured persons: Proposal and pilot study. Digit Health 2022;8:20552076211070724 [FREE Full text] [doi: 10.1177/20552076211070724] [Medline: 35140978]
26. Mahmood A, Blaizy V, Verma A, Stephen Sequeira J, Saha D, Ramachandran S, et al. Acceptability and attitude towards a mobile-based home exercise program among stroke survivors and caregivers: a cross-sectional study. Int J Telemed Appl 2019;2019:5903106 [FREE Full text] [doi: 10.1155/2019/5903106] [Medline: 31186627]
27. Hardiker NR, Grant MJ. Factors that influence public engagement with eHealth: A literature review. Int J Med Inform 2011 Jan;80(1):1-12 [FREE Full text] [doi: 10.1016/j.ijmedinf.2010.10.017] [Medline: 21112244]
28. Talboom-Kamp EP, Verdijk NA, Kasteleyn MJ, Numans ME, Chavannes NH. From chronic disease management to person-centered eHealth; a review on the necessity for blended care. Clin eHealth 2018 Mar;1(1):3-7 [FREE Full text] [doi: 10.1016/j.ceh.2018.01.001]
29. Koopman FS, Voorn EL, Beelen A, Bleijenberg G, de Visser M, Brehm MA, et al. No reduction of severe fatigue in patients with postpolio syndrome by exercise therapy or cognitive behavioral therapy: results of an RCT. Neurorehabil Neural Repair 2016 Jun;30(5):402-410. [doi: 10.1177/1545968315600271] [Medline: 26253175]
30. Voet N, Bleijenberg G, Hendriks J, de Groot I, Padberg G, van Engelen B, et al. Both aerobic exercise and cognitive-behavioral therapy reduce chronic fatigue in FSHD: an RCT. Neurology 2014 Nov 18;83(21):1914-1922. [doi: 10.1212/WNL.0000000000001008] [Medline: 25339206]
31. Sveen ML, Jeppesen TD, Hauerslev S, Køber L, Krag TO, Vissing J. Endurance training improves fitness and strength in patients with Becker muscular dystrophy. Brain 2008 Nov;131(Pt 11):2824-2831. [doi: 10.1093/brain/awn189] [Medline: 18776212]
32. Ernstoff B, Wetterqvist H, Kvist H, Grimby G. Endurance training effect on individuals with postpoliomyelitis. Arch Phys Med Rehabil 1996 Sep;77(9):843-848. [doi: 10.1016/s0003-9993(96)90268-3] [Medline: 8822672]
33. El Mhandi L, Millet GY, Calmels P, Richard A, Oullion R, Gautheron V, et al. Benefits of interval-training on fatigue and functional capacities in Charcot-Marie-Tooth disease. Muscle Nerve 2008 May;37(5):601-610. [doi: 10.1002/mus.20959] [Medline: 18335470]
34. Andersen G, Heje K, Buch AE, Vissing J. High-intensity interval training in facioscapulohumeral muscular dystrophy type 1: a randomized clinical trial. J Neurol 2017 Jun;264(6):1099-1106. [doi: 10.1007/s00415-017-8497-9] [Medline: 28470591]
35. Andersen G, Prahm KP, Dahlqvist JR, Citirak G, Vissing J. Aerobic training and postexercise protein in facioscapulohumeral muscular dystrophy: RCT study. Neurology 2015 Aug 04;85(5):396-403. [doi: 10.1212/WNL.0000000000001808] [Medline: 26156512]
36. Orngreen MC, Olsen DB, Vissing J. Aerobic training in patients with myotonic dystrophy type 1. Ann Neurol 2005 May;57(5):754-757. [doi: 10.1002/ana.20460] [Medline: 15852373]
37. Sveen M, Jeppesen TD, Hauerslev S, Krag TO, Vissing J. Endurance training: an effective and safe treatment for patients with LGMD2I. Neurology 2007 Jan 02;68(1):59-61. [doi: 10.1212/01.wnl.0000250358.32199.24] [Medline: 17200494]
38. Alemo Munters L, Dastmalchi M, Andgren V, Emilson C, Bergegård J, Regardt M, et al. Improvement in health and possible reduction in disease activity using endurance exercise in patients with established polymyositis and dermatomyositis: a multicenter randomized controlled trial with a 1-year open extension followup. Arthritis Care Res (Hoboken) 2013 Dec;65(12):1959-1968 [FREE Full text] [doi: 10.1002/acr.22068] [Medline: 23861241]
39. Cejudo P, Bautista J, Montemayor T, Villagómez R, Jiménez L, Ortega F, et al. Exercise training in mitochondrial myopathy: a randomized controlled trial. Muscle Nerve 2005 Sep;32(3):342-350. [doi: 10.1002/mus.20368] [Medline: 15962332]
40. Galetsi P, Katsaliaki K, Kumar S. Exploring benefits and ethical challenges in the rise of mHealth (mobile healthcare) technology for the common good: An analysis of mobile applications for health specialists. Technovation 2023 Mar;121:102598 [FREE Full text] [doi: 10.1016/j.technovation.2022.102598]
41. DeRuyter F, Jones M, Morris J. Mobile Health and mobile rehabilitation for people with disabilities: current state, challenges and opportunities. In: Computers Helping People with Special Needs: 17th International Conference. 2020 Presented at: ICCHP 2020; September 9-11, 2020; Lecco, Italy p. 17. [doi: 10.1007/978-3-030-58805-2_26]
42. Giannotta M, Petrelli C, Pini A. Telemedicine applied to neuromuscular disorders: focus on the COVID-19 pandemic era. Acta Myol 2022 Mar;41(1):30-36 [FREE Full text] [doi: 10.36185/2532-1900-066] [Medline: 35465343]

43. Costamagna G, Abati E, Bresolin N, Comi GP, Corti S. Management of patients with neuromuscular disorders at the time of the SARS-CoV-2 pandemic. J Neurol 2021 May;268(5):1580-1591 [FREE Full text] [doi: 10.1007/s00415-020-10149-2] [Medline: 32804279]

Abbreviations

AMC: Amsterdam Medical Center
AT: anaerobic threshold
GDPR: General Data Protection Regulation
ISO: International Organization for Standardization
mHealth: mobile health
NMD: neuromuscular disease
RPE: rating of perceived exertion

Edited by A Kushniruk; submitted 12.06.23; peer-reviewed by Y Chu, G Ramdharry; comments to author 31.08.23; revised version received 21.09.23; accepted 20.01.24; published 15.03.24.

Please cite as:
Veneman T, Koopman FS, Oorschot S, Koomen PG, Nollet F, Voorn EL
A Mobile Health App to Support Home-Based Aerobic Exercise in Neuromuscular Diseases: Usability Study
JMIR Hum Factors 2024;11:e49808
URL: https://humanfactors.jmir.org/2024/1/e49808
doi:10.2196/49808
PMID:38488838

Short Paper

Designing a Novel Digitally Delivered Antiracism Intervention for Mental Health Clinicians: Exploratory Analysis of Acceptability

Tashalee Rushell Brown[1], MD, PhD; Habiba Amir[1]; Drew Hirsch[1]; Madeline Owens Jansen[1], MD, MPH

Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, Los Angeles, CA, United States

Corresponding Author:
Tashalee Rushell Brown, MD, PhD
Department of Psychiatry and Biobehavioral Sciences
University of California, Los Angeles
760 Westwood Plaza, 48-257
Los Angeles, CA, 90024
United States
Phone: 1 310 206 0188
Email: tashaleebrown@mednet.ucla.edu

Abstract

Background: There is a great need for evidence-based antiracism interventions targeting mental health clinicians to help mitigate mental health disparities in racially and ethnically minoritized groups.

Objective: This study provides an exploratory analysis of mental health clinicians' perspectives on the acceptability of a web-based antiracism intervention.

Methods: Mental health clinicians were recruited from a single academic medical center through outreach emails. Data were collected through individual 30-minute semistructured remote video interviews with participants, then recorded, transcribed, and analyzed using content analysis.

Results: A total of 12 mental health clinicians completed the study; 10 out of 12 (83%) were female candidates. Over half (7/12, 58%) of the respondents desired more robust antiracism training in mental health care. Regarding the web-based antiracism intervention, (8/12, 67%) enjoyed the digitally delivered demo module, (7/12, 58%) of respondents suggested web-based content would be further enhanced with the addition of in-person or online group components.

Conclusions: Our results suggest a strong need for additional antiracist training for mental health clinicians. Overall, participants responded favorably to novel web-based delivery methods for an antiracism intervention. These findings provide important support for future development and pilot testing of a large-scale digitally enhanced antiracist curriculum targeting mental health clinicians.

(JMIR Hum Factors 2024;11:e52561) doi:10.2196/52561

KEYWORDS

acceptability; antiracism; clinicians; intervention; interview study; mental health; psychiatry residents; racism; social workers; web-based technology

Introduction

Racism or expressions of discrimination are often rooted in implicit bias and stigmatizing beliefs [1]. Currently, racism is known to be a key driver of mental health inequities in ethnoracially minoritized groups who may be victims of discrimination [2]. Such experiences often lead to negative mental health outcomes [2]. Current evidence suggests that Black, Indigenous, and people of color youth and adults experience highly disproportionate rates of delayed diagnosis and treatment of autism spectrum disorder, overdiagnosis of conduct disorder, and underdiagnosis of attention-deficit/hyperactivity disorder [3], overdiagnosis of schizophrenia, overuse of antipsychotics with long-term medical consequences, and the underdiagnosis and treatment of depression [4]. Antiracism is the practice of actively opposing the effects of racism through institutional policies and individual behaviors [5]. Several recent systematic and scoping reviews on antiracism interventions in mental health professions have identified only one relevant randomized pilot study to date [5,6]. Of additional importance is that the authors found significant variability in training methodology, variability of intervention duration, and a lack of sufficient efficacy measurements to evaluate existent antiracism interventions [5,6]. Thus, despite the strong need for evidence-based antiracism interventions targeting racial bias among mental health clinicians, such interventions remain underdeveloped and understudied in the literature. Within this context, evidence-based strategies, such as those based in cognitive-behavioral frameworks, have shown promise in addressing prejudiced thoughts, feelings, and behaviors but have yet to be applied to clinicians [4]. Notably, the delivery of tailored psychoeducational content such as this, has the potential to be greatly enhanced by digital design and delivery methods [3,7]. This is especially poignant given that web-based technologies are known to further augment interventional implementation structures with regard to both flexibility and sustainability [3,7].

Against the backdrop of a profound dearth of evidence-based antiracism interventions targeting mental health professionals, this study aims to explore aspects of the acceptability of a novel digitally delivered intervention of this sort [7]. Grounded in a strongly evidenced implementation science framework and through a dynamic and iterative process of evaluation, we explored facets of intervention acceptability regarding content, delivery, and implementation strategies [8]. Semistructured interviews were designed to elicit additional perceptions and attitudes among mental health professionals regarding gaps and opportunities in their current training on antiracism. Findings have the potential to be incorporated into future modifications of the intervention in order to optimize the feasibility and acceptability of large-scale randomized control pilot trials.

Methods

Overview

Participants were residents, fellows, and social workers specializing in mental health care. They were recruited from a single academic medical center in California through a remote method, which included outreach emails. Written, informed consent was obtained from all participants. Participants were compensated through US $50 gift cards and water bottles. Data were collected through individual 30-minute semistructured remote video interviews with participants, which were recorded and transcribed for analysis. Semistructured interview questions were developed based on the clinical experience and literature review conducted by MOJ and TRB (Table 1).

The semistructured interview featured a presentation of a digital demo module of the cognitive behavioral therapy (CBT)–based intervention, which discussed core beliefs that may be harmful in the treatment of patients with mental health conditions. The module features real-world examples, teaches a key concept of intervention, presents examples of self-monitoring, and provides a visual outline of the engagement and reward components. Data were qualitatively analyzed using inductive coding and thematic analysis methods [9] using the Atlas.ti (Scientific Software Development GmbH) software by 2 independent coders (HA and DH). Identified codes and themes were reviewed and consolidated by the leading authors (MOJ and TRB) until consensus was achieved. The number of respondents mentioning each code or theme was reported.

Table 1. Semistructured interview questions and probes.

Domain	Questions	Follow-up probes
Sociodemographic information	• What is your profession? • Where are you in your training?	• How would you describe your race and gender?
Definitions and thoughts of antiracism	• How would you define antiracism? • What terms do you like to use to describe or discuss racism and antiracism?	• Are you comfortable talking with coworkers or supervisors about racism and antiracism?
Strengths of the current antiracism training	• What are the strengths of your medical training thus far with regard to antiracism?	• What training, educational tools or courses have you benefited from in medical school or at the postgraduate level?
Weaknesses of the current antiracism training	• What are the weaknesses of your medical training thus far with regard to antiracism?	• What additional support, educational tools, or resources would help elevate your clinical skills to provide equitable care to diverse populations?

Feedback on the demo module	• How would you describe your experience going through the demo module?
	• What format would you prefer for antiracism training (in person, online, zoom)?
	• Would you have 10-15 minutes to dedicate to this specific type of antiracism learning?
	• Would seeing a report of potential bias in your electronic health record make you more or less likely to complete antiracism training? Why or why not?
		• What did you find helpful about the demo module?
		• What would make it more helpful?
		• How would you feel about your organization using digital means (online) in the form of self-directed modules to provide antiracism training?
		• Would 10-15 mins, once per week, for 6 weeks seem manageable?
		• A report of potential bias may include:
		  • Frequency of biased statements in notes
		  • Racial disparities in prescribing patterns

Ethical Considerations

This study was approved by the University of California, Los Angeles Institutional Review Board (IRB#22-001632-AM-00002).

Results

A total of 12 mental health clinicians (psychiatry residents, fellows, and social workers) completed the semistructured interviews. The participant characteristics included: female candidates (10/12, 83%), male candidates (2/12, 17%), and Asian (5/12, 42%), Black (2/12, 17%), Hispanic or Latinx (1/12, 8%), Middle Eastern (1/12, 8%), multiracial (1/12, 8%), White (1/12, 8%), and other (1/12, 8%) candidates.

The results of the content and thematic analysis are summarized in Table 2, but major themes are highlighted as follows: the majority of participants (7/12, 58%) desired more robust antiracism training in mental health care. With regard to the demo module, the majority (8/12, 67%) enjoyed the module, (6/12, 50%) found it to be well-organized, and (11/12, 92%) felt the time commitment to be manageable. Many participants particularly enjoyed the CBT-based content (4/12, 33%), especially the daily self-reflection log (4/12, 33%). About 4 participants expressed a preference for an online self-directed structure, and 7/12 (58%) participants suggested that online content could be enhanced with an in-person or group component. Lastly, 4 participants communicated ways to improve participant engagement through the digital modality, including offering incentives, sharing personal experiences, and recording progress.

Table 2. Themes and representative quotes from semistructured interviews.

Question	Themes	Quotes
Discussions about race	• Comfortable (5 mentions) • Somewhat comfortable (2 mentions) • Not very comfortable (4 mentions)	• I think so … I have to admit that oftentimes in the face of authority figures, it can be challenging…, it can get tiring though, when you're one of the few faces of color, or if you're like, the only Black person in the room…
Definitions of racism and antiracism	• Active advocacy against racism (11 mentions) • Racism as all-encompassing and systemic (5 mentions) • Self-awareness of antiracism (4 mentions)	• Antiracism, specifically, is a life-long journey, being aware of racial dynamics and disparities and power dynamics, I see it as, like, a modifiable factor.
Previous antiracism training	• Beneficial, in-depth discussions and courses at some point (11 mentions) • Limitations of training format and practicality (8 mentions) • Strong training in residency (7 mentions) • Minimal or no antiracism training (6 mentions) • Insufficient institutional support (5) • Beneficial scenario training (4 mentions)	• A lot of it is very theoretical; less of it is practical in the sense of, you know, in a specific situation. • I feel like there's a lot of, like, resident-driven antiracism efforts … justice, equity, diversity, and inclusion groups… • Anti-racist work has been performative, ...there was too high a burden on faculty of color…
Antiracism training needs	• More robust training and resources (7 mentions) • Integration of representation and lived experiences (4 mentions) • Accessible language (4 mentions) • Integrate translational social sciences in curriculum (3 mentions) • Increase cultural competency (2 mentions) • Mitigate minority tax (2 mentions)	• Just hiring, you know, more faculty of color, I feel that the best ways I've learned have been when developing relationships outside of academia bubbles and being with people with lived experience. • Having more, like, role-playing kind of activities might be great because for me, it's like if I'm in a situation where I have to speak up, my mind goes blank.
Digital demo module experience	• Enjoyed digital module (8 mentions) • Clear and organized web-based structure (6 mentions) • Particularly liked CBT-based examples of core beliefs (4 mentions) • Particularly liked online daily self-reflection logging (4 mentions)	• I really like the module. That's just like what happens at the hospital. • It was clear and I thought the structure was very helpful and consistent while going through the four examples of core beliefs • The good clinician one in particular led me to think about how there are so many ways the system rewards not thinking and not challenging biases, and I think it was nice that you provided that example.
Digital demo module time	• Feels that 15 minutes/week of web-based intervention content for six weeks is manageable (11 mentions)	• Yes, I think we can definitely make that time.
Demo module digital format	• Prefers online content with addition of in-person or group setting (7 mentions) • Concerns about exclusively online, self-directed formats (5 mentions) • Prefers self-directed online-only modules (4 mentions) • Open to conducting over a Zoom call (3 mentions)	• In-person is generally always the most effective. I think we tend to have short attention spans, and it becomes just an online module you have to do. • If you really want people to be an active participant and really engage with it, I don't know how good self-directed modules are … I'm just like clicking through it.
Digital demo module improvements	• Enhance resident participation and engagement in format (4 mentions) • Include web-based incentive to track growth (2 mentions)	• I think that it might be helpful to allow us space to bring up our own examples, but I know that it takes a lot of vulnerability for us to sit there and reflect. • If there's some sort of incentivization structure for people to check back in or record progress into, like a diary, I think that could be effective.
Potential report of EHR bias	• Yes, it would be helpful (11 mentions)	• Yeah, it would overall. I think it would be cool, because in the same way that they make us look at how often we are prescribing benzos, why can't we also be explicit, you know. in terms of antiracism? • Yeah, it would make me more wary. It would make me sit down and think.

Question	Themes	Quotes
Across questions	• Minority tax (5 mentions)	• Everyone has the responsibility to care for, like, a language diverse community. It [shouldn't] just fall on certain individuals just because of their background. • I remember there were some moments in medical school where I felt like there was too high a burden on faculty of color and also students of color.

Discussion

Overview

Using semistructured interviews with mental health professionals, our results indicate favorable acceptability of antiracist intervention content and digital delivery methods. The web-based demo module of the antiracism intervention received a high level of positive feedback, with participants finding it relevant, well-structured, and generally effective in teaching CBT principles. For example, participants enjoyed learning how to identify, react to, and consciously correct core beliefs that propagate racism in health care. Regarding acceptability, participants felt the time commitment would be feasible, especially the convenience of being able to access web-based modules for short periods of time over the course of several weeks. Online self-directed training was well-received, with a recommendation for the addition of a group, in-person, or zoom component to solidify and expand upon web-based self-directed learning. Participants also felt that this would improve engagement, especially with opportunities to share their own experiences. Such findings are in line with previous research suggesting that personalization and increased social connectedness facilitated by digital health intervention components can enhance user engagement [10].

In the context of existing literature, there is a need for targeted evidence-based antiracist strategies addressing the unique and specific needs of clinicians operating in any given health care specialty, as the needs of most mental health professionals will differ greatly from those of general health practitioners [11]. Unfortunately, most antiracism interventions to date have focused only on general health professionals, resulting in the existence of far less tailored interventions addressing a specific health care context or specialty. Furthermore, there are limited discussions of methods for enhancing engagement in antiracism training other than mandating antiracism work [11]. Findings from this study fill this critical gap in the literature by investigating needed aspects of antiracist intervention dedicated to specialized mental health care, with the added benefit of using novel digital-based design elements promoting enhanced acceptability and participant engagement.

Limitations

Limitations include the fact that this study was conducted at a single academic center, which limits its generalizability to other institutions. However, this is a targeted approach to be applied to the study population of mental health professionals. A similar approach can be applied to other health specialty areas, using interviews targeting clinicians of interest. Such methods may further be used to tailor digital antiracism training to other clinical specialties. Another limitation is that the current study focuses on the acceptability of the intervention rather than its efficacy. Lastly, another important limitation lies in the lack of community engagement in the intervention design process, an aspect known to enhance the health equity of digital health interventions [12,13]. Future iterations will therefore aim to involve the systematic incorporation of the voices of community members served by mental health professionals.

Conclusions

Taken together, these results provide important guidelines for the implementation of a targeted intervention for mental health clinicians. They suggest favorable acceptability regarding the use of CBT principles in antiracism education and delivery in a web-based format. Such synthesized findings and insights from mental health professionals may be used to tailor and guide practical aspects of the further development and piloting of a future large-scale web-based antiracism intervention.

Acknowledgments

Funding for this project was provided by the University of California, Los Angeles Initiative to Study Hate (TRB), the University of California, Los Angeles Department of Psychiatry Justice, Equity, Diversity, and Inclusion Research Support Funds, and the American Psychiatric Association Substance Abuse and Mental Health Services Administration Minority Fellowship program (#H79SM080388; TRB). The contents of this publication are solely the responsibility of the author and do not necessarily represent the official views of the Department of Health and Human Services, APA, or APA Foundation.

Data Availability

The data used for this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

References

1. Li H, Eack SM, Montrose DM, Miewald JM, Keshavan M. Longitudinal treatment outcome of African American and Caucasian patients with first episode psychosis. Asian J Psychiatr 2011;4(4):266-271 [FREE Full text] [doi: 10.1016/j.ajp.2011.08.004] [Medline: 23051160]
2. Bailey ZD, Krieger N, Agénor M, Graves J, Linos N, Bassett MT. Structural racism and health inequities in the USA: evidence and interventions. Lancet 2017;389(10077):1453-1463. [doi: 10.1016/S0140-6736(17)30569-X] [Medline: 28402827]
3. Jansen MO, Brown TR, Xu KY, Glowinski AL. Using digital technology to overcome racial disparities in child and adolescent psychiatry. J Am Acad Child Adolesc Psychiatry 2022;61(10):1211-1217 [FREE Full text] [doi: 10.1016/j.jaac.2022.03.013] [Medline: 35358663]
4. Brown TR, Xu KY, Glowinski AL. Cognitive behavioral therapy and the implementation of antiracism. JAMA Psychiatry 2021;78(8):819-820 [FREE Full text] [doi: 10.1001/jamapsychiatry.2021.0487] [Medline: 33950160]
5. Cénat JM, Broussard C, Jacob G, Kogan C, Corace K, Ukwu G, et al. Antiracist training programs for mental health professionals: a scoping review. Clin Psychol Rev 2024;108:102373 [FREE Full text] [doi: 10.1016/j.cpr.2023.102373] [Medline: 38232574]
6. Chu W, Wippold G, Becker KD. A systematic review of cultural competence trainings for mental health providers. Prof Psychol Res Pr 2022;53(4):362-371 [FREE Full text] [doi: 10.1037/pro0000469] [Medline: 37332624]
7. Friis-Healy EA, Nagy GA, Kollins SH. It is time to REACT: opportunities for digital mental health apps to reduce mental health disparities in racially and ethnically minoritized groups. JMIR Ment Health 2021;8(1):e25456 [FREE Full text] [doi: 10.2196/25456] [Medline: 33406050]
8. Lyon AR, Munson SA, Renn BN, Atkins DC, Pullmann MD, Friedman E, et al. Use of human-centered design to improve implementation of evidence-based psychotherapies in low-resource communities: protocol for studies applying a framework to assess usability? JMIR Res Protoc 2019;8(10):e14990 [FREE Full text] [doi: 10.2196/14990] [Medline: 31599736]
9. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
10. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
11. Hassen N, Lofters A, Michael S, Mall A, Pinto AD, Rackal J. Implementing anti-racism interventions in healthcare settings: a scoping review. Int J Environ Res Public Health 2021;18(6):2993 [FREE Full text] [doi: 10.3390/ijerph18062993] [Medline: 33803942]
12. Brewer LC, Fortuna KL, Jones C, Walker R, Hayes SN, Patten CA, et al. Back to the future: achieving health equity through health informatics and digital health. JMIR Mhealth Uhealth 2020;8(1):e14512 [FREE Full text] [doi: 10.2196/14512] [Medline: 31934874]
13. Nguyen-Truong CKY, Waters SF, Richardson M, Barrow N, Seia J, Eti DU, et al. An antiracism community-based participatory research with organizations serving immigrant and marginalized communities, including Asian Americans and Native Hawaiians/Pacific Islanders in the United States Pacific Northwest: qualitative description study with key informants. Asian Pac Isl Nurs J 2023;7:e43150 [FREE Full text] [doi: 10.2196/43150] [Medline: 36648292]

Abbreviations

CBT: cognitive behavioral therapy

Edited by D Gooch; submitted 07.09.23; peer-reviewed by K Xu, A AL-Asadi; comments to author 02.01.24; revised version received 04.02.24; accepted 06.02.24; published 03.04.24.

Please cite as:
Brown TR, Amir H, Hirsch D, Jansen MO
Designing a Novel Digitally Delivered Antiracism Intervention for Mental Health Clinicians: Exploratory Analysis of Acceptability
JMIR Hum Factors 2024;11:e52561
URL: https://humanfactors.jmir.org/2024/1/e52561
doi:10.2196/52561
PMID:38568730

Original Paper

A Geospatial Drug Abuse Risk Assessment and Monitoring Dashboard Tailored for School Students: Development Study With Requirement Analysis and Acceptance Evaluation

Ahmad Mustafa Al-Aboosi[1], MSc; Siti Norul Huda Sheikh Abdullah[1], PhD; Rozmi Ismail[2], PhD; Khairul Nizam Abdul Maulud[3], PhD; Lutfun Nahar[4], PhD; Khairul Akram Zainol Ariffin[1], PhD; Meng Chun Lam[1], PhD; Muhamad Lazim bin Talib[5], PhD; Suzaily Wahab[6], PhD; Mahadzir Elias[7], BA

[1]Faculty of Information Science & Technology, Universiti Kebangsaan Malaysia, Selangor, Malaysia
[2]Centre for Research in Psychology and Human Well-Being, Universiti Kebangsaan Malaysia, Selangor, Malaysia
[3]Department of Civil Engineering, Universiti Kebangsaan Malaysia, Selangor, Malaysia
[4]School of Computing and Data Science, Xiamen University Malaysia, Selangor, Malaysia
[5]Fakulti Sains & Teknologi Pertahanan, Pertahanan Nasional Malaysia, Kuala Lumpur, Malaysia
[6]Department of Psychiatry, Universiti Kebangsaan Malaysia, Selangor, Malaysia
[7]Agensi Antidadah Kebangsaan, Selangor, Malaysia

Corresponding Author:
Siti Norul Huda Sheikh Abdullah, PhD
Faculty of Information Science & Technology
Universiti Kebangsaan Malaysia
Jln Bangi, Bangi
Selangor, 43600
Malaysia
Phone: 60 03 8921 6179
Email: snhsabdullah@ukm.edu.my

Abstract

Background: The enormous consequences of drugs include suicides, traffic accidents, and violence, affecting the individual, family, society, and country. Therefore, it is necessary to constantly identify and monitor the drug abuse rate among school-going youth. A geospatial dashboard is vital for the monitoring of drug abuse and related crime incidence in a decision support system.

Objective: This paper mainly focuses on developing MyAsriGeo, a geospatial drug abuse risk assessment and monitoring dashboard tailored for school students. It introduces innovative functionality, seamlessly orchestrating the assessment of drug abuse usage patterns and risks using multivariate student data.

Methods: A geospatial drug abuse dashboard for monitoring and analysis was designed and developed in this study based on agile methodology and prototyping. Using focus group and interviews, we first examined and gathered the requirements, feedback, and user approval of the MyAsriGeo dashboard. Experts and stakeholders such as the National Anti-Drugs Agency, police, the Federal Department of Town and Country Planning, school instructors, students, and researchers were among those who responded. A total of 20 specialists were involved in the requirement analysis and acceptance evaluation of the pilot and final version of the dashboard. The evaluation sought to identify various user acceptance aspects, such as ease of use and usefulness, for both the pilot and final versions, and 2 additional factors based on the Post-Study System Usability Questionnaire and Task-Technology Fit models were enlisted to assess the interface quality and dashboard sufficiency for the final version.

Results: The MyAsriGeo geospatial dashboard was designed to meet the needs of all user types, as identified through a requirement gathering process. It includes several key functions, such as a geospatial map that shows the locations of high-risk areas for drug abuse, data on drug abuse among students, tools for assessing the risk of drug abuse in different areas, demographic information, and a self-problem test. It also includes the Alcohol, Smoking, and Substance Involvement Screening Test and its risk assessment to help users understand and interpret the results of student risk. The initial prototype and final version of the dashboard were evaluated by 20 experts, which revealed a significant improvement in the ease of use (P=.047) and usefulness (P=.02) factors and showed a high acceptance mean scores for ease of use (4.2), usefulness (4.46), interface quality (4.29), and sufficiency (4.13).

Conclusions: The MyAsriGeo geospatial dashboard is useful for monitoring and analyzing drug abuse among school-going youth in Malaysia. It was developed based on the needs of various stakeholders and includes a range of functions. The dashboard was evaluated by a group of experts. Overall, the MyAsriGeo geospatial dashboard is a valuable resource for helping stakeholders understand and respond to the issue of drug abuse among youth.

(JMIR Hum Factors 2024;11:e48139) doi:10.2196/48139

KEYWORDS

geospatial; statistics; map; youth; drugs; dashboard; evaluation; drug abuse; monitoring; risk assessment

Introduction

Background

According to the United Nations Office on Drugs and Crime, over 275 million people used drugs globally in 2020, resulting in 36 million individuals with drug use disorders. From 2010, there was a 22% increase in people who use drug, with an 11% increase projected by 2030 [1]. The National Anti-Drug Agency Malaysia arrested 20,643 people addicted to drugs in 2020, showing a decrease of 20.8% from 2019. The agency provides information on awareness programs, medicine, intervention, and rehabilitation, creating hot spots for drug abuse activities. In 2020, people who use and are addicted to drugs accounted for 2% of adolescents aged between 13 and 18 years. A total of 59.4% of all people who use drugs had a secondary education, but a recent study reported that approximately 5.5% of Malaysian youths are lifetime people who use drugs, which is relatively high compared with previous findings [2,3]. To combat the increasing number of adolescent people who use drugs in Malaysia, the government has focused on implementing measures to reduce drug use [4-6]. Drug abuse severely affects individuals, families, and society [7]. Consistent monitoring of drug abuse rates among Malaysian school-going youth is essential.

The need for information and statistics about people addicted to drugs, accessibility, usage, effects, and prevention is vital for reducing drug abuse [8]. However, local communities lack effective prevention tools despite drug abuse reaching epidemic levels [4,9]. The increasing drug use among school students contributes to the spread of drug abuse epidemics [10]. To address this challenge, geospatial analytics can provide location-based tools with statistical charts to assess prevalence and risk factors associated with substance abuse, aiding stakeholders' understanding of the epidemic and promoting monitoring and prevention activities. Because of the substantial amount of assessment results that need to be compiled, it will be necessary to develop a digital system capable of managing, integrating, and synthesizing assessment data [11]. However, developing a user-friendly and effective dashboard can be a complex task that requires input from various stakeholders and careful consideration of user acceptance factors.

This paper is organized as follows. First, we provide a brief overview of the literature and related work. Next, we describe the methodology used in the study, including requirement gathering, pilot study, agile development, and data analysis. Then, we present the results of the study and discuss their implications. Finally, we conclude with suggestions for future research directions.

Related Work

A digital dashboard tracks, analyzes, and manages information to indicate key performance related to a monitored subject matter [12,13]. However, sometimes a simple design is insufficient for understanding a specific topic that requires spatial analysis instead of simple visual inspection [14,15]. A geospatial dashboard provides a solution with a web-based interactive interface that combines mapping, spatial analysis, and visualization with dashboard functions [16]. In the case of drug abuse, geographical location needs to be considered to explore the location and distribution [17]. Though several dashboards are introduced in other sectors [18-20], few dashboards are found in the drug abuse sector.

As Muhamad et al [5] reported, psychoactive drugs have become a burden on public health globally, causing social difficulties and intentional overdose fatalities [21]. The government must focus on reducing teens who use drugs, and a geospatial dashboard can help through monitoring and prevention activities. However, previous dashboards were either agency-specific or not practical. This paper reports developing a geospatial dashboard for monitoring drug or substance abuse in Malaysian schools.

Several tools are available to help monitor and manage drug abuse, each with its unique strengths (see Multimedia Appendix 1) [8,9,21-30]. One such tool is DrugTracker [9], which combines social media and geographic data to detect and monitor drug abuse in near real time. This tool can help identify emerging drug trends and monitor drug abuse hot spots. Another tool, the Drug Abuse Information System [8], is designed to improve the storage and reporting of drug abuse information in Tanzania, making it easier to manage and prevent it. The Drug Abuse Information System is particularly useful for providing detailed data and reports on drug abuse, making tracking trends and identifying at-risk populations easier.

The VISN 22 dashboard [21] is another powerful tool to help detect veterans taking high-dose opioids and monitor and control for concomitant suicide risk factors in the United States. The dashboard helps review references to high-dose opioids prescribed with psychiatric illness and suicide risk factors. The Dashboard for Substance Abuse Prevention and Control [25] is a tool that uses survey data to help monitor, evaluate, and manage strategic performance in Los Angeles. It is beneficial for its ability to provide up-to-date raw data on patient age, employment, and criminal status risk factors, helping to identify at-risk populations and prevent drug abuse. Clinical Dashboards for Addiction Treatment [26] is another tool that arranges and presents patients' data so clinicians can make informed, tailored medical decisions. This tool is valuable for its ability to provide detailed patient data as a pilot study in a Midwestern substance use disorder treatment center, allowing clinicians to tailor treatment to individual patients.

The Drug Control Data Dashboard [27] is a tool that allows users to search topics by year, agency, drug, and, to a limited degree, geographic location, providing a machine-readable and interactive collection of drug data from many sources in the United States. This tool is particularly useful for its ability to provide a wealth of data on many specific drug topics. The opioid-related adverse drug events dashboard [28] allows hospitals to access their opioid-related adverse drug events and local benchmark data against national trends in the United States. The visual analytics dashboard [29] is another tool that presents hospital-wide electronic health record medication alerts in Philadelphia to help reduce alert fatigue and improve medication safety. Finally, the Substance Abuse InstantAtlas dashboard [30] presents data on the health impacts of alcohol and selected drugs concerning public health in Alaska.

However, there are also limitations to these tools, such as incomplete and shallow social media data, lack of risk assessment information, and concerns about design flexibility and the ability to feed up-to-date raw data. Also, they have limitations regarding the scope of data, such as only reporting data on specific substance categories or a limited patient population. It is also important to note that some of the tools discussed in our study have not undergone user acceptance testing. Acceptance testing is an essential component for evaluating the user acceptance of any tool or product. Without proper acceptance testing, it is difficult to determine whether a tool is easy to use, useful, and effective in meeting the needs of its intended audience.

MyAsriGeo, a geospatial drug abuse risk assessment and monitoring dashboard tailored for school students, addresses the shortcomings of previous related works by offering unique features specifically designed to monitor drug use in Malaysian schools. In contrast to existing tools, it overcomes limitations in data comprehensiveness, depth of analysis, and spatial relevance by providing dynamic visualizations and seamless data connectivity. The dashboard focuses on the Malaysian school context, coupled with innovative features, and ensures a more comprehensive understanding of drug use patterns. It is validated using the Technology Acceptance Model (TAM), Post-Study System Usability Questionnaire (PSSUQ), and Task-Technology Fit (TTF) metrics. These assessments provided comprehensive insights into the ease of use, usefulness, interface quality, and sufficiency of MyAsriGeo. The process confirms that the geospatial dashboard transcends limitations, emerging as an innovative solution adept at meeting the diverse needs of users.

Overall, we provided a useful overview of different tools available for drug abuse detection and management, but it is important to carefully consider the strengths and limitations of each tool before selecting one for a specific use case.

Methods

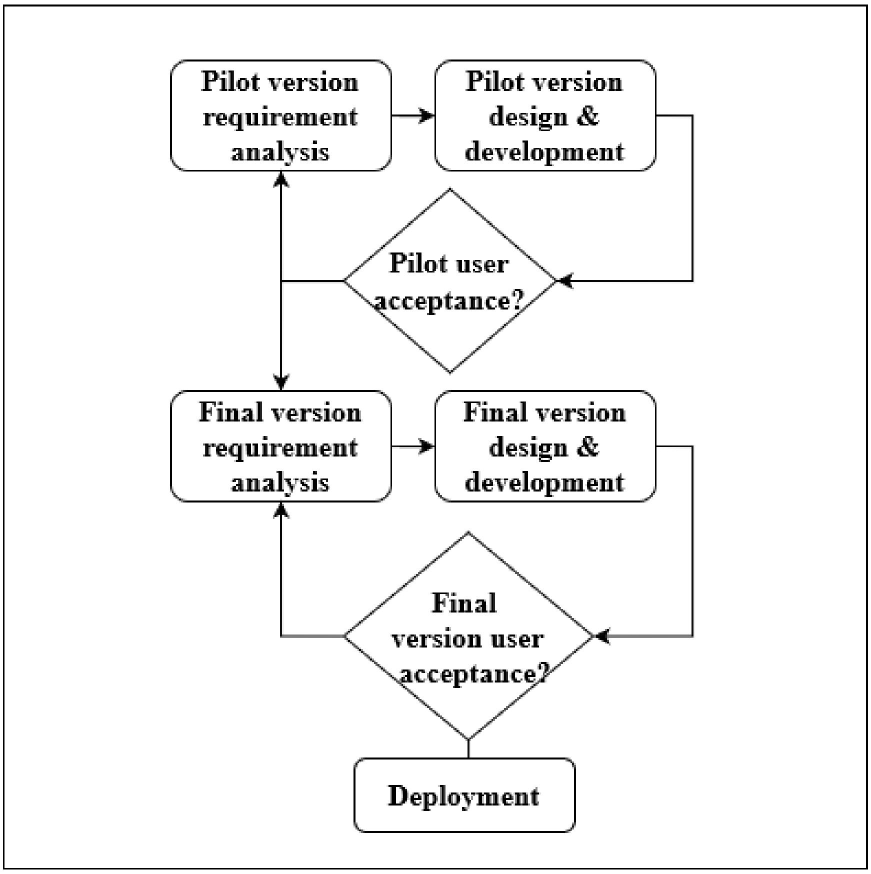
Procedures

The methodology for the MyAsriGeo dashboard was based on agile prototyping and included iterative cycles of requirement analysis, dashboard design, development, and acceptance testing [31,32]. The use of the prototyping technique in the agile development process is a common practice among various studies and is highly favored by the agile community; this technique is likely well suited to be implemented in this type of process [33].

A student survey on adolescent drug abuse in Malaysia was conducted in collaboration with 85 schools, the Ministry of Education, Universiti Kebangsaan Malaysia, and the National Anti-Drugs Agency, comprising psychiatrists, psychologists, counselors, and experienced practitioners, through the MyAsri platform of the SEGAR project. It highlights around 3000 students to understand risk factors, prevalence, and elements of trajectories among adolescents involved in polydrugs and amphetamine-type stimulants abuse at selected hot spots in Malaysia. Drug abuse surveys offer a valuable understanding of the sociobehavioral dimensions related to adolescent drug abuse activities in Malaysia.

The MyAsriGeo dashboard underwent 2 rounds of requirement analysis and evaluation, as shown in Figure 1. Experts from different fields, including medical, research, development, education, geospatial, and drug rehabilitation domains, participated in 2 rounds of focus groups and interviews and 2 rounds of user acceptance.

Figure 1. Methodology flowchart.



Ethical Considerations

With necessary approvals from relevant ethics committees (UKM PPI/111/8 JEP-2020-174(2)), parental consent was diligently obtained to ensure compliance with ethical standards. All data collected in this study were anonymized to protect the privacy and confidentiality of participants. Additionally, participants were compensated MYR RM20 (US $4.29) for their involvement in the research.

Pilot Version Sprint

The pilot version of the MyAsriGeo geospatial drug abuse dashboard was designed and developed based on the basic requirements collected from the experts and stakeholders.

Pilot Version Requirement Analysis

The requirement analysis phase involved gathering comprehensive feedback from a diverse group of experts and stakeholders, including 20 experts spanning medical, research, development, education, geospatial, and drug rehabilitation domains. Aged between 30 and 59 years and possessing more than 5 years of experience in their respective fields, participants actively engaged in both focus group discussions and interviews, providing valuable insights. Thematic analysis was used during subsequent data analysis, systematically reviewing and coding detailed notes to identify recurring themes, patterns, and key insights, such as the selection of the contents of the results of the student surveys, demographics, notable trends in drug usage, specific challenges faced by students, and key points. This rigorous approach ensured a nuanced understanding of stakeholders' perspectives. The resulting feedback served as a robust foundation for crafting the pilot version of the dashboard, strategically incorporating basic functionality to address critical needs identified in this collaborative and insightful process.

Pilot Version Design and Development

Based on the thematic analysis, the pilot version consisted of 3 pages, with no authentication or authorization system. The first page of the pilot version presented a basic view of the map, displaying the locations of schools without providing detailed summaries of key points. This page was designed to provide users with a simple and clear overview of the schools' locations. The second and third pages of the pilot version were used to display the student survey results and the drug and substance abuse in the form of charts. These charts provided an overview of the prevalence of drug abuse among students, the types of drugs used, and other related information. The charts were designed to be visually appealing and easily interpreted, giving users a clear understanding of the survey results.

Pilot Version User Acceptance

The first TAM questionnaire was conducted for the 20 experts and stakeholders to evaluate the pilot version's user acceptance such as ease of use and usefulness, and to identify areas for improvement. TAM is a theoretical framework used to predict how users will adopt and use technology. TAM suggests that usefulness and ease of use are important determinants of user acceptance and subsequent behavior. TAM has been widely used in various fields to analyze the adoption and usage of various technologies [22].

Initially, we crafted a survey using both a web and a paper-based questionnaire for individuals unable to access the online version. This survey aimed to explore the acceptance of the pilot version of the MyAsriGeo dashboard among the 20 experts and stakeholders, each with expertise in different specialty areas as mentioned earlier. The survey questions were designed, incorporating 5 items each for ease of use and usefulness (based on TAM). Responses were gathered using a Likert scale ranging from 1 to 5.

The feedback was used to refine the design and features of the dashboard. This process continued through multiple iterations until the MyAsriGeo dashboard was deemed effective in meeting the stakeholders' needs and goals and could effectively support their efforts to address drug abuse in their communities.

Final Version Sprint

Final Version Requirement Analysis

The feedback and acceptance results received from the experts and stakeholders who evaluated the pilot version were critical in identifying areas for improvement and ensuring that the final version of the dashboard met the needs of its intended users.

To define the end users and the new requirement for the MyAsriGeo dashboard, we conducted another focus group discussion and interviewed the experts and the stakeholders with our research team. During these sessions, we discussed the various demographic groups using the dashboard, including school-going youth, educators, and health care professionals. We also discussed the importance of including a variety of data and information on the dashboard, such as students' demography and the Mooney Problem Checklist [34]; information on drug and substance abuse; and the Alcohol, Smoking, and Substance Involvement Screening Test (ASSIST) [35].

Final Version Design and Development

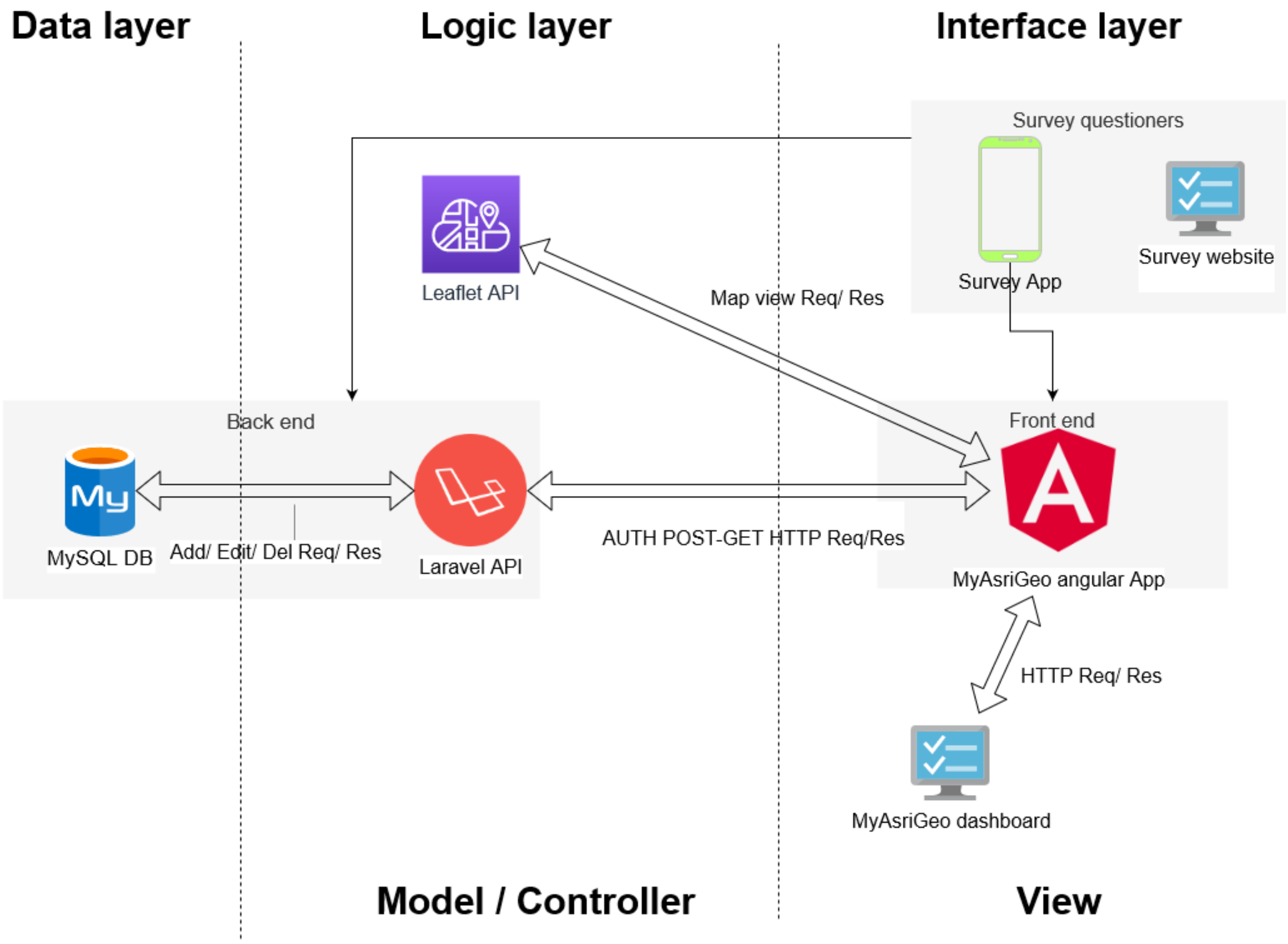
Based on the thematic analysis of the second requirement analysis, the results were presented in the dashboard, which were crucial in providing a comprehensive understanding of the drug abuse problem and helping users make informed decisions. Additionally, experts were concerned about the importance of including information on economic affordability and types of drugs being used and data on the time and place of drug use. These factors are considered in designing the MyAsriGeo dashboard to ensure that it effectively addresses the issue of drug abuse among students. The dashboard design phase focused on refining the pilot version through iterative feedback cycles, testing, and refinement. The development phase involved implementing the design and building the dashboard features, with each iteration building upon the previous one.

MyAsriGeo Architecture

The MyAsriGeo geospatial dashboard was developed using the Angular framework and TypeScript for the front end, with the Leaflet application programming interface (API) used for geomapping locations (as shown in Figure 2). The dashboard includes in-app charts to display students' survey responses collected from mobile app surveys (which is not in the scope of this study), and aggregated data are stored in a MySQL database shared between survey applications and the Laravel API. The dashboard has a 3-tiered architecture, including the data, logic, and interface layers. The data layer holds geospatial and nongeographical data, and the logical layer contains models mapped to corresponding database tables in an object-oriented manner. The interface layer interacts directly with the client, communicating every client interaction with the logical layer containing the Leaflet API and Laravel API.

In crafting the MyAsriGeo architecture, our intentional choice to develop it as a web system is grounded in a forward-thinking approach, prioritizing future collaboration and sustainability. The incorporation of an API serves a dual purpose—facilitating seamless integration with external partners for potential collaborations and ensuring adaptability for future initiatives. The use of free services is pivotal for the project's sustainability, providing scalability and maintenance without imposing significant financial constraints. Furthermore, MyAsriGeo's tailored design specifically caters to the unique needs of drug abuse risk assessment for educational contexts, distinguishing it from other solutions. This strategic architectural approach positions MyAsriGeo as a versatile and enduring solution, adept at meeting the evolving demands of current and prospective stakeholders within the targeted field.

Figure 2. Architecture design of MyAsriGeo. API: application programming interface; DB: database.



Final Version User Acceptance

Finally, the second Acceptance questionnaire was conducted to evaluate the final version of the dashboard; 2 additional factors added based on the PSSUQ and TTF models were enlisted to assess the interface quality and dashboard sufficiency for the final version by end-users and stakeholders. PSSUQ is a widely used measure of the usability of a system [24], while TTF measures the extent to which technology helps users perform their tasks effectively and efficiently [23].

Similar to the pilot version, we initiated the user acceptance for the final version of MyAsriGeo with the aforementioned experts. The survey questions were designed incorporating 5 items each for ease of use and usefulness (based on TAM), 4 items for interface quality (from PSSUQ), and 4 items for sufficiency (from TTF). Responses were gathered using a Likert scale ranging from 1 to 5. Subsequently, we presented the finalized dashboard to the aforementioned 20 experts and stakeholders. They were instructed to interact with and test the dashboard. After this hands-on experience, we administered the designed questionnaire to gather their feedback and insights.

Results

Requirement Analysis

The student survey application had 4 contents: demography, Mooney Problem Checklist, information on drug and substance abuse, and ASSIST. The first content, demography, gives an overview of the student's life, including name, age, area, school, race, gender, religion, and economic status. The second content, the Mooney Problem Checklist [34], helps adolescents and adults to express personal difficulties, which is useful in enhancing teachers' understanding of pupils and preparing children for counseling interviews. The topics of the checklist vary according to age group, health and physical development, home and family, morals and religion, romance, gender, and marriage. The third content, information on drug and substance abuse, identifies the information regarding drug and substance use, history, occurrence use, drug source, budget for drugs, medical history, prison history, abuse friends, and drug influencers of students. Finally, a multinational group of drug abuse researchers has created the World Health Organization's ASSIST to screen for psychoactive substance use and associated problems in primary care patients [35]. ASSIST was considered the fourth content of the survey. The risk and its representation on the geospatial map are the indicators with the result of the previous survey that will give a big picture of the whole drug abuse information.

Textbox 1 outlines the components of the initial version of the dashboard, incorporating selected results from a student survey conducted. This pilot version primarily includes demographic information, providing a comprehensive overview of students' demography, encompassing details such as name, age, area, school, race, gender, religion, and economic status. Additionally, it features information on drug and substance abuse, offering insights into students' history, occurrence of use, drug sources, budget for drugs, medical history, prison history, associations with friends who abuse substances, and key influencers in their drug-related decisions.

Textbox 1. Contents of the results of the student survey, which were presented in the pilot version of the dashboard.

1. Demography: Demographic information of the students.
2. Information on drug and substance abuse: Identify the information regarding drug and substance use of students.

Textbox 2 expands on the content integrated into the final version of the dashboard, building upon the initial selection. Based on the results of the second round of the focus group and interviews, additional features were incorporated. In addition to demographic data and information on drug and substance abuse, the final version includes the Mooney Problem Checklist [34] and the ASSIST tool. The Mooney Problem Checklist serves as a valuable tool to explore the problems students have experienced over the years, aiding in teachers' understanding of students and facilitating counseling interviews. The ASSIST tool, designed by a multinational group of drug abuse researchers, is used to screen for psychoactive substance use and related problems in primary care patients [35]. These enhancements contribute to a more holistic understanding of the challenges faced by students, offering decision makers a comprehensive view of effective planning and intervention strategies. The inclusion of risk indicators and their representation on a geospatial map further enhances the dashboard's capacity to provide a comprehensive picture of drug abuse patterns, enabling informed decision-making based on the survey results.

Textbox 2. Contents of the results of the student survey, which were presented in the final version of the dashboard.

1. Demography: Demographic information of the students.
2. Mooney Problem Checklist [34]: Investigate the problems that students have experienced over the years.
3. Information on drug and substance abuse: Identify the information regarding drug and substance use of students.
4. The Alcohol, Smoking, and Substance Involvement Screening Test (ASSIST) [35]: It finds out the prohibited substances used by students in the last 3 months.

After the thematic analysis of the requirement analysis from the focus group discussion and interviews, we identified the key users for the MyAsriGeo dashboard who will use the dashboard (Textbox 3).

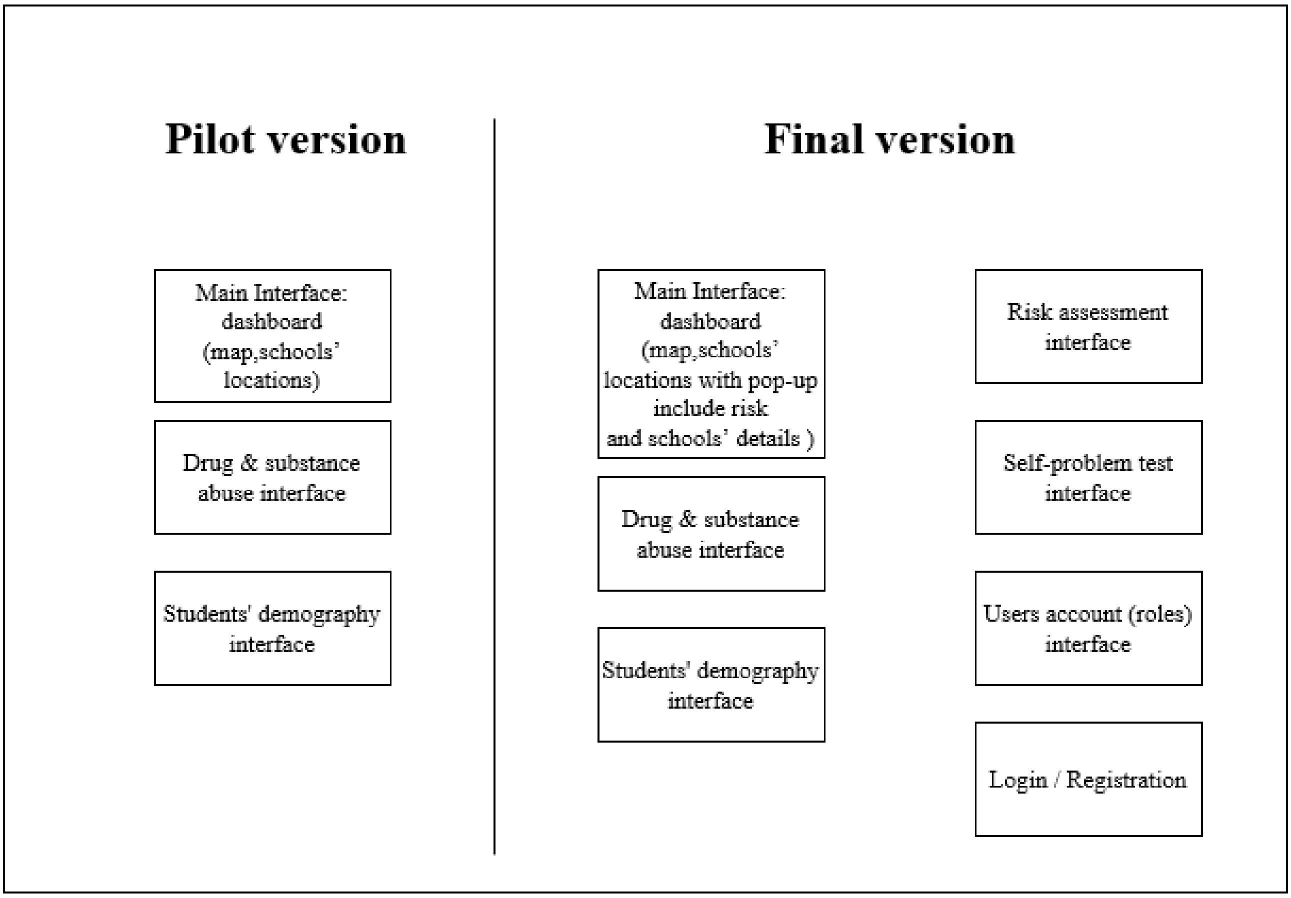
Textbox 3. The key users for the MyAsriGeo dashboard.

1. Researchers benefit from the ability to access and analyze data on drug abuse patterns and trends in schools, which is crucial in informing their research and studies on the subject.
2. Police: The dashboard allows them to identify hot spot schools concerning drug abuse and crime, aiding their investigations and targeted enforcement efforts.
3. National Anti-Drugs Agency: They can access detailed information on hot spot areas, schools, and nearby amenities through the dashboard to assess risk and evaluate the effectiveness of student intervention programs.
4. Community: By accessing information on drug abuse patterns and trends, they can increase their awareness and aid in prevention efforts.
5. Teachers: The dashboard provides access to information on drug abuse patterns and trends in their schools, which can increase their awareness and aid in prevention efforts.
6. Administrators: The dashboard gives them the ability to access, control, and manage confidential data, assign roles, and manage accounts, which ensures data security and compliance.
7. Decision makers: By accessing data and analyzing drug abuse patterns and trends, they can inform and support decision-making to minimize risk, increase awareness, and minimize crime rates.

MyAsriGeo Dashboard

The web application is structured with 7 distinct interfaces to effectively represent and interact with the collected data. In the pilot version (as shown in Figure 3), 3 primary interfaces are featured: the dashboard main page, providing an overarching view of key information with a basic map of the hot spot schools; the drug and substance use page, focusing on details related to substance use among students; and the students page, offering insights into the student population.

Figure 3. Pilot versus final version comparison.



In the final version (as shown in Figure 3), the main page of the dashboard has been enhanced to include a dynamic map feature that provides detailed summaries of key points. This addition augments the visual representation of the data, allowing users to geospatially explore and analyze critical information. Additional interfaces are introduced to enhance the functionality of the application. The risk assessment page provides a detailed breakdown of drug abuse risks, allowing for a more nuanced understanding of specific challenges. The self-problem test page facilitates self-assessment for students, aiding in the identification of personal difficulties and contributing to a proactive approach to addressing concerns. The users page and authentication interface are incorporated to manage user access and permissions securely, ensuring the confidentiality and integrity of the data.

The synchronization of information from the main database back to the geospatial web across these interfaces strengthens the application's ability to provide a comprehensive and accessible representation of the survey results. Together, these interfaces empower users with the tools they need to analyze, understand, and respond effectively to the collected data, fostering informed decision-making and intervention planning.

Main Interface: Dashboard

The MyAsriGeo dashboard has a main interface (as shown in Figure 4) with a menu bar containing 7 options for easy navigation. The interface also displays a geolocation map of hot spot schools in Malaysia with markers colored red where at least one of the students is at a high risk for drug abuse, yellow where at least one of the students is at a medium risk, or green for otherwise. The pop-up for each school provides general statistics such as school name, student count, and school risk. The interface also includes a drop-down list to filter results by drug type. The school risk equation is displayed based on equation 1:

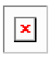
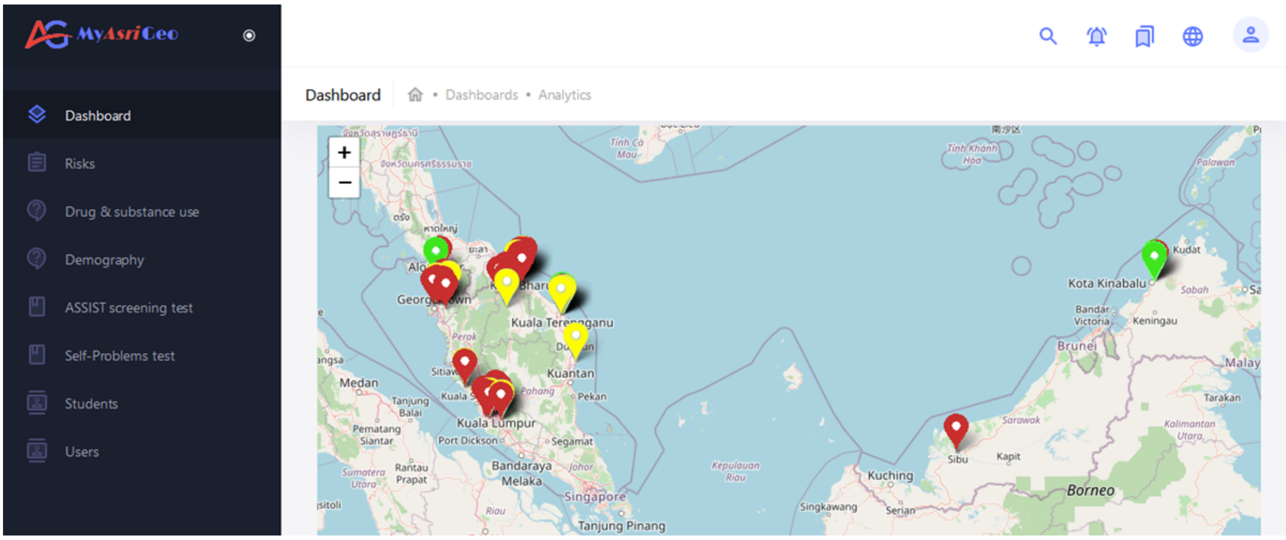


Figure 4. The main interface design of the MyAsriGeo dashboard.



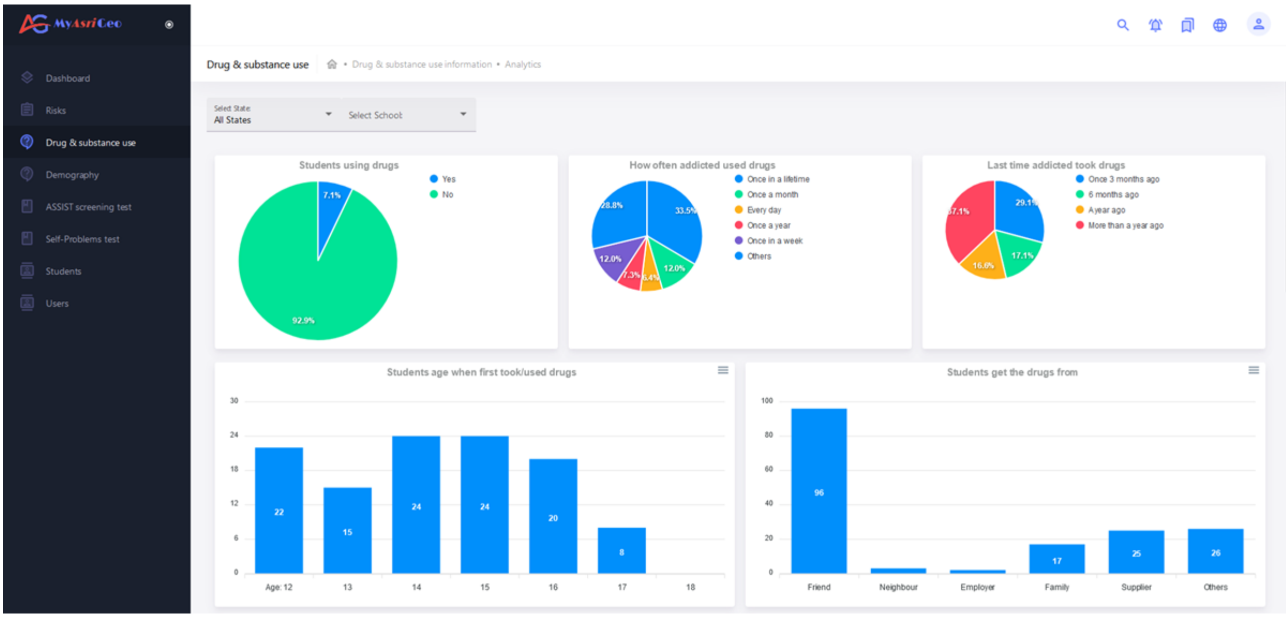
Risk Assessment Interface

The risk assessment page displays individual risks for different types of substance abuse in pie charts, which can be filtered by state and school using drop-down menus. The platform uses the ASSIST screening tool to assess the severity of substance abuse and assigns a numerical score to each question [35], which is then used to determine a low-, medium-, or high-risk level for each substance. Various factors such as family history and socioeconomic level are considered when assessing a student's risk of substance abuse; we represent the calculation of student risk as follows [35]:

Drug and Substance Abuse Interface

The drug and substance abuse page of MyAsriGeo (as shown in Figure 5) displays charts related to drug and substance use, including students' age, gender, medical history, parental marital status, financial status, frequency of drug use, age of first drug use, source of drugs, expenses of drug use, hospitalizations and imprisonment due to drug use, and drug-addicted friends. Users can filter the charts by state and school using drop-down lists. Pie charts and bar charts represent the data whereby the user can scroll down to view more charts.

Figure 5. The drug and substance use interface of MyAsriGeo.



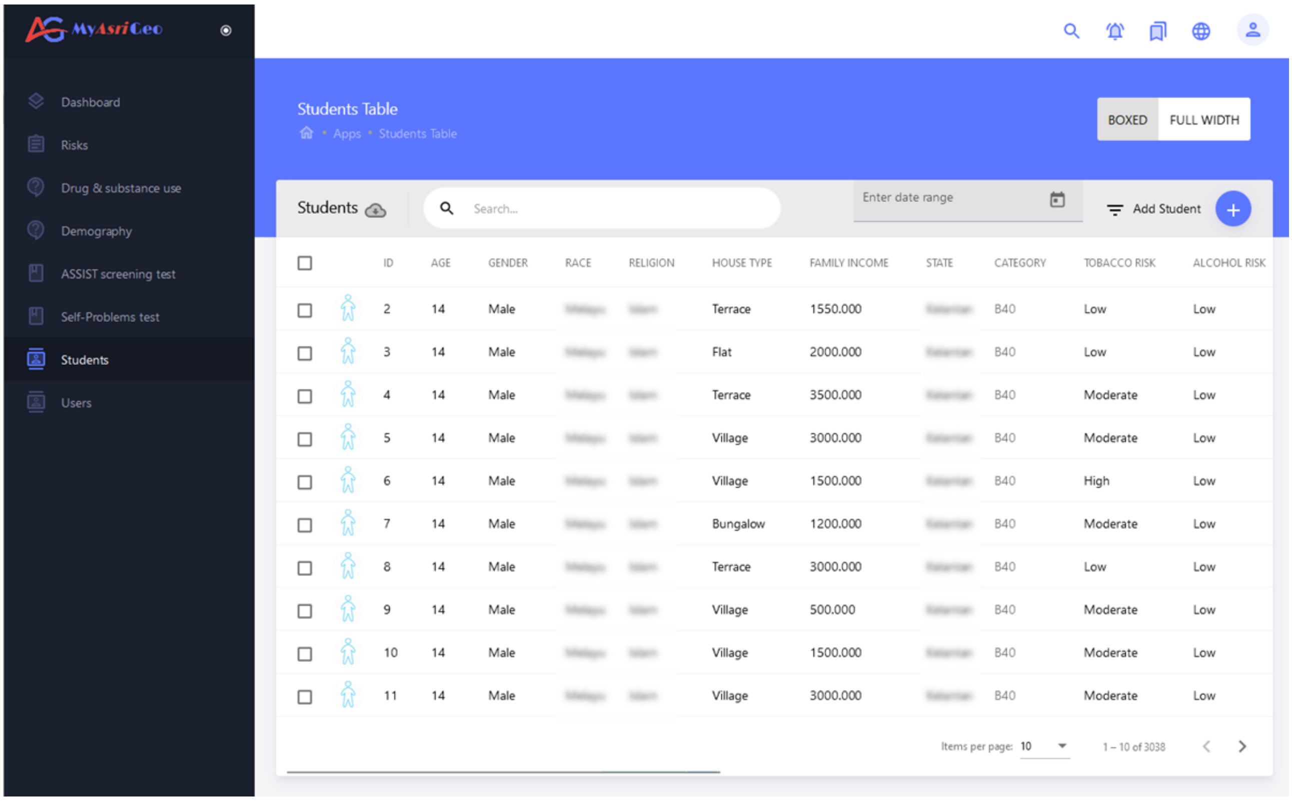
Self-Problem Test Interface

This page presents the Mooney self-problem test result in bar charts with 2 state and school selection filters. Around 60 charts are presented here to cover the questions about money, communication, life, family, faith, job, school, future, and learning. The answers are categorized based on the options (strongly agree, agree, disagree, do not agree, and very not agree) given during survey data collection. The charts can be exported to CSV files, PNG images, and SVG vector-based images.

Students' Demography Interface

On this page (as shown in Figure 6), students' demographic information is listed in a table. Individual students' risk levels for each drug type can be viewed on this page. Each row has an edit and delete button to edit and delete the student information. Apart from that, there is a button on the top of the table to search for any field type. The table can be sorted column-wise for each field. Ascending and descending options, including eliminating field columns, are also available. Options such as exporting tables in Excel form and adding new students are available on this page.

Figure 6. Students page of MyAsriGeo.



Users Account Interface

The MyAsriGeo system has 3 main user roles: administrator, country access user, and state access user. The roles determine the level of access to information and privileges granted to each user; the administrator has all the privileges, including managing user accounts by giving access to newly created accounts and editing and deleting accounts. The country access user can access everything except managing user accounts, while the state access user can access state-level schools, students, risks, and other information. A confirmation email is sent to new account holders to verify their email, and the administrator will review the account and assign appropriate access based on their assigned role. The users' information is listed in a table with add, edit, and delete buttons, search, and filters. The passport token scope allows users to access information based on their role, ensuring that only authorized users can access sensitive data.

User Acceptance Results

Table 1 shows the statistical analysis results comparing the means of 2 factors (ease of use and usefulness) between the pilot version and the final version of a dashboard. For ease of use, the mean score was 3.77 in the pilot version and 4.2 in the final version. The 2-tailed *P* value of .047 indicates a statistically significant difference between the 2 means. The *T* value of –2.1 shows the direction and magnitude of the difference. The effect size, measured by partial η^2, was 0.191, indicating a large effect size. This suggests that the changes made to the dashboard substantially impacted users' perception of its ease of use.

Table 1. Summary of the user acceptance results after the pilot prototype and the final version (N=20).

Factor	Pilot prototype, mean (SD)	Final version, mean (SD)	2-tailed *P* value	*T* value *(df)*	η[a]
Ease of use	3.77 (0.93)	4.2 (0.55)	.047	–2.1 (19)	0.191
Usefulness	3.9 (1.09)	4.46 (0.43)	.02	–2.5 (19)	0.248
Interface quality	—[a]	4.29 (0.49)	—	—	—
Sufficiency	—[a]	4.13 (0.65)	—	—	—

[a]Not applicable (these 2 factors were only designed for the final version evaluation).

For usefulness, the mean score was 3.9 in the pilot version and 4.46 in the final version. The 2-tailed *P* value of .02 indicates a statistically significant difference between the 2 means. The *T* value of –2.5 shows the direction and magnitude of the difference. The effect size, measured by partial η^2, was 0.248, indicating a large effect size. This suggests that the changes made to the dashboard substantially impacted users' perception of its usefulness.

The table shows the mean and SD scores for 2 different measures of user interface quality [24] and sufficiency [23]. PSSUQ is a widely used measure of the usability of a system, while TTF measures the extent to which technology helps users perform their tasks effectively and efficiently. The mean score for interface quality on PSSUQ was 4.29, indicating that users generally found the interface to be of high quality. The SD for PSSUQ was 0.49, indicating that the ratings were relatively consistent. The mean score for sufficiency on TTF was 4.13, indicating that users generally found the system sufficient for their needs. The SD for TTF was slightly higher at 0.65, indicating that ratings were slightly more varied.

To conclude, these results suggest that the changes made to the dashboard between the pilot and final versions significantly impacted users' perceptions of its ease of use and usefulness. The effect size was large for ease of use and large for usefulness, indicating that the changes were meaningful and had a noticeable impact on users' experience with the dashboard. The scores for interface quality and sufficiency of the final version show that users are satisfied with the quality of the MyAsriGeo interface and have found its content sufficient.

Discussion

Principal Findings

Dashboards have become a popular tool in various health care systems to assist in making informed decisions about patient care by presenting a vast amount of information concisely and easily comprehensibly. Their implementation is on the rise as they enhance the quality and safety of care while reducing drug abuse [36,37]. The MyAsriGeo geospatial dashboard, developed to handle and analyze multivariate students' data regarding drug usage, is a valuable resource for stakeholders in Malaysia to understand and respond to the issue of drug abuse among youth. The development process involved continuous feedback from experts and stakeholders, ensuring that the dashboard was designed to meet the needs of all user types. The acceptance evaluation revealed that the dashboard was user-friendly and useful, indicating that it can be a valuable resource for stakeholders involved in addressing the issue of drug abuse among youth.

The high level of interface quality and sufficiency achieved by the MyAsriGeo dashboard is an important outcome, as a well-designed and well-functioning dashboard can greatly enhance the quality and usefulness of the tool [23]. By providing a geospatial map of high-risk areas, data on drug abuse among students, and risk assessment tools, the dashboard helps stakeholders understand the extent of the problem, identify areas of concern, and make informed decisions to address the issue.

The involvement of various stakeholders in developing and evaluating the MyAsriGeo dashboard is also noteworthy. By gathering the feedback and requirements of the National Anti-Drugs Agency, police, schoolteachers, students, and researchers, the dashboard was designed to meet the diverse needs of the users. It enhances the acceptance and adoption of the tool, as it is tailored to the specific needs of the stakeholders.

The evaluation results of the MyAsriGeo dashboard with experts and stakeholders suggest that the changes made to the dashboard between the initial prototype and the final version significantly impacted users' perceptions of its ease of use and usefulness [22]. Specifically, the results indicate that changes to the dashboard greatly affected users' perception of its ease of use and usefulness. The use of agile methodology and prototyping in the design and development of the dashboard allows for close collaboration with experts and stakeholders throughout the development process. This iterative approach facilitated identifying and incorporating feedback and requirements from these key stakeholders [38], which likely contributed to the dashboard's success.

The TAM, PSSUQ, and TTF were used to assess user acceptance of the dashboard. The TAM enabled the evaluation of key aspects of the dashboard's acceptance and provided insight into users' attitudes toward the dashboard. The involvement of 20 experts in evaluating the dashboard further supports the reliability and validity of the findings.

The dashboard supports enhancing decision-making abilities of decision makers, providing insights not only on which drugs are abused by students in hot spot locations but also on the impact those drugs have on the students' resulting health risk. Thus, it allows the decision maker to get precise information and use it accordingly [39].

The results of this study suggest that the MyAsriGeo dashboard, designed and developed using agile methodology and prototyping, is a promising tool for monitoring and analyzing drug abuse trends. The significant impact of the changes made to the dashboard between the initial prototype and final version on users' perceptions of its ease of use and usefulness highlights the importance of ongoing evaluation and refinement of such tools in collaboration with key stakeholders.

To ensure the effectiveness of the MyAsriGeo dashboard, continuous updates and improvements are necessary to keep up with the changing landscape of drug abuse among students. As such, future research can explore including more data sources and real-time data analysis to enhance the accuracy and effectiveness of the dashboard. It can include integrating social media data or information from drug rehabilitation centers.

The review highlighted several limitations of the tools, including incomplete and shallow social media data, lack of risk assessment information, concerns about design flexibility, and the ability to feed up-to-date raw data, and limitations regarding the scope of data they cover. Additionally, some tools have not undergone user acceptance, which is crucial in evaluating the effectiveness of any tool or product [21,25-30].

Therefore, before selecting a tool for a specific use case, it is important to consider its strengths and limitations carefully.

The review provided a useful overview of different tools available for drug abuse detection and management, but it is important to understand that no tool is perfect, and each has its own set of limitations. It is also crucial to conduct user acceptance to ensure that a tool is user-friendly, efficient, and effective in meeting the needs of its intended audience. Overall, the discussion emphasizes the need for a careful and critical evaluation of the available tools to make informed decisions about their use.

Additionally, the MyAsriGeo dashboard can be extended to cover other areas beyond drug abuse, such as cyberbullying, mental health issues, and academic performance. It can help stakeholders to identify potential risk factors that may contribute to drug abuse among students and take preventive measures.

It is important to note that the MyAsriGeo dashboard is not a stand-alone solution for addressing drug abuse among youth. A comprehensive strategy must complement it, including education, prevention, treatment, and law enforcement efforts [40]. The dashboard can provide stakeholders with the necessary information and insights to make informed decisions, but it is up to them to take action and address the issue of drug abuse among youth.

Limitations

The involvement of 20 experts in the evaluation provides some evidence for the reliability and validity of the findings; a larger sample size would strengthen the generalizability of the results. Additionally, the study was limited to evaluating the dashboard's user acceptance, and future research could explore the impact of the dashboard on actual decision-making processes and outcomes.

Future Work

Future research could explore the effectiveness of the MyAsriGeo dashboard in detecting and monitoring drug abuse trends over time. Longitudinal studies could assess whether the dashboard provides decision makers with accurate and timely information to identify and respond to emerging drug abuse patterns. Additionally, future work could investigate the potential for the dashboard to be adapted and implemented in different contexts, such as other regions or countries, to assess its scalability and generalizability. Finally, integrating additional data sources, such as social media, could enhance the dashboard's ability to provide decision makers with a comprehensive view of drug abuse trends.

Conclusions

The geospatial drug abuse dashboard used in this study was developed and evaluated with the help of experts and stakeholders. The results showed that the changes made between the initial and final versions of the dashboard significantly impacted users' perceptions of its ease of use and usefulness. This study demonstrates the importance of geospatial dashboards for drug abuse monitoring and analysis among students.

Acknowledgments

This work was supported in part by the Ministry of Higher Education, Malaysia, under the Long Research Grant Scheme (LRGS/1/2019/UKM-UKM/2/7) and the Modal Insan Postdoctoral fund from the Centre for Research and Instrumentation Management, Universiti Kebangsaan Malaysia.

Authors' Contributions

AMA-A contributed to the investigation, methodology, development, and preparation of the original draft. SNHSA provided supervision, conceptualized the project, and was responsible for project administration, as well as reviewing and editing the manuscript. RI was involved in the conceptualization and administration of the project. KNAM offered supervision, project administration, and contributed to reviewing and editing the manuscript. LN took part in conceptualizing the project, preparing the original draft, reviewing and editing, as well as developing the methodology. KAZA was responsible for reviewing and editing the manuscript and contributing to the methodology. MCL contributed to the conceptualization and administration of the project. MLBT was involved in conceptualizing the project. SW contributed to conceptualization and administration of the project. ME was responsible for data curation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Evaluation questionnaire and related work comparison.
[PDF File (Adobe PDF File), 316 KB - humanfactors_v11i1e48139_app1.pdf]

References

1. World Drug Report 2021: pandemic effects ramp up drug risks, as youth underestimate cannabis dangers. United Nations Office on Drugs and Crime. 2021. URL: https://www.un-ilibrary.org/content/books/9789210058032 [accessed 2024-05-08]
2. Ismail R, Manaf MRA, Hassan MR, Nawi AM, Ibrahim N, Lyndon N, et al. Prevalence of drug and substance use among malaysian youth: a nationwide survey. Int J Environ Res Public Health 2022;19(8):4684 [FREE Full text] [doi: 10.3390/ijerph19084684] [Medline: 35457562]

3. National Health and Morbidity Survey (NHMS) 2019: non-communicable diseases, healthcare demand, and health literacy-key findings. Institute for Public Health. 2019. URL: https://iptk.moh.gov.my/images/technical_report/2020/4_Infographic_Booklet_NHMS_2019_-_English.pdf [accessed 2024-05-08]
4. Drug and substance abusers and addicts. National Anti-Drug Agency. 2020. URL: https://www.adk.gov.my/en/public/drugs-statistics/ [accessed 2024-05-08]
5. Muhamad NA, Mihat O, Ramly R, Aziz AA, Kamaruddin R, Mansor WNAW, et al. Translation, cross-cultural adaptation, reliability and validity of the malay version of alcohol, smoking and substance involvement screening test (ASSIST) V3.1. Health 2018;10(07):985-997 [FREE Full text] [doi: 10.4236/health.2018.107073]
6. Drus ZAM, Singh D, Mokhtar MR, Rashid RA. Review of computerized cognitive behavioural therapy based on culture centered design for substance abuse in Malaysia. Asia Pac J Inf Technol Multimed 2018;07(01):119-132 [FREE Full text] [doi: 10.17576/apjitm-2018-0701-10]
7. Eric P. Socioeconomic effects of drug abuse among Nigerian youths. Can Sci Med 2017;13(1):49-53 [FREE Full text]
8. Mnunguli JP. Development of a web and mobile application for drug abuse information awareness. Nelson Mandela African Institution of Science and Technology. 2019. URL: http://dspace.nm-aist.ac.tz/handle/20.500.12479/259 [accessed 2024-05-08]
9. Hu H, Phan N, Ye X, Jin R, Ding K, Dou D. DrugTracker: a community-focused drug abuse monitoring and supporting system using social media and geospatial data (Demo Paper). 2019 Presented at: SIGSPATIAL '19: 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems; November 5-8, 2019; Chicago, IL, USA URL: https://dl.acm.org/doi/10.1145/3347146.3359076 [doi: 10.1145/3347146.3359076]
10. Chie QT, Tam CL, Bonn G, Wong CP, Dang HM, Khairuddin R. Drug abuse, relapse, and prevention education in Malaysia: perspective of university students through a mixed methods approach. Front Psychiatry 2015;6:65 [FREE Full text] [doi: 10.3389/fpsyt.2015.00065] [Medline: 25999867]
11. Hsiao CT, Chou FC, Hsieh CC, Chang LC, Hsu CM. Developing a competency-based learning and assessment system for residency training: analysis study of user requirements and acceptance. J Med Internet Res 2020;22(4):e15655 [FREE Full text] [doi: 10.2196/15655] [Medline: 32286233]
12. Liu S. Design and implementation of a geospatial dashboard for crime analysis and prediction. Toronto Metropolitan University. 2021. URL: https://rshare.library.torontomu.ca/articles/thesis/Design_And_Implementation_Of_A_Geospatial_Dashboard_For_Crime_Analysis_And_Prediction/16834417 [accessed 2024-05-08]
13. Few S. Information Dashboard Design: The Effective Visual Communication of Data. Beijing: O'Reilly Media, Inc; 2006.
14. Abdullah MA, Abdullah SNHS, Nordin MJ. Additional feet-on-the-street deployment method for indexed crime prevention initiative. J Pengur 2018;53:181-193 [FREE Full text] [doi: 10.17576/pengurusan-2018-53-16]
15. Alzahrani NA, Abdullah SNHS, Mohamed I, Mukred M. The adoption of geographic information systems in the public sector of Saudi Arabia: a conceptual model. Math Probl Eng 2021;2021:1-14 [FREE Full text] [doi: 10.1155/2021/1099256]
16. Jing C, Du M, Li S, Liu S. Geospatial dashboards for monitoring smart city performance. Sustainability 2019;11(20):5648 [FREE Full text] [doi: 10.3390/su11205648]
17. Ghazvini A, Abdullah SNHS, Hasan MK, Bin Kasim DZA. Crime spatiotemporal prediction with fused objective function in time delay neural network. IEEE Access 2020;8:115167-115183 [FREE Full text] [doi: 10.1109/access.2020.3002766]
18. Rochin MAE, Gutierrez-Garcia JO, Rosales JJ, Rodriguez LF. Design and evaluation of a dashboard to support the comprehension of the progression of patients with dementia in day centers. Int J Med Inform 2021;156:104617 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104617] [Medline: 34710725]
19. Niu Y, Ying L, Yang J, Bao M, Sivaparthipan CB. Organizational business intelligence and decision making using big data analytics. Inf Process Manag 2021;58(6):102725 [FREE Full text] [doi: 10.1016/j.ipm.2021.102725]
20. Yang J, Xiu P, Sun L, Ying L, Muthu B. Social media data analytics for business decision making system to competitive analysis. Inf Process Manag 2022;59(1):102751 [FREE Full text] [doi: 10.1016/j.ipm.2021.102751]
21. Kryskalla J, Kern S, Gray D, Hauser P. Using dashboard technology to monitor overdose risk. Fed Pract 2014;31(9):8-14 [FREE Full text]
22. Davis F. User acceptance of information systems: the technology acceptance model (TAM). Graduate School of Business.: University of Michigan; 1987. URL: https://deepblue.lib.umich.edu/bitstream/handle/2027.42/35547/b1409190.0001.001.pdf?seque [accessed 2024-05-08]
23. De Wulf K, Schillewaert N, Muylle S, Rangarajan D. The role of pleasure in web site success. Inf Manag 2006;43(4):434-446 [FREE Full text] [doi: 10.1016/j.im.2005.10.005]
24. Lewis JR. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int J Hum Comput 1995;7(1):57-78 [FREE Full text] [doi: 10.1080/10447319509526110]
25. Alibrahim A, Wu S, Guerrero E. Performance dashboard for the substance abuse treatment system in Los Angeles. 2014 Presented at: IIE Annual Conference and Expo 2014; May 31-June 3, 2014; Montréal, QC, Canada.
26. Wolf DAPS, BlackDeer AA, Beeler-Stinn S, Zheng K, Stazrad K. Performance-based practice: clinical dashboards for addiction treatment retention. Res Social Work Prac 2020;31(2):205-211 [FREE Full text] [doi: 10.1177/1049731520972798]

27. Tableau Software. Office of national drug control policy, drug control data dashboard. The White House. 2022. URL: https://www.whitehouse.gov/ondcp/drug-control-data-dashboard [accessed 2024-05-08]
28. Stone AB, Jones MR, Rao N, Urman RD. A dashboard for monitoring opioid-related adverse drug events following surgery using a national administrative database. Am J Med Qual 2019;34(1):45-52 [FREE Full text] [doi: 10.1177/1062860618782646] [Medline: 29938518]
29. Simpao AF, Ahumada LM, Desai BR, Bonafide CP, Gálvez JA, Rehman MA, et al. Optimization of drug-drug interaction alert rules in a pediatric hospital's electronic health record system using a visual analytics dashboard. J Am Med Inform Assoc 2015;22(2):361-369 [FREE Full text] [doi: 10.1136/amiajnl-2013-002538] [Medline: 25318641]
30. Coughlin L, Hull-Jilly D, Saxon S. Creating an instantatlas™ dashboard for Alaska substance use and abuse data. 2019 Presented at: 2019 CSTE Annual Conference; June 2-6, 2019; Raleigh, North Carolina.
31. Cohen D, Lindvall M, Costa P. An introduction to agile methods. In: Advances in Computers: Advances in Software Engineering. San Diego, CA, USA: Elsevier Science; 2004:1-66.
32. Budde R, Kautz K, Kuhlenkamp K, Züllighoven H. What is prototyping? Inf Technol People 1990;6(2/3):89-95 [FREE Full text] [doi: 10.1108/EUM0000000003546]
33. de Oliveira Sousa A, Valentim NMC. Prototyping usability and user experience: a simple technique to agile teams. 2019 Presented at: SBQS'19: XVIII Brazilian Symposium on Software Quality; October 28-November 1, 2019; Fortaleza Brazil p. 222-227 URL: https://dl.acm.org/doi/abs/10.1145/3364641.3364667 [doi: 10.1145/3364641.3364667]
34. Mooney RL, Gordon LV. Mooney problem check lists. PsycTESTS Dataset. 1950. URL: https://psycnet.apa.org/doiLanding?doi=10.1037%2Ft15123-000 [accessed 2024-05-08]
35. Humeniuk R, Henry-Edwards S, Ali R, Poznyak V, Monteiro MG. The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): manual for use in primary care. World Health Organization. 2010. URL: https://iris.who.int/handle/10665/44320 [accessed 2024-05-08]
36. Dowding D, Randell R, Gardner P, Fitzpatrick G, Dykes P, Favela J, et al. Dashboards for improving patient care: review of the literature. Int J Med Inform 2015;84(2):87-100 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.10.001] [Medline: 25453274]
37. Buttigieg SC, Pace A, Rathert C. Hospital performance dashboards: a literature review. J Health Organ Manag 2017;31(3):385-406 [FREE Full text] [doi: 10.1108/JHOM-04-2017-0088] [Medline: 28686130]
38. Cao L, Ramesh B. Agile requirements engineering practices: an empirical study. IEEE Softw 2008;25(1):60-67 [FREE Full text] [doi: 10.1109/ms.2008.1]
39. Wu E, Villani J, Davis A, Fareed N, Harris DR, Huerta TR, et al. Community dashboards to support data-informed decision-making in the HEALing communities study. Drug Alcohol Depend 2020;217:108331 [FREE Full text] [doi: 10.1016/j.drugalcdep.2020.108331] [Medline: 33070058]
40. Ismail R, Shafurdin NS, Shukor MS, Mohammed Nawi A, Abdul Manaf MR, Ibrahim N, et al. Predictors of drug and substance abuse among school-going adolescents living in drug hotspot in Malaysia. PLoS One 2024;19(6):e0305460 [FREE Full text] [doi: 10.1371/journal.pone.0305460] [Medline: 38875293]

Abbreviations

API: application programming interface
ASSIST: Alcohol, Smoking, and Substance Involvement Screening Test
PSSUQ: Post-Study System Usability Questionnaire
TAM: Technology Acceptance Model
TTF: Task-Technology Fit

Edited by A Kushniruk; submitted 12.04.23; peer-reviewed by L Zheng, E Prochaska, D Carvalho; comments to author 09.01.24; revised version received 06.02.24; accepted 02.03.24; published 30.07.24.

Please cite as:
Al-Aboosi AM, Sheikh Abdullah SNH, Ismail R, Abdul Maulud KN, Nahar L, Zainol Ariffin KA, Lam MC, bin Talib ML, Wahab S, Elias M
A Geospatial Drug Abuse Risk Assessment and Monitoring Dashboard Tailored for School Students: Development Study With Requirement Analysis and Acceptance Evaluation
JMIR Hum Factors 2024;11:e48139
URL: https://humanfactors.jmir.org/2024/1/e48139
doi:10.2196/48139
PMID:39078685

Original Paper

Assessing Patient Trust in Automation in Health Care Systems: Within-Subjects Experimental Study

Matthew Nare[1*], BSc, MSc; Katherina Jurewicz[1*], BSc, MSc, PhD

School of Industrial Engineering and Management, Oklahoma State University, Stillwater, OK, United States

[*]all authors contributed equally

Corresponding Author:
Katherina Jurewicz, BSc, MSc, PhD
School of Industrial Engineering and Management
Oklahoma State University
329 Engineering North
Stillwater, OK, 74078
United States
Phone: 1 405 744 4167
Email: katie.jurewicz@okstate.edu

Abstract

Background: Health care technology has the ability to change patient outcomes for the betterment when designed appropriately. Automation is becoming smarter and is increasingly being integrated into health care work systems.

Objective: This study focuses on investigating trust between patients and an automated cardiac risk assessment tool (CRAT) in a simulated emergency department setting.

Methods: A within-subjects experimental study was performed to investigate differences in automation modes for the CRAT: (1) no automation, (2) automation only, and (3) semiautomation. Participants were asked to enter their simulated symptoms for each scenario into the CRAT as instructed by the experimenter, and they would automatically be classified as high, medium, or low risk depending on the symptoms entered. Participants were asked to provide their trust ratings for each combination of risk classification and automation mode on a scale of 1 to 10 (1=absolutely no trust and 10=complete trust).

Results: Results from this study indicate that the participants significantly trusted the semiautomation condition more compared to the automation-only condition (P=.002), and they trusted the no automation condition significantly more than the automation-only condition (P=.03). Additionally, participants significantly trusted the CRAT more in the high-severity scenario compared to the medium-severity scenario (P=.004).

Conclusions: The findings from this study emphasize the importance of the human component of automation when designing automated technology in health care systems. Automation and artificially intelligent systems are becoming more prevalent in health care systems, and this work emphasizes the need to consider the human element when designing automation into care delivery.

(JMIR Hum Factors 2024;11:e48584) doi:10.2196/48584

KEYWORDS

automation; emergency department; trust; health care; artificial intelligence; emergency; perceptions; attitude; opinions; belief; automated; trust ratings

Introduction

Delays in care can come from any number of factors that influence the health care work system, including technology and other automated computing tools that support patient care [1]. Understanding the dynamics between humans and technology, such as automated technology, allows designers to develop the appropriate equipment for completing patient care tasks in different clinical environments. Advancements in technology (eg, artificial intelligence and smart medical devices) bring new applications in health care with the intent to improve the delivery of care [2]. New technologies that are currently being implemented in health care systems change the way hospitals maintain a patient's health record [3] and surgeons conduct surgery [4,5] and assist with patient mobility [6]. Advancements in smart health care technology have allowed health care work systems to become increasingly interconnected.

The adoption of electronic health records or electronic medical records opened communication pathways between patients and clinicians [7]. Videoconferencing capabilities have advanced the opportunities for remotely conducting appointments and inpatient communication [8]. With the increased use of technology within health care work systems, it is imperative to understand the short- and long-term effects of that technology.

One area where technology has made significant strides is with automation. Automation has been defined in many ways, but the consistent theme across the definitions is a technology completing a task for humans [9,10]. Parasuraman and Riley [10] stated as part of their definition of automation that "what is considered automation will therefore change with time." The evolution of automation is portrayed through the types of technologies and industries integrating automation into daily workflows. However, in order for automated technology to be effectively used, it is important for the human user to trust the output of the automated system. Trust, like automation, has received many operational definitions depending on the context and field [11,12]. Development of trust in automation is a function of many factors including personality and system performance. However, humans are known to overtrust or become overreliant on automation, which has been known to cause unwanted outcomes [9]. Automation must be designed to promote trust in a way that adapts to its users' needs while lending itself to allow for the appropriate amount of reliance [9,13]. Chiou and Lee [13] described the importance of designing adaptive and resilient automation to improve the trust dynamics between automation and its users. Automation errors occur when there are discrepancies between the amount of trust and the capabilities of the system [9].

There are several areas within the health care system that can benefit from automated technologies including high-risk environments, such as emergency medicine. Emergency medicine is a dynamic work environment, and there is a constant turnover of patients presenting with varying symptoms ranging in severity in emergency medicine. Health care providers use a triage system to prioritize arriving patients based on the severity of the diagnosis. Health conditions such as myocardial infarctions (ie, heart attacks) and cerebrovascular attacks (ie, strokes) require immediate medical intervention. A delay in care can have irrevocable effects on a patient's health. Heart disease is a leading cause of death in the United States, and it is estimated over 800,000 people in the United States experience a heart attack every year [14]. Despite the prevalence of myocardial infarctions in the United States, the symptoms continue to be missed within emergency departments (EDs) [15]. Integrating automation into EDs capable of making an accurate assessment of a patient's symptoms has the potential to assist one aspect of a clinician's patient intake workflow.

Health care automation has the capability of improving clinical outcomes and hospital environments; however, for this to occur, automation must be designed in a way that allows for seamless integration into the work environment. To improve both health care providers' and patients' acceptance of automated technology, automation must be designed to promote the appropriate level of trust within its users [9,13]. The first step to designing automation that promotes the appropriate level of trust is to understand which aspects of the technology enhance or detract trust in all health care providers responsible for operating or interpreting the results of automated technology (ie, physicians and nurses). While the health care providers are the primary user group when it comes to health care automation, it is vital to understand the patient's perspective, as patients may have reservations for integrating automation into health care work systems [16]. Although health care consists of both providers and patients, there is little literature investigating the automation needs for both parties. Therefore, the objective of this study was to study the effect of automation mode on a patient's trust in a cardiac risk assessment tool (CRAT). A secondary objective was to study the effect of the severity of symptoms on a patient's trust in the CRAT.

Methods

Study Design

A within-subjects experimental study was performed to assess a patient's trust in the CRAT. The CRAT was designed to have three automation modes: (1) no automation, (2) automation only, and (3) semiautomation. No automation was defined as a physician acting as the cardiac risk assessor without any technology, the automation-only condition was defined as a fully automated risk assessment tool that had no human intervention, and the semiautomation condition consisted of the risk assessment tool that was validated by the experimenter.

Participants

In total, 12 participants were recruited from undergraduate and graduate students at Oklahoma State University and the general Stillwater, Oklahoma community to participate in this study.

Ethical Considerations

This study received exempt status from the Oklahoma State University Institutional Review Board (IRB-22-391-STW). All participants completed the informed consent process with the experimenter, and all data were anonymized. Each participant received a $10 Amazon gift card for their time.

Equipment and Materials

A digital application prototype was developed with Adobe XD (Adobe). The prototype displayed 16 potential health status symptoms with a touch-activated check box (Figure 1). Participants received a risk classification after checking their symptoms and selecting the submit button (Figure 2). Participants navigated the application using a Microsoft Surface tablet (Microsoft Corp).

Figure 1. Digital symptom application prototype showing 16 symptom options and checkboxes.

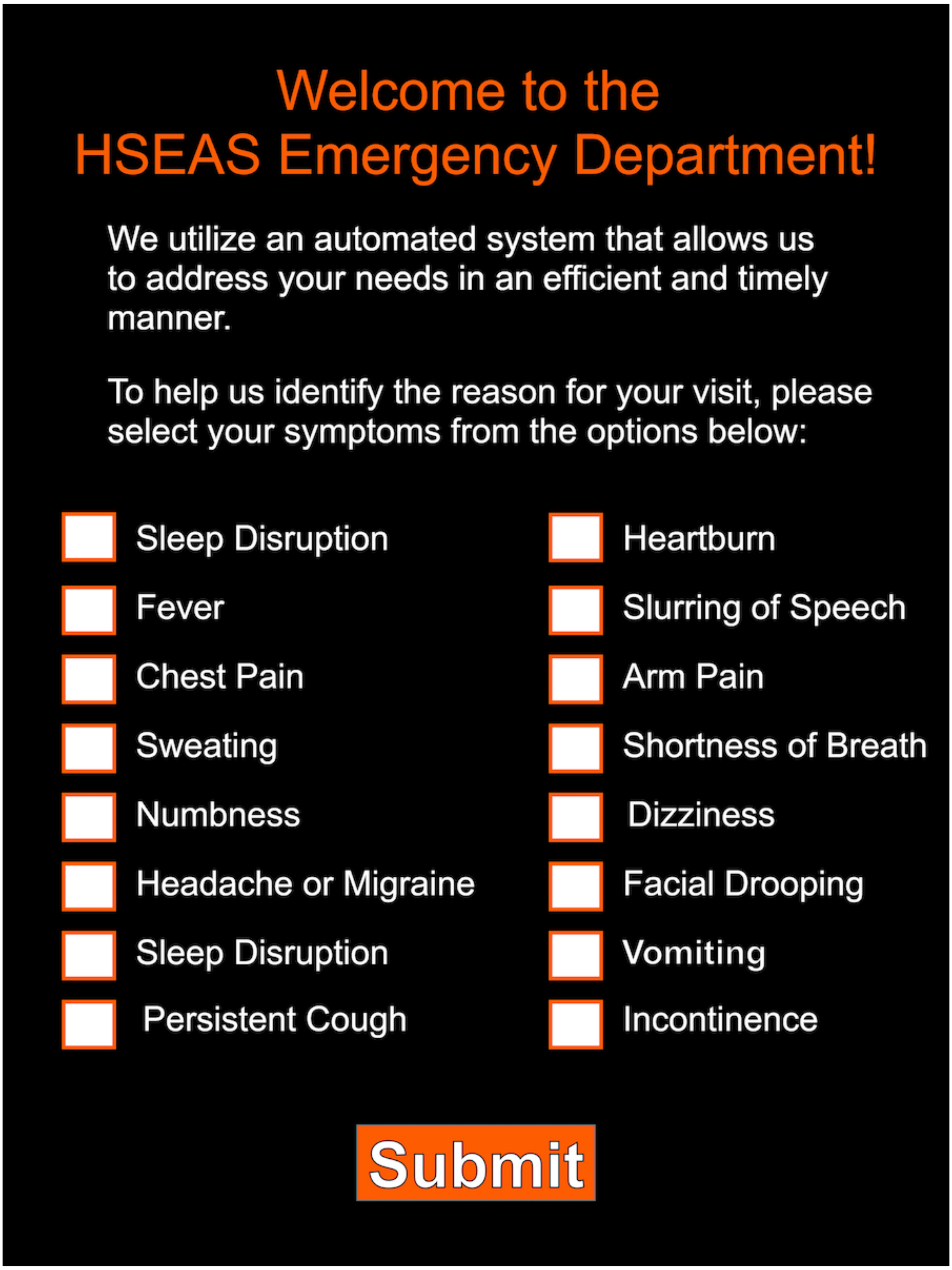
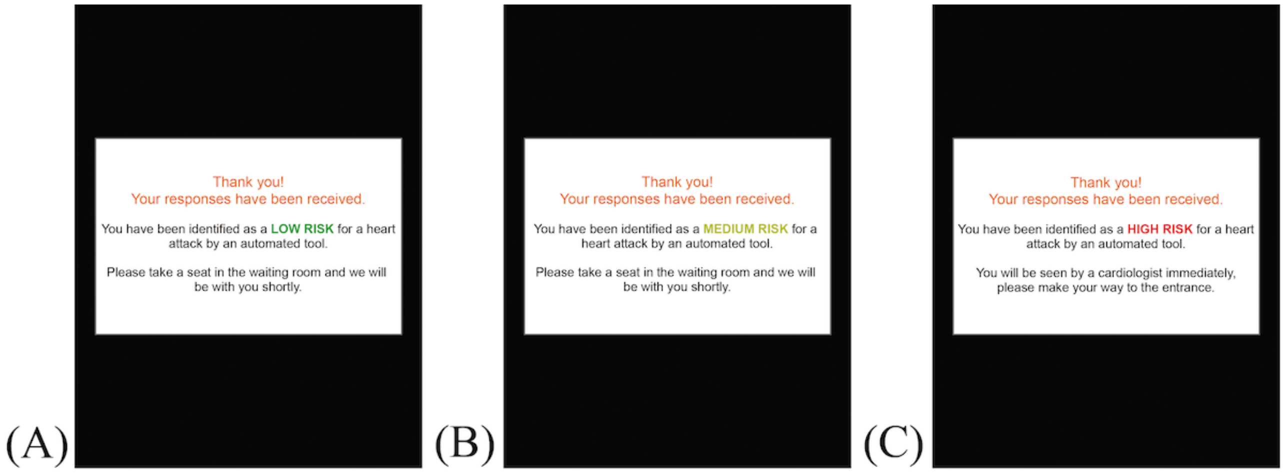


Figure 2. Digital symptom application prototype risk classification message for each risk classification level: (A) low, (B) medium, and (C) high risks.



Procedure

Upon arriving, the participants were welcomed to the Human-Systems Engineering Applied Statistics (HSEAS) Lab, and the informed consent process was completed with the participants. Participants were then seated at a table facing the computer monitor to complete the demographics questionnaire to collect data about the participants' personal characteristics (eg, school status, age, and ethnicity), personal or family professional health care experience, experience as a patient or family member in an ED, the tendency to research medical information on the web, and subjective ratings on their quality of health.

Participants completed a training session on heart attack symptoms by reading a document titled "What are the warning signs of a heart attack?" [17]. This document ensured that participants understood the symptoms of a heart attack before completing the experiment. Participants were then relocated to a simulated waiting room area. For the remainder of the experimental session, the participant assumed the role of a patient arriving at the HSEAS Lab ED (ie, the simulated waiting room at the HSEAS Lab), and the experimenter performed the role of the health care provider. Each participant gave trust ratings for 3 different scenarios for each mode of information.

Automation mode (eg, automation only, no automation, or semiautomation) was presented across 3 blocks. In the automation-only condition, the participant used the tablet to independently submit their symptoms and only received their risk classification from the tablet. The semiautomation condition consisted of the patient (ie, the participant) working with the health care provider (ie, the researcher) to submit the correct symptoms and then receiving a risk classification from the tablet that was confirmed by the health care provider. Thus, the experimenter and the participant worked together to enter the symptoms. In the no automation condition, participants verbally stated their symptoms to the health care provider and were verbally provided their risk classification. The automation modes (eg, automation only, no automation, or semiautomation) were based on the trust in the automation framework presented by Chiou and Lee [13].

The order of the blocks, and thus automation mode, was counterbalanced across the participants. Participants were then classified as high risk, medium risk, or low risk based on the symptoms provided in the symptom severity scenarios (Table 1). Participants were provided the written scenario (ie, symptom severity scenario) that included the present symptoms and a subjective description of how they were feeling as a result of those symptoms. The risk levels were randomized within the block; therefore, each participant completed 9 different scenarios for each automation mode and risk level classification.

Table 1. Risk classification symptoms and subjective description.

Risk severity	Present symptoms	Lacking symptoms	Subjective description
High risk	• Sweating • Shortness of breath • Severe chest pain • Nausea	• None	• Feel "like an elephant is sitting on my chest"
Medium risk	• Sweating • Nausea • Mild chest discomfort	• Severe chest pain • Shortness of breath	• Feel "like there is a mild pain in my chest"
Low risk	• Rapid heart rate	• Shortness of breath • Sweating • Severe chest pain • Nausea	• Feel "like something is wrong because my heart is beating fast"

During this study, the prototype provided the correct risk classification for each symptom presentation scenario using a "Wizard of Oz" technique; thus, the automation was always correct in the risk classification. Participants were informed that there may be errors in the risk classification and that the submitted symptoms will help fine-tune the output of the CRAT. At the end of each scenario, participants were asked to provide their rating of trust in the provided risk classification on a scale of 1 to 10 (1=absolutely no trust and 10=complete trust); therefore, each participant had 1 trust rating for every combination of automation mode and risk level for a total of 9 ratings per participant. The participants' subjective ratings of trust were collected following a similar approach presented in previous literature on measuring trust in automation via Likert scale ratings [18-20].

Data Analysis

A mixed effects linear regression model, with trust as the dependent variable, was fitted to the data, and stepwise deletion was performed until the best fit linear regression model was determined. Tukey contrasts were calculated to perform pairwise comparisons for all variables with multiple levels in the regression model. All statistical analyses were performed in RStudio (version 2024.04.1+748; Posit Software, PBC) using the *lme4* package (version 1.1-26) [21] and the *multcomp* package (version 1.4-16) [22].

Results

A majority of participants identified as female (n=8), and the rest identified as male (n=4). The age of participants ranged between 18 and 52 (mean 24.56, SD 9.10) years. In total, 6 participants were current students pursuing either an undergraduate or graduate degree at Oklahoma State University, and 6 participants were members of the greater Stillwater community (ie, nonstudents). There was 1 participant who reported working in the medical field, and 7 of the participants reported having family members who worked in the medical field. A majority of 11 participants reported having been to a doctor in the past year with 10 participants having between 2 and 4 doctor visits in the past year. In total, 7 participants had previously visited an ED as a patient and 6 participants had visited an ED with a family member. Overall, 3 participants had visited an ED in the past year. The frequency for all trust ratings and the percentage of the total response are provided in Table 2. The average trust rating overall across all conditions was 8.57 (SD 1.56).

Table 2. Summary of trust ratings by the frequency of occurrence (n=108).

Trust rating	Values, n (%)
0	—[a]
1	—
2	—
3	1 (0.9)
4	1 (0.9)
5	3 (2.8)
6	6 (5.6)
7	16 (14.8)
8	13 (12)
9	28 (25.9)
10	40 (37)

[a]Not available.

The average trust ratings for automation modes and risk levels are provided in Table 3. Between automation modes, the lowest trust ratings were provided for automation-only condition (mean 7.68, SD 1.42) and the highest trust ratings for the semiautomated condition (mean 9.08, SD 1.44). Overall, the highest trust ratings were for the high risk level in the semiautomated condition (mean 9.58, SD 0.86), and the lowest trust ratings were for the medium risk in the automation-only condition (mean 6.91, SD 1.38).

Table 3. Summary of trust rating.

Automation mode and risk level	Average trust rating, mean (SD)	Overall average trust rating, mean (SD)
Automation only		
Low	8.08 (1.19)	7.86 (1.42)
Medium	6.91 (1.38)	7.86 (1.42)
High	8.58 (1.11)	7.86 (1.42)
Semiautomation		
Low	9.25 (1.01)	9.08 (1.44)
Medium	8.42 (1.93)	9.08 (1.44)
High	9.58 (0.86)	9.08 (1.44)
No automation		
Low	8.67 (1.93)	8.78 (1.53)
Medium	8.50 (1.38)	8.78 (1.53)
High	9.17 (1.07)	8.78 (1.53)

The mixed effects model showed that the random effect was nonsignificant in explaining any additional variability; however, to provide the most accurate representation of the data, the random effect was included in the final model. The results of the final regression model are shown in Table 4. The Akaike information criterion for the linear mixed effects model was 391.155. There was no interaction term included in this model.

Table 4. Linear regression table for the final linear mixed effects model.

Coefficients	Estimate (SE)	*t* test (*df*)	*P* value
Intercept	8.791 (0.372)	23.636 (92)	<.001
Automation only	–1.222 (0.316)	–3.869 (92)	<.001
No automation	–0.306 (0.316)	–0.967 (92)	.34
High risk level	1.167 (0.316)	–3.693 (92)	<.001
Low risk level	0.722 (0.316)	2.286 (92)	.02
EDPatientYes[a]	–0.578 (0.351)	–1.647 (10)	.13

[a]ED: emergency department.

Automation mode (ie, automation only; *P*<.001) and risk level (ie, low and high risk; *P*=.02 and *P*<.001) were significantly associated with the participants' trust ratings. The Tukey pairwise comparisons for automation mode indicated that there were significant differences in the participants trust between no automation and automation only (*P*=.03) as well as automation only and semiautomation (*P*=.002).

Trust in the risk classification was also significantly lower when participants were classified as medium risk compared to a high-risk classification (*P*=.004). The Tukey pairwise comparisons showed no other significant differences between risk classification levels.

Discussion

Principal Findings

As automation becomes more integrated into health care work systems, it is important to understand the implications of automation on the patients and clinicians. Investigating patients' perceptions toward the use of automated technology, such as their trust, represents one such area of research. The purpose of this study was to investigate patients' perceptions of an automated CRAT prototype developed to improve the triage process for patients with cardiac symptoms arriving at an ED. Findings from this study indicate that the factors significantly associated with trust were the automation mode and risk level.

Participants reported significantly higher ratings of trust in the risk classification when a human participated (ie, no automation and semiautomation conditions) in passing on the information to the patient compared to when the risk classification was only presented with the CRAT (ie, automation-only condition). These findings support the value of keeping the human, both patients and clinicians, involved when integrating health care automation into hospitals. Health care traditionally includes a close patient-provider interaction [2]; however, as technology has evolved, patient-provider interactions adapt to incorporate more technology into patient care. Unlike other industries where the roles of humans decrease with automation, health care technology requires more human involvement to properly monitor large amounts of data introduced during Health Care 4.0 [2].

There was a significant difference in trust ratings for the high-risk scenarios and the medium-risk scenarios. The high-risk scenario provided symptoms that were easily interpreted by the participants; however, the symptoms for the medium-risk scenario were more ambiguous. While the high-risk scenario incorporated all the traditional heart attack symptoms (eg, complaints of severe chest pain, shortness of breath, and nausea), the medium-risk scenario presented symptoms with multiple interpretations (eg, mild chest discomfort could be interpreted as heartburn). The difference in trust ratings between the high- and medium-risk scenarios suggests that there may be a relationship between the quality of information used by the automation to draw its conclusion and trust. Given a participant's lack of trust based on their individual differences, it is important to understand how ambiguous situations, such as the medium-risk scenario, can challenge their trust in the output [23,24]. It is possible that the participants trusted the automation more during the clear scenarios compared to the ambiguous scenario because of their own automation complacency or bias [25]. The medium-risk scenario may have been ambiguous since the participant could have interpreted the symptoms as more or less severe; therefore, it may have required more thought on whether the risk classification was correct. The ambiguity may have increased the level of doubt the participants experienced in the CRAT's output as trust is dynamic and constantly changing as different situations unfold [11,13].

All of the components of the health care work system are interconnected; therefore, this means that the performance of one work system element influences the other elements [1]. If we remove the human operator from the work system, there may be a negative effect on the care processes, which influence patient and organization outcomes. With hospitals focused on patient-centered care and patient outcomes, it is important to consider how patients perceive the technology used in care environments [8,26]. As new technology, such as artificial intelligence and automated technology, is integrated into health care work systems, it is important to appropriately design the interactions between humans and technology, especially designing for trust in automation [13].

The use of automated technology to assist in the intake and triage of patients has the potential to simultaneously support providers' work and the patient experience. The full extent of this potential is dependent on developing automation capable of providing trustworthy and accurate information to both parties. Chiou and Lee [13] described a framework where automation and the human operator were influenced by the context of repeated interactions between single human operator and one form of automation. In health care, interaction with automation may include simultaneous use effects for both patients and clinicians. Patients are primarily on the receiving end for treatments provided by automation, while clinicians are responsible for the actions that lead to automation-driven treatment outcomes. Given the clinician and patient are both involved in the treatment process, both parties experience consequences when errors occur. During one ED experience, patients may receive treatment from several devices from several clinicians, and clinicians may operate several devices while delivering care to several different patients. Automation needs to be designed in a way to mitigate the risk of errors occurring due to the work pressures of clinical environments.

Limitations

This study emphasized the importance of humans in the development of automated technology for health care, and there are several areas of future work. The CRAT, while developed based on input from health care providers and material, was not validated as being a system that would be integrated into an ED, which means there is an opportunity to continue to develop the CRAT's viability and usability using ED provider feedback. Future work should also investigate the trust and perceptions of ED providers to understand how the use of automated technology such as the CRAT changes provider workload and care quality from their perspective as well as further investigations into patients' trust. This study focused on studying automation mode; however, individual differences, such as language, ethnicity, race, or cultural diversity, all of which may influence how a person trusts technology, may be important to study in future work. This study has a limited sample size of 12 participants, and we would like to continue this work with more participants in the future. Additionally, improving the diversity of future studies, such as expanding to a wider age range and number of participants, will provide the opportunity to gain more insights into the challenges of trusting automation.

To our knowledge, this is the first study investigating trust in a risk assessment tool such as the one studied here. The small sample size and simulated setting (ie, a simulated ED rather than a live ED) were selected to test the viability of this work while being able to remove the risk of negative impact on patient care in a live ED. Selecting heart attack risk as the diagnosis of focus for this study allowed for trust ratings to be gathered on an adverse health event, which has widely known consequences across the general populace. With that in mind, future work should investigate the trust in risk assessment tools for other adverse diagnoses (eg, neurological events). Furthermore, as automation is integrated into health care systems, designing for failures in automation (eg, misclassification of heart attack risk) should be studied in future work as well as the effect of automation failures on trust. Future investigation into the effect of automation failures is vital to understand how trust changes within a health care environment, where each care decision carries positive and negative outcomes for the patient and the provider. This work offers a foundation for trust in automation within health care to build on as future research with larger, more diverse sample sizes and greater ecologically valid environments to improve the generalizability of this work.

Conclusions

Automation and artificial intelligence are used to support clinical decision-making and ease the workload of health care providers. Using automation and artificial intelligence to support accurate and time-efficient decisions demonstrates new opportunities for advancing patient-centered care. Designing trustworthy automation can provide additional support when a clinical environment is understaffed or without access to specialty care such as rural clinics. This work represents an important first step at quantifying how patients trust automated technology within health care. The results from this study indicate that patients trust technology most when there is a combination of human and automation interaction throughout the care process and when there is little to no ambiguity based on their symptoms. Keeping the human involved in the process enhances the transparency of the automation with the patient, which can improve trust. Designing automation with the patient and clinician in mind is important when attempting to integrate automation into health care work environments, where errors can lead to irrevocable consequences such as patient death.

Conflicts of Interest

None declared.

References

1. Carayon P, Hundt AS, Karsh B, Gurses AP, Alvarado CJ, Smith M, et al. Work system design for patient safety: the SEIPS model. BMJ Qual Saf 2006;15(Suppl 1):i50-i58. [doi: 10.1136/qshc.2005.015842] [Medline: 17142610]
2. Li J, Carayon P. Health care 4.0: a vision for smart and connected health care. IISE Trans Healthc Syst Eng 2021;11(3):171-180 [FREE Full text] [doi: 10.1080/24725579.2021.1884627] [Medline: 34497970]
3. Kohli R, Tan SSL. Electronic health records: how can IS researchers contribute to transforming healthcare? MIS Q 2016;40(3):553-573. [doi: 10.25300/misq/2016/40.3.02]
4. Brodie A, Vasdev N. The future of robotic surgery: how robotics could help shape the future of surgical care. Ann R Coll Surg Engl 2018;100(Suppl 7):4-13 [FREE Full text] [doi: 10.1308/rcsann.supp2.4] [Medline: 30179048]
5. Long E, Kew F. Patient satisfaction with robotic surgery. J Robot Surg 2018;12(3):493-499. [doi: 10.1007/s11701-017-0772-3] [Medline: 29288373]
6. Krishnan RH, Pugazhenthi S. Mobility assistive devices and self-transfer robotic systems for elderly, a review. Intell Serv Robot 2013;7(1):37-49. [doi: 10.1007/s11370-013-0142-6]
7. Jha AK, DesRoches CM, Campbell EG, Donelan K, Rao SR, Ferris TG, et al. Use of electronic health records in U.S. hospitals. N Engl J Med 2009;360(16):1628-1638 [FREE Full text] [doi: 10.1056/NEJMsa0900592] [Medline: 19321858]
8. Lazzara EH, Benishek LE, Patzer B, Gregory ME, Hughes AM, Heyne K, et al. Utilizing telemedicine in the trauma intensive care unit: does it impact teamwork? Telemed e-Health 2015;21(8):670-676. [doi: 10.1089/tmj.2014.0074] [Medline: 25885369]
9. Lee JD, See KA. Trust in automation: designing for appropriate reliance. Hum Factors 2004;46(1):50-80. [doi: 10.1518/hfes.46.1.50_30392] [Medline: 15151155]
10. Parasuraman R, Riley V. Humans and automation: use, misuse, disuse, abuse. Hum Factors 1997;39(2):230-253. [doi: 10.1518/001872097778543886]
11. Hoff KA, Bashir M. Trust in automation: integrating empirical evidence on factors that influence trust. Hum Factors 2015;57(3):407-434. [doi: 10.1177/0018720814547570] [Medline: 25875432]
12. Schaefer KE, Chen JYC, Szalma JL, Hancock PA. A meta-analysis of factors influencing the development of trust in automation: implications for understanding autonomy in future systems. Hum Factors 2016;58(3):377-400. [doi: 10.1177/0018720816634228] [Medline: 27005902]
13. Chiou EK, Lee JD. Trusting automation: designing for responsivity and resilience. Hum Factors 2023;65(1):137-165 [FREE Full text] [doi: 10.1177/00187208211009995] [Medline: 33906505]
14. Tsao CW, Aday AW, Almarzooq ZI, Alonso A, Beaton AZ, Bittencourt MS, et al. Heart Disease and Stroke Statistics-2022 update: a report from the American Heart Association. Circulation 2022;145(8):e153-e639 [FREE Full text] [doi: 10.1161/CIR.0000000000001052] [Medline: 35078371]
15. Schull MJ, Vermeulen MJ, Stukel TA. The risk of missed diagnosis of acute myocardial infarction associated with emergency department volume. Ann Emerg Med 2006;48(6):647-655. [doi: 10.1016/j.annemergmed.2006.03.025] [Medline: 17112926]
16. Richardson JP, Smith C, Curtis S, Watson S, Zhu X, Barry B, et al. Patient apprehensions about the use of artificial intelligence in healthcare. NPJ Digit Med 2021;4(1):140 [FREE Full text] [doi: 10.1038/s41746-021-00509-1] [Medline: 34548621]
17. What are the warning signs of a heart attack? American Heart Association. 2015. URL: https://www.heart.org/-/media/Files/Health-Topics/Answers-by-Heart/What-Are-the-Warning-Signs-of-Heart-Attack.pdf [accessed 2024-04-20]
18. Merritt SM. Affective processes in human-automation interactions. Hum Factors 2011;53(4):356-370. [doi: 10.1177/0018720811411912] [Medline: 21901933]
19. Elder H, Canfield C, Shank DB, Rieger T, Hines C. Knowing when to pass: the effect of AI reliability in risky decision contexts. Hum Factors 2024;66(2):348-362 [FREE Full text] [doi: 10.1177/00187208221100691] [Medline: 35603703]
20. Lee JD, Moray N. Trust, self-confidence, and operators' adaptation to automation. Int J Hum-Comput Stud 1994;40(1):153-184. [doi: 10.1006/ijhc.1994.1007]
21. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67(1):1-48 [FREE Full text] [doi: 10.18637/jss.v067.i01]

22. Hothorn T, Bretz F, Westfall P. Simultaneous inference in general parametric models. Biom J 2008;50(3):346-363. [doi: 10.1002/bimj.200810425] [Medline: 18481363]
23. Gillath O, Ai T, Branicky MS, Keshmiri S, Davison RB, Spaulding R. Attachment and trust in artificial intelligence. Comput Hum Behav 2021;115:106607. [doi: 10.1016/j.chb.2020.106607]
24. Huo W, Zheng G, Yan J, Sun L, Han L. Interacting with medical artificial intelligence: integrating self-responsibility attribution, human–computer trust, and personality. Comput Hum Behav 2022;132:107253. [doi: 10.1016/j.chb.2022.107253]
25. Parasuraman R, Manzey DH. Complacency and bias in human use of automation: an attentional integration. Hum Factors 2010;52(3):381-410. [doi: 10.1177/0018720810376055] [Medline: 21077562]
26. Keebler JR, Rosen MA, Sittig DF, Thomas E, Salas E. Human factors and ergonomics in healthcare: industry demands and a path forward. Hum Factors 2022;64(1):250-258 [FREE Full text] [doi: 10.1177/00187208211073623] [Medline: 35000407]

Abbreviations

CRAT: cardiac risk assessment tool
ED: emergency department
HSEAS: Human-Systems Engineering Applied Statistics

Edited by A Kushniruk; submitted 30.04.23; peer-reviewed by D Neyens, D Kohen; comments to author 24.06.23; revised version received 11.10.23; accepted 15.12.23; published 06.08.24.

Please cite as:
Nare M, Jurewicz K
Assessing Patient Trust in Automation in Health Care Systems: Within-Subjects Experimental Study
JMIR Hum Factors 2024;11:e48584
URL: https://humanfactors.jmir.org/2024/1/e48584
doi:10.2196/48584
PMID:39106096

Original Paper

Validating the Effectiveness of the Patient-Centered Cancer Care Framework by Assessing the Impact of Work System Factors on Patient-Centered Care and Quality of Care: Interview Study With Newly Diagnosed Cancer Patients

Safa Elkefi[1], PhD; Onur Asan[2], PhD

[1]School of Nursing, Columbia University, New York, NY, United States

[2]School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ, United States

Corresponding Author:
Onur Asan, PhD
School of Systems and Enterprises
Stevens Institute of Technology
1 Castle Point Terrace
Hoboken, NJ, 07030
United States
Phone: 1 4145264330
Email: oasan@stevens.edu

Abstract

Background: Patients with cancer who have recently been diagnosed have distinct requirements compared to cancer survivors. It is crucial to take into account their unique needs to ensure that they make informed decisions and are receptive to the care provided.

Objective: This study suggested a framework titled Effectiveness of Patient-Centered Cancer Care that considers the needs of newly diagnosed patients with cancer and related work system factors. This study investigated how work system factors influence the perceptions of patient-centered care, quality of care, and associated outcomes among newly diagnosed patients with cancer. Patient-centered care is defined in terms of workload and communication considerations, whereas the quality of care is assessed through indicators such as trust in physicians, satisfaction with care, and perceptions of technology.

Methods: This study used qualitative data collected through interviews with newly diagnosed patients with cancer (N=20) right after their first visits with their physicians. Thematic analysis was conducted to validate the 5 hypotheses of the framework, mapping the interactions among quality of care, patient-centered care, and work system factors.

Results: We found that workload and patient-centered communication impact the quality of care and that the work system elements impact the patient-centeredness (workload and communication) and the quality of care (trust in physicians, satisfaction with care, and perception of technology use).

Conclusions: Qualitatively validating the proposed Effectiveness of Patient-Centered Cancer Care framework, this study demonstrated its efficacy in elucidating the interplay of various factors. The framework holds promise for informing interventions geared toward enhancing patients' experiences during their initial visits after diagnosis. There is a pressing need for heightened attention to the organizational design, patient processes, and collaborative efforts among diverse stakeholders and providers to optimize the overall patient experience.

(JMIR Hum Factors 2024;11:e53053) doi:10.2196/53053

KEYWORDS

cancer; communication; trust; satisfaction; technology; workload; work system factors

Introduction

Background

Improving the quality of care (QOC), coordination, and quality of life are essential goals of chronic care [1]. Patient-centered care (PCC) is one of the approaches used to assure the primacy of the individual's health and life goals in their care management [1]. In recent years, the concept of having the person be the driving force in their health care decisions has evolved and gained momentum, and it is now largely considered the gold standard for health care worldwide [1,2].

The initial physician visits after a cancer diagnosis are a critical period in which patients face a range of challenges that can significantly disrupt their lives. Symptoms of the disease and the overwhelming decisions related to treatment can pose a threat to their physical, cognitive, and emotional well-being [3]. Patients often struggle to comprehend medical information and express frustration with prolonged waiting periods for prognoses and follow-ups [3]. This can lead to psychosocial concerns, including high levels of distress, emotional strain, uncertainty surrounding mortality, and disruptions to social life [3]. The cognitive and emotional burdens can be overwhelming, potentially leading to nonadherence to treatment plans [4].

PCC approaches are considered crucial for the delivery of high-quality care to patients. However, there is considerable ambiguity concerning the exact meaning of the term and the optimal method for measuring the process and outcomes of PCC [5]. Despite the concept's popularity in the past 30 years, there has been a slight argument of perspective in the literature about the definition of PCC [5]. It has been an evolving concept, originally presented by Balint [6], who described patient-centered medicine as understanding the patient as a unique human being, whereas for Levenstein et al [7], it is an approach in which the "physician enters the patient's world to see the illness through his eyes." In 1998, Delbanco et al [8] developed a self-described utopian vision for a patient-centered health care system called People Power. The relationship is supported by "computer-based guidance and communication systems." Don Berwick, a former administrator for the Centers for Medicare and Medicaid Services, has popularized the slogan Delbanco and his group adopted, "Nothing about me without me," acknowledging that PCC is not always evidence based. In his 2009 *Health Affairs* article, he emphasized that PCC relates to one's set of decisions and choices of circumstances and relationships in health care.

This concept has received increased attention since the 2001 Institute of Medicine report, "Crossing the Quality Chasm" [9], where health care quality and system-of-care improvement efforts were linked to the 6 core values: safe, effective, efficient, patient centered, timely, and equitable. Since then, myriad clinical, policy, and research initiatives have been launched to promote the study, advancement, and implementation of PCC. Research later presented 8 primary dimensions of the PCC model (respect of values, physical comfort, coordination and integration of care, information and education, access to care, involvement of family and friends, and transition and continuity) [10]. In 2015, the World Health Organization released its framework on "people-centered health services" [11], emphasizing a focus on a system that adopts individuals', careers', families', and communities' perspectives into a trusted health care system.

PCC frameworks have proved to change the behavior of patients with cancer as they successfully engage the patient by incorporating his biopsychosocial support system into care delivery and ensuring sustainable development [12]. Involving patients with cancer meaningfully in the processes and responding to their emotions as part of PCC adoption have been linked to better health outcomes, more trust, and better engagement of the patient in their care [13]. Thus, to evaluate the effectiveness of PCC initiatives, the cognitive perception of patients with cancer needs to be studied in relation to their behavior within the care settings (eg, trust, satisfaction, anxiety, and engagement). On the other hand, achieving high-quality care is a complex pursuit in any setting, especially for cancer care. Improving the patient journey requires an integrated system of care and productive interactions among many system levels. By understanding the work system components, the design and integration of tasks, technology, and clinical processes can be reviewed to better support the needs of individuals while optimizing system performance. A supportive work environment and a highly engaged workforce correlate with improved quality of PCC and hospital performance [14]. Case managers, navigators, quality officers, and administrators may track patient outcomes at the population level. A study conducted in 2017 on postdiagnosis treatment communication with patients with cancer highlighted the importance of coordination among specialists, primary care, and other people involved in the care processes with patients to deliver necessary care as problems in coordination can lead to fragmentation in health outcomes and processes. However, existing initiatives and care-planning processes face barriers to adoption and implementation. To sum up, tools and initiatives designed to improve health care delivery through PCC need to be inspired by systems engineering principles as recommended by the Institute of Medicine and the National Academy of Engineering to identify, develop, and sustain best practices informed by the needs of survivors, caregivers, clinicians, organizations, and communities [13].

Due to the complex nature of the health care system, it remains hard to provide patients with care that meets their expectations without accounting for the work system in which they are receiving the care services [15]. However, to our knowledge, no framework focuses on PCC from a systems perspective. Human factors engineering interventions need to take into account issues across the whole system (system approach) with macro-ergonomic considerations, including organizational factors, to be more likely to significantly impact QOC. The Systems Engineering Initiative for Patient Safety (SEIPS) model of work system and patient safety, for example, emphasizes the principle of "balance" and focuses on system interactions that need to be considered to make significant progress in health care quality, linking the work system factors to health outcomes [16]. In addition, although many studies have focused on the workload of physicians and staff, no study has focused on the workload of patients with cancer. In this qualitative study, we explored the impact of work system factors on newly diagnosed patients with cancer's perceptions of PCC and QOC and the impact of PCC on the QOC outcomes among newly diagnosed patients with cancer following a suggested conceptual framework.

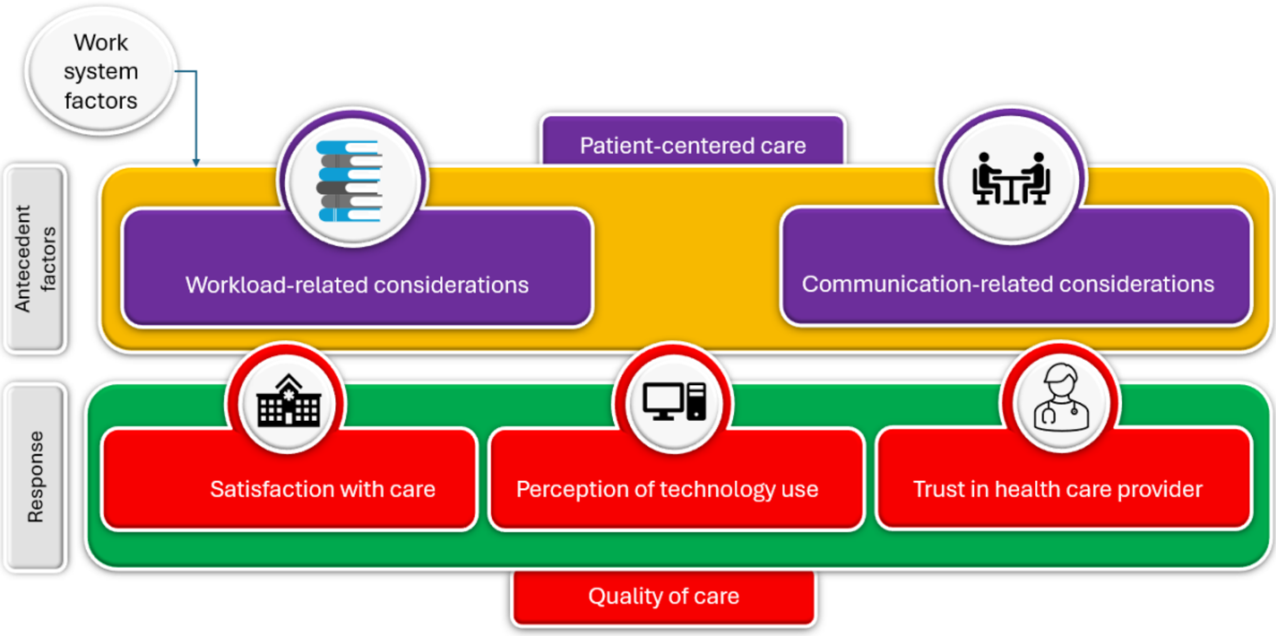
Theoretical Background

Overview

The framework built was inspired by different human factors models such as social cognitive theory [17], which conceptualizes the behavior of a person as a result of mental, personal, and social and environmental factors, therefore we considered behavior as a sum of a patient's perceptional cognitive input (patient-centeredness perception) and the response. Our patient-centered effectiveness components were inspired by the patient-centered communication in cancer care that defines communication through 6 functions [18] and the technology acceptance model. The technology acceptance model links technology perception to the attitude of the user toward the perceived usefulness and ease of use and the external variables [19,20]. The last model that inspired our framework is the SEIPS [21]. From a sociotechnical perspective, patients' experience, especially with chronic diseases, is a function of many coordination challenges [22]. Therefore, we need to go beyond the typical focus on a patient's single health care encounter and understand a patient's journey from a broader perspective through their interactions with other stakeholders in a system where not only patients and physicians are actors but the work system and the tools used are also important impactors of the perceptions and decision-making processes. Thus, we look at the systems' factors impacting patients' perception of patient-centeredness.

Our framework in Figure 1 emphasizes the relationship between patients' cognitive perceptions of patient-centeredness and QOC. We account for the impact of work system factors on these perceptions. We define patient-centeredness as a combination of workload support and communication and interrelationship support. Workload-related consideration characterizes the effective engagement of patients in their care experience. Communication and interrelationship improvement describes the communication effectiveness between patients and their providers. The dependent variables are related to the action tendency of patients: *satisfaction*, *perception of technology*, and *trust*. Exposure to the work system is considered a covariate in the model. To unpack this conceptual framework for evaluating patient-centeredness effectiveness, each independent variable has operational precedent in the human cognitive factors and behavioral economics literature.

Figure 1. Effectiveness of Patient-Centered Cancer Care framework.



Patient-Centeredness (Perceptional Cognitive Input)

Effective communication with patients with cancer can help meet information needs, improve physical and mental health, promote intimacy, and reduce burden [23]. In addition, patients diagnosed with cancer spend a lot of time and effort receiving treatment. Sometimes, patients have to deal with complex tasks related to medication taking and treatment in addition to rehabilitation activities that exceed their abilities, which engenders an overburden that has been proven to cause problems with adherence to treatment plans [24]. We define patient-centeredness in this conceptual framework as workload-related consideration and communication and interrelationship-related considerations.

Workload-Related Considerations (Ensuring Effective Engagement and Task Load Improvement)

Patient ergonomics is the application of human factors or related disciplines to study and improve patients' and nonprofessionals' performance of effortful work activities in pursuit of health goals [16,25]. A central emerging concept of societal views of health care considers that patients actively perform "work" to achieve health-related goals and objectives [26]. This way, human factors position patients at the center of the work system, aiming to improve their experience with the workload assigned [25,27]. In highly sensitive situations such as cancer care, this paradigm can help us better understand the dynamics among the 3 actors of the visits (physician, patient, and technology) and how their interaction can influence critical outcomes such as QOC, trust of physicians, and acceptability and perception of technology use. We define the role of patient-centeredness as a booster to the effective engagement and performance of patients in their care through *task load improvement perception*. Thus, the effectiveness is measured through task load improvement. Cognitive task load or workload is used in human factors or organizational psychology. It operationally refers to the levels of difficulty that an individual encounters during the performance of a task and is a measure of human performance [28]. Subjective methods commonly used in research include rating perceived task difficulty, engagement, or effort made by research participants [29]. There are 3 types of workload measurement: physiological, performance based, and subjective [30]. The physiological workload measures concern the continuous size of the body's physical responses [30].

Communication-Related Considerations (Communication and Interrelationship Improvement)

Compared to other health care settings, communicating information during oncology visits, especially initial ones, is critically important but can be particularly challenging due to the substantial amount of information provided, complex treatment decision options, involvement of multiple different providers (surgical, medical, and radiation oncology), and highly emotional situation with high patient workload [31]. Patients might not recall information accurately and might face difficulties understanding the information given. When information is particularly upsetting, many patients are too stunned to register further information [32]. Patients report leaving initial visits feeling that their informational needs (particularly about treatment, side effects, and prognosis) are not always met [32], which can lead to uncertainty, anxiety, and depression [31]. In one study with newly diagnosed patient with cancer–oncologist dyads, agreement on the content of the topics discussed ranged from only 37.5% for treatment side effects to 60% for prognosis [33]. Incomplete or inaccurate information about the disease process and treatment options increases the likelihood of patients receiving a suboptimal QOC [34]. Misunderstanding resulting from lack of communication has impacted health care outcomes such as decision-making, trust, and effective treatment [35]. Many countries have opened their accreditation, certification, and quality improvement programs for the past decade to examine physicians in training and communication skills [36]. Interpersonal and communication skills are 1 of the 6 general competencies for physicians identified by the Accreditation Council for Graduate Medical Education and the American Board of Medical Specialties in the United States [37,38]. "While communication skills are specific tasks and behaviors performed by individuals, interpersonal skills are relationship-oriented and process-driven, as noted by Duffy and colleagues" [39].

Response (QOC Perception)

Emotional distress is an average expected reaction to a cancer diagnosis. The diagnosis causes psychiatric complications (eg, anxiety, stress, and depression) induced by the patient's perceptions of the stigma commonly attached to cancer [40]. However, it is widely recognized that patient-centered interactions have the potential to influence patients' behavior and well-being [41-45]. Thus, we model patient-centeredness as an influencer of the behavior, which is patients' perception of QOC (*satisfaction* with the care offered, *perception of health ITuse*, and *trust* in health care providers).

QOC Perception: Perception of Health IT Use

It has been long promoted that health ITs (HITs) will improve efficiency and QOC, support health care delivery, and reduce costs for the health care industry [46]. Much of the work has assessed how health care providers and organizations can use HITs to deliver health care services [47,48]. However, a growing awareness exists that consumers also want to participate in their health care [49]. For chronic disease settings such as oncology, patients must participate in the monitoring and managing of chronic diseases [50]. Several factors contribute to the widespread use of eHealth in chronic care; acceptance and capability of using ITs are vital components of understanding the disease and treatment options [51]. Advancements in digital communication and medical technologies have led to digitalizing health care [52,53]. The increasing adoption of various HITs has created new channels for physician-patient communication beyond the walls of physicians' offices. With the increased adoption and use rate of electronic health records in cancer care, oncologists can use the provided data in the critical decision-making process and support their workload [54]. In a study by Mazur et al [55], the enhancement of electronic health record systems' usability was associated with better oncologist cognitive workload and performance. However, little attention has been paid to technology support for newly diagnosed patients with cancer. Therefore, extending the existing knowledge base is essential to better understand how technology impacts newly diagnosed patients with cancer. Research on the mechanism of patient-centeredness shows that it is necessary to ensure patients' engagement with their health and their providers over the treatment time [56] as it impacts patients' lifestyles, quality of life, and behavior in the context of cancer care.

QOC Perception: Trust in Health Care Providers

Extensive literature supports the importance of trust in physicians for patients with cancer as it has been linked to improving QOC and other treatment outcomes such as adherence to treatment [57]. On the basis of a review by Hillen et al [57], trust is needed to ensure a good interaction between physicians and patients. Trust has also been shown to be impacted by communication among newly diagnosed patients with cancer [58]. Thus, we consider trust as one of the QOC factors affected by the communication and workload of newly diagnosed patients with cancer.

QOC Perception: Satisfaction With Care

Patients demand excellent care services from their providers. It is becoming a competitive edge in health care to control the quality outcomes and patients' satisfaction with the services, the providers, and the organizations in which they receive care [59]. Satisfaction is an outcome of utmost importance in cancer care [60]. It was shown to be related to physicians' ability to elicit the concerns of patients' with cancer, consider their psychosocial needs, and involve them in treatment decision-making, which are the techniques of "patient-centered" care and communication [60,61]. We consider satisfaction with care to be the third main component of the framework.

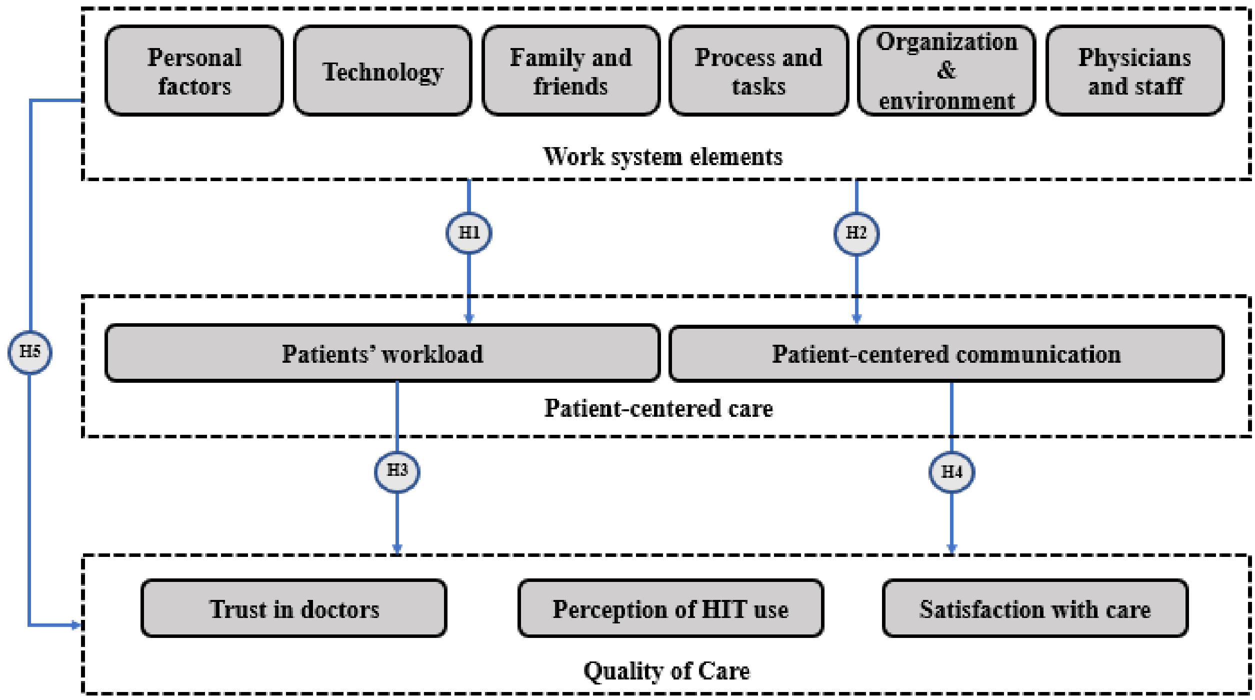
Hypotheses and Tests

Effective communication can prevent lapses in QOC and can mitigate harm when problems occur [62]. In cancer care, it is even more important to provide patients with the suitable communication needed [63]. Improving communication with patients with cancer in the first few visits requires a better understanding of patients' experiences of breakdowns in care and their needs in the early stage of their experiences [64]. In addition, patients frequently experience high load and feel overwhelmed due to their confusion about the treatment plans and their uncertainties about their options, which compromises their perception of QOC [34].

The complexity of cancer care, typified by the financial, emotional, and physical challenges, makes patient care challenging [65,66]. In addition, the complexity of the cancer care work system is reflected in the multiple clinicians that are involved in the processes, the long therapies, and the uncertainty of the outcomes [66]. Thus, we considered the following hypotheses: (1) work system elements—work system factors impact newly diagnosed patients with cancer's perception of their workload (hypothesis 1); (2) work system elements (communication)—work system factors impact newly diagnosed patients with cancer's perception of their communication with physicians (hypothesis 2); (3) workload (QOC)—workload impacts newly diagnosed patients with cancer's perception of QOC (trust in physicians, satisfaction with care, and perception of HIT use; hypothesis 3); (4) communication (QOC)—physician-patient communication impacts newly diagnosed patients with cancer's perception of QOC (trust in physicians, satisfaction with care, and perception of HIT use; hypothesis 4); and (5) work system elements (QOC)—work system factors impact newly diagnosed patients with cancer's perception of QOC (trust in physicians, satisfaction with care, and perception of HIT use; hypothesis 5);

We consider the following as work system factors: physicians and staff, organization and environment, family and friends, and processes and tasks. All hypotheses are summarized in Figure 2.

Figure 2. Hypothesis framework guiding the qualitative analysis. H: hypothesis; HIT: health IT.



Methods

Overview

This study used qualitative data from semistructured interviews to explore the impact of work system factors on newly diagnosed patients with cancer's perceptions of PCC and QOC, as well as the impact of PCC on QOC and technology preferences. We used semistructured interviews to facilitate candid disclosure of personal experiences. The interview questions used in this study were guided by the SEIPS 2.0 mode [26] and validated by existing literature. The full interview guide has been published elsewhere [67]. The SEIPS 2.0 model provides a framework that helps comprehend the work system (people, tools and technologies, tasks, working environment, and organization); process (clinical process, patient outcome, and organizational outcome) in the health care domain [26]. It also helps assess and understand the complex interaction among work system elements [26]. The interview guide was developed iteratively with the research team. Subsequent revisions of the interview guide were informed by emerging themes and sensitizing concepts generated through data collection and analysis. We also revised the interview guide questions based on expert feedback and the results of quantitative research conducted previously in the same center as part of the same project. Our sample included patients who (1) had been recently diagnosed with cancer, (2) were in their first few visits to the cancer center where the study took place, and (3) were adults aged ≥18 years.

We conducted a total of 20 in-depth semistructured interviews. Sampling continued until theoretical saturation was achieved, defined as the point at which further interviews did not advance the conceptual depth of the developed categories or reveal new dimensions of the relationships among categories [68,69]. The interviews lasted approximately between 30 and 60 minutes and were facilitated by a trained expert in clinical research management. The length of each interview was determined by the patient's level of comfort in disclosing their perceptions and sharing their experiences. We completed 20 interviews, resulting in 989 minutes of recording that were used for data analysis. All participants provided informed consent for the interviews to be audiotaped and professionally transcribed.

Analysis of interview transcripts was iterative and used a deductive and inductive approach. The deductive approach used focused coding, applying predetermined codes or themes resulting from the preset hypotheses made regarding the different interactions among the perceptions of work system factors, QOC, and patient-centeredness. Throughout the study, we incorporated memo writing to reflect on individual cases, interview settings, participants' responses, and emerging concepts and assess preconceived notions that were discussed weekly with the research team. The coding was done and visualized using a Microsoft Excel (Microsoft Corp) spreadsheet. We also prepared the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist and provided it in a supplementary file (Multimedia Appendix 1).

To mitigate the risk of bias caused by qualitative research, our study initially used a triangulation design with findings reported in different studies [3]. For more transparency and accuracy, different participants with different backgrounds reviewed and confirmed the transcripts and interpretations. In addition, a clear documentation of the analytical process was conducted.

Ethics Approval

This study received ethics approval (institutional review board ID 00011536) from the Stevens Institute of Technology and from Hackensack Meridian Health, John Theurer Cancer Center, New Jersey, where it took place.

Results

Overview

The distribution of demographics is shown in Table 1. Most participants were female (12/17, 71%), White (9/17, 53%), and aged >40 years (15/17, 88%).

Table 1. Sociodemographic characteristics of the study sample (N=17).

	Participants, n (%)
Gender	
Male	5 (29)
Female	12 (71)
Race	
Black	2 (12)
Hispanic	4 (24)
White	9 (53)
Other	2 (12)
Age (years)	
18-39	2 (12)
40-59	10 (59)
≥60	5 (29)
Education	
High school or lower	7 (41)
Bachelor's degree level	5 (29)
Graduate school	5 (29)

Impact of PCC on QOC

This section presents the findings related to the testing of hypotheses 3 and 4.

Workload of Newly Diagnosed Patients With Cancer

Newly diagnosed patients with cancer expressed their perception of the workload and reported experiencing a high mental workload due to frustration and emotions when they were first diagnosed:

> *At first you're all nervous and upset with your condition. So you're like, I don't know how to do this. You're freaking out.* [Patient 13]

The mental workload increases as patients are required to remember many details (eg, appointments, information, and medication) and are required to understand the results and options they are given with the little information shared with them:

> *There's a lot to remember. There's a lot you got to remember which medicine to take and when.* [Patient 12]

> *I think sometimes the interpretation of the test results is challenging and a bit anxiety producing.* [Patient 06]

In addition, patients experience high temporal and physical load as their tasks require a lot of effort and are time and energy consuming:

> *It's unfortunate that I have to get drawn so many time...You just get tired of sitting, you get tired of being in there.* [Patient 02]

> *Very demanding. I go get the blood work done first, and then when my blood work is done, I go up and wait to see the doctor. Then I go to the infusion room when I'm done with him.* [Patient 13]

Impact of Workload on Trust in Physicians

High workload negatively impacts newly diagnosed patients with cancer's trust in their physicians. For instance, patients who had less workload, felt cared about, were given enough time, or were less rushed were more likely to trust their physicians:

> *My doctor really care about me. I gage that because I'm not being rushed to leave right away. I see the attention that they pay attention to all my questions that I have, and also, I feel comfortable.* [Patient 03]

> *I think they do help with my emotions because they are trying to address every question that I have. I am not getting very emotional; I'm not getting extremely upset. So what I'm trying to say is that they might not even see any emotional reactions from me because I do trust that the doctor said he will fix me and I know that he will. So, I'm not emotional. I'm going very strong because of their support and my knowledge that they are there, and they will fix the situation.* [Patient 05]

Impact of Workload on Satisfaction With Care

The findings also showed that the workload perception among newly diagnosed patients with cancer impacts their satisfaction with the care services received:

> *As far as the demands, the time and effort it took to ensure the entirety of the visit was completed well in its entirety. Look, it was very effortless on my part.* [Patient 06]

> *They just gave you time to digest and let you just sit there and think about what the doctor was saying, and that really does make a difference. So basing out the information across the appointment instead of just kind of ramming it all down your throat at the beginning and asking any questions, it always moves quickly. But I do think that the taking of time in between providing information and providing time to digest was very helpful.* [Patient 12]

However, patients who felt that the care services were demanding of time or effort were less satisfied with their visits:

> *The treatment itself, when I'm there for chemo, does take a long time. That's like 4 hours at least. And then I'm on a pump for the next two days, so that does take a decent amount of my time. And then when I'm on the pump, I'm basically laying down because I really can't do much because I'm still getting treated. So, the treatments are demanding.* [Patient 13]

Impact of Workload on Perception of Technology Use

Finally, the high workload impacts patients' approval of physicians using technology during the visits. Patients who felt that using technology made their tasks less effort, memory, and time consuming were more likely to accept it. They felt that having all the information available in one place saved them from looking for information everywhere and trying to understand it and share it again with the physicians:

> *Technology can help keep me cooperating with the doctors and the treatment without much effort...Usually the nurse has the computer and she's checking in that all the details are correct and verifying information with me before the doctor even walks into the room. And she can answer questions about when I start the next cycle, et cetera. Because she has all the information with her. That saves me time. I don't need to remember everything.* [Patient 02]

> *It will impact only positively because I see an order, everything in one place in my app. I see my history, I see all my tests, I see all my appointments, all notes from the doctor, I see all the scans and everything is in one place. For me it's positive because again, I'm technology savvy and for me, that's great to be utilized, to have everything in one place.* [Patient 05]

To conclude, these qualitative findings showed that having less workload (eg, physical, temporal, and effort based) helped newly diagnosed patients with cancer trust their physicians more, be more satisfied with the care visits, and accept technology use during the visits (hypothesis 3 was supported).

Communication Needs and Impact of Communication on Trust in Physicians

The quality of the communication in cancer visits is critical. Physicians share a lot of information with patients in the first visit, and it is essential for them that physicians explain things clearly in an easy, comforting, and understandable way and that their questions are answered without being rushed. Good communication of the information needed made patients trust their physicians:

> *She made it very easy for me to understand very difficult, very complex procedures and explain to me clearly why, in my particular situation, she would recommend the clinical trial treatment that I'm on. And the appointment didn't last more than maybe 20 minutes, but in that time, I felt like I understood what I was getting myself into. She provided things for me to read when I got home and explained clearly that other options were not as good as this one...If I was concerned about something, she would try to reinforce that. Everything that she explained made sense because she was trying to keep me at ease and not worried about more than I should have.* [Patient 02]

> *Our communication me and the doctor is genuine, accessible, nurturing, informative.* [Patient 17]

Different patients also have different levels of understanding. Patients trust their physicians if they respect their pace and health literacy levels:

> *Yeah, basically pausing and making sure I understand things as she's saying them, making sure I'm caught up in the discussion, because you could go very fast over a lot of things and there's a lot of information to digest at an appointment like that. So she would take her time to pause and say, do you understand what I'm talking about? So that sort of thing helped.* [Patient 17]

Another important communication factor that impacted patients' trust in their physicians was feeling like they were treated equally without any bias independent of the type of insurance they had or their age or race. Patients liked to be treated as human beings, not as numbers:

> *It felt good. I didn't feel like I was going to be just thrown in there and just done whatever, and then no explanation of anything, so it felt really good. And the mere fact that I have Medicaid and I didn't feel like a patient, that felt phenomenal for me. It was very powerful. Well, I mean, it seems that we get lost in the shuffle. We're just like, nobody basically. We seem to be like, I don't know, nobody, we're not treated like we just don't matter. I didn't feel that way at all. I felt like I have a chance just like everybody else who has like [insurance name] or whatever. So I just felt like they cared about me and the process was they were going to do whatever they had to do, no matter what insurance I had. So that meant a lot to me.* [Patient 07]

Furthermore, newly diagnosed patients with cancer like to be given full attention by their physicians, communicating both verbally and nonverbally, exchanging eye contact, and paying enough attention to every question they ask. This impacts their trust in physicians:

> *Well, they always examine me, obviously. They talk to me. They come up to me, and they look at me not everywhere in the room, just looking at me eye to eye, and explain to me exactly what happened this week...They look at me straight in the eyes, and for the time that we're there, her attention is focused on me. And when I ask the question, they usually don't mind repeating themselves, because sometimes when you're in treatment, you don't hear well. And I'm taking notes when I'm there. And I sometimes repeat questions that she may have already answered. And she is very happy to follow up and expand a little more so that I can understand in more detail what she's trying to tell me about follow up questions, answering follow up questions.* [Patient 02]

Finally, patients want to be treated in a personalized way as a special person and to build a strong relationship based on empathy with their physicians by talking about their personal life and not only about treatment and visits:

> *My doctor is extremely approachable despite his busy schedule. He shared his cell phone number with me, but of course I'm not going to communicate to him. But those are things that you understand that you're not just a patient, you're a special individual for him and for his staff, and everything is personalized. I think that's my belief, honestly.* [Patient 05]

> *I like that the doctor talks to me in general not only about my treatment. Sometimes we talk about family and things like that, really getting to know the doctor. He was very communicative; he keeps being positive and that helps.* [Patient 01]

> *I think in particular, she has quite a good amount of empathy, which I think a lot of doctors don't. So she treats you like a human being and trying to think what else on the medical side. It's basically kind of explaining your condition well and giving you an idea of what's to come.* [Patient 10]

In fact, patients who do not have a strong bond with their physicians and only talk about treatments without discussing their options may have low trust in the physicians' opinion and feel that they are treated in an unfair way:

> *Well, yeah, I feel that, but I think this is more and more interested in selling products that are profitable for the hospital than necessarily what care I need. She just seems obsessed with selling an expensive procedure that I'm not ready for. I'd like to see more programs tailored to my situation and some options. There's a lot of options and treatments available*

today. And to say it's this or nothing, I don't think it's appropriate. [Patient 11]

Impact of Communication on Satisfaction With Care

The communication between newly diagnosed patients with cancer and their physicians also impacts their satisfaction with the care received during the visits. In fact, patients prefer to be told the truth about what is happening to them and for it to be stated that the physicians are doing their best:

> *Well, anything the doctor noticed, any concerns she has, she always meant that she just wants to be sure and just saying that she wanted to stay on top of things and that was pretty good to me. I would say I had really good care both in first visit and follow-up ones.* [Patient 01]

They also want physicians to explain the goal of the visit clearly and to be walked through every step of the procedures. On the basis of patients' needs, physicians need to ask them whether they are satisfied or not and allow families to be part of the visit for more support:

> *When I have to come visit, they know exactly what happened before, and they can specifically tell me where we are today and what we need to do today. They also keep some time for me to ask questions, and if I don't understand something that she explained, she'll explain it again. My daughter participates in those visits as well, and she always has questions too. As long as we have questions. If we are all clear and everything's just a routine visit, then we need less time, I guess. But if we need more time to answer questions, they're willing to answer until we're satisfied...I like it when they explain things to me, even if they're technical, because I can look it up and kind of make sure that I'm getting the right information from my doctors and it's consistent with what the best health organizations treating cancer are doing.* [Patient 02]

Although some prefer to be told everything, others do not understand much when it is too technical. Physicians need to explain things to them in an understandable way:

> *The communication was excellent. The answers that I was seeking were given and the questions that I had were answered. The care was given, felt comfortable. The doctor communicated with me in a way I could understand.* [Patient 06]

However, paying attention to special cases is important to gain patients' satisfaction. Some patients have comorbidities and need to be taken care of in a more careful way. For example, one of the patients whom we interviewed was blind. It is more challenging to communicate with such patients:

> *I am blind, so they always print documents out for me though but they tell me everything verbally. To make sure I understand and then they tell me it's on the printout.* [Patient 16]

Some other patients are skeptical about health care systems. It is important to know how to handle them and how to communicate information to them without losing their trust not only in physicians but also in the system itself:

> *Cancer care is a profit center for these medical centers. The doctor is trying to push a very expensive procedure that's very invasive that I don't have the support network to do. And it seems to be like she wants an all or nothing for that procedure. So, I think I really need a second opinion on this stuff. Well, like I said, I think that the cancer centers seem to be out to maximize profits because I see them advertised all over. I don't know, that's just what I seem to find out.* [Patient 11]

Finally, despite the focus on visits being very important, follow-up needs to cover home care for the first visits as patients need more support at the beginning of their experience and building a bridge of communication with them beyond the care visits would help them feel more cared about:

> *So, the communication during the treatment and while I am in the hospital is really good and I feel that I can ask any question and I always get the answers. As for communication when I am at home, I think I am still learning the system. After my first treatment, I had some adverse effects. I did document and I did write up my observations. Science in me did that. But I didn't know how to communicate that to the doctor until my next follow up with the doctor. And then we did discuss those adverse effects and they did adjust my dosing regimen.* [Patient 05]

Impact of Communication on Perception of Technology Use

Newly diagnosed patients with cancer need their physicians to communicate with them without any distractions:

> *No, no them using a computer is of no distraction at all. The doctors still attend to me.... They always check my blood work and put it on the computer and if there is anything they communicate after the visit. They always call.* [Patient 01]

If patients are made the center of the visits and the computer is used for documentation purposes, patients feel satisfied with its use by physicians:

> *I don't think technology is disturbing my communication because they are still there. It's not that we are communicating through computers only. They are still there in the room they are personally discussing with you. But then they document everything in the computer. And I think at this time and era, you do expect that everything will be documented on the computer? That's my expectation.* [Patient 05]

> *She occasionally doesn't always use the computer, but occasionally does to look at test results. But I never really found it distracting, and I don't feel like she was paying attention to the computer more than me. It was just there as a tool as part of the*

> *appointment. Never did I feel like it was computer first, patient second sort of thing.* [Patient 12]

However, if the use of technology made patients feel that they could not communicate well with their physician, they were not happy with it being used during the visits:

> *Not every time, but yeah it was distracting, sometimes. Actually, now that I think about it, I think that was where I could see a few instances where that was where their focus was. I think I would ask a question and there would be like a two-minute pause because he was in the middle of typing stuff on the computer and then he would answer after. So, the doctors was distracted with whatever he was doing on the computer.* [Patient 08]

> *I would prefer them not to be on their computer and rather making eye contact and communicating directly with the patient rather than typing. So I didn't feel as connected with the doctor.* [Patient 17]

To conclude, these findings showed that, if the communication between physicians and patients is built in such a way that patients are the center of the care and using technology does not distract physicians from building a bond with their patients, technology use during the visits is accepted and not judged as distracting. Thus, hypothesis 4 is supported.

Impact of Work System Factors on Communication

The work system factors in health care impact patients' communication with physicians. In fact, patients were more satisfied with the communication with health care staff when they felt that the organization was empowering nurses to intervene and raise issues related to their health. Thus, the *organization and environment* impact the communication perception among newly diagnosed patients with cancer:

> *There was a nurse, also a night nurse, who noticed that there was fluid in my lung and she put a note in for the doctor to see if they could remove it because she thought it was too big. The day after I got in from the emergency room and the nurse was the one who raised the flag to the doctors and the next day they removed the fluid. So I think they're empowered, but at the same time looking she didn't have to go back and look at the X rays in my lung because she was surprised that I was breathing so badly. So, she was just curious and checked in and brought it to everyone's attention. I also like that they called me directly to check on me. They make sure that we are cared about and that we know everything about our situation.* [Patient 02]

In addition, frequent follow-ups can help patients share their concerns and issues with their physicians and help them communicate well in a continued way, which shows the impact of the *processes and tasks* on communication for newly diagnosed patients with cancer:

> *So, I think follow-ups are very important, especially after the first treatment. When my first treatment was very miserable, I felt very miserable afterwards, I had very significant adverse effects to the drugs. And then in a week I had follow up with the doctor and we had really good communications for the second treatment.* [Patient 05]

The way in which physicians, nurses, and staff interact with patients impacts their perception of communication. For instance, patients were more satisfied with physicians and other staff if they felt cared about and if their questions were answered. In addition, allowing family members to attend visits may help patients feel more reassured:

> *The people in the lab are amazing. They understand that we get pinched a lot, and they try to work with you, and they help each other, too, because they have to get the results stat, how they like to say. And they look thoroughly at the request from the doctor. And if I have questions about what they're doing, they'll answer them intelligently...When I first saw my doctor, she knew the record just as well, but she asked me to tell her my experience so she would know firsthand from me how I was feeling now and what had happened in the past. So, I felt very well taken care of and the communication between my doctor and me was excellent really.* [Patient 02]

> *My daughter participates in those visits as well, and she always has questions too.* [Patient 02]

Thus, we conclude that the work system elements impact the perception of newly diagnosed patients with cancer regarding communication, which supports hypothesis 2.

Impact of Work System Factors on the Workload

The work system elements impact patients' workload. In fact, the process of detailed documentation in the records and providers accessing that information easily also reduces patients' mental workload and frustration. Patients also like being guided through every step at the clinic as it helps them feel better:

> *I do feel supported, even though we meet for short periods of time...I felt that in every visit, in the few minutes that she was maybe 15 minutes that she's in the room, she knows everything about what happened the previous weeks. I don't know how she does it, but if I forget to tell her about something that I was feeling the week before, she would ask me about it. So, I understand. However, they do it to go into the room and remember exactly this patient in particular, it makes me feel very reassured that they've done their homework when they walk into the room to talk to me.... I think they have a pretty good system. Once you register, they have someone already greeting you, walking you over to do any lab tests that you need. They kind of wait around and guide you to the elevator so that you can go up to the waiting room waiting area.* [Patient 02]

> *You come and you get greeted by a person that sits on the first floor. Then you go to lab. In the lab everyone is very attentive. Sometimes you have to wait a couple of minutes, but usually it's not very long wait time and they are very attentive to ensure that they are doing very good.* [Patient 05]

However, the long waiting time in the process makes patients anxious and nervous, which adds more workload to the physically demanding processes and procedures that they are experiencing:

> *I just wait in between seeing the nurse and the doctor a little bit too long, I thought. So, the wait was a bit lengthy, a bit long. That was nervous for me. And not only that, but we had to leave at a certain time because we had to go pick up my nephew after school. So that was my appointment was at I think it was at 110, and I didn't see the doctor to, like, almost 230 or something like that.* [Patient 07]

> *I sat there, and I waited, and I waited, and I waited for my first biopsy results and to get them at 02:00 p.m. I was calling and calling and calling, trying to see if anything came in...It's a lot and it takes a toll on you.* [Patient 13]

To reduce this load, physicians and staff members need to explain matters clearly to their patients to comfort them, reassure them, and make them feel cared about through personalized services:

> *I would say the doctor knew how much information I needed to avoid being overwhelmed. Just telling me the options of what we need to do and I think that pretty much helped. Not feeling overwhelmed, it's like, let's do this and get past it. So that was pretty much my feelings.* [Patient 06]

> *I was very nervous about what the nurse had told me that was going to happen once. I didn't want to need to have a tube in my lungs. But luckily, before we got to the procedure, they had already taken care of that and she put it in capital letters so that the radiologist didn't miss it, that they didn't need to put it too, because the treatment would resolve that over time. So, I think reassurance is what they tried to do and being attentive to the details, which in medicine, I think is very important because each case is a different case. And I felt very comforted that I'm not just a number, I'm a patient that they're trying to get out of the hospital.* [Patient 02]

> *What I do when I am overwhelmed is I call the nurses all the time, and they're so helpful. I was calling them multiple times a week, and whether it was a new side effect, or I just had follow-up questions. So, I definitely have been utilizing them, and they've been so incredibly helpful.* [Patient 08]

In addition to the health care actors, the organization impacts patients' workload. Patients need a relaxing, calm, clean, and organized environment. Using comforting colors and decorative signs that motivate patients can also give them more hope and reduce their load:

> *So, when we look at the physical layout of it and all those processes, it's very nice, very organized place, very relaxing when you have to wait, so it's no problem.... Everything was very comfortable. No noise at all. Very calm and especially very accommodating.* [Patient 01]

> *The environment is really good. Everything flows nicely. Everything is nice and clean. Everything the colors and the walls and everything is very calming. As far as decorations and stuff like that, there's like a passageway that sometimes I'll pass through that I see like puzzles of past people that I have done from cancer. They do like these jigsaw puzzles, and they've hung them all across this hallway that I pass, which I find very endearing when I pass, and I see that. So pretty much in the sitting areas and everything where everyone sits and waits nice, valid and everything like that.* [Patient 03]

Construction work, long walking distances between the rooms, parking, and many other issues may cause patients to experience more physical and temporal load:

> *Let's be honest here. Like, it's completely overbooked, and I know they're under construction or what have you, which is stressful to have that many people crowding in the hallway.... Like if I was upstairs and I couldn't make it down in time, they would call and say, hey, she's running late. And it was accommodated. But it's definitely overwhelming to navigate.* [Patient 17]

> *You can build a bigger parking lot if you have room. Yesterday, yeah, I was riding because the appointment was at 01:00 at that time was all packed. They had to drive around all the way up to the roof and start coming down...there are definitely not enough spots there.* [Patient 09]

Furthermore, family and friends can help support patients in their tasks, which reduces their workload by facilitating processes:

> *My daughter is there for me. I moved in with her so that she could drive me to my appointments, and I can have support when I'm not feeling so good.... It makes things way easier.* [Patient 02]

> *Wife, family, friends, taking me to the treatment, stopping by and visiting me, phone calls. Just a lot of support.* [Patient 08]

> *My husband is always next to me. I am 90% self-cared, but sometimes I need his help to move around. For example, during the chemo to go to the bathroom.* [Patient 05]

On the basis of these findings, we showed that work system elements impact patients' workload. Thus, hypothesis 1 is supported.

Impact of Work System Factors on Satisfaction With Care

The work system elements impact patients' perceptions of QOC. They impact patients' trust in their physicians, their perception of HIT use, and the satisfaction with the care received during the visits. In fact, patients appreciated nurses checking on them frequently and being nice to them, which made them feel more satisfied with the care received:

> *I think specifically the nurses in the infusion center, they were so kind and so nice, and they definitely were always asking how I was feeling during the infusions. They come check on me every ten minutes, pretty much. So, they were very accommodating and made me feel very comfortable.* [Patient 08]

> *Before seeing the doctor, you see the nurse nurses always welcoming even the staff that you go for copay or just approach to announce that you arrived. They are very attentive. You can see that they are feeling your pain. And that's very comforting, let's say...Usually my chemo is very long. So, the people who delivers lunch are so attentive to every single person. They are taking time for every single patient to repeat whole menu and convince you that this is very delicious to take. It's really warm and nice atmosphere. So I think they are taking extra steps to make you feel as comfortable as possible given the heaviness of the disease.* [Patient 05]

Receiving less attention from physicians would make patients unsatisfied with the visits:

> *My doctor is very busy. He's the head of the cancer center and he has tons of patients. But I would have liked to see him maybe sometimes not at the end of the treatment, but also in the middle of the treatment.* [Patient 05]

In addition, having a good organizational environment increases patients' satisfaction with care. Patients want to be in a well-designed environment where they can access better care services:

> *So, I think all is very accessible, very well designed, that you stop by first in the lab, then you can immediately pick up your pharmacy needs and then go to the second floor to visit the doctor and then move to the chemo center. So, I think everything was designed well. I love the sunny side and shady side of the infusion room. They have all these blankets, very nice people always asking what you want, more water, more anything to make you feel better. I think, as you say, from the organizational and structural perspective, is designed very well.* [Patient 05]

Impact of Work System Factors on Trust in Physicians and Staff

When physicians provide them with personalized services that are not based on generic information and that speak to their needs and situations, patients tend to trust them more:

> *It's just from reading the reports that the doctor gives me after the visit summary, I can tell that it's not generic. It's definitely speaking to my condition. I can see that what they're writing, and my evaluations are definitely about me, and I can see the reports, and it's definitely very much personalized.* [Patient 12]

The mix of personal and professional interaction makes them trust their physicians and the nurses delivering the services. Patients need people who listen to them, a friendly environment, and practices in which the main goal is to deliver the safest care to them:

> *What I did notice you have very good clinical practices that when the nurse has to introduce chemotherapy, they have second pair of eyes verification, which speaks of high quality and regulatory compliance of your organization, and that is incredible to see.* [Patient 05]

> *Every time I have a discrepancy, they always double-check. Either the nurse in the infusion room double-checks with the research nurse, the research nurse double-checks with the doctor, and everybody double-checks to make sure that we're doing it the right way, and that gives me comfort as well.* [Patient 19]

> *I think it's a good team. They listened to their own people, and they acted on it. That makes me trust the whole thing. So, I'm like I said, very professional and very personable. I really like that. It sums it up so beautifully. Like the two pieces in health care. Professional yet personable. I really do like that.... The hospital has great practice, and we have a number we can call. Twenty-four, seven. And they told us exactly how to behave and where to go when we got to the hospital. So that kept me more of these. But I was very scared. It was a Sunday night, so the doctor didn't physically come, but the emergency doctor had spoken to her. He knew my case.* [Patient 02]

> *They had a social worker contact me, which was nice to see if I had any opportunities or anything. I thought that was a good guess here. Any helping hand is an essential hand.* [Patient 11]

In addition, more trust is built among patients if the team's communication is healthy and professional:

> *The communication between the staff, I see that it is good and very professional. And the place is amazing. They walk with me and make sure I have all what I need to start the next step. The infusion, I prefer the one where I get sun and they already know that.* [Patient 02]

Impact of Work System Factors on Technology Perception

If physicians let themselves be distracted during the visits or do not pay enough attention to patients, patients may consider technology as a source of distraction and disruption to the visit:

> *The doctors was distracted with whatever he was doing on the computer.* [Patient 08]

Finally, the organized process of the visit, the good collaboration between nurses and physicians, and note-taking to make patients the center of the visits made patients trust their physicians more, be more satisfied with the visits, and accept the potential that technology may have in the success of the care processes:

> *Doctor showed up even with nurse practitioner. And normally the doctor talks to you, she examines you, explains things. And then normally the nurse practitioner fills up whatever they need to do in a database. And normally it's not distracting at all. It's all adequately it happens in the background, and you concentrate on the conversation with a doctor, and someone else is filling out all the paperwork. Something needs to be like sending the prescriptions to my pharmacy and setting up another test. Everything was done at the same time, but I don't feel it was destructive at all. It was good.* [Patient 09]

Thus, to sum up, the health system elements impact patients' perception of technology use, trust in physicians, and satisfaction with care, which supports hypothesis 5.

Discussion

Principal Findings

In this study, we qualitatively explored the impact of work system elements on QOC and PCC and how PCC also impacts QOC among newly diagnosed patients with cancer in the first follow-up visits after the diagnosis. We found that newly diagnosed patients with cancer experience a high workload (mental, physical, temporal, effort based, performance based, and emotional) resulting from the frustrating diagnosis and the load of information that they receive in the first visits. This load impacted patients' trust in their physicians, satisfaction with care, and perception of technology use during the visits.

A diagnosis of cancer is a threat to one's sense of security, whereas feelings and emotions accompanying the disease uproot everyday existence [70]. Patients find themselves unpredictably facing a high emotional load and under the obligation to cope with the stress and anxiety caused by their diagnosis [70,71], which explains the high emotional and mental workload faced by our participants.

In addition to that, patients with cancer have to deal not only with the physical ailments resulting from the illness and its treatment but also with the thoughts of permanent health impairment, disability, fatigue, and pain that may result from their diagnosis [72], which correlates with our finding of high physical and effort-based workload perception among the participants. This may explain the dissatisfaction of patients with the quality of the care received. Emotional stress and mental problems can cause difficulties in everyday life, such as not being able to work, financial problems, and a lack of social support. This has been shown to impact quality of life perception among patients with cancer in other studies [73]. The literature also shows that patients with cancer can experience a variety of needs as each person reacts individually to the hardships of illness depending on their personality traits and understanding of their new situation [70]. With the substantial incoming flow of information, patients may find themselves unable to trust physicians and may consider technology as a distraction to their visits at that stage.

We also found that newly diagnosed patients with cancer can be very needy when it comes to communication with their physicians and that their communication with physicians impacts their perception of QOC. Communication is the cornerstone of the relationship with the patient in all medical settings, specifically chronic care, with the main aims of creating a good interpersonal relationship, exchanging information, and making treatment-related decisions [74]. Certain attitudes, behaviors, and skills (eg, ability to impart confidence, empathy, "human touch," relating on a personal level, being forthright, being respectful, and being thorough) are part of effective communication, which was validated by our findings in this study [74]. A poor physician-patient communication in cancer care negatively affects psychological well-being and patients' decisions and perceptions regarding treatments. This validates our findings of the impact of communication with physicians among newly diagnosed patients with cancer on their perception of technology use during visits and trust in physicians [75].

In addition, we found that the work system elements impact patients' workload, communication, and QOC perceptions. This correlates with the findings of other studies in which the environment design was shown to impact patients with cancer's perception of QOC. A recent review of evidence-based design also found that a conscious design adapted to patient needs had an impact on a decrease in infection spreading, length of stay, pharmacological needs, and perceived stress among patients [76]. Furthermore, symbolic objects found in the environment have been shown to impact patients' sense of self and well-being [76].

In addition, a recently published Cochrane review on environmental impact on health stressed the profound need for well-designed studies following intervention in health care environments [77]. This correlates with our finding that patients who liked the decoration of the hospital, the motivational signs, the colors, the cleanness, the organized processes, the lighting, and the care of the nurses were more satisfied with the QOC and felt less overwhelmed. These findings lead to the expectation that major considerations ought to be taken when designing health care environments to meet quality requirements while considering patients' needs and supporting patients' sense of control, autonomy, and independence.

Theoretical and Practical Implications of the Findings

In the previous section, we validated the preset hypotheses that correlate with the findings of the quantitative studies from the greater project. This framework can help inform patient-centered interventions that aim to provide newly diagnosed patients with cancer with the support needed and ensure their satisfaction with the QOC offered. More empathy and human bond links between physicians and patients should be considered as patients want to be treated in a more patient-centered way and to feel that they are not receiving the same care as everyone else in the same way.

Patients also want to have the chance to ask as many questions as possible and be given as many follow-up visits in the beginning as possible to receive comfort and reassurance that everything will be fine. Empowering workers (nurses and staff) to intervene in case of emergency would help patients trust the health care organization. In addition, allowing patients to be accompanied by their family members would help them be emotionally comfortable. Another point to consider is to share a second screen with them in case a computer is used when the physician is communicating with them to comfort them regarding what the physicians are doing when they are not talking to them.

Pausing in the middle of the discussion to do other tasks would result in losing the patients' attention. Physicians should consider continuous communication where they pay as much attention as possible to the patients in a friendly way and where they listen to their concerns without rushing them even if there is a time limit as the time given can influence their decision-making process importantly.

This study's findings can also inform the organization's design. It should be considered that patients cannot move a lot between the laboratories and the visit rooms and it would be easier to assign them to rooms that are close to each other to minimize their physical effort during the visits. Better scheduling and allocation strategies should be considered to minimize the waiting time inside the hospital for each patient. Comforting colors, relaxing decoration, and motivational signs would help reassure patients while in the hospital. In addition, having any construction work when patients are coming in and out should be avoided as that can add more load to what they are already experiencing.

Limitations

Despite the useful insights garnered from this research, certain limitations must be addressed. First, the study's narrow geographic reach, which included only 1 cancer center, may limit the findings' generalizability to other cancer populations or health care settings. Patients' experiences and technology preferences in this facility may not represent those in other cancer centers or varied communities with different demographics or cultural backgrounds. Second, selection bias is possible as patients who chose to participate in the interviews may have different characteristics or opinions from those who declined or were unavailable. This may introduce bias in the findings and reduce the study's external validity. Furthermore, interviewing patients within a few visits following their initial diagnosis may not completely capture the dynamic character of their technological choices, which may change over time as patients adjust to their diagnosis and treatment. The reliance on patients' recollection of their technology preferences at this early time point may also be subject to recall bias. Furthermore, contextual factors particular to the cancer center where the research was conducted, such as local health care policies and the availability and accessibility of technology, may not be applicable or may vary in different contexts. Finally, social desirability bias and interviewer prejudice throughout the data collection process may have an impact on the data's authenticity and veracity. Despite these limitations, the findings of this study provide valuable insights into the technology preferences of newly diagnosed patients with cancer, and additional research with larger and more diverse samples, longer follow-up periods, and considerations of contextual factors is required to strengthen the findings' generalizability and validity.

Conclusions

In this study, we suggested a framework called Effectiveness of Patient-Centered Cancer Care and tested its validity in cancer visits to support PCC among newly diagnosed patients with cancer using qualitative data. We found that workload and patient-centered communication impact QOC and that the work system elements impact the patient-centeredness (workload and communication) and QOC (trust in physicians, satisfaction with care, and perception of technology use). To improve patients' experiences in the first visits after diagnosis, more interest needs to be given to the design of the organization, the processes that the patients have to go through, and the collaboration among the different actors and providers. This study's findings can also inform the organization's design. It should be considered that patients cannot move a lot between the laboratories and the visit rooms and it would be easier to assign them to rooms that are close to each other to minimize their physical effort during the visits. Better scheduling and allocation strategies should be considered to minimize the waiting time inside the hospital for each patient. Comforting colors, relaxing decoration, and motivational signs would help reassure patients while in the hospital. In addition, having any construction work when patients are coming in and out should be avoided as that can add more load to what they are already experiencing.

Acknowledgments

The research reported in this publication was supported by the National Institute of Nursing Research of the National Institutes of Health under award 1R15NR018965-01. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors also thank Dr Scott Rowley, Mariska Kop, Alene Mathurin, Farah Elkourdi, and Trina Mcclendon for their support in the data collection process.

Data Availability

The data sets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.
[DOCX File , 19 KB - humanfactors_v11i1e53053_app1.docx]

References

1. American Geriatrics Society Expert Panel on Person-Centered Care. Person-centered care: a definition and essential elements. J Am Geriatr Soc 2016 Jan;64(1):15-18. [doi: 10.1111/jgs.13866] [Medline: 26626262]
2. Li J, Porock D. Resident outcomes of person-centered care in long-term care: a narrative review of interventional research. Int J Nurs Stud 2014 Oct;51(10):1395-1415. [doi: 10.1016/j.ijnurstu.2014.04.003] [Medline: 24815772]
3. Elkefi S, Asan O. Perceived patient workload and its impact on outcomes during new cancer patient visits: analysis of a convenience sample. JMIR Hum Factors 2023 Aug 18;10:e49490 [FREE Full text] [doi: 10.2196/49490] [Medline: 37594798]
4. Elkefi S, Asan O. The impact of patient-centered care on cancer patients' QOC, self-efficacy, and trust towards doctors: analysis of a national survey. J Patient Exp 2023 Jan 17;10:23743735231151533 [FREE Full text] [doi: 10.1177/23743735231151533] [Medline: 36698621]
5. Mead N, Bower P. Measuring patient-centredness: a comparison of three observation-based instruments. Patient Educ Couns 2000 Jan;39(1):71-80. [doi: 10.1016/s0738-3991(99)00092-0] [Medline: 11013549]
6. Balint E. The possibilities of patient-centered medicine. J R Coll Gen Pract 1969 May;17(82):269-276 [FREE Full text] [Medline: 5770926]
7. Levenstein JH, McCracken EC, McWhinney IR, Stewart MA, Brown JB. The patient-centred clinical method. 1. A model for the doctor-patient interaction in family medicine. Fam Pract 1986 Mar;3(1):24-30. [doi: 10.1093/fampra/3.1.24] [Medline: 3956899]
8. Delbanco T, Berwick DM, Boufford JI, Edgman-Levitan S, Ollenschläger G, Plamping D, et al. Healthcare in a land called PeoplePower: nothing about me without me. Health Expect 2001 Sep 20;4(3):144-150 [FREE Full text] [doi: 10.1046/j.1369-6513.2001.00145.x] [Medline: 11493320]
9. National Academy of Sciences. The National Academies Collection: Reports Funded by National Institutes of Health. Washington, DC: National Academies Press; 2008.
10. Constand MK, MacDermid JC, Dal Bello-Haas D, Law M. Scoping review of patient-centered care approaches in healthcare. BMC Health Serv Res 2014 Jun 19;14(1):271 [FREE Full text] [doi: 10.1186/1472-6963-14-271] [Medline: 24947822]
11. Framework on integrated, people-centred health services. World Health Organization. 2016. URL: https://apps.who.int/gb/ebwha/pdf_files/WHA69/A69_39-en.pdf [accessed 2024-04-05]
12. Miller KL. Patient centered care: a path to better health outcomes through engagement and activation. NeuroRehabilitation 2016 Oct 14;39(4):465-470. [doi: 10.3233/NRE-161378] [Medline: 27689606]
13. Tevaarwerk AJ, Klemp JR, van Londen GJ, Hesse BW, Sesto ME. Moving beyond static survivorship care plans: a systems engineering approach to population health management for cancer survivors. Cancer 2018 Nov 15;124(22):4292-4300 [FREE Full text] [doi: 10.1002/cncr.31546] [Medline: 30277575]
14. Balbale SN, Turcios S, LaVela SL. Health care employee perceptions of patient-centered care. Qual Health Res 2015 Mar 01;25(3):417-425 [FREE Full text] [doi: 10.1177/1049732314553011] [Medline: 25274626]
15. Bergeson SC, Dean JD. A systems approach to patient-centered care. JAMA 2006 Dec 20;296(23):2848-2851. [doi: 10.1001/jama.296.23.2848] [Medline: 17179462]
16. Carayon P, Wetterneck TB, Rivera-Rodriguez AJ, Hundt AS, Hoonakker P, Holden R, et al. Human factors systems approach to healthcare quality and patient safety. Appl Ergon 2014 Jan;45(1):14-25 [FREE Full text] [doi: 10.1016/j.apergo.2013.04.023] [Medline: 23845724]
17. Schunk DH. Social cognitive theory. In: Harris KR, Graham S, Urdan T, McCormick CB, Sinatra GM, Sweller J, editors. APA Educational Psychology Handbook, Vol. 1. Theories, Constructs, and Critical Issues. New York, NY: American Psychological Association; 2012.
18. Epstein R, Street R. Patient-centered communication. In: Wittenberg E, Ferrell BR, Goldsmith J, Smith T, Glajchen M, Handzo RG, editors. Textbook of Palliative Care Communication. Oxford, UK: Oxford University Press; 2007:17-38.
19. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manag Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
20. Marangunić N, Granić A. Technology acceptance model: a literature review from 1986 to 2013. Univ Access Inf Soc 2014 Feb 16;14(1):81-95. [doi: 10.1007/s10209-014-0348-1]
21. Carayon P, Schoofs Hundt A, Karsh BT, Gurses AP, Alvarado CJ, Smith M, et al. Work system design for patient safety: the SEIPS model. Qual Saf Health Care 2006 Dec;15 Suppl 1(Suppl 1):i50-i58 [FREE Full text] [doi: 10.1136/qshc.2005.015842] [Medline: 17142610]
22. Carayon P, Wooldridge A, Hoonakker P, Hundt AS, Kelly MM. SEIPS 3.0: Human-centered design of the patient journey for patient safety. Appl Ergon 2020 Apr;84:103033 [FREE Full text] [doi: 10.1016/j.apergo.2019.103033] [Medline: 31987516]
23. Li J, Luo X, Cao Q, Lin Y, Xu Y, Li Q. Communication needs of cancer patients and/or caregivers: a critical literature review. J Oncol 2020;2020:7432849 [FREE Full text] [doi: 10.1155/2020/7432849] [Medline: 32454826]

24. Cheng AC, Levy MA. Measures of treatment workload for patients with breast cancer. JCO Clin Cancer Inform 2019 Feb;3:1-10 [FREE Full text] [doi: 10.1200/CCI.18.00122] [Medline: 30715929]
25. Holden RJ, Cornet VP, Valdez RS. Patient ergonomics: 10-year mapping review of patient-centered human factors. Appl Ergon 2020 Jan;82:102972 [FREE Full text] [doi: 10.1016/j.apergo.2019.102972] [Medline: 31654954]
26. Holden RJ, Carayon P, Gurses AP, Hoonakker P, Hundt AS, Ozok AA, et al. SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. Ergonomics 2013 Nov;56(11):1669-1686 [FREE Full text] [doi: 10.1080/00140139.2013.838643] [Medline: 24088063]
27. Norris B, West J, Anderson O, Davey G, Brodie A. Taking ergonomics to the bedside--a multi-disciplinary approach to designing safer healthcare. Appl Ergon 2014 May;45(3):629-638. [doi: 10.1016/j.apergo.2013.09.004] [Medline: 24135560]
28. Zimmerman ME. Task load. In: Kreutzer JS, DeLuca J, Caplan B, editors. Encyclopedia of Clinical Neuropsychology. New York, NY: Springer; 2011:2469-2470.
29. Korbach A, Brünken R, Park B. Differentiating different types of cognitive load: a comparison of different measures. Educ Psychol Rev 2017 Mar 31;30(2):503-529. [doi: 10.1007/s10648-017-9404-8]
30. Miller S. Workload measures: national advanced driving simulator. The University of Iowa. 2001. URL: https://www.nads-sc.uiowa.edu/publicationStorage/200501251347060.N01-006.pdf [accessed 2024-04-05]
31. LeBlanc TW, Back AL, Danis M, Abernethy AP. Electronic Health Records (EHRs) in the oncology clinic: how clinician interaction with EHRs can improve communication with the patient. J Oncol Pract 2014 Sep;10(5):317-321 [FREE Full text] [doi: 10.1200/JOP.2014.001385] [Medline: 25027025]
32. Hack TF, Degner LF, Parker PA, SCRN Communication Team. The communication goals and needs of cancer patients: a review. Psychooncology 2005 Oct;14(10):831-846. [doi: 10.1002/pon.949] [Medline: 16200519]
33. Eggly S, Penner LA, Hagiwara N, Gonzalez R, Harper FW, Heath EI, et al. Patient, companion, and oncologist agreement regarding information discussed during triadic oncology clinical interactions. Psychooncology 2013 Mar 15;22(3):637-645 [FREE Full text] [doi: 10.1002/pon.3045] [Medline: 22337320]
34. Wagner EH, Aiello Bowles EJ, Greene SM, Tuzzio L, Wiese CJ, Kirlin B, et al. The quality of cancer patient experience: perspectives of patients, family members, providers and experts. Qual Saf Health Care 2010 Dec 02;19(6):484-489. [doi: 10.1136/qshc.2010.042374] [Medline: 21127109]
35. Klasnja P, Hartzler AC, Unruh KT, Pratt W. Blowing in the wind: unanchored patient information work during cancer care. Proc SIGCHI Conf Hum Factor Comput Syst 2010 Apr;2010:193-202 [FREE Full text] [doi: 10.1145/1753326.1753355] [Medline: 20697453]
36. Makoul G, Krupat E, Chang CH. Measuring patient views of physician communication skills: development and testing of the Communication Assessment Tool. Patient Educ Couns 2007 Aug;67(3):333-342 [FREE Full text] [doi: 10.1016/j.pec.2007.05.005] [Medline: 17574367]
37. Batalden P, Leach D, Swing S, Dreyfus H, Dreyfus S. General competencies and accreditation in graduate medical education. Health Aff (Millwood) 2002 Sep;21(5):103-111. [doi: 10.1377/hlthaff.21.5.103] [Medline: 12224871]
38. Horowitz SD. Evaluation of clinical competencies: basic certification, subspecialty certification, and recertification. Am J Phys Med Rehabil 2000;79(5):478-480. [doi: 10.1097/00002060-200009000-00016] [Medline: 10994892]
39. Duffy FD, Gordon GH, Whelan G, Cole-Kelly K, Frankel R, Buffone N, Participants in the American Academy on Physician and Patient's Conference on Education and Evaluation of Competence in Communication and Interpersonal Skills. Assessing competence in communication and interpersonal skills: the Kalamazoo II report. Acad Med 2004 Jun;79(6):495-507. [doi: 10.1097/00001888-200406000-00002] [Medline: 15165967]
40. Massie MJ, Holland JC. The cancer patient with pain: psychiatric complications and their management. J Pain Symptom Manage 1992 Feb;7(2):99-109 [FREE Full text] [doi: 10.1016/0885-3924(92)90121-w] [Medline: 1573292]
41. Ong LM, Visser MR, Lammes FB, de Haes JC. Doctor-patient communication and cancer patients' quality of life and satisfaction. Patient Educ Couns 2000 Sep;41(2):145-156. [doi: 10.1016/s0738-3991(99)00108-1] [Medline: 12024540]
42. Stewart MA. What is a successful doctor-patient interview? A study of interactions and outcomes. Soc Sci Med 1984;19(2):167-175 [FREE Full text] [doi: 10.1016/0277-9536(84)90284-3] [Medline: 6474233]
43. Roter DL, Hall JA, Katz NR. Relations between physicians' behaviors and analogue patients' satisfaction, recall, and impressions. Med Care 1987 May;25(5):437-451. [doi: 10.1097/00005650-198705000-00007] [Medline: 3695654]
44. Kaplan SH, Greenfield S, Ware Jr JE. Assessing the effects of physician-patient interactions on the outcomes of chronic disease. Med Care 1989 Mar;27(3 Suppl):S110-S127. [doi: 10.1097/00005650-198903001-00010] [Medline: 2646486]
45. Henbest RJ, Stewart M. Patient-centredness in the consultation. 2: does it really make a difference? Fam Pract 1990 Mar;7(1):28-33. [doi: 10.1093/fampra/7.1.28] [Medline: 2318368]
46. Buntin MB, Burke MF, Hoaglin MC, Blumenthal D. The benefits of health information technology: a review of the recent literature shows predominantly positive results. Health Aff (Millwood) 2011 Mar;30(3):464-471. [doi: 10.1377/hlthaff.2011.0178] [Medline: 21383365]
47. Jones SS, Rudin RS, Perry T, Shekelle PG. Health information technology: an updated systematic review with a focus on meaningful use. Ann Intern Med 2014 Jan 07;160(1):48-54 [FREE Full text] [doi: 10.7326/M13-1531] [Medline: 24573664]

48. Sligo J, Gauld R, Roberts V, Villa L. A literature review for large-scale health information system project planning, implementation and evaluation. Int J Med Inform 2017 Jan;97:86-97. [doi: 10.1016/j.ijmedinf.2016.09.007] [Medline: 27919399]
49. Epstein RM, Street RL. The values and value of patient-centered care. Ann Fam Med 2011 Mar 14;9(2):100-103 [FREE Full text] [doi: 10.1370/afm.1239] [Medline: 21403134]
50. ElKefi S, Asan O. How technology impacts communication between cancer patients and their health care providers: a systematic literature review. Int J Med Inform 2021 May;149:104430 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104430] [Medline: 33684711]
51. Hardiker NR, Grant MJ. Factors that influence public engagement with eHealth: a literature review. Int J Med Inform 2011 Jan;80(1):1-12 [FREE Full text] [doi: 10.1016/j.ijmedinf.2010.10.017] [Medline: 21112244]
52. Meskó B, Drobni Z, Bényei É, Gergely B, Győrffy Z. Digital health is a cultural transformation of traditional healthcare. Mhealth 2017 Sep;3:38 [FREE Full text] [doi: 10.21037/mhealth.2017.08.07] [Medline: 29184890]
53. Asan O. Providers' perceived facilitators and barriers to EHR screen sharing in outpatient settings. Appl Ergon 2017 Jan;58:301-307. [doi: 10.1016/j.apergo.2016.07.005] [Medline: 27633226]
54. Aapro M, Bossi P, Dasari A, Fallowfield L, Gascón P, Geller M, et al. Digital health for optimal supportive care in oncology: benefits, limits, and future perspectives. Support Care Cancer 2020 Oct 12;28(10):4589-4612 [FREE Full text] [doi: 10.1007/s00520-020-05539-1] [Medline: 32533435]
55. Mazur LM, Mosaly PR, Moore C, Marks L. Association of the usability of electronic health records with cognitive workload and performance levels among physicians. JAMA Netw Open 2019 Apr 05;2(4):e191709 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.1709] [Medline: 30951160]
56. Mazor KM, Beard RL, Alexander GL, Arora NK, Firneno C, Gaglio B, et al. Patients' and family members' views on patient-centered communication during cancer care. Psychooncology 2013 Nov;22(11):2487-2495 [FREE Full text] [doi: 10.1002/pon.3317] [Medline: 23780672]
57. Hillen MA, de Haes HC, Smets EM. Cancer patients' trust in their physician-a review. Psychooncology 2011 Mar 23;20(3):227-241. [doi: 10.1002/pon.1745] [Medline: 20878840]
58. Cao W, Qi X, Yao T, Han X, Feng X. How doctors communicate the initial diagnosis of cancer matters: cancer disclosure and its relationship with Patients' hope and trust. Psychooncology 2017 May 17;26(5):640-648. [doi: 10.1002/pon.4063] [Medline: 26776310]
59. Urden LD. Patient satisfaction measurement: current issues and implications. Lippincotts Case Manag 2002;7(5):194-200. [doi: 10.1097/00129234-200209000-00006] [Medline: 12394558]
60. Mallinger JB, Griggs JJ, Shields CG. Patient-centered care and breast cancer survivors' satisfaction with information. Patient Educ Couns 2005 Jun;57(3):342-349. [doi: 10.1016/j.pec.2004.09.009] [Medline: 15893218]
61. Venetis MK, Robinson JD, Turkiewicz KL, Allen M. An evidence base for patient-centered cancer care: a meta-analysis of studies of observed communication between cancer specialists and their patients. Patient Educ Couns 2009 Dec;77(3):379-383. [doi: 10.1016/j.pec.2009.09.015] [Medline: 19836920]
62. Mazor KM, Roblin DW, Greene SM, Lemay CA, Firneno CL, Calvi J, et al. Toward patient-centered cancer care: patient perceptions of problematic events, impact, and response. J Clin Oncol 2012 May 20;30(15):1784-1790 [FREE Full text] [doi: 10.1200/JCO.2011.38.1384] [Medline: 22508828]
63. Peppercorn JM, Smith TJ, Helft PR, Debono DJ, Berry SR, Wollins DS, American Society of Clinical Oncology. American society of clinical oncology statement: toward individualized care for patients with advanced cancer. J Clin Oncol 2011 Feb 20;29(6):755-760. [doi: 10.1200/JCO.2010.33.1744] [Medline: 21263086]
64. Gallagher TH. A 62-year-old woman with skin cancer who experienced wrong-site surgery: review of medical error. JAMA 2009 Aug 12;302(6):669-677. [doi: 10.1001/jama.2009.1011] [Medline: 19584321]
65. Surbone A. Telling the truth to patients with cancer: what is the truth? Lancet Oncol 2006 Nov;7(11):944-950. [doi: 10.1016/s1470-2045(06)70941-x]
66. Ayanian JZ, Zaslavsky AM, Guadagnoli E, Fuchs CS, Yost KJ, Creech CM, et al. Patients' perceptions of quality of care for colorectal cancer by race, ethnicity, and language. J Clin Oncol 2005 Sep 20;23(27):6576-6586. [doi: 10.1200/JCO.2005.06.102] [Medline: 16116149]
67. Elkefi S. Evaluation of patient-centeredness in cancer care: a mixed methods study with a human factors approach. Stevens Institute of Technology. URL: https://www.proquest.com/openview/620d49593ddb80f77648a87fc6c0e905/1?pq-origsite=gscholar&cbl=18750&diss=y [accessed 2024-04-05]
68. Charmaz K. Constructing Grounded Theory: A Practical Guide through Qualitative Analysis. New York, NY: Sage Publications; 2006.
69. Nelson J. Using conceptual depth criteria: addressing the challenge of reaching saturation in qualitative research. Qual Res 2016 Dec 14;17(5):554-570. [doi: 10.1177/1468794116679873]
70. Lewandowska A, Rudzki G, Lewandowski T, Rudzki S. The problems and needs of patients diagnosed with cancer and their caregivers. Int J Environ Res Public Health 2020 Dec 24;18(1):87 [FREE Full text] [doi: 10.3390/ijerph18010087] [Medline: 33374440]

71. Page AE, Adler NE. Cancer Care for the Whole Patient Meeting: Psychosocial Health Needs. Washington, DC: National Academies Press; 2021.
72. Avancini A, Pala V, Trestini I, Tregnago D, Mariani L, Sieri S, et al. Exercise levels and preferences in cancer patients: a cross-sectional study. Int J Environ Res Public Health 2020 Jul 24;17(15):5351 [FREE Full text] [doi: 10.3390/ijerph17155351] [Medline: 32722265]
73. Bager L, Elsbernd A, Nissen A, Daugaard G, Pappot H. Danish translation and pilot testing of the European Organization for Research and Treatment of Cancer QLQ-TC 26 (EORTC QLQ-TC26) questionnaire to assess health-related quality of life in patients with testicular cancer. Health Qual Life Outcomes 2018 Jun 18;16(1):128 [FREE Full text] [doi: 10.1186/s12955-018-0954-3] [Medline: 29914500]
74. Grassi L, Caruso R, Costantini A. Communication with patients suffering from serious physical illness. Adv Psychosom Med 2015;34:10-23. [doi: 10.1159/000369050] [Medline: 25832510]
75. Lehmann C, Koch U, Mehnert A. [Impact of the doctor-patient-communication on distress and utilization of psychosocial services among cancer patients. A review of the current literature]. Psychother Psychosom Med Psychol 2009 Jul;59(7):e3-27. [doi: 10.1055/s-2008-1067443] [Medline: 18666041]
76. Browall M, Koinberg I, Falk H, Wijk H. Patients' experience of important factors in the healthcare environment in oncology care. Int J Qual Stud Health Well-being 2013 Aug 06;8(1):20870 [FREE Full text] [doi: 10.3402/qhw.v8i0.20870] [Medline: 23924604]
77. Drahota A, Ward D, Mackenzie H, Stores R, Higgins B, Gal D, et al. Sensory environment on health-related outcomes of hospital patients. Cochrane Database Syst Rev 2012 Mar 14;2012(3):CD005315. [doi: 10.1002/14651858.CD005315.pub2] [Medline: 22419308]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
HIT: health IT
PCC: patient-centered care
QOC: quality of care
SEIPS: Systems Engineering Initiative for Patient Safety

Edited by A Choudhury; submitted 24.09.23; peer-reviewed by Y Luo, F Albalwy; comments to author 08.02.24; revised version received 06.03.24; accepted 15.03.24; published 24.04.24.

Please cite as:
Elkefi S, Asan O
Validating the Effectiveness of the Patient-Centered Cancer Care Framework by Assessing the Impact of Work System Factors on Patient-Centered Care and Quality of Care: Interview Study With Newly Diagnosed Cancer Patients
JMIR Hum Factors 2024;11:e53053
URL: https://humanfactors.jmir.org/2024/1/e53053
doi:10.2196/53053
PMID:38656776

Original Paper

Development of a Real-Time Dashboard for Overdose Touchpoints: User-Centered Design Approach

Amey Salvi[1], MS; Logan A Gillenwater[2], BS; Brandon P Cockrum[3], MFA; Sarah E Wiehe[2,3], MPH, MD; Kaitlyn Christian[4], MPA; John Cayton[4], MAEd, MS; Timothy Bailey[4], MS; Katherine Schwartz[2], MPA, JD; Allyson L Dir[5], PhD; Bradley Ray[6], PhD; Matthew C Aalsma[2], PhD; Khairi Reda[1], PhD

[1]School of Informatics, Computing, and Engineering, Indiana University Indianapolis, Indianapolis, IN, United States
[2]Children's Health Services Research, Department of Pediatrics, Indiana University School of Medicine, Indianapolis, IN, United States
[3]Indiana Clinical and Translational Science Institute, Indianapolis, IN, United States
[4]Indiana Management Performance Hub, Indianapolis, IN, United States
[5]Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN, United States
[6]RTI International, Research Triangle Park, NC, United States

Corresponding Author:
Khairi Reda, PhD
School of Informatics, Computing, and Engineering
Indiana University Indianapolis
535 W Michigan St
Indianapolis, IN, 46202
United States
Phone: 1 317 274 5788
Email: redak@iu.edu

Abstract

Background: Overdose Fatality Review (OFR) is an important public health tool for shaping overdose prevention strategies in communities. However, OFR teams review only a few cases at a time, which typically represent a small fraction of the total fatalities in their jurisdiction. Such limited review could result in a partial understanding of local overdose patterns, leading to policy recommendations that do not fully address the broader community needs.

Objective: This study explored the potential to enhance conventional OFRs with a data dashboard, incorporating visualizations of touchpoints—events that precede overdoses—to highlight prevention opportunities.

Methods: We conducted 2 focus groups and a survey of OFR experts to characterize their information needs and design a real-time dashboard that tracks and measures decedents' past interactions with services in Indiana. Experts (N=27) were engaged, yielding insights on essential data features to incorporate and providing feedback to guide the development of visualizations.

Results: The findings highlighted the importance of showing decedents' interactions with health services (emergency medical services) and the justice system (incarcerations). Emphasis was also placed on maintaining decedent anonymity, particularly in small communities, and the need for training OFR members in data interpretation. The developed dashboard summarizes key touchpoint metrics, including prevalence, interaction frequency, and time intervals between touchpoints and overdoses, with data viewable at the county and state levels. In an initial evaluation, the dashboard was well received for its comprehensive data coverage and its potential for enhancing OFR recommendations and case selection.

Conclusions: The Indiana touchpoints dashboard is the first to display real-time visualizations that link administrative and overdose mortality data across the state. This resource equips local health officials and OFRs with timely, quantitative, and spatiotemporal insights into overdose risk factors in their communities, facilitating data-driven interventions and policy changes. However, fully integrating the dashboard into OFR practices will likely require training teams in data interpretation and decision-making.

(JMIR Hum Factors 2024;11:e57239) doi:10.2196/57239

KEYWORDS

overdose prevention; dashboards; fatality review; data integration; visualizations; visualization; dashboard; fatality; death; overdose; overdoses; overdosing; prevention; develop; development; design; interview; interviews; focus group; focus groups; touchpoints; touchpoint; substance abuse; drug abuse

Introduction

Background

The escalating drug overdose epidemic in the United States continues to pose a major public health challenge. Previous research has identified general risk factors that are linked to increased overdose rates [1-3], including unstable housing [4,5], recent release from incarceration [6,7], and frequent visits to the emergency department (ED) [8-11]. However, overdose risk factors exhibit considerable variation across communities and are influenced heavily by geographic and demographic disparities, particularly in access to health care and prevention services [9,12]. Moreover, the evolving nature of the epidemic has led to shifting risk profiles among different subpopulations [13]. These disparities underscore the need for timely and data-driven interventions that are tailored to the specific needs and challenges of local communities.

One mechanism for implementing targeted, community-specific interventions is through local Overdose Fatality Reviews (OFRs). Modeled after child fatality reviews [14,15], OFR teams comprise reviewers from multiple agencies who conduct collaborative, in-depth reviews of case files for individuals who have died of overdose [16,17]. Through these detailed case reviews, OFRs identify service gaps and recommend strategies to prevent future overdoses in their communities. The use of OFRs has gained momentum, with teams operating across various US localities [18]. However, current OFR practices primarily focus on reviewing only a handful of cases, typically 2 to 5 monthly or quarterly [19]. These cases typically represent a small fraction of the total fatalities occurring in their jurisdiction. While informative, the emphasis on a few individual cases could skew the review process, leading to OFRs making recommendations that do not fully address broader overdose trends.

As local governments continue to collect data on overdose events, there is an opportunity to leverage these data to enhance the OFR process. Previous research demonstrates the value of linking administrative data sets routinely collected by state governments (eg, calls to emergency services and incarceration records) with overdose mortality data [20-24]. For example, cross-referencing the records of decedents who experienced overdoses from across various data sets allows for uncovering their "touchpoints"—interactions with health and social services and other local systems they had before their overdose. When brought to light, touchpoints offer key opportunities to engage at-risk individuals and connect them with prevention services and treatments [25-27]. Analyses to identify touchpoints have so far been performed manually by researchers. However, the process is amenable to automation, enabling continuous assessment of touchpoint characteristics. The results can then be communicated in real time to local OFRs through a dashboard, providing review teams with up-to-date, quantitative information on the trajectories of decedents in their communities.

Dashboards have proven invaluable in public health settings [28,29] owing to their ability to visually summarize key metrics and statistics [30,31], thereby aiding surveillance and fostering evidence-based responses to emerging health threats [32,33]. Furthermore, dashboards are conducive to collaborative sense making among multiple individuals [34-36]. This feature makes them particularly suited to fatality review meetings, which are designed to be collaborative and deliberative in nature. Numerous dashboards have been developed to visualize drug overdose–related data [37-39]. However, existing solutions are primarily intended to surveil the level and distribution of overdoses as opposed to understanding events that *precede* them. Few of the earlier dashboards showcase touchpoints at the local level or update data in real time, making them less suited for understanding system-level gaps or for deriving prevention-oriented insights.

Aims

This study presents findings from human-centered research, design, development, and initial evaluation of a dashboard aimed at supporting OFR teams by visualizing overdose touchpoint statistics. The objective was to provide county-level OFR teams with timely and actionable data on events that consistently *precede* fatal overdoses in their communities. In doing so, we aimed to illuminate additional opportunities for interventions at the population level beyond what can be gleaned from individual fatality case reviews. The goal was to increase the chance of successful targeting and implementation of OFR recommendations. This stands to improve overdose prevention and reduce the number of preventable deaths.

Methods

Overview

To design a dashboard suitable for the needs of OFRs, we adopted a user-centered design framework [40,41] drawing on participatory methods to engage stakeholders in the process [42,43]. Specifically, we conducted focus groups with a panel of OFR experts to elicit perspectives on requirements and data needs, envision design possibilities, and document potential challenges. The elicited requirements were then used to develop exploratory visualizations of touchpoints data. The initial visualizations were further refined based on feedback from the expert panel. Subsequently, the revised visualizations were used to develop a web-based dashboard that is hosted by the Indiana state government.

Study Setting and Data Sources

We partnered with the state government of Indiana to prototype and develop the sought touchpoints dashboard. Indiana has a nationally recognized role in organizing and convening OFRs, with 28 active review teams organized at the county level and supported by the Indiana Department of Health. Similar to many other states, Indiana maintains a comprehensive and up-to-date database of fatal overdoses. This database includes all suspected accidental poisonings (coded as X40-X44), intentional poisonings (X60-X64), assaults by drug (X85), and cases of undetermined intent (Y10-Y14) that occurred among Indiana residents. In addition to overdose data, the state maintains administrative data sets from various agencies, including incarceration records, emergency and medical service use, and prescription dispensation. Importantly, these administrative data sets are linkable to the overdose mortality records. The Indiana Management Performance Hub (MPH), a state-level agency, serves as a central repository for these data sets, which are gathered from the corresponding agencies.

To identify events that precede drug-related fatalities, overdose cases are linked to administrative data sets at the individual level. This linking procedure is performed by the MPH using a probabilistic matching algorithm that considers identifiers such as the decedent's name, date of birth, and social security number, among others. This process allows for the reconstruction of past interactions with various touchpoints for each identifiable decedent. Subsequently, deidentified statistics about these interactions are pushed to the dashboard for visualization. This linkage process is performed weekly, enabling (near) real-time updates of the visualizations.

User-Centered Design Process

To inform the design of the dashboard, we conducted 2 focus groups with a panel of OFR experts. We recruited participants via email, inviting experienced OFR practitioners and early developers from across the United States. Our goal in these focus groups was to understand OFR information needs and leverage the panel's experience in conceptualizing, co-designing, and refining visualizations. The focus groups took place virtually using Zoom videoconferencing software (Zoom Video Communications). A virtual whiteboard was used to place and arrange "Post-it"–style notes. Participating experts were recruited from the same pool, with later focus groups involving fewer participants to allow for convergence and facilitate more in-depth feedback. The focus groups were video recorded, transcribed, and analyzed using thematic analysis techniques [44].

The first focus group sought to uncover data access barriers and needs for OFR teams. A total of 13 experts participated in the discussion. Participants were first prompted to share challenges and "pain points" regarding access to data. In a second activity, participants were divided into 2 breakout groups to identify key data attributes essential for review teams. They also gave high-level design parameters for the dashboard. Finally, participants reflected on their hopes and concerns for the dashboard's integration into OFR processes, emphasizing potential positive outcomes and addressing apprehensions.

On the basis of the findings of the initial focus group, we created a series of 6 initial visualizations that illustrate overdose touchpoints using a static snapshot of the MPH-linked data set described previously. These initial visualizations served as the foundation for a second focus group with the participation of 6 experts. During this session, a facilitator presented each of the 6 visualizations and prompted participants for feedback. Specifically, participants were asked to evaluate the ease of understanding of these visualizations and their potential usefulness in the OFR process. We sought additional input by conducting a survey of 5 experts. The survey presented the same initial visualizations and requested open-ended comments on their intuitiveness and utility. Insights gathered from the survey along with feedback obtained during the second focus group were used to refine the visualizations and develop an interactive dashboard.

Dashboard Evaluation

To obtain feedback on the final dashboard, we conducted an initial assessment with 3 OFR experts. Participants were asked to perform a series of data extraction tasks (eg, identifying the touchpoint with the highest prevalence). In addition, they were prompted to make recommendations based on the observed touchpoint patterns, simulating the use of the dashboard within a typical OFR meeting.

Ethical Considerations

This human-centered research was reviewed and approved by the Indiana University institutional review board (approval 17809). Participants received an information sheet explaining the study goals and procedures before agreeing to take part. The analysis of state mortality and administrative data sets, while not considered human participant research, followed state legal and ethical procedures. The dashboard displays only aggregate, population-level visualizations. No individual records are released or displayed to preserve anonymity. Furthermore, special care was taken to minimize the risk of reidentification by withholding actual event counts and substituting with percentages. Participants received a US $100 gift card as compensation.

Results

Overview

Participants highlighted barriers faced by OFRs in accessing and interpreting data within the context of fatality reviews. They also provided insights on what data attributes and features would be most useful for OFRs to look at. We report these findings and discuss how we incorporated them to create a real-time dashboard for visualizing overdose touchpoints.

Barriers to Accessing and Using Data

Data Accessibility

Several participants highlighted the lack of access to data as one of the major barriers in fatality reviews. Some of these barriers stem from challenges in sharing available data due to legal restrictions, data security, and privacy concerns:

> *Asking our state offices for data would result in, "Sorry, we can't share on the state level." There [needs to] be intergovernmental agreements between state police or our mental health or our human services or our health department.* [P2]

> *[Gaining access] is always an issue, and especially without laws that allow for the OFRs to get this. I*

> *know we had a lot of laws related to the child death review teams that I worked with that allowed us access to data, but it wasn't always the same for other death review teams.* [P11]

While recognizing existing regulatory and logistical obstacles, participants anticipated that increasing data access could empower OFRs to make more informed decisions:

> *We're trying to drive positive change that could maybe be implemented statewide, and they just give us a little bit. It [data] would give us the power to make better decisions.* [P2]

In addition to data access, the quality and accuracy of the data were also brought up as a prominent issue for OFRs, especially because of acknowledged variations in how data are coded and measured across different organizations. For example, 1 participant cited different standards for classifying services, noting that such inconsistencies could lead to misinterpretation:

> *When it's really law enforcement heavy, they're not understanding the public health ramifications of criminal justice involvement. It affects the lens from which data's being collected. So, when I go through the qualitative data...we've got people identifying jail substance use services as harm reduction, [and] you end up collecting some inaccurate data, which then misinforms the big picture.* [P7]

Influx of Case-Specific Data

While obtaining population-level data in certain arenas proved challenging, another concern was the vast amount of case-specific data that OFRs must already contend with. Participants noted that review teams are increasingly tasked with handling large volumes of individual reports from multiple systems, which often need to be manually and qualitatively analyzed at considerable time and effort:

> *OFRs collect an enormous amount of data, but you really need a whole army of researchers to be able to analyze it, especially the qualitative data. When the teams are putting forth all of these recommendations, it's just so hard to go through all the information and make a meaningful plan of it.* [P7]

Extensive data on individual death circumstances (as opposed to population-level statistics) reflect a conventional OFR focus on in-depth reviews of a few strategically selected cases. However, with the sheer number of overdose fatalities, it becomes difficult for OFRs to ensure that the selected cases represent the broader overdose patterns and risk factors prevalent in their community. One participant put it as follows:

> *[My experience] is that they would just randomly pick cases and then do a really deep dive into those cases, but you have no way to actually ensure that those are representative...And so, my hope had been that we would have certain [data] fields that we could have someone enter, and then that would allow us to do really large-scale analysis over the course of multiple years...[This] would have allowed us to really have a good sense as it relates to a variety of factors, but there just wasn't capacity. So, then we're just picking cases that look good or meet some theme to be able to have a more robust conversation at any given meeting. But again, they're not necessarily representative and you don't end up having the whole picture.* [P18]

Key Data Types and Attributes for the Dashboard

Participants identified key data attributes that they deemed essential for inclusion in a dashboard. We divided these attributes into 3 categories: touchpoints, social determinants of overdose risk, and case-specific data.

Touchpoints

Touchpoints represent interactions with systems and services before overdose. Thus, they serve as opportunities to connect people who use drugs with additional prevention services and treatments, potentially mitigating the risk of future overdoses. A frequently recurring set of touchpoints identified by experts was interaction with the justice system. For instance, the duration between a decedent's overdose and their last incarceration or residential treatment was cited as particularly important:

> *Were they justice involved or not at any point, but also the average distance in time from their last incarceration...So, to see were they in that window of high risk. And same if they were in residential treatments as average number of days.* [P3]
>
> *Average days out from treatment and incarceration because I feel like those are solid spaces that action can be taken.* [P5]

Several participants pointed to interactions with justice systems broadly as key touchpoints. Agencies such as county sheriffs, local police departments, and child protective services were thought to play a crucial role in an individual's risk of overdose both positively and negatively:

> *Justice systems can either be a force of treatment or a barrier to treatment. I think that involvement is really important...the extent of involvement can be really helpful to inform the justice system and the legislative changes that could help.* [P4]

Participants noted that data on criminal justice touchpoints might reveal new prevention opportunities or support policy recommendations, such as facilitating continued treatment for institutionalized individuals:

> *...keep people engaged in treatment, [such that] we're not disrupting treatment by violating [ie, rearresting] people and incarcerating them...It's a fruitful area for policy change. Most of our policy changes and recommendations from our OFR have been in the justice space.* [P3]

In addition to justice systems, participants noted interactions with health and medical facilities as crucial touchpoints. This included visits to the ED and emergency medical services (EMS):

> *Do we have one [attribute] here [on] the last date of medical intervention? Maybe like an ED visit or anything like that?* [P4]
>
> *There's an ED and EMS interaction right at the center there.* [P5]

Overall, three primary touchpoint categories emerged: (1) encounters with the justice system, such as incarceration; (2) engagement with health services, including ED and EMS interactions; and (3) involvement with residential treatment services. These touchpoints were recognized by participants as crucial opportunities for understanding risk factors and implementing services to close treatment gaps. Importantly, participants emphasized the typical interval between these touchpoints and overdose events as a critical feature to emphasize in the dashboard.

Social Determinants of Health

A second set of data attributes identified pertained to the social condition of the individuals themselves, which could shed light on factors that contribute to elevated overdose risk. For example, one of these factors was demographics:

> *Basic demographic information like poverty level, education level, homelessness. Anything that would affect those social determinants of health.* [P4]

A second factor was individuals' access to harm reduction services, as the same participant noted:

> *I was going to add...access to harm reduction services. So, what an environmental scan of resources or access to naloxone, treatment centers, syringe service programs, all those different community level access points.* [P4]

A third factor was housing, encompassing the shelter system and housing agencies:

> *Access to housing. Or maybe it's access to shelter because it could be both. There's housing policy, but then there's also the shelter systems.* [P5]

A fourth factor was the availability of transportation, which, according to participants, could influence an individual's access to treatment and harm reduction services:

> *Transportation between places: how easy is it for someone to get from point A to point B? Even if there's a syringe service program down the street, can they get to it? That kind of thing.* [P5]

Finally, participants also identified upstream social determinants such adverse childhood experiences as potentially relevant factors in assessing overdose risk:

> *...and some of that I think would fall under ACEs too because even if they're an adult, finding out if they were involved in that system as a child, trying to make some of those associations maybe.* [P5]

Case-Specific Data

Alongside touchpoints and social determinants of health, participants cited certain case-specific data, including toxicology reports, interviews with next of kin, and the decedent's circumstances at the time of death (eg, their position and whether they were alone). While these attributes are relevant to reviewing individual cases, they were not considered for inclusion in the dashboard as our primary objective was to offer population-level data that complement rather than supplant the conventional OFR case review model.

Apprehensions and Foreseen Challenges

Although participants were positive about the potential of the dashboard to enhance the OFR process, there were a few apprehensions. A major concern was the risk of unintentional identification of decedents in smaller counties, where there are fewer overdose deaths:

> *I've been aware of a couple different cases in relatively small communities where all the data says one thing, and of course, as a small community, we know exactly who we're talking about.* [P15]
>
> *I think one [concern] would be that the information might be too identifiable, especially for small communities.* [P8]

Participants discussed the ethics of displaying data that might be inaccurate or that could be misused (eg, by law enforcement) to target at-risk individuals:

> *...that it has inaccurate and bad data. And that it is used for evil rather than for good...That it's not used for bad downstream consequences kind of thing.* [P6]

Finally, participants raised the risk of misinterpreting data, noting that, while OFRs have expertise in studying individual histories of decedents to formulate recommendations, they are less familiar with analyzing population-level statistics. Some voiced reservations about OFR teams' data literacy and their ability to draw appropriate inferences from such quantitative data. For instance, 1 participant gave an example of how a decrease in emergency medical events could be erroneously interpreted as a reduction in overdoses when it might only reflect fewer 911 calls:

> *That [error] where you have a number and you think it means one thing, but it means another thing...You have measured something, but not the thing that you are taking that thing to be.* [P1]

Others commented on the potential downstream consequences of misinterpreting data, which could manifest as inappropriate or even detrimental recommendations:

> *We've seen this trend in our data. That probably means X, Y, Z. And you might be right. You might be very wrong, and the data might be used to justify a policy or programmatic intervention that could in fact exacerbate it.* [P17]

Helping users interpret data accurately was deemed by participants as a critical consideration for the dashboard. Equally important was not to inundate OFRs with even more (population-level) data that teams may lack the bandwidth or data literacy skills to act upon. These insights underscore the need to craft intuitive data visualizations that can be comprehended accurately with minimal effort. Moreover, such displays should actively guide OFR teams into making valid inferences from the data presented.

Touchpoints Selection

Our observations point to a longstanding limitation of current OFR practices, which focus on reviewing a handful of overdose cases at every meeting. OFR experts appeared to recognize the shortcomings of this model when pitted against the sheer volume of overdoses. Simultaneously, participants expressed strong interest in accessing additional data sets that would paint a broader picture of overdose risk factors and touchpoints in their community, provided that these data were consistently coded, intuitively summarized, and presented in a manner that did not overburden review teams.

Among the data emphasized by participants, touchpoints emerged as particularly actionable as they represent system interactions preceding overdose events. For instance, the proportion of decedents who used various touchpoints offers predictive power to identify the most effective points within the system for targeting at-risk individuals with prevention services. Moreover, understanding the typical time window between a touchpoint and an overdose event, along with the frequency of touchpoint use, can assist in designing interventions, including their timing and regularity.

Drawing on the insights of the expert panel and data availability in Indiana, we incorporated 5 touchpoint types into the dashboard: jail bookings, prison releases, visits to the ED, encounters with EMS, and prescriptions for controlled substances (eg, opioid analgesics). We excluded ED and EMS encounters occurring within a 24-hour window of death as those are likely to represent interactions directly related to the overdose event as opposed to potential touchpoints for prevention purposes. Interactions with both justice and medical systems were identified as key by the expert panel. Prescriptions for scheduled drugs, such as opioid analgesics, were included as touchpoints due to their established association with overdose risk [25]. We also included the dispensation of buprenorphine prescriptions as a touchpoint in the initial dashboard design. However, concerns were raised that singling medication for opioid use disorder as a separate touchpoint could cause it to be misconstrued as a causal risk factor for overdose. Consequently, buprenorphine data were merged and included among the general prescription dispensation touchpoint for scheduled drugs. Table 1 provides a summary of these touchpoints as highlighted by participants and featured in the dashboard. Although interactions with residential treatment services were identified as an important touchpoint by participants, related data are not centrally tracked by the state and, hence, were not available for inclusion in the dashboard. Moreover, social determinants of health are not currently included despite their relevance as the dashboard was intended to prioritize opportunities for immediate as opposed to upstream prevention. Case-specific attributes were also not considered for inclusion because they would be redundant to the traditional OFR case review process.

Table 1. Data types and attributes as identified by experts and featured in the dashboard.

Data type and attribute	Identified by expert panel?	Included in dashboard?
Touchpoint		
Jail booking	No	Yes
Release from prison	Yes	Yes
Visit to the ED[a]	Yes	Yes
Encounter with EMS[b]	Yes	Yes
Interaction with residential treatment services	Yes	No
Prescription dispensation for scheduled drugs, including opioid analgesics and MOUD[c]	No	Yes
Social determinants		
Demographics	Yes	No
Educational level	Yes	No
Poverty	Yes	No
Access to harm reduction services	Yes	No
Housing	Yes	No
Access to transportation	Yes	No
Adverse childhood experiences	Yes	No
Case-specific attributes		
Toxicology report	Yes	No
Next-of-kin interviews	Yes	No
Circumstances of death (eg, body position and presence of witnesses)	Yes	No

[a]ED: emergency department.

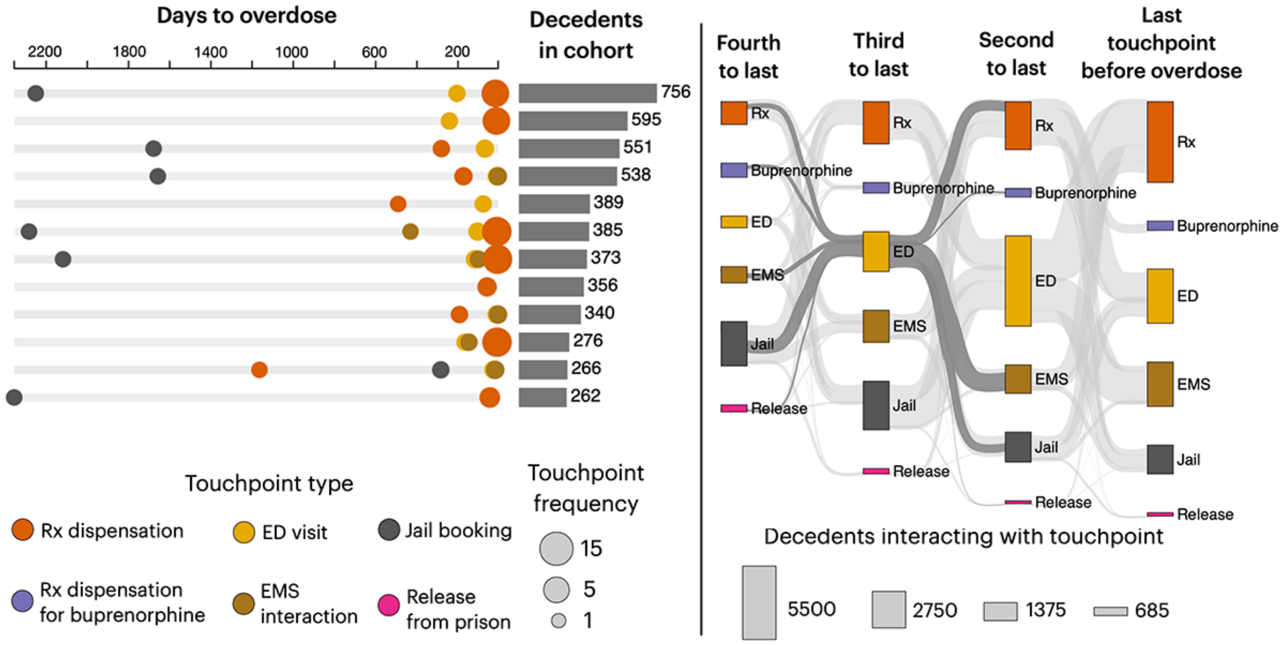
[b]EMS: emergency medical services.

[c]MOUD: medication for opioid use disorder.

Initial Visualization Attempts

Our initial visualization focused on timelines, illustrating cohorts of decedents who exhibited similar patterns of touchpoints before overdosing. For example, in Figure 1 (left), each row represents hundreds of decedents who exhibited a similar touchpoint sequence (eg, jail booking followed by one or more ED visits and then a series of prescriptions). This particular visualization was inspired by OFR teams' use of timelines to represent the histories of individuals discussed during case reviews. However, these initial visualizations received mixed reviews from the expert panel—while they were considered appealing and "interesting," the focus on cohorts was seen as providing excessive detail for OFRs. This feedback was used to revise the visualizations and develop a final dashboard.

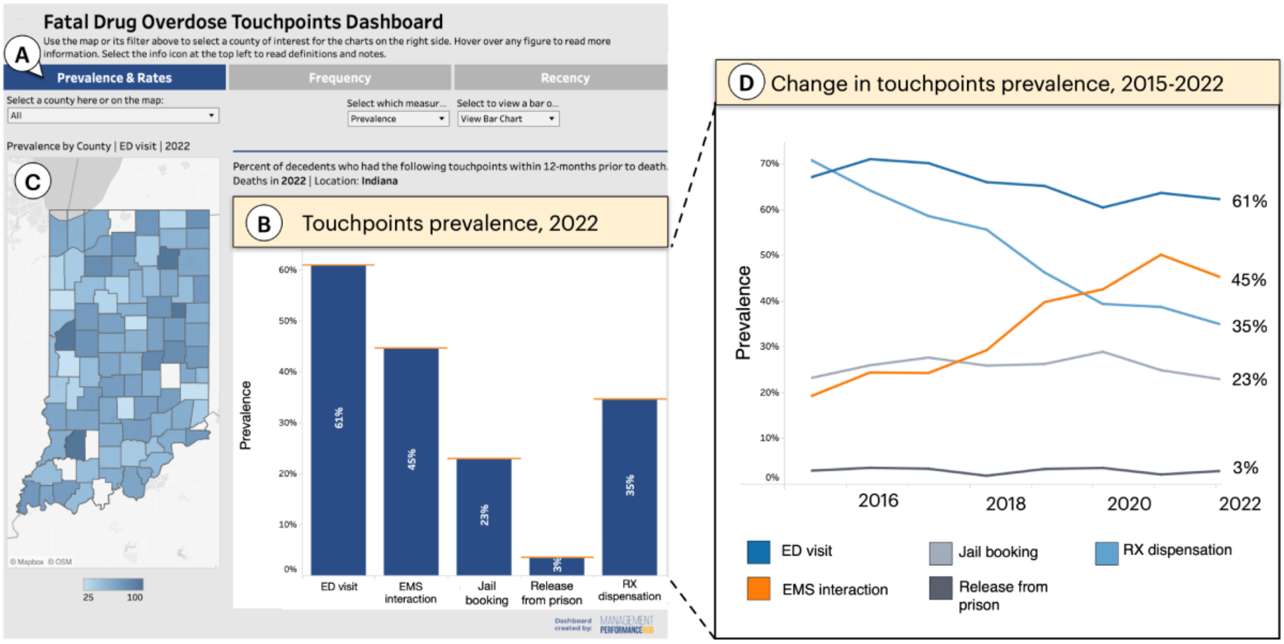
Figure 1. A total of 2 initial visual representations of touchpoints in Indiana (aggregate data from 2015 to 2022). On the left, a timeline-based visualization illustrates the cohorts of decedents with distinct sequences of touchpoints. The visualization depicts the average number of days to fatal overdose (circle position) and frequency of interaction with a touchpoint (circle diameter). For example, the first row shows 756 individuals who experienced a jail booking approximately 6 years before overdose, followed by a sequence of emergency department (ED) visits and medical prescription (Rx) dispensations, the last of which typically occurred approximately 200 and 90 days before overdose, respectively. A Sankey diagram (right) displays the temporal ordering of (up to 4) touchpoints but without showing durations. EMS: emergency medical services.



Final Dashboard

The dashboard consists of 3 primary displays (Figure 2A) showing the prevalence and rates, frequency, and recency for the 5 touchpoints. The dashboard can be accessed at the MPH website [45].

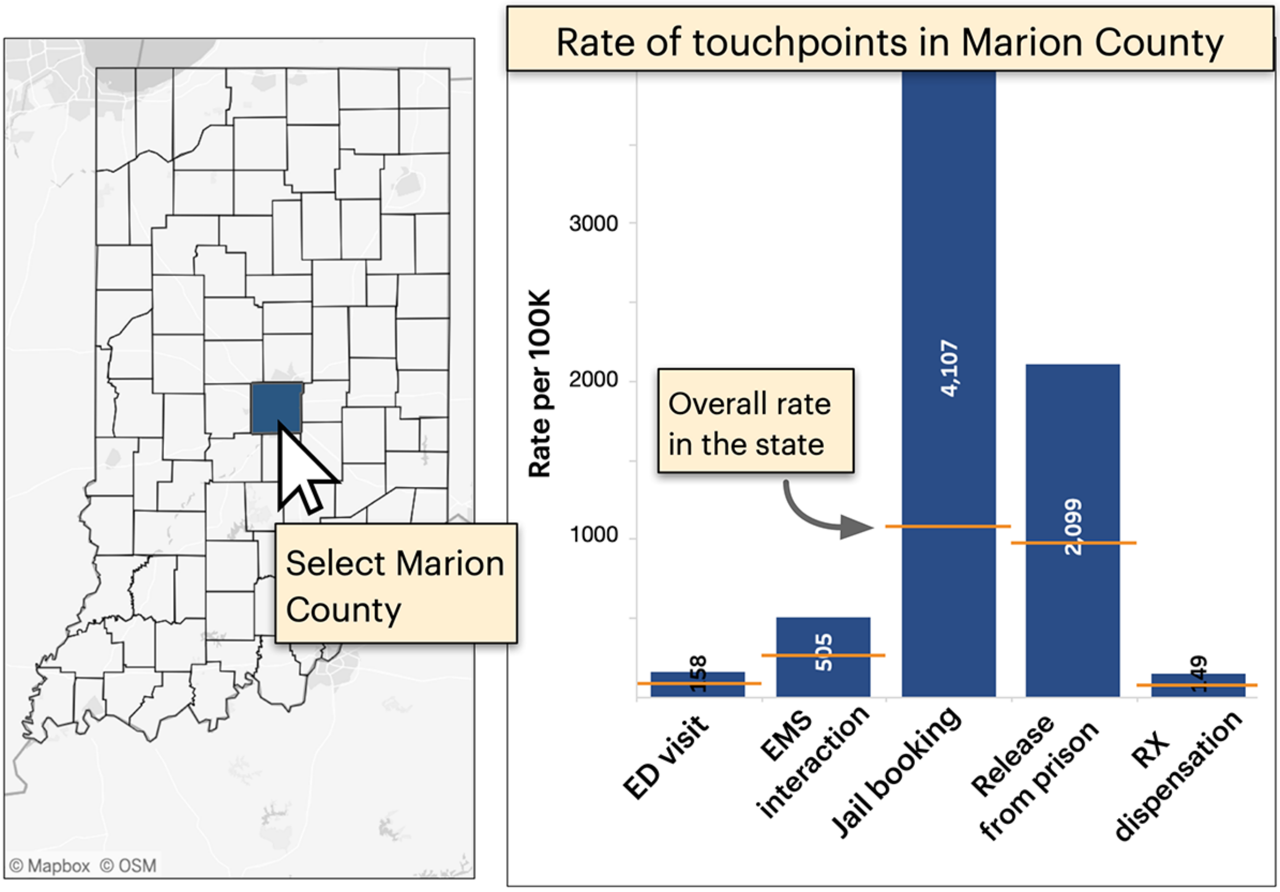
Figure 2. The final dashboard showing overall touchpoints prevalence in Indiana. (A) Buttons enable the user to switch among 4 measures: prevalence, rates, frequency, and recency of touchpoints. (B) The selected measure is visualized here as a bar chart comparing touchpoint prevalence (ie, the percentage of decedents who used each of the 5 touchpoints). (C) A map shows touchpoint prevalence (in this case for emergency department [ED] visits) by county, where darker shades of blue indicate higher prevalence. (D) As an alternative to the bar chart, a line graph allows users to observe how the prevalence of the touchpoints changes from year to year. EMS: emergency medical services; Rx: medical prescription.



Prevalence and Rates

By default, the dashboard displays touchpoint *prevalence*, depicting the percentage of decedents who used various touchpoints in the 12 months preceding overdose. For instance, in 2022, the highest-prevalence touchpoint was the ED, with 61% of individuals who overdosed in Indiana having visited the ED within a year before dying (Figure 2B). The user can also see the change in prevalence over time. For example, the data show that the prevalence of ED visits decreased over time, whereas the proportion of decedents who use EMS increased >2 times between 2015 and 2022 (Figure 2D). In addition to showing state levels, the dashboard can break down the data by county. For instance, the user can see the prevalence of ED visits in different counties on a map (Figure 2C). Notably, the map shows 4 counties in which practically all decedents had visited the ED a year before their overdose. The map can also be used to filter the bar or line graph displays. For example, clicking on Marion County, the most populous in Indiana, updates the display to show statistics for Marion only (Figure 3).

Figure 3. Rates showing the fraction of individuals who experienced a fatal overdose for every 100,000 people who use a touchpoint (right). A map allows the user to filter the data by county, in this example, to show rates for Marion County only. Orange dash marks depict the state average for context. ED: emergency department; EMS: emergency medical services; Rx: medical prescription.

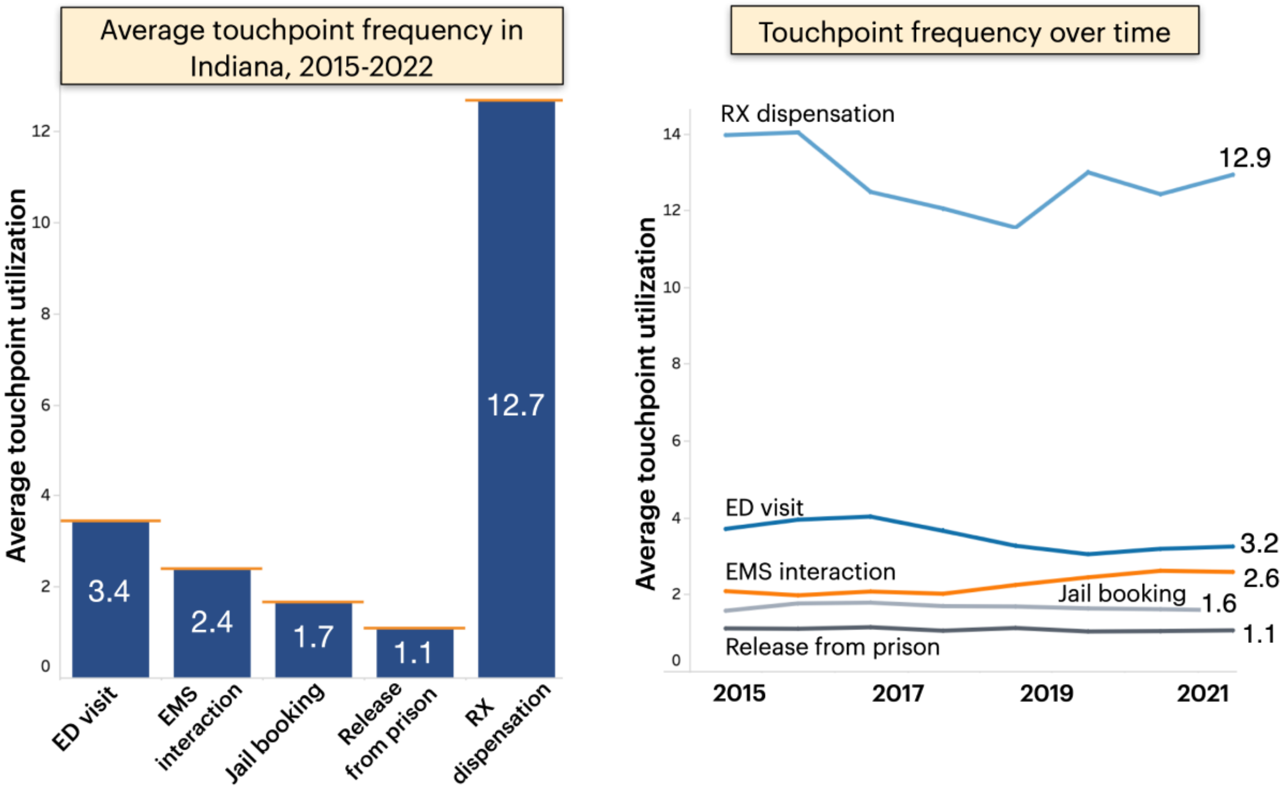


In addition to prevalence, the dashboard visualizes the *rate* of touchpoints among decedents. These rates depict the number of fatal overdose cases per 100,000 individuals who typically use services such as the ED. Unlike prevalence, which indicates the likelihood of a decedent using a touchpoint, rates reveal the probability of a fatal overdose after using 1 of the 5 legal or medical touchpoints included in the dashboard. Both measures are important for resource allocation—while prevalence helps users identify touchpoints with the broadest reach, rates can reveal more "efficient" touchpoints for targeted interventions. For example, consider jail bookings and releases from prison (Figure 3 [right]), which exhibit the highest rates among touchpoints in Marion County. This offers a high-specificity opportunity to focus on individuals at a greater risk of overdosing despite these touchpoints exhibiting relatively moderate to low prevalence at the state level (23% and 3%, respectively, as depicted in Figure 2 [left]).

Touchpoint Frequency

The second display summarizes the average number of interactions a decedent had with a touchpoint in the year preceding their overdose (Figure 4). Notably, the most frequently used touchpoint in the state is medical prescription (Rx) dispensation for controlled substances, such as an opioid analgesic (12.7 events on average at the time of writing). The user can also see how this frequency changes yearly (Figure 4 [right]). The line graph shows relatively stable use for ED, EMS, and criminal justice services, with the average number of Rx dispensations trending down slightly.

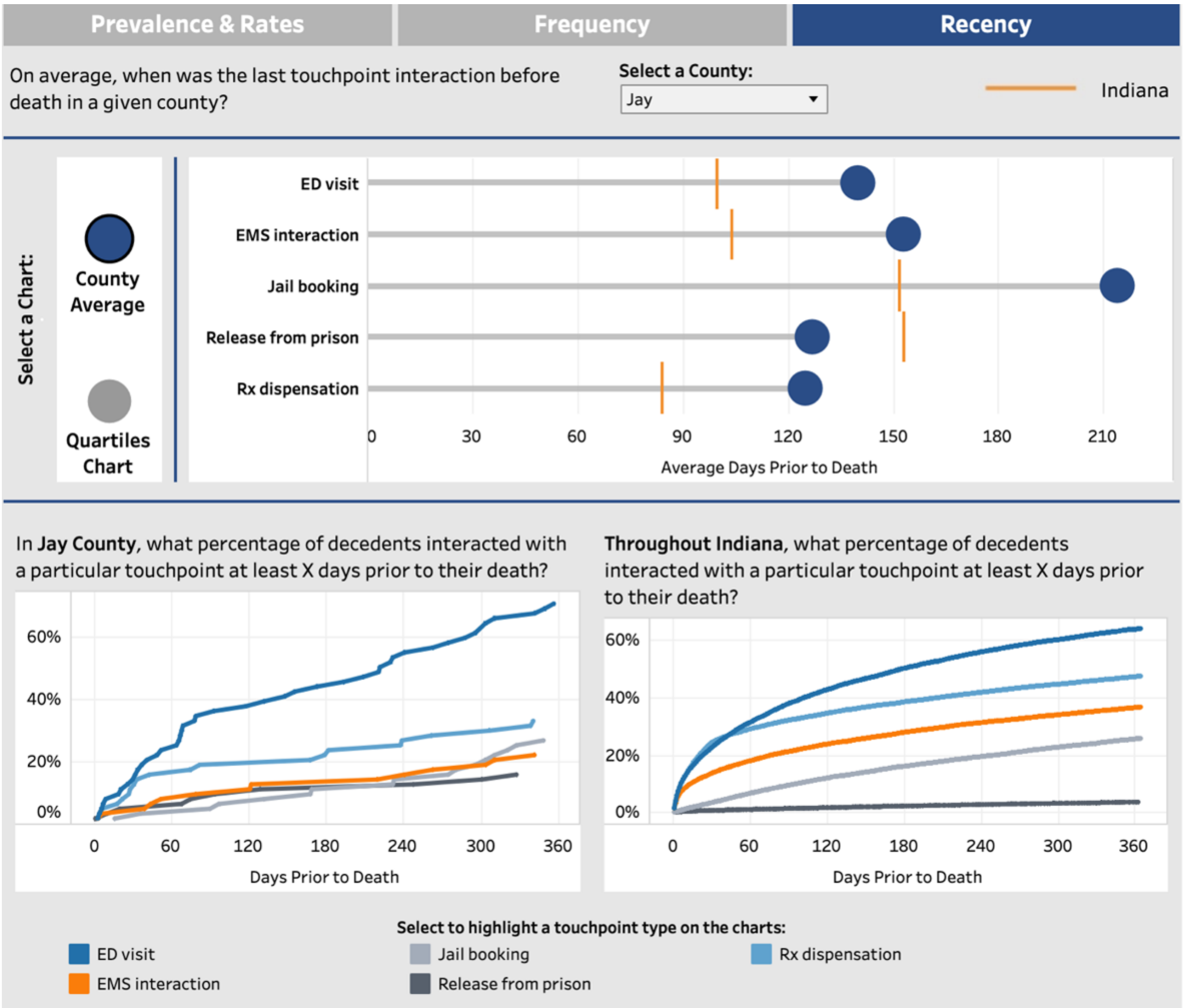
Figure 4. Average number of interactions with the 5 touchpoints from 2015 to 2022 (left) alongside a year-by-year breakdown. ED: emergency department; EMS: emergency medical services; Rx: medical prescription.



Recency

The timing of interaction with services was identified as a key factor for OFRs. Accordingly, the *recency* display illustrates the typical time intervals between final touchpoints and overdose events (Figure 5). The top features a “lollipop” chart depicting the number of days on average between the most recent interaction and the overdose (Figure 5 [top]). In this example, jail bookings in Jay County (selectable by the user) occur approximately 210 days on average before a fatal overdose compared to approximately 150 days for the entirety of Indiana. Conversely, releases from prison tend to happen approximately 120 days before the overdose, closer relative to the state average. The bottom visualizations show a curve for each touchpoint representing the cumulative percentage of individuals who could have been engaged at various time points relative to their time of death. In this case, approximately 27% of decedents in Jay County could have been engaged through an Rx dispensation touchpoint 30 days before an overdose.

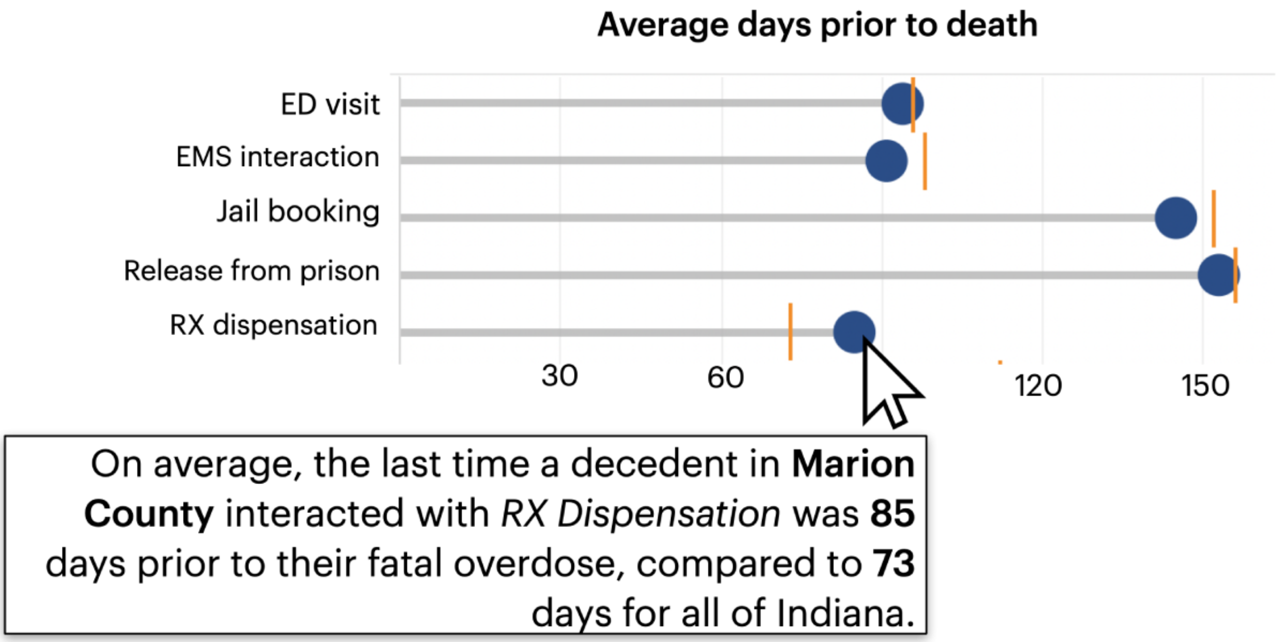
Figure 5. The average time gap between the final interaction and overdose events across different touchpoints (top). The lower section comprises 2 charts demonstrating the cumulative reach of touchpoints at varying time intervals, comparing the selected county (bottom left) with the state average (bottom right). ED: emergency department; EMS: emergency medical services; Rx: medical prescription.



Aiding Data Interpretation

One concern that emerged during the focus groups regarded OFR teams’ ability to interpret population-level statistics. To aid users in making sense of these data, the dashboard provides tooltips in the form of short text annotations that explain the interpretation of each visualization. For instance, in the recency chart, the text clarifies that the points depict the average number of days between a touchpoint and an overdose event (Figure 6).

Figure 6. Tooltips appear throughout the dashboard to promote accurate data interpretation. ED: emergency department; EMS: emergency medical services; Rx: medical prescription.



Initial Evaluation Results

We invited 3 OFR experts to review and provide feedback on the dashboard. They commented on features they thought were beneficial. They also provided suggestions on how to ensure dashboard integration into OFR practices. One of the notable strengths of the dashboard was its comprehensive data coverage, a feature that was highly appreciated by all participants. They specifically praised the breakdown of touchpoints on a county basis, a level of granularity that is often lacking in existing dashboards. The inclusion of small counties, the data on which can be especially difficult to obtain, was recognized as a significant advantage. Participants also appreciated the ability to compare different counties through the map, along with the ability to juxtapose county-specific data against state averages.

Among the various visualizations, the recency chart (referred to as the "timeline") stood out for its depiction of events leading up to overdoses. Participants thought that these temporal data, which can be difficult to obtain at the population level, can help in tailoring interventions:

It is interesting to see this [chart], and to know what can be done with data. We can check the timeline and help implement a strategy. Through these strategies, we can outline short, medium, and long-term goals.

In thinking about how the dashboard might complement existing OFR practices, participants highlighted its usefulness in guiding case selection for review and helping OFRs build a representative case profile. One participant specifically noted the potential of the dashboard in conducting "community data review" to explore "what is going on in my community." Moreover, the dashboard's availability on a publicly accessible URL was lauded as "a wonderful resource," extending its value to audiences beyond OFRs. The discussion opened the door for offering some form of training or educational support to OFR members, equipping review teams with skills to interpret quantitative data. One participant suggested the addition of a "demo video to help interpret and apply the data." Another suggested the need to specifically focus on OFR facilitators as crucial personnel for communicating data insights to review teams:

> *I don't think they [members of the review teams], will be able to fully understand the data, so training the facilitator will be key.*

Discussion

Principal Findings

OFR teams are proliferating in the United States, becoming an important public health tool to combat the drug overdose crisis. Traditional fatality reviews, often limited to a few cases, do not fully capture the broader overdose trends, especially in communities with numerous drug-related fatalities. This research aimed to enhance OFR data use by addressing data access barriers, identifying information needs, and creating actionable visualizations of population-level overdose data.

Our findings shed light on challenges that OFR teams face in accessing timely data, frequently impeded by legal constraints. When available, these data can often be inconsistent, for example, in the coding of events and classification of services. Despite these challenges, OFR teams seemed keen on incorporating a wider range of data into their review to better understand the factors contributing to overdose risks in their communities. Notably, the expert panel highlighted several key touchpoints, including incarcerations, interactions with substance treatment services, and visits to medical facilities such as EDs.

Some of these touchpoints have been previously recognized as opportunities for delivering prevention services [25,46,47]. For example, the time window following a prison release has been identified as a particularly critical and risky period, making this touchpoint a highly specific and valuable opportunity for administering prevention services [48-50]. However, effectively sharing these data insights with OFRs remains a challenge. Our findings suggest that a dashboard linking state administrative and mortality data could effectively provide local OFRs with insights on the timing and distribution of touchpoints. To explore this potential, we partnered with the Indiana state government and developed a dashboard that collates and visualizes data on 5 touchpoints at the county level, enabling OFR teams to see statistics and patterns on events that *precede* fatal overdoses in their community. To our knowledge, this is the first system to automatically analyze touchpoint characteristics and offer (near) real-time visualizations of their prevalence, frequency, and timing tailored to the local scale of OFR teams. In designing the dashboard, we specifically focused on this user group and prioritized actionable data that shed light on local prevention opportunities. The developed touchpoint dashboard stands in contrast to earlier dashboards for opioid prescription and overdose data, which are meant for the public or nonspecified stakeholders.

Our OFR expert panel suggested that one of the most crucial pieces of information is the timing of touchpoints—specifically, the average duration between an individual's last encounter and their overdose. The dashboard prominently features these data in a lollipop chart comparing the *recency* of various touchpoints. In addition, we incorporated displays of touchpoint prevalence and rates, providing insights into the reach of touchpoints and the specificity they afford for targeting individuals who are at high risk of overdose. The dashboard purposely uses familiar visualizations, including bar and line graphs and choropleth maps, to appeal to review teams who may be novice visualization users [51]. Importantly, the dashboard breaks down these statistics at the county level, aligning with how OFRs are organized in Indiana. By visualizing data "close to home," we aimed to improve the actionability of the dashboard [52]. However, users can easily compare county data to state averages or those of other similar counties.

Our initial evaluations show promise for the dashboard's usefulness. However, successfully integrating the dashboard into OFR practices will likely require training for OFR members, many of whom lack expertise in data analysis—a point that was notably underscored by the expert panel. In particular, teams may need educational support in how to interpret population-level features, such as the difference between the prevalence and rates of touchpoints. Regular meetings with OFR users could also help uncover usability issues and gauge dashboard adoption by review teams.

While the dashboard offers detailed insights into community touchpoints, it omits data on social determinants such as race, educational level, and access to housing and harm reduction services. These factors can be important for understanding overdose risks, as per our expert panel and research findings [53,54]. Future versions of the dashboard could incorporate local statistics on these risk factors. Furthermore, it is possible to expand the current list of touchpoints to include specific events associated with social determinants, such as loss of housing or employment. These additional touchpoints could offer further intervention avenues to disrupt pathways from marginalization to overdose [55]. Another limitation is that, while the dashboard includes critical touchpoints such as ED and EMS encounters, these events currently lack classification. Adding a breakdown of these touchpoints, for example, by distinguishing between substance-related versus other EMS encounters, could enable OFR teams to further tailor their recommendations.

The experts interviewed also sought demographic breakdowns of touchpoint data, in part to ensure that diverse populations would benefit from interventions at touchpoints. Unfortunately, this feature was not included in the current dashboard due to reidentification risks, particularly in rural areas that have fewer overdoses. In the future, the dashboard could be modified to provide a demographic breakdown of touchpoints at the aggregate (eg, state) level to substantially decrease the risk of reidentification instead of withholding these data altogether. To further protect individual confidentiality, which was a key concern of our expert panel, the dashboard presents data as percentages (eg, the proportion of decedents who were released from prison within a year before their overdose) and rates. Withholding the actual counts for events helps prevent the inference of individual identity in places where those counts are low. The dashboard provides a visual warning for statistics based on <20 cases, cautioning users against drawing strong conclusions from small samples. Future work could use more advanced privacy-preserving techniques [56,57], thus allowing for the display of a wider range of attributes without jeopardizing anonymity.

Although our dashboard is specific to Indiana, we believe that the approach could be adapted for other US states and localities. This expansion requires access to overdose mortality records that can be algorithmically cross-referenced with other administrative data sets. Many states already have data infrastructure for such linked analyses [58,59]. We estimate that the development and maintenance of the dashboard over 2 years will require approximately 350 personnel hours assuming the availability of data. The prevalence of overdose dashboards [39,60] indicates both the technical feasibility of creating such tools and the interest in them from the public health community. Our research demonstrates that dashboards can go beyond surveillance to directly visualize actionable prevention opportunities.

Conclusions

OFRs can play a crucial public health role in understanding overdose cases and recommending prevention strategies. This study explored the potential for enhancing these reviews with population-level data for broader, quantitative insights into risk factors. Following a user-centered design process, we developed a dashboard that tracks and visualizes decedents' encounters with medical and justice systems at the county level. Although initially designed for Indiana, the dashboard can be adapted to other localities, leveraging administrative and mortality data typically collected by local governments. Preliminary evaluation shows the potential utility of the dashboard for analysis and case selection but emphasizes the need for training OFR members in data interpretation and decision-making.

Acknowledgments

The authors thank the Indiana Department of Health; the Department of Homeland Security; the Indiana State Police; the Professional Licensing Agency; the Department of Correction; and the Governor's Drug Prevention, Treatment, and Enforcement Office for providing data for use in the dashboard and for feedback. This research was supported by the Helping to End Addiction Long-term Initiative through the Data 2 Action program (award R61DA057660). The work was also supported in part by Research Jam, the Indiana Clinical and Translational Sciences Institute Patient Engagement Core through an award from the National Institutes of Health, National Center for Advancing Translational Sciences Clinical and Translational Science Award (award UM1TR004402). The authors acknowledge Michelle Tran, BS, for her contributions to this work. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health or participating organizations.

Data Availability

The video recording data sets generated during and analyzed during this study are not publicly available to protect participants' confidentiality but are available from the corresponding author on reasonable request. The aggregate touchpoints data sets are available in the Management Performance Hub (MPH) data repository [61].

Authors' Contributions

AS and LAG analyzed the data and wrote the manuscript. BPC conducted the focus group and analyzed the data. SEW directed the human-centered research process and provided manuscript revisions. KC, JC, and TB developed the dashboard and the touchpoints data integration process. KS, ALD, BR, MCA, and KR edited and provided manuscript revisions. BR, MCA, and KR conceptualized the study and designed the research.

Conflicts of Interest

None declared.

References

1. Brady JE, Giglio R, Keyes KM, DiMaggio C, Li G. Risk markers for fatal and non-fatal prescription drug overdose: a meta-analysis. Inj Epidemiol 2017 Dec 7;4(1):24 [FREE Full text] [doi: 10.1186/s40621-017-0118-7] [Medline: 28762157]
2. Davoli M, Perucci CA, Forastiere F, Doyle P, Rapiti E, Zaccarelli M, et al. Risk factors for overdose mortality: a case-control study within a cohort of intravenous drug users. Int J Epidemiol 1993 Apr;22(2):273-277. [doi: 10.1093/ije/22.2.273] [Medline: 8505184]
3. Webster LR. Risk factors for opioid-use disorder and overdose. Anesth Analg 2017 Nov;125(5):1741-1748. [doi: 10.1213/ANE.0000000000002496] [Medline: 29049118]
4. Lyons RM, Yule AM, Schiff D, Bagley SM, Wilens TE. Risk factors for drug overdose in young people: a systematic review of the literature. J Child Adolesc Psychopharmacol 2019 Aug 01;29(7):487-497 [FREE Full text] [doi: 10.1089/cap.2019.0013] [Medline: 31246496]
5. Yamamoto A, Needleman J, Gelberg L, Kominski G, Shoptaw S, Tsugawa Y. Association between homelessness and opioid overdose and opioid-related hospital admissions/emergency department visits. Soc Sci Med 2019 Dec;242:112585 [FREE Full text] [doi: 10.1016/j.socscimed.2019.112585] [Medline: 31634808]
6. Binswanger IA, Blatchford PJ, Mueller SR, Stern MF. Mortality after prison release: opioid overdose and other causes of death, risk factors, and time trends from 1999 to 2009. Ann Intern Med 2013 Nov 05;159(9):592-600 [FREE Full text] [doi: 10.7326/0003-4819-159-9-201311050-00005] [Medline: 24189594]
7. Bukten A, Stavseth MR, Skurtveit S, Tverdal A, Strang J, Clausen T. High risk of overdose death following release from prison: variations in mortality during a 15-year observation period. Addiction 2017 Aug 07;112(8):1432-1439. [doi: 10.1111/add.13803] [Medline: 28319291]
8. Barboza G, Angulski K, Hines L, Brown P. Variability in opioid-related drug overdoses, social distancing, and area-level deprivation during the COVID-19 pandemic: a Bayesian spatiotemporal analysis. J Urban Health 2022 Oct 06;99(5):873-886 [FREE Full text] [doi: 10.1007/s11524-022-00675-x] [Medline: 36068454]
9. Forati AM, Ghose R, Mantsch JR. Examining opioid overdose deaths across communities defined by racial composition: a multiscale geographically weighted regression approach. J Urban Health 2021 Aug 06;98(4):551-562 [FREE Full text] [doi: 10.1007/s11524-021-00554-x] [Medline: 34231120]
10. Ghose R, Forati AM, Mantsch JR. Impact of the COVID-19 pandemic on opioid overdose deaths: a spatiotemporal analysis. J Urban Health 2022 Apr 18;99(2):316-327 [FREE Full text] [doi: 10.1007/s11524-022-00610-0] [Medline: 35181834]

11. Brady JE, DiMaggio CJ, Keyes KM, Doyle JJ, Richardson LD, Li G. Emergency department utilization and subsequent prescription drug overdose death. Ann Epidemiol 2015 Aug;25(8):613-9.e2 [FREE Full text] [doi: 10.1016/j.annepidem.2015.03.018] [Medline: 25935710]
12. Choe K, Zinn E, Lu K, Hoang D, Yang LH. Impact of COVID-19 pandemic on chronic pain and opioid use in marginalized populations: a scoping review. Front Public Health 2023 Apr 17;11:1046683 [FREE Full text] [doi: 10.3389/fpubh.2023.1046683] [Medline: 37139395]
13. Foglia R, Kline A, Cooperman NA. New and emerging opioid overdose risk factors. Curr Addict Rep 2021 Apr 22;8(2):319-329 [FREE Full text] [doi: 10.1007/s40429-021-00368-6] [Medline: 33907663]
14. Durfee M, Parra JM, Alexander R. Child fatality review teams. Pediatr Clin North Am 2009 Apr;56(2):379-387. [doi: 10.1016/j.pcl.2009.01.004] [Medline: 19358922]
15. Rebbert-Franklin K, Haas E, Singal P, Cherico-Hsii S, Baier M, Collins K, et al. Development of Maryland local overdose fatality review teams: a localized, interdisciplinary approach to combat the growing problem of drug overdose deaths. Health Promot Pract 2016 Jul 18;17(4):596-600. [doi: 10.1177/1524839916632549] [Medline: 27091609]
16. Jones C, Bracewell K, Clegg A, Stanley N, Chantler K. Domestic homicide review committees' recommendations and impacts: a systematic review. Homicide Stud 2022 Apr 08;28(1):78-98 [FREE Full text] [doi: 10.1177/10887679221081788]
17. Parrish JW, Schnitzer PG, Lanier P, Shanahan ME, Daniels JL, Marshall SW. Classification of maltreatment-related mortality by child death review teams: how reliable are they? Child Abuse Negl 2017 May;67:362-370. [doi: 10.1016/j.chiabu.2017.03.003] [Medline: 28365427]
18. Evaluation profile for overdose fatality reviews. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/overdose-prevention/media/pdfs/OD2A_EvalProfile_OverdoseFatalityReviews_508.pdf [accessed 2024-01-30]
19. Haas E, Truong C, Bartolomei-Hill L, Baier M, Bazron B, Rebbert-Franklin K. Local overdose fatality review team recommendations for overdose death prevention. Health Promot Pract 2019 Jul 11;20(4):553-564. [doi: 10.1177/1524839918797617] [Medline: 30205737]
20. Keen C, Young JT, Borschmann R, Kinner SA. Non-fatal drug overdose after release from prison: a prospective data linkage study. Drug Alcohol Depend 2020 Jan 01;206:107707. [doi: 10.1016/j.drugalcdep.2019.107707] [Medline: 31757517]
21. McLeod KE, Slaunwhite AK, Zhao B, Moe J, Purssell R, Gan W, et al. Comparing mortality and healthcare utilization in the year following a paramedic-attended non-fatal overdose among people who were and were not transported to hospital: a prospective cohort study using linked administrative health data. Drug Alcohol Depend 2021 Jan 01;218:108381. [doi: 10.1016/j.drugalcdep.2020.108381] [Medline: 33158663]
22. Shefner RT, Sloan JS, Sandler KR, Anderson ED. Missed opportunities: arrest and court touchpoints for individuals who fatally overdosed in Philadelphia in 2016. Int J Drug Policy 2020 Apr;78:102724. [doi: 10.1016/j.drugpo.2020.102724] [Medline: 32279054]
23. Stoové MA, Dietze PM, Jolley D. Overdose deaths following previous non-fatal heroin overdose: record linkage of ambulance attendance and death registry data. Drug Alcohol Rev 2009 Jul 06;28(4):347-352. [doi: 10.1111/j.1465-3362.2009.00057.x] [Medline: 19594787]
24. Tseregounis IE, Henry SG. Assessing opioid overdose risk: a review of clinical prediction models utilizing patient-level data. Transl Res 2021 Aug;234:74-87 [FREE Full text] [doi: 10.1016/j.trsl.2021.03.012] [Medline: 33762186]
25. Larochelle MR, Bernstein R, Bernson D, Land T, Stopka TJ, Rose AJ, et al. Touchpoints - opportunities to predict and prevent opioid overdose: a cohort study. Drug Alcohol Depend 2019 Nov 01;204:107537 [FREE Full text] [doi: 10.1016/j.drugalcdep.2019.06.039] [Medline: 31521956]
26. Ray B, Christian K, Bailey T, Alton M, Proctor A, Haggerty J, et al. Antecedents of fatal overdose in an adult cohort identified through administrative record linkage in Indiana, 2015-2022. Drug Alcohol Depend 2023 Jun 01;247:109891 [FREE Full text] [doi: 10.1016/j.drugalcdep.2023.109891] [Medline: 37120921]
27. Krawczyk N, Schneider KE, Eisenberg MD, Richards TM, Ferris L, Mojtabai R, et al. Opioid overdose death following criminal justice involvement: linking statewide corrections and hospital databases to detect individuals at highest risk. Drug Alcohol Depend 2020 Aug 01;213:107997. [doi: 10.1016/j.drugalcdep.2020.107997] [Medline: 32534407]
28. Dasgupta N, Kapadia F. The future of the public health data dashboard. Am J Public Health 2022 Jun;112(6):886-888. [doi: 10.2105/ajph.2022.306871]
29. Thorpe LE, Gourevitch MN. Data dashboards for advancing health and equity: proving their promise? Am J Public Health 2022 Jun;112(6):889-892. [doi: 10.2105/ajph.2022.306847]
30. Bach B, Freeman E, Abdul-Rahman A, Turkay C, Khan S, Fan Y, et al. Dashboard design patterns. IEEE Trans Vis Comput Graph 2023;29(1):342-352. [doi: 10.1109/tvcg.2022.3209448]
31. Few S. Information Dashboard Design: The Effective Visual Communication of Data. Sebastopol, CA: O'Reilly Media; 2006.
32. Gleeson J, Kitchin R, McCarthy E. Dashboards and public health: the development, impacts, and lessons from the Irish government COVID-19 dashboards. Am J Public Health 2022 Jun;112(6):896-899. [doi: 10.2105/ajph.2022.306848]
33. Lechner B, Fruhling A. Towards public health dashboard design guidelines. In: Proceedings of the 1st International Conference, HCIB 2014, Held as Part of HCI International on HCI in Business. 2014 Presented at: HCIB '14; June 22-27,

2014; Crete, Greece p. 49-59 URL: https://link.springer.com/chapter/10.1007/978-3-319-07293-7_5 [doi: 10.1007/978-3-319-07293-7_5]

34. Amarasinghe I, Hernandez-Leo D, Michos K, Vujovic M. An actionable orchestration dashboard to enhance collaboration in the classroom. IEEE Trans Learn Technol 2020 Oct 1;13(4):662-675. [doi: 10.1109/tlt.2020.3028597]
35. Lock O, Bednarz T, Leao SZ, Pettit C. A review and reframing of participatory urban dashboards. City Cult Soc 2020 Mar;20:100294. [doi: 10.1016/j.ccs.2019.100294]
36. Setlur V, Correll M, Satyanarayan A, Tory M. Heuristics for supporting cooperative dashboard design. IEEE Trans Vis Comput Graph 2024 Jan;30(1):370-380. [doi: 10.1109/TVCG.2023.3327158] [Medline: 37883257]
37. Anderson JA, Demeter N, Pasquires M, Wirtz S. Using the CA opioid overdose surveillance dashboard to track opioid overdose deaths. Online J Public Health Inform 2019 May 30;11(1):e449 [FREE Full text] [doi: 10.5210/ojphi.v11i1.9938]
38. Safranek CW, Feitzinger L, Joyner AK, Woo N, Smith V, Souza ED, et al. Visualizing opioid-use variation in a pediatric perioperative dashboard. Appl Clin Inform 2022 Mar 23;13(2):370-379 [FREE Full text] [doi: 10.1055/s-0042-1744387] [Medline: 35322398]
39. Wu E, Villani J, Davis A, Fareed N, Harris DR, Huerta TR, et al. Community dashboards to support data-informed decision-making in the HEALing communities study. Drug Alcohol Depend 2020 Dec 01;217:108331 [FREE Full text] [doi: 10.1016/j.drugalcdep.2020.108331] [Medline: 33070058]
40. Abras C, Maloney-Krichmar D, Preece J. User-centered design. In: Ghaoui C, editor. Encyclopedia of Human Computer Interaction. Thousand Oaks, CA: Sage Publications; 2004:445-456.
41. Vredenburg K, Mao JY, Smith PW, Carey T. A survey of user-centered design practice. In: Proceedings of the 2002 SIGCHI Conference on Human Factors in Computing Systems. 2002 Presented at: CHI '02; April 20-25, 2002; Minneapolis, MN p. 471-478 URL: https://dl.acm.org/doi/10.1145/503376.503460
42. Muller MJ, Haslwanter JH, Dayton T. Participatory practices in the software lifecycle. In: Helander MG, Landauer TK, Prabhu PV, editors. Handbook of Human-Computer Interaction. Amsterdam, The Netherlands: North Holland; 1997:255-297.
43. Schuler D, Namioka A. Participatory Design: Principles and Practices. Boca Raton, FL: CRC Press; 1993.
44. Guest G, MacQueen KM, Namey EE. Applied Thematic Analysis. Thousand Oaks, CA: Sage Publications; 2011.
45. Fatal drug overdose touchpoints. Indiana Management Performance Hub. URL: https://www.in.gov/mph/projects/fortress/ [accessed 2024-01-30]
46. Bennett AS, Elliott L. Naloxone's role in the national opioid crisis-past struggles, current efforts, and future opportunities. Transl Res 2021 Aug;234:43-57 [FREE Full text] [doi: 10.1016/j.trsl.2021.03.001] [Medline: 33684591]
47. Snyder H, Kalmin MM, Moulin A, Campbell A, Goodman-Meza D, Padwa H, et al. Rapid adoption of low-threshold buprenorphine treatment at California emergency departments participating in the CA bridge program. Ann Emerg Med 2021 Dec;78(6):759-772 [FREE Full text] [doi: 10.1016/j.annemergmed.2021.05.024] [Medline: 34353655]
48. Binswanger IA, Nowels C, Corsi KF, Glanz J, Long J, Booth RE, et al. Return to drug use and overdose after release from prison: a qualitative study of risk and protective factors. Addict Sci Clin Pract 2012 Mar 15;7(1):3 [FREE Full text] [doi: 10.1186/1940-0640-7-3] [Medline: 22966409]
49. Cooper JA, Onyeka I, Cardwell C, Paterson E, Kirk R, O'Reilly D, et al. Record linkage studies of drug-related deaths among adults who were released from prison to the community: a scoping review. BMC Public Health 2023 May 05;23(1):826 [FREE Full text] [doi: 10.1186/s12889-023-15673-0] [Medline: 37147595]
50. Merrall EL, Kariminia A, Binswanger IA, Hobbs MS, Farrell M, Marsden J, et al. Meta-analysis of drug-related deaths soon after release from prison. Addiction 2010 Sep 05;105(9):1545-1554 [FREE Full text] [doi: 10.1111/j.1360-0443.2010.02990.x] [Medline: 20579009]
51. Grammel L, Tory M, Storey MA. How information visualization novices construct visualizations. IEEE Trans Vis Comput Graph 2010 Nov;16(6):943-952. [doi: 10.1109/TVCG.2010.164] [Medline: 20975131]
52. Ivanković D, Barbazza E, Bos V, Brito Fernandes Ó, Jamieson Gilmore K, Jansen T, et al. Features constituting actionable COVID-19 dashboards: descriptive assessment and expert appraisal of 158 public web-based COVID-19 dashboards. J Med Internet Res 2021 Feb 24;23(2):e25682 [FREE Full text] [doi: 10.2196/25682] [Medline: 33577467]
53. Kariisa M, Davis NL, Kumar S, Seth P, Mattson CL, Chowdhury F, et al. Vital signs: drug overdose deaths, by selected sociodemographic and social determinants of health characteristics - 25 states and the district of Columbia, 2019-2020. MMWR Morb Mortal Wkly Rep 2022 Jul 22;71(29):940-947 [FREE Full text] [doi: 10.15585/mmwr.mm7129e2] [Medline: 35862289]
54. Park JN, Rouhani S, Beletsky L, Vincent L, Saloner B, Sherman SG. Situating the continuum of overdose risk in the social determinants of health: a new conceptual framework. Milbank Q 2020 Sep 18;98(3):700-746 [FREE Full text] [doi: 10.1111/1468-0009.12470] [Medline: 32808709]
55. van Draanen J, Jamula R, Karamouzian M, Mitra S, Richardson L. Pathways connecting socioeconomic marginalization and overdose: a qualitative narrative synthesis. Int J Drug Policy 2023 Mar;113:103971. [doi: 10.1016/j.drugpo.2023.103971] [Medline: 36822011]
56. Agrawal R, Srikant R. Privacy-preserving data mining. In: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data. 2000 Presented at: SIGMOD '00; May 15-18, 2000; Dallas, TX p. 439-450 URL: https://dl.acm.org/doi/10.1145/335191.335438 [doi: 10.1145/342009.335438]

57. Bhattacharjee K, Chen M, Dasgupta A. Privacy-preserving data visualization: reflections on the state of the art and research opportunities. Comput Graph Forum 2020 Jul 18;39(3):675-692. [doi: 10.1111/cgf.14032]
58. Evans EA, Delorme E, Cyr K, Goldstein DM. A qualitative study of big data and the opioid epidemic: recommendations for data governance. BMC Med Ethics 2020 Oct 21;21(1):101 [FREE Full text] [doi: 10.1186/s12910-020-00544-9] [Medline: 33087123]
59. Evans EA, Delorme E, Cyr KD, Geissler KH. The Massachusetts public health data warehouse and the opioid epidemic: a qualitative study of perceived strengths and limitations for advancing research. Prev Med Rep 2022 Aug;28:101847 [FREE Full text] [doi: 10.1016/j.pmedr.2022.101847] [Medline: 35669857]
60. Waye KM, Yedinak JL, Koziol J, Marshall BD. Action-focused, plain language communication for overdose prevention: a qualitative analysis of Rhode Island's overdose surveillance and information dashboard. Int J Drug Policy 2018 Dec;62:86-93. [doi: 10.1016/j.drugpo.2018.08.010] [Medline: 30384027]
61. Fatal drug overdose touchpoints. Management Performance Hub, Indiana MPH. URL: https://www.in.gov/mph/projects/fortress/ [accessed 2024-04-29]

Abbreviations

ED: emergency department
EMS: emergency medical services
MPH: Management Performance Hub
OFR: Overdose Fatality Review
Rx: medical prescription

Edited by A Kushniruk; submitted 08.02.24; peer-reviewed by K Wheeler; comments to author 08.03.24; revised version received 24.03.24; accepted 13.05.24; published 11.06.24.

Please cite as:
Salvi A, Gillenwater LA, Cockrum BP, Wiehe SE, Christian K, Cayton J, Bailey T, Schwartz K, Dir AL, Ray B, Aalsma MC, Reda K
Development of a Real-Time Dashboard for Overdose Touchpoints: User-Centered Design Approach
JMIR Hum Factors 2024;11:e57239
URL: https://humanfactors.jmir.org/2024/1/e57239
doi:10.2196/57239
PMID:38861717

Co-Designing a Digital App to Support Young People's Patient and Public Involvement and Engagement (VoiceIn): Development and Usability Study

Alison Branitsky[1,2,3], BA, MRes; Penny Bee[4,5], GIBiol, BSc (Hons), PhD; Sandra Bucci[1,2,5], BSc (Hons), ClinPsyD; Karina Lovell[4,5], RN, BA (Hons), MSc, PhD; Simon Foster[6], BA; Pauline Whelan[5,6,7], BA Mod (Hons), MSc, PhD

[1]Division of Psychology and Mental Health, School of Health Sciences, Faculty of Biology Medicine and Health, Manchester Academic Health Science, The University of Manchester, Oxford Road, Manchester, United Kingdom

[2]Complex Trauma and Resilience Research Unit, Greater Manchester Mental Health NHS Foundation Trust, Manchester, United Kingdom

[3]Psychosis Research Unit, Greater Manchester Mental Health NHS Foundation Trust, Manchester, United Kingdom

[4]Division of Nursing, Midwifery and Social Work, School of Health Sciences, Faculty of Biology, Medicine and Health, Manchester Academic Health Science, The University of Manchester, Manchester, United Kingdom

[5]NIHR Applied Research Collaborative - Greater Manchester, Manchester, United Kingdom

[6]Division of Informatics, Imaging & Data Sciences, School of Health Sciences, Faculty of Biology, Medicine and Health, Manchester Academic Health Science, The University of Manchester, Oxford Road, Manchester, United Kingdom

[7]GM.Digital Research Unit, Greater Manchester Mental Health NHS Foundation Trust, Manchester, United Kingdom

Corresponding Author:
Pauline Whelan, BA Mod (Hons), MSc, PhD
NIHR Applied Research Collaborative - Greater Manchester, , Manchester, , United Kingdom

Abstract

Background: While patient and public involvement and engagement (PPIE) is now seen as a cornerstone of mental health research, young people's involvement in PPIE faces limitations. Work and school demands and more limited independence can make it challenging for young people to engage with PPIE. Lack of ability or desire to attend face-to-face meetings or group discussions can further compound this difficulty. The VoiceIn app and digital platform were codeveloped by a multidisciplinary team of young people, mental health researchers, and software designers, and enables young people to engage directly with PPIE opportunities via a mobile app.

Objective: This paper aims to describe how VoiceIn was developed through a series of co-design workshops with relevant stakeholders, specifically (1) how the initial design of VoiceIn was informed and driven by focus groups with young people, mental health professionals, and PPIE leads; (2) how VoiceIn was refined through collaboration with the aforementioned stakeholders; (3) the priorities for an app to support PPIE; (4) the key features necessary in the PPIE app; and (5) the recommended next steps in testing and deploying the digital platform.

Methods: Initial co-design workshops took place with young people, mental health professionals, and PPIE leads to identify key features of an app to support PPIE. A series of VoiceIn design prototypes were developed and iterated based on the priorities and preferences of the stakeholders. The MoSCoW (must have, should have, could have, won't have) prioritization method was used throughout the process to identify priorities across the different stakeholder groups.

Results: Co-design with young people, mental health professionals, and PPIE leads supported the successful development and improvement of the VoiceIn app. As a result of this process, key features were identified, including allowing for various modes of providing feedback (eg, polls and comments), reviewing project updates, and expressing interest in categories of research. The researcher platform was developed to support multimedia uploads for project descriptions; a jargon detector; a dedicated section for providing project updates; and a visually appealing, user-friendly design. While all stakeholder groups emphasized the importance of allowing app users to engage with the app in various ways and for there to be ongoing progress updates, group differences were also noticed. Young people expressed a desire for incentives and rewards for engaging with the app (eg, to post on their public social media profiles), and mental health professionals and PPIE leads prioritized flexibility in describing the project and its PPIE needs.

Conclusions: A co-design approach was pivotal to the development of the VoiceIn app. This collaborative approach enabled the app to meet the divergent needs of young people, mental health professionals, and PPIE leads. This process mirrored the aspirations of PPIE initiatives by cocreating a digital health research tool with key stakeholders.

(JMIR Hum Factors 2024;11:e53394) doi:10.2196/53394

KEYWORDS

patient and public involvement and engagement; PPIE; digital mental health; young people; co-design; mental health

Introduction

Patient and public involvement and engagement (PPIE) has become a cornerstone of best practice in mental health research. PPIE is premised on the notion that research should be carried out "by" or "with" those who the research is intended to benefit, rather than "to" or "for" them [1]. At its core, PPIE values the primacy of subjective lived experience in knowledge construction and emphasizes a way of producing science where experts by experience are active co-designers and co-researchers through the entire research process [2,3]. PPIE provides valuable opportunities for addressing the democratic deficit and power imbalance that exist in much of health research by giving an equal voice to the intended beneficiaries of such research [4].

PPIE has numerous demonstrable benefits, including increasing the impact of research by homing in on the questions most relevant to stakeholders, designing more appropriate and targeted methodologies, addressing ethical tensions that exist between various stakeholder groups, and exploring novel ideas that may not have been generated by researchers alone [5-10]. By rooting findings and reports in user experience, outcomes are likely to be more relevant, relatable, and understandable to the public, thus enhancing both dissemination and implementation [6,7].

Young people can be valuable and engaged research partners. However, recent scoping research has demonstrated that young people's expertise may be underused. A 2018 United Kingdom–based survey of public contributors to the National Institute for Health and Social Care Research found that only 2% of those involved were younger than 25 years [11]. Similarly, Rouncefield-Swales et al [12] found that where young people were involved in PPIE, their level of involvement was varied, the impact of their involvement was often unknown, and details of precisely how PPIE contributions were integrated into projects were lacking in written reports. However, in regions where a specific effort was made to include young people in PPIE, their involvement surpassed those of any other age group [13]. Given that the principle of young people's involvement is clearly enshrined in international policy [14,15], there is a strong impetus to further develop ways in which young people can meaningfully engage with all stages of health research.

Involving young people in traditional, face-to-face PPIE can prove challenging. School and work commitments, limitations on independence and flexibility, resource and financial constraints, and rapid developmental and lifestyle changes can make it difficult for young people to engage with PPIE long-term [12,16,17]. Furthermore, meeting face-to-face with various stakeholders is not a suitable environment for all young people to feel comfortable and empowered to contribute. Consequently, researchers often struggle to recruit [18] and retain [19] young people to be involved in PPIE. Research projects may be further limited by recruiting young people from small geographic areas, thereby limiting the generalizability of research input and findings [20]. As a result, there is a clear need to re-evaluate the current adult-centric modes of PPIE participation and spearhead innovative means of youth participation [21,22].

Shimmin et al [23] argue that many current PPIE practices are limited in their recognition of the real complexities of people's lives. As a result, many individuals who carry the greatest burden of illness, particularly those with marginalized or excluded social identities and those experiencing various forms of systemic oppression, are less likely to have their voices heard in the sphere of PPIE. When looking at PPIE through the lens of health equity, there is a strong impetus to maximize diversity both in terms of social identity and in modes of meaningful participation.

Digital health technology provides a promising avenue for complementing existing PPIE methods by expanding participation opportunities to an audience that may not otherwise be reached. To our knowledge, there is currently no dedicated digital platform to provide PPIE opportunities to young people in real-time. To facilitate young people's involvement in mental health research and to make PPIE opportunities more accessible and inclusive, we built the "VoiceIn" PPIE digital platform, which we introduce here. VoiceIn acts as a 2-way PPIE toolkit whereby researchers can solicit PPIE input from young people, and young people provide feedback in their own time via a mobile app.

VoiceIn is a digital platform aimed at enabling young people to contribute easily and quickly to PPIE activities. It particularly aims to support young people who may ordinarily be unable or unwilling to attend traditional, face-to-face PPIE groups or activities. VoiceIn has been specifically designed to be user-friendly, fun, and easy to fit into young people's lifestyles. The app allows young people to give feedback on projects that align with their interests or lived experiences. Researchers can ask for specific input from PPIE participants on various aspects of the project, from the initial shaping of research topics and design to methodological feedback. Participants can provide feedback in the form of polls or free text. VoiceIn emphasizes ongoing collaboration between participants and researchers: researchers provide frequent updates to highlight the impact of PPIE on project development, and participants have the ability to track their contributions across various projects.

VoiceIn was co-designed with young people, mental health researchers, and leaders of community-based PPIE groups. The purpose of this paper is to detail (1) how the initial design of VoiceIn was informed and driven by focus groups with young people, mental health professionals, and PPIE leads; (2) how VoiceIn was refined through collaboration with the aforementioned stakeholders; (3) the central priorities for an app to support PPIE; (4) the key features necessary in the PPIE app; and (5) the recommended next steps in testing the digital platform.

Methods

Study Setting, Participants, and Recruitment

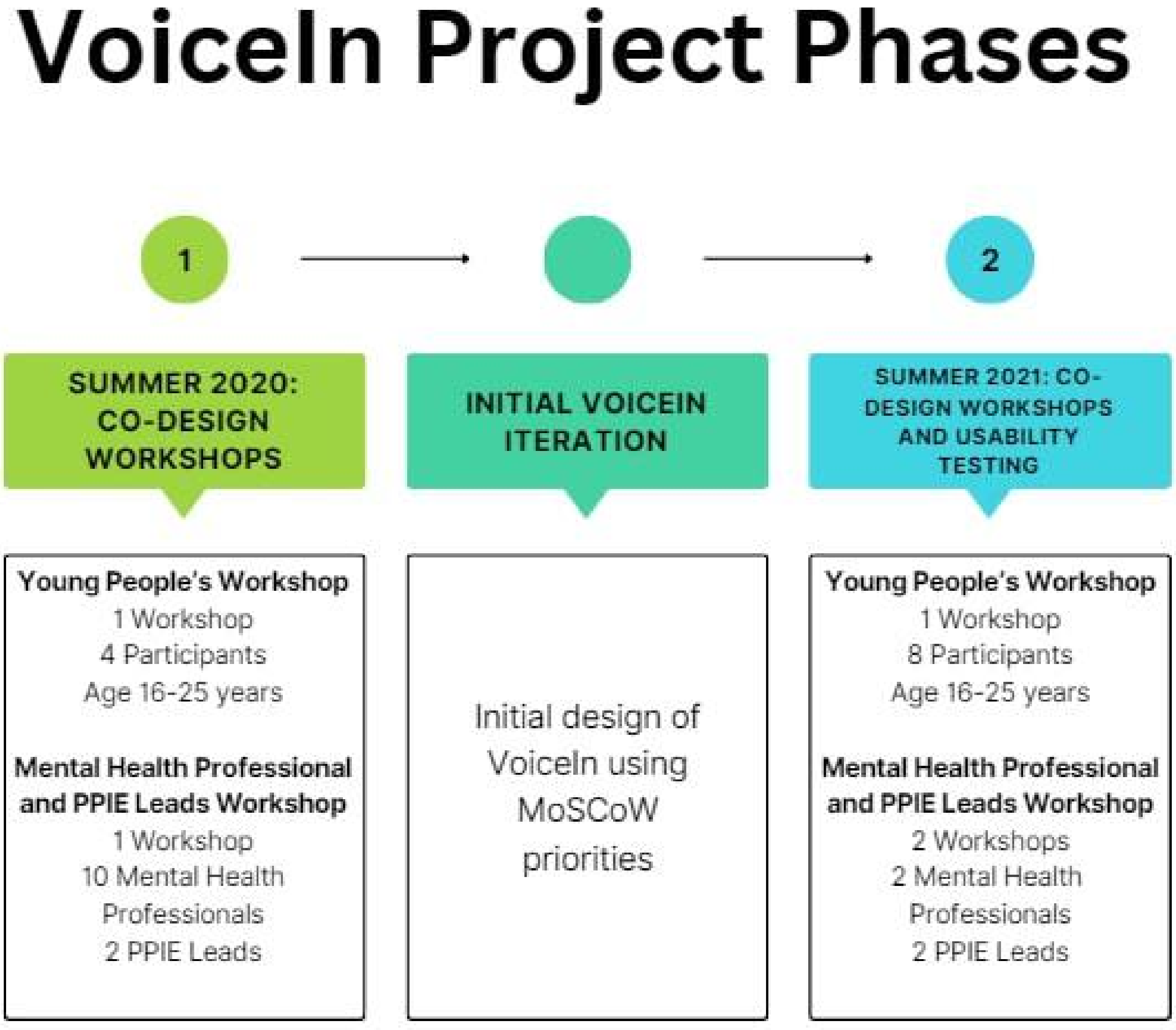
A convenience sampling approach was adopted for the phase 1 workshops, with participants recruited from established local young people's advisory groups and through team networks. The phase 1 co-design workshops were run as exploratory patient and public involvement activities with locally established young people's advisory groups, mental health staff, and PPIE leads. Convenience sampling was adopted at this stage to assess initial interest in the idea of the VoiceIn app and to determine high-level requirements. For the phase 2 co-design workshops, a purposive sampling approach was adopted. For phase 2, recruitment targeted individuals outside of immediate research team networks and aimed to reach people who were not previously involved in the project. The inclusion criterion for young people's groups was a requirement to be older than 16 years with the capacity to consent. The inclusion criterion for mental health professionals was a requirement to be adult staff working in mental health services for children or young people. For PPIE leads, the inclusion criterion was to have a role in running and coordinating PPIE activities for young people in a public sector or voluntary organization. Input from young people was seen as key to the development of a usable and engaging app, and for phase 2, we sought to increase the number of young people involved in the workshops (from 4 to 8). Around 8 people were deemed the maximum number we could support through a web-based workshop and sufficient for capturing ideas to progress the development of the app. Mental health professionals and PPIE leads for phase 2 were approached through national networks for mental health and PPIE. National organizations and third-sector or charitable organizations were approached by email at this stage.

Co-Design Workshops

Overview

Between 2020 and 2021, a series of co-design workshops were conducted with young people, mental health professionals, and PPIE leads. An overview of the timeline of the workshops and their participants is outlined in Figure 1.

Figure 1. Co-design workshops timeline. MoSCoW: must have, should have, could have, won't have; PPIE: patient and public involvement and engagement.



Workshops were held on Zoom and lasted 90 minutes. All workshops followed a similar structure: participants and the research team introduced themselves, participants were provided with an overview of the VoiceIn project, and the rest of the meeting was dedicated to exploring specific topics related to app co-design. The specific workshop structure is provided in Figures 2 and 3. For larger group sizes, participants were assigned to breakout rooms with 2 to 3 people and 1 member of the research team. For smaller groups, a single group was maintained throughout. Topics of discussion were decided a priori and are described in Figure 4. Time was also allocated to allow for discussions of topics that arose organically during the meeting. The software collaboration tool Mural [24] was used in the young people's workshop to enable participants to share their thoughts on a collaborative whiteboard.

Figure 2. Structure of phase 1 co-design workshops. MoSCoW: must have, should have, could have, won't have.

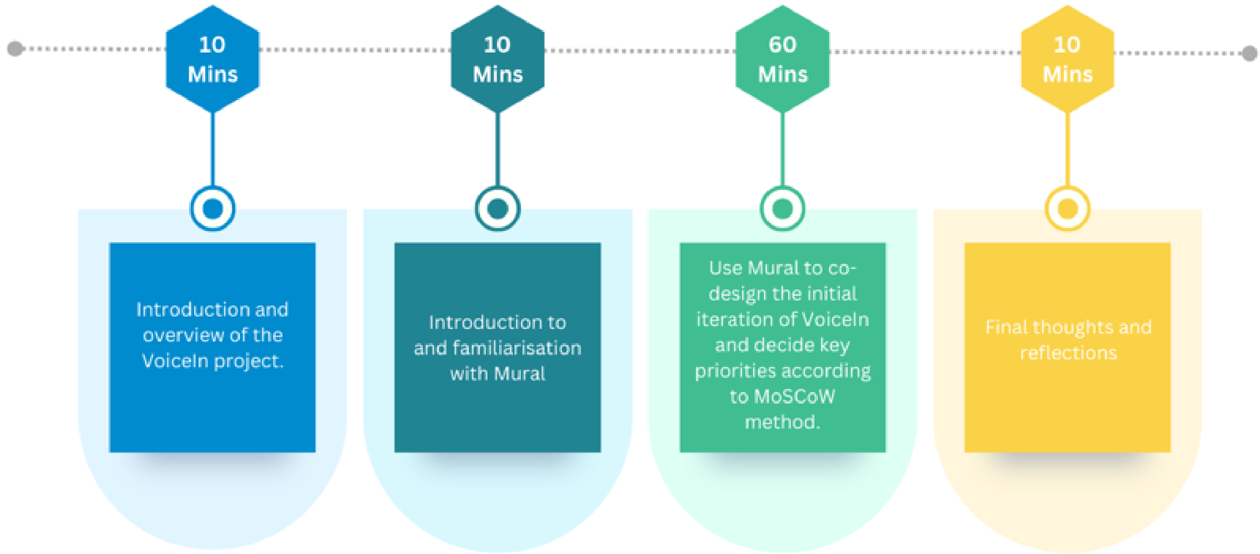


Figure 3. Structure of phase 2 co-design workshops. MoSCoW: must have, should have, could have, won't have

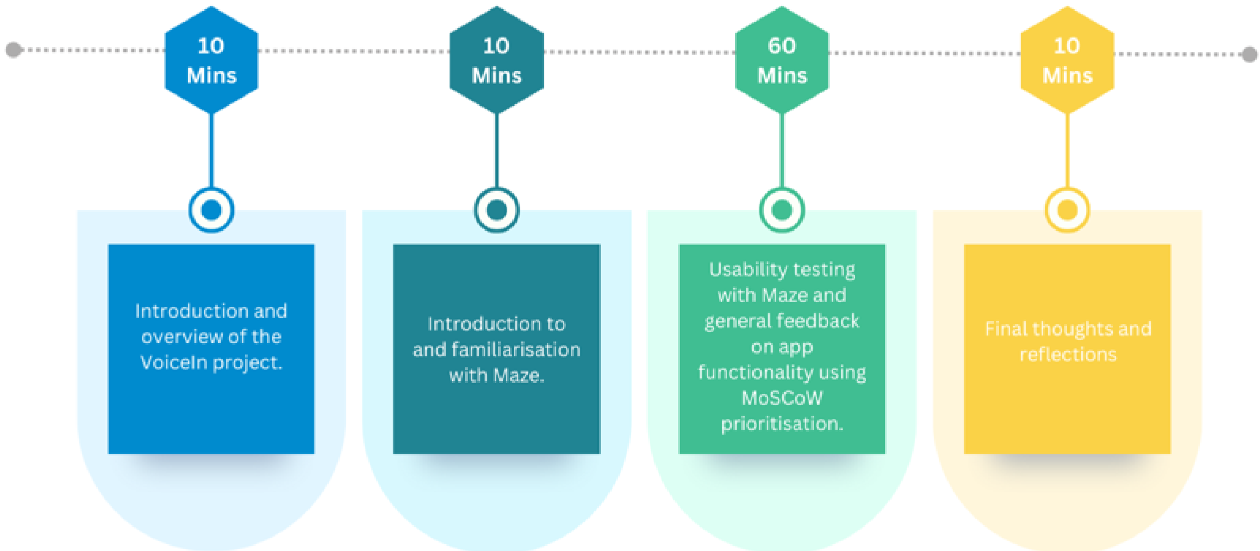
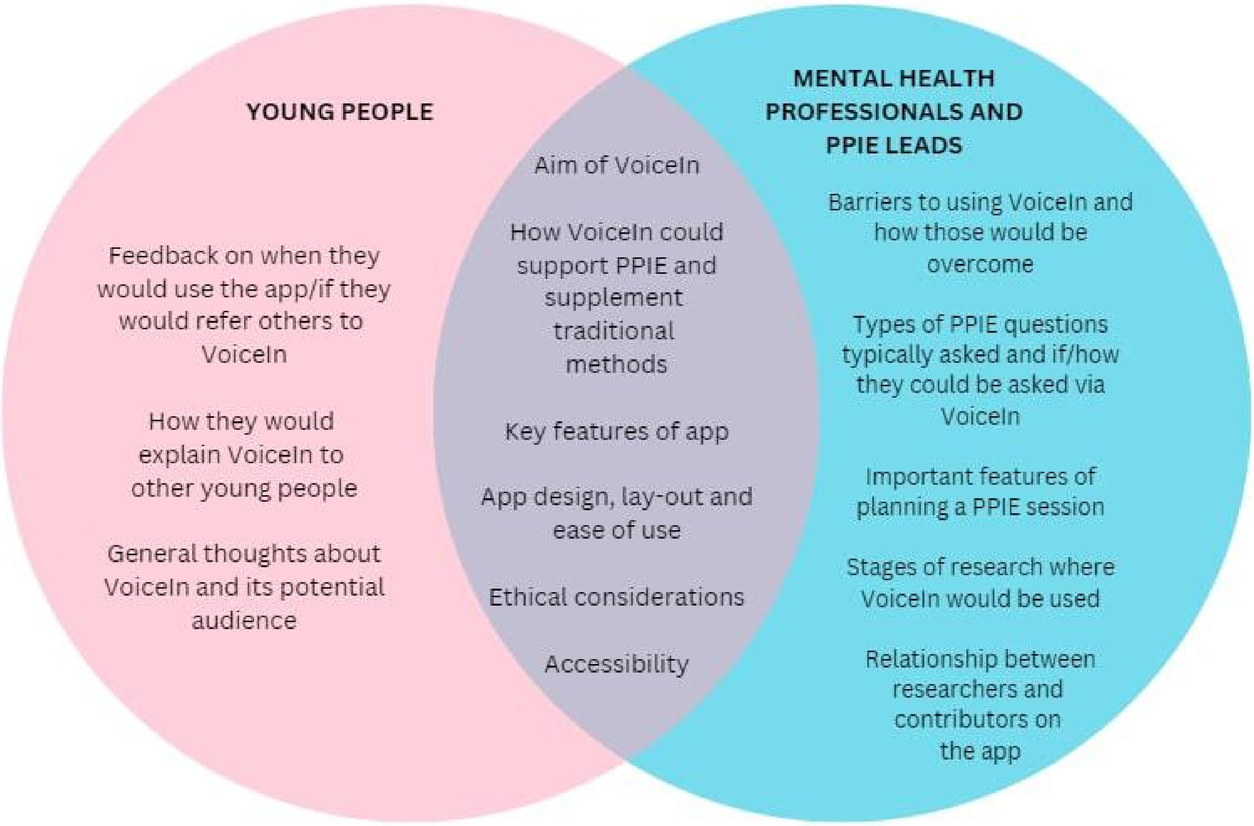


Figure 4. Topics discussed in co-design workshops. PPIE: patient and public involvement and engagement.



MoSCoW Prioritization

To enable the identification of priorities for the app, the MoSCoW (must have, should have, could have, won't have) prioritization method was used [25]. This method is an oft-used method in software project management to determine the most crucial deliverables across groups of stakeholders. In the MoSCoW approach, app features were ranked by each stakeholder group as "must have," "should have," "could have," and "won't have." "Must have" items were considered nonnegotiable by stakeholders; "should have" items were those considered to have significant value but would not affect the key functionality of the app; "could have" items were additional desirable features that were not necessarily related to the core goals of VoiceIn, while items classified as "won't have" were not priorities for the current VoiceIn release but may be considered for implementation in the future. Focus group facilitators supported discussion when different priorities were identified, aiming to arrive at a consensus decision. Where this was not possible, the majority of contributed opinions were used to determine the priority level assigned to each feature.

Agile Software Development

The software team adopted an Agile approach to software development, building the software iteratively and refining the platform based on feedback at the workshops. Following each co-design workshop, the software engineering team discussed the technical feasibility of identified requirements; some items were put on hold for implementation due to the complexity of the implementation involved and the time or budget constraints of the project. Software engineers then worked to incorporate the features ranked "must have" and "should have" in the MoSCoW process into the next iteration of the app for presentation at the subsequent workshop.

App Technical Development

The VoiceIn app and digital platform were developed by the specialist Digital Health Software and Platforms team based at the University of Manchester [26]. The software team has expertise in developing digital health research technology. An Agile software development approach was adopted with the software team delivering the digital platform iteratively across a number of Agile sprint cycles. A hybrid app framework (Capacitor by Ionic) was used for ease of deployment to multiple devices and to align with expertise within the software team [27]. The web back end was supported by an open-source relational database (MySQL) [28] with a lightweight open-source framework Vue.js used for the front end [29].

Usability Tasks

In the phase 2 co-design workshops, participants were asked to complete a usability testing exercise on a platform called Maze [30]. Maze is a commercial product development platform that enables usability and prototype testing of new products, including web and mobile apps. Participants used their own devices to access Maze; Maze was accessible with all Android and iOS devices. In the young people's workshops, participants were asked to provide project feedback on a prototype of the app. In the mental health professionals and PPIE leads' workshops, participants were asked to create a mock project. After completing their respective task, participants were asked to rate the difficulty of the assignment and provide feedback about what, if anything, they found difficult and if anything was missing.

Ethical Considerations

Informed consent was obtained from all workshop participants. Only participants older than 16 years with the capacity to consent were included in the workshops. Ethical approval for the phase 2 co-design workshops was granted by the University of Manchester Research Ethics Committee (REF 2021-10864-17849). Ground rules were established by the workshop facilitators at the start of each workshop to clarify the workshop's purpose and guide the conduct of the workshop.

Results

Phase 1 Co-Design Workshops: Young People

Preliminary co-design workshops with young people demonstrated a diverse array of preferences for engaging with the VoiceIn app. There was a general consensus that young people preferred to be involved throughout the entire research process, stating that PPIE should be the "driving force" behind research projects. Some participants felt it would be particularly interesting and rewarding to be involved in the design of research projects and identification of key research questions and priorities; others preferred more discrete and concrete tasks such as providing specific feedback on topic guides, picking questionnaires, and coproducing the terms of reference for ongoing PPIE support during projects. Participants highlighted the need for the app to fit into the day-to-day lives of young people, noting that quick responses such as polls and surveys should be included since some young people will not have the time or ability to review lengthy documents and give detailed feedback. Young people also suggested including rewards and incentives to encourage engagement with the app, considering ways for young people to engage with one another within the app, and including sections such as "meet the researcher" to make projects more contextualized and relational.

Young people identified a desire and need for training on research processes generally when long-term involvement PPIE involvement was necessary. Although this feedback fell outside the scope of the app (which was primarily intended to capture "quick" feedback on early-stage research ideas or short responses to specific questions during the lifetime of a project), it was noted that a training module within the app describing the research life cycle and processes would be a useful addition to consider for future implementation. Additional training material (eg, a research handbook) has been developed as a result of these conversations, but likewise falls beyond the scope of what could be included in the app.

Phase 1 Co-Design Workshops: Mental Health Researchers and PPIE Leads

Similar to the young people, mental health researchers and PPIE leads aimed to seek ongoing PPIE input throughout the research process. In particular, they predicted using the app to explore possible areas of research and define research questions, seek comments on recruitment methods and data collection tools, and disseminate results through the co-development of communication plans. Researchers noted, however, that they would not use the app to seek PPIE input for determining specific outcome measures or data analysis, due to the specialist knowledge required for both tasks.

MoSCoW Prioritization and Initial VoiceIn Iteration

Overview

During both the phase 1 and phase 2 co-design workshops, participants were asked to rate different app features according to the MoSCoW prioritization method in order to determine their priority. Every effort was made to ensure that items ranked as "must have" and "should have" were included in the initial iteration of VoiceIn. Some items were deemed outside of scope for this version (due to time or budget limitations) but were noted for future iterations of the platform.

General Capabilities

A summary of stakeholders' priorities regarding the general capabilities of VoiceIn is described in Table 1. Across stakeholder groups, providing project progress updates was identified as a clear priority; as such, VoiceIn was built to prompt mental health researchers to provide relevant updates, and a dedicated "project updates" section was built into the mobile app. Stakeholders likewise noted both global (eg, technical support section, terms of reference) and project-specific (eg, having the ability to decide on research questions and topics) as key priorities.

Table . MoSCoW[a] general capabilities priorities for user-facing mobile app interface capabilities.

Description of general capabilities	Group				
	Young people	Mental health researchers	PPIE[b] leads	Included?	Reason for exclusion
Input to decide on research topics or questions	Must Have	Must Have	Should Have	Yes	—[c]
Information on the role of PPIE in shaping the research	Must Have	Must Have	—	Yes	—
Project progress updates	Must Have	Must Have	Must Have	Yes	—
Technical support section	Must Have	Must Have	—	Yes	—
Notifications for new research opportunities	Must Have	—	—	Yes	—
Ability to collect research data within the app	Must Have	—	Should Have	No	As the app is specifically to aid with PPIE, rather than carry out research procedures, this falls beyond the scope of the app
Contributor leader board to demonstrate users who have provided the most contributions	Must Have	Could Have	—	No	Leader board ultimately excluded to demonstrate that all contributions are valuable, regardless of the number of contributions per user
Credit or time-based rewards for contributions	Must Have	Must Have	—	No	Complexity of implementation and variety of reward types discussed precluded the provision of rewards in the initial version. Further discussion is required to scope the requirements, planned for the next phase
Recognition for user inputs	—	Should Have	—	No	PPIE contributions recognized on a project-by-project basis and therefore fall beyond the scope of the app
Terms of reference or expectations	—	Must Have	Must Have	Yes	—
Data access or GDPR[d] information	—	Must Have	—	Yes	—
Users to assist with publicizing study or helping with recruitment	Should Have	Must Have	Should Have	No	Formal recruitment activities are outside the current PPIE scope of VoiceIn but project information and contact details can be provided
Glossary of research terminology	—	Must Have	—	No	Resource limitations
Notice that content may cause distress	—	Must Have	—	Yes	—

Description of general capabilities	Group			Included?	Reason for exclusion
	Young people	Mental health researchers	PPIE[b] leads		
Project categories are lay-friendly and not diagnosis-based	—	Must Have	—	Yes	—
Notifications for researcher to put project updates on the app	—	Must Have	—	Yes	Researchers are reminded on login to the website
Payment policies	—	—	Must Have	No	This is actively under consideration of how to recognize contributions fairly and safely within the app
Contributions exportable to LinkedIn or CV[e]	Should Have	—	Must Have	No	Technical constraints limit the ability to create exportable contributions at this time. But under consideration for future
Reminders for research milestones and dates for feedback to be returned	Should Have	—	—	Yes	—
Ideas lab to generate novel research ideas	—	Should Have	—	No	Resource limitations, but under consideration for future iteration
Demonstrate impact of research projects	—	Should Have	Should Have	Yes	—
Create a database of people interested in research	—	—	Should Have	No	Data protection limitations prohibit the creation of this type of database
Hear about new funding and related work	Could Have	—	—	Yes	—
Certificates of involvement	Could Have	Could Have	—	No	Technical constraints prevent the creation of certificates within the app at this time
Questions to assess users' mood and offer support	—	Could Have	—	No	A support section is included in the app, but ongoing monitoring of users' moods is beyond the scope of the app and may not be acceptable for some users
Debrief area	—	Could Have	—	Yes	—
Area to share lessons learned	—	—	Could Have	No	Social features between researchers are not currently enabled

Description of general capabilities	Group				
	Young people	Mental health researchers	PPIE[b] leads	Included?	Reason for exclusion
Translation into other languages	—	—	Could Have	No	Current limitations on project resources prevent translating the app into other languages

[a]MoSCoW: must have, should have, could have, won't have.

[b]PPIE: patient and public involvement and engagement.

[c]Not applicable.

[d]GDPR: General Data Protection Regulation.

[e]CV: curriculum vitae.

Project Details

There was no clear consensus across stakeholder groups on the priorities pertaining to study details; however, sharing information about funding, including the project dates, and having the ability to upload documents were all identified as priorities among mental health researchers. Project detail priorities across stakeholder groups are outlined in Table 2.

Table . MoSCoW[a] project details priorities for user-facing mobile app interface capabilities.

Description of project details	Young people	Mental health researchers	PPIE[b] leads	Included?	Reason for exclusion
Information about funding	Could Have	Must Have	—[c]	Yes	—
Study date ranges as part of the project details	—	Must Have	—	Yes	—
Ability to upload documents	—	Must Have	Should Have	Yes	—
Addition of images/video/graphics to project description	—	Should Have	—	Yes	—
Researcher profile	—	—	Should Have	Yes	—
Flexible options for entering project details	—	Could Have	—	Yes	—
Addition of inclusion or exclusion criteria	—	Won't Have	—	No	Inclusion or exclusion criteria can be included as free text; there will not be a separate section to input these criteria. Participant recruitment is not the goal of the app

[a]MoSCoW: must have, should have, could have, won't have.

[b]PPIE: patient and public involvement and engagement.

[c]Not applicable.

Project Feedback

Stakeholders agreed that it was vital for app users to provide quick feedback via polls and likewise indicated that having the ability to give feedback on both participant and staff-facing material was a key priority. Project details priorities across stakeholders are outlined in Table 3.

Table . MoSCoW[a] project feedback priorities for user-facing mobile app interface capabilities.

Description of project feedback	Young people	Mental health researchers	PPIE[b] leads	Included?	Reason for exclusion
Ability to provide feedback via polls	Must Have	Must Have	—[c]	Yes	—
Sharing project materials (eg, PIS[d], consent forms)	Must Have	—	—	Yes	This is possible but sharing lengthy documents via the app is discouraged as this is not the goal of the app
Providing feedback on project materials (eg, PIS, consent forms)	Must Have	—	Should Have	Yes	—
Method to select data collection tools	Should Have	—	Should Have	Yes	—
Area to review advertisements for staff, steering groups, etc	Should Have	—	Should Have	Yes	—
Area to analyze and interpret data	Should Have	—	—	Yes	—
Help provide feedback on qualitative interview questions	Should Have	—	—	Yes	—
Ability for users to pause and come back later to contribute	—	Should Have	—	—	—
Help decide who may be interested in hearing about the results	Could Have	—	—	Yes	—
Ways for contributors to share ideas (eg, recruitment, dissemination)	—	Could Have	—	Yes	—
Reviewing previous research	Won't Have	—	—	No	Not desired by young people

[a]MoSCoW: must have, should have, could have, won't have.

[b]PPIE: patient and public involvement and engagement.

[c]Not applicable.

[d]PIS: participant information sheet.

Training and External Links

Participants across stakeholder groups indicated a desire for the app to have the capability to both provide project-specific training and provide links to relevant resources and opportunities beyond the app. Resource limitations preclude maintaining a comprehensive list of external training or providing internal training via the app. However, links to additional opportunities and resources can be provided on a project-specific level via the project updates page. Training and external links priorities across stakeholder groups are outlined in Table 4.

Table . MoSCoW[a] training and external links priorities for user-facing mobile app interface capabilities.

Description of training and external links	Young people	Mental health researchers	PPIE[b] leads	Included?	Reason for exclusion
Project-specific training	Must Have	—[c]	—	No	However, contextual background project information can be provided
Tailored training for participants and researchers	Should Have	—	Should Have	No	—
Hear about opportunities for future training or learning	Should Have	—	—	Yes	Can be supported per project updates
Job opportunities and information about other ways to be involved in research	Should Have	—	Could Have	Yes	Can be supported per project updates
Advertise internal or external development opportunities	Could Have	—	—	Yes	Can be supported per project updates
Training on research methodology	Won't Have	—	—	No	—
Information on meeting with peers and community groups	—	—	Should Have	No	—
Links to wider PPIE networks and opportunities	—	Could Have	Could Have	No	Planned for future iteration
Expand app to other audiences or ages	—	—	Must Have	Yes	—
Stakeholder consensus meeting	Won't Have	—	Must Have	No	—

[a]MoSCoW: must have, should have, could have, won't have.

[b]PPIE: patient and public involvement and engagement.

[c]Not applicable.

Social Features

The various stakeholder groups identified different priorities for social features within VoiceIn, with the "ability for users to connect with one another, the public and other researchers" being rated as the top priority. While resource limitations preclude the ongoing monitoring necessary to enable social features, mental health researchers are encouraged to include information about the research team on the project details page, and provide ongoing updates to foster engagement with app users. Social features priorities across stakeholder groups are outlined in Table 5.

Table . MoSCoW[a] social features priorities for user-facing mobile app interface capabilities.

Description		Young people	Mental health researchers	PPIE[b] leads	Included?	Reason for exclusion
Social Features						Resource limitations preclude the ongoing monitoring necessary to enable social features
	Ability for users to connect with one another, the public and other researchers	—[c]	Must Have	—	No	—
	Private messages	—	Should Have	—	No	—
	Area for users to share experiences, blogs, stories, etc	—	Should Have	—	No	—
	Ability to have conversations between researchers and contributors	—	Should Have	—	No	—
	Ability to share outside the platform	—	Should Have	—	No	—
	Linkable to social media outcomes	—	—	Should Have	No	—
	Area to share findings with the public	—	—	Should Have	No	—
	Monitored chatroom for personal interaction	Could Have	—	—	No	—
	Areas for more creative outputs	—	Could Have	—	No	—
	Input to write blog articles	—	—	Could Have	No	—
	Ability to connect with people with similar interests	—	—	Could Have	No	—

[a]MoSCoW: must have, should have, could have, won’t have.

[b]PPIE: patient and public involvement and engagement.

[c]Not applicable.

Design Features

Participants emphasized that VoiceIn should have a visually appealing and easy-to-navigate design. Specific design feature priorities across stakeholder groups are outlined in Table 6.

Table . MoSCoW[a] design features priorities for user-facing mobile app interface capabilities.

Description of design features	Young people	Mental health researchers	PPIE[b] leads	Included?	Reason for exclusion
Customizable notifications	Must Have	—[c]	—	No	Considered for future iterations
Easy to navigate back to the home screen	—	Must Have	—	Yes	—
Academic diary	Should Have	Should Have	—	No	Falls beyond the scope of the app
Video guide on how to use app	—	—	Could Have	No	Resource limitations preclude the creation of a video guide

[a]MoSCoW: must have, should have, could have, won't have.

[b]PPIE: patient and public involvement and engagement.

[c]Not applicable.

Mental Health Researchers' Priorities for App Functionality

Mental health researchers were also asked to rate their priorities for VoiceIn app functionality. The priorities are outlined in Table 7.

Table . MoSCoW[a] priorities for app functionality.

Feature	Mental health researchers	Included in the app?
Users can sign up on a mobile app	Must Have	Yes
Users have a password	Must Have	Yes
User selects interests from pre-populated list	Must Have	Yes
Users can view research projects added by research on the web interface in the app	Must Have	Yes
Easy to navigate back to the home screen	Must Have	Yes
Users can “like” or “dislike” project cards to express interest	Must Have	Yes
Users can “like” a project to see more details	Must Have	Yes
Users who have “liked” a project and have chosen to contribute can respond to researcher polls & comments	Must Have	Yes
Users can view projects they have fed back on	Must have	Yes
Users contribute using both free text and polls	Must Have	Yes
Users can sign up using a handle of their choice	Should Have	No
User confirms birth year to verify age group	Should Have	Yes
Users can view progress on projects they have fed back on	Should Have	Yes
Users that have “liked” a project and have chosen to contribute can choose if they want to receive updates on the project.	Should Have	Yes
Users have an area where they can change preferences (topic interests)	Should Have	Yes
Users can choose to contribute anonymously (feedback does not have a user ID visible, but feedback can be seen by other users)	Could Have	Yes
Users can choose to contribute publicly (feedback has user ID visible)	Could Have	Not Yet
Users can change privacy preferences for visibility	Could Have	Not Yet
Users that have “liked” a project and chosen to contribute can see other responses to researcher polls	Could Have	Not Yet
Users that have “liked” a project can save it for later	Could Have	Yes
Users that have “disliked” a project have it removed from their card stack and do not receive notifications about this project again	Could Have	Yes
Users that have “liked” a project then can choose not to contribute to the project; project will be removed from their card stack and they will not receive notifications	Could Have	Yes
Sign up requires identifiable information	Won’t Have	No
Users can choose to contribute directly to the researcher (only researcher can see feedback; may or may not have user ID)	Won’t Have	No

[a]MoSCoW: must have, should have, could have, won’t have.

Phase 2 Co-Design Workshops

The phase 2 workshops focused on asking participants to try out aspects of the tool and discuss the refined design prototypes. The prototypes showed how the mobile app and web-based researcher platform would look and function. In the workshops, participants were offered the opportunity to create a mock VoiceIn project, provide feedback on the ease of use of the VoiceIn platform, and identify areas of improvement.

Results of written feedback on the ease of use and areas of improvement from the usability testing workshops are presented in Table 8.

Table . Summary of usability feedback across stakeholder groups.

Feedback category	Young people	Mental health researchers	Changes incorporated
Ease of use	Easy to understand	First cohort rated the website as 5/5 in terms of ease of use	—[a]
Ease of use	Easy to engage with	Second cohort rated the website as 4.9/5 in terms of ease of use	—
Features that were good about the app or website	Use of videos in project descriptions	Clean and tidy design	—
Features that were good about the app or website	Choice about whether to contribute to each project	Straightforward to use	—
Features that were good about the app or website	"Quick and fun" to sort through project cards	Guides users through the process of setting up PPIE[b] events	—
Features that were good about the app or website	Project updates section	—	—
Features that were good about the app or website	Ability to save projects for later	—	—
Features that were missing from the app or website or could be improved	Include description along with title on the project cards	Funder details should be included	Description incorporated
Features that were missing from the app or website or could be improved	—	Project start or end dates	Project start or end dates added to project details page
Features that were missing from the app or website or could be improved	—	Inclusion and exclusion criteria as separate boxes	Can be defined on per project basis by the researcher
Features that were missing from the app or website or could be improved	—	Distress and debrief section	Support page added signposting users to mental health support
Features that were missing from the app or website or could be improved	—	Use service-user defined categories like "self-harm" and "low mood" instead of diagnostic labels like "depression" to categorize studies	Categories expanded to include nondiagnostic categories like "hearing voices" and "loneliness"
Features that were missing from the app or website or could be improved	—	Add "potential impact of project" section	Researchers have option to include this as part of the project description
Features that were missing from the app or website or could be improved	—	Section to outline expectation about time commitment and necessary skills	—
Features that were missing from the app or website or could be improved	—	More nuanced age range for study participants	—
Features that were missing from the app or website or could be improved	—	"Home" button that returns users to home screen	"Home" button added
Features that were missing from the app or website or could be improved	—	More colorful design	—
Additional suggestions	—	Flexible project details page so it is fit-for-purpose for each project	Flexible project description section which enables the use of additional headings, photos, and videos

[a]Not applicable.

[b]PPIE: patient and public involvement and engagement.

In the phase 2 workshops, participants were also invited to give general feedback about the app and its potential use. A summary of this discussion and the changes incorporated in subsequent iterations of the app are described in Figures 5-7.

Figure 5. Summary of young people's feedback on VoiceIn and its potential use. PPIE: patient and public involvement and engagement.

Figure 6. Summary of mental health researchers' feedback on VoiceIn and its potential use. PPIE: patient and public involvement and engagement.

XSL•FO
RenderX

Figure 7. Summary of PPIE leads' feedback about VoiceIn and its potential use. CV: curriculum vitae; PPIE: patient and public involvement and engagement.

Final VoiceIn Iteration

VoiceIn Mobile App

The final iteration of the VoiceIn mobile app allows users to choose their research interests from a range of categories (eg, mental health and public health). Users are then brought to the home screen, in which relevant projects appear and users choose whether they are interested in finding out more about the project.

When a user indicates they are interested in a project, they are shown the project description and asked if they would like to provide feedback on the project. Users then have the ability to answer questions (both poll and free text) and view project updates. In instances where feedback was provided via a poll, users will have the opportunity to see the distribution of responses. Interests and privacy settings can be changed by the user at any point. Sample images of the VoiceIn app are displayed in Figures 8-11.

Figure 8. VoiceIn home screen where participants can view all projects.

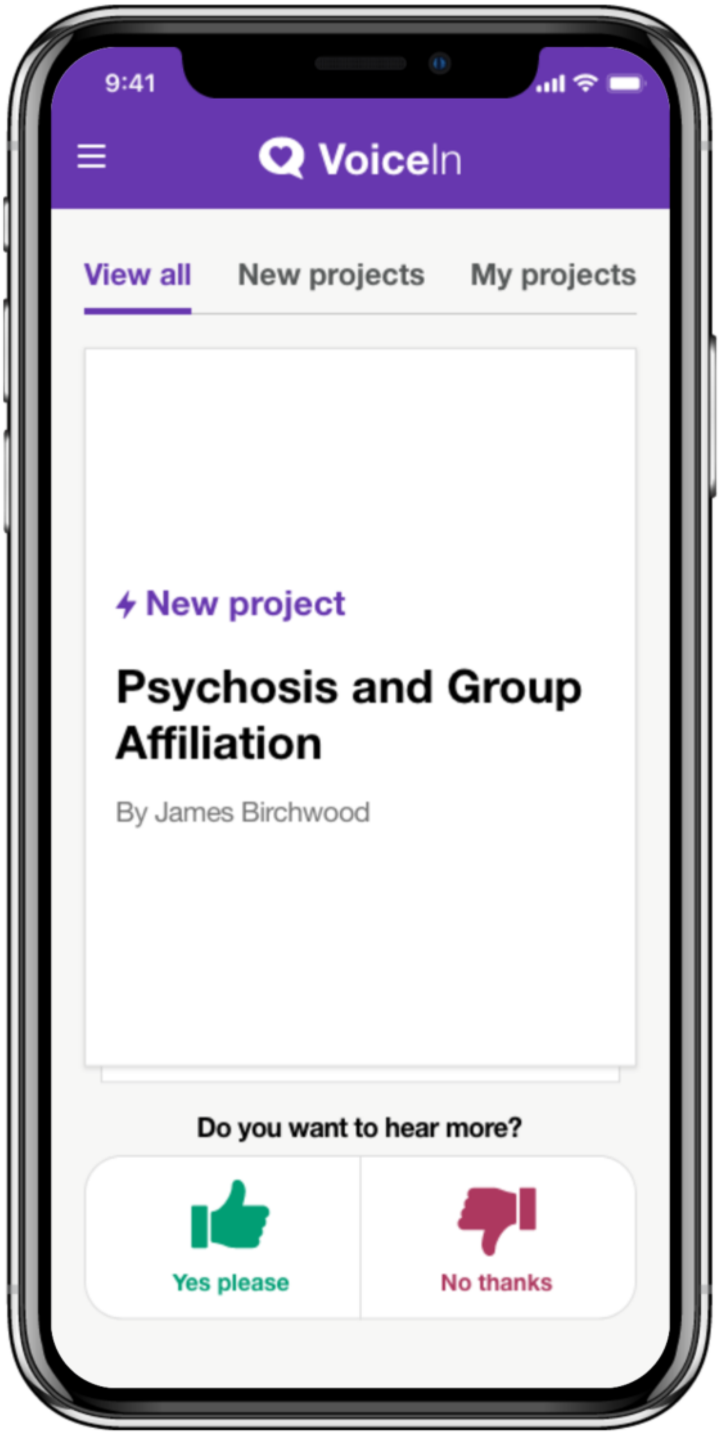


Figure 9. “My projects” page.

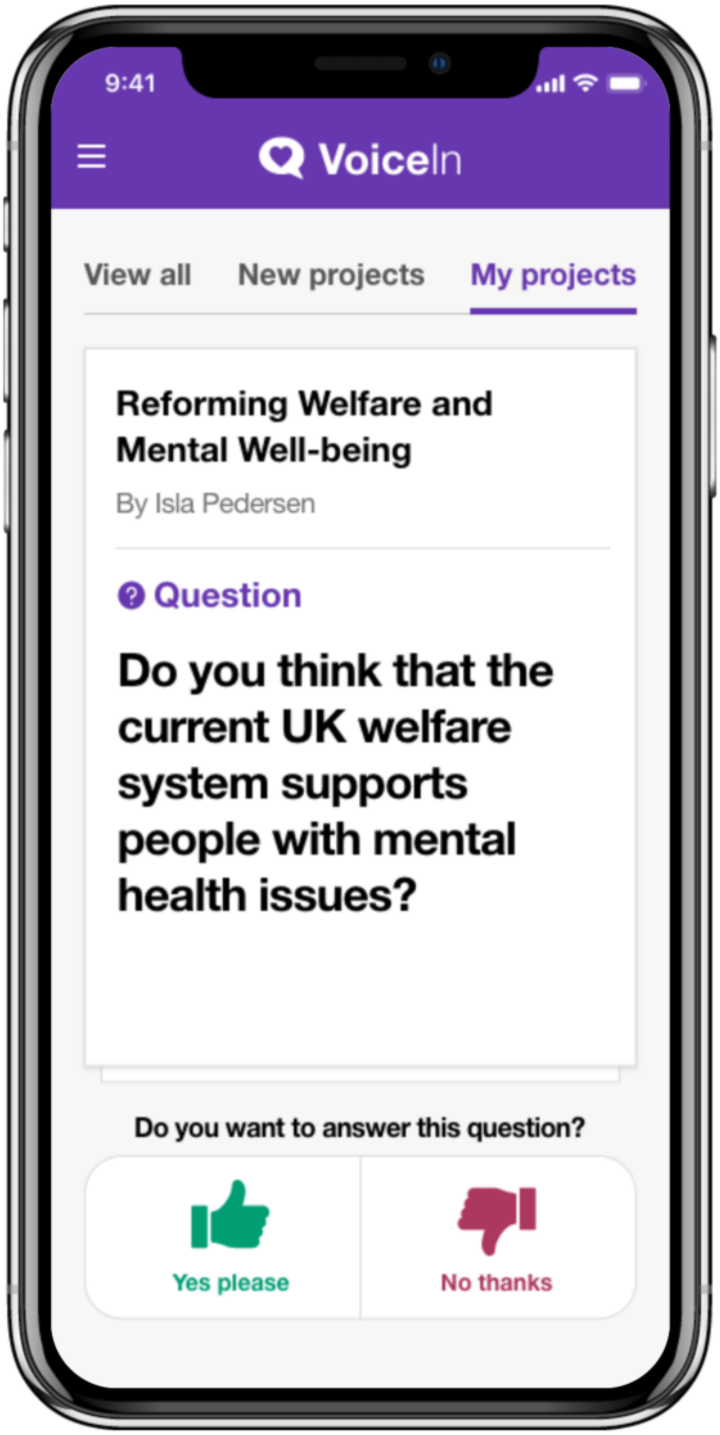
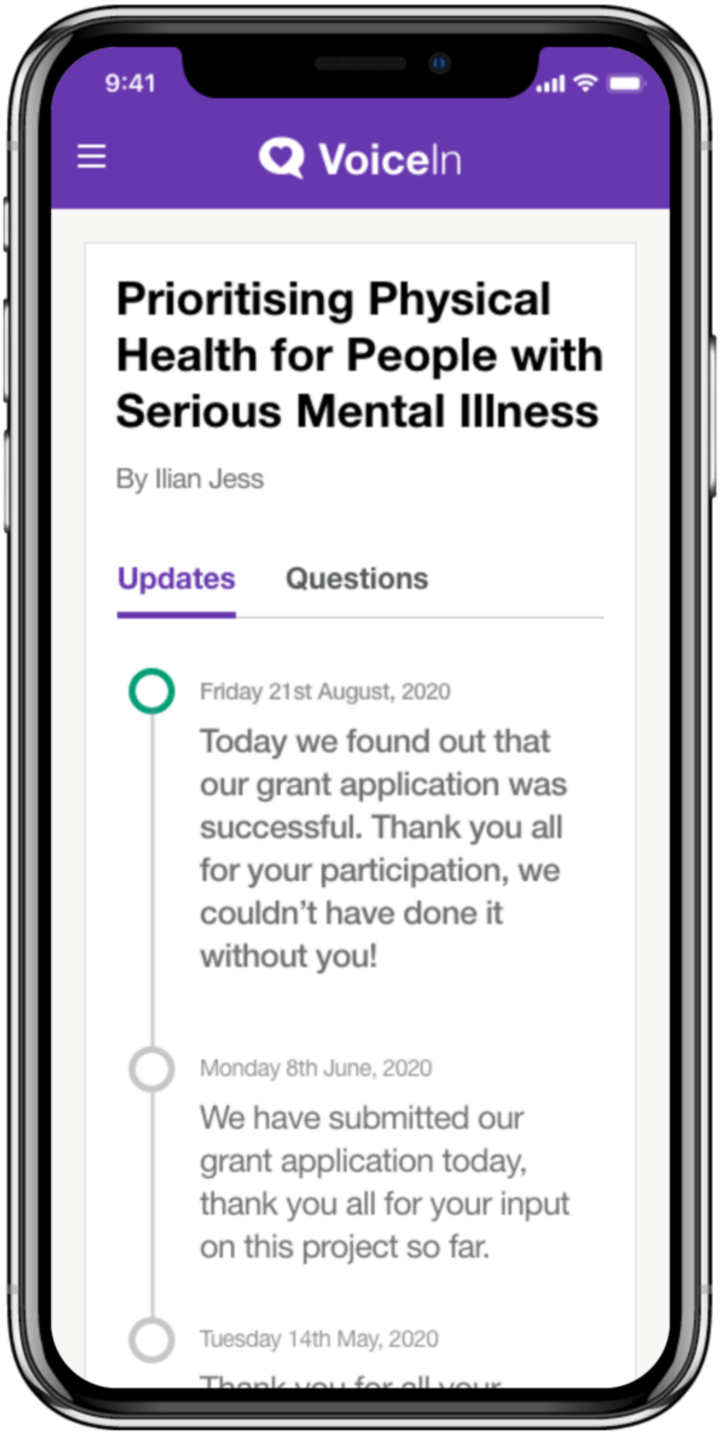


Figure 10. Sample distribution of responses to a poll.

Figure 11. Sample "project updates" page.



Researcher-Facing Web Platform

VoiceIn was designed to enable researchers to have flexibility with how they present and elicit feedback for their projects. The home screen displays all of a researcher's active projects and the amount of feedback received per project. When creating a new project, researchers insert a project title and indicate the intended audience (eg, age and interest), start and end dates, and terms of reference for the project. Project details can be displayed in the form of text, images, or videos; feedback can be requested as a poll or free text; and additional study documents can be uploaded. Once responses have been collected, researchers can view the demographic breakdown of participants, as well as see the poll and free text responses. Throughout the project, researchers can ask new questions, provide project updates, and review feedback to be displayed publicly on the app. Sample images of the web interface are displayed in Figures 12 and 13.

Figure 12. Sample “current projects” page for researchers.

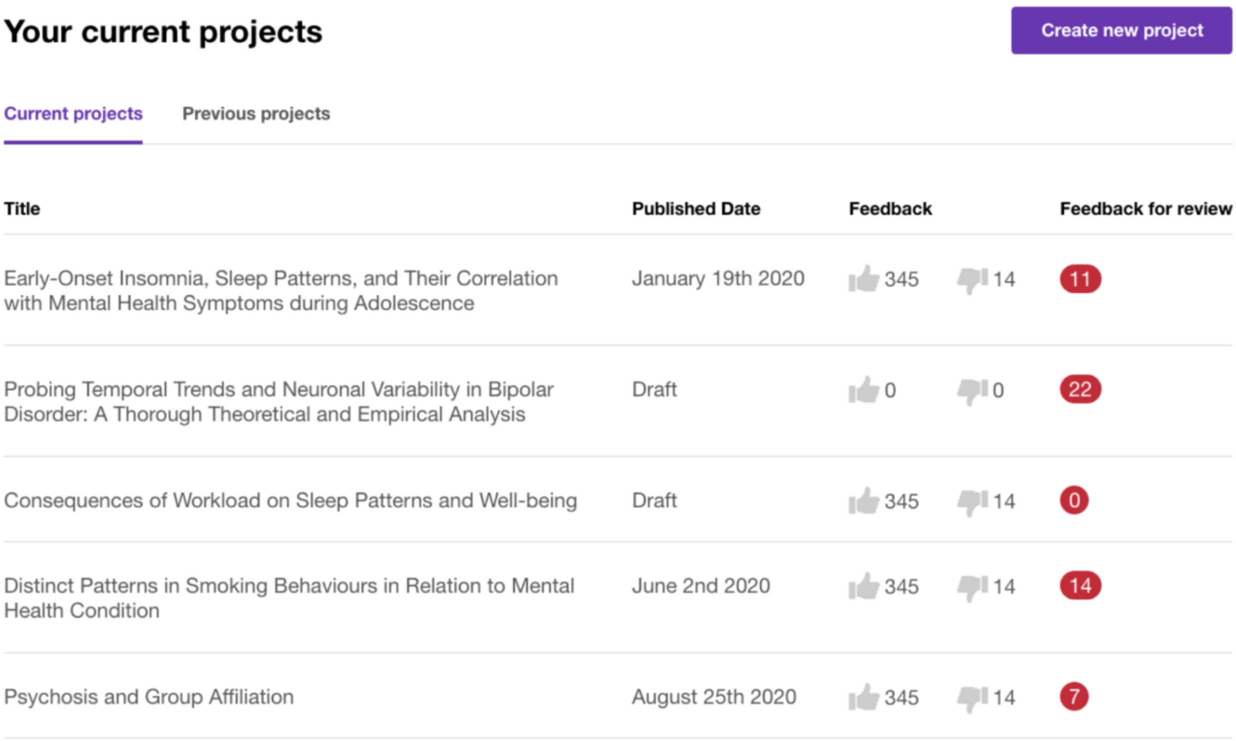
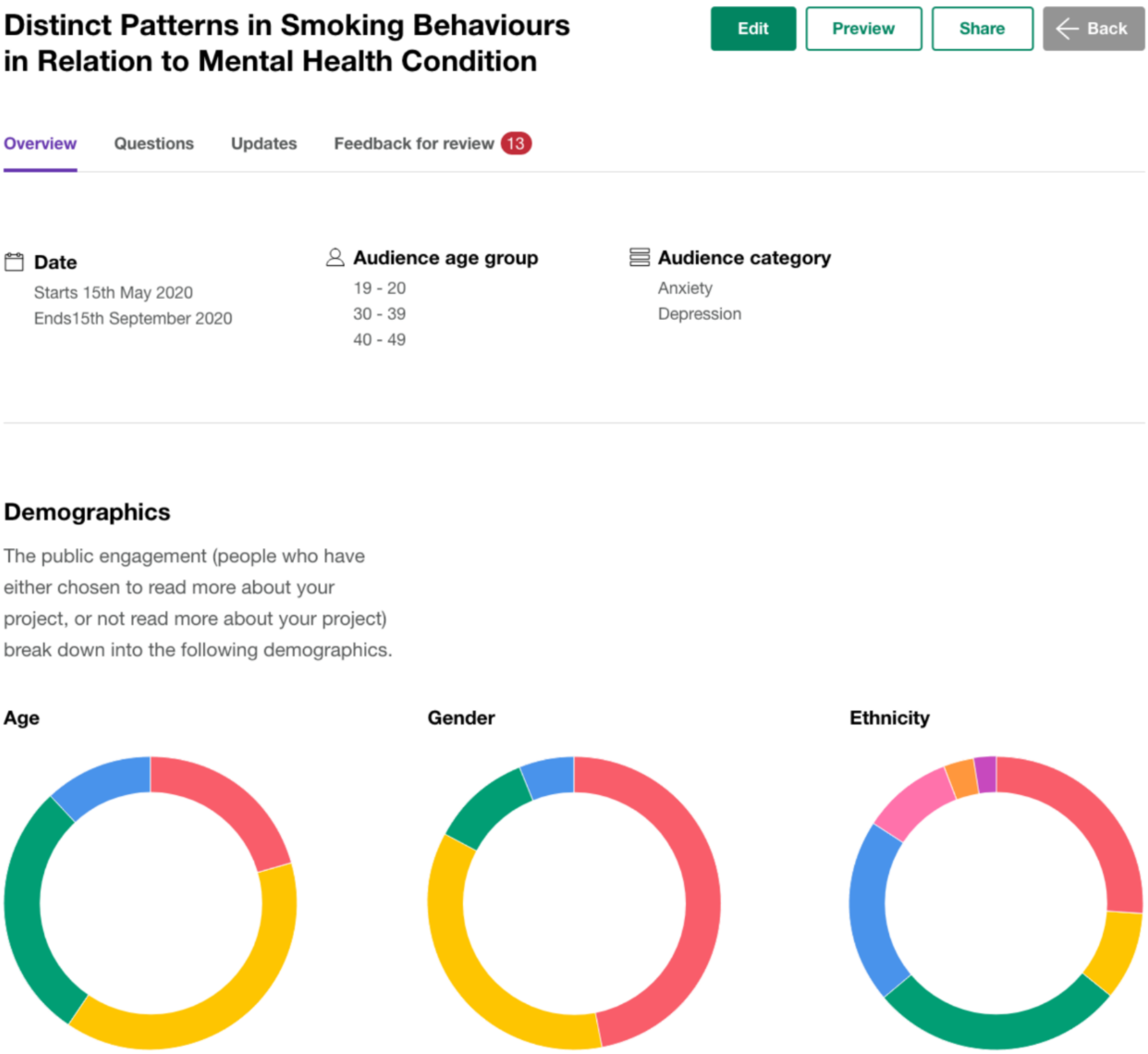


Figure 13. Sample demographics breakdown.



Discussion

Principal Findings

This paper outlines the process of co-designing the first digital tool to our knowledge that has been developed to enable real-time digital PPIE in mental health research. It likewise details the features that were identified as most important for young people, mental health professionals, and PPIE leads alike. Across stakeholder groups, participants highlighted that the VoiceIn app needed to be accessible, flexible, and easy to fit into the lives of young people. The key features that were included in the app were a flexible description of the research project, which can include the use of images and videos; numerous ways of engaging with the app and providing feedback on projects, including polls and free-text response options; and progress updates so that users can see how their input impacted the project and hear about how the research is making a difference in the world.

Co-designing the app with key stakeholders enabled the team to explore unique perspectives on developing a purely digital form of PPIE. The co-design process supported the development of an app that responded to the needs of our target user groups. It identified key features of the platform that were required for a usable and useful experience; supported feedback on the design and layout of the mobile app and web-based platform; established how a digital form could both supplant and supplement face-to-face PPIE activities; and highlighted the necessity of face-to-face modes of PPIE for in-depth discussion and exploration. It identified the potential for a mobile app that is accessible 24/7, and without geographical or spatial boundaries, to reach audiences often not represented in face-to-face activities.

The use of Agile software development methods enabled iterative versions of the app to be taken to app co-design workshops. This allowed end users to see the development of the app and the integration of features and changes that had been requested at previous workshops. Working iteratively with the software and research team, the Agile approach supported an ongoing process of iterative refinements and integration of feedback from the groups.

Time and budget inevitably constrained the breadth and depth of requirements that the software team was able to implement. Difficult choices on which features and functions to include in the app had to be made to ensure that the project met key milestones for delivery. Moreover, the co-design workshops, which were run by videoconference, encountered some of the problems that the VoiceIn app is designed to tackle. The co-design workshops required a reasonably long time commitment from people (90 mins); PPIE was conducted in a shared group space, which we recognize some may find challenging; and the research team struggled in some instances to recruit participants from diverse backgrounds. In the future, traditional modes of requesting co-design input could be supplemented by using the VoiceIn app to request feedback on itself by polling users.

Some challenges described during the co-design workshops were difficult to fully resolve and may require further iteration after live deployment. The challenge of remuneration of public users or contributors was difficult to resolve. Unlike traditional PPIE where fixed fees are usually paid for attending PPIE sessions, the groups suggested that payment for short contributions via the app was difficult to cost and to administer. Future versions of VoiceIn will consider remuneration in light of the philosophies of both PPIE, which promotes deep involvement in research, and citizen science, which may privilege rapid feedback [31]. VoiceIn will be more easily harmonized with rapid feedback models, many of which rely on voluntary involvement and nonmonetary incentives for involvement.

Next Steps

The potential of the app to support PPIE for mental health research has not yet been tested. The next phase of the project is to test the app. User needs will continue to be identified as the app is rolled out and piloted in the United Kingdom. The co-design and stakeholder involvement will continue to capture user needs after the app is in real-world use. A benefit of VoiceIn is that we can use the app to capture feedback in real-time, as well as track data analytics on how people use the app. Real-time feedback and analytics allow researchers to respond to problems or improvements identified. At the time of writing, we believe this is a world-first platform for digital PPIE; the potential for VoiceIn to reach a wider audience for PPIE than more traditional or face-to-face methods and to develop a learning health system for PPIE remains to be tested.

Our overarching goal is for VoiceIn to support PPIE activities for anyone, anywhere who would find it helpful. Our initial aim is for the platform to be widely used across the United Kingdom to support PPIE in health research broadly. Ultimately, we aim to scale the platform to support global research studies.

Conclusion

VoiceIn can support PPIE by providing an easy-to-use, digital interface that enables public contributors to share ideas and feedback on research projects from the comfort of their smartphone, at times that are convenient to them. For researchers, VoiceIn offers a new way to engage public contributors who otherwise may be difficult to access or who may not like to participate in face-to-face PPIE activities.

VoiceIn is not a solution for all PPIE needs and activities. There remains an important need for traditional and face-to-face ways of involving and engaging public contributors that can support sustained and meaningful conversations and more detailed exploration and discussion of ideas. Digital inequalities, such as lack of access to technology and lack of skills or confidence to use digital technology, also mean that VoiceIn is not an accessible solution for all. However, for VoiceIn's target audience of young people aged 16 - 25 years, providing an option to contribute to research via their smartphone was seen as convenient and appealing.

While VoiceIn was initially developed for mental health research, it has now been adapted to accommodate health research more broadly. The next phase of the project involves releasing the VoiceIn app onto the public app marketplace and enrolling research projects onto the platform. Our co-design work indicates that there is both a need and an appetite for this digital mode for PPIE. However, it remains to be seen if VoiceIn will be widely used by researchers and public contributors. The potential for VoiceIn to revolutionize PPIE by enabling individuals and groups who may have previously felt excluded from face-to-face or traditional PPIE activities and by supporting researchers to reach a wider public audience now needs to be tested.

Acknowledgments

Vocal, hosted by Manchester University NHS Foundation Trust in partnership with the University of Manchester, offers support to bring people and researchers together to connect with health research. We would like to extend our thanks to both Vocal and the GM Digital Young People's Digital Advisory Group for their participation in our co-design workshops and for offering invaluable feedback on all iterations of the app. This paper presents research funded by the National Institute for Health and Care Research (NIHR) Applied Research Collaboration Greater Manchester (ARC-GM). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Authors' Contributions

PB and KL conceived of the idea, which was developed further in discussion with PW and SB. SF designed the VoiceIn app. PW led the technical development and design of the app and led all co-design and usability activities. AB collated workshop findings and wrote the first draft of the manuscript. All authors critically reviewed the manuscript and approved the final draft for publication.

Conflicts of Interest

PW is a director and shareholder of a for-profit digital mental health company, CareLoop Health Ltd, and a director and shareholder of Prism Life Ltd, a small consultancy company. SB is a director and shareholder of CareLoop Health Ltd.

References

1. What is public involvement in research? INVOLVE. URL: https://www.invo.org.uk/find-out-more/what-is-public-involvement-in-research-2/?print=print [accessed 2013-07-25]
2. Minkler M, Wallerstein N. Introduction to CBPR: new issues and emphasis. In: Minkler M, Wallerstein N, editors. Community-Based Participatory Research for Health: From Process to Outcome: Jossey-Bass; 2008:5-23.
3. Sales CMD, Martins F, Alves MM, et al. Patient and public involvement in youth mental health research: protocol for a systematic review of practices and impact. Front Psychol 2021;12:703624. [doi: 10.3389/fpsyg.2021.703624] [Medline: 34803797]
4. Boivin A, Currie K, Fervers B, et al. Patient and public involvement in clinical guidelines: international experiences and future perspectives. BMJ Qual Saf 2010 Oct 1;19(5):e22. [doi: 10.1136/qshc.2009.034835]
5. Boote J, Baird W, Sutton A. Public involvement in the systematic review process in health and social care: a narrative review of case examples. Health Policy 2011 Oct;102(2-3):105-116. [doi: 10.1016/j.healthpol.2011.05.002] [Medline: 21641075]
6. Brett J, Staniszewska S, Mockford C, et al. Mapping the impact of patient and public involvement on health and social care research: a systematic review. Health Expect 2014 Oct;17(5):637-650. [doi: 10.1111/j.1369-7625.2012.00795.x] [Medline: 22809132]
7. Brett J, Staniszewska S, Mockford C, et al. A systematic review of the impact of patient and public involvement on service users, researchers and communities. Patient 2014;7(4):387-395. [doi: 10.1007/s40271-014-0065-0] [Medline: 25034612]
8. Knowles SE, Allen D, Donnelly A, et al. More than a method: trusting relationships, productive tensions, and two-way learning as mechanisms of authentic co-production. Res Involv Engagem 2021 May 31;7(1):34. [doi: 10.1186/s40900-021-00262-5] [Medline: 34059159]
9. Mitchell SJ, Slowther AM, Coad J, et al. Ethics and patient and public involvement with children and young people. Arch Dis Child Educ Pract Ed 2019 Aug;104(4):195-200. [doi: 10.1136/archdischild-2017-313480]
10. Rose D. Patient and public involvement in health research: ethical imperative and/or radical challenge? J Health Psychol 2014 Jan;19(1):149-158. [doi: 10.1177/1359105313500249] [Medline: 24058120]
11. Taking stock – NIHR public involvement and engagement. National Institute for Health and Care Research. 2019. URL: https://www.nihr.ac.uk/taking-stock-nihr-public-involvement-and-engagement [accessed 2023-06-18]
12. Rouncefield-Swales A, Harris J, Carter B, Bray L, Bewley T, Martin R. Children and young people's contributions to public involvement and engagement activities in health-related research: a scoping review. PLoS One 2021;16(6):e0252774. [doi: 10.1371/journal.pone.0252774] [Medline: 34106978]
13. Keane A, Islam S, Parsons S, et al. Understanding who is and isn't involved and engaged in health research: capturing and analysing demographic data to diversify patient and public involvement and engagement. Res Involv Engagem 2023 May 8;9(1):30. [doi: 10.1186/s40900-023-00434-5] [Medline: 37158951]
14. UN General Assembly. Convention on the Rights of the Child. : United Nations, Treaty Series; 1989.
15. World Health Organization. Global standards for quality health care services for adolescents. : World Health Organization; 2015.
16. Kendal SE, Milnes L, Welsby H, Pryjmachuk S, Co-Researchers' Group. Prioritizing young people's emotional health support needs via participatory research. J Psychiatr Ment Health Nurs 2017 Jun;24(5):263-271. [doi: 10.1111/jpm.12373] [Medline: 28170132]
17. Mawn L, Welsh P, Stain HJ, Windebank P. Youth Speak: increasing engagement of young people in mental health research. J Ment Health 2015;24(5):271-275. [doi: 10.3109/09638237.2014.998810] [Medline: 26193175]
18. Tume LN, Preston J, Blackwood B. Parents' and young people's involvement in designing a trial of ventilator weaning. Nurs Crit Care 2016 May;21(3):e10-e18. [doi: 10.1111/nicc.12221] [Medline: 26486094]
19. Chopel A, Lee RE, Ortiz-Matute E, et al. The META-Oak project: using photovoice to investigate youth perspectives on tobacco companies' marketing of e-cigarettes toward adolescents in Oakland. SAGE Open 2019 Jul;9(3):215824401985742. [doi: 10.1177/2158244019857420]
20. Hannon TS, Moore CM, Cheng ER, et al. Codesigned shared decision-making diabetes management plan tool for adolescents with type 1 diabetes mellitus and their parents: prototype development and pilot test. J Particip Med 2018 May 10;10(2):e8. [doi: 10.2196/jopm.9652] [Medline: 33052121]
21. Collin P, Lala G, Fieldgrass L. Participation, empowerment and democracy: engaging with young people's views. In: Alldred P, Cullen F, Edwards K, Fusco D, editors. The SAGE Handbook of Youth Work Practice: SAGE Publications Ltd; 2018:183-196. [doi: 10.4135/9781526416438.n15]
22. Collin P, Swist T. From products to publics? the potential of participatory design for research on youth, safety and well-being. J Youth Stud 2016 Mar 15;19(3):305-318. [doi: 10.1080/13676261.2015.1098774]

23. Shimmin C, Wittmeier KDM, Lavoie JG, Wicklund ED, Sibley KM. Moving towards a more inclusive patient and public involvement in health research paradigm: the incorporation of a trauma-informed intersectional analysis. BMC Health Serv Res 2017 Aug 7;17(1):539. [doi: 10.1186/s12913-017-2463-1] [Medline: 28784138]
24. Mural. URL: https://www.mural.co [accessed 2023-05-15]
25. Clegg D, Barker R. CASE Method Fast-Track: A RAD Approach: Addison-Wesley; 1994.
26. Digital health software and platforms. University of Manchester. URL: https://sites.manchester.ac.uk/digital-health-software-and-platforms/ [accessed 2024-04-09]
27. Capacitor. URL: https://capacitorjs.com/ [accessed 2024-04-09]
28. MySQL. URL: https://www.mysql.com [accessed 2024-04-09]
29. Vue.js. URL: https://vuejs.org [accessed 2024-04-09]
30. Maze. URL: https://www.maze.co [accessed 2023-05-15]
31. Wiggins A, Wilbanks J. The rise of citizen science in health and biomedical research. Am J Bioeth 2019 Aug;19(8):3-14. [doi: 10.1080/15265161.2019.1619859] [Medline: 31339831]

Abbreviations

MoSCoW: must have, should have, could have, won't have
PPIE: Patient and public involvement and engagement

Edited by A Kushniruk, D Kaufman; submitted 05.10.23; peer-reviewed by C Baxter, J Preston, PD Haghighi; revised version received 02.05.24; accepted 07.07.24; published 24.10.24.

Please cite as:
Branitsky A, Bee P, Bucci S, Lovell K, Foster S, Whelan P
Co-Designing a Digital App to Support Young People's Patient and Public Involvement and Engagement (VoiceIn): Development and Usability Study
JMIR Hum Factors 2024;11:e53394
URL: https://humanfactors.jmir.org/2024/1/e53394
doi:10.2196/53394

Corrigenda and Addenda

Correction: An Artificial Therapist (Manage Your Life Online) to Support the Mental Health of Youth: Co-Design and Case Series

Aimee-Rose Wrightson-Hester[1,2,3], BA; Georgia Anderson[4]; Joel Dunstan[5], MIT; Peter M McEvoy[1,2,6], PhD; Christopher J Sutton[7], PhD; Bronwyn Myers[1,8,9], PhD; Sarah Egan[1,2], PhD; Sara Tai[10], DClinPsy; Melanie Johnston-Hollitt[5], PhD; Wai Chen[1,11,12,13], PhD; Tom Gedeon[14], PhD; Warren Mansell[1,2,10], DClinPsy

[1]Curtin enAble Institute, Faculty of Health Sciences, Curtin University, Perth, Australia
[2]Discipline of Psychology, School of Population Health, Curtin University, Perth, Australia
[3]School of Arts and Humanities, Edith Cowan University, Perth, Australia
[4]Mental Health Commission of Western Australia, Perth, Australia
[5]Curtin Institute for Data Science, Curtin University, Perth, Australia
[6]Centre for Clinical Interventions, North Metropolitan Health Service, Nedlands, Australia
[7]Centre for Biostatistics, School of Health Sciences, The University of Manchester, Manchester, United Kingdom
[8]Alcohol, Tobacco and Other Drug Research Unit, South African Medical Research Council, Parow, South Africa
[9]Division of Addiction Psychiatry, Department of Psychiatry and Mental Health, University of Cape Town, Cape Town, South Africa
[10]Department of Clinical Psychology, School of Health Sciences, The University of Manchester, Manchester, United Kingdom
[11]Mental Health Service, Fiona Stanley Hospital, Perth, Australia
[12]Curtin Medical School, Curtin University, Perth, Australia
[13]Centre of Excellence in Medical Biotechnology, Faculty of Medical Science, Naresuan University, Phitsanulok, Thailand
[14]Optus-Curtin Centre of Excellence in AI, School of Electronic Engineering, Computing and Mathematical Sciences, Curtin University, Perth, Australia

Corresponding Author:
Aimee-Rose Wrightson-Hester, BA
Curtin enAble Institute
Faculty of Health Sciences
Curtin University
Kent St
Perth, WA 6102
Australia
Phone: 61 452491337
Email: aimee.wrightson-hester@curtin.edu.au

Related Article:

Correction of: https://humanfactors.jmir.org/2023/1/e46849/

(JMIR Hum Factors 2024;11:e69628) doi:10.2196/69628

In "An Artificial Therapist (Manage Your Life Online) to Support the Mental Health of Youth: Co-Design and Case Series" (JMIR Hum Factors 2023;10:e46849) the authors noted an error.

In the Results section of the text, a variable was wrongly described as the number of words typed by users, instead of the number of characters typed by users. Therefore, all mention of this variable has been revised.

The following sentence:

> *The word count of these conversations ranged from 58 to 2104 words and participants sent 2 to 20 texts.*

has been revised to:

> *The character count of these conversations ranged from 58 to 2104 characters including spaces and participants sent 2 to 20 texts.*

The following sentence:

> *As shown in Figure 3, in total 100% of the conversations 1000 words were rated as helpful, and all the remaining conversations 1000 words were rated as either unhelpful or neither.*

has been revised to:

As shown in Figure 3, in total 100% of the conversations in which participants typed over 1000 characters were rated as helpful, and the remaining conversations were rated as either unhelpful or neither.

The caption and axis title of Figure 3 have also been revised from:

> *Figure 3. Number of participant-generated words in each conversation and overall helpfulness rating.*

to:

> *Figure 3. Number of participant-generated characters in each conversation and overall helpfulness rating.*

In the discussion section the following sentence:

> *We identified a potential threshold of 1000 words for a conversation with MYLO to be rated as helpful, as opposed to unhelpful or "neither."*

has been revised to:

> *We identified a potential threshold of 1000 characters for a conversation with MYLO to be rated as helpful, as opposed to unhelpful or "neither."*

The correction will appear in the online version of the paper on the JMIR Publications website on December 16, 2024, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 04.12.24; this is a non–peer-reviewed article;accepted 04.12.24; published 16.12.24.

Please cite as:
Wrightson-Hester AR, Anderson G, Dunstan J, McEvoy PM, Sutton CJ, Myers B, Egan S, Tai S, Johnston-Hollitt M, Chen W, Gedeon T, Mansell W
Correction: An Artificial Therapist (Manage Your Life Online) to Support the Mental Health of Youth: Co-Design and Case Series
JMIR Hum Factors 2024;11:e69628
URL: https://humanfactors.jmir.org/2024/1/e69628
doi:10.2196/69628
PMID:

Development of a Tablet-Based Outpatient Care Application for People With Dementia: Interview and Workshop Study

Juliana Supplieth[1], BSc; Sonia Lech[2], MSc, PhD; Julie Lorraine O'Sullivan[1], PhD; Robert Spang[3], PhD; Jan‑Niklas Voigt-Antons[4], PhD; Johanna Schuster[1], PhD

[1]Institute of Medical Sociology and Rehabilitation Science, Charité–Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Charitéplatz 1, Berlin, Germany

[2]Department of Psychiatry and Psychotherapy, Charité–Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Berlin, Germany

[3]Quality and Usability Lab, Faculty IV Electrical Engineering and Computer Science, Technische Universität Berlin, Berlin, Germany

[4]Immersive Reality Lab, Hamm-Lippstadt University of Applied Sciences, Hamm, Germany

Corresponding Author:
Juliana Supplieth, BSc
Institute of Medical Sociology and Rehabilitation Science, Charité–Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Charitéplatz 1, Berlin, Germany

Abstract

Background: Dementia management presents a significant challenge for individuals affected by dementia, as well as their families, caregivers, and health care providers. Digital applications may support those living with dementia; however only a few dementia-friendly applications exist.

Objective: This paper emphasizes the necessity of considering multiple perspectives to ensure the high-quality development of supportive health care applications. The findings underscore the importance of incorporating input from stakeholders and the needs of affected families into application development.

Method: A qualitative approach was chosen, consisting of three interviews and an expert workshop. The interviews and the workshop were recorded and transcribed, and qualitative content analysis was carried out according to the methodology described by Kuckartz with the support of MAXQDA.

Results: During the development phases of the application, team meetings and discussions took place. We found that general practitioners and family caregivers play pivotal roles in the treatment and care of people with dementia, often expressing specific preferences and suggestions regarding supportive and assistive technologies. Moreover, the successful development of a useful tablet application requires robust scientific and multidisciplinary discussions and teamwork within the health care community.

Conclusion: This paper underscores the necessity of including multiple scientific, clinical, and technical perspectives to ensure the high-quality development of supportive health care applications. Furthermore, adopting a spiral development approach inclusive of feedback loops is imperative for iterative refinement and enhancement of the application.

International Registered Report Identifier (IRRID): RR2-10.1024/1662-9647/a000210

(JMIR Hum Factors 2024;11:e59865) doi:10.2196/59865

KEYWORDS
dementia; tablet application development; multidisciplinary health care; feasibility study; general practitioners; digital health care

Introduction

Dementia and Health Care

Dementia impacts approximately 47 million individuals globally, with nearly 10 million new cases emerging each year [1]. In Germany, around 1.6 million people live with dementia [2], posing a significant challenge not only for those affected but also for their families, caregivers, and health care providers. Dementia is characterized as "a syndrome resulting from mostly chronic or progressive brain disease, disrupting various higher cortical functions, including memory, cognition, orientation, perception, arithmetic, learning capacity, language, and judgment. Consciousness remains unaffected. Cognitive impairments often coincide with changes in emotional control, social behavior, or motivation..." [3]. These extensive symptoms coincide with a gradual decline in the ability to independently perform daily activities. Consequently, there are physical, psychological, social, and economic effects, leading to disability and dependence among affected individuals [1]. The substantial burden on caregivers, including family, has been widely recognized [1,4,5]. The majority of dementia caregivers are family members of people with dementia, who experience physical and psychological strain due to the declining health status and increasing health needs of the person with dementia [2,6]. For instance, the German College of General Practitioners and Family Physicians [5] provides a brief overview of various health issues faced by caregivers, such as physical health problems (eg, back injuries), mental health issues (eg, depression), and a constant feeling of being overwhelmed. Moreover, previous research has revealed that family caregivers also grapple with challenging behaviors from people with dementia [7].

Information and communication technologies (ICTs) may represent a promising approach to improve dementia treatment, especially for people with dementia that live at home. A systematic review of ICT studies by D'Onofrio et al [8] showed that ICTs can support several daily life activities for people with dementia, allow people with dementia to remain in their own home, increase the quality of life of their caregivers, and decrease health care costs. Evidence-based guidelines and training for general practitioners (GPs) [9] regarding treatment pathways and communication [10], as well as a multiprofessional team approach [11], may represent helpful mechanisms to support GPs.

GPs play a pivotal role in providing health care for people with dementia. Bohlken and Kostev [12] demonstrated that in 2015, GPs treated an average of 29.9 people with dementia, marking a 40% increase compared to 2005 (21.3 people with dementia). However, the role of GPs in diagnosing and treating dementia is subject to debate. Despite their commitment to holistic care and long-term relationships [13], there are significant challenges, including knowledge gaps, reluctance to diagnose, and cost concerns [14], alongside differing care priorities and limitations within the health care system [13]. For instance, in terms of communication, GPs' decisions to disclose a dementia diagnosis to a patient or their families are influenced by personal beliefs, patient circumstances, systemic care factors, and cultural norms [9]. Additionally, euphemistic terms like "memory problems" are often used instead of medical terminology to describe dementia [9]. Wangler and Jansky [10] observed that GPs' distant or negative attitudes, as well as their reluctance toward dementia testing and diagnoses, directly impact practical diagnostic outcomes. As a result, GPs may deny responsibility, fear stigmatization of their patients, and avoid confrontations with patients, citing concerns over the effectiveness of interventions, low compensation, and lack of recognition [10]. In summary, existing literature recognizes the indispensable role of GPs in caring for people with dementia and highlights the need to improve primary dementia care.

Digitization and Technologies

In Germany, current political initiatives aim to expedite the digitalization of the health care system, such as the Digital Act, with the goal of enhancing health care [15]. However, for these efforts to be successful, the technical solutions must align with the needs of the target group. Developing and evaluating assistive technologies necessitates an understanding of the diverse interactions between technology, users, and the contextual environment, as well as knowledge translation [16]. Dugstad et al [17] highlight crucial aspects for the successful implementation of digital technology in dementia care, emphasizing the involvement of key stakeholders from the very beginning, allowing time for exchanges and participation during the initial stages, and advocating for iterative refinement and skill development. Particularly in the development of assistive technologies for individuals with dementia, given their complex circumstances and requirements, extensive collaborative and transdisciplinary engagement is essential, as demonstrated by Boger et al [16].

The research project DemTab (tablet-based outpatient care for people with dementia) aimed to develop and evaluate a tablet-based intervention to enhance treatment for people with dementia in primary care by promoting guideline-based treatment [18,19]. To achieve this goal, a tablet-based intervention was created within the research project. The project was a collaboration between the Institute of Medical Sociology and Rehabilitation Science (ISMR) at Charité–Universitätsmedizin Berlin and the Quality and Usability Lab (QULab) at Technische Universität Berlin. Additionally, a cluster-randomized controlled trial was conducted to investigate the impact of this tablet-based intervention on guideline adherence (primary outcome) and various health-related outcomes for patients and caregivers (secondary outcomes). The tablet-based intervention offers several key functions for people with dementia and their informal caregivers [18]. The final application version contained serious games and included cognitive training games (eg, quizzes), activities focusing on daily living skills, an interactive music program, and a picture gallery for biography work. These tools were designed to engage and stimulate users and enrich their daily lives. Interactive location services provided by the Federal Ministry for Family Affairs, Senior Citizens, Women and Youth highlighted nearby consulting and support services for people with dementia, such as dementia-specific occupational therapy and daycare facilities, as well as resources for informal caregivers like support networks, counseling points, and self-help groups [20]. The communication platform allows for both indirect and direct communication with the GP through health information documentation and messaging, respectively. The care plan developed in collaboration with the GP can be accessed on the tablet by both the person with dementia and their caregivers and provides regular notifications and reminders (eg, for medication intake). Finally, a guided audio-relaxation program, which can be used by both people with dementia and their caregivers, supported relaxation and stress management.

The following functions were designed to support GPs in providing comprehensive, guideline-based care and improving communication with patients and caregivers [18]: (1) a checklist for guideline-based treatment based on the German Dementia Guideline [13]; (2) a prescription support function for antidementia drugs providing guideline-driven treatment recommendations, including automatic reminders for correct medication intake and adjustments; (3) a care development plan that assisted in creating a care plan and determining treatment plans and aims; (4) health information recorded by people with dementia and caregivers displayed on the GPs device, such as general condition or blood pressure values and direct messaging between GPs, people with dementia, and caregivers to enhance communication and action, such as adjusting medication or scheduling appointments; and (5) access to educational material on dementia and dementia care. This also included access for GPs to an electronic version of the German Dementia Guideline [13] and additional information about outpatient dementia care. People with different stages of dementia were included in the study. The application should therefore contain offers for these different stages of the disease. However, the majority (87.9%) of participants in this study had a mild to moderate dementia [19]. For more comprehensive information about the DemTab study, its participants, interventions, and expected outcomes, we refer to the study protocol [18].

Aim of This Study

This study focused on the initial phase of the DemTab study: the development of a tablet application designed for people with dementia, their caregivers, and GPs in Berlin and the surrounding regions. The objective was to outline the development process of the application within the framework of a feasibility study. The primary aim was to illustrate how health care providers (GPs and experts from supportive organizations in dementia care) contributed to the content development process and how collaborative and interdisciplinary research efforts between the ISMR and the QULab led to the cocreation of a functional health care application. A key objective was to ensure the incorporation of diverse perspectives and professional insights, as well as to facilitate a continuous and iterative research process through feedback loops and team discussions.

Methods

Ethical Considerations

Ethical approval was obtained by the ethics committee of the Charité–Universitätsmedizin Berlin (EA1/085/19). Written informed consent was obtained from all participants or legal guardians prior to data collection. All participants had the opportunity to drop out at any time. All data from the interviewees and workshop participants were pseudonymized. The list of reidentifying data was stored separately from the analyzed data, and only authorized individuals have access to them. A data protection concept and a data protection impact assessment were created for the entire project. No compensation was paid for the interviews and the workshop. At the end of the workshop, there was a small buffet lunch for all participants. No one was paid for the pretest application testing in the day clinic or the general practice.

Research Process and the International Organization for Standardization Standards

Our methodological approach adheres to the International Organization for Standardization (ISO) 9241 - 210 standard, titled “Ergonomics of human-system interaction—Part 210: Human-centered design for interactive systems” [21]. The ISO has devised a framework aimed at ensuring systems are both usable and useful, with a focus on users and their needs and requirements, by leveraging human factors, ergonomics, and usability knowledge and techniques [21]. Similarly, the user-centered design (UCD) framework provides a structured methodology for translating ideas into products, prioritizing user preferences and interactions with the final product [22]. The process aims to facilitate natural interactions without altering user behavior or expectations. UCD methodologies are rooted in the ISO 13.180 for ergonomic standards, with the standard 9241 - 210 playing a central role in the UCD framework [21]. Our approach to developing the DemTab application aligns with the 4 general phases of the UCD process.

First, we delineated the application’s context, identifying primary user groups and their typical environments based on the method outlined by the International Organization for Standardization [21]. Second, we identified the requirements that our solution must fulfill, encompassing content, usability, and regulatory compliance such as adherence to the General Data Protection Regulation principles [23]. Third, we formulated concepts and prototypes based on the identified requirements, marking the starting point of the design and development phase. Last, we concurrently and continually evaluated our prototype through expert interviews and usability tests with the target user group. This iterative process provided feedback that we could incorporate into subsequent iterations of the development process. Following the process [21], we provided a detailed account of the feedback activities that took place. Additionally, this process exhibits a resemblance to the widely recognized qualitative circular research process, which permits some degree of flexibility and refers back to previous research steps, as it embodies a more dialogical approach [24].

Process and Data Collection

The DemTab project and the application development process were based on the foundation laid by the previous project PflegeTab. PflegeTab was a tablet-based intervention designed to engage nursing home residents with dementia, with the aim of enhancing quality of life and addressing behavioral symptoms [25]. PflegeTab consisted of various adaptive tablet applications tailored to the participants’ cognitive, functional, and emotional self-regulation abilities. Based on this previous work, the main goal of the DemTab project was to extend this application and create a comprehensive platform for outpatient dementia care. In addition to the existing functionalities for people with dementia, the DemTab application also included additional functionalities developed specifically for GPs and informal caregivers [18,26].

In the initial phase of application development, two GPs from Berlin with experience in dementia primary care were interviewed (step 1): one with expertise in technological approaches in primary care and the other operating in rural areas. Another interview was conducted with a psychologist who worked as a leading manager for the National Dementia Association and was responsible for providing consultations and support to people with dementia and their relatives. These participants were selected based on their valuable expertise and were directly approached by the researchers. A semistructured interview guide was collaboratively developed by the research team. The interviewees were prompted to discuss key aspects of medical and social needs for people with dementia and their informal caregivers, as well as the role of technical support systems in practice. The objective was to gain a deeper understanding of the potential benefits of a tablet-based application for people with dementia, their caregivers, and GPs. The interviews were fully transcribed, and a content analysis was conducted following the methodology outlined by Kuckartz [27], using the MAXQDA 2018 software (VERBI Software) for data analysis.

For the expert workshop (step 1), the researchers extended invitations to two practicing GPs in Berlin, one researcher from the Institute of General Medicine at Charité–Universitätsmedizin Berlin, and two people in managing positions from a support organization based in Berlin that offers voluntary assistance for people with dementia. The workshop was held in September 2018, with participants selected based on their expertise, and they were directly approached by the researchers. While three GPs were initially invited, one was unable to attend. The workshop commenced with a presentation outlining the DemTab study, followed by participants testing version 0.5 of the application and engaging in a focus group discussion. This discussion included a presentation of the results pertaining to the 4 application functions: dementia guidelines, data representation, messaging, and games. Additionally, all workshop attendees completed a questionnaire detailing their personal information and their respective contexts of working with people with dementia and informal caregivers. Following the workshop, members of the research team also completed a brief written survey to capture immediate reflections. The survey comprised of 3 questions including:

- How did you perceive the workshop?
- Which application features or discussion topics were quickly dismissed?
- What topics were extensively discussed?

Team talks and discussions, as well as mutual agreements between researchers from the ISMR and the QULab, were integral (step 2). Biweekly team meetings were held throughout the development period, following a structured procedure. First, the QULab presented new feature additions, then the ISMR researchers tested and evaluated them, then all team members engaged in discussions regarding the user-friendliness of the tool, and finally potential changes and additions informed by insights from interviews and the expert workshop were instituted. The central questions of this study included:

- What valuable knowledge from the previous project PflegeTab should inform our approach?
- What insights from qualitative interviews and workshops should shape specific application functions?
- How can we enhance the implementation of the application in real-world settings?
- What technical features can feasibly be implemented within the timeframe and how?

Additionally, a pretest (step 4) of the DemTab application (version 0.5) was conducted in a geriatric day clinic in Berlin, where people with dementia and staff used the application for 7 days. Furthermore, the application was tested once by a GP who provided feedback and recommendations on content and visualization.

Results

The main results are presented following the steps explained in this section.

First Step: Preparation—Interviews and Workshop Findings

General Practitioners

Through the interviews and workshop, it became evident that GPs play a crucial role in the comprehensive care of individuals living with dementia and their families. Serving as primary care providers, they are trusted figures who facilitate communication and act as mediators. GP offices also serve as hubs for interdisciplinary and intersectoral communication. It was emphasized that collegial exchanges with specialists and collaboration with family members are vital for ensuring comprehensive home-based care. Moreover, there was a call for increased usage of dementia-specific advisory and support services. One GP highlighted the need for quick access to reliable information, noting the impracticality of consulting extensive guidelines on short notice. Describing the factors influencing care, one GP stated, “Accessing fast, secure information is essential; however, it’s impractical to review all the extensive guidelines on short notice” (Interviewee 1, passage 205).

Technology

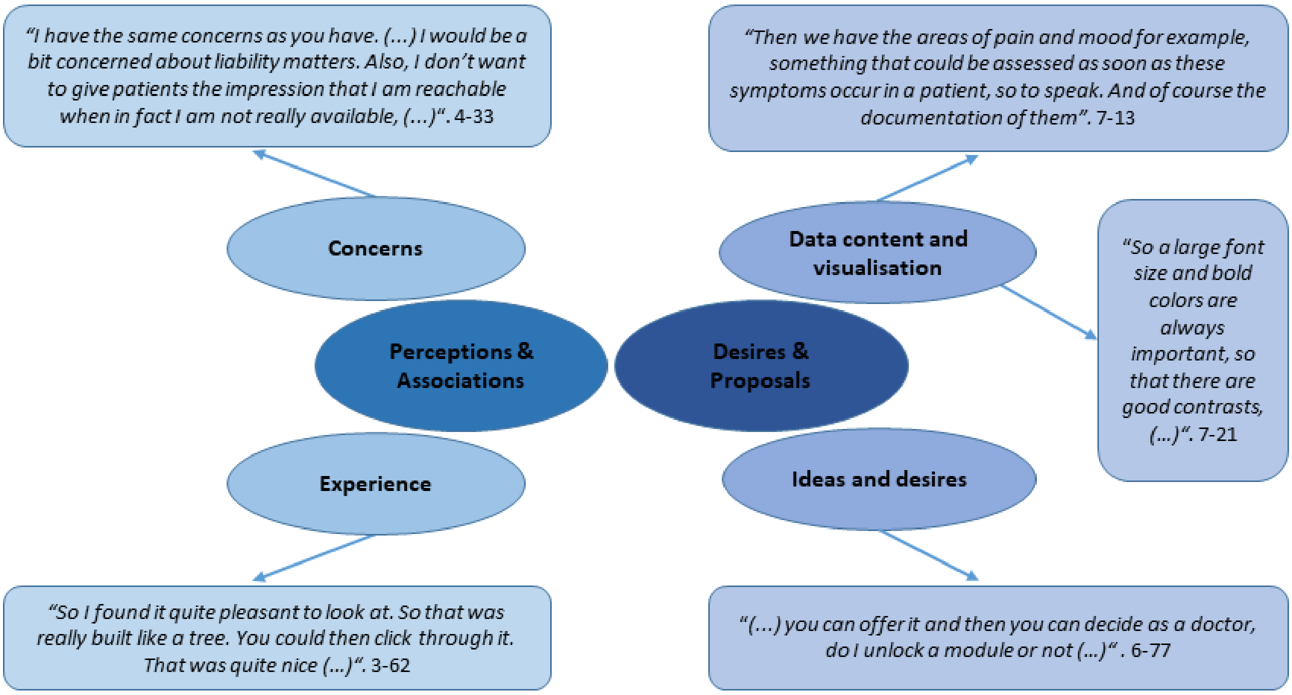
GPs emphasized the importance of having concise, practical, and filtered information to ensure guideline-oriented treatment and cooperative specialist care. They suggested that data collected by the application should be prioritized and presented graphically to improve comprehension. GPs stressed the significance of choice and self-determination in using the tablet, as well as decisions regarding its functionality. They also advocated for greater integration and networking to facilitate the sharing of information on available assistance and support services. Additionally, there were concerns about the potential for relatives to overload the technical application, particularly with communication functions, which could inadvertently burden GPs. This was discussed critically along with the advantages of video consultations, but they were ultimately rejected by the participating GPs. One GP also commented on potential future technologies stating, “The advantage of video calling is being able to see the person; that adds a different level compared to just talking” (Interviewee 2, passage 130).

Relatives and Caregivers

Relatives and caregivers play a significant and multifaceted role in the care of people with dementia at home. Family caregiving occurs under diverse circumstances and can be accompanied by significant stress. Consequently, family members of people with dementia require guidance on medical, psychosocial, and legal matters, as well as access to support services aimed at alleviating stress and fostering social connections tailored to their individual situations and needs. Empowering individuals to help themselves and enhancing their ability to act independently were recognized as crucial objectives. This is illustrated by the following quote stating, “The psychosocial counseling of relatives, aiming to empower them to help themselves and to raise awareness of their own competencies" (Interviewee 8, passage 15). Conclusively, caregivers should be actively engaged in diagnostic and therapeutic processes.

The workshop results encompassed information within two primary categories: perceptions and associations, which encompassed experiences and concerns, as well as desires and proposals, which included considerations regarding data content, visualization, and innovative ideas. The principal findings and examples are detailed in Figure 1.

Figure 1. Main categories and subcategories of the workshop with examples. Interviewee ID and passage number for each quote are presented in parentheses, respectively.



Second and Third Step: Application Development—Accompanying Team Talks and Discussions

The outcomes of the second step represent the culmination of frequent interdisciplinary team discussions and represent the core of application development. During these team deliberations, in conjunction with the findings of qualitative content analyses, the team concentrated on the study interests and objectives, feasibility, implementation of content, and presentation styles, taking into consideration time and resource constraints. The discussions were marked by their vibrant and emotionally charged nature, especially when deliberating the development of the application's features. This negotiation of feature development emerged as the primary focus within the team and was carried out in an interdisciplinary manner. For instance, decisions regarding medical content originated from the ISMR, while those concerning technical aspects were made by the QULab.

A substantial portion of the discussions drew upon insights gathered from the interviews and workshop. Based on these insights, the team opted to incorporate ten questions related to the state of health that people with dementia, with or without the assistance of their caregivers, should respond to at least weekly. These questions centered on daily activities such as well-being, hydration, and sleep patterns. Formulated by the IMSR team, these questions aimed to provide GPs with valuable health information and were crafted to be easily accessible by participating individuals. Further, the interdisciplinary discussions with GPs and caregiving experts led to several critical modifications in the application's design and functionality, including:

- An emergency disclaimer in the chat: A disclaimer was included in the chat function to inform users that the application is not intended for emergencies and that emergency services should be contacted in case of an emergency, with the appropriate emergency number prominently displayed.
- Patient records: The application was updated to include detailed patient profiles, allowing health care providers to access comprehensive patient information quickly and efficiently.
- A single patient view interface: The application's interface was redesigned to enable a single patient view, where selecting a patient displays all relevant data across various application functions, streamlining the process for health care providers.
- Checklist integration: Based on medical guidelines for dementia care, a checklist feature was added. This checklist allows items to be checked off as completed, with the system automatically resetting the item after three months to ensure regular follow-ups as per the guidelines.
- User interface enhancements: Various suggestions were incorporated to improve the application's usability and accessibility, ensuring it was intuitive and user-friendly for all health care providers.

Fourth Step: Application Testing—Feedback and Finalization

The pretest occurred in a geriatric day clinic, involving people with dementia, a social worker, physicians, and one GP who had previously participated in interviews. They provided comprehensive feedback on individual subapplications, focusing on aspects such as user-friendliness, effects on people with dementia and themselves, identification of missing or redundant content, and suggestions for useful additions. Given the time constraints of the project schedule and limited financial resources, discussions on prioritization and timelines were impactful. Throughout the development process, there was a consistent emphasis on ensuring good usability and an appealing presentation style.

Discussion

The findings of this study indicate that GPs and relatives bear the primary responsibility for the treatment and care of people with dementia, and they harbor specific desires and proposals concerning supportive and assistive technologies. Interdisciplinary discussions are essential to address these needs and develop a practical tablet application. The primary objective of this study was to delineate a multidisciplinary and interdisciplinary approach to the development of an application aimed at enhancing primary care for people with dementia and their caregivers. People with dementia and their caregivers should be supported in their daily lives and disease management, and GPs should be provided with support for their health care. The application should support the joint and interlocking care of those involved.

In conclusion, this study yields various results and implications. First, the study highlights the significant contribution of individuals with diverse disciplinary and professional backgrounds in the development of supportive technical health care applications. In addition to researchers, professionals working in different areas of dementia care provided invaluable insights into care aspects for people with dementia and their family caregivers. These experts should be integrated into the development process as specialists, as they possess implicit knowledge and contributory expertise [28], and experts in local knowledge should be included to leverage the potentials and mitigate the risks of such a study [29]. It is essential for potential users and beneficiaries of the application to be involved from the outset of the development process [21].

We chose to interview employees of the Alzheimer's Association instead of individuals with dementia and their caregivers. While this approach does not replace the deep, individual perceptions and experiences of affected individuals and their caregivers, it allowed us to gather consolidated information and comprehensive insights accumulated over years of work. Moreover, factors such as time constraints, personal resources, and the challenges of accessing interviewees influenced this decision. While much is known about the progressive and multifaceted challenges and needs of people with dementia, the health care system, professional care institutions, and society still fall short in providing full-time emotional and practical support for them, their caregivers, and their GPs [1]. Additionally, there remains ambiguity surrounding technical solutions [8]. The participation of GPs and employees from the Alzheimer's Association provided practical insights from health care providers engaged in daily interactions with people with dementia and their caregivers. They shared professional experiences and highlighted daily challenges and needs for improvement. This collaborative process can enhance the acceptance and usage of the application [21].

Our interdisciplinary approach not only facilitated a comprehensive understanding of the needs of GPs and caregivers but also directly influenced the application's development. The inclusion of features such as the emergency button, patient records, and a checklist system were direct results of feedback from health care professionals. These features were critical in ensuring that the application provided practical, guideline-oriented support in real-world health care settings. Moreover, the interface redesign to include a single patient view and other usability improvements underscored the importance of integrating diverse perspectives to create a tool that is both functional and user-friendly.

Second, the process of developing a health application necessitated multidisciplinary and interdisciplinary teamwork. It is crucial to differentiate between intermittent multidisciplinary teamwork, characterized by juxtaposed disciplinary work, and interdisciplinary teamwork, which entails proactive and interactive collaboration [30]. However, both multidisciplinary and interdisciplinary teamwork are circular scientific endeavors in social care research and involve a process of social negotiation. Effective collaboration involves not only managing coworking due to separate work locations, formal team meetings, and the exchange of information and data but also engaging in discussions and informal team meetings to address preferences, relevancies, and disagreements. Given that each scientific discipline possesses its own specific domain, concepts, and methodologies, multiple constraints and barriers between different disciplines must be dismantled and integrated [31]. Successful multidisciplinary and interdisciplinary teamwork necessitates openness to the knowledge, structures, and practices of other disciplines [32] while consistently advocating for one's own discipline. Continuous communication and compromise are essential requirements for effective collaboration [21].

Third, future generations are expected to rely more heavily on technical applications for various aspects of life, including health care [33]. Therefore, it is crucial to examine and delve into the development of technical tools, even though the current older generation may not be as technologically savvy as today's adults and youth. However, it is anticipated that the usage and acceptance of digital technologies will grow in the future [33].

Overall, this study underscores the necessity of multiple scientific, clinical, and technical perspectives to ensure the high-quality development of supportive health care applications. The content and design of such applications for GPs, people with dementia, and their caregivers must be informed by the expertise and experiences of various professional disciplines involved in health care provisions and the usage of such applications in real health care settings. Therefore, multidisciplinary and interdisciplinary collaboration is indispensable in the realm of health care–supporting applications. The development of this dementia care application exemplifies the importance of an interdisciplinary approach in health care technology. The modifications made based on feedback from GPs and other caregiving experts, such as the emergency button, patient profiles, and the checklist feature, highlight how diverse professional insights can lead to a more effective and user-centered design. These changes not only improved the application's functionality but also enhanced its acceptance and usability among health care providers. The overarching goal is to systematically improve health care for those affected and those involved in their care. Everyday life and care for people with dementia should be supported. Medical disciplines, social sciences, technical expertise, and supporting organizations are all essential to integrate relevant knowledge and competencies into the design and development of useful products, as well as facilitating daily use and active support for health care needs. Furthermore, such development processes require a spiral approach and feedback loops, as well as critical professional and emotional discussions, to reach a consensus and make informed decisions.

Acknowledgments

The authors would like to express their special appreciation to Sebastian Möller and Adelheid Kuhlmey for their support in the application process of the DemTab study, as well as to all workshop participants, employees of the day clinic, and all study participants. We used the generative artificial intelligence tool ChatGPT 3.5 (OpenAI) to correct and improve the English language. The original ChatGPT transcripts are provided as Multimedia Appendix 1. The research is funded by the Innovation Fund of the Federal Joint Committee (G-BA), (Grant Project "DemTab" Grant-Number 01VSF17039).

Data Availability

Data are stored in a nonpublicly available repository. Data are available from the corresponding author upon request.

Authors' Contributions

J Supplieth was the primary contributor to data analysis, data interpretation, and manuscript drafting. SL, JLOS, JNVA, and RS assisted with writing and interpreting findings. J Schuster and JNVA designed and led the DemTab study in which this study was embedded. All authors contributed to this study and critically revised and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Chat GPT transcript.
[DOCX File, 60 KB - humanfactors_v11i1e59865_app1.docx]

References

1. Global action plan on the public health response to dementia 2017 - 2025. World Health Organization. 2017. URL: https://www.who.int/publications/i/item/global-action-plan-on-the-public-health-response-to-dementia-2017---2025 [accessed 2024-11-19]
2. Dementia in europe yearbook 2019: estimating the prevalence of dementia in europe. Alzheimer Europe. 2019. URL: https://www.alzheimer-europe.org/sites/default/files/alzheimer_europe_dementia_in_europe_yearbook_2019.pdf [accessed 2024-11-19]
3. F00-F09 Organische, einschließlich symptomatischer psychischer Störungen [Article in German]. ICD-Code. URL: https://www.icd-code.de/icd/code/F03.html [accessed 2024-11-19]
4. Informationsblatt 7: Die Entlastung pflegender Angehöriger [Article in German]. Deutsche Alzheimer Gesellschaft eV Selbsthilfe Demenz. 2000. URL: https://www.deutsche-alzheimer.de/fileadmin/Alz/pdf/factsheets/infoblatt7_entlastung.pdf [accessed 2024-11-25]
5. Deutsche Gesellschaft für Allgemeinmedizin und Familienmedizin e.V. Definition und Epidemiologie des Gesundheitsproblems. In: Pflegende Angehörige von Erwachsenen S3-Leitlinie AWMF-Register-Nr 053-006 DEGAM-Leitlinie Nr 6: German College of General Practitioners and Family Physicians; 2018:24-29 URL: https://register.awmf.org/assets/guidelines/053-006l_S3_Pflegende-Angehoerige-von-Erwachsenen_2019-03-abgelaufen.pdf [accessed 2024-11-25]
6. Zwingmann I, Hoffmann W, Michalowsky B, et al. Offene Versorgungsbedarfe pflegender Angehöriger von Menschen mit Demenz. Nervenarzt 2018 May;89(5):495-499. [doi: 10.1007/s00115-018-0509-1]
7. Eggert S, Schnapp P, Sulmann D. Aggressionen und Gewalt in der Pflege. Zentrum für Qualität in der Pflege. 2018. URL: https://www.zqp.de/wp-content/uploads/ZQP_Analyse_Gewalt_informelle_Pflege.pdf [accessed 2024-11-25]

8. D'Onofrio G, Sancarlo D, Ricciardi F, et al. Information and communication technologies for the activities of daily living in older patients with dementia: a systematic review. J Alzheimers Dis 2017;57(3):927-935. [doi: 10.3233/JAD-161145] [Medline: 28304297]
9. Low LF, McGrath M, Swaffer K, Brodaty H. Communicating a diagnosis of dementia: a systematic mixed studies review of attitudes and practices of health practitioners. Dem (Lond) 2019;18(7-8):2856-2905. [doi: 10.1177/1471301218761911] [Medline: 29544345]
10. Wangler J, Jansky M. Dementia diagnostics in general practitioner care. Wien Med Wochenschr 2020 Jun;170(9-10):230-237. [doi: 10.1007/s10354-019-00722-4]
11. Pentzek M, Vollmar HC, Wilm S, Leve V. Putting dementia awareness into general practice: the CADIF approach. Z Gerontol Geriatr 2017 May;50(Suppl 2):44-47. [doi: 10.1007/s00391-017-1206-6] [Medline: 28315047]
12. Bohlken J, Kostev K. Diagnostic and prescription behavior of general practitioners and specialist physicians in patients with dementia in 2005 and 2015 in Germany. Psychiatr Prax 2018 Apr;45(3):154-159. [doi: 10.1055/s-0043-110032] [Medline: 28499316]
13. Deutsche Gesellschaft für Psychiatrie und Psychotherapie, Psychosomatik und Nervenheilkunde e. V. (DGPPN), Deutsche Gesellschaft für Neurologie e. V. (DGN). S3-Leitlinie Demenzen Langfassung: Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e. V; 2023. URL: https://register.awmf.org/assets/guidelines/038-013l_S3_KF_Demenzen-2023-09.pdf [accessed 2024-11-25]
14. Deutsche Gesellschaft für Allgemeinmedizin und Familienmedizin e.V. DEGAM-Leitlinie Nr12: Demenz: Omikron publishing; 2008. URL: https://www.yumpu.com/de/document/view/5708562/degam-leitlinie-zur-demenz-arztbibliothek [accessed 2024-12-02]
15. Digitalisation in healthcare. Federal Ministry of Health. 2023. URL: https://www.bundesgesundheitsministerium.de/en/themen/digitalisierung/digitalisation-in-healthcare.html [accessed 2024-11-19]
16. Boger J, Jackson P, Mulvenna M, et al. Principles for fostering the transdisciplinary development of assistive technologies. Disabil Rehabil Assist Technol 2017 Jul;12(5):480-490. [doi: 10.3109/17483107.2016.1151953] [Medline: 27052793]
17. Dugstad J, Eide T, Nilsen ER, Eide H. Towards successful digital transformation through co-creation: a longitudinal study of a four-year implementation of digital monitoring technology in residential care for persons with dementia. BMC Health Serv Res 2019 Jun 10;19(1):366. [doi: 10.1186/s12913-019-4191-1] [Medline: 31182093]
18. Lech S, O'Sullivan JL, Gellert P, Voigt-Antons JN, Greinacher R, Nordheim J. Tablet-based outpatient care for people with dementia. GeroPsych (Bern) 2019 Sep;32(3):135-144. [doi: 10.1024/1662-9647/a000210]
19. Lech S, O'Sullivan JL, Drewelies J, et al. Dementia care and the role of guideline adherence in primary care: cross-sectional findings from the DemTab study. BMC Geriatr 2021 Dec 18;21(1):717. [doi: 10.1186/s12877-021-02650-8] [Medline: 34922486]
20. Wegweiser Demenz. 2024. URL: https://www.wegweiser-demenz.de/ [accessed 2024-11-25]
21. Ergonomics of human-system interaction — part 210: human-centred design for interactive systems. International Organization for Standardization. 2019. URL: https://www.iso.org/standard/77520.html [accessed 2024-11-19]
22. Abras C, Maloney-Krichmar D, Preece J. User-centered design. In: Bainbridge WS, editor. Encyclopedia of Human-Computer Interaction: Berkshire Publishing Company; 2004:445-456.
23. Principles of the GDPR. European Commission. URL: https://commission.europa.eu/law/law-topic/data-protection/reform/rules-business-and-organisations/principles-gdpr_en [accessed 2024-11-19]
24. Witt H. Strategies in qualitative and quantitative research forum qualitative sozialforschung/forum: qualitative social research. Forum Qualitative Sozialforschung. 2001. URL: https://www.qualitative-research.net/index.php/fqs/article/view/969 [accessed 2024-11-19]
25. O'Sullivan JL, Lech S, Gellert P, et al. A tablet-based intervention for activating nursing home residents with dementia: results from a cluster-randomized controlled trial. Int Psychogeriatr 2022 Feb;34(2):129-141. [doi: 10.1017/S1041610221000818] [Medline: 34183087]
26. Lech S, O'Sullivan JL, Wellmann L, et al. Recruiting general practitioners and patients with dementia into a cluster randomised controlled trial: strategies, barriers and facilitators. BMC Med Res Methodol 2021 Mar 30;21(1):61. [doi: 10.1186/s12874-021-01253-6] [Medline: 33784967]
27. Kuckartz U. Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung 2, Durchges Aufl: Beltz Juventa; 2014.
28. Collins H, Evans R. Rethinking Expertise: University of Chicago Press; 2007. [doi: 10.7208/chicago/9780226113623.001.0001]
29. Barrotta P, Montuschi E. Expertise, relevance and types of knowledge. Soc Epistemol 2018 Nov 2;32(6):387-396. [doi: 10.1080/02691728.2018.1546345]
30. Klein JT. A taxonomy of interdisciplinarity. In: The Oxford Handbook of Interdisciplinarity: Oxford University Press; 2010:15-30.
31. MacLeod M. What makes interdisciplinarity difficult? Some consequences of domain specificity in interdisciplinary practice. Synthese 2018 Feb;195(2):697-720. [doi: 10.1007/s11229-016-1236-4]

32. Woiwode H, Froese A. Two hearts beating in a research centers' chest: how scholars in interdisciplinary research settings cope with monodisciplinary deep structures. Stud Higher Educ 2021 Nov 2;46(11):2230-2244. [doi: 10.1080/03075079.2020.1716321]
33. Achter Altersbericht: Ältere Menschen und Digitalisierung. Bundesministerium für Familie, Senioren, Frauen und Jugend. 2020. URL: https://www.bmfsfj.de/resource/blob/159916/9f488c2a406ccc42cb1a694944230c96/achter-altersbericht-bundestagsdrucksache-data.pdf [accessed 2024-11-25]

Abbreviations

GP: general practitioner
ICT: information and communication technology
ISMR: Institute of Medical Sociology and Rehabilitation Science
ISO: International Organization for Standardization
QULab: Quality and Usability Lab
UCD: user-centered design

Edited by A Choudhury; submitted 24.04.24; peer-reviewed by MB Ellur; revised version received 11.09.24; accepted 16.09.24; published 19.12.24.

Please cite as:
Supplieth J, Lech S, O'Sullivan JL, Spang R, Voigt-Antons J, Schuster J
Development of a Tablet-Based Outpatient Care Application for People With Dementia: Interview and Workshop Study
JMIR Hum Factors 2024;11:e59865
URL: https://humanfactors.jmir.org/2024/1/e59865
doi:10.2196/59865

AI Hesitancy and Acceptability—Perceptions of AI Chatbots for Chronic Health Management and Long COVID Support: Survey Study

Philip Fei Wu[1], PhD; Charlotte Summers[2,3], BSc; Arjun Panesar[2,3], MEng; Amit Kaura[2,4], MSc, MBChB, PhD; Li Zhang[5], BEng, MSc, PhD

[1]School of Business and Management, Royal Holloway, University of London, Egham, United Kingdom

[2]DDM Health, Coventry, United Kingdom

[3]Warwick Medical School, University of Warwick, Coventry, United Kingdom

[4]Imperial College Healthcare NHS Trust, London, United Kingdom

[5]Department of Computer Science, Royal Holloway, University of London, Egham, United Kingdom

Corresponding Author:
Philip Fei Wu, PhD
School of Business and Management, Royal Holloway, University of London, , Egham, , United Kingdom

Abstract

Background: Artificial intelligence (AI) chatbots have the potential to assist individuals with chronic health conditions by providing tailored information, monitoring symptoms, and offering mental health support. Despite their potential benefits, research on public attitudes toward health care chatbots is still limited. To effectively support individuals with long-term health conditions like long COVID (or post–COVID-19 condition), it is crucial to understand their perspectives and preferences regarding the use of AI chatbots.

Objective: This study has two main objectives: (1) provide insights into AI chatbot acceptance among people with chronic health conditions, particularly adults older than 55 years and (2) explore the perceptions of using AI chatbots for health self-management and long COVID support.

Methods: A web-based survey study was conducted between January and March 2023, specifically targeting individuals with diabetes and other chronic conditions. This particular population was chosen due to their potential awareness and ability to self-manage their condition. The survey aimed to capture data at multiple intervals, taking into consideration the public launch of ChatGPT, which could have potentially impacted public opinions during the project timeline. The survey received 1310 clicks and garnered 900 responses, resulting in a total of 888 usable data points.

Results: Although past experience with chatbots (P<.001, 95% CI .110-.302) and online information seeking (P<.001, 95% CI .039-.084) are strong indicators of respondents' future adoption of health chatbots, they are in general skeptical or unsure about the use of AI chatbots for health care purposes. Less than one-third of the respondents (n=203, 30.1%) indicated that they were likely to use a health chatbot in the next 12 months if available. Most were uncertain about a chatbot's capability to provide accurate medical advice. However, people seemed more receptive to using voice-based chatbots for mental well-being, health data collection, and analysis. Half of the respondents with long COVID showed interest in using emotionally intelligent chatbots.

Conclusions: AI hesitancy is not uniform across all health domains and user groups. Despite persistent AI hesitancy, there are promising opportunities for chatbots to offer support for chronic conditions in areas of lifestyle enhancement and mental well-being, potentially through voice-based user interfaces.

(JMIR Hum Factors 2024;11:e51086) doi:10.2196/51086

KEYWORDS

AI hesitancy; chatbot; long COVID; diabetes; chronic disease management; technology acceptance; post–COVID-19 condition; artificial intelligence

Introduction

Artificial intelligence (AI) chatbots are programs designed to simulate human conversations and provide tailored responses to users' questions and concerns. Chatbots can provide a range of services, including appointment scheduling, medication reminders, and various types of health support. AI chatbots have the potential to support individuals with chronic health conditions by providing tailored information and resources, monitoring symptoms, and offering emotional support [1]. While there are some limitations to chatbots' use, they could be a valuable tool for individuals with long-term conditions such as long COVID (or post–COVID-19 condition) [2] and those living in remote or rural areas [3,4].

Researchers have long investigated the use of chatbots in managing various chronic illnesses. For example, past studies have documented how chatbots improved medication adherence rates of patients with breast cancer [5]; enhanced the quality of life for people with type 2 diabetes [6]; reduced the severity of panic disorder symptoms [7]; and helped health care professionals, patients with asthma, and their family members build collaborative relationships [8]. Several systematic literature reviews on the topic suggest that conversational agents are generally effective in supporting the self-management of chronic conditions, particularly for mental health [9-11]. Hence, "empathic" chatbots that demonstrate "emotional intelligence" seem particularly relevant and useful. Although some researchers use the term emotional intelligence to denote a chatbot's ability to express a full range of human sentiments (positive and negative) [12], in the health care context, an emotionally intelligent chatbot usually refers to a conversational agent being able to recognize and respond to emotions a user expresses in their interaction and that "uses evidence-based self-help practices such as CBT, DBT, motivational interviewing, positive behavior support, behavioral reinforcement, mindfulness, and guided microactions and tools to encourage users to build emotional resilience skills" [13].

Despite the potential benefits of AI chatbots in health care, empirical research on public attitudes toward health care chatbots is still in its early stages [14]. Some early studies have suggested that users are generally positive about the use of AI chatbots [15]. For example, Bickmore et al [16] found that participants were generally satisfied with a health care chatbot that provided them with medication reminders and lifestyle advice. Similarly, a study by Crutzen et al [17] found that a health promotion chatbot targeting adolescents was used intensively and evaluated positively, especially in comparison with information lines and search engines. However, recent research has also highlighted many challenges associated with the use of AI chatbots in health care. There are concerns about the ability of chatbots to understand complex medical issues and provide accurate advice [11,18]. Patients and medical researchers alike were skeptical about the use of chatbots for mental health support, citing concerns about the lack of empathetic communication and the potential for the chatbot to misunderstand their emotional states [18,19].

Research on voice-based chatbots for health management is also in its infancy. Medical professionals' views on voice-based chatbots echo views on text-based chatbots in terms of the technology's potential and limitations. A 2-round Delphi study [20] surveying experts on the future of voice-controlled AI agents in health care anticipates significant technological development and increased user trust. The study focused on how voice-controlled agents could support health care professionals through applications like remote real-time interviews with patients, hands-free instructions for medical staff, and communication between staff and patients. However, the authors concluded that chatbots are not expected to outperform or replace human health care workers despite a more intuitive speech interaction.

A systematic review conducted in 2020 examining the literature on voice-based conversational agents for chronic health conditions only found 12 papers [21]. In another scoping review conducted in 2021, only 4 studies among 32 reviewed focused on voice-based chatbots in health care [11]. The consensus in the literature seems to be that the technology shows feasibility and acceptability for managing chronic diseases, but more research is still needed on their real-world efficacy. Importantly, these literature reviews highlight several limitations in the literature such as small sample sizes, questionable sample compositions (healthy or convenience samples instead of samples of patient groups), and not controlling for participants' previous experience with voice-based intelligent agents.

In summary, AI chatbots have the potential to provide targeted support and improve the management of various chronic diseases, but only if they are designed to meet the specific needs and preferences of their users. It is essential to understand target users' perspectives, preferences, and experiences of using chatbots for health purposes so that chatbot solutions address the needs of their intended users. Individuals with long-term health conditions often face complex challenges that require ongoing tailored support, while the extant research on using chatbots for health care support provides limited insights into the acceptance (or resistance) among people with chronic conditions. To address the limitations identified in previous studies [11,21], this study gathered a large sample of people with chronic conditions and delved into their past experiences of and future preferences for interacting with AI chatbots.

Methods

Overview

A web-based survey study was conducted between January 1 and March 31, 2023, targeting the diabetes.co.uk user population. The site is the largest web-based community of people with diabetes in Europe with hundreds of thousands of registered users [22]. We chose to target this population because our previous research collaborations showed the community's wide awareness and practice of using technological solutions to self-manage their long-term health conditions (people with diabetes often experience other chronic conditions) [23,24]. For the survey, we defined AI chatbots broadly as computer programs designed to interact in humanlike conversation, and referred to Alexa and Siri on smart devices as examples of AI chatbots. As ChatGPT was launched on November 30, 2022, and quickly gained popularity, public understanding and opinions of chatbots might have changed during our project timeline. Hence, we aimed to capture the survey responses at multiple intervals. Several social media advertisements with the survey link were posted in January, February, and March 2023 on the Facebook page for diabetes.co.uk and clicked on by 1310 people.

As part of a research project funded by an Innovate UK grant, the survey was administered through the Qualtrics software by the digital health company leading the project. The purpose of the study and a consent form were presented on the landing page of the web survey. After the survey was closed, we exported response data from Qualtrics to SPSS v.28 (IBM Corp) for a quality check, data recoding, and variable labeling. We carefully examined the initial data set to remove duplicate records (mainly generated in the survey setup and testing process) and poor-quality responses such as those who completed (or abandoned) the survey in less than 120 seconds. The final data set for analysis contained 888 records. We also used SPSS to assign numeric values to all the nominal variables in the survey (eg, 1=male, 2=female). After the data set was cleaned, we exported it as a .csv file to R version 4.2.3 (The R Foundation for Statistical Computing) for frequency, crosstabulation, and regression tests.

The survey contained 30 questions: 24 closed questions, 2 open questions, and 4 demographic questions. Participant consent was provided at the start of the survey before completion. Quantitative information (closed and multiple-choice questions) was collected on four topic areas: (1) demographic characteristics; (2) long COVID symptoms and clinical diagnoses; (3) health apps, websites, and chatbot use; and (4) opinions about chatbots. The majority of the questions were adapted from the digital health literature (eg, [14,25]). To address the two research objectives, we first asked questions on general attitude and acceptance toward chatbots, such as "How familiar are you with AI chatbots?" and "How likely are you to use a health chatbot within the next 12 months if available?"; then we focus on questions on using chatbots in specific health management scenarios, such as "Do you think chatbots have the capability of delivering accurate medical advice?" and "Would you like a chatbot to understand your stress levels and emotional states?" We also included questions on widely cited factors in the literature that might affect chatbot adoption such as trust and privacy.

Ethical Considerations

The survey instrument, along with other details of the methodology, was approved by the Royal Holloway, University of London (Full-Review-3509-2022-11-18-13-35-UATM024). The participants were presented with a consent page before starting the web-based questionnaire. The purpose and the method of the research were briefly explained, along with an informed consent form asking participants to confirm that they were 18 years or older and that they were voluntarily taking part in the study. The survey did not collect personally identifiable information. The IP addresses were monitored to prevent multiple entries from the same computer. However, all IP addresses were removed from the data when the survey closed. Each survey response was assigned a unique ID, and the encrypted data were stored in the United Kingdom on Microsoft Azure servers. No compensation was provided to participants for completing the survey.

Results

Demographic Characteristics

Of the 888 individuals who started the survey, 729 (82.1%) responded to the sex question, of which 471 (64.6%) were female, 252 (34.6%) were male, and the remaining 6 (0.8%) were nonbinary/third sex or "prefer not to say." Of the 741 respondents who provided their age, 556 (75%) were 55 years or older, with a median age of 63 (IQR 55-70) years. The sample consisted predominantly of White (n=466,64.2%) individuals; other ethnicities were represented in smaller percentages and scattered across different categories of ethnicity (eg, Indian and Pakistani: n=25, 3.5%; Black n=7, 1.2%).

In relation to chronic health conditions, almost half of the 888 respondents reported having type 2 diabetes (n=437, 49.2%). Table 1 provides an overview of the top 10 chronic conditions identified in the survey responses, plus long COVID. Of the 740 individuals who responded to the question "Would you describe yourself as having Long COVID?" 170 (23%) answered yes. While a majority of the respondents utilized health apps (n=500, 73.5%), a much smaller portion (n=259, 38.1%) made use of voice-assisted apps or devices like Amazon Alexa.

Table . Most common chronic health conditions (n=888).

Frequency rank	Condition	Frequency, n (%)
1	Type 2 diabetes	437 (49.2)
2	High blood pressure/hypertension	330 (37.2)
3	Alzheimer's disease	240 (27.0)
4	Long COVID	170 (19.1)
5	Arthritis	195 (22.0)
6	Allergies	187 (21.1)
7	Anxiety	157 (17.7)
8	Depression	137 (15.4)
9	Asthma	114 (12.8)
10	Type 1 diabetes	106 (11.9)
11	Chronic pain	102 (11.5)

Attitudes Toward Health Chatbots

Although the survey was conducted at a time when ChatGPT was beginning to receive wide public attention, a significant number of respondents were "not familiar at all" (n=272, 40.5%) or only "slightly familiar" (n=175, 26%) with AI chatbots. There was an overall hesitancy about using a health chatbot, with less than one-third of respondents (n=203, 30.1%) indicating that they were "somewhat likely" or "very likely" to use a health chatbot in the next 12 months if available.

However, a deeper dive into the survey data reveals a more nuanced picture. There seems to be a great deal of uncertainty among people about AI chatbots' capability of providing accurate medical advice. When asked if they believe chatbots have the capability of providing accurate medical advice, 396 of 677 (58.5%) respondents answered "unsure," while only 77 (11.4%) answered "yes" and 204 (30.1%) chose "no."

On the other hand, people seem to be more open to the idea of chatbots supporting mental well-being: 272 (40.2%) would like a chatbot to understand their stress levels and emotional states, 211 (31.2%) were unsure, and 194 (28.7%) indicated no interest. A further cross-tabulation and χ^2 analysis using the chisq function in R suggests that people with long COVID in our sample were more likely to be interested in an emotionally intelligent chatbot than those without long COVID (n=673; χ^2_2=13.73; *P*=.001), although nearly one-third of the former group were still "unsure" (Table 2).

Table . Cross-tabulation: long COVID chatbot understands stress and emotion.

	Long COVID?		Total (n=673)
	Yes (n=162)	No (n=511)	
Would you like a chatbot to understand your stress and emotional states?			
Yes	81 (50.0)	190 (37.2)	271 (40.3)
No	29 (17.9)	164 (32.1)	193 (28.7)
Unsure	52 (32.1)	157 (30.7)	209 (31.1)
Did not respond	—[a]	—	215 (24.2[b])

[a]Not applicable.

[b]Percentage based on the 888 total respondents.

This "chatbot hesitancy" is also evident when comparing people's preferences of a bot and a real person in various health scenarios. Overall, our survey respondents overwhelmingly prefer to speak to a real person rather than a bot about physical and mental health. However, people seem to not mind speaking with a bot about nutrition and sleep as much or letting it collect symptom data and conduct some preliminary analysis as indicated in Table 3.

Table . People's preferences of a bot and a real person in various health scenarios.

Would you prefer a bot or a real person when...	Bot, n (%)	Person, n (%)	Don't mind, n (%)	Did not respond, n (%)[a]
Speaking about general health (n=600)	18 (3.0)	484 (80.7)	98 (16.3)	288 (32.4)
Speaking about mental health (n=596)	26 (4.4)	483 (81.0)	87 (14.6)	292 (32.9)
Speaking about sleep (n=595)	46 (7.7)	370 (62.2)	179 (30.1)	293 (33.0)
Speaking about nutrition (n=599)	64 (10.7)	343 (57.3)	192 (32.1)	289 (32.5)
Collecting symptoms (n=590)	63 (10.7)	270 (45.8)	257 (43.6)	298 (33.6)
Conducting preliminary analysis (n=596)	72 (12.1)	314 (52.7)	210 (35.2)	292 (32.9)

[a]Percentages in this column are based on the total 888 respondents.

Consistent with the observations above, an encouraging sign for health chatbot developers is that people are willing to try voice-based health chatbots despite the overwhelming hesitance toward bots. Of 679 respondents, 309 (45.5%) expressed willingness to use a voice-based chatbot to record their health symptoms on a mobile device, and 278 (41.1%) would let their voice be analyzed to diagnose health problems. When asked if they would like to trial a voice-based health chatbot that the research team is developing, 364 of 560 (65%) respondents answered "yes," as illustrated in Table 4.

Table . Attitude toward voice-based health chatbot.

	Yes, n (%)	No, n (%)	Unsure, n (%)	Didn't respond, n (%[a])
Would you use your voice to record health symptoms on a mobile device? (n=679)	309 (45.5)	182 (26.8)	188 (27.7)	209 (23.5)
Would you use an app that analyzes your voice to diagnose potential health problems? (n=679)	278 (41.1)	158 (23.3)	241 (35.6)	209 (23.5)
Would you like to trial the voice-based health chatbot we are developing? (n=560)	364 (65.0)	196 (35.0)	—[b]	328 (36.9)

[a]Percentages in this column are based on the total 888 respondents.

[b]Option not provided.

Factors Predicting Health Chatbot Adoption

We ran linear regression analyses in R to explore factors that could predict individuals' likelihood to use a health chatbot in the next 12 months. We categorized the variables into three groups: demographic, experience, and attitudinal. Table 5 presents our findings.

Table . Predictors analysis using regression. Likelihood of adopting a health chatbot in the next 12 months (n=485).[a]

		β	SE	*P* value	95% CI
Demographic					
	Age	–.075	.005	.06	–.021 to .000
	Sex	*.078* [b]	*.097*	*.03*	*.039 to .460*
	Long COVID?	–.064	.115	.08	–.428 to .023
Experience					
	Familiarity with AI[c] chatbot	*.169*	*.048*	*<.001*	*.110 to .302*
	Frequency of online health information seeking	*.198*	*.012*	*<.001*	*.039 to .084*
Attitudinal					
	Comfortable with reporting symptoms to health chatbot	*.228*	*.060*	*<.001*	*.142 to .377*
	Worry about privacy using health chatbot	–.032	.042	.40	–.128 to .042
	Trust health chatbot	*.255*	*.069*	*<.001*	*.219 to .489*

[a]Likelihood of adoption was measured on a 5-point Likert scale (1=extremely unlikely and 5=extremely likely). R^2=0.388.

[b]Italics indicate statistical significance.

[c]AI: artificial intelligence.

We considered *P* values <.05 as statistically significant in our regression analysis results. The results showed that age (β=–.075; *P*=.06) and long COVID status (β=−.064; *P*=.08) have little to do with participants' tendency to use a health chatbot. Sex, dummy-coded as male=1 and female=2 with other sex categories excluded from this analysis due to a small number in each of the categories, seems to have a marginal effect (β=.078; *P*=.03), with female respondents potentially being more inclined to adopt a health chatbot than male respondents. Past experience with an AI chatbot ("familiarity with AI chatbots"; β=.169; *P*<.001) and online health information-seeking frequency (aggregated frequencies across Google, social media, and professional health sites; β=.198; *P*<.001) show strong associations with chatbot adoption likelihood. Similarly, two attitudinal items measuring a person's comfort in outlining symptoms to a health chatbot (β=.228; *P*<.001) and their trust in a chatbot for advice (β=.255; *P*<.001) also strongly predicted their likelihood of adopting a health chatbot. Interestingly, privacy concerns, despite being widely reported in the academic literature as a deterrent to chatbot or virtual assistant adoption [20,26], did not seem to affect the likelihood of survey respondents adopting a health chatbot (β=−.032; *P*=.40).

Discussion

Principal Findings

The survey results summarized above present a nuanced portrayal of public attitudes toward health care chatbots. The findings indicate that trust continues to be a crucial element in predicting people's inclination to embrace health chatbots, aligning with prior research on user acceptance of digital health technologies [11,27,28]. It seems that for our sample of predominantly female adults older than 55 years, most of them do not trust a chatbot to provide accurate diagnosis and professional medical advice. This echoes findings in previous studies [11,29] that while patients were generally receptive to the use of AI chatbots in health care, they had concerns about the accuracy of information provided and the ability of chatbots to understand complex medical issues. Although past experience with chatbots and online information seeking are strong indicators of respondents' future adoption of health chatbots, they are in general skeptical or unsure about the use of AI chatbots for health care purposes. Because of this "AI hesitancy" [14], it is unsurprising that most people show an overall preference for a real person (clinician) over a chatbot in health care encounters.

On the other hand, this study contributes fresh insights into overcoming AI hesitancy and the potential use of AI chatbots in supporting long-term health conditions like long COVID. A key finding from the survey is that AI hesitancy is not uniform across all health domains and user groups. A significant proportion of survey respondents expressed willingness to engage with a health chatbot regarding nutrition and sleep, as well as allowing it to collect symptom data. Furthermore, although doubts about the medical capabilities of AI chatbots persist, people are more receptive to utilizing them for stress detection and emotional enhancement. Notably, individuals with long COVID in our sample exhibited a particular interest in emotionally intelligent chatbots, highlighting the mental health needs of those with long COVID and the potential of using conversational agents as an intervention [30]. It is also surprising that privacy concerns did not correlate with the likelihood of health chatbot adoption in our study, a finding contrary to conclusions in many previous studies [20,31]. Leveraging these positive attitudes toward AI chatbots could enhance public familiarity and increase the likelihood of chatbot adoption for health care purposes, as evidenced by our regression analysis. For instance, an AI chatbot focusing on lifestyle or emotional well-being could pave the way for broader acceptance of health chatbots that are reliable and highly personalized [32].

This research also provides preliminary insights into the potential of voice-based interaction with health chatbots. Despite the popularity of voice-based AI agents such as Siri and Alexa on smart devices, there are only a handful of academic studies on the public's attitude toward an alternative voice-based interface for health chatbots [11,21]. Traditional text-based chatbots on mobile phones present challenges to older adults in terms of vision and dexterity, as typing on a smartphone can be difficult and the screen size is often too small for them [33]. From the technology acceptance research, we understand that the usability of a health chatbot plays a role in its perceived usefulness [32]. Therefore, the relatively high acceptability and enthusiasm toward voice-based health chatbots expressed in our survey responses indicate a potential avenue for reaching a wider, often neglected population of adults older than 55 years. In addition to the usability benefits, voice input can be captured and analyzed for symptom tracking and medical diagnosis [34], complementing other data inputs to enable a more accurate assessment of the user's health.

Limitations

It is noteworthy that this study had several limitations. The use of a web-based survey for the empirical study introduces the potential for response bias. The sample primarily consisted of female adults older than 55 years with diabetes and other long-term health conditions, rather than representing the general population. As this population may have been more attuned to ongoing health concerns, they might have been more prone to reporting long COVID symptoms, potentially explaining the significantly higher proportion of those with long COVID in our sample compared to the national data from the National Health Service. The UK COVID-19 Infection National Survey (monthly, terminated in March 2023) reported that around 3% of the UK population were experiencing symptoms 4 weeks after they first had COVID-19 [35].

As cross-sectional survey research, this study is unable to provide a deep understanding of attitudes and opinions. For example, we do not know exactly why people are more receptive to voice-based AI chatbots aside from an educated guess that a voice interface may be easier to use and more natural than typed text for adults older than 55 years. Future research based on in-depth interviews or experimental methods might help unpack these user attitudes reported in the survey and identify possible causal factors.

Conclusion

With the rapid development of AI and chatbot technologies, the utilization of chatbots in health care is primed for substantial expansion in the forthcoming years. The potential benefits offered by these technologies, such as enhanced health care accessibility [4,36], cost reduction [15], and improved patient outcomes [37], are too substantial to disregard. However, health care providers and technology developers must acknowledge AI hesitancy [14] among patients and ensure the inclusive and effective utilization of AI chatbots.

This study contributes valuable insights into the acceptability of health chatbots among a population identified as requiring continuous long-term health care support, such as individuals at high risk for conditions like long COVID [38]. First, we augment the evidence in the literature that there exists a general skepticism toward health chatbots among people with chronic diseases. Second, notwithstanding this persistent AI hesitancy, we found that individuals are more receptive toward chatbots supporting lifestyle enhancement than those aiding disease diagnosis and health care management. Third, compared to other subgroups in the study, those with long COVID were more amenable to using chatbots for emotional support. Lastly, while popular AI chatbots on the market are text-based, this study demonstrates that individuals with chronic conditions exhibit a high interest in voice-based conversational agents despite their general AI hesitancy. Moving forward, it is paramount to continue exploring the potential applications of health chatbots in addressing the unique needs of specific patient populations, including those with chronic health conditions.

Acknowledgments

This research was funded by Innovate UK Smart Grants (grant 10032047).

Data Availability

The original data set supporting this study cannot be made publicly available due to data protection considerations. Partial, anonymized, or aggregated data may be available from the corresponding author upon reasonable request.

Conflicts of Interest

CS, AP, and AK are senior executives of the digital health company that implemented the empirical study.

References

1. Montenegro JLZ, da Costa CA, da Rosa Righi R. Survey of conversational agents in health. Expert Syst Applications 2019 Sep 1;129:56-67. [doi: 10.1016/j.eswa.2019.03.054]
2. Almalki M, Azeez F. Health chatbots for fighting COVID-19: a scoping review. Acta Inform Med 2020 Dec;28(4):241-247. [doi: 10.5455/aim.2020.28.241-247] [Medline: 33627924]
3. Battineni G, Chintalapudi N, Amenta F. AI chatbot design during an epidemic like the novel coronavirus. Healthcare (Basel) 2020 Jun 3;8(2):154. [doi: 10.3390/healthcare8020154] [Medline: 32503298]
4. Bharti U, Bajaj D, Batra H, Lalit S, Lalit S, Gangwani A. Medbot: conversational artificial intelligence powered chatbot for delivering tele-health after COVID-19. Presented at: 2020 5th International Conference on Communication and Electronics Systems; Jun 10-12, 2020; Coimbatore, India. [doi: 10.1109/ICCES48766.2020.9137944]
5. Chaix B, Bibault JE, Pienkowski A, et al. When chatbots meet patients: one-year prospective study of conversations between patients with breast cancer and a chatbot. JMIR Cancer 2019 May 2;5(1):e12856. [doi: 10.2196/12856] [Medline: 31045505]
6. Gong E, Baptista S, Russell A, et al. My Diabetes Coach, a mobile app-based interactive conversational agent to support type 2 diabetes self-management: randomized effectiveness-implementation trial. J Med Internet Res 2020 Nov 5;22(11):e20322. [doi: 10.2196/20322] [Medline: 33151154]
7. Oh J, Jang S, Kim H, Kim JJ. Efficacy of mobile app-based interactive cognitive behavioral therapy using a chatbot for panic disorder. Int J Med Inform 2020 Aug;140:104171. [doi: 10.1016/j.ijmedinf.2020.104171] [Medline: 32446158]
8. Kowatsch T, Schachner T, Harperink S, et al. Conversational agents as mediating social actors in chronic disease management involving health care professionals, patients, and family members: multisite single-arm feasibility study. J Med Internet Res 2021 Feb 17;23(2):e25060. [doi: 10.2196/25060] [Medline: 33484114]
9. Griffin AC, Xing Z, Khairat S, et al. Conversational agents for chronic disease self-management: a systematic review. AMIA Annu Symp Proc 2021 Jan 25;2020:504-513. [Medline: 33936424]
10. Schachner T, Keller R, Wangenheim FV. Artificial intelligence-based conversational agents for chronic conditions: systematic literature review. J Med Internet Res 2020 Sep 14;22(9):e20701. [doi: 10.2196/20701] [Medline: 32924957]
11. Wilson L, Marasoiu M. The development and use of chatbots in public health: scoping review. JMIR Hum Factors 2022 Oct 5;9(4):e35882. [doi: 10.2196/35882] [Medline: 36197708]
12. Chaves AP, Gerosa MA. How should my chatbot interact? A survey on social characteristics in human–chatbot interaction design. Int J Hum Computer Interaction 2021 May 9;37(8):729-758. [doi: 10.1080/10447318.2020.1841438]
13. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (WYSA) for digital mental well-being: real-world data evaluation mixed-methods study. JMIR Mhealth Uhealth 2018 Nov 23;6(11):e12106. [doi: 10.2196/12106] [Medline: 30470676]
14. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: a mixed-methods study. Digit Health 2019 Aug 21;5:2055207619871808. [doi: 10.1177/2055207619871808] [Medline: 31467682]
15. Laranjo L, Dunn AG, Tong HL, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 1;25(9):1248-1258. [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
16. Bickmore TW, Caruso L, Clough-Gorr K, Heeren T. 'It's just like you talk to a friend' relational agents for older adults. Interacting Comput 2005 Dec;17(6):711-735. [doi: 10.1016/j.intcom.2005.09.002]

17. Crutzen R, Peters GJY, Portugal SD, Fisser EM, Grolleman JJ. An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: an exploratory study. J Adolesc Health 2011 May;48(5):514-519. [doi: 10.1016/j.jadohealth.2010.09.002] [Medline: 21501812]
18. Brown JEH, Halpern J. AI chatbots cannot replace human interactions in the pursuit of more inclusive mental healthcare. SSM Ment Health 2021 Dec;1:100017. [doi: 10.1016/j.ssmmh.2021.100017] [Medline: 34494013]
19. Sweeney C, Potts C, Ennis E, et al. Can chatbots help support a person's mental health? Perceptions and views from mental healthcare professionals and experts. ACM Trans Computing Healthcare 2021 Jul 15;2(3):1-15. [doi: 10.1145/3453175]
20. Ermolina A, Tiberius V. Voice-controlled intelligent personal assistants in health care: international Delphi study. J Med Internet Res 2021 Apr 9;23(4):e25312. [doi: 10.2196/25312] [Medline: 33835032]
21. Bérubé C, Schachner T, Keller R, et al. Voice-based conversational agents for the prevention and management of chronic and mental health conditions: systematic literature review. J Med Internet Res 2021 Mar 29;23(3):e25933. [doi: 10.2196/25933] [Medline: 33658174]
22. About Diabetes.co.uk. Diabetes.co.uk. URL: https://www.diabetes.co.uk/about.html [accessed 2024-02-19]
23. Summers C, Wu P, Taylor AJG. Supporting mental health during the COVID-19 pandemic using a digital behavior change intervention: an open-label, single-arm, pre-post intervention study. JMIR Form Res 2021 Oct 6;5(10):e31273. [doi: 10.2196/31273] [Medline: 34459740]
24. Bernardi R, Wu PF. Online health communities and the patient-doctor relationship: an institutional logics perspective. Soc Sci Med 2022 Dec;314:115494. [doi: 10.1016/j.socscimed.2022.115494] [Medline: 36334494]
25. Chi OH, Jia S, Li Y, Gursoy D. Developing a formative scale to measure consumers' trust toward interaction with artificially intelligent (AI) social robots in service delivery. Comput Hum Behav 2021 May;118:106700. [doi: 10.1016/j.chb.2021.106700]
26. May R, Denecke K. Security, privacy, and healthcare-related conversational agents: a scoping review. Inform Health Soc Care 2022 Apr 3;47(2):194-210. [doi: 10.1080/17538157.2021.1983578] [Medline: 34617857]
27. Jermutus E, Kneale D, Thomas J, Michie S. Influences on user trust in healthcare artificial intelligence: a systematic review. Wellcome Open Res 2022;7:65. [doi: 10.12688/wellcomeopenres.17550.1]
28. Seitz L, Bekmeier-Feuerhahn S, Gohil K. Can we trust a chatbot like a physician? A qualitative study on understanding the emergence of trust toward diagnostic chatbots. Int J Hum Computer Stud 2022 Sep;165:102848. [doi: 10.1016/j.ijhcs.2022.102848]
29. Young AT, Amara D, Bhattacharya A, Wei ML. Patient and general public attitudes towards clinical artificial intelligence: a mixed methods systematic review. Lancet Digit Health 2021 Sep;3(9):e599-e611. [doi: 10.1016/S2589-7500(21)00132-1] [Medline: 34446266]
30. He Y, Yang L, Qian C, et al. Conversational agent interventions for mental health problems: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2023 Apr 28;25(1):e43862. [doi: 10.2196/43862] [Medline: 37115595]
31. Coghlan S, Leins K, Sheldrick S, Cheong M, Gooding P, D'Alfonso S. To chat or bot to chat: ethical issues with using chatbots in mental health. Digit Health 2023 Jun 22;9:20552076231183542. [doi: 10.1177/20552076231183542] [Medline: 37377565]
32. Abd-Alrazaq AA, Alajlani M, Ali N, Denecke K, Bewick BM, Househ M. Perceptions and opinions of patients about mental health chatbots: scoping review. J Med Internet Res 2021 Jan 13;23(1):e17828. [doi: 10.2196/17828] [Medline: 33439133]
33. Gudala M, Ross MET, Mogalla S, Lyons M, Ramaswamy P, Roberts K. Benefits of, barriers to, and needs for an artificial intelligence-powered medication information voice chatbot for older adults: interview study with geriatrics experts. JMIR Aging 2022 Apr 28;5(2):e32169. [doi: 10.2196/32169] [Medline: 35482367]
34. Chen Z, Li M, Wang R, et al. Diagnosis of COVID-19 via acoustic analysis and artificial intelligence by monitoring breath sounds on smartphones. J Biomed Inform 2022 Jun;130:104078. [doi: 10.1016/j.jbi.2022.104078] [Medline: 35489595]
35. Prevalence of ongoing symptoms following coronavirus (COVID-19) infection in the UK: 30 March 2023. Office for National Statistics. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/prevalenceofongoingsymptomsfollowingcoronaviruscovid19infectionintheuk/30march2023 [accessed 2024-02-12]
36. Habicht J, Viswanathan S, Carrington B, Hauser TU, Harper R, Rollwage M. Closing the accessibility gap to mental health treatment with a personalized self-referral chatbot. Nat Med 2024 Feb;30(2):595-602. [doi: 10.1038/s41591-023-02766-x] [Medline: 38317020]
37. Aggarwal A, Tam CC, Wu D, Li X, Qiao S. Artificial intelligence–based chatbots for promoting health behavioral changes: systematic review. J Med Internet Res 2023 Feb 24;25(1):e40789. [doi: 10.2196/40789] [Medline: 36826990]
38. Crook H, Raza S, Nowell J, Young M, Edison P. Long covid-mechanisms, risk factors, and management. BMJ 2021 Jul 26;374:n1648. [doi: 10.1136/bmj.n1648] [Medline: 34312178]

Abbreviations

AI: artificial intelligence

Edited by A Kushniruk; submitted 20.07.23; peer-reviewed by E Vashishtha, I Degen, L Zhu; revised version received 22.02.24; accepted 09.05.24; published 23.07.24.

Please cite as:
Wu PF, Summers C, Panesar A, Kaura A, Zhang L
AI Hesitancy and Acceptability—Perceptions of AI Chatbots for Chronic Health Management and Long COVID Support: Survey Study
JMIR Hum Factors 2024;11:e51086
URL: https://humanfactors.jmir.org/2024/1/e51086
doi:10.2196/51086

Investigating Users' Attitudes Toward Automated Smartwatch Cardiac Arrest Detection: Cross-Sectional Survey Study

Wisse M F van den Beuken[1], MD; Hans van Schuppen[1,2], MD, PhD; Derya Demirtas[3,4], MMath, PhD; Vokko P van Halm[5], MD, PhD; Patrick van der Geest[6,7], MD, PhD; Stephan A Loer[1], MD, PhD; Lothar A Schwarte[1,2], MD, PhD; Patrick Schober[1,2], MD, PhD

[1]Department of Anesthesiology, Amsterdam UMC, Amsterdam, Netherlands
[2]Helicopter Emergency Medical Service Lifeliner 1, Amsterdam, Netherlands
[3]Center for Healthcare Operations Improvement and Research, University of Twente, Enschede, Netherlands
[4]Industrial Engineering and Business Information Systems, University of Twente, Enschede, Netherlands
[5]Department of Cardiology, Amsterdam UMC, Amsterdam, Netherlands
[6]Ambulance Rotterdam-Rijnmond, Barendrecht, Netherlands
[7]Department of Intensive Care Medicine, Spaarne Gasthuis, Haarlem, Netherlands

Corresponding Author:
Patrick Schober, MD, PhD
Department of Anesthesiology, Amsterdam UMC, , Amsterdam, , Netherlands

Abstract

Background: Out-of-hospital cardiac arrest (OHCA) is a leading cause of mortality in the developed world. Timely detection of cardiac arrest and prompt activation of emergency medical services (EMS) are essential, yet challenging. Automated cardiac arrest detection using sensor signals from smartwatches has the potential to shorten the interval between cardiac arrest and activation of EMS, thereby increasing the likelihood of survival.

Objective: This cross-sectional survey study aims to investigate users' perspectives on aspects of continuous monitoring such as privacy and data protection, as well as other implications, and to collect insights into their attitudes toward the technology.

Methods: We conducted a cross-sectional web-based survey in the Netherlands among 2 groups of potential users of automated cardiac arrest technology: consumers who already own a smartwatch and patients at risk of cardiac arrest. Surveys primarily consisted of closed-ended questions with some additional open-ended questions to provide supplementary insight. The quantitative data were analyzed descriptively, and a content analysis of the open-ended questions was conducted.

Results: In the consumer group (n=1005), 90.2% (n=906; 95% CI 88.1%-91.9%) of participants expressed an interest in the technology, and 89% (n=1196; 95% CI 87.3%-90.7%) of the patient group (n=1344) showed interest. More than 75% (consumer group: n= 756; patient group: n=1004) of the participants in both groups indicated they were willing to use the technology. The main concerns raised by participants regarding the technology included privacy, data protection, reliability, and accessibility.

Conclusions: The vast majority of potential users expressed a strong interest in and positive attitude toward automated cardiac arrest detection using smartwatch technology. However, a number of concerns were identified, which should be addressed in the development and implementation process to optimize acceptance and effectiveness of the technology.

(JMIR Hum Factors 2024;11:e57574) doi:10.2196/57574

KEYWORDS

out-of-hospital cardiac arrest; wearables; wearable; digital health; smartwatch; automated cardiac arrest detection; emergency medicine; emergency; cardiology; heart; cardiac; cross sectional; survey; surveys; questionnaire; questionnaires; experience; experiences; attitude; attitudes; opinion; perception; perceptions; perspective; perspectives; acceptance; adoption; willingness; intent; intention

Introduction

Out-of-hospital cardiac arrest (OHCA) is a leading cause of mortality in the developed world [1,2]. The chain of survival after OHCA starts with timely detection of the cardiac arrest and prompt activation of emergency medical services (EMS) [3,4]. While this step is crucial to prevent the death of the patient, it is also the most fragile link in the chain of survival because it requires the presence of a witness. Reliance on bystander activation often introduces a significant delay that markedly reduces the chances of survival [5-7].

Several research groups, including our own, are developing a possible technical solution to automatically detect OHCA and to automate the activation of EMS using wearables and smart devices [8-10]. Smartwatches are computing devices that resemble wristwatches, with functionalities comparable to smartphones. Smartwatches are one of the most prevalent wearable technologies [11] and incorporate a wide array of sensors. Among these sensors are GPS and photoplethysmography (PPG). GPS is a global navigation satellite system that provides location, velocity, and time. This could potentially be used to track the location of patients experiencing cardiac arrest. PPG is used to detect changes in light absorption due to pulsatile blood flow [12] and hence can be used to measure the heartbeat, allowing smartwatches to accurately detect cardiac arrhythmias [13-15]. PPG and other sensors integrated in smartwatches could also potentially be used to detect cardiac arrest by measuring the cessation of pulsatile blood flow.

Such a technical solution, which is currently being developed, should be well aligned with the needs of its potential users in order to enable successful implementation [8]. We therefore aim to investigate users' perspectives on aspects of continuous monitoring such as privacy and data protection, as well as other implications, and to collect insights into their attitude toward the technology.

Methods

Study Design

We conducted a cross-sectional web-based survey aiming to investigate the perceptions and attitudes of potential users toward automated cardiac arrest diagnosis using smartwatches. We identified 2 groups of potential users: the first was consumers who already own a smartwatch (the consumer group), as these individuals could instantly make use of this technology as soon as it becomes available. The second was patients at increased risk of experiencing cardiac arrest (the patient group), as these individuals can potentially benefit the most from using this technology. A survey was deployed among both groups as described below, and participants had up to 4 weeks to respond. Data were collected between October 27, 2022, and March 17, 2023.

Ethical Considerations

The study was assessed by the Medical Ethics Review Committee of VU University Medical Center (2022.0544), which declared on November 29, 2022, that the study was not subject to the Medical Research Involving Human Subjects Act (WMO), such that formal approval was not required. Participants received written information about the purpose of the research and consented to the use of their provided answers for research purposes. Participants in the consumer group were compensated with €0.30 to €0.50 (US $0.32 to $0.53) for their participation. Participants in the patient group did not receive any compensation. The data collected from both groups were completely anonymous.

Participant Recruitment

Consumer Group

The consumers were recruited by a leading Dutch market research agency, Markteffect. This agency has several panels comprising approximately 225,000 consumers from different sectors in the Netherlands. For our survey, we recruited consumers who were aged 18 years or older and owned a smartwatch.

Patient Group

The at-risk patients were recruited through the health panel of the Netherlands Patient Federation (NPF). This health panel comprises approximately 23,000 individuals from different patient associations. All patients had to be aged at least 18 years and at increased risk for cardiac arrest based on self-reported medical history and comorbidities. Patients were considered at increased risk of cardiac arrest if they had 1 or more of the following: cardiovascular disease (eg, hypertension, heart failure, angina pectoris, myocardial infarction), severe renal insufficiencies, diabetes mellitus, cerebrovascular accident, or severe pulmonary disease (eg, chronic obstructive pulmonary disease or lung emphysema).

Sample Size Considerations

An a priori sample size analysis revealed that a minimum sample size of 385 participants would be needed to attain a margin of error of no more than 5% at a 95% confidence level. To attain an even higher precision while still allowing for dropouts and subgroup analyses, we targeted approximately 1000 participants per user group.

Survey Development and Data Collection

The surveys were developed by WMFvdB and PS in collaboration with the research experts from Markteffect and the NPF. Both surveys underwent a comprehensive review and testing by members of our research group and Markteffect or the NPF, respectively, and the surveys can be found in Multimedia Appendix 1. Both surveys largely comprised the same introductory text and questions, but slight modifications were made to tailor the surveys to each group. The surveys included an introductory text explaining the prevalence of witnessed and unwitnessed OHCAs in the Netherlands, emphasizing the need for an alerting system to ensure prompt intervention in case of cardiac arrest. We explained that we are currently developing a novel technology that could potentially address this issue, as it is capable of automatically diagnosing cardiac arrest using smartwatches. We underscored our interest in obtaining their opinions and perspectives on this technology. If participants agreed to participate by continuing to the online survey, we would proceed to ask questions regarding automated cardiac arrest diagnosis using smartwatches. In addition, we asked questions about gender, age, income, and other demographic characteristics. The questions primarily consisted of closed-ended questions and included 5-point Likert-scale questions to assess their agreement with specific aspects of the technology. Limited open-ended questions were provided to enable supplementary insights. Markteffect hosted and distributed the online survey to the consumer group using Collecthor (Collecthor BV). The online survey sent to the patient group was created and hosted Castor Electronic Data Capture (Ciwit BV). The NPF distributed the survey to the patients who were at increased risk of cardiac arrest and had indicated interest to participate in the survey.

Data Screening and Statistical Analysis

Markteffect cleaned the consumer data set, removing surveys with any of the following characteristics, according to their internal standard operating procedures: (1) speeders—surveys that were completed within an exceptionally implausibly fast timeframe; (2) double respondents—multiple surveys submitted by the same individual; (3) straight liners—participants who consistently selected the same answer choice (eg, always chose a neutral response or always chose the first option) or did not answer the questions; and (4) surveys with consistent unintelligible language. The cleaned data set was provided for further analysis. The data from at-risk patients were cleaned by removing completely empty surveys, speeders, and surveys in which less than 50% of the questions were completed.

Quantitative data were analyzed descriptively. Measures of central tendency, measures of variation, and measures of distribution were calculated [16]. The analyses were performed using R (version 4.2.1) and RStudio (version 2022.2.3.492; R Foundation for Statistical Computing).

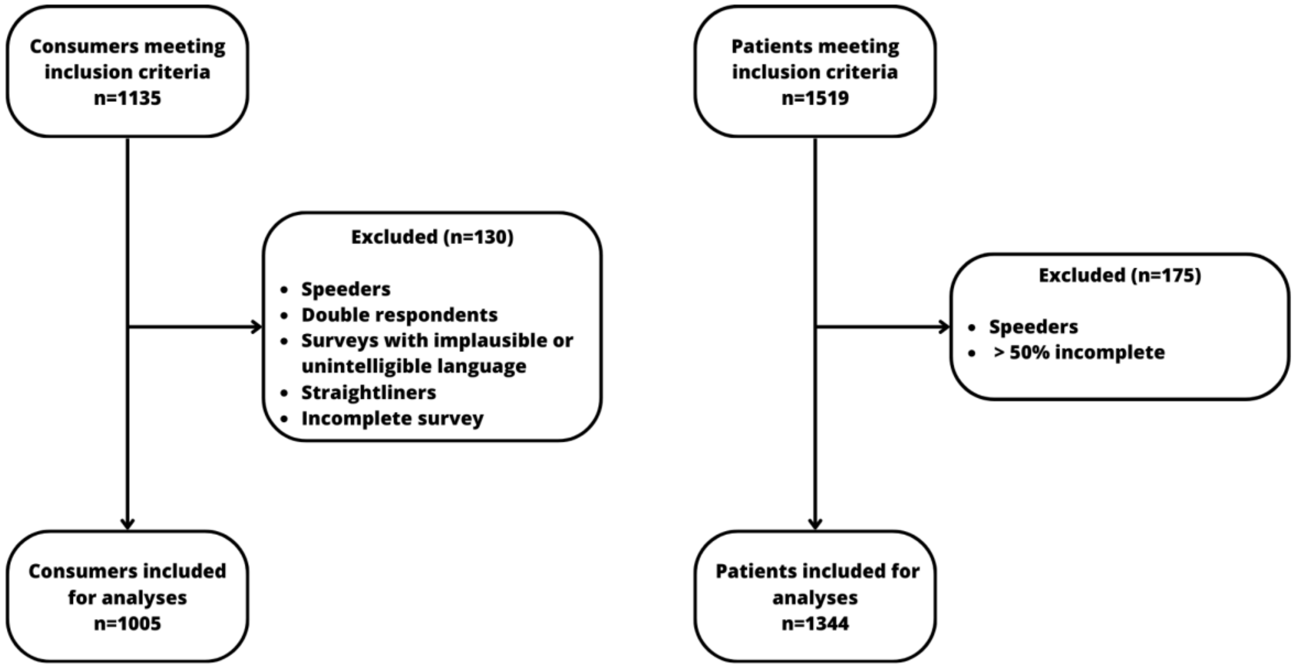
A content analysis of the open-ended questions was conducted to provide deeper insights into the participants' perspectives. WMFvdB created a coding framework consisting of inductive and deductive codes. The coding was reviewed by PS and any disagreements were resolved in consensus. The codes were categorized to identify emerging themes. The content analysis was performed using MAXQDA 2022 (VERBI Software).

Results

Quantitative Analyses

In the consumer group and the patient group, 1135 and 1519 participants, respectively, met the inclusion criteria. In both groups, 88.5% (n=1005 and n=1344, respectively) of the participants were included for analysis (Figure 1).

Figure 1. Flowchart of participant inclusion.



The mean age in the consumer group was 48.6 (SD 13.2) years, and 48.7% (n=489) of the population was male. In the patient group, the mean age was 67.7 (SD 10.0) years, and 56.4% (n=750) of the population was male. Regarding level of education, smartwatch preference, and ethnic background, both groups followed a similar pattern. In the consumer and patient groups, 58.5% (n=588) and 56.1% (n=731), respectively, had a higher level of education. Apple and Samsung smartwatches were most prevalent, and the majority of the population had a Dutch background (Table 1). Both groups had a similar geographical distribution throughout the Netherlands as the general Dutch population (Figure 2).

Table . Population characteristics.

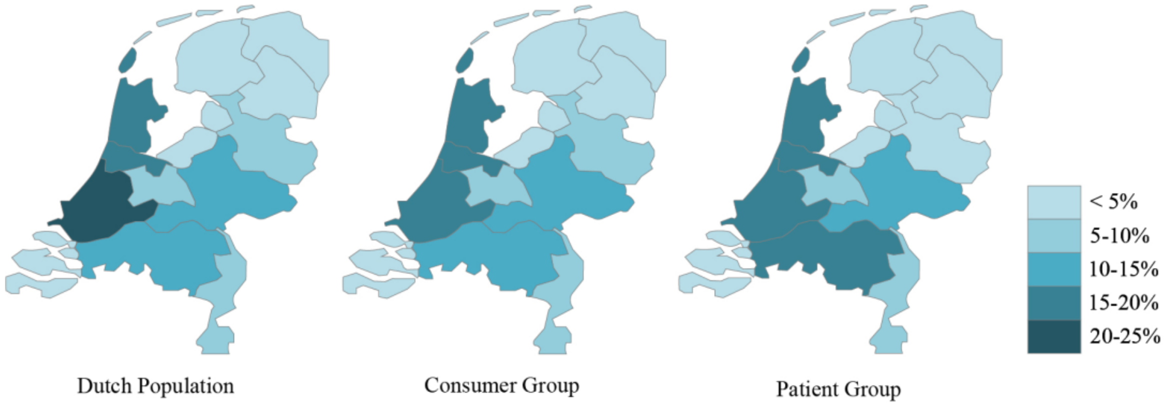
		Consumers (n=1005)	Patients (n=1344)	Dutch population[a] (n=17,475,415)
Age (years), mean (SD)		48.6 (13.2)	67.7 (10.0)	42.3 (N/A[b])
Male sex, n (%)		489 (48.7)	750 (56.4)	8,686,536 (49.7)
Higher level of education, n (%)		588 (58.5)	731 (56.1)	6,203,772 (35.5)
Smartwatch brands n (%)				
	Apple	258 (25.7)	112 (25.5)	N/A
	Samsung	252 (25.1)	107(24.4)	N/A
	Garmin	144 (14.3)	48 (10.9)	N/A
	Fitbit	127 (12.6)	63 (14.4)	N/A
	Other	224 (22.3)	120 (27.3)	N/A
Household income [c] , n (%)				
	<€1600 per month	50 (5)	93 (7.5)	N/A
	€1600-€2600 per month	97 (9.7)	204 (16.5)	N/A
	€2600-€3000 per per month	119 (11.8)	211 (17.1)	N/A
	€3000-€4000 per month	228 (22.7)	229 (18.5)	N/A
	€4000-€8000 per month	233 (23.2)	227 (18.4)	N/A
	>€8000 per month	64 (6.4)	45 (3.6)	N/A
	I do not know/I prefer not to state	214 (21.2)	226 (18.4)	N/A
Region of origin n (%)				
	Dutch	947 (94.2)	1164 (96.1)	N/A
	Western	23 (2.3)	24 (2)	N/A
	Non-Western	24 (2.4)	10 (0.8)	N/A
	I prefer not to state	11 (1.1)	10 (0.8)	N/A
Medical history, n (%)				
	High blood pressure	N/A	864 (64.3)	N/A
	Diabetes mellitus	N/A	324 (24.1)	N/A
	Cardiovascular disease	N/A	752 (56)	N/A
	Severe renal disease	N/A	59 (4.4)	N/A
	Cerebrovascular accident	N/A	109 (8.1)	N/A
	Severe lung disease	N/A	274 (20.4)	N/A
	History of cardiac arrest	N/A	78 (6.2)	N/A
	Other	N/A	204 (15.2)	N/A

[a]Data from 2021, acquired from the Central Bureau of Statistics (CBS) in the Netherlands [17,18].

[b]N/A: not applicable.

[c]In 2021, according to CBS, the median primary income of a household in the Netherlands was €3525 per month [19]. An exchange rate of €1=US $1.06054 applied at the time of the study.

Figure 2. Population density in percentages. Data for the Dutch population for 2021 comes from the Central Bureau of Statistics [18].



Both groups expressed interest in the technology; 90.2% (n=906; 95% CI 88.1%-91.9%) of the consumer group and 89% (n=1196; 95% CI 87.3%-90.7%) of the patient group considered the technology as "interesting" or "very interesting" (Figure 3). Moreover, 75.2% (n=756; 95% CI 72.5%-77.9%) of the participants in the consumer group and 77.6% (n=1004; 95% CI 75.3%-79.8%) of the patient group indicated their willingness to use the technology (Figure 4). The most frequently cited reason for abstaining from or expressing uncertainty about adopting the technology in both groups was "not wanting to be resuscitated." The second most common reason given by the consumers was cultural or religious objections, whereas none of the patients provided this as a reason for expressing uncertainty or abstaining from using the technology. In the patient group, 31.7% (n=92) of participants who had indicated abstaining from or expressed uncertainty about adopting the technology stated not wanting to use a smartwatch as a reason (Figure 5).

Figure 3. Participants' interest in the technology according to a 5-point Likert scale.

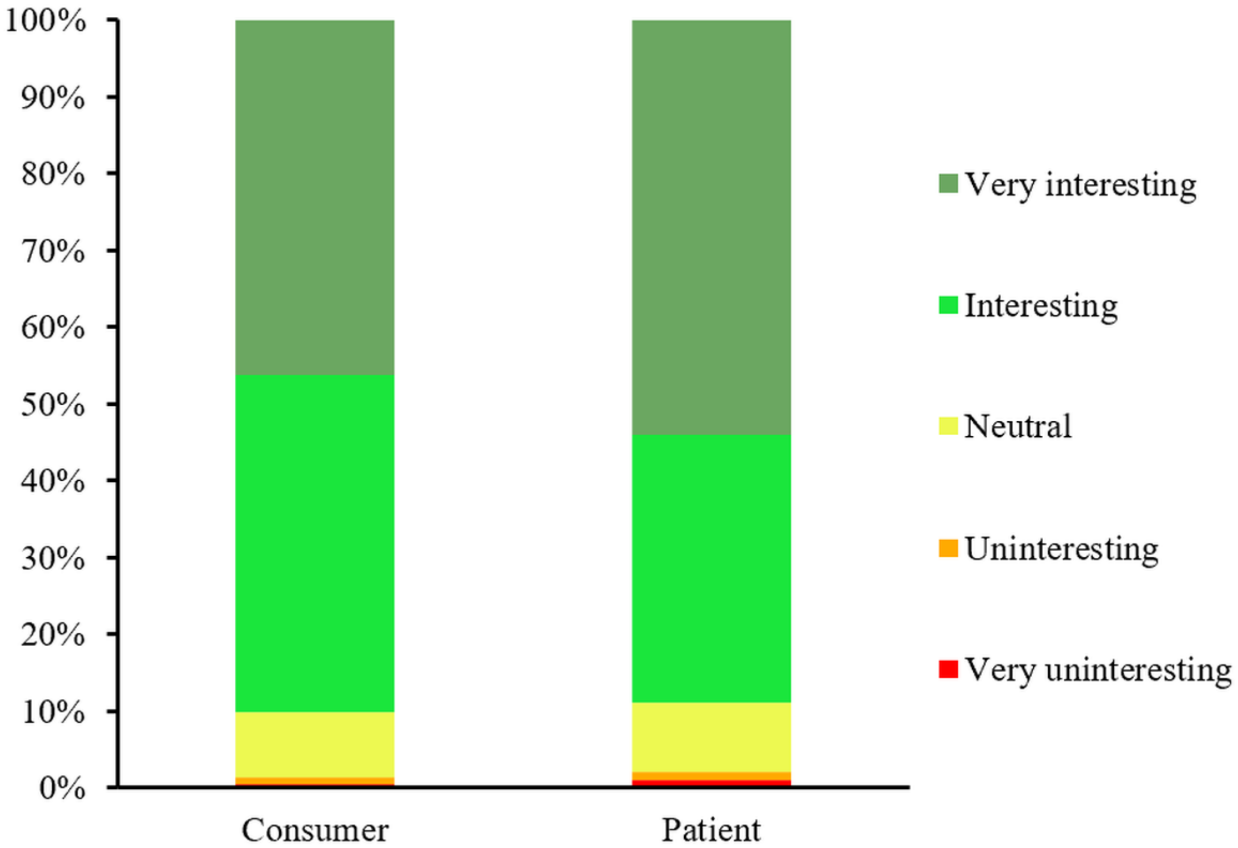


Figure 4. Participants' willingness to use the technology. Error bars represent 95% CIs. Consumer group n=1005, patient group n=1294.

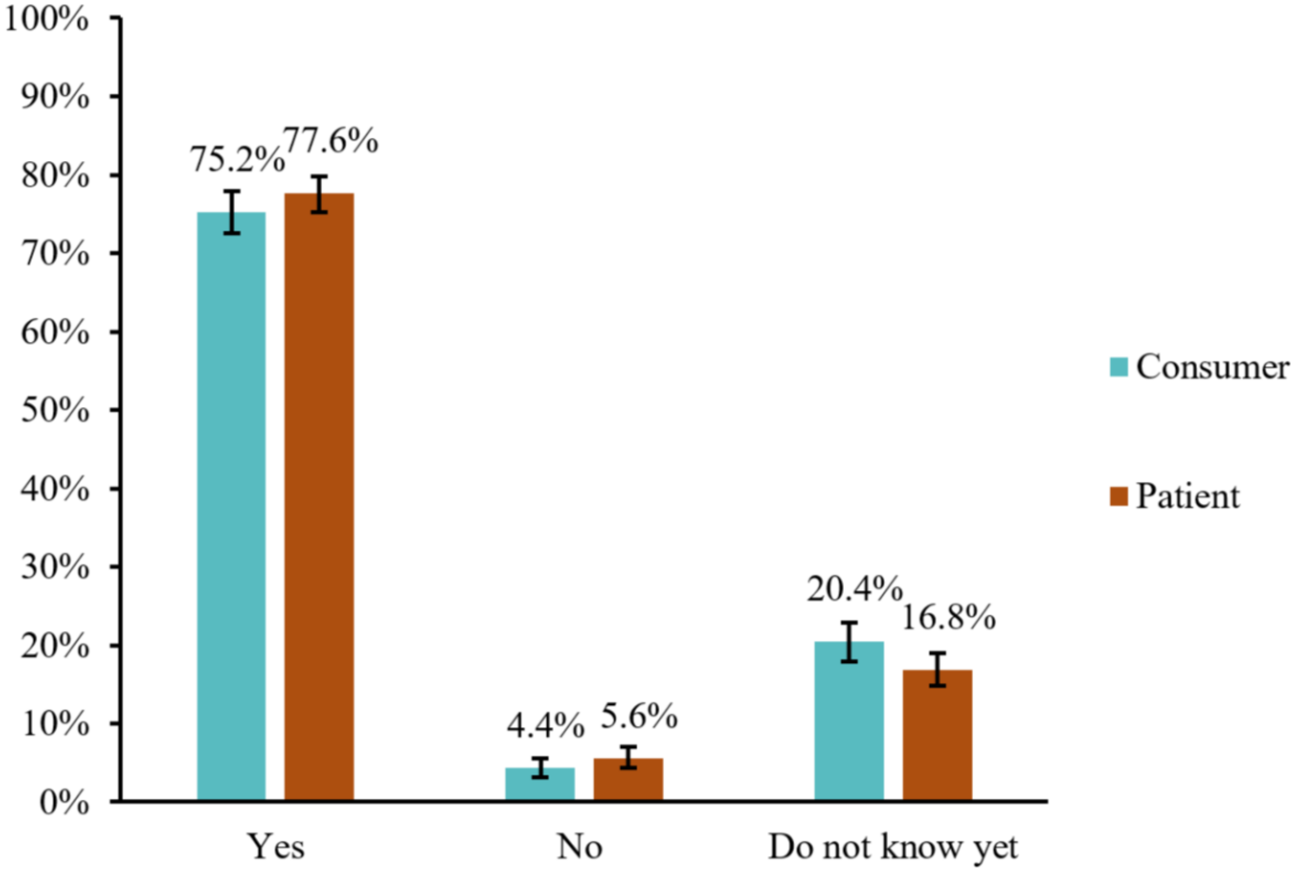
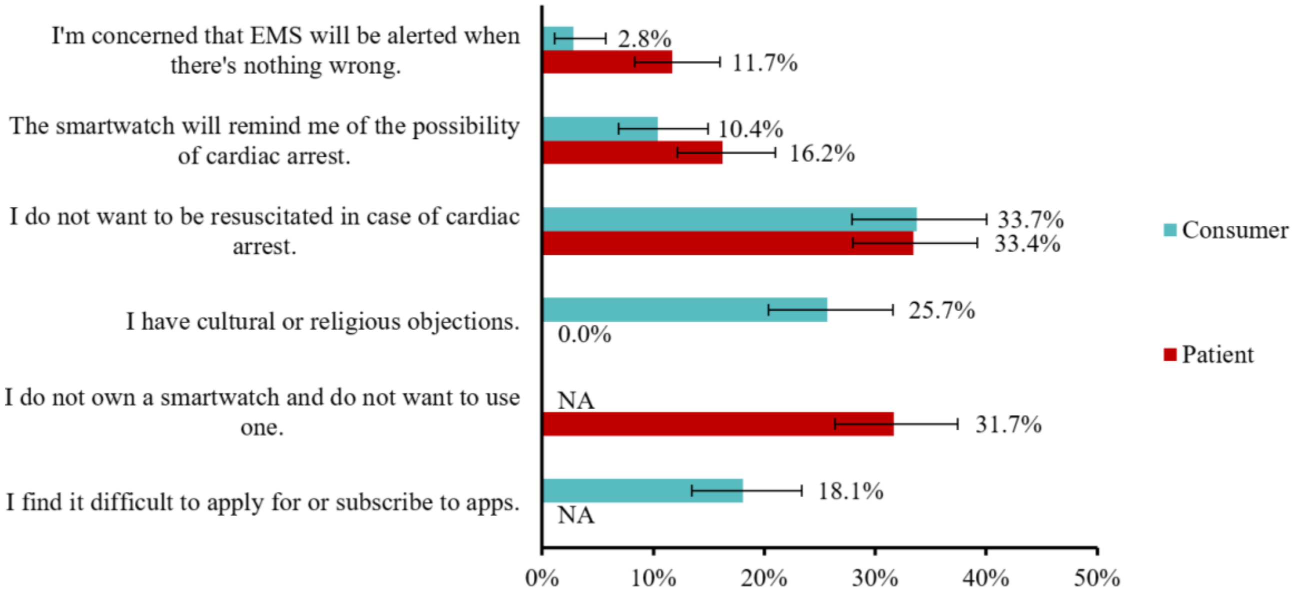
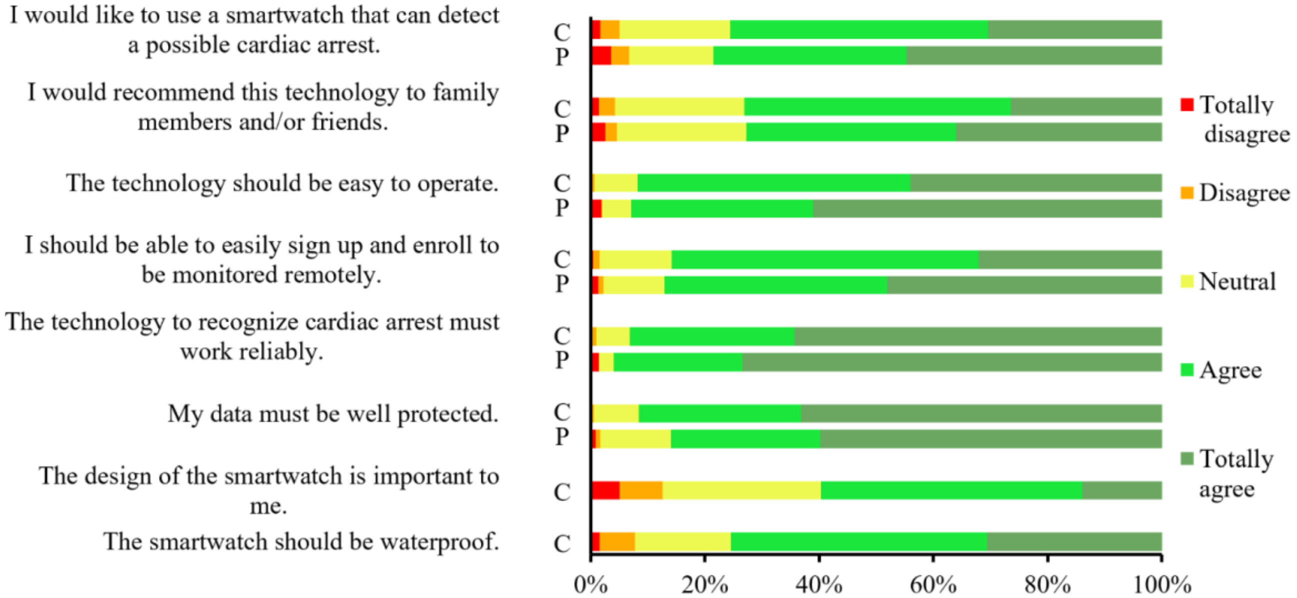


Figure 5. Most cited reasons for abstaining from or expressing uncertainty about adopting the technology. Error bars represent 95% CIs. Consumer group n=249, patient group n=290. EMS: emergency medical services.



In both groups, around 75% (consumer group: n=760; patient group: n=1045) agreed that they would like to use a smartwatch that can be used to detect cardiac arrests, about 72% (consumer group: n=735; patient group: n=964) agreed they would recommend the technology to family and friends, and around 86% (consumer group: n=863; patient group: n=1116) agreed that it should be easy to sign up to use the technology. Both groups agreed (>93%) that the technology should be reliable (consumer group: n=936; patient group: n=1204), and agreed (around 92%) that the technology should be easy to operate (consumer group: n= 923; patient group: n=1223). The consumer group felt a little more strongly that data should be well protected; 91.6% (n=920) agreed compared to 86% (n=1071) in the patient group (Figure 6).

Figure 6. Participants' agreement with general statements regarding the technology, ranked according to a 5-point Likert scale. C: consumer group; P: patient group.



Participants were also surveyed regarding their willingness to make monthly payments for the technology, expressed in euros. The median amount the consumer group and patient group were willing to allocate, given they indicated that they would use the technology, was €5.0 (IQR €2.0 - €10.0) and €10.0 (IQR €5.0 - €20.0), respectively, per month (an exchange rate of €1 = US $1.06054 applied at the time of the study).

Content Analyses

The open-ended questions were categorized into the most prevalent themes, as summarized below. Table 2 provides a quantitative overview of the content analysis.

Table . Content analysis.[a]

Themes and subthemes		Consumers, n	Patients, n	Total, n
Life-saving potential				
	Overall	278	402	680
	Time	107	257	364
	Experience/medical history	72	195	267
Peace of mind				
	Overall	49	72	121
	Reduce stress	39	59	98
	Induce stress	10	13	23
Prevention		48	92	140
Affordability and accessibility				
	Overall	96	145	241
	Health inequity	22	53	75
	Ease of use	30	44	74
Accuracy and reliability		55	59	114
Data and privacy protection		24	11	35

[a]Participants could give answers that were applicable to multiple themes and subthemes. The table shows the number of individuals that mentioned the specific theme in the open-ended questions.

Life-Saving Potential

Both groups acknowledged the potential life-saving capabilities of the technology; this was mentioned by 680 participants. Participants underscored the importance of swift intervention during critical events such as cardiac arrest and expressed that this technology could help shorten the time to resuscitation by EMS; 364 participants mentioned this. They acknowledged that rapid intervention is associated with increased survival and recognized the potential benefit from using automated cardiac arrest detection.

In total, 267 participants mentioned their personal experiences or medical history. Some shared accounts of friends or family members who experienced cardiac arrest, some of whom did not survive. These experiences influenced their favorable perception toward automated cardiac arrest diagnosis. Additionally, participants, especially in the patient group, often referred to their own medical history, noting their increased risk of cardiac arrest. They expressed a strong desire to use every tool available to enhance their chances of survival in such an event.

Peace of Mind

A total of 98 participants mentioned that the technology could also provide peace of mind for users, offering the assurance that EMS will be alerted in the event of cardiac arrest, even if no witnesses are present. Potential users mentioned an increased feeling of security and confidence if such a technology were available, potentially leading to greater physical activity. However, 23 individuals also expressed concerns that this technology could serve as a constant reminder of the possibility of cardiac arrest, potentially leading to stress.

Prevention

A total of 140 respondents also mentioned that such a technology could potentially be used for preventive purposes, such as detecting a cardiac problem before it manifests in cardiac arrest. Participants also indicated that using this technology may raise awareness for heart problems and may also promote a healthier lifestyle.

Affordability and Accessibility

A total of 241 individuals expressed concerns about, or mentioned the importance of, the accessibility and affordability of the technology. Among them, 75 highlighted that disparities in access could potentially lead to or exacerbate health care inequities. The participants indicated that researchers and product manufacturers should identify solutions for widespread accessibility.

Moreover, participants mentioned the importance of integrating the technology into standard health insurance as an option to enhance accessibility. Finally, to increase accessibility, the technology should be intuitive, simple, and easy to use. This was emphasized by 74 individuals. This is especially important when making the technology accessible for users with a low level of technical proficiency.

Accuracy and Reliability

A total of 114 participants expressed concerns regarding the accuracy and reliability of the technology. They emphasized that thorough testing is needed to minimize false alarms. False alarms could potentially strain the existing health care system, induce anxiety, and erode trust in the technology.

Data and Privacy Protection

Another concern raised by 35 participants was related to data collection and the secure handling of medical and personal data. Mainly, the consumers placed significant emphasis on safeguarding their data and privacy and the need to obtain consent regarding data collection while ensuring that only essential data are collected. Some participants expressed reservations about the involvement of smartwatch companies in managing medical data.

Discussion

Principal Findings

We assessed attitudes and perceptions of potential users toward automated cardiac arrest detection using smartwatch sensor data. We found that the vast majority of the participants expressed their willingness to adopt this innovative technology. However, 1 of 20 individuals indicated that they did not want to adopt this technology, and a considerable number of individuals were yet undecided. Our findings revealed several barriers and concerns that warrant careful consideration.

Previous research assessed the acceptability of the use of wrist-worn wearables, mainly focusing on activity tracking and time spent using the device as a measurement of acceptance [20]. However, to our knowledge, our research is the first to focus on the perceptions of potential users and their willingness to accept automated cardiac arrest detection. Understanding users' perspectives is imperative for research groups developing such technology, enabling them to address concerns early in the development process. Moreover, health care professionals involved in counseling patients at risk, as well as any stakeholders involved in the implementation, distribution, or marketing of the technology, should have a clear understanding of users' perspectives. This is paramount in order to increase acceptability [21-23] and to ensure effective implementation [24].

In this context, the insights gained from this study have several important implications. One key factor contributing to reluctance in adopting the technology is that some individuals do not want to be resuscitated in the event of a cardiac arrest. Remarkably, this reluctance was not limited to older people or those with significant comorbidities in the patient group but also extended to the relatively young population in the consumer group. While there are legitimate reasons for refusing resuscitation attempts, it is likely that this reluctance partially stems from a lack of understanding about the prognosis of cardiac arrest, particularly when detected and treated early. Fear of being severely handicapped or incapacitated may have played a role. However, in the Netherlands, approximately 90%‐95% of cardiac arrest survivors are known to survive with a favorable neurologic outcome [25,26]. Education and awareness campaigns may play a pivotal role in addressing misconceptions, ensuring that potential users are well-informed about the life-saving potential of the technology [27].

In the consumer group, cultural or religious objections to adopting the smartwatch-based cardiac arrest technology were frequently reported. Although our survey did not delve into the specific nature of these objections, the evident heterogeneity among potential users highlights the need for customized product development and implementation strategies. Acknowledging and understanding the cultural and religious dynamics that influence technology adoption decisions is critical and should be investigated in more depth in future studies.

In the patient group, a significant barrier to adopting smartwatch-based cardiac arrest detection technology was the reluctance to use a smartwatch. This may partly be attributed to a lack of digital literacy. In particular, older individuals may be less accustomed to digital technology and may perceive a high complexity of operating such devices. Emphasizing simplicity and intuitive operation in the development process is therefore crucial.

Potential users also expressed concerns regarding the potentially high cost of the technology and that financial inaccessibility may exacerbate disparities in health equity [28]. It is worth noting that low socioeconomic status is associated with a higher incidence of cardiac arrest [29], suggesting that this demographic may potentially benefit the most from this technology. Therefore, future research on automatic cardiac arrest detection should include a sufficient number of participants from lower socioeconomic backgrounds. Given that low socioeconomic status is associated with a higher incidence of cardiac arrest, it is crucial to consider cost-effectiveness and affordability early in the development phase. Proactive collaboration with health insurance providers, public social welfare systems, nonprofit health care foundations, and government organizations is essential to ensure financial accessibility.

The analysis also highlighted concerns about privacy, data protection, and data use, which are common concerns when introducing novel medical technologies [30]. Governmental and regulatory bodies have been developing guidelines for software used as a medical device to ensure privacy and data protection [31-34]. For developers and researchers involved in creating such technologies, strict adherence to these regulations plays a vital role in building user trust by ensuring the robust protection of their privacy.

Another concern identified was the reliability of automated cardiac arrest detection systems. Developers are thus challenged to achieve exceptionally high levels of sensitivity and specificity [35]. The goal is to create a system capable of accurately identifying cardiac arrest events while simultaneously minimizing false alarms, which are detrimental to both user trust and system efficacy.

Strengths and Limitations

This study features a considerable sample size encompassing a diverse range of participants who might use this technology. The questionnaire was carefully designed using the joint expertise of a leading professional market research organization, a patient federation experienced in researching patient perceptions, and a medical research group with documented experience in survey methodology [36-38].

Our research not only focuses on patients at increased risk of cardiac arrest, but also includes consumers who already own a smartwatch, which positions them to be early adopters as soon as the technology is implemented. It is noteworthy that approximately 50% of individuals experiencing a cardiac arrest have no prior history of cardiac symptoms or events, and OHCA also frequently affects middle-aged adults [39]. This underscores the applicability of this technology across a broad demographic, including those perceived as "healthy." By encompassing both high-risk patients and regular smartwatch users, our study captures a wide spectrum of perspectives, enhancing the relevance and applicability of our findings.

We acknowledge some limitations. First, a few participants (n=30) seemed to conflate the terms cardiac arrest and heart attack, suggesting potential misunderstandings about what the technology monitors and detects. This confusion may have influenced their responses. However, both heart attack and cardiac arrest are serious medical conditions that benefit from early detection, and the rationale for continuous monitoring should logically extend from one condition to the other. This confusion also underscores the necessity for enhanced public education to improve understanding of these distinct medical events [27].

Second, the research was conducted with patients and consumers in the Dutch population, where local culture and health care systems may shape attitudes toward technology adoption [40]. This aspect has to be considered when extrapolating our results to other regions, with different health care systems or cultures.

Conclusion

The vast majority of potential users expressed a positive attitude toward automated cardiac arrest detection using smartwatch technology. The primary concerns raised by participants included privacy, data protection, reliability, and accessibility of the technology. Despite such concerns, the vast majority indicated that they would be willing to use the technology. This indicates a strong potential user base but also underscores the importance of addressing the identified concerns to optimize acceptance and effectiveness of the technology.

Conflicts of Interest

The project is supported by a grant from the Top Consortia for Knowledge and Innovation (TKI) office of the Dutch Life Sciences & Health Top Sector (Health Holland). HvS reports grants to his institution from the Zoll Foundation and Stryker Emergency Care, both outside the submitted work.

Multimedia Appendix 1
Translated surveys.
[DOCX File, 46 KB - humanfactors_v11i1e57574_app1.docx]

References

1. Berdowski J, Berg RA, Tijssen JGP, Koster RW. Global incidences of out-of-hospital cardiac arrest and survival rates: systematic review of 67 prospective studies. Resuscitation 2010 Nov;81(11):1479-1487. [doi: 10.1016/j.resuscitation.2010.08.006] [Medline: 20828914]
2. Gräsner JT, Wnent J, Herlitz J, et al. Survival after out-of-hospital cardiac arrest in Europe - results of the Eureca TWO study. Resuscitation 2020 Mar 1;148:218-226. [doi: 10.1016/j.resuscitation.2019.12.042] [Medline: 32027980]
3. Myat A, Song KJ, Rea T. Out-of-hospital cardiac arrest: current concepts. Lancet 2018 Mar 10;391(10124):970-979. [doi: 10.1016/S0140-6736(18)30472-0] [Medline: 29536861]
4. Iwami T, Nichol G, Hiraide A, et al. Continuous improvements in "chain of survival" increased survival after out-of-hospital cardiac arrests: a large-scale population-based study. Circulation 2009 Feb 10;119(5):728-734. [doi: 10.1161/CIRCULATIONAHA.108.802058] [Medline: 19171854]
5. Dami F, Heymann E, Pasquier M, Fuchs V, Carron PN, Hugli O. Time to identify cardiac arrest and provide dispatch-assisted cardio-pulmonary resuscitation in a criteria-based dispatch system. Resuscitation 2015 Dec;97:27-33. [doi: 10.1016/j.resuscitation.2015.09.390] [Medline: 26433118]
6. Larsen MP, Eisenberg MS, Cummins RO, Hallstrom AP. Predicting survival from out-of-hospital cardiac arrest: a graphic model. Ann Emerg Med 1993 Nov;22(11):1652-1658. [doi: 10.1016/s0196-0644(05)81302-2] [Medline: 8214853]
7. Bürger A, Wnent J, Bohn A, et al. The effect of ambulance response time on survival following out-of-hospital cardiac arrest. Dtsch Arztebl Int 2018 Aug 20;115(33-34):541-548. [doi: 10.3238/arztebl.2018.0541] [Medline: 30189973]
8. Schober P, van den Beuken WMF, Nideröst B, et al. Smartwatch based automatic detection of out-of-hospital cardiac arrest: study rationale and protocol of the HEART-SAFE project. Resusc Plus 2022 Dec;12:100324. [doi: 10.1016/j.resplu.2022.100324] [Medline: 36386769]
9. Hutton J, Lingawi S, Puyat JH, et al. Sensor technologies to detect out-of-hospital cardiac arrest: a systematic review of diagnostic test performance. Resusc Plus 2022 Sep;11:100277. [doi: 10.1016/j.resplu.2022.100277] [Medline: 35935174]
10. van den Beuken WMF, Sayre MR, Olasveengen TM, Sunshine JE. Wolf Creek XVII part 3: automated cardiac arrest diagnosis. Resusc Plus 2023 Dec;16:100499. [doi: 10.1016/j.resplu.2023.100499] [Medline: 38059269]
11. Shewale R. Smartwatch statistics 2024: worldwide market data. Demandsage. 2024. URL: https://www.demandsage.com/smartwatch-statistics/ [accessed 2024-06-28]
12. Schober P, Schwarte LA. From system to organ to cell: oxygenation and perfusion measurement in anesthesia and critical care. J Clin Monit Comput 2012 Aug;26(4):255-265. [doi: 10.1007/s10877-012-9350-4] [Medline: 22437884]
13. Nazarian S, Lam K, Darzi A, Ashrafian H. Diagnostic accuracy of smartwatches for the detection of cardiac arrhythmia: systematic review and meta-analysis. J Med Internet Res 2021 Aug 27;23(8):e28974. [doi: 10.2196/28974] [Medline: 34448706]
14. Strik M, Ploux S, Ramirez FD, et al. Smartwatch-based detection of cardiac arrhythmias: beyond the differentiation between sinus rhythm and atrial fibrillation. Heart Rhythm 2021 Sep;18(9):1524-1532. [doi: 10.1016/j.hrthm.2021.06.1176] [Medline: 34147700]
15. Han D, Bashar SK, Zieneddin F, et al. Digital image processing features of smartwatch photoplethysmography for cardiac arrhythmia detection. Annu Int Conf IEEE Eng Med Biol Soc 2020 Jul:4071-4074. [doi: 10.1109/EMBC44109.2020.9176142] [Medline: 33018893]
16. Schober P, Vetter TR. Descriptive statistics in medical research. Anesth Analg 2019 Dec;129(6):1445. [doi: 10.1213/ANE.0000000000004480] [Medline: 31743158]
17. van der Mooren F, de Vries R. Steeds meer hoogopgeleiden in Nederland: wat voor beroep hebben ze? Centraal Bureau voor de Statistiek. 2022. URL: https://www.cbs.nl/nl-nl/longread/statistische-trends/2022/steeds-meer-hoogopgeleiden-in-nederland-wat-voor-beroep-hebben-ze-?onepage=true [accessed 2024-06-28]
18. Bevolking; kerncijfers. Centraal Bureau voor de Statistiek. 2024. URL: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/85496NED/table?ts=1715176085413 [accessed 2024-06-28]
19. Inkomen van huishoudens; inkomensklassen, huishoudenskenmerken. Centraal Bureau voor de Statistiek. 2022. URL: https://opendata.cbs.nl/#/CBS/nl/dataset/83932NED/table?searchKeywords=inkomen%20huishoudens%20nederland [accessed 2024-06-28]
20. Germini F, Noronha N, Borg Debono V, et al. Accuracy and acceptability of wrist-wearable activity-tracking devices: systematic review of the literature. J Med Internet Res 2022 Jan 21;24(1):e30791. [doi: 10.2196/30791] [Medline: 35060915]
21. Dwivedi YK, Rana NP, Tamilmani K, Raman R. A meta-analysis based modified unified theory of acceptance and use of technology (meta-UTAUT): a review of emerging literature. Curr Opin Psychol 2020 Dec;36:13-18. [doi: 10.1016/j.copsyc.2020.03.008] [Medline: 32339928]
22. Rahimi B, Nadri H, Lotfnezhad Afshar H, Timpka T. A systematic review of the technology acceptance model in health informatics. Appl Clin Inform 2018 Jul;9(3):604-634. [doi: 10.1055/s-0038-1668091] [Medline: 30112741]
23. Marikyan M, Papagiannidis P. Unified theory of acceptance and use of technology. In: TheoryHub Book: TheoryHub; 2021. URL: https://open.ncl.ac.uk/theory-library/unified-theory-of-acceptance-and-use-of-technology.pdf [accessed 2024-06-28]

24. Jacobs JV, Hettinger LJ, Huang YH, et al. Employee acceptance of wearable technology in the workplace. Appl Ergon 2019 Jul;78:148-156. [doi: 10.1016/j.apergo.2019.03.003] [Medline: 31046946]
25. Zijlstra J, Radstok A, Pijls R, Nas J, Beesems S, Hulleman M. Overleving na een reanimatie buiten het ziekenhuis: vergelijking van de resultaten van 6 verschillende Nederlandse regio's. In: Reanimatie in Nederland: Hartstichting; 2016:9-24.
26. Beesems SG, Blom MT, van der Pas MHA, et al. Comorbidity and favorable neurologic outcome after out-of-hospital cardiac arrest in patients of 70 years and older. Resuscitation 2015 Sep;94:33-39. [doi: 10.1016/j.resuscitation.2015.06.017] [Medline: 26116780]
27. Donohoe RT, Haefeli K, Moore F. Public perceptions and experiences of myocardial infarction, cardiac arrest and CPR in London. Resuscitation 2006 Oct;71(1):70-79. [doi: 10.1016/j.resuscitation.2006.03.003] [Medline: 16945467]
28. Oliver A, Mossialos E. Equity of access to health care: outlining the foundations for action. J Epidemiol Community Health 2004 Aug;58(8):655-658. [doi: 10.1136/jech.2003.017731] [Medline: 15252067]
29. Reinier K, Stecker EC, Vickers C, Gunson K, Jui J, Chugh SS. Incidence of sudden cardiac arrest is higher in areas of low socioeconomic status: a prospective two year study in a large United States community. Resuscitation 2006 Aug;70(2):186-192. [doi: 10.1016/j.resuscitation.2005.11.018] [Medline: 16814445]
30. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019 Jan;25(1):30-36. [doi: 10.1038/s41591-018-0307-0] [Medline: 30617336]
31. Goodman B, Flaxman S. European Union regulations on algorithmic decision-making and a "right to explanation". AI Magazine 2017 Sep;38(3):50-57. [doi: 10.1609/aimag.v38i3.2741]
32. Software as a medical device (SaMD): application of quality management system. IMDRF International Medical Device Regulators Forum. 2015. URL: https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-151002-samd-qms.pdf [accessed 2024-07-09]
33. Software as a medical device (SAMD): clinical evaluation. IMDRF International Medical Device Regulators Forum. 2017. URL: https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-170921-samd-n41-clinical-evaluation_1.pdf [accessed 2024-07-09]
34. "Software as a medical device": possible framework for risk categorization and corresponding considerations. IMDRF International Medical Device Regulators Forum. 2014. URL: https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-140918-samd-framework-risk-categorization-141013.pdf [accessed 2024-07-09]
35. Vetter TR, Schober P, Mascha EJ. Diagnostic testing and decision-making: beauty is not just in the eye of the beholder. Anesth Analg 2018 Oct;127(4):1085-1091. [doi: 10.1213/ANE.0000000000003698] [Medline: 30096083]
36. Krage R, van Rijn C, van Groeningen D, Loer SA, Schwarte LA, Schober P. Cormack-Lehane classification revisited. Br J Anaesth 2010 Aug;105(2):220-227. [doi: 10.1093/bja/aeq136] [Medline: 20554633]
37. Schober P, Bossers SM, Dong PV, Boer C, Schwarte LA. What do anesthesiologists know about P values, confidence intervals, and correlations: a pilot survey. Anesthesiol Res Pract 2017;2017:4201289. [doi: 10.1155/2017/4201289] [Medline: 29158732]
38. Schober P, van Dehn FB, Bierens J, Loer SA, Schwarte LA. Public access defibrillation: time to access the public. Ann Emerg Med 2011 Sep;58(3):240-247. [doi: 10.1016/j.annemergmed.2010.12.016] [Medline: 21295376]
39. Marijon E, Uy-Evanado A, Dumas F, et al. Warning symptoms are associated with survival from sudden cardiac arrest. Ann Intern Med 2016 Jan 5;164(1):23-29. [doi: 10.7326/M14-2342] [Medline: 26720493]
40. Yang Meier D, Barthelmess P, Sun W, Liberatore F. Wearable technology acceptance in health care based on national culture differences: cross-country analysis between Chinese and Swiss consumers. J Med Internet Res 2020 Oct 22;22(10):e18801. [doi: 10.2196/18801] [Medline: 33090108]

Abbreviations

CBS: Central Bureau of Statistics
EMS: emergency medical services
NPF: Netherlands Patient Federation
OHCA: out-of-hospital cardiac arrest
PPG: photoplethysmography

Edited by K Adapa; submitted 20.02.24; peer-reviewed by H Isto, M Fong; revised version received 19.05.24; accepted 02.06.24; published 25.07.24.

Please cite as:
van den Beuken WMF, van Schuppen H, Demirtas D, van Halm VP, van der Geest P, Loer SA, Schwarte LA, Schober P
Investigating Users' Attitudes Toward Automated Smartwatch Cardiac Arrest Detection: Cross-Sectional Survey Study
JMIR Hum Factors 2024;11:e57574
URL: https://humanfactors.jmir.org/2024/1/e57574
doi:10.2196/57574

Designing a Smartphone-Based Pulse Oximeter for Children in South Africa (Phefumla Project): Qualitative Analysis of Human-Centered Design Workshops With Health Care Workers

Elif I Ilhan[1], MA; Lucia N Jola[1], MA; Marieke M van der Zalm[1], PhD; Mike Bernstein[2], BSc; Pierre Goussard[3], PhD; Andrew Redfern[3], MBCHB, MPhil; Anneke C Hesseling[1], PhD; Graeme Hoddinott[1,4], PhD; Eric D McCollum[1,5], MPH, MD; Carina King[6], PhD

[1]Desmond Tutu TB Centre, Department of Paediatrics and Child Health, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

[2]PhysioMonitor, San Roman, CA, United States

[3]Department of Paediatrics and Child Health, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

[4]School of Public Health, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

[5]Global Program for Pediatric Respiratory Sciences, Eudowood Division of Pediatric Respiratory Sciences, Johns Hopkins School of Medicine, Baltimore, MD, United States

[6]Department of Global Public Health, Karolinska Institutet, Stockholm, Sweden

Corresponding Author:
Carina King, PhD
Department of Global Public Health, Karolinska Institutet, , Stockholm, , Sweden

Abstract

Background: Pulse oximeters noninvasively measure blood oxygen levels, but these devices have rarely been designed for low-resource settings and are inconsistently available at outpatient clinics.

Objective: The *Phefumla* project aims to develop and validate a pediatric smartphone-based pulse oximeter designed specifically for this context. We present the process of human-centered oximeter design with health care workers in South Africa.

Methods: We purposively sampled 19 health care workers from 5 clinics in Khayelitsha, Cape Town. Using a human-centered design approach, we conducted participatory workshops with four activities with health care workers: (1) they received 3D-printed prototypes of potential oximeter designs to provide feedback; (2) we demonstrated on dolls how they would use the novel oximeter; (3) they used pile sorting to rank design features and suggest additional features they desired; and (4) they designed their preferred user interface using a whiteboard, marker, and magnetized features that could be repositioned. We audio recorded the workshops, photographed outputs, and took detailed field notes. Analysis involved iterative review of these data to describe preferences, identify key design updates, and provide modifications.

Results: Participants expressed a positive sentiment toward the idea of a smartphone pulse oximeter and suggested that a pediatric device would address an important gap in outpatient care. Specifically, participants expressed a preference for the prototype that they felt enabled more diversity in the way it could be used. There was a strong tendency to prioritize pragmatic design features, such as robustness, which was largely dictated by health care worker context. They also added features that would allow the oximeter device to serve other clinical functions in addition to oxygen saturation measurement, such as temperature and respiratory rate measurements.

Conclusions: Our end user–centered rapid participatory approach led to tangible design changes and prompted design discussions that the team had not previously considered. Overall, health care workers prioritized pragmatism for pediatric pulse oximeter device design.

(JMIR Hum Factors 2024;11:e54983) doi:10.2196/54983

KEYWORDS

pediatrics; human-centered design; participatory design; pulse oximeter; South Africa; smartphone; mobile phone

Introduction

Hypoxemia, defined as an abnormally low peripheral arterial oxyhemoglobin saturation (SpO_2) of <90%, is an important risk factor for death among children with lower respiratory infections in low- and middle-income countries (LMICs) [1-4]. An estimated 7 million children were hospitalized with hypoxemic pneumonia in 2019, and in sub-Saharan Africa, 28% (95% CI

25%-35%) of children with acute respiratory diseases were hypoxemic [5]. During outpatient care, the burden of hypoxemia may be considerable, with 2019 estimates suggesting a 23.1% prevalence among children with respiratory illnesses [5]. Pulse oximeters are medical devices that noninvasively measure SpO_2 and can therefore detect hypoxemia.

Frequent pulse oximeter use is associated with positive health outcomes such as reducing mortality rates and has been found to be cost-effective in low-resource settings [1,6]. Although oximeters are commonly used in pediatric clinical care in high-income countries, they are not consistently available in LMICs [4], especially during outpatient care where most children first access care and their illness may be more treatment responsive. The COVID-19 pandemic led to large investments being made into oxygen ecosystems, including pulse oximetry [7]; however, it did not focus on overcoming key implementation challenges for children. Pediatric pulse oximetry implementation in LMICs is restricted by barriers such as cost, lack of appropriately designed pediatric devices and probes, disruptive movements of small children, unavailability of devices, lack of training and supervision, lack of maintenance, lack of electricity, and health care provider misconceptions [8-14]. A pediatric-specialized, low-cost, smartphone-based pulse oximeter device could potentially address many of these implementation barriers and serve as a valuable tool in outpatient LMIC settings.

The *Phefumla* project aims to cocreate a low-cost, smartphone-based, reflectance pulse oximeter device for children that is optimized for LMIC outpatient contexts. Reflectance oximetry, unlike transmittance oximetry, measures the relative ratio of unabsorbed red and infrared light that is reflected off of tissues rather than through tissues to produce an estimate of SpO_2 [15]. A key source of inequities in health is access to diagnostic services, with almost half of the global population having little or no access to diagnostics [16]. Part of this inequality stems from devices designed for high-income and inpatient settings that are cost-prohibitive to purchase, sustain consumables, and maintain. To address this challenge, there is a need for a holistic framework to guide the design of medical devices so that they may be contextually appropriate for the settings in which they will be used [17].

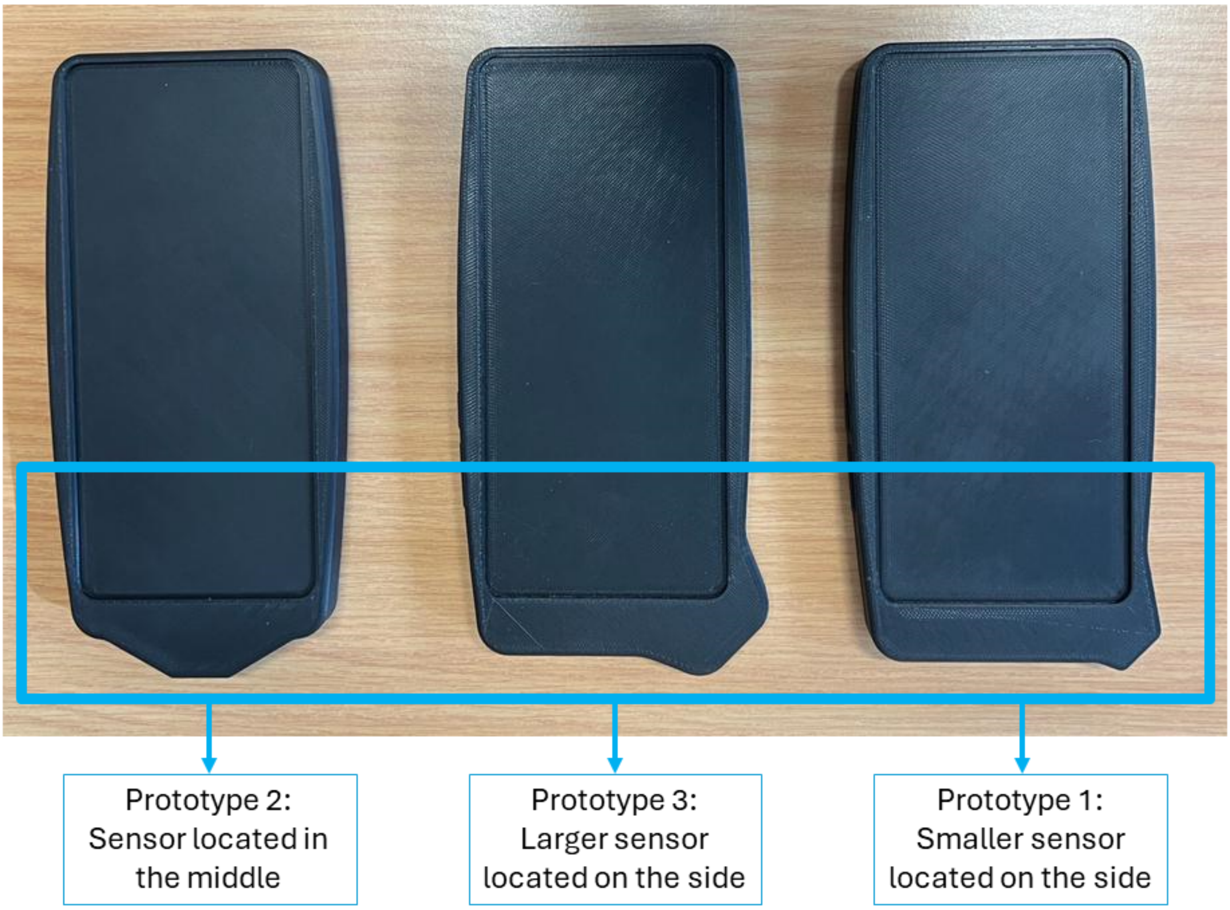
We used a human-centered design (HCD) approach, with the aim of achieving a contextually appropriate device that meets the specific health care needs of the population [18]. The HCD method is one of many approaches to co-design and was chosen for this study given its successful application in previous global health intervention and medical device development projects [19-25]. In this paper, we describe the participatory HCD processes with health care workers (HCWs) and how this led to design changes, as an example of a rapid approach to medical device development that centers inclusion.

Methods

Overview

We conducted a qualitative observation study of participatory workshops that drew on the HCD approach, with HCWs in Khayelitsha, Cape Town, South Africa, from September 1-16, 2022. For these workshops, we had 3D-printed 3 prototype reflectance devices, all based on the same smartphone model being housed inside a case that would contain the oximeter sensor and additional hardware for processing (Figure 1). These prototypes were developed by the *Phefumla* team to prompt HCW reflections on the size and positioning of the sensor while keeping all other factors consistent.

Figure 1. 3D-printed *Phefumla* reflectance oximeter prototypes.



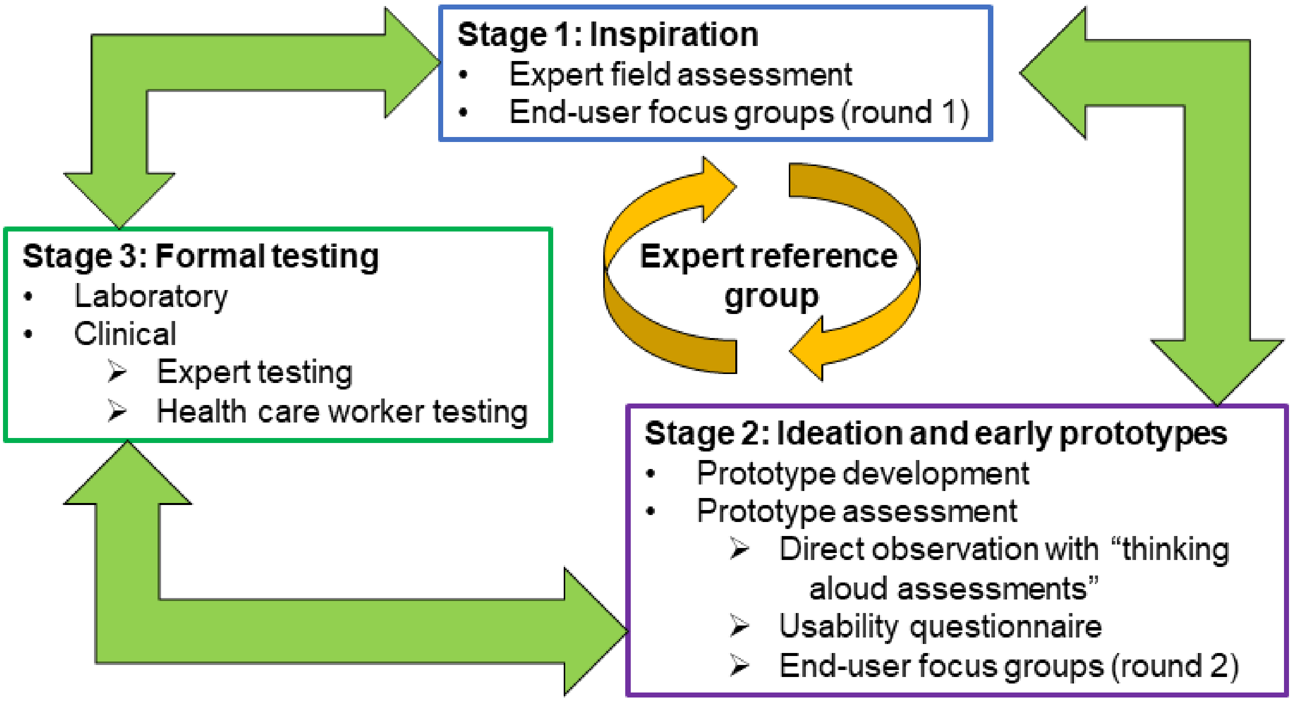
Setting

The East subdistrict of Khayelitsha is a low-income and low-resource area in Cape Town, South Africa, often referred to as a township. It has an estimated population of 450,000 people [26], who are predominantly Black African (99%) and the majority of whom live in informal housing [27]. First-tier primary health care (PHC) in South Africa is provided primarily through nurse-led clinics and community health centers, which are available within 5 km of 90% of the population and is often the first point of contact [28]. These facilities are free of cost and provide comprehensive basic services such as maternal, child, and reproductive health; HIV and tuberculosis testing and treatment; and care for noncommunicable diseases and common ailments [28]. Secondary care is delivered at district hospitals, which conduct minor procedures, and the third tier consists of tertiary hospitals that have the infrastructure, specialists, and equipment for major surgeries [29]. Many obstacles limit adequate implementation of health services at the PHC level in South Africa, including the HIV/AIDs pandemic, shortages of HCWs, unequal distribution of resources, and the legacy of the apartheid era [30]. This study was conducted at PHCs.

HCD Approach

HCD is based on using “techniques which communicate, interact, empathize and stimulate the people involved, obtaining an understanding of their needs, desires and experiences which often transcends that which the people themselves actually realized” [31]. This approach encourages stakeholders, experts, and end users—in our case, HCWs—to generate knowledge collaboratively to co-design a medical device [32]. Through involving end users in the design process, HCD allows for the development of devices that are locally and contextually appropriate and can meet the specific health care needs of the population [18]. The key principles of HCD are the active involvement of users and a clear understanding of user and task requirements; iteration of design solutions, where end users provide feedback on design solutions starting early in the process; and making use of multidisciplinary design teams [33]. The HCD approach consists of three iterative stages: (1) inspiration, (2) ideation and prototyping, and (3) formal testing. In this study, we report activities conducted in stage 2, the ideation and prototyping of the *Phefumla* smartphone oximeter development (Figure 2). This builds on our stage-1 findings that explored HCWs’ current experience with pediatric pulse oximetry, which will be published elsewhere.

Figure 2. The human-centered design approach.



Participatory Workshops

We conducted participatory workshops, consisting of 4 activities, to facilitate the process of having HCWs co-design a smartphone-based reflectance oximeter (Table 1). The discussion guide is available in Multimedia Appendix 1.

Table . Summary of participatory design workshop activities.

Activities	Resources and tools	Description
1. Design preference	Three 3D-printed plastic prototypes of potential device designs	Participants were given 3 different rapid prototypes and asked questions about the devices' design. These included how they felt about the placement of the sensor on the device; how confident they felt placing the device on a child for a reading; how easy they believe the device would be to clean; and how durable the device was. Participants were encouraged to give suggestions and state preferences.
2. Device use	Two dolls (1 infant sized and 1 slightly larger sized)	Participants were asked to demonstrate how they would use the 3 prototypes on a child, specifically where they would place the sensor for taking a measurement on 2 different sized dolls to represent younger and older infants.
3. Feature ranking	A pile of cards with design features written on them	Participants were asked to rank 11 design features, which were deemed important from stage 1 of the HCD[a] process (eg, battery life), from most to least important. They were also given an opportunity to add their own features on blank cards with markers.
4. User interface	A magnetic board and magnets of different design features of the interface (SpO_2[b] reading, waveform, pulse, bouncing bar, buttons, charging symbols, date, and time)	Participants were asked to arrange interface components as they would like the screen of the device to look. They were provided with a whiteboard marker to include any other features.

[a]HCD: human-centered design.

[b]SpO_2: oxyhemoglobin saturation.

We conducted a pilot design workshop with 3 research nurses from the Desmond Tutu TB Centre to check the quality and coherency of the planned activities. Following the relevant consenting procedures, 1-hour-long design workshops were held with 7 small groups of our sample's HCWs (2-4 HCWs per group). These workshops were conducted by 2 female postgraduate research assistants with comprehensive knowledge and experience of qualitative data collection (EII and LNJ) under the supervision and with the assistance of a pediatric pulmonologist (EDM). Small groups were chosen for pragmatic reasons to minimize disruption to clinical service at the facilities and were conducted in the clinics.

Sampling and Participants

Participants were sampled from a larger pool of participants who had taken part in the previous stage of the study. Stage 1 of the HCD process (inspiration) involved small group discussions with HCWs focusing on barriers and challenges to routine pediatric oximetry use. These HCWs were therefore primed before the co-design workshops to think about the pros and cons of pulse oximeter features. Five clinics in the East subdistrict of Khayelitsha were eligible, and HCWs were purposefully sampled (rich case) using the following inclusion criteria: (1) having experience taking pulse oximeter measurements in children and (2) having taken part in the previous stage of the *Phefumla* study. Participants who had consented in stage 1 to be contacted were followed up to setup face-to-face meetings. Participants were given a small monetary voucher (worth approximately US $15) and provided with refreshments as reimbursement for their time.

Data Collection

Data were collected via audio recordings and photographs taken of activity end results, as well as through comprehensive observation notes. A semistructured workshop guide was developed and used in English, the predominant working language in health care settings in South Africa. However, most participants were native Xhosa speakers, and some discussions were held in Xhosa; LNJ is a native Xhosa speaker and acted as a translator for these sections. The 2 researchers who facilitated the workshops alternated between (1) asking questions and leading facilitation and (2) keeping detailed field notes.

Analysis

Data were analyzed using the framework of exploratory qualitative analysis. Exploratory research is concerned with exploring a phenomenon more deeply to gain a granular understanding of it and has 2 key aspects: open-mindedness and flexibility [34]. Recordings were repeatedly listened to by the 2 researchers who conducted the workshops (EII and LNJ), alongside looking through the captured pictures and written field notes. The wider research team had preidentified key design features of interest based on a rapid scoping review, stage-1 small group discussions, and team expertise. Quotes and notes taken during the workshops were mapped together by EII and LNJ under these categories of design features, using Microsoft Excel. This initial mapping framework was shared and discussed with the entire research team, where findings were discussed and probed. This was done iteratively until the team decided on actionable feedback for the pulse oximeter prototype and shared them with the engineer (MB).

Ethical Considerations

Approval was obtained from the City of Cape Town to recruit HCWs and conduct the project at 7 clinics in the East subdistrict of Khayelitsha. Institutional approval was obtained from Johns Hopkins University (IRB00294436), Stellenbosch University (N22/01/009), and the Swedish Ethics Board (Dnr 2022-01897-01). Facility managers and other relevant gatekeepers were approached after receiving approval to ask for permission to access HCWs. All participants provided written, informed consent. Field notes were anonymized and did not record any identifiable data from HCWs, and recordings were stored in secure local servers to safeguard participant information.

Results

Overview

A total of 7 workshops were conducted with 19 HCWs. The most common reasons for participants from stage 1 not taking part in these stage-2 workshop were being ill, on leave, or absent at the clinic on the scheduled days; having been rotated to a different PHC; or having resigned. All 19 participants retained were nursing staff (including a range of nursing cadres), with 18 (95%) female participants and 1 (5%) male participant.

Activity 1: Selecting a Preferred Prototype

Three 3D-printed prototypes were presented to the groups (Figure 1). The strongest preference was shown for prototype 2 with the sensor in the middle, with 4 (57%) of the 7 groups reaching a consensus on preferring this design. However, this was not unanimous, with 1 (14%) group preferring prototype 1, one (14%) group preferring prototype 3, and 1 (14%) group wanting a combination of prototype 3's larger sensor size with prototype 2's sensor location.

Participants primarily liked prototype 2 because of its sensor location being in the middle, noting that the device would be easier to use on a child as you would not have to angle it to get a reading, it did not matter if you were right- or left-handed, and some participants liked the larger size of the overall device and smaller size of the sensor (compared to the larger sensor of prototype 3). There were some concerns that the device itself was too large and that a smaller device would be easier to use, as well as concerns about having to hold the device without dropping it.

> *It's easier for me to get grasp of the monitor and put it on the child, rather than using the corner.* [HCW, clinic 3]

When discussing the sensor, prototype 3's larger sensor elicited a range of responses, with some HCWs stating that it would be too difficult to use on an infant or young child (eg "it's too big"), whereas others thought the larger sensor size was a benefit, for example:

> *Very much easy [to use] because the sensor is bigger.* [HCW, clinic 1]
>
> *The sensor is nice and big.* [HCW, clinic 5]

Overall, participants displayed a positive sentiment to this style of device being easy to use on a child, and most participants felt comfortable placing the sensor correctly on a child. Robustness was a concern in several groups, as it was noted that a smartphone screen can break when dropped, and participants offered several modifications in relation to this:

> *It would be better if it were rubber or had a pouch, so it does not break.* [HCW, clinic 1]

> *The back must be rubbery, and the outer part is rubbery too.* [HCW, clinic 3]

> *If it is a glass screen it will [break easily].* [HCW, clinic 4]

All groups stated the device would be very easy to clean, with the most common suggestion for cleaning the device being wiping it with a disinfectant and cloth after each use. All groups felt it would be easy to store as well, with suggestions such as to keep it in a locked drawer, cabinet, or room or to include a storage pouch with the device:

> *Important that it's got a pouch—a bag, so it doesn't get too dusty.* [HCW, clinic 1]

Activity 2: Using the Device on a Child

For activity 2, we asked participants to indicate for each prototype where they would take the pulse oximeter measurements on an infant-sized doll and a larger toddler-sized doll. The purpose was to understand how this novel device would be instinctively applied. The most common location of measurements for infants included the sole of foot, followed by the palm of hand, the hand, the thumb, and toes. These were similar for the older toddler–sized doll, with the sole of foot also being preferred, although HCWs noted that toddlers can kick. Infrequent answers included the wrist, the chest, the forehead, and the neck, which a participant noted would be beneficial as it would not require a child to be undressed. We deliberately did not prompt HCWs to consider specific locations, and it is likely that HCWs defaulted to appendages (ie, hands and feet) that are the most commonly used with a standard pulse oximeter, even if the positioning on those appendages (eg, the palm of the hand) differed.

Activity 3: Design Feature Selection

Table 2 present the results from the feature pile sorting activity, showing the features considered as the 5 most important among the groups. When asked to elaborate on their ranking, participants stated that they first considered what would be essential for the device to function (eg, battery lasting) and that the rest were add-ons (eg, apps installed) that would be nice but were not necessary for core functioning. There was a strong preference displayed for pragmatism in this context:

> *The ones on top are the most important because they're going to sustain the device.* [HCW, clinic 3]

Table . Features ranked among the top 5 for each group.

Feature	Groups (n=7), n (%)	Example quotations for prioritizing features
Portable device	7 (100)	• "You can take it anywhere, for example if it is needed in emergency...then you can take it there." (HCW[a], clinic 4) • "So you can take the sick baby to another room and take the device to the next room" (HCW, clinic 2) • "Because we have three triages in this clinic" (HCW, clinic 3)
Does not break when dropped	6 (86)	• "We are working with kids. It's inevitable that it will fall. It is important that it doesn't break easily when it falls. The kids might not want it and push it away from them and then it falls." (HCW, clinic 4) • "Because we are designing for a small baby not an adult so there's high chances of it falling" (HCW, clinic 1) • "Maybe you're gonna be busy with an emergency so you're going to be scared so you're gonna be shivering or shaking maybe, and the baby will also be fighting you, so at least if it drops it mustn't break easily" (HCW, clinic 2)
Long battery life	6 (86)	• "We're seeing more than 30 children a day and sometimes we don't have time to charge—there's no break when they come. They start to come as early as half past 7 to 4 o'clock so there's no time to say we're still waiting for the battery to get full." (HCW, clinic 4) • "If there's no battery, there's no device." (HCW, clinic 1) • "Loadshedding [of electricity] is happening so it must be charged, and the battery must last" (HCW, clinic 2)
Can measure different parts of the body	5 (71)	• "It doesn't limit you so you can use it on whatever part of the body that you want" (HCW, clinic 2) • "We've got limited sites where you can do accurate readings" (HCW, clinic 1)
Easy to clean	4 (57)	• "Hygiene is very important because we're dealing with kids, so if it's easy to clean then it's more safe." (HCW, clinic 4) • "Because it's used between many patients" (HCW, clinic 5)

[a]HCW: health care worker.

We asked if there were any disagreements in the group, but all groups indicated that they were happy with the consensus reached after discussion.

Participants added the following tangible features: small size, protective cover and storage bag, device holder or stand, and time stamp. However, more generic statements such as "easy to use" and "user friendly" were also added. Some added distractions to the child, such as a colorful screen or pictures. In 1 group, the HCWs also added the inclusion of a probe—the key design feature we were proposing to move away from.

Participants who included "other apps installed" in their overall ranking of design features were asked which apps they desired. The predominant suggestion was to have an app that could also measure temperature and respiratory rate, with a high preference for a multimodal device being displayed. Other app suggestions included an app for referral to hospital emergency wards (known as the *Vula* mobile app in this setting) [35] and an app that referred participants to the emergency medicine practice guidelines [36]. However, there were mixed feelings toward additional apps, and these often were not ranked among the most important features, with the exception of 1 (14%) of the 7 groups. Some of reasons given were that other apps would not be used; that they would negatively affect the battery life of the device; or that people would use the apps for personal reasons. As some HCWs noted:

> *We're not gonna use other apps.* [HCW, clinic 4]

> *I wonder if it's not gonna affect the battery life.* [HCW, clinic 3]

> *People overuse it for personal things.* [HCW, clinic 2]

Activity 4: Interface Design

For the user-interface design activity, participants tended to place pulse and SpO_2 readings together (6/7, 86% groups), although there was variability in where on the screen these were placed as well as variability in the size of the icons. Further, 6 (86%) out of the 7 groups included both the waveform as well as the bouncing bar, with the following reasons: if one is not working, the other will; each feature gives you different information; and it makes the device more accessible in the case someone is only familiar with either the bouncing bar or the waveform. The majority (5/7, 71%) of the groups included icons indicating temperature and respiratory rate, further indicating their preference for a multimodal device. Two (29%) groups added a distraction for the child, such as a moving video with sound. The battery was mostly placed at the top of the screen so that HCWs could immediately see whether the device needed to be charged when switched on. When asked what alarms and sounds were wanted, the main preference was for a sound when there was an abnormal reading. Furthermore, the preference was for a loud volume given the noisy environment of the clinics.

Discussion

In this study, we conducted design workshops with South African HCWs to develop a novel, pediatric-specialized pulse oximeter device, to ensure the device is context appropriate. Through the design workshops, we found that HCWs displayed an overall positive and enthusiastic sentiment toward such a device, seeing its value in clinical use with children with hypoxemia. The findings from these workshops were used to select oximeter prototype 2 (Figure 1), with the sensor in the middle, to take forward into the prototype testing stage, with key updates to the robustness and planned user-interface incorporated.

Participants displayed the strongest preference for a device design with a sensor in the middle, feeling that it was overall the easiest to work with. Although participants felt that a smartphone device would be easy to use on a child, clean, and store and felt confident in placing the sensor correctly, some had concerns over the robustness of the device. They provided multiple suggestions to overcome this, such as a pouch, case, or rubber casing, and as a result, we increased the robustness of the device to be able to resist a drop test. However, this may point toward a potential limitation in using a smartphone interface, which HCWs are largely familiar with and have likely had experiences of breakages. This prompted a discussion within the study team on whether the smartphone inside the casing could be replaced with a locally available phone, allowing for a more sustainable repair solution than most traditional medical devices. Although this was not dealt with at this stage of design, the HCW feedback triggered us to reflect on this aspect of the device in more depth and to plan for future prototypes.

We received the least in-depth feedback on the mock placement and use of the prototypes with dolls of infants. One issue may have been the design of the activity, using infant dolls to prompt discussion. As a key challenge in pediatric oximetry, as noted by the HCWs as well, is the children's movement and them becoming agitated with measurements being done, using a real child may have resulted in more reflective responses. The locations that the HCWs largely defaulted to were the thumb, toes, hands, and feet—where oximeter probes are generally used currently, although not with the same versatility. We had hypothesized that a benefit of reflectance oximetry is the range of locations that could be used, which reduces both the HCWs' need to disturb the child and restrict their movement (eg, their forehead or upper back).

Pragmatic concerns arose most strongly during the activity where design features were ranked. These findings speak to the context in which HCWs in LMICs work, where having usable, durable, and long-lasting devices is of the essence—with participants noting that once devices break, they are unlikely to be replaced. This is due to factors such as limited technical and biomedical support and ties into other literature regarding "medical equipment graveyards"—composed of obsolete or otherwise broken biomedical, donated equipment—which are a common occurrence across LMICs [37]. These findings also speak to similar findings in other literature, where opportunities for redesign in pulse oximeters in LMICs included similar themes such as battery charging and durability, probe fit, and sensitivity in pediatric populations [11].

Participants liked the idea of a multimodal device. Although there were various suggestions given for additional design features, a device that could take temperature and respiratory rate readings in addition to SpO_2 was by far the most desirable design feature proposed. This was desirable to participants as one device with multiple modalities is pragmatically beneficial. This speaks to possible opportunities for integration in future device designs and further developments in the field of eHealth; however, this needs to be weighed against risks. There is the risk that more complex devices will be more expensive, have reduced usability, and not be optimized for the oximetry function. Therefore, the benefits need to be weighed against the added value of additional functions, as a device performing a core functionality well could be beneficial over a device that performs poorly across various functionalities [11,38].

Some of the themes raised by our participants were raised in other studies in LMICs, indicating a degree of generalizability. Khayelitsha is considered to be fairly representative of other low-resource, sub-Saharan African settings when considering HIV exposure, tuberculosis mortality rates, and quality of care. However, some contextual factors may be unique to a particular context. For example, Khayelitsha has access to electricity but frequently experiences power supply blackouts (loadshedding), which happens nationwide, meaning that mains-charged devices are acceptable but need a long battery life. In contrast, in other settings, solar-powered charging was prioritized as access to electricity in health facilities was not universal [11].

Our study has several strengths in being able to rapidly engage with a range of HCWs. However, we also had 3 key limitations. First, we came up with the initial idea for a reflectance pulse oximeter, hypothesizing that this could solve several usability issues for LMIC outpatient settings. Our participants were therefore restricted in their first prototypes to 1 type of oximeter that the research team had chosen. It is possible participants may have preferred an alternative design or traditional transmittance pulse oximeter. It may also have biased our team's presentation of the device and interpretation of the data. However, the workshop researchers had no prior experience in oximetry and led the data collection and analysis process in an attempt to mitigate potential researcher bias. Second, given that our data collection and analysis process were designed to be rapid and pragmatic, we did not extensively pilot the instruments. Lastly, the workshops were conducted primarily in English. Clinical training is done in English and is the language spoken in most professional South African environments. However, it was not the majority of participants' home or first language, which could be a potential limitation. We allowed participants to answer in whatever language they wanted to, and we always had a Xhosa-speaking researcher available to mitigate this limitation.

A contextually appropriate, low-cost, pediatric-specialized, smartphone-based reflectance pulse oximeter was seen to have potential clinical value in the South African context. The process of HCD allowed us to explore HCW's design preferences qualitatively to design a prototype device that would address their specific needs. The overall preference was for a multimodal and pragmatic device, with our rapid participatory approach successfully leading to changes in the oximeter design executed by our engineer.

Acknowledgments

We would like to thank all the health care workers who took part for giving both their time and their wisdom, and the research nurses for their initial feedback when piloting the workshops. The research reported in this publication was supported by the Fogarty International Centre of the National Institutes of Health under award 1R21TW012212-01. In addition, GH receives the financial assistance of the European Union (grant DCI-PANAF/2020/420-028), through the African Research Initiative for Scientific Excellence (ARISE) Pilot Programme. ARISE is implemented by the African Academy of Sciences with support from the European Commission and the African Union Commission. The contents of this document are the sole responsibility of the authors and can under no circumstances be regarded as reflecting the position of the European Union, the African Academy of Sciences, and the African Union Commission. MMvdZ is supported by a career development grant from the EDCTP2 (European and Developing Countries Clinical Trials Partnership) program supported by the European Union (TMA2019SFP-2836 TB lung-FACT2) and the Fogarty International Centre of the National Institutes of Health under award K43TW011028.

Conflicts of Interest

CK and EDM act as independent scientific advisors to the Lifebox Foundation.

Multimedia Appendix 1
Design workshop discussion guide.
[DOCX File, 16 KB - humanfactors_v11i1e54983_app1.docx]

References

1. Lazzerini M, Sonego M, Pellegrin MC. Hypoxaemia as a mortality risk factor in acute lower respiratory infections in children in low and middle-income countries: systematic review and meta-analysis. PLoS One 2015 Sep 15;10(9):e0136166. [doi: 10.1371/journal.pone.0136166] [Medline: 26372640]
2. McCollum ED, King C, Ahmed S, et al. Defining hypoxaemia from pulse oximeter measurements of oxygen saturation in well children at low altitude in Bangladesh: an observational study. BMJ Open Respir Res 2021 Nov;8(1):e001023. [doi: 10.1136/bmjresp-2021-001023] [Medline: 34728475]
3. King C, McCollum ED. Trends in the global burden of paediatric lower respiratory infections. Lancet Infect Dis 2020 Jan;20(1):4-5. [doi: 10.1016/S1473-3099(19)30557-2] [Medline: 31678028]
4. McCollum ED, King C, Colbourn T, et al. Pulse oximetry in paediatric primary care in low-income and middle-income countries. Lancet Respir Med 2019 Dec;7(12):1001-1002. [doi: 10.1016/S2213-2600(19)30358-3] [Medline: 31631030]
5. Rahman AE, Hossain AT, Nair H, et al. Prevalence of hypoxaemia in children with pneumonia in low-income and middle-income countries: a systematic review and meta-analysis. Lancet Glob Health 2022 Mar;10(3):e348-e359. [doi: 10.1016/S2214-109X(21)00586-6] [Medline: 35180418]
6. Lam F, Stegmuller A, Chou VB, Graham HR. Oxygen systems strengthening as an intervention to prevent childhood deaths due to pneumonia in low-resource settings: systematic review, meta-analysis and cost-effectiveness. BMJ Glob Health 2021 Dec;6(12):e007468. [doi: 10.1136/bmjgh-2021-007468] [Medline: 34930758]

7. Badgujar KC, Badgujar AB, Dhangar DV, Badgujar VC. Importance and use of pulse oximeter in COVID-19 pandemic: general factors affecting the sensitivity of pulse oximeter. Indian Chemical Engineer 2020 Nov 27;62(4):374-384. [doi: 10.1080/00194506.2020.1845988]
8. Enoch AJ, English M, Clinical Information Network, McGivern G, Shepperd S. Variability in the use of pulse oximeters with children in Kenyan hospitals: a mixed-methods analysis. PLoS Med 2019 Dec 31;16(12):e1002987. [doi: 10.1371/journal.pmed.1002987] [Medline: 31891572]
9. Ginsburg AS, Tawiah Agyemang C, Ambler G, et al. mPneumonia, an innovation for diagnosing and treating childhood pneumonia in low-resource settings: a feasibility, usability and acceptability study in Ghana. PLoS One 2016 Oct 27;11(10):e0165201. [doi: 10.1371/journal.pone.0165201] [Medline: 27788179]
10. Graham HR, Bakare AA, Gray A, et al. Adoption of paediatric and neonatal pulse oximetry by 12 hospitals in Nigeria: a mixed-methods realist evaluation. BMJ Glob Health 2018 Jun 26;3(3):e000812. [doi: 10.1136/bmjgh-2018-000812] [Medline: 29989086]
11. King C, Boyd N, Walker I, et al. Opportunities and barriers in paediatric pulse oximetry for pneumonia in low-resource clinical settings: a qualitative evaluation from Malawi and Bangladesh. BMJ Open 2018 Jan 30;8(1):e019177. [doi: 10.1136/bmjopen-2017-019177] [Medline: 29382679]
12. Sheikh M, Ahmad H, Ibrahim R, Nisar I, Jehan F. Pulse oximetry: why oxygen saturation is still not a part of standard pediatric guidelines in low-and-middle-income countries (LMICs). Pneumonia (Nathan) 2023 Feb 5;15(1):3. [doi: 10.1186/s41479-023-00108-6] [Medline: 36739442]
13. Zhang J, Tüshaus L, Nuño Martínez N, et al. Data integrity–based methodology and checklist for identifying implementation risks of physiological sensing in mobile health projects: quantitative and qualitative analysis. JMIR Mhealth Uhealth 2018 Dec 14;6(12):e11896. [doi: 10.2196/11896] [Medline: 30552079]
14. Sarin E, Kumar A, Alwadhi V, Saboth P, Kumar H. Experiences with use of a pulse oximeter multimodal device in outpatient management of children with acute respiratory infection during COVID pandemic. J Family Med Prim Care 2021 Feb;10(2):631-635. [doi: 10.4103/jfmpc.jfmpc_1410_20] [Medline: 34041052]
15. Keogh BF, Kopotic RJ. Recent findings in the use of reflectance oximetry: a critical review. Curr Opin Anaesthesiol 2005 Dec;18(6):649-654. [doi: 10.1097/01.aco.0000191893.21444.56] [Medline: 16534307]
16. Fleming KA, Horton S, Wilson ML, et al. The Lancet Commission on diagnostics: transforming access to diagnostics. Lancet 2021 Nov 27;398(10315):1997-2050. [doi: 10.1016/S0140-6736(21)00673-5] [Medline: 34626542]
17. Aranda-Jan CB, Jagtap S, Moultrie J. Towards a framework for holistic contextual design for low-resource settings. International Journal of Design 2016 Dec 31;10(3):43-63 [FREE Full text]
18. Bartlett R, Boyle JA, Simons Smith J, Khan N, Robinson T, Ramaswamy R. Evaluating human-centred design for public health: a case study on developing a healthcare app with refugee communities. Res Involv Engagem 2021 May 30;7(1):32. [doi: 10.1186/s40900-021-00273-2] [Medline: 34053451]
19. Boyd N, King C, Walker IA, et al. Usability testing of a reusable pulse oximeter probe developed for health-care workers caring for children <5 years old in low-resource settings. Am J Trop Med Hyg 2018 Oct;99(4):1096-1104. [doi: 10.4269/ajtmh.18-0016] [Medline: 30141389]
20. Bazzano AN, Martin J, Hicks E, Faughnan M, Murphy L. Human-centred design in global health: a scoping review of applications and contexts. PLoS One 2017 Nov 1;12(11):e0186744. [doi: 10.1371/journal.pone.0186744] [Medline: 29091935]
21. Fisher M, Johansen E. Human-centered design for medical devices and diagnostics in global health. Global Health Innovation 2020 Mar 22;3(1):1-15. [doi: 10.15641/ghi.v3i1.762]
22. Taylor Salisbury T, Atmore KH, Nhambongo I, et al. Integrating human-centred design into the development of an intervention to improve the mental wellbeing of young women in the perinatal period: the Catalyst project. BMC Pregnancy Childbirth 2021 Mar 5;21(1):183. [doi: 10.1186/s12884-021-03675-y] [Medline: 33673826]
23. Asiedu A, Nelson AR, Gomez PP, Tappis H, Effah F, Allen C. “It builds your confidence… you've done well”: healthcare workers' experiences of participating in a low-dose, high-frequency training to improve newborn survival on the day of birth in Ghana. Gates Open Res 2019 May 22;3:1470. [doi: 10.12688/gatesopenres.12936.1] [Medline: 31410394]
24. Triplett NS, Munson S, Mbwayo A, et al. Applying human-centered design to maximize acceptability, feasibility, and usability of mobile technology supervision in Kenya: a mixed methods pilot study protocol. Implement Sci Commun 2021 Jan 7;2(1):2. [doi: 10.1186/s43058-020-00102-9] [Medline: 33413688]
25. Liu C, Lee JH, Gupta AJ, et al. Cost-effectiveness analysis of human-centred design for global health interventions: a quantitative framework. BMJ Glob Health 2022 Mar;7(3):e007912. [doi: 10.1136/bmjgh-2021-007912] [Medline: 35346954]
26. Population data. Western Cape Government. 2020. URL: https://www.westerncape.gov.za/assets/departments/health/h_102_2020_covid-19_population_data.pdf [accessed 2024-05-17]
27. City of Cape Town - 2011 Census suburb Khayelitsha. City of Cape Town. 2013 Jul. URL: https://resource.capetown.gov.za/documentcentre/Documents/Maps%20and%20statistics/2011_Census_CT_Suburb_Khayelitsha_Profile.pdf [accessed 2024-05-17]

28. World Health Organization, Alliance for Health Policy and Systems Research. Primary Health Care Systems (PRIMASYS) : case study from South Africa. World Health Organization. 2017. URL: https://iris.who.int/handle/10665/341145 [accessed 2024-05-17]
29. Lim J, van Loggerenberg E, Chater R. Country profile: South Africa. Bertha Centre for Social Innovation & Entrepreneurship. 2016. URL: https://socialinnovationinhealth.org/2016/12/01/country-profile-south-africa/ [accessed 2024-05-20]
30. Kautzky K, Tollman SM. A perspective on primary health care in South Africa. In: Barron P, Roma-Reardon J, editors. South African Health Review 2008: Health Systems Trust; 2008:17-30 URL: https://www.hst.org.za/publications/South%20African%20Health%20Reviews/sahr2008.pdf [accessed 2024-05-17]
31. Giacomin J. What is human centred design? Des J (Aldershot) 2014 Dec;17(4):606-623. [doi: 10.2752/175630614X14056185480186]
32. Sandholdt CT, Cunningham J, Westendorp RGJ, Kristiansen M. Towards inclusive healthcare delivery: potentials and challenges of human-centred design in health innovation processes to increase healthy aging. Int J Environ Res Public Health 2020 Jun 24;17(12):4551. [doi: 10.3390/ijerph17124551] [Medline: 32599806]
33. Maguire M. Methods to support human-centred design. Int J Hum Comput Stud 2001 Oct;55(4):587-634. [doi: 10.1006/ijhc.2001.0503]
34. Stebbins RA. Exploratory Research in the Social Sciences: SAGE Publications, Inc; 2001. [doi: 10.4135/9781412984249]
35. Secure medical chat & patient referrals. Vula mobile. URL: https://www.vulamobile.com/ [accessed 2023-10-24]
36. Practice guidelines. Emercency Medicine Society South Africa. URL: https://emssa.org.za/practice-guidelines/ [accessed 2023-10-24]
37. Marks IH, Thomas H, Bakhet M, Fitzgerald E. Medical equipment donation in low-resource settings: a review of the literature and guidelines for surgery and anaesthesia in low-income and middle-income countries. BMJ Glob Health 2019 Sep 29;4(5):e001785. [doi: 10.1136/bmjgh-2019-001785] [Medline: 31637029]
38. Baker K, Ward C, Maurel A, et al. Usability and acceptability of a multimodal respiratory rate and pulse oximeter device in case management of children with symptoms of pneumonia: a cross-sectional study in Ethiopia. Acta Paediatr 2021 May;110(5):1620-1632. [doi: 10.1111/apa.15682] [Medline: 33220086]

Abbreviations

HCD: human-centered design
HCW: health care worker
LMIC: low- and middle-income country
PHC: primary health care
SpO_2: oxyhemoglobin saturation

Edited by C Jacob; submitted 29.11.23; peer-reviewed by D Dunsmuir, K Bochtler; revised version received 23.04.24; accepted 01.05.24; published 30.05.24.

Please cite as:
Ilhan EI, Jola LN, van der Zalm MM, Bernstein M, Goussard P, Redfern A, Hesseling AC, Hoddinott G, McCollum ED, King C
Designing a Smartphone-Based Pulse Oximeter for Children in South Africa (Phefumla Project): Qualitative Analysis of Human-Centered Design Workshops With Health Care Workers
JMIR Hum Factors 2024;11:e54983
URL: https://humanfactors.jmir.org/2024/1/e54983
doi:10.2196/54983

Original Paper

User-Centered Design and Usability of a Culturally Adapted Virtual Survivorship Care App for Chinese Canadian Prostate Cancer Survivors: Qualitative Descriptive Study

Karen Young[1,2], BCom, MSc; Ting Xiong[1,2], BSc, MSc; Rachel Lee[1], BSc, MHI; Ananya Tina Banerjee[3], PhD; Myles Leslie[4], PhD; Wellam Yu Ko[5], PhD; Quynh Pham[1,2,6,7,8], PhD

[1]Centre for Digital Therapeutics, Techna Institute, University Health Network, Toronto, ON, Canada
[2]Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[3]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada
[4]School of Public Policy, University of Calgary, Calgary, AB, Canada
[5]Men's Health Research Program, University of British Columbia, Vancouver, BC, Canada
[6]Toronto General Hospital Research Institute, University Health Network, Toronto, ON, Canada
[7]Telfer School of Management, University of Ottawa, Ottawa, ON, Canada
[8]School of Public Health Sciences, University of Waterloo, Waterloo, ON, Canada

Corresponding Author:
Quynh Pham, PhD
Centre for Digital Therapeutics
Techna Institute
University Health Network
Toronto General Hospital, R. Fraser Elliott Building, 4th Floor
190 Elizabeth Street
Toronto, ON, M5G 2C4
Canada
Phone: 1 (416) 340 4800 ext 4765
Email: q.pham@uhn.ca

Abstract

Background: Cultural adaptations of digital health innovations are a growing field. However, digital health innovations can increase health inequities. While completing exploratory work for the cultural adaptation of the *Ned* Clinic virtual survivorship app, we identified structural considerations that provided a space to design digitally connected and collective care.

Objective: This study used a community-based participatory research and user-centered design process to develop a cultural adaptation of the *Ned* Clinic app while designing to intervene in structural inequities.

Methods: The design process included primary data collection and qualitative analysis to explore and distill design principles, an iterative design phase with a multidisciplinary team, and a final evaluation phase with participants throughout the design process as a form of member checking and validation.

Results: Participants indicated that they found the final adapted prototype to be acceptable, appropriate, and feasible for their use. The changes made to adapt the prototype were not specifically culturally Chinese. Instead, we identified ways to strengthen connections between the survivor and their providers; improve accessibility to resources; and honor participants' desires for relationality, accountability, and care.

Conclusions: We grounded the use of user-centered design to develop a prototype design that supports the acts of caring through digital technology by identifying and designing to resist structures that create health inequities in the lives of this community of survivors. By designing for collective justice, we can provide accessible, feasible, and relational care with digital health through the application of Indigenous and Black feminist ways of being and knowing.

(JMIR Hum Factors 2024;11:e49353) doi:10.2196/49353

KEYWORDS

digital health; virtual care; digital therapeutics; prostate cancer; cancer survivorship; user-centred design; usability; supportive care; cultural adaptation; Chinese Canadians

Introduction

Digital health has been posited as a pathway to more equitable and holistic care [1,2]. However, the digital divide, or the capacity for digital technology to exacerbate inequities, has been widely described [3]. Its differential impacts on the social determinants of health are known as the digital determinants of health [4]. Recent years have seen an acceleration of digital health innovations (DHIs) such as digital therapeutics into health care systems, which was supercharged by the COVID-19 pandemic and the resulting widespread implementation of telemedicine [4]. One such digital therapeutic is the *Ned* Clinic ("No Evidence of Disease"), which aims to optimize clinical care and patient self-management through virtual asynchronous care delivery for prostate cancer (PCa) survivors [5]. The *Ned* Clinic platforms, including clinician-led (*Specialist Ned*) and nurse-led (*Ned Nurse*) interventions, were developed at the University Health Network in Toronto, Canada, by a consortium of stakeholders [5].

PCa is the most commonly diagnosed nonskin cancer for Canadian male individuals, and most (99%) are estimated to be diagnosed in male individuals aged 50 years and older [6]. Older adults are negatively impacted by the digital divide [7]. Race, a social determinant of health, is also linked to worse survivorship and care outcomes for PCa survivors, most notably for Black male individuals [8]. Asian (generally defined as East Asian and South Asian ethnicity) male individuals have been found to have better survival rates than the median but are more likely to present with advanced PCa, suggesting systemic issues with identifying health issues and obtaining timely appropriate care [9]. These differences carry over into the delivery of follow-up care, as PCa survivors' care needs and access to care are affected by the complex intersection of ethnicity, culture, and other social and structural factors [10,11].

Cultural adaptation is the process of applying changes to existing health interventions based on "surface" (social and behavioral characteristics) and "deep" (worldview, norms, beliefs, and values) cultural structures [12]. As these structures are known to impact beliefs about illness and well-being, the intent is to provide intervention benefits for communities that have experienced health inequities [13]. Culturally adapted DHIs appear to have been most widely reported in the field of mental health; in contrast, cultural adaptations of cancer survivorship apps have not been published, likely owing to the few interventions in this area [2,14]. The frameworks that appear to be most widely used to adapt health interventions were developed by Bernal et al [15], Resnicow et al [16] (an adaptation of the model by Bernal et al [15]), and Barrera and Castro [17].

However, these guidelines and models often use framings of cultural sensitivity and competency (eg, Resnicow et al [16] and Castro et al [18]), continuing to place the burden of change on individuals rather than addressing the upstream structural determinants of health. These framings can serve to "museumize" and problematize identity categories and culture as causes of ill-health, echoing the long-standing use of culture as a scapegoat to fault specific communities for health inequities. Moreover, defining "culture" for such adaptations can be a complex process in Canada, where culture, race, ethnicity, settler colonialism, and white supremacy (ie, the social and structural determinants of health) all create intersectional and differential lived experiences under a putatively shared identity—Canadian [19-21].

This research reports on the second and final phase of a project to design a cultural adaptation of the patient-facing *Ned* Clinic virtual follow-up care app for Chinese Canadian PCa survivors. In phase 1, we completed formative work distilling a set of themes relevant to survivors' user needs for follow-up and virtual care. Following the user-centered design (UCD) framework, we describe the results of the design and formative evaluation of a culturally adapted prototype of the app.

Methods

Study Design

The overall qualitative descriptive study design was structured using the community-based participatory research (CBPR) and UCD frameworks [22-24]. This study was conducted at the University of Toronto between December 2022 and March 2023 during the COVID-19 pandemic. For communities that face barriers to care, it was found that CBPR practices such as our engagement of a key informant and invitations to community members to share their lived experiences through open-ended interviews are appropriate [1,25]. CBPR concepts were applied to meaningfully involve the community (including several authors of this study) and return the results for their benefit. Here, *community* represents a "symbolic totality as well as a practical multiplicity," as the Chinese Canadian community is highly heterogeneous [26]. We view our participants as a coalition of self-identified Chinese Canadian individuals impacted by PCa survivorship to attend to their differences.

The Chinese Canadian community is an immigrant community that exists as a result of settler colonialism. In recognizing this, we redefine "immigrants" as "people with ancestral roots outside of Indigenous lands, who are beholden to Indigenous laws and epistemologies" [27]. This definition led us to apply a relational paradigm to this project and an axiology of relational accountability. It also provided a pathway to apply several multilevel Indigenous and Black feminist theorizations, guiding principles, and tools [27-29]. These included decolonial theory, *Etuaptmumk* (two-eyed seeing), intersectionality, and cultural safety to inform our conceptualization of digital space as intimately related to land [27,30,31]. This approach allowed us to contextualize the place-related experiences of our participants and uncover their desires for relational and culturally safe care [32]. We noted that these desires are not specifically Chinese, and this presented an opportunity to design for relationally connected digital health.

UCD is a flexible, iterative, and evidence-based 3-step design process framework that consults, involves, and considers the needs of the end user throughout the entire project [23]. Phase 1 of this study encompasses steps 1 and 2; phase 2 encompasses steps 2 and 3. We present this study according to the Consolidated Criteria for Reporting Qualitative Research (COREQ) guidelines [33].

Step 1: Ideation and Concept Generation

To contextualize the potential use of this app, we sought to understand the structures that impact Chinese Canadian PCa survivors' experiences with follow-up care and virtual care. The results of this phenomenologically informed exploratory-descriptive qualitative study are described elsewhere [34]. Based on the findings of this formative research, we synthesized a list of design principles (Table 1), which we then categorized into the cultural adaptation taxonomy created by Spanhel et al [14] to systematically adapt the patient-facing prototype.

Table 1. Summary of design principles for the adaptation of the *Ned Nurse* patient-facing app.

Research finding	Design principle	Taxonomic classification [14][a]
PHI[b] freedom: patients felt that they were expected to track and remember overwhelming amounts of information.	The system should automatically update, store, and provide access to PHI on demand.	Content components: • (9) Goals of treatment • (10) Methods of treatment
Access to personalized education and information: patients felt that they were unable to access information about their care options and disease status.	The system should provide access to personalized and evidence-based information regarding staging, self-management, and treatment options.	Content components: • (9) Goals of treatment • (10) Methods of treatment
Continuity of care: patients desired a connection with their provider and the ability to communicate during times of need.	The system should improve accessibility and continuity of care, as strong care relationships create a sense of safety.	Content components: • (9) Goals of treatment • (10) Methods of treatment
Security: patients expressed suspicion about digital health because they had concerns about surveillance and security.	The system should be architected and built with a high level of security and privacy.	Methodological components: • (12) Functionality
Accessibility: patients wanted to access care in readable and accessible language formats.	The system should provide readable language and accessible language formats.	Content components: • (5) Language translation • (6) Language tailoring
Digital literacy: patients felt comfortable with their device of choice but desired simplicity, form over function, and accessible help and documentation.	The system should prioritize usability, provide straightforward instruction and support, and maintain simple user interface and user experience design.	Methodological components: • (11) Structure • (12) Functionality • (13) Design and aesthetics
Care coordination: patients felt like they were expected to coordinate their care, as communication between specialists, primary care, and other services were fragmented.	The system should coordinate and provide a clear follow-up appointment schedule.	Content components: • (9) Goals of treatment • (10) Methods of treatment
Resources: patients felt unable to access, refused, or unaware of needed resources such as mental health support.	The system should provide accessible pathways to resources, such as psychological support, supportive care, and financial support.	Content components: • (8) Difference in concepts of mental health and its treatment • (9) Goals of treatment • (10) Methods of treatment

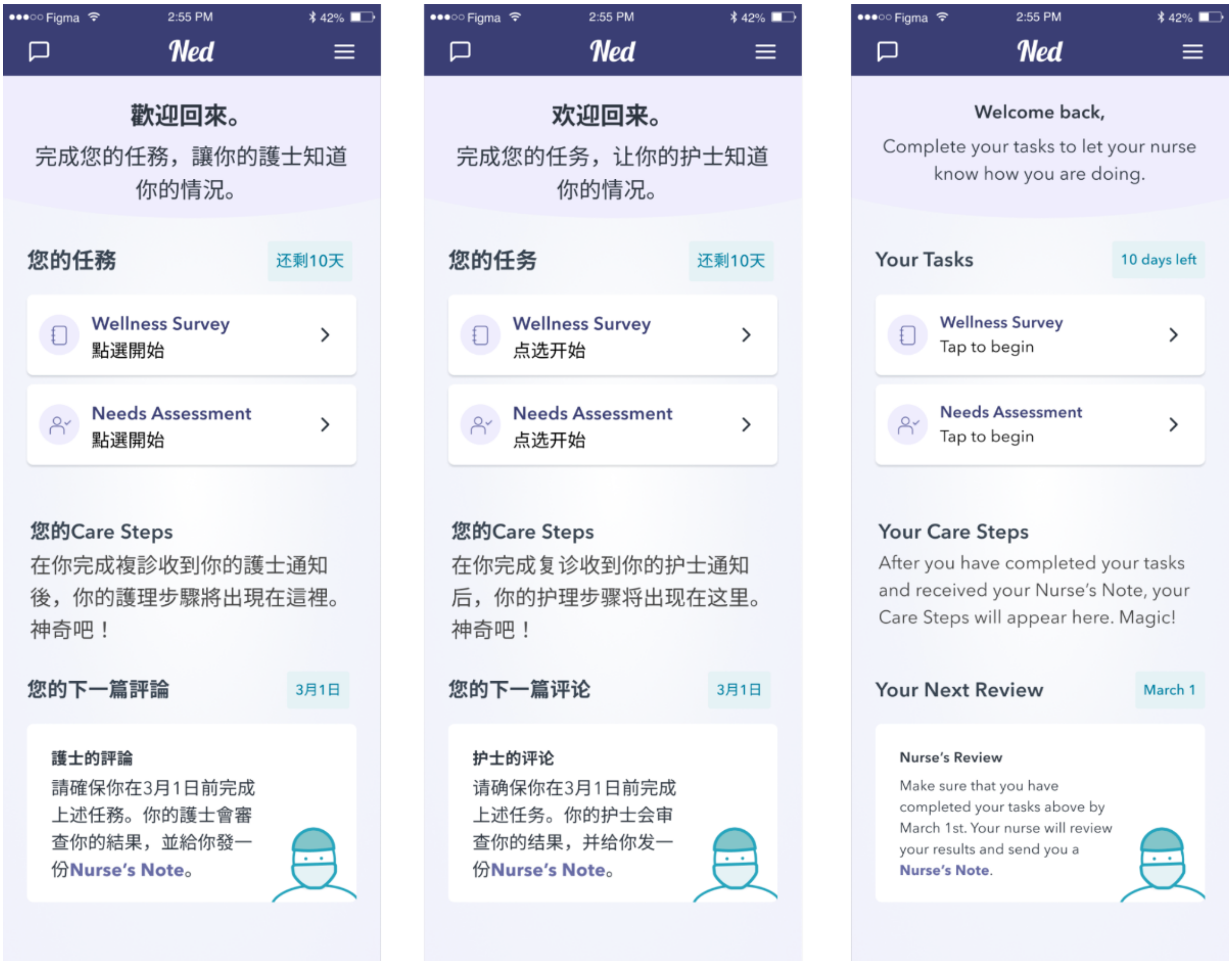
[a]The design principles used to adapt the *Ned* Clinic patient app identified here are classified to the corresponding taxonomic components found in Spanhel et al [14].

[b]PHI: personal health information.

Step 2: Design and Development

We applied these design principles to adapt the *Ned Nurse* patient app for Chinese Canadian survivors. A composite profile of a sample representative user was created to situate the design team during the development of the wireframes. A list of 5 use scenarios was created to guide the adaptation. These scenarios encompassed the design principles created in step 1 and included actions such as completing follow-up tasks, accessing a follow-up care schedule, and using the app to chat with a clinician. All use scenarios are described in the interview guide (Multimedia Appendix 1). Then, the original *Ned Nurse* app wireframes were redesigned to reflect the features required to operationalize these scenarios through the app, resulting in a new prototype. The prototype was created in Figma (Figma Inc.) on an iPhone 13 (Apple Inc) interface. This initial adaptation was iteratively critiqued by a team of researchers and human factors designers to refine the content, user interface, and user experience. Once the adapted prototype was finalized, it was translated from English into written Chinese via the translation process outlined in Haldane et al [35]. This resulted in 3 versions of the adapted prototype in English, Simplified Chinese, and Traditional Chinese. The app home page in each language version is shown in Figure 1.

Figure 1. Wireframes of the adapted *Ned Nurse* homepage in all 3 language versions.



Step 3: Evaluation

Overview

We empirically evaluated the acceptability, appropriateness, and feasibility of the adaptation through a cultural safety lens [22]. These dimensions are early-stage implementation outcomes and have also been found to be core to the success of DHIs [36,37]. A moderated cognitive walkthrough approach and the think-aloud protocol were used to construct a semistructured interview guide encompassing the 5 scenarios describing usual tasks that an end user might complete through the app [38].

Usability Testing

Facilitators began each test by outlining the usability testing procedure and think-aloud protocol. Context regarding the intended use and deployment of the *Ned Nurse* system was provided. Participants were asked to complete a series of actions for each scenario on the prototype to evaluate its design and functionality. We asked participants to think and speak about improvements they desired during their evaluations. In situations where the participant was unable to access the prototype on their device, they were asked to state their intended actions using the think-aloud protocol to the facilitator, who completed the action in the prototype on their behalf.

Interviews were completed through Microsoft Teams or Zoom (Zoom Technologies Inc). Informed consent for this work was previously obtained as part of overall study consent from participants. Participants were provided with the choice of completing their interview in Cantonese, Mandarin, or English and were also able to choose which language they wished to test the prototype in. The results of each usability test were iteratively analyzed via content analysis. Audio recordings of the participant interviews were translated into English as needed, according to the translation process described previously. A deductive and inductive content analysis approach was used, in which analysis of the data was completed by coders (TX and KY) through a process including open coding, creating categories, and abstraction [39]. Recommendations were applied in real time to create a final prototype that incorporated feedback from each user over the course of usability testing.

Positionality

An important marker of excellent qualitative research is "sincerity" or positionality, which indicates that the researcher has thought about and is reflective and aware about their values, experiences, biases, and inclinations within their research [40]. Here, the lead researcher reports on their social position, personal experiences, and political and professional beliefs to center the active role that the researcher plays in the framing of the research problem, interpretation of data, methods used, and the reporting of the results [41].

KY is a health informatics trainee and second-generation Chinese Canadian settler who was born and raised in the Greater Vancouver Regional District (GVRD) by a working-class, first-generation immigrant family with roots in southeastern China. She does not have any direct experience with PCa and has not previously provided care for PCa survivors. KY works primarily from a relational paradigm, focusing on the structures, contexts, and relationships that shape the design, development, and implementation of digital therapeutics and health technologies. She led and participated in all study activities.

Setting and Place

This study was conducted in the GVRD, located on the current, unceded, and future territories of the (Tsleil-Waututh, Squamish, and Musqueam) First Nations. The GVRD is home to one of Canada's oldest and largest living Chinese communities, including persons and families whose stories and identities span multiple geographies and generations [42]. The lead (KY) and senior author (QP) established relations with a supportive care program that provides care for Chinese Canadian PCa survivors and a Chinese PCa support group in this area. A key community informant agreed to guide this study and review and approve study materials.

Ethical Considerations

Research ethics approval for this study was obtained from the University of Toronto research ethics board (Human Protocol #43145). Written and verbal informed consent to participate in both phases of the project was obtained from all participants prior to interviews via the REDCap tool (Research Electronic Data Capture; Vanderbilt University), hosted at the University of Toronto. All data collected and disseminated here have been de-identified. Participants were provided with an honorarium of $50.00 CAD ($37.65 USD) per hour in appreciation of their time.

Results

Demographics

Usability testing was performed by 6 user testers, convenience sampled from the pool of 14 survivors and partner-caregivers who participated in the first phase of work as a form of member checking. This sample was also informed by Nielsen-Norman usability testing guidelines [43]. The reasons for nonparticipation were not collected. To protect the privacy of the participants involved in this phase, a demographic overview of the overall research project is provided here. Of the 14 participants in the first phase of this project, all survivors identified as men (n=12, 86%), and all partner-caregivers identified as women (n=2, 14%). A total of 13 (93%) participants indicated that they spoke English as an additional language. Most made an income between CAD $15,000 (US $11,048) and CAD $100,000 (US $73,653; n=12, 86%), lived in an urban area (n=13, 93%), were married (n=12, 86%), were educated beyond high school (n=13, 93%), and were retired (n=9, 64%). A 50/50 split emerged between preferences for smartphone or desktop or laptop use. Most (n=10, 71%) self-rated as being comfortable with their device. Participants indicated that they had 2 or fewer smartphone health apps (n=13, 93%).

Phase 1: Ideation

Table 1 summarizes the user requirement findings that emerged from previous formative research in phase 1 of this project and their subsequent translation to design principles.

Phase 2: Design and Development

Overview of Ned Nurse

An overview of the *Ned Nurse* clinical trial protocol is described by Pham et al [5]. The findings from formative work on the perspectives of health care providers, patients from the wider PCa survivor community, and the service design of the platform are forthcoming. Briefly, *Ned Nurse* digitally operationalizes a nurse-led model of survivorship care. Patients complete a series of tasks or access resources designed to support them in their survivorship. The platform aims to facilitate holistic care for patient quality of life.

Overview of the Adapted Patient-Facing System

The patient-facing adaptation set 2 user-input "care tasks," a validated questionnaire (Expanded Prostate Cancer Index Composite-Clinical Practice [EPIC-CP]) and a needs assessment survey, to constitute a single *Ned Nurse* "review" [5,44]. Language within the app avoided wording such as appointment, visit, and so forth to clarify the differences between synchronous and asynchronous care encounters. The user interface and user experience were designed to draw the user's attention to these tasks on the homepage immediately after login. All features were accessible via an in-app hamburger menu.

User inputs to the questionnaire were triaged via a decision-tree algorithm [45]. The algorithm was designed to return in-app self-management resources within a progress note ("Nurse's Note") automatically available to the user after input submission. If the algorithm detected that the patient required further support, they were prompted to specify domains for follow-up and asked to select their preferred contact method. This action would flag this patient to the nurse for follow-up. Resource links would appear on the homepage after the note was read and cleared.

To ensure that patients were aware of their review schedule, a feature was designed to display the last date, frequency, and next date of their expected reviews. The name of the nurse in charge and an explanation of their *Ned Nurse* role were provided to strengthen the perceived connection between the user and the nurse. This feature also set expectations for manual response times and included a link to users' previous submissions for on-demand access.

Resources were made available in 3 separate categories: symptom self-management advice, PCa information and education, and support and programmatic resources. Within each category, resources were further categorized. For example, symptom management included resources for symptoms such as anxiety, urinary incontinence, and hot flashes. Each resource provided an overview; relevant self-management steps; off-app links; and the ability to email, print, or save the resource. The feature home page also sectioned resources saved by the user ("Saved Resources") and resources picked for the user ("Picked for Me") by their nurse.

All available and historical prostate-specific antigen and testosterone blood work results were made available in chronological order to the user on-demand in a separate feature. Finally, a chat feature was designed to explore whether users might find it useful. It incorporated both responses in English from an automated support assistant (chatbot) and manually submitted by the nurse. This feature was simulated for evaluation.

Phase 3: Evaluation

Of the 6 participants, 2 (33%) tested in Cantonese, 3 (50%) tested in English, and 1 (17%) tested in Mandarin. These ratios correspond to testing of the Traditional Chinese, English, and Simplified Chinese versions. We note that patients who completed their testing in 1 language were functional to fluent in 1 or all of the other languages and provided critique for multiple versions.

Overall, there was strong agreement that the adaptation presented here would be acceptable, appropriate, and feasible for use, with the exception of the chat feature. Participants agreed that this app would make them feel comfortable and safe by allowing them to have more control over their care, access to resources, and stronger connections to their providers. They were encouraged by its perceived ability to meet their needs by protecting their connection with their providers, leveraging the functional flexibility of digital health, and providing resources beyond what they currently accessed. It was particularly valuable that features could be accessed at their convenience, as some felt that their follow-ups were far too short to meet their needs. Overall, 5 (83%) of 6 participants indicated that the level of support provided by this app was beneficial enough that it should be offered to patients prior to beginning treatment, or even at the point of diagnosis.

Participants' critiques centered on expanding flexibility, access to information, and streamlining responses. They felt that responses for some assessment questions (from 4 to 8 options) were overwhelming and should be reduced (3/6, 50%). English-Chinese translations would increase self-confidence in navigating the health care system. Medication names were spotlighted as particularly difficult. This was noted as an opportunity to expand the app's personal health information (PHI) storage, as a feature containing self-reported PHI (including medications) would be helpful to reference. Pictures and videos were desired instead of textual explanations. Laboratory results were asked to be displayed graphed or with severity indicators by 1 participant, and a text size adjustment function was requested by another.

Support for sexual dysfunction was not requested explicitly but appeared to be implied (3/6, 50%). A sexual therapy resource section was requested by 1 participant. Another noted that they would be more comfortable with nurses gendered as men as they felt uneasy when discussing sexual dysfunction with women. A final participant was keen to indicate that sexual dysfunction was a major area of concern when completing the EPIC-CP questionnaire.

As resources could be accessed on demand, some indicated that more would be beneficial. However, other participants expressed that the number displayed in the prototype were more than sufficient, reflecting our previous study findings on the bifurcated information-seeking behaviors of Chinese Canadian PCa survivors. Participants were also asked if they might find having their imaging results helpful. Although the majority (4/6, 67%) said no, those who said yes (2/6, 33%) were keen on having this information, especially if they needed to travel outside of Canada.

The questionnaire and assessment were generally deemed to be acceptable by most participants (4/6, 33%), with several notable dissents (2/6, 33%). The EPIC-CP question regarding hormonal function was highlighted as confusing by some because the connection between hormonal function and fatigue was not readily apparent. The spiritual domain in the needs assessment was flagged, as some thought that it would not be appropriately addressed by the nurse. Those who felt uncomfortable with this domain noted that they would prefer speaking about these needs to a spiritual leader. Agreement on appropriate response times also varied.

The chat function was deemed possibly helpful but likely unnecessary (4/6, 67%). As all chat interactions were in English, participants who were not confident in their English communication skills felt that their use of this feature would be limited (3/6, 50%). Others felt reminded of troubleshooting cable services rather than feeling connected to their provider. It was emphasized that any opportunity to improve connections to their providers through the app would be appreciated.

Discussion

Principal Findings and Implications

This study provides an applied example of a DHI for Chinese Canadian PCa survivors, which is based on broader principles of collectivism and relationality from Indigenous and Black feminist theory. Our initial aim was to co-design a cultural adaptation of the *Ned* Clinic to provide compassionate care and meet the unmet needs of Chinese Canadian PCa survivors via digital health.

However, attending to cultural adaptation theory and the lived realities of settler colonialism identified gaps to interweave Indigenous and Black feminist teachings. We began by synthesizing design principles that surfaced as critical to our participants and their feelings of comfort and safety when receiving follow-up care. This allowed us to leverage digital health to strengthen relations between the survivor and their providers; improve accessibility to resources; and honor desires for relationality, accountability, and care [46,47]. Rather than adapting by defining Chinese Canadian culture, we co-designed to intervene in structural causes of health inequities created by settler colonial culture instead [21,48].

We applied *Etuaptmumk* by interweaving strengths from different ways of being and knowing, including those from Indigenous, Western, Chinese, and Black feminist traditions in relation to PCa follow-up and virtual care [27,30,31]. These included prioritizing relational care, accounting for the use of prostate-specific antigen screening as a recurrence monitoring tool, and the benefits of supportive care programs to create adaptation features [30,49]. The EPIC-CP validated questionnaire is a key part of clinical follow-up care, as it allows clinicians to identify possible areas of concern during follow-up [50]. The needs assessment addresses domains beyond clinical care, reflecting the holistic nature of the medicine wheel [51]. Access to resources includes education and guidance for the self-management of concerns across multiple domains. The app presents a "care contract" in the form of a schedule that clearly states the "terms" and dates of the user's follow-ups [52]. It also respects the user's privacy by providing access and allowing them to share their PHI on their terms [53]. Only key inputs are communicated for triage and response. Finally, language access is built into the app as a question of communication accessibility, rather than only culture.

This design approach and these features do not deny the fact that culture is a real influence and can be a source of strength in many peoples' lives. However, we must go beyond implicating culture when designing DHIs for communities made vulnerable and instead address the overarching and underlying structures that create health inequities. Our design approach looked "up" at these structural causes rather than looking "down" and museumizing culture for participants through cultural sensitivity and competency. We demonstrate that a structural approach that applies teachings such as cultural safety and intersectionality can result in DHIs that are found to be acceptable, appropriate, and feasible for use while still leaving room for users to self-define and practice culture on their own terms. We are supporting, not replacing, the *labor* and *acts* of caring with digital health. Beginning with a paradigm shift opened a window to design for collective care, a scalable opportunity to benefit communities beyond Chinese Canadians with this *Ned Nurse* patient-facing app adaptation.

Strengths and Limitations

We have created the first "cultural" adaptation of a PCa follow-up care application for Chinese Canadian survivors. We extended the accessibility of this prototype by offering it in 3 language versions and tested its validity through member checking by returning it to participants who had provided their experiences and expertise as part of the first phase of this project. The findings should be considered with some limitations. Our sample does not fully represent the Chinese Canadian PCa community, as the heterogeneity of the community makes it difficult to recruit a fully representative sample [42]. User testing did not differentiate between results derived from users who interacted with the app themselves and users who directed a facilitator to perform actions on their behalf. However, all participants received the same set of instructions to apply the think-aloud method. A broad description of our theoretical stance, setting and place, methods, and results are provided to enhance understanding. We think of and encourage the transferability of this research as to how it might be made meaningful (ie, valid) for other communities in places where they may be subject to similar constructs and patterns of oppression [32]. Finally, this study does not include the provider perspective, although *Ned* was developed with clinicians who provide follow-up care for patients from this community. Future studies should examine the clinician's perspective on the design and development of similar DHIs, including provision of care through these apps, acceptability and feasibility, and implementation readiness.

Conclusions

This study demonstrates the relationality of Indigenous and Black feminist ontologies, epistemologies, and methodologies to digital health design by providing a worked example of its empirical use for an adaptation of a PCa follow-up care app, the *Ned Nurse* Clinic, for Chinese Canadian PCa survivors. We applied UCD principles to develop a prototype design that supports the relational act of caring through digital technology by identifying structures that create inequities in the experiences of this community of survivors and designing to intervene and provide accessible, connected care instead. We hope that this prototype serves as a tool to help regenerate places of caring, as we have learned from Indigenous and Black feminist scholars' teachings on power, place, and digital technologies.

Acknowledgments

We are deeply grateful for the time and effort gifted to us by the community members who shared their stories. We acknowledge and appreciate the staff and members of the Prostate Cancer Supportive Care Program, including Dr C Higano and Ms M Sundar, and the Richmond Chinese Prostate Cancer Support Group (列治文華人前列腺癌支援網絡), led by Mr D Leung and advised by Dr W Yu Ko. We thank the *Ned* team at the Centre for Digital Therapeutics (University Health Network) for their efforts and commitment to delivering a DHI for Canadian prostate cancer survivors. We give special thanks to Laura Parente for her incisive and human-centered design guidance. The study was funded by the Canadian Institutes of Health Research (CIHR) through the CIHR Canadian Cancer Society Survivorship Grant (CCS: 706713; CIHR: 168606) and the Canada Graduate Scholarship program, as well as by AGE-WELL Networks of Centres of Excellence through the Early Professionals & Inspired Careers in AgeTech (EPIC-AT) Fellowship and Graduate Award. These agencies were not involved in the design or analysis of this study.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

KY, QP, ATB, ML, and WYK contributed to project conceptualization and study design. KY, RL, and TX contributed to data collection and analysis. KY prepared the first manuscript draft, with contributions from TX, WYK, ATB, ML, and QP. All authors contributed, reviewed, and approved the manuscript.

Conflicts of Interest

QP and the University Health Network (Toronto, Ontario) jointly own intellectual property rights to the *Ned* app. Under the respective agreements with their organizations, QP is entitled to personally benefit from any commercial use of the intellectual property.

Multimedia Appendix 1
Semistructured usability testing interview guide.
[DOCX File , 12 KB - humanfactors_v11i1e49353_app1.docx]

References

1. Pham Q, El-Dassouki N, Lohani R, Jebanesan A, Young K. The future of virtual care for older ethnic adults beyond the COVID-19 pandemic. J Med Internet Res 2022;24(1):e29876 [FREE Full text] [doi: 10.2196/29876] [Medline: 34994707]
2. Pham Q, Hearn J, Gao B, Brown I, Hamilton RJ, Berlin A, et al. Virtual care models for cancer survivorship. NPJ Digit Med 2020;3(1):113 [FREE Full text] [doi: 10.1038/s41746-020-00321-3] [Medline: 32923690]
3. Shaw J, Chandra S, Gordon D, Bickford J, Fujioka J, Yang R, et al. Digital health technologies for more equitable health systems: a discussion paper. Toronto (ON): University of Toronto Joint Centre for Bioethics, Women's College Hospital Institute for Health Systems Solutions and Virtual Care; 2020. URL: https://www.wchwihv.ca/assets/uploads/Digital_Health_and_Equity_Report___September_2020FINAL.pdf [accessed 2023-12-08]
4. Crawford A, Serhal E. Digital health equity and COVID-19: the innovation curve cannot reinforce the social gradient of health. J Med Internet Res 2020;22(6):e19361 [FREE Full text] [doi: 10.2196/19361] [Medline: 32452816]
5. Pham Q, Hearn J, Bender JL, Berlin A, Brown I, Bryant-Lukosius D, et al. Virtual care for prostate cancer survivorship: protocol for an evaluation of a nurse-led algorithm-enhanced virtual clinic implemented at five cancer centres across Canada. BMJ Open 2021;11(4):e045806 [FREE Full text] [doi: 10.1136/bmjopen-2020-045806] [Medline: 33883153]
6. Canadian Cancer Statistics Advisory Committee in collaboration with the Canadian Cancer Society, Statistics Canada and the Public Health Agency of Canada. Canadian Cancer Statistics 2021. Canadian Cancer Society. 2021. URL: http://cancer.ca/Canadian-Cancer-Statistics-2021-EN [accessed 2023-12-08]
7. Latulippe K, Hamel C, Giroux D. Social health inequalities and eHealth: a literature review with qualitative synthesis of theoretical and empirical studies. J Med Internet Res 2017;19(4):e136 [FREE Full text] [doi: 10.2196/jmir.6731] [Medline: 28450271]
8. Bernard B, Muralidhar V, Chen YH, Sridhar SS, Mitchell EP, Pettaway CA, et al. Impact of ethnicity on the outcome of men with metastatic, hormone-sensitive prostate cancer. Cancer 2017;123(9):1536-1544 [FREE Full text] [doi: 10.1002/cncr.30503] [Medline: 28055108]
9. Chao GF, Krishna N, Aizer AA, Dalela D, Hanske J, Li H, et al. Asian Americans and prostate cancer: a nationwide population-based analysis. Urol Oncol 2016;34(5):233.e7-233.e15 [FREE Full text] [doi: 10.1016/j.urolonc.2015.11.013] [Medline: 26725248]
10. Rivas C, Matheson L, Nayoan J, Glaser A, Gavin A, Wright P, et al. Ethnicity and the prostate cancer experience: a qualitative metasynthesis. Psychooncology 2016;25(10):1147-1156 [FREE Full text] [doi: 10.1002/pon.4222] [Medline: 27416079]
11. Wall D, Kristjanson L. Men, culture and hegemonic masculinity: understanding the experience of prostate cancer. Nurs Inq 2005;12(2):87-97. [doi: 10.1111/j.1440-1800.2005.00258.x] [Medline: 15892724]
12. Castro FG, Barrera M, Martinez CR. The cultural adaptation of prevention interventions: resolving tensions between fidelity and fit. Prev Sci 2004;5(1):41-45. [doi: 10.1023/b:prev.0000013980.12412.cd] [Medline: 15058911]
13. Barrera M, Castro FG, Strycker LA, Toobert DJ. Cultural adaptations of behavioral health interventions: a progress report. J Consult Clin Psychol 2013;81(2):196-205 [FREE Full text] [doi: 10.1037/a0027085] [Medline: 22289132]
14. Spanhel K, Balci S, Feldhahn F, Bengel J, Baumeister H, Sander LB. Cultural adaptation of internet- and mobile-based interventions for mental disorders: a systematic review. NPJ Digit Med 2021;4(1):128 [FREE Full text] [doi: 10.1038/s41746-021-00498-1] [Medline: 34433875]
15. Bernal G, Bonilla J, Bellido C. Ecological validity and cultural sensitivity for outcome research: issues for the cultural adaptation and development of psychosocial treatments with Hispanics. J Abnorm Child Psychol 1995;23(1):67-82. [doi: 10.1007/BF01447045] [Medline: 7759675]

16. Resnicow K, Baranowski T, Ahluwalia JS, Braithwaite RL. Cultural sensitivity in public health: defined and demystified. Ethn Dis 1999;9(1):10-21. [Medline: 10355471]
17. Barrera M, Castro FG. A heuristic framework for the cultural adaptation of interventions. Clin Psychol: Sci Pract 2006;13(4):311-316. [doi: 10.1111/j.1468-2850.2006.00043.x]
18. Castro FG, Barrera M, Steiker LKH. Issues and challenges in the design of culturally adapted evidence-based interventions. Annu Rev Clin Psychol 2010;6:213-239 [FREE Full text] [doi: 10.1146/annurev-clinpsy-033109-132032] [Medline: 20192800]
19. Snelgrove C, Dhamoon R, Corntassel J. Unsettling settler colonialism: the discourse and politics of settlers, and solidarity with Indigenous nations. Decolonization 2014;3(2):1-32 [FREE Full text]
20. Harvey M, Neff J, Knight KR, Mukherjee JS, Shamasunder S, Le PV, et al. Structural competency and global health education. Glob Public Health 2022;17(3):341-362. [doi: 10.1080/17441692.2020.1864751] [Medline: 33351721]
21. Czyzewski K. Colonialism as a broader social determinant of health. Int Indig Policy J 2011;2(1):1-16 [FREE Full text] [doi: 10.18584/iipj.2011.2.1.5]
22. Osei-Twum JA, Pulfer E, Banerjee AT. Exploring the applicability of indigenous cultural safety to immigrant health research. In: A Research Agenda for Migration and Health. UK, USA: Edward Elgar Publishing; 2019:138-158.
23. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;Suppl:49-56 [FREE Full text] [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
24. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000;23(4):334-340. [doi: 10.1002/1098-240x(200008)23:4<334::aid-nur9>3.0.co;2-g] [Medline: 10940958]
25. Pekmezaris R, Kozikowski A, Pascarelli B, Wolf-Klein G, Boye-Codjoe E, Jacome S, et al. A telehealth-delivered pulmonary rehabilitation intervention in underserved hispanic and African American patients with chronic obstructive pulmonary disease: a community-based participatory research approach. JMIR Form Res 2020;4(1):e13197 [FREE Full text] [doi: 10.2196/13197] [Medline: 32012039]
26. Edelman EA. Beyond resilience. Soc Text Duke University Press 2020;38(1):109-130. [doi: 10.1215/01642472-7971127]
27. Tuck E, Wayne YK. Decolonization is not a metaphor. Decolonization 2012;1(1):1-40 [FREE Full text]
28. Wilson S. Research is Ceremony: Indigenous Research Methods. Black Point, Nova Scotia, Canada: Fernwood Publishing Co Ltd; 2008.
29. King TL. The Black Shoals: Offshore Formations of Black and Native Studies. Durham, North Carolina, United States: Duke University Press; 2019.
30. Bartlett C, Marshall M, Marshall A. Two-eyed seeing and other lessons learned within a co-learning journey of bringing together Indigenous and mainstream knowledges and ways of knowing. J Environ Stud Sci 2012;2(4):331-340. [doi: 10.1007/s13412-012-0086-8]
31. Nash JC, Pinto S, editors. The Routledge Companion to Intersectionalities. 1st Edition. Milton Park, Abingdon-on-Thames, Oxfordshire United Kingdom: Taylor & Francis Group; 2022.
32. Tuck E, McKenzie M. Place in Research: Theory, Methodology, and Methods. Milton Park, Abingdon-on-Thames, Oxfordshire, England, UK: Routledge; 2014.
33. Tong A, Sainsbury P, Craig J. Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
34. Young K, Xiong T, Lee R, Banerjee AT, Leslie M, Ko WY, et al. Honouring the care experiences of Chinese Canadian prostate cancer survivors to cultivate cultural safety and relationality in digital healthxploratory-descriptive qualitative study. J Med Internet Res Preprint posted online on November 28, 2023. [FREE Full text] [doi: 10.2196/49349]
35. Haldane V, Li BP, Ge S, Huang JZ, Huang H, Sadutshang L, et al. Exploring the translation process for multilingual implementation research studies: a collaborative autoethnography. BMJ Glob Health 2022;7(5):e008674 [FREE Full text] [doi: 10.1136/bmjgh-2022-008674] [Medline: 35636804]
36. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
37. Perski O, Short CE. Acceptability of digital health interventions: embracing the complexity. Transl Behav Med 2021;11(7):1473-1480 [FREE Full text] [doi: 10.1093/tbm/ibab048] [Medline: 33963864]
38. Jaspers MWM. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. Int J Med Inform 2009;78(5):340-353 [FREE Full text] [doi: 10.1016/j.ijmedinf.2008.10.002] [Medline: 19046928]
39. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
40. Tracy SJ. Qualitative quality: eight “big-tent” criteria for excellent qualitative research. Qual Inq 2010;16(10):837-851. [doi: 10.1177/1077800410383121]
41. Creswell JW, Poth CN. Qualitative Inquiry and Research Design: Choosing Among Five Approaches. Thousand Oaks, California, United States: SAGE Publications; 2016.

42. Li JTY. The Transcultural Streams of Chinese Canadian Identities. Quebec, Canada: McGill-Queen's Press - MQUP; 2019.
43. Nielsen J. Why you only need to test with 5 users. Nielsen Norman Group. 2000. URL: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/ [accessed 2023-11-26]
44. Chang P, Szymanski KM, Dunn RL, Chipman JJ, Litwin MS, Nguyen PL, et al. Expanded prostate cancer index composite for clinical practice: development and validation of a practical health related quality of life instrument for use in the routine clinical care of patients with prostate cancer. J Urol 2011;186(3):865-872. [doi: 10.1016/j.juro.2011.04.085] [Medline: 21788038]
45. Pfisterer KJ, Lohani R, Janes E, Ng D, Wang D, Bryant-Lukosius D, et al. An actionable expert-system algorithm to support nurse-led cancer survivorship care: algorithm development study. JMIR Cancer 2023;9:e44332 [FREE Full text] [doi: 10.2196/44332] [Medline: 37792435]
46. Arvin M, Tuck E, Morrill A. Decolonizing feminism: challenging connections between settler colonialism and heteropatriarchy. ff 2013;25(1):8-34. [doi: 10.1353/ff.2013.0006]
47. Gaudet JC. Keeoukaywin: the visiting way—fostering an Indigenous research methodology. aps 2019;7(2):47-64 [FREE Full text] [doi: 10.5663/aps.v7i2.29336]
48. Siddiqi A, Hertzman C, Smith BT. The fundamental role of socioeconomic resources for health and health behaviors. In: Fisher EB, Cameron LD, Christensen AJ, Ehlert U, Guo Y, Oldenburg B, et al, editors. Principles and Concepts of Behavioral Medicine: A Global Handbook. New York, NY: Springer New York; 2018:389-413.
49. Roher SIG, Yu Z, Martin DH, Benoit AC. How is Etuaptmumk/Two-Eyed Seeing characterized in Indigenous health research? a scoping review. PLoS One 2021;16(7):e0254612 [FREE Full text] [doi: 10.1371/journal.pone.0254612] [Medline: 34283831]
50. Lee TK, Poon DMC, Ng ACF, Ho T, Singh-Carlson S, Joffres M, et al. Cultural adaptation and validation of the Chinese version of the expanded prostate cancer index composite. Asia Pac J Clin Oncol 2018;14(Suppl 1):10-15. [doi: 10.1111/ajco.12855] [Medline: 29508938]
51. Dapice AN. The medicine wheel. J Transcult Nurs 2006;17(3):251-260. [doi: 10.1177/1043659606288383] [Medline: 16757664]
52. Crosby C, Jakobsen JR. Disability, debility, and caring queerly. Soc Text 2020;38(4):77-103. [doi: 10.1215/01642472-8680454]
53. Ware P, Bartlett SJ, Paré G, Symeonidis I, Tannenbaum C, Bartlett G, et al. Using eHealth technologies: interests, preferences, and concerns of older adults. Interact J Med Res 2017;6(1):e3 [FREE Full text] [doi: 10.2196/ijmr.4447] [Medline: 28336506]

Abbreviations

CBPR: community-based participatory research
COREQ: Consolidated Criteria for Reporting Qualitative Research
DHI: digital health innovation
EPIC-CP: Expanded Prostate Cancer Index Composite-Clinical Practice
GVRD: Greater Vancouver Regional District
PCa: prostate cancer
PHI: personal health information
UCD: user-centered design

Edited by A Kushniruk; submitted 25.05.23; peer-reviewed by J Sharp, X Yan, Y Jain; comments to author 02.08.23; revised version received 05.11.23; accepted 20.11.23; published 01.01.24.

Please cite as:
Young K, Xiong T, Lee R, Banerjee AT, Leslie M, Ko WY, Pham Q
User-Centered Design and Usability of a Culturally Adapted Virtual Survivorship Care App for Chinese Canadian Prostate Cancer Survivors: Qualitative Descriptive Study
JMIR Hum Factors 2024;11:e49353
URL: https://humanfactors.jmir.org/2024/1/e49353
doi:10.2196/49353
PMID:38163295

Original Paper

Effectiveness and User Perception of an In-Vehicle Voice Warning for Hypoglycemia: Development and Feasibility Trial

Caterina Bérubé[1], PhD; Vera Franziska Lehmann[1,2], PhD; Martin Maritsch[1], PhD; Mathias Kraus[3], Prof Dr; Stefan Feuerriegel[4], Prof Dr; Felix Wortmann[1,5], Prof Dr; Thomas Züger[1,2,6], MD; Christoph Stettler[1,2], Prof Dr; Elgar Fleisch[1,5], Prof Dr; A Baki Kocaballi[7], PhD; Tobias Kowatsch[1,5,8,9], Prof Dr

[1]Centre for Digital Health Interventions, Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland
[2]Department of Diabetes, Endocrinology, Nutritional Medicine and Metabolism, Bern University Hospital, University of Bern, Bern, Switzerland
[3]School of Business, Economics and Society, Friedrich-Alexander-Universität Erlangen-Nürnberg, Nürnberg, Germany
[4]School of Management, Ludwig-Maximilians-Universität München, Munich, Germany
[5]Centre for Digital Health Interventions, Institute of Technology Management, University of St Gallen, St Gallen, Switzerland
[6]Department of Endocrinology and Metabolic Diseases, Kantonsspital Olten, Olten, Switzerland
[7]School of Computer Science, University of Technology Sydney, Sydney, Australia
[8]Institute for Implementation Science in Health Care, University of Zurich, Zurich, Switzerland
[9]School of Medicine, University of St Gallen, St Gallen, Switzerland

Corresponding Author:
Tobias Kowatsch, Prof Dr
Centre for Digital Health Interventions
Department of Management, Technology, and Economics
ETH Zurich
WEV G 216
Weinbergstrasse 56/58
Zurich, 8092
Switzerland
Phone: 41 44 632 94 88
Email: tkowatsch@ethz.ch

Abstract

Background: Hypoglycemia is a frequent and acute complication in type 1 diabetes mellitus (T1DM) and is associated with a higher risk of car mishaps. Currently, hypoglycemia can be detected and signaled through flash glucose monitoring or continuous glucose monitoring devices, which require manual and visual interaction, thereby removing the focus of attention from the driving task. Hypoglycemia causes a decrease in attention, thereby challenging the safety of using such devices behind the wheel. Here, we present an investigation of a hands-free technology—a voice warning that can potentially be delivered via an in-vehicle voice assistant.

Objective: This study aims to investigate the feasibility of an in-vehicle voice warning for hypoglycemia, evaluating both its effectiveness and user perception.

Methods: We designed a voice warning and evaluated it in 3 studies. In all studies, participants received a voice warning while driving. Study 0 (n=10) assessed the feasibility of using a voice warning with healthy participants driving in a simulator. Study 1 (n=18) assessed the voice warning in participants with T1DM. Study 2 (n=20) assessed the voice warning in participants with T1DM undergoing hypoglycemia while driving in a real car. We measured participants' self-reported perception of the voice warning (with a user experience scale in study 0 and with acceptance, alliance, and trust scales in studies 1 and 2) and compliance behavior (whether they stopped the car and reaction time). In addition, we assessed technology affinity and collected the participants' verbal feedback.

Results: Technology affinity was similar across studies and approximately 70% of the maximal value. Perception measure of the voice warning was approximately 62% to 78% in the simulated driving and 34% to 56% in real-world driving. Perception correlated with technology affinity on specific constructs (eg, Affinity for Technology Interaction score and intention to use, optimism and performance expectancy, behavioral intention, Session Alliance Inventory score, innovativeness and hedonic motivation, and negative correlations between discomfort and behavioral intention and discomfort and competence trust; all

P<.05). Compliance was 100% in all studies, whereas reaction time was higher in study 1 (mean 23, SD 5.2 seconds) than in study 0 (mean 12.6, SD 5.7 seconds) and study 2 (mean 14.6, SD 4.3 seconds). Finally, verbal feedback showed that the participants preferred the voice warning to be less verbose and interactive.

Conclusions: This is the first study to investigate the feasibility of an in-vehicle voice warning for hypoglycemia. Drivers find such an implementation useful and effective in a simulated environment, but improvements are needed in the real-world driving context. This study is a kickoff for the use of in-vehicle voice assistants for digital health interventions.

(JMIR Hum Factors 2024;11:e42823) doi:10.2196/42823

KEYWORDS

hypoglycemia; type-1 diabetes mellitus; in-vehicle voice assistant; voice interface; voice warning; digital health intervention; mobile phone

Introduction

Background

Type 1 diabetes mellitus (T1DM) is a chronic condition caused by an inability of the pancreas to produce insulin and requires lifelong insulin therapy [1]. Hypoglycemia, also known as low blood glucose, is a frequent and acute complication in patients with T1DM [2,3]. Symptoms range from autonomic reactions such as trembling, anxiety, and hunger (ie, mild hypoglycemia) to neuroglycopenic reactions such as vision impairment, weakness, or cognitive impairments (ie, severe hypoglycemia) [2,4-6]. Hypoglycemia is a major issue in the context of driving: research has shown that hypoglycemia is associated with a higher risk of car mishaps [7-9]. In fact, drivers experiencing hypoglycemia are recommended by the local authorities [10] to stop the car and treat their condition. However, drivers do not always comply with these recommendations [11,12]. Thus, to help reduce hypoglycemia-related car accidents, there should be an effective warning that informs the driver about an upcoming hypoglycemic episode and supports the driver in coping with the situation. Currently, hypoglycemia can be detected and signaled through flash glucose monitoring (FGM) or continuous glucose monitoring (CGM) devices (ie, wearable receivers connected to a sensor inserted in the subcutaneous tissue of the arm or abdomen) [13]. These allow for glucose monitoring by displaying the values either continuously (ie, CGM) or upon active retrieval (ie, FGM) and deliver alerts in the form of a tone or vibration in case of out-of-range values. However, these devices present limitations in the context of driving. For instance, FGM needs to be held close to the sensor to transfer the data from the subcutaneous sensors to the monitoring device, that is, the driver needs to actively engage in a manual gesture to access the glucose value and to look at a visual display moving the focus of attention from the driving task. In contrast, allowing the drivers to receive an alert in a hands-free mode will facilitate warning reception [14] and lower worry associated with driving with T1DM [15]. However, hypoglycemia is known to cause a decrease in attention [2,4-6], thereby challenging the effectiveness of such devices. As 90% of road accidents are caused by human error, the European Commission has set new safety technologies as mandatory equipment for vehicles as of 2022 (eg, driver drowsiness and distraction warnings and speed assistance) [16]. In-vehicle warning systems for impaired driver states, such as fatigue [17], distraction [18], and breath alcohol concentration [19], are increasingly being developed. However, to the best of our knowledge, there is no existing implementation for hypoglycemia. Such technology would be aligned with the "healing car" concept [20], where vehicles become environments promoting well-being for passengers, including ergonomic seats, ambient lighting, relaxation exercises [21], and detection of health-critical states [22]. This concept is still in its early stages, but it may become a standard in car manufacturing in the future. So far, the only attempts of in-vehicle glucose monitoring are either only proof of concept without user validation [23] or conceptual work [24]. However, the online community clearly expressed a need for in-vehicle glucose monitoring and warning [25].

A growing number of automotive companies are introducing voice assistance technology into their products [26,27]. Voice assistants add value not only for the associated consumer experience but also for their greater safety. Indeed, vocal interactions have been observed to be the least cognitively demanding while driving compared with visual and haptic interactions [28,29]. Moreover, voice assistants are increasingly being implemented to deliver digital health interventions [30-33]. Although research is still in its infancy, efforts have been made to develop voice-based conversational agents to monitor and support individuals with chronic diseases such as cancer, cardiovascular diseases, cognitive disorders, or diabetes [30]. Other recent examples include prevention of excessive alcohol consumption [34], health education and monitoring, physical and mental exercise, and nutrition [35]. Furthermore, a voice assistant delivering a warning is a form of proactive behavior initiated by the computer rather than the user [36,37]. In-vehicle voice assistants can provide personalized and adaptive suggestions, but users may ignore proactive behavior if it is inopportune, violates privacy, or distracts from driving [38-40]. However, emergencies are the most suitable context for proactive behavior that violates privacy [39].

Objectives

Therefore, we investigate the feasibility of an in-vehicle voice warning delivered by a built-in voice assistant to alert and support drivers with T1DM during hypoglycemia. To the best of our knowledge, there have been no investigations on safe and effective in-vehicle hypoglycemia warnings to support drivers with T1DM or on the perception of such technology. Thus, we sought to answer the following research questions (RQs):

- RQ1: How do drivers perceive an in-vehicle voice warning for hypoglycemia while driving?
- RQ2: How effective is an in-vehicle voice warning in prompting drivers to cope with hypoglycemia?

RQ1 refers to the attitude of drivers toward the warning, whereas RQ2 refers to the driver's compliance behavior once the warning is delivered. Answering these RQs will allow us to conclude on the feasibility of an in-vehicle voice warning for hypoglycemia. To control for individual factors influencing the perception of the warning [41], we also assessed technology affinity.

Methods

Study 0: Preliminary Assessment With Healthy Individuals in Simulated Driving

Driving Setting

Participants performed the task in a driving simulator (Carnetsoft Inc) with 3 monitors displaying the front, left, and right views. The central monitor also showed the cockpit and navigation arrows. The participants used a steering wheel and pedals (Logitech Driving Force G29) to control the simulator, which was set to automatic (ie, no clutch or gear shifter). The simulator's computer was connected to a stereo speaker with a subwoofer, which was kept at a constant volume. To control for driving difficulty, 3 environments were used: highway, countryside, and town, with the first and last being the least and most difficult, respectively.

In-Vehicle Voice Warning Simulation

Before testing a hypoglycemia voice warning with people with T1DM, we tested the concept of a car voice assistant as an interface between a dedicated monitoring system and the user with healthy participants. As the participants were not affected by hypoglycemia, the first version of the warning was a simulated low fuel warning ("The car needs a refill. Please pull over and turn off the engine"). Although not health related, it signaled an event of reasonable urgency that required safely stopping the car. Note that the participants were informed that this message aimed to ask them to stop the car as soon as possible and that they did not need to look for a gas station.

The warning was simulated using the Wizard-of-Oz method, where the conversational turns produced by the voice assistant were played by the experimenter [42] from a laptop using predefined keyboard keys. The turns were based on the Google Cloud text-to-speech engine, with a de-DE-Wavenet-C voice, a speed of 1.11 times the normal native speed of the specific voice, and a pitch of −1.20 semitones from the original pitch. The experimenter's computer was connected to the same sound system as the driving simulators so that the voice warning could be heard as part of the driving simulation. No visuals were included.

Voice Warning Evaluation Measures

To assess the RQs, we assessed participants' perception of the warning (self-reported through the modular evaluation of key Components of User Experience [meCUE]; 10 constructs evaluated on a 7-point Likert scale and a general evaluation evaluated on a 10-point scale [43,44]) and participant compliance with the warning (measured by the experimenter manually assessing if the participant would pull over and stop the car following the warning, and reaction time in seconds from the timestamp of the warning to the timestamp of the car fully stopped). As the perception of technology can be influenced by technology-related personality [45], we also measured technology affinity (measured by the Affinity for Technology Interaction [ATI], a 6-point Likert scale [46]). Finally, qualitative feedback was collected informally.

Evaluation Procedure

The participants were welcomed, informed about the procedure, and invited to sit in the simulator. The voice assistant introduced itself and invited the participants to familiarize themselves with the setting, including the 3 environments. The training also screened for motion sickness.

In the experimental session, participants drove 12 times, with 4 blocks of 3 drives each, for approximately 5 minutes per drive. The driving environment's order and starting point varied to minimize habituation. The drive began when the voice assistant prompted participants to start the engine. A timer started to deliver the low fuel warning at either 100 or 200 seconds to add variation and minimize habituation effects. At the end of the session, participants completed the meCUE.

Data Analysis

Participants were characterized by sex, age, and driver's license duration. The ATI was aggregated as a whole, and meCUE items were aggregated per construct. All reports were aggregated across the sample, with mean and SD. Compliance was coded as binary (0=not compliant, 1=compliant) and reported in terms of frequency. Reaction time was aggregated in seconds across participants and phases, with mean and SD.

Study 1: Assessment With Individuals With T1DM in Simulated Driving

Following the iterative approach described earlier, we conducted 3 exploratory iterations. This study was part of a clinical trial registered at ClinicalTrials.gov (NCT04035993).

Driving Setting

The driving setting was the same as in study 0.

In-Vehicle Voice Warning Simulation

On the basis of the results of study 0, we adapted the warning to hypoglycemia instead of low fuel, using the fewest conversational turns possible [47]. To ensure that the drivers were available, the voice assistant started with a receptivity check: "May I disturb you?"

We designed the warning based on the guidelines of the Swiss Diabetes Association [10], which recommends taking carbohydrates and stopping the car as soon as signs of hypoglycemia are noticed. To give the driver a sense of autonomy [48], we designed the warning to suggest eating carbohydrates rather than directly engaging in stopping the car. However, if the driver did not have carbohydrates, they were asked to pull over. On the basis of the feedback, we enhanced the voice warning used in the following study to recommend pulling over directly (detailed conversation flow is available in Multimedia Appendix 1).

As in study 0, the warning was simulated with a Wizard-of-Oz method [42], and the turns were generated by recording the same voice. However, to reduce fatigue and cognitive load, we decreased the speed and pitch to 0.93 times the normal speed and −4.8 semitones from the original pitch, respectively. As in study 0, the experimenter would play the turns from a Microsoft Windows laptop using predefined keyboard keys to play prerecorded voice sounds. However, in study 1, the laptop program included a visualization mirrored on a smartphone. The visuals consisted of a blue circle that gradually faded in and out when the voice assistant was speaking. As in study 0, the experimenter's computer was connected to the same sound system as the driving simulators, so that the voice assistant could be heard as part of the driving simulation.

Voice Warning Evaluation Measures

Perception assessment focused on evaluating the voice assistant as a trustworthy driving companion. Specifically, participants completed the Acceptance and Use of Technology (AUT) questionnaire [49,50], the Session Alliance Inventory (SAI) [51], and the Emotional Trust and Competence Trust subscales (henceforth Trust) of the Trust and Adoption questionnaire [52].

To assess technology affinity, participants completed the streamlined scale of the Technology Readiness Index (TRI 2.0) [53]. Items were rated on a 5-point Likert scale (ie, 1=totally disagree, 5=totally agree). We also added a question on whether the participants had previous experience with in-vehicle voice assistants (ie, "Have you already had experience with in-vehicle voice assistants?" with a yes or no answer).

Finally, to obtain qualitative and more in-depth feedback for improvement, we conducted a semistructured interview about their experience with the warning (the interview questions are provided in Multimedia Appendix 2).

Evaluation Procedure

The procedure was the same as in study 0, except that participants drove only once for 5 minutes (the evaluation procedure is detailed in Multimedia Appendix 3). Before driving, we ensured that the participants had normal blood glucose levels (5-8 mmol/L).

Data Analysis

The sample of participants was characterized by sex, mean age, and mean duration of their driver's license before the study.

TRI and SAI were aggregated as a whole, and the AUT and Trust items were aggregated per construct. Scores from the negatively formulated questionnaire items were inverted. Previous experience with an in-vehicle voice assistant was reported in terms of frequency. All these reports were aggregated across participants of each iteration, with mean and SD.

To further explain results in perception, they were associated with technology affinity measures. The difference in perception between participants with and without experience with an in-vehicle voice assistant was tested using a 2-sided *t* test, and it was correlated with the TRI constructs using a Pearson test.

Compliance was defined as whether the participant would comply with the warning and was coded as binary (0=did not comply, 1=complied). Reaction time was aggregated in seconds with mean and SD. Compliance behavior was aggregated across participants of each iteration.

Feedback was summarized in positive and negative topics, with a focus on the most prominent suggestions for improvement. Feedback was aggregated across participants of each iteration.

Study 2: Assessment With Individuals With T1DM in Real-World Driving Undergoing Hypoglycemia

Following the iterative approach described earlier, we conducted 2 exploratory iterations. This study was part of a clinical trial registered at ClinicalTrials.gov (NCT04569630).

Driving Setting

Participants drove in Volkswagen Touran on a closed circuit accompanied by a driving instructor. Dual pedals allowed the driving instructor to intervene and stop the car if necessary. The driving environments on the test track corresponded to the environments of the driving simulator used in the previous studies. Straight paths, turns, crossroads, stop signs, and a pedestrian crossing with a doll were used to implement the highway, countryside, and town scenarios. Artificial obstacles (eg, boxes and lines of traffic pylons) were used to simulate the traffic.

In-Vehicle Voice Warning Simulation

On the basis of the participant feedback from study 1, we revised the voice warning and addressed low trust ratings by explaining the cause of the warning. We simulated driving behavior as a trigger to detect hypoglycemia while driving, as in the study by Lehmann et al [54]. We created 2 variations of the simplified hypoglycemia notification—one with a statement of the cause (driving behavior) and one without. The final recommendation was reformulated as stricter but less directive than that in study 1.

In the second iteration, we simplified the conversational flow by removing the receptivity check ("May I disturb you?") and the final recommendation (Multimedia Appendix 1 provides the conversation flow).

We used the Wizard-of-Oz method to simulate the warning, as in studies 0 and 1. We implemented the voice assistant in a smartphone with the same voice as in study 1. However, the experimenter had to control it remotely (outside the car), so we implemented the interaction in a smartphone app controlled by a remote desktop application. The experimenter used the smartphone screen to control the voice warning delivery; therefore, no visualization was included. Because of network-related slowdowns in the remote control, we used a combination of remote control and speech-to-text programing.

Voice Warning Evaluation Measures

All measures were the same as in study 1. Reaction time was calculated from the warning onset until the car reached a velocity of 0. In addition, at the end of the experiment, we included a questionnaire item asking which of the 2 types of warning they preferred, that is, the warning including a statement of the cause that triggered the warning or the one without it, or if they would not use either of them.

Evaluation Procedure

After welcoming participants and explaining the procedure and simulated voice assistant, the voice assistant introduced itself as an in-vehicle assistant to support drivers with hypoglycemia. The participants then completed a training drive.

The warning was delivered at different stages of hypoglycemia (see the study by Lehmann et al [54]). Drive blocks were defined based on blood glucose levels. In the first phase, the participants drove at normal glucose (5-8 mmol/L). In the second phase, blood glucose level was progressively lowered below the moderate hypoglycemia threshold (3.0 mmol/L) to a target range of 2.0 to 2.5 mmol/L. In the third phase, moderate hypoglycemia was maintained. In the fourth phase, participants drove again with normal blood glucose levels (Multimedia Appendix 3).

To explore the effect of blood glucose level on warning perception and compliance, we delivered a warning at the end of the last drive of each phase. Participants received 2 warnings with an explanation and 2 without, in randomized order.

Data Analysis

Data analysis was carried out as in study 1.

Ethical Considerations

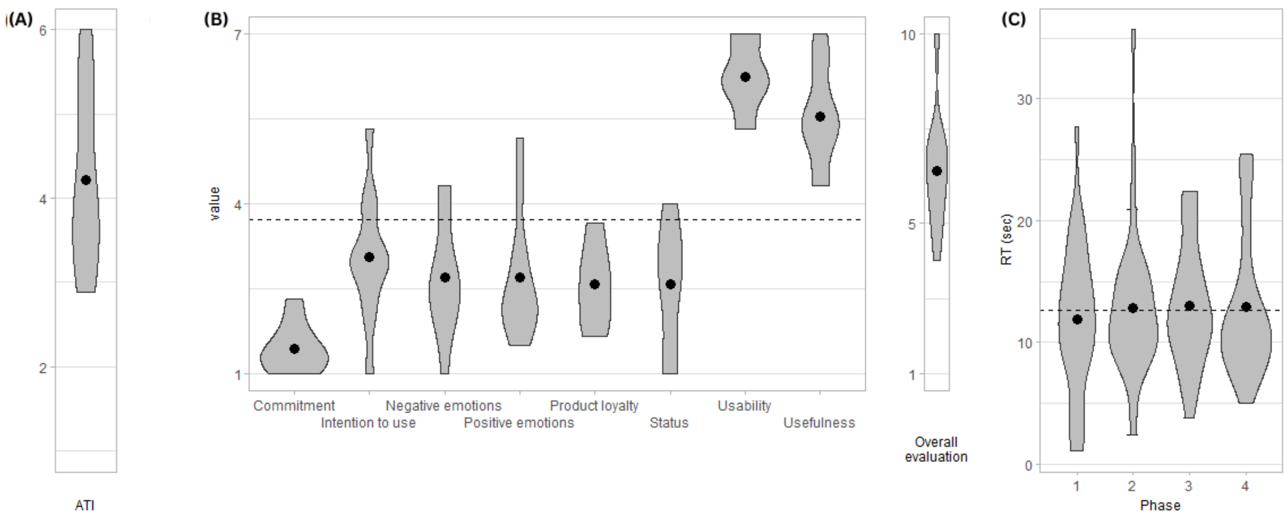
Study 0 was approved by the Ethics Board of ETH Zürich, Switzerland (2019-N-32), and study 1 and study 2 were approved within the context of the HEADWIND study by the cantonal ethics commission of Bern, Switzerland (2020-00685 and 2021-02381, respectively). Study 1 and study 2 are available at ClinicalTrials.gov (NCT04035993 and NCT04569630, respectively). All participants provided written informed consent.

Results

Study 0: Preliminary Assessment With Healthy Individuals in Simulated Driving

Results are summarized in Figure 1.

Figure 1. Violin plots of (A) Affinity for Technology Interaction (ATI; min=1, max=6), (B) score values across the constructs of modular evaluation of key Components of User Experience (meCUE; min=1, max=7, except for overall evaluation, which is min=1, max=10), and (C) reaction time across phases in study 0 (n=11). The dots represent the group mean; the dashed line represents the overall mean. RT: reaction time; sec: seconds.



Recruitment and Participants

We recruited 11 healthy individuals with a valid driver's license via a web advertisement (ie, University of Zurich marketplace). One participant was excluded owing to simulator sickness. Thus, we included 10 participants (n=4, 40% female; n=6, 60% male) with an average age of 30.4 (SD 7.8; range 23-47) years and holding a license for 11 (SD 7.5; range 2-26) years, on average.

Technology Affinity Measure

Participants showed a mean ATI of 4.2 (SD 1; Cronbach α=.91), which is 70% of the maximal value.

Perception Measure

The meCUE (Cronbach α=.7) revealed a mean overall evaluation of 6.4 (SD 1.6), which is 64% of the maximal value. Moreover, the highest mean values were achieved for usability (mean 6.2, SD 0.6, 89%) and usefulness (mean 5.6, SD 0.9, 80%), whereas lower values were observed for commitment (mean 1.5, SD 0.4, 21%), positive emotions (mean 2.7, SD 1.1, 39%), negative emotions (mean 2.7, SD 1, 39%), intention to use (mean 3.1, SD 1.1, 44%), and product loyalty (mean 2.6, SD 0.7, 37%). A low value for negative emotions reflects a more positive evaluation.

To explain the perception results with the technology affinity measure, we correlated each meCUE construct with ATI. We observed a correlation between ATI and intention to use (ρ=0.70; *P*=.02). All the other correlations were not significant at the .05 level.

Compliance Measure

All the participants complied with the warning and stopped the car. Participants took 12.6 (SD 5.7) seconds on average.

Qualitative Feedback

Finally, some participants reported that the voice assistant spoke too fast to deliver information during a driving task without being distracting.

Study 1: Assessment With Individuals With T1DM in Simulated Driving

Results are summarized in Figures 2 and 3.

Figure 2. Violin plots of (A) count of previous experience, (B) score values across the constructs of Technology Readiness Index (TRI; min=1, max=5), (C) score values across the constructs of Acceptance and Use of Technology (AUT; min=1, max=5), (D) Session Alliance Inventory (SAI) scores (min=1, max=5), (E) Trust scores (min=1, max=5), and (F) reaction time across iterations in study 1 (n=18). The dots represent the group means; the dashed line represents the overall mean within an iteration. RT: reaction time; sec: seconds.

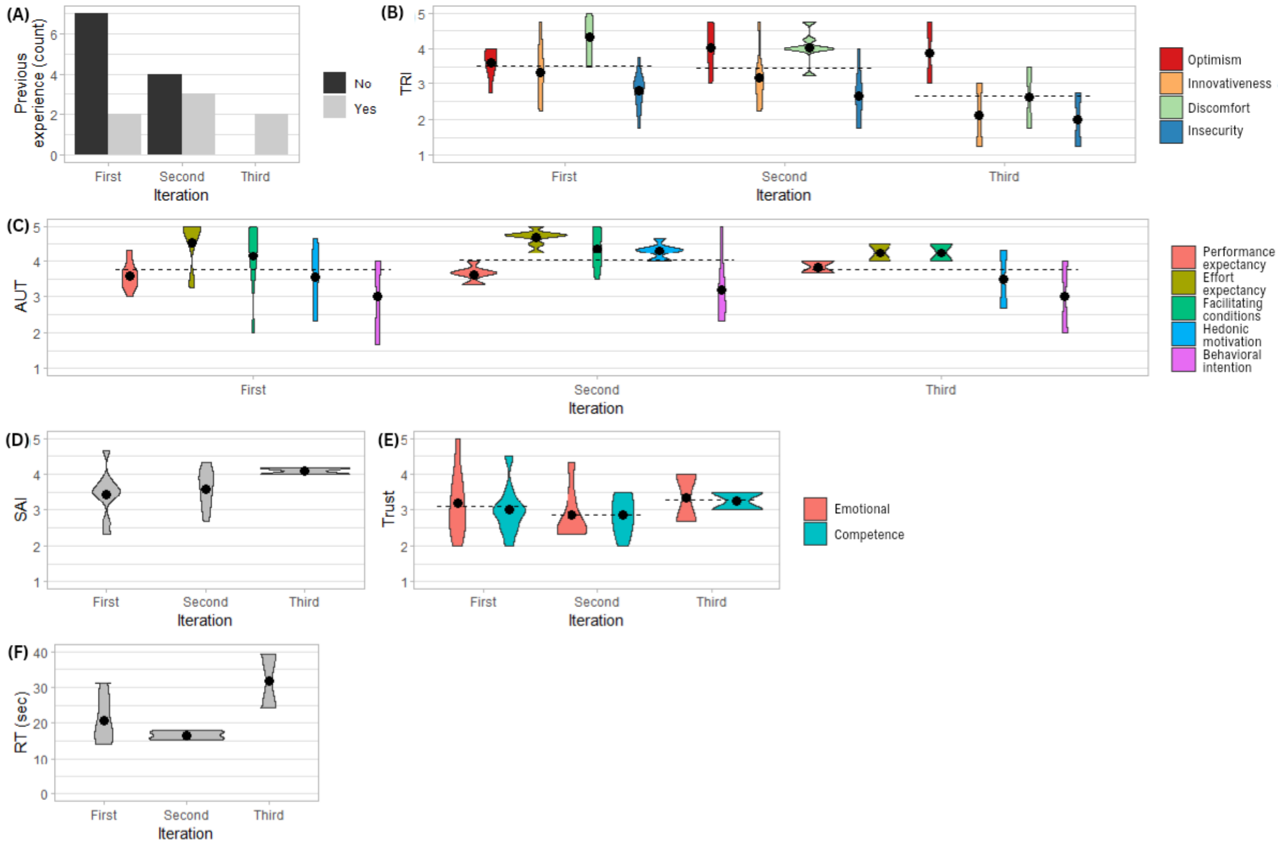
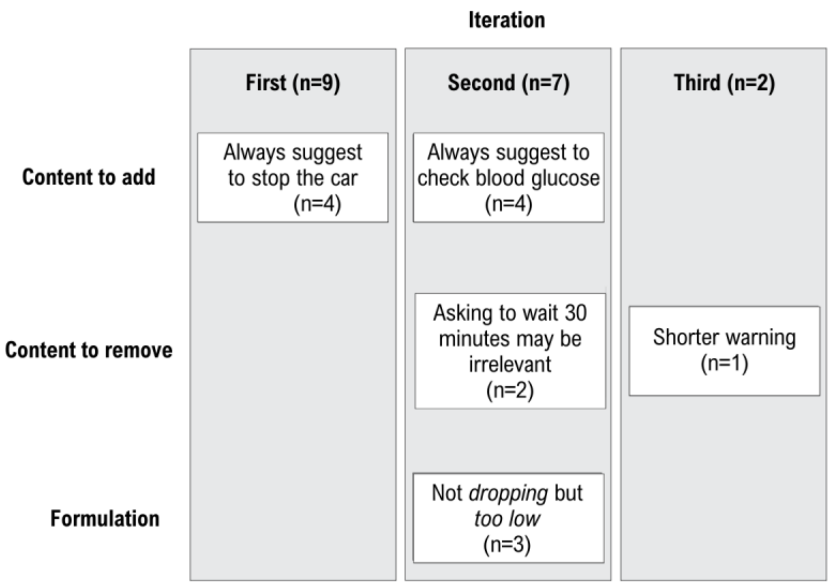


Figure 3. Thematic summary of participants' feedback in study 1 (n=18).



Recruitment and Participants

We recruited 20 patients with T1DM from the Department of Diabetes, Endocrinology, Nutritional Medicine, and Metabolism at the Bern University Hospital. Participants needed functional insulin treatment, good insulin self-management knowledge, a driver's license, and active driving in the past 6 months. We excluded one participant owing to simulator sickness and one participant owing to technical errors in the warning delivery. This resulted in a total of 18 participants (n=6, 33% female and n=12, 67% male; mean age 31.4, SD 7, range 24-44 years; mean driver's license age 13, SD 7.5, range 4.5-28.6 years). The first iteration had 9 participants, the second iteration had 7, and the third iteration had 2 participants. Although the last iteration's sample size was small, it provided useful feedback to improve the warning for study 2.

Technology Affinity Measure

Seven participants had previous experience with an in-vehicle voice assistant (n=2, 28% in the first iteration; n=3, 43% in the second iteration, and n=2, 28% in the third iteration). TRI was 3.4 (SD 0.6, Cronbach α=.85), or 68% of the maximum. Specifically, TRI was 3.5 in the first and second iterations (SD 0.6 and 0.7, respectively) and 2.7 in the third iteration (SD 0.9).

Perception Measure

Perception averaged 3.9 out of 5 (SD 0.8, 78%) and remained stable across iterations. Average AUT (Cronbach α=.81) values were 3.8 (SD 1) in the first iteration, 4 (SD 0.7) in the second iteration, and 3.8 (SD 0.8) in the third iteration. Effort expectancy and facilitating conditions had the highest values across all iterations, whereas behavioral intention always had the lowest values.

SAI (Cronbach α=.79) averaged 3.7 out of 5 (SD 0.5, 74%) and increased slightly over the iterations, from 3.4 (SD 0.7) in the first iteration to 3.6 (SD 0.6) in the second iteration to 4.1 (SD 0.1) in the third iteration.

Trust (Cronbach α=.8) averaged 3.1 out of 5 (SD 0.7, 62%), was stable across constructs, and had the lowest values of the 3 perception measures. Trust averaged 3.1 (SD 0.8) in the first iteration, 2.9 (SD 0.6) in the second iteration, and 3.3 (SD 0.6) in the third iteration.

To explain the perception results with technology affinity, we tested the difference in perception (AUT, SAI, and Trust) between participants with and without previous experience with in-vehicle voice assistants. The means of all constructs, excluding facilitating conditions, were slightly higher for participants with previous experience. However, a 2-sided *t* test revealed no significant result (ie, *P*>.05).

We also correlated perception with TRI and observed a correlation between the optimism construct and performance expectancy (ρ=0.49; *P*=.04), behavioral intention (ρ=0.52; *P*=.03), and SAI (ρ=0.57; *P*=.01). All the other correlations were not significant (*P*>.05).

Compliance Measure

All the participants complied with the warning. In the first iteration, all participants answered *yes* or *no* to the receptivity check ("May I disturb you?") and when asked if they had carbohydrates on hand. Five of the 9 participants answered yes to the latter question, although they did not. Two of those 9 participants stopped the car although they were not explicitly advised to do so. In the second iteration, all participants answered the prompts with yes and stopped the car as advised. One participant gave an affirmative *mhm* when asked, "May I disturb you?" during the hypoglycemic phase but were otherwise compliant. Because we used the Wizard-of-Oz method, the experimenter interpreted the affirmation. However, a current voice assistant might have interpreted it as an error. In the third iteration, both participants answered the prompts with yes and stopped the car. Across iterations, compliance took approximately 22 seconds. In particular, compliance took approximately 20 (mean 20.7, SD 6.2) seconds in the first iteration, approximately 17 (mean 16.7, SD 1.2) seconds in the second iteration, and approximately 31 (mean 31.7, SD 10.6) seconds in the third iteration.

Qualitative Feedback

Participants judged the voice warning as pleasant, simple, and as clear and efficient (n=15, n=11, and n=13, respectively). The topics for improvement are summarized in Figure 3. Note that these results are best understood when compared with Multimedia Appendix 1.

Given that Trust showed the lowest values in the first iteration, in comparison with the other perception measures, we decided to specifically ask participants, in our second and third iterations, what would help them trust the warning more. Of the 9 participants included in both the second and third iterations, 5 (55%) said they would just need to have a prolonged experience with the warning, whereas 3 (33%) said they would need to know what kind of data is used to infer that the driver is about to experience hypoglycemia. One participant did not know what would improve their trust.

Study 2: Assessment With Individuals With T1DM in Real-World Driving Undergoing Hypoglycemia

Results are summarized in Figures 4 and 5.

Figure 4. Violin plots of (A) count of previous experience, (B) score values across the constructs of Technology Readiness Index (TRI; min=1, max=5), (C) score values across the constructs of Acceptance and Use of Technology (AUT; min=1, max=5), (D) Session Alliance Inventory (SAI) scores (min=1, max=5), (E) Trust scores (min=1, max=5), and (F) reaction time across iterations in study 2 (n=20). The dots represent the group means; the dashed line represents the overall mean within an iteration. RT: reaction time; sec: seconds.

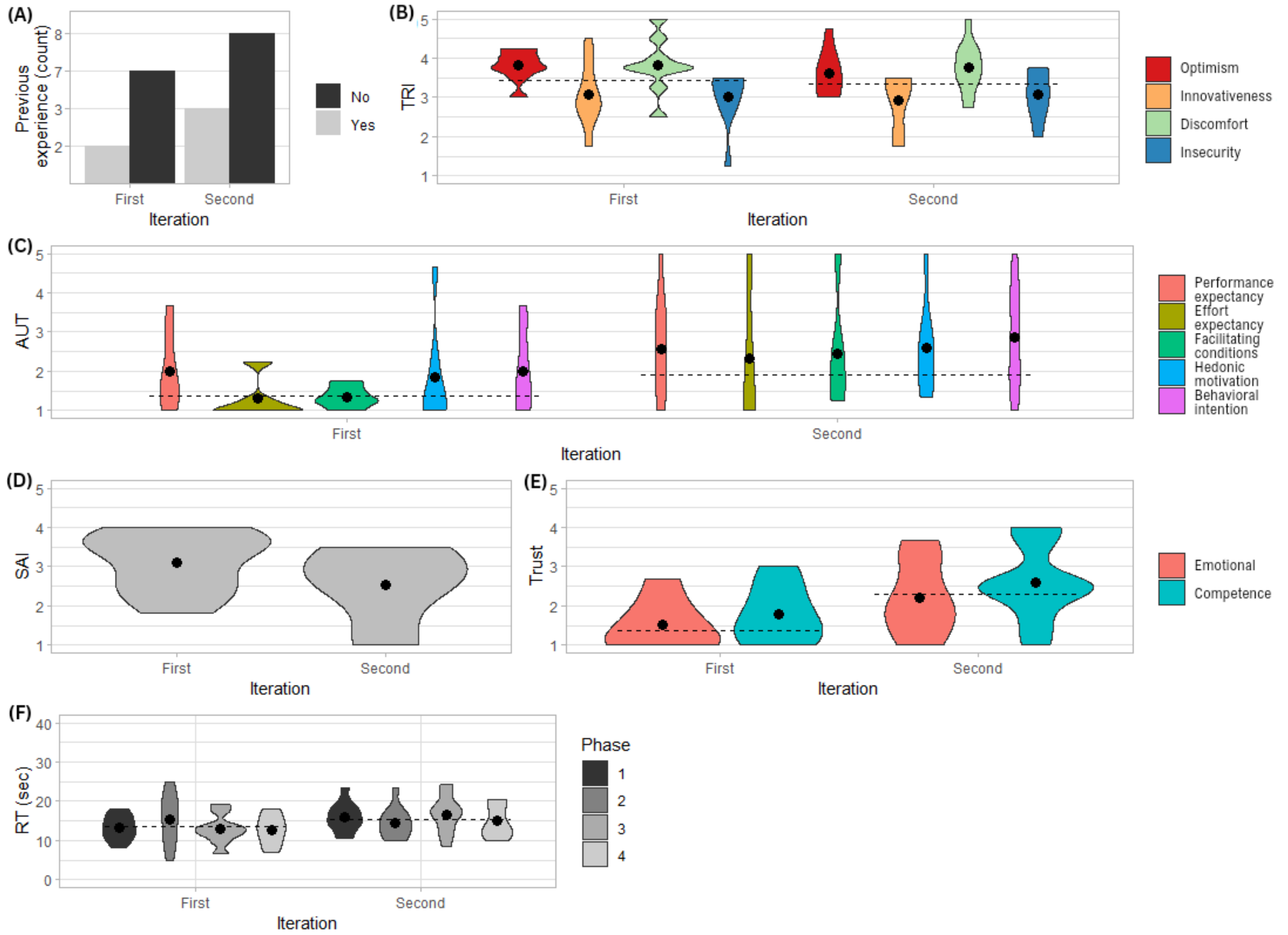
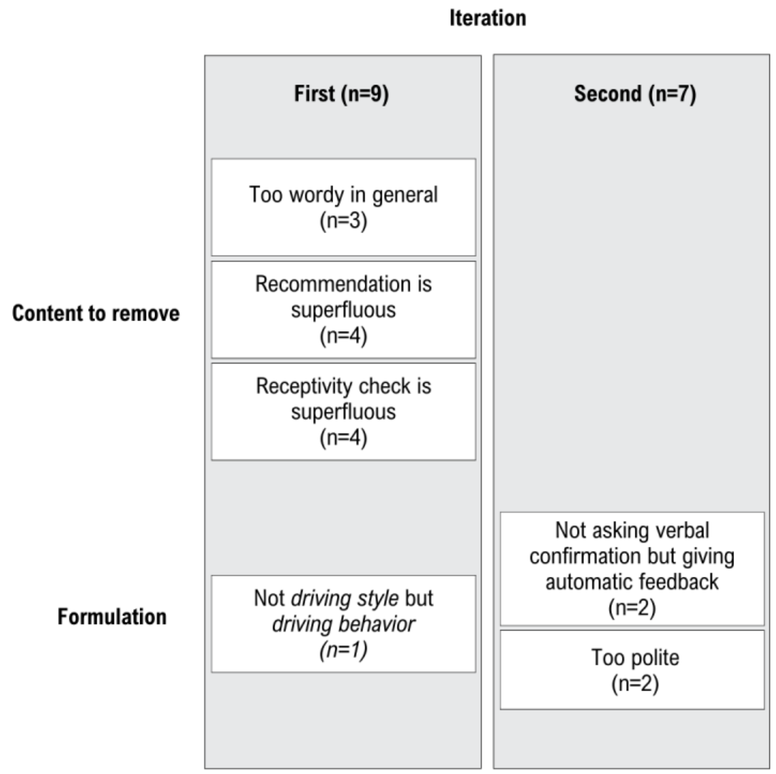


Figure 5. Thematic summary of participants' feedback in study 2 (n=20).



Recruitment and Participants

The recruitment procedure was the same as in study 2. We recruited 21 individuals, and 1 participant was excluded owing to data loss. Thus, we included 20 participants (n=3, 15% female and n=17, 85% male; mean age 40.9, SD 10.6, range 23-57 years; and holding a license on average since 23.7, SD 11.1, range 3.1-42.4 years). The first iteration included a sample of 9 participants and the second iteration included a sample of 11 participants.

Technology Affinity Measure

The pretest measurements revealed that 25% (5/20) of the participants had previous experience with an in-vehicle voice assistant (2 in the first iteration, and 3 in the second iteration), whereas TRI was on average 3.4 (SD 0.7; Cronbach α=.44), which is 68% of the maximal value. In particular, TRI was 3.4 in the first iteration (SD 0.8), and 3.3 in the second iteration (SD 0.7).

Perception Measure

The overall perception score was 1.7 out of 5 (SD 1.3, 34%). The results showed a slight increase in mean AUT (Cronbach α=.95) and Trust (Cronbach α=.85) values between the first and the second iteration, whereas SAI (Cronbach α=.80) showed a slight decrease. AUT also showed a considerable increase in SD. In particular, AUT values were on average 1.4 (SD 1) in the first iteration, and 1.9 (SD 1.6) in the second iteration; SAI was overall 2.8 out of 5 (SD 0.8, 56%). Values were on average 3.1 (SD 0.7) in the first iteration and 2.5 (SD 0.9) in the second iteration; Trust values were on average 1.4 (SD 0.8) in the first iteration, and 2.3 (SD 1.1) in the second iteration.

Similar to study 1, to explain the perception results with the technology affinity measure, we tested the difference in perception (ie, AUT, SAI, and Trust) among participants who had previous experience with in-vehicle voice assistants and those who did not. The means of all perception measures, excluding SAI, were consistently slightly higher in the second iteration. The means of all constructs, excluding SAI, were consistently slightly higher for participants who had previous experience with in-vehicle voice assistants. However, a 2-sided *t* test revealed no significant result (ie, $P>.05$). When correlating each perception measure with TRI, we observed a correlation between innovativeness and hedonic motivation ($\rho=0.52$; $P=.02$), a negative correlation between discomfort and behavioral intention ($\rho=-0.46$; $P=.04$), and a negative correlation between discomfort and competence trust ($\rho=-0.45$; $P=.05$). All the other correlations were not significant ($P>.05$).

Compliance Measure

All the participants complied with the warning. Two drives were excluded: one participant stopped once before the warning was delivered and data from one drive of one participant was lost. The results showed that the reaction time does not seem to vary across glycemic phases and, although minimal, there is a tendency for the reaction time to increase in the second iteration. Participants took 13.6 (SD 4.5) seconds in the first iteration and 15.5 (SD 4.1) seconds in the second iteration.

Preference for the Disclosure of the Triggering Cause

One participant was excluded because of data loss. The results showed that although 10 participants preferred when the warning was delivered with an explanation for the warning being triggered (in this case, driving behavior), 8 participants preferred it without the explanation. One participant stated that they would not use this in-vehicle voice warning either way.

Qualitative Feedback

In general, and similar to study 1, the participants found the communication style pleasant and efficient (n=4 and n=5, respectively). The topics for improvement are summarized in Figure 5. Note that these results are best understood when compared with Multimedia Appendix 1.

Discussion

Principal Findings

Most participants had not previously used an in-vehicle voice assistant, and technology affinity was similar across studies. In general, the voice warning elicited a positive perception, although the perception values were lower in the real-car study. In addition, participants complied with the warning in all studies, and reaction times were shorter in the real-car study than in the simulator study. Finally, the participants preferred the voice warning to be less verbose and prompt fewer interactions with the driver.

Technology Affinity

Although we did not observe a significant effect on the perception of the warning, we suspect that the participants may have experienced a double novelty: using a voice assistant while driving and experiencing a warning from an in-vehicle voice assistant. Thus, future research should include a more balanced sample and compare the perception of a voice assistant–based warning with a standard warning (eg, an acoustic tone). Moreover, although we cannot directly compare ATI (used in study 0) with TRI (used in study 1 and study 2), we can observe that technology affinity was similar across studies. Although ATI showed a mean of 4.2 over 7 (60%), TRI showed a mean of 3.4 over 5 both in study 1 and study 2 (68%). The change in technology affinity measure was the result of an internal discussion between the coauthors, and we recommend the scientific community to use TRI in future research, as it is more widely used and focuses not only on the interaction but also on the general attitudes toward new technologies.

Perception

We observed that AUT, SAI, and Trust values were higher in study 1 (simulated driving) than in study 2 (real-world driving). This evaluation might have been influenced by the driving setting. There can be 2 possible reasons. First, participants may have found the warning to be more distracting in the real car than in the simulator. However, research shows that drivers are more in control in real-world driving than in simulated driving [55]. Second, the technical difficulties in controlling the driver-assistant interaction owing to network slowdowns might have affected the user experience, and thus the perception measures. Future Wizard-of-Oz studies may account for this methodological weakness with a more accurate text-to-speech technology, avoiding remote control, and reducing interactions.

In addition, TRI seemed to have influenced behavioral intention (AUT) but did not consistently influence the other perception measures (ie, other constructs of AUT, SAI, and Trust). Thus, participants may have been excited about the potential of the voice warning, but they may not have been happy with the actual experience of using it.

Compliance

The reaction times were short enough to ensure a timely reaction to the critical event. Blood glucose can change with a maximum rate of 0.22 mmol/L/min [56]. This means that someone driving with a normal glucose of 5.5 mmol/L might reach hypoglycemia (ie, 3.9 mmol/L) within a minimum of 7.5 minutes. Thus, although experiencing hypoglycemia while driving does not require an abrupt stop but rather a careful pullover maneuver and treating the condition, measuring reaction time provided an insight into the time required to take the first measure (ie, pullover). Interestingly, the reaction time was shorter in the real car (study 2) than in the simulator (study 0 and study 1). This difference may be attributed to the lack of traffic in study 2, which allowed the driver to pull over faster.

Feedback

Although we aimed to keep the warning conversational, participants preferred a more direct notification of the problem without specific recommendations (eg, recommending waiting until the blood glucose is at its normal level) or polite formulations (eg, asking for permission to talk). To the best of the author's knowledge, there was no in-vehicle voice warning at the time of the study, and we mostly relied on the guidelines of the Swiss Diabetes Association [10], while keeping the conversation as simple as possible. The participants' feedback allowed us to improve the warning in this direction.

Implications and Future Directions

Hypoglycemia Warnings

Reportedly, no research has been conducted for in-vehicle applications providing a hypoglycemia warning. However, smartphone apps for hypoglycemic events tracking have been investigated [57]. Although most of the research on glucose monitoring solutions conducted so far focused on diary apps rather than warning delivery, a pilot study on a smartphone-based hypoglycemia warning showed an improved hypoglycemia awareness and a reduction in daytime hypoglycemia [58] (other research is still in the phase of validation [59]). Future research should investigate such outcomes with an in-vehicle extension of this type of application.

In-Vehicle Warnings

Although there seems to be no related work testing the voice assistant of a private vehicle to deliver hypoglycemia warnings, there is a need for "driver-friendly" in-vehicle glucose monitoring solutions, expressed by the online community [25]. In particular, drivers with T1DM have contributed to the Nightscout Foundation [60], a nonprofit organization founded in 2014 and supporting open source technology for T1DM management, with the development of a data-sharing app, able to connect a car to a CGM, and display the glucose trends while driving on the dashboard of the private vehicle [25]. Moreover, there has been conceptual work manifesting the need for collaboration between automotive and medical industries to improve the safety of drivers with T1DM [24]. However, this work has not been followed by any implementation. Furthermore, no testing with the actual users has been conducted. Our work provides preliminary evidence, both in a simulated and a real-world environment.

Needless to say, recognizing hypoglycemia is only one part of glucose monitoring while driving; general imbalance of blood glucose (including hyperglycemia) can be problematic for the driver, if not dangerous [61]. Our work can be extended to hyperglycemia and, therefore, support further the safety of drivers with T1DM.

Finally, using the in-vehicle voice assistant to deliver a warning is compatible with current technology: not only are cars increasingly equipped with voice assistants [26,27] but also the automotive industry is aware of the relevance of using the upcoming "in-car proactivity" [62].

Warning Escalation

Our results showed 100% compliance in all 3 studies. This can only mean that the warning was clear enough for the participant to understand that it was time to pull over. That is, as all studies were run in a controlled setting, where an experimental team was present, and the participant knew they would be recommended to pull over eventually, we can safely assume that the experiments experienced a participant bias [63]. Thus, we cannot conclude that the warning was compelling enough to motivate the participants to comply (see the Limitations section). Nevertheless, the warning should be designed to allow for escalation, whereas in case the driver does not pull over in due time (eg, 2-3 min [56]) or explicitly rejects the warning, delayed reprompts with an increasingly severe tone would be delivered by the voice assistant (eg, "You are at risk of hypoglycemia. Please stop the car safely and check your blood sugar, then risk of hypoglycemia. Pull over now").

Hypoglycemia Detection

Finally, in this paper, we focus on the interface between the hypoglycemia detection system and the driver, with the aim of visually distracting them as little as possible. Although the detection side is beyond the scope of this study, the designed warning is intended to be produced by a voice assistant built into the vehicle. Therefore, how a vehicle monitors blood sugar depends on the technology of the car. For instance, the aforementioned open source app displaying the glucose levels on the dashboard of a private vehicle [25] could be enhanced to connect with the in-vehicle voice assistant and use a voice warning instead of a visual one. Furthermore, research has been conducted on how to detect hypoglycemia from the car's data [54] and from consumer-available wearable devices [64], with the argument that CGM devices can impose a social and financial burden on the individual.

Limitations and Strengths

Despite our best efforts, this research has 3 main limitations.

First, the studies included a relatively small sample size. However, this study includes 3 feasibility studies (ie, a preliminary study with healthy individuals and 2 feasibility studies with individuals with T1DM), and the research presented in this paper is intended to be understood as an iterative development of a hypoglycemic warning. As such, this research aimed to pioneer the use of in-vehicle voice assistants for a driver health-related warning, rather than draw conclusions to be generalized to the population with T1DM. Thus, although we included a total sample size of 48 individuals, each feasibility study provides insight into the changes required by the users, and we provide the scientific community with an opening to the design of in-vehicle voice assistant–based health-related warning. Furthermore, previous studies on digital health systems used a similar sample size [65-67]. Thus, we believe that although the sample size does not allow drawing conclusions on the interaction of drivers with T1DM with in-vehicle hypoglycemia warnings, it still reports pioneer research.

Second, the studies were conducted over a short period. The participants had only a short-term experience with the warning. Perception and compliance may therefore be influenced by the novelty of such an experience, whereas perception may stabilize with repeated experience [68]. Future research should investigate the user experience of the warning in a longitudinal study. Third, these studies did not control for all potentially confounding variables related to real-world traffic and driver's priorities. For instance, both simulator and real-car experiments involved disadvantages: while assessing the warning in a simulator allowed a controlled and safe experiment, such a setting remains artificial and lacks external validity. In contrast, while testing it in a real car increased the ecological validity of the human-machine interaction, it did not allow for as much traffic and speed variation as was possible in the simulator. Future research should investigate the effects of real-world traffic on the perception of the warning and compliance behavior. Moreover, receiving a warning in the presence of a team of experimenters may have influenced the participant's verbal and behavioral responses; participants knew they would receive the warning sooner or later and had no reason not to follow it (eg, ignoring the warning because of being late for an appointment). In a real situation, drivers may not respond as expected or may even ignore the warning. Future research should test such a warning in a more ecological context, for instance, in a field study where the driver may not fall for a participant bias [69].

Finally, as we aimed to test a voice warning, our studies used a Wizard-of-Oz methodology to avoid problems related to natural language processing. Note that our studies were conducted in German-speaking Switzerland, where the German accents easily vary from region to region. As this aspect was beyond the scope of our research, we did not implement a working voice assistant or account for potential fallback intents triggered by the voice assistant's failure to understand the user. Future research should push this research further and examine the potential danger of delayed treatment of hypoglycemia owing to the voice assistant's natural language processing errors.

Conclusions

Although hypoglycemia increases the risk of car mishaps [7,8], current solutions (eg, CGM and FGM) require visual human-machine interaction, which is inappropriate for an in-vehicle context. As voice assistants are increasingly present in private vehicles [26,27] and the European Commission fosters safety technologies inside the car [16], we propose to warn the driver of their critical health state through a voice assistant–based health warning. This paper reports on an iterative development and assessment of a hypoglycemia warning. In particular, we conducted in 3 studies: a preliminary study using a simulator with healthy participants, a test with individuals with T1DM in a simulator, and a test with individuals with T1DM in a real car. This gradual increase in authenticity in the experimental design allowed us to increase the ecological validity of our results while keeping experimental control. To the best of our knowledge, this is the first attempt of such a comprehensive feasibility assessment of an in-vehicle voice warning for hypoglycemia. Our results suggest that a voice warning can be useful, but that proactive behavior in voice assistants is still emerging and unfamiliar. We hope that these preliminary findings will foster future research to further develop in-vehicle hypoglycemia warnings.

Acknowledgments

The authors would like to thank the Swiss Federal Department of Defense, Civil Protection and Sports for providing the closed circuit used in study 2.

This work was partly funded by the Swiss National Science Foundation, Sinergia Project (CRSII5 183569).

Data Availability

The data supporting the findings of this study are available from the last author, TK, upon reasonable request.

Authors' Contributions

CS and EF were responsible for the oversight, leadership, and management of the research activity and for funding acquisition. CB, MM, VFL, MK, SF, TZ, FW, CS, EF, and TK were responsible for the methodology. CB and TK were responsible for the formulation of the research goals and aims, conceptualizing the voice warning. CB and MM were responsible for developing the voice warnings. CB, MM, FW, and TK were responsible for developing the driving scenarios. FW, TK, and TZ were responsible for providing the driving simulator, the real car, and the closed circuit. CB and MM were responsible for recruiting participants for study 0. VFL, TZ, and CS were responsible for recruiting participants for studies 2 and 3. CB, MM, and VFL were responsible for data collection. CB was responsible for the data analysis and presentation and the first draft of this manuscript. VFL, MK, SF, TZ, CS, EF, ABK, and TK were responsible for critical feedback and final revisions of the manuscript.

Conflicts of Interest

CB, VFL, SF, FW, TZ, CS, EF, and TK are affiliated with the Centre for Digital Health Interventions, a joint initiative of the Department of Management, Technology, and Economics at ETH Zürich, and the Institute of Technology Management at the University of St Gallen, which is funded in part by the Swiss health insurer CSS. EF and TK are also the cofounders of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways. However, neither CSS nor Pathmate Technologies were involved in any way in the design, interpretation, analysis, or writing. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Original (German) and translated version of the conversation flow of the hypoglycemia voice warning in study 1 and study 2.
[PDF File (Adobe PDF File), 554 KB - humanfactors_v11i1e42823_app1.pdf]

Multimedia Appendix 2
Questions used in the semistructured interview about participants' experience with the warning conducted in study 1 and study 2.
[PDF File (Adobe PDF File), 209 KB - humanfactors_v11i1e42823_app2.pdf]

Multimedia Appendix 3
Illustration of the procedure across studies.
[PDF File (Adobe PDF File), 174 KB - humanfactors_v11i1e42823_app3.pdf]

References

1. Moini J. Type 1 diabetes. Epidemiol Diabetes 2019:75-90. [doi: 10.1016/b978-0-12- 816864-6.00006-7]
2. Cryer PE, Davis SN, Shamoon H. Hypoglycemia in diabetes. Diabetes Care 2003 Jun;26(6):1902-1912 [FREE Full text] [doi: 10.2337/diacare.26.6.1902] [Medline: 12766131]
3. Malkani S, Kotwal A. Frequency and predictors of self-reported hypoglycemia in insulin-treated diabetes. J Diabetes Res 2017;2017:7425925 [FREE Full text] [doi: 10.1155/2017/7425925] [Medline: 28913365]
4. Allen KV, Pickering MJ, Zammitt NN, Hartsuiker RJ, Traxler MJ, Frier BM, et al. Effects of acute hypoglycemia on working memory and language processing in adults with and without type 1 diabetes. Diabetes Care 2015 Jun;38(6):1108-1115 [FREE Full text] [doi: 10.2337/dc14-1657] [Medline: 25758768]
5. Leung E. New and not so new information on hypoglycemia and prevention strategies. Journal: The Official Publication of the Ontario Occupational Health Nurses Association 2018;37(1):20-22 [FREE Full text]
6. Warren RE, Frier BM. Hypoglycaemia and cognitive function. Diabetes Obes Metab 2005 Sep;7(5):493-503 [FREE Full text] [doi: 10.1111/j.1463-1326.2004.00421.x] [Medline: 16050942]
7. Blumer I, Kenshole AB, Stilman J, Lewis GF. Insulin-treated diabetes and driving: legal jeopardy and consequences of hypoglycemia. Can J Diabetes 2019 Apr;43(3):221-223 [FREE Full text] [doi: 10.1016/j.jcjd.2018.08.201] [Medline: 30348593]
8. Cox DJ, Penberthy JK, Zrebiec J, Weinger K, Aikens JE, Frier B, et al. Diabetes and driving mishaps: frequency and correlations from a multinational survey. Diabetes Care 2003 Aug;26(8):2329-2334 [FREE Full text] [doi: 10.2337/diacare.26.8.2329] [Medline: 12882857]
9. Waris A, Jose R, Mathew N, Jacob JJ. Attitudes to driving among patients with diabetes mellitus in Punjab (ADD-Punjab study). Prim Care Diabetes 2022 Feb;16(1):34-40 [FREE Full text] [doi: 10.1016/j.pcd.2021.11.007] [Medline: 34872840]
10. Fankhauser B. Diabetes und autofahren: worauf sie am steuer achten sollten. d-journal. 2018 Nov 2. URL: https://www.d-journal.ch/uebriges-keine/diabetes-und-autofahren-worauf-sie-am-steuer-achten-sollten/ [accessed 2023-12-22]
11. Cox DJ, Gonder-Frederick L, Clarke W. Driving decrements in type I diabetes during moderate hypoglycemia. Diabetes 1993 Feb;42(2):239-243. [doi: 10.2337/diab.42.2.239] [Medline: 8425660]
12. Roberts AJ, Moss A, Malik FS, Taplin C, Pihoker C, Hirsch IB, et al. Driving safety in adolescents and young adults with type 1 diabetes. Diabetes Spectr 2020 Nov;33(4):352-357 [FREE Full text] [doi: 10.2337/ds20-0021] [Medline: 33223774]
13. Cappon G, Vettoretti M, Sparacino G, Facchinetti A. Continuous glucose monitoring sensors for diabetes management: a review of technologies and applications. Diabetes Metab J 2019 Aug;43(4):383-397 [FREE Full text] [doi: 10.4093/dmj.2019.0121] [Medline: 31441246]
14. Landry S, Jeon M, Lautala P, Nelson D. Design and assessment of in-vehicle auditory alerts for highway-rail grade crossings. Transp Res Part F Traffic Psychol Behav 2019 Apr;62:228-245 [FREE Full text] [doi: 10.1016/j.trf.2018.12.024]

15. Böhme P, Bertin E, Cosson E, Chevalier N, GEODE group. Fear of hypoglycaemia in patients with type 1 diabetes: do patients and diabetologists feel the same way? Diabetes Metab 2013 Feb;39(1):63-70 [FREE Full text] [doi: 10.1016/j.diabet.2012.10.006] [Medline: 23266467]
16. Road safety: commission welcomes agreement on new EU rules to help save lives. European Commission. 2019 Mar 26. URL: https://ec.europa.eu/commission/presscorner/detail/en/IP_19_1793 [accessed 2023-12-22]
17. Sikander G, Anwar S. Driver fatigue detection systems: a review. IEEE Trans Intell Transport Syst 2019 Jun;20(6):2339-2352 [FREE Full text] [doi: 10.1109/tits.2018.2868499]
18. Fridman L, Toyoda H, Seaman S, Seppelt B, Angell L, Lee J, et al. What can be predicted from six seconds of driver glances? In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI '17: CHI Conference on Human Factors in Computing Systems; May 6-11, 2017; Denver, CO URL: https://doi.org/10.1145/3025453.3025929 [doi: 10.1145/3025453.3025929]
19. Ljungblad J, Hök B, Allalou A, Pettersson H. Passive in-vehicle driver breath alcohol detection using advanced sensor signal acquisition and fusion. Traffic Inj Prev 2017 May 29;18(sup1):S31-S36 [FREE Full text] [doi: 10.1080/15389588.2017.1312688] [Medline: 28368660]
20. Koch K, Liu S, Berger T, Wortmann F. Towards the healing car: investigating the potential of psychotherapeutic in-vehicle interventions. In: Proceedings of the European Conference on Information Systems. 2020 Presented at: ECIS 2020; June 15-17, 2020; Marrakech, Morocco URL: https://aisel.aisnet.org/ecis2020_rip/8
21. Paredes PE, Zhou Y, Hamdan NA, Balters S, Murnane E, Ju W, et al. Just breathe: in-car interventions for guided slow breathing. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Mar 26;2(1):1-23 [FREE Full text] [doi: 10.1145/3191760]
22. Albadawi Y, Takruri M, Awad M. A review of recent developments in driver drowsiness detection systems. Sensors (Basel) 2022 Mar 07;22(5):2069 [FREE Full text] [doi: 10.3390/s22052069] [Medline: 35271215]
23. Merickel J, High R, Smith L, Wichman C, Frankel E, Smits K, et al. Driving safety and real-time glucose monitoring in insulin-dependent diabetes. Int J Automot Eng 2019;10(1):34-40 [FREE Full text] [doi: 10.20485/jsaeijae.10.1_34] [Medline: 34306907]
24. Kerr D, Olateju T. Driving with diabetes in the future: in-vehicle medical monitoring. J Diabetes Sci Technol 2010 Mar 01;4(2):464-469 [FREE Full text] [doi: 10.1177/193229681000400228] [Medline: 20307408]
25. Hoskins M. Glooko’s role in revolutionizing diabetes data usability. Healthline. 2022 Apr 7. URL: https://www.healthline.com/diabetesmine/diabetes-data-in-cars-wearenotwaiting [accessed 2023-12-22]
26. In-car voice assistant consumer adoption report 2019. Voicebot.ai. URL: https://voicebot.ai/in-car-voice-assistant-consumer-adoption-report-2019/ [accessed 2023-12-22]
27. In-car voice assistant consumer adoption report 2020. Voicebot.ai. URL: https://voicebot.ai/in-car-voice-assistant-consumer-adoption-report-2020/ [accessed 2023-12-22]
28. Baron A, Green P. Safety and usability of speech interfaces for in-vehicle tasks while driving: a brief literature review. The National Academies of Sciences, Engineering, and Medicine. 2006. URL: https://trid.trb.org/view/806083 [accessed 2023-12-22]
29. Strayer DL, Cooper JM, Goethe RM, McCarty MM, Getty DJ, Biondi F. Assessing the visual and cognitive demands of in-vehicle information systems. Cogn Res Princ Implic 2019 Jun 21;4(1):18 [FREE Full text] [doi: 10.1186/s41235-019-0166-3] [Medline: 31227955]
30. Bérubé C, Schachner T, Keller R, Fleisch E, V Wangenheim F, Barata F, et al. Voice-based conversational agents for the prevention and management of chronic and mental health conditions: systematic literature review. J Med Internet Res 2021 Mar 29;23(3):e25933 [FREE Full text] [doi: 10.2196/25933] [Medline: 33658174]
31. Bérubé C, Fleisch E. Voice-based conversational agents for sensing and support: examples from academia and industry. In: Marsch LA, Jacobson NC, Kowatsch T, editors. Digital Therapeutics for Mental Health and Addiction: The State of the Science and Vision for the Future. Amsterdam, Netherlands: Elsevier Science; 2022:113-134.
32. Baki Kocaballi A, Laranjo L, Clark L, Kocielnik R, Moore RJ, Vera Liao Q, et al. Special issue on conversational agents for healthcare and wellbeing. ACM Trans Interact Intell Syst 2022 Jul 12;12(2):1-3 [FREE Full text] [doi: 10.1145/3532860]
33. Sezgin E, Militello LK, Huang Y, Lin S. A scoping review of patient-facing, behavioral health interventions with voice assistant technology targeting self-management and healthy lifestyle behaviors. Transl Behav Med 2020 Aug 07;10(3):606-628. [doi: 10.1093/tbm/ibz141] [Medline: 32766865]
34. Kahler CW, Lechner WJ, MacGlashan J, Wray TB, Littman ML. Initial progress toward development of a voice-based computer-delivered motivational intervention for heavy drinking college students: an experimental study. JMIR Ment Health 2017 Jun 28;4(2):e25 [FREE Full text] [doi: 10.2196/mental.7571] [Medline: 28659259]
35. Chung AE, Griffin AC, Selezneva D, Gotz D. Health and fitness apps for hands-free voice-activated assistants: content analysis. JMIR Mhealth Uhealth 2018 Sep 24;6(9):e174 [FREE Full text] [doi: 10.2196/mhealth.9705] [Medline: 30249581]
36. Kraus M, Wagner N, Minker W. Effects of proactive dialogue strategies on human-computer trust. In: Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization. 2020 Presented at: UMAP '20; July 14-17, 2020; Genoa, Italy URL: https://dl.acm.org/doi/abs/10.1145/3340631.3394840 [doi: 10.1145/3340631.3394840]

37. Dey AK, Abowd GD, Salber D. A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. Hum Comput Interact 2001;16(2-4) [FREE Full text] [doi: 10.1207/S15327051HCI16234_02]
38. Cha N, Kim A, Park CY, Kang S, Park M, Lee JG, et al. Hello there! Is now a good time to talk?: opportune moments for proactive interactions with smart speakers. Proc ACM Interact Mob Wearable Ubiquitous Technol 2020 Sep 04;4(3):1-28 [FREE Full text] [doi: 10.1145/3411810]
39. Zargham N, Reicherts L, Bonfert M, Voelkel ST, Schoening J, Malaka R, et al. Understanding circumstances for desirable proactive behaviour of voice assistants: the proactivity dilemma. In: Proceedings of the 4th Conference on Conversational User Interfaces. 2022 Presented at: CUI '22; July 26-28, 2022; Glasgow, UK URL: https://doi.org/10.1145/3543829.3543834 [doi: 10.1145/3543829.3543834]
40. Wu T, Martelaro N, Stent S, Ortiz J, Ju W. Learning when agents can talk to drivers using the INAGT dataset and multisensor fusion. Proc ACM Interact Mob Wearable Ubiquitous Technol 2021 Sep 14;5(3):1-28 [FREE Full text] [doi: 10.1145/3478125]
41. Lin CH, Shih HY, Sher PJ. Integrating technology readiness into technology acceptance: the TRAM model. Psychol Market 2007 May 24;24(7):641-657 [FREE Full text] [doi: 10.1002/mar.20177]
42. Dahlbäck N, Jönsson A, Ahrenberg L. Wizard of Oz studies: why and how. In: Proceedings of the 1st international conference on Intelligent user interfaces. 1993 Presented at: IUI '93; January 4-7, 1993; Orlando, FL URL: https://dl.acm.org/doi/10.1145/169891.169968 [doi: 10.1016/0950-7051(93)90017-n]
43. Minge M, Thuering M, Wagner I, Kuhr CV. The meCUE questionnaire: a modular tool for measuring user experience. In: Proceedings of the 7th Applied Human Factors and Ergonomics Society Conference 2016. 2016 Presented at: 7th Applied Human Factors and Ergonomics Society Conference 2016; July 27-31, 2016; Orlando, FL URL: https://www.researchgate.net/publication/305877075_The_meCUE_Questionnaire_A_Modular_Tool_for_Measuring_User_Experience [doi: 10.1007/978-3-319-41685-4_11]
44. Minge M, Thüring T. The MeCUE questionnaire (2.0): meeting five basic requirements for lean and standardized UX assessment. In: Proceedings of the Design, User Experience, and Usability: Theory and Practice. 2018 Presented at: Design, User Experience, and Usability: Theory and Practice; July 15-20, 2018; Las Vegas, NV URL: https://doi.org/10.1007/978-3-319-91797-9 [doi: 10.1007/978-3-319-91797-9_33]
45. Attig C, Wessel D, Franke T. Assessing personality differences in human-technology interaction: an overview of key self-report scales to predict successful interaction. In: Proceedings of the HCI International 2017. 2017 Presented at: HCI International 2017; July 9-14, 2017; Vancouver, BC. [doi: 10.1007/978-3-319-58750-9_3]
46. Attig C, Mach S, Wessel D, Franke T, Schmalfuß F, Krems JF. Technikaffinität als ressource für die Arbeit in industrie 4.0. In: Proceedings of the Innteract 2018 Conference. 2018 Jun 28 Presented at: Innteract 2018 Conference; 28/06/2018; Chemnitz, Germany. [doi: 10.14464/awic.v3i0.251]
47. Ren GJ, Szymanski MH, Arar R, Moore RJ. Studies in Conversational UX Design. Cham, Switzerland: Springer International Publishing; 2018.
48. Deci EL, Ryan RM. Self-determination theory: a macrotheory of human motivation, development, and health. Can Psychol 2008 Aug;49(3):182-185. [doi: 10.1037/a0012801]
49. Adell E. Acceptance of driver support systems. In: Proceedings of European Conference on Human Centred Design for Intelligent Transport Systems. 2010 Presented at: European Conference on Human Centred Design for Intelligent Transport Systems; April 29-30, 2010; Berlin, Germany.
50. Venkatesh V, Thong JY, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157-178 [FREE Full text] [doi: 10.2307/41410412]
51. Falkenström F, Hatcher RL, Skjulsvik T, Larsson MH, Holmqvist R. Development and validation of a 6-item working alliance questionnaire for repeated administrations during psychotherapy. Psychol Assess 2015 Mar;27(1):169-183 [FREE Full text] [doi: 10.1037/pas0000038] [Medline: 25346997]
52. Komiak SY, Benbasat I. The effects of personalization and familiarity on trust and adoption of recommendation agents. MIS Q 2006 Dec;30(4):941-960 [FREE Full text] [doi: 10.2307/25148760]
53. Parasuraman A, Colby CL. An updated and streamlined technology readiness index: TRI 2.0. J Serv Res 2014 Jun 22;18(1):59-74 [FREE Full text] [doi: 10.1177/1094670514539730]
54. Lehmann V, Zueger T, Maritsch M, Notter M, Bérubé C, Albrecht C, et al. 234-OR: noninvasive hypoglycemia detection during real car driving using in-vehicle data. Diabetes 2022 Jun 1;71(Supplement_1):234. [doi: 10.2337/db22-234-OR]
55. Bach KM, Jæger MG, Skov MB, Thomassen NG. Evaluating driver attention and driving behaviour: comparing controlled driving and simulated driving. In: Proceedings of the People and Computers XXII Culture, Creativity, Interaction. 2008 Presented at: People and Computers XXII Culture, Creativity, Interaction; September 1-8, 2008; Liverpool, UK. [doi: 10.14236/ewic/HCI2008.19]
56. Heise T, Koschinsky T, Heinemann L, Lodwig V, Glucose Monitoring Study Group. Hypoglycemia warning signal and glucose sensors: requirements and concepts. Diabetes Technol Ther 2003 Aug;5(4):563-571. [doi: 10.1089/152091503322250587] [Medline: 14511411]

57. Sun C, Malcolm JC, Wong B, Shorr R, Doyle MA. Improving glycemic control in adults and children with type 1 diabetes with the use of smartphone-based mobile applications: a systematic review. Can J Diabetes 2019 Feb;43(1):51-8.e3 [FREE Full text] [doi: 10.1016/j.jcjd.2018.03.010] [Medline: 30026048]
58. Feuerstein-Simon C, Bzdick S, Padmanabhuni A, Bains P, Roe C, Weinstock RS. Use of a smartphone application to reduce hypoglycemia in type 1 diabetes: a pilot study. J Diabetes Sci Technol 2018 Nov;12(6):1192-1199 [FREE Full text] [doi: 10.1177/1932296817749859] [Medline: 29291641]
59. Pagiatakis C, Rivest-Hénault D, Roy D, Thibault F, Jiang D. Intelligent interaction interface for medical emergencies: application to mobile hypoglycemia management. Smart Health 2020 Mar;15:100091. [doi: 10.1016/j.smhl.2019.100091]
60. The Nightscout Foundation. URL: https://www.nightscoutfoundation.org/ [accessed 2023-12-22]
61. Cox DJ, Singh H, Lorber D, Hermayer K. Diabetes and driving safety: science, ethics, legality and practice. Am J Med Sci 2013 Apr;345(4):263-265 [FREE Full text] [doi: 10.1097/maj.0b013e31828bf8d7]
62. Tobisch V. Why proactive voice AI will determine the Next In-Car UX. Cerence. 2020 Dec 9. URL: https://cerence.com/news-releases/news-release-details/why-proactive-voice-ai-will-determine-next-car-ux#:~:text=The%20concept%20of%20proactivity%20not,but%20have%20never%20reached%20them [accessed 2023-12-22]
63. Brito CF. Demonstrating experimenter and participant bias. In: Stowell JR, Addison WE, editors. Activities for Teaching Statistics and Research Methods: A Guide for Psychology Instructors. Washington, DC: American Psychological Association; 2017.
64. Maritsch M, Föll S, Lehmann V, Bérubé C, Kraus M, Feuerriegel S, et al. Towards wearable-based hypoglycemia detection and warning in diabetes. In: Proceedings of the Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI EA '20; April 25-30, 2020; Honolulu, HI. [doi: 10.1145/3334480.3382808]
65. Heijmans M, Habets JG, Herff C, Aarts J, Stevens A, Kuijf ML, et al. Monitoring Parkinson's disease symptoms during daily life: a feasibility study. NPJ Parkinsons Dis 2019 Sep 30;5(1):21 [FREE Full text] [doi: 10.1038/s41531-019-0093-5] [Medline: 31583270]
66. Papathanail I, Vasiloglou MF, Stathopoulou T, Ghosh A, Baumann M, Faeh D, et al. A feasibility study to assess Mediterranean diet adherence using an AI-powered system. Sci Rep 2022 Oct 11;12(1):17008 [FREE Full text] [doi: 10.1038/s41598-022-21421-y] [Medline: 36220998]
67. Payne O, Smith V, Rolnik DL, Davies-Tuck M, Warty R, Seo D, et al. Virtual reality and its use in post-operative pain following laparoscopy: a feasibility study. Sci Rep 2022 Jul 30;12(1):13137 [FREE Full text] [doi: 10.1038/s41598-022-17183-2] [Medline: 35907935]
68. Kujala S, Roto V, Väänänen-Vainio-Mattila K, Karapanos E, Sinnelä A. UX Curve: a method for evaluating long-term user experience. Interact Comput 2011 Sep;23(5):473-483. [doi: 10.1016/j.intcom.2011.06.005]
69. Dell N, Vaidyanathan V, Medhi I, Cutrell E, Thies W. "Yours is better!": participant response bias in HCI. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2012 Presented at: CHI '12; May 5-10, 2012; Austin, Texas URL: https://doi.org/10.1145/2207676.2208589 [doi: 10.1145/2207676.2208589]

Abbreviations

ATI: Affinity for Technology Interaction
AUT: Acceptance and Use of Technology
CGM: continuous glucose monitoring
FGM: flash glucose monitoring
meCUE: modular evaluation of key Components of User Experience
RQ: research question
SAI: Session Alliance Inventory
T1DM: type 1 diabetes mellitus
TRI: Technology Readiness Index

Edited by A Kushniruk; submitted 15.12.22; peer-reviewed by M Chatzimina; comments to author 31.01.23; revised version received 06.02.23; accepted 04.12.23; published 09.01.24.

Please cite as:
Bérubé C, Lehmann VF, Maritsch M, Kraus M, Feuerriegel S, Wortmann F, Züger T, Stettler C, Fleisch E, Kocaballi AB, Kowatsch T
Effectiveness and User Perception of an In-Vehicle Voice Warning for Hypoglycemia: Development and Feasibility Trial
JMIR Hum Factors 2024;11:e42823
URL: https://humanfactors.jmir.org/2024/1/e42823
doi:10.2196/42823
PMID:38194257

Original Paper

A Closed-Loop Falls Monitoring and Prevention App for Multiple Sclerosis Clinical Practice: Human-Centered Design of the Multiple Sclerosis Falls InsightTrack

Valerie J Block[1,2], PT, DPTSc; Kanishka Koshal[1], BDS, MSc; Jaeleene Wijangco[1], BS; Nicolette Miller[1], BS; Narender Sara[1], MBA, MS; Kyra Henderson[1], BA; Jennifer Reihm[1], MPA; Arpita Gopal[2], DPT, PhD; Sonam D Mohan[1], MD, PhD; Jeffrey M Gelfand[1], MAS, MD; Chu-Yueh Guo[1], MD; Lauren Oommen[1], DO; Alyssa Nylander[1], MD, PhD; James A Rowson[1], PhD; Ethan Brown[1], MD; Stephen Sanders[1], PhD; Katherine Rankin[1], PhD; Courtney R Lyles[3,4], PhD; Ida Sim[5], MD, PhD; Riley Bove[1], MMSc, MD

[1]Department of Neurology, University of California San Francisco Weill Institute, University of California San Francisco, San Francisco, CA, United States

[2]Department of Physical Therapy and Rehabilitation Science, University of California San Francisco, San Francisco, CA, United States

[3]University of California San Francisco Division of General Internal Medicine, Zuckerberg San Francisco General Hospital, San Francisco, CA, United States

[4]Center for Vulnerable Populations, University of California San Francisco, San Francisco, CA, United States

[5]Department of Medicine, University of California San Francisco, San Francisco, CA, United States

Corresponding Author:
Riley Bove, MMSc, MD
Department of Neurology
University of California San Francisco Weill Institute
University of California San Francisco
Box 3126 1651 4th St
Room 612A
San Francisco, CA, 94143
United States
Phone: 1 (415) 353 2069
Email: riley.bove@ucsf.edu

Abstract

Background: Falls are common in people with multiple sclerosis (MS), causing injuries, fear of falling, and loss of independence. Although targeted interventions (physical therapy) can help, patients underreport and clinicians undertreat this issue. Patient-generated data, combined with clinical data, can support the prediction of falls and lead to timely intervention (including referral to specialized physical therapy). To be actionable, such data must be efficiently delivered to clinicians, with care customized to the patient's specific context.

Objective: This study aims to describe the iterative process of the design and development of Multiple Sclerosis Falls InsightTrack (MS-FIT), identifying the clinical and technological features of this closed-loop app designed to support streamlined falls reporting, timely falls evaluation, and comprehensive and sustained falls prevention efforts.

Methods: Stakeholders were engaged in a *double diamond* process of human-centered design to ensure that technological features aligned with users' needs. Patient and clinician interviews were designed to elicit insight around ability blockers and boosters using the capability, opportunity, motivation, and behavior (COM-B) framework to facilitate subsequent mapping to the Behavior Change Wheel. To support generalizability, patients and experts from other clinical conditions associated with falls (geriatrics, orthopedics, and Parkinson disease) were also engaged. Designs were iterated based on each round of feedback, and final mock-ups were tested during routine clinical visits.

Results: A sample of 30 patients and 14 clinicians provided at least 1 round of feedback. To support falls reporting, patients favored a simple biweekly survey built using REDCap (Research Electronic Data Capture; Vanderbilt University) to support *bring-your-own-device* accessibility—with optional additional context (the severity and location of falls). To support the evaluation and prevention of falls, clinicians favored a clinical dashboard featuring several key visualization widgets: a longitudinal falls display coded by the time of data capture, severity, and context; a comprehensive, multidisciplinary, and evidence-based checklist of actions intended to evaluate and prevent falls; and MS resources local to a patient's community. In-basket messaging alerts clinicians of severe falls. The tool scored highly for usability, likability, usefulness, and perceived effectiveness (based on the Health IT Usability Evaluation Model scoring).

Conclusions: To our knowledge, this is the first falls app designed using human-centered design to prioritize behavior change and, while being accessible at home for patients, to deliver actionable data to clinicians at the point of care. MS-FIT streamlines data delivery to clinicians via an electronic health record–embedded window, aligning with the *5 rights* approach. Leveraging MS-FIT for data processing and algorithms minimizes clinician load while boosting care quality. Our innovation seamlessly integrates real-world patient-generated data as well as clinical and community-level factors, empowering self-care and addressing the impact of falls in people with MS. Preliminary findings indicate wider relevance, extending to other neurological conditions associated with falls and their consequences.

(JMIR Hum Factors 2024;11:e49331) doi:10.2196/49331

KEYWORDS

digital health; mobile tools; falls; prevention; behavioral medicine; implementation science; closed-loop monitoring; multiple sclerosis; mobile phone

Introduction

Background

Falls are common in patients with multiple sclerosis (MS), occurring in 50% to 70% of published cohorts, a rate similar to that of older adults [1]. Falls often lead to injury, result in significant health care costs [2-5], and increase *thefear of falling* [6,7]; furthermore, they lead to a decline in physical activity and participation in daily life as well as cause loss of independence [8,9]. Targeted interventions such as physical therapy (PT) can reduce falls and the fear of falling [10-12], but patients often underreport and clinicians undertreat this issue. Indeed, fewer than half of the people with MS who report falls receive falls prevention information from their clinician [13], and there is a lack of self-management apps to engage and empower people with MS about falls prevention [14-16].

To address this gap, multimodal closed-loop tools hold promise. Closed-loop tools can use real-time feedback and patient-generated data (PGD; such as those already validated in MS [17-22]) to continuously monitor and adjust interventions to improve outcomes. Such an approach has been used in biological functions and symptoms, such as insulin delivery or depression [23-25]. Unfortunately, in MS, apps on the commercial market exist outside of the health system, that is, away from the point of care. To close these gaps in care, a tool should close the loop of information flow from the patient to the appropriate clinician (depending on the diagnosis and symptoms being treated, ie, neurologist) at the point of care and back to the patient to support patient-centered care. Furthermore, the tool must address the behavioral barriers to change to promote the behaviors (eg, reporting, screening, treatment recommendations, and follow-up with timely refills or referral scheduling) likely to lead to falls prevention. From previous work, real-time PGD such as prospective near-falls reports, patient-reported outcomes, and changes in step count captured by wearable sensors all provide useful input for the closed-loop models [26,27]. The integration of these in a multimodal tool would enhance falls prediction accuracy and could act as an early warning system for timely PT referrals, reducing falls risk and related injuries [28,29]. However, challenges lie in delivering PGD to the point of care, granting access for prompt intervention, and active self-management. To be actionable, these PGD, generated from remote devices or patient-reported outcomes, must be delivered according to the *5 rights* [30]: the right information, to the right person, in the right format, through the right channel, at the right time in the workflow. This is a hurdle that health systems have for the most part not yet overcome, and PGD are not typically integrated into care systems.

To address these challenges, we developed Multiple Sclerosis Falls InsightTrack (MS-FIT), a closed-loop falls monitoring and prevention app. MS-FIT enables seamless information exchange between patients and clinicians, driven by stakeholder input and human-centered design (HCD) principles [31,32]. It empowers individuals with MS to *track* falls, enhances clinician decision-making by providing real-world *insights*, and fills a crucial gap in self-management for falls monitoring and prevention.

Objectives

This paper describes the iterative process of the design and development of MS-FIT. MS-FIT is designed to integrate various data types to personalize falls risk assessments and interventions for individuals with MS. To achieve this, a planned process of engagement of patients and clinicians (ie, neurologists) was performed to ensure that MS-FIT aligns with user needs, whereas usability evaluations validated its potential impact on falls prevention. Subsequently, we will test the feasibility of implementation and effectiveness of MS-FIT in a larger clinical trial.

Methods

Study Setting

The primary clinical setting is the University of California San Francisco (UCSF) Multiple Sclerosis and Neuroinflammation Center, which provides specialized care to >6000 adults with MS annually. Clinician stakeholders were approached via email or in person and invited to participate in the study. Patients who had given permission to be contacted for research participation or who had sustained falls in the past year were invited via secure email to participate as stakeholders.

Ethical Considerations

The University of California San Francisco Institutional Review Board approved all study activities (22-36680). Informed consent forms and Health Insurance Portability and Accountability Act documents were signed by each study participant (patients, clinicians, and other interviewees). Patients received US $50 (1-time compensation) for their participation in the study.

Study Design

The overarching approach was grounded in the principles and phases of HCD [31]. This process focuses on the usability and needs of those whom the tool is meant to serve, in this case, patients and clinicians. The development protocols included (1) thorough engagement from a comprehensive range of stakeholders, (2) models based on HCD approaches to ensure alignment with the needs of the intended users (patients and clinicians), (3) an evaluation of the tool's usability using an established framework: the Health IT Usability Evaluation Model (Health-ITUEM) [33], and (4) plans to support the generalizability and scalability of the tool to other clinical settings associated with falls.

HCD involves a series of steps, articulated initially in the context of design [34] and expanded to health care [35]: inspire (empathize with all stakeholders), ideate (define the problem and conceptualize in an open-minded manner), implement (prototype solutions and test), and iterate. Figure 1 illustrates these phases in a *modified double diamond* approach as they were undertaken in the current project, depicting the iterative broadening and narrowing of content and layout throughout the phases [36]. Figure 2 shows the trajectory of MS-FIT and the assimilation of insights obtained from user interviews (involving patients and clinicians) throughout the phases of discover, define, develop (iterative), and deliver.

The initial prototype (prediscover) was developed based on feedback from extensive HCD of the BRIDGE point-of-care clinical dashboard (refer to the Technological Building Blocks subsection) summarized elsewhere [37,38], where both patients and clinicians expressed a desire for the integrations of features and episodes of falls to be incorporated into the design. The study team initially identified key elements for MS-FIT through a combination of clinical expertise and literature review [39,40] (Figure 3). These elements were then amalgamated into mock app screens using PowerPoint (Microsoft Corp) for the first round of patient interviews. Figure 3 illustrates the inaugural prototype, which was informed by valuable insights from observational [41] and interventional [39] studies that used PGD to monitor walking and falls in individuals with MS. In addition, the prototype draws inspiration from clinician-facing [42] and patient-facing [43] apps designed using HCD principles to promote shared decision-making and evidence-based practice in MS.

Figure 1. Modified double diamond approach: phases of development and stakeholder engagement. The double diamond depicts the human-centered design principles and framework, with iterations through the discover, define, develop, and delivery phases. The timeline and workflow of the human-centered design phases depict corresponding interviews and products. The curved arrows between "Define" and "Develop" indicate an iterative process between these 2 phases. MS: multiple sclerosis; MS-FIT: Multiple Sclerosis Falls InsightTrack; MVP: minimum viable product; PD: Parkinson disease; REDCap: Research Electronic Data Capture.

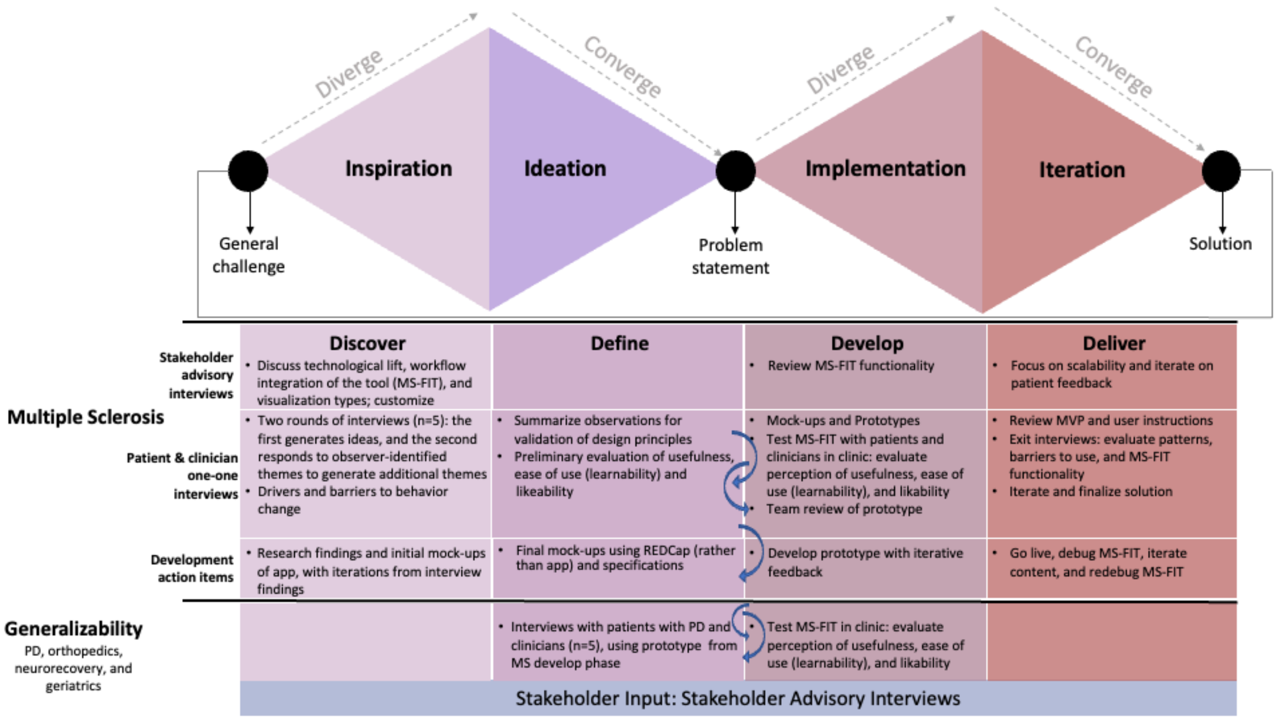


Figure 2. The trajectory of Multiple Sclerosis Falls InsightTrack (MS-FIT) though the phases of development and stakeholder engagement. The final tool components include a patient survey (MS-FIT patient survey) and a clinical dashboard (MS-FIT BRIDGE). The trajectory integrates feedback from user (patient and clinician) interviews through the phases of discover, define, develop (iterative), and deliver. The version numbers indicate a revised version of the patient- or clinician-facing prototype. “Other patients” refers to patients with Parkinson disease as well as orthopedics, neurorecovery, and geriatrics populations. “Full test” refers to the prototype testing in the contextual environment. MS: multiple sclerosis.

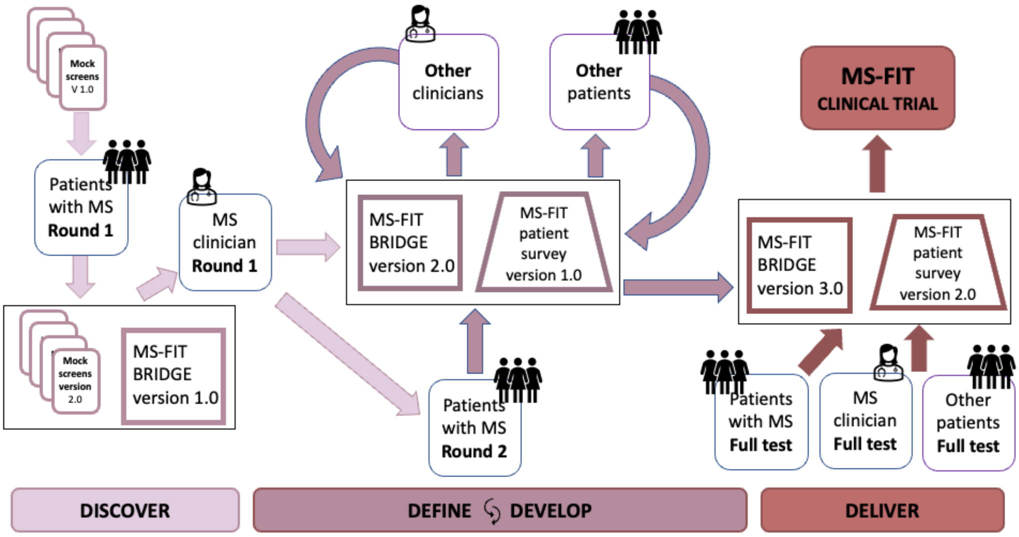
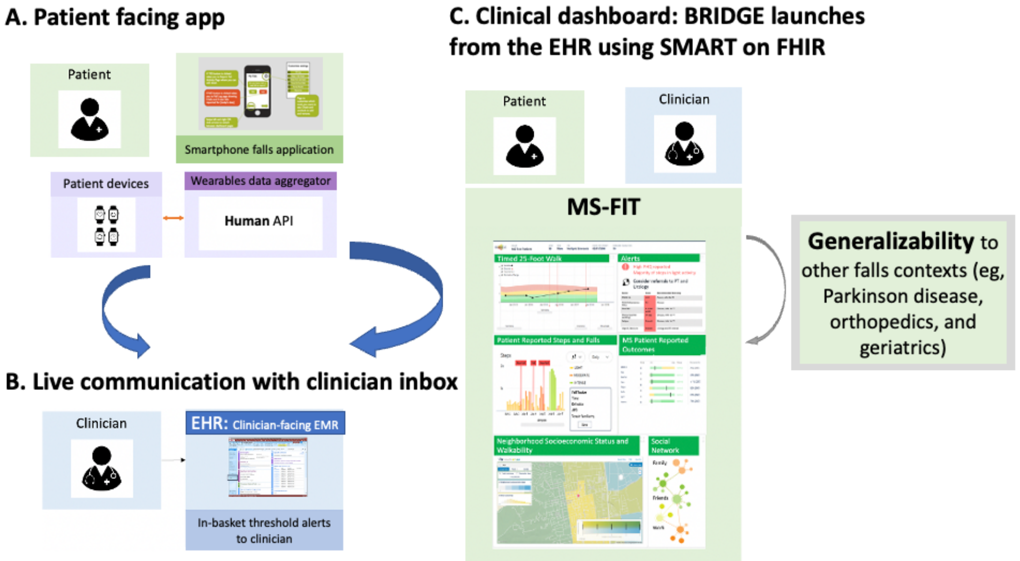


Figure 3. Initial proposal for Multiple Sclerosis Falls InsightTrack (MS-FIT), which involved designing a closed-loop integrated MS-FIT personal health library. MS-FIT is designed to enable patients to track their falls in the context of their lived experience, report them to their care team, and gain insight into multimodal contributors to falls, falls’ impact on daily life participation, and interventions likely to prevent falls. Clinicians, by using BRIDGE, can gain insight into which patients are falling between clinical encounters and how best to personalize risk reduction interventions for the individual patient. This prototype was generated from a number of insights from observational and interventional studies that used patient-generated data to monitor walking and falls in people with multiple sclerosis (MS) and from clinician-facing [42] and patient-facing [43] apps designed using human-centered design to facilitate shared decision-making and evidence-based practice in MS. (A) Patient facing app; (B) Live communication with clinician inbox; (C) Clinical dashboard: BRIDGE launches from the EHR using SMART or FHIR. API: application programming interface; EHR: electronic health record; EMR: electronic medical record; FHIR: Fast Healthcare Interoperability Resources; MS-FIT: Multiple Sclerosis Falls InsightTrack; SMART: Substitutable Medical Apps and Reusable Technologies.



Framework for Tool Evaluation

The Health-ITUEM framework appraises both subjective and objective outcomes that inform a tool’s usability [44]. In the design phases described herein, the subjective outcomes (satisfaction measured by the perceived ease of use and perceived usefulness) were primarily evaluated. Furthermore, the 4 key variables proposed by Mathews et al [45] to determine both (1) whether the tool (MS-FIT) reflects HCD principles and (2) whether it is likely to engage patients were applied. These four domains encompass (1) usefulness, (2) ease of use or learnability, (3) likability, and (4) effectiveness. These frameworks were used to categorize critical data and visualization elements, as well as the technological and clinical workflow aspects of MS-FIT [46].

Technological Building Blocks

The architecture of the tool was built leveraging existing tools, primarily BRIDGE and REDCap (Research Electronic Data Capture; Vanderbilt University).

BRIDGE

The BRIDGE precision medicine platform at UCSF is an application programming interface (API) that assembles clinical and research data from a variety of sources into a dashboard customized for a given clinical context, displaying a series of digestible, actionable visualizations [38]. BRIDGE is integrated with the Epic electronic health record (EHR; Epic Systems Corporation), launches from Epic using Substitutable Medical Apps and Reusable Technologies (SMART) on Fast Healthcare Interoperability Resources (FHIR; a standard approach for building reusable and extendable EHR-integrated apps), and is integrated with Epic FHIR APIs and other data integrations. The back-end of BRIDGE is built using Python, the flask framework, and PostgreSQL to store configuration data. Although individual-level data will populate the tool, cohort-level data can become the reference cohort against which an individual's data can be contextualized. BRIDGE pulls data not only from the EHR but also from a range of custom research databases as well as other APIs, such as REDCap [38]. BRIDGE was developed based on extensive HCD processes both within the field of MS [42,43] and beyond [37]. The data visualizations can be developed using HTML, cascading style sheets (CSS), JavaScript, Data-Driven Documents–JavaScript, and other front-end libraries. Each front-end visualization is modular, allowing for asynchronous loading, and is a parameterized JavaScript component, allowing us to extend the code to additional platforms and data sources. Data formatting standards are also applied to make all visualizations and data inputs modular. All API calls are made in real time; BRIDGE does not store patient data, but there is an option to write back to the EHR by pasting the visuals into a clinical note. Furthermore, the development team follows universal design principles, influenced by the Agency for Healthcare Research and Quality Toolkit for Designing Consumer Health IT [47].

REDCap Tool

REDCap [38] includes editable or annotatable functions to enable patients to keep track of, and annotate, their PGD. Design choices reflect digital health literacy principles and feedback provided from diverse patients. Together, these enhancements make the data understandable and actionable.

Investigator Team

The core team included an MS neurologist with HCD expertise (RB), software engineers (NM and NS), a health literacy and patient engagement expert (JR), and an MS physical therapist with remote ambulatory and falls monitoring expertise (VJB). Additional key scientific input was provided by a digital health cloud infrastructure expert (IS), an implementation science expert (CL), a health disparities and population health expert (CL), and an expert in large-scale mobile health (IS). Patient stakeholders included National Multiple Sclerosis Society advocates (LG) and patients (3 core stakeholders). Research team members included a program manager (KK) and clinical coordinators (JW and KH). Before starting the project, this team met to determine the phases of research and design an initial mock-up of the tool that could be used during the discover phase. Volunteer consultants included a software engineer (JR) and user interface or user experience experts.

Phases of Design

Phase 1: Discover

Stakeholder Advisory Team

An initial stakeholder meeting took place, during which the goals and phases of the project were outlined. Next, the core team met biweekly as a group or as subgroups to discuss an agenda that included the development of patient and clinician interview guides, interview coding schemes and thematic analysis, the practical aspects of the technological lift, the workflow integration of MS-FIT, and the visualization types and customizations. The iterations of mock-ups were revised based on patient and clinician interview feedback.

Interviews

One-on-one interviews were conducted by the health literacy and patient engagement expert (hereinafter referred to as the interviewer) with patients (round 1) and clinicians. Because of ongoing COVID-19 restrictions on in-person engagements, interviews were conducted via the UCSF Zoom video platform (Zoom Video Communications, Inc) using interview guides developed for each audience to elucidate how a tool might be designed to promote behavior change around falls ascertainment, reporting, and prevention. All questions were administered verbally, and interviews lasted between 45 and 60 minutes. With participant consent, interviews were simultaneously recorded and transcribed using Zoom's video transcription feature.

Interview guides included qualitative and quantitative components. Open-ended questions probed around the domains of the capability, opportunity, motivation, and behavior (COM-B) framework to facilitate subsequent mapping to the Behavioral Change Wheel (BCW) proposed by Michie et al [48]. Quantitative questions with Likert-style responses (ranging from 1=lowest to 5=highest) were administered verbally throughout each interview to assess specific aspects of patient and clinician experience related to capability, opportunity, and motivation, as well as the perceived usefulness of mock screen views and workflows. Participants were asked to comment on their Likert-style responses.

Patient interviews were semistructured around 2 key thematic topics: patient experience with (1) falls and activity, including ability to be active, knowledge, communication with care team, experience, feelings, and expectations; and (2) use of technology, including smartphone, tracking devices, apps, and communication with care team. To complement qualitative insights, patients were asked to use a Likert scale to rate the perceived usefulness of each of 3 app screen views featuring different design elements.

Semistructured interviews with clinicians started with a review of the activity blockers and boosters identified during the discover phase interviews with patients. With this insight, clinicians were asked a series of open-ended qualitative questions to elicit their perspectives on whether a falls reporting tool might promote sustainable falls prevention, as well as gather feedback on the initial closed-loop design (Figure 3) intended to support falls treatment and clinical decision support. To assess each design feature, clinicians were asked to rate perceived usefulness on a Likert scale.

Analysis

After all interviews were concluded for each audience, the interviewer reviewed each transcript and used inductive coding to develop a coding scheme on the basis of responses to the open-ended questions [46]. Frequently occurring topics were assigned a unique thematic category, and less frequent topics were coded *other*. Categories were defined by the interviewer, and quotations from the transcript were used to illustrate the type of text coded into the category. Although the interviewer was the sole coder, the stakeholder advisory team provided ongoing consultation on the coding scheme and how to code less frequently occurring responses.

The interviewer transferred Likert-style response data to a spreadsheet to calculate means and SDs for each question. To analyze questions designed to map to the COM-B framework, the interviewer created a data grid where the rows were COM-B categories with subthemes of ability blocker and booster types, and the columns were evidence (quotes) of specific blockers or boosters [49]. Evidence of blockers or boosters that spanned >1 category were placed in all relevant categories to ensure that they would be represented when considering BCW-guided interventions.

After developing the initial COM-B data grid, the interviewer, in consultation with the stakeholder advisory group, expanded the grid to include (1) BCW intervention functions to help users overcome barriers to performing target behaviors and (2) potential intervention solution features designed to be effective for each corresponding blocker category. Intervention solution features were subsequently added to the design road map for immediate or future implementation.

Phases 2 and 3: Define and Develop (Iterative)

Stakeholder Advisory Team

In these phases, the team reviewed qualitative and quantitative findings from additional patient (2 rounds) and clinician (1 round) interviews and used this feedback to further refine MS-FIT tool functionality, including design and technological features. Changes were prioritized according to the strength of feedback (occurrence of themes and usability scores) and technical feasibility.

Interviews

The define and develop phases encompassed a second round of patient interviews, followed by 2 rounds of interviews with clinicians and patients designed to assess MS-FIT generalizability to other high-risk clinical contexts. The same process was followed as that described in phase 1 (discover). One-on-one interviews were conducted by the interviewer via the UCSF Zoom video platform using interview guides. All questions were administered verbally, and interviews lasted between 45 and 60 minutes. With participant consent, interviews were simultaneously recorded and transcribed using Zoom's video transcription feature.

Patient Interviews (Round 2)

Interview guides included qualitative and quantitative components. In an effort to validate the patient experience findings from round 1 interviews, patients interviewed during round 2 were similarly asked to share qualitative feedback around personal experiences with falls, falls and near-falls reporting, perceived benefits and concerns around using a falls tracking app, and thoughts on what supports would be helpful between appointments. Quantitative questions with Likert-style responses (ranging from 1=lowest to 5=highest) were used to rate 9 mock screens for usefulness, understandability, and importance for each view. Mock screens had been iterated after the discover phase; therefore, patient feedback during this second round further validated and helped refine the designs.

Generalizability to Other High-Risk Clinical Contexts

To ensure that the technological build was not *overdesigned* for MS and to support the scalability of the tool to other clinical settings, interviews were expanded to intended users in other clinical specialties associated with falls, including geriatrics, orthopedics, neurorecovery (after stroke or traumatic brain injury), and Parkinson disease (PD). Clinicians from each discipline and patients with PD were interviewed. Interview protocols used during the discover phase were adapted to reference specific disciplines and diseases, whereas the questions (qualitative and quantitative) remained the same to yield a parallel assessment of each audience's experiences, preferences, capabilities, opportunities, and motivations.

Analysis

Qualitative and quantitative interview analysis used the same inductive coding and calculation techniques, respectively, used during the discover phase. The results were analyzed by the interviewer, with ongoing thematic consultation with the stakeholder advisory team, and used to inform and prioritize design and content iterations.

Phase 4: Deliver

Stakeholder Advisory Team

The core team met with stakeholders on an ad hoc small-group basis during this phase to plan observation and tool-scoring protocols, specifically to identify a subset of questions from the Health IT Usability Evaluation Scale (Health-ITUES) derived from the Health-ITUEM to assess the 2 *subjective* components of usability—usefulness and ease of use [33]—as well as the Patient Education Materials Assessment Tool for Audiovisual Materials to assess understandability and actionability [50]. As recommended for digital tool validation [45], a single survey question—Net Promoter Score (NPS)—was asked regarding the likelihood that users (patients and clinicians) would recommend the MS-FIT to colleagues or friends. Additional conversations focused on the scalability of the tool, as well as the qualitative and quantitative feedback received.

Observations and Scoring

Observations and scoring for the patient-facing falls assessment survey took place with 2 audiences: people with MS and people with PD. Patients scheduled for a routine upcoming in-person clinical visit with their neurologist were contacted and invited to participate in testing and evaluating the tool. After providing informed written consent, and while being observed by the interviewer, participants were asked to engage with the MS-FIT minimum viable product consisting of the falls assessment survey and accompanying patient instructions while being observed by the interviewer. Patients were specifically asked to complete the falls assessment survey by entering up to 5 falls (real or hypothetical) that had occurred in the prior 2 weeks and responding to on-screen prompts to provide context about each reported fall. Patients could ask questions of the interviewer, if needed. After survey submission, each patient was asked to complete an 18-item survey about their experience to assess usability, usefulness and ease of use, likability, understandability, actionability, and NPS. Patients were subsequently asked if they had any feedback about their experience. Feedback was documented in field notes captured contemporaneously.

Clinicians seeing people with MS and those with PD who had just been observed entering data in the falls assessment survey were asked to launch the MS-FIT BRIDGE app in real-time clinical encounters with these patients to review the falls and contextual data the patient had entered and to engage with the various widgets designed to help evaluate and address reported falls. The interviewer met with the clinicians immediately after the encounters to conduct in-person exit interviews and administer a 9-item survey to assess usability, usefulness and ease of use, likability, understandability, actionability, and NPS. Clinicians were subsequently asked whether they had any feedback about their experience, including any barriers to use and functionality challenges. Feedback was documented in contemporaneous field notes.

Analysis

Qualitative feedback, although limited, was analyzed by the interviewer using the same inductive coding technique used during the previous 2 interview phases. Quantitative questions with Likert-style responses (ranging from 1=lowest to 5=highest) were used to score likability, usability, usefulness, and ease of use. Understandability and actionability were assessed using a binary *agree* or *disagree* scale. Another member of the research team entered quantitative responses into REDCap, which was used to calculate means and SDs for all Likert-style responses and total binary responses. NPS responses (0-10 scale) were calculated by the interviewer by subtracting the percentage who were detractors (those who scored 0 to 6) from the percentage who were promoters (those who scored 9 or 10). An NPS >0 was considered good, >20 was considered favorable, and >50 was considered excellent.

Development Action Items

Once the tool was live, the developer was able to debug MS-FIT; iterate based on patient, clinician, and stakeholder feedback; and redebug as needed.

Results

Overview

Demographic information about each interview panel is shown in Multimedia Appendix 1 [42]. Altogether, 30 patients of diverse ages, disability levels, and technological literacy as well as 14 clinicians provided at least 1 round of feedback. The level of involvement from the users ranged from testers to informants [32]. Feedback from both rounds of interviews with people with MS, MS clinician comments, and feedback from other high-risk clinical context patient and clinician interviews were integrated into the final MS-FIT design. Iterative interview feedback was categorized into activity blockers (what keeps people from performing a behavior) and boosters (what is already working well that we can build on) in the COM-B model. Examples of how interview feedback findings fit into the COM-B and BCW, along with intervention function solution features integrated into MS-FIT, are shown in Figure 4. Details are provided in Table 1. Further discussions with clinicians and patients in other high-risk clinical contexts confirmed the findings from the MS context. Across these specialties, the main barriers to falls prevention efforts included access to specialized PT (availability and physical ability to access it), insurance coverage, ability to adapt the home to improve safety, the adequate use of assistive devices, and COVID-19–related restrictions to community exercise areas.

The overview of findings from interviews with clinicians and participants with MS, highlighting areas that block or boost patient and clinician behavior change with regard to falls and falls prevention, are shown in Multimedia Appendix 2.

Figure 4. Example of mapping blockers and boosters relating to falls prevention (findings from interviews with patients with multiple sclerosis [MS]) to the Behavioral Change Wheel [49] and associated behaviorally informed intervention solution features. Examples for each of the sections of the capability, opportunity, motivation, and behavior (COM-B) model are highlighted, showing how these integrate into the Behavioral Change Wheel. The examples provided relate to patients' reported goals, blockers (features that block falls prevention behavior), and boosters (features that boost behaviors related to falls prevention).

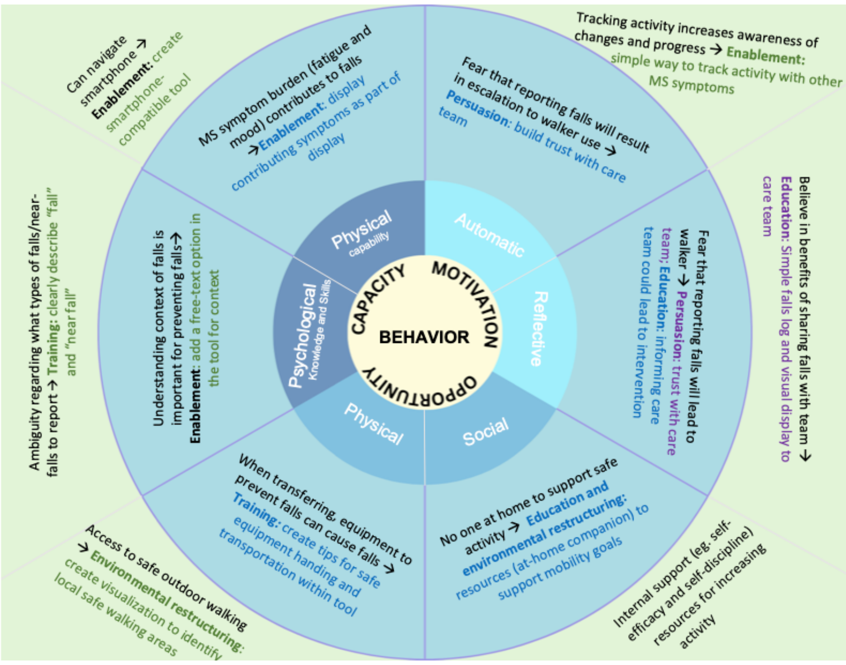


Table 1. Scoring of the final University of California San Francisco Multiple Sclerosis Falls InsightTrack app (REDCap [Research Electronic Data Capture]) by patients with multiple sclerosis: usability, ease of use, and likability (n=10).

Health-ITUES[a]–based questions for usability, ease of use, and likability	Score, mean (SD)	Score <4 out of 5, n (%)
"It is useful to report if I've had any falls or near falls every 2 weeks"	4.80 (0.42)	0 (0)
"It is useful to have my survey answers sent to my care team"	4.90 (0.32)	0 (0)
"The survey asks about important topics"	4.70 (0.48)	0 (0)
"I am comfortable with my ability to complete the survey"	4.80 (0.42)	0 (0)
"I find the survey easy to use"	4.80 (0.42)	0 (0)
"I can easily remember how to access the survey through my email"	4.60 (0.70)	1 (10)
"I like the survey"	4.80 (0.42)	0 (0)

[a]Health-ITUES: Health IT Usability Evaluation Scale (scores range from 1=strongly disagree to 5=strongly agree).

Tool Components

Thematic saturation was reached after 5 patient interviews (round 1), and we incorporated these insights into prototypes for an additional 5 patient interviews (round 2), which were then iteratively reviewed.

UCSF Support Self-Monitoring: A Patient-Facing Tool to Track Falls and Self-Monitor

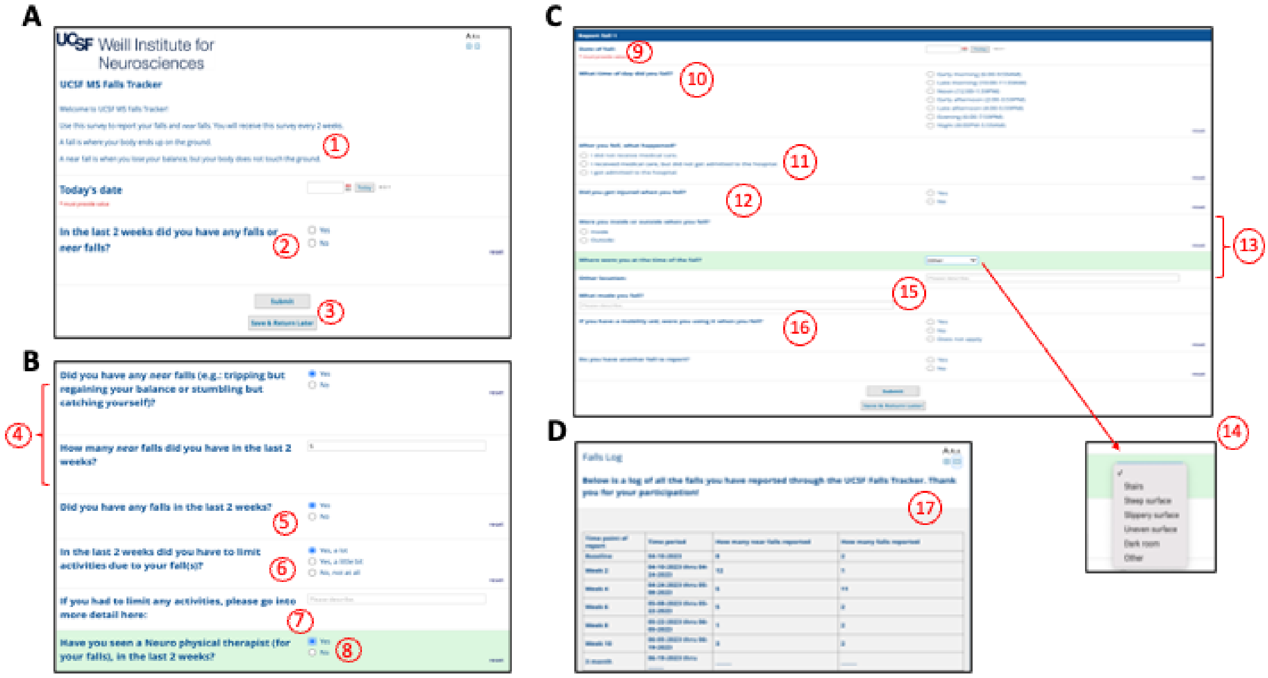
Tool Architecture

One key and consistent theme emerging from patient interviews was a preference for a simpler design for the patient-facing tool than had been initially conceived. Combined with a goal of maintaining confidentiality and keeping personal information within our university firewall, the study team opted for a tailored REDCap app rather than a custom new app.

Tool Components

Key features informed by patient and clinician feedback are detailed in Figure 5. Key features mapping to the COM-B framework (Multimedia Appendix 2 and Figure 4) are denoted by a red number and described in Textbox 1.

Figure 5. University of California San Francisco (UCSF) Multiple Sclerosis (MS) Falls Tracker: a patient-facing tool to track falls and support self-monitoring. This is the "MS-FIT [Multiple Sclerosis Falls InsightTrack] patient survey V2.0," sent via email to patients with a REDCap (Research Electronic Data Capture) survey link. Key features mapping to user-generated perspectives and feedback and to the capability, opportunity, motivation, and behavior (COM-B) framework are denoted by a red number and described in Textbox 1.



Textbox 1. University of California San Francisco Multiple Sclerosis Falls InsightTrack: key features. The numbers correspond to the red numbers in Figure 5, which denote key features mapping to the capability, opportunity, motivation, and behavior (COM-B) framework.

Concise and precise falls screening

1. Clear definitions were preferred to distinguish between a fall and a near fall to support the reporting of meaningful data.
2. An easy-to-use and simple 1-question tool that could be completed frequently (every 2 wk) was preferred to relying on "flawed memory" to report falls during sporadic clinic visits: if "No," then the survey ends at this point; if "Yes," then branching logic continues.
3. The ability to easily report each fall or near fall separately was preferred. The ability to edit (return later) was important for reducing burden.
4. Simple reporting for near falls (yes or no and overall number) was preferred, given the large volume of near falls experienced by some patients and the potential burden and time commitment of providing details.
5. The 2-wk epoch between reporting was determined feasible (balance between memory and overburdening).
6. The ability to report activity limitations was preferred because these pertain to primary goals with regard to the "ability to continue independence for activities of daily living" and to "stay active."
7. Because of the heterogeneity in answers, a free-text option would allow patients to add further details regarding activity limitations.
8. Indicating whether the patient has seen a neurorehabilitation specialist could help clinicians triage the continued plan of care.

Detailed context of falls (optional)

1. Recording the date of the fall using a simple button allows the tool to display each fall into the longitudinal representation (refer to Textbox 2; Figure 6).
2. The time of falls can also inform falls context (eg, in the dark or when fatigued). The 24-h day was divided into time blocks for clarity and to reduce recall error of exact time.
3. Information regarding the medical consequences of a fall can inform both its severity and the clinical follow-up needed.
4. Injury after a fall is considered distinct from seeking and receiving medical attention.
5. Fall location can inform prevention efforts, including home safety; "some falls inside the home can be avoided through modifications such as removing a rug, better lighting etc."
6. Other details of the fall location can also inform home safety and prevention (eg, curb, stairs, and poor lighting).
7. Specifying whether falls occur because of factors related to multiple sclerosis or other factors (obstacle, etc) is important owing to the heterogeneity of fall triggers and of clinical responses.
8. The question "If you have a mobility aid, were you using it when you fell?" can remind patients to use the assistive device and can cue clinicians of the need to modify or change the current assistive device.
9. A falls log is provided to patients and shows the reported falls over time.

Closing the loop: real-time in-basket messaging

1. Enabling the reporting (patient) and ascertainment (clinician) of falls at regular intervals optimizes timeliness (vs periodic visits) while maintaining low burden (vs daily or "at time of fall"). If a severe fall is reported on the biweekly survey, an in-basket message to the electronic medical record alerts the care team in a manner integrated into the clinical workflow (Figure 6, #15).

Textbox 2. Multiple Sclerosis Falls InsightTrack clinical management dashboard integrated into the Epic electronic health record: key features. The numbers correspond to the red numbers in Figure 6, which denote key features mapping to the capability, opportunity, motivation, and behavior (COM-B) framework.

Longitudinal multiple sclerosis trajectory widget (visualizes patients' disease and medication trajectory over time with integrated normative ranges)

1. Ability to toggle through disability measures (eg, Expanded Disability Status Scale [EDSS] and Timed 25-Foot Walk)
2. Succinct overview of patient's longitudinal MS trajectory, including relapses, disability, medications, and normative data

Longitudinal falls widget (visualizes falls reported every 2 wk by the patient using their patient-facing app [Figure 5] data regarding date, time, and severity of each fall on 1 display)

1. Fall severity visualized by color shade (grading falls by severity considered important to trigger an alert to the care team and to inform type of clinical response)
2. Ability to include a way to visualize the falls log with falls over time
3. Estimated time of day of the fall can inform further interventions needed, including vision check, home safety evaluation, and medication review (especially for Parkinson disease)
4. Time of day visualized with colors for daytime (lighter: yellow) and nighttime (darker: blue) preferred by all stakeholders

Community resources widget (map automatically displays the patient's home community and allows for web-based identification of MS health care professionals in their community)

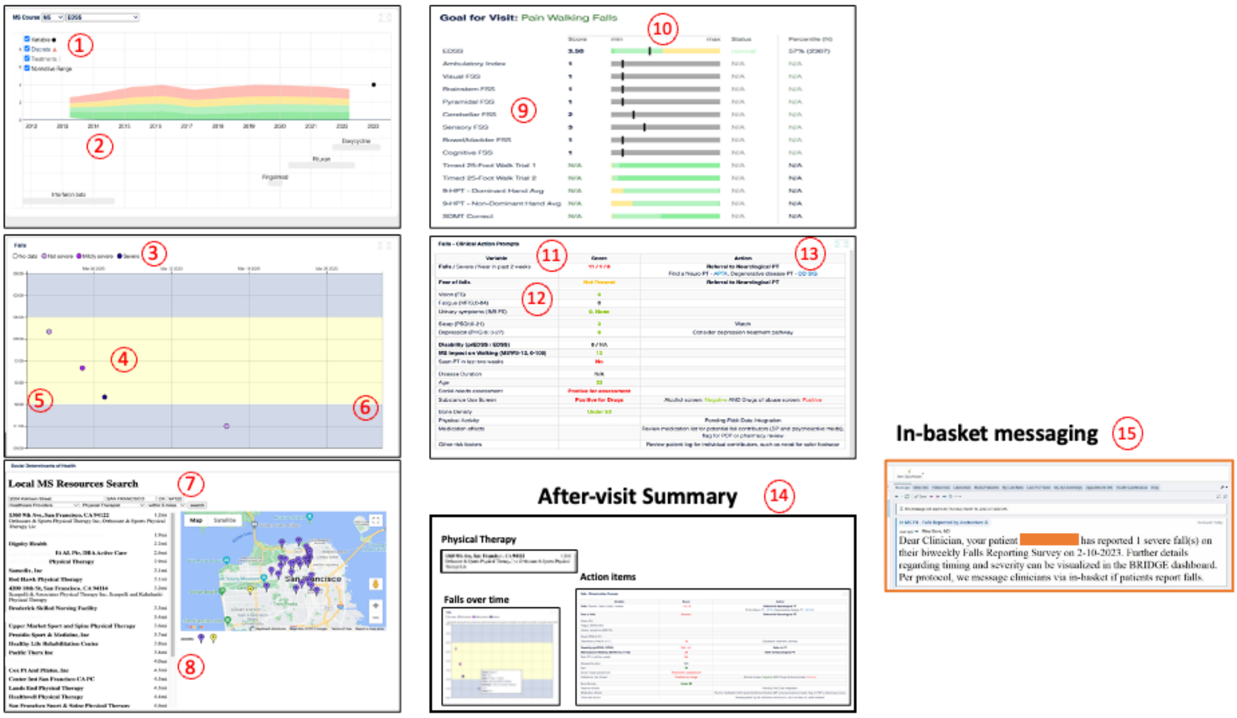
1. Automated display of MS professionals (physical therapist, occupational therapist, and talk therapist) in the patient's community, which reduces barriers for patient to identify local recourses once physical therapy or other referrals have been placed
2. Contact information and driving navigations between the patient's home and the resources automated, can be pasted into the patient's after-visit summary or the clinician's note

Cross-sectional widget (summary display with 2 tabs displaying clinical disability outcomes and patient-reported outcomes [PROs]; clinician can toggle between time points)

1. Clinic-based performance measures (walking speed, hand function, and cognition) and disability outcomes (EDSS with separate functional system scores) as well as PROs can inform a more global assessment of the patient at given time point
2. Color-coded normative ranges can provide rapid assessment of whether patient's given function is within "normal" range

Falls treatment– and action-prompt widget (tabulates core data needed for a comprehensive assessment of falls risk and prevention; for each category, patient's score is colored according to severity, and possible action prompts are displayed)

Figure 6. Multiple Sclerosis Falls InsightTrack (MS-FIT) clinical management dashboard, integrated into the Epic electronic health record (Epic Systems Corporation). This is the "MS-FIT BRIDGE version 3.0," which is viewable from Epic in the electronic health record at the time of the clinic visit. Key features mapping to user-generated perspectives and feedback and to the capability, opportunity, motivation, and behavior (COM-B) framework are denoted by a red number and described in Textbox 2.



Phase 4: Deliver

Altogether, 15 patients (10 with MS and 5 with PD) with an age range of 34 to 79 years and 6 MS clinicians with a clinical experience range of 2 to 22 years (Multimedia Appendix 1) launched the tool components *live* and provided feedback.

Patient-Facing UCSF MS-FIT

People With MS

Of the 15 patients, 10 (67%) had been diagnosed with MS; they had a mean age of 48.8 (SD 8.8; range 34-60) years, with disability level (EDSS score) ranging from 1.5 to 6.0 and a median disease duration of 14.5 (IQR 6.3-24; range 2-27) years. The feedback from people with MS was overwhelmingly positive (Table 1). Likability scores were all NPS≥100 (all promoters). The survey was found to be brief and clear. Patients appreciated the benefit of the closed-loop system and the overall impact on clinical encounters.

Patients With PD

Of the 15 patients, 5 (33%) had been diagnosed with PD; they had a mean age of 60.6 (SD 13.2; range 46-79) years, with a median disability level (Unified Parkinson's Disease Rating Scale score) of 31 (IQR 30.3-8.5; range 17-42) and a median disease duration of 4 (IQR 1.5-8.5, range 1-10) years. Overall, the NPS was found to be 0 (20%-20%, with 1/5, 20% detractor, 1/5, 20% promoter, and 3/5, 60% passive scores that trended toward promoters), indicating that patients with PD could be easily swayed to use MS-FIT. The mean scores on the Health-ITUES questions were all >4 (ie, agree or strongly agree), and only 1 score was <3 out of 5 (Table 2).

Qualitative insights from the interviews revealed that *falling*, *fear of falling*, and *thinking about falling* were "not at the top of their list," in contrast to people with MS. Nevertheless, patients with PD found the tracker "easy to fill out," and they "liked the idea of reporting falls and reporting if [they] experienced fear of falling." Patients with PD felt that it was important to have the ability to increase the font size (incorporated into MS-FIT patient survey v 2.0; Figure 5).

For future use in PD, patients reported that it would be important for ease of use and usability to have the ability to report motor vehicle accidents and specific PD symptoms as they relate to falls risk. Patients with PD also reported greater issues with using an iPad (motor or tremor issues).

Table 2. Scoring of the final University of California San Francisco Multiple Sclerosis Falls Tracker (REDCap [Research Electronic Data Capture]) by patients with Parkinson disease: usability, ease of use, and likability (n=5).

Health-ITUES[a]–based questions for usability, ease of use, and likability	Score, mean (SD)	Score <4 out of 5, n (%)
"It is useful to report if I've had any falls or near falls every 2 weeks"	4.40 (0.89)	1 (20)
"It is useful to have my survey answers sent to my care team"	4.20 (1.22)	1 (20)
"The survey asks about important topics"	4.00 (1.00)	2 (40)
"I am comfortable with my ability to complete the survey"	4.60 (0.89)	1 (20)
"I find the survey easy to use"	4.60 (0.89)	1 (20)
"I can easily remember how to access the survey through my email"	4.20 (1.10)	2 (40)
"I like the survey"	4.20 (1.10)	2 (40)

[a]Health-ITUES: Health IT Usability Evaluation Scale (scores range from 1=strongly disagree to 5=strongly agree).

MS-FIT Clinical Management Dashboard

Overall, the MS clinicians (n=6) rated the dashboard highly (NPS=16.67; Table 3):

> *I like that [the app] summarizes important clinical information in an easily digestible format, and the new widget that includes an MS [multiple sclerosis]-specific review of systems and actionable items seems like it will help ensure well-rounded MS care!* [Clinician 1]

With regard to reporting falls and near falls, the MS clinicians noted multiple benefits to aiding with patient care:

> *You can infer a lot from [fall data] in terms of disease activity, disease course, changes in a patient's life, their living setting, their support. If you see a jump in falls or the onset of falls in a patient who wasn't falling—it is worthy of clinical attention and needs to be addressed. It would give us an objective way to know if interventions are helping to reduce falls.* [Clinician 2]

> *Near falls are particularly underscreened, so any granularity on near falls would be helpful.* [Clinician 3]

> *For some patients, near falls may not be worth reporting—may just be part of life. But other patients it could make sense for. Any change from baseline has potential to be significant. Near falls can be [a] canary in a coal mine.* [Clinician 5]

Table 3. Scoring of the final University of California San Francisco Multiple Sclerosis Falls BRIDGE dashboard by multiple sclerosis clinicians: usability, ease of use, and likability.

Health-ITUES[a]–based questions for usability, ease of use, and likability	Score, mean (SD)	Score <4 out of 5, n (%)
"The information that appears in BRIDGE is useful to me."	4.80 (0.41)	0 (0)
"It is useful to be updated on my patient's significant fall activity between appointments."	4.50 (0.84)	1 (17)
"I find BRIDGE easy to use."	4.20 (0.75)	1 (20)
"I can always remember how to access BRIDGE."	4.00 (1.10)	1 (20)
"I like BRIDGE."	4.50 (0.55)	0 (0)

[a]Health-ITUES: Health IT Usability Evaluation Scale (scores range from 1=strongly disagree to 5=strongly agree).

Discussion

Principal Findings

To our knowledge, this is the first tool designed using the HCD framework, anchored in the COM-B approach to behavior change, and capable of delivering relevant information at the point of care in line with the *5 rights* with the aim of preventing falls in people with MS. Other apps have been developed, although the focus has mainly been on 1 component of falls (eg, evaluating falls risk [51]) at a time. In addition, many large-scale clinical research projects, such as those conducted at the Stanford Center for Digital Health and the Remote Assessment of Disease and Relapse–Central Nervous System program, are exploring applications of wearable data. However, most of the collected wearable data remain inaccessible for visualization or integration within a clinic's EHR. This limitation can impede the effective use of PGD by clinicians and compromise patient-physician collaboration related to PGD [52]. MS-FIT fills a critical gap in multimodal closed-loop self-management apps for falls monitoring and prevention.

Through extensive stakeholder engagement, MS-FIT offers novel aspects of customization, generalizability, and scalability, integrating multiple data streams relevant to reducing falls. It provides rapid personalized in-basket notifications, limited to severe falls, and digitally displays PGD through the EHR, increasing the likelihood of adoption by patients and clinicians.

Designed in collaboration with patients and clinicians, MS-FIT has emerged as a well-received closed-loop tool for tracking falls and reducing falls risk in individuals with MS. Patients liked its brevity, simplicity, and overall utility, recognizing its potential to enhance clinical discussions. The utility of between-visit reporting and contextualized information for identifying modifiable falls risks was acknowledged by both patients and clinicians. The trial phase aims to validate its low-burden design in practice. Clinicians welcomed the closed-loop system, foreseeing proactive interventions and streamlined implementation. Longitudinal falls visualization, incorporating time and severity, along with clinician prompts targeting MS symptoms and medication effects, was favored for its ability to capture often overlooked components during regular visits.

Another noteworthy finding was the minimal number of interviews required to attain thematic saturation in our initial discover phase, indicating that some clear guidance for potentially high-value initial design features was achieved with a minimal sample size. This could be due to the fact that MS-FIT was based on an initial prototype developed during a prediscover phase using patient and clinician feedback. It could also be attributed to homogeneous samples of study participants consulted throughout the discover and design phases. Overall design efficiency was likely aided by the experience and regular input of interprofessional teams.

The ongoing process involves testing MS-FIT in a prospective longitudinal study in a cohort of 100 adults with MS over 12 months. The primary objectives of this larger study include assessing the adoption rate of the tool, evaluating the level of sustained use of the tool, monitoring adherence to falls reporting, and assessing study retention over the 12-month period. Secondary and exploratory analyses will center around the prediction of adoption, sustained use, adherence to action prompts, and study retention. To determine effectiveness, the study will compare in-study falls with a prior falls data set (Fitbit remote monitoring in MS) [41], and patient satisfaction will be assessed during an exit interview.

Scalability

Our approach, characterized by the selection of key technological and clinical features, allows for the scalability and generalizability of the tool's modular infrastructure to various symptoms, conditions, and clinical settings for other high fall-risk diseases as well as other symptoms within MS (eg, bladder dysfunction). Technological factors for scalability include (1) high-quality, widely shareable static visualizations; and (2) optimized industry standards for code sharing with clinicians in other health care settings, such as other MS centers using Epic EHR. However, successful integration into other health systems depends on the internal governance and motivation within each system.

Limitations

All interviews were conducted remotely, using the UCSF Zoom video platform, which may have biased the patient stakeholders to people who are technologically literate and have access to the internet. However, 92% of people in the United States have access to the internet [53], and given that MS-FIT is an app, users (patients or caregivers) are expected to possess a certain level of technical proficiency. Only clinicians at UCSF and patients seen by this (broad) group of clinicians were interviewed; therefore, we may have missed important feedback from a wider cohort of users. Although HCD is favored for user-driven eHealth innovations, certain limitations exist [32], including a narrow focus; thus, exploring alternatives such as value-sensitive design, citizen science, and more-than-human design could enhance inclusivity and impact within eHealth innovation [54]. Finally, having the interviewer serve as the primary coder could have introduced bias into the qualitative analysis process. Stakeholder advisory group engagement in the coding process was an effort to reduce any potential bias.

Conclusions

MS-FIT delivers relevant data to clinicians through an embedded window within the EHR, following the *5 rights* approach. By using MS-FIT for data processing and algorithms, we reduce clinician burden while enhancing care. Our innovation extends to enabling and integrating real-world PGD as well as clinical and community-level factors, providing actionable information to empower self-care and addressing the impact of falls in people with MS. Our preliminary data indicate that this tool and design extend beyond MS and can be applied to other conditions associated with falls as well as the fear of falls and their associated consequences. To test the feasibility and effectiveness of the app, a clinical trial is ongoing (University of California San Francisco Clinical Trials identifier: NCT05837949).

Acknowledgments

This study was funded by the National Library of Medicine award (R01LM013396; principal investigator: RB). VJB is supported by the National Multiple Sclerosis Society Career Transition Award. RB is funded by the National Multiple Sclerosis Society Harry Weaver Scholar Award.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

VJB is funded by the National Multiple Sclerosis Society Career Transition Award. CYG provides medical consulting for EMD Serono, Genentech, and Horizon Therapeutics. JMG receives research support to University of California San Francisco from Genentech, Hoffmann-La Roche, Vigil Neuroscience, and consulting for Arialys. EB reports research funding from the Michael J Fox Foundation, the Gateway Foundation for Brain Research, the National Institutes of Health, and Biogen Inc. EB has also received honoraria for his work as neurology section editor of NEJM Knowledge+ and consulting fees from Rune Labs, Inc. IS is the cofounder of Open mHealth; general assembly member for The Commons Project; serves on the medical advisory board for 98point6 Technologies; is cofounder, board director, and consultant for Vivli; is on the scientific advisory board for Myovant; and holds stocks in Myia. RB is a recipient of funding from the National Multiple Sclerosis Society Harry Weaver Award, the National Institutes of Health, DOD, and NSF, as well as Biogen, Novartis, and Roche Genentech. She has received personal fees for consulting from Alexion, EMD Serono, Horizon, Jansen, Genzyme Sanofi, and TG Therapeutics. All other authors declare no other conflicts of interest.

Multimedia Appendix 1

Demographic information for each round of interviews that led to the development of Multiple Sclerosis Falls InsightTrack.

[DOCX File , 16 KB - humanfactors_v11i1e49331_app1.docx]

Multimedia Appendix 2

Overview of findings from interviews with clinicians and participants with multiple sclerosis (MS), highlighting areas that block or boost patient and clinician behavior change with regard to falls and falls prevention. The table indicates whether intervention solution features were incorporated into the Multiple Sclerosis Falls InsightTrack (MS-FIT) patient survey, the clinician dashboard, or both.

[DOCX File , 33 KB - humanfactors_v11i1e49331_app2.docx]

References

1. Coote S, Comber L, Quinn G, Santoyo-Medina C, Kalron A, Gunn H. Falls in people with multiple sclerosis: risk identification, intervention, and future directions. Int J MS Care 2020;22(6):247-255 [FREE Full text] [doi: 10.7224/1537-2073.2020-014] [Medline: 33424479]
2. Cattaneo D, De Nuzzo C, Fascia T, Macalli M, Pisoni I, Cardini R. Risks of falls in subjects with multiple sclerosis. Arch Phys Med Rehabil 2002 Jun;83(6):864-867. [doi: 10.1053/apmr.2002.32825] [Medline: 12048669]
3. Tajali S, Shaterzadeh-Yazdi MJ, Negahban H, van Dieën JH, Mehravar M, Majdinasab N, et al. Predicting falls among patients with multiple sclerosis: comparison of patient-reported outcomes and performance-based measures of lower extremity functions. Mult Scler Relat Disord 2017 Oct;17:69-74. [doi: 10.1016/j.msard.2017.06.014] [Medline: 29055478]
4. Cameron MH, Nilsagard Y. Balance, gait, and falls in multiple sclerosis. Handb Clin Neurol 2018;159:237-250. [doi: 10.1016/B978-0-444-63916-5.00015-X] [Medline: 30482317]
5. Haddad YK, Bergen G, Florence CS. Estimating the economic burden related to older adult falls by state. J Public Health Manag Pract 2019;25(2):E17-E24 [FREE Full text] [doi: 10.1097/PHH.0000000000000816] [Medline: 29757813]
6. Peterson EW, Cho CC, von Koch L, Finlayson ML. Injurious falls among middle aged and older adults with multiple sclerosis. Arch Phys Med Rehabil 2008 Jun;89(6):1031-1037. [doi: 10.1016/j.apmr.2007.10.043] [Medline: 18503796]
7. Kalron A, Aloni R, Givon U, Menascu S. Fear of falling, not falls, impacts leisure-time physical activity in people with multiple sclerosis. Gait Posture 2018 Sep;65:33-38. [doi: 10.1016/j.gaitpost.2018.06.174] [Medline: 30558943]
8. Choi K, Jeon GS, Cho SI. Prospective study on the impact of fear of falling on functional decline among community dwelling elderly women. Int J Environ Res Public Health 2017 Apr 27;14(5):469 [FREE Full text] [doi: 10.3390/ijerph14050469] [Medline: 28448461]
9. Schoene D, Heller C, Aung YN, Sieber CC, Kemmler W, Freiberger E. A systematic review on the influence of fear of falling on quality of life in older people: is there a role for falls? Clin Interv Aging 2019 Apr 24;14:701-719 [FREE Full text] [doi: 10.2147/CIA.S197857] [Medline: 31190764]
10. Coote S, Sosnoff JJ, Gunn H. Fall incidence as the primary outcome in multiple sclerosis falls-prevention trials: recommendation from the international MS falls prevention research network. Int J MS Care 2014;16(4):178-184 [FREE Full text] [doi: 10.7224/1537-2073.2014-059] [Medline: 25694776]
11. Gunn H, Markevics S, Haas B, Marsden J, Freeman J. Systematic review: the effectiveness of interventions to reduce falls and improve balance in adults with multiple sclerosis. Arch Phys Med Rehabil 2015 Oct;96(10):1898-1912. [doi: 10.1016/j.apmr.2015.05.018] [Medline: 26070975]
12. Lee HS, Park YJ, Park SW. The effects of virtual reality training on function in chronic stroke patients: a systematic review and meta-analysis. Biomed Res Int 2019 Jun 18;2019:7595639 [FREE Full text] [doi: 10.1155/2019/7595639] [Medline: 31317037]
13. Matsuda PN, Shumway-Cook A, Bamer AM, Johnson SL, Amtmann D, Kraft GH. Falls in multiple sclerosis. PM R 2011 Jul 21;3(7):624-32; quiz 632. [doi: 10.1016/j.pmrj.2011.04.015] [Medline: 21777861]
14. Mazuz K, Biswas S, Lindner U. Developing self-management application of fall prevention among older adults: a content and usability evaluation. Front Digit Health 2020 Sep 2;2:11 [FREE Full text] [doi: 10.3389/fdgth.2020.00011] [Medline: 34713024]
15. Rajagopalan R, Litvan I, Jung TP. Fall prediction and prevention systems: recent trends, challenges, and future research directions. Sensors (Basel) 2017 Nov 01;17(11):2509 [FREE Full text] [doi: 10.3390/s17112509] [Medline: 29104256]

16. Bet P, Castro PC, Ponti MA. Fall detection and fall risk assessment in older person using wearable sensors: a systematic review. Int J Med Inform 2019 Oct;130:103946. [doi: 10.1016/j.ijmedinf.2019.08.006] [Medline: 31450081]
17. Klaren RE, Motl RW, Dlugonski D, Sandroff BM, Pilutti LA. Objectively quantified physical activity in persons with multiple sclerosis. Arch Phys Med Rehabil 2013 Dec;94(12):2342-2348. [doi: 10.1016/j.apmr.2013.07.011] [Medline: 23906692]
18. Gelfand JM. Multiple sclerosis: diagnosis, differential diagnosis, and clinical presentation. Handb Clin Neurol 2014;122:269-290. [doi: 10.1016/B978-0-444-52001-2.00011-X] [Medline: 24507522]
19. Hennessey A, Robertson NP, Swingler R, Compston DA. Urinary, faecal and sexual dysfunction in patients with multiple sclerosis. J Neurol 1999 Nov;246(11):1027-1032. [doi: 10.1007/s004150050508] [Medline: 10631634]
20. Motl RW, McAuley E, Dlugonski D. Reactivity in baseline accelerometer data from a physical activity behavioral intervention. Health Psychol 2012 Mar;31(2):172-175. [doi: 10.1037/a0025965] [Medline: 22023436]
21. Motl RW, Pilutti LA, Learmonth YC, Goldman MD, Brown T. Clinical importance of steps taken per day among persons with multiple sclerosis. PLoS One 2013 Sep 4;8(9):e73247 [FREE Full text] [doi: 10.1371/journal.pone.0073247] [Medline: 24023843]
22. Sandroff BM, Dlugonski D, Weikert M, Suh Y, Balantrapu S, Motl RW. Physical activity and multiple sclerosis: new insights regarding inactivity. Acta Neurol Scand 2012 Oct 03;126(4):256-262. [doi: 10.1111/j.1600-0404.2011.01634.x] [Medline: 22211941]
23. Adams SD, Kouzani AZ, Tye SJ, Bennet KE, Berk M. An investigation into closed-loop treatment of neurological disorders based on sensing mitochondrial dysfunction. J Neuroeng Rehabil 2018 Feb 13;15(1):8 [FREE Full text] [doi: 10.1186/s12984-018-0349-z] [Medline: 29439744]
24. Wadwa RP, Reed ZW, Buckingham BA, DeBoer MD, Ekhlaspour L, Forlenza GP, et al. Trial of hybrid closed-loop control in young children with type 1 diabetes. N Engl J Med 2023 Mar 16;388(11):991-1001 [FREE Full text] [doi: 10.1056/NEJMoa2210834] [Medline: 36920756]
25. Lo MC, Widge AS. Closed-loop neuromodulation systems: next-generation treatments for psychiatric illness. Int Rev Psychiatry 2017 Apr;29(2):191-204 [FREE Full text] [doi: 10.1080/09540261.2017.1282438] [Medline: 28523978]
26. Quinn G, Comber L, Galvin R, Coote S. The ability of clinical balance measures to identify falls risk in multiple sclerosis: a systematic review and meta-analysis. Clin Rehabil 2018 May 20;32(5):571-582. [doi: 10.1177/0269215517748714] [Medline: 29260583]
27. Quinn G, Comber L, McGuigan C, Galvin R, Coote S. Discriminative ability and clinical utility of the Timed Up and Go (TUG) in identifying falls risk in people with multiple sclerosis: a prospective cohort study. Clin Rehabil 2019 Feb 14;33(2):317-326. [doi: 10.1177/0269215518793481] [Medline: 30103642]
28. Block V, Pitsch E, Gopal A, Zhao C, Pletcher MJ, Marcus GM, et al. Identifying falling risk factors remotely in multiple sclerosis (2597). Neurology 2021;96(15 Supplement):2597.
29. Sebastião E, Learmonth YC, Motl RW. Lower physical activity in persons with multiple sclerosis at increased fall risk: a cross-sectional study. Am J Phys Med Rehabil 2017 May;96(5):357-361. [doi: 10.1097/PHM.0000000000000581] [Medline: 28415071]
30. Campbell JD. The five "rights" of clinical decision support. J AHIMA 2013 Oct;84(10):42-7; quiz 48. [Medline: 24245088]
31. Brown T. Design thinking. Harv Bus Rev 2008 Jun;86(6):84-92, 141. [Medline: 18605031]
32. Göttgens I, Oertelt-Prigione S. The application of human-centered design approaches in health research and innovation: a narrative review of current practices. JMIR Mhealth Uhealth 2021 Dec 06;9(12):e28102 [FREE Full text] [doi: 10.2196/28102] [Medline: 34874893]
33. Brown W3, Yen PY, Rojas M, Schnall R. Assessment of the health IT usability evaluation model (Health-ITUEM) for evaluating mobile health (mHealth) technology. J Biomed Inform 2013 Dec;46(6):1080-1087 [FREE Full text] [doi: 10.1016/j.jbi.2013.08.001] [Medline: 23973872]
34. Design kit: the human-centered design toolkit. IDEO. 2015 Jun 11. URL: https://www.ideo.com/post/design-kit [accessed 2023-07-07]
35. Matheson GO, Pacione C, Shultz RK, Klügl M. Leveraging human-centered design in chronic disease prevention. Am J Prev Med 2015 Apr;48(4):472-479. [doi: 10.1016/j.amepre.2014.10.014] [Medline: 25700655]
36. The design process: what is the double diamond? Design Council. URL: https://www.designcouncil.org.uk/news-opinion/design-process-what-double-diamond [accessed 2023-07-24]
37. Brown EG, Schleimer E, Bledsoe IO, Rowles W, Miller NA, Sanders SJ, et al. Enhancing clinical information display to improve patient encounters: human-centered design and evaluation of the Parkinson disease-BRIDGE platform. JMIR Hum Factors 2022 May 06;9(2):e33967 [FREE Full text] [doi: 10.2196/33967] [Medline: 35522472]
38. Bove R, Schleimer E, Sukhanov P, Gilson M, Law SM, Barnecut A, et al. Building a precision medicine delivery platform for clinics: the University of California, San Francisco, BRIDGE experience. J Med Internet Res 2022 Feb 15;24(2):e34560 [FREE Full text] [doi: 10.2196/34560] [Medline: 35166689]
39. Block VJ, Gopal A, Rowles W, Yueh C, Gelfand JM, Bove R. CoachMS, an innovative closed-loop, interdisciplinary platform to monitor and proactively treat MS symptoms: a pilot study. Mult Scler J Exp Transl Clin 2021 Feb 02;7(1):2055217321988937 [FREE Full text] [doi: 10.1177/2055217321988937] [Medline: 33796329]

40. Gopal A, Gelfand JM, Bove R, Block VJ. Fall assessment and monitoring in people with multiple sclerosis: a practical evidence-based review for clinicians. Neurol Clin Pract 2023 Oct;13(5):e200184 [FREE Full text] [doi: 10.1212/CPJ.0000000000200184] [Medline: 37720138]
41. Block VJ, Pitsch EA, Gopal A, Zhao C, Pletcher MJ, Marcus GM, et al. Identifying falls remotely in people with multiple sclerosis. J Neurol 2022 Apr;269(4):1889-1898 [FREE Full text] [doi: 10.1007/s00415-021-10743-y] [Medline: 34405267]
42. Bove R, Bruce CA, Lunders CK, Pearce JR, Liu J, Schleimer E, et al. Electronic health record technology designed for the clinical encounter: MS NeuroShare. Neurol Clin Pract 2021 Aug;11(4):318-326 [FREE Full text] [doi: 10.1212/CPJ.0000000000000986] [Medline: 34484932]
43. Schleimer E, Pearce J, Barnecut A, Rowles W, Lizee A, Klein A, et al. A precision medicine tool for patients with multiple sclerosis (the open MS BioScreen): human-centered design and development. J Med Internet Res 2020 Jul 06;22(7):e15605 [FREE Full text] [doi: 10.2196/15605] [Medline: 32628124]
44. Schnall R, Cho H, Liu J. Health information technology usability evaluation scale (health-ITUES) for usability assessment of mobile health technology: validation study. JMIR Mhealth Uhealth 2018 Jan 05;6(1):e4 [FREE Full text] [doi: 10.2196/mhealth.8851] [Medline: 29305343]
45. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Digital health: a path to validation. NPJ Digit Med 2019 May 13;2:38 [FREE Full text] [doi: 10.1038/s41746-019-0111-3] [Medline: 31304384]
46. Frankfort-Nachmias C, Nachmias D. Research Methods in the Social Sciences 7th edition. New York, NY: Worth Publishers; Mar 6, 2007.
47. Designing consumer health IT: a guide for developers and systems designers. Agency for Healthcare Research and Quality. 2012 Sep. URL: https://digital.ahrq.gov/sites/default/files/docs/page/designing-consumer-health-it-a-guide-for-developers-and-systems-designers.pdf [accessed 2023-07-01]
48. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
49. Bucher A. Engaged: Designing for Behavior Change. Brooklyn, NY: Rosenfeld Media; Mar 03, 2020.
50. The patient education materials assessment tool (PEMAT) and user's guide. Agency for Healthcare Research and Quality. URL: https://www.ahrq.gov/health-literacy/patient-education/pemat.html [accessed 2023-04-09]
51. Hsieh K, Fanning J, Frechette M, Sosnoff J. Usability of a fall risk mHealth app for people with multiple sclerosis: mixed methods study. JMIR Hum Factors 2021 Mar 22;8(1):e25604 [FREE Full text] [doi: 10.2196/25604] [Medline: 33749609]
52. Lordon RJ, Mikles SP, Kneale L, Evans HL, Munson SA, Backonja U, et al. How patient-generated health data and patient-reported outcomes affect patient-clinician relationships: a systematic review. Health Informatics J 2020 Dec;26(4):2689-2706 [FREE Full text] [doi: 10.1177/1460458220928184] [Medline: 32567460]
53. Percentage of population using the internet in the United States from 2000 to 2023. Statista. URL: https://www.statista.com/statistics/209117/us-internet-penetration/#:~:text=As%20of%202023%2C%20approximately%2092,internet%20users%20in%20the%20country [accessed 2023-07-01]
54. van Velsen L, Ludden G, Grünloh C. The limitations of user-and human-centered design in an eHealth context and how to move beyond them. J Med Internet Res 2022 Oct 05;24(10):e37341 [FREE Full text] [doi: 10.2196/37341] [Medline: 36197718]

Abbreviations

API: application programming interface
BCW: Behavioral Change Wheel
COM-B: capability, opportunity, motivation, and behavior
CSS: cascading style sheets
EHR: electronic health record
FHIR: Fast Healthcare Interoperability Resources
HCD: human-centered design
Health-ITUEM: Health IT Usability Evaluation Model
Health-ITUES: Health IT Usability Evaluation Scale
MS: multiple sclerosis
MS-FIT: Multiple Sclerosis Falls InsightTrack
PD: Parkinson disease
PGD: patient-generated data
PT: physical therapy
REDCap: Research Electronic Data Capture
SMART: Substitutable Medical Apps and Reusable Technologies
UCSF: University of California San Francisco

Edited by A Choudhury; submitted 26.05.23; peer-reviewed by H Gooch, E Prochaska; comments to author 27.07.23; revised version received 14.08.23; accepted 19.10.23; published 11.01.24.

Please cite as:
Block VJ, Koshal K, Wijangco J, Miller N, Sara N, Henderson K, Reihm J, Gopal A, Mohan SD, Gelfand JM, Guo CY, Oommen L, Nylander A, Rowson JA, Brown E, Sanders S, Rankin K, Lyles CR, Sim I, Bove R
A Closed-Loop Falls Monitoring and Prevention App for Multiple Sclerosis Clinical Practice: Human-Centered Design of the Multiple Sclerosis Falls InsightTrack
JMIR Hum Factors 2024;11:e49331
URL: https://humanfactors.jmir.org/2024/1/e49331
doi:10.2196/49331
PMID:38206662

Original Paper

Digital Triage Tools for Sexually Transmitted Infection Testing Compared With General Practitioners' Advice: Vignette-Based Qualitative Study With Interviews Among General Practitioners

Kyma Schnoor[1,2], MSc; Anke Versluis[1,2], PhD; Niels H Chavannes[1,2], MD, PhD; Esther P W A Talboom-Kamp[1,2,3], MBA, MD, PhD

[1]Public Health and Primary Care, Leiden University Medical Center, Leiden, Netherlands

[2]National eHealth Living Lab, Leiden University Medical Center, Leiden, Netherlands

[3]Zuyderland, Sittard-Geleen, Netherlands

Corresponding Author:
Kyma Schnoor, MSc
Public Health and Primary Care
Leiden University Medical Center
Hippocratespad 21
Leiden, 2333 ZA
Netherlands
Phone: 31 71526 8433
Email: k.schnoor@lumc.nl

Abstract

Background: Digital triage tools for sexually transmitted infection (STI) testing can potentially be used as a substitute for the triage that general practitioners (GPs) perform to lower their work pressure. The studied tool is based on medical guidelines. The same guidelines support GPs' decision-making process. However, research has shown that GPs make decisions from a holistic perspective and, therefore, do not always adhere to those guidelines. To have a high-quality digital triage tool that results in an efficient care process, it is important to learn more about GPs' decision-making process.

Objective: The first objective was to identify whether the advice of the studied digital triage tool aligned with GPs' daily medical practice. The second objective was to learn which factors influence GPs' decisions regarding referral for diagnostic testing. In addition, this study provides insights into GPs' decision-making process.

Methods: A qualitative vignette-based study using semistructured interviews was conducted. In total, 6 vignettes representing patient cases were discussed with the participants (GPs). The participants needed to think aloud whether they would advise an STI test for the patient and why. A thematic analysis was conducted on the transcripts of the interviews. The vignette patient cases were also passed through the digital triage tool, resulting in advice to test or not for an STI. A comparison was made between the advice of the tool and that of the participants.

Results: In total, 10 interviews were conducted. Participants (GPs) had a mean age of 48.30 (SD 11.88) years. For 3 vignettes, the advice of the digital triage tool and of all participants was the same. In those vignettes, the patients' risk factors were sufficiently clear for the participants to advise the same as the digital tool. For 3 vignettes, the advice of the digital tool differed from that of the participants. Patient-related factors that influenced the participants' decision-making process were the patient's anxiety, young age, and willingness to be tested. Participants would test at a lower threshold than the triage tool because of those factors. Sometimes, participants wanted more information than was provided in the vignette or would like to conduct a physical examination. These elements were not part of the digital triage tool.

Conclusions: The advice to conduct a diagnostic STI test differed between a digital triage tool and GPs. The digital triage tool considered only medical guidelines, whereas GPs were open to discussion reasoning from a holistic perspective. The GPs' decision-making process was influenced by patients' anxiety, willingness to be tested, and age. On the basis of these results, we believe that the digital triage tool for STI testing could support GPs and even replace consultations in the future. Further research must substantiate how this can be done safely.

(JMIR Hum Factors 2024;11:e49221) doi:10.2196/49221

KEYWORDS

eHealth; digital triage tool; sexually transmitted infection; STI; human immunodeficiency virus; general practitioners; GPs decision-making; digital health; diagnostic; sexually transmitted disease; STD; sexually transmitted; sexual transmission; triage; artificial intelligence; HIV; diagnostics; diagnosis; vignette; vignettes; interview; interviews; best practice; best practices; thematic analysis; referral; medical advice

Introduction

Background

The use of eHealth, health services delivered through the internet or related technologies, is increasing, especially since the COVID-19 pandemic [1,2]. The COVID-19 pandemic has shed light on the crucial role of digitization in health care [2]. An important and promising element of digitization in health care are digital triage tools consisting of a questionnaire for patients to identify the risk of a medical problem. These tools use a digital questionnaire typically administered by a health care professional, and an algorithm based on a medical decision tree generates automatic advice for follow-up, for example, a web-based symptom checker. In this paper, we discuss a digital triage tool that advises whether a specific diagnostic test for a specific combination of symptoms is necessary. This specific digital triage tool is based on Dutch medical guidelines.

Such a digital triage tool for different problems and symptoms could be an efficient and accessible method for citizens with medical questions. In addition, this digital triage tool could possibly lower the workload of general practitioners (GPs) as it can replace the triage that health care professionals would do themselves [3]. However, it is important that triage leads to responsible and appropriate care given the situation. Digital triage tools should not result in "over-triage" or "under-triage" [4]. Overtriage is when a patient is advised to undergo a medical treatment or diagnostic test when they do not have an (urgent) medical problem [4]. Undertriage is when a patient is told that they do not have an (urgent) medical problem when they do, with the advice that a diagnostic test or medical treatment is not necessary [4]. It is important to know whether the digital triage tool for diagnostic tests is in line with daily medical practice to maximize its validity.

In daily practice at GPs' offices, medical guidelines are used to support their decision-making. GPs following guidelines has been an important research subject into the decision-making process of GPs in dermatology has shown that GPs do not always adhere to medical guidelines [5]. For example, concerns about the patient or the relationship between the GP and the patient were sometimes part of the decision-making process [5]. Furthermore, a meta-synthesis of qualitative studies identified GPs' attitudes toward and experiences with clinical guidelines [6]. First, this study showed that GPs experience tension between their own experiences and the guidelines they must adhere to as guidelines do not consider personal circumstances. Second, GPs are afraid of missing a patient diagnosis. Third, GPs experience that the guidelines do not always fit with patients' needs, and therefore, GPs act differently from what the guidelines instruct them to do. Earlier reviews have revealed other factors that play a role in the decision-making process of GPs in referrals for diagnostic tests [7-9]. These are, among others, demographic and nonclinical factors such as patient characteristics (eg, age, sex, and social class [8]). In addition, the patient's quality of life and wishes are nonclinical factors that influence the decision-making process of the GP [7]. Not all those factors are included in medical guidelines and, consequently, in digital triage. All these factors clearly show that the GP makes decisions from a holistic perspective, which makes it even more interesting and important to critically consider decision-making using digital tools from the perspective of the GP. Regarding diagnostic testing, to our knowledge, our study is the first one that compares the advice of GPs with that of a web-based tool. At the same time, this study identifies what factors influence a GP's decision-making process for a diagnostic test.

Objectives

If a digital triage tool is of high quality and the patient is adequately advised, a consultation with the GP could be avoided, resulting in an efficient care process for the patient. The GP can also be supported in the hectic daily workload as the patient uses the tool independently [9]. The first objective of this study was to identify whether the advice of the studied digital triage tool aligned with the daily medical practice of the GP. The second objective was to learn which factors influenced the GP's decision regarding a referral for diagnostic testing. In addition, this research provides insights into the GP's decision-making process and whether factors are possibly missing from a digital triage tool. As a starting point, we investigated these research questions for sexually transmitted infection (STI) triage as the medical guidelines are straightforward (eg, clear risk factors and answer categories). Much research has been conducted on digital applications for STI testing, such as websites in which tests can be ordered, with positive feedback from patients about their usability [10]. Moreover, research has shown that a digital triage tool can potentially lower the threshold for STI testing [10] as this problem can be associated with feelings of shame [11]. To answer the research questions, a vignette-based qualitative study was conducted based on different STI-related patient cases [12].

Methods

Study Design and Participants

A qualitative vignette study was conducted using semistructured interviews with GPs as participants. Data saturation was expected after 10 interviews [13]. There were no specific exclusion criteria. GPs in training, practicing, or retired (for ≤5 y) could participate. In the interviews, the participants were presented with different patient vignettes (see the *Materials* section for details). After each vignette, the participants were asked about their clinical decision regarding STI diagnostic testing and to describe their thinking and decision-making process. This approach is called the "Think Aloud" method, which allows for a description of how information is structured during a problem-solving task [14]. In addition, it provides rich data for analysis [15].

Ethical Considerations

This study was declared not to fall within the scope of the Dutch Medical Research Involving Human Subjects Act by the departmental ethics committee of the Leiden University Medical Center (reference 22-3002).

Materials

A vignette is a short hypothetical description of a patient representing a standardized combination of specific characteristics [16]. Vignettes made it possible to present patients with the same characteristics to every participant (eg, complaints, relationship status, and age) and, in this way, minimize variations between patients, which is not possible in real life. In this study, the vignettes were based on different aspects of the Dutch medical guidelines for STI testing [17]. In the medical guidelines, different aspects are taken into account to calculate the risk of an STI, such as endemic areas, unsafe sex, and different complaints. The following factors were incorporated into the vignettes: age, gender, sexuality, relationship status, employment (eg, full-time job or student), history of unsafe sex and how long ago it took place, number of sexual partners, frequency of unsafe sex, frequent GP visits, symptoms, and ethnicity. Some of these factors are not in the guidelines but were included to research whether they influenced the decision-making process of the GP (eg, situation and if the GP was visited often by that patient). In addition, the vignettes were designed in such a way that they would lead to advice from participants to undergo a diagnostic test for STIs or not. In total, 6 different vignettes were created and used (Multimedia Appendix 1). In Textbox 1, a short description of the vignettes is provided. The Dutch vignettes were designed with a GP and checked by another GP. An example of a translated vignette can be found in Textbox 2.

Textbox 1. Short description of the vignettes.

Vignette 1

- Woman, aged 20 years, from Spain, student, had unsafe sex multiple times >3 weeks ago, itching of the vagina, does not visit her general practitioner (GP) often

Vignette 2

- Man, aged 26 years, plumber, steady relationship, has irritation at the urethra and sensitivity when urinating, visits GP often

Vignette 3

- Woman, aged 17 years, high school student, had unsafe sex <3 weeks ago with no complaints, the first time she comes to the practice

Vignette 4

- Man, aged 24 years, has a relationship with a man, his partner has sexual contact with other men, has difficulty urinating

Vignette 5

- Woman, aged 45 years, has a steady relationship but thinks her partner cheated 6 months ago, has contact bleeding, visits the GP often

Vignette 6

- Woman, aged 35 years, has a steady relationship, comes from Surinam, has a burning sensation when urinating, visits her GP often

Textbox 2. Vignette 1 translated from Dutch to English.

- Mrs A is aged 20 years and studies in the Netherlands but comes from Spain originally. She has not visited you at the practice often. She is not in a committed relationship and has had unprotected sex several times in the past 6 months for more than 3 weeks. She experiences vaginal discharge and itching and irritation in her vagina. She wonders whether she might have a sexually transmitted infection.

Procedure

Participants were recruited via a LinkedIn post that included the email address of the researcher. Interested participants were instructed to send an email if they wanted to take part. In addition, participants were emailed from the network of the researchers, and the GPs could reply to the email if they wanted to participate. Interested participants were sent information and the informed consent form. In addition, different data and time points were included in the interviews, which could be face-to-face or digital (based on the preference of the participant). Participants had the right to withdraw at any time.

An interview protocol guided the semistructured interviews (Multimedia Appendix 2). All interviews were audio recorded. Each interview started with a short explanation of the study. The first vignette was then read out loud to the participant. They were asked whether they would advise undergoing diagnostic tests for STIs. Next, they were asked to share their reasoning process. These 2 steps were repeated for each vignette (ie, 6 in total). The first interviews were conducted with both interviewers present (KS and Fleur Rekveld), and KS was the lead. The other interviews were conducted by KS, Fleur Rekveld, or both.

Service: Digital Triage Tool

The digital triage tool was developed by a Dutch diagnostic center [18] based on a decision tree with Dutch medical guidelines [17]. The digital triage tool was developed in cocreation with GPs and clinical chemists. A Dutch academic knowledge center assessed the digital triage [19]. During triage, users first go through a series of questions. Their answers determine what question they have to answer next and, in the end, what advice is given. For example, the first question is "Did you have unsafe sex?" If the answer is "no," the advice is not to be tested. If the answer is "yes," a follow-up question appears: *what is your gender?* Gender is asked about as differences in gender result in different advice (eg, for women users who are advised to undergo a chlamydia test, it means that the service could advise doing a vaginal swab). Ultimately, the digital triage tool advises whether a diagnostic test for STIs is necessary and, if yes, which one (eg, chlamydia, gonorrhea, or HIV). The digital triage tool is now used in 2 digital services of the diagnostic company where patients can order diagnostic tests themselves with or without a health care professional. These diagnostic services are Directlab, where users can order web-based diagnostic test packages independent of a health care professional, and Homelab, where patients in the digital environment of their GP can order diagnostic test packages. In regular daily practice in the Netherlands, the patient needs to ask for a consultation with the GP (on the phone or in person) and ask for a diagnostic test for STIs. In this situation, the GP performs triage to identify whether it is necessary to conduct an STI test.

Data Analysis

To determine the diagnostic test advice of the digital triage tool, the characteristics of each vignette were entered into it. The ensuing advice was compared with the test advice of the GPs per vignette. To learn which factors influenced the GPs' decision-making process, the combination of the think-aloud process, vignettes, and semistructured interviews was used as a triangulation method to obtain a complete range of data to result in a strong conclusion [12,20]. All interviews were transcribed (intelligent) verbatim. When the transcripts were completed and uploaded to ATLAS.ti (version 22; ATLAS.ti Scientific Software Development GmbH), the audio recordings were deleted. In total, 2 authors (Fleur Rekveld and KS) conducted the qualitative data analysis according to the principles of thematic analysis. Fleur Rekveld and KS developed a preliminary coding scheme based on the coded data from the first 8 participants. The final coding scheme emerged after all the coding was performed by the 2 authors independently. The codes were grouped into themes and subthemes.

Results

Characteristics of the Study Population

Data saturation was reached after 10 interviews. The characteristics of the participants are presented in Table 1. Their ages ranged from 32 to 70 years, with a mean of 48.30 (SD 11.88) years. The number of men and women was almost equal (6/10, 60% and 4/10, 40%, respectively). Of the 10 GPs, 1 (10%) was retired, 3 (30%) were working part time as GPs, and 6 (60%) were working full time.

Table 1. Characteristics of the participants.

Participant	Age (y)	Gender	Employment status
1	32	Woman	Part time
2	55	Man	Full time
3	38	Man	Part time
4	59	Man	Full time
5	70	Man	Retired
6	53	Man	Full time
7	55	Woman	Full time
8	43	Man	Full time
9	38	Woman	Part time
10	40	Woman	Full time

Testing Advice of Digital Triage Tool Versus GPs

Table 2 shows, for each vignette, whether the digital tool would advise conducting an STI test and what each participant would advise to do. For 50% (3/6) of the vignettes (ie, numbers 1, 4, and 5), the digital triage tool's advice aligned with all participants' advice. For all 3 vignettes, the advice was to conduct a diagnostic test for STIs. For those 3 vignettes, the patients' risk factors were sufficiently clear for the participants to advise to conduct a test.

In vignette 1, the most important decision-making factor was the patient's age; young age combined with women was an important factor influencing the participants' test advice as having an STI could make this woman infertile. Participant 7 answered the following:

> *I would test her, always with women of her age who are sexually active.*

In addition, unsafe sex was an important factor in the decision to test.

For vignette 4, the main factor in advising to test was the “men having sex with men” risk factor. Participant 5 answered the following:

> *It is male-male contact, and in addition, there are changes in sexual contacts so that he can do an STI test.*

For vignette 5, all participants would advise conducting an STI test as well. Furthermore, 80% (8/10) mentioned that they would also conduct cervical cancer diagnostic tests because of the symptom of contact bleeding. Participant 9 mentioned the following:

> *In the case of contact bleeding, more research than only an STI is needed. It could be Chlamydia, but a smear test is needed to exclude cervical cancer.*

For the other 50% (3/6) of the vignettes, not all participants gave the same advice as each other or as the digital triage tool. For vignette 2, a total of 60% (6/10) of the participants agreed with the advice of the digital tool, and for vignettes 3 and 6, the proportions were 70% (7/10) and 80% (8/10), respectively. It is important to mention that the initial answer of the participants is presented in Table 2. It could be the case that participants answered “no” to advising an STI test for the patient initially. However, the participants mentioned that they would advise conducting an STI test after excluding other diseases. In addition, sometimes, the participants wanted more information about the patient’s situation before advising to conduct an STI test.

For vignette 2, most participants wanted to know more about the patient’s case before giving the advice to test for an STI. In addition, they wanted to conduct a physical examination or other tests, such as a test to exclude urinary infection, as the patient’s symptoms seemed not totally compliant with those of an STI. Participant 2 said the following:

> *I would like to know a little more; why does he think he has an STI? Does he have other contacts next to his current relationship or an open relationship? Has he heard anything from his wife?*

Participant 4 answered the following:

> *I would check his urine.*

Participants answered that the symptoms and risk factors were too unclear to advise an STI test. A minority of the participants would test for an STI to exclude it or to satisfy the patient’s request. Participant 2 answered the following:

> *He asked for an STI test so I would do one.*

The participants mentioned that, sometimes, a patient does not have an apparent reason for wanting to take an STI test or the patient has no symptoms that fit with those of an STI. However, sometimes patients do not want to discuss this in detail, and participants found it important to allow for testing at a low threshold if patients asked for it themselves. Participant 9 mentioned the following:

> *Maybe he (or his wife) is cheating, and they do not want to tell you that directly...It is always the question if the patient is honest with you, so I would test at a low threshold after I did a urine infection test, and then I think he would accept that.*

Table 2. Advice of the digital tool and the participants to test for a sexually transmitted infection.

	Digital triage tool	P[a]1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Agreement, n (%)[b]
Vignette 1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10 (100)
Vignette 2	No	No	Yes	No	No	Yes	No	No	Yes	No	Yes	6 (60)
Vignette 3	Later	Later	Later	Later	Later	No	Yes	Yes	Later	Later	Later	7 (70)
Vignette 4	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10 (100)
Vignette 5	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10 (100)
Vignette 6	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	8 (80)

[a]P: participant.

[b]Percentage of participants who agreed with the advice of the digital triage tool.

For vignette 3, most participants (7/10, 70%) answered that the patient could take an STI diagnostic test but at a later time. At this time, it was too early to detect an STI. A total of 20% (2/10) of the participants also mentioned that they would talk to the patient about her contraception and provide education about safe sex. Participant 2 said the following:

> *She had unsafe sex, so I would do two things. Maybe check if she uses birth control, and I would tell her that she can do an STI test after two weeks.*

Vignette 6 involved a patient from an endemic area. In total, 25% (2/8) of the participants who agreed with the advice of the digital tool mentioned the endemic area as a reason for testing. Participant 10 mentioned the following:

> *I would ask her some more questions; however, she is from Surinam, a risk area. So I would test her at a low threshold, especially for a serological test.*

The other 62% (6/8) of the participants mentioned low-threshold testing because of the patient’s symptoms. Most participants (6/10, 60%) mentioned that they would check for a urinary infection, some before conducting an STI test and others in addition to it. Participant 1 mentioned the following:

> *I would check her urine first to ensure she has no urinary infection.*

It is important to note that almost all participants mentioned that, if a patient requested an STI test, they would meet the request. They also mentioned that, in some cases, they would also give patients more information about safe sex or conduct a physical examination. The decision to do so often depended on age or other risk factors such as contact bleeding. Especially in the case of younger patients, GPs educated them about safe sex and birth control. However, this information provision was not part of their decision-making process but rather of their consultation.

Extra Factors That Influenced the Decision of the GPs

There were several factors that the participants considered in their decision that were not included in the digital triage tool. The most important additional patient-related factors were anxiety about infection, the wishes of the patient, and age. Among all participants (10/10, 100%), the patient's anxiety was an additional reason for referring them to an STI test. The participants reasoned that a request for an STI test is not made easily and that there may be an unknown reason behind it. In their opinion, when patients experience fear-related stress, it might harm their health. Participant 10 mentioned the following:

> *Sometimes you feel that there is more than they want to say, and then you decide to test at a low threshold.*

Age played a role in the decision-making process of the GPs. This was especially the case in vignettes 1 and 3. The GPs mentioned that checking for STIs was important at a fertile age, especially for women. In the Dutch medical guidelines, it is noted that, below the age of 25 years, there needs to be a low threshold for STI testing even if patients report no complaints. Participant 6 answered the following in the interview about vignette 3:

> *Especially in younger patients, you want to know what they know about sex and the transmission of STIs.*

In 2 vignettes, the GPs felt the need to ask additional questions or conduct a physical examination. The digital triage tool only provides advice on an STI test. However, the symptoms may also indicate a urinary tract infection or a stage of cervical cancer. These tests are not advised via the digital tool but were advised by the participants in this study for those 2 vignettes.

One GP also considered who had to pay for the test and whether it was affordable. Participant 3 mentioned the role of the payer or possible reimbursement in the decision. He answered the following about vignette 6:

> *If she wants to pay for a test and she wants to do a test...Then, she can do a test.*

In summary, it can be generally said that GPs in this study paid extra attention to patient-related factors such as fear of infection, desire to undergo the test, and young age when deciding whether to request an STI test.

Discussion

Principal Findings

In this study, we tried to identify whether the advice of a digital triage tool based on medical guidelines aligned with GPs' medical practice. The results showed that other factors, which are not part of the guidelines, played a role in the GPs' decision-making process when determining whether to advise an STI test for a patient. The most important additional patient-related factors were the patient's anxiety, wishes, and age. The GPs also considered who had to pay for the test and whether it was affordable. Finally, the GPs were willing in some vignettes to ask additional questions or conduct a physical examination. The most notable factors are discussed in this section and compared with the literature.

In line with other research, the GPs' decision to test depends sometimes on the anxiety and wishes of the patient [7]; these factors were not included in the studied digital triage tool. This additional aspect aligns with the research by Hajjaj et al [5,7]. In addition, our results align with those of a study that researched the barriers to following guidelines among GPs [6] that showed that the patient's preferences were considered more important than following guidelines.

The interviews showed that the age of the patients was an important factor that influenced the GPs' advice. Specifically, younger age was an important reason to advise an STI test because of the risk of infertility and the sexual activity in this group. Age was not included as a factor in the digital triage tool. As STIs mainly occur under the age of 30 years, it is not surprising that GPs tend to advise testing more for patients in this age group [21].

From the literature, it was found that the factor "knowing the patient" influences the decision-making process of GPs [22]. Accumulated knowledge about the patient influences the context and interpretation of the conversation between the patient and the health care professional, especially in the case of psychosocial or unspecific problems such as fatigue. However, in this study, knowing the patient was not a factor that was considered in the vignettes. For this reason, the decisions that the GPs made in this study could be different in real life as they might know the patients.

In addition to patient-related factors (eg, the wishes of the patient), GP-related factors also influenced the decision-making process. The extent to which GPs were open to discussion with patients about why they wanted an STI test or to which GPs were willing to address patients' concerns influenced the decision. In addition, based on the findings of this study, it seems that the GPs expressed a preference for obtaining a complete set of information before deciding. For example, some GPs wanted to have more information about the situation of the patients and their partners. In some cases, GPs wanted to conduct a physical examination or other diagnostic tests (eg, urinary infection) to exclude other diseases. The digital triage tool is strictly bound to the guidelines set up without paying attention to, for example, the anxiety of the patient or the need for additional information. Other guidelines have been developed for possible symptoms of urinary tract infection or cervical problems, which have not yet been combined on the internet.

The advice of the digital triage tool is straightforward and always in line with a strict algorithm. In this study, GPs were found to recommend a diagnostic test for STIs more often than the digital tool. In the Netherlands, a study showed that unnecessary diagnostics (overdiagnostics) are a common problem among Dutch GPs; slightly more than half of the participating GPs indicated that patients could submit a complaint for not requesting an examination that was indicated and that this played a role to some or a significant extent in the request for diagnostic testing [23].

Our study did not investigate whether the digital tool can prevent overdiagnostics, but we assume that it can be a powerful decision support tool for daily general practice, just as tools for pharmacotherapy are already in use. More research is needed to confirm this.

Another possible reason why GPs are more inclined to test seems to be that it could save them time [24]. For example, if a patient has vague symptoms, it would be easy to request some tests first without having a thorough conversation. Another possible reason specifically for low-threshold STI testing could be feelings of embarrassment to ask about sexual behavior [25]. Recently, a Dutch center for sexual health found that talking about sexual behavior is not done as often as it should by health care professionals [26]. This could be seen as an additional justification for supporting GPs with digital tools for STI testing.

This study does not suggest that digital triage is the holy grail to prevent overdiagnostics or that it is *the* solution to lower the work pressure of GPs. However, this vignette study confirms that GPs have a more holistic approach to their patients compared with a digital triage tool. A digital triage tool primarily relies on specific responses to predefined questions, whereas a GP can consider more factors such as social factors, lifestyle, and personal context. On the one hand, the comprehensive perspective of GPs might result in a higher frequency of diagnostics when compared with a digital triage tool. This is due to the GPs considering additional factors. Given the high workload and time constraints of GPs, the investigated digital tool can play a helpful role in daily decision-making. In contrast, this holistic approach by GPs could potentially lead to fewer diagnostics. Given their deep understanding of the patients' condition, GPs are better positioned to assess the necessity of tests.

This study has several limitations. It could be that social desirability influenced the GPs' answers on the vignettes and interviews. Potentially, the advice of the GPs was more in line with the guidelines compared with that in their daily practice as they were aware of the fact that they were part of research on this topic [12]. It is also worth mentioning that there could be a disparity between what people think they would do in a particular situation and their actual behavior [27]. In addition, this study is not generalizable to the entire field of diagnostics at general practices because of its focus on STI testing. As a starting point, this study identified factors that influenced the decision-making process of GPs for STI testing. In future research, we recommend investigating digital tools and the decision-making process of GPs for other common diagnostic tests.

A strength of this study is the combination of the vignette method, the think-aloud process, and the semistructured interviews, which aimed to obtain a complete range of data on the topic (triangulation). Although no actual patients were included in this study, we aimed to make the vignettes as valid as possible by developing and testing them with GPs. In addition, providing the same vignettes to different GPs made it easier to compare patients within different general practices instead of comparing real-life patients with different complaints and characteristics. Currently, we are working on a real-life study in which patients in the waiting room of a GP's office complete digital triage for STI testing (the result of the digital triage tool is not shown to the patient), after which they go on to have their planned consultation with the GP. At this consultation, the GP will also advise whether to test for an STI; the advice of the digital tool and of the GP will be compared. We expect more detailed and practical information to further refine this working method using a digital tool.

A qualitative study in which GPs were interviewed about their general attitude toward the use of digital tools by patients in their practice showed that GPs' attitudes toward digital STI diagnostic services were positive, and they acknowledged that the use of eHealth in their practice could result in a more efficient workflow [28].

It will be interesting to further investigate whether GPs are also willing to use digital triage tools as a standard gateway for their practice for some diagnostic tests. When a digital triage tool is implemented and integrated into the care pathway, it is important to investigate what users think of this integration and whether they are satisfied with this change in their way of working. For future research, it could be beneficial to make a comparison of the experiences of patients with a digital triage tool, triage at the GP's office, and a mix. Notably, recent studies on digital chatbots for medical questions have shown that patients perceived the chatbot's responses to be superior to those provided by GPs [29]. For future applications, it is essential to consider patients' eHealth literacy before using a digital triage tool as the primary tool in daily general practice [30,31]; hybrid care might be a solution to address all types of patients. Finally, it is important to realize that the tool in the care pathway needs to stay up-to-date and needs to be changed when the medical guidelines are updated [32]. This study showed that (holistic) factors that are not part of the digital triage tool affect GPs' decision-making. This is an interesting topic for future research as digital tools and artificial intelligence are increasingly being used in health care. Nowadays, GPs use digital medication prescription tools to support their decision-making, which could help with handwriting errors but also with poor treatment decisions [33]. Another example is an artificial intelligence system that could help GPs decide on the early detection of skin cancer [34,35]. Digital technologies such as these should be researched carefully to see what the impact and consequences are for both GPs and patients.

Conclusions

This study shows that, in some cases, patients receive different advice to undergo an STI test from a digital tool and from a GP. Other factors that are not part of medical guidelines play a role in the GPs' decision-making process when deciding whether to request an STI test. The most important additional patient-related factors were the patient's anxiety, wishes, and age. One GP also considered who had to pay for the test and whether it was affordable. Finally, some GPs expressed a desire to ask additional questions or conduct a physical examination in certain vignettes. In comparison, the digital triage tool adhered more closely to the medical guidelines, with GPs being more inclined than the digital tool to recommend an STI test for the same patient case. Alignment between the digital tool and GP advice only occurred when the risk factors for STI testing were unequivocally evident. This confirms that GPs decide from a holistic perspective. On the basis of these initial findings, we cautiously posit that a digital triage tool for STI testing can potentially support GPs and may even serve as a substitute for in-person consultations in the future. However, it is imperative to conduct further research to establish safe and effective methods for implementing such a transition.

These conclusions should be approached carefully, recognizing that this study represents an initial exploration and that additional research is required to substantiate and refine these findings.

Acknowledgments

The authors would like to thank all general practitioners who participated in this study. They appreciate the time they invested in this research. In addition, the authors want to thank Fleur Rekveld for her contribution to this paper. Funding was provided by Unilabs. Unilabs played no role in the research design or data collection, analysis, or interpretation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Translated vignettes from Dutch to English.
[DOCX File , 13 KB - humanfactors_v11i1e49221_app1.docx]

Multimedia Appendix 2
Semistructured interview protocol.
[DOCX File , 16 KB - humanfactors_v11i1e49221_app2.docx]

References

1. Eysenbach G. What is e-health? J Med Internet Res 2001 Jun 18;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
2. Fang ML, Walker M, Wong KL, Sixsmith J, Remund L, Sixsmith A. Future of digital health and community care: exploring intended positive impacts and unintended negative consequences of COVID-19. Healthc Manage Forum 2022 Sep 01;35(5):279-285 [FREE Full text] [doi: 10.1177/08404704221107362] [Medline: 35775162]
3. Gottliebsen K, Petersson G. Limited evidence of benefits of patient operated intelligent primary care triage tools: findings of a literature review. BMJ Health Care Inform 2020 May 07;27(1):e100114 [FREE Full text] [doi: 10.1136/bmjhci-2019-100114] [Medline: 32385041]
4. Schmieding ML, Kopka M, Schmidt K, Schulz-Niethammer S, Balzer F, Feufel MA. Triage accuracy of symptom checker apps: 5-year follow-up evaluation. J Med Internet Res 2022 May 10;24(5):e31810 [FREE Full text] [doi: 10.2196/31810] [Medline: 35536633]
5. Hajjaj FM, Salek MS, Basra MK, Finlay AY. Nonclinical influences, beyond diagnosis and severity, on clinical decision making in dermatology: understanding the gap between guidelines and practice. Br J Dermatol 2010 Oct;163(4):789-799. [doi: 10.1111/j.1365-2133.2010.09868.x] [Medline: 20854402]
6. Carlsen B, Glenton C, Pope C. Thou shalt versus thou shalt not: a meta-synthesis of GPs' attitudes to clinical practice guidelines. Br J Gen Pract 2007 Dec 01;57(545):971-978 [FREE Full text] [doi: 10.3399/096016407782604820] [Medline: 18252073]
7. Hajjaj FM, Salek MS, Basra MK, Finlay AY. Non-clinical influences on clinical decision-making: a major challenge to evidence-based practice. J R Soc Med 2010 May 30;103(5):178-187 [FREE Full text] [doi: 10.1258/jrsm.2010.100104] [Medline: 20436026]
8. O'Donnell CA. Variation in GP referral rates: what can we learn from the literature? Fam Pract 2000 Dec;17(6):462-471. [doi: 10.1093/fampra/17.6.462] [Medline: 11120716]
9. Duddy C, Wong G. Efficiency over thoroughness in laboratory testing decision making in primary care: findings from a realist review. BJGP Open 2021 Apr;5(2):bjgpopen20X101146 [FREE Full text] [doi: 10.3399/bjgpopen20X101146] [Medline: 33293413]
10. Versluis A, Schnoor K, Chavannes NH, Talboom-Kamp EP. Direct access for patients to diagnostic testing and results using eHealth: systematic review on ehealth and diagnostics. J Med Internet Res 2022 Jan 12;24(1):e29303 [FREE Full text] [doi: 10.2196/29303] [Medline: 35019848]

11. Balfe M, Brugha R. Disclosure of STI testing activities by young adults: the influence of emotions and social networks. Sociol Health Illn 2010 Nov 11;32(7):1041-1058 [FREE Full text] [doi: 10.1111/j.1467-9566.2010.01281.x] [Medline: 20937054]
12. Daemers DO, van Limbeek EB, Wijnen HA, Nieuwenhuijze MJ, de Vries RG. Factors influencing the clinical decision-making of midwives: a qualitative study. BMC Pregnancy Childbirth 2017 Oct 06;17(1):345 [FREE Full text] [doi: 10.1186/s12884-017-1511-5] [Medline: 28985725]
13. Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: a systematic review of empirical tests. Soc Sci Med 2022 Jan;292:114523 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114523] [Medline: 34785096]
14. Fonteyn ME, Kuipers B, Grobe SJ. A description of think aloud method and protocol analysis. Qual Health Res 2016 Jul 01;3(4):430-441 [FREE Full text] [doi: 10.1177/104973239300300403]
15. Lundgrén-Laine H, Salanterä S. Think-aloud technique and protocol analysis in clinical decision-making research. Qual Health Res 2010 Apr 03;20(4):565-575. [doi: 10.1177/1049732309354278] [Medline: 19959822]
16. Atzmüller C, Steiner PM. Experimental vignette studies in survey research. Methodol 2010 Jan;6(3):128-138 [FREE Full text] [doi: 10.1027/1614-2241/a000014]
17. Van Bergen J, Dekker J, Boeke A, Kronenberg E, Van der Spruit R, Burgers J. NHG-Standaard Het soa-consult (eerste herziening). Huisarts Wet 2013;56(9):450-463. [doi: 10.1007/978-90-313-8279-8_88]
18. Home page. Saltro. URL: https://saltro.nl/ [accessed 2024-01-12]
19. Home page. National eHealth Living Lab. URL: https://nell.eu/ [accessed 2024-01-12]
20. Noble H, Heale R. Triangulation in research, with examples. Evid Based Nurs 2019 Jul 14;22(3):67-68. [doi: 10.1136/ebnurs-2019-103145] [Medline: 31201209]
21. Zorg V. SOA: Leeftijd & Geslacht. VZinfo. 2017. URL: https://www.vzinfo.nl/soa/leeftijd-en-geslacht/algemeen [accessed 2024-01-12]
22. Hjortdahl P. The influence of general practitioners' knowledge about their patients on the clinical decision-making process. Scand J Prim Health Care 1992 Dec 12;10(4):290-294. [doi: 10.3109/02813439209014076] [Medline: 1480869]
23. Wammes J, Verhoef L, Westert G, Assendelft P, Jeurissen P, Faber M. Onnodige zorg in de Nederlandse gezondheidszorg, gezien vanuit het perspectief van de huisarts. Celsus Academie voor Betaalbare Zorg. 2013. URL: https://betaalbaarheidvanzorg.nl/images/Publicaties/CELSUS_Werkdocument_Onnodige%20zorg%20vanuit%20perspectief%20huisarts_DEF.pdf [accessed 2024-01-12]
24. Vervloet M, Bomhoff M, Schellevis F, van Dijk L. Niet te veel en niet te weinig: de balans tussen nodige en onnodige zorg in de huisartsenpraktijk. NIVEL. 2015. URL: https://www.nivel.nl/sites/default/files/bestanden/Overzichtsstudie-niet-te-veel-en-niet-te-weinig.pdf [accessed 2024-01-12]
25. Hinchliff S, Gott M, Galena E. GPs' perceptions of the gender-related barriers to discussing sexual health in consultations--a qualitative study. Eur J Gen Pract 2004 Jun 11;10(2):56-60. [doi: 10.3109/13814780409094233] [Medline: 15232525]
26. Zuil W. Rutgers: seksualiteit onvoldoende besproken binnen de zorg. Skipr. URL: https://www.skipr.nl/nieuws/rutgers-seksualiteit-onvoldoende-besproken-binnen-de-zorg/?daily=1&utm_medium=email&utm_source=20230425+skipr+daily+nieuwsbrief-daily&utm_campaign=NB_SKIPR&tid=TIDP1993360X0AD5FC84E9664D39893DF4F2104FF76CYI4 [accessed 2024-01-12]
27. Erfanian F, Roudsari RL, Heydari A, Bahmani MN. A narrative on using vignettes: its advantages and drawbacks. J Midwifery Reproductive Health 2020;8(2):2134-2145 [FREE Full text] [doi: 10.22038/JMRH.2020.41650.1472]
28. de Wilt T, Versluis A, Goedhart A, Talboom-Kamp E, van Delft S. General practitioners attitude towards the use of eHealth and online testing in primary care. Clin eHealth 2020;3:16-22 [FREE Full text] [doi: 10.1016/j.ceh.2020.02.002]
29. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 2023 Jun 01;183(6):589-596. [doi: 10.1001/jamainternmed.2023.1838] [Medline: 37115527]
30. van der Kleij RM, Kasteleyn MJ, Meijer E, Bonten TN, Houwink EJ, Teichert M, et al. SERIES: eHealth in primary care. Part 1: concepts, conditions and challenges. Eur J Gen Pract 2019 Oct 10;25(4):179-189 [FREE Full text] [doi: 10.1080/13814788.2019.1658190] [Medline: 31597502]
31. Heijsters F, van Loon G, Santema J, Mullender M, Bouman M, de Bruijne M, et al. A usability evaluation of the perceived user friendliness, accessibility, and inclusiveness of a personalized digital care pathway tool. Int J Med Inform 2023 Jul;175:105070 [FREE Full text] [doi: 10.1016/j.ijmedinf.2023.105070] [Medline: 37121138]
32. Stahl L, Spatz M. Quality assurance in eHealth for consumers. J Consum Health Internet 2003 Jan;7(1):33-42. [doi: 10.1300/j381v07n01_03]
33. Armando LG, Miglio G, de Cosmo P, Cena C. Clinical decision support systems to improve drug prescription and therapy optimisation in clinical practice: a scoping review. BMJ Health Care Inform 2023 May 02;30(1):e100683 [FREE Full text] [doi: 10.1136/bmjhci-2022-100683] [Medline: 37130626]
34. Micocci M, Borsci S, Thakerar V, Walne S, Manshadi Y, Edridge F, et al. Attitudes towards trusting artificial intelligence insights and factors to prevent the passive adherence of GPs: a pilot study. J Clin Med 2021 Jul 14;10(14):3101 [FREE Full text] [doi: 10.3390/jcm10143101] [Medline: 34300267]

35. Sangers TE, Wakkee M, Moolenburgh FJ, Nijsten T, Lugtenberg M. Towards successful implementation of artificial intelligence in skin cancer care: a qualitative study exploring the views of dermatologists and general practitioners. Arch Dermatol Res 2023 Jul 07;315(5):1187-1195 [FREE Full text] [doi: 10.1007/s00403-022-02492-3] [Medline: 36477587]

Abbreviations

GP: general practitioner
STI: sexually transmitted infection

Edited by A Kushniruk; submitted 30.05.23; peer-reviewed by S Hinder; comments to author 30.06.23; revised version received 05.07.23; accepted 20.11.23; published 22.01.24.

Please cite as:
Schnoor K, Versluis A, Chavannes NH, Talboom-Kamp EPWA
Digital Triage Tools for Sexually Transmitted Infection Testing Compared With General Practitioners' Advice: Vignette-Based Qualitative Study With Interviews Among General Practitioners
JMIR Hum Factors 2024;11:e49221
URL: https://humanfactors.jmir.org/2024/1/e49221
doi:10.2196/49221
PMID:38252474

Original Paper

Evaluating Users' Experiences of a Child Multimodal Wearable Device: Mixed Methods Approach

Nancy L McElwain[1,2], DPhil; Meghan C Fisher[1], DPhil; Camille Nebeker[3], MS, EdD; Jordan M Bodway[1], BSc; Bashima Islam[4], DPhil; Mark Hasegawa-Johnson[2,5], DPhil

[1]Department of Human Development and Family Studies, University of Illinois Urbana-Champaign, Urbana, IL, United States

[2]Beckman Institute for Advanced Science & Technology, University of Illinois Urbana-Champaign, Urbana, IL, United States

[3]Herbert Wertheim School of Public Health and Human Longevity Science, University of California, San Diego, La Jolla, CA, United States

[4]Department of Electrical and Computer Engineering, Worcester Polytechnic Institute, Worcester, MA, United States

[5]Department of Electrical and Computer Engineering, University of Illinois Urbana-Champaign, Urbana, IL, United States

Corresponding Author:
Nancy L McElwain, DPhil
Department of Human Development and Family Studies
University of Illinois Urbana-Champaign
2004 Doris Christopher Hall
904 West Nevada Street
Urbana, IL, 61801
United States
Phone: 1 217 244 7168
Email: mcelwn@illinois.edu

Abstract

Background: Wearable devices permit the continuous, unobtrusive collection of data from children in their natural environments and can transform our understanding of child development. Although the use of wearable devices has begun to emerge in research involving children, few studies have considered families' experiences and perspectives of participating in research of this kind.

Objective: Through a mixed methods approach, we assessed parents' and children's experiences of using a new wearable device in the home environment. The wearable device was designed specifically for use with infants and young children, and it integrates audio, electrocardiogram, and motion sensors.

Methods: In study 1, semistructured phone interviews were conducted with 42 parents of children aged 1 month to 9.5 years who completed 2 day-long recordings using the device, which the children wore on a specially designed shirt. In study 2, a total of 110 parents of children aged 2 months to 5.5 years responded to a questionnaire assessing their experience of completing 3 day-long device recordings in the home. Guided by the Digital Health Checklist, we assessed parental responses from both studies in relation to the following three key domains: (1) access and usability, (2) privacy, and (3) risks and benefits.

Results: In study 1, most parents viewed the device as easy to use and safe and remote visits as convenient. Parents' views on privacy related to the audio recordings were more varied. The use of machine learning algorithms (vs human annotators) in the analysis of the audio data, the ability to stop recordings at any time, and the view that the recordings reflected ordinary family life were some reasons cited by parents who expressed minimal, if any, privacy concerns. Varied risks and benefits were also reported, including perceived child comfort or discomfort, the need to adjust routines to accommodate the study, the understanding gained from the study procedures, and the parent's and child's enjoyment of study participation. In study 2, parents' ratings on 5 close-ended items yielded a similar pattern of findings. Compared with a "neutral" rating, parents were significantly more likely to agree that (1) device instructions were helpful and clear (t_{109}=−45.98; P<.001), (2) they felt comfortable putting the device on their child (t_{109}=−22.22; P<.001), and (3) they felt their child was safe while wearing the device (t_{109}=−34.48; P<.001). They were also less likely to worry about the audio recordings gathered by the device (t_{108}=6.14; P<.001), whereas parents' rating of the burden of the study procedures did not differ significantly from a "neutral" rating (t_{109}=−0.16; P=.87).

Conclusions: On the basis of parents' feedback, several concrete changes can be implemented to improve this new wearable platform and, ultimately, parents' and children's experiences of using child wearable devices in the home setting.

(JMIR Hum Factors 2024;11:e49316) doi:10.2196/49316

KEYWORDS

wearable devices; multimodal sensing; user experience; usability; privacy; children; mobile phone

Introduction

Background

Advances in pervasive sensing, internet of medical things, and digital health strategies more broadly [1-6] have rapidly accelerated over the past decade. Although digital health research among adults and adolescents has predominantly used smartphones [7-9], parallel work with infants and children tends to use wearable devices [10], including motion sensors to detect body posture and physical activity [11], audio recorders to assess language environment and development [12,13], heart rate sensors to assess psychophysiology [14], and head-mounted cameras to capture infants' visual perspective of the physical and social environment [15]. Such wearable technology, especially when paired with machine learning algorithms, permits the automated detection of children's behavioral and physiological states, as well as caregivers' responses, and has the potential to transform the field of child development through the collection of big data in real-world environments [16].

At the same time, the use of wearable devices among infants and young children in home environments raises unique ethical, legal, and social implications and logistical challenges. As such, careful attention to the perspectives and experiences of end users of such technology, in this case, parents and their children, is required. In this study, we assessed parents' perceptions of and experiences with a novel wearable device, LittleBeats, developed specifically for use with infants and young children. Little Beats, which is not Food and Drug Administration approved and used only for research purposes, integrates a microphone, a 3-lead electrocardiogram (ECG) sensor, and an inertial motion sensor to synchronously collect information about infant vocalizations, cardiac physiology (heart rate and respiratory sinus arrhythmia), and motion (eg, physical activity level, position, and discrete movements). The electronics are housed in a 3D-printed case (55×57×13 mm), which is placed on a specially designed shirt that the child wears. Data can be collected throughout the day at home, without the researchers present. In prior papers, we reported on machine learning algorithms used to detect and classify child and parent vocalizations using audio data from the LittleBeats device [17] and child sleep states using all 3 sensor modalities [18]. We also conducted technical validation studies to assess the signal quality of each sensor modality in relation to established laboratory protocols and gold-standard equipment [19]. Complementing these prior reports, we focus here on the critical issue of "user experience" among families and their children aged 1 month to 9.5 years. Using semistructured interviews and parent questionnaires to assess parents' experiences and perceptions, our mixed methods investigation examined usability, privacy, and perceived risks and benefits.

The "Digital Health Checklist" for Use in Child Development Research

The proliferation of digital health technologies has spurred a parallel examination of ethical practices and related decision-making processes around the use of such technologies with human participants. To evaluate the LittleBeats platform, we used the Digital Health Checklist developed by Nebeker et al [20,21]; it is grounded in the ethical principles of the *Belmont Report* [22], which speaks to beneficence, respect for persons (or autonomy), and justice, and the *Menlo Report* [23], which added the principle of respect for law and public trust. These principles form the foundation of a 4-domain framework that includes privacy, access and usability, data management (eg, collection, storage, interoperability, and sharing), and assessment of risks and benefits (Figure 1).

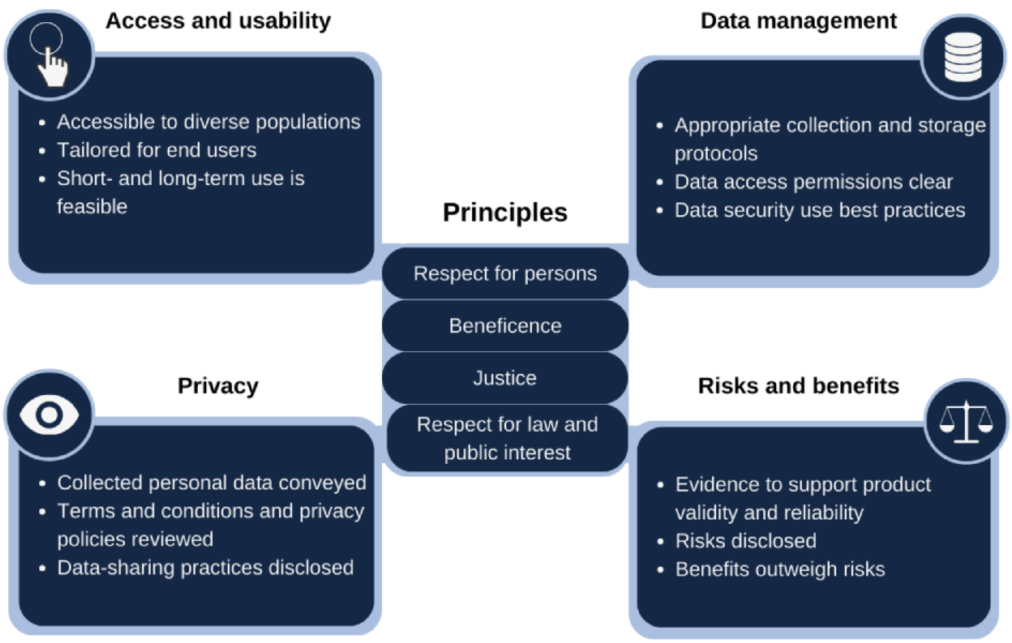
To date, the research and development of the Digital Health Checklist has been applied to digital health protocols in adult samples, including for use in cardiovascular disease prevention [24]; studies of human emotion [25]; and improvement of informed consent communications [26]. The current investigation extended the use of the Digital Health Checklist to research involving parents of infants and children. In doing so, we integrated ethical considerations specific to research with children [27]. Specifically, children are a heterogeneous group, and the potential benefits and risks to child participants need to be understood within the context of the child's age and related physical, cognitive, and socioemotional abilities.

For instance, infants and toddlers may be more susceptible to risks related to emotionally stressful procedures because their coping abilities are less well developed and depend, in part, on support from caregivers. By contrast, older children may be better able to regulate emotions and exert their autonomy, although they might be at an increased risk in other domains. For instance, owing to their growing self-awareness and other awareness, preschool- and school-aged children may be increasingly susceptible to experiencing shame and embarrassment, heightened concerns about privacy, and other related risks to the child's self-concept.

With developmental differences in risk assessment in mind, we assessed LittleBeats user experience among children representing a large age range (infancy through middle childhood). Although we did not interview children about their study experiences, we considered the children's age in our analysis of parents' open-ended responses and parents' perspectives regarding how their children felt about and responded to the research procedures. Research with children requires parental consent and, depending on the child's age, the child's assent to affirm their willingness to participate in the research. The consent process related to the LittleBeats technology has been addressed in a prior report [26]; therefore, we did not consider issues related to the provision of parental consent before participating in this research. Instead, our focus here was on parents' perceptions of and reflections on their own and their children's experiences following the use of LittleBeats at home.

Although child development research incorporating the use of wearable devices is rapidly expanding [28-32], systematic assessment of parents' perspectives and experiences (or ethical considerations more broadly) of such research has been sparse. A notable exception is a report by Levin et al [33], which outlines several key concerns parents may have about participating in research using wearable or remote sensing devices. These concerns focus on privacy expectations, particularly regarding audio or video data (considered "high fidelity data streams"), data management, and data use (eg, for scientific vs commercial purposes). Although we know of no study that assessed parents' perceptions and experiences of using wearable devices at home *after* data collection, Levin et al [33] provided valuable insights into parents' general willingness to participate in such research. Among a nationally representative sample of 210 parents (n=105, 50% mothers) with at least 1 child aged ≤5 years, 71.4% (n=150) of parents responding to hypothetical scenarios indicated at least some willingness to participate in studies involving motion or physiological sensors (low fidelity), whereas a significantly lower percentage of parents (n=99, 47.1%) endorsed willingness to participate in studies gathering audio recordings at home. It remains unknown whether the concerns expressed in the study by Levin et al [33], in which parents hypothetically considered participating in different types of remote sensing research, would also be voiced among parents who participated in research in which their children wore a wearable device with multiple sensor types (eg, motion, physiology, and audio).

Figure 1. Four-domain framework of the Digital Health Checklist for researchers. The Digital Health Checklist for researchers depicts the 4 ethical principles undergirding the 4 key domains of access and usability, privacy, risks and benefits, and data management. Source: this figure is published with permission and reflects an adaptation of the Digital Health Checklist Developed for Researchers (DHC-R) [34,35].



This Study

Guided by the domains of the Digital Health Checklist [20], we assessed parents' experiences with and perceptions of using LittleBeats at home using a mixed methods approach. In study 1, we conducted a qualitative (thematic) analysis of parental responses to a semistructured interview following the completion of 2 day-long LittleBeats recordings at home; children in this study were aged between 1 month and 9.5 years. In study 2, we collected data on parents' perspectives of using LittleBeats (again, following the completion of several day-long recordings at home) from a separate, larger sample. In this second study, we administered close-ended questionnaire items developed considering the qualitative themes identified in study 1. The parents in study 2 also had the opportunity to provide open-ended comments. In study 2, we narrowed our developmental focus to children aged 1 month to 5 years because our substantive interests focused on early childhood, and analytic tools are currently being developed for LittleBeats data collected among children aged ≤5 years.

Study 1

Methods

Participants

A total of 47 families with children aged 1 month to 9.5 years were recruited through web-based forums (eg, Facebook [Meta Platforms, Inc] parenting groups) and flyers distributed to local organizations (eg, libraries and day care centers) in a small Midwestern city. Because the larger study from which data were drawn included assessments of child stress physiology, families were excluded if their children had any known cardiac abnormalities. Of the 47 families that participated in the larger study, 42 (89%) completed the follow-up interview about their experience of using LittleBeats at home. Interviews were not completed with 5 (11%) families because of losing contact with them or because interview procedures were not finalized at the time of their study participation.

From these 42 families, 43 children (n=20, 47% female) participated. In 1 instance, 2 (5%) children (aged 13 and 71 mo) were from the same family. Children were aged 1.1 month to

9.5 years (mean 44.9, SD 38.36 mo) and represented 6 age groups: young infants (aged 1-5 mo; 7/43, 16%), older infants (aged 6-17 mo; 10/43, 23%), toddlers (aged 18-35 mo; 7/43, 16%), preschool-aged children (aged 36-59 mo; 6/43, 14%), early school-aged children (aged 5-7 y; 7/43, 16%), and school-aged children (aged 8-10 y; 6/43, 14%). Overall, 22 (51%) children were first born, 11 (26%) were second born, and 9 (21%) were third or later born. Mothers were aged, on average, 35.04 (SD 4.09) years, and fathers were aged, on average, 37.42 (SD 4.48) years. Across mothers and fathers, the highest level of education reported included a high-school degree (1/79, 1%), some college or 2-year degree (18/79, 23%), a bachelor's degree (22/79, 28%), or an advanced degree (38/79, 48%). Parents identified as Black (2/79, 3%), Asian (3/79, 4%), White non-Hispanic (70/79, 89%), Hispanic (2/79, 3%), or >1 race (2/79, 3%). These demographic data were missing for 2 (5%) of the 42 mothers and 3 (7%) of the 42 fathers. The mean family income was US $79,500 (SD US $25,000).

Ethical Considerations

This study was approved by the institutional review board at the University of Illinois Urbana-Champaign (protocol #21032).

Overview of LittleBeats Procedures

LittleBeats collects 3 streams of data (ECG, motion, and audio data) simultaneously while participants go about their everyday routines (Figure 2). Owing to COVID-19 protocols, all participant engagement was remote. LittleBeats kits (ie, LittleBeats device and shirt, ECG leads, disposable ECG electrodes, alcohol swabs to remove residue from electrodes, medical tape to secure wires on the child's chest, charging cable and block, and setup instruction cards) were either mailed or delivered by a research coordinator to the family's home. After receiving the kit, the mother and child met with the study coordinator through Zoom, a secure video web-conferencing platform. During this 40-minute Zoom visit, the study coordinator guided the mother through the LittleBeats setup (described in more detail subsequently), and the mother-child dyad participated in a series of tasks (video recorded for subsequent coding), including a baseline assessment of child stress physiology at rest and a mother-child play session.

For child participants aged <7 years, the mother-child dyads were also asked to complete a brief series of age-appropriate motion interaction tasks, such as the mother picking up her child (aged 1-4 mo), the mother and child (aged 11 mo) clapping together, or the mother and child (aged 6 y) playing "Simon Says." Toward the end of the Zoom visit, the study coordinator provided instructions for completing the LittleBeats home recordings. Families were asked to complete 2 day-long recordings (approximately 8 hours per day). All adults present at home during the recordings (eg, parents, grandparents, and babysitters) were required to provide consent to the LittleBeats recordings using a secure web-based form provided by the research team. If any nonconsenting adults were at home, parents were asked to turn off the device while these individuals were present. At the end of each day of recording, parents (usually mothers) completed a brief questionnaire about the day's recording (eg, recording start and stop times). To compensate the families for their time, parents were sent a US $100 e-gift card.

With regard to setting up LittleBeats, the research coordinator walked the mother through the following setup steps at the beginning of the Zoom visit: (1) threading a set of ECG lead wires (20 cm) through the back of the shirt pocket, (2) connecting ECG leads jack (2.5 mm) to the LittleBeats device, (3) turning the device on by sliding the switch to the "on" position (confirmation that the beginning of the recording is indicated by a red flashing light displayed on the device), (4) placing the device in a snug, specially designed shirt pocket, which is secured using 2 snaps, (5) snapping leads to 3 repositionable latex-free gel electrodes, (6) putting the LittleBeats shirt on the child, (7) cleaning the skin (where the electrodes will be placed) with an alcohol prep pad and then placing the electrodes on the child's skin, and (8) applying a small strip of 3M Micropore medical tape to each ECG wire approximately 5.1 cm below each ECG sticker to help secure the wires in place.

At the end of the Zoom visit, the research coordinator also walked the mother through how the LittleBeats device should be removed. The removal steps include (1) removing electrodes from the child's skin and using provided alcohol wipes, as needed, to remove residual gel from the electrodes; (2) unsnapping the electrodes from the ECG wires; (3) taking off the LittleBeats shirt; (4) removing the device from the shirt pocket; (5) sliding the slide switch to the "off" position; and (6) plugging the device into the provided charging cable (microUSB cable).

Figure 2. (A) LittleBeats device case; (B) LittleBeats supplies, including electrocardiogram leads, electrodes, charger, and shirt; and (C) an infant wearing LittleBeats at home.

LittleBeats Device Design and Study Implementation for End Users

LittleBeats was developed with parents and children (ie, end users) in mind. To provide a context for parents' interview responses about their study experiences, we noted several aspects of the device design and study implementation intended to proactively increase usability and decrease concerns about privacy. With respect to usability, we provided participants with clear, illustrated instructions in several formats (eg, hard copy and on the web). The device was also designed to be simple to use, with an on-off switch and a charging port, and we provided parents with the all the materials in the LittleBeats kit (refer to the *Overview of LittleBeats Procedures* section) that they would need to set up and use LittleBeats at home. For the child's comfort, the device is compact (55×57×14 mm) and lightweight (1.48 oz), with foam padding lining the inside of the shirt pocket in which the device is to be placed. The shirts are adorned with a variety of pocket designs (eg, hearts, animals, trucks, and dinosaurs) to appeal to toddlers and preschool-aged children, and, as part of the LittleBeats kit, families received 2 shirts with different designs. For older children, we provided more age-appropriate solid shirt pockets.

With respect to privacy, audio recordings provide high-fidelity information regarding participants' lives and require special considerations related to participant privacy, data confidentiality, and recording bystanders (for review, refer to the study by Cychosz et al [13]). Our approach to protecting participant privacy aligns with user-centered privacy protections recommended for mobile health research [36] and a "rights-based" approach adopted increasingly in the United States and used by the European Union (ie, General Data Protection Regulation), according to which individuals have the right to control their personal data, including but not limited to consent, erasure, secure data management practices, and transparency. For example, an important strategy to minimize privacy risks includes giving participants control over recordings [37,38]. In this vein, parents were told at several points during their participation (eg, consent process and consent form, verbally during the Zoom visit, and written instruction card) that they were free to turn off or pause the device at any time and that they could request that their recordings be partially or fully destroyed and not be used in the research. With respect to third-party individuals, parents were also instructed to use the device at home when only immediate family members or other consenting adults are present. Parents were informed that all data files were marked only by identification numbers, machine learning algorithms would be used to process the audio data, research personnel would listen to only snippets of the audio files as part of checks on algorithm development and accuracy, and research personnel were trained to protect participant privacy and would immediately cease listening to audio snippets in instances where personal information (discussion of medical, financial, or other personal issues) is being relayed. To minimize the risk of data being intercepted during transfer (ie, uploading data via wireless or Bluetooth networks), data were stored directly on a microSD card on the physical device, and files were configured in such a way that only study personnel could access the data in a human-readable format (eg, wav files for audio) using a data processing pipeline developed specifically for LittleBeats. Because LittleBeats is not a commercial device, simple modifications can be made to the device firmware (eg, "turning off" ≥1 of the sensors) to suit research goals (refer to the study by Islam et al [19] for details about technical specifications).

Parent Interview and Coding Procedures

Upon the completion of the LittleBeats recordings, parents (41 mothers and 1 father) completed a brief phone interview about their experiences of using LittleBeats in the home. To help minimize social desirability biases in parental responses, such as parents' reports of positive experiences with LittleBeats instructions received during the Zoom visit, these interviews were conducted by a second study coordinator who was not present during the Zoom visit. Guided by the dimensions outlined in the Digital Health Checklist [20], as well as special considerations related to research with children [27], our semistructured interview was designed for the purpose of this research to capture information about parents' experiences and perspectives regarding access, usability, privacy concerns, and risks and benefits with respect to the use of the LittleBeats device and the process of carrying out home recordings. Participants rarely provided information specific to the fourth domain of the Digital Health Checklist, data management, which encompasses how data are collected, stored, and shared and the extent to which the data are accessible to other systems or interoperability. Given the nature of the LittleBeats data (ie, they are not shared outside the research team, not accessible or integrated with other systems, and not transferred via a wireless or Bluetooth network that might be susceptible to security breaches), the data management theme is somewhat less relevant to LittleBeats than to health applications that might be accessed by multiple users (eg, patients, health care providers, and insurance providers). When parents expressed their views on the processes of data collection, storage, and security in the interviews, they almost exclusively focused on the audio recordings and privacy considerations. Therefore, we coded these responses under the privacy domain.

The interview included 11 open-ended questions, and the study coordinator conducting the interviews used standard probes to gain more insight into parents' experiences, perceptions, concerns, and questions (Multimedia Appendix 1). The interview questions allowed for feedback from all family members' perspectives (ie, the participating child, participating parents, and any other children or adults in the home). All parent interviews, conducted by the same study coordinator to ensure consistency, were audio recorded with the participant's permission. Interview recordings were manually transcribed, and identifiable information (eg, names and birth dates) and conversational placeholders (eg, "uh-huh") were omitted from the transcripts.

We used Taguette [39], an open-source web-based tool for coding textual qualitative data, to capture prevalent themes in our interview data and followed the 6-step approach to thematic analysis defined by Braun and Clarke [40]. At step 1, a review of the transcripts provided preliminary ideas for codes. At step 2, initial codes were generated based on the data from 5 interview transcripts of parents with children from different age groups. Through a series of team discussions, we developed an initial codebook focusing on areas that fell into the larger categories outlined in the Digital Health Checklist [20]. Three transcripts were then used for training purposes, and 3 researchers individually coded the transcripts. Discrepancies were discussed, and additional changes were made to the codebook. Upon the completion of the training, 1 researcher (who was not informed of the specific study objectives) coded all the transcripts using the refined codebook. Reliability was assessed by having the fourth author code 8 randomly chosen transcripts, and among the parent responses that both coders deemed codable, agreement was excellent (Cohen κ=0.967). At step 3, the research team met on a regular basis throughout the coding process to identify and discuss potential themes. At step 4 and after the completion of coding, final themes were reviewed by checking themes in relation to the entire data set to ensure an accurate representation of the data. At step 5, themes were refined and finalized by providing descriptive labels and definitions. At the final step, we organized the results based on the key domains of the Digital Health Checklist and created a summary table of themes with selected interview excerpts to illustrate the findings.

Results

Overview

Themes identified under the major categories of access and usability, privacy, and risks and benefits are summarized in the subsequent sections. Overall, similar themes were identified across developmental periods, although specific examples illustrating a given theme often differed depending on whether the parent reported on their infant, toddler, preschool-aged child, or school-aged child.

Access and Usability

According to the Digital Health Checklist, the domain of access and usability prompts researchers to consider whether the participant will be able to use the device as intended. This may involve evaluating whether the product has infrastructure requirements, such as internet access, as well as whether the device has been successfully used in the target population. In this study, usability refers to parents knowing how and being able to successfully use the LittleBeats device and materials (eg, ECG leads). Furthermore, usability encompasses families' experience of and ability to adhere to the study procedures more generally (ie, participant burden, eg, completing multiple day-long recordings), beyond the use of the device itself (Table 1).

A majority of parents expressed sentiments regarding their ability to easily operate the device (ie, turning the device on-off and charging the device). Some parents indicated feeling comfortable given their previous experience with comparable equipment, yet other parents with no such prior experience expressed similar views about the ease of use. Parents also commented that the instructions were helpful and appreciated having a variety of resources to refer to, if needed (eg, written instruction card, website, and study personnel contact). Aside from operating the device itself, parents had varying views on the materials needed to place the device on their child. Some parents noted that the design was well thought out and that setting up the electrodes was not complicated. However, other parents indicated some challenges with the materials, such as with threading the electrodes through the back of the shirt pocket.

Parents also expressed differing perspectives about the ease of setting up (and removing) the device. Although many parents felt comfortable placing LittleBeats on their children, some parents noted that gaining their children's cooperation was sometimes a challenge. For instance, some parents reported difficulty putting the device on their "wiggly, squiggly" infants. Other parents reported reluctance on the part of their toddlers or preschool-aged children, who could express their opinions and desires verbally. Typically, if challenges related to child cooperation were experienced, it was during the setup phase, and parents suggested that once their child was wearing LittleBeats, it was quickly forgotten. Parents expressed that the placement of the device on the upper anterior torso (ie, chest) may be disruptive to some activities, such as napping for a child who is a tummy sleeper. Relatedly, the device being concealed in the shirt pocket, with the ECG leads underneath the shirt, was viewed as a disadvantage by some parents who wanted to know whether the device was recording properly or whether there was a malfunction (eg, device turned off or ECG electrodes fell off).

With respect to participant burden, parents expressed a mix of perspectives. Many parents described day-long recordings (ie, >8 h/d) as feasible but challenging. However, parents noted factors that mitigated this challenge, such as the need to record for only a limited number of days spaced across multiple weeks, the ability to schedule their recordings when it worked for them, and the reduction in other competing activities due to the COVID-19 pandemic. In the same vein, parents expressed wanting more features to help them fulfill project expectations. Currently, the device provides no information to the user beyond an indicator light showing that the device is powered on. Parents found it difficult to know how long they had recorded for or how much battery charge was left when using the device.

In addition, many parents described the project as convenient, indicating that the remote data collection procedures were appealing. Being able to collect data at home, on their family's own schedule, made it relatively easy to participate. Parents were not burdened by the need to travel to a research laboratory, and they could set up the device and start recording when it fit their schedule. Concerns about being able to keep the device on securely or ensure that the device was collecting data were voiced by some parents of older and more active children (eg, increased unsupervised time and gel adhesive weakening owing to perspiration). Other parents expressed their worry that their children would damage the device during data collection.

Table 1. Themes, subthemes, and example excerpts related to the access and usability of the LittleBeats device and study procedures (study 1).

Themes and subthemes	Example excerpts
Operating the device	• "Everything was pretty easy. It was easy to charge, it was easy to you know put the stickers on and attach, and like I said I don't think she really felt like it was on. The first day after she asked after an hour 'how long have I had it on' I was like 'why is it uncomfortable' she was like 'no I was just wondering' and I was like 'oh okay.' I don't think she even realized she had it on half the time." (Parent of a school-aged child)
Instructions	• "They [the instructions] were very clear. I mean they made it so that I felt confident putting it on her and doing what I was supposed to do." (Parent of an older infant)
Support materials (specially designed t-shirt, wires, and electrodes)	
Ease of use	• "I think the t-shirt definitely made it easier to use. That was a nice little set up, and it made it, you know, stay in place and like see where it [the device] needed to be for it to be hooked up and stay in place...And then even with the hole on the inside [of the shirt] to make it easy to get all the cords. That was really a unique design tool but effective." (Parent of a school-aged child)
Challenges	• "It was a little hard getting the black metal piece through the back of the shirt. Like I needed that hole to be a little bigger. So, I'm sure I ripped mine just a little bit...But I just made it a little bit looser." (Parent of an older infant)
Setting up and removal of LittleBeats	
Comfort with set up	• "I'm pretty comfortable getting it set up and turning it on. It seems pretty straightforward." (Parent of a school-aged child)
Child cooperation	• "It was mostly the initial putting it on. She didn't want to cooperate with letting us get it on...but after a little bit she forgot it was there because she didn't have any issues messing with it and then when it was time to take it off she was fine." (Parent of a preschool-aged child)
Location of device	• "I wish the device itself was a little more discreet. Well, he's a stomach sleeper so for naps I had to take it off but if it was a little more discreet or was not in front of the t-shirt but maybe on the arm it would be more convenient." (Parent of an older infant)
Participant burden	
Time commitment	• "You know once we broke it up a little bit we could [complete recordings]. I was more worried about you know were rarely all home just the four of us especially now that quarantine is over...We're just more on the go than we were a year ago." (Parent of an early school–aged child)
Convenience	• "It was really easy for me as a parent. I drive my other son like I said to [research lab in different city] a bunch...and so that is just a drag, a lot of back and forth. But for I would say from a parent's standpoint, this was very easy for me to do." (Parent of an early school–aged child)
Worry about recordings	• "My son's pretty active, so he sweats a lot over the course of the day. The little stickers would kind of migrate a little bit...So, I worry a little bit that the first recording like the second half of the day might not be as accurate as it was supposed to be." (Parent of a preschool-aged child) • "It would be nice if there were some kind of indicator of battery more visible. And it was also, you know, since I had to take it on and off then count the time, that was also kind of challenging...so some kind of indication of time would also be awesome but I don't know how complicated it would be to make it." (Parent of an older infant)
Worry about device	• "A lot of the activities that she wants to do involve painting or drinking water...those kinds of worrying me every time she picks them up. I was more concerned about the hardware." (Parent of a preschool-aged child)

Privacy

The privacy domain focuses on the types of personal information that are or will be collected about participants. In this study, privacy relates to participants' expectations about and understanding of the process of data collection, in general, and the audio recordings, specifically. Furthermore, this category encompasses the control that participants had over the data collected (Table 2).

Many parents commented that they were initially apprehensive about the home audio recordings but that their worries subsided when provided with more details during the initial informational call with the study coordinator. Other parents noted feeling more comfortable with the audio recordings over time as they participated in the study. Some parents discussed that although they had no concerns, their spouse or partner did. Typically, only 1 parent (usually the mother) was present for the initial informational call with the study coordinator, and this parent then conveyed information to the other parent, which often sufficed to relieve privacy concerns.

By contrast, for some parents, positive views of research, such as having trust or placing value in research, negated concerns about privacy. Other participants described not being concerned with the audio recordings because they "had nothing to hide." From this view, the audio would capture a typical day in their life, and participants elaborated by describing that the recordings would include everyday family discussions as well as arguments, which participants conveyed as just part of ordinary family life. Others' lack of concern regarding the audio recordings stemmed from their ability to control when they were recording and, consequently, what was being recorded. They described the process of turning the device on and off as relatively easy and, therefore, reported turning the device off when they were discussing private matters. Some participants mentioned developing ground rules ahead of time to ensure that private information was not discussed when recordings were taking place and, if needed, would alert or remind other family members of the recordings.

The possibility of recording other individuals beyond immediate family members was considered. In working to respect others' privacy, the participants mentioned several challenges. Some participants expressed that they altered their typical day to avoid interacting with others so that they would not have to worry about unintentionally recording a nonconsenting individual. Other participants stated that although they had planned to record at convenient times when no nonconsenting individuals were around, unexpected situations arose. In addition, although parents had the ability to control when the device recorded, some parents acknowledged that remembering to turn off the device when others were around could be challenging.

Table 2. Themes, subthemes, and example excerpts related to privacy concerns about the LittleBeats audio recordings (study 1).

Themes and subthemes	Example excerpts
Initial apprehension about audio recordings	• "Cause that was my husband's big question like 'are they just going to sit and listen to our day?' So, he was a little worried about that but once it was explained [that machine learning algorithms would be used to analyze the audio data] he was more comfortable and on board." (Parent of an early school–aged child)
Unconcerned about audio recordings	
Former views or experiences of research	• "She [study coordinator] also told me that it is only used for research purpose and nothing else...I actually love to participate [in research studies]. It is only used for research purposes, so that's okay."(Parent of a preschool-aged child) • "I was in a study when I was pregnant and we did something similar...my understanding was that the recordings just gets run through the software so we really don't have anything any interesting happening here so I wasn't terribly concerned about that [the audio recordings]." (Parent of an older infant)
Just an ordinary family context	• "I explained everything to everybody [family members, including older children in home]. I do remember there was one particular situation where my 10-year-old was getting into trouble and afterwards he said, 'Well, they're gonna hear that!' And I said this is just a regular family, there's nothing to be embarrassed about or whatever." (Parent of a toddler)
Ability to control the recordings	• "My husband's a veteran, and he works at the V.A..., so we had to make sure we turned it off before he came home from work because a lot of times he talks about his day." (Parent of a toddler)
Respecting others' privacy	
Adjusting routines or activities to accommodate the study	• "I think the only thing is that we didn't go play with some friends across the street those days where we would've otherwise. Like it impeded a little bit of our typical routine, but it felt pretty unobtrusive." (Parent of a preschool-aged child)
Unexpected situations	• "When something was happening that I wasn't expecting, like when I would get a phone call or something like that, and I was just a little concerned about remembering to turn off the device." (Parent of an older infant)

Risks and Benefits

Evaluating the risks of possible harms in relation to the possible benefits resulting from the knowledge to be gained from the research is linked to the principle of beneficence. Study benefits should outweigh the possible harm to participants and the groups they represent. Risk assessment includes evaluating the type of harm, psychological, physical, reputational, or economic. In addition, researchers must consider the duration, severity, and intensity of the possible harm. Specific to the risks associated with the use of LittleBeats at home, parents expressed varying views along several dimensions, including safety, child comfort, and understanding of the research and its direct outcomes for participants (Table 3).

Many parents expressed that they thought the device was safe for their children to wear. These parents described not being concerned about safety because of the design of the device and the protective features built into it (eg, device was enclosed, tape-covered wires, fitted shirt, and pocket with secure snaps).

Some parents indicated that they initially had safety concerns (eg, the device being close to the skin and use of Bluetooth to transfer data) before learning more about the device and its setup (eg, the device itself is not in contact with the skin but is placed in a padded pocket, data are stored directly on the device, and Bluetooth is not used for data transfer). In some instances, parents detailed concerns about their children wearing the device in unsupervised contexts, such as during naptime, and they preemptively removed the device before naps.

Parents also commented on their children's level of comfort or discomfort. Several parents mentioned that they observed their child functioning normally, such as engaging in typical routines and activities. Parents also stated that their children did not express any discomfort and did not seem to notice that they were wearing the device after a while. Other parents noted their children's discomfort in putting on or removing the electrodes and medical tape used to secure the wires on the chest. Some parents worried about how comfortable it would be if the child were to hit the device on another object, such as the edge of a table.

Finally, parents' understanding of the research and its direct outcomes for their families may confer risks and benefits. Some parents revealed a limited understanding of how the data would be used (ie, the ultimate outcome of the research process) or wanted direct feedback on their children's development, which could pose unintended risks (eg, unfulfilled expectations of direct benefits). Other parents voiced the benefits attributed to participating in the research project itself. For instance, participation provided dedicated time spent together as a family, or completing the surveys was an opportunity to reflect on their children's activities and development. Several parents expressed their desire to contribute to the project because they recognized the importance of the research. Some parents indicated that they had enjoyed participating in previous studies, and others stated that this project's description seemed interesting and fun. Other parents of older children revealed that when they initially talked to their children about the study, their children seemed interested in participating, so they signed up. Some participants communicated that their children enjoyed participating in the project, with one parent acknowledging that their children felt special for a day while wearing the LittleBeats shirt.

Table 3. Themes, subthemes, and example excerpts related to the risks and benefits of participating in the LittleBeats study (study 1).

Themes and subthemes	Example excerpts
Safety	• "No [safety concerns] because all of the wires were covered by her shirt and taped down." (Parent of an older infant) • "Not really [any safety concerns]. I mean the wires were short enough that I wasn't worried about them." (Parent of an early school–aged child) • "I thought it will get like hot because I recorded for the 8 hours straight, I didn't stop it at all, I was worried maybe it's gonna be hot or something, but it wasn't hot at all. That was my main concern only." (Parent of a younger infant) • "And then I did have an initial concern...about the safety of having that device running on Bluetooth. I'm not sure how it communicates data and that being so close to skin." (Parent of an older infant)
Child's comfort or discomfort	• "I guess putting them [electrocardiogram electrodes] on wasn't the hard part. The hard part was taking them off, especially the was a little bit hard, and my son is also not very fond of changing clothes." (Parent of an older infant) • "I'd probably take it off especially because my little one is about 10 1/2 months and she's a tummy sleeper so that would be uncomfortable." (Parent of a preschool-aged child) • "I mean it seemed it was fine. My sons were playing outside you know riding their bikes and everything and they didn't...say anything was uncomfortable." (Parent of an early school–aged child)
Outcomes of participating in the research	
Limited understanding	• "I would love to know what kind of information. I know what kind of information they collected with the device and I'm just curious what they are going to use it for in the future." (Parent of an older infant)
Understanding gained	• "[Filling out] this survey, I found that I am pretty lucky that my son is more adaptable. The question, was for example, 'when you want him to go to bed, he just cried or tantrum' but he never does that." (Parent of a preschool-age child)
Parent's enjoyment or satisfaction	• "I just like participating in research and helping out the scholars. In my undergrad, I was doing some research and I know how important it is and how hard it can be so...I think it's good to help." (Parent of an older infant) • "I actually like to spend time with my son. He goes to school every day, so I like to do something with him like the zoom interview. And also I want to show him new technologies." (Parent of a preschool-aged child)
Child's enjoyment	• "I didn't mind the surveys or anything, and my son loved wearing the LittleBeats. He kept asking if he could put them on. So, I think it captured the kid's interest too." (Parent of an early school–aged child) • "We had fun doing it [the study], and I think [my son] enjoyed being special, wearing his special shirt for a day." (Parent of a toddler)

Study 2

Overview

Building on the key themes of access and usability, privacy, and risks and benefits identified in study 1, we administered a brief survey among a larger sample of parents participating in a different LittleBeats study with children aged 0 to 5 years. Although our main interest was to complement the qualitative findings of study 1 with a quantitative assessment of parents' perceptions using close-ended rating scales, parents were also able to provide open-ended comments. Therefore, we have also summarized the main themes reflected in these open-ended comments.

Methods

Participants

In study 2, a total of 110 parents (n=108, 98% mothers and n=2, 2% fathers) completed a user experience survey after completing 3 days of LittleBeats recordings at home. Recruitment procedures were similar to those described in study 1. Children (60/110, 54.5% female) were aged, on average, 23.4 months (SD 16.87 mo; range: 2-65 mo) and were identified by parents as Black (n=5, 4.7%), Asian (n=8, 7.5%), White non-Hispanic (n=67, 63.2%), Hispanic (n=15, 14.2%), or >1 race (n=11, 10.4%). Children were first born (n=50, 47%), second born (n=39, 38%), and third or later born (n=17, 15%). Parents were aged, on average, 34.85 (SD 5.01) years, and their highest level of education reported included some high school or high-school degree (4/106, 3.8%), some college or 2-year degree (9/106, 8.5%), a bachelor's degree (33/106, 31.1%), or an advanced degree (60/106, 56.6%). Parents identified as Black (7/106, 6.6%), Asian (13/106, 12.3%), White non-Hispanic (75/106, 70.8%), Hispanic (8/106, 7.5%), or >1 race (3/106, 2.8%). The mean family income was US $83,250 (SD US $26,470). Of the 110 parents, 4 (4%) were missing responses on the demographic survey but did complete the LittleBeats user experience survey described subsequently.

Ethical Considerations

This study was approved by the institutional review board at the UIUC (protocol #22631).

Procedure

Families were mailed a LittleBeats kit and participated in a Zoom visit, during which a study coordinator walked the parent through the LittleBeats setup and a visit procedure consisting of a baseline assessment of child stress physiology and parent-child interaction tasks (eg, play). At the end of the visit, parents received instructions about completing the day-long recordings and were asked to complete 3 day-long recordings over the course of 2 weeks. Parents also completed a series of web-based questionnaires about family demographics, child behavior, and family functioning. Parent questionnaires were administered either via Qualtrics or REDCap (Research Electronic Data Capture; Vanderbilt University [41,42]) hosted at the UIUC, with the support of the Interdisciplinary Health Sciences Institute and Research IT—Technology Services at the UIUC. Both web-based software platforms are designed to support secure data capture for research studies. Once parents returned the LittleBeats kit by mail, 1 parent in the household (who had been involved in setting up and carrying out the LittleBeats recordings) was asked to rate 5 items about their experience of using LittleBeats, including setting up LittleBeats, along with their perceptions of safety, privacy, and participant burden. Each item was rated on a 5-point scale ranging from 1 (*strongly agree*) to 5 (*strongly disagree*). Following each item, parents had the opportunity to add comments or elaborate on their rating. A final open-ended item also asked parents whether there was anything else they would like to share about their experience or anything they would tell someone who was considering joining a LittleBeats study.

Data Analytic Plan

Descriptive statistics, including the frequency distribution, for parental ratings on each of the LittleBeats user experience items were examined. For each close-ended item, we conducted a single-sample *t* test (2-tailed) to determine whether the mean rating significantly differed from the midpoint of the scale (ie, value of 3="neutral"). Finally, using the coding scheme developed in study 1, we assessed themes from parents' responses to the open-ended items.

Results

Parents' Ratings on User Experience Items

Percentage frequency distributions of parents' ratings on the user experience items are shown in Table 4. Single-sample *t* tests indicated a significant difference between the item average (lower ratings indicated greater agreement; higher rating indicated greater disagreement) and the midpoint of the rating scale (3="neutral") for 4 (80%) of the 5 items. Compared with a "neutral" response, parents were significantly more likely to *agree* that (1) the LittleBeats instructions were helpful and clear (mean 1.21, SD 0.41; t_{109}=−45.98; *P*<.001), (2) they felt comfortable setting up LittleBeats on their child (mean 1.42, SD 0.75; t_{109}=−22.22; *P*<.001), and (3) they felt their child was safe while wearing LittleBeats (mean 1.33, SD 0.51; t_{109}=−34.48; *P*<.001). Compared with a "neutral" response, parents were significantly more likely to *disagree* that they worried about being recorded by the LittleBeats device (mean 3.62, SD 1.06; t_{108}=6.14; *P*<.001). The final item tapped parents' perceptions of burden ("I felt that completing LittleBeats recordings for full 3 days was challenging"), and the item average (mean 2.98, SD 1.17) did not significantly differ from "neutral" (t_{109}=−.16; *P*=.87).

Table 4. Frequency distributions of parental rating of the LittleBeats user experience survey (study 2; n=110).

Survey item	Strongly agree, n (%)	Agree, n (%)	Neutral, n (%)	Disagree, n (%)	Strongly disagree, n (%)
The instructions to setup LittleBeats were helpful and clear.	87 (79.1)	23 (20.9)	0 (0)	0 (0)	0 (0)
I felt comfortable setting up LittleBeats on my child.	75 (68.2)	29 (26.4)	2 (1.8)	3 (2.7)	1 (0.9)
I felt my child was safe while wearing LittleBeats.	76 (69.1)	32 (29.1)	2 (1.8)	0 (0)	0 (0)
I was worried about being recorded when the LittleBeats device was on.	3 (2.7)	13 (11.8)	32 (29.1)	35 (31.8)	26 (23.6)
I felt that completing LittleBeats recordings for 3 full days was challenging.	11 (10)	32 (29.1)	26 (23.6)	30 (27.3)	11 (10)

Parents' Responses to Open-Ended Items

A review of parents' responses to the optional item to add further comments following each of the rating scales revealed themes that closely mirrored study 1 findings. Regarding the ease-of-use item, 29 (26.4%) of the 110 parents added comments. Most parents noted that having an instruction card included in the kit, as well as a QR code to easily link to the website for more detailed instructions, increased usability.

Regarding comfort in setting up the device, 23 (20.9%) of the 110 parents added comments. Parents noted that they felt comfortable and that the setting up of the device was easy. However, parents also noted that the process of setting up the device was difficult when their child moved around. Other parents mentioned the comfort level of their child (eg, noting that their child felt discomfort when removing the ECG electrodes).

Regarding safety, 25 (22.7%) of the 110 parents added comments. Parents noted few concerns because the device was concealed in a pocket and not easily accessible to the child. Parents who expressed a concern commented on the placement of the device on their child's chest.

Regarding concerns about being recorded, 32 (29.1%) of the 110 parents added comments. Some parents noted feeling self-conscious about their parenting or other family members' language choices. Typically, these comments were followed by comments about feeling relieved that the audio would be processed by a machine (vs a human coder). By contrast, many parents explained that they went about their day as usual, which typically contained some sort of sibling argument or other family disagreements.

Regarding participant burden, 68 (61.8%) of the 110 parents added comments. Unlike in study 1, where participants were asked to use the device for 2 days, study 2 participants were asked to use the device for 3 full days (or a total of about 24 hours) over the course of 2 weeks. Several parents commented on their families' busy schedules and difficulty finding 3 full days when only immediate family members were present.

Finally, a number of parents (46/110, 41.8%) responded to the final open-ended question asking whether they had any other comments they would like to share. Responses mirrored study 1 themes in several respects, including parents' and children's enjoyment in participating in the study (eg, "fun and easy" and "I would recommend to my friends"), children's ability to forget about the device and go about their usual day (eg, "did not interfere with our day"; "[Child] did not notice the device...he was able to nap with it on and so it was really pretty simple to participate!"; and "once the shirt was on, she forgot it was there and so did I!"), and suggestions for ways to minimize burden and improve the experience (eg, adding a display on the device that provides more information about battery charge, power status, and recording length).

Discussion

Summary

Digital health technologies have largely been developed with adults in mind. Interest in and attention to the use of wearable devices among infants and young children, however, has been growing, and data collection using wearable devices provides several advantages over traditional data collection methods, including continuous assessment, greater ecological validity, and the automated detection of behaviors using machine learning algorithms. Given these advantages, combined with rapid technological advances, it is likely that the use of wearables in child development research will burgeon in the coming years. Therefore, assessing how such devices and related data collection protocols are perceived and experienced by parents and their children is critical. User experience studies not only address ethical considerations but can also lead to important changes in research protocols that address parents' concerns and increase the benefits for future families who participate. Indeed, our mixed methods investigation across 2 studies yielded consistent findings that shed light on parents' experiences and perceptions of LittleBeats' usability and safety, the privacy of the audio recordings, and potential risks and benefits of participating in research of this kind. A large majority of parents indicated that device instructions were helpful and clear, the device was easy to use and safe, and remote visits were convenient. Parents' views about privacy, risks, and benefits were more varied, although, on average, parents reported feeling comfortable with the audio recordings. In summarizing the major themes identified within the major categories, we consider ways in which the findings can inform the future design and implementation of wearable platforms in child development research.

Key Findings

Results across all themes underscored the variability in parents' (mostly mothers') perspectives and experiences. With respect to access and usability, some parents expressed interest in having access to information that indicated the cumulative time recorded as well as the battery charge remaining. Such additions to the platform would eliminate parents' need to track the recording length and minimize parents' concerns about whether the device was sufficiently charged and recording. Some parents also noted difficulty with threading the ECG lead wires through the back of the shirt or were worried that their child would tug on the wires. These challenges can be remedied by changing the shirt design such that the ECG wires would be more fully integrated into the shirt fabric or design. Although parents indicated that day-long recordings (ie, >8 h/d) were feasible, some parents noted challenges. To alleviate the burden of day-long recordings, the time requirements can be adjusted to be more flexible. For instance, parents can be asked to complete recordings for fewer hours per day across multiple days (ie, 3 to 4 h/d across 4 to 5 d), although the optimal length and frequency of recordings needed to reliably capture the constructs of interest will vary as a function of the research questions being addressed. Importantly, such burdens were balanced by parents' comments regarding the convenience of remote visit procedures and the ease of using LittleBeats.

Privacy was a theme that also garnered a variety of responses. Some parents indicated few concerns about the privacy of the home audio recordings, whereas other parents worried that the recordings captured private conversations. In the latter case, some families used rules or reminders to control or limit when audio recordings were collected. It is also notable that parents within the same family sometimes expressed differing levels of comfort or concern with the audio recordings. When this pattern emerged, it was largely fathers who voiced concern about invasion of privacy, perhaps because they were not present for initial conversations with the study coordinator, who detailed how the data would be collected and used.

We consider 2 main ways to address parents' privacy concerns about the home audio recordings (also refer to the study by Cychosz et al [13]). First, providing specific and concrete examples of how the audio recordings are processed and analyzed, perhaps by illustrating a hypothetical example of the data collection, processing, and analysis steps, may help ease privacy concerns. Indeed, some parents noted that the use of machine learning algorithms to analyze the data alleviated their concerns about the audio recordings and privacy-related issues. Thus, describing the machine learning algorithms in a detailed yet accessible manner for nontechnical users and stating ways in which the data will not be used or analyzed (eg, no transcriptions of speech) may help reassure parents. Such information should be provided to all family members participating in the home recordings, including older siblings, and should be presented in various formats (eg, brief informational videos, hard copy pamphlets, interactive web page), along with multiple ways to contact study personnel for questions or comments. As part of this solution and building on some parents' perspectives that the recordings were just capturing "typical family life," researchers conducting day-long recordings may also explicitly highlight the family as an important context for development, coupled with appreciation for the fact that all families are different, and that, as researchers, we want to capture what life is like for each family and infant.

A second solution to alleviate parents' concerns about privacy could involve technological innovations, such as collecting audio recordings in which speech content is not intelligible (refer to the study by Levin et al [33]) or data processing (eg, machine learning algorithms) that occurs on the device or hub in the home so that the audio recordings are not stored or released to the researcher. However, these solutions require further technological advances in audio signal processing and raise issues regarding data-quality assurance. That is, without high-fidelity recordings, the validation and quality checks of machine learning algorithms become difficult. Furthermore, when parents were presented with several hypothetical scenarios for collecting child sensor data in the home environment, parent-reported willingness to participate did not significantly differ between study scenarios in which lower resolution audio data were collected (eg, recording 1-min snippets every 20 min and processing audio data automatically so that raw audio data are not stored) and study scenarios in which higher resolution data (eg, continuous audio recordings) were collected [33]. Taken together, although technological solutions aimed at increasing privacy protection seem to be a reasonable avenue to pursue, future studies on users' experiences of child wearables, particularly home audio or video recordings, should systematically assess parents' concerns, needs, and desires when it comes to balancing the privacy of day-long home recordings with the benefits of participation.

Third-party or bystander privacy is also a complex issue [37,38]. In this study, there were two categories of potential third parties: (1) nonparental caregivers or relatives at home who were part of the child's regular routine and (2) individuals who were not part of the home environment (eg, delivery persons and neighbors). In the first case, nonparental caregivers can be included in the recording if they provided consent. In the second case, the parent would need to turn off the device while the individual is present or change their routine to avoid third parties, which may have consequences for ecological validity. Concerns about third-party recordings can also be resolved by the same types of technical solutions outlined earlier.

The principle of beneficence yielded a variety of responses regarding the risks and benefits of the study procedures. First and foremost, safety was a key theme, and across both samples, parents predominantly expressed views that LittleBeats was safe. When concerns about safety were mentioned, parents often presented hypothetical concerns (eg, the device being close to the skin, the device radiating heat, and the child accidentally falling on the device; the last scenario is mentioned as a potential risk in the parental consent form), which were usually alleviated once the parent learned more about the study. Some parents also mentioned concerns about the child wearing the device during unsupervised times, such as naps, and removed the device during these times. Because infants and young children are much more likely to take ≥1 naps over the course of the day, this subtheme differed across age groups, with parents of children in younger age groups being more likely to mention device use with respect to nap times. Another set of risks is related to the child's discomfort, particularly around the application and removal of the ECG electrodes. This potential risk is also mentioned in the parental consent form, and we aimed to ameliorate this risk using latex-free electrodes designed specifically for pediatric populations.

Potential or perceived risks were balanced by parents' perceived benefits, including increased understanding of their child's development through the completion of the parent surveys, parents' satisfaction in contributing to the scientific process, children's enjoyment of the study procedures (eg, play session with parents), and wearing the novel LittleBeats shirt and device. We note that we did not ask directly about perceived benefits in study 2 close-ended items, although parents in this study did indicate the benefits of participation in the final open-ended question asking whether they had any other comments they would like to share. These responses often paralleled the positive sentiments that study 1 parents expressed. Nevertheless, items that assess the perceived benefits of study participation will be important to include in future studies.

With respect to increasing direct benefits to participants, we gave families personalized books summarizing information that we have collected about their children (eg, height and weight at different ages) in prior studies. Such summaries have been well received and appreciated. Similar types of summaries can be made from data extracted from day-long recordings (eg, frequency and duration of infant babbling or crying). Providing this type of study feedback to parents may also promote effective participant recruitment and retention, particularly among studies that involve high-fidelity data, such as audio recordings. As noted by Levin et al [33], individuals are likely to evaluate intrusiveness and data privacy, on the one hand, and direct benefits to themselves and their children (such as receiving useful, personalized information or feedback from the data collected), on the other hand, when making decisions about whether to participate in such research.

Study Limitations and Future Directions

We note several limitations of our user experience studies. First, we did not ask our older child participants about their experiences directly, although parents reported on a variety of child experiences, including compliance with putting on the device, excitement in wearing the shirt, feeling special while wearing the shirt, and comfort or discomfort. The device hardware was relatively compact and lightweight, and parents reported that children tended to forget about it once it was on. Nonetheless, these reflections clearly highlight the need to directly assess not only parents' perspectives but also children's perspectives. Thus, parental reports of their child's experiences should be augmented by direct observations of infants and younger children while wearing the device as well as interviews with older children. Second, we tracked parents' reported experiences based on the child's developmental stage. Similar themes were found across developmental periods, although specific examples of how themes manifested often differed by the child's age. However, because the subsamples of children in different age groups were relatively small, future research with larger subsamples is needed to more thoroughly investigate developmental considerations related to user experiences in the context of research using child wearables. However, an age-specific consideration that did clearly emerge relates to daytime sleep. Third, in both samples, parents reported high levels of educational attainment. Future research on parents' perspectives of using child wearable devices in the home setting should include families with diverse demographic characteristics. Including samples characterized by sociodemographic factors in user experience studies is especially critical for child wearables developed for the purposes of mobile health interventions.

Conclusions

Wearable sensors designed for and validated with infants and young children present researchers and clinicians with tremendous opportunities to assess developmental processes and outcomes in more ecologically valid and potentially less burdensome ways than laboratory assessments. Furthermore, LittleBeats' multiple modalities provide especially rich data to assess an array of constructs central to child development researchers and clinicians, including parent-child vocal turn-taking, regulation of stress, sleep-wake cycles, physical activity, and developmental disorders. At the same time, although we have validated LittleBeats sensors and machine learning algorithms to accurately capture some of these key constructs [17-19,43], the degree to which LittleBeats and similar child wearables deliver benefits (eg, high ecological validity and low burden) will largely depend on acceptance by the end users (eg, parents and children), making user experience studies critical to this research space. In short, if the technology is not acceptable to the end user, it is less likely to be adopted and used as intended. The user experience assessment presented in this paper goes hand in hand with technical validations of the device, and both are critical for successful implementation. The current results suggest that parents predominantly view LittleBeats as easy to set up and use at home, although views regarding privacy and burden were more varied. On the basis of parents' thoughtful and specific feedback, several concrete changes can be implemented to improve the LittleBeats platform and, ultimately, parents' and children's experiences.

Acknowledgments

The authors are grateful to the families who participated in this research and to Jenny Baldwin for her assistance with participant recruitment and data collection. This work was supported by funding from the National Institute on Drug Abuse (R34DA050256), National Institute of Mental Health (R21MH112578), and USDA National Institute of Food and Agriculture (ILLU-793-368).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Parent user experience interview.

[DOCX File , 16 KB - humanfactors_v11i1e49316_app1.docx]

References

1. Peake JM, Kerr G, Sullivan JP. A critical review of consumer wearables, mobile applications, and equipment for providing biofeedback, monitoring stress, and sleep in physically active populations. Front Physiol 2018 Jun 28;9:743 [FREE Full text] [doi: 10.3389/fphys.2018.00743] [Medline: 30002629]
2. Sawka MN, Friedl KE. Emerging wearable physiological monitoring technologies and decision aids for health and performance. J Appl Physiol (1985) 2018 Feb 01;124(2):430-431 [FREE Full text] [doi: 10.1152/japplphysiol.00964.2017] [Medline: 29097633]
3. Muhammad G, Alshehri F, Karray F, Saddik AE, Alsulaiman M, Falk TH. A comprehensive survey on multimodal medical signals fusion for smart healthcare systems. Inf Fusion 2021 Dec;76:355-375 [FREE Full text] [doi: 10.1016/j.inffus.2021.06.007]
4. Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annu Rev Clin Psychol 2017 May 08;13(1):23-47 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
5. Jaworski DJ, Park A, Park EJ. Internet of things for sleep monitoring. IEEE Instrum Meas Mag 2021 Apr;24(2):30-36 [FREE Full text] [doi: 10.1109/mim.2021.9400950]
6. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR Mhealth Uhealth 2020 Sep 08;8(9):e18694 [FREE Full text] [doi: 10.2196/18694] [Medline: 32897239]
7. Harari GM, Lane ND, Wang R, Crosier BS, Campbell AT, Gosling SD. Using smartphones to collect behavioral data in psychological science: opportunities, practical considerations, and challenges. Perspect Psychol Sci 2016 Nov 28;11(6):838-854 [FREE Full text] [doi: 10.1177/1745691616650285] [Medline: 27899727]
8. Seifert A, Hofer M, Allemand M. Mobile data collection: smart, but not (yet) smart enough. Front Neurosci 2018 Dec 18;12:971 [FREE Full text] [doi: 10.3389/fnins.2018.00971] [Medline: 30618590]
9. Stewart E, Milton A, Yee HF, Song MJ, Roberts A, Davenport T, et al. eHealth tools that assess and track health and well-being in children and young people: systematic review. J Med Internet Res 2022 May 12;24(5):e26015 [FREE Full text] [doi: 10.2196/26015] [Medline: 35550285]
10. Memon SF, Memon M, Bhatti S. Wearable technology for infant health monitoring: a survey. IET Circuits Devices Syst 2020 Jan 29;14(2):115-129 [FREE Full text] [doi: 10.1049/iet-cds.2018.5447]
11. Airaksinen M, Räsänen O, Ilén E, Häyrinen T, Kivi A, Marchi V, et al. Automatic posture and movement tracking of infants with wearable movement sensors. Sci Rep 2020 Jan 13;10(1):169 [FREE Full text] [doi: 10.1038/s41598-019-56862-5] [Medline: 31932616]
12. Wang Y, Williams R, Dilley L, Houston DM. A meta-analysis of the predictability of LENA™ automated measures for child language development. Dev Rev 2020 Sep;57:100921 [FREE Full text] [doi: 10.1016/j.dr.2020.100921] [Medline: 32632339]
13. Cychosz M, Romeo R, Soderstrom M, Scaff C, Ganek H, Cristia A, et al. Longform recordings of everyday life: ethics for best practices. Behav Res Methods 2020 Oct 26;52(5):1951-1969 [FREE Full text] [doi: 10.3758/s13428-020-01365-9] [Medline: 32103465]
14. Nuske HJ, Goodwin MS, Kushleyeva Y, Forsyth D, Pennington JW, Masino AJ, et al. Evaluating commercially available wireless cardiovascular monitors for measuring and transmitting real-time physiological responses in children with autism. Autism Res 2022 Jan;15(1):117-130 [FREE Full text] [doi: 10.1002/aur.2633] [Medline: 34741438]
15. Lee R, Skinner A, Bornstein MH, Radford AN, Campbell A, Graham K, et al. Through babies' eyes: practical and theoretical considerations of using wearable technology to measure parent-infant behaviour from the mothers' and infants' view points. Infant Behav Dev 2017 May;47:62-71 [FREE Full text] [doi: 10.1016/j.infbeh.2017.02.006] [Medline: 28347907]
16. de Barbaro K. Automated sensing of daily activity: a new lens into development. Dev Psychobiol 2019 Apr 18;61(3):444-464 [FREE Full text] [doi: 10.1002/dev.21831] [Medline: 30883745]
17. Li J, Hasegawa-Johnson M, McElwain NL. Towards robust family-infant audio analysis based on unsupervised pretraining of Wav2vec 2.0 on large-scale unlabeled family audio. arXiv Preprint posted online on May 21, 2023 [FREE Full text] [doi: 10.21437/interspeech.2023-460]
18. Chang KC, Hasegawa-Johnson M, McElwain NL, Islam B. Classification of infant sleep/wake states: cross-attention among large scale pretrained transformer networks using audio, ECG, and IMU data. arXiv Preprint posted online on June 27, 2023 [FREE Full text] [doi: 10.1109/apsipaasc58517.2023.10317201]

19. Islam B, McElwain NL, Li J, Davila M, Hu Y, Hu K, et al. Preliminary Technical Validation of LittleBeats™: A Multimodal Sensing Platform to Capture Cardiac Physiology, Motion, and Vocalizations. Sensors 2024 Jan 30;24(3):901. [doi: 10.3390/s24030901]
20. Nebeker C, Torous J, Bartlett Ellis RJ. Building the case for actionable ethics in digital health research supported by artificial intelligence. BMC Med 2019 Jul 17;17(1):137 [FREE Full text] [doi: 10.1186/s12916-019-1377-7] [Medline: 31311535]
21. Nebeker C, Bartlett Ellis RJ, Torous J. Development of a decision-making checklist tool to support technology selection in digital health research. Transl Behav Med 2020 Oct 08;10(4):1004-1015 [FREE Full text] [doi: 10.1093/tbm/ibz074] [Medline: 31120511]
22. The Belmont report: ethical principles and guidelines for the protection of human subjects of research. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. 1979. URL: https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html [accessed 2023-12-14]
23. Dittrich D, Bailey M, Maughan D. The Menlo report: ethical principles guiding information and communication technology research. IEEE Secur Priv 2012 Mar;10(2):71-75 [FREE Full text] [doi: 10.2139/ssrn.2445102]
24. Nebeker C, Zlatar ZZ. Learning from older adults to promote independent physical activity using mobile health (mHealth). Front Public Health 2021 Aug 12;9:703910 [FREE Full text] [doi: 10.3389/fpubh.2021.703910] [Medline: 34476229]
25. Behnke M, Saganowski S, Kunc D, Kazienko P. Ethical considerations and checklist for affective research with wearables. IEEE Trans Affect Comput 2022:1-14 [FREE Full text] [doi: 10.1109/taffc.2022.3222524]
26. Nebeker C, Gholami M, Kareem D, Kim E. Applying a digital health checklist and readability tools to improve informed consent for digital health research. Front Digit Health 2021 Jul 15;3:690901 [FREE Full text] [doi: 10.3389/fdgth.2021.690901] [Medline: 34713167]
27. Thompson RA. Vulnerability in research: a developmental perspective on research risk. Child Dev 1990 Feb;61(1):1-16 [FREE Full text] [doi: 10.1111/j.1467-8624.1990.tb02756.x]
28. Tavakoulnia A, Lakes K, Guzman KD, Hayes GR, Cibrian FL, Schuck SE. Designing a wearable technology application for enhancing executive functioning skills in children with ADHD. In: Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers. 2019 Presented at: UbiComp/ISWC '19; September 9-13, 2019; London, UK p. 222-225 URL: https://dl.acm.org/doi/10.1145/3341162.3343819 [doi: 10.1145/3341162.3343819]
29. Díaz-Quesada G, Bahamonde-Pérez C, Giménez-Egido JM, Torres-Luque G. Use of wearable devices to study physical activity in early childhood education. Sustainability 2021 Dec 18;13(24):13998 [FREE Full text] [doi: 10.3390/su132413998]
30. Bagot KS, Matthews SA, Mason M, Squeglia LM, Fowler J, Gray K, et al. Current, future and potential use of mobile and wearable technologies and social media data in the ABCD study to increase understanding of contributors to child health. Dev Cogn Neurosci 2018 Aug;32:121-129 [FREE Full text] [doi: 10.1016/j.dcn.2018.03.008] [Medline: 29636283]
31. Nakib M, Hasan U, Negulescu II. Wearable technology for baby monitoring: a review. J Text Eng Fash Technol 2020 Jul 14;6(4):112-120 [FREE Full text] [doi: 10.15406/jteft.2020.06.00239]
32. Geršak V, Vitulić HS, Prosen S, Starc G, Humar I, Geršak G. Use of wearable devices to study activity of children in classroom; case study — learning geometry using movement. Comput Commun 2020 Jan;150:581-588 [FREE Full text] [doi: 10.1016/j.comcom.2019.12.019]
33. Levin HI, Egger D, Andres L, Johnson M, Bearman SK, de Barbaro K. Sensing everyday activity: parent perceptions and feasibility. Infant Behav Dev 2021 Feb;62:101511 [FREE Full text] [doi: 10.1016/j.infbeh.2020.101511] [Medline: 33465730]
34. Nebeker C, Ellis RB, Torous J. Digital health checklist for researchers (DHC_R). ReCODE Health. 2019. URL: https://recode.health/tools/ [accessed 2024-01-04]
35. Health tools are licensed under a creative commons attribution-non-commercial 4.0 international license 2018-2024. ReCODE. URL: https://recode.health/tools/ [accessed 2024-01-08]
36. Zhang D, Lim J, Zhou L, Dahl AA. Breaking the data value-privacy paradox in mobile mental health systems through user-centered privacy protection: a web-based survey study. JMIR Ment Health 2021 Dec 24;8(12):e31633 [FREE Full text] [doi: 10.2196/31633] [Medline: 34951604]
37. Kelly P, Marshall SJ, Badland H, Kerr J, Oliver M, Doherty AR, et al. An ethical framework for automated, wearable cameras in health behavior research. Am J Prev Med 2013 Mar;44(3):314-319 [FREE Full text] [doi: 10.1016/j.amepre.2012.11.006] [Medline: 23415131]
38. Nebeker C, Lagare T, Takemoto M, Lewars B, Crist K, Bloss CS, et al. Engaging research participants to inform the ethical conduct of mobile imaging, pervasive sensing, and location tracking research. Transl Behav Med 2016 Dec;6(4):577-586 [FREE Full text] [doi: 10.1007/s13142-016-0426-4] [Medline: 27688250]
39. Rampin R, Rampin V. Taguette: open-source qualitative data analysis. J Open Source Softw 2021 Dec 10;6(68):3522 [FREE Full text] [doi: 10.21105/joss.03522]
40. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101 [FREE Full text] [doi: 10.1191/1478088706qp063oa]
41. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]

42. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
43. Li J, Hasegawa-Johnson M, McElwain NL. Analysis of acoustic and voice quality features for the classification of infant and mother vocalizations. Speech Commun 2021 Oct;133:41-61 [FREE Full text] [doi: 10.1016/j.specom.2021.07.010] [Medline: 36062214]

Abbreviations

ECG: electrocardiogram
REDCap: Research Electronic Data Capture
UIUC: University of Illinois Urbana-Champaign

Edited by C Jacob; submitted 24.05.23; peer-reviewed by A Angelucci, J Wang; comments to author 04.08.23; revised version received 26.09.23; accepted 29.09.23; published 08.02.24.

Please cite as:
McElwain NL, Fisher MC, Nebeker C, Bodway JM, Islam B, Hasegawa-Johnson M
Evaluating Users' Experiences of a Child Multimodal Wearable Device: Mixed Methods Approach
JMIR Hum Factors 2024;11:e49316
URL: https://humanfactors.jmir.org/2024/1/e49316
doi:10.2196/49316
PMID:38329785

Original Paper

Usability and User Experience of an mHealth App for Therapy Support of Patients With Breast Cancer: Mixed Methods Study Using Eye Tracking

Carolin Anders[1*], MSc; Preetha Moorthy[2*], MSc; Laura Svensson[1], MSc; Julia Müller[1], MSc; Oliver Heinze[1], Dr sc hum; Petra Knaup[1], Dr sc hum; Markus Wallwiener[3], Dr med; Thomas M Deutsch[3], Dr med; Thao-Vy Le[3]; Lina Weinert[1,4], MSc

[1]Institute of Medical Informatics, Heidelberg University Hospital, Heidelberg, Germany

[2]Department of Biomedical Informatics at the Center for Preventive Medicine and Digital Health (CPD), Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

[3]Department of Obstetrics and Gynecology, Heidelberg University Hospital, Heidelberg, Germany

[4]Section for Oral Health, Heidelberg Institute of Global Health, Heidelberg University Hospital, Heidelberg, Germany

[*]these authors contributed equally

Corresponding Author:
Carolin Anders, MSc
Institute of Medical Informatics
Heidelberg University Hospital
Im Neuenheimer Feld 130.3
Heidelberg, 69120
Germany
Phone: 49 622156 ext 34651
Email: Carolin.Anders@med.uni-heidelberg.de

Abstract

Background: Early identification of quality of life (QoL) loss and side effects is a key challenge in breast cancer therapy. Digital tools can be helpful components of therapeutic support. *Enable*, a smartphone app, was used in a multicenter, prospective randomized controlled trial in 3 breast cancer centers. The app simultaneously serves as a therapy companion (eg, by displaying appointments), a tool for documenting QoL (eg, by enabling data collection for QoL questionnaires), and documentation of patient-reported side effects. The need for digital tools is continually rising. However, evidence of the effects of long-term use of mobile health (mHealth) apps in aftercare for patients with breast cancer is limited. Therefore, evaluating the usability and understanding the user experience of this mHealth app could potentially contribute valuable insights in this field.

Objective: A usability study was conducted to explore how patients with breast cancer receiving neoadjuvant, adjuvant, or palliative outpatient treatment rated their engagement with the app , the user experience, and the benefits of using the app.

Methods: A mixed methods approach was chosen to combine subjective and objective measures, including an eye-tracking procedure, a standardized usability questionnaire (mHealth App Usability Questionnaire), and semistructured interviews. Participants were surveyed twice during the study period. Interviews were transcribed verbatim and analyzed using thematic analysis. Analysis of the eye-tracking data was carried out using the tracker-integrated software. Descriptive analysis was conducted for the quantitative data.

Results: The mHealth App Usability Questionnaire results (n=105) indicated good overall usability for 2 different time points (4 wk: mean 89.15, SD 9.65; 20 wk: mean 85.57, SD 12.88). The qualitative analysis of the eye-tracking recordings (n=10) and interviews (n=16) showed that users found the *Enable* app easy to use. The design of the app, information about therapies and side effects, and usefulness of the app as a therapy companion were rated positively. Additionally, participants contributed requests for additional app features and suggestions for improving the content and usability of the app. Relevant themes included optimization of the appointment feature, updating the app's content regularly, and self-administration. In contrast to the app's current passive method of operation, participants expressed a desire for more active engagement through messaging, alarms, or emails.

Conclusions: The results of this study demonstrate the good usability of the *Enable* app as well as the potential for further development. We concluded from patients' feedback and requests that mHealth apps could benefit from giving patients a more active role (eg, being able to actively document side effects as they occur). Additionally, regular updates of app content could further contribute to encouraging continued use of mHealth apps. Our findings may also assist other researchers in tailoring their mHealth apps to the actual needs of patients undergoing breast cancer therapy.

(JMIR Hum Factors 2024;11:e50926) doi:10.2196/50926

KEYWORDS

mobile health; mHealth; usability; breast cancer; eye tracking; user interface; mixed methods; mobile phone

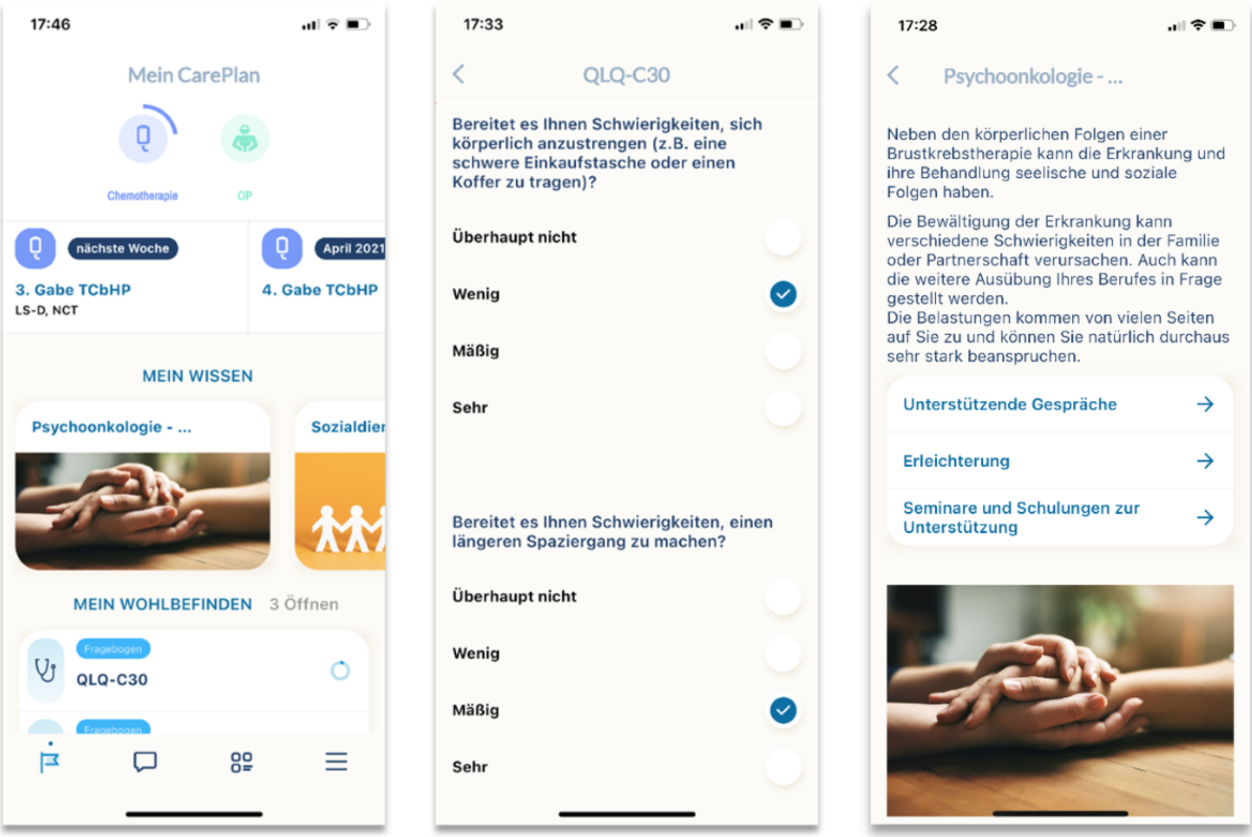
Introduction

Background

Breast cancer is the most common type of cancer detected in women in the Western world. One in 8 women will develop breast cancer during her lifetime. In Germany, there are 69,000 new cases per year [1]. The diagnosis is a drastic event in the lives of those affected. Although the mortality rate has decreased in recent years, processing and dealing with the new life situation is a great challenge for patients and their social environment [2]. At the onset of therapy, patients can have a strong desire for education and information. Therefore, providing patients with reliable sources of information and support services is a major and important task for the treatment team. Digitalization in medicine offers great potential for supporting the exchange of information and communication between patients and health care providers [3-5]. These benefits can be realized through the use of mobile health (mHealth) apps, which can encompass several helpful functions for patients, such as the provision of educational materials, appointment or medication reminders, and diaries. For the cohort of patients with breast cancer, many of these mHealth apps are already available or are in development [6]. This cohort also shows a high readiness for using health technology, indicating that mHealth apps are an appropriate means of support in the early phase of breast cancer treatment.

A recent study by Chen et al [7] also found that remote monitoring of symptoms between clinical visits could not only improve patient-provider communication but also prepare patients for subsequent chemotherapy cycles and support symptom management. Within the joint *Center for Innovative Care* project, a network of 5 university hospitals in southwest Germany, a new mHealth app for patients with breast cancer was developed. This therapy support tool, called the *Enable* app, aims to combine known benefits of mHealth tools with an innovative reactive assessment of patient-reported outcomes (PROs). It was conceptualized as an iOS or Android mobile app for smartphones and developed by members of the research team with the support of software developers. It includes educational content, information about the side effects of therapies and medications, and information about other support services such as psycho-oncology or nutritional counseling in the form of static text and images. A progress bar illustrates the patient's individual therapy status in terms of clinical treatment over time (ie, cycles of treatment). In addition to its role as a therapy companion, the app serves as a measurement tool to systematically record patient satisfaction, health-related quality of life (QoL), and patient-reported adverse events. It monitors the neoadjuvant, adjuvant, and follow-up situations in patients with indications for surgery, chemotherapy, radiation, or systemic therapy with primary or metastatic breast cancer. Figure 1 shows exemplary screenshots of the *Enable* app's start page, the questionnaire display, and information about treatments. As studies have shown that physicians generally underestimate a large proportion of relevant side effects, patients are empowered to report PRO data and side effects directly through the app. In cases of significant treatment-related deterioration, the care team is alerted, and recommendations are sent to the patient. This more relevant treatment information, in turn, helps improve therapy monitoring, treatment quality, and patient satisfaction [8,9].

Figure 1. Exemplary content of the Enable app. Left: On the start page of the app, patients can see their current therapy status, a selection of articles from the "My Knowledge" collection, and upcoming questionnaires; Middle: View of the QLQ-C30 questionnaire to assess the quality of life of cancer patients; Right: Sample article on psycho-oncology with the aim of patient education.



The clinical outcomes of the use of the *Enable* app were studied in the *ENABLE* randomized controlled trial (RCT). Other research questions addressed in the *ENABLE* RCT related to improving patients' adherence to therapy, recognizing and treating critical side effects in a timely manner, and measuring the health-related QoL of different therapy strategies. All study participants underwent QoL assessments at 6 time points during and after adjuvant or neoadjuvant chemotherapy. In the intervention group, an additional short weekly EuroQol Visual Analogue Scale questionnaire was administered. In case of deteriorating results, further screening for side effects was triggered, alerting study staff and enabling immediate contact with the patient to provide support in all phases of breast cancer therapy (reactive PRO assessment). The control group received only the app without the reactive PRO assessment.

The body of scientific literature shows that good usability is an important factor for the success of an mHealth app. More specifically, usability can influence patients' acceptance and adoption of mHealth [10,11]. Usability is defined by Nielsen [12] as a "quality attribute that assesses how easy interfaces are to use." According to the International Organization for Standardization (ISO) 9241-1, usability is the "extent to which a system, product, or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" [13]. Although usability focuses exclusively on the process of using an app or device, user experience involves the users' subjective feelings that result from the use or anticipated use of a system or a product. For the evaluation of mHealth, both concepts are relevant to obtain a comprehensive view of influencing factors [14].

Good usability can help ensure that the app can be used intuitively by patients and health care providers, which in turn improves compliance and increases the effectiveness of the app. A review by Zapata et al [10] demonstrated the importance of adapting mHealth apps to patients' needs. Relevant usability themes of similar apps were, for example, streamlining of the navigation paths, a clearer information architecture, or the desire for personalization [15,16]. Recent research has also shown that usability assessment is an essential step in the mHealth app development process [17,18]. It is important to ensure that the app is easy to use for the target group and provides the desired benefits [12]. However, a systematic review by Jongerius et al [6] showed that only 1 of 29 mHealth apps for breast cancer care that were studied in their work underwent and published a usability assessment. To address the aforementioned requirements and achieve sustainable and effective use of the *Enable* app, the investigation of usability and user experience is indispensable. Therefore, the study presented in this paper intended to gain an understanding of how patients use the app. The aim was to investigate how patients evaluate their engagement with the app, the user experience, and the benefits of using the app. These findings will serve as a basis for further optimization and adaptation of the app to the patients' needs.

A mixed methods approach provides the opportunity to collect, triangulate, and analyze qualitative and quantitative data, allowing for the possibility of interpreting the findings from one research approach (ie, qualitative and quantitative) to explain the data generated from the other research approaches. Furthermore, it allows for the use of a qualitative approach to illustrate quantitative findings or the integration of various research approaches to provide a thorough and comprehensive picture of the study [19,20]. Previous studies [15,21,22] have indicated that interviews and usability questionnaires are prevailing methods used for assessing the usability of mHealth apps. However, there are limited studies regarding the real-time capture of users' visual interactions and the subsequent retrospective analysis of user engagement with mHealth apps through techniques such as eye tracking. Eye tracking, a sensor technology, is used to ascertain an individual's presence and record their real-time eye movements. This approach is also used to assess the usability of technologies by showcasing decision-making processes through the analysis of eye movement patterns [23,24].

Objectives

Developing new mHealth apps can be time-consuming and requires several iterations of testing and evaluation. The *ENABLE* project aims to evaluate both the usability and clinical outcomes of the *Enable* app within the same RCT, which could be a promising approach to speed up development, testing, and planning for further implementation. This paper presents a usability study nested within the *ENABLE* RCT and following a mixed methods approach incorporating the eye-tracking method. The objective of this usability study was to explore how patients with breast cancer receiving neoadjuvant, adjuvant, or palliative outpatient treatment rated their engagement with the app, the user experience, and the benefits of using the app.

Methods

Study Design

This study was designed following a mixed methods approach combining real-world user experience and standardized observations in a laboratory setting. The study took place at the Department of Obstetrics and Gynecology, Heidelberg University Hospital, Germany.

Procedure

Study Population and Recruitment

The study participants were recruited from the intervention and control groups of the *ENABLE* RCT patient cohort (German Clinical Trials Register—DRKS ID: DRKS00025611). The *ENABLE* RCT had the following inclusion criteria: diagnosis of invasive or metastatic breast cancer and planning of neoadjuvant, adjuvant, or palliative therapy in an outpatient treatment setting (indications for surgery or chemo-, radio-, or systemic therapy); minimum age of 18 years; German language skills; and possession of a smartphone with internet access. Owing to technical requirements for eye tracking, patients wearing bifocals were excluded from participation. At study enrollment, patients were asked about their interest in participating in the usability study. All interested patients at the Department of Obstetrics and Gynecology, Heidelberg University Hospital, Germany, received written and verbal information regarding the content and aim of the study and the respective data protection regulations. On the informed consent form, patients could indicate whether they were interested in participating in the usability aspect of the *ENABLE* RCT. Patients who consented to participate in the nested usability study were contacted individually to schedule appointments for participation following a convenience sampling strategy. No reimbursement was provided. The target sample size was 100 questionnaires, 15 qualitative interviews, and 10 eye-tracking studies. Patient recruitment took place from March 2021 to September 2023.

Instruments

The German translation of the mHealth App Usability Questionnaire (MAUQ) [25] was chosen to quantitatively assess the usability of the *Enable* app [26]. The MAUQ enables the usability assessment of mHealth apps from the user's perspective. The MAUQ stand-alone version was formulated to evaluate 3 constructs of usability—ease of use, interface and satisfaction, and usefulness—as well as the overall usability score for the app through descriptive statistics. Each of the items of the MAUQ is rated on a Likert scale ranging from 1 (*strongly agree*) to 7 (*strongly disagree*), with the overall score ranging from 0 to 100. In addition, the questionnaires were complemented with a set of questions developed by the authors. Newly added questions concerned the use of other mHealth apps, smartphone ownership, sociodemographic information, and a free-text field to be able to describe the study sample more precisely. The target sample size was 100.

In addition to the questionnaire, open-ended, semistructured, and guide-based interviews with patients were conducted to explore their perspectives on the usability of the *Enable* app. The interviews were conducted by 2 female researchers (CA and LW) with a professional background in health services research and implementation science. Both researchers have profound experience with qualitative interviewing. The interview guide (Multimedia Appendix 1) was developed by a team of health services researchers (LW and JM) based on an extensive literature review and recommendations from the app developers. Afterward, the interview guide was pretested. This study is reported according to the COREQ (Consolidated Criteria for Reporting Qualitative Research) guidelines (Multimedia Appendix 2 [27]).

Furthermore, to objectively assess how patients interact with the app and identify potential usability issues, an eye-tracking study was conducted. The eye-tracking study was conducted by a usability expert (PM) and a team of health services researchers (CA and LW). A total of 5 tasks were formulated for the eye-tracking study (Multimedia Appendix 3): app log-in, filling in a questionnaire, searching and reading an article, and logging out from the app. To determine the comprehensibility of the tasks, the duration of the study, and the workings of the *Enable* app, 2 pilot tests were conducted. Following the pilot test outcome, the eye-tracking studies were carried out for 60 minutes with each participant, including the eye tracker setup and the retrospective interview.

The chosen mixed methods approach is designed to systematically collect, cross-validate, and analyze both qualitative data (derived from semistructured interviews and eye tracking) and quantitative data (obtained through the MAUQ). The inclusion of the eye-tracking method in the usability study enriches the capacity to integrate subjective and objective metrics. The qualitative aspect of the eye-tracking analysis enhances the understanding of the user's app perception within the context of individual interactions and app usability. Simultaneously, semistructured interviews enable an assessment of the practicality of integrating the *Enable* app into daily routines. In contrast, the quantitative data derived from the questionnaire provide precise metrics related to usability measurements.

Hence, the mixed methods approach investigates the *why* and *how* aspects through qualitative inquiry, supplementing conventional quantitative and visual data analyses. The fusion of direct observations of user interactions with the app, poststudy retrospective interviews, semistructured interviews, and the usability questionnaire collectively supports the contextualization and comprehensive interpretation of the gathered data.

Data Collection and Analysis

Quantitative Measures

The MAUQ and sociodemographic questionnaire were mailed twice to all patients after inclusion in the RCT. Data collection lasted from May 2021 to October 2022. Study data were collected and managed using REDCap (Research Electronic Data Capture; Vanderbilt University) tools [28] hosted at Heidelberg University Hospital. REDCap is a secure, web-based software platform designed to support data capture for research studies. After completion, all data were exported from REDCap to the R statistical software (version 4.0.4; R Foundation for Statistical Computing). All data were checked for completeness and analyzed by study team members. A descriptive analysis of the questionnaires was performed using R. Means and absolute and relative frequencies were calculated.

Qualitative Measures

Interviews were conducted after participants had used the app for 8 weeks. The interviews took place partly face-to-face at the clinic and by telephone in consideration of current guidelines for preventing infections with SARS-CoV-2 (ie, participants and researchers wore appropriate masks and distance was kept at all times). Nonparticipants were not present during the interviews. No relationship with participants was established before taking part in the study. No repeated interviews were conducted. No field notes were taken. All interviews were audiotaped, pseudonymized, and transcribed verbatim. Transcripts were not returned to participants for verification. Data were transcribed, managed, and analyzed using MAXQDA Standard 2020 (version 20.4.1; VERBI GmbH). After 16 interviews, data saturation was discussed among the researchers. As no new themes emerged in later interviews, the researchers agreed that data saturation had been reached and no additional interviews were necessary. After completion of data collection, thematic analysis of the data was conducted independently by 2 researchers (CA and LW) [29]. First, the researchers reviewed the transcripts independently and identified themes from the literature and the interview guide and inductively from the data. Second, discrepancies were discussed in iterative cycles until a consensus on themes and the final coding scheme was reached. All themes were organized into main themes and subthemes. Each theme was clearly defined by a quote from the interview transcripts (Multimedia Appendix 4). Quantitative and qualitative data were analyzed separately.

For the eye-tracking data collection process, an assigned room where the Tobii Pro Nano (Tobii AB) was installed at the hospital was used; the Tobii Pro Nano is an eye-tracking device specifically designed for small screens, including smartphones. This hardware features a sampling rate of 60 Hz, measures 17 × 1.8 × 1.3 cm, and includes a USB type-A connector. The Tobii Pro Nano was securely affixed to the mobile phone stand, and the *Enable* app was installed on a smartphone. To facilitate data capture, both the smartphone and the eye tracker were connected to a laptop running the Tobii Pro Lab software (version 1.194) via USB cables. For the purposes of this study, both an Android device (Samsung Galaxy 10, Android version 11) and an iOS device (iPhone 11, iOS version 14.6) were available to users. The choice of smartphone was contingent upon the user's preferred operating system. The eye tracker recorded the participants' interactions with the *Enable* app, such as task completion time, participants' navigation, gaze plots, and heat maps [30-32]. A heat map was used when fixation duration data were collected [30,31], and a gaze plot was used when location of eye movement data were collected [33,34]. For this study, after the completion of tasks, the study moderators composed post hoc questions pertaining to the interactions, participants' experiences, and usability issues observed during the procedure. The post hoc questions were discussed with the participants in a short debrief. The debriefing sessions were held to gather direct feedback from participants after interacting with the *Enable* app, allowing for a deeper understanding of the participants' behavior and interaction with the app. Through these debriefing sessions, participants could provide context and commentary on their behavior and interaction [35]. Engaging users using post hoc questions, such as using images or live content from recorded sessions, allowed for a better understanding of the real-life context with minimal disruption as it facilitated the recall of situational information prompted by data, sound, or visual imagery.

The data analysis was based on the recordings of the study sessions concurrent with the eye movements of participants. The retrospective analysis involved transcribing participants' feedback from the audio recordings obtained during the debriefing sessions. Data analysis also included the completion of predefined tasks by the participants, task completion time, and completion status of the tasks. The analysis focused on task performance analysis and the problem analysis of eye-tracking metrics and participants' feedback.

Ethical Considerations

The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Heidelberg University Hospital (S-685/2020). All participants provided written informed consent for taking part, audio recording of the interviews, and video recordings during the eye-tracking procedures. Confidentiality and anonymity were ensured throughout the study. The data was protected against unauthorized access. No incentives or compensation was provided to participants for study participation.

Results

Overview

The MAUQ was sent to 165 patients recruited from the *ENABLE* RCT. The response rate was 63.6% (105/165) for the MAUQ at week 4 and 56.4% (93/165) for the MAUQ at week 20. A total of 105 questionnaires for the MAUQ at week 4 (including sociodemographic data) and 93 questionnaires for the MAUQ at week 20 were analyzed. In total, 16 patients were recruited for the interviews, and 10 were recruited for the eye-tracking procedure. The mean duration of the interviews was 25 (SD 7.34) minutes.

Sociodemographic Characteristics

The sociodemographic data of the participants in the *ENABLE* usability study are shown in Table 1, and additional characteristics of the participants regarding smartphone and app use are shown in Table 2. The mean age of all participants (n=105) was 51.3 (SD 10.9) years.

Table 1. Sociodemographic characteristics of the participants.

Characteristic	Interview participants (n=16), n (%)	Eye-tracking study participants (n=10), n (%)	Questionnaire participants (n=105), n (%)
Gender			
Woman	16 (100)	10 (100)	105 (100)
Age group (y)			
<30	2 (12.5)	2 (20)	2 (1.9)
30-40	2 (12.5)	1 (10)	16 (15.2)
41-50	6 (37.5)	3 (30)	32 (30.5)
51-60	4 (25)	2 (20)	33 (31.4)
61-70	1 (6.3)	0 (0)	16 (15.2)
71-80	1 (6.3)	0 (0)	6 (5.7)
Education			
Academic degree	9 (56.3)	6 (60)	37 (35.2)
High school education	0 (0)	0 (0)	13 (12.4)
Lower or intermediate secondary school	5 (31.3)	4 (40)	54 (51.4)
Prefer not to say	2 (12.5)	0 (0)	1 (1)
Employment			
Employed	11 (68.8)	9 (90)	66 (62.9)
Unemployed	0 (0)	0 (0)	17 (16.2)
Studying or vocational training	1 (6.3)	1 (10)	1 (1)
Retired	2 (12.5)	0 (0)	18 (17.1)
Prefer not to say	2 (12.5)	0 (0)	3 (2.9)

Table 2. Additional participant characteristics on smartphone and app use.

Characteristic	Interview participants (n=16), n (%)	Eye-tracking study participants (n=10), n (%)	Questionnaire participants (n=105), n (%)
Use of smartphone (y)			
≤10	9 (69.2)[a]	5 (62.5)[b]	56 (53.3)
>10	4 (30.8)[a]	3 (37.5)[b]	44 (41.9)
Prefer not to say	0 (0)	0 (0)	5 (4.8)
Use of other mHealth[c] apps			
Yes	4 (28.6)[d]	5 (50)[e]	33 (31.4)
No	10 (71.4)[d]	5 (50)[e]	71 (67.6)
Prefer not to say	0 (0)	0 (0)	1 (1)
Frequency of app use			
Daily or several days a week	5 (45.5)[f]	3 (37.5)[b]	0 (0)
Once a week	5 (45.5)[f]	4 (50)[b]	0 (0)
Once a month or less	1 (9.1)[f]	1 (12.5)[b]	0 (0)

[a]n=13.
[b]n=8.
[c]mHealth: mobile health.
[d]n=14.
[e]n=10.
[f]n=11.

Quantitative Measures

The MAUQ [25] was used to collect quantitative data on the usability of the *Enable* app. The data were collected at weeks 4 and 20 starting from the baseline of the study. Quantitative data gathered from the MAUQ were analyzed using descriptive statistics. Only complete questionnaires for which the MAUQ score could be calculated were evaluated. Hence, 32.4% (34/105) of incomplete questionnaires collected at week 4 and 29% (27/93) of incomplete questionnaires collected at week 20 were excluded from the analysis. According to Zhou et al [25], the usability of an app is calculated based on the average of the responses to all statements. The higher the overall average, the higher the usability of the app. In this study, the overall usability scores for weeks 4 and 20 were 89.15 (SD 9.65) and 85.57 (SD 12.88), respectively. The mean for each of the subscales from week 4 to week 20 was also calculated and is presented in Table 3. The results show that the usefulness score declined over time from week 4 (80.89) to week 20 (77.33). In addition, the *interface and satisfaction* score also decreased but not as much as that of the *usefulness* subscale. The *ease of use* score, in contrast, remained constant at both weeks 4 and 20.

Table 3. Quantitative analysis of the mHealth App Usability Questionnaire and subscales.

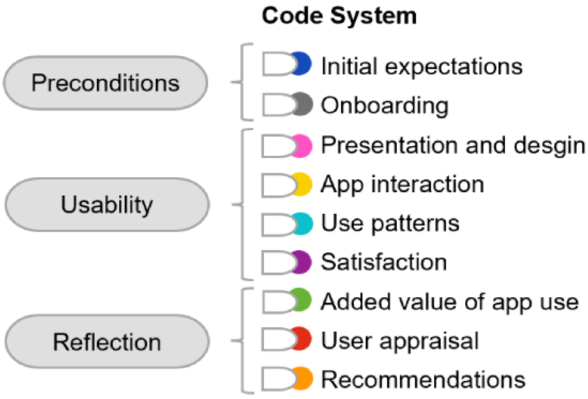
Time point	Overall, mean (SD)	Ease of use, mean (SD)	Interface and satisfaction, mean (SD)	Usefulness, mean (SD)
Wk 4 (n=71)	89.15 (9.65)	92.41 (11)	91.6 (10.15)	80.89 (15.67)
Wk 20 (n=66)	85.57 (12.88)	92.27 (11.91)	88.23 (13.6)	77.33 (16.63)

Qualitative Measures

Interviews

In total, 527 text passages were coded during the interviews. A total of 9 themes and 60 subthemes were identified, each of which could still be categorized under the superordinate themes of preconditions for app use, usability, and reflection. These themes are summarized in Figure 2.

Figure 2. Overview of identified themes.



Preconditions

Initial Expectations

As an opening question in the interview, patients were asked about their initial thoughts when they first heard about the *Enable* app. The most frequently mentioned expectations were related to the quality of information. Patients expected the information in the app to be updated regularly, understandable, and in line with the latest research. Another expectation was that the app would provide contemporary therapy support and be perceived as modern, including replacing printed brochures. Patients expected the app to provide guidance over the course of therapy, contact options, and easy access to relevant information. Approximately half of the participants had neutral expectations for the app:

> *Yes, I already thought that it [the app] would support me through everyday life and therapy, that I can also use it to organize myself a bit.* [Interview 9; transcript position 2]

Onboarding

The aspect of the onboarding process was not part of the interview guide. However, individual participants reported that they felt well supported by the study staff at the beginning of their app use. Even if they were initially overwhelmed by the app or experienced technical difficulties, participants expressed that they received the necessary support and were able to handle the app:

> *Oh dear, now I have to dig into yet another app. I don't know if I can handle it. But the more I got a grip on it, the better it worked.* [Interview 3; transcript position 2]

Usability

Presentation and Design

Patients were asked to describe their impressions of specified design aspects. Overall, patients were content with the color scheme and perceived it as pleasant without being boring or flashy. For some patients, this cheerful esthetic contributed to a sense of joy when using the app and encouraged them to use it more often:

> *[The design is] very friendly. Very beautifully visualized. I always enjoy opening the app. It is also well designed, you always have the feeling that it is not draining in any way, it is more playful with all these images and visualizations. I find it very very clear.* [Interview 11; transcript position 34]

Regarding the layout of the written information, patients appreciated how the most relevant parts were highlighted through the positioning of boxes. The font size, design, and structuring of the information were seen as adequate. The selection of accompanying images was described as empathetic and not too explicit. The app included a personalized visual representation of the therapy progress. This display was also rated as clear and useful. Patients explained that the presence of this display motivated them:

> *I found this progress bar, which shows me how long I will be in therapy for, especially beautiful. It...motivated me, showing me that there is always a path forward and that the therapy will soon be over.* [Interview 13; transcript position 43]

App Interaction

Regarding usability, 4 important aspects emerged while interacting with the app. Neither the positioning nor the design of the app icons were perceived as entirely intuitive. However, patients grew acclimated to the icons, and thus, this did not further impede usability:

> *Yes, the icons that were down in this bar. In the beginning, I didn't know the meaning of each icon. But when I took a closer look once, I knew it for the next time.* [Interview 7; transcript position 39]

Log-in and log-out procedures were described as easy and quick and did not pose any problems for the patients in this study. Most patients had no issues working with the app's structure. They could easily navigate within the app and were able to find what they were looking for:

> *I found my way around the app really quickly. I haven't tried all the features yet, I haven't clicked on everything because I don't need it all. But I have always been able to find the things that I wanted very quickly, and everything is right there when you click on it.* [Interview 11; transcript position 44]

Overall, patients liked using the app as it enabled them to access information *on the go*. Patients described that having their smartphones with them at all times allowed them to read information given the absence of other electronic devices such as laptops or tablets. However, a few patients mentioned the additional benefits of having a web-based version of the *Enable* app.

Use Patterns

This code encompasses descriptions of how and when patients used the app. Most patients experienced changes in the frequency of app use. In the beginning, they used the app often, and some patients used it multiple times per day:

> *In the beginning, shortly after my diagnosis, I had a lot of questions—for my physicians, how things work and so on. During this time, it (the app) really helped me a lot.* [Interview 10; transcript position 24]

Over time, use declined. This development was mostly due to lower demand for support and information as patients became used to therapy proceedings. Patients also used the app less as they felt that they had already read everything.

After this initial phase, patients reported using the app whenever they needed to look up appointments, had free time (eg, during waiting times before physician's appointments), had or experienced new side effects from their treatment, were prescribed new medications, or were prompted by push notifications:

> *I always used it shortly before my [chemotherapy] appointments. Or when I had questions regarding diet and exercise. And sometimes there were questionnaires I had to fill in. And yes, as soon as the app said "there is news," I opened it...And to look up times for my appointments.* [Interview 10; transcript position 12]

Satisfaction

Patients praised the general aspects of the app and liked the idea of having a digital tool accompanying them throughout their therapy; for example, the app provides a good overview of relevant topics, especially at the beginning of the disease. Except for 1 interviewee, all participants (15/16, 94%) would recommend the app to others:

> *...because it really provides a great overview...because so many aspects are addressed. Not only the type of therapy, but also just different things about cancer. Especially at the beginning these keywords—Yes, these terms in the boxes from tiredness to fatigue and polyneuropathy and different things.* [Interview 7; transcript position 47]

Reflection

Added Value of App Use

When asked about the concrete benefits of the app in everyday life, several aspects were mentioned. The most important aspect for the participants was the information on therapies and side effects, which was perceived as helpful, especially in the initial phase of therapy. The quality of the information was praised as the app's information was considered understandable and its origin was considered reliable:

> *You feel informed, you feel—that gives you a form of security, because you say to yourself: Well, if I have the information from here [the app], then it was completely clear to me: I don't have to look it up again. That's true for me because these are reliable information providers who wrote this.* [Interview 12; transcript position 81]

The comprehensibility and language level were also perceived as adequate. Statements on the amount of information were heterogeneous according to individual information needs. However, the amount of information was predominantly perceived as sufficient in the context of the app. Furthermore, the appointment display, contact information, and progress bar were found to be helpful and clear. With regard to the contact information provided in the app, the fact that it was easy to find was rated positively.

Some patients reported that the questionnaires in the app gave them a positive feeling as they reflected on their condition and (in the intervention group) it was experienced positively that the questionnaires were read by the study staff and that staff could react proactively to them if necessary. Overall, patients perceived the app as a good therapy companion that guided and supported them through the various phases of the disease and therapies.

User Appraisal

Users' opinions on the existing functions and features of the app were added to this category. Most patients complained about the appointment display as the date and time on the app did not always correspond to the actual clinic appointments (eg, in the case of last-minute postponements):

> *It's a shame that the—I don't know how the appointments displayed in the app, how often those are matched. I've had frequent differences there. Especially when appointments had to be postponed, the chronology was no longer correct for me.* [Interview 9; transcript position 2]

Regarding the quantity of information, some patients wished for more in-depth information or links to other information platforms. It was remarked that the amount of information available varied depending on the topic. Regarding the quality of the content, patients noted that the listed side effects or drugs were grouped differently. For instance, the patients were unable to locate paclitaxel as it belonged to the taxane drug class. In total, 12% (2/16) of the patients in particular perceived errors in spelling, punctuation, and grammar as distracting. The presentation of the contact information on the app was described as difficult to find, especially in emergencies. The additional pop-up notifications of the app updates were rated negatively as it was not apparent to the user what exactly was new in the app. Furthermore, respondents ascertained that the menu navigation was not intuitive enough and, therefore, needed to be improved.

Recommendations

Statements about features of the app that are not yet offered were classified as recommendations or wishes. Most wishes were mentioned in relation to the appointment display. Patients would like to have additional information about appointments, such as directions, a reminder function, the ability to export appointments from the app to private calendars (eg, Google or Outlook calendars), or the ability to make appointments directly from the app. The desire for self-administration (ie, areas such as appointments, questionnaires, or therapy progress that can be actively managed by the patient) was also frequently voiced. In addition, some patients wished to view the questionnaires that had already been completed to be able to monitor their condition over the course of therapy:

> *With the exception of filling in the questionnaires, you can't work with the app yourself. Therefore, if you could manage things in the app by yourself, then of course I would think that would be great.* [Interview 9; transcript position 2]

Patients also wanted the content of the app to be updated, expanded, and adapted to new scientific findings. In this context, there was a desire for more explanatory videos to be included in the app. Patients also suggested that the app should offer more information about current and upcoming clinical trials for patients with breast cancer. To see what content in the app has already been read, patients suggested a read status, where content that has already been read is highlighted. Emergency contacts should also be highlighted in the app to make them easier to find, for example, by displaying them on the home page:

> *Especially the emergency numbers, I don't know how to get something like that into the app, but that might be an idea, because I've been looking a lot for the right contact person. Maybe that would also be something that you could highlight a little bit or display as a button.* [Interview 4; transcript position 32]

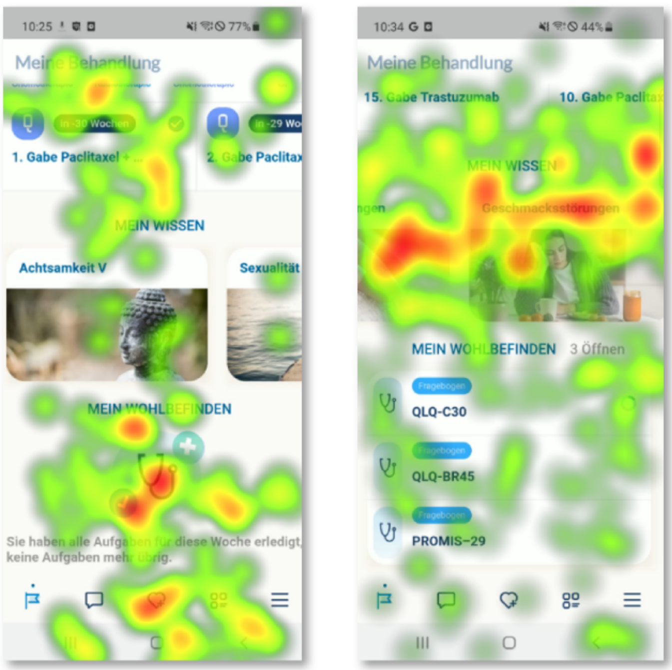
To be able to find certain topics more quickly, the need for a search function was mentioned several times. Furthermore, to improve the readability of the content, patients would like to be able to adjust the font size. It was also suggested that the app could be used on other devices, such as tablets.

Eye Tracking

Overview

The analysis of the data collected from the eye-tracking recordings as well as the retrospective interviews showed that the participants found the app easy to use. We observed that most participants completed the given tasks, although the time taken to complete a few tasks proved to be challenging. On the observations and retrospective interviews during the eye-tracking study, we discovered 3 noticeable patterns related to the design and layout of the app, content and navigation through the app, and additional features the participants would like to have in the app. Figure 3 shows exemplary heat maps from the eye-tracking analysis. The data collected during the task performance, such as the task completion rate and task completion times, are provided in Multimedia Appendix 5.

Figure 3. Heat maps from the eye-tracking analysis.



Design and Layout of the App

Many of the participants had problems understanding and interpreting the icon at the bottom of the screen. The eye-tracking data showed fixations at the bottom of the screen while the patients clicked each of the icons displayed to view the content of the page. Patients expressed a preference for finding the most important information, such as appointment dates and the progress of a questionnaire, at the top of the screen. This finding indicates that patients expect important information to be located at the top of the app's layout. Furthermore, the patients actively mentioned that the retrievability and visibility of the questionnaire were low. Although the questionnaires were available on the home screen of the app, patients believed that the questionnaires were available on the menu. In contrast, patients found the overall layout of the app to be acceptable.

Content and Navigation of the App

Regarding the content of the app, patients showed more interest in the titles of the articles (eg, topics such as symptoms or side effects) than in the images displayed. When asked during the retrospective interviews, patients mentioned that they did not pay attention to the images as they provided no information on what the article was about. Patients preferred to read the title of the article as it gave them information about its content, as shown by the red areas of the heat maps in Figure 3. Moreover, many participants explored the app to find the right information or icon to perform the tasks. However, this correlates with how frequently patients used the app. During the interviews, some patients said that they used the app frequently, for example, every day, to read articles on side effects or symptoms and fill out questionnaires regularly, whereas some patients used the app frequently at the start phase of the *ENABLE* RCT and later minimized the use of the app except to fill out questionnaires. The data showed that patients also had issues navigating through the app, especially related to the task of finding a specific article. Analysis of the recorded data of the participants' navigation and gaze plots from the Tobii Eye Tracker showed that patients looked for a search function. Most patients clicked the menu icon; however, they did not proceed further to find the article nested under the *Symptoms* category on the menu. In addition, some patients searched for the article on the start page along with the other articles already displayed.

"Would Like to Have" (Wishes)

Participants identified a need for additional features in the *Enable* app as a consequence of the challenges they encountered during the eye-tracking study tasks. These suggested features were considered as *nice-to-have* options and were based on the specific problems faced by the participants during the study. The first was the availability of an option to mark an article as a favorite and be able to view the favorite article on the start page. Second, patients desired to have more articles or information about the symptoms and side effects of breast cancer and its treatments. Third, the icon currently representing contact information for health care providers (*My Care Team*) was misleading. Patients preferred to have another icon that indicates contact or communication as this would enable them to contact the study nurses more quickly. Finally, a search option was suggested by all participants.

Discussion

Principal Findings

The aim of this study was to investigate how patients with breast cancer rated their engagement with the *Enable* app, the user experience, and the benefits of using the app. In particular, the design, layout, navigation, content, and requests for new features were identified as important outcomes of interest for evaluating the app and further improving it to meet user needs. The interviews provided valuable suggestions for optimizing the app and the implementation process. The design and color scheme were rated very positively overall. In terms of use patterns, it was noticeable that the frequency of app use decreased over the therapy period.

Patients found the app easy to navigate. However, there was some criticism that the menu icons were not intuitive enough, especially at the onset of use. Perceived benefits were discussed extensively in the interviews. Patients found the information on therapies and side effects very useful. The appointment display and progress bar were also found to be helpful and motivating. At the same time, the appointment display was most often criticized, and it was the feature for which there were the most recommendations for change (eg, to be able to manage appointments autonomously in the app or set reminders). In terms of content, it was mentioned that there was a lot of information on some topics and not enough on others. Patients also wanted more content updates within the app (eg, on current topics such as the COVID-19 pandemic) and a search function to access specific content.

A study by Ansaar et al [36] showed that nearly 78% of all usability evaluation studies in their systematic review used a questionnaire-based method. However, using mixed methods approaches in usability evaluation studies provides benefits such as the possibility to balance the advantages and disadvantages of the different methods. Moreover, by applying the mixed methods approach, both subjective and objective aspects can be combined to assess usability [36]. In many aspects, such as the *navigation*, *recommendations*, and *perceived benefits* codes, the results of the different survey methods support each other. However, the interviews and eye-tracking study sometimes provided different findings. For example, the importance of images within the app was positively highlighted in the interviews. In contrast, the eye-tracking study and retrospective interviews revealed that images played a subordinate role for patients, with titles being more important for finding relevant content in the app. Although participants reported in the interviews that they were able to navigate easily within the app and find the content they were looking for, we observed in the eye-tracking study that there were difficulties with finding specific content. Furthermore, the interview inquiries primarily centered on the practicality of incorporating the *Enable* app as a follow-up intervention in daily life. Meanwhile, the use of eye-tracking technology allowed for direct, real-time observation of user behavior while engaging with the app through task performance. Despite patients reporting the ability to regularly use the app without difficulty, the eye-tracking study's direct observation unveiled valuable insights into their actual use patterns within their everyday routines. In this context, disparities between the results obtained from the 2 methods emerged, possibly stemming from users' lack of awareness regarding any issues until they were prompted with specific inquiries.

Comparison With Prior Work

Our results on the MAUQ indicate good usability. The results for the total scale showed that usability decreased from weeks 4 to 20. A decrease in usability over time has also been observed in previous studies [37-39]. Possible explanations for this decline in our study can be found in the interviews, indicating that the extent of app use also decreased over the course of therapy. Patients found the app to be particularly advantageous at the start of their therapy because of their great need for information. However, as they gained more knowledge about the disease and its treatment, their demand for information decreased. In addition, patients reported that the app lost its appeal once all the available articles had been read, often leading to a desire for new content to be added. Patients also expressed a need for additional features or improvements as they continued to use the app. As a result, the decrease in the app's usability score could be attributed to patients perceiving it to be less useful after an extended period of use owing to the lack of content updates and unmet desires.

Looking more closely at the subscales of the MAUQ, *usefulness* had the lowest score compared with *ease of use* and *user interface and satisfaction*. These items assess whether the app is helpful and useful for patients' health and well-being. This relationship is also apparent when looking at the *usage patterns* category from the interview analysis. It appears that patients are less likely to use the app because of the lack of new content. This is consistent with the findings of other studies on mHealth apps for patients with breast cancer [16,40,41]. As an implication for similar apps for other chronic conditions, it seems important to update the app content on a regular basis to provide patients with an incentive to continue using the app as well as strengthening patients' satisfaction and information needs. Consistent with the findings from the interviews and eye-tracking study, only the *ease of use* subscale remained almost stable over the duration of app use.

In the context of other usability studies on mHealth apps, the importance of paying more attention to the user group of older adults is emphasized. The different age ranges of patients and the different levels of technical affinity for older patients are mentioned as possible factors causing usability problems. Some studies emphasize that these factors are often overlooked and need to be considered when developing mHealth apps [42,43]. In our study, these aspects were less evident. With an average age of 51 years, our study participants do not represent a predominantly older population but are close to the German population average for women, which is 46 years [44]. In contrast, the study participants were also far below the average age of 64 years for patients with breast cancer. Therefore, further research on app development and usability with a focus on older participants should be conducted to more adequately represent the typical population of patients with breast cancer.

Considering the preferred device for using the *Enable* app, most participants were content with using the app on their smartphones. However, there were isolated requests to be able to increase the font size of the content and use the app on a larger-scale device, such as a tablet or PC. This issue was also mentioned by participants in a usability study by Jessen et al [45], in which an mHealth app for self-management of chronic diseases was evaluated.

Although the onboarding process was not part of the interview guide, some patients actively recalled how they were introduced to the app as well as how they perceived the technical onboarding process. The patients did not experience issues with these steps and reported being content with the process, mostly because of the strong support of the study team. Previous research has pointed out that complex registration and log-in procedures can be perceived as especially cumbersome by patients and can lead to stopping app use [46-48]. Our study identified the strong interpersonal connection with and continued support from the study team as a positive influence on the perceived ease of onboarding. This support took place in the context of a research study and is not viable in a real-world implementation. However, the issue of technical support arose exclusively during the qualitative interviews. We did not collect any quantitative data on this topic. Thus, further streamlining of the onboarding process while being mindful of health care workers' limited time resources should be an area for future research.

Strengths and Limitations

The chosen mixed methods approach can positively support the further development of the app. The expansion of the classic social science method spectrum to include technical methods such as eye tracking made it possible to combine the subjective patient perceptions reported in interviews and questionnaires during everyday use with objective measurements under laboratory conditions.

However, the integration of qualitative results and the objective measurement from the eye-tracking procedure introduced discrepancies. As noted previously, interviewees appreciated the use of images in the app, whereas eye-tracking results showed that more time was spent on the article titles than on the images. Another example is that the interviews and the questionnaire produced good ratings of usability, but the eye-tracking study showed that patients found it difficult to find defined content. Although difficult to analyze, these discrepancies are common in mixed methods studies [19]. In our study, these discrepancies could be explained by methodological differences. For example, reading a title naturally takes longer than glancing at an image, leading to a long fixation time. Therefore, this result does not allow for the conclusion that titles are more important than images. Here, the qualitative interviews were helpful in interpreting this finding. Regarding the second example—overall good usability scores in comparison with eye-tracking times—several interpretations appear plausible. First, it is possible that social desirability led patients to rate the usability more favorably in both the interviews and the questionnaire. Consequently, the objective measure via eye tracking revealed that usability was worse than in subjective measures. Second, the setting of the eye-tracking procedure (eg, unusual or uncomfortable sitting position, being observed by ≥2 researchers, or using a different device) could have led to changed patterns in (app use) behavior. Although we acknowledge these discrepancies, we conclude that the mixed methods approach and its results deepened the understanding of the studied topic and produced valuable insights, with discrepancies leading to vigorous and fruitful discussions among the researchers.

However, the generalizability of the study results is limited by several factors. To ensure that patients with lower digital health literacy could participate in the quantitative data collection without constraints, we decided to use printed surveys sent by mail. Patients returned them at their discretion. Hence, it cannot be verified whether the surveys were filled out at the correct time. In addition, some values were missing from the returned surveys, and manual data entry could have led to documentation errors. Incomplete or inconclusive questionnaires had to be completely excluded from the analysis as it was not possible to calculate the score. Although all necessary steps were taken to ensure high-quality and reliable data (eg, data entry was always checked by another researcher), using a web-based survey instead of a printed survey could have made data collection easier, faster, and more reliable. These trade-offs have to be balanced in future research projects.

This study population contained an above-average proportion of academics, especially among the subgroups of interviewees and eye-tracking study participants. This should be taken into account when interpreting these results. A systematic review by Niazkhani et al [49] showed that patients with lower educational attainment and limited health literacy were less likely to intend to use an electronic patient health record and were more likely to use it ineffectively. Moreover, previous experience with computers or health technology has been associated with increased acceptance, and acceptance increases with higher education [7]. Although these results refer to electronic health records, they indicate that this aspect should be further investigated in future studies. Given the median age at breast cancer diagnosis of 64 years and the relatively younger median age of this study cohort, conclusions from this study must be interpreted with caution as they may not represent the views and digital literacy of older women with breast cancer [50].

The *Enable* app was developed specifically for patients with breast cancer. Consequently, our study sample included only female patients with breast cancer. Some of our results and recommendations may have limited generalizability to other patient populations. Nevertheless, we think that aspects such as the relevance of content updates, the accuracy of displayed appointments, or the intuitiveness of the app navigation might also be relevant beyond the target group. This should be verified in further research.

As part of the *ENABLE* RCT, reasons for dropping out were documented where available. These reasons were examined to see whether there were any indications of usability problems. A small proportion of the included study participants in the RCT dropped out because of physical exertion or feelings of being overwhelmed by the app. In this respect, further research is needed to understand how patients in later stages of the disease or with greater disease burden perceive the usability and benefits of the intervention. Furthermore, mHealth apps should be designed to be usable and helpful for these patient groups as well, especially in the context of patients living with cancer. As the mean age of participants in this study was relatively low, it can be assumed that there is a risk of selection bias. It is possible that younger patients decided to participate in the study and use the app because of a higher affinity for smartphones [11].

In addition, using the eye-tracking device led to further limitations. Potential participants in the eye-tracking study had to undergo an additional screening process to exclude patients wearing bifocal glasses. Although patients were recruited for the study, this criterion did not allow us to cast a wider net for the participant recruitment process. Furthermore, we also had the challenge of asking patients to sit still so that the eye-tracking data could be captured without breaks. However, this request is generally against the natural way in which users sit and interact with mobile devices. Another point to note is that the execution of the tasks on the app by the patients was deviated as the tasks were presented on paper and this retracted some of the gaze points of the patients. This is, in general, a common problem when tasks are not integrated into mobile apps during development for testing purposes.

Conclusions

The results of this usability study demonstrate good usability of the studied app and potential for purposeful development. The design and color scheme were rated very positively overall. However, there was some criticism that the menu icons were not intuitive enough, especially at the onset of use. Noticeably, the frequency of app use decreased over the therapy period. Perceived benefits of the app were information on therapies and side effects. The appointment display and progress bar were also found to be helpful and motivating. Still, participants offered recommendations for changing the appointment display (eg, to be able to manage appointments autonomously in the app or set reminders). In terms of content, it was mentioned that there was a lot of information on some topics and not enough on others. Patients also wanted more content updates within the app (eg, on current topics such as the COVID-19 pandemic) and a search function to access specific content. The interviews and eye-tracking study revealed valuable suggestions for improvement as well as requests for additional app features. An important point is that the app currently provides information to the patient mainly passively. The patients' wishes indicate that the app needs to be further developed so that they can actively enter information into the app and work with it. The overlap between decreasing usability and decreasing usefulness also suggests that the app needs to be regularly updated with new content to maintain its usefulness over time. These findings will be incorporated into the further development of the *Enable* app. We concluded from patients' feedback and requests that similar mHealth apps could benefit from giving patients a more active role (eg, being able to actively document side effects as they show up instead of being prompted to do so). In addition, regular updates to app content (eg, adding new informational pieces) could further contribute to and, thus, encourage the continued use of mHealth apps.

Acknowledgments

This study was funded by the Baden-Württemberg (Germany) Ministry of Science, Research, and Arts under reference 42-04HV.MED (19)/15/1 as part of the Center for Innovative Care project. The authors would like to thank all patients in this study for their participation and valuable input. They would also like to thank Corinna Hartmann for her support in recruiting patients for this study. Thank you to Bendix Harms for his support with the quantitative analysis.

Data Availability

The data sets used and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

CA, PM, and LW collaborated on the draft of the manuscript, PK contributed. MW is the principal investigator of the *ENABLE* project. LW, JM, and PM were responsible for the study design and protocol, and OH contributed. MW and TMD prepared and submitted the study protocol, and LW, JM, and PM contributed. LW, JM, and PM collaborated on the construction and testing of the interview guides and the quantitative data collection tools, and LS supported the finalization of these instruments. CA and LW conducted the interviews, analyzed transcripts, and interpreted the interview data. PM, CA, and LW conducted the eye-tracking study and retrospective interviews, TL contributed. PM analyzed and interpreted the eye-tracking data. PM analyzed and interpreted the survey data, and CA and LW supported data interpretation. OH contributed to the acquisition of funding. All the authors provided substantial comments and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.
[DOCX File , 27 KB - humanfactors_v11i1e50926_app1.docx]

Multimedia Appendix 2
COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.
[PDF File (Adobe PDF File), 480 KB - humanfactors_v11i1e50926_app2.pdf]

Multimedia Appendix 3
Eye-tracking tasks.
[DOCX File , 18 KB - humanfactors_v11i1e50926_app3.docx]

Multimedia Appendix 4
Definition of themes.
[DOCX File , 34 KB - humanfactors_v11i1e50926_app4.docx]

Multimedia Appendix 5
Task performance data.
[XLSX File (Microsoft Excel File), 15 KB - humanfactors_v11i1e50926_app5.xlsx]

References

1. Erdmann F, Spix C, Katalinic A, Christ M, Folkerts J, Hansmann J, et al. Krebs in Deutschland für 2017/2018. Robert Koch Institut. 2021. URL: https://edoc.rki.de/handle/176904/9042 [accessed 2024-02-19]
2. Tighe M, Molassiotis A, Morris J, Richardson J. Coping, meaning and symptom experience: a narrative approach to the overwhelming impacts of breast cancer in the first year following diagnosis. Eur J Oncol Nurs 2011 Jul;15(3):226-232 [FREE Full text] [doi: 10.1016/j.ejon.2011.03.004] [Medline: 21511530]
3. ElKefi S, Asan O. How technology impacts communication between cancer patients and their health care providers: a systematic literature review. Int J Med Inform 2021 May;149:104430 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104430] [Medline: 33684711]
4. Kashgary A, Alsolaimani R, Mosli M, Faraj S. The role of mobile devices in doctor-patient communication: a systematic review and meta-analysis. J Telemed Telecare 2017 Sep;23(8):693-700. [doi: 10.1177/1357633X16661604] [Medline: 27632990]
5. Qudah B, Luetsch K. The influence of mobile health applications on patient - healthcare provider relationships: a systematic, narrative review. Patient Educ Couns 2019 Jun;102(6):1080-1089. [doi: 10.1016/j.pec.2019.01.021] [Medline: 30745178]

6. Jongerius C, Russo S, Mazzocco K, Pravettoni G. Research-tested mobile apps for breast cancer care: systematic review. JMIR Mhealth Uhealth 2019 Feb 11;7(2):e10930 [FREE Full text] [doi: 10.2196/10930] [Medline: 30741644]
7. Chen L, Bartel C, Cai X, Cheng Y, Perer A, McClaine S, et al. Patient and provider perspectives on symptom monitoring during outpatient chemotherapy: interview study. JMIR Form Res 2023 Apr 17;7:e46001 [FREE Full text] [doi: 10.2196/46001] [Medline: 37067857]
8. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. J Clin Oncol 2016 Feb 20;34(6):557-565 [FREE Full text] [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
9. Di Maio M, Gallo C, Leighl NB, Piccirillo MC, Daniele G, Nuzzo F, et al. Symptomatic toxicities experienced during anticancer treatment: agreement between patient and physician reporting in three randomized trials. J Clin Oncol 2015 Mar 10;33(8):910-915 [FREE Full text] [doi: 10.1200/JCO.2014.57.9334] [Medline: 25624439]
10. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Feb;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
11. Or C, Tao D. Usability study of a computer-based self-management system for older adults with chronic diseases. JMIR Res Protoc 2012 Nov 08;1(2):e13 [FREE Full text] [doi: 10.2196/resprot.2184] [Medline: 23612015]
12. Nielsen J. What is usability? In: User Experience Re-Mastered: Your Guide to Getting the Right Design. Cambridge, MA: Morgan Kaufmann; 2010:3-22.
13. ISO 9241-11:2018 Ergonomics of human-system interaction part 11: usability: definitions and concepts. International Organization for Standardization. 2018 Mar. URL: https://www.iso.org/standard/63500.html [accessed 2024-02-19]
14. Law EL, Abrahão S. Interplay between user experience (UX) evaluation and system development. Int J Hum Comput Stud 2014 Jun;72(6):523-525. [doi: 10.1016/j.ijhcs.2014.03.003]
15. Mendes-Santos C, Nunes F, Weiderpass E, Santana R, Andersson G. Development and evaluation of the usefulness, usability, and feasibility of iNNOV breast cancer: mixed methods study. JMIR Cancer 2022 Feb 15;8(1):e33550 [FREE Full text] [doi: 10.2196/33550] [Medline: 35166682]
16. Cai T, Huang Y, Zhang Y, Lu Z, Huang Q, Yuan C. Mobile health applications for the care of patients with breast cancer: a scoping review. Int J Nurs Sci 2021 Aug 25;8(4):470-476 [FREE Full text] [doi: 10.1016/j.ijnss.2021.07.003] [Medline: 34631997]
17. Sadegh SS, Khakshour Saadat P, Sepehri MM, Assadi V. A framework for m-health service development and success evaluation. Int J Med Inform 2018 Apr;112:123-130. [doi: 10.1016/j.ijmedinf.2018.01.003] [Medline: 29500009]
18. Harte R, Glynn L, Rodríguez-Molinero A, Baker PM, Scharf T, Quinlan LR, et al. A human-centered design methodology to enhance the usability, human factors, and user experience of connected health systems: a three-phase methodology. JMIR Hum Factors 2017 Mar 16;4(1):e8 [FREE Full text] [doi: 10.2196/humanfactors.5443] [Medline: 28302594]
19. Doyle L, Brady AM, Byrne G. An overview of mixed methods research – revisited. J Res Nurs 2016;21(8):623-635. [doi: 10.1177/1744987116674257]
20. Bryman A. Integrating quantitative and qualitative research: how is it done? Qual Res 2006;6(1):97-113. [doi: 10.1177/1468794106058877]
21. Lim JY, Kim Y, Yeo SM, Chae BJ, Yu J, Hwang JH. Feasibility and usability of a personalized mHealth app for self-management in the first year following breast cancer surgery. Health Informatics J 2023;29(1):14604582231156476 [FREE Full text] [doi: 10.1177/14604582231156476] [Medline: 36772832]
22. Stubbins R, He T, Yu X, Puppala M, Ezeana CF, Chen S, et al. A behavior-modification, clinical-grade mobile application to improve breast cancer survivors' accountability and health outcomes. JCO Clin Cancer Inform 2018 Dec;2:1-11 [FREE Full text] [doi: 10.1200/CCI.18.00054] [Medline: 30652617]
23. Cooke L. Is eye tracking the next step in usability testing? In: Proceedings of the 2006 IEEE International Professional Communication Conference. 2006 Presented at: 2006 IEEE; October 23-25, 2006; Saragota Springs, NY. [doi: 10.1109/IPCC.2006.320355]
24. Jacob RJ, Karn KS. Commentary on section 4 - eye tracking in human-computer interaction and usability research: ready to deliver the promises. In: The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research. Amsterdam, The Netherlands: Elsevier Science; 2003:573-605.
25. Zhou L, Bao J, Setiawan IM, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500 [FREE Full text] [doi: 10.2196/11500] [Medline: 30973342]
26. Moorthy P, Weinert L, Harms BC, Anders C, Siegel F. German version of the mHealth app usability questionnaire in a cohort of patients with cancer: translation and validation study. JMIR Hum Factors 2023 Nov 01;10:e51090 [FREE Full text] [doi: 10.2196/51090] [Medline: 37910144]
27. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]

28. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
29. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
30. Bansback N, Li LC, Lynd L, Bryan S. Development and preliminary user testing of the DCIDA (dynamic computer interactive decision application) for 'nudging' patients towards high quality decisions. BMC Med Inform Decis Mak 2014 Aug 01;14:62 [FREE Full text] [doi: 10.1186/1472-6947-14-62] [Medline: 25084808]
31. Erol Barkana D, Açık A, Duru DG, Duru AD. Improvement of design of a surgical interface using an eye tracking device. Theor Biol Med Model 2014 May 07;11 Suppl 1(Suppl 1):S4 [FREE Full text] [doi: 10.1186/1742-4682-11-S1-S4] [Medline: 25080176]
32. Rashid S, Soo ST, Sivaji A, Naeni HS, Bahri S. Preliminary usability testing with eye tracking and FCAT analysis on occupational safety and health websites. Procedia Soc Behav Sci 2013 Nov 06;97:737-744. [doi: 10.1016/j.sbspro.2013.10.295]
33. Eghdam A, Forsman J, Falkenhav M, Lind M, Koch S. Combining usability testing with eye-tracking technology: evaluation of a visualization support for antibiotic use in intensive care. Stud Health Technol Inform 2011;169:945-949. [Medline: 21893885]
34. Forsman J, Anani N, Eghdam A, Falkenhav M, Koch S. Integrated information visualization to support decision making for use of antibiotics in intensive care: design and usability evaluation. Inform Health Soc Care 2013 Dec;38(4):330-353. [doi: 10.3109/17538157.2013.812649] [Medline: 23957739]
35. Russell DM, Chi EH. Looking back: retrospective study methods for HCI. In: Olson J, Kellogg W, editors. Ways of Knowing in HCI. New York, NY: Springer; 2014.
36. Ansaar MZ, Hussain J, Bang J, Lee S, Shin KY, Woo KY. The mHealth applications usability evaluation review. In: Proceedings of the 2020 International Conference on Information Networking (ICOIN). 2020 Presented at: ICOIN; January 7-10, 2020; Barcelona, Spain. [doi: 10.1109/icoin48656.2020.9016509]
37. Kjeldskov J, Skov MB, Stage J. A longitudinal study of usability in health care: does time heal? Int J Med Inform 2010 Jun;79(6):e135-e143 [FREE Full text] [doi: 10.1016/j.ijmedinf.2008.07.008] [Medline: 18757234]
38. Dunn Lopez K, Chin CL, Leitão Azevedo RF, Kaushik V, Roy B, Schuh W, et al. Electronic health record usability and workload changes over time for provider and nursing staff following transition to new EHR. Appl Ergon 2021 May;93:103359. [doi: 10.1016/j.apergo.2021.103359] [Medline: 33556884]
39. Sonderegger A, Zbinden G, Uebelbacher A, Sauer J. The influence of product aesthetics and usability over the course of time: a longitudinal field experiment. Ergonomics 2012 Apr 16;55(7):713-730. [doi: 10.1080/00140139.2012.672658] [Medline: 22506866]
40. Nápoles AM, Santoyo-Olsson J, Chacón L, Stewart AL, Dixit N, Ortiz C. Feasibility of a mobile phone app and telephone coaching survivorship care planning program among Spanish-speaking breast cancer survivors. JMIR Cancer 2019 Jul 09;5(2):e13543 [FREE Full text] [doi: 10.2196/13543] [Medline: 31290395]
41. Petrocchi S, Filipponi C, Montagna G, Bonollo M, Pagani O, Meani F. A breast cancer smartphone app to navigate the breast cancer journey: mixed methods study. JMIR Form Res 2021 May 10;5(5):e28668 [FREE Full text] [doi: 10.2196/28668] [Medline: 33970120]
42. Wildenbos GA, Jaspers MW, Schijven MP, Dusseljee-Peute LW. Mobile health for older adult patients: using an aging barriers framework to classify usability problems. Int J Med Inform 2019 Apr;124:68-77. [doi: 10.1016/j.ijmedinf.2019.01.006] [Medline: 30784429]
43. Morey SA, Barg-Walkow LH, Rogers WA. Managing heart failure on the go: usability issues with mHealth apps for older adults. Proc Hum Factors Ergon Soc Annu Meet 2017 Sep 28;61(1):1-5. [doi: 10.1177/1541931213601496]
44. Durchschnittsalter der bevölkerung in Deutschland (1871-2021). Bundesinstitut für Bevölkerungsforschung. URL: https://www.bib.bund.de/DE/Fakten/Fakt/B19-Durchschnittsalter-Bevoelkerung-ab-1871.html [accessed 2024-02-19]
45. Jessen S, Mirkovic J, Halvorsen Brendmo E, Solberg Nes L. Evaluating a strengths-based mHealth tool (MyStrengths): explorative feasibility trial. JMIR Form Res 2021 Nov 17;5(11):e30572 [FREE Full text] [doi: 10.2196/30572] [Medline: 34787580]
46. Graham TA, Ali S, Avdagovska M, Ballermann M. Effects of a web-based patient portal on patient satisfaction and missed appointment rates: survey study. J Med Internet Res 2020 May 19;22(5):e17955 [FREE Full text] [doi: 10.2196/17955] [Medline: 32427109]
47. Avdagovska M, Ballermann M, Olson K, Graham T, Menon D, Stafinski T. Patient portal implementation and uptake: qualitative comparative case study. J Med Internet Res 2020 Jul 27;22(7):e18973 [FREE Full text] [doi: 10.2196/18973] [Medline: 32716308]
48. Müller J, Weinert L, Svensson L, Rivinius R, Kreusser MM, Heinze O. Mobile access to medical records in heart transplantation aftercare: mixed-methods study assessing usability, feasibility and effects of a mobile application. Life (Basel) 2022 Aug 08;12(8):1204 [FREE Full text] [doi: 10.3390/life12081204] [Medline: 36013383]

49. Niazkhani Z, Toni E, Cheshmekaboodi M, Georgiou A, Pirnejad H. Barriers to patient, provider, and caregiver adoption and use of electronic personal health records in chronic care: a systematic review. BMC Med Inform Decis Mak 2020 Jul 08;20(1):153 [FREE Full text] [doi: 10.1186/s12911-020-01159-1] [Medline: 32641128]
50. Sollie M, Hansen M, Thomsen JB. Health technology readiness amongst patients with suspected breast cancer using the READHY-tool - a cross-sectional study. J Med Syst 2023 Nov 16;47(1):118 [FREE Full text] [doi: 10.1007/s10916-023-02016-0] [Medline: 37971517]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
ISO: International Organization for Standardization
MAUQ: mHealth App Usability Questionnaire
mHealth: mobile health
PRO: patient-reported outcome
QoL: quality of life
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture

Edited by P Santana-Mancilla; submitted 18.07.23; peer-reviewed by A Carrillo, M Moreno, A Wada, C Baxter; comments to author 25.08.23; revised version received 30.11.23; accepted 17.12.23; published 05.03.24.

Please cite as:
Anders C, Moorthy P, Svensson L, Müller J, Heinze O, Knaup P, Wallwiener M, Deutsch TM, Le TV, Weinert L
Usability and User Experience of an mHealth App for Therapy Support of Patients With Breast Cancer: Mixed Methods Study Using Eye Tracking
JMIR Hum Factors 2024;11:e50926
URL: https://humanfactors.jmir.org/2024/1/e50926
doi:10.2196/50926
PMID:38441959

Original Paper

Feasibility and Design Factors for Home-Based Pulmonary Rehabilitation of Patients With Chronic Obstructive Pulmonary Disease and Chronic Lung Diseases Based on a People-Object-Environment Framework: Qualitative Interview Study

Shih-Ying Chien[1,2], PhD; Alice May-Kuen Wong[2], MD; Winston Tseng[3], PhD; Han-Chung Hu[4,5], MD; Hsiu-Ying Cho[5], MA

[1]Department of Industrial Design, Chang Gung University, Taoyuan, Taiwan

[2]Department of Physical Medicine and Rehabilitation, Chang Gung Memorial Hospital, Taoyuan, Taiwan

[3]Division of Community Health Sciences, School of Public Health, University of California, Berkeley, CA, United States

[4]Department of Thoracic Medicine, Chang Gung Memorial Hospital, Taoyuan, Taiwan

[5]Department of Respiratory Therapy, Lin-Kou Chang Gung Memorial Hospital, Taoyuan, Taiwan

Corresponding Author:
Shih-Ying Chien, PhD
Department of Industrial Design
Chang Gung University
259 Wen-Hwa 1st Rd, Kweishan
Taoyuan, 333
Taiwan
Phone: 886 3 2118800 ext 5669
Email: shihying@mail.cgu.edu.tw

Abstract

Background: The feasibility of implementing home-based pulmonary rehabilitation (PR) can be assessed from the perspectives of patients with chronic lung disease and health care professionals involved in PR.

Objective: Through a qualitative inquiry using interviews and the adoption of the people-object-environment framework, this study aims to understand the influences of interpersonal, environmental, and situational factors on the perceptions and considerations of individuals involved in home-based PR for patients with chronic lung disease.

Methods: One-on-one interviews were conducted with 20 patients with chronic lung disease and 20 health care professionals for investigating their attitudes and opinions based on their experiences regarding home-based PR as well as for identifying the key factors affecting the benefits and drawbacks of such therapies. This study further evaluates the feasibility of using digital tools for medical diagnosis and treatment by examining the technology usage of both parties.

Results: The 4 key issues that all participants were the most concerned about were as follows: distance to outpatient medical care, medical efficiency, internet connectivity and equipment, and physical space for diagnosis and treatment. Interviews with patients and health care professionals revealed that the use of technology and internet was perceived differently depending on age and area of residence. Most participants reported that digital tools and internet connectivity had many benefits but still could not solve all the problems; moreover, these same digital tools and network transmission could lead to problems such as information security and digital divide concerns. This study also emphasizes the significant impact of human behavior and thinking on shaping the design of health care interventions and technologies. Understanding user perspectives and experiences is crucial for developing effective solutions for unmet needs.

Conclusions: The results of this study indicate that despite the different perspectives of patients and health care professionals, their considerations of the key issues are very similar. Therefore, the implementation of plans related to telemedicine diagnosis, treatment, or rehabilitation should take the suggestions and considerations of both parties into account as crucial factors for telehealth care design.

(JMIR Hum Factors 2024;11:e51150) doi:10.2196/51150

KEYWORDS

chronic lung diseases; home-based pulmonary rehabilitation; telehealth; remote health care

Introduction

As the third leading cause of morbidity and mortality worldwide, chronic obstructive pulmonary disease (COPD) is a significant public health issue [1-4]. In 2019, the number of individuals diagnosed with COPD exceeded 328 million worldwide [5-8]. A significant correlation between physical activity and lung function [9-12] emphasizes the importance of regular exercise for individuals with COPD who require pulmonary rehabilitation (PR) [13-19]. However, patients with COPD often report reluctance to engage in physical activities due to dyspnea, the effects of which include chronic cough, exacerbations, reduced exercise capacity, and impaired quality of life [20-25]. PR is a tailored and comprehensive intervention conducted via a thorough assessment of the patient. In individuals with chronic pulmonary diseases, the primary objective of the pulmonary intervention is to improve not only their overall health but also their psychosocial well-being in the long term [26-28]. Typically, PR programs are customized for personal symptomatic conditions [29-31]; hence, PR interventions entail tailored exercises and educational sessions aimed at enhancing activity tolerance, mitigating symptoms, and augmenting skills that aid in managing chronic respiratory diseases [31,32]. The majority of PR treatments usually require one-on-one sessions and the assistance of a therapist [33-35]. However, the one-on-one care approach is limited due to shortages in health care personnel, elevated work-related stress, and prolonged working hours [36]. Moreover, when the COVID-19 pandemic hit, lockdowns and personnel restrictions forced the interruption of PR for many patients with chronic lung disease, which posed a threat to their lives [37-40].

Due to the COVID-19 pandemic, telehealth has become increasingly attractive owing to its functionality, importance, and prospects [41,42]. In addition to reducing human contact and easing the burden on health care workers, telehealth leverages technology communication and transmission to alleviate the workload of respiratory therapists and improve the accuracy of respiratory rehabilitation records [43-45]. Using telehealth, patients can undergo rehabilitation at home and be monitored remotely by medical personnel [46,47]. Home-based PR can also mitigate the difficulties of outpatient care for patients living in remote areas and those with physical disabilities [48-50]. Furthermore, it can be used as an auxiliary means of physical PR to assist in self-management and precisely modify behavior, thereby reducing hospitalization and medical costs [51,52].

Traditional PR usually relies on one-on-one human monitoring through observation or physiological monitors to examine a patient's health condition. Remote health care has the advantage of prescribing home-based PR, enabling patients who are unable to leave their homes due to physical conditions such as disability or living in rural areas to partake in rehabilitation programs at home [53,54]. However, there are also many limitations and considerations of remote health care, as follows:

1. Lack of security and limited interpersonal interaction: The safety of patients is the primary concern of clinical physicians [55,56]. The biggest challenge of home rehabilitation is emergency treatment, which has been the main hurdle for remote health care since many years [57]. In addition, remote therapy can only provide limited physical and mental assessments [58,59]. Due to the lack of face-to-face interpersonal interactions, patients may develop loneliness, helplessness, and frustration, which may reduce the effectiveness of treatments and the speed of recovery [60,61].
2. Privacy and security issues: Most remote health care is performed through network transmission. Many clinical physicians believe that network transmission may lead to data leakage or theft of medical records or personal information of patients [62,63].
3. Technological and equipment limitations: The implementation of remote health care requires specific technological equipment such as smartphones or computers with network functions. However, for many remote users or special groups such as older persons, lack of equipment, poor network communication quality, or unfamiliarity with network-related technology hinder utilization [64].
4. Insurance payment limitations: Different regions or countries have different standards for remote health care services. Therefore, many insurance companies do not have a remote health care reimbursement system or only cover specific services [65-68].

Despite its limitations and by taking people, object, and environment into consideration, telemedicine remains a valuable tool for the provision of health care services, especially for patients who have difficulty visiting medical facilities in person or those affected by infectious diseases and related restrictions such as lockdowns and quarantine. Telemedicine enables uninterrupted treatment and continued assistance for patients in their recovery. However, in establishing a home-based PR, it is essential to consider the various environments of participants to effectively maximize the benefits of this medical service.

Methods

Ethics Approval

This study was approved by the institutional review board of Chang Gung Memorial Hospital (approval 202200070B0). The participants were patients with chronic lung disease and respiratory health care professionals who had provided written informed consent from both urban and rural areas. Due to the COVID-19 pandemic, all one-on-one interviews were conducted by videoconferencing.

Participants and Procedures

The 20 patients recruited for the interviews included those who had participated in PR programs and those who had not. During the interviews, the patients provided insights into the implementation of PR programs from a patient-centric standpoint. All interviewees had a medical history of 5 years or more.

The 20 respiratory health care professionals included registered thoracic surgeons, respiratory therapists, physical therapists, and PR specialists. Most of these professionals had experience treating patients with chronic lung disease and had participated in designing exercise prescriptions, patient tracking and monitoring, and disease progression research in PR programs. Furthermore, the majority of the interviewees had treated a specific proportion of patients with chronic lung disease within the past 3 years (Table 1).

Table 1. Characteristics of the health care professionals (n=20).

	Values
Gender (male:female)	8:12
Age (years), min-max; mean (SD)	27-65; 46 (11)
Experience in pulmonary rehabilitation (years), mean (SD)	5.2 (8.47)
Type of health care professional in pulmonary rehabilitation, n (%)	
Thoracic surgeons	3 (15)
Respiratory therapists	12 (60)
Physical therapists	2 (10)
Pulmonary rehabilitation specialists	3 (15)

Prior to the interviews, all participants were required to complete a survey questionnaire, which included demographic information and details of their use of smart devices and the internet. Daily use was defined as regular usage. The patients provided information about their pulmonary disease status, duration of illness, and a self-assessment of their health status (on a 5-point scale ranging from excellent to poor) as well as recalled their activity frequency over the past 7 days. Health care personnel were required to answer questions related to their primary clinical responsibilities. Each participant took part in a 1.5- to 2-hour interview session conducted by the primary author, who was also a clinical researcher and an assistant professor affiliated with the Chang Gung Medical Foundation. In-depth interviews were primarily used to collect the data. After collecting the interview data, all identifiable personal information was removed from the transcripts. The data were then coded, organized, and analyzed using NVivo 12.0 software (Lumivero) for qualitative data analysis. For accurate and detailed data interpretation, the transcripts were provided to the interviewees for review and cross-checked with relevant researchers to confirm the accuracy of data interpretation.

Results

Characteristics of the Participants

This study consisted of 40 participants: 20 health care professionals specializing in PR and 20 patients with chronic lung diseases. The background characteristics of the 20 health care professionals are shown in Table 1; nearly 60% (12/20) were respiratory therapists, and the remaining health care professionals were pulmonary surgeons, physical therapists, and rehabilitation physicians. Their mean age was 46 years, and all had more than 3 years of experience in PR and treatment (mean 5.2 years). The background characteristics and activity habits of the 20 patients interviewed are shown in Table 2; the majority of the patients had COPD (12/20, 60%), and 25% (5/20) were lung transplant recipients. The majority of the participants (15/20, 74%) had never participated in a PR program, and 70% (14/20) of the patients rated their physical condition as poor. Regarding exercise over the past week, 65% (13/20) of the patients chose a 10-minute walk as their exercise indicator, followed by strength training (5/20, 25%). Notably, 55% (11/20) of the patients reported preferring to sit rather than stand and to stand rather than move.

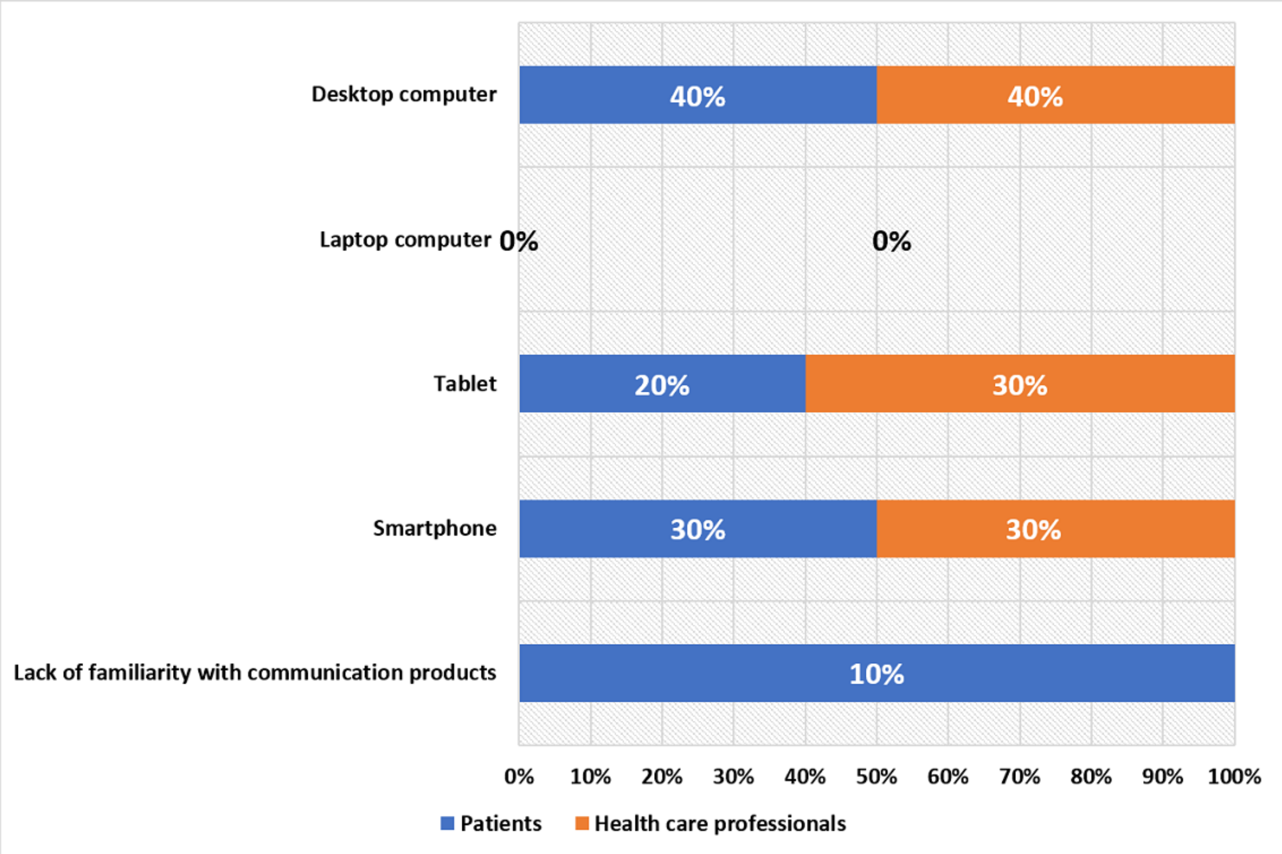
Table 2. Characteristics of the patients (n=20).

	Values
Gender (male:female)	10:10
Age (years), min-max; mean (SD)	51-85; 68 (9.8)
Participation in pulmonary rehabilitation programs, n (%)	
Yes	5 (26)
No	15 (74)
Chronic lung diseases, n (%)	
Chronic obstructive pulmonary disease	12 (60)
Asthma	3 (15)
Lung transplantation	5 (25)
Self-assessment of their health, n (%)	
Excellent	0 (0)
Very good	0 (0)
Good	1 (5)
Fair	5 (25)
Poor	14 (70)
Exercise frequency and quantity over the past week, n (%)	
I have engaged in high-intensity strength training, including aerobic exercise, fast cycling, and swimming.	2 (10)
I have participated in moderate physical activities such as stretching exercises and flexibility training.	5 (25)
I have walked for at least 10 minutes every day.	13 (65)
What statement best characterizes my exercise habits? n (%)	
Given the option, I will opt to sit rather than stand.	11 (55)
I frequently require standing but not for the purpose of lifting heavy objects.	5 (25)
Climbing slopes and stairs is a common necessity for me.	4 (20)
I often transport heavy objects and engage in manual labor.	0 (0)

Survey Results

In order to better examine the potential of telehealth, we conducted a survey targeting contemporary electronic communication tools, specifically computers, cellphones, and tablets, which were widely utilized by both patients and health care professionals, as shown in Figure 1. The majority of the patients and health care professionals reported using desktop computers as their most frequently used electronic communication device, constituting the largest proportion at 40% (8/20), followed by smartphones at 30% (6/20). Notably, health care professionals reported a higher usage rate (by 10%) of tablet computers compared to patients. Note that neither group of participants reported habitually using laptops. Overall, the patients were less proficient with technology compared to the health care professionals, which is a crucial determinant in implementing telehealth programs.

Figure 1. Survey results of the usage of electronic communication devices among patients and health care professionals in this study.



People-Object-Environment Framework

This study adopts the people-object-environment framework as the focal point for the interview investigation to improve the understanding of the feasibility of home-based PR. This study analyzes the advantages and disadvantages of remote PR in the current context, with the aim of bridging the gap between ideal use and reality (Table 3). Activities involving various elements such as individuals, entities, and environmental factors often result in the emergence of diverse concerns among different participants. Through the analysis presented in Table 3, we identified the gaps in home-based rehabilitation services from the perspectives of health care professionals and patients. Subsequently, this facilitated a thorough discussion of potential solutions to meet the needs and expectations of all involved parties. Our research findings reveal that the use of telehealth for home-based PR programs had both advantages and disadvantages. Using a people-object-environment framework to analyze the results, we describe 4 dimensions: reduced time and transportation constraints to access medical care; improved medical efficiency; changes in equipment, network, and physical space; and information transmission security, about which health care professionals particularly raised concerns in telehealth.

Table 3. Analysis of the pros and cons of telerehabilitation with the people-object-environment framework.

	People (patient)	Object (patient)	Environment (patient)
People (health care professionals)			
Pros	• Promoting health care access in remote areas: Telehealth facilitates convenient health care services, which enhance medical care for patients in remote areas and promote community health. • Boosting patient involvement: Telehealth enables interactions with health care professionals via web-based platforms, providing medical information and guidance as well as fostering active engagement of patients in their own health management.	• Improving health care resource allocation: Telehealth enables physicians to diagnose and treat patients across different geographical areas, alleviating shortages in local health care resources and enhancing the efficiency of health care resource allocation.	• Reducing health care burden: Telehealth reduces the health care burden for long-term patients or those requiring regular follow-ups; this minimizes the time and effort associated with transportation and waiting as well as provides cost-effective health care options. • Decreasing cross-infection risks: Telehealth minimizes contact between patients and health care professionals, thereby lowering the risk of cross-infection and promoting the health and safety of both health care professionals and patients.
Cons	• Bridging communication barriers: Telehealth reduces physical interactions and social contact between patients and physicians, which may have long-term effects on patients' psychological and social well-being.	• Operational and communication barriers: Older or technologically inexperienced patients may encounter difficulties in understanding instructions from remote health care professionals through telehealth.	• Environmental limitations: Home environments often impose spatial constraints that may limit various rehabilitation, diagnostic, and treatment activities.
Object (health care professionals)			
Pros	• Advantages of telehealth: Telehealth provides a convenient health care model, particularly beneficial for regions facing constraints related to time, geographical location, and transportation. The utilization of basic computer equipment enables the provision of medical consultations, making health care more accessible and efficient for patients in such areas.	• Wireless transmission: Wireless transmission significantly reduces the workload of health care professionals and makes health care services more efficient by transitioning from a one-on-one service to a one-to-many format.	• Digital health care infrastructure: Telehealth accelerates the transmission and exchange speed of health care information, leading to improved overall health care efficiency.
Cons	• Digital divide: Individuals who are from lower socioeconomic backgrounds or have limited access to digital resources may face barriers to participation in telehealth due to the lack of appropriate technological equipment or internet connectivity. This highlights inequalities in the distribution of health care resources. • Technological dependency: Users with limited technological skills or resources may encounter difficulties in operating telehealth, which requires adequate knowledge of technology, suitable equipment, and stable internet connectivity. • Health care quality and patient experience: Although telehealth provides convenient remote health care options for certain diseases or conditions, in-person consultations or measurements from medical instruments may offer more accurate health care services.	• Disparity in health care resources between urban and rural areas: Despite the convenience of telehealth for remote consultations, operational difficulties may still exist for areas lacking proper equipment.	• Equipment and infrastructure requirements: Telehealth relies on high-speed internet and appropriate equipment, which can still pose challenges in certain rural areas.
Environment (health care professionals)			

	People (patient)	Object (patient)	Environment (patient)
Pros	• Expansion of health care service areas: Through telehealth, physicians can diagnose and treat patients remotely without being limited by geographical location while providing real-time medical services. • Expansion of professional scope: Telehealth enables physicians to engage in remote meetings and collaborations with other health care experts, enhancing medical efficiency. • Enhancement of diagnosis and treatment efficiency: Telehealth reduces time and space limitations between physicians and patients, improving the efficiency of the overall health care services. • Increased convenience: Telehealth offers patients greater convenience, particularly for those residing in remote areas or facing mobility challenges, thereby reducing the time and costs associated with hospital visits.	• Enhancement of diagnostic and treatment capabilities: Through remote imaging and information sharing, physicians can access additional support and assistance, which improve diagnostic accuracy and treatment outcomes. • Improvement of health care resource utilization: Telehealth aids physicians in managing and allocating regional health care resources more effectively, enhancing utilization efficiency and reducing unnecessary health care costs.	• Reducing reliance on physical space: Telehealth reduces the need for physical space such as clinics and hospitals, thereby lowering costs and burdens associated with facilities and resources for health care institutions.
Cons	• Lack of physical contact: Telehealth may not provide opportunities for face-to-face contact with patients, which can make it challenging for physicians to conduct comprehensive physical examinations or assessments. • Limitations in comprehensive treatment: Some diagnoses and treatments may require physical contact and assistance from specific equipment, which cannot necessarily be substituted by telehealth.	• Technical requirements: The use of telehealth requires stable internet connectivity and appropriate device support, which may be challenging for users who are not familiar with technology. • Medical responsibility and risk management: Telehealth may involve issues of medical responsibility and risk management, such as misdiagnosis, treatment errors, or incomplete medical records, which may result in medical disputes and litigation. Physicians and health care institutions need to ensure compliance with relevant medical responsibility and risk management principles in telehealth as well as maintain a high level of medical practice.	• Health care security and privacy risks: The use of telehealth may involve health care security and digital privacy risks such as patient identity verification and medical record protection. Physicians should exercise caution in handling such issues.

Dimensions

Dimension 1: Distance, Time, and Transportation Issues

Both patients and health care professionals acknowledged the significant benefits of telehealth in addressing the challenges of distance in accessing medical care. Reductions in travel and wait times due to telehealth allowed patients to actively participate in their health care decision-making through web-based platforms. The digitization of medical records for better disease management was also facilitated. For residents in remote areas, telehealth eliminated geographical barriers to health care, promoted more efficient allocation of medical resources, and prevented the closure of regional hospitals and the "medical deserts" phenomenon. Nevertheless, health care professionals identified potential risks and quality-of-care issues associated with telehealth. Due to the absence of physical interactions, physicians were limited to relying on surface-level symptoms for diagnosis. Comprehensive physiological examinations and physical evaluations were also limited due to the lack of suitable equipment, which could in fact jeopardize the safety of critically ill patients who require urgent or emergency treatment. Moreover, prolonged social isolation resulting from telehealth may have adverse effects on patients' psychological and social well-being.

Dimension 2: Enhancing Medical Efficiency Through the Use of Telehealth

Health care professionals described that the popularization of telehealth was due to its advantages in improving the efficiency of disseminating medical information. Through online platforms, health care professionals can access patients' medical history instantaneously and collaborate to provide optimized treatment. This approach reduced constraints on patients' time and space, expanded the scope of medical services, reduced the workload of health care professionals, and transformed the traditional one-on-one respiratory treatment mode into a one-to-many model. Health care professionals also noted that the advantages of data and imaging arising from telehealth actually improved diagnostic and treatment efficiency, which could as a result achieve precision medicine. However, health care professionals also raised concerns about telehealth. The mode of transmitting medical information through data still harbored many risks and considerations such as diagnostic and treatment errors. Additionally, injuries (such as falls or respiratory distress) that could occur during treatment raised issues related to medical responsibility and risk management. Therefore, handling patient identity verification and medical records with caution when administering telehealth is crucial in order to safeguard patient privacy during medical treatments.

Dimension 3: Leveraging Internet Connectivity and Device-Based Solutions for Rehabilitation and Health Monitoring

Patients and health care professionals both identified that digital health care had the potential to significantly reduce medical wait times and enhance efficiency, which complemented the services of regional hospitals, lowered the medical burden of chronic patients, and eliminated limitations due to transportation, geography, and time. However, many challenges still remain with the use of digital tools for rehabilitation and monitoring systems. Most patients who require PR are older, aged ≥65 years, and unfamiliar with digital devices and networks, and they often lack the knowledge and understanding of how to install and configure such devices and applications. In addition, remote areas lack stable networks, technology, and equipment. This digital divide inhibits a subset of patients in certain areas from fully utilizing relevant medical services, highlighting the problem of an uneven distribution of medical resources.

Dimension 4: Advantages and Disadvantages of Converting Medical Spaces

The changing medical environment has brought many advantages to patients and health care institutions through telehealth, particularly during the COVID-19 pandemic, by reducing hospital-acquired infections and patients' reliance on medical space and resources, thereby alleviating the burden of health care costs. However, as previously mentioned, some health care professionals reported that not all diagnoses and treatments could be properly conducted remotely due to the availability or operation of equipment and the limitation of patients' home space and environment, which restrict the implementation of many treatment regimens. All the 4 dimensions supported by quotes from patients and health care professionals are shown in Table 4.

Table 4. Verbatim quotes supporting the main dimensions by patients and health care professionals.

Dimensions	Patient	Health care professional
Dimension 1: Distance to outpatient care		
Positive	…*During the pandemic, being able to have online consultations reduced a lot of my stress. I heard many of my friends around me were infected while at hospital.* [Female, 48 years old]	…*Telemedicine has reduced a lot of transport-related issues. Through online connections, we can access all of the patient's data and make more accurate assessments.* [HCP[a] #12]
Negative	…*I live in a very rural area where there are no taxis, so every time I see a doctor, I have to take four different buses. The journey alone takes me over three hours, so I avoid seeing a doctor if I can help it.* [Female, 64 years old]	…*To be honest, not every patient is suitable for telemedicine. For example, older patients may have difficulty understanding what I ask them to do. Also, some patients' conditions cannot be determined solely by questioning and require examination using medical instruments and devices, so it is difficult for me to make a diagnosis without a proper examination.* [HCP #8]
Dimension 2: Medical efficiency		
Positive	…*Because I get breathless when I walk, I try not to go out if I don't have to. I'm also afraid of falling when I go out, and I don't want to bother my children. So if I can see a doctor through a computer, I prefer that.* [Male, 66 years old]	…*The biggest advantage of telehealth is saving a lot of time and manpower. Of course, this refers to medical work that is more repetitive and lower risk. But I hope that in the future, online systems will have warning functions that can quickly let me know which patient has an issue that needs special attention.* [HCP #3]
Negative	…*I would rather see a real doctor. Just talking on the phone doesn't give me a feeling that I've really seen a doctor.* [Female, 72 years old]	…*To be honest, although the internet is convenient, I feel that its effectiveness is sometimes limited. Perhaps respiratory therapy needs to be divided into stages, and not every stage is suitable for being done at home. It may need to be classified/graded.* [HCP #17]
Dimension 3: Internet connectivity and equipment		
Positive	…*Every time I go out to see a doctor, I'm always in a rush and get so nervous that I forget to ask the doctor any questions. By seeing the doctor through a computer, I have more time to chat with the doctor.* [Female, 70 years old]	…*The internet connection is very convenient. As long as the health insurance card is inserted, all the patient's information can be accessed. Telehealth has not only changed a patient's medical treatment mode but also prevented many regional hospitals from closing down.* [HCP #11]
Negative	…*To be honest, I don't really understand the internet. If no one helps me set it up, I won't know how to see a doctor online. And if the doctor doesn't see me, how will they know what's wrong with me?* [Male, 76 years old]	…*Sometimes, the reason why I cannot wait for a patient is because the foreign caregiver has not set up the computer properly. When communicating with the patient through the computer, sometimes the elderly cannot understand, and it is also difficult to communicate with the caregiver. If I were there in person, I could still teach them how to do it.* [HCP #8]
Dimension 4: Space for diagnosis and treatment		
Positive	…*I am old and unable to move around easily. It would be best for me to see the doctor at home.* [Male, 86 years old]	…*A patient receiving online medical care at home will require much less space for us, such as waiting rooms and registration areas. It will also significantly reduce the demand and burden on staff.* [HCP #1]
Negative	…*There are many things that I cannot do at home. I need to have my blood pressure measured, but there is no one to help me at home. Also, I like to chat with people, but at home, there's only me.* [Female, 70 years old]	…*Online medical care now is quite good, with many complete functions such as registration, appointment progress, and electronic medical records. However, I personally have reservations about having many medical records stored in the cloud, as there is no absolute security. Also, if a patient falls at home, how to allocate responsibility and the risks involved are also concerns.* [HCP #14]

[a]HCP: health care professional.

Discussion

This study explores the perspectives and barriers of respiratory health care professionals and patients toward telehealth for rehabilitation for respiratory diseases. Based on the participant interviews, the use of home-based telerehabilitation for patients with lung diseases was perceived to have both advantages and disadvantages, which could be categorized under 4 domains: location, digital technology, internet connectivity, and physical space requirements. Unlike previous research [69], we adopted the people-object-environment framework and interviewed patients as well as health care professionals, with the aim of obtaining feedback from all participants in the same context. The 4 aspects mentioned above were found to be the most important concerns for health care institutions and patients.

Previous studies have reported that distance to outpatient care has a profound impact on chronic patients; in other words, the farther away from home, the lower is a patient's willingness to seek medical care [70]. However, Bhatt et al [71,72] highlighted that patients often exhibit a reluctance to engage in PR, regardless of proximity, for several reasons. Both groups of interviewees in our study reported that the provision of alternative options would reduce the number of hospital visits, which would benefit patients and health care institutions. Although digital health care has its limitations, leveraging the internet to expand regional hospital services is not only beneficial to the public but also makes medical services more effective [73]. In special circumstances such as the outbreak of a pandemic, issues such as patients being unable to attend in-person treatments cannot be ignored. It is undeniable that digital health care, in particular, spawns numerous benefits in such situations. However, there are still limitations to digital health care for people (doctors or the public), equipment and network, and the environment (urban or rural). For example, most older patients feel that only consultations in person with doctors generate the feeling of being treated. Furthermore, without physical examinations, it could be difficult for doctors to diagnose the cause of symptoms. Nonetheless, digital technology remains a good choice for patients with respiratory disease who do not want to venture outside or exercise; however, not every patient's home is equipped with remote medical devices or equipment, especially in rural areas [74]. Most patients with respiratory disease are also older, and without caregiver assistance, operating such devices can be difficult. In addition, due to limited professional knowledge, equipment, network, and living space, home care cannot replace all hospital diagnoses and rehabilitation. Therefore, in consideration of the findings from our study and a previous study [75], the implementation of hierarchical medical care requires that patients first undergo video consultations and then be referred to nearby medical institutions for appropriate treatment based on the severity of their condition. Patients can be referred to a larger medical center when necessary for treatment through an electronic referral platform between institutions. This approach not only effectively improves the utilization efficiency of medical resources but also significantly reduces medical expenses and transportation costs for patients.

This study has some limitations. First, as our study was conducted in a specific health care institution, our findings may not be generalizable to other regions or institutions. Second, most patients had poor health conditions and no prior experience with PR; thus, they relied only on limited experience and information. Lastly, the majority of the participants were older, which may have influenced their responses regarding computer use or internet issues. Additionally, health care service needs likely vary between urban and rural areas, and our study does not distinguish between the challenges and differences in home-based PR between these 2 types of areas. Future research should consider this aspect in their study design.

Most study participants reported that telehealth could greatly benefit patients with chronic pulmonary diseases; however, these benefits were not without limitations. Reflections on these limitations by patients or health care professionals revealed that telehealth is not suitable for all patients. For example, diagnosis and treatment via telehealth can only accomplish certain tasks and merely serve as a tool for preliminary diagnostic assessments. Nonetheless, preliminary assessments can determine whether a referral to a regional hospital or a large teaching hospital is necessary. This classification and referral system will also be applicable to rehabilitation therapy. Not all patients are suitable for home-based PR, considering patient safety, the required space and equipment, or the need for further precision testing, among other factors. In addition, although telehealth brought many conveniences to patients and health care professionals, both parties still faced significant psychological pressure. Patients noted that digital medicine lacked warmth, and they tended to prefer human care, while physicians had doubts about medical decision-making without the ability to perform physical examinations. The degree of control over digital technology was also an issue. Both parties lacked confidence that effective treatment could be achieved solely through the internet. Even though digital care has the advantage of long-term patient monitoring, some patients were unfamiliar with internet devices, and health care professionals were concerned that patients may not always respond correctly to instructions. Moreover, patients often neglect their physician's advice (such as following prescribed exercise schedules) due to lack of motivation and the need to physically meet with the physician. Both groups of study participants indicated that significant improvements in telerehabilitation technology were still needed, particularly for patients in rural areas or those who were older and living alone, who require more support and services.

Acknowledgments

This research was supported by the National Science and Technology Council (grant 111-2222-E-182-002-MY2). We also thank all the interviewees who provided insight and expertise that greatly assisted this research.

Conflicts of Interest

None declared.

References

1. Viegi G, Maio S, Pistelli F, Baldacci S, Carrozzi L. Epidemiology of chronic obstructive pulmonary disease: health effects of air pollution. Respirology 2006 Sep;11(5):523-532. [doi: 10.1111/j.1440-1843.2006.00886.x] [Medline: 16916323]

2. Cukic V, Lovre V, Dragisic D, Ustamujic A. Asthma and chronic obstructive pulmonary disease (COPD) - differences and similarities. Mater Sociomed 2012;24(2):100-105 [FREE Full text] [doi: 10.5455/msm.2012.24.100-105] [Medline: 23678316]
3. Chapman KR, Mannino DM, Soriano JB, Vermeire PA, Buist AS, Thun MJ, et al. Epidemiology and costs of chronic obstructive pulmonary disease. Eur Respir J 2006 Jan;27(1):188-207 [FREE Full text] [doi: 10.1183/09031936.06.00024505] [Medline: 16387952]
4. Pauwels R. Global initiative for chronic obstructive lung diseases (GOLD): time to act. Eur Respir J 2001 Dec;18(6):901-902 [FREE Full text] [Medline: 11829093]
5. Whittaker Brown S, Braman S. Recent advances in the management of acute exacerbations of chronic obstructive pulmonary disease. Med Clin North Am 2020 Jul;104(4):615-630. [doi: 10.1016/j.mcna.2020.02.003] [Medline: 32505256]
6. López-Campos JL, Tan W, Soriano JB. Global burden of COPD. Respirology 2016 Jan;21(1):14-23 [FREE Full text] [doi: 10.1111/resp.12660] [Medline: 26494423]
7. Nguyen HT, Collins PF, Pavey TG, Nguyen NV, Pham TD, Gallegos DL. Nutritional status, dietary intake, and health-related quality of life in outpatients with COPD. Int J Chron Obstruct Pulmon Dis 2019;14:215-226 [FREE Full text] [doi: 10.2147/COPD.S181322] [Medline: 30666102]
8. GBD 2015 Chronic Respiratory Disease Collaborators. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet Respir Med 2017 Sep;5(9):691-706 [FREE Full text] [doi: 10.1016/S2213-2600(17)30293-X] [Medline: 28822787]
9. Hartman JE, Boezen HM, de Greef MH, Bossenbroek L, ten Hacken NH. Consequences of physical inactivity in chronic obstructive pulmonary disease. Expert Rev Respir Med 2010 Dec;4(6):735-745. [doi: 10.1586/ers.10.76] [Medline: 21128749]
10. Pitta F, Troosters T, Spruit MA, Probst VS, Decramer M, Gosselink R. Characteristics of physical activities in daily life in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2005 May 01;171(9):972-977. [doi: 10.1164/rccm.200407-855OC] [Medline: 15665324]
11. Han Y, Heo Y, Hong Y, Kwon SO, Kim WJ. Correlation between physical activity and lung function in dusty areas: results from the chronic obstructive pulmonary disease in dusty areas (CODA) cohort. Tuberc Respir Dis (Seoul) 2019 Oct;82(4):311-318 [FREE Full text] [doi: 10.4046/trd.2019.0006] [Medline: 31172706]
12. Garcia-Aymerich J, Serra I, Gómez FP, Farrero E, Balcells E, Rodríguez DA, PhenotypeCourse of COPD (PAC-COPD) Study Group. Physical activity and clinical and functional status in COPD. Chest 2009 Jul;136(1):62-70. [doi: 10.1378/chest.08-2532] [Medline: 19255291]
13. Jenkins S, Hill K, Cecins NM. State of the art: how to set up a pulmonary rehabilitation program. Respirology 2010 Nov;15(8):1157-1173. [doi: 10.1111/j.1440-1843.2010.01849.x] [Medline: 20920127]
14. Bourbeau J. Making pulmonary rehabilitation a success in COPD. Swiss Med Wkly 2010;140:w13067 [FREE Full text] [doi: 10.4414/smw.2010.13067] [Medline: 20809435]
15. Spruit MA, Pitta F, McAuley E, ZuWallack RL, Nici L. Pulmonary rehabilitation and physical activity in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2015 Oct 15;192(8):924-933. [doi: 10.1164/rccm.201505-0929CI] [Medline: 26161676]
16. Williams V, Bruton A, Ellis-Hill C, McPherson K. The effect of pulmonary rehabilitation on perceptions of breathlessness and activity in COPD patients: a qualitative study. Prim Care Respir J 2010 Mar;19(1):45-51 [FREE Full text] [doi: 10.4104/pcrj.2009.00044] [Medline: 19603125]
17. Donner C, Muir J. Selection criteria and programmes for pulmonary rehabilitation in COPD patients. Rehabilitation and Chronic Care Scientific Group of the European Respiratory Society. Eur Respir J 1997 Mar;10(3):744-757 [FREE Full text] [Medline: 9073018]
18. Fishman AP. Pulmonary rehabilitation research. Am J Respir Crit Care Med 1994 Mar;149(3 Pt 1):825-833. [doi: 10.1164/ajrccm.149.3.8118655] [Medline: 8118655]
19. Troosters T, Gosselink R, Janssens W, Decramer M. Exercise training and pulmonary rehabilitation: new insights and remaining challenges. Eur Respir Rev 2010 Mar;19(115):24-29 [FREE Full text] [doi: 10.1183/09059180.00007809] [Medline: 20956162]
20. Dürr S, Zogg S, Miedinger D, Steveling EH, Maier S, Leuppi JD. Daily physical activity, functional capacity and quality of life in patients with COPD. COPD 2014 Dec;11(6):689-696. [doi: 10.3109/15412555.2014.898050] [Medline: 24945972]
21. José A, Dal Corso S. Inpatient rehabilitation improves functional capacity, peripheral muscle strength and quality of life in patients with community-acquired pneumonia: a randomised trial. J Physiother 2016 Apr;62(2):96-102 [FREE Full text] [doi: 10.1016/j.jphys.2016.02.014] [Medline: 26996093]
22. Holland AE, Wadell K, Spruit MA. How to adapt the pulmonary rehabilitation programme to patients with chronic respiratory disease other than COPD. Eur Respir Rev 2013 Dec;22(130):577-586 [FREE Full text] [doi: 10.1183/09059180.00005613] [Medline: 24293474]
23. Blumenthal JA, Keefe FJ, Babyak MA, Fenwick CV, Johnson JM, Stott K, et al. Caregiver-assisted coping skills training for patients with COPD: background, design, and methodological issues for the INSPIRE-II study. Clin Trials 2009 Apr;6(2):172-184 [FREE Full text] [doi: 10.1177/1740774509102565] [Medline: 19342470]

24. Maurer J, Rebbapragada V, Borson S, Goldstein R, Kunik ME, Yohannes AM, ACCP Workshop Panel on AnxietyDepression in COPD. Anxiety and depression in COPD: current understanding, unanswered questions, and research needs. Chest 2008 Oct;134(4 Suppl):43S-56S [FREE Full text] [doi: 10.1378/chest.08-0342] [Medline: 18842932]
25. Kessler R, Partridge MR, Miravitlles M, Cazzola M, Vogelmeier C, Leynaud D, et al. Symptom variability in patients with severe COPD: a pan-European cross-sectional study. Eur Respir J 2011 Feb;37(2):264-272 [FREE Full text] [doi: 10.1183/09031936.00051110] [Medline: 21115606]
26. Spruit MA, Singh SJ, Garvey C, ZuWallack R, Nici L, Rochester C, ATS/ERS Task Force on Pulmonary Rehabilitation. An official American Thoracic Society/European Respiratory Society statement: key concepts and advances in pulmonary rehabilitation. Am J Respir Crit Care Med 2013 Oct 15;188(8):e13-e64. [doi: 10.1164/rccm.201309-1634ST] [Medline: 24127811]
27. Bolton CE, Bevan-Smith EF, Blakey JD, Crowe P, Elkin SL, Garrod R, British Thoracic Society Pulmonary Rehabilitation Guideline Development Group, British Thoracic Society Standards of Care Committee. British Thoracic Society guideline on pulmonary rehabilitation in adults. Thorax 2013 Sep;68 Suppl 2:ii1-i30. [doi: 10.1136/thoraxjnl-2013-203808] [Medline: 23880483]
28. Donner C, Ambrosino N, Goldstein RS. Pulmonary Rehabilitation. London: Taylor & Francis group; Aug 04, 2020.
29. Billings PR, Kohn MA, de Cuevas M, Beckwith J, Alper JS, Natowicz MR. Discrimination as a consequence of genetic testing. Am J Hum Genet 1992 Mar;50(3):476-482 [FREE Full text] [Medline: 1539589]
30. Olivier C, Grosbois J, Cortot AB, Peres S, Heron C, Delourme J, et al. Real-life feasibility of home-based pulmonary rehabilitation in chemotherapy-treated patients with thoracic cancers: a pilot study. BMC Cancer 2018 Feb 13;18(1):178 [FREE Full text] [doi: 10.1186/s12885-018-4102-6] [Medline: 29433474]
31. Ryan P, Sawin KJ. The individual and family self-management theory: background and perspectives on context, process, and outcomes. Nurs Outlook 2009;57(4):217-225 [FREE Full text] [doi: 10.1016/j.outlook.2008.10.004] [Medline: 19631064]
32. Sohanpal R. Understanding the reasons for non participation in self-management interventions amongst patients with chronic conditions: addressing and increasing opportunities for patients with advanced chronic obstructive pulmonary disease to access self-management [theses]. Queen Mary University of London. 2015 Apr 29. URL: http://qmro.qmul.ac.uk/xmlui/handle/123456789/15031 [accessed 2015-04-29]
33. Hall C, Nici L, Sood S, ZuWallack R, Castro M. Nonpharmacologic therapy for severe persistent asthma. J Allergy Clin Immunol Pract 2017;5(4):928-935. [doi: 10.1016/j.jaip.2017.04.030] [Medline: 28689843]
34. Ruppel GL, Enright PL. Pulmonary function testing. Respir Care 2012 Jan;57(1):165-175 [FREE Full text] [doi: 10.4187/respcare.01640] [Medline: 22222135]
35. Shah S, Nasb M, Lu M, Huang L, Wang Y, Chen H. Scaling the need, benefits, and risks associated with COVID-19 acute and postacute care rehabilitation: a review. Rehabil Res Pract 2020;2020:3642143 [FREE Full text] [doi: 10.1155/2020/3642143] [Medline: 32908705]
36. Czaja S, Sharit J. Human resource management. In: Designing Training and Instructional Programs for Older Adults. Boca Raton: Taylor & Francis Group; Oct 24, 2012.
37. Moubarak S, Merheb D, Basbous L, Chamseddine N, Bou Zerdan M, Assi HI. COVID-19 and lung cancer: update on the latest screening, diagnosis, management and challenges. J Int Med Res 2022 Sep;50(9):3000605221125047 [FREE Full text] [doi: 10.1177/03000605221125047] [Medline: 36154328]
38. Tabassum T, Rahman A, Araf Y, Ullah MA, Hosen MJ. Management of asthma patients during the COVID-19 pandemic: pathophysiological considerations to address the challenges. Beni Suef Univ J Basic Appl Sci 2022;11(1):20 [FREE Full text] [doi: 10.1186/s43088-022-00204-4] [Medline: 35155689]
39. Gharibo C, Sharma A, Soin A, Shah S, Diwan S, Buenaventura R, et al. Triaging interventional pain procedures during COVID-19 or related elective surgery restrictions: evidence-informed guidance from the American Society of Interventional Pain Physicians (ASIPP). Pain Physician 2020 Aug;23(4S):S183-S204 [FREE Full text] [Medline: 32942785]
40. Coker M, Folayan MO, Michelow IC, Oladokun RE, Torbunde N, Sam-Agudu NA. Things must not fall apart: the ripple effects of the COVID-19 pandemic on children in sub-Saharan Africa. Pediatr Res 2021 Apr;89(5):1078-1086 [FREE Full text] [doi: 10.1038/s41390-020-01174-y] [Medline: 32971527]
41. Bokolo AJ. Exploring the adoption of telemedicine and virtual software for care of outpatients during and after COVID-19 pandemic. Ir J Med Sci 2021 Feb;190(1):1-10 [FREE Full text] [doi: 10.1007/s11845-020-02299-z] [Medline: 32642981]
42. Lukas H, Xu C, Yu Y, Gao W. Emerging telemedicine tools for remote COVID-19 diagnosis, monitoring, and management. ACS Nano 2020 Dec 22;14(12):16180-16193. [doi: 10.1021/acsnano.0c08494] [Medline: 33314910]
43. Tan JY, Conceicao EP, Wee LE, Sim XYJ, Venkatachalam I. COVID-19 public health measures: a reduction in hospital admissions for COPD exacerbations. Thorax 2021 May;76(5):512-513. [doi: 10.1136/thoraxjnl-2020-216083] [Medline: 33273024]
44. Sharma P, Pandey AK, Bhattacharyya DK. Determining crucial genes associated with COVID-19 based on COPD findings. Comput Biol Med 2021 Jan;128:104126 [FREE Full text] [doi: 10.1016/j.compbiomed.2020.104126] [Medline: 33260035]
45. Radzikowska U, Ding M, Tan G, Zhakparov D, Peng Y, Wawrzyniak P, et al. Distribution of ACE2, CD147, CD26, and other SARS-CoV-2 associated molecules in tissues and immune cells in health and in asthma, COPD, obesity, hypertension,

and COVID-19 risk factors. Allergy 2020 Nov;75(11):2829-2845 [FREE Full text] [doi: 10.1111/all.14429] [Medline: 32496587]
46. Sakai T, Hoshino C, Yamaguchi R, Hirao M, Nakahara R, Okawa A. Remote rehabilitation for patients with COVID-19. J Rehabil Med 2020 Sep 04;52(9):jrm00095 [FREE Full text] [doi: 10.2340/16501977-2731] [Medline: 32871014]
47. Wickerson L, Helm D, Gottesman C, Rozenberg D, Singer LG, Keshavjee S, et al. Telerehabilitation for lung transplant candidates and recipients during the COVID-19 pandemic: program evaluation. JMIR Mhealth Uhealth 2021 Jun 17;9(6):e28708 [FREE Full text] [doi: 10.2196/28708] [Medline: 34048354]
48. Brundisini F, Giacomini M, DeJean D, Vanstone M, Winsor S, Smith A. Chronic disease patients' experiences with accessing health care in rural and remote areas: a systematic review and qualitative meta-synthesis. Ont Health Technol Assess Ser 2013;13(15):1-33 [FREE Full text] [Medline: 24228078]
49. Russell TG, Buttrum P, Wootton R, Jull GA. Internet-based outpatient telerehabilitation for patients following total knee arthroplasty. Journal of Bone and Joint Surgery 2011;93(2):113-120 [FREE Full text] [doi: 10.2106/jbjs.i.01375]
50. Roberts P. Staffing an empty schoolhouse: attracting and retaining teachers in rural, remote and isolated communities. ERIC. 2004. URL: https://eric.ed.gov/?id=ED507343 [accessed 2004-09-01]
51. Fitzner K, Moss G. Telehealth--an effective delivery method for diabetes self-management education? Popul Health Manag 2013 Jun;16(3):169-177. [doi: 10.1089/pop.2012.0054] [Medline: 23216062]
52. Martin MA, Catrambone CD, Kee RA, Evans AT, Sharp LK, Lyttle C, Chicago Initiative to Raise Asthma Health Equity Investigative Team. Improving asthma self-efficacy: developing and testing a pilot community-based asthma intervention for African American adults. J Allergy Clin Immunol 2009 Jan;123(1):153-159.e3 [FREE Full text] [doi: 10.1016/j.jaci.2008.10.057] [Medline: 19130936]
53. Brennan D, Mawson S, Brownsell S. Telerehabilitation: enabling the remote delivery of healthcare, rehabilitation, and self management. Stud Health Technol Inform 2009;145:231-248. [Medline: 19592797]
54. Dew A, Bulkeley K, Veitch C, Bundy A, Gallego G, Lincoln M, et al. Addressing the barriers to accessing therapy services in rural and remote areas. Disabil Rehabil 2013 Aug;35(18):1564-1570. [doi: 10.3109/09638288.2012.720346] [Medline: 23009191]
55. Hasson S, Waissengrin B, Shachar E, Hodruj M, Fayngor R, Brezis M, et al. Rapid implementation of telemedicine during the COVID-19 pandemic: perspectives and preferences of patients with cancer. Oncologist 2021 Apr;26(4):e679-e685 [FREE Full text] [doi: 10.1002/onco.13676] [Medline: 33453121]
56. Al Ameen M, Liu J, Kwak K. Security and privacy issues in wireless sensor networks for healthcare applications. J Med Syst 2012 Feb;36(1):93-101 [FREE Full text] [doi: 10.1007/s10916-010-9449-4] [Medline: 20703745]
57. Patel S, Park H, Bonato P, Chan L, Rodgers M. A review of wearable sensors and systems with application in rehabilitation. J Neuroeng Rehabil 2012 Apr 20;9:21 [FREE Full text] [doi: 10.1186/1743-0003-9-21] [Medline: 22520559]
58. Barnett P, Goulding L, Casetta C, Jordan H, Sheridan-Rains L, Steare T, et al. Implementation of telemental health services before COVID-19: rapid umbrella review of systematic reviews. J Med Internet Res 2021 Jul 20;23(7):e26492 [FREE Full text] [doi: 10.2196/26492] [Medline: 34061758]
59. Sampson JP, Kolodinsky RW, Greeno BP. Counseling on the information highway: future possibilities and potential problems. Jour of Counseling & Develop 2011 Dec 23;75(3):203-212 [FREE Full text] [doi: 10.1002/j.1556-6676.1997.tb02334.x]
60. White M, Dorman SM. Receiving social support online: implications for health education. Health Educ Res 2001 Dec;16(6):693-707. [doi: 10.1093/her/16.6.693] [Medline: 11780708]
61. Dolev-Amit T, Leibovich L, Zilcha-Mano S. Repairing alliance ruptures using supportive techniques in telepsychotherapy during the COVID-19 pandemic. Counselling Psychology Quarterly 2020 Jun 14;34(3-4):485-498 [FREE Full text] [doi: 10.1080/09515070.2020.1777089]
62. Simon SR, Evans JS, Benjamin A, Delano D, Bates DW. Patients' attitudes toward electronic health information exchange: qualitative study. J Med Internet Res 2009 Aug 06;11(3):e30 [FREE Full text] [doi: 10.2196/jmir.1164] [Medline: 19674960]
63. Shah SM, Khan RA. Secondary use of electronic health record: opportunities and challenges. IEEE Access 2020;8:136947-136965 [FREE Full text] [doi: 10.1109/access.2020.3011099]
64. Staring K. Organizational open source in the Global South [PhD thesis]. University of Oslo. 2011 May. URL: http://tinyurl.com/4bbfvt9u [accessed 2024-02-14]
65. Kichloo A, Albosta M, Dettloff K, Wani F, El-Amir Z, Singh J, et al. Telemedicine, the current COVID-19 pandemic and the future: a narrative review and perspectives moving forward in the USA. Fam Med Community Health 2020 Aug;8(3):e000530 [FREE Full text] [doi: 10.1136/fmch-2020-000530] [Medline: 32816942]
66. Tyree PT, Lind BK, Lafferty WE. Challenges of using medical insurance claims data for utilization analysis. Am J Med Qual 2006;21(4):269-275 [FREE Full text] [doi: 10.1177/1062860606288774] [Medline: 16849784]
67. Ho Chan WS. Taiwan's healthcare report 2010. EPMA J 2010 Dec;1(4):563-585 [FREE Full text] [doi: 10.1007/s13167-010-0056-8] [Medline: 23199110]
68. Health financing challenges and institutional options to move towards universal coverage in Nicaragua. World Health Report. 2010. URL: https://cdn.who.int/media/docs/default-source/health-financing/technical-briefs-background-papers/24nicaragua.pdf?sfvrsn=1cd08511_3&download=true [accessed 2021-06-11]

69. Lahham A, McDonald CF, Mahal A, Lee AL, Hill CJ, Burge AT, et al. Home-based pulmonary rehabilitation for people with COPD: A qualitative study reporting the patient perspective. Chron Respir Dis 2018 May;15(2):123-130 [FREE Full text] [doi: 10.1177/1479972317729050] [Medline: 28868892]
70. Ensor T, Cooper S. Overcoming barriers to health service access: influencing the demand side. Health Policy Plan 2004 Mar;19(2):69-79. [doi: 10.1093/heapol/czh009] [Medline: 14982885]
71. Bhatt SP, Patel SB, Anderson EM, Baugh D, Givens T, Schumann C, et al. Video telehealth pulmonary rehabilitation intervention in chronic obstructive pulmonary disease reduces 30-day readmissions. Am J Respir Crit Care Med 2019 Aug 15;200(4):511-513 [FREE Full text] [doi: 10.1164/rccm.201902-0314LE] [Medline: 30978302]
72. Bhatt SP, Baugh D, Hitchcock J, Kim Y, Cutter G, Aban I, et al. Video telehealth pulmonary rehabilitation for chronic obstructive pulmonary disease is associated with clinical improvement similar to center-based pulmonary rehabilitation. Ann Am Thorac Soc 2022 Feb;19(2):331-333 [FREE Full text] [doi: 10.1513/AnnalsATS.202104-419RL] [Medline: 34388079]
73. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection--harnessing the web for public health surveillance. N Engl J Med 2009 May 21;360(21):2153-5, 2157 [FREE Full text] [doi: 10.1056/NEJMp0900702] [Medline: 19423867]
74. Lavin B, Dormond C, Scantlebury MH, Frouin P, Brodie MJ. Bridging the healthcare gap: Building the case for epilepsy virtual clinics in the current healthcare environment. Epilepsy Behav 2020 Oct;111:107262 [FREE Full text] [doi: 10.1016/j.yebeh.2020.107262] [Medline: 32645620]
75. Delitto A, Erhard RE, Bowling RW. A treatment-based classification approach to low back syndrome: identifying and staging patients for conservative treatment. Phys Ther 1995 Jun;75(6):470-485. [doi: 10.1093/ptj/75.6.470] [Medline: 7770494]

Abbreviations

COPD: chronic obstructive pulmonary disease
PR: pulmonary rehabilitation

Edited by A Kushniruk; submitted 22.07.23; peer-reviewed by HH Chen, UK Chalwadi; comments to author 20.11.23; revised version received 01.12.23; accepted 24.01.24; published 07.03.24.

Please cite as:
Chien SY, Wong AMK, Tseng W, Hu HC, Cho HY
Feasibility and Design Factors for Home-Based Pulmonary Rehabilitation of Patients With Chronic Obstructive Pulmonary Disease and Chronic Lung Diseases Based on a People-Object-Environment Framework: Qualitative Interview Study
JMIR Hum Factors 2024;11:e51150
URL: https://humanfactors.jmir.org/2024/1/e51150
doi:10.2196/51150
PMID:38452366

Original Paper

A Technological Tool Aimed at Self-Care in Patients With Multimorbidity: Cross-Sectional Usability Study

Rodrigo Medina-García[1,2,3,4*], MD; Juan A López-Rodríguez[1,2,4,5,6*], MSc, MD, PhD; Cristina María Lozano-Hernández[1,3,4,5*], MSc; Verónica Ruiz Bejerano[7*], MSc; Paride Criscio[8*]; Isabel Del Cura-González[1,5,6*], MD, PhD; TeNDER Atención Primaria Group[1*]

[1]Primary Care Research Unit, Madrid Health Service, Madrid, Spain
[2]General Ricardos Primary Health Care Centre, Madrid Health Service, Madrid, Spain
[3]Interuniversity Doctoral Program in Epidemiology and Public Health, Rey Juan Carlos University, Alcorcón, Madrid, Spain
[4]Biosanitary Research and Innovation Foundation of Primary Care, Madrid, Spain
[5]Research Network on Chronicity, Primary Care and Health Promotion, Madrid, Spain
[6]Department of Medical Specialties and Public Health, Faculty of Health Sciences, Rey Juan Carlos University, Alcorcón, Madrid, Spain
[7]Visual Telecommunications Application Research Group, Universidad Politécnica de Madrid, Madrid, Spain
[8]DataWizard Srl, Rome, Italy
[*]all authors contributed equally

Corresponding Author:
Rodrigo Medina-García, MD
General Ricardos Primary Health Care Centre
Madrid Health Service
Calle General Ricardos, 131
Madrid, 28019
Spain
Phone: 34 660802663
Email: rodrigo.medina@salud.madrid.org

Abstract

Background: Information and communication technologies (ICTs) have been positioned as useful tools to facilitate self-care. The interaction between a patient and technology, known as usability, is particularly important for achieving positive health outcomes. Specific characteristics of patients with chronic diseases, including multimorbidity, can affect their interaction with different technologies. Thus, studying the usability of ICTs in the field of multimorbidity has become a key element to ensure their relevant role in promoting self-care.

Objective: The aim of this study was to analyze the usability of a technological tool dedicated to health and self-care in patients with multimorbidity in primary care.

Methods: A descriptive observational cross-sectional usability study was performed framed in the clinical trial in the primary care health centers of Madrid Health Service of the TeNDER (Affective Based Integrated Care for Better Quality of Life) project. The TeNDER technological tool integrates sensors for monitoring physical and sleep activity along with a mobile app for consulting the data collected and working with self-management tools. This project included patients over 60 years of age who had one or more chronic diseases, at least one of which was mild-moderate cognitive impairment, Parkinson disease, or cardiovascular disease. From the 250 patients included in the project, 38 agreed to participate in the usability study. The usability variables investigated were effectiveness, which was determined by the degree of completion and the total number of errors per task; efficiency, evaluated as the average time to perform each task; and satisfaction, quantified by the System Usability Scale. Five tasks were evaluated based on real case scenarios. Usability variables were analyzed according to the sociodemographic and clinical characteristics of patients. A logistic regression model was constructed to estimate the factors associated with the type of support provided for task completion.

Results: The median age of the 38 participants was 75 (IQR 72.0-79.0) years. There was a slight majority of women (20/38, 52.6%) and the participants had a median of 8 (IQR 7.0-11.0) chronic diseases. Thirty patients completed the usability study, with a usability effectiveness result of 89.3% (134/150 tasks completed). Among the 30 patients, 66.7% (n=20) completed all tasks and 56.7% (17/30) required personalized help on at least one task. In the multivariate analysis, educational level emerged as a facilitating factor for independent task completion (odds ratio 1.79, 95% CI 0.47-6.83). The median time to complete the total tasks was 296 seconds (IQR 210.0-397.0) and the median satisfaction score was 55 (IQR 45.0-62.5) out of 100.

Conclusions: Although usability effectiveness was high, the poor efficiency and usability satisfaction scores suggest that there are other factors that may interfere with the results. Multimorbidity was not confirmed to be a key factor affecting the usability of the technological tool.

Trial Registration: Clinicaltrials.gov NCT05681065; https://clinicaltrials.gov/study/NCT05681065

(JMIR Hum Factors 2024;11:e46811) doi:10.2196/46811

KEYWORDS

user-centered design; multimorbidity; comorbid; self-care; medical informatics; primary health care; chronic disease; chronic condition; chronic illness; primary care; usability; telemedicne; telehealth; information and communication technologies; ICT; digital health; eHealth; human-computer interaction

Introduction

Multimorbidity, which is generally defined as the presence of two or more simultaneous chronic diseases in a patient, is a major challenge for health systems [1]. In the European Union, up to 50 million people are estimated to have multimorbidity [2]. Barnett et al [3] estimated a multimorbidity prevalence of 64.9% among patients aged 65-84 years and of 81.5% for those 85 years or older [3]. In recent years, patient-centered care models [4] and, more specifically, interventions aimed at self-care education have made it possible to optimize how patients with multimorbidity manage their chronic diseases [5]. This type of intervention makes it easier for patients to identify their health-problem needs and to identify techniques that can help them make decisions, take appropriate actions, and modify them as they present changes in their diseases [6].

Information and communication technologies (ICTs) have been positioned as useful tools to facilitate self-care [7]. The different self-care strategies using ICTs include those dedicated to the monitoring of biometric parameters through wearable technologies or portable devices and mobile apps [8,9]. To achieve positive health outcomes from these interventions, the interaction between a patient and technology is particularly important. The description of this interaction between the technology, the specific tasks to be developed, and the end user is a property known as usability [10].

Research on usability has grown in parallel with the development of ICTs in health [11]. Reports from international organizations such as the 2012-2020 eHealth Action Plan of the European Commission [12] and the World Health Organization Global Strategy on Digital Health 2020-2025 [13] summarized the importance of the development of technological tools that take into account their interaction with the special conditions of older adults. The study of usability can help determine the reasons for low patient adherence and adoption of a specific technological tool. The improvements in usability could facilitate interaction through several mechanisms: reducing anxiety related to the use of new tools, increasing accessibility and distribution among a greater number of users, and reducing the possible risks derived from misuse [10].

In evaluating usability, the International Organization for Standardization (ISO) 9126 standard [14] assesses the quality of the product [15] and the ISO 9241 standard focuses on processes, referred to as “the extent to which users in a specific environment can use a product to achieve objectives of effectiveness, efficiency and satisfaction in a particular task” [16]. Thus, usability comprises the effectiveness in usability, defined as the degree of completion [11,17,18] and the total number of errors per task; efficiency in usability, defined as the average time to perform the task; and satisfaction in usability, defined as the degree to which the user’s physical and emotional responses resulting from the use of a product satisfy their needs and expectations. The most commonly used methods for usability evaluation are questionnaires and interviews carried out after the use of the technological tool for a certain period of time.

The systematic reviews of Saeed et al [17] and Zapata et al [18] analyzed the definitions of the ISO standards and the methods used in evaluating the usability of health-related technological tools. Their results indicate that the most frequent usability problems are those related to visual aspects of the system and the ability to learn and use specific features [17,18]. However, because the results are limited to a specific technology and may not be generalizable, their interpretation should take into account the special characteristics of the end users, including their health conditions.

The specific characteristics of patients with chronic diseases can affect their interaction with different technologies. For example, in the usability evaluation studies of an app for diabetes self-management [19,20] and that conducted on an automatic drug dispenser for patients with dementia [21], characteristics of the patients were identified that interact with different aspects of usability. Relatedly, Wildenbos et al [22] differentiated four traits related to aging and chronic diseases that act as barriers to usability: cognitive, physical, motivational, and perceptual. Although the development of technological tools aimed at self-care is increasing, their usability has thus far mainly been evaluated in patients with specific isolated pathologies such as in the previous examples. Research from the perspective of patients with multimorbidity has been increasing in recent years [23,24] but remains insufficient [25], even though multimorbidity is the most common way of reporting chronic diseases in the population over 60 years of age [3].

Thus, studying the usability of ICTs has become a key element in the field of multimorbidity [26] to ensure its relevant role in promoting self-care [7]. Along these lines, the TeNDER (Affective Based Integrated Care for Better Quality of Life) project [27,28] was a multisectoral project funded by Horizon 2020, the EU Framework Programme for research and innovation. The TeNDER project developed an integrated care model to manage multimorbidity in patients with dementia, Parkinson disease, and cardiovascular disease in four European countries: Spain, Germany, Italy, and Slovenia. One of the clinical studies related to the TeNDER project was a multicenter, randomized, parallel-group clinical trial carried out in Spain with the main objective of evaluating the effectiveness of the TeNDER system to improve quality of life in patients with chronic diseases. Secondary aims were to describe the satisfaction of patients and their caregivers and the usability of the TeNDER system [29].

The objective of this study was to analyze the usability of a technological tool (TeNDER) dedicated to health and self-care in patients with multimorbidity in primary care.

Methods

Design

This was a descriptive observational cross-sectional study of usability. This study was framed in the clinical trial in the primary care health centers of Madrid Health Service of the TeNDER project (ClinicalTrials.gov NCT05681065) [29].

Ethical Considerations

This study respects the basic ethical principles of autonomy, beneficence, justice, and nonmaleficence, and its development followed the norms of Good Clinical Practice and the principles enunciated in the latest Declaration of Helsinki (Seoul 2013). The study obtained a favorable report from the Research Ethics Board of the Hospital Universitario 12 de Octubre (20/450) and was approved by the Central Research Commission of the Community of Madrid (PC:39/20). Informed consent was obtained from all participants involved in the study. No camera recording or any other identification was made. Patients were included with an anonymous identifier in the data collection logbook (DCL). All data were processed based on the provisions of the EU General Data Protection Regulation 2016/679 of the European Parliament and the Council (April 27, 2016) and the Organic Law on Data Protection and Guarantee of Digital Rights in the Spanish territory (LOPDGDD 3/2018 of 5 December). Participants did not receive any financial compensation for their participation in the study. The only compensation was that received through the user experience during the use of the technological tool.

Population and Sample

The study population included patients with one or more chronic diseases recruited from four primary care health centers in the Community of Madrid that had been included in the TeNDER project by their referring professionals [27,28].

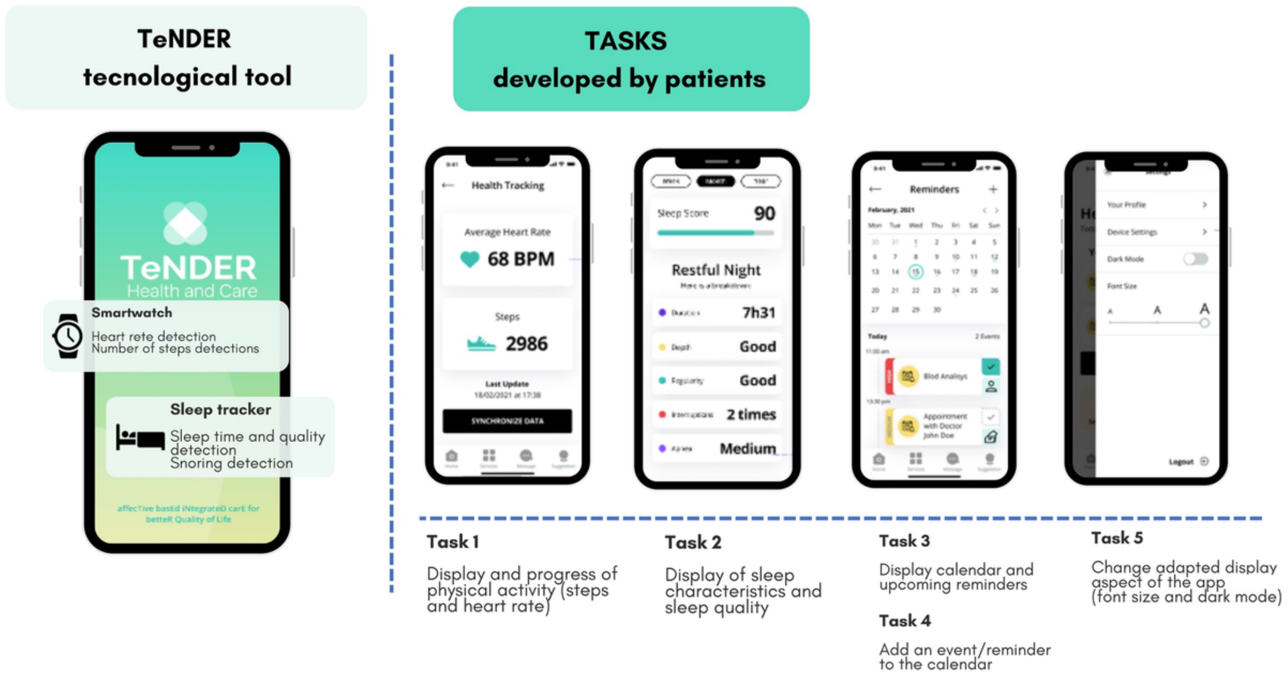
Patients over 60 years of age who had visited their health center in the last year and who had any of the following chronic diseases were included: mild-moderate cognitive impairment (Mini-Mental State Evaluation [MMSE] score 19-28 points), Parkinson disease, or cardiovascular disease, which includes patients with heart failure, chronic ischemic heart disease, or atrial fibrillation. Patients with a life expectancy of less than 6 months based on the opinion of their health care professionals, severe mental illness, incapacity for autonomous movement, or an MMSE score of less than 19 points were excluded.

The 250 patients included in the primary care health centers of Madrid Health Service for the TeNDER project were invited to participate via text messages with a mobile instant messaging app. Thirty-eight patients with multimorbidity (≥2 chronic diseases) agreed to participate. Considering an approximate 90% completion rate of tasks in previous usability evaluation studies of monitoring tools [19,25,30], with this sample size, we report a precision of 9.6% with the 95% CI.

TeNDER Technological Tool

The TeNDER technological tool is a web-based platform that included integrating sensors such as a smartwatch for monitoring physical activity, a sleep tracker to study sleep activity, and a mobile app in which the data collected are displayed and tools for self-management are offered (Figure 1). All of the TeNDER ecosystem technology was developed through a co-design process with all relevant stakeholders using a patient-centered approach. During the project, the functionalities and the mobile app were validated and released after user validation within an incremental development approach, ensuring a feedback framework that provided iterative refinement and improvements of the mobile app.

Figure 1. TeNDER technological tool and tasks to be performed by patients for the usability evaluation.



Variables

The main variable of the study was usability effectiveness, which was determined by the degree of completion and the total number of errors per task. Five tasks were designed using the TeNDER system (Textbox 1). The tasks to be evaluated were based on real case scenarios to simulate how patients would interact with the system in a real-life situation according to the care and self-management process based on the main functionalities of the TeNDER app [29]. The tasks were validated by a panel of three health professionals with experience in the study of usability to verify the accuracy of the content and the context. The degree of completion of the tasks was coded using three categories: (0) not completed when the subject was unable to complete the task (inability to progress or to request advanced help or interruption in task execution), (1) completed with personalized help in the form of comments or directed indications, and (2) completed independently when the user was able to carry out the task either without any help from the person in charge of the test or with the aid of minor indications. An error was coded when the subject made errors that could not be solved or that prevented further progress.

Textbox 1. Tasks evaluated for the usability study.

Task 1: Display and progress of physical activity (steps and heart rate).

Task 2: Display of sleep characteristics and sleep quality.

Task 3: Display calendar and upcoming reminders.

Task 4: Add an event/reminder to the calendar.

Task 5: Change-adapted display aspects of the app (font size and dark mode).

As secondary variables, usability efficiency was determined by timing each individual task and calculating the average time in each task. Usability satisfaction was quantified by administering the System Usability Scale (SUS) (Multimedia Appendix 1) in its Spanish-validated version [31]. For this scale, the global score ranges from 0 to 100, where higher values indicate greater usability satisfaction. According to Bangor et al [32], SUS scores of 70-100 indicate acceptable, whereas scores of 0-50 indicate not acceptable; scores between 50 and 70 are considered to indicate marginally acceptable results.

Sociodemographic variables collected included age, sex, and education level, and clinical variables included type and number of chronic diseases. Chronic pathologies were identified according to the proposals in the O'Halloran classification [33] (Multimedia Appendix 2).

Technology-related variables included previous use of touch screens and the affinity for technology interaction (ATI) scale [34]. For this scale, the global score ranges between 1 and 6, where higher values indicate a greater affinity for the technology (Multimedia Appendix 3).

Data Collection

The variables were collected by interview with the patient in consultation with their referring professional and were recorded in an electronic DCL designed ad hoc with the Research Electronic Data Capture (REDCap) tool hosted on the secure storage server of the institution. REDCap is a secure, web-based software platform designed to support data capture for research studies [35,36].

The patients received the TeNDER technological tool. The usability study was carried out 48 hours afterward based on the execution of tasks in a face-to-face session with a member of the research team who could provide assistance. To record the variables that measure usability, a real-time screen recording of the mobile device was performed during the entirety of task performance. One member of the research team analyzed the recordings. The start and end times were determined from the time the instructions were offered until the moment each task was completed; that information was subsequently transferred to the DCL.

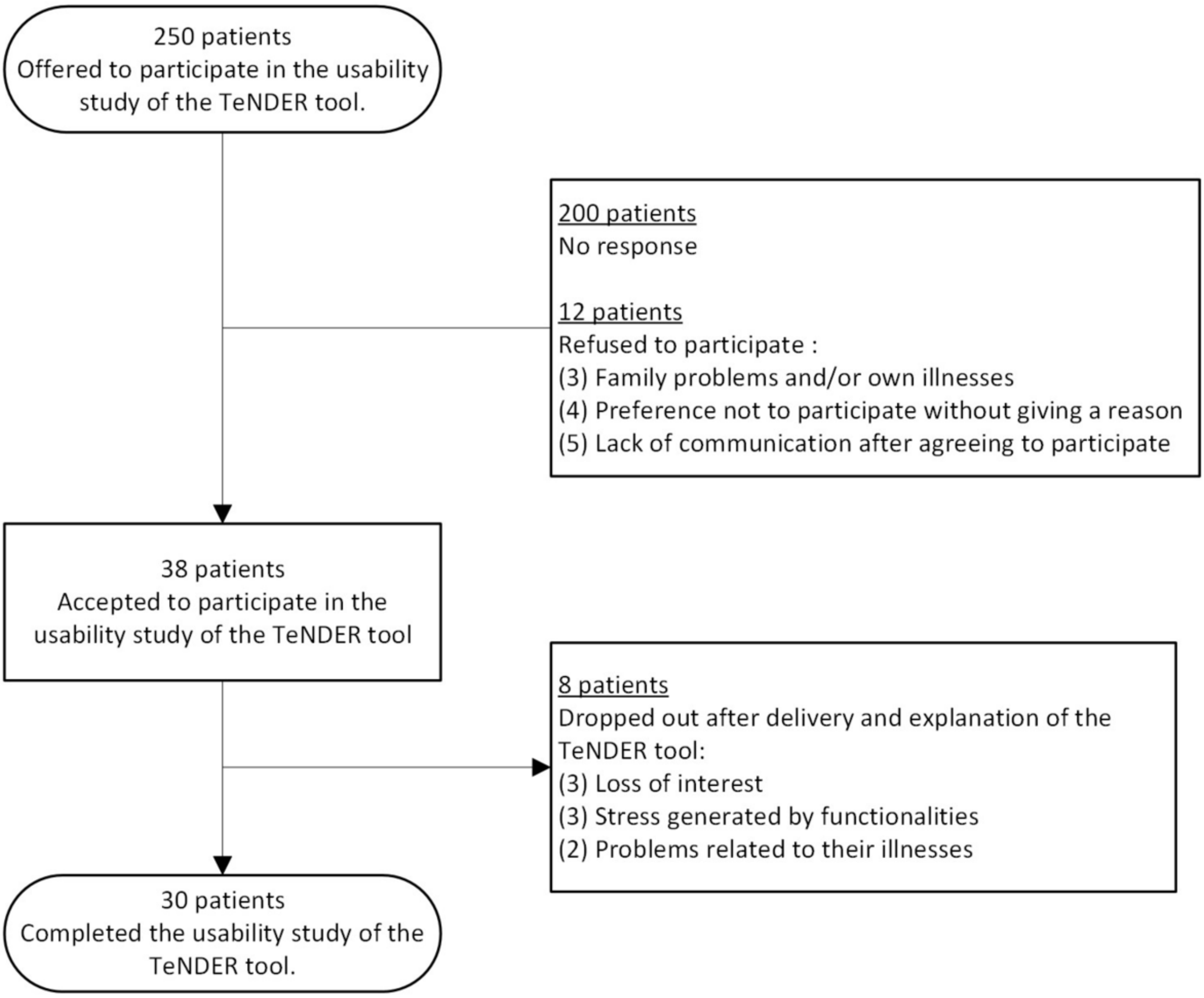
Statistical Analysis

The categorical variables are described as frequencies and percentages. The quantitative variables are described as medians and IQR, as they were nonnormally distributed for the number of patients under study. The main result variable was the proportion of completed tasks (usability effectiveness) with its 95% CI. As secondary outcome variables, the mean effective time to perform each of the tasks (usability efficiency) and the mean score in the SUS questionnaire (usability satisfaction) were estimated. The association of the different usability components (efficacy, effectiveness, and satisfaction) with the sociodemographic and clinical variables was evaluated using the χ^2 test for categorical variables and the Student *t* test for quantitative variables (the Mann-Whitney *U* test was used for comparison of variables that did not follow a normal distribution). The factors associated with completing the task in an independent manner were analyzed using a multiple logistic regression model with robust estimators. The dependent variable was completing the task autonomously. The independent variables were those found to be statistically significant in the bivariable analyses or variables that are otherwise considered to be clinically important. STATA 14 software was used for all statistical analyses.

Results

Among the 250 patients included in the TeNDER project invited, 38 (15.2%) agreed to participate in this study (Figure 2). There were no differences in sociodemographic characteristics between those who refused to participate and the final sample. Finally, 30 patients completed the usability study.

Figure 2. Flowchart of the participants.



The median age of the participants was 75 (IQR 72.0-79.0) years and 20/38 (52.6%) were women. With a median of 8 (IQR 7.0-11.0) chronic diseases, 89.5% of the patients had at least one cardiovascular risk factor. Syndromes that include anxiety and depression occurred at significantly different rates between women (11/20, 55%) and men (3/18, 16.7%). A total of 83.8% of the patients had previously interacted with touch screens, and the median result on the ATI scale was 3.4 (IQR 3.0-3.8), which differed between men and women.

The baseline characteristics of the patients are presented in Table 1.

Thirty patients completed the usability study. Among them, 66.7% (20/30) completed all the tasks to be evaluated. All patients were able to complete at least three of the five proposed tasks and 10 patients did not complete at least one task. At least one mistake was made while carrying out the tasks in 28/30 patients. A total of 66.7% (10/15) of the women required personalized help in at least one of the tasks for their completion. The median usability satisfaction in the SUS questionnaire was 55 (IQR 45.0-62.5).

A total of 150 tasks were carried out among all users and 89.3% (134/150) of the tasks were completed. Tasks 1 and 2 were completed 100% (60/60) of the time. Task 4 was completed at a lower proportion than the other tasks (22/30, 73.3%) and presented the highest number of errors (mean 2.5, SD 0.47). Task 3 and task 4 required personalized help to be completed (10/30, 33.3% and 8/30, 26.7%, respectively). The results of the different usability components are shown in Table 2 (also see Multimedia Appendix 4) and the details according to the different tasks are shown in Table 3 (also see Multimedia Appendix 5). The results of the subgroup usability analysis considering patients with cognitive deficits are provided in Multimedia Appendix 6.

In the multivariate analysis of the characteristics of the total tasks evaluated that were completed (Table 4), education level emerged as a facilitating factor to complete the task in an independent manner. Being male, having diseases related to cognition, and age hindered the completion of the task without help, with the latter factor being statistically significant.

Table 1. Patient baseline characteristics.

Characteristics	Total	Women	Men	*P* value
Participants, n (%)	38 (100)	20 (52.6)	18 (47.4)	N/A[a]
Age (years), median (IQR)	75.0 (72.0-79.0)	75.0 (69.0-80.0)	76.0 (73.0-78.0)	.67
Education level, n (%)				.19
Up to primary studies	17 (44.7)	8 (40.0)	9 (50.0)	
Secondary studies	7 (18.4)	6 (30.0)	1 (5.6)	
Higher education	14 (36.8)	6 (30.0)	8 (44.4)	
Number of chronic diseases, median (IQR)	8.0 (7.0-11.0)	9.0 (8.0-11.5)	8.0 (5.0-10.0)	.05
Cardiovascular risk factors, n (%)				
Total	34 (89.5)	19 (95.0)	15 (83.3)	.33
Arterial hypertension	24 (63.2)	13 (65.0)	11 (61.1)	>.99
Lipid metabolism disorders	24 (63.2)	16 (80.0)	8 (44.4)	.04
Type 2 diabetes mellitus	13 (34.2)	8 (40.0)	5 (27.8)	.51
Overweight/obesity	13 (34.2)	7 (35.0)	6 (33.3)	>.99
Cardiovascular disease	28 (73.7)	4 (20.0)	4 (22.2)	>.99
Perception problems, n (%)				
Total	21 (55.3)	9 (45.0)	12 (66.7)	.21
Vision problems	16 (42.1)	8 (40.0)	8 (44.4)	>.99
Musculoskeletal problems, n (%)	32 (84.2)	19 (95.0)	13 (72.2)	.08
Cognition problems, n (%)				
Total	21 (55.3)	13 (65.0)	8 (44.4)	.33
Cognitive impairment	9 (23.7)	5 (25.0)	4 (22.2)	>.99
Anxiety-depression	14 (36.8)	11 (55.0)	3 (16.7)	.02
Sleep disorders, n (%)	12 (31.6)	4 (20.0)	8 (44.4)	.16
Previous interaction with touch screens, n (%)	31 (83.8)	15 (78.9)	16 (88.9)	.66
Affinity for technology interaction scale, median (IQR)	3.4 (3.0-3.8)	3.3 (2.7-3.8)	3.6 (3.2-3.9)	.16

[a]N/A: not applicable.

Table 2. Usability results according to the total number of patients who completed the study and total number of tasks completed.

Usability metric	Total	≤10 CDs[a]	>10 CDs	Women	Men
Per patient					
Number of patients	30	16	14	15	15
Usability effectiveness					
Number of tasks completed, median (IQR)	5 (4.0-5.0)	5 (4.0-5.0)	5 (4.0-5.0)	5 (4.0-5.0)	5 (5.0-5.0)
At least one task with personalized help, n (%)	17 (56.7)	9 (56.2)	8 (57.1)	10 (66.7)	7 (46.7)
All tasks completed independently, n (%)	12 (40.0)	6 (37.5)	6 (42.8)	5 (33.3)	7 (46.7)
Number of errors made, median (IQR)	4.0 (2.0-8.0)	5 (2.0-9.0)	4 (2.0-6.0)	6 (3.0-9.0)	3.5 (1.0-5.0)
Usability efficiency: time to complete all tasks (seconds), median (IQR)	296.0 (210.0-397.0)	300.0 (236.0-397.0)	284.0 (205.0-431.0)	296.0 (201.0-350.0)	293.5 (211.0-447.0)
Usability satisfaction: SUS[b] questionnaire score, median (IQR)	55.0 (45.0-62.5)	50.0 (45.0-58.8)	61.2 (42.5-70.0)	52.5 (40.0-62.5)	60.0 (47.5-67.5)
Tasks					
Number of tasks	150	80	70	75	75
Usability effectiveness					
Proportion of tasks completed, n (%)	134 (89.3)	71 (88.7)	63 (90.0)	65 (86.7)	69 (92.0)
Number of errors per task, mean (SD)	1.0 (1.7)	1.2 (2.0)	0.8 (1.3)	1.4 (2.0)[c]	0.6 (1.2)[c]
Usability efficiency: time per task (seconds), mean (SD)	65.4 (92.7)	66.8 (95.1)	63.8 (90.7)	60.1 (79.3)	70.9 (105.1)

[a]CD: chronic disease.

[b]SUS: System Usability Scale.

[c]*P*=.007. This is the only comparison in which significant differences were found. The table with all *P* values is provided in Multimedia Appendix 4.

Table 3. Usability results by task.[a]

Usability metric	Total (N=30)	Women (n=15)	Men (n=15)	≤10 CDs[b] (n=16)	>10 CDs (n=14)	Up to secondary education (n=18)	Postsecondary education (n=12)
Task 3							
Usability effectiveness							
Number of patients completing the task, n (%)	28 (93.3)	14 (93.3)	14 (93.3)	14 (87.5)	14 (100.0)	16 (88.9)	12 (100.0)
Completed the task with personalized help, n (%)	10 (33.3)	5 (33.3)	5 (33.3)	5 (31.2)	5 (35.7)	7 (38.9)	3 (25.0)
Number of errors made, median (IQR)	0.0 (0.0-1.0)	0.0 (0.0-1.0)	0.0 (0.0-1.0)	1.0 (0.0-1.5)	0.0 (0.0-1.0)	0.0 (0.0-1.0)	0.5 (0.0-2.0)
Usability efficiency: time to perform the task (seconds), median (IQR)	57.0 (33.0-90.0)	40.0 (30.0-70.0)	65.0 (33.0-105.0)	50.0 (23.0-90.0)	58.5 (37.0-90.0)	60.0 (30.0-90.0)	48.5 (34.0-80.0)
Task 4							
Usability effectiveness							
Number of patients completing the task, n (%)	22 (73.3)	9 (60.0)	13 (86.7)	12 (75.0)	10 (71.4)	10 (55.6)[d]	12 (100.0)[d]
Completed the task with personalized help, n (%)	8 (26.7)	4 (26.7)	4 (26.7)	4 (25.0)	4 (28.6)	5 (27.8)	3 (25.0)
Number of errors made, median (IQR)	2.0 (0.0-4.0)	3.0 (1.0-5.0)[e]	1.0 (0.0-3.0)[e]	2.0 (0.0-4.0)	1.5 (0.0-3.0)	3.0 (0.0-5.0)	1.0 (0.0-2.0)
Usability efficiency: time to perform the task (seconds), median (IQR)	182.5 (0.0-280.0)	185.0 (0.0-220.0)	180.0 (137.0-300.0)	186.0 (80.0-265.0)	174.0 (0.0-280.0)	176.0 (0.0-300.0)	186.0 (161.5-225.0)
Task 5							
Usability effectiveness							
Number of patients completing the task, n (%)	26 (86.7)	12 (80.0)	14 (93.3)	15 (93.8)	11 (78.6)	14 (77.8)	12 (100.0)
Completed the task with personalized help, n (%)	5 (16.7)	2 (13.3)	3 (20.0)	3 (18.8)	2 (14.3)	4 (22.2)	1 (8.3)
Number of errors made, median (IQR)	1.0 (0.0-2.0)	1.0 (1.0-3.0)[c]	0.0 (0.0-1.0)[c]	0.5 (0.0-2.0)	1.0 (0.0-2.0)	1.5 (0.0-3.0)	0.5 (0.0-1.0)
Usability efficiency: time to perform the task (seconds), median (IQR)	20.0 (10.0-50.0)	30.0 (6.0-75.0)	18.0 (10.0-21.0)	20.0 (14.0-75.0)	19.0 (6.0-31.0)	20.5 (10.0-80.0)	19.0 (12.0-27.5)

[a]The usability results for tasks 1 and 2 are not included because they showed 100% effectiveness in usability; *P* values are only indicated for comparisons in which significant differences were found. The table with all *P* values is provided in Multimedia Appendix 5.

[b]CD: chronic disease.

[c]*P*=.01.

[d]*P*=.01.

[e]*P*=.04.

Table 4. Factors associated with completing a task in an independent manner.

Associated factors	Odds ratio (95% CI)	*P* value
Male sex	0.81 (0.24-2.74)	.74
Age	0.85 (0.77-0.94)	.002
Education level		
Up to secondary education	Reference	N/A[a]
Postsecondary education	1.79 (0.47-6.83)	.39
Diseases related to cognition	0.18 (0.04-0.81)	.03

[a]N/A: not applicable.

Discussion

Main Findings

The usability effectiveness of the TeNDER technological tool was 89.3% (134/150). Overall, 40% (12/30) of the patients completed all tasks independently. Task 4 was completed at a lower proportion than the rest of the tasks (22/30, 73.3%) and presented the highest number of errors (mean 2.5, SD 0.47). The usability efficiency, evaluated as the median time to complete the total tasks, was 296.0 seconds (IQR 210.0-397.0), with an average value per task of 65.4 seconds (SD 92.7). The satisfaction in usability perceived by the patients was acceptable (mean 52.2, SD 16.9). Being male, having diseases related to cognition, and age were factors that hindered the completion of the task without personalized help, among which only age was statistically significant.

Comparison With Other Studies

The usability effectiveness of the TeNDER technological tool was 89.3%, which is similar to the results of previous studies carried out on different categories of patients for similar technological developments. Sánchez-Morillo et al [30] evaluated the usability of a technological tool aimed at monitoring the symptoms of patients with chronic obstructive pulmonary disease, and Georgsson et al [19] evaluated a system designed for the management of self-care in patients with diabetes. The proportion of tasks completed was 88% and 91%, respectively, despite the opposite characteristics of the participants with respect to the level of education and affinity for technology in each of the studies. As in our study, the degree of task complexity could have been adapted to the characteristics of the potential users: tasks 1 and 2 were completed by all patients, whereas the rest of the tasks, of greater complexity, were completed by only those with higher education. For those who did not have higher education, the task completion rate reached up to 55.6%. These differences in the use of technology depending on the level of education have been confirmed in previous studies [37].

Despite the high proportion of completed tasks, 56.7% of the patients required personalized help to complete at least one of the tasks. Older age and cognition-related diseases were risk factors for requiring personalized help to complete the tasks. Previous experience in evaluating the usability of a computerized system for self-care management aimed at patients with chronic diseases yielded similar percentages of effectiveness in usability and help for task completion [25], which points to the importance of having family members or professionals assist patients with chronic diseases to interact with a mobile app [38].

The median value for usability satisfaction was 55.0 (IQR 45.0-62.5), which is a low marginal score over not acceptable [32]. Ligons et al [21] obtained similar results and indicated that the degree of response in satisfaction with a technology or system may not be related to the ability for the completion of its tasks. That is, patients may be able to complete tasks without knowing why they have completed them or how they can benefit from them in their day-to-day lives. Other studies, including that of Sánchez-Morillo et al [30], suggested that high levels of satisfaction may be caused by the presence of qualified professionals who assisted during the usability evaluation.

The median age of the patients in our study was 75 years and a high degree of multimorbidity was notable, with a median of up to eight chronic diseases. Previous studies have analyzed the usability of a technology from the perspective of patients with a chronic index disease in particular [19,20,30,39]. For example, Wildenbos et al [22] analyzed how chronic diseases can affect the usability of technological tools. Thus, a single chronic disease can be the cause of physical, cognitive, and perception barriers [22]. Medical conditions that could favor the appearance of these barriers are represented in our study: diabetes, cardiovascular disease, cognitive impairment, and vision problems. However, as in previous studies, no differences were found in the different aspects of usability based on the number of chronic diseases.

Only 38 patients out of a total of 250 who signed prior informed consent agreed to participate in this usability study. It should be noted that a mobile instant messaging app was used as the method of offering participation and there was a nonresponse rate of 80% (200/250). This means of communication, although common in current society, could have caused a lack of confidence or security in patients [40]. Among the 38 patients who agreed to participate, 6 (15.2%) decided to leave the study as a result of the stress generated by the proposed tasks or due to lack of interest. Few previous usability studies have reported the number of losses [41], perhaps due to the small number of patients involved. Wildenbos et al [22] mentioned lack of motivation as a key element to achieving acceptance of technology by older people. The benefits of using a technology should be made evident quickly and easily; otherwise, feelings of frustration and of giving up its use are likely. In a time-limited usability evaluation, these benefits are not evident, and their nonparticipation can help to avoid feelings of uncertainty, wasting the time invested in learning a technology, or the shame of making mistakes.

Differences based on sex in the use of ICTs have been described in previous studies [37,42]. Among older people, access to technology and their degree of involvement in daily activity is greater in men than in women [37]. These differences are also identified in the different aspects of usability. In our study, the average number of errors committed per task was significantly higher in women (mean 1.4, SD 2.0) than in men (mean 0.6, SD 1.2). In addition, the proportion of women who needed personalized help to complete tasks was higher (10/15, 67%) than that in men (7/15, 47%). These differences have been largely justified by the fact that the labor participation of women has been lower, particularly in computerized jobs due to less training over the years [43].

Limitations

Although the number of participants in our study is similar to that of other studies and the findings obtained provide valuable information, a larger sample size would provide a larger data set to conduct more sophisticated and detailed statistical analyses. Moreover, given the characteristics of the research, it has not been possible to collect opinions, sensations, and emotions in relation to the technological tool that the patients experienced during task execution. For this reason, including a qualitative methodology such as focus groups [44], think-aloud tasks [45], or a user-centered cognitive walkthrough [20] could provide essential information for understanding the decision-making of patients with multimorbidity when faced with a mobile app aimed at health.

Another limitation identified is the time of tool use being limited to 48 hours. Studies such as those of Tahsin et al [46] and Baek et al [47] showed how usability results can change at different times over longer intervals of use for up to 1 year.

Conclusions

Although usability effectiveness was high, the poor efficiency and usability satisfaction results suggest that there are other factors that may interfere with these results. Sex and education level can influence the degree of completion of tasks. It has not been possible to show that multimorbidity is a key factor in the usability results of a technological tool.

Acknowledgments

This work received funding from the European Commission under the EU Horizon 2020 program (grant agreement number 875325; TeNDER project, Affective Based Integrated Care for Better Quality of Life). The paper solely reflects the views of the authors. The Commission is not responsible for the contents of this paper, or any use made thereof. Funding was also received from the semFYC private foundation to IDCG who won a 2022 doctoral thesis award, and from the Foundation for Biosanitary Research and Innovation in Primary Care (FIIBAP) for the translation and publication of this paper. The following technical partners and primary care health centers (PCHCs) and family physicians (Grupo TeNDER Atención Primaria) participated in this work: Polytechnic University of Madrid (UPM): Javier Serrano-Romero, Alberto Del Rio-Ponce, Verónica Ruiz-Bejarano; PCHC Barrio del Pilar (Madrid): Esther Labrador-Arranz, Azucena Sáez-Berlana, Elena Pérez-Gonzalez, Marta Yagüe-López; Rosa Badenes, Virtudes Enguita, Araceli Garrido-Barral; PCHC Perales del Río (Getafe): Cristina Marcos-De Frutos, Maribel León-Coronado, Victoria Ruiz-Pascual, Cesar Minué-Lorenzo, Eloína Delgado-Silván; PCHC Pedro Laín Entralgo (Alcorcón): María Esther García-Talavera, Susana Benito-Alarcón, María Luisa Muñoz-Balsa, Pablo Amich-Alemany, María Teresa Cobos-Hinojal, Lorena Córdova-Castaño, Andrea Vinssac-Rayado, Angelica Fajardo-Alcántara, Esther Alonso-Herrador, Mercedes Gea; María Mendoza-Paico, Sandra Pérez-Santos, Marta Pinedo-Hoyos, Esther Martínez-Largo, María Aránzazu Murciano-Antón, Teresa Mazarro-Enrique, Canto De Hoyos-Alonso; PCHC General Ricardos (Madrid): Esther Cabrera, Daniel Prieto-Barranco, Celeste García-Galeano, Isaías Utiel, Álvaro López, Laura Jaurena, Yolanda Rodríguez; PCHC La Rivota (Alcorcón): Inmaculada Sánchez, Esther García-Esperalta; PCHC Arroyomolinos (Arroyomolinos): Tania Beatriz Ruiz-Molina, Jesús Sánchez-Montoya, Noemí Arraña-Gallego, Susana Domínguez-Mateos; Field researchers: Rocío Acosta-Cano, Susana Marugán-Rubio, Elisa Martínez-Núñez, Laura Calderón-Jiménez.

Data Availability

The data sets generated during and/or analyzed during this study are available in the Zenodo repository [48].

Conflicts of Interest

None declared.

Multimedia Appendix 1
System Usability Scale (SUS) questionnaire.
[DOCX File , 17 KB - humanfactors_v11i1e46811_app1.docx]

Multimedia Appendix 2
Chronic diseases classified according to O’Halloran et al [33].

[DOCX File , 16 KB - humanfactors_v11i1e46811_app2.docx]

Multimedia Appendix 3
Affinity for technology interaction (ATI) scale.
[DOCX File , 16 KB - humanfactors_v11i1e46811_app3.docx]

Multimedia Appendix 4
Usability results according to the total number of patients who completed the study and total number of tasks completed with complete *P* values for all comparisons (related to Table 2).
[XLSX File (Microsoft Excel File), 12 KB - humanfactors_v11i1e46811_app4.xlsx]

Multimedia Appendix 5
Usability results by task with complete *P* values for all comparisons (related to Table 3).
[XLSX File (Microsoft Excel File), 11 KB - humanfactors_v11i1e46811_app5.xlsx]

Multimedia Appendix 6
Subgroup analysis considering patients with and without cognitive deficit.
[DOCX File , 21 KB - humanfactors_v11i1e46811_app6.docx]

References

1. Gijsen R, Hoeymans N, Schellevis FG, Ruwaard D, Satariano WA, van den Bos GA. Causes and consequences of comorbidity: a review. J Clin Epidemiol 2001 Jul;54(7):661-674. [doi: 10.1016/s0895-4356(00)00363-2] [Medline: 11438406]
2. Multimorbidity: a priority for global health research. The Academy of Medical Science. 2018 Apr. URL: https://acmedsci.ac.uk/file-download/82222577 [accessed 2022-12-03]
3. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet 2012 Jul 07;380(9836):37-43 [FREE Full text] [doi: 10.1016/S0140-6736(12)60240-2] [Medline: 22579043]
4. Buetow S. Making the improbable probable: communication across models of medical practice. Health Care Anal 2014 Jun 29;22(2):160-173. [doi: 10.1007/s10728-012-0214-3] [Medline: 22743693]
5. Smith SM, Wallace E, O'Dowd T, Fortin M. Interventions for improving outcomes in patients with multimorbidity in primary care and community settings. Cochrane Database Syst Rev 2016 Mar 14;3(3):CD006560 [FREE Full text] [doi: 10.1002/14651858.CD006560.pub3] [Medline: 26976529]
6. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. JAMA 2002 Nov 20;288(19):2469-2475. [doi: 10.1001/jama.288.19.2469] [Medline: 12435261]
7. Leo DG, Buckley BJR, Chowdhury M, Harrison SL, Isanejad M, Lip GYH, TAILOR investigators. Interactive remote patient monitoring devices for managing chronic health conditions: systematic review and meta-analysis. J Med Internet Res 2022 Nov 03;24(11):e35508 [FREE Full text] [doi: 10.2196/35508] [Medline: 36326818]
8. Kim BY, Lee J. Smart devices for older adults managing chronic disease: a scoping review. JMIR Mhealth Uhealth 2017 May 23;5(5):e69 [FREE Full text] [doi: 10.2196/mhealth.7141] [Medline: 28536089]
9. Linn N, Goetzinger C, Regnaux J, Schmitz S, Dessenne C, Fagherazzi G, et al. Digital health interventions among people living with frailty: a scoping review. J Am Med Dir Assoc 2021 Sep;22(9):1802-1812 [FREE Full text] [doi: 10.1016/j.jamda.2021.04.012] [Medline: 34000266]
10. Jordan PW. An Introduction to Usability. London: Taylor & Francis; 1998.
11. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: a scoping review. Int J Med Inform 2019 Jun;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
12. REPORT on the eHealth Action Plan 2012-2020 - Innovative healthcare for the 21st century. European Commission. URL: https://www.europarl.europa.eu/doceo/document/A-7-2013-0443_EN.html [accessed 2022-11-06]
13. Global Strategy on Digital Health 2020-2025. World Health Organization. URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf [accessed 2022-11-15]
14. Abran A, Khelifi A, Suryn W, Seffah A. Usability meanings and interpretations in ISO standards. Softw Qual J 2003;11(4):325-338. [doi: 10.1023/A:1025869312943]
15. ISO/IEC 25022:2016 Systems and software engineering. Systems and software quality requirements and evaluation (SQuaRE). Measurement of quality in use. International Organization for Standardization (ISO). 2016. URL: https://www.iso.org/standard/35746.html [accessed 2022-12-03]
16. Ergonomics of human-system interaction - Part 11: Usability: Definitions and concepts (ISO 9241-11). Tienda. 2018. URL: https://tienda.aenor.com/norma-iso-9241-11-2018-063500 [accessed 2022-10-26]

17. Saeed N, Manzoor M, Khosravi P. An exploration of usability issues in telecare monitoring systems and possible solutions: a systematic literature review. Disabil Rehabil Assist Technol 2020 Apr;15(3):271-281. [doi: 10.1080/17483107.2019.1578998] [Medline: 30794009]
18. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Feb;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
19. Georgsson M, Staggers N. Quantifying usability: an evaluation of a diabetes mHealth system on effectiveness, efficiency, and satisfaction metrics with associated user characteristics. J Am Med Inform Assoc 2016 Jan;23(1):5-11 [FREE Full text] [doi: 10.1093/jamia/ocv099] [Medline: 26377990]
20. Georgsson M, Staggers N, Årsand E, Kushniruk A. Employing a user-centered cognitive walkthrough to evaluate a mHealth diabetes self-management application: A case study and beginning method validation. J Biomed Inform 2019 Mar;91:103110 [FREE Full text] [doi: 10.1016/j.jbi.2019.103110] [Medline: 30721757]
21. Ligons FM, Mello-Thoms C, Handler SM, Romagnoli KM, Hochheiser H. Assessing the impact of cognitive impairment on the usability of an electronic medication delivery unit in an assisted living population. Int J Med Inform 2014 Nov;83(11):841-848 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.07.004] [Medline: 25153770]
22. Wildenbos GA, Peute L, Jaspers M. Aging barriers influencing mobile health usability for older adults: a literature based framework (MOLD-US). Int J Med Inform 2018 Jun;114:66-75. [doi: 10.1016/j.ijmedinf.2018.03.012] [Medline: 29673606]
23. Doyle J, Murphy E, Kuiper J, Smith S, Hannigan C, Jacobs A, et al. Managing multimorbidity: identifying design requirements for a digital self-management tool to support older adults with multiple chronic conditions. 2019 Presented at: 2019 CHI Conference on Human Factors in Computing Systems; Glasgow, Scotland; May 4-9, 2019. [doi: 10.1145/3290605.3300629]
24. Chaudhry BM, Dasgupta D, Chawla N. Formative evaluation of a tablet application to support goal-oriented care in community-dwelling older adults. Proc ACM Hum Comput Interact 2022 Sep 20;6(MHCI):1-21. [doi: 10.1145/3546743]
25. Or C, Tao D. Usability study of a computer-based self-management system for older adults with chronic diseases. JMIR Res Protoc 2012 Nov 08;1(2):e13 [FREE Full text] [doi: 10.2196/resprot.2184] [Medline: 23612015]
26. Peute LW, Wildenbos G, Engelsma T, Lesselroth BJ, Lichtner V, Monkman H, et al. Overcoming challenges to inclusive user-based testing of health information technology with vulnerable older adults: recommendations from a human factors engineering expert inquiry. Yearb Med Inform 2022 Aug 02;31(1):74-81 [FREE Full text] [doi: 10.1055/s-0042-1742499] [Medline: 35654432]
27. Affective Based Integrated Care for Better quality of life (TeNDER Project). European Commission. URL: https://cordis.europa.eu/project/id/875325 [accessed 2022-11-02]
28. TeNDER. URL: https://www.tender-health.eu/project/ [accessed 2022-11-02]
29. Lozano Hernández CM, Medina-García R, de Hoyos-Alonso MC, Garrido-Barral A, Minué Lorenzo C, Sanz-Cuesta T, TeNDER Group Primary Care. Improvement in quality of life with the use of a technological system among patients with chronic disease followed up in primary care (TeNDER Project): protocol for a randomized controlled trial. JMIR Res Protoc 2023 Jul 03;12:e47331 [FREE Full text] [doi: 10.2196/47331] [Medline: 37399054]
30. Sánchez-Morillo D, Crespo M, León A, Crespo Foix LF. A novel multimodal tool for telemonitoring patients with COPD. Inform Health Soc Care 2015 Jan;40(1):1-22. [doi: 10.3109/17538157.2013.872114] [Medline: 24380372]
31. Sevilla-Gonzalez MDR, Moreno Loaeza L, Lazaro-Carrera LS, Bourguet Ramirez B, Vázquez Rodríguez A, Peralta-Pedrero ML, et al. Spanish version of the system usability scale for the assessment of electronic tools: development and validation. JMIR Hum Factors 2020 Dec 16;7(4):e21161 [FREE Full text] [doi: 10.2196/21161] [Medline: 33325828]
32. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
33. O'Halloran J, Miller G, Britt H. Defining chronic conditions for primary care with ICPC-2. Fam Pract 2004 Aug;21(4):381-386. [doi: 10.1093/fampra/cmh407] [Medline: 15249526]
34. Franke T, Attig C, Wessel D. A personal resource for technology interaction: development and validation of the Affinity for Technology Interaction (ATI) scale. Int J Hum Comput Interact 2018 Mar 30;35(6):456-467. [doi: 10.1080/10447318.2018.1456150]
35. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
36. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
37. Kim J, Lee HY, Christensen MC, Merighi JR. Technology access and use, and their associations with social engagement among older adults: do women and men differ? J Gerontol B Psychol Sci Soc Sci 2017 Sep 01;72(5):836-845. [doi: 10.1093/geronb/gbw123] [Medline: 28073816]
38. Smith A. Older adults and technology use. Pew Research Center. 2014 Apr. URL: http://www.pewinternet.org/2014/04/03/older-adults-and-technology-use/ [accessed 2022-11-02]

39. Wildenbos GA, Jaspers MWM, Schijven MP, Dusseljee-Peute LW. Mobile health for older adult patients: using an aging barriers framework to classify usability problems. Int J Med Inform 2019 Apr;124:68-77 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.01.006] [Medline: 30784429]
40. Kiat BW, Chen W. Mobile instant messaging for the elderly. Proced Comput Sci 2015;67:28-37. [doi: 10.1016/j.procs.2015.09.246]
41. Czaja SJ, Sharit J, Lee CC, Nair SN, Hernández MA, Arana N, et al. Factors influencing use of an e-health website in a community sample of older adults. J Am Med Inform Assoc 2013 Mar 01;20(2):277-284 [FREE Full text] [doi: 10.1136/amiajnl-2012-000876] [Medline: 22802269]
42. Milagros S, Lidia A, Cecilia C. Mujeres y digitalización. De las brechas a los algoritmos. Instituto de la Mujer y para la Igualdad de Oportunidades. Ministerio de Igualdad. 2020. URL: https://www.inmujeres.gob.es/diseno/novedades/M_MUJERES_Y_DIGITALIZACION_DE_LAS_BRECHAS_A_LOS_ALGORITMOS_04.pdf [accessed 2022-11-02]
43. Hargittai E, Dobransky K. Old dogs, new clicks: digital inequality in skills and uses among older adults. Can J Commun 2017 May 23;42(2):195-212. [doi: 10.22230/cjc.2017v42n2a3176]
44. Brown W, Yen P, Rojas M, Schnall R. Assessment of the Health IT Usability Evaluation Model (Health-ITUEM) for evaluating mobile health (mHealth) technology. J Biomed Inform 2013 Dec;46(6):1080-1087 [FREE Full text] [doi: 10.1016/j.jbi.2013.08.001] [Medline: 23973872]
45. Bolle S, Romijn G, Smets EMA, Loos EF, Kunneman M, van Weert JCM. Older cancer patients' user experiences with web-based health information tools: a think-aloud study. J Med Internet Res 2016 Jul 25;18(7):e208 [FREE Full text] [doi: 10.2196/jmir.5618] [Medline: 27457709]
46. Tahsin F, Tracy S, Chau E, Harvey S, Loganathan M, McKinstry B, et al. Exploring the relationship between the usability of a goal-oriented mobile health application and non-usage attrition in patients with multimorbidity: a blended data analysis approach. Digit Health 2021 Oct 05;7:20552076211045579 [FREE Full text] [doi: 10.1177/20552076211045579] [Medline: 34868614]
47. Baek JY, Na SH, Lee H, Jung H, Lee E, Jo M, et al. Implementation of an integrated home internet of things system for vulnerable older adults using a frailty-centered approach. Sci Rep 2022 Feb 04;12(1):1922. [doi: 10.1038/s41598-022-05963-9] [Medline: 35121795]
48. Medina-García R, Lopez-Rodríguez JA. A technological tool aimed at self-care in patients with multimorbidity: cross-sectional usability study (complete study data). Zenodo. URL: https://zenodo.org/records/10848551 [accessed 2024-03-21]

Abbreviations

ATI: affinity for technology interaction
DCL: data collection logbook
ICT: Information and communication technology
ISO: International Organization for Standardization
MMSE: Mini-Mental State Evaluation
REDCap: Research Electronic Data Capture
SUS: System Usability Scale
TeNDER: Affective Based Integrated Care for Better Quality of Life

Edited by Y Senathirajah; submitted 26.02.23; peer-reviewed by N Mungoli, B Chaudhry; comments to author 22.08.23; revised version received 05.09.23; accepted 28.02.24; published 05.04.24.

Please cite as:
Medina-García R, López-Rodríguez JA, Lozano-Hernández CM, Ruiz Bejerano V, Criscio P, Del Cura-González I, TeNDER Atención Primaria Group
A Technological Tool Aimed at Self-Care in Patients With Multimorbidity: Cross-Sectional Usability Study
JMIR Hum Factors 2024;11:e46811
URL: https://humanfactors.jmir.org/2024/1/e46811
doi:10.2196/46811
PMID:38578675

Original Paper

Multimodal In-Vehicle Hypoglycemia Warning for Drivers With Type 1 Diabetes: Design and Evaluation in Simulated and Real-World Driving

Caterina Bérubé[1], PhD; Martin Maritsch[1], PhD; Vera Franziska Lehmann[1,2], Dr med, PhD; Mathias Kraus[3], Prof Dr; Stefan Feuerriegel[1,4], Prof Dr; Thomas Züger[1,2,5], Dr med, PhD; Felix Wortmann[1,6], Prof Dr; Christoph Stettler[1,2], Prof Dr; Elgar Fleisch[1,6], Prof Dr; A Baki Kocaballi[7], PhD; Tobias Kowatsch[1,6,8], Prof Dr

[1]Centre for Digital Health Interventions, Department of Management, Technology and Economics, ETH Zurich, Zurich, Switzerland
[2]Department of Diabetes, Endocrinology, Nutritional Medicine and Metabolism, Bern University Hospital, University of Bern, Bern, Switzerland
[3]School of Business, Economics and Society, Friedrich-Alexander-Universit¨at Erlangen-Nürnberg, Nürnberg, Germany
[4]LMU Munich School of Management, LMU Munich, Munich, Germany
[5]Department of Endocrinology and Metabolic Diseases, Kantonsspital Olten, Olten, Switzerland
[6]Institute of Technology Management, University of St.Gallen, St Gallen, Switzerland
[7]School of Computer Science, University of Technology Sydney, Sydney, Australia
[8]Institute for Implementation Science in Health Care, University of Zurich, Zurich, Switzerland

Corresponding Author:
Tobias Kowatsch, Prof Dr
Institute of Technology Management
University of St.Gallen
Office 62-314
St.Jakob-Strasse 21
St Gallen, 9000
Switzerland
Phone: 41 71 224 7244
Email: tobias.kowatsch@unisg.ch

Abstract

Background: Hypoglycemia threatens cognitive function and driving safety. Previous research investigated in-vehicle voice assistants as hypoglycemia warnings. However, they could startle drivers. To address this, we combine voice warnings with ambient LEDs.

Objective: The study assesses the effect of in-vehicle multimodal warning on emotional reaction and technology acceptance among drivers with type 1 diabetes.

Methods: Two studies were conducted, one in simulated driving and the other in real-world driving. A quasi-experimental design included 2 independent variables (blood glucose phase and warning modality) and 1 main dependent variable (emotional reaction). Blood glucose was manipulated via intravenous catheters, and warning modality was manipulated by combining a tablet voice warning app and LEDs. Emotional reaction was measured physiologically via skin conductance response and subjectively with the Affective Slider and tested with a mixed-effect linear model. Secondary outcomes included self-reported technology acceptance. Participants were recruited from Bern University Hospital, Switzerland.

Results: The simulated and real-world driving studies involved 9 and 10 participants with type 1 diabetes, respectively. Both studies showed significant results in self-reported emotional reactions (P<.001). In simulated driving, neither warning modality nor blood glucose phase significantly affected self-reported arousal, but in real-world driving, both did ($F_{2,68}$=4.3; P<.05 and $F_{2,76}$=4.1; P=.03). Warning modality affected self-reported valence in simulated driving ($F_{2,68}$=3.9; P<.05), while blood glucose phase affected it in real-world driving ($F_{2,76}$=9.3; P<.001). Skin conductance response did not yield significant results neither in the simulated driving study (modality: $F_{2,68}$=2.46; P=.09, blood glucose phase: $F_{2,68}$=0.3; P=.74), nor in the real-world driving study (modality: $F_{2,76}$=0.8; P=.47, blood glucose phase: $F_{2,76}$=0.7; P=.5). In both simulated and real-world driving studies, the voice+LED warning modality was the most effective (simulated: mean 3.38, SD 1.06 and real-world: mean 3.5, SD 0.71) and urgent (simulated: mean 3.12, SD 0.64 and real-world: mean 3.6, SD 0.52). Annoyance varied across settings. The standard warning modality was the least effective (simulated: mean 2.25, SD 1.16 and real-world: mean 3.3, SD 1.06) and urgent (simulated: mean 1.88, SD 1.55 and real-world: mean 2.6, SD 1.26) and the most annoying (simulated: mean 2.25, SD 1.16 and real-world: mean 1.7, SD 0.95). In terms of preference, the voice warning modality outperformed the standard warning modality. In simulated driving, the voice+LED warning modality (mean rank 1.5, SD rank 0.82) was preferred over the voice (mean rank 2.2, SD rank 0.6) and standard (mean rank 2.4, SD rank 0.81) warning modalities, while in real-world driving, the voice+LED and voice warning modalities were equally preferred (mean rank 1.8, SD rank 0.79) to the standard warning modality (mean rank 2.4, SD rank 0.84).

Conclusions: Despite the mixed results, this paper highlights the potential of implementing voice assistant–based health warnings in cars and advocates for multimodal alerts to enhance hypoglycemia management while driving.

Trial Registration: ClinicalTrials.gov NCT05183191; https://classic.clinicaltrials.gov/ct2/show/NCT05183191, ClinicalTrials.gov NCT05308095; https://classic.clinicaltrials.gov/ct2/show/NCT05308095

(JMIR Hum Factors 2024;11:e46967) doi:10.2196/46967

KEYWORDS

digital health; voice assistant; ambient lighting; in-vehicle technology; health state; diabetes; hypoglycemia; warning; emotional reaction; technology acceptance; mobile phone; diabetes; implementation

Introduction

Overview

Around 9 million adults worldwide experience type 1 diabetes mellitus (T1DM) [1]. One of the most relevant acute complications associated with T1DM is hypoglycemia (ie, low blood glucose). This condition is associated with impaired cognitive, executive, and psychomotor function [2-4] and is linked to driving mishaps [5-7].

Previous work introduced the development of a voice warning for hypoglycemia while behind the wheel, whereas the voice assistant (VA) would work as a warning interface [8]. The hypoglycemia warning was intended as an app compatible with the VA that is already available in the car and that would allow delivering an alert in a hands-free manner. The study reported on the iterative development and evaluation of an in-vehicle hypoglycemia voice warning. It demonstrated that it is deemed useful and effective by drivers with T1DM, especially if the warning is kept simple and direct (ie, avoiding initiating a conversation with the driver). However, the paper did not investigate the effect of proactive behavior in the VA in such a context. Proactivity in VAs can cause a startling reaction, which is prone to annoyance [9] and driving impairments [10-12].

Ambient lighting can communicate with drivers without distracting them from their main task [13] or eliciting a strong emotional response [14]. This interface has been investigated as an indicator of several driving-relevant events, such as obstacle warnings or vehicle-state communication [15]. However, to the best of our knowledge, in-vehicle ambient lighting through LEDs has never been investigated as an indicator of a critical health state.

Our previous work [8] tested the concept of an in-vehicle voice warning delivered by the VA with healthy participants and then with individuals with T1DM, both in a driving simulator and a real car. The concept was developed following an iterative approach, and study participants provided feedback that we used to enhance the voice warning and test it on new participants. Thus, our previous work focused on technology acceptance and on improving it through user feedback. However, the voice warning was not evaluated against a standard warning (ie, beep with text), which could be used as a benchmark. Moreover, our previous work did not focus on the emotional reaction generated by getting a warning while driving. Therefore, this work investigates the effect of LEDs (ie, a possible solution to alleviate emotional reaction) and is to be understood as a continuation of our previous work. To foster experimental control and external validity, the same procedure is replicated in a simulated driving setting (ie, a computer simulator) and a real-world driving setting (ie, a car in a closed circuit).

Background

Hypoglycemia Warnings

Hypoglycemia is a common complication of diabetes. The monitoring of blood glucose is essential to prevent hypoglycemia. Intermittent self-monitoring of blood glucose, flash glucose monitoring, and continuous glucose monitoring are commonly used methods. However, these methods are not adapted to the in-vehicle context, as they require the driver to visually attend to a handheld mobile device displaying the current blood glucose value. This behavior is known to impair driving performance [16], thus leading to dangerous situations while driving.

Tentative hands-free solutions have been proposed to address this issue in academia [17] and in the community of individuals with T1DM [18]. Specifically, prior research [17] suggested using vehicles as a platform to display blood glucose data on infotainment screens. Moreover, a digital community [18] created an open-source program to show their continuous glucose monitoring data on infotainment screens. However, these solutions are limited to visual information display, thus failing to be ergonomically suitable for the in-vehicle context while driving. Therefore, solutions must be developed, which can provide hypoglycemia warnings while driving. One approach is to use voice-first warnings (ie, delivered by the built-in in-vehicle VA), where the driver can be informed of the issue without having to attend to a display.

In-Vehicle Warnings

In-vehicle health-state warning systems are a part of advanced driver–assistance systems [19]. From a human-computer interaction perspective, in-vehicle warnings should be effective and communicate urgency without being annoying [20]. Currently, in-vehicle warnings vary from classic car warnings, visually presented on the dashboard with traffic-light colors and unspecific tones, to advanced driver–assistance system warnings that use visual, auditory, and haptic modalities [21]. Even though the visual signal should be redundant to the auditory and haptic signals, some driver-state warnings, such as the driver attention alert, are predominantly visual (eg, mug symbol with an indicative text such as "Time for a break").

To decrease the demand for drivers' visual attention, it is necessary to develop attention-attractive warnings without relying on visual displays as the main source of information. One approach would be to use the in-vehicle VA already built into the car to warn or alert drivers for critical situations, such as hypoglycemia (or drowsiness). This approach could ensure warnings' effectiveness while reducing drivers' visual distraction.

In-Vehicle VAs

VAs are increasingly being integrated into cars [22-24], allowing digital health interventions to be delivered via such an interface in a scalable way. In the in-vehicle context, VAs have ergonomic and experiential advantages: they reduce visual distraction compared to other infotainment technologies [25], foster a natural interaction [26], create a sense of social presence [27], and increase engagement [28]. Ultimately, VAs can create a sense of being in the presence of a copilot [29]. Therefore, in-vehicle VAs have great potential for delivering real-time and effective hypoglycemia warnings to drivers while driving.

Proactive VAs and the Risk of Startle

Proactivity is not part of the current common mental model of a VA. However, it does not necessarily affect driving performance [30] and is well-accepted by drivers [31,32]. Nevertheless, a sudden auditory stimulus can create a startling reaction, which could interfere with driving performance [10,33]. Hence, when it comes to critical situations such as hypoglycemia [34], it is important to develop warnings that gradually prepare the driver to be receptive to them. Ambient lighting can be used to gradually prepare the driver and add information without consequentially distracting the driver [13]. This technology has been previously investigated for in-vehicle driving behavior support such as collision and blind warnings, lane change decision support, and speed and attention direction recommendations [15]. In addition, it has been investigated to inform the driver about the vehicle's decision-making in autonomous cars [35].

Objectives

To this end, a hypoglycemia warning delivered by a mock-up of a built-in in-vehicle VA is designed and tested with individuals with T1DM and compared to a standard format of in-car warning (ie, unspecific alert tone with visual information) regarding driver experience. Specifically, this work aims (1) to design a hands-free multimodal health intervention for hypoglycemia (ie, warning) compatible with the in-vehicle context and (2) to investigate the effect of warning modality (visual, vocal, and vocal with ambient lighting) on the emotional reaction and the acceptance of such technology.

Methods

Overview

Two studies were carried out, a simulated driving study and a real-world driving study. Across these studies, participant recruitment, study design, material and apparatus, procedure, and data analysis were the same. The difference lied in the setting. For this reason, all the following subsections, except for *Setting*, are described only once.

Participants

Sampling, Inclusion Criteria, and Compensation

Patients diagnosed with T1DM attending the diabetes outpatient clinic of the Bern University Hospital were recruited. A physician (VFL) of the study team performed recruitment during regular outpatient visits with a face-to-face assessment. For the simulated driving study, participants were recruited between November 2021 and March 2022. For the real-world driving study, participants were recruited between April and June 2022. Inclusion criteria were age between 21 and 60 years, hemoglobin A_{1c}≤9% (ie, a blood test indicating how well the patient's diabetes is being controlled), functional insulin treatment (with insulin pump therapy or multiple daily injections) for at least 3 months with good knowledge of insulin self-management, possession of a Swiss driver's license at least 3 years before study inclusion, and have driven at least once in the last 6 months. Each participant received an expense allowance of US $209.62 to cover general expenses caused by study participation (eg, transport).

Experience and Beliefs Questionnaires

Upon inclusion in the study, participants were asked to report the frequency of driving per week, their previous use of in-vehicle VAs, and their technology affinity. Technology affinity was assessed with the 16-item Technology Readiness Index [36]. This scale measures constructs susceptible to influencing the adoption of cutting-edge technology, such as optimism, innovativeness, discomfort, and insecurity.

Study Design

The study was designed as quasi-experimental with 2 independent variables, that is, blood glucose phase and warning modality, and 1 main dependent variable, that is, emotional reaction. The blood glucose variable had 3 levels, that is, euglycemia, decreasing, and hypoglycemia (see the *Procedure* section). The warning modality variable had 3 levels as well, that is, standard, voice, and voice+LED (see the *Warning* section). The blood glucose phase was varied in a nonrandomized fashion (see the *Procedure* section), while the warning modality was pseudorandomized, and each modality was crossed with each blood glucose phase. Secondary outcomes included self-reported user experience measures, such as warning acceptance, perceived urgency, alerting effectiveness, annoyance, and preference.

Material and Apparatus

Overview

In this section, the operationalization of the design variables is described. An overview is listed in Table 1.

Table 1. Study design variables.

Variable	Tool	Levels or values
Blood glucose[a]	Controlled hypoglycemia protocol [37]	Normal, decreasing, and hypoglycemia
Warning modality[a]	Hypoglycemia warning app [38]	Standard, voice, and voice+LED
Emotional reaction (objective)[b]	Empatica E4	Skin conductance response
Emotional reaction (subjective)[b]	Affective Slider [39]	Score (0-100)
Warning perceived urgency[b]	Baldwin and Moore scale [20]	Score (1=very urgent and 5=very insignificant)
Warning alerting effectiveness[b]	Baldwin and Moore scale [20]	Score (1=effective and 5=ineffective)
Warning annoyance[b]	Baldwin and Moore scale [20]	Score (1=I dislike it very much and 5=I like it very much)
Warning acceptance[b]	Van Der Laan Acceptance Scale [40]	Score (–2=negative extreme and +2=positive extreme)
Warning preference[b]	Arbitrary 3-point scale	Rank (1=best and 3=worst)

[a]Manipulation.

[b]Measure.

Blood Glucose

Blood glucose was manipulated by inserting 2 intravenous catheters: one for blood glucose measurement with an interval of 5-10 minutes and the other for the infusion of a combination of insulin and glucose, according to the patient's current blood glucose and the experimental target blood glucose range. Euglycemia (ie, normal blood glucose) was defined as a concentration of 5-8 mmol/L; decreasing blood glucose was identified when blood glucose was below the euglycemia range (5-8 mmol/L) and progressing toward a target hypoglycemic range (3-3.5 mmol/L); and hypoglycemia was defined as a concentration of 3-3.5 mmol/L. For more technical details, refer to related research [37].

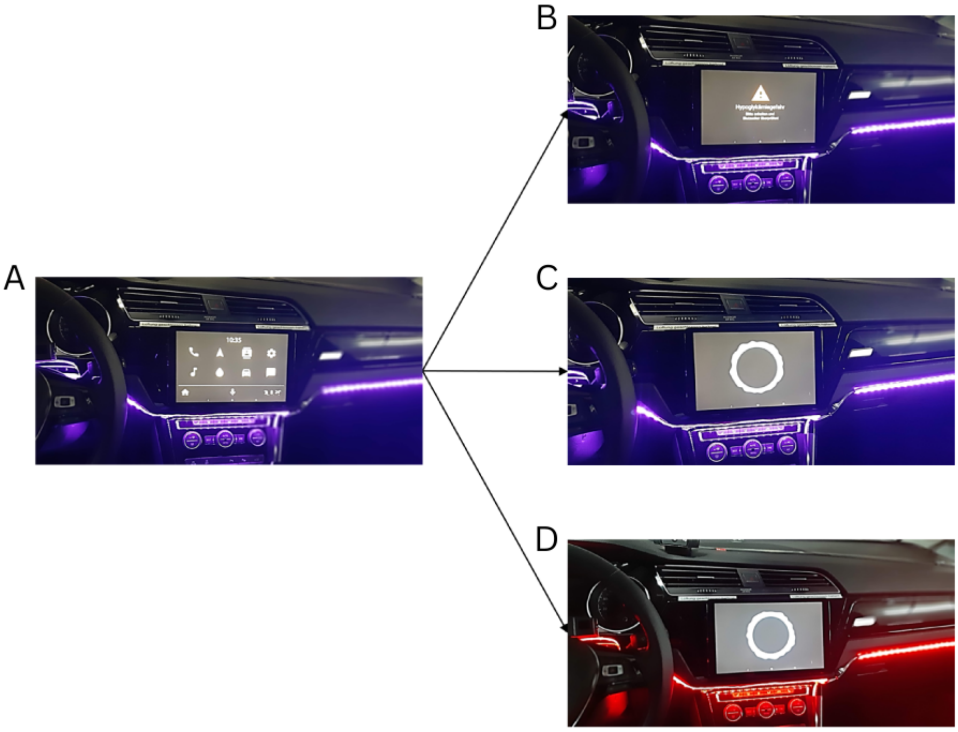
Warning

The Hypoglycemia warning app operationalized into a tablet (model SM-T590, Samsung) simulated the infotainment system's touchscreen, and LED strips (RGB Light Strip Pro, Cololight) simulated the interior ambient lighting. The warning system was controlled through the Wizard of Oz method [41], that is, the system was controlled by the experimenter behind the scene and acting as if it was fully automated.

The LED strips had 60 LEDs per meter and could be remotely controlled with the Cololight app (Klaus Stephan GmbH). The tablet was used to run the Hypoglycemia warning app (publicly available on GitHub [38]), remotely controlled via a remote desktop app (AnyDesk, AnyDesk Software GmbH).

The warning system had 4 possible states: a default state and 3 intervention modalities (standard, voice, and voice+LED). The default state involved the LED strip being turned on in blue and the tablet showing a fake navigation menu (Figure 1A). The standard modality displayed a yellow warning sign with an informative text and was accompanied by an earcon. The text said "Risk of hypoglycemia. Please pull over and verify blood sugar." The LED strips remained blue (Figure 1B). The voice modality displayed a VA animation accompanied by a prerecorded synthesized female voice (de-DE-Wavenet-C with speed=0.85 and pitch=–3.20, Google Inc). The voice said "I have detected a risk of hypoglycemia. Please pull over safely and verify your blood sugar" (translation from the German formulation "*Ich habe eine Hypogefahr erkannt. Bitte sicher anhalten und deinen Blutzucker überprüfen*"). The voice warning was designed based on the results reported in our previous work [8]. Once again, the LED strips remained blue (Figure 1C). The voice+LED modality displayed the same VA animation and prerecorded synthesized female voice but, before the onset of the voice warning, the LED strips turned red (Figure 1D).

Figure 1. Illustration of the warning app and modalities. (A) The default state was shown during the drive and simulated the infotainment menu of the car. The (B) standard, (C) voice, and (D) voice+LED modalities were activated when a warning was delivered.



Participants knew that they would receive a warning during each drive. Still, the warning presentation was pseudorandomized, where they would receive a complete permutation of the 3 warning modalities within a blood glucose phase. Participants were not explicitly informed about which warning was "the intervention of interest" and which one was the "comparator."

Objective Emotional Reaction

Emotional reaction was measured physiologically through skin conductance response (SCR). SCR is the result of the sympathetic nervous system promptly regulating the activity of the sweat glands in response to a stimulus. This measure is associated with emotional arousal [42] and can be used to measure event-related emotional reactions objectively [43]. In this study, the Empatica E4 (Empatica Inc), a Conformité Européenne–certified wristband collecting physiological data in real time, was used. Participants wore the E4 during the main visit (see the *Procedure* section). Note that this measure provided solely the arousal dimension of emotion.

Subjective Emotional Reaction

Emotional reaction was also measured subjectively through the Affective Slider [39]. This digital scale is a self-reporting tool measuring valence and arousal on 2 separate sliders. Participants did not see any numerical anchor, but the score ranged from 0 to 100. Thus, valence is rated between a frowning and a smiling face (0=frowning and 100=smiling), and arousal between a sleepy and a widely awake face (0=sleepy and 100=widely awake).

Warning Perceived Urgency, Alerting Effectiveness, and Annoyance

To measure the perceived urgency mapping of the 3 modalities, for each modality, participants rated the perceived urgency, alerting effectiveness, and annoyance according to a scale from prior work [20]. The 3 dimensions were rated using a 5-point Likert scale (1=very urgent or effective or I like it very much and 5=very insignificant or ineffective or I dislike it very much). This questionnaire was filled out during the posttest visit (see the *Procedure* section).

Warning Acceptance

To compare the acceptance of the 3 modalities, participants filled out the Van Der Laan Acceptance Scale [40], once per modality. This scale consists of the 2 constructs, usefulness and satisfaction, with items answered on a 5-point semantic differential from –2 to +2, which means participants had to select a point between 2 opposite adjectives (eg, unpleasant or pleasant). This questionnaire was filled out during the posttest visit (see the *Procedure* section).

Warning Preference

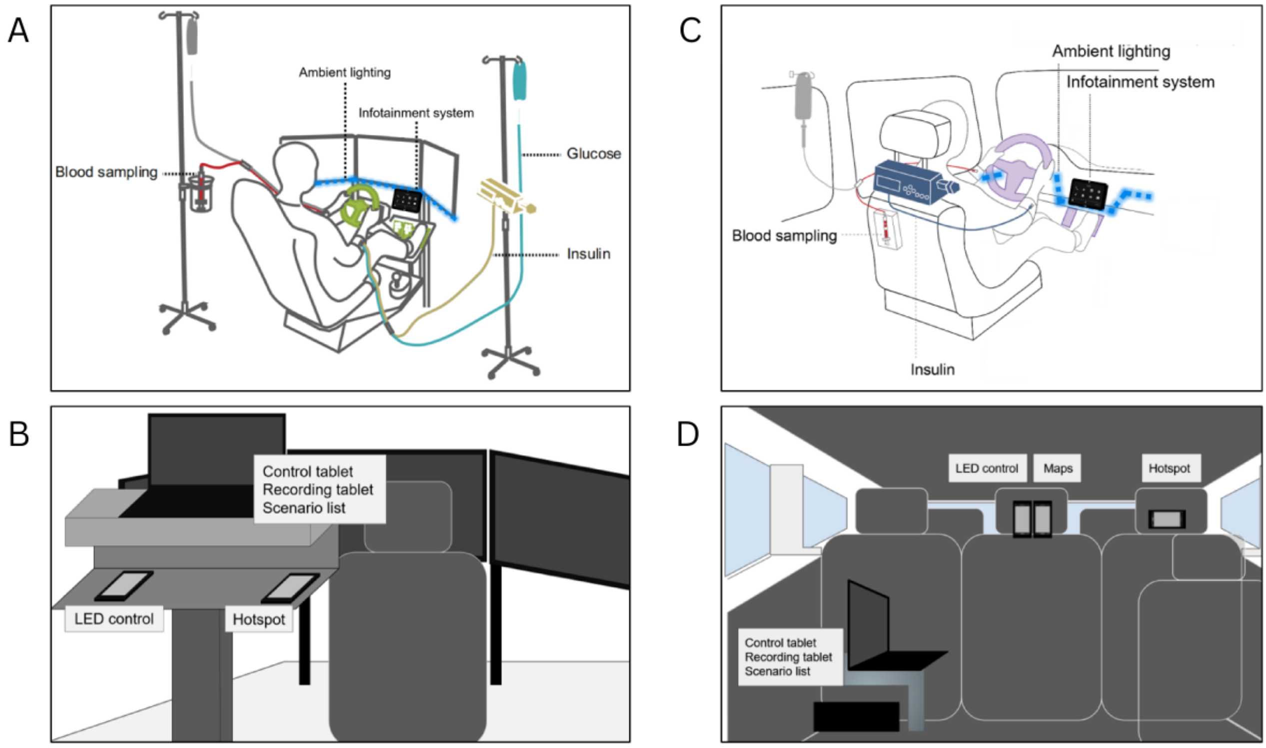
To formalize their preference, patients were asked to rank the 3 modalities from best to worst. The scale was implemented as a radio button questionnaire with 1 item per modality (ie, beep with a warning sign and text, voice, and LED with voice) and a 3-point scale (ie, 1=best and 3=worst). Participants were also encouraged to provide comments to their answers, which were topically (ie, without verbatim transcription) recorded in written form. This questionnaire was filled out during the posttest visit (see the *Procedure* section).

Setting

Simulated Driving

Patients used a driving simulator (Carnetsoft Inc), featuring 3 monitors displaying the front, left, and right views of the driving cabin. The central monitor also displayed the cockpit and navigation arrows, which directed the patient through the environment. To control the simulator, participants used a steering wheel and pedals (Logitech Driving Force G29, Logitech), set to the automatic transmission. The simulator was connected to a stereo speaker and maintained at a constant volume (Figure 2A and C). The infotainment system simulator tablet (see the *Warning* section) was placed under the right side of the central monitor and connected to the simulator's sound system. The LED lights were attached at the bottom of the 3 monitors. Figure 2A and 2C illustrates the patient's setup.

Figure 2. Comparison of patient's and experimenter's setup between simulated and real-world driving. The left figures represent the simulated driving setting, the right figures represent the real-world driving setting, the top figures (A and C) show the patient's setup, and the bottom figures (B and D) show the experimenter's setup.



The experimenter was standing behind the patient and controlled the LED stripes with a smartphone (Redmi Note, Xiaomi Inc) and the tablet via a laptop computer (ThinkPad X1 Carbon, Lenovo PC HK Ltd). A stopwatch app was used to manually onset the warning. Figure 2B and 2D illustrates the experimenter's setup.

Three environments were used, namely, highway, countryside, and town, with the highway being the easiest to navigate due to variable traffic but no turns, the countryside having a moderated amount of traffic and turns, and the town being the most difficult with the most turns and traffic. Participants drove in the environments for about 5 minutes before receiving a hypoglycemia warning (run-in phase).

Real-World Driving

Patients drove in a minivan (Touran, Volkswagen) with an automatic transmission. The car was equipped with dual pedals to allow for intervention from a trained driving instructor, in case of emergency. In case the instructor needed to intervene, the event was recorded.

The infotainment system simulator tablet (see the *Warning* section) was placed on top of the infotainment screen and connected to the car's sound system, maintained at a constant volume. The LED lights were attached along the cockpit from the left to the right extremities, passing by under the steering wheel, the infotainment system, and above the aperture of the glove compartment. The experimenter was sitting in the third row of the car and controlled the LED stripes with a smartphone and the tablet via a laptop computer (6th Gen ThinkPad X1 Carbon, Lenovo PC HK Ltd). A Google Map was used to manually onset the warning.

Patients were exposed to real-world driving on a test track provided by the Swiss Federal Department of Defense, Civil Protection and Sports. The driving scenarios on the track were designed to correspond with simulated environments used in the simulator setting (ie, highway, countryside, and town), featuring various driving elements such as turns, crossroads, stop signs, and a pedestrian crossing equipped with a dummy. As traffic simulation was not feasible, artificial obstacles, including boxes and traffic pylons, were used. Participants drove in the environments for 5-7 minutes before receiving a hypoglycemia warning (run-in phase).

Ethical Considerations

The experiments were approved within the context of this project by the cantonal ethics commission of Bern, Switzerland

(BASEC2020-00685 and BASEC2021-02381). Before any study-related procedure, informed consent specifying the analysis and the study protocol presented in this paper was obtained in written form from all participants. All collected data were deidentified by associating individual data to a numerical identification number. The data reported in this paper are part of the HEADWIND Study, a clinical trial registered under ClinicalTrials.gov (Part 3: NCT05183191 and Part 4: NCT05308095).

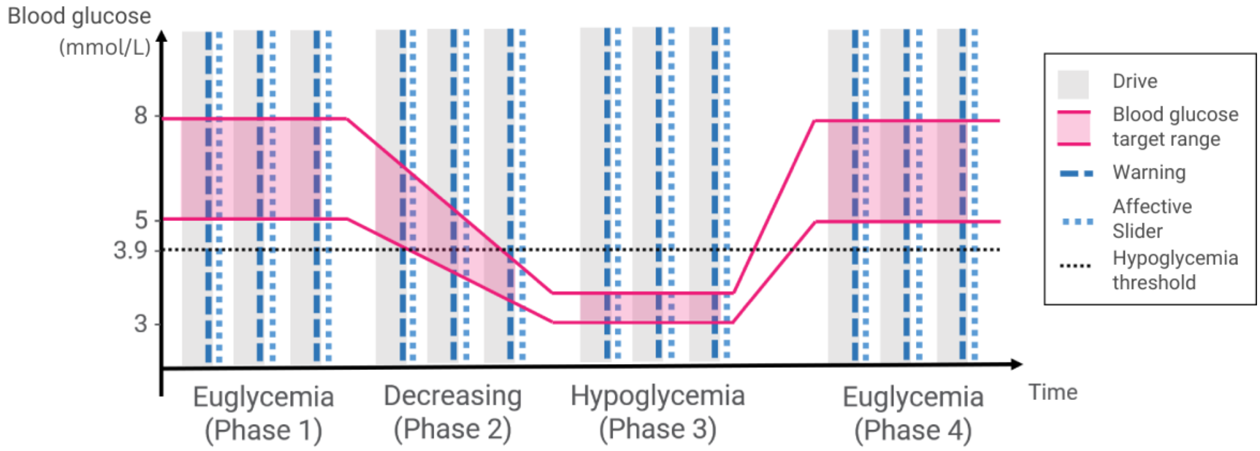
Procedure

The procedure was divided into 3 visits. In a pretest visit, patients were welcomed to the Bern University Hospital, informed about the procedure and the warnings, and asked to fill out demographic and experience and beliefs questionnaires.

In the main visit, participants were welcomed to the relevant setting for the blood glucose manipulation and the objective emotional reaction measurements. Participants were aware that their blood sugar would be manipulated to reach specific goal ranges, but they were not informed of their blood glucose during the experiment. In the real-world driving setting, the driving instructor was also aware of the blood glucose manipulation and was blinded to the current blood glucose. The experimental team was not blinded to the blood glucose level.

Each participant went through a fixed sequence of blood glucose phases: (1) a first drive with normal blood glucose (phase 1: euglycemia), where participants were first experiencing all types of environment and warning (and was thus considered as training); (2) a phase where blood glucose was progressively decreasing toward the moderate hypoglycemia threshold (ie, <3.0 mmol/L) with a target range of 3-3.5 mmol/L (phase 2: decreasing); (3) a phase with stable moderate hypoglycemia (phase 3: hypoglycemia); and (4) a final phase with normal blood glucose (phase 4: euglycemia). A warning was delivered at the end of each drive to explore the effect of the blood glucose phase. Participants drove in the 3 types of environments in each phase. The sequence of environment type was pseudorandomized [8], that is, participants were exposed to all 3 warning modalities within each phase, but the sequence of modalities within 1 phase was random. Similarly, the warning modality was pseudorandomized to balance modality with environments. Once participants received a warning, they were expected to stop, and the drive came to an end. At the end of each drive, participants filled out the Affective Slider referring to the warning they just received. Figure 3 shows an overview of the procedure.

Figure 3. Overview of the procedure. The vertical gray bars represent the drives, the ribbon delimited by solid lines represents the blood glucose manipulation, the vertical dashed lines represent the warning deliveries, the vertical dotted lines represent the Affective Slider submission, and the horizontal dotted line represents the hypoglycemia threshold.



In a posttest visit, participants were once more exposed to the 3 warning modalities and were required, after each exposure, to fill the Baldwin and Moore scale [20] and the Van Der Laan Acceptance Scale. Finally, they ranked the 3 warning modalities.

Data Analysis

The continuous variables of sample characteristics (ie, demographics and previous experience) are presented with mean and SD. Frequency variables of sample characteristics are presented in count numbers (ie, n) and percentages of the total experiment sample.

Emotional reaction (objective and subjective) measures were analyzed as a function of blood glucose (excluding phase 1) and warning modality and verified with a mixed-effects linear model, ANOVA test, and a significance threshold of P=.05. Effect size was calculated with partial η^2 (0.01 indicates a small effect, 0.06 indicates a medium effect, and 0.14 indicates a large effect). Moreover, the objective emotional reaction was analyzed following established guidelines [43]: SCR (ie, rapid phasic component) was standardized for individual differences by dividing the SCR signal by the individual maximum SCR and by reducing the noise. In addition, SCR was calculated by considering the change in skin conductance between the average skin conductance in the 5-second window before the warning onset and the average skin conductance in the 5-second window after the warning onset itself (including latency of 1 second). For each measure of emotional reaction, a mixed-effects linear model was estimated with warning modality (3 levels: standard, voice, and voice+LED) and blood glucose (3 levels: normal, decreasing, and hypoglycemia) as independent variables and with emotional reaction measure (ie, either self-reported arousal, self-reported valence, or SCR) as the dependent variable.

Warning evaluations (ie, perceived urgency, alerting effectiveness, annoyance, and acceptance) were aggregated with means and SDs, presented in a table. Moreover, perceived urgency, alerting effectiveness, and annoyance were centered on 2 to match the acceptance scores, for the sake of comparison.

Preference ranking was aggregated across modalities in the frequency of rank as best, middle, or worst (ie, how many times 1 modality was ranked as the best, middle, or worst in comparison to the other 2). If participants were commenting on their choices, highlight note-taking was performed.

Data analysis and graphical representations were performed using RStudio (Posit Software) packages such as *lmerTest* or mixed-effects linear modeling and *ggplot2* and patchwork for data visualization. All results are separated by experiment (ie, simulated vs real-world driving) and juxtaposed to allow direct comparison.

Results

Overview

The data of 2 participants of the simulated driving study were excluded due to partial data loss. In the real-world driving setting, the driving instruction had to intervene in 1 instance, as the participants did not follow the driving path (ie, did not turn left) during phase 4 (ie, while in euglycemia).

Sample Characteristics

Overall, the majority were male participants, who drove multiple times per week and did not have previous experience with in-vehicle VAs. Participants were approximately 40 years of age and had a Technology Readiness Index between 3.5 and 4 (over a maximum of 5). Details are shown in Table 2.

Table 2. Sample characteristics across studies.

Characteristics	Simulated (n=9)	Real-world (n=10)
Sex, n (%)		
Male	8 (89)	7 (67)
Female	1 (11)	3 (33)
Age (years), mean (SD)	45.7 (11.79)	37.3 (11.1)
Frequency of driving, n (%)		
1 time per month	1 (11)	1 (10)
2-5 times per month	—[a]	2 (20)
2-5 times per week	5 (56)	4 (40)
Every day	3 (33)	3 (30)
Previous use of in-vehicle voice assistants, n (%)		
Never	4 (44)	6 (60)
Rarely	1 (11)	2 (20)
Sometimes	2 (22)	1 (10)
Often	2 (22)	1 (10)
TRI[b] (over a maximum of 5), mean (SD)	3.6 (0.62)	3.9 (0.5)

[a]Not available.

[b]TRI: Technology Readiness Index.

Emotional Reaction

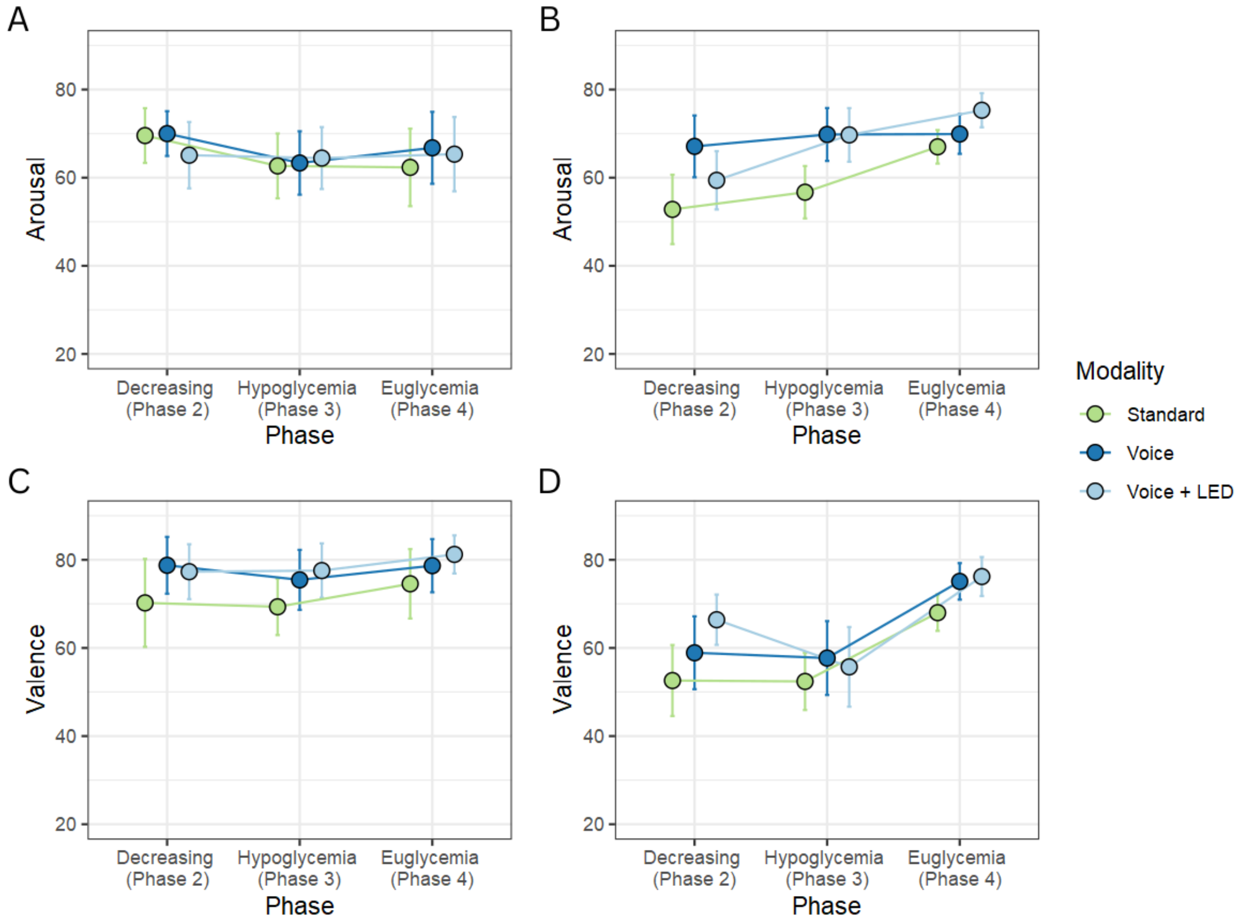
Overview

In this section, the results of the self-reported and physiological measures of emotional reaction (ie, self-reported valence and arousal and SCR) are described. A mixed-effect model was run on all these measures, with warning modality and blood glucose phase as independent variables.

Self-Reported Arousal and Valence

According to our results, the mixed-effect models were significant for both valence and arousal in both studies (P<.001). Figure 4 shows the means and SEs for arousal and valence.

Figure 4. Line plots of arousal and valence across warning modality and blood glucose phases across (A and C) simulated and (B and D) real-world driving. Error bars represent SEs.



In the simulated driving study, arousal was not significantly affected by either of the independent variables (modality: $F_{2,68}$=0.1; P=.91; partial η^2=0, blood glucose phase: $F_{2,68}$=1.1; P=.35; partial η^2=0.03). Valence was significantly affected by warning modality ($F_{2,68}$=3.9; P=.03; partial η^2=0.10) but not by blood glucose phase ($F_{2,68}$=1.1; P=.35; partial η^2=0.03).

In the real-world driving study, arousal was significantly affected both by warning modality ($F_{2,76}$=4.3; P=.02; partial η^2=0.10) and blood glucose phase ($F_{2,76}$=4.1; P=.02; partial η^2=0.10). Valence was significantly affected by blood glucose phase ($F_{2,76}$=9.3; P<.001; partial η^2=0.20) but not by warning modality ($F_{2,76}$=2; P=.14; partial η^2=0.05).

Physiological Arousal

According to our results, the mixed-effect models were not significant in both studies. Hence, the warning modality and blood glucose phase did not significantly affect physiological arousal measured via SCR neither in the simulated driving study (modality: $F_{2,68}$=2.46; P=.09; partial η^2=0.1, blood glucose phase: $F_{2,68}$=0.3; P=.74; partial η^2=0). nor in the real-world driving study (modality: $F_{2,76}$=0.8; P=.47; partial η^2=0, blood glucose phase: $F_{2,76}$=0.7; P=.5; partial η^2=0).

Technology Acceptance

In this section, the technology acceptance results (ie, Baldwin and Moore scales of urgency, effectiveness, and annoyance and the Van Der Laan Acceptance Scale) are described. Details are available in Table 3 [44].

Table 3. Technology acceptance measure across studies.

Measure and warning modalities	Simulated driving, mean (SD)	Real-world driving, mean (SD)
Urgency		
Standard	1.88 (1.55)	2.6 (1.26)
Voice	2.88 (1)	3 (0.82)
Voice+LED	3.12 (0.64)	3.6 (0.52)
Effectiveness		
Standard	2.25 (1.16)	3.3 (1.06)
Voice	2.75 (1.28)	3.6 (0.7)
Voice+LED	3.38 (1.06)	3.5 (0.71)
Annoyance		
Standard	2.25 (1.16)	1.7 (0.95)
Voice	0.88 (0.84)	1 (0.82)
Voice+LED	0.5 (0.76)	1.1 (1.2)
Acceptance		
Standard	3.35 (0.55)	3.52 (0.53)
Voice	3.74 (0.47)	3.88 (0.41)
Voice+LED	3.86 (0.47)	3.77 (0.61)

In the simulated driving study, the voice+LED modality elicited the highest sense of urgency and effectiveness, the least annoyance, and the highest acceptance, followed by the voice modality. In real-world driving, the voice+LED modality elicited the highest sense of urgency and least annoyance, while the voice modality elicited the most sense of effectiveness and the highest acceptance.

Preference Ranking

The average rank in the simulated driving study was 1.5 (SD 0.82) for the voice+LED modality, 2.2 (SD 0.6) for the voice modality, and 2.4 (SD 0.81) for the standard modality. The average rank in real-world driving was 1.8 (SD 0.79) for both the voice+LED and voice modalities and 2.4 (SD 0.84) for the standard modality.

In the real-world driving study, topical feedback showed that 6 participants mentioned that the light was not noticeable while driving (eg, "I have not noticed the light but at night, it certainly works better" [Participant 4]).

Discussion

Principal Findings

This study investigated the effect of warning modality (visual, vocal, and vocal with ambient lighting) on the emotional reaction and the acceptance of such a technology. Our results showed that voice warnings are more appreciated and considered more effective than standard warnings. However, the ambient lighting did affect such judgments.

Effects of Warning Modality on Emotional Reaction

Effect on SCR

No significant effect of warning modality (or blood glucose phase) on skin conductance was found. SCR measured through Empatica E4 has been previously shown to be linked with response to stimuli [45-47]. Moreover, it has been associated with blood glucose variation [48,49]. Therefore, we may consider the possibility that the measurement protocol used in this research may have experienced certain weaknesses and have affected the validity of the obtained results. Hence, we cannot consider the lack of significant results as negative evidence. Nevertheless, future research should further investigate the startling effect of the voice warning while driving, either by replicating our experimental setting, by using alternative electrodermal activity measurement tools [50], or by using other measures of emotional reaction, such as eye blinks [44].

Self-Reports

Our results showed that although in simulated driving the effect of modality on self-reported arousal was not significant, this was the case in real-world driving. In particular, higher arousal was observed during decreasing glucose and hypoglycemia for the voice and voice+LED modalities. Moreover, in the simulated driving study, the results on self-reported valence show a significant effect of modality, with voice+LED and voice warnings eliciting higher valence than a standard warning, particularly during decreasing glucose and hypoglycemia. This was not the case in the real-world driving study.

Despite the mixed results, the warning modality had a significant effect in the critical moments, that is, when the participants were about to experience or already experiencing hypoglycemia. Thus, our results showed the relevance of measuring emotional reaction at different levels of blood glucose. While the Affective Slider is a very efficient measurement tool, it is important to note that some participants expressed a lack of confidence in self-reporting their emotions with it. Thus, future research might benefit from using alternative self-reported measures of emotion, such as the Positive and Negative Affect Schedule [51] or the Discreet Emotions Questionnaire [52].

Mixed Results

These mixed results preclude the formulation of definitive conclusions regarding the effect of modality on emotional reaction based on this study. As a sudden auditory stimulus can create a startling reaction, which can interfere with driving performance [10,33], future research should consider our recommendations and further investigate the design of a warning that is both effective and nonstartling.

Effects of Warning Modality on Acceptance and Preference

Acceptance

Our results demonstrated that the voice+LED modality tended to be the most valued regarding acceptance and preference. However, in the real-world driving study, this advantage, compared to the voice modality, seemed to decrease compared to the simulated driving study. This change might be due to the setup, where the ambient lighting was more visible in the laboratory than outdoors. Therefore, the advantage of the ambient lighting might have decreased in the real-world driving study. Thus, from the results, it is clear that the voice warning had an advantage over the standard warning, while the addition of ambient lighting (ie, voice+LED modality) did not bring a substantial advantage.

Preference

Finally, when asking participants for a posttest ranking of the warning modalities, results showed that voice (ie, both voice and voice+LED modalities) had a constant advantage over a beep with text (ie, standard modality). However, in the simulated driving study, the voice+LED modality was ranked first more often than in the real-world driving study, leading to interpret these results similar to the technology acceptance results, that is, adding ambient lighting to a voice warning was not considered substantially advantageous. These results might be influenced by the perception of the lights in daylight conditions, which differed between the simulated and the real-world driving settings. Based on the topical feedback from the participants experiencing the real-world driving setting, it might be that the contrast between the exterior and interior luminance was too low. As we did not measure the contrast in luminance between the daylight and the LEDs, future research should further investigate the use of ambient lighting as a component of in-vehicle warnings with greater control on luminance [53].

Implications

This paper involves implications both from health and automotive perceptive. First, our investigation represents a step forward in managing a real-life hazard associated with hypoglycemia. Our previous work [8] focused on designing an effective voice warning. This work compared it to a standard warning with an unspecified auditory signal with a text, and with an addition of ambient lighting, to make the voice warning less abrupt. Our results show an advantage for a voice warning (ie, spoken) over a tone with text but not for ambient lighting. While participants showed a higher preference for adding ambient lighting in the simulated driving study, the way the ambient lighting was set in the real-world driving study was not noticeable enough to replicate the results. Other work has investigated more attention-grabbing ambient lighting, such as blinking lights [54]. Future research should investigate if different typologies of ambient light patterns could affect emotional reaction and acceptance.

Second, our investigation aims to inspire in-vehicle technology designers to develop in-vehicle health warnings using the in-vehicle VA. While there are driver-state warnings, they predominantly alarm the driver with an unspecific beep and a text on the cockpit. Future research should investigate how using the in-vehicle VA to warn the driver about dizziness or lack of attention would be accepted by drivers. Moreover, providing health-related warnings while driving fits the concept of “health-conscious” cars [55,56]. Along the lines of our investigation, there have been some attempts to develop blood glucose monitoring interfaces for the car [17,18]. However, they primarily rely on visual displays and are ill-adapted to the context of driving.

Finally, this study assumed the delivery of a hypoglycemia warning in a car with an autonomy of level 0 (ie, no automation) or level 1 (ie, with driver assistance). As cars are becoming increasingly automated, a hypoglycemia warning should be compatible with cars with a higher level of autonomy. However, the warning designed in this work is compatible with higher levels of automation. For instance, during autonomous driving, the in-vehicle VA could alert the driver that hypoglycemia has been detected and trigger the car to autonomously stop. During manual driving, the in-vehicle VA could warn the driver and trigger the car to take over (switch from manual to autonomous driving) and pull over.

Limitations and Future Research

Despite our best efforts, this investigation involved certain limitations. First, the sample size was rather small. Nevertheless, valuable insights on digital solutions can be provided with a small sample size [57-59]. Thus, although it does not allow drawing conclusions on the interaction of drivers with T1DM with in-vehicle hypoglycemia warnings, it motivates further research in this domain.

Second, emotional response to the warning was evaluated in different blood glucose states and a controlled setting (ie, simulator and closed circuit). However, the warning was not delivered when relevant (ie, only when the driver was actually undergoing hypoglycemia) or while the participant was driving on a public road unaware of the upcoming critical state. While our method allowed controlling for the blood glucose in the emotional response, it did not allow us to find the opportune moment for intervention delivery, that is, at what point of upcoming hypoglycemia is the warning most appropriate (both in terms of emotional reaction and acceptance). Future research should investigate the effectiveness of such an intervention in a more realistic context, where the driver does not expect to be warned and is actually about to undergo hypoglycemia while on a public road.

Conclusions

This paper proposes the use of the in-vehicle VA and ambient lighting system to deliver a hypoglycemia warning, ensuring a hands-free alert. The investigation focused on the extent to which warning modality could affect emotional response and acceptance both in a simulated and real-world environment. Although further investigations are needed, our results suggest, together with our previous work [8], that implementing multimodal warnings can improve the management of hypoglycemia in cars and also emphasize the potential of in-vehicle VA for delivering health-related warnings.

Acknowledgments

This work was part-funded by the Swiss National Science Foundation, Sinergia-Project CRSII5_183569 "Design and Evaluation of a Vehicle Hypoglycaemia Warning System in Diabetes (HEADWIND Project)."

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

CB, ABK, and TK were responsible for the warning design and development. CB, MM, FW, and TK were responsible for the design of the driving scenarios. CB, MM, and VFL were responsible for the data collection on the intervention. CB was responsible for the data analysis and the first draft of this paper. All authors were responsible for critical feedback and final revisions of the paper.

Conflicts of Interest

TK is affiliated with the Centre for Digital Health Interventions, a joint initiative of the Institute for Implementation Science in Health Care, University of Zurich; the Department of Management, Technology, and Economics at ETH Zurich; and the Institute of Technology Management and School of Medicine at the University of St.Gallen. The Centre for Digital Health Interventions is funded in part by CSS, a Swiss health insurer; Mavie Next, an Austrian health insurer; and MTIP, a Swiss digital health investor. TK is also a cofounder of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways. However, neither CSS, Mavie Next, MTIP nor Pathmate Technologies were involved in this research.

References

1. Green A, Hede SM, Patterson CC, Wild SH, Imperatore G, Roglic G, et al. Type 1 diabetes in 2017: global estimates of incident and prevalent cases in children and adults. Diabetologia 2021;64(12):2741-2750 [FREE Full text] [doi: 10.1007/s00125-021-05571-8] [Medline: 34599655]
2. Allen KV, Pickering MJ, Zammitt NN, Hartsuiker RJ, Traxler MJ, Frier BM, et al. Effects of acute hypoglycemia on working memory and language processing in adults with and without type 1 diabetes. Diabetes Care 2015;38(6):1108-1115 [FREE Full text] [doi: 10.2337/dc14-1657] [Medline: 25758768]
3. Warren RE, Frier BM. Hypoglycaemia and cognitive function. Diabetes Obes Metab 2005;7(5):493-503. [doi: 10.1111/j.1463-1326.2004.00421.x] [Medline: 16050942]
4. Cryer PE. Hypoglycemia: Pathophysiology, Diagnosis, and Treatment. Oxford: Oxford University Press; 1997.
5. Blumer I, Kenshole AB, Stilman J, Lewis GF. Insulin-treated diabetes and driving: legal jeopardy and consequences of hypoglycemia. Can J Diabetes 2019;43(3):221-223. [doi: 10.1016/j.jcjd.2018.08.201] [Medline: 30348593]
6. Cox DJ, Gonder-Frederick LA, Kovatchev BP, Clarke WL. Self-treatment of hypoglycemia while driving. Diabetes Res Clin Pract 2001;54(1):17-26. [doi: 10.1016/s0168-8227(01)00274-1] [Medline: 11532326]
7. Cox DJ, Penberthy JK, Zrebiec J, Weinger K, Aikens JE, Frier B, et al. Diabetes and driving mishaps: frequency and correlations from a multinational survey. Diabetes Care 2003;26(8):2329-2334 [FREE Full text] [doi: 10.2337/diacare.26.8.2329] [Medline: 12882857]
8. Bérubé C, Lehmann VF, Maritsch M, Kraus M, Feuerriegel S, Wortmann F, et al. Effectiveness and user perception of an in-vehicle voice warning for hypoglycemia: development and feasibility trial. JMIR Hum Factors 2024 Jan 09;11:e42823 [FREE Full text] [doi: 10.2196/42823] [Medline: 38194257]
9. Landry S, Jeon M, Lautala P, Nelson D. Design and assessment of in-vehicle auditory alerts for highway-rail grade crossings. Transp Res F: Traffic Psychol Behav 2019;62:228-245. [doi: 10.1016/j.trf.2018.12.024]
10. Biondi F, Rossi R, Gastaldi M, Mulatti C. Beeping ADAS: reflexive effect on drivers' behavior. Transp Res F: Traffic Psychol Behav 2014;25(A):27-33 [FREE Full text] [doi: 10.1016/j.trf.2014.04.020]

11. Larsson P, Opperud A, Fredriksson K, Västfjäll D. Emotional and behavioural response to auditory icons and earcons in driver-vehicle interfaces. 2009 Presented at: Proceedings of the 21st International Technical Conference on the Enhanced Safety of Vehicles; June 2009; Stuttgart, Germany URL: https://tinyurl.com/msz254ve
12. Larsson P, Västfjäll D. Emotional and behavioural responses to auditory interfaces in commercial vehicles. IJVNV 2013;9(1-2):75-95. [doi: 10.1504/ijvnv.2013.053818]
13. Caberletti L, Elfmann K, Kummel M, Schierz C. Influence of ambient lighting in a vehicle interior on the driver's perceptions. Light Res Technol 2010;42(3):297-311. [doi: 10.1177/1477153510370554]
14. Kim T, Kim Y, Jeon H, Choi CS, Suk HJC. Emotional response to in-car dynamic lighting. Int J Automot Technol 2021;22(4):1035-1043. [doi: 10.1007/s12239-021-0093-4]
15. Löcken A, Heuten W, Boll S. Enlightening drivers: a survey on in-vehicle light displays. 2016 Presented at: Automotive'UI 16: Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications; October 24-26, 2016; Ann Arbor, MI, USA p. 97-104 URL: https://dl.acm.org/doi/proceedings/10.1145/3003715 [doi: 10.1145/3003715.3005416]
16. McCartt AT, Hellinga LA, Bratiman KA. Cell phones and driving: review of research. Traffic Inj Prev 2006;7(2):89-106. [doi: 10.1080/15389580600651103] [Medline: 16854702]
17. Kerr D, Olateju T. Driving with diabetes in the future: in-vehicle medical monitoring. J Diabetes Sci Technol 2010;4(2):464-469 [FREE Full text] [doi: 10.1177/193229681000400228] [Medline: 20307408]
18. Hoskins M. Diabetes data behind the wheel... #WeAreNotWaiting. Healthline Media. 2018. URL: https://web.archive.org/web/20201112015408/https://www.healthline.com/diabetesmine/diabetes-data-in-cars-wearenotwaiting#1 [accessed 2024-05-04]
19. Advanced driver assistance systems. European Commission. 2016. URL: https://web.archive.org/web/20231225024440/https://road-safety.transport.ec.europa.eu/system/files/2021-07/ersosynthesis2016-adas15_en.pdf [accessed 2024-04-05]
20. Baldwin CL, Moore C. Perceived urgency, alerting effectiveness and annoyance of verbal collision avoidance system messages. Proc Hum Factors Ergon Soc Annu Meet 2016;46(22):1848-1852. [doi: 10.1177/154193120204602215]
21. Statement of principles on the design of high-priority warning signals for in-vehicle intelligent transport systems—draft. International Harmonized Research Activities (IHRA) Working Group on Intelligent Transport Systems (ITS). 2009. URL: https://unece.org/DAM/trans/doc/2009/wp29/ITS-17-03e.pdf [accessed 2024-05-04]
22. Kinsella B, Mutchler A. In-car voice assistant consumer adoption report 2019. Voicebot.ai. 2019. URL: https://web.archive.org/web/20231002110008/https://voicebot.ai/in-car-voice-assistant-consumer-adoption-report-2019/ [accessed 2024-04-05]
23. Kinsella B, Mutchler A. In-car voice assistant consumer adoption report 2020. Voicebot.ai. 2020. URL: https://web.archive.org/web/20231002082700/https://voicebot.ai/in-car-voice-assistant-consumer-adoption-report-2020/ [accessed 2024-04-05]
24. Kinsella B. In-car voice assistant users 127 million in the U.S. with strong adoption in the UK and Germany as well—new report. Voicebot.ai. 2021. URL: https://web.archive.org/web/20230221102553/https://voicebot.ai/2021/08/23/in-car-voice-assistant-users-127-million-in-the-u-s-with-strong-adoption-in-the-uk-and-germany-as-well-new-report/ [accessed 2021-08-23]
25. Strayer DL, Cooper JM, Goethe RM, McCarty MM, Getty DJ, Biondi F. Assessing the visual and cognitive demands of in-vehicle information systems. Cogn Res Princ Implic 2019;4(1):18 [FREE Full text] [doi: 10.1186/s41235-019-0166-3] [Medline: 31227955]
26. Nass C, Steuer J, Tauber ER. Computers are social actors. 1994 Presented at: CHI '94: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; April 24-28, 1994; Boston, MA, USA p. 72-78 URL: https://dl.acm.org/doi/proceedings/10.1145/191666 [doi: 10.1145/191666.191703]
27. Cho E, Molina MD, Wang J. The effects of modality, device, and task differences on perceived human likeness of voice-activated virtual assistants. Cyberpsychol Behav Soc Netw 2019;22(8):515-520. [doi: 10.1089/cyber.2018.0571] [Medline: 31314579]
28. Mahajan K, Large DR, Burnett G, Velaga NR. Exploring the effectiveness of a digital voice assistant to maintain driver alertness in partially automated vehicles. Traffic Inj Prev 2021;22(5):378-383. [doi: 10.1080/15389588.2021.1904138] [Medline: 33881365]
29. Voice-powered assistants make way for integrated, intuitive co-pilots. Cerence. 2022. URL: https://tinyurl.com/45v8cfau [accessed 2022-01-12]
30. Hofmann H, Hermanutz M, Tobisch V, Ehrlich U, Berton A, Minker W. Evaluation of in-car SDS notification concepts for incoming proactive events. In: Situated Dialog in Speech-Based Human-Computer Interaction. Cham: Springer; 2016:111-124.
31. Schmidt M, Minker W, Werner S. How users react to proactive voice assistant behavior while driving. 2020 Presented at: Proceedings of the Twelfth Language Resources and Evaluation Conference; May 2020; Marseille, France p. 485-490 URL: https://api.semanticscholar.org/CorpusID:218973946
32. Ringfort-Felner R, Laschke M, Sadeghian S, Hassenzahl M. Kiro: a design fiction to explore social conversation with voice assistants. Proc ACM Hum-Comput Interact 2022;6(GROUP):1-21. [doi: 10.1145/3492852]

33. Woodhead MM. Effect of brief loud noise on decision making. J Acoust Soc Am 1959;31(10):1329-1331. [doi: 10.1121/1.1907629]
34. Heise T, Koschinsky T, Heinemann L, Lodwig V, Glucose Monitoring Study Group. Hypoglycemia warning signal and glucose sensors: requirements and concepts. Diabetes Technol Ther 2003;5(4):563-571. [doi: 10.1089/152091503322250587] [Medline: 14511411]
35. Löcken A, Heuten W, Boll SC. AutoAmbiCar: using ambient light to inform drivers about intentions of their automated cars. 2016 Presented at: AutomotiveUI '16 Adjunct: Adjunct Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications; October 24-26, 2016; Ann Arbor, MI, USA p. 57-62 URL: https://dl.acm.org/doi/proceedings/10.1145/3004323 [doi: 10.1145/3004323.3004329]
36. Parasuraman A, Colby CL. An updated and streamlined Technology Readiness Index: TRI 2.0. J Serv Res 2014;18(1):59-74. [doi: 10.1177/1094670514539730]
37. Lehmann V, Zueger T, Maritsch M, Kraus M, Albrecht C, Bérubé C, et al. Machine learning for non-invasive sensing of hypoglycaemia while driving in people with diabetes. Diabetes Obes Metab 2023;25(6):1668-1676 [FREE Full text] [doi: 10.1111/dom.15021] [Medline: 36789962]
38. Bérubé C. GitHub—hypogycemia-warning. catbot0. 2021. URL: https://github.com/catbot0/hypogycemia-warning [accessed 2024-03-15]
39. Betella A, Verschure PFMJ. The Affective Slider: a digital self-assessment scale for the measurement of human emotions. PLoS One 2016;11(2):e0148037 [FREE Full text] [doi: 10.1371/journal.pone.0148037] [Medline: 26849361]
40. Van Der Laan JD, Heino A, De Waard D. A simple procedure for the assessment of acceptance of advanced transport telematics. Transp Res Part C Emerg Technol 1997;5(1):1-10. [doi: 10.1016/s0968-090x(96)00025-3]
41. Dahlbäck N, Jönsson A, Ahrenberg L. Wizard of Oz studies—why and how. Knowl-Based Syst 1993;6(4):258-266. [doi: 10.1016/0950-7051(93)90017-n]
42. Boucsein W. Electrodermal Activity. 2nd Edition. New York: Springer; 2012.
43. Braithwaite JJ, Watson DG, Jones R, Rowe ML. A guide for analysing electrodermal activity (EDA) and skin conductance responses (SCRs) for psychological experiments. Psychophysiology. 2013. URL: https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/psychology/ABriefGuideforAnalysingElectrodermalActivityv4.pdf) [accessed 2024-03-15]
44. Brown P, Rothwell JC, Thompson PD, Britton TC, Day BL, Marsden CD. New observations on the normal auditory startle reflex in man. Brain 1991;114((Pt 4)):1891-1902. [doi: 10.1093/brain/114.4.1891] [Medline: 1884184]
45. Borrego A, Latorre J, Alcañiz M, Llorens R. Reliability of the Empatica E4 wristband to measure electrodermal activity to emotional stimuli. 2019 Presented at: 2019 International Conference on Virtual Rehabilitation (ICVR); July 21-24, 2019; Tel Aviv, Israel p. 1-2 URL: https://ieeexplore.ieee.org/document/8994546/ [doi: 10.1109/icvr46560.2019.8994546]
46. Milstein N, Gordon I. Validating measures of electrodermal activity and heart rate variability derived from the Empatica E4 utilized in research settings that involve interactive dyadic states. Front Behav Neurosci 2020;14:148 [FREE Full text] [doi: 10.3389/fnbeh.2020.00148] [Medline: 33013337]
47. Ragot M, Martin N, Em S, Pallamin N, Diverrez JM. Emotion recognition using physiological signals: laboratory vs. wearable sensors. 2017 Presented at: Advances in Human Factors in Wearable Technologies and Game Design: Proceedings of the AHFE 2017 International Conference on Advances in Human Factors and Wearable Technologies; July 17-21, 2017; The Westin Bonaventure Hotel, Los Angeles, CA, USA p. 15-220 URL: https://doi.org/10.1007/978-3-319-60639-2_2 [doi: 10.1007/978-3-319-60639-2_2]
48. Bent B, Cho PJ, Henriquez M, Wittmann A, Thacker C, Feinglos M, et al. Engineering digital biomarkers of interstitial glucose from noninvasive smartwatches. NPJ Digit Med 2021;4(1):89 [FREE Full text] [doi: 10.1038/s41746-021-00465-w] [Medline: 34079049]
49. Bogue-Jimenez B, Huang X, Powell D, Doblas A. Selection of noninvasive features in wrist-based wearable sensors to predict blood glucose concentrations using machine learning algorithms. Sensors (Basel) 2022;22(9):3534 [FREE Full text] [doi: 10.3390/s22093534] [Medline: 35591223]
50. Kappeler-Setz C, Gravenhorst F, Schumm J, Arnrich B, Tröster G. Towards long term monitoring of electrodermal activity in daily life. Pers Ubiquitous Comput 2013;17(2):261-271 [FREE Full text] [doi: 10.1007/s00779-011-0463-4]
51. Crawford JR, Henry JD. The Positive and Negative Affect Schedule (PANAS): construct validity, measurement properties and normative data in a large non-clinical sample. Br J Clin Psychol 2004;43(Pt 3):245-265. [doi: 10.1348/0144665031752934] [Medline: 15333231]
52. Harmon-Jones C, Bastian B, Harmon-Jones E. The Discrete Emotions Questionnaire: a new tool for measuring state self-reported emotions. PLoS One 2016;11(8):e0159915 [FREE Full text] [doi: 10.1371/journal.pone.0159915] [Medline: 27500829]
53. Blankenbach K, Hertlein F, Hoffmann S. Advances in automotive interior lighting concerning new LED approach and optical performance. J Soc Info Display 2020;28(8):655-667 [FREE Full text] [doi: 10.1002/jsid.887]
54. Kim T, Lee G, Park M, Lee HM, Park JW, Suk HJ. User responses to dynamic light in automobiles with EEG and self-assessments. IEEE Access 2022;10:123847-123857 [FREE Full text] [doi: 10.1109/access.2022.3223665]

55. Koch K, Mishra V, Liu S, Berger T, Fleisch E, Kotz D, et al. When do drivers interact with in-vehicle well-being interventions? An exploratory analysis of a longitudinal study on public roads. Proc ACM Interact Mob Wearable Ubiquitous Technol 2021;5(1):19 [FREE Full text] [doi: 10.1145/3448116] [Medline: 35178497]
56. Koch K, Tiefenbeck V, Liu S, Berger T, Fleisch E, Wortmann F. Taking mental health and well-being to the streets: an exploratory evaluation of in-vehicle interventions in the wild. 2021 Presented at: CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems; May 8-13, 2021; Yokohama, Japan p. 1-15 URL: https://dl.acm.org/doi/10.1145/3411764.3446865 [doi: 10.1145/3411764.3446865]
57. Papathanail I, Vasiloglou MF, Stathopoulou T, Ghosh A, Baumann M, Faeh D, et al. A feasibility study to assess Mediterranean diet adherence using an AI-powered system. Sci Rep 2022;12(1):17008 [FREE Full text] [doi: 10.1038/s41598-022-21421-y] [Medline: 36220998]
58. Payne O, Smith V, Rolnik DL, Davies-Tuck M, Warty R, Seo D, et al. Virtual reality and its use in post-operative pain following laparoscopy: a feasibility study. Sci Rep 2022;12(1):13137 [FREE Full text] [doi: 10.1038/s41598-022-17183-2] [Medline: 35907935]
59. Heijmans M, Habets JGV, Herff C, Aarts J, Stevens A, Kuijf ML, et al. Monitoring Parkinson's disease symptoms during daily life: a feasibility study. NPJ Parkinsons Dis 2019;5:21 [FREE Full text] [doi: 10.1038/s41531-019-0093-5] [Medline: 31583270]

Abbreviations

SCR: skin conductance response
T1DM: type 1 diabetes mellitus
VA: voice assistant

Edited by A Kushniruk; submitted 08.03.23; peer-reviewed by K Baker, Y Jeem; comments to author 03.05.23; revised version received 23.05.23; accepted 02.03.24; published 18.04.24.

Please cite as:
Bérubé C, Maritsch M, Lehmann VF, Kraus M, Feuerriegel S, Züger T, Wortmann F, Stettler C, Fleisch E, Kocaballi AB, Kowatsch T
Multimodal In-Vehicle Hypoglycemia Warning for Drivers With Type 1 Diabetes: Design and Evaluation in Simulated and Real-World Driving
JMIR Hum Factors 2024;11:e46967
URL: https://humanfactors.jmir.org/2024/1/e46967
doi:10.2196/46967
PMID:38635313

Original Paper

Usability Comparison Among Healthy Participants of an Anthropomorphic Digital Human and a Text-Based Chatbot as a Responder to Questions on Mental Health: Randomized Controlled Trial

Almira Osmanovic Thunström[1,2], MSc; Hanne Krage Carlsen[2,3], MPH, PhD; Lilas Ali[1,4,5], RN, PhD; Tomas Larson[1,2], MSc, PhD; Andreas Hellström[6], PhD; Steinn Steingrimsson[1,2], BSc, MD, PhD

[1]Region Västra Götaland, Psychiatric Department, Sahlgrenska University Hospital, Gothenburg, Sweden

[2]Section of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

[3]Region Västra Götaland, Centre of Registers, Gothenburg, Sweden

[4]Institute of Health Care Sciences, Centre for Person-Centred Care, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

[5]Centre for Person-Centred Care, University of Gothenburg, Gothenburg, Sweden

[6]Department of Technology Management and Economics, Chalmers University of Technology, Gothenburg, Sweden

Corresponding Author:
Steinn Steingrimsson, BSc, MD, PhD
Region Västra Götaland
Psychiatric Department
Sahlgrenska University Hospital
Journalvägen 5
Gothenburg, 41650
Sweden
Phone: 46 313421000
Email: steinn.steingrimsson@gu.se

Abstract

Background: The use of chatbots in mental health support has increased exponentially in recent years, with studies showing that they may be effective in treating mental health problems. More recently, the use of visual avatars called digital humans has been introduced. Digital humans have the capability to use facial expressions as another dimension in human-computer interactions. It is important to study the difference in emotional response and usability preferences between text-based chatbots and digital humans for interacting with mental health services.

Objective: This study aims to explore to what extent a digital human interface and a text-only chatbot interface differed in usability when tested by healthy participants, using BETSY (Behavior, Emotion, Therapy System, and You) which uses 2 distinct interfaces: a digital human with anthropomorphic features and a text-only user interface. We also set out to explore how chatbot-generated conversations on mental health (specific to each interface) affected self-reported feelings and biometrics.

Methods: We explored to what extent a digital human with anthropomorphic features differed from a traditional text-only chatbot regarding perception of usability through the System Usability Scale, emotional reactions through electroencephalography, and feelings of closeness. Healthy participants (n=45) were randomized to 2 groups that used a digital human with anthropomorphic features (n=25) or a text-only chatbot with no such features (n=20). The groups were compared by linear regression analysis and *t* tests.

Results: No differences were observed between the text-only and digital human groups regarding demographic features. The mean System Usability Scale score was 75.34 (SD 10.01; range 57-90) for the text-only chatbot versus 64.80 (SD 14.14; range 40-90) for the digital human interface. Both groups scored their respective chatbot interfaces as average or above average in usability. Women were more likely to report feeling annoyed by BETSY.

Conclusions: The text-only chatbot was perceived as significantly more user-friendly than the digital human, although there were no significant differences in electroencephalography measurements. Male participants exhibited lower levels of annoyance with both interfaces, contrary to previously reported findings.

(JMIR Hum Factors 2024;11:e54581) doi:10.2196/54581

KEYWORDS

chatbot; chatbots; chat-bot; chat-bots; text-only chatbot, voice-only chatbot; mental health; mental illness; mental disease; mental diseases; mental illnesses; mental health service; mental health services; interface; system usability; usability; digital health; machine learning; ML; artificial intelligence; AI; algorithm; algorithms; NLP; natural language processing

Introduction

Conversational user interfaces, also known as chatbots, have been a part of human-computer interactions since the 1960s. Most notable and one of the earliest examples is the ELIZA system, which aimed at simulating a human psychologist [1,2]. A subsequent system named PARRY was implemented in 1972, in which the conversational agent was designed to emulate a patient experiencing schizophrenia [3]. It is by no means a coincidence that the 2 earliest systems targeting the replication of human behavior through natural language processing were both derived from the field of psychiatry. The use of conversational agents has increased exponentially in the past decade [4]. With the availability of systems and the increasing need for 24-hour availability due to globalization, Radziwill and Benton [5] found that perhaps as many as 1 of 3 web-based conversations were conducted with a chatbot or a system moderated by language models, of which some have garnered more than 100 million users [4-8].

Previous research on rule-based conversational agents has shown promise with respect to the alleviation of mental health problems [4,9-11]. In a study by Oh et al [12], patients with panic disorder were randomized to support via a chatbot or support via a self-help book. The patients who were assigned to a chatbot as a support system for exercises in cognitive behavioral therapy were more likely to show symptom alleviation [12]. Digital evaluations as well as digital deliverance of mental health aid were more intensively explored following the COVID-19 pandemic [13]. In a study by Islam et al [14], the authors explored a similar design to that of Oh et al [12] and randomized a set of participants to either book or chatbot intervention for support regarding mental health issues [14]. The group of participants allocated to the chatbot intervention also significantly improved control of helplessness and social phobia scores. Some studies have shown that even a single exposure to a chatbot therapist can have a positive influence on the current state of well-being and repeated exposure can be a good complementary treatment for anxiety [9,10,15,16].

In recent years, a novel facet has been introduced into the evaluative framework for mental health chatbots: the incorporation of voice-controlled visual avatars embodying humanoid countenances colloquially referred to as "digital humans." These digital entities harness the power of machine learning, emotion-infused linguistics, and adept emulation of facial expressions to cultivate a profound emotional rapport with their users. Research shows that human features elicit more social engagement and can trigger a stronger emotional bond [17]. This has primarily been measured through electroencephalography (EEG) with a specific focus being placed on the importance of increased α and θ wave activity as indicators of overall emotional stability and positive response to stimuli [18-20], while β wave activity has largely been associated with less desirable states of mind such as anxiety and an active stress response [19,21].

While acknowledging the inherent complexity of brain states and wave activity, delving into the extent to which distinct brain wave frequencies exert influence during a chat session presents an intriguing avenue for investigating the emotional states of the user. In a study by Bos et al [22], the authors explored capturing vigilance and states of emotion with EEG in usability testing of chatbot technology. The study findings revealed that EEG effectively captured the facets of user experience and conversation that piqued interest. This was accomplished through the delineation of γ wave activity, predominantly linked with positivity and problem-solving. Consequently, this approach affords researchers a more objective means of apprehending user experience. A study by Ciechanowski et al [23] indicated that there is a difference in emotional response and usability preference between text-based chatbots and digital humans, with text-based chatbots eliciting more positive interactions.

Although EEG has served as a proficient tool for quantifying objective assessments of emotional responses to chatbot interventions, it is customary to use usability scales for capturing the subjective dimensions pertaining to emotions and experiences in the context of chatbot systems. While a universally accepted benchmark for conducting usability tests on chatbots remains elusive, numerous studies have gravitated toward the adoption of the System Usability Scale (SUS-10) [24-29] and the Speech User Interface Service Quality scale. SUS-10 captures the overall usability of a system independently of the platform or interface. The score ranges from 0 to 100, indicating higher usability with increasing scores [26]. A score of 68 is considered as a passing grade, while a score below 50 is considered as indicating that the system has less optimal usability. For a system to be considered as exceptionally good in terms of its design and usability, a score of 85 on average should be applied [29-31]. In the previously mentioned study by Oh et al [12], mean SUS-10 was not significantly worse or better comparing a chatbot and a book: 64.5 (SD 17.0) versus 69.5 (SD 17.2), respectively (*P*=.35). Several studies have advocated the idea that chatbots represent user-friendly alternatives to conventional analog methods or standard digital tools, such as forms [4]. Nonetheless, there is research that suggests the design flaws in a chatbot system can markedly diminish its effectiveness, potentially leading to perceptions of unhelpfulness among users [3]. Chatbots that are perceived as unhelpful, repetitive, or lacking the users' trust tend to receive a lower SUS-10 score [32].

Many social chatbots aim to comfort, support, and advise their users [3]. Studies show that the availability of chatbot technology is what is central to its perception of usefulness compared with human therapists. However, studies have also noted that most users prefer human therapists and are more interested in using the system as a complementary tool when a human therapist is not available [33-35]. While mental health chatbots are generally viewed positively by the user, there are many issues that can lead to decreased usability, lower SUS-10 scores, and undesirable outcomes such as irritation or worsened mental health. The propensity for misunderstanding, miscommunication, and annoyance are frequently reported in qualitative assessments of social support chatbots [33-35]. Feeling annoyed by repetitive messaging, incoherent conversations, and inability to comprehend the user's needs are frequently named as issues that increase the feeling of annoyance in users of social support chatbots [34]. The selection of an interface can wield a considerable influence on both the effectiveness and user-friendliness of a system. Users exhibit disparate reactions to chatbots depending on whether they incorporate an avatar, particularly one with humanoid attributes capable of evoking emotions. Although our understanding of chatbot usability and user preference is somewhat limited, investigations into anthropomorphic interfaces do underscore their ability to affect our emotional states [36].

The chatbot used in our study, known as BETSY (Behavior, Emotion, Therapy System, and You), uses 2 distinct interfaces: a digital human, voice-activated user interface with anthropomorphic features and a text-only user interface. Within the scope of our investigation, we aimed to thoroughly examine both interface modalities. Phase 1 of usability testing involves enlisting the participation of healthy volunteers.

The aim of this study was to explore to what extent a digital human and a text-only chatbot interface differed in usability when tested by healthy participants. We also set out to explore how chatbot-generated conversations on mental health (specific to each interface) affected self-reported feelings and biometrics.

Methods

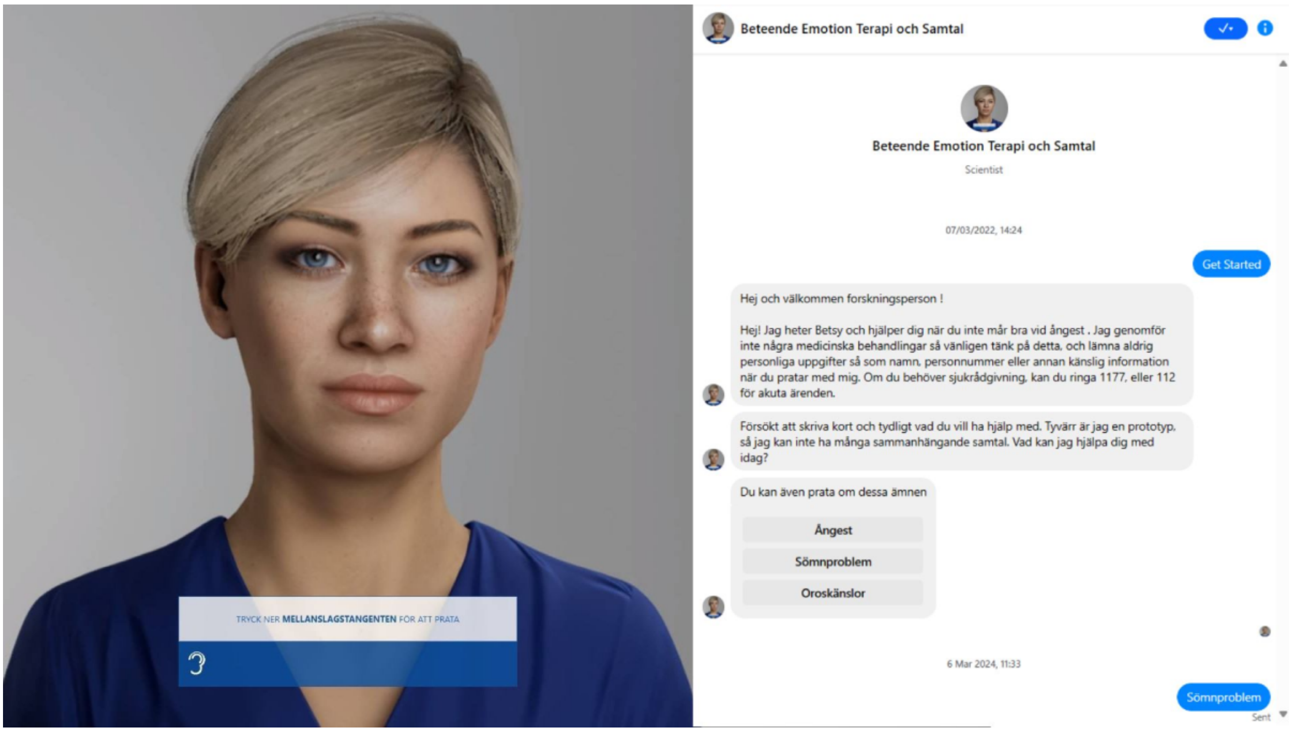
Construction of the User Interface (BETSY)

This project adopted a participatory design approach to ensure the broad involvement of health care professionals, patients, and the public. A multidisciplinary team consisting of 2 psychiatrists, 2 psychiatric nurses, 4 clinical psychologists, 1 user of health care services, and 1 engineer was assembled to comprehensively address ethical, medical, and legal considerations for a potential chatbot. Team members were selected for their expertise in digitalization and psychiatry. Before the initial workshop, where the algorithm's preliminary outline was presented, the engineer created a survey. This survey drew partly from Radziwill and Benton [5] quality attribute listing, which synthesized findings from various chatbot usability projects.

A survey was distributed to the public via the secure research platform Psytoolkit.org, offering heightened anonymity by omitting the collection of metadata such as IP addresses and locations. The survey comprised 8 multiple-choice questions and 4 open-ended free-text questions, covering demographic information, design requirements, functionality suggestions, and overall attitudes toward mental health chatbots. It was accessible for 14 days and disseminated through various social media channels. Subsequently, the collected data were analyzed to inform a series of 4 workshops conducted by the group between June 2020 and December 2020. During these workshops, the chatbot's design, encompassing appearance, content, and personality, underwent iterative development based on input from the general public and co-designer feedback, with the latter representing a patient perspective. A comprehensive account of this process will be available in a separate publication.

Two versions of the chatbot (Figure 1) were created: one enabling voice interaction with a facial expression and an avatar component, and another relying solely on text-based communication with an avatar image. The digital human was implemented using Dialogflow (Google) for conversation logic and connected to the UNEEQ platform for the human-avatar interface. Data infrastructure was hosted by Deloitte Digital and VästraGötalandsregionen/VGR-IT. In contrast, the text-only BETSY chatbot was developed on the Itsalive.io platform and deployed to a research and development account on Facebook that was closed to the public. Importantly, no personal metadata were collected during on-site testing via digital platforms. The users did not use their personal social media accounts to talk to the chatbot.

Figure 1. Two versions of BETSY (Behavior, Emotion, Therapy System, and You): digital human, voice-activated (left) and text-only (right).



Both versions of BETSY encompassed 24 topics (detailed in Multimedia Appendix 1) related to mental health, including anxiety, depression, stress, sleep, addiction, eating disorders, anger, hopelessness, helplessness, loneliness, sadness, suicidal ideation, and suicidality, among others. These chatbots were designed in the Swedish language. An assessment was conducted to evaluate the alignment of the text-only and digital human algorithms. Specifically, testers posed identical questions to both systems within various domains, with only 1 instance revealing a discrepancy when the digital human could not provide an appropriate response while the text-based bot could, indicating the need for further refinement.

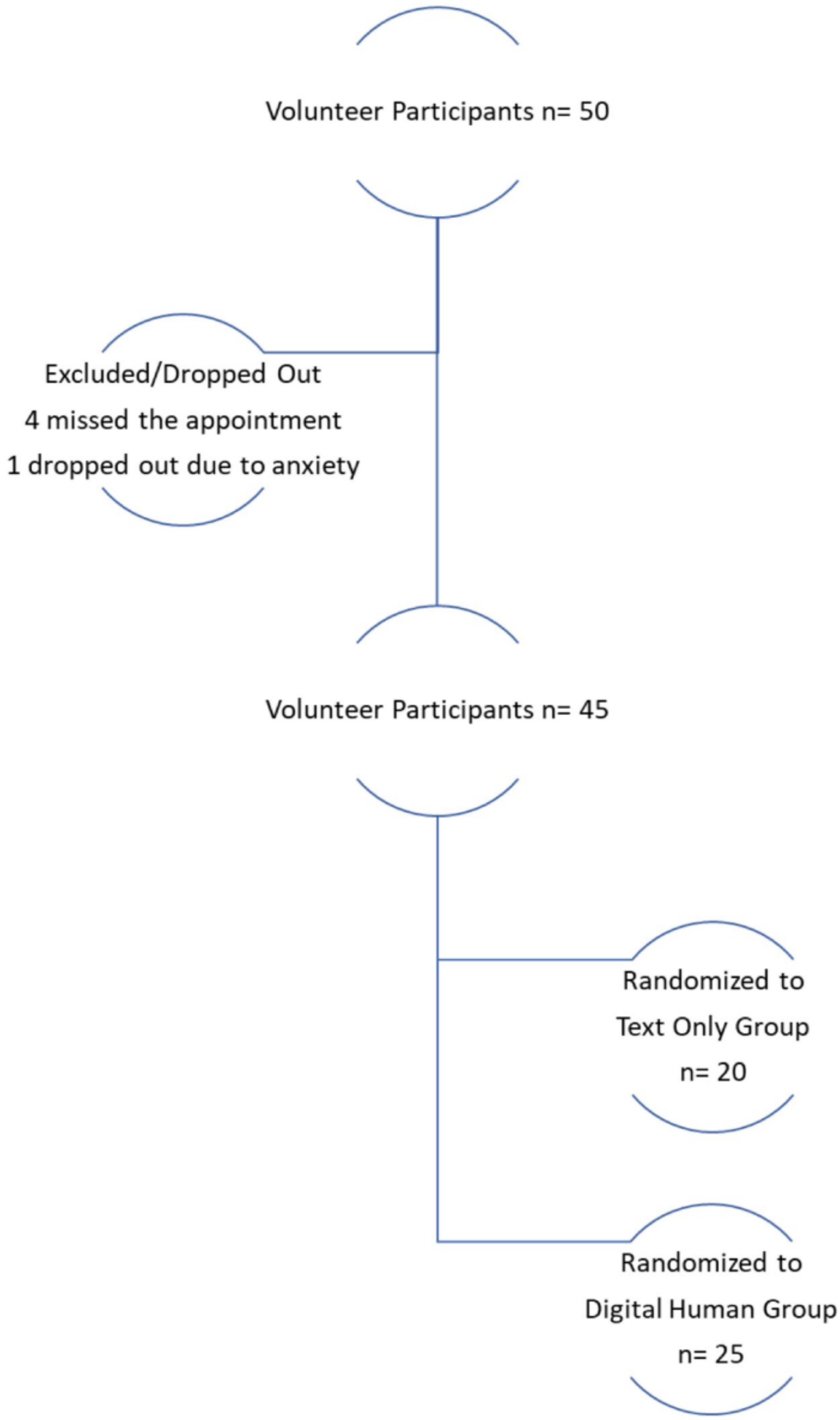
Recruitment

In this initial phase of system exploration, our focus was on evaluating the system's capabilities using volunteers who did not exhibit severe anxiety. As the system is still in its prototype stage, we exercised caution to avoid any potential exacerbation of symptoms in individuals with severe anxiety. Our recruitment announcement, disseminated through various social media channels associated with Sahlgrenska University's official account, specified that participants should be 18 years or older, free from any current mental health disorders, and willing to physically attend the testing facility in Gothenburg, Sweden.

Participants

Of the 50 individuals who initially volunteered, 5 participants (2 men and 3 women) opted out before providing their consent (Figure 2). Subsequently, 45 individuals attended the screening at the test facility. Each participant was required to provide informed consent before undergoing the Generalized Anxiety Disorder (GAD-7) scale assessment for anxiety symptoms. Those scoring 14 or higher on the GAD-7 were excluded from the study (Figure 2). Eligible participants were then randomly assigned to one of two groups: (1) engaging in text-based conversations with the text-only BETSY or (2) participating in voice-based interactions with the digital human (Figure 2).

Figure 2. Flow chart of participation and exclusion.



The randomization process was conducted with strict double-blind procedures overseen by an independent researcher not affiliated with this project and facilitated by an automated randomization system, ensuring the impartiality of the allocation.

Prechat Procedure

The experiments were performed during the COVID-19 pandemic (June 2021-November 2021). Due to safety precautions, the participants were greeted by a tester wearing protective gear, including a surgical R-II mask, gloves, a face visor, and hospital scrubs. Protective gear was also offered to participants upon their arrival. Participants were then placed in a sanitized room equipped with a screen, which underwent thorough sanitization with medical-grade disinfectants and a sterilizing UV lamp before and after each participant's session.

Before starting the chat with BETSY, participants were outfitted with a mobile dry-sensor EEG device to record their brain wave activity. Additionally, their blood pressure and pulse were recorded on the left arm after a 5-minute seated rest. Systolic and diastolic blood pressure were measured using a digital sphygmomanometer, and the pulse was monitored with a pulse oximeter.

Despite relatively relaxed COVID-19 restrictions during the testing period, the tester opted not to be physically present in the room to minimize any potential risk of contagion. Each participant sat alone in the room, with the tester observing remotely via a nonrecordable streaming camera. This camera served to facilitate real-time communication and allowed the tester to monitor the participant's reactions and anticipate any need for assistance. The participants were made aware of this procedure.

Following the measurement of biometric data (blood pressure and pulse), participants were instructed to complete a questionnaire. This questionnaire covered their prior experiences with mental health chatbots as well as their demographic information, including sex, occupation, and marital status. Additionally, participants rated their overall well-being on a visual analog scale ranging from 1 (not good at all) to 10 (feeling excellent) before starting their session with BETSY. Participants were given instructions by the tester along with an accompanying sheet that provided potential chat scenarios and specified the topics within BETSY's scope. Each participant had a maximum of 30 minutes to engage with the chatbot version they were assigned to.

Chat Session Protocol and EEG Data Collection Process

EEG recording commenced simultaneously with the participant's initiation of their session with BETSY. We used a dry-sensor mobile EEG system, specifically the MUSE headband from Interaxon, which incorporates 7 sensors. These sensors include 3 frontal reference sensors and active sensors situated at Fp1, Fp2, Tp9, and Tp10.

The MUSE headband was seamlessly connected to a smartphone via Bluetooth and data collection was facilitated through the Mind Monitor app on an Android smartphone. It is noteworthy that this app neither necessitates user registration nor collects data that can identify or pinpoint individual users or their locations. Consequently, the data were recorded in an anonymous fashion and stored as a CSV file within the smartphone's document section.

Upon placing the EEG headband on the participant, they were requested to close their eyes to enable the Mind Monitor to perform calibration. After calibration was successfully established, the participant was left alone in the room to initiate a conversation with BETSY. Importantly, the EEG recording remained active throughout the entire conversation and was terminated when the participant indicated they had concluded their interaction with the chatbot.

Postchat Procedure

Upon reaching a point of satisfaction with the conversation or upon the completion of their allocated chat time, participants were directed to complete supplementary questionnaires and scales. Participants were administered the SUS-10, developed by Lewis and Sauro [26]. The SUS-10 calculates an average score derived from a 10-item questionnaire with response options ranging from 0 to 4, resulting in a total score between 0 and 100, as outlined by Bangor et al [31].

Participants were also presented with multiple-choice questions regarding their emotional state during the chat. Additionally, in the digital human group, participants were instructed to fill out the Standardized Questionnaires for Voice Interaction Design Short Version (SUISQ-MR). SUSIQ is a questionnaire tailored to assessing critical usability attributes of Interactive Voice Response, as outlined by Lewis and Sauro [26]. The original scale comprises 25 items categorized into 4 factors: user goal orientation, customer service behaviors, speech characteristics, and verbosity. SUISQ-MR, a shortened version, encompasses 9 items rated on a 7-point Likert scale, ranging from "strongly disagree" to "strongly agree." Higher scores on this scale indicate a more favorable assessment of the system's usability [24].

Furthermore, participants were provided with an open-ended questionnaire to gather their suggestions and insights regarding their session experience. It should be noted that qualitative data from this survey will be reported separately.

EEG Monitoring and Analysis

The monitoring of EEG activity entails the use of the Mind Monitor app, which captures and visually represents EEG brain wave data. The quantification of absolute brain wave values is predicated on the computation of absolute band powers. These powers are derived from the logarithm of the power spectral density calculated from the EEG data for each channel.

The frequency spectrum categories used for this analysis encompass the following bandwidths: δ (1-4 Hz), θ (4-8 Hz), α (7.5-13 Hz), β (13-30 Hz), and γ (30-44 Hz). Notably, the EEG power spectral density values acquired from the sensors typically fall within the {–1:+1} range, which is subsequently transformed into a more intelligible {0:100} range for text-based display purposes.

Subsequently, the collected EEG data underwent an analytical process facilitated by the Mind Monitor online graphing tool. Within this tool, the values are presented as average (dB) per session. It is imperative to note that, in the context of this study, we exclusively used absolute data for our analyses (information sourced from Mindmonitor.com and Choosemuse.com).

Statistical Analysis

All data were entered and processed in SPSS Statistics (version 28.0.1.1; IBM Corp). For group differences, means analysis was used using Pearson χ^2 asymptotic significance (2-sided) set at .05 as the significance level. For continuous outcome variables such as SUS-10, SUISQ-MR, brain wave activity, positivity, and GAD-7, linear regression analyses were used. The data were tested for kurtosis and skewness. Based on the results, *t* tests were performed. All results were analyzed according to group.

Ethical Considerations

All ethical decisions were guided by the Declaration of Helsinki and its subsequent amendments. The study protocol was reviewed and approved by the central ethical review board; Etikprövningsmyndigheten, Sweden (DRN 2021-02771). All precautions were taken in order to avoid any possible contagion from COVID-19. Due to the nature of the prototype, no patients were used in this initial examination of the chatbot in order to ensure that vulnerable participants would not be negatively affected by errors and flaws that might be present in a prototype-stage system.

Results

Characteristics of Participants

There were no statistically significant differences in the demographic variables between the digital human and text-only groups (Table 1). No participants were excluded due to a high GAD-7 score. The age of the participants ranged from 24 to 68 years, and as only 12 participants registered their age, this variable was consequently excluded from more advanced analyses (Table 1).

Table 1. Demographic characteristics of the study population.

Characteristic	Text-only, n (%)	Digital human, n (%)	*P* value[a]
Sex			.62
Male	8 (40)	10 (40)	
Female	12 (60)	15 (60)	
Marital status			.49
Married	11 (55)	11 (44)	
Single	5 (25)	3 (12)	
Divorced	1 (5)	2 (8)	
Domestic partnership	2 (10)	7 (28)	
Other	1 (5)	2 (8)	
Educational level			.74
High school/trade school	1 (5)	3 (12)	
Bachelor’s	7 (35)	8 (32)	
Master’s	7 (35)	11 (44)	
PhD	3 (15)	1 (4)	
Other/higher than master’s or PhD	2 (10)	3 (8)	
Occupation			.30
Sick leave/sick leave part-time	0 (0)	1 (4)	
Working part-time	0 (0)	2 (8)	
Working full time	18 (90)	19 (76)	
Student	2 (10)	1 (4)	
Retired	0 (0)	2 (8)	
Housing			.35
Living alone	4 (20)	2 (8)	
Cohabitation	16 (80)	23 (92)	

[a]Pearson χ^2 test.

Comparison Between Digital Human and Text-Only Chatbots

When comparing self-reported emotional states between the digital human and the text-only chatbot groups, it was observed that participants using the digital human exhibited a notably higher propensity to report feelings of nervousness versus the text-only chatbot group (Table 2). The mean GAD-7 score for the text-only chatbot group was 2.32 (SD 2.52) compared with 2.80 (SD 2.60) for the digital human chatbot group, with no statistically significant difference between the groups.

Table 2. Self-reported prior therapy experience, emotions, biometrics, and electroencephalography.

	Text-only	Digital human	*P* value[a]
Therapy, n (%)			.47
Yes	2 (10)	4 (16)	
No	18 (90)	20 (80)	
Do not remember	0	1 (4)	
"Have you talked to a chatbot about mental health before?", n (%)			.84
Yes	1 (5)	1 (4)	
No	19 (95)	24 (96)	
Do not remember	0	0	
GAD-7[b] score, mean (SD)[a]	2.3 (2.5)	2.8 (2.6)	.56
Positivity toward chatbot, mean (SD)[c]	7.1 (2.1)	7.5 (2.1)	.69
"Do you feel closeness to BETSY[d]?", n (%)			.46
Yes	7 (35)	11 (45.8)	
No	13 (65)	13 (54.2)	
"Did you feel relaxed?", n (%)			.23
Yes or sometimes	17 (89.5)	17 (73.9)	
No	2 (10.5)	6 (26.1)	
"Did you feel nervous?", n (%)			.02
Yes or sometimes	0	6 (26.1)	
No	19 (100)	17 (73.9)	
"Did you feel sad?", n (%)			.1
Yes or sometimes	0	3 (13.0)	
No	19 (100)	20 (87.0)	
"Did you feel annoyance?", n (%)			.8
Yes or sometimes	9 (47.4)	10 (43.5)	
No	10 (52.6)	13 (56.5)	
VAS-W[e] presession, mean (SD)[f]	8.8 (1.32)	8.4 (1.41)	.33
VAS-W postsession, mean (SD)[f]	8.8 (1.23)	8.3 (1.27)	.14
Pulse presession, mean (SD)	72.2 (10.7)	71.6 (10.9)	.77
Pulse postsession, mean (SD)	68.5 (8.8)	70.2 (11.1)	.58
Average δ wave activity, mean (SD)	114 (30)	97 (25)	.06
Average θ wave activity, mean (SD)	86 (23)	74 (21)	.08
Average α wave activity, mean (SD)	97 (27)	82 (24)	.03
Average β wave activity, mean (SD)	81 (17)	76 (20)	.34
Average γ wave activity, mean (SD)	65 (15)	66 (21)	.98
SUS-10[g], mean (SD)	74.82 (10)	64.80 (14)	.01
SUISQ-MR[h], mean (range)	N/A[i]	4.92 (2.83-6.75)	N/A

[a]Pearson χ^2 test for categorical variables and ANOVA for continuous variables.

[b]GAD-7: Generalized Anxiety Disorder Scale.

[c]Participants were asked to what extent they felt positive about talking to BETSY about mental health with scores ranging from 1 (not positive at all) to 10 (very positive).

[d]BETSY: Behavior, Emotion, Therapy System, and You.

[e]VAS-W: Visual Analogue Scale for Well-Being.

[f]Range from 1 (not feeling well at all) to 10 (feeling very good).

[g]SUS-10: System Usability Scale.

[h]SUISQ-MR: Standardized Questionnaires for Voice Interaction Design Short Version.

[i]N/A: not applicable.

Conversely, the evaluation of system usability as gauged by SUS-10 showed a significant (*P*=.01) difference between the groups. Notably, the mean SUS-10 score was higher in the text-only chatbot group at 75.34 (SD 10.01; range 57-90) compared with the digital human group at 64.80 (SD 14.14; range 40-90). In addition, the digital human group underwent assessment using SUISQ-MR: BETSY had a mean score of 4.92 (SD 0.83; range 2.83-6.75), as depicted in Table 2, which is indicative of a commendable level of usability for BETSY's voice interface in accordance with the framework presented by Lewis [24].

Biometric Measures

There were no statistically significant distinctions for mean values of blood pressure or pulse between the groups either at baseline or following exposure to the interventions. Specifically, the mean pulse rate showed no discernible variations between the groups both before and after exposure, reflecting consistent values across the groups on average (data not shown).

The EEG signals collected during the study exhibited suboptimal quality, which was primarily attributed to participant movement and signal acquisition sensitivity. These challenges occasionally disrupted signal continuity during the sessions. Nonetheless, the data yielded adequate information to calculate mean values pertaining to δ, θ, α, β, and γ frequency bands, as facilitated by the web-based graphing module within the MindMonitor's platform. Only 1 significant difference was found in terms of means: the average α was significantly higher in the text-only group (Table 2).

System Usability Scale and Outcomes

Predictors of SUS-10 usability were used as a dependent variable in linear regression analysis and matched against biometric and subjective variables. Each variable was independently analyzed in a model together with SUS-10 as the dependent variable. Analysis showed that there was a significant positive relationship between average α and θ wave activity and SUS-10 in the chat-only group. A significant positive relationship was seen between SUISQ-MR scores and SUS-10 (Table 3).

Table 3. Linear regression analysis between usability and biometric variables.

	Unstandardized coefficients		*P* value[a]
	β	SE	
SUS-10[b] × positivity			
Text (n=19)	1.82	1.02	.09
Voice (n=24)	1.313	1.361	.35
SUS-10 × average δ wave activity			
Text (n=17)	0.153	0.08	.07
Voice (n=23)	0.062	0.12	.61
SUS-10 × average θ wave activity			
Text (n=17)	0.212	0.1	.05
Voice (n=23)	0.083	0.146	.57
SUS-10 × average α wave activity			
Text (n=17)	0.196	0.083	.03
Voice (n=23)	0.054	0.124	.67
SUS-10 × average β wave activity			
Text (n=17)	0.251	0.143	.10
Voice (n=23)	0.03	0.152	.85
SUS-10 × average γ wave activity			
Text (n=17)	0.148	0.177	.42
Voice (n=23)	–0.017	0.146	.91
SUS-10 × SUISQ-MR[c] (n=24)	8.100	2.976	.01

[a]Pearson χ^2 test.

[b]SUS-10: System Usability Scale.

[c]SUISQ-MR: Standardized Questionnaires for Voice Interaction Design Short Version.

Self-Reported Feelings and Gender

Furthermore, our investigation sought to discern whether significant gender disparities existed in terms of self-reported emotions. Notably, we observed a significant difference between men and women, with men exhibiting a notably lower tendency to report feeling annoyed by BETSY in contrast to women. No other statistically significant distinctions were identified (Table 4).

Table 4. Sex difference in emotional expression toward BETSY (Behavior, Emotion, Therapy System, and You).

Self-reported feeling	Men, n (%)	Women, n (%)	*P* value[a]
Did you feel annoyed?			
Chat			.03
Yes	1 (11)	8 (67)	
No	6 (88)	4 (33)	
Voice			.03
Yes	1 (12.5)	9 (60)	
No	7 (84.5)	6 (40)	
Did you feel relaxed?			
Chat			.68
Yes	6 (86)	11 (92)	
No	1 (14)	1 (8)	
Voice			.29
Yes	7 (87.5)	10 (67)	
No	1 (12.5)	5 (33)	
Did you feel closeness or connection to BETSY[b]?			
Chat			.25
Yes	4 (50)	3 (25)	
No	4 (50)	9 (75)	
Voice			.63
Yes	4 (40)	7 (50)	
No	6 (60)	7 (50)	
Did you feel nervous?			
Chat			
Yes	N/A[c]	N/A	
No	7 (100)	12 (100)	
Voice			.93
Yes	2 (25)	4 (27)	
No	6 (75)	11 (73)	
Did you feel sadness?			
Chat			
Yes	N/A	N/A	
No	7 (100)	12 (100)	
Voice			.17
Yes	0	3 (20)	
No	8 (100)	12 (80)	

[a]Pearson χ^2 test.

[b]BETSY: Behavior, Emotion, Therapy System, and You.

[c]N/A: not applicable.

An analysis of feelings of closeness and positivity toward chatbot conversations was undertaken to explore differences between men and women. In mean score analyses, the results showed that men were significantly more positive toward talking to BETSY prior to the session: 8.16 (SD 1.50) for men and 6.81 (SD 2.30) for women (*P*=.34). Conversely, there were no discernible gender-based differences concerning feelings of closeness during chatbot interactions.

Discussion

Principal Findings

This study explored how a digital human versus text-only chatbot interface affected usability and user experience in healthy participants. We also examined how chatbot-generated conversations on mental health affected self-reported feelings and biometrics. The overall sample was small and, thus, should not act as a point of reference for generalization. This study was, however, not smaller than the average study in the investigative field of mental health chatbots [9,10,12,37-41].

While the text-only system scored higher on usability, both versions of the chatbot scored average or above average with respect to overall usability [31]. The mean text-only chatbot SUS-10 score of 75.34 falls between the threshold of good (a score of 70) and excellent (a score of 80 and above) [29-31]. However, the score for the digital human (64.8) indicates that the system is perceived to be usable, but has room for improvement. Usability can be affected by many factors such as user interface design, content layout, and overall user experience [42,43].

The digital human score indicates that there may be areas for improvement in terms of all of the aforementioned aspects. It should also be noted that the SUS-10 scale does not measure a specific feature or aspect of system design, but instead provides an overall assessment of user experience [31]. Using more elaborate scales that cover more dimensions across the system is more suitable for a more in-depth analysis of the usability of chatbots. It can also be noted that the range of scores was much higher for the text-only interface (lowest score for the text-only group was 57 and the equivalent for the voice-only chatbot was 40), which indicates much poorer usability.

Taking into account the specific usability of the digital human interface, usability was considered high with an average SUISQ-MR score of 4.92. This score indicates that the voice interaction design is likely to be perceived as intuitive and useful by users. A score of 4.92 falls within the range of 4.5-5.5, which has been classified as "very good" in previous studies [24]. In addition, higher scores on the SUISQ-MR have been associated with increased user engagement and task completion rates. Therefore, a score of 4.92 can be interpreted as an indication that the voice interaction design will likely provide positive user experiences [24].

Men were more likely to score higher on positivity and less likely to report feeling annoyed by BETSY independently of the interface, which is the opposite of other studies that indicate men are more likely to be annoyed or aggressive toward female avatars [44]. In a study by Luger and Sellen [45], the authors found that higher expectations of the system lead to a higher risk of disappointment and lower scores: this could possibly explain why female participants were more agitated as their expectations might have been higher [45], however, we have no data to explore this empirically in the frame of this study. Unlike other studies with similar designs and populations [40], we did not analyze the content of the conversations. The conversations between BETSY and the participants were deleted immediately after the session as the research question was geared toward usability and not the effect on the user's own mental health status. Much like the results from Hearst and Tory [46], the interface was the focus of this investigation. Hearst and Tory showed that a well-designed conversation tamped the choice of interface. The interface played into the perception of usability only when the system failed to respond or create barriers to conversation. In our study, we used biometric data to explore feelings of relaxation or excitement/agitation while using BETSY. Despite EEG data collection not being optimal, we were able to collect and compare some brain wave activity data in the study groups during the sessions. Even though the amplitude of brain wave activity can result in large intraindividual variation, the data were evenly distributed and there were no mean differences between the groups in our study. While we observed no significant association between scores of usability and β wave activity (more likely to be associated with frustration, agitation, or perhaps excitement), we did observe brain wave activity that is typically associated with relaxed states of mind [18-21] and this had a positive linear relationship with SUS-10 score. This indicated that the chat-only group was either more relaxed or less aroused (or both). The explanation could lie in the combination of the small sample and the fact that more individuals in the voice-only chat group reported feeling nervousness, a feeling that generally elicits higher brain wave activity and less relaxation [20,47]. With the low quality of data, combined with a limited sample, it is hard to draw any generalizable conclusions from the biometric data.

Feelings of closeness did not differ between the 2 interfaces and seem not to have been affected by the presence of anthropomorphic features. When gender was explored as a factor, there was no significant difference to what extent men and women reported feeling close to BETSY in the respective assigned interfaces. Due to the small sample size in our study, it was not possible to perform further and more elaborate designs looking at mediation of other demographic or biometric factors in a reliable way.

When devising chatbots for mental health, this study indicates that a mixed approach might be the best course of action, allowing the user to choose a preferred way of interacting with the chatbot.

Limitations

This study consisted of healthy volunteers. It is good to keep in mind that mental health issues can affect some parts of cognitive performance [48] and, thus, usability may not be equally perceived by a person in a state of emotional distress and a healthy volunteer. Further investigation and collaboration are needed in future studies to capture the usability aspects of individuals who are in an active state of distress.

The results of this study suggest that overall usability seems to be perceived as higher for the text-only chatbot interface and no significant emotional boost was present with the addition of anthropomorphic features to a digital human chatbot.

Further studies which include a larger sample of participants as well as participants who experience mild to moderate anxiety are needed to explore and further evaluate the research question posed in this paper. In this study, the age range was limited, and the variable was incomplete. In future studies, we will strive to include more young adults and adults older than 60 years.

Large language models and application programming interface models were not available at the time this chatbot was constructed and neither were Metahuman creator or more advanced voice-cloning or voice-generating options, which would have significantly improved the anthropomorphic features of the digital human. The first iteration of the generative pretrained transformer was not available to the public and the generative pretrained transformer-3 application programming interface had a limited release during the development of this project: it was not available to our team until a year after the project was completed. With large language models, repetitiveness and limitations in terms of variability of answers would have most likely been avoided; however, the aim of our investigation was not the general effect of the content but rather the perception of text- versus voice-driven interfaces.

Conclusions

In conclusion, the text-only chatbot was perceived as more user-friendly in terms of usability indicators for SUS-10. However, both the digital human and text-only interfaces scored average or above average in comparison to other studies performed on mental health chatbots. Although biometric data did not differ significantly, we saw significant gender differences in terms of prechat positivity and postchat annoyance, which is contrary to other studies. Male participants in our study were more likely to report higher prechat positivity toward BETSY and report less irritation postchat. SUISQ-MR also indicated that BETSY's overall usability and voice were highly ranked compared with other studies, indicating that there is great promise for mental health chatbots independently of the chosen user interface.

Acknowledgments

The authors would like to thank the participants of this study for their time and contribution to valuable knowledge. The authors also especially thank Mathias Alvidius, head of department, for contributing significantly to the possibilities of the creation of this research project. The authors would also like to express sincere gratitude to Linn Dahlén Ölander for her help in recruiting participants. SS received support from the Söderström König Foundation (SLS-974247), Stiftelsen Psykiatriska Forskningsfonden for this project. Parts of this project were funded by Innovationsfonden VGR (project VGRINN-936800) and Västra Götalandsregionen. The authors would also like to thank Peter Todd (Tajut Ltd) for third-party editorial assistance in the drafting of this manuscript, for which he received financial compensation.

Conflicts of Interest

SS receives fees for scientific consultation from Mindforce.

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 107 KB - humanfactors_v11i1e54581_app1.pdf]

References

1. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. J Am Med Inform Assoc 2011;18(5):544-551 [FREE Full text] [doi: 10.1136/amiajnl-2011-000464] [Medline: 21846786]
2. Weizenbaum J. ELIZA—a computer program for the study of natural language communication between man and machine. Commun ACM 1966;9(1):36-45 [FREE Full text] [doi: 10.1145/365153.365168]
3. Shum H, He X, Li D. From Eliza to XiaoIce: challenges and opportunities with social chatbots. Front Inf Technol Electronic Eng 2018;19(1):10-26 [FREE Full text] [doi: 10.1631/fitee.1700826]
4. Lim SM, Shiau CWC, Cheng LJ, Lau Y. Chatbot-delivered psychotherapy for adults with depressive and anxiety symptoms: a systematic review and meta-regression. Behav Ther 2022;53(2):334-347 [FREE Full text] [doi: 10.1016/j.beth.2021.09.007] [Medline: 35227408]
5. Radziwill NM, Benton MC. Evaluating quality of chatbots and intelligent conversational agents. ArXiv. Preprint posted online on April 15 2017 [FREE Full text]
6. Dale R. GPT-3: what's it good for? Nat Lang Eng 2020;27(1):113-118 [FREE Full text] [doi: 10.1017/S1351324920000601]
7. Homolak J. Opportunities and risks of ChatGPT in medicine, science, and academic publishing: a modern Promethean dilemma. Croat Med J 2023;64(1):1-3 [FREE Full text] [doi: 10.3325/cmj.2023.64.1] [Medline: 36864812]
8. Caldarini G, Jaf S, McGarry K. A literature survey of recent advances in chatbots. Information 2022;13(1):41. [doi: 10.3390/info13010041]
9. Cameron G, Cameron D, Megaw G, Bond R, Mulvenna M, O'Neill S, et al. Towards a chatbot for digital counselling. 2017 Presented at: Proceedings of the 31st International BCS Human Computer Interaction Conference (HCI 2017) (HCI); July 3-6, 2017; Sunderland, UK URL: https://www.scienceopen.com/hosted-document?doi=10.14236/ewic/HCI2017.24 [doi: 10.14236/ewic/hci2017.24]

10. Denecke K, Vaaheesan S, Arulnathan A. A mental health chatbot for regulating emotions (SERMO) - concept and usability test. IEEE Trans Emerg Topics Comput 2020;9(3):1170-1182 [FREE Full text] [doi: 10.1109/tetc.2020.2974478]
11. Torous J, Bucci S, Bell IH, Kessing LV, Faurholt-Jepsen M, Whelan P, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021;20(3):318-335 [FREE Full text] [doi: 10.1002/wps.20883] [Medline: 34505369]
12. Oh J, Jang S, Kim H, Kim JJ. Efficacy of mobile app-based interactive cognitive behavioral therapy using a chatbot for panic disorder. Int J Med Inform 2020;140:104171 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104171] [Medline: 32446158]
13. Kaywan P, Ahmed K, Ibaida A, Miao Y, Gu B. Early detection of depression using a conversational AI bot: a non-clinical trial. PLoS One 2023;18(2):e0279743 [FREE Full text] [doi: 10.1371/journal.pone.0279743] [Medline: 36735701]
14. Islam MN, Khan SR, Islam NN, Rownok RA, Zaman SR, Zaman SR. A mobile application for mental health care during COVID-19 pandemic: development and usability evaluation with System Usability Scale. : Springer; 2021 Presented at: Computational Intelligence in Information Systems: Proceedings of the Computational Intelligence in Information Systems Conference (CIIS 2020); January 25–27, 2021; Brunei Darussalam p. 33-42. [doi: 10.1007/978-3-030-68133-3_4]
15. Heller B, Proctor M, Mah D, Jewell L, Cheung B. Freudbot: an investigation of chatbot technology in distance education. In: EdMedia+ Innovate Learning. Paris, France: Association for the Advancement of Computing in Education (AACE); 2005:3913-3918.
16. Kadariya D, Venkataramanan R, Yip HY, Kalra M, Thirunarayanan K, Sheth A. kBot: knowledge-enabled personalized chatbot for asthma self-management. Proc Int Conf Smart Comput SMARTCOMP 2019;2019:138-143 [FREE Full text] [doi: 10.1109/smartcomp.2019.00043] [Medline: 32832938]
17. Sagar M, Seymour M, Henderson A. Creating connection with autonomous facial animation. Commun ACM 2016;59(12):82-91 [FREE Full text] [doi: 10.1145/2950041]
18. Allen JJB, Reznik SJ. Frontal EEG asymmetry as a promising marker of depression vulnerability: summary and methodological considerations. Curr Opin Psychol 2015;4:93-97 [FREE Full text] [doi: 10.1016/j.copsyc.2014.12.017] [Medline: 26462291]
19. Hagemann D, Naumann E, Lürken A, Becker G, Maier S, Bartussek D. EEG asymmetry, dispositional mood and personality. Pers Individ Dif 1999;27(3):541-568 [FREE Full text] [doi: 10.1016/s0191-8869(98)00263-3]
20. Lomas T, Ivtzan I, Fu CHY. A systematic review of the neurophysiology of mindfulness on EEG oscillations. Neurosci Biobehav Rev 2015;57:401-410 [FREE Full text] [doi: 10.1016/j.neubiorev.2015.09.018] [Medline: 26441373]
21. Al-Nafjan A, Hosny M, Al-Ohali Y, Al-Wabil A. Review and classification of emotion recognition based on EEG brain-computer interface system research: a systematic review. Appl Sci 2017;7(12):1239. [doi: 10.3390/app7121239]
22. Bos AS, Pizzato MC, Vettori M, Donato LG, Soares PP, Fagundes JG, et al. Empirical evidence during the implementation of an educational chatbot with the electroencephalogram metric. Creat Educ 2020;11(11):2337-2345 [FREE Full text] [doi: 10.4236/ce.2020.1111171]
23. Ciechanowski L, Przegalinska A, Magnuski M, Gloor P. In the shades of the uncanny valley: an experimental study of human–chatbot interaction. Future Gener Comput Syst 2019;92:539-548 [FREE Full text] [doi: 10.1016/j.future.2018.01.055]
24. Lewis JR. Standardized questionnaires for voice interaction design. Voice Interact Des 2016;1(1):1-16 [FREE Full text] [doi: 10.1016/b978-1-55860-768-2.x5000-2]
25. Lewis JR. The System Usability Scale: past, present, and future. Int J Hum Comput Interact 2018;34(7):577-590 [FREE Full text] [doi: 10.1080/10447318.2018.1455307]
26. Lewis JR, Sauro J. The factor structure of the System Usability Scale. 2009 Presented at: Human Centered Design: First International Conference, HCD 2009, held as part of HCI International; July 19-24, 2009; San Diego, CA. [doi: 10.1007/978-3-642-02806-9_12]
27. Lewis JR, Hardzinski ML. Investigating the psychometric properties of the speech user interface service quality questionnaire. Int J Speech Technol 2015;18(3):479-487 [FREE Full text] [doi: 10.1007/s10772-015-9289-1]
28. Peres SC, Pham T, Phillips R. Validation of the System Usability Scale (SUS): SUS in the wild. Proc Hum Factors Ergon Soc Annu Meet 2013;57(1):192-196 [FREE Full text] [doi: 10.1177/1541931213571043]
29. Grier RA, Bangor A, Kortum P, Peres SC. The System Usability Scale: beyond standard usability testing. Proc Hum Factors Ergon Soc Annu Meet 2013;57(1):187-191 [FREE Full text] [doi: 10.1177/1541931213571042]
30. Kortum PT, Bangor A. Usability ratings for everyday products measured with the System Usability Scale. Int J Hum Comput Interact 2013;29(2):67-76 [FREE Full text] [doi: 10.1080/10447318.2012.681221]
31. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008;24(6):574-594 [FREE Full text] [doi: 10.1080/10447310802205776]
32. Bennion MR, Hardy GE, Moore RK, Kellett S, Millings A. Usability, acceptability, and effectiveness of web-based conversational agents to facilitate problem solving in older adults: controlled study. J Med Internet Res 2020;22(5):e16794 [FREE Full text] [doi: 10.2196/16794] [Medline: 32384055]
33. Ly KH, Ly AM, Andersson G. A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. Internet Interv 2017;10:39-46 [FREE Full text] [doi: 10.1016/j.invent.2017.10.002] [Medline: 30135751]

34. Ta V, Griffith C, Boatfield C, Wang X, Civitello M, Bader H, et al. User experiences of social support from companion chatbots in everyday contexts: thematic analysis. J Med Internet Res 2020;22(3):e16235 [FREE Full text] [doi: 10.2196/16235] [Medline: 32141837]
35. van Wezel MMC, Croes EAJ, Antheunis ML. “I’m here for you”: can social chatbots truly support their users? a literature review. 2020 Presented at: Chatbot Research and Design: Proceedings of the 4th International Workshop, CONVERSATIONS 2020; November 23-24, 2020; Virtual Event p. 96-113. [doi: 10.1007/978-3-030-68288-0_7]
36. Crolic C, Thomaz F, Hadi R, Stephen AT. Blame the bot: anthropomorphism and anger in customer–chatbot interactions. J Mark 2021;86(1):132-148 [FREE Full text] [doi: 10.1177/00222429211045687]
37. Greer S, Ramo D, Chang YJ, Fu M, Moskowitz J, Haritatos J. Use of the chatbot "Vivibot" to deliver positive psychology skills and promote well-being among young people after cancer treatment: randomized controlled feasibility trial. JMIR Mhealth Uhealth 2019;7(10):e15018 [FREE Full text] [doi: 10.2196/15018] [Medline: 31674920]
38. Jang S, Kim JJ, Kim SJ, Hong J, Kim S, Kim E. Mobile app-based chatbot to deliver cognitive behavioral therapy and psychoeducation for adults with attention deficit: a development and feasibility/usability study. Int J Med Inform 2021;150:104440 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104440] [Medline: 33799055]
39. Loveys K, Sagar M, Pickering I, Broadbent E. A digital human for delivering a remote loneliness and stress intervention to at-risk younger and older adults during the COVID-19 pandemic: randomized pilot trial. JMIR Ment Health 2021;8(11):e31586 [FREE Full text] [doi: 10.2196/31586] [Medline: 34596572]
40. Ruane E, Farrell S, Ventresque A. User perception of text-based chatbot personality. 2020 Presented at: Chatbot Research and Design: Proceedings of the 4th International Workshop, CONVERSATIONS 2020; November 23-24, 2020; Virtual event p. 32-47. [doi: 10.1007/978-3-030-68288-0_3]
41. Yen C, Chiang MC. Trust me, if you can: a study on the factors that influence consumers’ purchase intention triggered by chatbots based on brain image evidence and self-reported assessments. Behav Inf Technol 2020;40(11):1177-1194 [FREE Full text] [doi: 10.1080/0144929x.2020.1743362]
42. Cheng Y, Jiang H. How do AI-driven chatbots impact user experience? examining gratifications, perceived privacy risk, satisfaction, loyalty, and continued use. J Broadcast Electron Media 2020;64(4):592-614 [FREE Full text] [doi: 10.1080/08838151.2020.1834296]
43. Smestad TL, Volden F. Chatbot personalities matters: improving the user experience of chatbot interfaces. 2018 Presented at: Internet Science: 5th International Conference, INSCI 2018; October 24–26, 2018; St. Petersburg, Russia p. 170-181. [doi: 10.1007/978-3-030-17705-8_15]
44. Silvervarg A, Raukola K, Haake M, Gulz A. The effect of visual gender on abuse in conversation with ECAs. 2012 Presented at: Intelligent Virtual Agents: 12th International Conference, IVA 2012; September 12-14, 2012; Santa Cruz, CA p. 153-160. [doi: 10.1007/978-3-642-33197-8_16]
45. Luger E, Sellen A. "Like having a really bad PA": the gulf between user expectation and experience of conversational agents. 2016 Presented at: CHI'16: CHI Conference on Human Factors in Computing Systems; May 7-12, 2016; San Jose, CA p. 5286-5297 URL: https://dl.acm.org/doi/10.1145/2858036.2858288 [doi: 10.1145/2858036.2858288]
46. Hearst M, Tory M. Would you like a chart with that? incorporating visualizations into conversational interfaces. 2019 Presented at: 2019 IEEE Visualization Conference (VIS); October 20-25, 2019; Vancouver, BC p. 1-5. [doi: 10.1109/visual.2019.8933766]
47. Díaz H, Cid FM, Otárola J, Rojas R, Alarcón O, Cañete L. EEG Beta band frequency domain evaluation for assessing stress and anxiety in resting, eyes closed, basal conditions. Procedia Comput Sci 2019;162:974-981 [FREE Full text] [doi: 10.1016/j.procs.2019.12.075]
48. Eysenck MW, Derakshan N, Santos R, Calvo MG. Anxiety and cognitive performance: attentional control theory. Emotion 2007;7(2):336-353. [doi: 10.1037/1528-3542.7.2.336] [Medline: 17516812]

Abbreviations

BETSY: Behavior, Emotion, Therapy System, and You
EEG: electroencephalography
GAD-7: Generalized Anxiety Disorder-7
SUS-10: System Usability Scale-10
SUISQ-MR: Standardized Questionnaires for Voice Interaction Design Short Version

Edited by A Kushniruk; submitted 15.11.23; peer-reviewed by R Weeks, R Pryss; comments to author 22.12.23; revised version received 27.01.24; accepted 18.02.24; published 29.04.24.

Please cite as:
Thunström AO, Carlsen HK, Ali L, Larson T, Hellström A, Steingrimsson S
Usability Comparison Among Healthy Participants of an Anthropomorphic Digital Human and a Text-Based Chatbot as a Responder to Questions on Mental Health: Randomized Controlled Trial
JMIR Hum Factors 2024;11:e54581
URL: https://humanfactors.jmir.org/2024/1/e54581
doi:10.2196/54581
PMID:38683664

Original Paper

Identifying Factors of User Acceptance of a Drone-Based Medication Delivery: User-Centered Design Approach

Franziska Fink[1,2*], Dr rer nat; Ivonne Kalter[1,2*], MA; Jenny-Victoria Steindorff[1,2], MSc; Hans Konrad Helmbold[3], MSc; Denny Paulicke[1,2,4], Prof Dr; Patrick Jahn[1,2], Prof Dr

[1]Translation Region for Digitalised Healthcare, Department of Internal Medicine, Faculty of Medicine, University Medicine, Martin-Luther-University Halle-Wittenberg, Halle (Saale), Germany

[2]Health Service Research Working Group | Acute Care, Department of Internal Medicine, Faculty of Medicine, University Medicine, Martin-Luther-University Halle-Wittenberg, Halle (Saale), Germany

[3]Department of Economics, Anhalt University of Applied Sciences, Bernburg, Germany

[4]Department of Medical Pedagogy, Akkon University of Human Sciences, Berlin, Germany

[*]these authors contributed equally

Corresponding Author:
Patrick Jahn, Prof Dr
Translation Region for Digitalised Healthcare
Department of Internal Medicine, Faculty of Medicine, University Medicine
Martin-Luther-University Halle-Wittenberg
Magdeburger Str. 12
Halle (Saale), 06112
Germany
Phone: 49 3455575493
Email: patrick.jahn@uk-halle.de

Abstract

Background: The use of drones in the health care sector is increasingly being discussed against the background of the aging population and the growing shortage of skilled workers. In particular, the use of drones to provide medication in rural areas could bring advantages for the care of people with and without a need for care. However, there are hardly any data available that focus on the interaction between humans and drones.

Objective: This study aims to disclose and analyze factors associated with user acceptance of drone-based medication delivery to derive practice-relevant guidance points for participatory technology development (for apps and drones).

Methods: A controlled mixed methods study was conducted that supports the technical development process of an app design for drone-assisted drug delivery based on a participatory research design. For the quantitative analysis, established and standardized survey instruments to capture technology acceptance, such as the System Usability Scale; Technology Usage Inventory (TUI); and the Motivation, Engagement, and Thriving in User Experience model, were used. To avoid possible biasing effects from a continuous user development (eg, response shifts and learning effects), an ad hoc group was formed at each of the 3 iterative development steps and was subsequently compared with the consisting core group, which went through all 3 iterations.

Results: The study found a positive correlation between the usability of a pharmacy drone app and participants' willingness to use it (r=0.833). Participants' perception of usefulness positively influenced their willingness to use the app (r=0.487; TUI). Skepticism had a negative impact on perceived usability and willingness to use it (r=−0.542; System Usability Scale and r=−0.446; TUI). The study found that usefulness, skepticism, and curiosity explained most of the intention to use the app ($F_{3,17}$=21.12; P<.001; R^2=0.788; adjusted R^2=0.751). The core group showed higher ratings on the intention to use the pharmacy drone app than the ad hoc groups. Results of the 2-tailed t tests showed a higher rating on usability for the third iteration of the core group compared with the first iteration.

Conclusions: With the help of the participatory design, important aspects of acceptance could be revealed by the people involved in relation to drone-assisted drug delivery. For example, the length of time spent using the technology is an important factor for the intention to use the app. Technology-specific factors such as user-friendliness or curiosity are directly related to the use acceptance of the drone app. Results of this study showed that the more participants perceived their own competence in handling the app, the more they were willing to use the technology and the more they rated the app as usable.

(JMIR Hum Factors 2024;11:e51587) doi:10.2196/51587

KEYWORDS
human-drone interaction; medical supplies; participative research; user-centered design; technology acceptance

Introduction

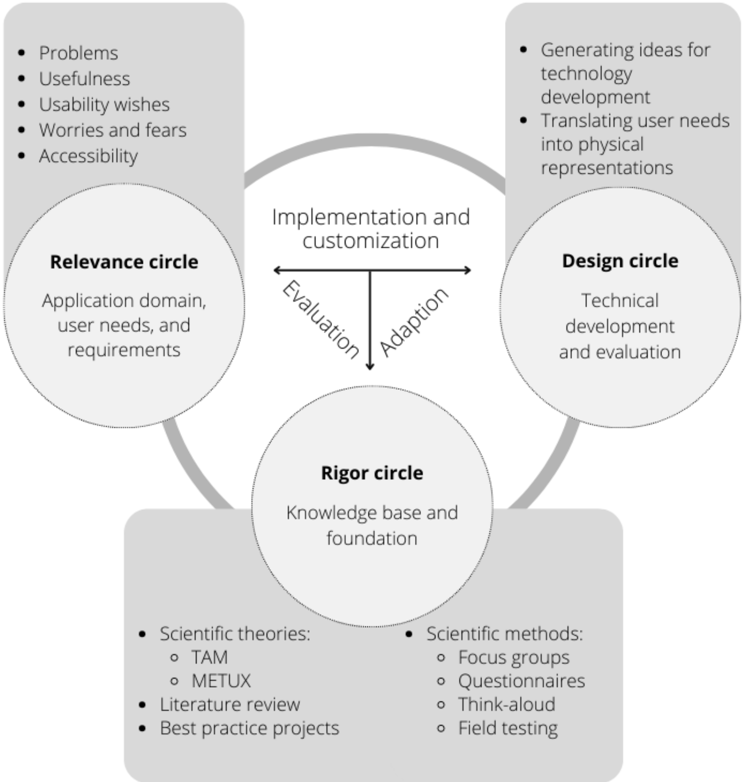
Background

The health care system faces challenges such as a rural exodus, aging populations, and increasing shortages in the health care workforce; drones have the potential to increase the efficiency and capacity of the health care system [1]. The COVID-19 pandemic intensified the demand for new logistic solutions, such as fast and contactless delivery strategies [2]. It is also important to understand the attitude of the civilian population and public opinion on the use of drones [3]. In this vein, delivering medications with drones is the most identified application in health care [1,4]. There are a few studies showing that usability, lack of user skills and expertise, and negative perceptions affect user acceptance and hinder drone use [1,5,6]. Therefore, it is particularly important that all user groups are involved at an early stage. Furthermore, after a recent scoping review showed that there were no empirical studies on user acceptance of drone-based medication delivery [7], we could only find 1 study in Asia that investigated user acceptance among health care professionals in the delivery of drone-based medication [8]. The successful application of technology is predominantly determined by the type and extent of acceptance [9,10]. Acceptance in this context refers to the positive acceptance decision of an innovation by its users, which is described in the technology acceptance model (TAM; perceived usability, usefulness, immersion, and accessibility: TAM 1) [11,12]. It proposes that users tend to use a technology when they believe it will help them to perform a better job (perceived usefulness) and when they believe that the system can be handled without effort (perceived ease of use). These variables were found to correlate with the intention to use, wherein usefulness was substantially more strongly related to frequency of use than ease of use. Nevertheless, both are strong correlates of user acceptance and should not be ignored by designing and implementing successful technologies [12]. In other words, the greater the benefit of a technology and the simpler its usability, the more the users are willing to use the new system. However, some more variables that affect user acceptance such as social influences (subjective norm, image, and voluntariness), cognitive instrumental processes (job relevance, output quality, and result demonstrability; TAM 2), and psychological foundations (self-efficacy, external control, playfulness, anxiety, enjoyment, and usability; TAM 3) can be stated [13,14]. Thus, Peters et al [15] argued that this model alone does not indicate whether people would actually use a technology. In this context, basic human needs according to Ryan and Deci [16,17] and Deci and Ryan [18] play an important role. Following the Basic Psychological Need Theory [17], the more the interaction with the system satisfies basic psychological needs, the more the users will engage with a technology. Ryan and Deci [16,17] and Deci and Ryan [18], defined 3 basic psychological needs in their self-determination theory that are crucial to whether a person is proactive and engaged or passive and demotivated. These needs include competence (ie, feeling capable), relatedness (ie, feeling connected to others), and autonomy (ie, feeling self-determined). On the basis of this, Peters et al [15] defined the Motivation, Engagement, Thriving in User Experience (METUX) model, which can be used for the evaluation and iterative design process of technologies to increase motivation, engagement, and well-being. In this case, the 3 basic needs are mediating variables between a technical product and the well-being of the users. This implies that as autonomy increases, engagement increases; as competence increases, motivation increases; and as relatedness increases, well-being increases.

This Study

Acceptance building is a process that starts before the initial contact with the innovation and continues into the application phase, which was addressed within the pharmacy drone study (Apotheken-Drohnen-App; ADApp) [19]. This study represents a section of the whole ADApp project by investigating factors that are associated with the user acceptance of a drone-based medication delivery to be able to derive practice-relevant orientation points for participatory technology development (for apps and drones). Stephan et al [7] described that little attention is paid to the design phases of drones including the delivery process. Thus, this study used a mixed methods design and followed the methodological guidelines of the cocreative user-centered design proposed by Farao et al [20]. User-centered design is used to help consider the context of technologies as well as their implementation consequences at the design stage (Figure 1) [20,21]. Implementing technologies without user involvement may compromise the desired outcome, which in turn can lead to unmet health goals and adverse outcomes [20,22,23]. It is an evidence-based approach that involves users in developing technologies and prioritizes their needs [20,24]. In contrast to classical user-centered designs, this study used a controlled design for the first time. Traditionally, small sample groups are observed or interviewed or participate in usability tests during the testing and development phases of new technologies (usually operationalized through the think-aloud [TA] method and questionnaires). These are important approaches to get insights into key needs of the target population [25]. However, repeated measurements cause a change in the meaning of test scores, which makes it difficult to compare repeated measures. In a measurement perspective, it can be considered as bias in the measurement of change [26]. It can be inferred that conducting repeated measures with the same sample group may alter their attitudes, expectations, and behaviors in interacting with the technology. This could potentially influence their acceptance of the technology, as they become aware that the technology will adjust to their needs. Moreover, they are not unfamiliar with dealing with the app, which might influence user acceptance as well. This is indeed the purpose of a user-centered design, but it loses insights into perspectives of inexperienced users without a concrete expectation about the changes in technology after giving feedback to it. By integrating a control group (called the *ad hocgroup*), this study aimed to investigate whether the assumption of the *core group* (ie, the same sample group at all measurement time points) is generalizable to a broader population. In this regard, a second purpose of this study was to answer the question of whether user acceptance differs between the core and ad hoc groups.

Figure 1. Implementation circle. METUX: Motivation, Engagement, Thriving in User Experience; TAM: technology acceptance model.



Methods

Study Design

The monocentric ADApp study aimed at an iterative and cocreative development of a pharmacy drone app with multiple measurement time points (Table 1). As this study design was embedded in the broader ADApp project, it was preceded by 2 research steps. As a first step, we conducted a scoping review of experimental studies examining the interaction between humans and drones during the delivery of drugs and defibrillators to identify research gaps and explore the scope of research activities [7]. In the next step, problems, needs, and requirements of the users were identified and concrete scenarios were outlined, which were necessary for the implementation of a first demonstrator of the app [27]. At this point, we decided to use focus groups instead of individual interviews because it allows participants to respond to each other's answers and gives us the most information. In this study, we conducted 3 iteration loops, where we tested the app demonstrators a well as the entire ADApp flow from order to delivery at an airport along with the user groups. The iterative process is one of the central features of the study and will be discussed in detail in the Study Setting section.

Table 1. The Apotheken-Drohnen-App design.

Goal	Methods	Participation	Time point	Reference
Knowledge	Scoping review	N/A[a]	August 2021	Stephan et al [7], 2022
Needs	Focus groups	User groups	October 6, 2021	Fink et al [27], 2023
Evaluation of functionalities	Questionnaires and think-aloud	First and second iteration loop: core group; 2 ad hoc groups	July 3, 2022	This study
Test flights	Questionnaire, think-aloud, and focus groups	Third iteration loop: core group; 1 ad hoc group	October 2022	This study

[a]N/A: not applicable.

To address potential biasing effects owing to a response shift through repeat interviews with the core group, a total of 3 ad hoc groups were acquired in this study. At each of the 3 development steps, a new and naive ad hoc group was used. This unique approach of adding ad hoc groups as control groups also allows for the generalizability of key needs identified in the development process [25].

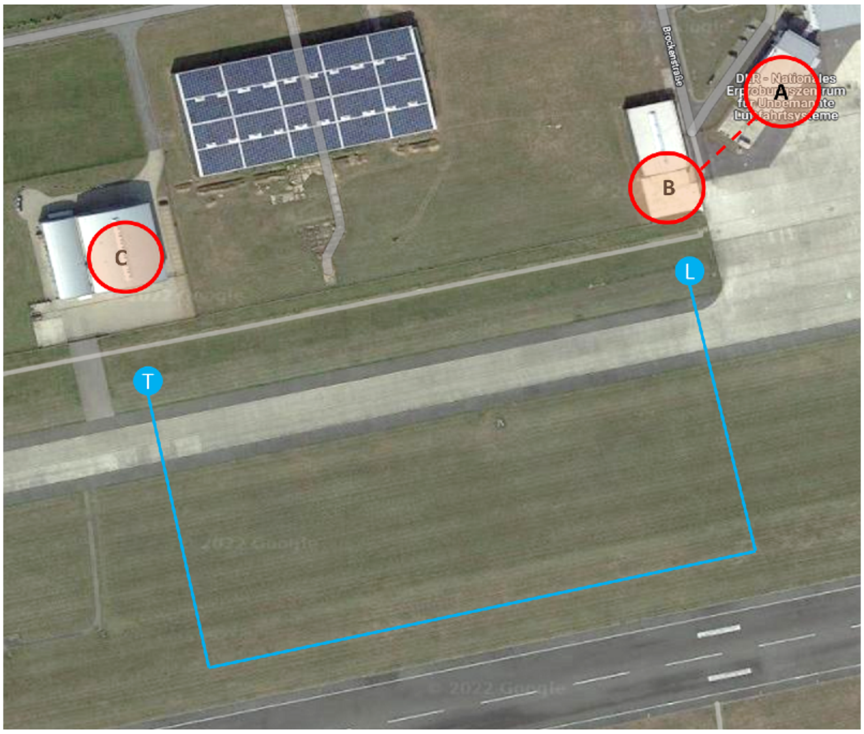
Study Setting

Our prior work showed that a pharmacy drone app, for example, must be lean and simple; must facilitate the user's performance (eg, software integration, shipment tracking, and plannability); must enable control (eg, handover identification and data sovereignty); and must include consultation and reconciliation features. The most frequently discussed problems associated with drones were the physical contact with the drone and the drone's noise [27]. Following the focus groups, the developers designed a demonstrator of the app including communication features (eg, an extra text field for communication with the pharmacy), plannability features (eg, time slots for drone delivery), shipment tracking, and features for enabling handover security, which was tested in 3 iteration loops with users. All loops were conducted separately per the core and ad hoc groups.

In the *first iteration loop* (conducted between March and April 2022), the usability of the app was reviewed. The app should be as easy as possible to use and allow users to create an account, add details such as delivery location, or submit an electronic prescription. The first iteration loop aimed at evaluating how intuitive the app is. For this purpose, users were shown the app and asked to think aloud while using it—without any introduction. For each participant, one experimenter prepared a protocol for taking notes. Finally, the prototypes were queried by using additional questionnaires: the Technology Usage Inventory (TUI) assessment was used to assess the intention to use, and the System Usability Scale (SUS) was used to measure usability. The basic psychological needs such as competence, autonomy, and connectedness were identified at the task level through the Technology-based Experience of Need Satisfaction (TENS) Task. After the first iteration loop, the demonstrator was adapted according to user feedback. The *second iteration loop* (conducted between June and July 2022) was dedicated to the design of the prototype and its technical development and evaluation. The decisive factor was how intuitive the design is and whether the tasks of the app are adequately represented. The procedure was the same as that in the first iteration loop, with the exception that instead of the TENS Task, basic psychological needs were queried via the TENS Interface. To gain deep insights into the participants' thoughts while using the technology, the core and ad hoc groups were again asked to think aloud while using the app. After receiving the second round of feedback from participants, a third demonstrator was developed. However, the *third iteration loop* (conducted in October 2022) aimed at testing the overall process starting with registration, setting the delivery location, submitting a prescription, and actually receiving a delivery. As focus groups showed concerns about injuries with the drone, this study wanted to test the handover via a winch, a parachute, or dropping to reduce physical contact with the drone, but unfortunately, owing to regulations and restrictions, testing the handover was not possible [27]. Thus, the drone had to land (Figure 2). Therefore, we divided participants into 2 groups: one group tested the app, whereas the other group talked about different handover scenarios and looked at the drone from close up. After each round, the groups were swapped. Similar to that in the first and second iteration loops, the intention to use and usability was queried using the TUI and SUS. To gauge the degree to which the technology fulfills users' needs in terms of behavior, participants completed the Basic Psychological Need Satisfaction and Frustration Scale (BPNSFS). Again, the participants were instructed to think aloud while going through the tasks in the app to gain more insights into the functionality of the app. After participants submitted their prescription, they went to the location in the airport where the drone was set to land (Figure 3). After landing, participants took the *medication* (a bag of gummy bears) out of the drone. After the testing, a short discussion was held with all participants to sum up their impressions.

Figure 2. Drone landing.

Figure 3. Airport landing zone. (A) Order location, (B) reception location, and (C) drone starting point. L: landing; T: takeoff.



Participants

Participants were chosen to represent potential user groups for both the drone-based medication delivery service and the supply chain, based on their respective role characteristics: physicians, nurses, pharmacists, and interested users, especially patients with COVID-19 and relatives of patients who need palliative care. Participants of the ad hoc group were recruited with the support of the Merseburger Innovations- und Technologiezentrum (innovation and technology center) and the ADApp project team. They were contacted via email or telephone. Core group participants were recruited from the focus group that was conducted in October 2021 [27]. A total of 3 pharmacists, 2 nurses, 3 general practitioners, and 1 patient with COVID-19 were recruited from the focus groups. Owing to the underrepresentation of interested users or patients, 2 more participants were recruited with the support of the ADApp project team.

All participants were informed about the general aim and reasons of the study and the procedure. They gave written informed consent before starting the iteration loops.

Ethics Approval

The ADApp study was approved by the Ethics Committee of the Martin-Luther University Halle-Wittenberg (protocol code 2021–069; date of approval May 6, 2021).

Measures

Quantitative Measures

SUS Score

The SUS measures the user's subjective perception of the usability of a system. It is technology independent, that is, it can be used for a wide range of systems and technologies [28-30]. Overall, 10 items were divided into 5 positively and 5 negatively directed statements, each represented on a 5-point Likert scale. The answers provided the SUS item score, which were then converted into the SUS overall score. The overall score ranges from 0 to 100. To calculate the SUS overall score, the first step was to calculate the raw score minus 1 for all odd items, whereas the raw score of 5 was subtracted for the even items. For example, if for item 1, the raw score is 4, the result is a score of 3 (4-1). If for item 2 the raw score is 2, the score is 3 (5-2). In the next step, the calculated scores are summed and multiplied by 2.5 [31].

Systems can be considered usable if they achieve the benchmark score of ≥68 [29,31]. In preliminary works, an adjective scale was developed for a more comprehensible classification of the SUS score, which ranges from outstanding (score 90-100) to very poor (score 0-34) [32].

TUI Assessment

The TUI assessment [33] aims to measure the intention to use and is based on the TAM [12]. The intention to use a technology is a comprehensive construct based on a variety of explanatory factors. As suggested in the TAM 2 and TAM 3, the TUI considers technology-specific factors and psychological factors. The TUI therefore supplements the classic technology acceptance factors of the TAM 1, such as perceived usability, usefulness, immersion, and accessibility (technology specific) with important psychological constructs, such as technology anxiety, curiosity, interest, and skepticism. The items, with the exception of the technology anxiety and interest scale, are related to a specific technology. Each scale consists of 3 to 4 items, each of which is to be rated on a 7-point Likert scale. The ninth scale measures the intention to use a specific technology with 3 items on a visual analog scale (each 100 mm). In total, the TUI consists of 33 items and has a modular design so that individual scales can be excluded and item formulations can be adapted to the circumstances (eg, concrete technology names). With the exception of the "immersion" scale, all scales were used this study. All answers of a scale were summed to a sum value. It starts with 1 as the lowest expression of the construct and goes up to 21 (3 items) or 28 (4 items) as the highest expression, depending on the number of items. For the intention to use scale, the distance from the right end point (full rejection) to the answer cross on the line are measured. The distance in mm is determined for all 3 items and summed. The maximum scale value to be achieved is 300. The determined scale sum values are converted into standard values (stanines). The stanines reach from 1 (strongly below average) to 9 (strongly above average) [33].

METUX Model

The "pure" usability does not necessarily predict higher use of a technology [34]. Although the SUS questionnaire focuses on usability, the TUI questionnaire also includes psychological factors. However, both neglect basic psychological needs, which are addressed in the METUX model. It aims to optimize engagement, motivation, and well-being of technologies in iterative design processes [15]. Within the model, different spheres of experience were assumed to influence well-being: interface (ie, interacting with the technology), tasks (ie, engaging with the technology), behavior (ie, the relation to the overarching technology-supported behavior), and life (ie, the overall experience outside and beyond the technology). There are different questionnaires for measuring the basic psychological needs in different spheres. The spheres adoption, interface, tasks, and behavior were tested within this study using the Autonomy and Competence in Technology Adoption Questionnaire for the first adoption process, the TENS [15] for interface and task, and the BPNSFS [35,36] for behavior. The sphere "life" was not relevant for this study. The Autonomy and Competence in Technology Adoption Questionnaire addresses the question of why people use a technology and to what extent they experience themselves as competent to use it. It consists of 2 parts: the first, self-regulation, includes 12 items; and the second, perceived competence, includes 2 items, which are represented on a 5-point Likert scale. The goal of the TENS Interface and Task questionnaires is to assess the extent to which direct interaction (via interface) with a technology and engagement in technology-specific tasks satisfies the basic psychological needs for autonomy, competence, and relatedness [15]. In the TENS questionnaires, the items are each assigned to the basic needs of competence, autonomy, and relatedness but are presented randomly in the questionnaire. All items were equally weighted, summed, and averaged per basic need. The TENS Interface questionnaire consists of 15 items with 5 items per need, whereas the TENS Task questionnaire consists of 12 items with 4 items per need. The items are each represented on a 5-point Likert scale. As the TENS questionnaires are only validated for English-speaking countries so far, a linguistic validation of the questionnaires was conducted. For this purpose, the questionnaires were translated into German by an interpreter whose native language is German and who is fluent in English. Nevertheless, both the TENS questionnaires are not standardized for the German population that has to be considered when interpreting the results.

The BPNSFS is intended to assess the extent to which a technology improves need satisfaction in relation to the behavior the technology is intended to support [15]. The BPNSFS measures the satisfaction and frustration of the basic psychological needs of autonomy, competence, and relatedness. This includes a balanced combination of subscales for satisfaction and frustration. This distinction is necessary because the absence of need satisfaction does not equate to frustration of the same [17,35]. On the basis of the original questionnaire, Heissel et al [36] identified 6 different, but intercorrelating, factors with 4 items each for the German version of the BPNSFS: Autonomy Satisfaction, Autonomy Frustration, Competence Satisfaction, Competence Frustration, Relatedness Satisfaction, and Relatedness Frustration. Each response is represented on a 5-point Likert scale. The evaluation of the BPNSFS can be handled differently; for this study, the items per basic need are summed, and the 12 items of the subscales satisfaction and frustration are summed. There exist several adaptations of the BPNSFS, which have been validated and subjected to reliability testing. These adaptations include language translations, adjustment for age (children or adults), domain (sports, work, and romantic relationships), and clinical status (HIV, intellectual impairment, and chronic pain). However, a questionnaire related to technological aspects does not yet exist. Thus, for this study, the German version of the BPNSFS was used and minimally adapted according to the wording.

Qualitative Measures: TA Method

TA has traditionally been used in psychology and education to identify cognitive processes that occur internalized in the context of problem-solving [37]. In the context of technology development, TA is equally used to gain deep insights about thinking while using a technology [38]. The advantage of the method is to capture problems and solutions as the technology is being used, as retrospective surveys can lead to incomplete information about the problems of a technology. This means TA is helpful in tracing user thinking strategies [39,40]. For this purpose, participants were instructed to think aloud constantly while using the demonstrator. If participants stopped thinking aloud, they are reminded by the experimenter to continue speaking aloud [38,39]. For understanding problems and solutions, we did not explain how the demonstrator is supposed to work. Instead, we asked them to experience the app with little direction by explaining the task they had to do: registration, set delivery location, and submit recipe [41]. Thus, during the TA situation, it was important that the experimenter interact with participants as little as possible to prevent interference with the users' thoughts [39]. During the TA situations, the statements of the participants were digitally recorded (audio recordings) and transcribed afterward. Moreover, experimenters prepared protocols for making notes and describing events that are not verbally made by participants but important for analyzing. For example, if a participant said, "That is confusing," the experimenter protocoled what exactly was confusing [42].

Data Analysis

Quantitative Analyses

Analyses were performed with the statistical program SPSS Statistics (version 25; IBM Corp). Bivariate correlation analyses were performed to assess the association between different user acceptance measurements of usability (SUS score), intention to use a technology (TUI factors using raw values), and psychological needs (the TENS Task and Interface as well as the BPNSFS using raw values). To investigate which factors are associated with the intention to use the pharmacy drone app, hierarchical regression analyses were performed among TUI factors as well as among all measurements. To answer the question whether TUI and SUS factors differ between the core and ad hoc groups, 2-tailed *t* tests were performed. Owing to the nature of the study design (small sample sizes), statistical analyses should not be overinterpreted; thus, the results of questionnaires were also analyzed descriptively according to the improvement of acceptance level of SUS and TUI scores.

Qualitative Analyses

The transcripts were analyzed according to the event sampling method of Berelson [43], where an utterance represents an event. Utterances are defined as a complete sentence, a sentence fragment, or any sequence of speech separated in time (eg, a pause of 2 seconds) or semantic (eg, a change in content) [39,44]. Utterances were analyzed by referring phrase analysis [39]. First, all nouns and noun phrases were identified, and utterances with the reference concept name were coded by the first author (FF). This coding shows which concepts the participants focused on during the task. After concepts have been identified, these concepts were defined by the investigator (Tables 2 and 3). The resulting coding scheme was used to code the statements of the participants. Another researcher (JS) who was familiar with the analyses analyzed randomly selected portions of transcripts (20%) to determine if there was a match. In case of disagreement, discussions were held between the 2 examiners until a consensus was reached. Cohen κ [45] was computed for all variables. The interrater reliability was κ=0.654 (*P*<.001) with a substantial agreement [46].

Table 2. Examples of the coded concepts.

Coded concept	Segment
Value and problem	"What is stupid now is this field. It does not disappear." (Doctor, aged 48 years)
Proposal	"Would be nicer: 'Your location and address has been confirmed.'" (Nurse, aged 51 years)
Value and conceptuality	"That black sign that irritates." (Patient, aged 55 years)
Status	"Now I have uploaded this successfully." (Pharmacist, aged 35 years)
Ambiguity	"Do I have to register again now?" (Nurse, aged 49 years)

Table 3. Definitions of coded concepts.

Concept	Definition
Value	Rating of usefulness, importance, or worth
Status	Information indicative of status or self-instruction
Problem	Technical inconvenience requiring action
Ambiguity	Incomprehensibility of the process, operation, or handling
Conceptuality	System of terms or concepts
Proposal	Recommendation for technical implementation

The experimenters who prepared the protocols during the iteration loops were asked to evaluate the accuracy of the definitions of coded concepts to ensure that no undefined concepts remained. After all utterances were coded, concepts were summarized for groups. The results were arranged in a table per task and discussed with the ADApp team to derive practice-relevant orientation points. To rank the participant's points, we classified the concepts according to criteria within the ADApp team. The criteria helped us to evaluate the relevance of the concepts for developing the technology. We defined 4 criteria: safety, risk in the delivery, optimization potential, and outside the capabilities (Multimedia Appendix 1).

Results

Participants

For the 3 iteration loops, between March 2021 and October 2022, we collected data from 18 participants (mean age 43.08, SD 12.44; range 25-65 years). Owing to the relatively large amount of time required, not all participants of the core group could always participate in the iteration loops. In total, 6 participants took part in the first core group (1 general practitioner, 2 pharmacists, 1 patient, and 2 nurses); 3 participants took part in the second core group (1 nurse, 1 pharmacist, and 1 patient); and 4 participants took part in the third core group (2 nurses, 1 pharmacist, and 1 patient). Although, we tried to balance the 2 groups, it was not always possible to equalize the core and ad hoc group. Four participants took part in the first ad hoc group (1 general practitioner, 1 nurse, and 2 relatives of patients who need palliative care); 4 participants took part in the second ad hoc group (3 patients and 1 nurse); and 4 participants took part in the third ad hoc group (1 nurse and 3 patients). Table 4 provides detailed demographics.

Table 4. Participants' demographics.

Characteristics	Ad hoc group 1 (n[a]=4)	Ad hoc group 2 (n=4)	Ad hoc group 3 (n=4)	Core group 1 (n=6)	Core group 2 (n=3)	Core group 3 (n=4)	All (n=18)
Age (years), mean (SD; range)	50.25 (3.59; 47-55)	47.25 (4.42; 43-52)	50.75 (16.19; 31-64)	39.17 (15.45; 25-65)	37.00 (11.53; 26-49)	34.5 (10.97; 26-49)	46.00 (12.19; 25-65)
Gender (female)							
Values, mean (SD)	1.00 (0.00)	1.25 (0.50)	1.25 (0.50)	1.5 (0.55)	1.67 (0.57)	1.50 (0.58)	1.28 (0.46)
Values, n (%)	4 (100)	3 (75)	3 (75)	3 (50)	1 (33)	2 (50)	13 (72)
COVID-19							
Values, mean (SD)	1.5 (0.53)	1.25 (0.50)	1.50 (0.58)	1.5 (0.55)	1.00 (0.00)	1.25 (0.50)	1.44 (0.51)
Values, n (%)	2 (50)	3 (75)	2 (50)	3 (50)	3 (100)	3 (75)	10 (56)
Drone competence							
Values, mean (SD)	1.75 (0.50)	1.75 (0.50)	2.00 (0.00)	2.00 (0.00)	2.00 (0.00)	1.75 (0.50)	1.89 (0.32)
Values, n (%)	1 (25)	1 (25)	0 (0)	0 (0)	0 (0)	1 (25)	2 (11)
Medication app competence							
Values, mean (SD)	2.00 (0.00)	2.00 (0.00)	2.00 (0.00)	1.5 (0.55)	1.67 (0.57)	1.5 (0.58)	1.83 (0.38)
Values, n (%)	0 (0)	0 (0)	0 (0)	3 (50)	1 (33)	2 (50)	3 (17)

[a]n has been adjusted by weighting.

Quantitative Results

Bivariate Correlations

Bivariate correlations showed that the more usable (SUS) the technology was, the more the participants were willing to use it (TUI; *r*=0.833). Moreover, the more they found the technology useful (TUI), the more the technology was rated as usable (*r*=0.711; SUS and *r*=0.487; TUI) and the more participants would use the pharmacy drone app (TUI; *r*=0.754). In addition, the intention to use (TUI) the app was positively correlated with curiosity (TUI; *r*=0.550). The more skeptical (TUI) the participants were, the more participants rated the app as unusable (*r*=−0.542; SUS and *r*=−0.479; TUI) and would not be willing to use it (*r*=−0.446; TUI).

Furthermore, the intention to use (TUI) the technology was positively correlated with the perceived task competence (*r*=0.784; first iteration loop) and the need satisfaction of participant's competence (*r*=0.788; third iteration loop). The more participants felt competent in handling the app in the spheres of task (*r*=0.829; ie, engaging with the app; first iteration loop) and interface (*r*=0.929; ie, interacting with the app; second iteration loop), the higher was perceived usability (SUS). However, the more participants felt competent in task (*r*=0.675) and behavior (*r*=0.875), the more they felt autonomous (ie, the app provided options and participants did not feel under pressure from the app). The more curious the participants were, the more they felt satisfied in psychological needs of competence, autonomy, and relatedness (*r*=0.744). However, the absence of need satisfaction does not equate with the presence of need frustration [17,35]. The results revealed that the more participants rated the app as usable, the lesser they felt autonomy (*r*=−0.792; SUS and *r*=−0.768; TUI) and competence (*r*=−0.751; TUI) frustration. The more they believed the technology would be useful for them, the lesser they felt autonomy frustration (*r*=−0.827). Overall frustration shows a negative correlation with usability (*r*=−0.822; SUS and *r*=−0.799; TUI), usefulness (*r*=−0.923; TUI), and intention to use (*r*=−0.730; TUI). Moreover, the results indicated that autonomy frustration is a relevant marker for overall need frustration (*r*=0.933).

Furthermore, correlations showed that the more time participants needed to solve the task within the app, the less usable the app was rated (*r*=−0.534; SUS), the lesser they were willing to use the app (*r*=−0.429; TUI), the lesser they felt competent (*r*=−0.805), the more skepticism they had (*r*=0.504), and the older the participants were (*r*=0.681). The older the participants were, the more skeptical they were (*r*=.525) and the lesser they believed that the technology was accessible (*r*=−0.510). Female participants showed more technology anxiety (*r*=−0.505), had more interest (*r*=0.497), and felt more related by using the app (*r*=−0.800). The results of 2-tailed *t* tests confirmed these differences between female and male participants (anxiety: t_{19}=0.003; interest: $t_{23=}$0.012; relatedness satisfaction: t_6=0.004).

TUI Assessment

To test whether anxiety, curiosity, interest, skepticism, usefulness, usability, and accessibility contributed to higher intention to use, we regressed participant's ratings of these variables on their intention to use and controlled for age, gender, and duration using the app. Results showed that TUI factors such as usefulness, skepticism, and curiosity explain most of the variance in intention to use the pharmacy drone app

($F_{3,17}$=21.12; P<.001; R^2=0.788; adjusted R^2=0.751; Table 5). Usefulness (β=.499; P=.001) and curiosity (β=.376; P=.008) were significantly and positively associated with intention to use, whereas skepticism (β=−.397; P=.004) was significantly and negatively related with intention to use.

Table 5. Hierarchical regression analysis predicting intention to use the technology per the Technology Usage Inventory factors.

Model and predictor	Unstandardized coefficients, B (SE)	Standardized coefficients		R^2	R^2 change	*F* test (*df*)	*P* value
		β	*P* value				
1				0.582	0.560	26.48 (1,19)	<.001
Usefulness	14.73 (2.86)	.763	<.001				
2				0.676	0.640	18.81 (2,18)	<.001
Usefulness	13.03 (2.69)	.675	<.001				
Skepticism	−9.82 (4.29)	−.319	.03				
3				0.788	0.751	21.12 (3,17)	<.001
Usefulness	9.638 (2.51)	.499	.001				
Skepticism	−12.20 (3.66)	−.397	.004				
Curiosity	6.36 (2.12)	.376	.008				

Descriptively, anxiety, curiosity, interest, usefulness, and accessibility did not vary much between the core group and ad hoc groups, as shown in Table 6. Ad hoc groups appeared to be slightly more skeptical about the pharmacy drone app compared to the core group. This could potentially be attributed to age differences, as those in the ad hoc groups were older than those in the core group, and skepticism was positively correlated with age (Multimedia Appendix 2). Although usability remained at slightly below average in the ad hoc groups, it increased from average to slightly above average in the core group from the first to third iteration. Moreover, the core group showed higher ratings on the intention to use the pharmacy drone app than the ad hoc groups.

Table 6. Mean values of the Technology Usage Inventory stanine of the ad hoc and core groups per iteration loop.

	Ad hoc group 1	Ad hoc group 2	Ad hoc group 3	Core group 1	Core group 2	Core group 3
Psychological factors, mean (SD)[a]						
Anxiety	4.75 (1.89)	4.67 (3.05)	5.25 (1.50)	3.50 (1.22)	—[b]	3.00 (1.16)
Curiosity	7.00 (1.41)	7.67 (0.58)	5.25 (2.36)	7.33 (1.03)	—	7.00 (0.00)
Interest	5.50 (1.91)	5.25 (1.89)	7.25 (1.50)	6.00 (1.41)	5.67 (1.53)	6.00 (0.82)
Skepticism	3.75 (1.89)	2.75 (0.96)	3.50 (1.00)	2.83 (0.98)	1.67 (0.58)	2.00 (0.82)
Technology-specific factors, mean (SD)[a]						
Usefulness	8.00 (1.41)	8.75 (0.50)	8.00 (2.00)	8.67 (0.82)	9.00 (0.00)	8.75 (0.50)
Usability	4.50 (1.91)	4.25 (1.50)	4.50 (0.58)	5.50 (1.05)	6.67 (0.58)	7.00 (0.82)
Accessibility	7.75 (1.26)	4.50 (1.00)	6.25 (1.26)	7.33 (1.03)	8.67 (0.58)	7.25 (0.96)
Intention to use, mean (SD)[a]	6.75 (1.89)	8.50 (0.58)	5.75 (0.96)	8.50 (0.55)	8.00 (1.00)	8.75 (0.50)

[a]Stanine: 1 to 2, strongly below average; 3 to 4, slightly below average; 5, average; 6 to 7, slightly above average; and 8 to 9, strongly above average [33].

[b]Missing data.

The results of *t* tests showed a higher rating on usability for core group 3 compared with core group 1 (t_8=−2.68; P=.03). Moreover, independent samples *t* tests showed a higher anxiety (t_{19}=2.88; P=.01) as well as a higher skepticism toward the technology (t_{23}=2.17; P=.04) within the ad hoc group compared with the core group over all iteration loops. Moreover, the core group rated the app more usable than the ad hoc group (t_{23}=−3.33; P=.003) over all iteration loops. Although descriptive data showed a higher rating on intention to use the pharmacy drone app within the core group (mean 90.36, SD 11.01) than the ad hoc groups (mean 76.00, SD 23.74) over all iteration loops, 2-tailed *t* tests became not significant (t_3=−1.91; P=.07).

SUS Score

Descriptively, perceived usability (SUS score) decreased between ad hoc group 1 (rated as marginal) and ad hoc group 3 (rated as poor). Within the core group, the usability increased from core group 1 (rated as good) to core group 3 (rated as excellent). However, within the second iteration loop, both groups (ad hoc and core groups) rated the app more usable than during the first and third iteration loops. Moreover, the core group showed higher SUS scores than the ad hoc groups (Table 7). Results of the *t* tests indicated a significant group difference in SUS score between ad hoc group 2 and ad hoc group 3 (t_6=3.35; *P*=.01). The results indicated lower SUS scores of ad hoc group 3 (rated as poor) than ad hoc group 2 (rated as excellent). Furthermore, independent samples *t* tests showed a significant group difference in the SUS score between the ad hoc and core groups over all iteration loops, with higher scores for the core group than the ad hoc groups (t_{23}=−2.87; *P*=.004).

Table 7. System Usability Score (SUS) scores (0 to 34: very poor; 35 to 49: poor; 50 to 67: marginal; 68 to 79: good; 80 to 89: excellent; and 90 to 100: outstanding [32]) in the ad hoc and core groups per iteration loop.

Group	SUS scores, mean (SD)
Ad hoc group 1	54.38 (23.22)
Ad hoc group 2	80.00 (14.43)
Ad hoc group 3	49.38 (11.25)
Core group 1	75.42 (16.84)
Core group 2	88.33 (10.10)
Core group 3	84.38 (5.54)
Ad hoc group 1 + core group 1	67.00 (21.34)
Ad hoc group 2 + core group 2	83.57 (11.57)
Ad hoc group 1 + core group 3	66.88 (20.43)
All groups	71.60 (19.75)

Overall Regressions

When regressing all factors of all usability and psychological needs, the results showed that the SUS usability score as well as the TUI factors such as curiosity and interest explained most of the variance on intention to use the pharmacy drone app ($F_{3,16}$=40.27; *P*<.001; R^2=.883; adjusted R^2=.861; Table 8). Usability (β=.845; *P*<.001), curiosity (β=.232; *P*=.02), and interest (β=.195; *P*=.04) were significantly and positively associated with intention to use.

Table 8. Hierarchical regression analysis predicting intention to use the technology per the Technology Usage Inventory, System Usability Scale (SUS), Technology-based Experience of Need Satisfaction Task, and Basic Psychological Need Satisfaction and Frustration Scale factors.

Model and predictor	Unstandardized coefficients, B (SE)	Standardized coefficients		R^2	R^2 change	*F* test (*df*)	*P* value
		β	*P* value				
1				0.789	0.777	67.16 (1,18)	<.001
SUS usability	0.999 (0.122)	.888	<.001				
2				0.848	0.830	47.31 (2,17)	<.001
SUS usability	0.890 (0.115)	.791	<.001				
Curiosity	4.45 (1.73)	.262	.02				
3				0.883	0.861	40.27 (3,16)	<.001
SUS usability	0.951 (0.107)	.845	<.001				
Curiosity	3.94 (1.58)	.232	.02				
Interest	2.88 (1.31)	.195	.04				

Qualitative Results

First Iteration

After the first iteration, the app received a new design according to user's feedback. Important points after the first iteration were providing more guidance through the app with information about next steps and reasons for doing these steps (eg, the importance of setting the delivery location, clear information about how to choose the delivery location, and more precise symbols; Figures 4 and 5); adaption of conceptualizations (eg, "location" [German: *Standort*] to "delivery point" [German: *Lieferort*]; Figure 4); automatizations (eg, transferring address data automatically to the map); minimizations (eg, reducing symbols within the map and information within each step); communication options (eg, integrating a field to formulate a message to the pharmacist); control features (eg, an order summary); and autonomy options (eg, to upload >1 prescription if necessary). However, participants missed a visualization of password requirements and a preview function of the uploaded recipe. Furthermore, users would rather preview individual pages per click than perform 1-page scrolling. They would also desire information about payment options within the app as well as details about medication availability. This necessitates integration of the interface with the pharmacy's merchandise management system, which requires additional technical and regulatory administration. A short solution for that was to integrate a comment field at the step of ordering the medication to describe further medication wishes as well as to ask questions to the pharmacist. Moreover, participants emphasized, at this point, the importance of shipment tracking, as Fink et al [27] described in their study. However, the most difficult step participants reported was setting of the delivery location. This was also shown in the amount of support needed while using the app (Multimedia Appendix 3). Although experimenters were instructed to not help participants, at some points the help was necessary so that the participants could finish the task.

Figure 4. User-centered app design (A) before and (B) after first iteration: delivery location.

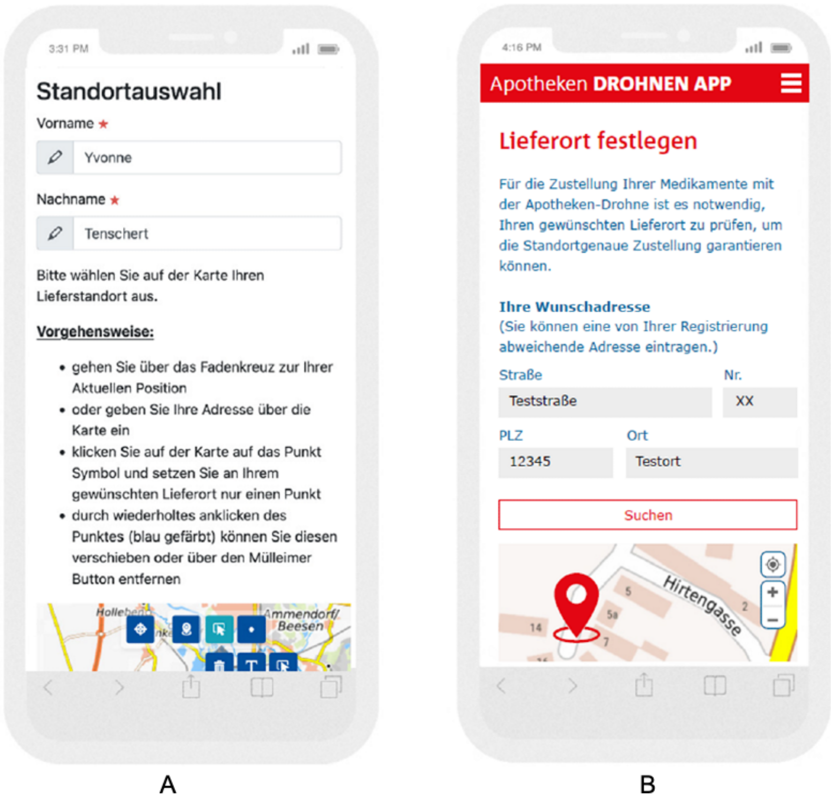
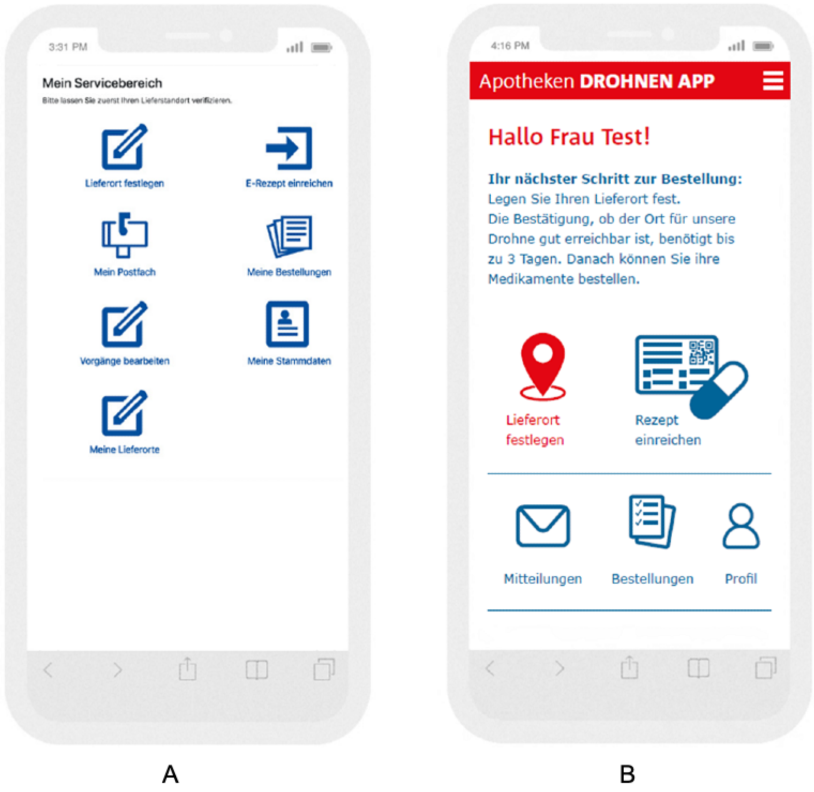


Figure 5. User-centered app design (A) before and (B) after first iteration: symbols.



Second Iteration

After the second iteration, participants in both the core group and the ad hoc group reported that registration was simplified, indicating that it was easy for them to register. Moreover, participants reported that the summary of order was clear, the texts were more comprehensible (core group), and the participants liked the option to upload >1 prescription (ad hoc group). However, participants missed a preview function of the uploaded prescription as a picture and the selection of push messages. Moreover, they suggested highlighting the icon, which is the next step to be clicked on (eg, after setting the delivery location, the icon for submitting the prescription should be highlighted), and highlighting the inbox when a new message has arrived. Important points after the second iteration were still reducing symbols within the map (eg, symbols were reduced to a minimum, and the text describing the symbols was shortened because the participants did not read the instructions above the card); more guidance through the app with information about next steps; and more transparency about how a response will be received from pharmacists (eg, integration of information about how the contact will take place, via mail or telephone), indications on password requirements (such as length, upper and lower case, and special characters), and information about shipment tracking. An important safety-relevant point was that flight slots were not up to date according to the original time. Similar to the first iteration loop, participants reported setting the delivery location as the most difficult step. They would like to have more guidance for the subsequent steps after determination of the delivery location.

Third Iteration

The core group reported that the app was more intuitive and had improved compared with the first iteration loop. They mentioned that the automated fill-in of address data in the map as well as the identification that the prescription was successfully uploaded was useful.

Important points after the third iteration were more guidance through a processing status within the app; adaption of conceptualization (eg, “mailbox” [German: *Postfach*] to “messages” [German: *Nachrichten*]); visualization of password requirements (eg, upper and lower case and special characters); automatization features (eg, automatic suggestions such as city when entering the postal code); differentiation (eg, distinction between the delivery and billing address); control (eg, adjustment of the amount of information within the app to control how much information the user wants, which might be configured via profile); push messages instead of emails because participants did not read the texts despite shortening them; a preview function of the uploaded prescription; and design features (eg, it was not clear that scrolling was necessary, thus more guidance is useful through individual pages [click to continue]). They also desired the inbox to be highlighted when new messages had arrived.

Although the app was adapted according to the participant’s feedback, the ad hoc group reported that the step of setting the delivery location remained too complex for them. They mentioned that this step was too bulky, time consuming, and not intuitive despite adjustments such as the reduction of map symbols, providing the most important information, shortening the text, and inserting address data automatically in the map. Participants of the ad hoc group felt lost and helpless during this step. However, during discussions after thinking aloud, participants suggested that the setting of delivery comes from the operator and thus must not be made by users. They only wanted confirmation of the delivery point to ensure its correctness. Excluding this step might decrease the likelihood of user errors.

However, across all iterations, criteria emerged that were repeatedly the focus of the participants' attention: automatization (ie, easy and fast use for avoiding redundancies), minimization (ie, as little information as possible and as much as necessary), differentiation (ie, clear distinctions), control (ie, options to choose), guiding (ie, concise and understandable instructions supporting guidance through the app), conceptualization (ie, easy and precise language), barrier-free design (ie, uniformity between different steps and intuitive visualizations), and transparency (about disclosures to be made or obtained information).

Although we could not test the handover scenarios, we have made modifications to the drone. Previous results of focus group testing within the ADApp project showed concerns about injuries caused by the drone [27]. Thus, the drone has now been given a flap underneath so that the medications can be dropped by ejection, using a parachute, or using a winch and a landing of the drone can be prevented (Figure 6). Further testing is planned to test different handover scenarios with participants to adapt the handover according to their needs.

Figure 6. Drone medication ejection.

Discussion

Factors of User Acceptance

This study aimed at investigating factors that are associated with the user acceptance of a drone-based medication delivery by using a mixed methods design to be able to derive practice-relevant orientation points for participatory technology development (for apps and drones).

First, an important point is duration handling with the app. Older participants needed more time to solve the tasks within the app. Furthermore, the longer the process took, the more the usability, intention to use the app, and feelings of competence decreased, while skepticism increased. Therefore, the duration of interaction with a technology appears to be a crucial factor for user acceptance.

Second, psychological factors such as skepticism and curiosity as well as technology-specific factors such as usefulness and usability are related with participants' intention to use the technology for a drone-based medication delivery. Regression analyses within the TUI factors revealed usefulness, curiosity, and skepticism as significant predictors for intention to use the technology, wherein usefulness explained the highest variance (49.9%), which is consistent with the findings of Güsken et al [47]. This implies a particular relevance of factor usefulness for the development of technologies in health care, especially in drone-based medication delivery. With the help of the TUI questionnaire, we can conclude that curiosity and skepticism affect user acceptance. The more the users were curious about the pharmacy drone app and the less skeptical they were, the more the users were willing to use and interact with the technology. This is consistent with the findings of Eißfeld et al [6] who found that a positive attitude toward drones and a general technical interest are related to improved information about it.

Third, basic human needs according to Ryan and Deci [16,17] and Deci and Ryan [18]also play an important role. Results of this study showed that the more participants perceived competence in handling the app, the more they are willing to use the technology and the more they rated the app as usable. This implies that, although competence satisfaction in all iteration loops was related with usability and the intention to use, autonomy and relatedness were not related. Nevertheless, results showed that the more participants felt competent, the more they felt autonomous. Moreover, the basic psychological needs (competence, autonomy, and relatedness) were positively correlated with curiosity. In addition, the lesser participants felt frustrated on psychological needs, the higher they rated the usability, usefulness, and intention to use. Thus, following the Basic Psychological Need Theory [17], the more the interaction with the system satisfies basic psychological needs, the more the users will engage with a technology. Following the METUX model, an increase in autonomy increases engagement and an increase in competence increases motivation of using the app, which is in accordance with the results of this study [15]. Interestingly, while competence and autonomy appear to be significant factors in explaining differences in intention to use and usability, relatedness does not play a role, despite focus group discussions emphasizing the importance of communication and consultation features within such technology [27]. One reason for this result might be the nature of the study task. Although participants used the app in a simulated context, communication aspects did not play a role in this developing step of technology. Therefore, supply studies in real-life contexts are necessary to test the impact of psychological and technological factors on real-life complex problems, which cannot be fully investigated in a simulated context such as in this study, where, for example, relatedness might hold greater significance in real-life scenarios than in simulations, as increased relatedness could potentially enhance overall well-being [15].

Fourth, overall regression analyses showed that usability, curiosity, and interest explain most of the variance on intention to use the pharmacy drone app, wherein usability showed the strongest effect (84.5%). This means, when adding all factors in one model, usability becomes the strongest predictor for intention to use the pharmacy drone app. Similar to other studies, this study found evidence for the importance of usability in using a technology [1,5]. This means that better usability of a technology leads to higher acceptance [12,33]. However, this study used a very small sample group, and the statistical results should be considered carefully. In conclusion, studies with higher sample sizes are necessary.

Taken together, the intention to use a drone-based medication delivery system is a comprehensive construct that is based on a large number of underlying, explanatory factors. This study showed that usability, curiosity, and interest had a considerable impact on intention to use, wherein usefulness, skepticism, and competence also played an important role. The failure to actively involve users in technology development can thus result in insufficient addressing of profession- and person-specific needs, thus resulting in a lack of intention to use the technology. For successful technology development, it is therefore crucial to develop an understanding of the necessary characteristics of health care technologies and to identify the determinants that ensure a high level of acceptance for improving the current supply situation [47].

In accordance with the previous scoping review [7], user feedback was collected iteratively and focused on user experience. The TA method [38] used in this process provided valuable insights that were taken into account when developing the app. In this way, important changes to the app were successfully implemented by user request such as a reduced design and automatic fill-in aids. It was found that communication with the dispatcher and shipment tracking are very important to users, which is consistent with the assumptions from the scoping review [7], and this also led to further adjustments. The changes made could be verified in the further iterations; for example, it turned out that the revision for the definition of the delivery location was not helpful: the process was adjusted based on participant feedback with more information, but this step remained too complex. In the third iteration, it became clear that the required texts were not being read at this point, leading to the ultimate decision to omit this step altogether as participants indicated that they only wanted to confirm the delivery location.

Differences in User Acceptance

A second purpose of this study was to assess group differences between a core group and ad hoc groups in user acceptance.

The results of this study indicated the importance of an ad hoc group in an iterative, cocreative process. Although within the core group, intention to use (TUI) was similar over all iteration loops (strongly above average), within the ad hoc group, intention to use varied from slightly above average to strongly above average to average. Although within the core group, the usability of the app slightly increased from “average resp. good” to “slightly above average resp. excellent,” the usability within the ad hoc group decreased from marginal to poor (SUS) and remained slightly above average (TUI). Results of *t* tests of both questionnaires showed a significant group difference between the ad hoc and core groups, with higher ratings in perceived usability for the core group. This suggests that repeated measurements induce a shift in the interpretation of test scores, potentially biasing the measurement of change [26]. Thus, this study shows that repeated measures with the same sample group might change their attitude, expectations, and behavior in dealing with the technology, which changed their ratings on usability. The core group then tended to evaluate the app better than the ad hoc group because they were not unfamiliar with dealing with the app. Thus, for naive users, the app is just not intuitive and easy enough to use. An additional explanation for this result is that the ad hoc groups were more anxious and skeptical than the core group, wherein a higher skepticism was found to be related to lower ratings on usability [47]. However, within the usability score (SUS), data showed an increase from the first to second iteration and a decrease from the second to third iteration in both groups, wherein the ad hoc groups rated the pharmacy drone app as less usable compared with the core group. One reason for the decrease from the second to third iteration could be the more complex setting during the third iteration: the participants had to run through the entire process from registration, setting the delivery location, and ordering the medication to receiving the medication. Meanwhile, the core group showed a learning effect and maybe thus rated the pharmacy drone process to be more usable from the first iteration to second iteration, and the ad hoc groups could stumble because of the complexity.

Limitations

This study shows for the first time the importance of ad hoc groups as a control group while developing and evaluating a technology in a user-centered design. When interpreting the results of this study, several methodological limitations must be considered. In terms of age and gender distribution, the sample can be classified as unrepresentative owing to the small number of participants. This is particularly evident in the statistical evaluation. Nevertheless, there is a basic tendency toward a clear effect, which is evident despite the small number of samples. The participants had a basic interest in new topics and in the topic itself. Although the risk of “positive selection” cannot be completely ruled out, it is not seen because the topic of drone-assisted medication delivery was largely unknown. Thus, the perspectives of participants who consistently reject technical systems in the context of care and delivery were as poorly represented as those who chose not to participate in the surveys for other reasons. Reasons for this could include a general dismissive attitude toward additional effort owing to time resources, heavy workloads, or other thematic priorities.

However, it can be concluded that the results obtained to assess the acceptance of the drone app for utility purposes have revealed important insights regarding technical development and its practical use. In this context, the findings exhibit similarities to surveys conducted for other target groups in health care.

Conclusions

The study highlights the significance of understanding the essential attributes of health care technologies and the factors that lead to their acceptance in improving the current supply situation. It offers valuable insights for practitioners to develop participatory technologies and recommends ad hoc groups as a complementary approach to control the process of a user-centered design. However, larger samples and real-world contexts are required to confirm these findings.

Acknowledgments

The authors would like to thank the Apotheken-Drohnen-App team for their involvement in the participatory development of the drone. In addition, the authors would like to thank Dr Anne-Marie Lachmund for her help with proofreading.

Authors' Contributions

FF contributed to conceptualization, formal analysis, investigation, methodology, visualization, and writing of the original draft. IK contributed to visualization and writing of the original draft. JVS helped with the formal analysis. PJ contributed to conceptualization, funding acquisition, supervision, and writing of the original draft (review and editing). HKH helped with the formal analysis and visualization. DP contributed with writing of the original draft (review and editing).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Examples of criteria.
[DOCX File , 13 KB - humanfactors_v11i1e51587_app1.docx]

Multimedia Appendix 2
Bivariate correlations between System Usability Scale (SUS), Technology Usage Inventory (TUI), Technology-based Experience of Need Satisfaction (TENS) Task, TENS Interface, and Basic Psychological Need Satisfaction and Frustration Scale (BPNSFS).
[DOCX File , 46 KB - humanfactors_v11i1e51587_app2.docx]

Multimedia Appendix 3
Support required per task.
[DOCX File , 13 KB - humanfactors_v11i1e51587_app3.docx]

References

1. Hiebert B, Nouvet E, Jeyabalan V, Donelle L. The application of drones in healthcare and health-related services in North America: a scoping review. Drones 2020 Jul 04;4(3):30. [doi: 10.3390/drones4030030]
2. Saponi M, Borboni A, Adamini R, Faglia R, Amici C. Embedded payload solutions in UAVs for medium and small package delivery. Machines 2022 Aug 27;10(9):737. [doi: 10.3390/machines10090737]
3. Sabino H, Almeida RV, Moraes LB, Silva WP, Guerra R, Malcher C, et al. A systematic literature review on the main factors for public acceptance of drones. Technol Soc 2022 Nov;71:102097. [doi: 10.1016/j.techsoc.2022.102097]
4. Rave A, Fontaine P, Kuhn H. Drone network design for emergency resupply of pharmacies and ambulances. SSRN. Preprint posted online September 14, 2023 [FREE Full text] [doi: 10.2139/ssrn.4569199]
5. Rejeb A, Rejeb K, Simske S, Treiblmaier H. Humanitarian drones: a review and research agenda. Internet Things 2021 Dec;16:100434. [doi: 10.1016/j.iot.2021.100434]
6. Eißfeldt H, Vogelpohl V, Stolz M, Papenfuß A, Biella M, Belz J, et al. The acceptance of civil drones in Germany. CEAS Aeronaut J 2020 Apr 04;11(3):665-676. [doi: 10.1007/s13272-020-00447-w]
7. Stephan F, Reinsperger N, Grünthal M, Paulicke D, Jahn P. Human drone interaction in delivery of medical supplies: a scoping review of experimental studies. PLoS One 2022;17(4):e0267664 [FREE Full text] [doi: 10.1371/journal.pone.0267664] [Medline: 35482656]

8. Sham R, Siau CS, Tan S, Kiu DC, Sabhi H, Thew HZ, et al. Drone usage for medicine and vaccine delivery during the COVID-19 pandemic: attitude of health care workers in rural medical centres. Drones 2022 Apr 27;6(5):109 [FREE Full text] [doi: 10.3390/drones6050109]
9. Milchrahm E. Entwicklung eines Modells zur Akzeptanzproblematik von Informationstechnologie. In: Proceedings of the Information und Mobilität, Optimierung und Vermeidung von Mobilität durch Information. 2002 Presented at: ISI 2002; October 8-10, 2002; Regensburg, Germany.
10. Krey M, Seiler R. Usage and acceptance of drone technology in healthcare: exploring patients and physicians perspective. In: Proceedings of the 52nd Hawaii International Conference on System Sciences. 2019 Presented at: HICSS 2020; January 8-11, 2019; Grand Wailea, HI URL: https://digitalcollection.zhaw.ch/handle/11475/15635
11. Simon B. E-Learning an Hochschulen. Gestaltungsräume und Erfolgsfaktoren von Wissensmedien. Lohmar, Germany: Josef Eul Verlag; 2001.
12. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
13. Venkatesh V, Bala H. Technology Acceptance Model 3 and a research agenda on interventions. Decis Sci 2008 May 09;39(2):273-315. [doi: 10.1111/j.1540-5915.2008.00192.x]
14. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manag Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
15. Peters D, Calvo RA, Ryan RM. Designing for motivation, engagement and wellbeing in digital experience. Front Psychol 2018;9:797 [FREE Full text] [doi: 10.3389/fpsyg.2018.00797] [Medline: 29892246]
16. Ryan RM, Deci EL. Intrinsic and extrinsic motivations: classic definitions and new directions. Contemp Educ Psychol 2000 Jan;25(1):54-67. [doi: 10.1006/ceps.1999.1020] [Medline: 10620381]
17. Ryan RM, Deci EL. Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness. New York, NY: Guilford Press; 2017.
18. Deci EL, Ryan RM. Intrinsic Motivation and Self-Determination in Human Behavior. New York, NY: Springer; 1985.
19. Jockisch M. Das ist gar kein Modell!. In: Bandow G, Holzmüller H, editors. Das Technologieakzeptanzmodell. Wiesbaden, Germany: Springer Gabler; 2010:233-254.
20. Farao J, Malila B, Conrad N, Mutsvangwa T, Rangaka MX, Douglas TS. A user-centred design framework for mHealth. PLoS One 2020;15(8):e0237910 [FREE Full text] [doi: 10.1371/journal.pone.0237910] [Medline: 32813711]
21. Risling TL, Risling DE. Advancing nursing participation in user-centred design. J Res Nurs 2020 May 03;25(3):226-238 [FREE Full text] [doi: 10.1177/1744987120913590] [Medline: 34394630]
22. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
23. Nilsen W, Kumar S, Shar A, Varoquiers C, Wiley T, Riley WT, et al. Advancing the science of mHealth. J Health Commun 2012;17 Suppl 1:5-10. [doi: 10.1080/10810730.2012.677394] [Medline: 22548593]
24. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;Suppl:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
25. Altman M, Huang TT, Breland JY. Design thinking in health care. Prev Chronic Dis 2018 Sep 27;15:E117 [FREE Full text] [doi: 10.5888/pcd15.180128] [Medline: 30264690]
26. Oort FJ, Visser MR, Sprangers MA. Formal definitions of measurement bias and explanation bias clarify measurement and conceptual perspectives on response shift. J Clin Epidemiol 2009 Nov;62(11):1126-1137. [doi: 10.1016/j.jclinepi.2009.03.013] [Medline: 19540722]
27. Fink F, Paulicke D, Grünthal M, Jahn P. "Of course, drones delivering urgent medicines are necessary. But I would not use them until…" Insights from a qualitative study on users' needs and requirements regarding the use of medical drones. PLoS One 2023;18(5):e0285393 [FREE Full text] [doi: 10.1371/journal.pone.0285393] [Medline: 37155650]
28. Brooke J. SUS: a 'quick and dirty' usability scale. In: Usability Evaluation In Industry. Boca Raton, FL: CRC Press; 1996.
29. Brooke J. SUS: a retrospective. J Usab Stud 2013 Feb;8(2):29-40.
30. Hyzy M, Bond R, Mulvenna M, Bai L, Dix A, Leigh S, et al. System usability scale benchmarking for digital health apps: meta-analysis. JMIR Mhealth Uhealth 2022 Aug 18;10(8):e37290 [FREE Full text] [doi: 10.2196/37290] [Medline: 35980732]
31. Sauro J. A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices. Denver, CO: Measuring Usability LLC; 2011.
32. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usab Stud 2009 May 1;4(3):114-123.
33. Kothgassner OD, Felnhofer A, Hauk N, Kastenhofer E, Gomm J, Ryspin-Exner I. TUI: Technology Usage Inventory. Vienna, Austria: ICARUS; 2012.
34. Febretti A, Garzotto F. Usability, playability, and long-term engagement in computer games. In: Proceedings of the CHI '09 Extended Abstracts on Human Factors in Computing Systems. 2009 Presented at: CHI EA '09; April 4-9, 2009; Boston, MA. [doi: 10.1145/1520340.1520618]

35. Chen B, Vansteenkiste M, Beyers W, Boone L, Deci EL, Van der Kaap-Deeder J, et al. Basic psychological need satisfaction, need frustration, and need strength across four cultures. Motiv Emot 2015;39(2):216-236. [doi: 10.1007/s11031-014-9450-1]
36. Heissel A, Pietrek A, Flunger B, Fydrich T, Rapp MA, Heinzel S, et al. The validation of the German basic psychological need satisfaction and frustration scale in the context of mental health. Eur J Health Psychol 2018 Oct;25(4):119-132. [doi: 10.1027/2512-8442/A000017]
37. Gunzenhauser C, Karbach J, Saalbach H. Function of verbal strategies in monolingual vs. bilingual students' planning performance: an experimental approach. Cognit Dev 2019 Apr;50:1-12. [doi: 10.1016/j.cogdev.2019.01.003]
38. Jaspers MW, Steen T, van den Bos C, Geenen M. The think aloud method: a guide to user interface design. Int J Med Inform 2004 Nov;73(11-12):781-795. [doi: 10.1016/j.ijmedinf.2004.08.003] [Medline: 15491929]
39. Fonteyn ME, Kuipers B, Grobe SJ. A description of think aloud method and protocol analysis. Qual Health Res 2016 Jul 01;3(4):430-441. [doi: 10.1177/104973239300300403]
40. Johnson WR, Artino ARJ, Durning SJ. Using the think aloud protocol in health professions education: an interview method for exploring thought processes: AMEE Guide No. 151. Med Teach 2022 Dec 19;45(9):937-948. [doi: 10.1080/0142159x.2022.2155123]
41. Roberts JP, Fisher TR, Trowbridge MJ, Bent C. A design thinking framework for healthcare management and innovation. Healthc (Amst) 2016 Mar;4(1):11-14. [doi: 10.1016/j.hjdsi.2015.12.002] [Medline: 27001093]
42. Ericsson KA. Protocol analysis and expert thought: concurrent verbalizations of thinking during experts' performance on representative tasks. In: Ericsson KA, Charness N, Feltovich PJ, Hoffman RR, editors. The Cambridge Handbook of Expertise and Expert Performance. Cambridge, MA: Cambridge University Press; 2006:223-242.
43. Berelson B. Content Analysis in Communication Research. Glencoe, Ill: Free Press; 1952.
44. Winsler A, Fernyhough C, McClaren EM, Way E. Private speech coding manual. Unpublished manuscript. George Mason University. 2005. URL: https://www.researchgate.net/profile/Adam-Winsler/publication/228654613_Private_speech_coding_manual/links/09e4150d729fe68323000000/Private-speech-coding-manual.pdf [accessed 2024-03-29]
45. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Measure 1960;20(1):37-46. [doi: 10.1177/001316446002000104]
46. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
47. Güsken SR, Frings K, Zafar F, Saltan T, Fuchs-Frohnhofen P, Bitter-Krahe J. Einflussfaktoren auf die Nutzungsintention von Pflegekräften zur Verwendung digitaler Technologien in der ambulanten Pflege. Fallstudie zur Einführung eines Sensortexils. Z Arbeitswiss 2021 Oct 13;75(4):470-490 [FREE Full text] [doi: 10.1007/s41449-021-00277-4] [Medline: 34658493]

Abbreviations

ADApp: Apotheken-Drohnen-App
BPNSFS: Basic Psychological Need Satisfaction and Frustration Scale
METUX: Motivation, Engagement, Thriving in User Experience
SUS: System Usability Scale
TAM: technology acceptance model
TA: think-aloud
TENS: Technology-based Experience of Need Satisfaction
TUI: Technology Usage Inventory

Edited by A Kushniruk; submitted 04.08.23; peer-reviewed by S Ariffin, C Baxter, J Brooke; comments to author 29.10.23; revised version received 21.12.23; accepted 06.01.24; published 30.04.24.

Please cite as:
Fink F, Kalter I, Steindorff JV, Helmbold HK, Paulicke D, Jahn P
Identifying Factors of User Acceptance of a Drone-Based Medication Delivery: User-Centered Design Approach
JMIR Hum Factors 2024;11:e51587
URL: https://humanfactors.jmir.org/2024/1/e51587
doi:10.2196/51587
PMID:38687589

Original Paper

Facilitators of and Barriers to the Use of a Digital Self-Management Service for Diagnostic Testing: Focus Group Study With Potential Users

Kyma Schnoor[1,2], MSc; Esther P W A Talboom-Kamp[1,2,3], MBA, MD, PhD; Muamer Hajtić[4], MSc; Niels H Chavannes[1,2], MD, PhD; Anke Versluis[1,2], PhD

[1]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, Netherlands

[2]National eHealth Living Lab, Leiden University Medical Center, Leiden, Netherlands

[3]Zuyderland, Sittard-Geleen, Netherlands

[4]Department of Medical informatics, University of Amsterdam, Amsterdam, Netherlands

Corresponding Author:
Kyma Schnoor, MSc
Department of Public Health and Primary Care
Leiden University Medical Center
Hippocratespad 21
Leiden
Netherlands
Phone: 31 71526 8433
Email: k.schnoor@lumc.nl

Abstract

Background: Health care lags in digital transformation, despite the potential of technology to improve the well-being of individuals. The COVID-19 pandemic has accelerated the uptake of technology in health care and increased individuals' willingness to perform self-management using technology. A web-based service, Directlab Online, provides consumers with direct digital access to diagnostic test packages, which can digitally support the self-management of health.

Objective: This study aims to identify the facilitators, barriers, and needs of Directlab Online, a self-management service for web-based access to diagnostic testing.

Methods: A qualitative method was used from a potential user's perspective. The needs and future needs for, facilitators of, and barriers to the use of Directlab Online were evaluated. Semistructured focus group meetings were conducted in 2022. Two focus groups were focused on sexually transmitted infection test packages and 2 were focused on prevention test packages. Data analysis was performed according to the principles of the Framework Method. The Consolidated Framework for Implementation Research was used to categorize the facilitators and barriers.

Results: In total, 19 participants, with a mean age of 34.32 (SD 14.70) years, participated in the focus groups. Important barriers were a lack of privacy information, too much and difficult information, and a commercial appearance. Important facilitators were the right amount of information, the right kind of tests, and the involvement of a health care professional. The need for a service such as Directlab Online was to ensure its availability for users' health and to maintain their health.

Conclusions: According to the participants, facilitators and barriers were comprehension of the information, the goal of the website, and the overall appearance of the service. Although the service was developed in cocreation with health care professionals and users, the needs did not align. The users preferred understandable and adequate, but not excessive, information. In addition, they preferred other types of tests to be available on the service. For future research, it would be beneficial to focus on cocreation between the involved medical professionals and users to develop, improve, and implement a service such as Directlab Online.

(JMIR Hum Factors 2024;11:e45115) doi:10.2196/45115

KEYWORDS

eHealth; usability; self-management; diagnostic test service; diagnostic; testing; test service; perspective; focus group; user need; user testing; implementation; qualitative; test result; laboratory test; laboratory result

Introduction

Background

Society is changing, and the world is becoming increasingly digital [1]. Health care lags in digital transformation, despite the potential of technology to improve the well-being of individuals [1,2]. The COVID-19 pandemic accelerated the development and use of technology in health care, also referred to as eHealth, with more digital consultations and increased use of home monitoring [3,4]. Furthermore, the COVID-19 pandemic, among others, has increased the need and willingness of individuals to perform self-management [5-7]. In patients with chronic diseases, self-management strategies are often used to support patients in dealing with treatment and lifestyle changes [8]. In addition, self-management strategies can be used to support individuals with home diagnostic tests [9]. The concept of self-management aligns with this positive health definition: “health as the ability to adapt and self-manage in the face of social, physical, and emotional challenges” [10,11].

eHealth can be used in 3 stages of laboratory diagnostic testing. The first stage is triage and advice on diagnostic testing, the second stage is the testing itself (ie, at home or a facility), and the third stage is the communication of the test results to the user. A systematic review by Versluis et al [9] showed that web-based diagnostic testing services were positively evaluated and preferred over clinic-based testing. However, most of the evaluated services only offered tests to detect sexually transmitted infections (STIs) [9].

eHealth services can support self-management, for example, with web-based services that support behavior and lifestyle changes (eg, Liva Healthcare) [12] and with websites where individuals can obtain health information (eg, Thuisarts.nl) [13]. In addition, there are multiple apps to support patients with chronic conditions such as hypertension, diabetes, or lower back pain [14-16].

In the Netherlands, a web-based service called Directlab Online (Saltro, part of Unilabs) offers individuals direct access to laboratory diagnostic tests independent of a health care provider [17]. It is a so-called direct-to-consumer platform. Directlab Online gives individuals direct digital access to diagnostic testing based on a triage that aligns with medical guidelines. Unlike the services identified in the systematic review by Versluis et al [9], Directlab Online offers a variety of diagnostic tests, for example, diagnostic tests for STIs, COVID-19, and vitamin deficiencies, as well as testing for health-related questions concerning fatigue and the prevention of heart disease. The results and the information on the website can give individuals insight into their health, which could support and motivate them to adopt healthier behaviors [12]. In addition, it supports users to be better informed about their health without the interference of a health care professional, which can lead to more efficient and accessible care [18]. Packages to test the health of individuals align with the patient-centered care approach, which can lead to a better quality of care [19]. Patient-centered care aims to empower patients to take charge of their health and actively participate in their health care [20]. Another term used is person-centered care, which is similar but does not solely focus on disease-related aspects, aligning better with the positive health definition [21].

To completely harness the potential and significance of Directlab Online, prioritizing high-quality and user-friendly service is paramount. Delving into the barriers and facilitators individuals encounter when using the service could provide invaluable insights, facilitating the enhancement of its user-friendliness and effectiveness. For example, known factors in dermatology that could influence the uptake of a digital service are, among others, financial aspects and accessibility for a digital service [22]. In a study by Vergouw et al [23], facilitators of and barriers to digital services for older adults in primary care were researched. Nonfamiliarities with web-based environments appeared to be a barrier, and efficiency was seen as an important facilitator for using a digital service in primary care [23]. In the review of STI testing by Versluis et al [9], concerns regarding complicated language and data handling insecurities were also discovered for ordering an STI test on the web. To our knowledge, no research has been conducted on facilitators, barriers, and needs of a direct-to-consumer platform that offers direct access to multiple diagnostic tests and web-based results. Identifying the needs, facilitators, and barriers will help determine what is necessary to optimize the use and improve the implementation of those services. This can give insight into the potential future directions for developing such services.

Objectives

This study aims to identify the facilitators of and barriers to using a service such as Directlab Online and to identify the needs regarding direct digital access to diagnostic testing. To achieve this, focus groups were conducted. Half of the focus groups focused on STI test packages and the other half on prevention test packages. STI tests and prevention test packages are the most ordered test packages on Directlab Online. The focus is on potential users, that is, those who have not used Directlab Online before, because we are interested in capturing people’s first impression of the service.

Methods

The Service: Directlab Online

Directlab Online is a Dutch, web-based service available for everyone, through which diagnostic tests can be ordered on the web [17]. The service was developed by a multidisciplinary innovation team of a diagnostic company (Saltro, part of Unilabs) and was launched in 2016 [24,25]. The process of using the service is presented in Figure 1. First, individuals undergo a web-based triage, based on medical guidelines, to determine whether the diagnostic tests are relevant and, if applicable, which tests are relevant. Second, individuals can order and buy associated tests. Depending on the diagnostic tests ordered, a self-sampling kit is sent to the individual’s home address, or an appointment is scheduled at a blood collection center or a laboratory for collecting a blood sample. Once the laboratory receives the collected specimen, high-quality analyses are conducted. The results of the tests are communicated through a web-based, secure patient portal. Furthermore, deviating results are communicated to the patient’s general practitioner but only if the patient has authorized it. The triage was based on medical guidelines, and the diagnostic test packages were developed in cocreation with general practitioners and tested by them and laboratory specialists referred to as health care professionals. Diagnostic test packages consist of different parameters for diagnostic testing. For example, a test package for cholesterol measures the following parameters: low-density lipoproteins, high-density lipoproteins, triglycerides, and total cholesterol. Multimedia Appendix 1 provides a complete overview of the test packages that could be ordered on Directlab Online during the focus group meetings. Table 1 provides an overview of the prevention and STI test packages that were part of the discussions with the focus groups.

Figure 1. Stages of using Directlab Online.

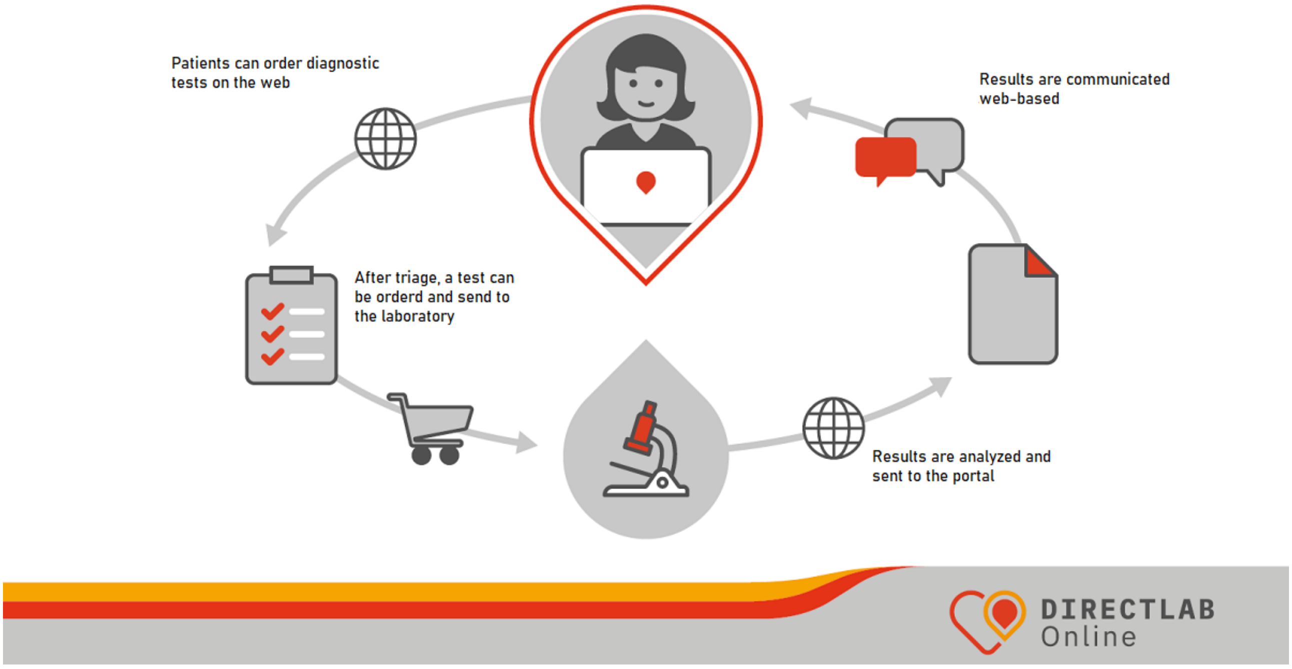


Table 1. Test packages that are available on Directlab Online.

Category	Parameters
Prevention tests	
Health checkup	Check total cholesterol[a], low-density lipoproteins (LDL)[a], high-density lipoproteins (HDL)[a], triglycerides[a], HbA_{1c}[a], and albumin and creatinine ratio[b]
Health checkup at home[c]	Measure parameters via self-sampling of blood: total cholesterol[d], LDL[d], HDL[d], triglycerides[d], HbA_{1c}[d], and albumin/creatinine ratio[b]
Cholesterol	Check total cholesterol[a], LDL[a], HDL[a], and triglycerides[a]
Cholesterol at home[c]	Measure parameters via self-sampling of blood: total cholesterol[d], LDL[d], HDL[d], and triglycerides[d]
Anemia	Check hemoglobin[a], mean corpuscular volume[a], ferritin[a], and C-reactive protein[a]
Diabetes	Check glucose[a] and HbA_{1c}[a]
Healthy bones[c]	Check calcium[a] and vitamin D[a]
Healthy kidneys[c]	Check creatinine[a], glomerular filtration rate[a], and albumin/creatinine ratio[b]
Thyroid check	Check thyroid function via thyroid-stimulating hormone[a] and free T4[a]
Sexually transmitted infection tests	
Chlamydia	Check for chlamydia[e]
Gonorrhea	Check for gonorrhea[e]
HIV	Check for HIV[a]
Syphilis	Check for syphilis[a]
Hepatitis B	Check for hepatitis B[a]

[a]Blood sample needed for diagnostics; HbA_{1c}: hemoglobin A_{1c}.

[b]Urine sample needed for diagnostics.

[c]These tests are not available any more on Directlab Online after the service update.

[d]Blood sample needed for diagnostics collected by self-sampling.

[e]Oral, anal, vaginal, or urine sample needed for diagnostic tests.

Study Design and Participants

Focus group meetings were conducted with potential users of the service. As the Directlab Online service offers a wide variety of test packages, we focused on 2 specific categories (ie, prevention test and STI test packages). These test packages were ordered most frequently. Half of the focus groups focused on STI test packages, and the other half focused on the prevention test packages. The general inclusion criteria for the focus groups were speaking Dutch and not having used Directlab Online earlier. In addition, there were specific inclusion criteria to ensure that the sociodemographic characteristics of the participants in the focus groups were consistent with the characteristics of the target population of the test packages. Notably, a specific inclusion criterion for the focus group about STI testing was that the participants were aged between 18 and 30 years. The specific inclusion criterion for the focus groups about prevention test packages was that the participants were aged between 18 and 65 years. It is important to note that there were no specific health or disease requirements to participate. Focus group meetings were held until data saturation was reached.

Ethical Considerations

The study was declared to not fall within the scope of the Dutch Medical Research Involving Human Subjects Act by the Leiden University Medical Center Medical Ethics Committee (N21.101).

Procedure and Data Collection

The recruitment period started on October 25, 2021, and lasted until February 20, 2022. Participants were recruited via different web-based channels (eg, LinkedIn [LinkedIn Corporation] and Facebook [Meta platforms, Inc]). Individuals were invited to contact KS via email when interested. Then, KS sent them more information. In addition, questions were asked regarding their birth year and if they could understand Dutch. A few date options for web-based meetings were sent if the individual met the inclusion criteria. When individuals could participate, they received an email with the date and time, a link to the Zoom (Zoom Video Communications) platform where the meeting would be conducted (on the web), and a link to a web-based informed consent form, which they were asked to sign before participation. All participants had the right to withdraw at any moment. The focus group meetings occurred between January

10 and March 2, 2022, in the presence of MH and KS [26]. KS led the focus groups, and MH managed the time and assisted with technical issues. The focus group meetings were in a semistructured format, following a predefined topic list with open-ended questions to leave space for discussion (Multimedia Appendix 2). First, general questions were asked regarding using eHealth to see how familiar participants were with eHealth. Second, participants were provided 10 minutes to view the website of Directlab Online and navigate through the website on a computer or mobile phone; no further instructions were given. When time was up, questions were asked regarding the website in general (eg, the first impression, whether they needed help when using the website, and whether they found the website attractive). While navigating the website, they had the option to write down notes or vocalize their impressions, expressing their observations, preferences, and feelings about the website [27]. Third, participants were instructed to go through Directlab Online, do some triages, and look at their test advice. Notably, we allowed participants to navigate through the process as normal users would. Hence, they were required to peruse informational materials, undergo a triage process involving medical inquiries concerning their symptoms, and obtain guidance regarding testing. Subsequently, questions were asked regarding the triage service, facilitators of and barriers to using Directlab Online, and the participants' needs for such a service. At the end of the focus groups, they received a digital gift card of €25 (US $27).

Data Analysis

All focus group meetings were audio recorded for subsequent analyses and were transcribed (intelligent) verbatim. When the transcripts were completed, the audio records were deleted. Two reviewers, MH and KS, conducted the qualitative data analysis according to the principles of the Framework Method [28]. The Framework Method is a systematic and flexible approach commonly used for the thematic analysis of health research semistructured interview data [29]. The method combines deductive and inductive techniques, which align with the aim of the study to identify specific issues regarding the use of Directlab Online and leave space to identify needs and opportunities that have not been formulated a priori. First, open coding was performed independently by the 2 reviewers, KS and MH. The interview data were coded using the software Atlas.ti 22 (Atlas.ti 22 Scientific Software Development). Second, the codes were compared between the 2 reviewers. Third, the codes were grouped into categories, resulting in the analytical framework. Fourth, for identifying the facilitators and barriers, the Consolidated Framework for Implementation Research (CFIR) was used [30]. The framework is widely used for the content analysis of qualitative data regarding the factors influencing implementation success [30]. Furthermore, the framework is comprehensive and makes it convenient to systematically study a wide array of facilitators and barriers [31]. In addition, using this framework made it possible to compare the findings and transfer them to other implementation studies [32]. The CFIR is a theory-driven model and comprises five domains: (1) the innovation domain, (2) the outer setting domain, (3) the inner setting domain, (4) the individuals' domain, and (5) the implementation process [30,33]. Identified facilitators and barriers were placed within the CFIR domains. Final themes were achieved via discussion and consensus between researchers KS and MH.

Results

Participant Characteristics

Data saturation was reached after forming 4 focus groups with 19 participants. The characteristics of the participants are shown in Table 2. The participants were aged 20 to 61 (mean 34.32, SD 14.70) years. The number of male (9/19, 47%) and female (10/19, 53%) participants was almost equal. The focus group meetings lasted around 90 minutes per group.

Age differed over the 2 different focus groups, as aligned with the target population of the diagnostic test packages. Overall, the experiences and choices of the focus groups regarding the website were the same. Therefore, in most cases, the focus group results were discussed together. When the results differed between the 2 groups, this was specified. Different themes around usability, facilitators, barriers, and needs emerged from the data and are elaborated in subsequent sections.

Table 2. Characteristics of the participants (N=19).

Participant	Gender	Age (years)	Focus group[a]
1	Woman	27	1
2	Woman	25	1
3	Man	24	1
4	Man	30	1
5	Woman	20	1
6	Woman	25	2
7	Woman	46	2
8	Woman	59	2
9	Man	24	2
10	Man	20	2
11	Woman	25	3
12	Man	25	3
13	Woman	30	3
14	Man	24	3
15	Man	39	4
16	Woman	58	4
17	Woman	59	4
18	Man	30	4
19	Man	62	4

[a]Groups 1 and 3 focused on sexually transmitted infection packages, and groups 2 and 4 focused on prevention packages.

Facilitators of and Barriers to the Uptake of Innovation

The identified barriers and facilitators were categorized specifically into the following 3 CFIR domains: innovation domain, outer setting domain, and individuals domain. The other 2 domains of the CFIR (ie, inner setting and implementation process) did not align with the facilitators and barriers mentioned by the participants and were therefore not discussed. Table 3 provides insight into the most essential and changeable facilitators and barriers identified. Therefore, it is not an exhaustive list of all potential barriers and facilitators that influenced the service uptake. It is notable that certain factors can be considered as a facilitator and barrier. For example, financial costs are frequently mentioned as a factor affecting the willingness to use digital health services [33]. When there are high user costs, it is a barrier; however, low costs can be considered a facilitator. The identified facilitators and barriers are explained in detail and explained per domain in subsequent sections.

Table 3. Facilitators and barriers derived from the focus groups embedded in the Conceptual Framework for Implementation Research (CFIR).

Domain of CFIR	Domain description	Results
Innovation domain		
Innovation source	The group that developed and visibly sponsored the use of the innovation is reputable, credible, and trustable.	• The general practitioner group that developed and visibly sponsored the service was reputable, credible, and trustable, which resulted in a reliable service. • Information about privacy and presenting good reviews improved reliability and credibility. • Commercial appearance influenced the credibility. Furthermore, stock pictures influenced the credibility.
Innovation relative advantage	The innovation is better than other available innovations or current practices.	• The service was easy to use, which made the service accessible. • It was easy to use the service without visiting the general practitioner.
Innovation complexity	The innovation is complicated, which may be reflected by its scope and the nature and number of connections and steps.	• Too many testing possibilities and too much information made the website less user-friendly. • The search bar and filters on the website increased the user-friendliness of the website. • Using multiple medical words made the service difficult to comprehend.
Outer setting domain		
Partnerships and connections	The inner setting is networked with external entities, including referral networks, academic affiliations, and professional organization networks.	• The service was linked with academic institutions and other medical professionals, which increased the reliability of the service for users.
Societal pressure	Mass media campaigns, advocacy groups, or social movements or protests drive the implementation and delivery of the innovation.	• Media campaigns, reviews, and blogs could help stimulate participants to use the service.
Individuals domain: subdomain patient characteristics		
Capability	The individual has interpersonal competence, knowledge, and skills to fulfill a “Role” (different characteristics of individuals).	• If participants had experience with a similar service, they felt more confident in using the service. Otherwise, feelings of anxiety or tension could have influenced their competence, knowledge, and skills.

Facilitators and Barriers in the Innovation Domain

Innovation Source

Participants mentioned different factors that were related to the innovation source of the innovation domain. These factors mainly influenced the credibility and trustworthiness in a positive (ie, facilitator) or negative (ie, barrier) way. First, the website’s commercial appearance were the most frequently mentioned barriers that influenced its reliability. For example, participants mentioned that the option to buy a gift card for a diagnostic test package did not align with a website designed for health. In addition, regarding the high prices for diagnostic test packages and the website’s general appearance, they said the following:

> *The website said: buy this. But I want to know why this test?* [Participant 4]
>
> *I found it a very commercial website; this lowers my enthusiasm.* [Participant 8]

Participants did not notice that health care professionals were involved in the service and partly developed the service, while this could increase the credibility of the website.

Second, the availability of reviews was frequently mentioned as a facilitator for reliability and credibility but as a barrier in some cases. Good reviews could be considered as a facilitator, and bad reviews could be considered as a barrier to experiencing the website as reliable and trustworthy. The following was said about this view:

> *Yes, [...] I found it important if I go to a new website to sell or buy something to see that others used the site and what they bought.* [Participant 13]

Third, 37% (7/19) of the participants mentioned the facilitator's "privacy." For the participants, it was important to know where the data were stored and for how long. However, this information was difficult to find on the website:

> *And then it is the question of how long data is stored and how that is important to know.* [Participant 8]

> *I want to know, what happens to the data and how long is it stored?* [Participant 16]

Participant 7 pointed out that a clear and transparent privacy statement could be a unique selling point of the service.

Finally, the most mentioned barrier in the innovation source was the presence of stock pictures on Directlab Online:

> *[...] those stock pictures on the website; they gave an image of unreliability.* [Participant 3]

Participants mentioned that pictures of real people or famous people who used the tests could be a facilitator and positively influence the reliability and use of the service. Furthermore, they mentioned that a short video with education and instructions about diagnostic test packages could improve the triage's clarity and the diagnostic packages' content.

Innovation Relative Advantage

Participants mentioned several factors regarding why they would use this innovative service. These factors were mostly related to the accessibility of the service compared to other services or normal practice. For example, the easiness of ordering a test on the web without going to the general practitioner was a relative advantage of the service, as mentioned by a participant:

> *Yes, I would rather order online because going to the general practitioner [...] it takes time.* [Participant 7]

Furthermore, another participant mentioned the benefit of ordering a test on the web without going to the general practitioner:

> *Hmm yes, I thought of a few things when I first saw the website... of the vitamin tests, STI tests, and COVID tests, I thought yes, you do not want to go to the general practitioner for that. Especially for STI testing, the threshold is high. In this way, you still test and see if you are healthy.* [Participant 1]

However, the relative advantage was negatively influenced by the high costs of the tests. One participant stated:

> *The costs will stop people from buying anything.* [Participant 17]

Innovation Complexity

Several facilitators and barriers that influenced the complexity of the service were mentioned by the participants.

First, the number of test packages and parameters available was confusing. It became clear from the focus groups that offering the "right" number of diagnostic tests was important; participants were not enthusiastic about a test package with many separate parameters. Participants mentioned that they were optimistic about the possibility of ordering STI testing, COVID-19 testing, and some prevention tests. However, they mentioned that after the triage, they received advice to select many different test packages. Recommending many diagnostic test packages to the participants was a barrier because they were confused about which test package was important for them. Furthermore, the high amount of information provided about these test packages was experienced as challenging by approximately half (11/19, 58%) of the participants:

> *When I open the website, a lot of information is present. Too many tests are available. Of course, this website wants to sell tests, but... I do not know. I found the home page too complicated, too unclear.* [Participant 13]

Second, the language used on the website was a factor that influenced the use of the service. The language on the home page was experienced as straightforward and was therefore a facilitator. However, when completing the triage and choosing the diagnostic package, the information was more challenging to understand. Notably, medical and incomprehensible terms were used:

> *I think you have a very broad target group of people who would like to use this, and I think it is written for the somewhat well-educated, reasonably well-informed citizen, shall we say [...] Offer more comfort to people by using less difficult vocabulary.* [Participant 8]

Third, the participants mentioned elements of the website that influenced user-friendliness. Participants were happy with the filters in the search bar to find a particular test, as well as the search function and the website's colors:

> *Personally, I found the website easy to use, and what I experienced as very positive were the filters [...].* [Participant 14]

However, approximately one-third (6/19, 32%) of the participants found the website unclear—among others, due to too much text—and complicated (eg, where to find what they were looking for), and they found the home page too busy.

Facilitators and Barriers in the Outer Setting Domain

Partnerships and Connections

The service was linked to academic institutions, which increased its reliability. Mentioning partners would increase the uptake according to the participants:

> *Yes, mentioning partners would be nice. And famous names always attract attention.* [Participant 13]

Societal Pressure

Participants mentioned that reviews and blogs could help in increasing the use of the service and its reliability:

> *You want to read reviews and experiences of others.* [Participant 5]

Facilitators and Barriers in the Individuals Domain

The individual's skills and knowledge regarding services such as Directlab Online influenced their willingness to use the service and their perception of potentially using it. The younger participants (aged 20 to 30 years) mentioned that they had experience with this type of website, which reassured them to use this service. However, some older participants (aged ≥39 years) had less experience with digital services in general and mentioned some anxiety and tension when they needed to order a test. Some (4/19, 21%) preferred to go to the general practitioner for diagnostic tests. However, participants of all age groups mentioned the benefit of ordering STI tests on the web without visiting the general practitioner.

Future Needs

Different needs were identified regarding services such as Directlab Online. First, the service's purpose must be more explicit for the participants. For them, it was unclear whether the service could help them self-manage their health:

> *And this is what I miss on the website; what is in it for me and my health as a patient or consumer?* [Participant 19]

Second, there was a need to understand the advantages of ordering diagnostic tests on the web (eg, more accessible than going to the general practitioner for tests). Participants wanted this information to be more evident on the website. Third, after receiving their results, the participants explained that they preferred to have more information regarding how they could remain healthy or what they could do to become healthier. It could help, according to the participants, to let them know more specifically that general practitioners make the diagnostic test packages designed for the service. All participants saw the benefit of ordering STI diagnostic test packages on the web and undergoing them at home. The current offer of diagnostic test packages does not meet the wishes of all participants. There was a need for additional tests, such as tests for food allergies, testosterone levels, fertility, or urinary infections. A participant mentioned as follows:

> *I want a urine tract infection test; those are relatively cheap, I think [...].* [Participant 1]

Discussion

Principal Findings

This qualitative study aimed to evaluate the facilitators of and barriers to web-based direct access to diagnostic test services from the perspective of potential users. In addition, the study tried to identify the need to use such services. The study showed that a tailored amount of information could benefit the service. Participants need to use a service such as Directlab Online to ensure that the website is available for their health. The participants needed to see the benefit of a diagnostic test package. Identified barriers and facilitators were categorized using the CFIR. The study showed that a lack of privacy, information overload, and a commercial appearance were important barriers. Facilitators included providing the right amount of information on the service and involving a health care professional in developing the service. In addition, the study showed that a tailored amount of information could benefit the use of the service. In short, we noticed that several facilitators and barriers were influencing the reliability or accessibility of the service. For example, the commercial appearance and lack of privacy information contributed to a less reliable service for the potential users, and ordering a test on the web without a health care professional influenced the accessibility.

Directlab Online is a service for users to support themselves in self-managing their health. An important quality-enhancing element for Directlab Online was that health care professionals were actively involved in developing this service. Health care professionals significantly influenced the content and information shown on the website. However, the focus groups with potential users identified needs and wishes that did not completely align with the ideas of the general practitioner. To illustrate, health care professionals preferred other types of diagnostic test packages on the web than those that the participants preferred to use. Furthermore, the general practitioners preferred detailed information on the website, whereas this information overload was not always beneficial for the participants. A study by Talboom-Kamp et al [34] regarding a web-based results portal discussed the complex balance between the general practitioner's necessities and participants' needs for the right amount of understandable information. Presenting information requires a balance between an overload of medical information and the information users need to understand test packages and results. A potential way to solve overwhelming participants with information is to not present all the information directly in one view to the participant but by offering clickable links or short videos [34].

This study used the CFIR to identify and categorize the facilitators and barriers. In another study, Versluis et al [33] performed an inventory to determine the obstacles that must be overcome and how to optimize eHealth in primary care using this framework. They found similar results to our study; costs and privacy issues were identified as important barriers. In addition, in line with other studies, the following facilitators were identified as having "experience with eHealth" and "easiness of use" [33,35]. In comparison with other studies using the CFIR to classify facilitators and barriers, similar factors were predominantly identified. A notable factor highlighted in the study by Verweij et al [36] involving patients with cancer using a digital self-monitoring system was the necessity to elucidate the service's added value, alongside concerns regarding privacy issues. However, other factors were also mentioned, such as the connection with health care professionals, which were not identified in our study. The target population (patients with cancer) could be an important explanation for this difference. The comparison with other literature revealed that, irrespective of the type of digital service or the user population, the facilitators and barriers remained quite consistent. This study's inventory could help determine what obstacles need to be overcome and how we might optimize an application such as Directlab Online.

Depending on the participants, mainly influenced by age, some would use a web-based website to organize their health. In contrast, other participants, mainly older participants, were more at ease with visiting the general practitioner and organizing their health directly via the general practitioner [37]. In this study, the older participants preferred to visit the general practitioner, which could lead to the cautious conclusion that web-based direct access to diagnostic services is not attractive to everyone [37]. In addition, this study showed that using a service such as Directlab Online is not only related to age but also related to the user's health problem and that the type of test package was important. Participants' needs were to feel the relevance of ordering a diagnostic test package on the web instead of visiting the general practitioner. The relevance was clear for the STI test packages but unclear for other diagnostic test packages. The study results showed that it remains important to involve all end users in the service to ensure that the service supports the needs of the target population [38]. Directlab Online was developed with general practitioners, and the elements that they found important were integrated into the service. While this study provided insight into the facilitators and barriers of potential users, it appeared that these things were not the same. It is important for a reliable and proper service that both perspectives of all stakeholders should be included in (further) development of such services. Finally, the facilitators and barriers to using a service such as Directlab Online found in this study could be used to optimize this service and other comparable services.

Strengths and Limitations

There is a lot of direct access to diagnostic testing services available, mainly when it entails STI diagnostic test packages. However, not many of them have a scientific basis or are developed by medical professionals. This is the first study that examined the facilitators of and barriers to a service that provides more diagnostic test packages than only STI tests and which is developed in cocreation with health care professionals. Another strength of the study was that the CFIR was used to analyze the facilitators and barriers mentioned in the focus groups. Embedding the facilitators and barriers in this framework made the comparison with other research easier. In addition, the domains identified by the CFIR can help to find the right implementation strategy [33,39].

This study focused on potential users because we were interested in their first impression of the service. The rationale was that, in the real world, such a service could be visited by many new users [40]. Previous experiences have not biased the impression of potential users. However, this could also be a limitation because participants who did use Directlab Online before could have another opinion regarding the service. This made the results less generalizable. Another limitation is that the mean age of participants was relatively low, making it more difficult to generalize the results to the general Dutch population. However, all participants, independent of age, mentioned the benefit of ordering STI tests on the web. The service showed benefits for participants who were ashamed to visit a general practitioner for a diagnostic package and for participants who wished to order tests in an accessible, nonbinding manner.

Future Research

Directlab Online is a service developed for a wide range of users. However, this study showed that it is important to include end users to ensure that the service aligns with the population's needs. Cocreation with end users and medical professionals could be a solution to solve disbalances in wishes and needs between them and to improve an eHealth application [38]. For future research, organizing cocreation sessions and analyzing their results could be beneficial to improve the service. Finally, in future research, information about the influence of the diagnostic test's result on the user's lifestyle could be analyzed. Namely, this could result in a preventive role for a service such as Directlab Online to improve the health of a population.

Conclusions

According to participants, information provision, comprehension, and the overall appearance of the website were the most important elements that influenced the use and uptake of a direct-to-consumer website for diagnostic test packages. Barriers, such as the commercial appearance and lack of privacy information, negatively influenced reliability and accessibility. The study showed that it is important to include relevant stakeholders in creating an eHealth intervention because there was a disbalance between the users' needs and what the involved general practitioners considered necessary. Future research could take a quantitative approach to further identify the needs regarding test packages and to identify the demographics of users and the influence of test results on the behavior of users. Directlab Online offers opportunities for more web-based self-management of health.

Acknowledgments

The authors would like to thank all participants for their input and contributions during the focus groups. Funding for this research was provided by Unilabs.

Conflicts of Interest

During the research, KS and EPWATK were employees at Unilabs.

Multimedia Appendix 1

Diagnostic test packages overview of Directlab Online.

[DOCX File , 20 KB - humanfactors_v11i1e45115_app1.docx]

Multimedia Appendix 2

Semistructured interview protocol.
[DOCX File , 15 KB - humanfactors_v11i1e45115_app2.docx]

References

1. Chui M, Roberts R, Yee L. McKinsey technology trends outlook 2022. McKinsey & Company. 2022 Aug. URL: https://tinyurl.com/3tv93phy [accessed 2024-04-28]
2. Rawstron E, Baulderstone M. Plugging healthcare workers into the digital future. KPMG. 2022. URL: https://tinyurl.com/bdm4dpu9 [accessed 2024-04-28]
3. Clipper B. The influence of the COVID-19 pandemic on technology: adoption in health care. Nurse Lead 2020 Oct;18(5):500-503 [FREE Full text] [doi: 10.1016/j.mnl.2020.06.008] [Medline: 32837346]
4. Keuper J, Batenburg R, Verheij R, van Tuyl L. Use of e-health in Dutch general practice during the COVID-19 pandemic. Int J Environ Res Public Health 2021 Nov 26;18(23):12479 [FREE Full text] [doi: 10.3390/ijerph182312479] [Medline: 34886204]
5. Menting J, Schelven F, Boeije H. Gevolgen van de coronapandemie voor gezondheid, behandeling en zelfmanagement van mensen met een chronische ziekte. Nivel. 2020. URL: https://nivel.nl/nl/publicatie/gevolgen-van-de-coronapandemie-voor-gezondheid-behandeling-en-zelfmanagement-van-mensen [accessed 2024-04-28]
6. Flinterman L, Groenewegen P, Verheij R. Zorglandschap en zorggebruik in een veranderende eerste lijn. Nivel. 2018. URL: https://www.nivel.nl/nl/publicatie/zorglandschap-en-zorggebruik-een-veranderende-eerste-lijn [accessed 2024-04-28]
7. Gauthier-Beaupré A, Battistini BJ, Kuziemsky C, Jutai JW. Policymaker perspectives on self-management of disease and disabilities using information and communication technologies. Health Res Policy Syst 2023 Jun 14;21(1):52 [FREE Full text] [doi: 10.1186/s12961-023-01004-7] [Medline: 37316838]
8. McGowan P. Self-management: a background paper. University of Victoria - Centre on Aging. 2005. URL: https://tinyurl.com/378my7xc [accessed 2024-04-28]
9. Versluis A, Schnoor K, Chavannes NH, Talboom-Kamp EP. Direct access for patients to diagnostic testing and results using eHealth: systematic review on eHealth and diagnostics. J Med Internet Res 2022 Jan 12;24(1):e29303 [FREE Full text] [doi: 10.2196/29303] [Medline: 35019848]
10. Huber M, Knottnerus JA, Green L, van der Horst H, Jadad AR, Kromhout D, et al. How should we define health? BMJ 2011 Jul 26;343(jul26 2):d4163. [doi: 10.1136/bmj.d4163] [Medline: 21791490]
11. Huber M, van Vliet M, Boers I. Heroverweeg uw opvatting van het begrip 'gezondheid'. Ned Tijdschr Geneeskd 2016;160:A7720 [FREE Full text]
12. Komkova A, Brandt CJ, Hansen Pedersen D, Emneus M, Sortsø C. Electronic health lifestyle coaching among diabetes patients in a real-life municipality setting: observational study. JMIR Diabetes 2019 Mar 12;4(1):e12140 [FREE Full text] [doi: 10.2196/12140] [Medline: 30860486]
13. Spoelman WA, Bonten TN, de Waal MW, Drenthen T, Smeele IJ, Nielen MM, et al. Effect of an evidence-based website on healthcare usage: an interrupted time-series study. BMJ Open 2016 Nov 09;6(11):e013166 [FREE Full text] [doi: 10.1136/bmjopen-2016-013166] [Medline: 28186945]
14. Machado GC, Pinheiro MB, Lee H, Ahmed OH, Hendrick P, Williams C, et al. Smartphone apps for the self-management of low back pain: a systematic review. Best Pract Res Clin Rheumatol 2016 Dec;30(6):1098-1109. [doi: 10.1016/j.berh.2017.04.002] [Medline: 29103552]
15. El-Gayar O, Timsina P, Nawar N, Eid W. Mobile applications for diabetes self-management: status and potential. J Diabetes Sci Technol 2013 Jan 01;7(1):247-262 [FREE Full text] [doi: 10.1177/193229681300700130] [Medline: 23439183]
16. Alessa T, Hawley MS, Hock ES, de Witte L. Smartphone apps to support self-management of hypertension: review and content analysis. JMIR Mhealth Uhealth 2019 May 28;7(5):e13645 [FREE Full text] [doi: 10.2196/13645] [Medline: 31140434]
17. Unilabs Online homepage. Unilabs Online. URL: https://directlabonline.nl/ [accessed 2024-04-24]
18. Holman H, Lorig K. Patient self-management: a key to effectiveness and efficiency in care of chronic disease. Public Health Rep 2004 Aug 02;119(3):239-243 [FREE Full text] [doi: 10.1016/j.phr.2004.04.002] [Medline: 15158102]
19. Bergeson SC, Dean JD. A systems approach to patient-centered care. JAMA 2006 Dec 20;296(23):2848-2851. [doi: 10.1001/jama.296.23.2848] [Medline: 17179462]
20. Reynolds A. Patient-centered care. Radiol Technol 2009;81(2):133-147. [Medline: 19901351]
21. Starfield B. Is patient-centered care the same as person-focused care? Perm J 2011;15(2):63-69 [FREE Full text] [doi: 10.7812/TPP/10-148] [Medline: 21841928]
22. Ariens LF, Schussler-Raymakers FM, Frima C, Flinterman A, Hamminga E, Arents BW, et al. Barriers and facilitators to eHealth use in daily practice: perspectives of patients and professionals in dermatology. J Med Internet Res 2017 Sep 05;19(9):e300 [FREE Full text] [doi: 10.2196/jmir.7512] [Medline: 28874336]
23. Vergouw JW, Smits-Pelser H, Kars MC, van Houwelingen T, van Os-Medendorp H, Kort H, et al. Needs, barriers and facilitators of older adults towards eHealth in general practice: a qualitative study. Prim Health Care Res Dev 2020 Dec 02;21:e54 [FREE Full text] [doi: 10.1017/S1463423620000547] [Medline: 33263272]
24. Over Saltro. Saltro. URL: https://saltro.nl/ [accessed 2024-04-28]

25. Unilabs homepage. Unilbas. URL: https://unilabs.com/ [accessed 2022-11-28]
26. Willemsen RF, Aardoom JJ, Chavannes NH, Versluis A. Online synchronous focus group interviews: practical considerations. Qual Res 2023;23(6):1810-1820. [doi: 10.1177/14687941221110161]
27. Bonten TN, Rauwerdink A, Wyatt JC, Kasteleyn MJ, Witkamp L, Riper H, et al. Online guide for electronic health evaluation approaches: systematic scoping review and concept mapping study. J Med Internet Res 2020 Aug 12;22(8):e17774 [FREE Full text] [doi: 10.2196/17774] [Medline: 32784173]
28. Spencer L, Ritchie J, Lewis J, Dillon L. Quality in qualitative evaluation: a framework for assessing research evidence. Government Chief Social Researcher's Office, United Kingdom. 2003 Aug. URL: https://assets.publishing.service.gov.uk/media/5a8179c1ed915d74e33fe69e/Quality-in-qualitative-evaulation_tcm6-38739.pdf [accessed 2024-04-28]
29. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13:117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
30. Breimaier HE, Heckemann B, Halfens RJ, Lohrmann C. The Consolidated Framework for Implementation Research (CFIR): a useful theoretical framework for guiding and evaluating a guideline implementation process in a hospital-based nursing practice. BMC Nurs 2015 Aug 12;14:43 [FREE Full text] [doi: 10.1186/s12912-015-0088-4] [Medline: 26269693]
31. Safaeinili N, Brown-Johnson C, Shaw JG, Mahoney M, Winget M. CFIR simplified: pragmatic application of and adaptations to the Consolidated Framework for Implementation Research (CFIR) for evaluation of a patient-centered care transformation within a learning health system. Learn Health Syst 2020 Sep 26;4(1):e10201 [FREE Full text] [doi: 10.1002/lrh2.10201] [Medline: 31989028]
32. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
33. Versluis A, van Luenen S, Meijer E, Honkoop PJ, Pinnock H, Mohr DC, et al. SERIES: eHealth in primary care. Part 4: addressing the challenges of implementation. Eur J Gen Pract 2020 Dec;26(1):140-145 [FREE Full text] [doi: 10.1080/13814788.2020.1826431] [Medline: 33025820]
34. Talboom-Kamp E, Tossaint-Schoenmakers R, Goedhart A, Versluis A, Kasteleyn M. Patients' attitudes toward an online patient portal for communicating laboratory test results: real-world study using the eHealth impact questionnaire. JMIR Form Res 2020 Mar 04;4(3):e17060 [FREE Full text] [doi: 10.2196/17060] [Medline: 32024632]
35. Cajita MI, Hodgson NA, Lam KW, Yoo S, Han HR. Facilitators of and barriers to mHealth adoption in older adults with heart failure. Comput Inform Nurs 2018 Aug;36(8):376-382 [FREE Full text] [doi: 10.1097/CIN.0000000000000442] [Medline: 29742549]
36. Verweij L, Smit Y, Blijlevens NM, Hermens RP. A comprehensive eHealth implementation guide constructed on a qualitative case study on barriers and facilitators of the digital care platform CMyLife. BMC Health Serv Res 2022 Jun 06;22(1):751 [FREE Full text] [doi: 10.1186/s12913-022-08020-3] [Medline: 35668491]
37. Wilson J, Heinsch M, Betts D, Booth D, Kay-Lambkin F. Barriers and facilitators to the use of e-health by older adults: a scoping review. BMC Public Health 2021 Aug 17;21(1):1556 [FREE Full text] [doi: 10.1186/s12889-021-11623-w] [Medline: 34399716]
38. Kasteleyn MJ, Versluis A, van Peet P, Kirk UB, van Dalfsen J, Meijer E, et al. SERIES: eHealth in primary care. Part 5: a critical appraisal of five widely used eHealth applications for primary care - opportunities and challenges. Eur J Gen Pract 2021 Dec;27(1):248-256 [FREE Full text] [doi: 10.1080/13814788.2021.1962845] [Medline: 34432601]
39. Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, Matthieu MM, et al. A refined compilation of implementation strategies: results from the expert recommendations for implementing change (ERIC) project. Implement Sci 2015 Feb 12;10:21 [FREE Full text] [doi: 10.1186/s13012-015-0209-1] [Medline: 25889199]
40. Grindrod KA, Li M, Gates A. Evaluating user perceptions of mobile medication management applications with older adults: a usability study. JMIR Mhealth Uhealth 2014 Mar 14;2(1):e11 [FREE Full text] [doi: 10.2196/mhealth.3048] [Medline: 25099993]

Abbreviations

CFIR: Consolidated Framework for Implementation Research
STI: sexually transmitted infection

Edited by A Kushniruk; submitted 30.01.23; peer-reviewed by S Choemprayong, Y Chu, CE Moe; comments to author 19.12.23; revised version received 02.01.24; accepted 20.03.24; published 10.05.24.

Please cite as:
Schnoor K, Talboom-Kamp EPWA, Hajtić M, Chavannes NH, Versluis A
Facilitators of and Barriers to the Use of a Digital Self-Management Service for Diagnostic Testing: Focus Group Study With Potential Users
JMIR Hum Factors 2024;11:e45115
URL: https://humanfactors.jmir.org/2024/1/e45115
doi:10.2196/45115
PMID:38728071

Original Paper

Evaluating the Energy Efficiency of Popular US Smartphone Health Care Apps: Comparative Analysis Study Toward Sustainable Health and Nutrition Apps Practices

Abdullah Almasri[1*], PhD; Tatyana Y El-Kour[2*], RDN, PhD; Liliana Silva[3*], PhD; Yousef Abdulfattah[4*], BSc

[1]College of Humanities and Sciences, Prince Sultan University, Riyadh, Saudi Arabia

[2]Media Psychology Department, Fielding Graduate University, Santa Barbara, CA, United States

[3]Faculty of Health Sciences, University Fernando Pessoa, Porto, Portugal

[4]College of Computer & Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia

[*]all authors contributed equally

Corresponding Author:
Abdullah Almasri, PhD
College of Humanities and Sciences
Prince Sultan University
PPP2+CGP, Nasser Bin Farhan Street
King Salman Neighborhood
Riyadh, 12435
Saudi Arabia
Phone: 966 4948000
Email: aalmasri@psu.edu.sa

Abstract

Background: The emergence of smartphones has sparked a transformation across multiple fields, with health care being one of the most notable due to the advent of mobile health (mHealth) apps. As mHealth apps have gained popularity, there is a need to understand their energy consumption patterns as an integral part of the evolving landscape of health care technologies.

Objective: This study aims to identify the key contributors to elevated energy consumption in mHealth apps and suggest methods for their optimization, addressing a significant void in our comprehension of the energy dynamics at play within mHealth apps.

Methods: Through quantitative comparative analysis of 10 prominent mHealth apps available on Android platforms within the United States, this study examined factors contributing to high energy consumption. The analysis included descriptive statistics, comparative analysis using ANOVA, and regression analysis to examine how certain factors impact energy use and consumption.

Results: Observed energy use variances in mHealth apps stemmed from user interactions, features, and underlying technology. Descriptive analysis revealed variability in app energy consumption (150-310 milliwatt-hours), highlighting the influence of user interaction and app complexity. ANOVA verified these findings, indicating the critical role of engagement and functionality. Regression modeling (energy consumption = $\beta + \beta_1 \times$ notification frequency + $\beta_2 \times$ GPS use + $\beta_3 \times$ app complexity + ε), with statistically significant *P* values (notification frequency with a *P* value of .01, GPS use with a *P* value of .05, and app complexity with a *P* value of .03), further quantified these bases' effects on energy use.

Conclusions: The observed differences in the energy consumption of dietary apps reaffirm the need for a multidisciplinary approach to bring together app developers, end users, and health care experts to foster improved energy conservation practice while achieving a balance between sustainable practice and user experience. More research is needed to better understand how to scale-up consumer engagement to achieve sustainable development goal 12 on responsible consumption and production.

(JMIR Hum Factors 2024;11:e58311) doi:10.2196/58311

KEYWORDS

mobile health; energy consumption in health care smartphone apps; dietary tracking apps; optimization and sustainability in mobile health; user engagement and experience; Android apps performance; digital health technologies; app; apps; applications; digital health; energy; consumption; sustainable; sustainability; environment; environmental; use; smartphone; smartphones; electricity; electrical; mobile phone

Introduction

Background

Nations worldwide and researchers from various disciplines are increasingly focusing on sustainable and energy-efficient techniques for energy production. The works of Bhaskar et al [1], Muthanna et al [2], and Ashfaq et al [3] exemplified the innovative approaches being developed in this domain, highlighting the significance of renewable energy applications, unmanned aerial vehicle path scheduling in the Internet of Things, and secure energy trading with machine learning and blockchain technology, respectively. In today's health care scene, smartphones stand as crucial companions, seamlessly connecting the realms of technology and wellness promotion. The surge in popularity of mobile health (mHealth) apps reflects a broader movement toward adopting energy-smart habits in all facets of mobile computing. This trend underscores the pivotal role of crafting sustainable software to lessen our ecological footprint, a goal echoed by the strides made in green computing and energy-saving innovations [4-6]. These apps mark a transformative step toward digital health, empowering people to proactively manage their health journeys. The growing focus on energy efficiency and the adoption of eco-friendly use habits emphasize the significance of these apps. Research by Choi et al [7] and Pop et al [8] shed light on the essential role that energy-efficient software plays in prolonging the lifespan of devices and mitigating environmental impacts, heralding a significant shift in digital health practices. The fusion of wearable technologies with these apps further highlights the importance of designing with energy mindfulness at the forefront, ensuring that our pursuit of health does not lead to unsustainable energy use. Table 1 illustrates the relationship between the popularity of mHealth apps and their user review scores. Apps with the highest user satisfaction were selected in this study to be assessed for energy efficiency.

Table 1. Correlation between app popularity, where popularity is determined by the number of downloads.

App name	Downloads (in millions)	User review (out of 5)
Ate Food Journal	2	4.4
Calorie Counter	5	4.2
Lifesum	10	4.6
My Plate	8	4.4
MyFitnessPal	45	4.4
Noom	15	4.2
Ovia	3	4.0
PlateJoy	1	4.6
Spokin	4	4.2
Yummly	20	4.6

Problem Statement

Even though mHealth and nutrition apps have become increasingly popular, there is a dearth of research on how much energy they are consumed on Android devices and practical guidance on what can users do about it. Almasri and Gouveia [9] studied the gap in sustainable practice using Android apps and highlighted the need given their popularity and potential for energy-saving practice and given the global priority and commitment toward creating sustainable smartphones to achieve sustainable development goal 12.

Objective

The objective of this study is to assess the energy consumption of popular mHealth and nutrition apps and identify key areas where improvements can be made.

Literature Review

mHealth Apps and Energy Consumption

The widespread use of mHealth apps in our everyday routines has underscored the need to better understand energy consumption. Awais et al [10] examined the direct link between the complexity of these apps and their energy demands. Their findings indicate that apps with advanced features, such as real-time monitoring and personalized recommendations, can consume up to 30% more energy compared with simpler apps. Additionally, Sahar et al [11] provided an in-depth analysis of how unnoticed background activities, such as continuous data syncing and location tracking, play a significant role in draining smartphone batteries. Their study revealed that background activities could account for up to 40% of an app's total energy consumption, underscoring the importance of both developers and users to understand the app architecture and appreciate the influence it plays on energy use.

User Behavior and Energy Efficiency

Understanding energy efficiency warrants an understanding of user behaviors around app use. Personal relationships, belief in one's abilities as presented by Rahman et al [12] (self-efficacy), and the collective confidence in our shared power to effect change are key to embracing and consistently using mHealth technologies [13-15]. How individuals use mHealth apps has a significant impact on energy consumption patterns. Al Nidawi et al [16] showed that regular app use, including entering data and syncing, significantly increases energy consumption. Acer et al [17] highlighted how notifications, a common feature in mHealth apps, significantly boost energy use.

Strategies for Energy Optimization

Isuwa et al [18] showed that using adaptive brightness settings and energy-saving modes can extend the battery life of mobile devices by up to 20%. Furthermore, Benkhelifa et al [19] explored the potential of leveraging software-defined networking for energy optimization in mobile cloud computing, resulting in a decrease of up to 25% in energy use.

Technological Advancements and Energy Consumption

Emerging technologies play a nuanced role in the story of mHealth apps' energy consumption, presenting a mix of hurdles and breakthroughs. On the one hand, advancements in app development frameworks, as outlined by Kelényi et al [20], opened fresh opportunities for energy efficiency. On the other hand, the growing complexity of those apps, as pointed out by Porter [21], introduces significant obstacles to keeping energy use in check. The potential of artificial intelligence (AI) in optimizing energy consumption for sustainability has been highlighted in a recent article by Ericsson [22]. Their research indicates that AI features, while enhancing app functionality, can lead to a 25% increase in energy consumption if not optimized properly.

Cross-Platform Analysis of Energy Consumption

Khan et al [23] conducted a comparative analysis of power consumption in mobile devices to inform the development of energy-efficient mobile apps. Their study introduced a methodology for assessing and evaluating power use, providing valuable insights and guidelines for developers aiming to create more sustainable mobile apps. Ciman and Gaggi [24] analyzed smartphone energy consumption using different sensors, including either only app, or by using GPS, accelerometer, compass, camera, or microphone. They found that cross-platform frameworks significantly increase energy consumption compared with native apps. They suggested that power consumption should be considered when choosing between native implementation and using a framework or between different frameworks for mobile app development.

Energy Consumption Metrics and Measurement Techniques

Ergasheva et al [25] explored metrics of energy consumption to evaluate the energy efficiency of apps. They introduced metrics, such as energy-per-function, which quantifies the energy consumed for each app function, and energy-per-user interaction, which measures the energy used per user interaction, providing a more granular understanding of app energy consumption. Pathak et al [26] used advanced methods for tracking app energy consumption in real time, offering insights into the variables that drive energy use. They developed a real-time energy monitoring framework that captures detailed energy use data at the component level, enabling developers to identify energy hotspots within the app. This approach allows for more targeted energy optimization strategies, focusing on the most energy-intensive components and interactions.

Methods

Ethical Considerations

The approach we took was a quantitative analysis study measuring energy use in popular US-based health and nutrition apps. This study did not require ethics board approval as it involved the quantitative analysis of publicly available data related to the energy consumption of mHealth apps. No human subjects were directly involved, and no personal or sensitive data were collected during the study. This approach aligns with the institutional guidelines and adheres to regional and local policies regarding research involving nonhuman subjects, ensuring that all analyses remain within ethical boundaries as per the existing frameworks.

Selection of mHealth Apps

The selection of mHealth apps was based on Almasri and Gouveia's [27] criteria and insights from Kelényi et al [20].

Popularity and User Base

Apps with a vast number of downloads and positive feedback from users were selected.

Functional Complexity

Apps featuring a spectrum of functionalities were selected, from the simplest to the most complex, aiming to understand how different features influence energy use following the concept by Isuwa et al [18].

Energy Consumption Potential

Apps known or suspected to be high on energy use, including features such as continuous data syncing or GPS tracking, based on preliminary evaluations and what developers have documented (Benkhelifa et al [19]), were selected.

Measurement of Energy Consumption

Overview

Ergasheva et al [25] and Pathak et al [26] both introduced metrics such as energy-per-function and energy-per-interaction, offering detailed insights into app energy efficiency by measuring energy use for specific functions and user interactions. The process unfolded in 3 key steps as given below.

Baseline Measurement

We first set a baseline for energy consumption for each app when it was not in use, providing a benchmark for comparing energy use during more active scenarios.

Feature-Specific Scenarios

We then measured energy use in scenarios that trigger specific features of the apps such as logging meals or syncing with wearable technology. This step was crucial for pinpointing and measuring the energy footprint of distinct functionalities within the apps. The flowchart depicted in Figure 1 outlines the sequential steps taken from the collection of energy consumption data to the identification of high-impact features and a review of the data collection methodology. Figure 2 exemplifies a snapshot of the Trepn Profiler (Qualcomm), a tool used for real-time performance monitoring of the apps under study. The graphs depict central processing unit frequency and graphics processing unit load over a session, demonstrating how various app features and user interactions can influence energy consumption. Such detailed monitoring is indispensable for identifying high-energy-demand periods, thereby informing our strategies for app optimization.

Figure 1. Flowchart of the energy consumption analysis methodology.

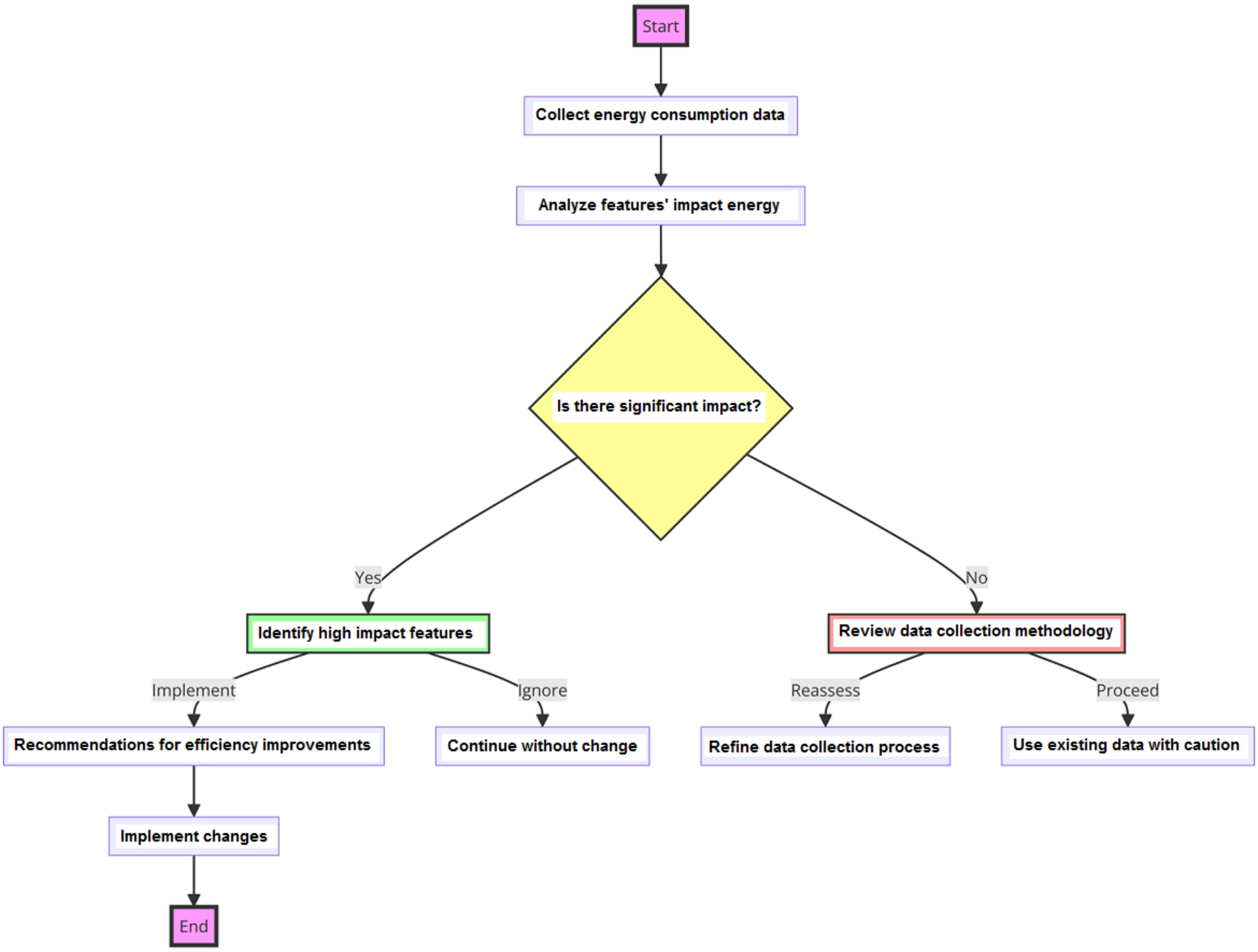
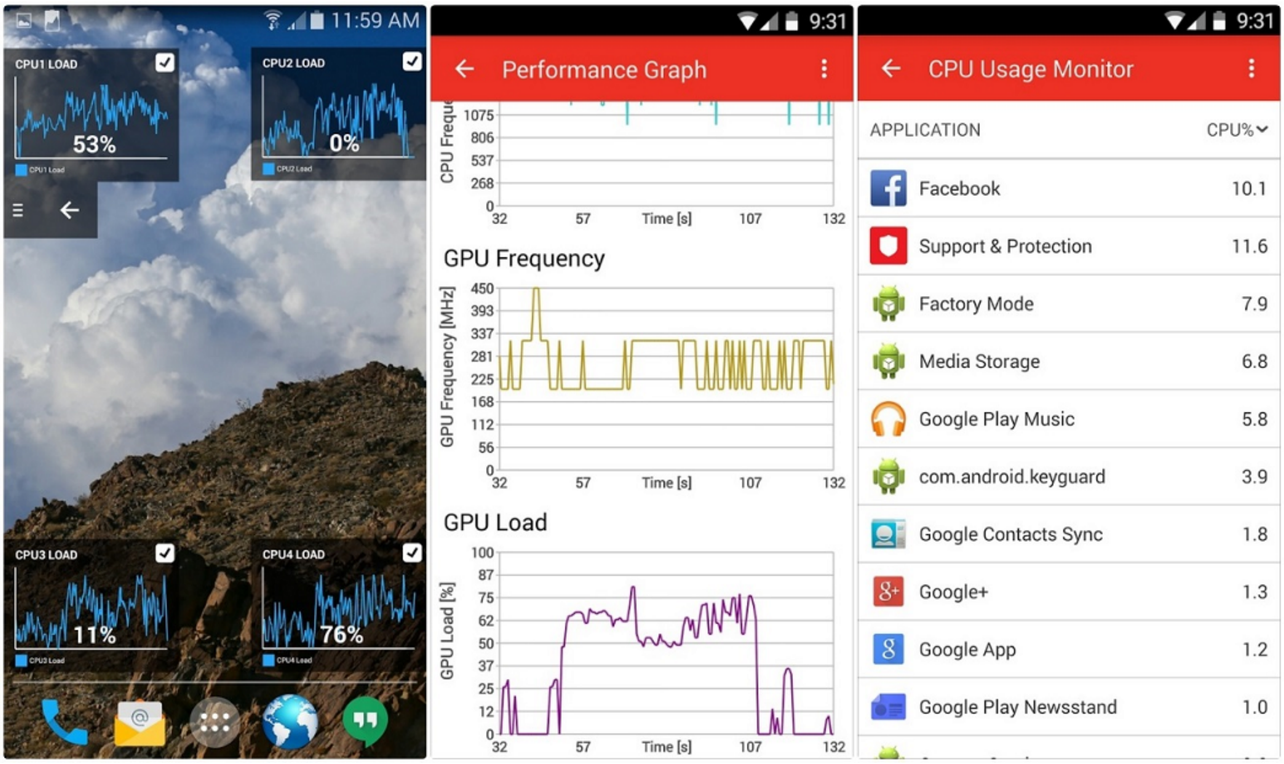


Figure 2. Example of real-time performance monitoring using Trepn Profiler. CPU: central processing unit; GPU: graphics processing unit.



User Interaction Patterns

Finally, by simulating a range of real-life user interactions, from minimal to extensive use, we were able to depict how different use patterns impact the app's energy consumption.

Data Analysis

Overview

This study examined data on energy consumption and the potential of behavior change interventions to cut energy use, drawing inspiration from Internet of Things–enabled tactics designed to boost energy efficiency consistent with recent findings that underscore the effectiveness of behavioral strategies in curbing energy use across different contexts [28,29]. We conducted our statistical analysis using MATLAB software (MathWorks), and our analysis approach included descriptive statistics, comparative analysis, and regression analysis.

Descriptive Statistics

Energy consumption for each app across various scenarios was quantified in milliwatt-hour (mWh) using real-time energy monitoring tools. This analysis provided insights into energy consumption patterns and fluctuations.

Comparative Analysis

ANOVAs were used to compare energy data across different apps and scenarios, identifying significant differences attributable to app features or user interactions.

Regression Analysis

Building regression models enabled us to measure how certain factors, such as how often notifications pop up or GPS tracking is used, impact energy use. This analysis helps understand the layers of what drives energy consumption in mHealth apps.

Results

Energy Consumption Patterns

Table 2 maps out a comparative analysis of the 10 popular mHealth apps. It elucidates each app's market presence, user experience, and estimated energy consumption, laying a foundation for understanding the interplay between app features and energy efficiency. We found a notable range in how much energy these apps use, with some consuming up to 3 times more energy than their counterparts in similar conditions. This disparity stemmed from various factors, such as the complexity of the app's features, how efficiently it runs in the background, and how often and in what ways users interact with the app. In our analysis, we conducted a descriptive statistical examination to highlight the energy consumption patterns of the selected apps. Our findings reveal a variance in average energy use, with apps consuming between 150 mWh and 310 mWh under typical use scenarios. The SD in energy consumption underscores the variability, ranging from 15 mWh to 31 mWh, which is indicative of how user interactions and background processing contribute to energy expenditure. The minimum and maximum energy use values further allocate the range of energy efficiency among these apps, from 135 mWh to 341 mWh, reflecting the impact of app features and optimization on battery life.

Table 2. Detailed comparative analysis of top mobile health apps.

App name	Popularity (downloads)	User review (out of 5)	Average energy use (CPU[a])	Feature complexity	Integration with wearables	Notification frequency	Support for multiple diets
Yummly	>5M[b]	4.3	Low	Medium	No	Low	Yes
My Plate Calorie Counter	>5M	4	Medium	Medium	No	Medium	Yes
MyFitnessPal	>50M	4.6	High	High	Yes	Medium	No
Spokin	>50K[c]	3.9	Low	Low	No	Low	No
PlateJoy	>500K	4.2	Low	Low	No	Low	Yes
Ovia	>1M	4.1	Medium	Low	No	Medium	No
Lifesum	>10M	4.5	Medium	High	Yes	High	Yes
Noom	>10M	4.4	High	High	Yes	High	No
Calorie Counter	>10M	4.7	High	High	Yes	High	Yes
Ate Food Journal	>100K	4.2	Low	Low	No	Low	No

[a]CPU: central processing unit.

[b]M: million or more.

[c]K: hundred thousand or more.

Table 3 illustrates the comparative energy consumption patterns of 10 popular mHealth apps under various use scenarios. This visualization underscores the substantial disparities in energy use, driven by factors such as app feature complexity, background processing efficiency, and user interaction methods. It highlights the critical need for targeted energy optimization strategies to mitigate the significant energy demands of feature-rich apps. The apps that demanded the most energy were those packed with sophisticated features such as live syncing with wearable tech and ongoing background updates. On the flip side, the more straightforward apps that relied on manual inputs and had fewer background processes were much kinder to battery life.

Table 3. Comparative energy consumption patterns of 10 popular mobile health apps.

App name	Use scenario and energy consumption (milliwatt-hour)		
	Baseline	GPS use	High use
Ovia	1	4	22
Calorie Counter	3	5	18
My Plate Calorie Counter	5	8	16
Yummly	5	10	20
Lifesum	4	9	17
Noom	3	7	15
My Fitness Pal	2	5	23
Ate Food Journal	1	6	13
PlateJoy	2	8	11
Spokin	3	6	10

Additionally, our research highlights how the use of notifications and alerts plays a significant role in energy consumption. Apps that leaned heavily on notifications to keep users engaged were more likely to use more energy, primarily due to the frequent lighting up of screens and the data exchanged over network services. It was found that on average, using an mHealth app for an hour each day could drain approximately 15% to 20% of a smartphone’s battery life, depending on the app’s complexity and background activity. This observation points to the critical need for fine-tuning notification strategies, and finding a sweet spot that maintains user interest without unnecessarily draining the battery.

Impact of App Features on Energy Use

Our findings suggest that certain app functions are linked to the amount of energy they use. GPS tracking—used for recording outdoor meals or activities—along with frequent data synchronization and sophisticated graphical interfaces, emerged as the main factors driving up energy consumption. GPS tracking was particularly notable for its high energy use, relying heavily on constant location services and data exchange. The research also brings to light how user behavior affects energy use, specifically how long and how often people use the apps. Apps designed to keep users engaged for longer periods, whether through fun gamification features or detailed dietary logging, were seen to consume more energy overall. This finding points to the need for thoughtfully crafting user engagement methods to avoid unnecessary energy consumption. Our findings are consistent with existing research. Choi et al [7] reaffirmed the considerable effect of screen brightness and network use on energy consumption. Our findings highlight the intense energy demands of certain features in mHealth apps, such as GPS tracking and frequent synchronization, areas not deeply researched by previous studies. While earlier research underscored the significance of hardware and system optimizations for lowering energy use, our research emphasizes the paramount role of optimizations at the app level where user-centered design and behavior will be critical. By concentrating on the architecture and capabilities of mHealth apps, developers have a profound opportunity to enhance the energy efficiency of their creations, as well as by involving both users and practitioners alike who can guide what features remain paramount for impactful and sustainable technology practice.

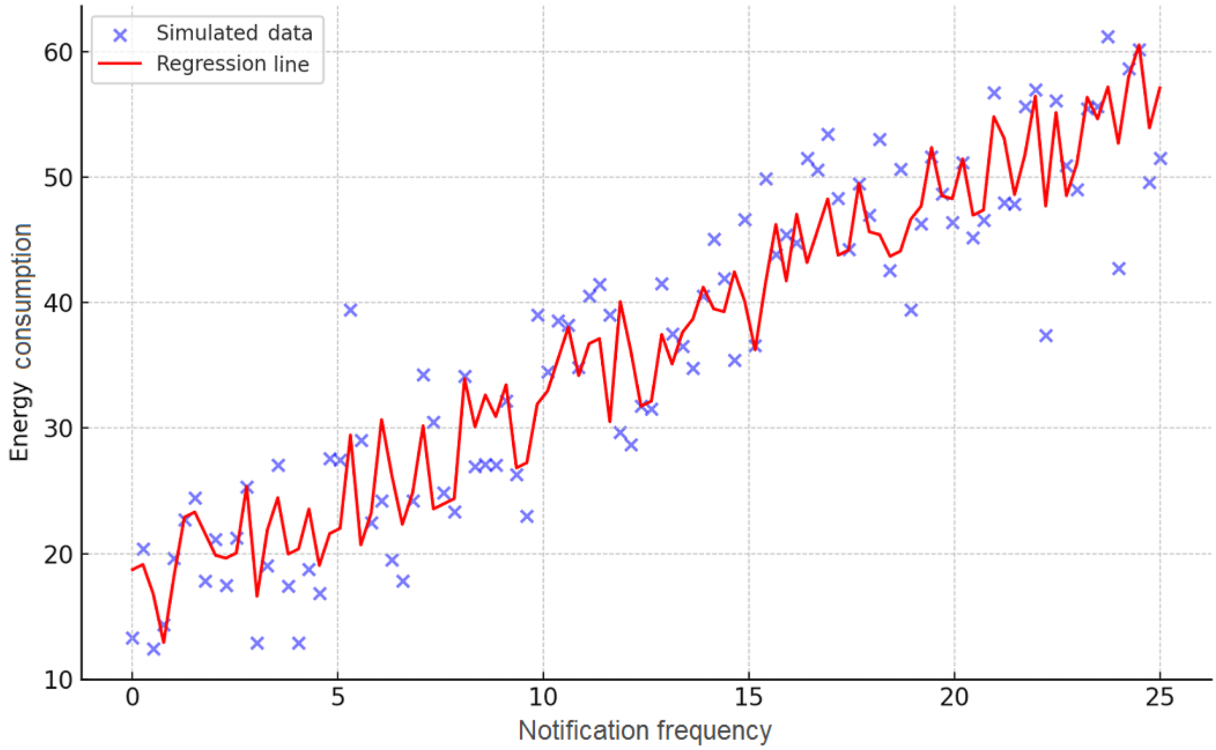
Insights From User Engagement and Energy Efficiency

Our regression analysis showed the relationship between user engagement, app features, and energy consumption. The analysis featured a key insight: while increased user engagement typically leads to higher energy consumption, strategic app design can mitigate this effect. Specifically, our findings highlight how certain app features, such as notification frequency, GPS use, and complexity level, influence the energy efficiency of mHealth apps. Assuming a linear relationship between these factors and energy consumption, our regression model is represented by the equation:

$$\text{Energy consumption} = \beta_0 + \beta_1 \times \text{notification frequency} + \beta_2 \times \text{GPS use} + \beta_3 \times \text{app complexity} + \varepsilon$$

where β_0 is the intercept, indicating the baseline energy consumption in the absence of the examined features. β_1, β_2, and β_3 are coefficients quantifying the impact of notification frequency, GPS use, and app complexity on energy consumption, respectively, and ε represents the error term, accounting for variability not explained by the model. Figure 3 clarifies this relationship, presenting a regression analysis that demonstrates the impact of notification frequency on energy consumption.

Figure 3. Regression analysis of notification frequency on energy consumption.



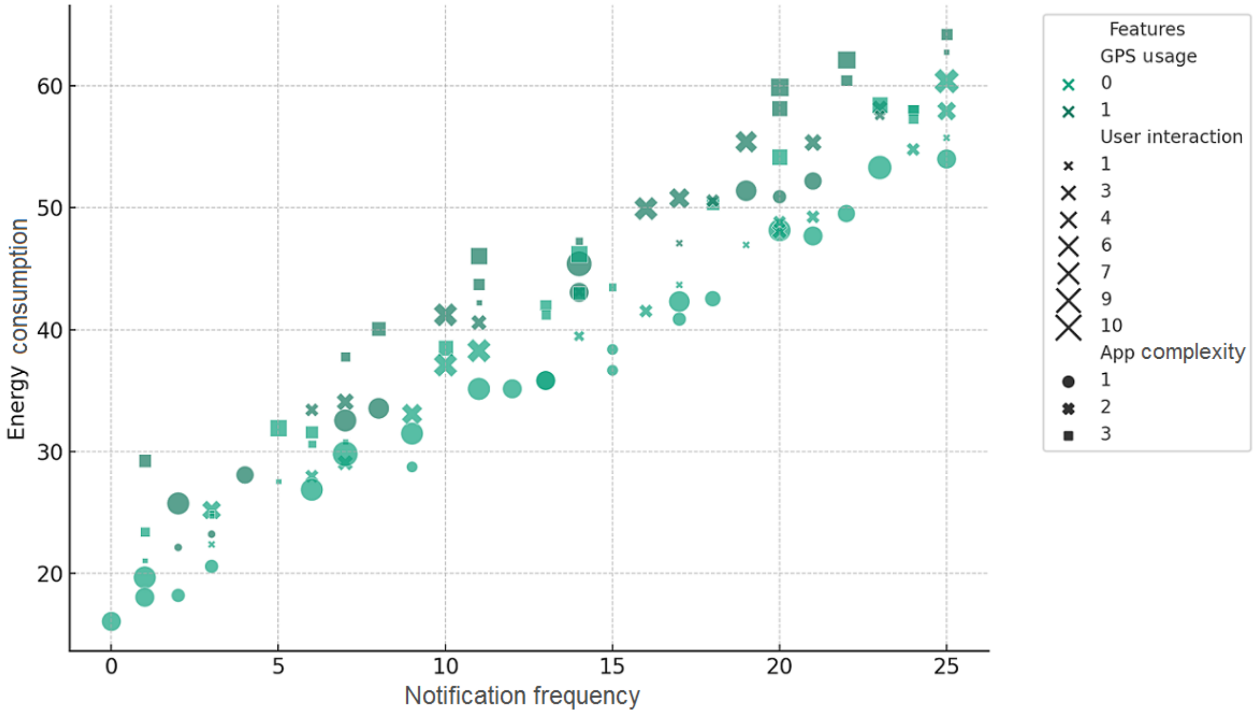
Our simulated analysis yielded the following equation:

$$\text{Energy consumption} = 9.55 + 1.62 \times \text{notification frequency}$$

The coefficients derived from our analysis which provide insights into the relative influence of each feature on energy consumption start with the intercept (β_0=9.55) representing the baseline energy consumption. Then, each unit increase in notification frequency (β_1=1.62) corresponds to a 1.62-unit increase in energy consumption, emphasizing its significant role. Besides, GPS use (β_2=5.00) suggests that activating GPS functionality contributes an additional 5 units to energy consumption, highlighting the substantial energy demand of location services. Finally, the app complexity (β_3=3.00) shows that higher complexity levels increase energy consumption by 3 units, indicating the impact of advanced features and functionalities. The statistical significance of each coefficient was evaluated through *P* values, confirming the strength of our findings. The *P* value for notification frequency is .01, indicating a highly significant relationship with energy consumption. The *P* value for GPS use is .05, suggesting an impact on energy consumption at the 5% level. The *P* value for app complexity is .03, demonstrating its significant effect on energy consumption.

For example, apps that adopt flexible synchronization schedules and energy-conscious notification strategies are more likely to succeed in keeping users engaged without a corresponding spike in energy use. This finding is also critical for developers aiming to refine the user experience while staying true to the principles of energy efficiency. Furthermore, we include an “integrated analysis approach” to examine the compounded effects of app features on energy consumption. This analysis builds upon our original regression model by openly considering the interactions between different app functionalities and their collective impact on energy use. To convey this concept, in Figure 4, we present an integrated analysis, contrasting the specific energy demands of app features against user interaction patterns, highlighting the potential for energy optimization. This visualization highlights the synergy between GPS use, notification frequency, app complexity, and their aggregate effect on energy use. Through this analysis, we aim to guide developers in identifying which combinations of features escalate energy demand and how thoughtful integration can mitigate such effects, fostering more energy-efficient app designs.

Figure 4. Integrated analysis of feature-specific energy consumption and user interaction patterns.



Discussion

Principal Findings

Finding the right equilibrium among app functionality, user satisfaction, and energy efficiency is critical for sustainable practice. Our research analysis enriches our understanding of the nuanced relationships within mHealth app use and highlights the broader consequences and opportunities for innovation in the digital health domain, underscoring the importance of a balanced approach to app development that honors both human and environmental considerations.

The diverse energy use among various mHealth apps, especially those dedicated to diet and meal tracking, reveals how an app is built and how users interact with it. Features, such as GPS tracking and constant data updates, significantly increase energy consumption, emphasizing the urgency for app creators to weave energy efficiency into the fabric of app development. User behaviors are also critical—including how often users interact with the app, respond to notifications, or use specific features. Behavior change modalities also need to be introduced to address user habits and smarter app configurations.

Our study findings bring forth the question of the need to identify modalities or to balance between incorporating features that boost user engagement and satisfaction and the essential task of reducing energy consumption. Such modalities are important to prolong battery life and lessen the ecological impact of mHealth app use.

Implications for Developers and Users

This study highlights for developers the crucial role of weaving energy efficiency into every stage of the app development cycle. This means going beyond just streamlining code and choosing low-power software development code libraries. It also means crafting app features and user interactions in ways that naturally lead to less energy use. Developers are urged to embrace smart algorithms that dynamically tweak app functions according to real-time energy use and battery status, ensuring the apps are as energy-efficient as possible without sacrificing the quality of the user experience.

On the user side, the research points out how a little awareness about how apps are set up and used can go a long way in reducing energy consumption. Users have the power to drive energy savings by tweaking their app settings, such as reducing how often apps search for new content or turning off unnecessary background activities. It is paramount to use apps that are designed from the ground up to be energy conscious with efficient battery life and for the purpose of encouraging a greener approach to leveraging digital health tools.

Broader Implications for Digital Health Technology

This research adds a valuable perspective to the conversation about making digital health technologies more sustainable, emphasizing the collective responsibility of consumption and production, including developers, users, health care professionals, and stakeholders, to put energy efficiency at the forefront. As mHealth apps play a more prominent role in enhancing health and nutrition outcomes and managing diseases, understanding and optimizing their energy use becomes essential for ensuring digital solutions can grow sustainably, and consumers and producers are both responsible for sustainable practice as well.

Moreover, the insights gathered here highlight the exciting possibilities of interdisciplinary studies that merge knowledge from software engineering, behavioral science, and environmental sustainability. This approach could lead to the creation of comprehensive guidelines and best practices for crafting mHealth apps that are not only effective but also energy efficient. By working together across fields, there is a tremendous opportunity to drive forward app innovations that serve the dual purpose of advancing health care while respecting our planet.

Conclusion

The mHealth apps within the mHealth sector consume energy, especially when app functionalities are governed by how we interact with these apps. It is a challenge for developers and users to find the right mix of features that drive engagement and health and nutrition benefits while also becoming cognizant of reducing energy use.

For developers, this means weaving energy efficiency more deeply into the fabric of app creation, from concept through to coding. This can be done by embracing flexible technologies and applying forward-thinking design philosophies that marry efficacy with energy savings. For users, it is about becoming more aware of how the choices they make in app settings and their daily use can affect energy consumption, moving toward a more conscious and deliberate use of these digital tools.

Acknowledgments

The authors would like to acknowledge the support of Prince Sultan University for paying the article processing charges of this publication and for their support.

Conflicts of Interest

None declared.

References

1. Bhaskar MS, Almakhles D, Padmanaban S, Ionel DM, Blaabjerg F, He J, et al. Investigation of a transistor clamped T-type multilevel H-bridge inverter with inverted double reference single carrier PWM technique for renewable energy applications. IEEE Access 2020;8:161787-161804 [FREE Full text] [doi: 10.1109/access.2020.3020625]
2. Muthanna MSA, Muthanna A, Nguyen TN, Alshahrani A, Abd El-Latif AA. Towards optimal positioning and energy-efficient UAV path scheduling in IoT applications. Comput Commun 2022;191:145-160 [FREE Full text] [doi: 10.1016/j.comcom.2022.04.029]
3. Ashfaq T, Khalid MI, Ali G, Affendi ME, Iqbal J, Hussain S, et al. An efficient and secure energy trading approach with machine learning technique and consortium blockchain. Sensors (Basel) 2022;22(19):7263 [FREE Full text] [doi: 10.3390/s22197263] [Medline: 36236363]
4. Chandani A, Wagholikar S, Prakash O. A bibliometric analysis of green computing. In: Xie J, Kaiser MS, Rathore VS, editors. Information and Communication Technology for Competitive Strategies (ICTCS 2021): Intelligent Strategies for ICT. Springer: Springer Nature; 2023:547-557.
5. Switzer J, Siu E, Ramesh S, Hu R, Zadorian E, Kastner R. Renée: new life for old phones. IEEE Embedded Syst Lett 2022;14(3):135-138 [FREE Full text] [doi: 10.1109/les.2022.3147409]
6. Muchinenyika SH, Muyingi HNN. An analysis of screen resolution effects on battery endurance: a case of smartphones. 2022 Presented at: 2022 IEEE 9th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT); May 28-30, 2022; Hammamet, Tunisia p. 356-360 URL: https://ieeexplore.ieee.org/abstract/document/9875592?casa_token=l9Pq-gdVmZ0AAAAA:nUzSENDhR1EmpEzMUPBuJr7lxSsSvP9_5HLUjrLJRkRhhh3tEPLshJNFOtJBw1A_IYFNpRVg [doi: 10.1109/setit54465.2022.9875592]
7. Choi Y, Park S, Jeon S, Ha R, Cha H. Optimizing energy consumption of mobile games. IEEE Trans Mob Comput 2022;21(10):3744-3756 [FREE Full text] [doi: 10.1109/tmc.2021.3058381]

8. Pop RA, Dabija DC, Pelău C, Dinu V. Usage intentions, attitudes, and behaviors towards energy-efficient applications during the COVID-19 pandemic. J Bus Econ Manag 2022;23(3):668-689 [FREE Full text] [doi: 10.3846/jbem.2022.16959]
9. Almasri AM, Gouveia LB. The evolution of energy requirements of smartphones based on user behaviour and implications of the COVID-19 era. Int J Electr Comput Eng 2021;11(3):2423-2431 [FREE Full text] [doi: 10.11591/ijece.v11i3.pp2423-2431]
10. Awais Q, Adeel M, Jawad H, Adnan K. Evaluating the impact of design pattern usage on energy consumption of applications for mobile platform. Appl Comput Syst 2021 Jun 04;26(1):1-11. [doi: 10.2478/acss-2021-0001]
11. Sahar H, Bangash AA, Beg MO. Towards energy aware object-oriented development of android applications. Sustain Comput Inform Syst 2019;21:28-46 [FREE Full text] [doi: 10.1016/j.suscom.2018.10.005]
12. Rahman MS, Ko M, Warren J, Carpenter D. Healthcare technology self-efficacy (HTSE) and its influence on individual attitude: an empirical study. Comput Human Behav 2016;58:12-24 [FREE Full text] [doi: 10.1016/j.chb.2015.12.016]
13. Sanakulov N, Karjaluoto H. Consumer adoption of mobile technologies: a literature review. Int J Mob Commun 2015;13(3):244-275 [FREE Full text] [doi: 10.1504/ijmc.2015.069120]
14. Rahimi N, Jetter A. Explaining health technology adoption: past, present, future. 2015 Presented at: 2015 Portland International Conference on Management of Engineering and Technology (PICMET); August 02-06, 2015; Portland, OR, USA p. 2465-2495. [doi: 10.1109/picmet.2015.7273030]
15. Goddard RD, Salloum SJ. Collective efficacy beliefs, organizational excellence, and leadership. In: Spreitzer GM, Cameron KS, editors. The Oxford Handbook of Positive Organizational Scholarship. Oxford: Oxford University Press; 2012:642-650.
16. Al Nidawi HSA, Koh TW, Dawood K, Khaleel A. Energy consumption patterns of mobile applications in android platform: a systematic literature review. J Theor Appl Inf Technol 2017;95(24):6776-6787 [FREE Full text]
17. Acer UG, Mashhadi A, Forlivesi C, Kawsar F. Energy efficient scheduling for mobile push notifications. EAI Endorsed Trans Energy Web 2015;2(5):100-109. [doi: 10.4108/eai.22-7-2015.2260067]
18. Isuwa S, Amos D, Singh AK, Al-Hashimi BM, Merrett GV. Maximising mobile user experience through self-adaptive content-and ambient-aware display brightness scaling. J Syst Architect 2023;145:103023 [FREE Full text] [doi: 10.1016/j.sysarc.2023.103023]
19. Benkhelifa E, Welsh T, Tawalbeh L, Jararweh Y, Al-Ayyoub M. Leveraging software-defined-networking for energy optimisation in mobile-cloud-computing. Procedia Comput Sci 2016;94:479-484 [FREE Full text] [doi: 10.1016/j.procs.2016.08.074]
20. Kelényi I, Nurminen JK, Siekkinen M, Lengyel L. Supporting energy-efficient mobile application development with model-driven code generation. In: van Do T, Le Thi HA, Nguyen NT, editors. Advanced Computational Methods for Knowledge Engineering: Proceedings of the 2nd International Conference on Computer Science, Applied Mathematics and Applications (ICCSAMA 2014). Cham: Springer International Publishing; 2014:143-156.
21. Porter C. Energy efficiency in high-performance computing: balancing speed and sustainability. NVIDIA Technical Blog. 2023. URL: https://developer.nvidia.com/blog/energy-efficiency-in-high-performance-computing-balancing-speed-and-sustainability/ [accessed 2024-02-29]
22. Ericsson. Intelligent sustainability: the role of AI in energy consumption, management and new revenues. Ericsson Blog. 2023. URL: https://www.ericsson.com/en/blog/2023/6/intelligent-sustainability-ai-energy-consumption [accessed 2024-02-25]
23. Khan MU, Abbas S, Lee SUJ, Abbas A. Measuring power consumption in mobile devices for energy sustainable app development: a comparative study and challenges. Sustain Comput Inform Syst 2021;31:100589 [FREE Full text]
24. Ciman M, Gaggi O. Evaluating impact of cross-platform frameworks in energy consumption of mobile applications. 2014 Presented at: 10th International Conference on Web Information Systems and Technologies (WEBIST-2014); 3-5 April, 2014; Barcelona, Spain URL: https://www.scitepress.org/Papers/2014/48576/48576.pdf [doi: 10.5220/0004857604230431]
25. Ergasheva S, Khomyakov I, Kruglov A, Succil G. Metrics of energy consumption in software systems: a systematic literature review. IOP Conf Ser Earth Environ Sci 2020;431(1):012051 [FREE Full text] [doi: 10.1088/1755-1315/431/1/012051]
26. Pathak A, Jindal A, Hu YC, Midkiff SP. What is keeping my phone awake?: characterizing and detecting no-sleep energy bugs in smartphone apps. 2012 Presented at: MobiSys'12: The 10th International Conference on Mobile Systems, Applications, and Services; June 25-29, 2012; Low Wood Bay Lake District UK p. 267-280. [doi: 10.1145/2307636.2307661]
27. Almasri AM, Gouveia LB. Star-rating evaluation model for rating the energy-efficiency level of android google play apps. Int J Electr Comput Eng 2021;11(2):1599-1612 [FREE Full text] [doi: 10.11591/ijece.v11i2.pp1599-1612]
28. Ramallo-González AP, Bardaki C, Kotsopoulos D, Tomat V, Vidal AG, Ruiz PJF, et al. Reducing energy consumption in the workplace via IoT-allowed behavioural change interventions. Buildings 2022;12(6):708 [FREE Full text] [doi: 10.3390/buildings12060708]
29. Varvello M, Livshits B. Shedding (Some) light on mobile browsers energy consumption. Traffic Monitoring and Analysis. 2021. URL: https://tma.ifip.org/2021/wp-content/uploads/sites/10/2021/08/tma2021-paper18.pdf [accessed 2024-04-30]

Abbreviations

AI: artificial intelligence
mHealth: mobile health

mWh: milliwatt-hour

Edited by G Eysenbach, A Kushniruk, E Borycki; submitted 12.03.24; peer-reviewed by M Ayiad, Y Boukili, G Nogueira, N Russ; accepted 07.04.24; published 10.05.24.

Please cite as:
Almasri A, El-Kour TY, Silva L, Abdulfattah Y
Evaluating the Energy Efficiency of Popular US Smartphone Health Care Apps: Comparative Analysis Study Toward Sustainable Health and Nutrition Apps Practices
JMIR Hum Factors 2024;11:e58311
URL: https://humanfactors.jmir.org/2024/1/e58311
doi:10.2196/58311
PMID:38729624

Original Paper

Accessibility, Relevance, and Impact of a Symptom Monitoring Tool for Home Hospice Care: Theory Elaboration and Qualitative Assessment

Karla T Washington[1], PhD; Debra Parker Oliver[1,2], PhD; Allison K Donehower[1,3], MS; Patrick White[1,4], MD, PhD; Jacquelyn J Benson[1], PhD; Patrick G Lyons[5,6], MSc, MD; George Demiris[7,8], PhD

[1]Department of Medicine, School of Medicine, Washington University in St. Louis, St. Louis, MO, United States
[2]Goldfarb School of Nursing, Barnes-Jewish College, St. Louis, MO, United States
[3]Department of Human Development and Family Science, College of Education and Human Development, University of Missouri, Columbia, MO, United States
[4]BJC Hospice, St. Louis, MO, United States
[5]Department of Medicine, School of Medicine, Oregon Health & Science University, Portland, OR, United States
[6]Department of Medical Informatics and Clinical Epidemiology, School of Medicine, Oregon Health & Science University, Portland, OR, United States
[7]Department of Biostatistics, Epidemiology, & Informatics, School of Medicine, University of Pennsylvania, Philadelphia, PA, United States
[8]Department of Biobehavioral Health Sciences, School of Nursing, University of Pennsylvania, Philadelphia, PA, United States

Corresponding Author:
Karla T Washington, PhD
Department of Medicine
School of Medicine
Washington University in St. Louis
Mid Campus Center
4590 Children's Place, Mailstop 90-29-931
St. Louis, MO, 63110
United States
Phone: 1 (314) 285 0905
Email: kwashington@wustl.edu

Abstract

Background: Early users found Engagement and Visualization to Improve Symptoms in Oncology Care (ENVISION), a web-based application designed to improve home management of hospice patients' symptoms and support patients' and family caregivers' well-being, to be generally useful and easy to use. However, they also raised concerns about potential challenges users with limited technological proficiency might experience.

Objective: We sought to concurrently accomplish two interrelated study aims: (1) to develop a conceptual framework of digital inclusivity for health information systems and (2) to apply the framework in evaluating the digital inclusivity of the ENVISION application.

Methods: We engaged ENVISION users (N=34) in a qualitative study in which data were collected via direct observation, think-aloud techniques, and responses to open-ended queries. Data were analyzed via theory elaboration and basic qualitative description.

Results: Accessibility, relevance, and impact were identified as 3 essential considerations in evaluating a health system's digital inclusivity. Study findings generally supported ENVISION's digital inclusivity, particularly concerning its perceived relevance to the work of family caregivers and hospice clinicians and its potentially positive impact on symptom management and quality of life. Limitations to ENVISION's digital inclusivity centered around issues of accessibility, particularly operability among individuals with limited technological knowledge and skills.

Conclusions: The Accessibility, Relevance, and Impact conceptual framework of digital inclusivity for health information systems can help identify opportunities to strengthen the digital inclusivity of tools, such as ENVISION, intended for use by a broad and diverse range of users.

(JMIR Hum Factors 2024;11:e51789) doi:10.2196/51789

KEYWORDS

caregivers; home care services; hospice care; signs and symptoms; technology; mobile phone

Introduction

Background

Hospice is a health care delivery model and a philosophy of care focused on reducing pain and promoting quality of life among patients who are terminally ill and their families [1]. In the United States, hospice care is most often provided in patients' homes [2]. While more intensive staffing is available during acute medical crises, routine home hospice care consists of only periodic visits from nurses, nursing aides, social workers, chaplains, and others operating under the direction of a hospice physician [3]. A total of 3- to 4-hour-long weekly home visits may be typical for an established patient. Thus, responsibility for the overwhelming majority of home hospice care falls to patients' family members and friends (referred to as *family caregivers*), who are typically unpaid and often lack formal health care training [4-6].

Hospice family caregivers are commonly tasked with in-home management of patients' symptoms, including pain, shortness of breath, anxiety, and fatigue. Recent population-based research indicates that >78% of family caregivers who assist with symptom management in the last month of a patient's life report difficulty doing so [7]. These findings are consistent with those of numerous other studies highlighting the reality that symptom management challenges are a significant source of stress for many hospice family caregivers [8-11]. These challenges, coupled with lack of a standardized processes for real-time symptom reporting and monitoring in home hospice care, commonly result in suboptimal home management of patients' symptoms [12].

Engagement and Visualization to Improve Symptoms in Oncology Care

Engagement and Visualization to Improve Symptoms in Oncology Care (ENVISION) is a secure, web-based application designed to improve home management of hospice patients' symptoms and support patients' and family caregivers' well-being by improving the exchange of information between family caregivers and hospice clinicians [13]. It uses daily symptom and well-being data entered on the internet by family caregivers to create simple visualizations summarized in a patient and caregiver scorecard (Figure 1), allowing hospice clinicians to quickly identify areas of concern. These scorecards are displayed during biweekly hospice interdisciplinary team meetings and are available on demand to hospice clinicians outside of regularly scheduled meetings. A workflow diagram illustrating ENVISION's use is provided in Figure 2. Optional views, including longitudinal graphs of individual or combined symptoms, are also available to clinician users. Although ENVISION was initially developed specifically for advanced cancer care, its use has been expanded to include care for hospice-eligible individuals experiencing any life-limiting illness.

Figure 1. Mobile version of the application Engagement and Visualization to Improve Symptoms in Oncology Care (ENVISION), showing a sample patient and caregiver scorecard.

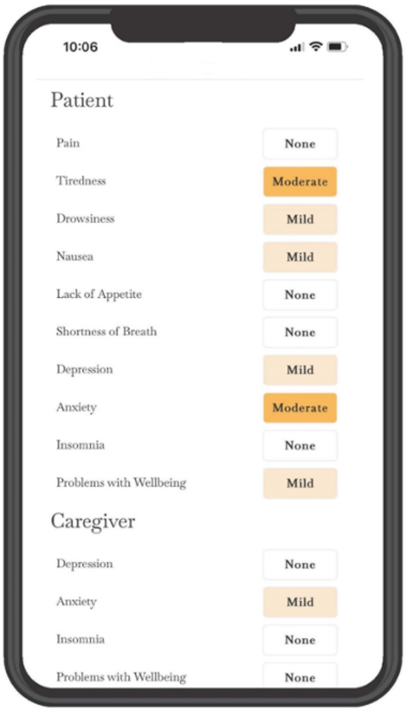
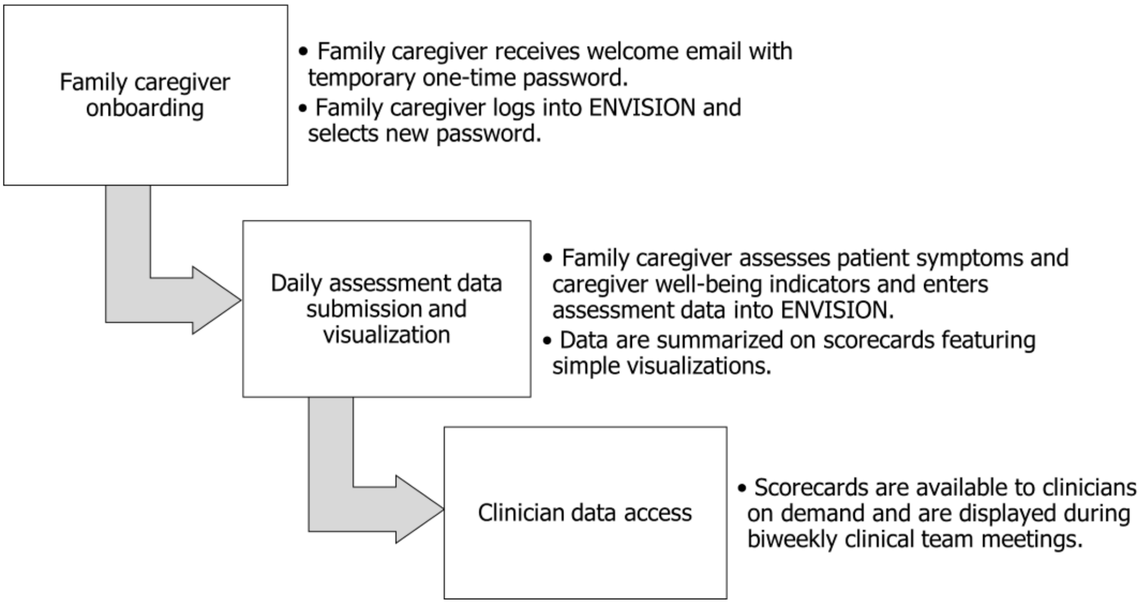


Figure 2. Engagement and Visualization to Improve Symptoms in Oncology Care (ENVISION) workflow.



Digital Inclusivity

ENVISION was created over several years with significant involvement of hospice family caregivers, clinicians, and administrators. While early research broadly supported ENVISION's usefulness and ease of use, it also raised concerns about potential barriers to use that might be experienced by family caregivers with limited technological skills or resources [13]. These concerns echo those voiced as part of an ongoing discussion in health care regarding digital inclusivity [14], broadly defined as "the ability of individuals and groups to access and use information and communication technologies [15]."

Digital inclusivity, particularly as it pertains to digital health technologies, is a salient concern on multiple social levels [16]. Individually, users vary with regard to their level of digital literacy and their ability to personally take advantage of available technological resources [17]. For example, an individual may struggle to discern differences between trustworthy and untrustworthy sources of health information, may have functional limitations (for example, vision or hearing impairment), or may be unable to afford home internet access. Similarly, families and other social groups differ in their degree of collective technological resources, a reality evident when less technologically equipped individuals benefit from the digital knowledge and skills of family members and friends. A family member assisting a patient in accessing their health care portal would be one example [18]. Communities can also be considered more or less digitally inclusive based on the adequacy of the infrastructure (such as home broadband connectivity or public Wi-Fi networks) available to meet residents' technological needs [15].

Study Aims

Our initial aim in conducting the study described herein was to better understand early ENVISION users' concerns regarding the application's digital inclusivity. However, in planning our study, we struggled to identify an existing conceptual framework to guide our research, given our plan to explore digital inclusivity as a quality of an individual application (rather than, for example, a community). Thus, we added a second study aim: to engage ENVISION users in a process of theory elaboration, resulting in a conceptual framework of digital inclusivity for health information systems. In this way, we sought to inform future ENVISION enhancements while contributing to the broader emerging science of digital inclusivity in health care. Thus, our finalized study aims were as follows: (1) to develop a conceptual framework of digital inclusivity for health information systems and (2) to apply the framework in evaluating the digital inclusivity of the ENVISION application.

Methods

Setting, Participants, and Recruitment

As part of ENVISION's ongoing, iterative, user-centered design [19], we recruited hospice family caregivers and clinicians (nurses, physicians, social workers, and chaplains) to participate in a qualitative research study [20]. We partnered with the university's Institute of Clinical and Translational Sciences' Recruitment Enhancement Core to recruit hospice family caregivers via flyers, targeted email blasts, social media posts, and listing of the study on a public-facing website. Family caregivers were eligible for inclusion if they were aged ≥18, able to speak and read English, and current or former (within the prior year) family caregivers of a patient receiving services from a Medicare-certified US hospice agency. We recruited hospice clinicians via social media posts and email blasts from professional hospice organizations, targeted emails to prior research partners, and presentation of the study opportunity at scheduled meetings of hospice agencies with which we had established partnerships. Hospice clinicians were eligible for study inclusion if they were aged ≥18, able to speak and read English, and currently employed or affiliated with a Medicare-certified US hospice agency.

Data Collection

All consenting participants met online individually with a research team member for approximately 30 to 45 minutes via a university-managed, Health Insurance Portability and Accountability Act–compliant Zoom (Zoom Video Communications) account [21]. At the start of each call, the researcher provided assistance in using Zoom's screen share feature, which the researcher later used to observe the participants completing a series of structured tasks in the ENVISION application. In addition, the researcher provided instruction in the think-aloud technique [22,23], explaining that the participants would be asked to verbalize their thinking as they navigated the application and completed specific tasks. The researcher also informed the participants that they would be asked to answer a series of open-ended questions about their perceptions of the application and its potential use in hospice care. Finally, the researcher confirmed the participants' understanding and began recording the session with the participants' knowledge and permission.

Structured Tasks

During the recorded Zoom session, family caregivers were sent a personalized email with a brief welcome message, a link to the ENVISION website, and a temporary one-time password (an automatically generated alphanumeric string of characters). As their first observed task, family caregivers were asked to navigate to the ENVISION website, enter the site using their email address and one-time password, and choose a new password. They were then asked to recall a typical caregiving day and enter corresponding symptom and well-being data for the patient and themselves into the ENVISION application. Next, they were asked to navigate to the patient and caregiver scorecard, which summarized the data they had just entered, and answer questions that required them to interpret simple data visualizations (labeled rectangles filled with different shades of orange ranging from none or white to dark or bright orange to reflect greater symptom intensity). Finally, they were asked to exit the application. After completing these structured tasks, they were asked a series of open-ended questions, including, for example, "What made it easy to use ENVISION?" "What made it challenging?" and "Which symptom(s) would be most important to communicate to the hospice team? Why?"

Hospice clinicians were also observed navigating to the ENVISION website, entering the site using their email address and temporary password, and selecting a new password. Because the researcher had entered them into the system as a clinician when generating their welcome email, clinicians were taken to a screen that included a list of fictitious patients' names and medical record numbers upon logging into the system. When they reached this screen, clinicians were asked to navigate to a specific patient's information (this required them to locate and click on the patient's name, but they were not given these specific instructions). Clicking on the patient's name took them to a screen that included a daily patient and caregiver scorecard that featured visualizations similar to those shown to family caregivers. This page also included a simple line graph that allowed clinician users to view the intensity of one or more symptoms or well-being indicators over time by clicking a box next to the appropriate symptoms or indicators (users were not provided with these specific instructions). While on this page, clinicians were asked questions that required them to interpret the colored boxes on the patient and caregiver scorecard; select and deselect specific symptoms on the longitudinal graph; and interpret trends, including symptom co-occurrence over time, shown via the graphed data. Finally, they were asked to exit the system and answer a series of open-ended questions, including, for example, "How, if at all, would having [information provided via ENVISION] affect how you do your job?" and "Describe how you would access ENVISION. For example, would you use a desktop computer, tablet, or smartphone? Would you use the application from the hospice agency office, patients' homes, or elsewhere?"

Data Preparation

In preparation for data analysis, we contracted with a third-party service to transcribe audio files of participants' recorded Zoom sessions verbatim. We then imported the resulting transcripts into NVivo qualitative analysis software (Lumivero). Complete copies of all audio and video files of participants' recorded Zoom sessions and corresponding field notes were stored in a secure Box folder made available to all institutional review board–approved research team members throughout data analysis.

Data Analysis

Our analysis was broadly informed by the work of organizational management researchers Fisher and Aguinis [24], who described a process they referred to as theory elaboration, defined as "the process of conceptualizing and executing empirical research using preexisting conceptual ideas or a preliminary model as a basis for developing new theoretical insights by contrasting, specifying, or structuring theoretical constructs and relations to account for and explain empirical observations." As part of this process, we engaged in vertical contrasting, which entailed adapting an existing conceptual framework (described in detail in the next paragraph) developed for one level of analysis to examine a phenomenon at another level. In doing so, we sought to determine which aspects of the framework functioned similarly on both levels of analysis and which functioned differently. We also engaged in construct specification, seeking to refine the constructs articulated in the original framework and to introduce new constructs when the existing constructs failed to capture important aspects of the phenomenon under investigation (in our case, ENVISION's digital inclusivity). At times, this involved construct splitting, a process whereby we split existing constructs into more specific dimensions if more conceptual specificity was needed to capture important aspects of ENVISION's digital inclusivity. Finally, we engaged in structuring, or identifying relationships between and among constructs, remaining open to new relation structures.

To accomplish these analytic activities (ie, contrasting, specifying, and structuring), 2 researchers (KTW and AKD) first reviewed all study transcripts, video files, and field notes. They then met to develop an initial codebook based on the elements of an existing framework: *Building Digital Communities: A Framework for Action* [15], created by the Institute of Museum and Library Services, the University of

Washington Technology and Social Change Group, and the International City or County Management Association. As its name suggests, this framework was created to promote digital inclusivity at the community level. Consistent with this purpose, it articulated 13 principles for community-wide digital inclusivity, including access principles (which addressed community infrastructure needs), adoption principles (which focused on community members' facilitators and barriers to use), and application principles (which specified areas where deployment of digital technologies would be likely to enhance community members' lives). The original framework's principles and corresponding definitions are provided in Multimedia Appendix 1.

We originally envisioned development of the initial codebook as a relatively straightforward process in which most, if not all, constructs articulated in the original framework would be initially retained and then refined in subsequent analytic steps. However, it soon became apparent that some of the original principles had limited applicability in the context of an individual application and should, therefore, be de-emphasized or excluded in the early stages of our analysis. For example, the application principles outlined in the original framework identified specific community sectors, such as education and public safety, where the deployment of technologies was deemed likely to benefit community well-being. However, digital health tools are, by definition, intended for deployment in health care (and, in the case of ENVISION, more specifically in hospice care). Thus, we omitted them from the initial codebook, feeling confident they would neither enrich our understanding of ENVISION's digital inclusivity nor ultimately represent constructs comprising our adapted conceptual framework.

After completing the initial codebook, KTW and AKD independently coded approximately 15% of the study transcripts, consulting video recordings and field notes as needed for context or clarification of transcribed data. KTW and AKD then met to make substantive modifications to the codebook to enhance its goodness of fit with the data. Examples of changes made at this stage included specifying that *affordability* referred to ENVISION's initial and ongoing costs and should, thus, be relabeled as *affordability and sustainability* (construct specification) and dividing *design for inclusion* into *perceivability*, *operability*, and *comprehensibility* (construct splitting), as the available data suggested that these were conceptually meaningful distinctions. We then used the modified codebook to code the entire data set, meeting afterward to compare individual coding decisions (resolving discrepancies via discussion and arriving at consensus), finalize our code definitions, and group related codes into broader categories that comprised our resulting conceptual framework and shed light on ENVISION's strengths and limitations with regard to digital inclusivity.

Ethical Considerations

All research activities were reviewed and approved by the Washington University in St Louis Institutional Review Board (#202105172).

Individuals interested in study participation were provided with contact information for our study coordinator, who screened potential participants for eligibility, obtained verbal informed consent for those interested in participating, and coordinated all subsequent research activities including participant payments of US $40 sent via check to the mailing address of the participants' choice.

Results

Overview

A total of 34 individuals participated in our qualitative research study, enabling the concurrent achievement of 2 interrelated study aims: (1) to develop a conceptual framework of digital inclusivity for health information systems and (2) to evaluate the digital inclusivity of the ENVISION application (participant characteristics are summarized in Table 1). In the following sections, we present our study findings, beginning with a brief overview of our conceptual framework and its essential elements. We then provide an in-depth description of the framework, illustrating its specific constructs with examples from our evaluation of ENVISION's digital inclusivity.

Table 1. Summary of participant characteristics (N=34).

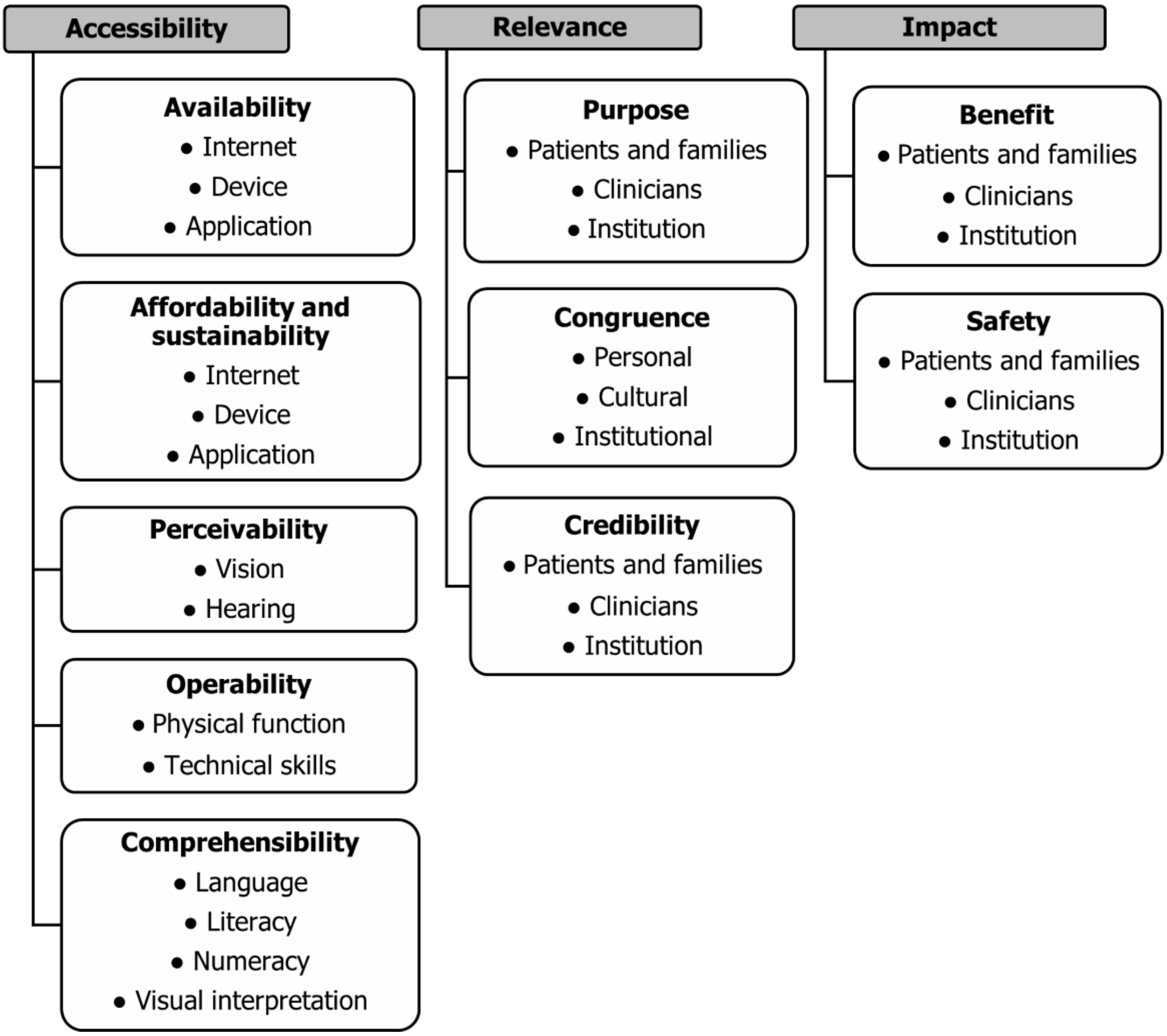
Characteristic	Family caregivers (n=10), n (%)	Hospice clinicians (n=24), n (%)
Age range (y)		
18-29	1 (10)	2 (8)
30-39	1 (10)	5 (21)
40-49	1 (10)	6 (25)
50-59	2 (20)	8 (33)
60-69	3 (30)	3 (13)
≥70	2 (20)	0 (0)
Gender		
Man	1 (10)	5 (21)
Woman	9 (90)	19 (79)
Race		
Black	3 (30)	0 (0)
White	7 (70)	24 (100)
Ethnicity		
Hispanic	0 (0)	1 (4)
Non-Hispanic	10 (100)	23 (96)
Relationship to patient		
Spouse or partner	2 (20)	N/A[a]
Adult child	5 (50)	N/A
Other	3 (30)	N/A
Highest formal education		
Some college or trade school	2 (20)	N/A
Associate's degree	2 (20)	N/A
Bachelor's degree	3 (30)	N/A
Graduate or professional degree	3 (30)	N/A
Profession		
Chaplain	N/A	5 (21)
Nurse	N/A	7 (29)
Other	N/A	1 (4)
Physician	N/A	3 (13)
Social worker	N/A	8 (33)
Professional experience (y)		
0-5	N/A	6 (25)
6-10	N/A	5 (21)
11-15	N/A	4 (17)
16-20	N/A	2 (8)
21-25	N/A	2 (8)
>25	N/A	5 (21)

[a]N/A: not applicable.

The Accessibility, Relevance, and Impact Conceptual Framework of Digital Inclusivity for Health Information Systems

Our analysis resulted in the development of a conceptual framework of digital inclusivity for health information systems that comprises 3 essential elements: accessibility, relevance, and impact (Figure 3). Per the Accessibility, Relevance, and Impact (ARI) framework, in evaluating a health information system's accessibility, the availability and affordability or sustainability of access to the internet (for web-based applications); necessary devices (eg, computers, smartphones, and tablets); and the application or system itself must be considered. Also relevant are a system's perceivability (the extent to which it can be used by individuals with different sensory abilities, such as visual or hearing impairments), operability (the extent to which it can be used by individuals with different physical abilities or technological proficiencies), and comprehensibility (the extent to which users can understand and accurately interpret the system's content). A health information system's relevance is also key to its digital inclusivity. Digitally inclusive systems are useful (ie, they fulfill a clear purpose); trustworthy (ie, they are viewed as credible); and aligned with users' values, beliefs, customs, and preferences (ie, they are congruent). Finally, the framework suggests that the evaluation of a health information system's digital inclusivity requires consideration of its impact, that is, the extent to which it improves (or would be expected to improve) users' lives (benefit) and the presence or absence of protection from web-based threats (eg, malware and data breaches) associated with the system's use (safety).

Figure 3. The Accessibility, Relevance, and Impact (ARI) conceptual framework of digital inclusivity for health information systems.



Evaluating ENVISION's Digital Inclusivity Using the ARI Framework

Accessibility

Data describing ENVISION's availability referenced the presence or lack of internet access, technological devices, or otherwise referred to potential users' ability to retrieve the application. Study participants' feedback and experiences related to availability were generally positive. Family caregivers either did not comment on ENVISION's availability or were positive in their responses; no family caregivers identified availability-related barriers to accessing the application. Although hospice clinicians also expressed generally positive perceptions of ENVISION's availability, some clinicians noted potential limitations. For example, one clinician stated that the application would be inaccessible for some, including "people without internet, without computers." Another explained, "Keep in mind that your patients or family members...may not have access to technology." With regard to the application's availability to clinicians, one participant emphasized the importance of integrating any new tools with the existing electronic health record:

> *Sometimes more tools are better, but sometimes more tools take more time...If this could be embedded into our current [EHR], I could find that helpful. To log out and then log in to something else or to toggle between two applications, I think, would be more cumbersome.*

Data describing ENVISION's affordability and sustainability referenced the cost of ENVISION itself or other resources (eg, internet access and technological devices) required to access and use the application. Participants who commented on ENVISION's affordability and sustainability expressed generally positive perceptions. Family caregivers, many of whom likely assumed the application would be included in routine hospice care and thus free of charge to patients and families, did not directly raise the issue. Two clinicians (one of whom also occupied an administrative role) noted the importance of making the application affordable and sustainable for hospice agencies, stressing the issue of cost-effectiveness, and linking the application to clinically relevant outcomes. When asked if they would recommend routine use of ENVISION in hospice care, they replied as follows:

> *I think you'd have to look at...cost, [but] this program will definitely improve outcomes for symptom management.*

Data describing ENVISION's perceivability referenced the extent to which the application could be used by individuals with different sensory abilities, particularly with regard to vision. Participants' comments regarding the perceivability of ENVISION varied. Some suggested that the font size was too small:

> *The print would have to be a little larger. I think, in general, anything to do with [older adults] should be larger.*

Others stated the opposite, describing ENVISION's text and images as "not too small." One participant who did not cut and paste the one-time password from their welcome email into the log-in screen noted that "for people [who] are...visually challenged, [entering] the password [could be] a bit of a headache." A clinician described ENVISION as a "wonderful option for patients and family members" but suggested that it not be required, as some may not be "able to see and hear...and all those sorts of things."

Data describing ENVISION's operability referenced the extent to which it could be used by individuals with different physical abilities or technological proficiencies. Participants provided mixed feedback on ENVISION's operability. Overall, for family caregivers, logging into the application for the first time (which required them to enter their email address and a one-time password emailed to them during the call) was more challenging than using it once they logged in. One family caregiver stated as follows:

> *If you're not a computer-savvy sort of person, [logging in the first time] could be a challenge.*

This was particularly the case for users (family caregivers and hospice clinicians) who did not cut and paste their one-time password from their welcome email into the log-in screen:

> *I have to write [the one-time password] down because I won't remember it.*

Another user provided a specific suggestion related to this issue:

> *Do you know, with this system, can you instruct people to copy the initial password and paste it in, just to make it easier for people?*

After entering the system, however, family caregivers could easily enter symptom and well-being data into ENVISION, describing this process as "pretty simple" and "really easy." One family caregiver stated, "That took less than 30 seconds," while another explained as follows:

> *I think it was really easy. I really liked how there's a definition [of each symptom or wellbeing indicator]...It's comparable to other applications I use for work...I think that would be pretty straightforward for the general population, too.*

Although this user saw the information symbol (lowercase "i" with a circle around it) located next to each symptom and well-being indicator and knew to click on it for more information, others required prompting before being able to do so. Ultimately, however, 100% of the users who expressed a desire for more information about a symptom or well-being indicator were able to successfully obtain that information by clicking on the information symbol independently or after receiving the following verbal prompt from the researcher: "Is there anything on the screen that might give you more information about that?"

Clinicians' comments regarding ENVISION's operability with regard to logging into the system and navigating the application generally mirrored those of family caregivers. They described the overall application as generally operable while emphasizing that it might be challenging for individuals to use if they were physically unable to type or "[could not] even use a smartphone" (the issue of family caregivers' ability to independently use the application is further described when discussing *congruence* under the *Relevance* section, as numerous clinicians expressed concern that they would be tasked with training and assisting technologically challenged family caregivers with ENVISION's use, requiring significant amounts of their already limited work time). Most of the unique data about ENVISION's operability for clinicians focused on using the interactive graphs that enabled longitudinal viewing of symptoms and well-being indicators (these graphs were available only to clinician users). To choose which symptoms or well-being indicators appeared on the longitudinal graphs, clinician users needed to click a box next to the appropriate symptoms or indicators, which multiple users failed to do without prompting or considerable thought.

For example, one clinician user's think-aloud data included the following:

> *I'm guessing maybe—I was looking at it on my computer—the little check boxes underneath the graph might affect the graph, I guess...Now, let's see...[begins clicking on boxes and noting changes to the graph].*

Another suggested that the application be modified to include "some education on what that graph is and how to utilize it." Other clinician users appeared to interact with the graph more intuitively and were observed easily manipulating it. One such clinician stated as follows:

> *I didn't have any problem with it. I'm middle-aged and...pretty computer-literate. I didn't have any problems with it at all.*

Data describing ENVISION's comprehensibility referenced the extent to which users could understand and accurately interpret the application's content. ENVISION's comprehensibility was determined to be mixed. Overall, family caregivers could easily comprehend ENVISION's content, successfully entering symptom and well-being indicators and accurately interpreting the data visualizations (labeled boxes shaded in different intensities to reflect symptom and indicator intensity) featured on patient and caregiver scorecards. Several described the content as easy to understand, using words and phrases such as "straightforward" and "written in plain English." Among family caregivers, comprehensibility challenges were limited to understanding the definition of specific symptoms or well-being indicators; however, most of these challenges were resolved when the users clicked on the information symbol and were shown a definition. Users commonly clicked on the information symbol next to "well-being," expressing confusion about what it entailed (eg, "Is that mental well-being or is that physical?"). Differentiating between "tiredness" and "drowsiness" was also challenging for numerous family caregivers. Among the comprehensibility challenges that were not resolved by clicking on the information symbol, nonspecificity (eg, uncertainty whether they were being asked to report on generalized anxiety or anxiety specific to the hospice experience and confusion about the insomnia indicator: "Is that insomnia [as in] you can't sleep, or is it just that you know you have to get up because you have to check [on the patient]?") was by far the most common. Clinicians recommended that longitudinal graphs be labeled with complete descriptions of symptoms and well-being indicators rather than shortened descriptors (eg, use "shortness of breath" rather than just "breath"). However, this may have been more of a design preference than an issue related to comprehensibility. One family caregiver recommended that features beyond the patient or caregiver name and uploaded photograph be included to remind the family caregiver when they were being asked to report on the patient's experience or their own:

> *You could say, "Now we're...addressing you, not the patient" or however you would want to say it...Make it clearer which page is for the patient and which is for the caregiver.*

Relevance

Data describing ENVISION's purpose referenced its perceived usefulness. Most users identified a clear and important purpose for ENVISION in their respective roles. Family caregivers repeatedly emphasized the importance of reporting symptoms and well-being indicators to the hospice team (patient pain was most commonly cited as a high-priority symptom to communicate). Feedback on the importance of the general well-being indicator, however, was mixed among family caregivers. Some family caregivers selected it as the most critical piece of data to communicate, while others saw it as redundant:

> *I feel that was a culmination of all the options that you gave me to begin with. If I'm already addressing each one of those issues individually…maybe I didn't necessarily need to rate it separately.*

One caregiver was unclear why caregiver insomnia was included as a well-being indicator:

> *If I had insomnia, how would the healthcare provider help me with that?*

With a few exceptions, clinician participants could readily identify a purpose for ENVISION in their clinical role, evident from representative statements as follows:

> *I think it would help me do my job better due to it being so precise, and I go back to the [patient and caregiver scorecard]. It's a lot easier for me to see what's going on with that patient the way that was presented than what I'm doing now in a chart, where I have to click and copy and paste and go here and there and everywhere [to] different notes and things like that.*

A chaplain explained how using ENVISION would enhance spiritual care:

> *[ENVISION might help me decide] how soon I might want to make another visit. Because if the person is very spiritual and prayer or listening to hymns or singing [helps] with pain or anxiety, [using ENVISION would allow me to] see if maybe another visit might be something that they might appreciate sooner than later.*

One chaplain user, who expressed generally positive perceptions of ENVISION's relevance, suggested that the application would be improved by the inclusion of an indicator for "some type of spiritual distress." With regard to the graph's usefulness, clinician feedback was generally positive. One clinician explained as follows:

> *[ENVISION] would be helpful to identify patterns without having to go back and read your notes, and it would also be helpful to measure how long a pattern's been happening when it might be hard to conceptualize that just through memory.*

More general comments described ENVISION as "a really cool tool and a really good idea [that would be] really useful" and "really helpful." Another stated as follows:

> *I would be eager for [ENVISION]. I think it would be great for patients' families [to feel] like they have another method of communicating with us.*

A clinician described the patient and caregiver scorecards as follows:

> *I think they are very helpful. It's easy, quick to identify, and you can see exactly what the problems are.*

A few clinicians specifically commented on including caregiver well-being indicators in addition to patient data, noting its usefulness:

> *What I really appreciate is that...it indicates an attention to continued review about how the caregiver is doing, and that isn't always done.*

A hospice physician stated as follows:

> *I think having access to this would really help, so I can get the patient perspective. As a hospice physician, a lot of times I'm getting just a third-party review from the nurse. I don't necessarily get this drilled-down of a rating scale on what's going on.*

Two clinician users were more negative than positive regarding ENVISION's purpose. One (the more ambivalent of the 2) user stated as follows:

> *I think it's helpful, but is it necessary? I don't know.*

The other user explained their reservations about the application:

> *[My] knee-jerk response to [ENVISION] is why the heck would I be looking at a computer and not talking to [the family caregiver]? I have no idea why we would add a layer between the hospice nurse and the [patient and family]...I'm struggling with the whole concept...I'm a [age in the 60-69 range]-year-old nurse, and I'm covering two different teams...[Even] with 21 patients, I would still want to have direct conversations with my patients and families. I would want them to feel like they have no barriers whatsoever to either calling the office or calling my work cell phone and saying, "Guess what's happening this morning?"*

Data describing ENVISION's congruence referenced the degree to which the application was aligned with users' values, beliefs, customs, and preferences. When discussing ENVISION's congruence, participants commented on qualities such as the application's goodness of fit or described what they liked and disliked about it. Overall, family caregivers generally reported ENVISION to be aligned with their values, beliefs, customs, and preferences. None of the family caregivers reported perceived or anticipated challenges with daily symptom and well-being data entry. Two family caregivers mentioned specific symptoms that seemed at odds with their expectations or understanding of hospice care. One questioned why the hospice team would need to know whether they (the family caregiver) were experiencing insomnia (as previously described), and the other thought asking about patients' lack of appetite might be problematic, as they understood decreased appetite to be a normal part of the dying process rather than something that required a clinical response: "I was just told, 'Don't try to make her [eat].'" Other data suggested that this family caregiver's concern might have been warranted, as one caregiver cited "lack of appetite" as among the most important symptoms to communicate to the hospice team, explaining that a hospice patient "needs to eat." While cultural congruence was infrequently discussed regarding the ENVISION application, the few comments provided were positive and related to cultural norms that might reduce the likelihood of unscheduled contacts with the hospice team in the absence of a tool such as ENVISION:

> *In an ideal world, every clinician would call [the family caregiver of a patient whose pain medications were increased] the next day to check in to see if this is working better, but I know that's not going to be the case. A lot of caregivers actually wait a full week until the...nurse comes back, and I'm like, 'Oh, don't do that. Let them know that it's working. Let them know if it's not working. Because [patients] don't need to suffer like that.' There are cultural values that limit how people communicate, and that's especially true in, like, Latino populations and other people who have been marginalized before who don't know that they're also an authority in this, in the reporting of patients' symptoms.*

Congruence pertaining to clinician data primarily related to clinicians' preferences and experiences as busy professionals with limited time to engage in additional work tasks. These data primarily addressed the provision of technical support or data entry reminders to family caregivers using ENVISION, something clinicians were almost universally disinclined to take on. In addition, the previously described response from the hospice nurse with a strong preference for nontechnologically mediated communication ("Why we would add a layer between the hospice nurse and the [patient and family]?") was identified as a likely example of perceived incongruence with the clinician's personal values (ie, an aversion to technology or belief that more traditional forms of communication are more effective or personal). Conversely, perceptions of ENVISION as a tool to increase efficiency were strongly related to perceptions of the application as a good fit for clinicians' workflow. For example, one clinician emphasized the timesaving value of ENVISION's patient and caregiver scorecards:

> *It doesn't seem like there's a lot of information on [the scorecards] that doesn't need to be there, so that's helpful...Whenever I'm reading people's [medical] records, I'm just like, "Where is the information I'm looking for?"*

Descriptions of the application as "a quick snapshot" and as allowing clinicians to quickly identify symptoms in need of attention were common.

Data describing ENVISION's credibility referenced the degree to which users perceived the application as trustworthy. Participants rarely commented on ENVISION's credibility. Furthermore, 100% of the data segments labeled with the code "credibility" were also labeled with the code "safety" and were found to pertain more directly to security issues than to the perceived trustworthiness of the application. Thus, to avoid duplicate reporting of findings, these data are described in the context of ENVISION's safety, which is discussed in the section describing findings related to ENVISION's impact.

Impact

Data describing ENVISION's benefit referenced the ways in which the application might improve users' lives. Both family caregivers and clinicians cited potential benefits of the ENVISION application, primarily centered around better symptom management and increased awareness of opportunities to improve patients' and family caregivers' quality of life. Much of the information labeled with code "benefit" was also labeled with the code "relevance" due to perceived improvement in individuals' ability to complete tasks associated with their respective roles, whether as family caregivers or hospice clinicians. For example, a hospice clinician indicated that ENVISION "would be a good communication tool [so] that...all the team is getting the same information in real time." Another stated, "It could allow for efficient follow-up and getting the care needed to the patient probably in...a faster manner." Several clinicians predicted that ENVISION use would increase patients' and family caregivers' satisfaction with the care they received. A clinician explained as follows:

> *It would make the patients and families feel like the hospice team is more competent, that we actually work together as a team, because we would know going in [to the home] what has been going on with them for the past few days or since we've been there. I do get that a lot. Patients are like, 'I don't want to go over it again. Don't you guys talk to each other?'*

Another clinician described ENVISION as potentially empowering:

> *When your patients and families are allowed to have input, it makes them feel empowered and a part of the care. I could see how [ENVISION] would be beneficial for the patients or their families to utilize.*

One clinician user identified benefits from 3 perspectives:

> *From the caregiver's perspective, I think it's helpful to have some sense of feeling like you have an outlet to discuss what symptoms you're having so that you can actually get help from the interdisciplinary team. I think it's helpful from the patient's perspective to kind of have a sense of control over how their symptoms are being managed...I think it's helpful from the provider's perspective for…symptom management, changes in medications, gauging how they're working, as well as helping guide that family with new symptoms that are showing up and education as well as prognostication.*

Comments describing potential drawbacks of the application were less frequent and often co-coded with other digital inclusivity elements. For example, clinicians worried that family caregivers with limited technological skills might feel frustrated when interacting with the application. Clinicians also worried that family caregivers would find it burdensome to enter daily symptom assessment data:

> *[Having daily symptom and well-being data] would be sweet. That might be a big ask for some caregivers. One more thing to do.*

However, no family caregivers cited daily data entry as likely burdensome. Clinicians also cautioned against using ENVISION data to reduce or "change the care that we otherwise would attempt to provide."

Data describing ENVISION's safety referenced the presence or absence of protection from online threats associated with the application's use, such as malware or data breaches. ENVISION's safety was rarely addressed. When the users did address it, they tended to focus on password-related hassles rather than concerns that using the application made them susceptible to digital threats. One exception was that clinicians emphasized the need for any application used in hospice to be Health Insurance Portability and Accountability Act compliant, as that would likely be required for adoption into routine care.

Discussion

Principal Findings

We developed a conceptual framework of digital inclusivity for health information systems and then applied the framework in evaluating the digital inclusivity of ENVISION, a symptom monitoring tool for home hospice care. Our analysis identified accessibility, relevance, and impact as essential considerations in assessing a health system's digital inclusivity; all 3 were incorporated into our newly created ARI framework. Study findings resulting from our application of the ARI framework generally supported ENVISION's digital inclusivity, particularly concerning its perceived relevance to the work of family caregivers and hospice clinicians and its potentially positive impact on symptom management and quality of life. Limitations to ENVISION's digital inclusivity centered around issues of accessibility, particularly operability among individuals with limited technological knowledge and skills.

The ARI framework is informed by and extends prior knowledge. It incorporates constructs from several existing models, most notably the community-oriented framework on which it was explicitly based [15]. Both frameworks place a strong emphasis on accessibility, including availability, affordability, and more standard accessibility features, conceptualized in the ARI framework as perceivability, operability, and comprehensibility (these closely mirror principles detailed in the widely referenced Web Content Accessibility Guidelines 2.0, authored by the World Wide Web Consortium) [25]. The ARI framework also echoes some of the principles highlighted in usability heuristics for user interface design given by Nielsen [26] (eg, the match between the system and the real world) and Technology Acceptance Model elaborated by Davis [27] (eg, usefulness and ease of use). In building on prior research, the ARI framework synthesizes relevant constructs from numerous bodies of existing work, setting the stage for meaningful assessment of the digital inclusivity of individual tools. In addition to providing a valuable synthesis of existing models pertinent to digital inclusivity, the ARI framework incorporates constructs uniquely relevant to the context of health information systems. It identifies patients and families, clinicians, and institutions as unique yet interdependent user types, each with potentially different cultures, responsibilities, needs, and concerns. Furthermore, it is grounded in data derived from home hospice care, a clinical context that highlights the extent to which patients and families are increasingly required to be both care providers, via family caregiving [28] and disease self-management [29], and care recipients, via patient- and family-centered models of care [30].

Importantly, the salience of specific constructs highlighted in the ARI framework will likely fluctuate over time. For example, limited internet availability may become less problematic in the United States, where the federal government's recent infrastructure investments are predicted to significantly expand rural internet availability [31]. Similarly, while health information systems' operability will likely continue to be important, tools requiring basic technological skills to operate may become more broadly operable due to demographic shifts, as the proportion of potential users who are digital natives (people who grew up regularly using digital technologies [32]) is expanding. Other issues, such as cost-related barriers to accessing digital health tools, seem likely to retain their importance over time, particularly in light of increased recognition of income inequality and other social determinants of health [33].

Limitations

Study findings should be interpreted in light of numerous limitations. First, our study sample was disproportionate in terms of having higher number of non-Hispanic, White, and female individuals. All family caregivers who participated in the study had at least some college education or trade school experience, and all could speak and read English. Furthermore, while our sample reflected some variability regarding functional abilities (eg, some participants reported mild visual impairment requiring corrective lenses), no participants reported significant physical disabilities. Additional research with more diverse participants, including individuals with varying degrees of literacy and functional ability, is needed to refine the ARI framework and better understand and ultimately enhance ENVISION's digital inclusivity. Notably, the ARI framework is in its infancy, and additional development and testing will likely be needed to maximize its potential impact on the field. In particular, noted conceptual links between relevance and the benefit subcategory of impact highlight the need for ongoing attention to issues of construct validity. With regard to ENVISION's potential for clinical adoption, recommended next steps include pilot testing in real-world scenarios, followed by more definitive efficacy testing to determine its effect on outcomes identified by users as areas of potential impact, such as symptom management and quality of life. In addition, in developing the ARI framework, we opted to include some elements of digital inclusivity even in the absence of data from the ENVISION evaluation supporting their inclusion if existing literature or expertise among team members suggested that they were essential to the concept of digital inclusivity. For example, although participants rarely discussed safety, it was retained from the original framework in light of the large and growing number of digital security threats in existence and noted disparities in individuals' knowledge of cybersecurity [34]. Thus, while the ARI framework is primarily grounded in data derived from our evaluation of ENVISION's digital inclusivity, some exceptions apply. Finally, we emphasize that all study participants used ENVISION in a hypothetical manner, interacting with data either from memory (as was the case for family caregiver participants) or from fictitious patients and caregivers (as was the case for clinician participants). We cannot conclude with certainty that individuals using the application in real-life situations would have similar experiences or provide feedback mirroring that provided by the study participants. The hypothetical nature of participants' application use also limited their ability to provide feedback on certain aspects of ENVISION, such as the application's actual costs, including the labor and other resources required to support and sustain its use. Additional research is needed to determine ENVISION's costs and its benefits to home hospice patients, family caregivers, and clinicians.

Conclusions

Our evaluation of ENVISION identified many ways by which the tool is digitally inclusive. Although specific users' experiences and feedback varied, ENVISION was determined to be generally accessible by individuals with the skills and resources required to access and operate typical web-based applications. This overall assessment was most explicitly reflected in users' comparisons of ENVISION's operability to that of applications that they regularly encountered in their work and personal lives. User data were most positive regarding ENVISION's relevance, with nearly all family caregivers and clinicians readily identifying multiple use case scenarios for the application in home hospice care. Although individuals participating in the study interacted with hypothetical patient and family caregiver data, most predicted numerous, meaningful, positive outcomes of ENVISION use, including improved symptom management and patient and caregiver well-being.

User data also provided insights into ways in which ENVISION's digital inclusivity is limited. While most Americans can access the internet from home, this capability remains limited among older adults, racial and ethnic minority groups, and individuals residing in rural communities and low-income households [35,36]. As an entirely web-based application requiring daily use, ENVISION would largely be inaccessible to individuals without high-speed internet access at home or nearby. Moreover, individuals with limited technological skills may be unable to use the application without training and support, which many hospice agencies lack the resources to provide. Minimally, our findings suggest that support would be needed to assist first-time users in logging into the system and creating a new password, as this proved to be the most challenging aspect of operating ENVISION for many users. Adoption of password alternatives (eg, biometrics, physical hardware, etc) may be considered as this technology evolves [37]. In addition, offering support in using existing tools to enhance accessibility (eg, the zoom-in feature or magnifying applications on mobile phones to enhance character visibility) may be needed. Incorporation of existing principles (eg, the usability heuristics by Nielsen) [26] into future design efforts would likely enhance operability and is thus supported by study findings. With regard to relevance, ENVISION may be a poor fit for family caregivers and clinicians who prefer face-to-face (or telephone) contact over more technologically mediated communication. Clinicians' concerns that the application might lead to decreased face-to-face contact might be assuaged by presenting ENVISION as a tool to supplement, not substitute, in-home patient and family care. Finally, findings clearly indicate that ENVISION must provide clinicians with a net gain in terms of efficiency, consistent with existing research highlighting time constraints as the most significant professional challenge for hospice clinicians [38].

Acknowledgments

The Engagement and Visualization to Improve Symptoms in Oncology Care (ENVISION) application was created in partnership with the software engineering and user experience team at PlatformSTL. Funding for ENVISION and this study was provided by the BJC HealthCare and Washington University in St Louis Healthcare Innovation Lab and the Washington University in St. Louis Institute for Informatics, Data Science, and Biostatistics through the Big Ideas Data-Driven Innovation Competition (principal investigator: KTW) and by the Foundation for Barnes-Jewish Hospital (principal investigator: PW). All works made for hire are the exclusive joint property of Washington University in St. Louis and BJC HealthCare. The Recruitment Enhancement Core in the Network Capacity component of the Washington University Institute of Clinical and Translational Sciences is supported by the National Center for Advancing Translational Sciences of the National Institutes of Health (UL1TR002345). The paper's content is solely the authors' responsibility and does not necessarily represent the official views of the National Institutes of Health or any other funding entities.

Authors' Contributions

All authors contributed to the writing of the article and provided substantive feedback on its content. KTW designed and directed the study and led data analysis and interpretation of findings. DPO and GD contributed to data analysis and interpretation of findings. AKD served as the study coordinator, collected data, and contributed to data analysis and interpretation of findings. PGL, PW, and JJB contributed to the study design and interpretation of findings.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Original community-level digital inclusivity principles and final codes.
[DOCX File , 19 KB - humanfactors_v11i1e51789_app1.docx]

References

1. Tatum PE, Mills SS. Hospice and palliative care: an overview. Med Clin North Am 2020 May;104(3):359-373. [doi: 10.1016/j.mcna.2020.01.001] [Medline: 32312403]
2. NHPCO facts and figures. National Hospice and Palliative Care Organization. 2022 Dec. URL: https://www.nhpco.org/wp-content/uploads/NHPCO-Facts-Figures-2022.pdf [accessed 2024-05-07]
3. Casarett D, Harrold J, Harris PS, Bender L, Farrington S, Smither E, et al. Does continuous hospice care help patients remain at home? J Pain Symptom Manage 2015 Sep;50(3):297-304 [FREE Full text] [doi: 10.1016/j.jpainsymman.2015.04.007] [Medline: 25936937]
4. Haley WE, Allen RS, Reynolds S, Chen H, Burton A, Gallagher-Thompson D. Family issues in end-of-life decision making and end-of-life care. Am Behav Sci 2016 Jul 27;46(2):284-298. [doi: 10.1177/000276402236680]
5. McMillan SC. Interventions to facilitate family caregiving at the end of life. J Palliat Med 2005;8 Suppl 1:S132-S139. [doi: 10.1089/jpm.2005.8.s-132] [Medline: 16499461]
6. Washington K. Commitment to care: a grounded theory of informal hospice caregivers' development as symptom managers. Qual Soc Work 2011 Dec 29;12(3):358-371. [doi: 10.1177/1473325011429047]
7. Mather H, Kleijwegt H, Bollens-Lund E, Kelley AS, Ornstein KA. Symptom management experience of end-of-life family caregivers: a population-based study. J Pain Symptom Manage 2022 Dec;64(6):513-520 [FREE Full text] [doi: 10.1016/j.jpainsymman.2022.07.017] [Medline: 35944883]
8. Chi NC, Demiris G. Family caregivers' pain management in end-of-life care: a systematic review. Am J Hosp Palliat Care 2017 Jun;34(5):470-485. [doi: 10.1177/1049909116637359] [Medline: 26975303]
9. Ratkowski KL, Washington KT, Craig KW, Albright DL. The stress of sadness: the most stressful symptoms for hospice family caregivers. Am J Hosp Palliat Care 2015 Nov;32(7):745-749. [doi: 10.1177/1049909114540034] [Medline: 24982302]
10. Tarter R, Demiris G, Pike K, Washington K, Parker Oliver D. Pain in hospice patients with dementia: the informal caregiver experience. Am J Alzheimers Dis Other Demen 2016 Sep 09;31(6):524-529 [FREE Full text] [doi: 10.1177/1533317516653825] [Medline: 27303062]

11. Washington KT, Wilkes CM, Rakes CR, Otten SJ, Parker Oliver D, Demiris G. Relationships among symptom management burden, coping responses, and caregiver psychological distress at end of life. J Palliat Med 2018 Sep;21(9):1234-1241. [doi: 10.1089/jpm.2017.0584] [Medline: 29727256]
12. Phongtankuel V, Teresi JA, Eimicke JP, Kong JX, Adelman RD, Prigerson HG, et al. Identifying the prevalence and correlates of caregiver-reported symptoms in home hospice patients at the end of life. J Palliat Med 2020 May;23(5):635-640 [FREE Full text] [doi: 10.1089/jpm.2019.0324] [Medline: 31873053]
13. Washington KT, Demiris G, Oliver DP, Backonja U, Norfleet M, Han CJ, et al. ENVISION: a tool to improve communication in hospice interdisciplinary team meetings. J Gerontol Nurs 2020 Jul 01;46(7):9-14. [doi: 10.3928/00989134-20200605-03] [Medline: 32597996]
14. Yao R, Zhang W, Evans R, Cao G, Rui T, Shen L. Inequities in health care services caused by the adoption of digital health technologies: scoping review. J Med Internet Res 2022 Mar 21;24(3):e34144 [FREE Full text] [doi: 10.2196/34144] [Medline: 35311682]
15. Building digital communities: a framework for action. Institute of Museum and Library Services. URL: https://www.imls.gov/publications/building-digital-communities-framework-action [accessed 2024-05-07]
16. van Kessel R, Roman-Urrestarazu A, Anderson M, Kyriopoulos I, Field S, Monti G, et al. Mapping factors that affect the uptake of digital therapeutics within health systems: scoping review. J Med Internet Res 2023 Jul 25;25:e48000 [FREE Full text] [doi: 10.2196/48000] [Medline: 37490322]
17. Sieck CJ, Sheon A, Ancker JS, Castek J, Callahan B, Siefer A. Digital inclusion as a social determinant of health. NPJ Digit Med 2021 Mar 17;4(1):52 [FREE Full text] [doi: 10.1038/s41746-021-00413-8] [Medline: 33731887]
18. Portz JD, Fruhauf C, Bull S, Boxer RS, Bekelman DB, Casillas A, et al. "Call a teenager… that's what I do!" - Grandchildren help older adults use new technologies: qualitative study. JMIR Aging 2019 Jun 06;2(1):e13713 [FREE Full text] [doi: 10.2196/13713] [Medline: 31518274]
19. Usability. U.S. General Services Administration. URL: https://www.usability.gov/how-to-and-tools/resources/ucd-map.html [accessed 2017-01-08]
20. Mik-Meyer N. Multimethod qualitative research. In: Silverman D, editor. Qualitative Research. London, UK: SAGE Publications; 2020.
21. Santhosh L, Rojas JC, Lyons PG. Zooming into focus groups: strategies for qualitative research in the era of social distancing. ATS Sch 2021 Jun;2(2):176-184 [FREE Full text] [doi: 10.34197/ats-scholar.2020-0127PS] [Medline: 34409412]
22. Lundgrén-Laine H, Salanterä S. Think-aloud technique and protocol analysis in clinical decision-making research. Qual Health Res 2010 Apr;20(4):565-575. [doi: 10.1177/1049732309354278] [Medline: 19959822]
23. Lunde P, Skoglund G, Olsen CF, Hilde G, Bong WK, Nilsson BB. Think aloud testing of a smartphone app for lifestyle change among persons at risk of type 2 diabetes: usability study. JMIR Hum Factors 2023 Nov 15;10:e48950 [FREE Full text] [doi: 10.2196/48950] [Medline: 37966894]
24. Fisher G, Aguinis H. Using theory elaboration to make theoretical advancements. Org Res Method 2017 Feb 03;20(3):438-464. [doi: 10.1177/1094428116689707]
25. Web Content Accessibility Guidelines (WCAG) 2.0. World Wide Web Consortium. 2008. URL: https://www.w3.org/TR/WCAG20/ [accessed 2023-08-11]
26. Nielsen J. Enhancing the explanatory power of usability heuristics. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1994 Presented at: CHI '94; April 24-28, 1994; Boston, MA. [doi: 10.1145/191666.191729]
27. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
28. Washington KT. Hospice and caregiving. In: Gu D, Dupre ME, editors. Encyclopedia of Gerontology and Population Aging. Cham, Switzerland: Springer; 2021.
29. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. JAMA 2002 Nov 20;288(19):2469-2475. [doi: 10.1001/jama.288.19.2469] [Medline: 12435261]
30. Rawson JV, Moretz J. Patient- and family-centered care: a primer. J Am Coll Radiol 2016 Dec;13(12 Pt B):1544-1549. [doi: 10.1016/j.jacr.2016.09.003] [Medline: 27888939]
31. Biden-Harris administration announces over $700 million to connect people in rural areas to high-speed internet. U.S. Department of Agriculture. 2023 Jun 12. URL: https://www.usda.gov/media/press-releases/2023/06/12/biden-harris-administration-announces-over-700-million-connect [accessed 2023-07-05]
32. Prensky M. Digital natives, digital immigrants part 1. On the Horizon 2001 Sep;9(5):1-6. [doi: 10.1108/10748120110424816]
33. Truesdale BC, Jencks C. The health effects of income inequality: averages and disparities. Annu Rev Public Health 2016;37:413-430. [doi: 10.1146/annurev-publhealth-032315-021606] [Medline: 26735427]
34. Vogels EA, Anderson M. Americans and digital knowledge in 2019. Pew Research Center. 2019. URL: https://www.pewresearch.org/internet/2019/10/09/americans-and-digital-knowledge/ [accessed 2023-06-01]
35. Internet/broadband fact sheet. Pew Research Center. 2024 Jan 31. URL: https://www.pewresearch.org/internet/fact-sheet/internet-broadband/ [accessed 2024-05-13]

36. Swenson K, Ghertner R. People in low-income households have less access to internet services. Office of the Assistant Secretary for Planning & Evaluation, U.S. Department of Health & Human Services. 2020 Apr. URL: https://aspe.hhs.gov/sites/default/files/private/pdf/263601/Internet_Access_Among_Low_Income.pdf [accessed 2023-07-14]
37. van Oorschot PC. User authentication—passwords, biometrics and alternatives. In: Computer Security and the Internet. Information Security and Cryptography. Cham, Switzerland: Springer; 2021.
38. Washington KT, Alaniz Staner LM, Collie BE, Craig KW, Demiris G, Oliver DP. Toward a practice-informed agenda for hospice intervention research: what are staff members' biggest challenges? Am J Hosp Palliat Care 2021 May;38(5):467-471. [doi: 10.1177/1049909120948222] [Medline: 32757823]

Abbreviations

ARI: Accessibility, Relevance, and Impact
ENVISION: Engagement and Visualization to Improve Symptoms in Oncology Care

Edited by A Kushniruk; submitted 11.08.23; peer-reviewed by D Chrimes, V Provencher; comments to author 18.01.24; revised version received 13.02.24; accepted 07.04.24; published 23.05.24.

Please cite as:
Washington KT, Oliver DP, Donehower AK, White P, Benson JJ, Lyons PG, Demiris G
Accessibility, Relevance, and Impact of a Symptom Monitoring Tool for Home Hospice Care: Theory Elaboration and Qualitative Assessment
JMIR Hum Factors 2024;11:e51789
URL: https://humanfactors.jmir.org/2024/1/e51789
doi:10.2196/51789
PMID:38781581

Original Paper

Usability, Acceptability, and Preliminary Effectiveness of a Peer-Delivered and Technology-Supported Mental Health Intervention for Family Caregivers of People With Dementia: Field Usability Study

Caroline Collins-Pisano[1,2], BA; Amanda N Leggett[3], PhD; David Gambee[4], BA; Karen L Fortuna[4], PhD, LICSW

[1]Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, United States

[2]Department of Psychology, University of Colorado, Colorado Springs, Colorado Springs, CO, United States

[3]Institute of Gerontology, Wayne State University, Detroit, MI, United States

[4]Department of Psychiatry, Dartmouth College, Hanover, NH, United States

Corresponding Author:
Caroline Collins-Pisano, BA
Department of Psychology
University of Colorado, Colorado Springs
1420 Austin Bluffs Pkwy
Colorado Springs, CO, 80918
United States
Phone: 1 (719) 255 8227
Email: ccolli20@uccs.edu

Abstract

Background: Family caregivers of people with dementia are critical to the quality of life of care recipients and the sustainability of health care systems but face an increased risk of emotional distress and negative physical and mental health outcomes.

Objective: The purpose of this study was to examine the usability, acceptability, and preliminary effectiveness of a technology-based and caregiver-delivered peer support program, the Caregiver Remote Education and Support (CARES) smartphone or tablet app.

Methods: A total of 9 adult family caregivers of people with dementia received the CARES intervention, and 3 former family caregivers of people with dementia were trained to deliver it. Quantitative data were collected at baseline and at the end of the 2-week field usability study. Qualitative data were also collected at the end of the 2-week field usability study.

Results: The field usability study demonstrated that a 2-week peer-delivered and technology-supported mental health intervention designed to improve burden, stress, and strain levels was experienced by former and current family caregivers of people with dementia as acceptable. Current family caregivers rated CARES as above average in usability, whereas the caregiver peer supporters rated CARES as marginally usable. CARES was associated with non–statistically significant improvements in burden, stress, and strain levels.

Conclusions: This field usability study demonstrated that it is possible to train former family caregivers of people with dementia to use technology to deliver a mental health intervention to current family caregivers of people with dementia. Future studies would benefit from a longer trial; a larger sample size; a randomized controlled design; and a control of covariables such as stages of dementia, years providing care, and severity of dementia symptoms.

(JMIR Hum Factors 2024;11:e41202) doi:10.2196/41202

KEYWORDS

family caregivers; dementia; peer support; technology; mobile phone

Introduction

Background

Family caregivers of people with dementia are critical to the quality of life of care recipients and to the sustainability of health care systems. Family caregivers provide US $257 billion in unpaid care to people living with dementia but face an increased risk of emotional distress and negative physical (eg, heart disease and hypertension) and mental (eg, depression and anxiety) health outcomes [1,2]. While positive gains are reported in the caregiving role, caregivers are more likely than their noncaregiving peers to report stress, burden, and strain. Approximately 46% of family caregivers of people with dementia are classified as having high levels of burden [1,2]. Burden is defined as the psychological, physical, emotional, and social challenges that family caregivers experience in response to the demands of providing care [3]. Caregivers with high levels of burden report more physical and psychological symptoms, use prescription medications and health care more frequently, and provide poorer quality of care to recipients, leading to an increased likelihood of premature institutional care [4,5]. In addition, 59% of family caregivers of people with dementia rate the emotional stress of caregiving as high or very high, and 38% rate the physical stress of caregiving as high or very high [1,2]. Stress is defined as individuals' emotional or physical responses to the challenges of caregiving, such as fatigue [6,7]. The stress of providing care to a family member with dementia increases caregivers' risk of health complications, increases their susceptibility to diseases such as hypertension, and negatively affects their quality of sleep [1,2]. Family caregivers have also reported greater levels of strain compared to caregivers of people without dementia. Strain is defined as caregivers' perception of the challenges of caregiving and their state of well-being [7]. Family caregivers who perceived themselves as having higher strain levels due to caregiving responsibilities were at a higher risk of death than those who perceived little or no strain [1,2].

Technology-Based Interventions for Caregivers

While psychosocial interventions have been shown in fully powered randomized controlled trials to reduce caregiver burden and delay nursing home admission for the care recipient, a recent meta-analysis by Walter and Pinquart [8] found that current interventions had a disappointingly small to moderate effect on caregiver well-being, burden, depression, and anxiety [4,8-10]. In addition, uptake of these interventions in the real world is limited due to caregiving obligations (between 69 and 117 hours of informal care are provided to people with dementia per week), geographical distance from the intervention, requirements to meet in person, and failure to address the personalized needs of the family caregivers.

Technology-based interventions may offset these challenges through improved accessibility of psychosocial interventions at any time or location [11,12]. A scoping review on existing technology-based peer support interventions for family caregivers found that web-based programs include websites that offer educational materials with the option to contact other informal caregivers, web-based portals with psychoeducation and peer-to-peer contact, asynchronous e-learning platforms, internet support forums and chat rooms, videoconferencing support groups facilitated by a health professional, and live virtual reality support groups facilitated by psychologists [11]. Online informal peer support groups for informal caregivers have shown high levels of engagement. Technology-based interventions enable caregivers to participate even if they are unable to leave the person with dementia unsupervised. For example, during the COVID-19 pandemic, telephone-, videoconference-, and chat room–based online support groups were the only media accessible to caregivers, a benefit due to the 24/7 nature of supporting a family member with dementia and the consequential challenges in accessing support [11,13]. Online informal peer support groups are an effective method of asynchronous web-based delivery when offered in combination with a structured psychosocial and educational intervention by skilled clinical professionals [11]. However, while some studies such as that by Han et al [14] have shown significant reductions in depression, stress, and helplessness, others such as the study by Marziali and Garcia [15] have shown only moderate effects in reducing burden and depression and increasing caregiver knowledge.

Significance of Technology-Delivered Caregiver Peer Support

To date, technology-based interventions have relied on a skilled workforce of geriatric mental health professionals. However, there are insufficient numbers of adequately trained geriatric mental health care providers [16]. As such, task shifting services from skilled clinical professionals to community members provides an emerging workforce of peer support workers (ie, interventionists without formal mental health education but with life experiences similar to those of the people they serve) [17]. Although the need for traditional clinical professionals remains, peer support services for individuals with mental health conditions have been shown to increase service accessibility without impacting service quality [17]. However, there is limited knowledge of caregiver-delivered peer support.

The use of caregiver-delivered digital peer support may promote the uptake of psychosocial interventions, reduce burden, and delay nursing home admission. Social and behavioral theories such as social support, experiential knowledge, and the helper therapy principle highlight how peer supporters have the unique ability to offer acceptance, understanding, and validation to the individuals they work with [18]. Because of their shared lived experiences, peers are often viewed as more credible and trustworthy than other health care providers and, therefore, encourage increased digital health engagement. Individuals are motivated to achieve their mental health goals (eg, reductions in burden, strain, and stress) because of the reciprocal accountability offered and modeled by their peers. Former family caregivers of people with dementia have the knowledge and skillset to deliver trained peer support to current family caregivers and could potentially benefit emotionally from the delivery of support. While there are positive outcomes associated with the end of caregiving, when family caregivers of people with dementia become former family caregivers, the detrimental effects of previous caregiving fail to improve [19]. Many former dementia caregivers experience persistent sleep disturbances, depression, anxiety, increased physician office visits, declining health, and feelings of guilt and regret [20]. Despite evidence that former caregivers who pursue new caring roles benefit emotionally, payers and health care providers have not hired and trained former caregivers to provide evidence-based digital interventions [20].

Research Aims

The purpose of this study was to examine the usability, acceptability, and preliminary effectiveness of a technology-based and caregiver-delivered peer support program, the Caregiver Remote Education and Support (CARES) smartphone or tablet app. In this study, 3 former family caregivers of people with dementia (caregiver peer supporters) were trained in the delivery of caregiver peer support and delivered the CARES intervention and app to 9 current family caregivers of people with dementia (caregiver participants) in a 2-week field usability study.

Design of the CARES App

The CARES app and intervention were adapted from the PeerTECH smartphone or tablet app and developed to facilitate the examination of the usability, acceptability, and preliminary effectiveness of the first technology-based and caregiver-delivered peer support program in a 2-week field usability study. The PeerTECH system was designed using universal design principles and for lay interventionists (not skilled) to deliver fidelity-adherent evidence-based interventions. The PeerTECH system has been successfully used with Certified Peer Specialists, home health aides, and Certified Older Adult Peer Specialists [21]. Certified Peer Specialists are people with a mental health diagnosis who are hired, trained, and certified to provide peer support services to individuals with a similar diagnosis [22].

PeerTECH was built on the stress-vulnerability model. According to the stress-vulnerability model, the outcomes of a mental health condition are connected to biological vulnerability, stress, and protective factors (eg, peer support) [23]. PeerTECH was designed to empower individuals to address the stress and vulnerability that lead to worsening medical, psychiatric, and social health conditions. Peer specialists are trained to deliver evidence-based practices through the PeerTECH app to help participants decrease stressors and increase protective factors.

The PeerTECH app was codeveloped with peer specialists and includes 2 features developed through community-engaged research. The first is a peer-facing smartphone or tablet app that guides peers in delivering evidence-based health self-management skill development interventions. The app includes prompts to share their lived experiences of health challenges and solutions (ie, peer support) and structured intervention delivery through scripted, evidence-based training on topics such as coping skills, psychoeducation, medical management, social skills, and self-advocacy. The second feature is a participant-facing app that offers self-management support through features such as a library of self-management resources (eg, peer-led videos) and a secure messaging feature to connect with their assigned peer specialists and reinforce goals [21].

Similar to PeerTECH, the CARES app includes 2 features: a peer-facing smartphone or tablet app and a participant-facing smartphone or tablet app (Multimedia Appendix 1 for illustrations of the CARES app). The content of the CARES app, similar to that of PeerTECH, was designed according to the stress-vulnerability model to help participants decrease stressors and increase protective factors. The CARES app and intervention were adapted from the PeerTECH app by a team of researchers with expertise in peer support to guide caregiver peers in the delivery of evidence-based mental health interventions and designed according to the techniques and principles of peer support and acceptance and commitment therapy (ACT). The participant-facing CARES app offers a library of resources designed to educate participants on psychological skills (eg, mindfulness) that empower them to manage difficult thoughts (eg, acceptance) and engage in activities and behaviors that are guided by their life values and boost their well-being (eg, setting goals and identifying values) [24]. Caregiver peer supporters were trained and educated on topics such as values, goal setting, acceptance, avoidance, and negative thoughts and connected with their assigned participants via a secure messaging feature to reinforce ACT principles, share their lived experience of caregiving challenges, and share practices to enhance caregivers' wellness and mental health in relation to caregiver-related stressors [25]. The mutual practice of goal setting, for example, can help link caregivers' values to concrete plans for behavior change [26]. Figure 1 illustrates the CARES app.

The participant-facing app includes access to direct messaging with an assigned caregiver peer supporter, goals, wellness, surveys, and a resource library (see the bottom right of Figure 1). The caregiver peer supporter app allows caregiver peer supporters to message assigned participants directly (see the bottom left of Figure 1), view participants' goals and wellness plans, and view their progress in the library resource feature. The principal investigator (PI) had access to data on the participants' and peer supporters' rate of engagement with library resources, messaging, and goals and wellness features (see the top of Figure 1).

Figure 1. Caregiver Remote Education and Support app.

Description of the CARES App

Fundamentally, CARES is a peer-instructed and mediated caregiver support program that uses a smartphone app–based mechanism for communication. The CARES app consists primarily of two features: (1) a former family caregiver (caregiver peer supporter)–facing app on a smartphone or tablet that includes the option to message or video chat with the current family caregivers (participants) they have been assigned to provide caregiver peer support (Figure 2) and (2) a participant-facing app that offers the option to review an on-demand library of educational resources and a HIPAA (Health Insurance Portability and Accountability Act)-compliant text and video chat feature to communicate with their assigned caregiver peer supporter (Figure 3). The participants and caregiver peer supporters also have the option to set goals and create wellness plans. Figure 4 shows the features seen by both the participants and caregiver peer supporters.

Figure 2A depicts the home page of the caregiver peer supporter–facing CARES app. The home page includes the option to select the individual's availability to offer caregiver peer support to their assigned participants (Figure 2B), set goals, and access information on their assigned participants and chats. Figure 2C shows the options to message and video chat with the assigned participant and track their goals and wellness plan.

Figure 3 depicts the home page of the participant-facing CARES app. The home page includes the option to send messages to the assigned caregiver peer supporter, video chat directly with the assigned caregiver peer supporter, set goals, create a wellness plan, and access a library of resources (Figure 5).

Figure 2. Screenshot A depicts the home page on the caregiver peer supporter facing CARES application. The homepage includes the option to select the individual’s availability to offer caregiver peer support to their assigned participants (see Screenshot B), set goals, and access information on their assigned participants and chats. Screenshot C provides options to message and video chat the assigned participant and track their goals and wellness plan.

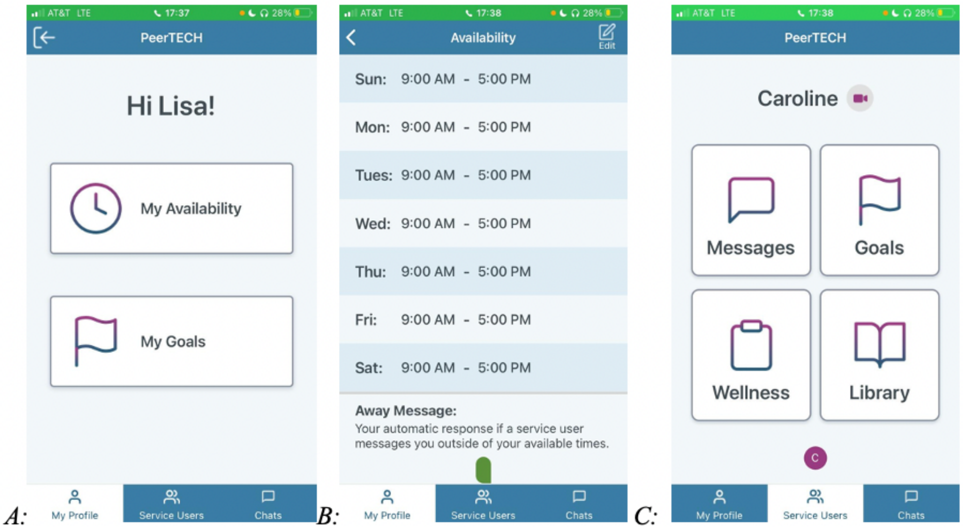
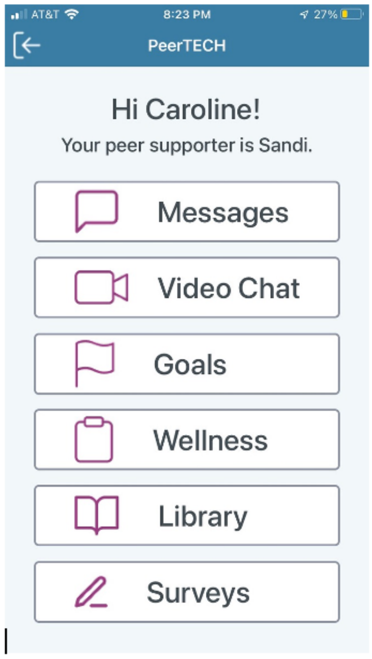


Figure 3. Caregiver Remote Education and Support participant-facing app.



The left panel in Figure 4 depicts the messaging option provided within the CARES app. Within the messaging section of the app, participants and their assigned caregiver peer supporters can send each other text-like messages. The panel in the center depicts the wellness plan. Within the wellness plan section, participants and caregiver peer supporters can add activities and strategies they wish to complete to enhance their wellness. The panel on the right depicts the goals section. Within the goals section, participants and caregiver peer supporters can set goals for themselves.

Figure 4. Caregiver Remote Education and Support features on both the participant- and caregiver peer supporter–facing apps.

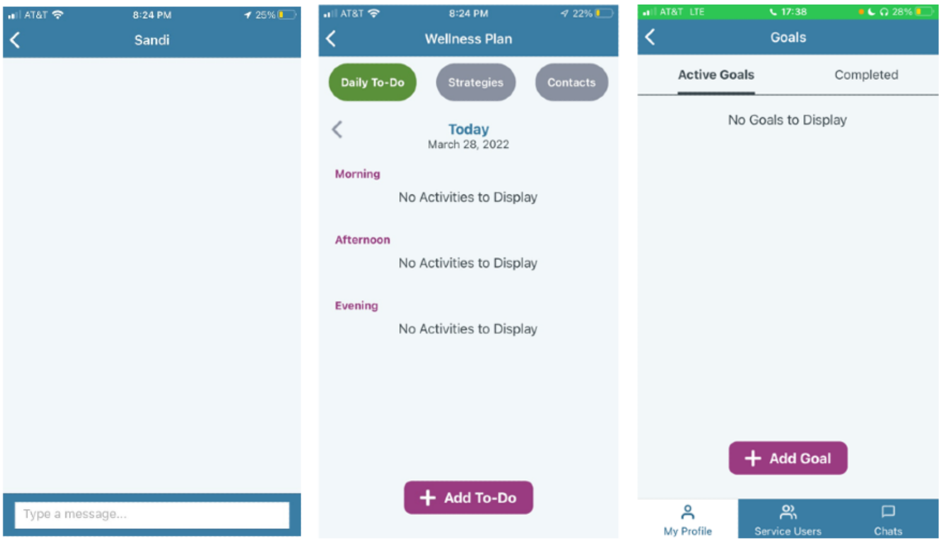
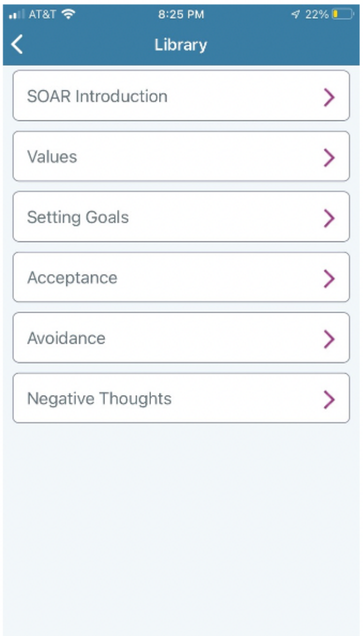


Figure 5. Caregiver Remote Education and Support library of resources.



Description of the CARES Resource Library

The CARES app includes a resource library with materials related to the principles of peer support and evidence-based practices and skills, such as ACT, to manage stress and promote mental health and well-being. CARES educational resources are designed to be reviewed by a participant and caregiver peer supporter together or individually (Figure 5). Each resource includes a combination of videos and text.

Figure 5 depicts the library of resources found in the CARES app. Each topic includes a scripted curriculum with evidence-based practices to improve mental health and well-being.

Methods

Overview

A field usability study was conducted in April 2022 to evaluate the usability, acceptability, and preliminary effectiveness of CARES as an assistive tool for guiding former family caregivers of people with dementia (n=3) in fidelity-adherent delivery to current family caregivers of people with dementia (n=12). The purpose of a field usability study is to assess the feasibility and acceptability of technology in users' natural and conceptual environments [27]. Through field usability studies, researchers gain an understanding of the problems users encounter while using the system and gain insights into how individuals use the product [27]. The field usability study was conducted for 2 weeks to provide caregiver peer supporters and participants with the time to use and assess the CARES app and establish a peer connection. The field usability study was a first step in assessing the usability and acceptability of CARES, the study design, and the training. Caregiver peer supporters provided 5 to 7 hours of peer support per week, including video chats, messaging, and supervision. Each caregiver peer supporter was assigned 4 participants. In total, 17% (2/12) of the participants dropped out of the study due to a delay in the start date, and 8% (1/12) of the participants were excluded for not using the app and failing to initiate the field usability testing process, resulting in a final sample of 9 current family caregivers and 3 caregiver peer supporters. Study measures of burden, stress, and strain levels were administered via Qualtrics (Qualtrics International Inc) at baseline and at the end of the 2-week field usability study. Study measures on usability and acceptability were administered in an hour-long HIPAA-compliant videoconference semistructured interview at the end of the 2-week field usability study. All assessments were conducted by the PI.

Ethical Considerations

The Committee for the Protection of Human Subjects at the Dartmouth Health Institutional Review Board approved the project (FP00002112). Participants provided their written informed consent. Participants were compensated for taking part in the study. Current family caregivers received US $20 for the baseline assessment, US $20 for the completion of the 2-week CARES field usability study, and US $20 for the completion of the semistructured interview conducted after the 2-week field usability study. Caregiver peer supporters received US $120 for the 6 hours of training, US $30 per hour for the 2-week field usability study, and US $30 for the completion of the semistructured interview conducted after the 2-week field usability study.

Participants

A total of 12 participants were recruited from the Dartmouth-Hitchcock Aging Resource Center, memory cafés, and senior centers across New England and via dementia caregiver Facebook support groups to participate in the study with the goal of recruiting between 10 and 20 participants. Previous research has found that 10 users report approximately 80% of usability problems and 20 users report approximately 95% of usability problems for a given product [28]. The pilot study included 9 current family caregivers of people with dementia. Participants were eligible if they (1) were a current family caregiver of an individual with dementia, (2) were aged ≥18 years, (3) spoke and read English, and (4) provided voluntary informed consent for participation in the study. The study also included 3 former family caregivers of people with dementia. Participants were eligible if they (1) were a former family caregiver of an individual with dementia, (2) were aged ≥18 years, (3) spoke and read English, (4) were willing to use technology to deliver an intervention, and (5) provided voluntary informed consent for participation in the study. All participants were excluded if they (1) had a chart diagnosis of dementia or documented cognitive impairment as indicated by a Mini-Mental State Examination score of <24; (2) had major visual, hearing, or motor impairment; (3) had a terminal illness expected to result in death within 1 year; or (4) were patients with ≥2 emergency room visits or hospitalizations in the previous 6 months or determined by the clinical team to be psychiatrically or medically unstable.

Measures

The usability of the CARES app was evaluated using the System Usability Scale (SUS). The SUS is a widely used, valid, reliable 10-item scale that assesses system usability [29]. Sample questions include “I think that I would like to use this system frequently” and “I thought the system was easy to use.” Response options range from 1 (*strongly disagree*) to 5 (*strongly agree*). Scores range from 0 to 100, with higher scores indicating better usability [30]. A mean SUS of ≥68 is indicative of an above-average user experience [30]. A system with an SUS score of >70 is considered acceptable. Scores of >85 are considered “excellent,” scores between 71 and 84 are considered “good,” and scores between 51 and 70 are considered “OK” [30].

Caregiver burden was assessed using the Zarit Burden Interview–Short Form. The Zarit Burden Interview–Short Form is a 12-item scale that evaluates caregivers’ physical burden, financial burden, interpersonal burden, and health [3]. The Zarit Burden Interview–Short Form is a valid scale for evaluating burden in caregivers of community-dwelling individuals with dementia [3]. Sample questions include “do you feel that because of the time you spend with your relative that you don’t have enough time for yourself” and “do you feel that you have lost control of your life since your relative’s illness.” Response options range from 0 (*never*) to 4 (*nearly always*). Scores range from 0 to 48, with higher scores indicating higher levels of burden.

The Modified Caregiver Strain Index is a 13-item scale that was used to assess strain and its consequences on caregivers’ overall health. The Modified Caregiver Strain Index is a stable and reliable measure of strain among long-term caregivers [7]. Sample questions include “caregiving is inconvenient” and “I feel completely overwhelmed.” Response options include “yes, on a regular basis,” “yes, sometimes,” and “no.” Scores range from 0 to 26, with higher scores indicating higher levels of strain.

Caregiver stress was assessed using the Caregiver Self-Assessment Questionnaire (CSAQ). The CSAQ is an 18-item scale that assesses the stress levels of family caregivers [31]. The CSAQ is a valid scale for individuals caring for people with dementia [31]. Sample questions include “during the past week or so, I have felt that I couldn’t leave my relative alone” and “during the past week or so, I have been satisfied with the support my family has given me.” Response options for questions 1 to 16 include “yes” or “no.” Caregivers are considered to have high levels of stress if they respond with “yes” to ≥10 questions. Question 17 asks the caregiver to rate their level of stress from “not stressful” to “extremely stressful” on a scale from 1 to 10. Caregivers are considered to have high levels of stress if they score ≥6 on question 17.

A semistructured interview was administered to assess the acceptability of the CARES app from the perspective of the participants. The interview questions were informed by the Consolidated Framework for Implementation Research (CFIR).

The CFIR is a meta-theoretical framework that guides implementation research and is used to systematically assess potential barriers to and facilitators of implementing an intervention [32,33]. Previous studies on the feasibility of web-based tools and interventions have used the CFIR to guide qualitative analysis, identify aspects of implementation feasibility, and determine areas of improvement and adaptation to better meet the needs of users (eg, see the study by Lawson et al [34]). The PI used the CFIR Interview Guide Tool to develop the interview questions. Interview questions covered CFIR domains such as intervention characteristics (what key attributes of the intervention influence the success of implementation), patients' needs and resources (the extent to which patient needs and barriers to and facilitators of meeting those needs are accurately known), implementation climate (shared receptivity of involved individuals to an intervention and the extent to which the use of that intervention will be supported), self-efficacy (individuals' beliefs in their own capacity to successfully implement the intervention), and evidence strength and quality (individuals' perceptions of the quality and validity of the intervention) [32]. Interview guide questions included the following: "what would you change about the CARES system and intervention?" "How well does the intervention align with your values and norms?" "What barriers will family caregivers of people with dementia face to delivering or participating in the intervention?"

Procedures

Recruitment

The PI (CCP) met with staff members at the Aging Resource Center to discuss the purpose of the study and the recruitment process. Together, they identified potential groups within and outside the Dartmouth-Hitchcock Aging Resource Center to recruit both former family caregivers of people with dementia to be trained in the delivery of the CARES app and current family caregivers of people with dementia to receive the CARES intervention. If the potential participants met the inclusion criteria, they were contacted via email and provided with a description of the study. Participants were told that the study was for an honors psychology thesis that aimed to assess the usability, acceptability, and preliminary effectiveness of a technology-based and caregiver-delivered peer support program, CARES, intended to help current family caregivers of people with dementia manage burden, strain, and stress levels. If they were interested in the study, they agreed to meet with the PI via HIPAA-compliant videoconferencing software or telephone for a baseline interview. The baseline interview lasted approximately 20 to 60 minutes. For the baseline interview, the PI read through the informed consent forms and answered participants' questions regarding the content of the study. After the baseline interview, participants who provided informed consent for the study were sent a copy of their informed consent form; sent a link to a baseline survey on Qualtrics with questions on demographic information and their current burden, stress, and strain levels; and provided with information on how to download and log in to the CARES app.

Training

Once 3 former caregivers of people with dementia were recruited, they completed 6 hours of CARES training over 2 days through HIPAA-compliant videoconferencing software. The CARES training was adapted from the Digital Peer Support Certification [35]. Fortuna et al [35,36] found that a combination of educational training (the Digital Peer Support Certification) and management of the PeerTECH system increased peer support specialists' capacity to use features of the digital peer support technology [35]. The training was based on adult learning theory and experiential learning theory. Adult learning theory suggests that adults learn best when they use past lived experiences and past developed skills and knowledge to enhance their learning process [37]. Experiential learning theory consists of four principles: (1) concrete experience, (2) observation and reflection, (3) abstract conceptualization, and (4) active experimentation [38]. In the CARES training, former caregivers were taught new skills; asked to reflect on and connect new knowledge and skills to past experiences and situations; and, finally, asked to practice the new skills they had learned. Techniques such as reinforcement, summation, and teach-back were used in the CARES training to promote the mastery of peer support skills [35].

The CARES training focused on instructing the former caregivers on the basic principles and competencies in the delivery of digital peer support and evidence-based practices to manage stress. The training included an overview of the following topics: peer support, health and aging, engaging older service users with technologies, teaching older adults how to use technology, life review, acceptance, mindfulness, coaching and making a plan of action, recognizing negative thoughts, the art and science of adult learning theory, and the role of family and caregivers in technology. Facilitated group discussions were paired with a PowerPoint (Microsoft Corp) presentation. The PowerPoint presentation was provided to all caregiver peer supporters at the end of the training. After the 6-hour training, participants who provided informed consent for the study were sent a copy of their informed consent form, a link to a baseline survey on Qualtrics with questions on demographic information, a copy of the caregiver peer support training, and information on how to download and log in to the CARES app. The PI was available for technological support from Monday to Saturday between the hours of 9 AM and 5 PM.

Fidelity

Over the course of the 2-week field usability study, a member of the research team tracked the CARES app messages between the caregiver peer supporters and assigned participants to evaluate whether the caregiver peer supporters were providing peer support in line with the training.

Informed Consent

Before the 2-week field usability study, the participant was provided with a description of the study, shown the CARES app, and read aloud the consent form word for word. Participants were evaluated according to the study criteria. If the participant was eligible and provided informed consent to take part in the study, they then completed the baseline survey on Qualtrics.

Quantitative Analyses

Descriptive statistics were used to describe the demographic characteristics of the study sample and the results of the SUS. A paired-sample 2-tailed *t* test was conducted to assess the difference between the baseline and 2-week burden, stress, and strain level scores for statistical significance. All incomplete survey responses were excluded from the analyses. Descriptive statistics and analyses were computed using the SPSS software (version 26; IBM Corp).

Qualitative Analyses

Interview data were analyzed using the rigorous and accelerated data reduction (RADar) technique. The RADar method helps streamline the process of qualitative data analysis through its ability to organize, reduce, and analyze data in user-friendly software packages such as Excel (Microsoft Corp) [39]. In accordance with the RADar methodology, the interview transcripts were formatted into an all-inclusive Excel spreadsheet. The Excel spreadsheet column headings included participant number, question, response, code, and notes. One researcher assigned codes to each response. A priori codes and themes related to the CFIR framework were identified. These codes included the acceptability of CARES, user needs and resources, intervention characteristics key to the success of the implementation, self-efficacy, quality and validity of the intervention, and receptivity of users. Codes were derived from the interview data by carefully reviewing the transcribed text. The all-inclusive data table was then reduced to include only content that answered the overarching research questions and was of primary interest to the analysis. The remaining text and codes were then organized into themes that applied the CFIR framework and were adjusted to best fit the content covered in the qualitative interviews. The percentage of each theme was determined by dividing the frequency with which the theme was present in the interview quotes by the total number of interview quotes.

Results

Overview

Table 1 presents the sociodemographic characteristics of the sample at baseline. The sample of current family caregivers (9/12, 75%; the participants) had a mean age of 67.3 (SD 15.1) years and was predominantly female (6/9, 67%) and White (9/9, 100%), and most of them (4/9, 44%) cared for a spouse. The sample of former family caregivers (3/12, 25%; the caregiver peer supporters) had a mean age of 68.3 (SD 11.0) years and was predominantly female (3/3, 100%) and White (3/3, 100%). One of the caregiver peer supporters had experience caring for a parent with dementia (1/3, 33%), one had experience caring for a spouse with dementia (1/3, 33%), and the other had experience caring for a sibling with dementia (1/3, 33%).

Table 1. Sociodemographic characteristics of study participants (N=12).

Characteristics	Participants (n=9)	Caregiver peer supporters (n=3)
Sex, n (%)		
Male	3 (33)	0 (0)
Female	6 (67)	3 (100)
Age (years), mean (SD; range)	67.3 (15.1; 42-87)	68.3 (11.0; 61-81)
Race (White), n (%)	9 (100)	3 (100)
State, n (%)		
Connecticut	1 (11)	1 (33)
Florida	1 (11)	0 (0)
Massachusetts	2 (22)	1 (33)
New Hampshire	3 (33)	0 (0)
Vermont	2 (22)	1 (33)
Smartphone owner, n (%)		
Yes	9 (100)	3 (100)
No	0 (0)	0 (0)
Relation to relative with dementia, n (%)		
Child	3 (33)	1 (33)
Parent	1 (11)	0 (0)
Sibling	1 (11)	1 (33)
Spouse	4 (44)	1 (33)

A total of 3 participants were excluded from the study. In total, 33% (1/3) of these participants did not complete the 2-week CARES app field usability study and interview session. A total of 67% (2/3) of these participants did not complete the 2-week CARES app due to a delay in the 2-week field usability study with one of the caregiver peer supporters. The caregiver peer supporter delayed the start of their 2-week field usability study because the CARES app was not functioning on their smartphone. The remaining 9 participants and all 3 caregiver peer supporters completed the CARES intervention.

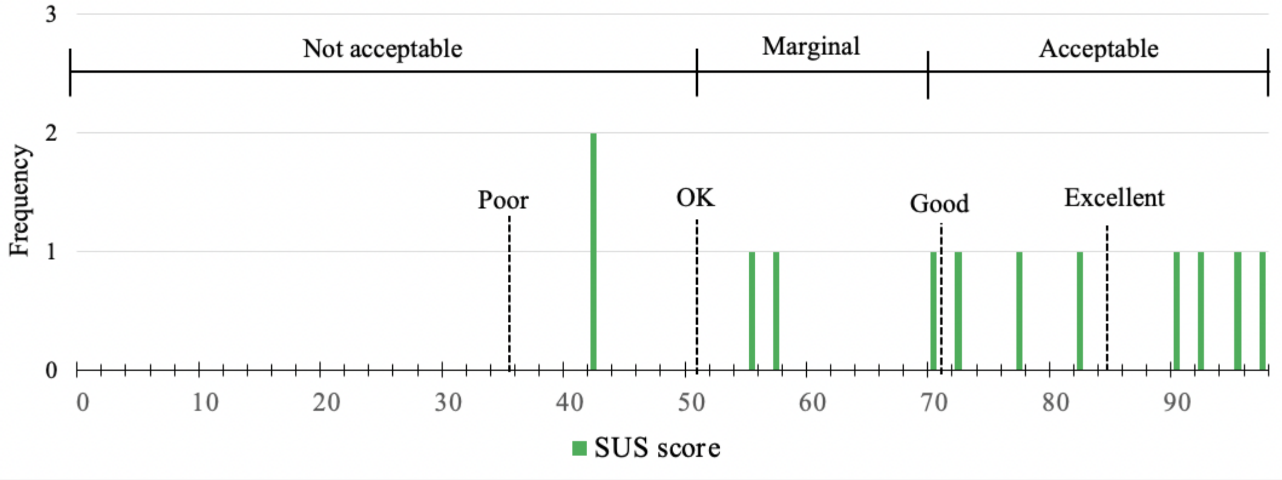
Usability of CARES

Overall, participants reported above-average system usability on the SUS, with a mean score of 72.92 (SD 19.77) and a range from 42.50 to 97.50. Most participants found CARES to be an acceptable (8/12, 67%) and good or excellent (7/12, 58%) system. Specifically, the current family caregivers receiving support reported above-average system usability, with a mean score of 76.94 (SD 19.03) and a range from 42.50 to 97.50. Caregiver peer supporters reported marginal usability, with a mean score of 60.83 (SD 20.21) and a range from 42.50 to 82.50 on the SUS. One current caregiver and one caregiver peer supporter rated CARES as below OK. The distribution of the acceptability and adjective ratings as indicated by Bangor et al [30] are shown in Figure 6.

On average, participants sent 27 (SD 6.88) messages, with a range from 15 to 36, to their assigned caregiver peer over the course of the 2 weeks. All participants engaged weekly with the app. On average, participants reviewed 44% of the library resources over the course of the intervention, with a range from 0% to 100%.

Figure 6 shows the distribution of participant and interventionist (N=12) responses to the SUS. The x-axis marks the individual SUS scores, and the y-axis marks the frequency of the scores. Acceptability ranges and adjective ratings are informed by the work by Bangor et al [30].

Figure 6. Distribution of results of the System Usability Scale (SUS) for study participants and interventionists.



Barriers to and Acceptability and Facilitators of Using CARES

Overview

Regarding the acceptability of the CARES app and intervention, 24 codes and 6 themes related to the acceptance of CARES were identified. The 6 themes were improving the CARES app and intervention, acceptability of the CARES app features and design, value of the caregiver-peer relationship, barriers and limitations of CARES, caregiver needs and preferences, and caregiver challenges. The themes are listed in Table 2.

Table 2. Themes from qualitative analysis of the semistructured interviews.

Theme	Description	Participant quotes
Improving the CARES[a] app and intervention	The participants and caregiver peer supporters provided input on how to improve the CARES app features and study protocol.	• "If you're dealing with some kind of messaging app...there's no point if there's no notification because nobody will think to go check it." [participant 5] • "One of my concerns was that a couple of the people didn't understand what they were supposed to be doing, or how to interact with the app. I think that there needs to be a little bit more explanation upfront before we start interacting with [participants]." [caregiver peer supporter 14]
Acceptability of the CARES app features and design	Most participants and caregiver peer supporters found the CARES app to be acceptable for providing support to caregivers of individuals with dementia.	• "I knew that right after, like the first couple of messages back and forth, I was like, this is a great idea. Because it's convenient. It's easy. It's, you know, nonjudgmental. It's just what you want from a support thing." [participant 7] • "I think that it's like a personal support group. That's what it is. And it's in your pocket, because it's on your phone and it's delivered in an app, you don't have to leave your home, you don't have to try to arrange coverage to somebody to sit with, you know, your loved one, so that you can sneak out for an hour and then worry the whole hour that you're out about what's going on at home." [caregiver peer supporter 13]
Value of the caregiver-peer relationship	Participants specifically highlighted their appreciation of the caregiver-peer relationship.	• "You really felt like you had somebody to reach out to in the times when things were really stressful and really felt overwhelming. It just was somebody that you could connect with that knew how you were feeling." [participant 7] • "...knowing that there's somebody out there who is thinking about me and my situation." [participant 10]
Barriers and limitations of CARES	This refers to challenges that users face in using the CARES app and delivering the CARES intervention.	• "The only barrier I see is, if somebody doesn't have access to an iPhone." [participant 8] • "I'm so busy doing the caregiver stuff, and all the other sort of managing, so that if I have any time to myself, I would want to be doing other things that, you know, that didn't involve caregiving. So I wouldn't be apt to wanting to take the time out of those personal times." [participant 10]
Caregiver needs and preferences	Most participants and caregiver peer supporters found that the CARES app met the needs and preferences of family caregivers of people with dementia.	• "I think that a lot of caregivers will love [CARES] too. You know, the doctors are doctors, they're doing the medical part of it. They don't even think about the emotional part that the caregiver is going through." [caregiver peer supporter 14] • "It's more I know that I have to take care of myself in order to be a better caregiver. And I can't do that if I'm not feeling good about myself. And yet, I didn't know how to do that...So I think I think [my peer supporter] was really good the way that she, she validated my feelings and, and was out there for me." [participant 6]
Caregiver challenges	Participants highlighted the overall challenges of caring for a family member with dementia.	• "My daughters were very concerned about me not getting out enough on my own." [participant 9] • "One of the biggest struggles that I have is finding people that understand what I'm going through...it's very difficult to find somebody that I could connect with that had been through what I was going through, and that I felt comfortable really voicing my feelings to. So I think the idea behind the app is like, great, actually because, I mean, it's like somebody that's always there that knows exactly what you're feeling and what you've been through." [participant 7]

[a]CARES: Caregiver Remote Education and Support.

Improving the CARES App and Intervention

The most prevalent theme was improving the CARES app and intervention. This theme comprised 3 subthemes: improving the study, improving the app features, and technological difficulties. Improving the CARES app and intervention constituted 26% of the themes discussed in the interviews. The participants and caregiver peer supporters provided input on how to improve the CARES app features and study protocol. For example, a few participants noted problems receiving notifications when their assigned peer sent them a message through the app. Participant 5 mentioned the following:

> *...if you're dealing with some kind of messaging app...there's no point if there's no notification because nobody will think to go check it.*

Some participants found aspects of the app difficult to maneuver. At times, the video chat would not connect properly, and the library resources would appear blank. Caregiver peer supporter 13 thought that the home page of the CARES app should clarify the content of the library resource feature to increase interaction with the materials contained within the library:

> *...you don't know that the resources are there, they're offering them to you, but you can't find them. It's like you're going on a board game, and you don't know where to get off.*

Other participants provided ideas on how to enhance the *app*. For example, caregiver peer supporter 15 suggested that the app include the option to communicate via telephone along with video chat and messaging as building rapport can be "very difficult to do with texting." Other participants suggested adding a support group–like feature where participants could connect with multiple peers rather than just 1. Participant 5 suggested the following:

> *...having kind of a group chat or a message board, where you could just be like, okay, just, you know,*

> *like, venting for a second or whatever, that could be really kind of an expansion from beyond just the one on one.*

The caregiver peer supporters also provided feedback on how to improve the peer support training. The caregiver peer supporters thought that clarification of the peer supporter and participant roles and expectations would have improved the participant-peer interaction and relationship. For example, caregiver peer supporter 14 stated the following:

> *...one of my concerns was that a couple of the people didn't understand what they were supposed to be doing, or how to interact with the app. I think that there needs to be a little bit more explanation upfront before we start interacting with* [participants]

Both participants and caregiver peer supporters recommended holding a training specifically on the features of the CARES app. Caregiver peer supporter 13 said the following:

> *I think that it would have also been beneficial to download the app, and then on part of the training, you walk through it with us, and we...just play with it...like hands on learning.*

The caregiver peer supporters suggested that the training should include additional practice using the CARES app and suggestions on how to initiate relationships with their assigned participants. Participants proposed that, in the future, a researcher should explain the features available on the CARES app and clarify their expectations for both the caregiver peer and the participant. Finally, participants recommended that future studies should match peers with participants based on their relationship to the individual with dementia they are caring for.

Acceptability of the CARES App Features and Design

The second theme, acceptability of the CARES *app* features and design, constituted 25% of the themes discussed in the interviews. Overall, most participants and caregiver peer supporters found the CARES app to be acceptable for providing support to caregivers of individuals with dementia. All participants agreed that the main purpose of the app was to connect caregivers with peers with a similar lived experience. Most participants interacted most with the messaging feature. Participant 5 mentioned the following:

> *I think the main point, or the main feature, is the connection to peers.*

Participants and caregiver peer supporters emphasized the convenience and accessibility of the CARES app. For example, participant 7 said the following:

> *I knew that right after, like the first couple of messages back and forth, I was like, this is a great idea. Because it's convenient. It's easy. It's, you know, non-judgmental. It's just what you want from a support thing. And it's also sort of on your own terms, though, because you get a message. So like, if you didn't want to respond instantly, you can kind of gather your thoughts, and you have time to respond, unlike a regular back and forth support group where if somebody asked me a question, I kind of have to have an answer...So. I think this was better, because I had a few minutes to really think through what she was asking me...and I had a minute to kind of gather my thoughts, and then I could just type it back.*

Participant 5 mentioned that, in contrast to in-person support groups where caregivers may "struggle with getting away even for an hour out of an apartment or their house for an hour," they "like the fact that they can [use CARES] over an iPad or an iPhone...I think that definitely makes it more accessible and easy." Caregiver peer supporters agreed that the CARES app was an appropriate tool for current caregivers of people with dementia. Caregiver peer supporter 13 shared the following:

> *I think that it's like a personal support group. That's what it is. And it's in your pocket, because it's on your phone and it's delivered in an app, you don't have to leave your home, you don't have to try to arrange coverage to somebody to sit with, you know, your loved one, so that you can sneak out for an hour and then worry the whole hour that you're out about what's going on at home.*

However, while all participants found the messaging feature of the CARES app beneficial, only some of the participants used the library resources (specifically the mindfulness, values, and negative thinking information). The "goals" and "wellness" features of the app were the least used by participants.

Value of the Caregiver-Peer Relationship

The third theme, the value of the caregiver-peer relationship, constituted 15% of the themes discussed in the interviews. Participants specifically highlighted their appreciation of the caregiver-peer relationship. Most participants emphasized that the purpose of the CARES app was the caregiver-peer connection. Participants found that their assigned caregiver peer supporters were knowledgeable, understanding, and validating. Participant 7 mentioned the following:

> *...you really felt like you had somebody to reach out to in the times when things were really stressful and really felt overwhelming. It just was somebody that you could connect with that knew how you were feeling.*

They also shared the following:

> *...it was also nice, because the person that I was connecting with chose to be connected with somebody. So it wasn't like you felt like you were burdening somebody else with your feelings.*

The caregiver peer supporters also found value in the caregiver-peer relationship. Caregiver peer supporter 13 said the following:

> *I think it's a great way to connect with caregivers. And it's easy because you can just type a message and somebody picks up that message at that point, and it's like having a support system at your fingertips...So I think when it works correctly, it would be a great effective tool for caregivers...because I found that caregivers in general, you know, they're*

> *not healthcare professionals. And they're expected to be in that role as part of the healthcare team. And they don't even know like, who to call, like, do I call you when my husband is running a fever and needs care?...I feel for these people that are just like plunked down in this role. And they don't have anyone. Even [one participant] said, I feel great, just knowing that you're there. And you provided this emotional support for me. And it's nice to know that you can reach out to someone that gets it. And that's exactly what it is...You know, it's a lifeline for people.*

Participants and caregivers found that participants appreciated the intervention "knowing that there's somebody out there who is thinking about me and my situation" (participant 10). The caregiver-peer connection was central to the CARES intervention.

Barriers and Limitations of CARES

The fourth theme, barriers and limitations of CARES, constituted 12% of the themes discussed in the interviews. Participants highlighted 2 main barriers and limitations: access to technology and time constraints. First, participants recognized that some caregivers may not have access to a smartphone or tablet or may not have adequate internet connection. Participant 8 mentioned that "the only barrier I see is, if somebody doesn't have access to an iPhone." Other participants noted that older adults may not be comfortable using technology. Time constraints and competing priorities were cited as other barriers that participants might face when using the CARES app. For example, participant 10 shared the following:

> *I'm so busy doing the caregiver stuff, and all the other sort of managing, so that if I have any time to myself, I would want to be doing other things that, you know, that didn't involve caregiving. So I wouldn't be apt to wanting to take the time out of those personal times.*

Finally, many participants and caregiver peer supporters thought that a 2-week period did not provide enough time to fully explore the features of the CARES app and peer relationship. For example, participant 7 said that "well, I didn't really check it out as much as I wanted to, because two weeks is not a lot of time."

Caregiver Needs and Preferences

The fifth theme, caregiver needs and preferences, constituted 11% of the themes discussed in the interviews. Overall, most participants and caregiver peer supporters found that the CARES app met the needs and preferences of family caregivers of people with dementia. Specifically, participants found that the CARES app and intervention provided social and emotional support. Participant 6 shared the following:

> *I think sometimes it's important for people who are caregivers to just be able to say how they're, how they're feeling.*

Caregiver peer supporter 14 mentioned the following:

> *I think that a lot of caregivers will love [CARES] too. You know, the doctors are doctors, they're doing the medical part of it. They don't even think about the emotional part that the caregiver is going through.*

For many participants, the CARES app and intervention provided individualized care and support. For example, participant 6 shared the following:

> *...it's more I know that I have to take care of myself in order to be a better caregiver. And I can't do that if I'm not feeling good about myself. And yet, I didn't know how to do that. Because I'm so tied up, so wrapped up in this. So I think I think she was really good the way that she, she validated my feelings and, and was out there for me.*

However, participants also stressed that caregiver peer supporters should have a general understanding of individuals' backgrounds and access to resources when providing support. For example, participant 5 mentioned the following:

> *...there's a wide range of resources that people have. I've seen this in the caregiver group I am a part of. Some people have planned well, and can afford help and support and some people don't have that, and they have no family around...The support person...should have an awareness of saying...you need to just hire somebody to come in for an hour every day.*

Caregiver peer supporters need to have an awareness of participants' available resources and priorities when providing support. Finally, most caregiver peer supporters also felt that CARES met their needs and preferences as former caregivers. For example, caregiver peer supporter 13 shared the following:

> *I always feel purposeful when giving back. That's most of the reason that I do coaching...I always feel a feeling of purpose. And there's a lot of emotional support and a feeling of gratification that comes from giving that emotional support, because you have the lived experience that you can share with these other caregivers. And if you don't have that experience, you don't get it like you can be in that role, but you don't truly get what they're going through.*

Caregiver peer supporters felt a sense of purpose while delivering the CARES intervention.

Caregiver Challenges

Finally, caregiver challenges was an emerging theme that constituted 5% of the themes discussed in the interviews. In their interviews, the participants highlighted the challenges of caring for a family member with dementia. Topics included difficulty taking time for oneself, frustration, anxiety about the unknown and upcoming changes, guilt, and the inability to find people who understand their situation. For example, participant 9 mentioned the following:

> *...my daughters were very concerned about me not getting out enough on my own.*

Participant 6 said the following:

> *I just wish with this disease they could say, well, in six months you may experience this and another six months you may experience that.*

Participant 2 stated the following:

> *...there are certain days you just feel more on top of things than others.*

Some participants shared how the CARES app and intervention addressed their challenges:

> *But one of the biggest struggles that I have is finding people that understand what I'm going through. Because if you have not provided care for somebody with dementia, you really don't understand like you can try to understand you can have the knowledge. But if you don't have the experience, it's very difficult to find somebody that I could connect with that had been through what I was going through, and that I felt comfortable really voicing my feelings to. So I think the idea behind the app is like, great, actually because, I mean, it's like somebody that's always there that knows exactly what you're feeling and what you've been through.* [participant 7]

Others did not voice whether CARES attended to the specific challenges they faced as caregivers.

Preliminary Effectiveness of CARES

Participants' decreases in burden, strain, and stress levels were not significant. However, we were not powered to find significance; rather, the purpose of this study was feasibility and acceptability. Non–statistically significant improvements were observed in all measures. The results of the baseline and 2-week posttreatment assessment for the 9 participants who completed the CARES intervention are shown in Table 3.

Table 3. Changes in outcomes from before to after the field usability study (2 weeks) for study participants (N=9)[a].

Measure	Participants, n (%)	Pretest assessment, mean (SD)	Posttest assessment, mean (SD)	*t* test (*df*)	*P* value	Effect size (Cohen *d*)
ZBI-12[b]	9 (100)	22.44 (10.04)	21.44 (11.34)	1.25 (8)	.25	0.42
MCSI[c]	9 (100)	11.22 (6.62)	11.00 (7.44)	0.41 (8)	.70	0.13
CSAQ[d] total	9 (100)	6.56 (3.97)	5.78 (4.12)	1.31 (8)	.23	0.44
CSAQ stress	9 (100)	5.28 (2.71)	4.89 (2.84)	1.00 (8)	.35	0.33

[a]A 2-tailed paired *t* test was used to assess statistical significance.

[b]ZBI-12: Zarit Burden Interview–Short Form.

[c]MCSI: Modified Caregiver Strain Index.

[d]CSAQ: Caregiver Self-Assessment Questionnaire. The CSAQ total is the average number of "yes" responses on the CSAQ, and the CSAQ stress score is the mean score on question 17 of the CSAQ, which asked participants to rate their level of stress on a scale from 1 to 10.

Discussion

Principal Findings

The purpose of this study was to evaluate the usability, acceptability, and preliminary effectiveness of the CARES app and intervention. The pilot study demonstrated that a 2-week peer-delivered and technology-supported mental health intervention (CARES) was acceptable for both former family caregivers of people with dementia who delivered peer support and current family caregivers of people with dementia who received peer support. Current caregivers reported above-average usability of CARES, and former caregivers reported marginal usability. The pilot study demonstrated that it is possible to train former caregivers in peer support and the delivery of CARES and integrated psychoeducation and mental health interventions using technology. CARES was associated with non–statistically significant improvements in burden, stress, and strain levels.

The usability of the CARES app was demonstrated using the SUS. Most caregivers found the CARES app to be an acceptable and good system with above-average usability. The CARES app allowed peer caregivers to provide individualized support and provided caregivers with access to evidence-based mental health resources on topics such as mindfulness and acceptance. The usability of CARES was also demonstrated through participants' capacity to use the smartphone and tablet app, completion of library resources on the app, and use of the messaging chat feature. Overall, most participants and caregiver peer supporters agreed that the CARES app and intervention were an acceptable tool to support family caregivers of people with dementia. However, participants and caregiver peer supporters identified areas in which the usability and acceptability of the app could be improved, and the caregiver peer supporters specifically reported marginal usability of the CARES app. Caregiver peer supporters may have reported below-average usability because of latency in messaging and other technological difficulties using the app. Future studies should examine the cause of differences in usability scores between current and former family caregivers and update the CARES app accordingly.

Participants suggested that improvement of technological features would strengthen the app's ability to achieve its purpose of connecting caregivers with peers. Some participants and peers faced technological difficulties while using the app. For example, at times, the app would not notify participants of new messages. This created a barrier in participants' and peers' ability to communicate efficiently and effectively. Participants also provided suggestions on how to improve the app features. For example, participants suggested adding a telephone feature to the app. Participants believed that adding a telephone feature would allow caregivers to communicate via the medium of their preference and, therefore, increase their comfort level with technology and the caregiver-peer relationship. Participants also suggested adding a feature through which they could connect with more than one caregiver with a similar lived experience and suggested adding more caregiver-specific resources to the library resource page.

Despite technological limitations, most participants and peers found that the CARES app was an acceptable support intervention for family caregivers of people with dementia. Participants identified the caregiver-peer connection as the principal feature of the CARES app. Participants labeled the message-based support as convenient, easy-to-use, accessible, and individualized. The caregiver peer supporters were described as knowledgeable, understanding, validating, and supporting. Participants felt that the shared lived experience offered by the former caregivers better matched their needs and preferences for emotional and social support compared to professional medical and health care workers. On the other hand, the former family caregivers felt a sense of purpose and gratification while delivering the CARES intervention.

These findings suggest that technology- and peer-based interventions are usable and acceptable to family caregivers of people with dementia and that a smartphone app is a promising tool to support the mental and emotional health and well-being of family caregivers outside of an in-person or clinical setting. Task shifting informal caregiver digital mental health services to community members with lived experience has the potential to provide acceptable mental health interventions to family caregivers of people with dementia while addressing the current barriers and challenges with respect to accessing support. While it is estimated that nearly 153 million older adults will have dementia worldwide by 2050, mental health services for caregivers and their family members are limited due to an insufficient number of adequately trained geriatric mental health care providers [16,40]. Peer-delivered and technology-supported interventions have the potential to provide mental health services to family caregivers of people with dementia that are easily accessible and effective [36]. While caregiver psychosocial interventions have faced limitations due to time commitments, geographical location, requirements to meet in person, and failure to address the individualized needs of the caregiver, former and current family caregivers of people with dementia reported that the CARES app and intervention addressed the unique needs and experiences of consumers.

The results of the study support the hypothesis that former family caregivers of people with dementia have the knowledge and skillset to deliver trained peer support to current family caregivers. In previous studies, peers have been reported to be particularly effective at engaging participants in interventions. By sharing a lived experience, peers have the ability to develop alliances with participants and are viewed as more credible than traditional clinicians and providers [21]. With the use of technology-based messaging and support, participants were able to provide individualized support to caregivers at any time and location.

However, there are barriers and limitations to consider when using and delivering the CARES app and intervention. First, the CARES app is not accessible to caregivers who do not have a smartphone or adequate internet connection. In addition, time constraints may limit caregivers' ability to interact with the app and their assigned peer. Offering peer support to family caregivers of people with dementia may place stress on the interventionists. Caregiver peer supporters involved in the study should be offered mental health support while delivering the intervention. Finally, participants and peer supporters suggested that the abrupt ending of the 2-week field usability study may leave caregivers without the resources and support systems they came to rely on to manage well-being. Future studies should provide caregiver participants with additional caregiver support resources at the end of the CARES field usability study.

This study is not without limitations. First, one member of the research and intervention development team conducted a qualitative analysis of the interview data and identified codes and themes, which may have biased the results. In addition, member checking was not used to validate the findings and assess the accuracy of the qualitative results. Second, caregivers' input was not included in the initial development of the CARES app. Stakeholder engagement in the early stages of intervention development is essential to ensure that the modality and components are relevant to the community [41]. Future studies will incorporate caregivers' feedback to further improve and adapt the CARES app and intervention. Third, some participants and caregiver peer supporters experienced technological difficulties with the CARES app. For example, 22% (2/9) of the participants and 33% (1/3) of the caregiver peer supporters were unable to receive notifications and, therefore, had delayed responses to messages. Another caregiver peer supporter experienced challenges sending messages and, consequently, had to delay the start of their 2-week field usability study. These technological difficulties could have impacted the results of the field usability study and SUS. Fourth, the participant response rate was not tracked during the recruitment process. Future studies should track the response rate to improve recruitment procedures and decrease bias (eg, nonresponse bias). Fifth, data on the frequency of use of CARES app features such as video chat, goal setting, and wellness plans were not tracked across participants. This information will be helpful for further understanding the acceptability and usability of CARES. Sixth, the caregiver peer supporters were not assessed regarding whether they had learned the topics presented in the training, and the competence of the training was not assessed. Therefore, it is unknown whether the training sufficiently educated caregivers on the delivery of peer support. Seventh, the fidelity of caregiver peer support was not systematically evaluated through audio interactions; rather, message data determined high fidelity to the peer support model. Future research will assess message and audio interactions to determine fidelity to the caregiver peer support model of care through audio recordings on the app. Caregiver peer supporters' deviation from the training may have biased the results. Eighth, the participant eligibility criteria were broad and included all family caregivers of individuals with dementia aged ≥18 years. Criteria such as whether the caregiver lived with their care recipient and the number of hours spent caring for their relative with dementia could impact the acceptability, usability, and effectiveness of the CARES app and intervention. Future studies should investigate the influence of differing caregiver roles and responsibilities on CARES. Finally, demographic information on factors such as education and income level was not collected, which may have affected the results and the perceived usability of the app.

While the aim of this study was to evaluate the usability of CARES and assess the acceptability and potential barriers to using the CARES technology, future studies would benefit from a larger sample size and a longer trial duration. In the qualitative interviews, participants shared that they wished they had a longer period to explore the app and the caregiver-peer relationship. While the length of the study and sample size were consistent with a field usability study, longer trials would allow participants to further assess the usability and acceptability of the app and whether it meets their needs and preferences as caregivers [42]. A longer trial would also more accurately reflect the length and fluctuation of the dementia caregiving experience. Future studies would also benefit from a randomized controlled design (eg, an intervention group with CARES and caregiver peer support vs a control group) and a control of covariables influencing the outcomes to evaluate the effectiveness of the CARES training, intervention, and app in reducing burden, strain, and stress in family caregivers of people with dementia. For example, future studies should examine the influence of age; relationship to the individual with dementia receiving care; stage of dementia; years providing care; severity of dementia symptoms; severity of the family caregiver's stress, strain, and burden levels; and the use of other caregiver support treatments and interventions on the effectiveness of the CARES app.

Future studies should also assign caregivers to peers based on dementia diagnosis or relationship to the individual with dementia. Matching caregivers with peers based on shared caregiver lived experiences may moderate the effectiveness of the CARES intervention. Future research would also benefit from a more diverse group of participants. Recruitment procedures should focus on recruiting a sample of participants representative of the demographics of the greater caregiver population. This includes recruiting more caregivers of color and caregivers who identify as male or nonbinary. Finally, future work should address the benefit of caregiver peer support for the family caregivers both delivering and receiving the mental health intervention. Caregiver peer support may have bidirectional effects. The caregivers providing support may see improvements in their mental health and well-being along with those of the participants they are supporting. As indicated by the study interview data, caregiver peer supporters felt a sense of purpose while delivering the intervention. Future studies should further assess the potential bidirectional influence of the CARES app and intervention.

Conclusions

This pilot study demonstrated that it is possible to train former family caregivers of people with dementia to use technology and deliver the CARES mental health intervention to current family caregivers of people with dementia. These findings provide preliminary evidence that a peer-delivered and technology-supported intervention designed to improve burden, stress, and strain levels is feasible and acceptable.

Acknowledgments

This study was funded by the Dartmouth Ethics Institute Sayles grant.

Data Availability

The data sets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

KLF is partners with Social Wellness LLC & Emissary Health, Inc.

Multimedia Appendix 1
Images of the Caregiver Remote Education and Support app features and distribution of the System Usability Scale participant results.
[DOCX File , 2987 KB - humanfactors_v11i1e41202_app1.docx]

References

1. Alzheimer's Association. 2020 Alzheimer's disease facts and figures. Alzheimers Dement 2020 Mar 10;16(3):391-460 [FREE Full text] [doi: 10.1002/alz.12068] [Medline: 32157811]
2. Alzheimer's disease facts and figures. Alzheimer's Association. 2024. URL: https://www.alz.org/media/documents/alzheimers-facts-and-figures.pdf [accessed 2024-03-17]
3. Gratão AC, Brigola AG, Ottaviani AC, Luchesi BM, Souza É, Rossetti ES, et al. Brief version of Zarit Burden Interview (ZBI) for burden assessment in older caregivers. Dement Neuropsychol 2019 Mar;13(1):122-129 [FREE Full text] [doi: 10.1590/1980-57642018dn13-010015] [Medline: 31073389]
4. Brodaty H, Donkin M. Family caregivers of people with dementia. Dialogues Clin Neurosci 2022 Apr 01;11(2):217-228. [doi: 10.31887/dcns.2009.11.2/hbrodaty]

5. Kim H, Chang M, Rose K, Kim S. Predictors of caregiver burden in caregivers of individuals with dementia. J Adv Nurs 2012 Apr 28;68(4):846-855. [doi: 10.1111/j.1365-2648.2011.05787.x] [Medline: 21793872]
6. Liu Z, Heffernan C, Tan J. Caregiver burden: a concept analysis. Int J Nurs Sci 2020 Oct 10;7(4):438-445 [FREE Full text] [doi: 10.1016/j.ijnss.2020.07.012] [Medline: 33195757]
7. Thornton M, Travis SS. Analysis of the reliability of the modified caregiver strain index. J Gerontol B Psychol Sci Soc Sci 2003 Mar 01;58(2):S127-S132. [doi: 10.1093/geronb/58.2.s127] [Medline: 12646602]
8. Walter E, Pinquart M. How effective are dementia caregiver interventions? An updated comprehensive meta-analysis. Gerontologist 2020 Nov 23;60(8):609-619. [doi: 10.1093/geront/gnz118] [Medline: 33226434]
9. Signe A, Elmståhl S. Psychosocial intervention for family caregivers of people with dementia reduces caregiver's burden: development and effect after 6 and 12 months. Scand J Caring Sci 2008 Mar;22(1):98-109. [doi: 10.1111/j.1471-6712.2007.00498.x] [Medline: 18269429]
10. Butler M, Gaugler JE, Talley KM, Abdi HI, Desai PJ, Duval S, et al. Care interventions for people living with dementia and their caregivers. Agency for Healthcare Research and Quality (AHRQ). 2020 Aug. URL: https://effectivehealthcare.ahrq.gov/products/care-interventions-pwd/report [accessed 2024-03-17]
11. Carter G, Monaghan C, Santin O. What is known from the existing literature about peer support interventions for carers of individuals living with dementia: a scoping review. Health Soc Care Community 2020 Jul 09;28(4):1134-1151. [doi: 10.1111/hsc.12944] [Medline: 31919974]
12. Ploeg J, Markle-Reid M, Valaitis R, McAiney C, Duggleby W, Bartholomew A, et al. Web-based interventions to improve mental health, general caregiving outcomes, and general health for informal caregivers of adults with chronic conditions living in the community: rapid evidence review. J Med Internet Res 2017 Jul 28;19(7):e263 [FREE Full text] [doi: 10.2196/jmir.7564] [Medline: 28754652]
13. Weems JA, Rhodes S, Powers JS. Dementia caregiver virtual support-an implementation evaluation of two pragmatic models during COVID-19. Geriatrics (Basel) 2021 Aug 19;6(3):80 [FREE Full text] [doi: 10.3390/geriatrics6030080] [Medline: 34449649]
14. Han J, Guo G, Hong L. Impact of professionally facilitated peer support for family carers of people with dementia in a WeChat virtual community. J Telemed Telecare 2022 Jan;28(1):68-76. [doi: 10.1177/1357633X20910830] [Medline: 32192386]
15. Marziali E, Garcia LJ. Dementia caregivers' responses to 2 Internet-based intervention programs. Am J Alzheimers Dis Other Demen 2011 Feb;26(1):36-43 [FREE Full text] [doi: 10.1177/1533317510387586] [Medline: 21282276]
16. Kunik ME, Mills WL, Amspoker AB, Cully JA, Kraus-Schuman C, Stanley M, et al. Expanding the geriatric mental health workforce through utilization of non-licensed providers. Aging Ment Health 2017 Sep 31;21(9):954-960 [FREE Full text] [doi: 10.1080/13607863.2016.1186150] [Medline: 27243369]
17. Barnett ML, Lau AS, Miranda J. Lay health worker involvement in evidence-based treatment delivery: a conceptual model to address disparities in care. Annu Rev Clin Psychol 2018 May 07;14(1):185-208 [FREE Full text] [doi: 10.1146/annurev-clinpsy-050817-084825] [Medline: 29401043]
18. Fortuna KL, Brooks JM, Umucu E, Walker R, Chow PI. Peer support: a human factor to enhance engagement in digital health behavior change interventions. J Technol Behav Sci 2019 Jun 15;4(2):152-161 [FREE Full text] [doi: 10.1007/s41347-019-00105-x] [Medline: 34337145]
19. Corey KL, McCurry MK, Sethares KA, Bourbonniere M, Hirschman KB, Meghani SH. Predictors of psychological distress and sleep quality in former family caregivers of people with dementia. Aging Ment Health 2020 Feb 27;24(2):233-241 [FREE Full text] [doi: 10.1080/13607863.2018.1531375] [Medline: 30588830]
20. Corey KL, McCurry MK. When caregiving ends: the experiences of former family caregivers of people with dementia. Gerontologist 2018 Mar 19;58(2):e87-e96. [doi: 10.1093/geront/gnw205] [Medline: 28064236]
21. Fortuna KL, DiMilia PR, Lohman MC, Bruce ML, Zubritsky CD, Halaby MR, et al. Feasibility, acceptability, and preliminary effectiveness of a peer-delivered and technology supported self-management intervention for older adults with serious mental illness. Psychiatr Q 2018 Jun 26;89(2):293-305 [FREE Full text] [doi: 10.1007/s11126-017-9534-7] [Medline: 28948424]
22. Solomon P. Peer support/peer provided services underlying processes, benefits, and critical ingredients. Psychiatr Rehabil J 2004;27(4):392-401. [doi: 10.2975/27.2004.392.401] [Medline: 15222150]
23. Zubin J, Spring B. Vulnerability--a new view of schizophrenia. J Abnorm Psychol 1977 Apr;86(2):103-126. [doi: 10.1037//0021-843x.86.2.103] [Medline: 858828]
24. Harris R. ACT Made Simple: An Easy-To-Read Primer on Acceptance and Commitment Therapy. Okland, CA: New Harbinger; 2009.
25. SAMHSA. Core competencies for peer workers in behavioral health services. SAMHSA. 2015. URL: https://www.samhsa.gov/sites/default/files/programs_campaigns/brss_tacs/core-competencies_508_12_13_18.pdf [accessed 2024-03-17]
26. Hayes SC, Pistorello J, Levin ME. Acceptance and commitment therapy as a unified model of behavior change. Couns Psychol 2012 Sep 26;40(7):976-1002. [doi: 10.1177/0011000012460836]

27. Rosenbaum S, Kantner L. Field usability testing: method, not compromise. In: Proceedings of the 2007 IEEE International Professional Communication Conference. 2007 Presented at: IPCC '07; October 1-3, 2007; Seattle, WA p. 1-7 URL: https://ieeexplore.ieee.org/document/4464060 [doi: 10.1109/ipcc.2007.4464060]
28. Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. Behav Res Methods Instrum Comput 2003 Aug;35(3):379-383. [doi: 10.3758/BF03195514]
29. Lewis JR. The system usability scale: past, present, and future. Int J Hum Comput Interact 2018 Mar 30;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
30. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
31. Epstein-Lubow G, Gaudiano BA, Hinckley M, Salloway S, Miller IW. Evidence for the validity of the American Medical Association's caregiver self-assessment questionnaire as a screening measure for depression. J Am Geriatr Soc 2010 Feb 27;58(2):387-388. [doi: 10.1111/j.1532-5415.2009.02701.x] [Medline: 20370867]
32. Technical assistance for users of the CFIR framework. The Consolidated Framework for Implementation Research. URL: https://cfirguide.org/ [accessed 2022-04-17]
33. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
34. Lawson B, Sampalli T, Wood S, Warner G, Moorhouse P, Gibson R, et al. Evaluating the implementation and feasibility of a web-based tool to support timely identification and care for the frail population in primary healthcare settings. Int J Health Policy Manag 2017 Jul 01;6(7):377-382 [FREE Full text] [doi: 10.15171/ijhpm.2017.32] [Medline: 28812833]
35. Fortuna KL, Myers AL, Walsh D, Walker R, Mois G, Brooks JM. Strategies to increase peer support specialists' capacity to use digital technology in the era of COVID-19: pre-post study. JMIR Ment Health 2020 Jul 23;7(7):e20429 [FREE Full text] [doi: 10.2196/20429] [Medline: 32629424]
36. Fortuna KL, Naslund JA, LaCroix JM, Bianco CL, Brooks JM, Zisman-Ilani Y, et al. Digital peer support mental health interventions for people with a lived experience of a serious mental illness: systematic review. JMIR Ment Health 2020 Apr 03;7(4):e16460 [FREE Full text] [doi: 10.2196/16460] [Medline: 32243256]
37. Fortuna KL, Torous J, Depp CA, Jimenez DE, Areán PA, Walker R, et al. A future research agenda for digital geriatric mental healthcare. Am J Geriatr Psychiatry 2019 Nov;27(11):1277-1285 [FREE Full text] [doi: 10.1016/j.jagp.2019.05.013] [Medline: 31196619]
38. Kolb DA, Fry R. Towards an applied theory of experiential learning. In: Cooper C, editor. Theories of Group Process. Hoboken, NJ: John Wiley & Sons; 1975:33-57.
39. Watkins DC. Rapid and rigorous qualitative data analysis. Int J Qual Methods 2017 Jun 08;16(1):160940691771213. [doi: 10.1177/1609406917712131]
40. GBD 2019 Dementia Forecasting Collaborators. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. Lancet Public Health 2022 Feb;7(2):e105-e125 [FREE Full text] [doi: 10.1016/S2468-2667(21)00249-8] [Medline: 34998485]
41. Petkovic J, Riddle A, Akl EA, Khabsa J, Lytvyn L, Atwere P, et al. Protocol for the development of guidance for stakeholder engagement in health and healthcare guideline development and implementation. Syst Rev 2020 Feb 01;9(1):21 [FREE Full text] [doi: 10.1186/s13643-020-1272-5] [Medline: 32007104]
42. Duh HB, Tan GC, Chen VH. Usability evaluation for mobile device: a comparison of laboratory and field tests. In: Proceedings of the 8th conference on Human-computer interaction with mobile devices and services. 2006 Presented at: MobileHCI '06; September 12-15, 2006; Helsinki, Finland p. 181-186 URL: https://dl.acm.org/doi/abs/10.1145/1152215.1152254 [doi: 10.1145/1152215.1152254]

Abbreviations

ACT: acceptance and commitment therapy
CARES: Caregiver Remote Education and Support
CFIR: Consolidated Framework for Implementation Research
CSAQ: Caregiver Self-Assessment Questionnaire
HIPAA: Health Insurance Portability and Accountability Act
PI: principal investigator
RADar: rigorous and accelerated data reduction
SUS: System Usability Scale

Edited by A Kushniruk; submitted 18.07.22; peer-reviewed by L Moo, F Puga, W Tang, H Perez, YH Tung; comments to author 08.11.22; revised version received 02.01.23; accepted 13.04.23; published 27.05.24.

Please cite as:
Collins-Pisano C, Leggett AN, Gambee D, Fortuna KL
Usability, Acceptability, and Preliminary Effectiveness of a Peer-Delivered and Technology-Supported Mental Health Intervention for Family Caregivers of People With Dementia: Field Usability Study
JMIR Hum Factors 2024;11:e41202
URL: https://humanfactors.jmir.org/2024/1/e41202
doi:10.2196/41202
PMID:38801660

Original Paper

Smartphone App Designed to Collect Health Information in Older Adults: Usability Study

Joanne M Murabito[1,2], MD, ScM; Jamie M Faro[3], PhD; Yuankai Zhang[4], MA; Angelo DeMalia[2], BA; Alexander Hamel[3], BA; Nakesha Agyapong[3], BA; Hongshan Liu[2], MA; Eric Schramm[5], PharmD; David D McManus[3,6], MD, ScM; Belinda Borrelli[7], PhD

[1]Section of General Internal Medicine, Department of Medicine, Boston University Chobanian & Avedisian School of Medicine, Boston, MA, United States

[2]Framingham Heart Study, Boston University, Framingham, MA, United States

[3]Department of Medicine, University of Massachusetts Chan Medical School, Worcester, MA, United States

[4]Department of Biostatistics, Boston University School of Public Health, Boston, MA, United States

[5]CareEvolution, Inc, Ann Arbour, MI, United States

[6]Cardiology Division, Department of Medicine, University of Massachusetts Chan Medical School, Worcester, MA, United States

[7]Center for Behavioral Science Research, Henry M. Goldman School of Dental Medicine, Boston University, Boston, MA, United States

Corresponding Author:
Joanne M Murabito, MD, ScM
Section of General Internal Medicine
Department of Medicine
Boston University Chobanian & Avedisian School of Medicine
72 E Concord St Suite L-516
Boston, MA, 02118
United States
Phone: 1 508 935 3500
Fax: 1 508 626 1262
Email: murabito@bu.edu

Abstract

Background: Studies evaluating the usability of mobile-phone assessments in older adults are limited.

Objective: This study aims to identify design-based barriers and facilitators to mobile app survey completion among 2 samples of older adults; those in the Framingham Heart Study and a more diverse sample from a hospital-based setting.

Methods: We used mixed methods to identify challenging and beneficial features of the mobile app in participants from the electronic Framingham Heart Study (n=15; mean age of 72 years; 6/15, 40% women; 15/15, 100% non-Hispanic and White) and among participants recruited from a hospital-based setting (n=15; mean age of 71 years; 7/15, 47% women; 3/15, 20% Hispanic; and 8/15, 53% non-White). A variety of app-based measures with different response formats were tested, including self-reported surveys, pictorial assessments (to indicate body pain sites), and cognitive testing tasks (eg, Trail Making Test and Stroop). Participants completed each measure using a think-aloud protocol, while being audio- and video-recorded with a qualitative interview conducted at the end of the session. Recordings were coded for participant usability errors by 2 pairs of coders. Participants completed the Mobile App Rating Scale to assess the app (response range 1=inadequate to 5=excellent).

Results: In electronic Framingham Heart Study participants, the average total Mobile App Rating Scale score was 7.6 (SD 1.1), with no significant differences in the hospital-based sample. In general, participants were pleased with the app and found it easy to use. A large minority had at least 1 navigational issue, most committed only once. Most older adults did not have difficulty completing the self-reported multiple-choice measures unless it included lengthy instructions but participants had usability issues with the Stroop and Trail Making Test.

Conclusions: Our methods and results help guide app development and app-based survey construction for older adults, while also giving consideration to sociodemographic differences.

(JMIR Hum Factors 2024;11:e56653) doi:10.2196/56653

KEYWORDS
mobile application surveys; mixed methods; electronic data collection; mHealth; mobile health; mobile application; mobile applications; app; apps; application; applications; digital health; digital technology; digital intervention; digital interventions; smartphone; smartphones; usability; usable; usableness; usefulness; utility; health information

Introduction

Adults aged 65 years and older are increasingly using tablets and smartphones and engaging with a range of technologies [1]. Technology use can reduce social isolation [2] and enhance communication with family members and health care providers, thereby, increasing well-being. Further, digital technologies have the potential to improve the health of older adults by facilitating symptom monitoring and self-care management as well as monitoring cognitive and mobility decline. However, older adults often lack confidence in their ability to use technology [3] and report needing assistance with new electronic devices and mobile apps [4]. They face unique challenges with technology including poor eyesight, hearing loss, fine motor skill and sensory limitations, and cognitive decline. These challenges make it essential to understand how technologies can be made more useful to older adults. Perceived value, usefulness, and impact on quality of life are important predictors of technology adoption in this age group [5,6]. In addition, a design that minimizes user frustration will enhance the use and lower the risk of leaving older users out of the technology revolution.

Older adults are often not well represented in user testing of technology [7] due to the restricted age range of research studies, physical or sensory impairments, or because technology studies may be less appealing to them. There are a growing number of smartphone apps that include opportunities for self-management of specific diseases and cognitive self-assessment but the quality and usability of the apps are often unknown especially among groups of older adults and adults from diverse race and ethnic populations [8,9]. In addition, health care providers and hospital systems are increasingly requesting that patients complete previsit health questionnaires electronically, which help with care efficiency and are preferred by providers [10,11]. However, older adults are less likely to access and use patient portals and may have unique needs influencing their use [12-14].

Usability information provides practical recommendations that can help increase patient responsiveness to electronically collected data. For example, studies that have evaluated the usability of mobile apps that assess fall risk demonstrated the importance of simple instructions and clear feedback such as a color change to indicate task completion [6,15,16] A mobile app designed for older adults with heart failure to report Patient Reported Outcomes Measurement Information System (PROMIS) measures demonstrated that these adults successfully returned the PROMIS data and an additional survey indicated high levels of usability [17].

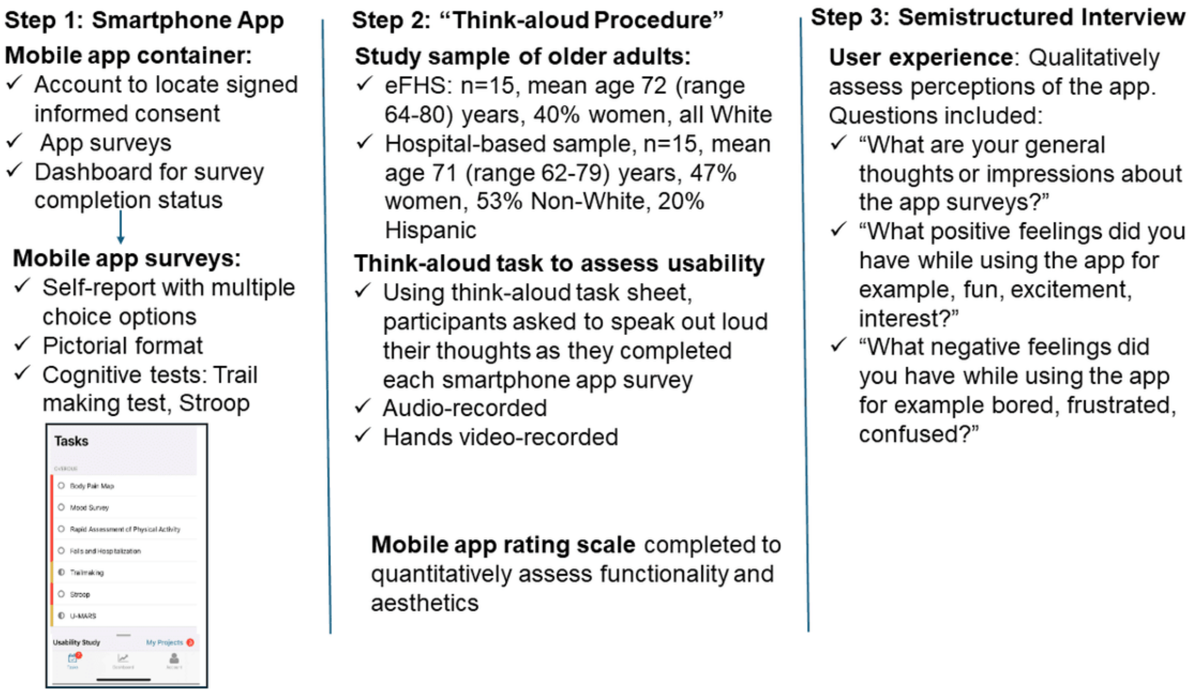
We designed a smartphone app for use by community-dwelling older adults who are participants in the Framingham Heart Study (FHS). The smartphone app consists of surveys with different response formats (eg, multiple choice, pictorial, and tasks). The aim of this study is to identify design-based barriers and facilitators to mobile app navigation and survey completion through usability testing. We also sought to understand whether participant feedback differed depending on the response format of each measure. Because FHS participants were White, we enrolled a diverse sample of older adults at a second site to understand if any additional barriers to mobile app survey completion were observed given that digital literacy and preferences for using technologies can vary across older adults of different races and ethnic backgrounds [12,18]. Importantly, little is known about the usability of mobile apps for racially or ethnically diverse populations [8].

Methods

Study Design

The study used mixed methods to conduct 1 usability testing session followed by a postsession interview with enrolled participants (Figure 1). Usability testing methods included using the "think-aloud" protocol while conducting a series of surveys and tasks on the smartphone app. This was followed by a semistructured interview using an interview guide, to solicit information on barriers encountered in the session.

Figure 1. Overall study design.



Study Sample

The study sample was drawn from 2 sites: the FHS Offspring study and a hospital-based site. The FHS Offspring participants were recruited in 1971 and are invited back to the research center for examination every 6 to 8 years [19]. The tenth examination of the FHS Offspring participants occurred from 2019 to 2022 and included a mobile health component called the electronic Framingham Heart Study (eFHS). For this study, we enrolled eFHS participants who were English-speaking, owned a smartphone (iPhone or Android), attended Offspring exam 10 before the eFHS began, and enrolled in eFHS between July and September 2022. The eFHS research technician assisted the participants with registration, informed consent, and app download. Because most of the eFHS sample had iPhones and were less racially and ethnically diverse, we recruited participants from a second site who were not part of eFHS. The second site was a hospital in an urban area with a racially and ethnically diverse patient population. At the second site, inclusion criteria were age 60 years and older, English speaking, and able to attend a study session between December 2022 and April 2023. Participants were recruited through flyers placed in clinic waiting rooms and community centers, participation in prior research studies, and through patient registry lists. Non-White and Hispanic/Latino patients were oversampled from patient registry lists. As with eFHS, the research technician assisted the participants with registration, informed consent, and app download on a study iPhone. Participants at both sites were using the app for the first time. We enrolled 15 participants at each study site to ensure the representation of men and women, iPhone and Android users, and older people below and above 75 years. In addition, the number of unique challenges identified with the study design proposed appears to asymptote 15 participants [20]. While eFHS participants were not compensated for their time in the study (because FHS participants have not been compensated for participation in the parent study), participants from the hospital-based cohort were provided a US $100 card for participating upon completion of the session.

Ethical Considerations

The institutional review board at Boston University Medical Campus approved the eFHS study (H-36586) and this study (H-42659). The institutional review board at the University of Massachusetts Chan Medical School approved the hospital-based study (approval number 00000567).

Measures

Study App

A mobile app hosted a compendium of different types of survey assessments and tasks that users could click on to complete. CareEvolution's MyDataHelps Designer platform was used to build the smartphone app surveys for iPhone (iOS 10.0 or higher) or Android (version 7.0 or higher) devices. The MyDataHelps mobile app container includes an account where participants can locate their signed consent form, tasks (app surveys), and a dashboard. The dashboard was created in the app to provide the participant with survey completion status and encouragement with a thank you message. Investigators and CareEvolution industry partners internally tested the app surveys and tasks with attention to consistency inspection, and user-centered design principles to ensure clear instructions, easy navigation, and simple words and sentences [21].

Because the goal of the study was to assess usability, we included a variety of app-based surveys and tasks with different response formats (Multimedia Appendix 1). First, we tested several self-report surveys with multiple choice options including (1) the short form of the PROMIS measure of mood (anxiety and depression, 8-items) [22] and cognitive function

(4-items) [23] that included 5 response options that ranged from "never" to "always" for the mood assessment and "not at all" to "very much" for the cognitive function assessment; (2) falls and hospitalizations with yes/no response options and a calendar wheel to ascertain the date of a fall occurrence; and (3) rapid assessment of physical activity [24] to report the frequency of physical activity and the level of intensity (light, moderate, and vigorous), assessing frequency, intensity, and time, with yes/no response choices. The hospital-based site did not complete the PROMIS measure of mood. Second, we tested a pictorial format measure to collect data on chronic pain, based on a modified version of the Michigan body pain map measure [25]. Participants were shown an outline of a human body and were asked to click on the different places of the body where they are currently experiencing chronic pain, defined as pain lasting 3 months or longer. The app first displays the image of the front of the body and next the back of the body. If the participant is not experiencing chronic pain, a box was provided so that the user can indicate "no chronic pain." The third and final smartphone tasks tested were 2 commonly used cognitive tests. The Trail Making Test [26,27] is a timed assessment that requires the user to consecutively tap dots in alternating order between numbers and letters by first tapping the number "1" followed by tapping the letter "A" and then "2" followed by "B" until the user reaches the number "7." Correct answers result in the appearance of lines between the dots, resulting in a "trail." Finally, we assessed the Stroop [28] on the smartphone that requires the user to complete a series of 4 increasingly more difficult tasks responding as quickly and accurately as possible to changes in color and instructions. Because the Stroop requires the user to be able to see colors (yellow, green, red, and blue), persons with color blindness are not eligible for this task. A practice session was provided for each set of the 4 tasks with the ability to repeat the practice session should the user desire to do so [29]. At the end of the testing session, participants completed the Mobile App Rating Scale (MARS) on the smartphone to assess app functionality and aesthetics including ease of use, navigation, visual appeal, performance, graphics, and layout [30]. Items were rated by the participant using a 5-point Likert scale from 1=inadequate to 5=excellent.

Demographic data were not collected within the study app. For eFHS participants, these data were collected as part of Offspring exam 10. Sociodemographics such as age, gender, and employment were assessed via the research study coordinators and entered into a secure web-based software platform, REDCap (Research Electronic Data Capture; developed initially at Vanderbilt University, now collaborative support from the REDCap consortium) at the hospital-based site.

Procedure

Study sessions were conducted in person at the FHS Research Center. eFHS participants were asked to complete the above assessments using their own smartphones, but hospital-based participants used a study smartphone (iPhone 7). In addition, they were also asked to navigate to different areas of the MyDataHelps app container (account, tasks, and dashboard). While doing these tasks, participants were asked to "think your thoughts" out loud, including feelings (positive like "fun" and negative like "frustrating"). Prior to beginning the usability testing, the research technician demonstrated the "think-aloud" procedure using the text app on a smartphone and then asked the participant to demonstrate the think-aloud procedure using the same app verbalizing every movement, feeling, and decision. Once the participant was ready to begin, the research technician encouraged the participant to use the think-aloud task sheet (Multimedia Appendix 1) that included a list of app-based tasks. Participants were asked to speak out loud their thoughts, feelings, and actions as they completed each task. The participant was audio-recorded, and the participant's hands were video-recorded throughout the think-aloud procedure. The participant also completed the MARS [30] on the smartphone after the think-aloud procedure. The technician was present to audio- and video-record the think-aloud procedure and to encourage participants to speak their thoughts out loud. The technician was explicitly trained not to assist the participant with the app unless the participant was irrevocably stuck.

After completion of the think-aloud procedure, the research technician conducted a 15-20–minute interview using open-ended questions and reflective listening to obtain participant feedback on their experiences using the app. Interview questions are available in the Multimedia Appendix 1 and include questions such as "What are your general thoughts or impressions about the app surveys?"; "What positive feelings did you have while using the app for example, fun, excitement, interest?"; and "What negative feelings did you have while using the app for example bored, frustrated, confused?" In addition, the interviewer asked the participants to what extent family, friends, and people of their own age would be able to use and enjoy the app and their thoughts on how to ensure the app would be acceptable to people of different cultures. The research technicians (AD, NA, and AH) are coauthors of this work.

Research Technician Training

In total, 4 interviewers were trained (by BB and JF), 3 were bachelor level and 1 was PhD level. Training consisted of learning the think-aloud procedure and also how to conduct the post think-aloud interview. Building rapport and communication skills (open-ended questions and reflective listening) were part of the training. Training included didactics and role plays and trainees were required to complete 3 sessions with "mock" participants, supervised by 1 or both trainers before being cleared to do the protocol with study participants. Feedback on study participants was provided to research technicians on an ongoing basis, by viewing the audio and video tapes together. The audiotaped portion of the think-aloud procedure and interview was professionally transcribed (Daily Transcription).

Process of Coding the Sessions

Investigators developed a coding sheet and accompanying coding manual for use by teams of coders when coding the video- and audio-recordings of the think-aloud procedure and postinterviews. The coding sheet included general items (eg, navigation between surveys, tapping in the wrong area to advance to the next task, and unclear instructions for surveys) as well as assessment-specific variables (clarity of instruction, understanding concepts, navigating within a survey, and "look and feel" eg, font size, line spacing, and color). For training purposes, all coders reviewed 3 participant recordings together and resolved any coding discrepancies before separating into the 2 coding teams to code in pairs. In order to maintain coding reliability over time, the 2 teams (team one: JF, JM; team 2: BB, DDM) independently coded the same participant recordings on 5 occasions throughout the coding process and came together to review the coding sheets for any discrepancies across teams to ensure all coders were following the coding guidelines. The average percent agreement between coders ranged from 80.5% to 98% across the 5 recordings that were coded in common by the 2 teams.

Statistical Analysis

Descriptive statistics of the 2 study samples used mean and SD for continuous variables, and numbers and percentages for nominal variables. For the comparison of continuous variables, 2-tailed *t* tests were applied, and chi-square tests or Fisher exact tests were used to compare nominal variables between the samples. The percent agreement between coding teams was calculated by the ratio of the number of discrepancies divided by the number of items in the coding sheet. For the MARS, all items used a 5-point Likert scale and the mean score for each domain (functionality and aesthetics) was calculated separately and an overall MARS score was computed for each of the 2 study samples. In addition, we calculated the mean score of each item within a domain (eg, layout, graphics, and visual appeal).

Results

Study Sample

In eFHS, 15 participants signed informed consent (mean age of 72, SD 4.2, range 64-80 years; 6/15, 40% were women; 1/15, 100% non-Hispanic, White). All eFHS participants owned a smartphone with 9 of 15 (60%) of the eFHS study sample participants owning an iPhone (Table 1). In the hospital-based site, 77 were contacted, 19 declined to participate, 3 deferred enrollment to a later date, and 15 participants signed informed consent (mean age 70.6, SD 6.2, range 62-79 years; 7/15, 47% women; 3/15, 20% Hispanic/Latino; and 8/15, 53% non-White). In the hospital-based sample, 1 participant did not own a smartphone, and among smartphone owners, 6 of 14 (42.9%) owned an iPhone. More than half of the participants at both sites had a college education or advanced degree. While nearly 90% (13 of 15 participants) of eFHS participants reported their health to be very good to excellent, only one-third (5 of 15 participants) of the participants at the hospital-based site did (*P*=.003).

Table 1. Participant characteristics.

Characteristics	eFHS[a] (n=15)	Hospital-based sample (n=15)	*P* value
Age (years)			.48
Mean (SD)	72 (4.2)	70.6 (6.2)	
Range	64-80	62-79	
Women, n (%)	6 (40)	7 (46.7)	.71
Non-White, n (%)	0 (0)	8 (53.3)	.002
Hispanic, n (%)	0 (0)	3 (20)	.22
Smartphone owner, n (%)[b]			
iPhone	9 (60)	6 (42.9)	.36
Android	6 (40)	8 (53.3)	.36
Bachelor degree and higher, n (%)	10 (66.7)	9 (60)	.71
Marital status (married, living as married), n (%)	12 (80)	8 (53.3)	.12
Income <US $55,000, n (%)[c]	4 (31)	5 (33)	.89
Subjective health, very good or excellent, n (%)	13 (86.7)	5 (33.3)	.003

[a]eFHS: electronic Framingham Heart Study.

[b]All offspring participants owned a smartphone (iPhone or Android); 1 hospital-based participant did not own a smartphone.

[c]Income of 2 participants in eFHS sample was unknown.

In the eFHS sample, the average length of time of the think-aloud procedure was 25.5 (range 12.4-44.0) minutes and the postprocedure interview time on average was 18.7 (range 12.4-20.9) minutes. Two participants declined the interview (think-aloud times were 28.5 and 33.35 minutes, respectively). At the hospital-based site, the average length of time of the think-aloud procedure was approximately 11 minutes longer (mean 36.5, range 24.1-55.1 minutes) while the postprocedure interview time was similar to eFHS with an average of 18.5 (range 10.6-35.1) minutes. No participant at the hospital-based site declined the post think-aloud interview.

Barriers to App Use Identified During Think-Aloud and Postprocedure Interview

Table 2 presents the themes identified from the think-aloud task and postprocedure interview along with sample participant quotes.

Table 2. Barriers and facilitators identified during the think-aloud and postprocedure interview.

Domain or survey type	Theme	Sample quotes
General structure, satisfaction	• Information about security of app • Use to raise awareness of health • Everyone could learn from it	• "The more I do, the easier it gets" • "I just think it's gonna help me monitor my health and go from there. I like technology that helps me and doesn't just amuse me or keep going and if this helps me stay healthy, stay fit, it makes sense."
Ease of use	• Easy, simple, fun • Questions easy to answer • Confusion related to not paying attention and not reading the app instructions carefully enough • Font color yellow difficult to see • Font could be bigger	• "The app spells everything perfectly clear after you read it a couple times to get it" • "I wasn't really paying attention to what I was reading. That was mostly my problem."
Navigation	• Navigating within and between surveys was easy • More detail on getting back to the home screen and using "next," "cancel," and "done"	• "The more I did it, the more I was able to figure out how to get from one place to another." • "You might want to give a little more detail about some of the sections, like what happens when you cancel, what happens when you hit done."
Multiple-choice survey formats	• Surveys were interesting and simple except physical activity survey with lengthy definitions and confusing response choices	• "It was easy the questions were simple and it was easy to find the answers."
Pictorial format	• Body pain map did not include all areas that can be a real problem • Checkbox worked differently than response choices for other surveys • Hard to find the "no pain" box—small font	• "Well, on the pain thing, I think they should have something near the anus. Because that can be a real headache. And then, of course, for women, it would also include, uh, the uterus."
Cognitive tasks: Trail Making Test and Stroop	• Stroop was confusing • Difficulty understanding practice session versus testing session • Confusion over 4 increasingly challenging sets of tests within the Stroop task • TMT[a] had difficulties with instructions and navigation but the eFHS sample did not	• "The all underling thing was a little confusing to me. I think because I was trying to get through it quick. I lost track of whether underline the word or is it the color or is it this color or the word?" • "The trail-making one I liked"
Friends or family	• Fear of technology; • Little interest in learning how to use technology • Learning curve for older people	• "I think they would find it very useful I think it's very useful, just for like, the cognitive part of it."
Different cultural backgrounds	• Need the app to be available in different languages	• "I come from the Indian community, and we place a lot of value on education, you know, I think this is the kind of thing that, you know—I would—we do—I would like to be quizzed, and this is a quiz." • "For example, I came from Burma. Burma used to be very underdeveloped country, and, also, still lots of problem. But they are very good at technology."

[a]TMT: Trail Making Test.

General Structure of the App

In general, participants were pleased with the general structure of the app. However, a few participants had not used an app previously and they said it was "not intuitive" but "once you run through it a few times it isn't a problem." Participants generally thought that the app had the right amount of surveys (and that additional surveys would make them bored). Participants suggested that a progress indicator be inserted to let users know the status of survey completion. One participant also raised the importance of conveying information regarding the security of the app.

Ease of Use

Overall, participants found the app "easy to use," "simple," and "fun." One participant said that it looked and functioned similarly to apps he was already familiar with, such as a social security app. A small minority of participants, however, wanted increased font size and more user-friendly colors. One participant was frustrated with the functionality of the app, but he blamed it on using an old phone. There were some suggestions for modifications, such as tailoring the app by age group and disability, so it is not so "one size fits all." Another participant suggested removing the free text space (which was needed for 1 question) or adding a digital keyboard so it is more intuitive that the person is supposed to type. For example, upon completion of the Stroop task, there was a query ("did you encounter any issues during the task?") that permitted the participant to freely type in a response.

Navigation

A large minority of participants reported at least 1 navigational issue, including navigating within and between surveys and also clicking the incorrect area in the app to navigate to the next screen (eg, navigating to "done"). The vast majority of these errors were committed only once. Participants suggested making the buttons look more like buttons or having them flash when you are supposed to press the button, "...making it extremely obvious what you are supposed to hit." Participants also suggested that more details should be added, such as what will happen after you hit a button (eg, after "cancel" or "done" is pressed), and more information about what the "icons at the bottom of the app" mean (eg, dashboard). A few participants mentioned that the app should avoid having to scroll to see more information, and instead put all of the information on one page if possible or continue to the next page.

Response Choice Formats

Multiple-Choice Formats

Participants reported that most of the multiple-choice surveys were easy and interesting to complete, with only a few sporadic usability errors by a few participants. For the most part, participants reported that the instructions were easy to understand, with one exception, which was "The Rapid Assessment of Physical Activity" survey, which included lengthy definitions of physical activity levels (eg, mild, moderate, and vigorous activity) which were needed to answer the subsequent questions. Participants were confused by the content and format of the definitions. With regard to the latter, both physical activity intensity level and physical activity frequency in a single question (eg, "I do some light or moderate physical activities but not every week," yes or no).

Pictorial Format

For the pictorial format assessment, (modified version of the Michigan Body Pain map), a small minority of participants noted the checkbox for response worked differently than the other surveys; instead of seeing a checkmark in the response box, the response box if selected was highlighted in color and was confusing for some. Some participants also noted that the map did not include all body areas where pain was present. For some participants, the "no chronic pain" response box was difficult to find due to the small font, and the small font used to label the body parts was also difficult for several participants (but necessary to eliminate scrolling to see the entire body outline). There were no participants who reported difficulty understanding the instructions or concepts. Only a couple of participants did not realize that they had to press "next" to go to the next page to view the back of the body to indicate pain areas there. One participant said that bowels and reproductive areas give people his age a lot of issues and that these areas should be added to the body pain map.

Cognitive Testing Formats

In terms of the formats used for cognitive testing, most older participants had some difficulty with the process of completing these measures. In terms of the Stroop, some participants had difficulty understanding that they were in the practice session versus the test session. Participants also thought that they needed to do the testing quickly and were confused if they did not read the instructions for each of the 4 test sets, as each had different instructions. Most participants did not take advantage of the opportunity to repeat the practice session when asked by the app, even when they were confused by the task. The yellow font color for the Stroop was also problematic for a few participants. One participant suggested that voice and animation be used to explain the instructions for the Stroop. For the Trail Making Test, the eFHS sample did not report any difficulties with instructions or navigating within the task, and seemed to have a good understanding of the concepts included in the task. However, a large minority of the hospital-based sample had difficulties in all 3 of these areas. There were no difficulties reported on the look and feel of the task (eg, font, line spacing, and color).

Relevance for Friends, Family, and Different Cultures

Participants noted that they knew older people who engaged with technology and others who were not interested in the "electronic age," did not use a computer or smartphone, and had no interest in learning. One participant said that she had a friend aged 90 years or older who would easily be able to interact with the app and another friend of the same age who would have more difficulty. Participants noted older adults may have a fear of technology, be less confident using technology, and need assistance or a training session given that, for older people who have not used computers or technology, using the app would "be like a foreign language to them." Participants provided their thoughts on using the app in older adults of different cultural backgrounds. Many noted that the current version of the app is available only in the English language and would need translation to other languages. One participant was from a country that they felt embraced technology, and another participant reported her culture valued education and felt like the app included an educational-like component like a "quiz" which would be viewed positively by her culture.

Satisfaction With the App

The mean total MARS score (7.6, SD 1.1), mean functionality score (3.8, SD 0.6), and mean aesthetics score (3.8, SD 0.6) in the eFHS sample did not significantly differ from the hospital-based sample (Figure 2). With the exception of ease of use, the individual items of the functionality and aesthetics scores (performance, navigation, interactions such as taps, swipes, layout, graphics, and visual appeal) did not significantly differ between the 2 samples (Figure 3). Ease of use may have differed as participants at the hospital-based site used a study iPhone whereas eFHS participants used their own smartphone. The performance item was rated the highest with a mean of 4.5 in both the samples. The mean overall star rating was 3.5 (SD

0.7) in the eFHS sample and 3.7 (SD 0.96) in the hospital-based sample indicating participants rated the app above average. In addition, during the interview, participants noted that the app could be used to raise awareness of health and people could learn from it, even though that was not the original purpose of the app. Some participants liked the app because it enabled one to “express yourself” through the surveys. Finally, a few participants suggested making the app “more entertaining” by adding narration.

Figure 2. MARS scores by study sample: overall MARS functionality and aesthetics scores. MARS: Mobile App Rating Scale.

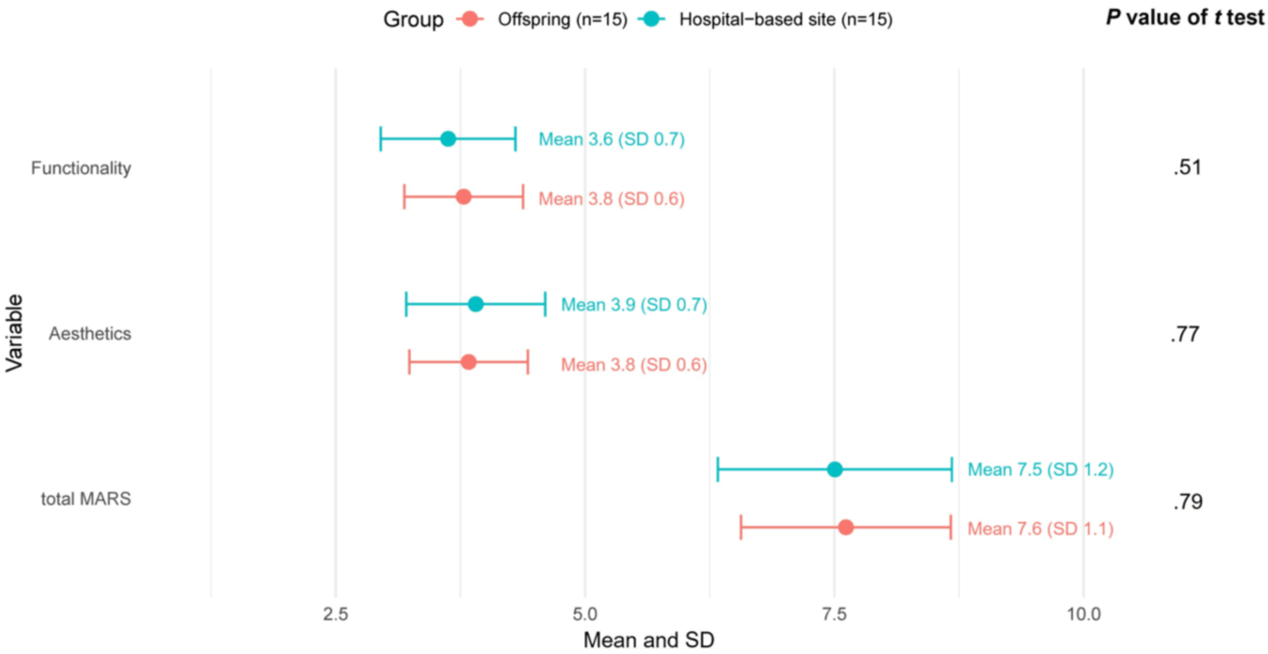
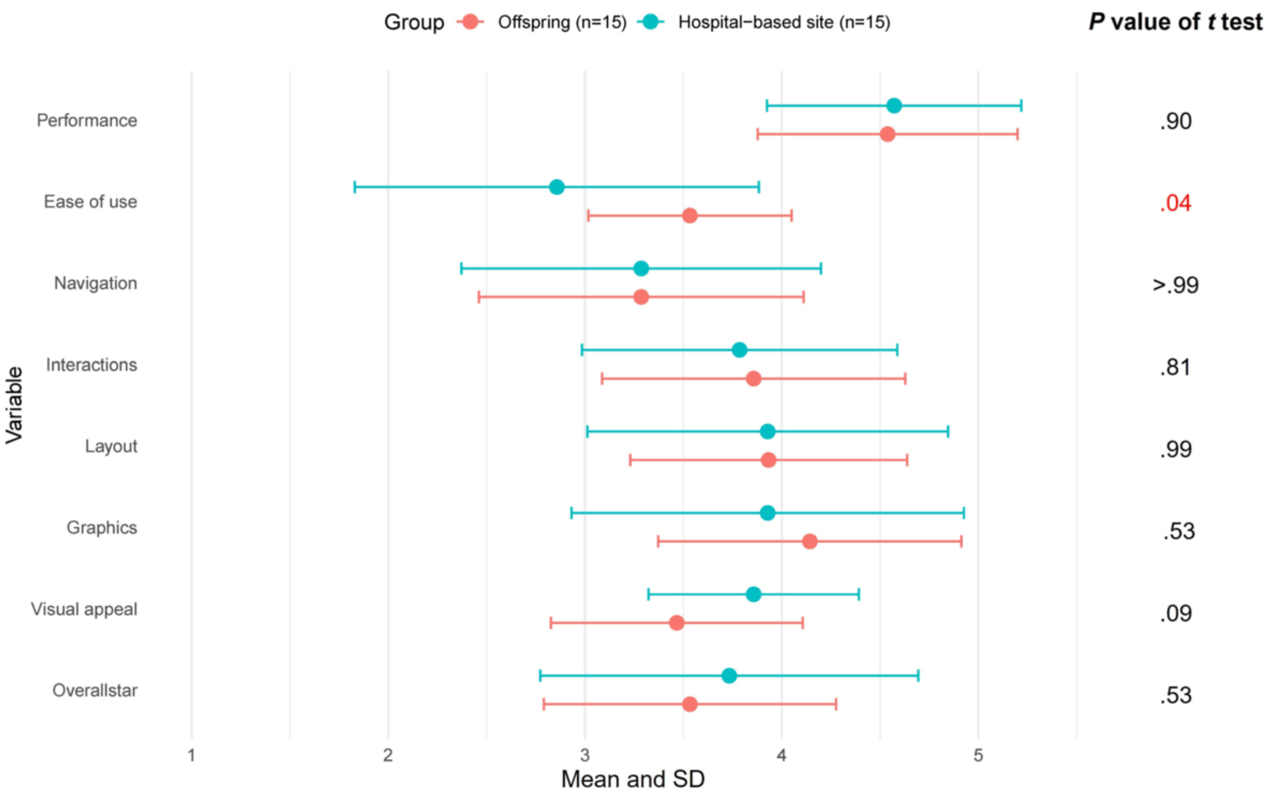


Figure 3. MARS (Mobile App Rating Scale) scores by study sample: individual items within the Functionality and Aesthetics domains.



Discussion

We tested the usability of a smartphone app designed to collect health information using complementary approaches in a community sample of older adult participants of the Framingham Offspring Study, and to understand generalizability we also tested older adults from a more diverse hospital-based sample. In general, participants liked the structure of the app and found the app was simple, fun, and easy to use. However, a large minority reported navigation issues that mostly occurred once with the ability to learn and figure out how to move within and between app-based surveys. A small minority of participants verbalized a preference for larger font sizes or more user-friendly colors. We observed that most older adults did not have difficulty with the multiple-choice app-based surveys unless the survey included lengthy instructions. Finally, most older adults experienced challenges with the app-based cognitive tasks especially the Stroop which required participants to read and understand a series of 4 increasingly more challenging tests with the Stroop task. Of importance, some participants noted that the app could be used to raise awareness of health and one could learn from it. Our observations confirm those of others that the involvement of older users can result in positive feelings among older adults, dispel stereotypes associated with older users, and the insights gained from older users can be used to enhance the quality of the design [31]. For example, our participants suggested using voice instructions and animated tutorials. To enhance usability, app designers and investigators should consider training that includes tutorials within the app provided by an older adult guide to boost confidence when designing smartphone apps for use by older adults.

This work has several implications for tailoring technology for older adult users. First, a guide within the app explaining the purpose of the app and highlighting key app functions including such functions as “next,” “back,” and “done” would enhance usability. Older adults are more likely to engage with technology that they perceive as useful [5]; therefore, having a clear understanding of the goals of technology is critical. Further, the addition of basic training in smartphone use in older adults less proficient with technology was associated with fewer errors and less cueing during a smartphone app–based health prevention program and may result in improved engagement [32]. In our sample, older adults who had not used an app before did not find it intuitive but after running through it a few times did not find it a problem. Our results support recently published app design guidelines for older adults advocating for initial training, if possible face-to-face, along with video instructions that are contextualized and provide step-by-step instructions to support older users [33]. Training may boost confidence and make the experience as frictionless as possible lowering the potential for abandonment. Second, streamlining and simplifying instructions may enhance understanding by inviting older adults to read them attentively. Participants in our study noted confusion that they attributed to not paying attention or related to the need to slow down and read directions more than once. Consistent with our observation, others have noted the need for clear and simple instructions when designing mHealth apps for older adults [6]. Some participants also requested features they enjoyed in their use of other apps. Gamification of functions, where possible, such as a flashing “done” button or a countdown to the start of the next task may improve engagement. Finally, consider voice narration and animated guides throughout the app surveys where features other than straightforward multiple-choice questions and responses will be encountered.

Our study may have important insights to help address the continued digital disparities observed in older adults. Older adults are increasingly using the internet and smartphones [34]; however, connection to the internet decreases across ages with nearly half of young adults almost constantly connected versus 8% of adults aged 65 years and older [4]. Other key digital health behaviors are also lower in a nationally representative sample of older adults including using health apps, using a digital device to track health or a health-related goal, and digitally communicating with a health care provider [35]. Digital technologies were lifesaving during the COVID-19 pandemic as the rapid transformation from in-person visits to televisits permitted access to health care in a setting that provided social distance and did not expose vulnerable older adults to the virus. Similarly, the ability to participate in digital interventions may provide several benefits, such as improved memory and independent living [36], physical functioning, physical activity [37], depression, and anxiety [38]. Including older adults in technology design and conducting usability testing may address digital health inequities by addressing digital health literacy and creating programs that are user-friendly to this population [39]. They may also improve implementation beyond pilot studies and achieve the needed sustainability of technology solutions [40] for chronic disease management and home care options for older adults and, at the same time, maybe one step in addressing digital disparities. We plan to use the smartphone app more widely in the Framingham cohorts as a tool to monitor health. We will be able to provide critical information on the characteristics of those who enroll and use the technology, as well as those who choose not to.

Our study had several strengths. There is no “best” method to assess usability [21]; therefore, we used both qualitative and quantitative methods. We tested older users with mean age of 70 years and older in both samples often not included in studies testing technology and included older adults from diverse race and ethnic backgrounds. The study sample included a range of older users. Participants without a smartphone or experience with app use and both iPhone and Android users were included. This strategy allowed us to uncover errors with the app beyond what would be observed with “regular” users and permit greater guidance in app redesign to benefit older users.

In addition, participants with health issues were included. Some older adults with health conditions or geriatric syndromes such as frailty have higher levels of nonuse of information communication technologies and more negative views on usefulness and usability [41]. Our study also had some limitations that merit comment. Participants at the hospital-based site used an iPhone only. This may have been a limitation if the participants were Android owners or had a different iPhone version; however, this is also a strength as we were able to include participants who were not smartphone owners. Our observations focus on the first interactions with the app-based surveys. It is beyond the scope of the study to examine other aspects of use such as efficiency (how quickly the survey or task is completed once the design is learned) and memorability (how easy to use after a period of not using the app) that may have important implications to research study designs. Continued engagement with technology changes over time in older adults but the factors related to continued technology use are unclear and require further investigation [42]. Our study took place in Massachusetts, and therefore, may not be generalizable to other geographic areas. Participants with color blindness were not eligible for the Stroop task and were excluded from testing. Therefore, results may not be generalizable to this group of older adults. FHS participants who enroll in eFHS are healthier and have higher levels of education than participants who chose not to enroll.

Our study of a diverse sample of older adults testing several different smartphone app survey types and response formats provides a guide to investigators and clinicians that can be used for future app development and app-based survey construction for older adults. Many older users are able to interact with and enjoy technology. Further work to enhance engagement among older users and diminish digital disparities in this group is needed if the potential of technology to improve well-being, functioning, and health in older adults is to be realized.

Acknowledgments

This work was supported by the National Institutes of Health, National Institute of Aging (R01AG076757) to JMM. The Framingham Heart Study is supported by the National Heart, Lung, and Blood Institute at the National Institutes of Health (contracts N01-HC-25195, HHSN268201500001I, and 75N92019D00031).

Conflicts of Interest

DDM has received research support from Fitbit, Apple Inc, Bristol–Myers Squibb, Boehringer–Ingelheim, Pfizer, Flexcon, Samsung, Philips Healthcare, and Biotronik, and has received consultancy fees from Heart Rhythm Society, Bristol–Myers Squibb, Pfizer, Fitbit, Flexcon, Boston Biomedical Associates, VentureWell, Avania, and Rose Consulting. DDM also declares financial support for serving on the steering committee for the GUARD-AF study (NCT04126486) and the advisory committee for the Fitbit Heart study (NCT04176926). ES is an employee of CareEvolution, Inc, a health care technology company. The remaining authors have no conflicts of interest to declare.

Multimedia Appendix 1
Screenshots of app-based surveys and tasks, think- aloud task sheet, and post procedure interview questions.
[DOCX File , 3864 KB - humanfactors_v11i1e56653_app1.docx]

References

1. Faverio M. Share of those 65 and older who are tech users has grown in the past decade. Pew Research Center. 2022. URL: https://tinyurl.com/2fza68h7 [accessed 2022-04-11]
2. Umoh ME, Prichett L, Boyd CM, Cudjoe TKM. Impact of technology on social isolation: longitudinal analysis from the National Health Aging Trends Study. J Am Geriatr Soc 2023;71(4):1117-1123. [doi: 10.1111/jgs.18179] [Medline: 36519748]
3. Lee CC, Czaja SJ, Moxley JH, Sharit J, Boot WR, Charness N, et al. Attitudes toward computers across adulthood from 1994 to 2013. Gerontologist 2019;59(1):22-33 [FREE Full text] [doi: 10.1093/geront/gny081] [Medline: 29982458]
4. Anderson M, Perrin A. 2. Barriers to adoption and attitudes towards technology. Pew Research Center. 2017. URL: https://www.pewinternet.org/2017/05/17/barriers-to-adoption-and-attitudes-towards-technology/ [accessed 2019-09-14]
5. Berkowsky RW, Sharit J, Czaja SJ. Factors predicting decisions about technology adoption among older adults. Innov Aging 2018;2(1):igy002 [FREE Full text] [doi: 10.1093/geroni/igy002] [Medline: 30480129]
6. Hsieh KL, Fanning JT, Rogers WA, Wood TA, Sosnoff JJ. A fall risk mHealth app for older adults: development and usability study. JMIR Aging 2018;1(2):e11569 [FREE Full text] [doi: 10.2196/11569] [Medline: 31518234]
7. Schumacher RM. Gran got tech: inclusivity and older adults. J Usability Stud 2023;18(2):62-67 [FREE Full text] [doi: 10.1016/b978-0-12-408138-3.00012-1]
8. Sarkar U, Gourley GI, Lyles CR, Tieu L, Clarity C, Newmark L, et al. Usability of commercially available mobile applications for diverse patients. J Gen Intern Med 2016;31(12):1417-1426 [FREE Full text] [doi: 10.1007/s11606-016-3771-6] [Medline: 27418347]
9. Charalambous AP, Pye A, Yeung WK, Leroi I, Neil M, Thodi C, et al. Tools for app- and web-based self-testing of cognitive impairment: systematic search and evaluation. J Med Internet Res 2020;22(1):e14551 [FREE Full text] [doi: 10.2196/14551] [Medline: 31951218]
10. Wald JS, Businger A, Gandhi TK, Grant RW, Poon EG, Schnipper JL, et al. Implementing practice-linked pre-visit electronic journals in primary care: patient and physician use and satisfaction. J Am Med Inform Assoc 2010;17(5):502-506 [FREE Full text] [doi: 10.1136/jamia.2009.001362] [Medline: 20819852]

11. Cusatis R, Holt JM, Williams J, Nukuna S, Asan O, Flynn KE, et al. The impact of patient-generated contextual data on communication in clinical practice: a qualitative assessment of patient and clinician perspectives. Patient Educ Couns 2020;103(4):734-740. [doi: 10.1016/j.pec.2019.10.020] [Medline: 31744702]
12. Gordon NP, Hornbrook MC. Differences in access to and preferences for using patient portals and other eHealth technologies based on race, ethnicity, and age: a database and survey study of seniors in a large health plan. J Med Internet Res 2016;18(3):e50 [FREE Full text] [doi: 10.2196/jmir.5105] [Medline: 26944212]
13. Crouch E, Gordon NP. Prevalence and factors influencing use of internet and electronic health resources by middle-aged and older adults in a US Health Plan population: cross-sectional survey study. JMIR Aging 2019;2(1):e11451 [FREE Full text] [doi: 10.2196/11451] [Medline: 31518256]
14. Irizarry T, Shoemake J, Nilsen ML, Czaja S, Beach S, DeVito Dabbs A. Patient portals as a tool for health care engagement: a mixed-method study of older adults with varying levels of health literacy and prior patient portal use. J Med Internet Res 2017;19(3):e99 [FREE Full text] [doi: 10.2196/jmir.7099] [Medline: 28360022]
15. Harte R, Quinlan LR, Glynn L, Rodríguez-Molinero A, Baker PM, Scharf T, et al. Human-centered design study: enhancing the usability of a mobile phone app in an integrated falls risk detection system for use by older adult users. JMIR Mhealth Uhealth 2017;5(5):e71 [FREE Full text] [doi: 10.2196/mhealth.7046] [Medline: 28559227]
16. Mansson L, Wiklund M, Öhberg F, Danielsson K, Sandlund M. Co-creation with older adults to improve user-experience of a smartphone self-test application to assess balance function. Int J Environ Res Public Health 2020;17(11):3768 [FREE Full text] [doi: 10.3390/ijerph17113768] [Medline: 32466484]
17. Reading Turchioe M, Grossman LV, Baik D, Lee CS, Maurer MS, Goyal P, et al. Older adults can successfully monitor symptoms using an inclusively designed mobile application. J Am Geriatr Soc 2020;68(6):1313-1318 [FREE Full text] [doi: 10.1111/jgs.16403] [Medline: 32157679]
18. Greer DB, Abel WM. Exploring feasibility of mHealth to manage hypertension in rural Black older adults: a convergent parallel mixed method study. Patient Prefer Adherence 2022;16:2135-2148 [FREE Full text] [doi: 10.2147/PPA.S361032] [Medline: 35999840]
19. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham offspring study. design and preliminary data. Prev Med 1975;4(4):518-525. [doi: 10.1016/0091-7435(75)90037-7] [Medline: 1208363]
20. Blair J, Conrad F, Ackermann AC, Claxton G. The effect of sample size on interview findings. AAPOR-American Statistical Association, Section on Survey Research Methods. 2005. URL: http://www.asasrms.org/Proceedings/y2006/Files/JSM2006-000127.pdf [accessed 2024-05-14]
21. Jaspers MWM. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. Int J Med Inform 2009;78(5):340-353. [doi: 10.1016/j.ijmedinf.2008.10.002] [Medline: 19046928]
22. Cella D, Choi SW, Condon DM, Schalet B, Hays RD, Rothrock NE, et al. PROMIS adult health profiles: efficient short-form measures of seven health domains. Value Health 2019;22(5):537-544 [FREE Full text] [doi: 10.1016/j.jval.2019.02.004] [Medline: 31104731]
23. Lai JS, Wagner LI, Jacobsen PB, Cella D. Self-reported cognitive concerns and abilities: two sides of one coin? Psychooncology 2014;23(10):1133-1141 [FREE Full text] [doi: 10.1002/pon.3522] [Medline: 24700645]
24. Department of Health Systems and Population Health. Rapid assessment of physical activity (RAPA). University of Washington. 2006. URL: https://depts.washington.edu/hprc/programs-tools/tools-guides/rapa/ [accessed 2024-05-14]
25. Brummett CM, Bakshi RR, Goesling J, Leung D, Moser SE, Zollars JW, et al. Preliminary validation of the Michigan Body Map. Pain 2016;157(6):1205-1212 [FREE Full text] [doi: 10.1097/j.pain.0000000000000506] [Medline: 26835782]
26. Dahmen J, Cook D, Fellows R, Schmitter-Edgecombe M. An analysis of a digital variant of the trail making test using machine learning techniques. Technol Health Care 2017;25(2):251-264 [FREE Full text] [doi: 10.3233/THC-161274] [Medline: 27886019]
27. Ashendorf L, Jefferson AL, O'Connor MK, Chaisson C, Green RC, Stern RA. Trail making test errors in normal aging, mild cognitive impairment, and dementia. Arch Clin Neuropsychol 2008;23(2):129-137 [FREE Full text] [doi: 10.1016/j.acn.2007.11.005] [Medline: 18178372]
28. Hankee LD, Preis SR, Piers RJ, Beiser AS, Devine SA, Liu Y, et al. Population normative data for the CERAD word list and victoria stroop test in younger- and middle-aged adults: cross-sectional analyses from the Framingham Heart Study. Exp Aging Res 2016;42(4):315-328 [FREE Full text] [doi: 10.1080/0361073X.2016.1191838] [Medline: 27410241]
29. Jongstra S, Wijsman LW, Cachucho R, Hoevenaar-Blom MP, Mooijaart SP, Richard E. Cognitive testing in people at increased risk of dementia using a smartphone app: the iVitality proof-of-principle study. JMIR Mhealth Uhealth 2017;5(5):e68 [FREE Full text] [doi: 10.2196/mhealth.6939] [Medline: 28546139]
30. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
31. Fischer B, Peine A, Östlund B. The importance of user involvement: a systematic review of involving older users in technology design. Gerontologist 2020;60(7):e513-e523 [FREE Full text] [doi: 10.1093/geront/gnz163] [Medline: 31773145]

32. Harte R, Hall T, Glynn L, Rodríguez-Molinero A, Scharf T, Quinlan LR, et al. Enhancing home health mobile phone app usability through general smartphone training: usability and learnability case study. JMIR Hum Factors 2018;5(2):e18 [FREE Full text] [doi: 10.2196/humanfactors.7718] [Medline: 29699969]
33. Gomez-Hernandez M, Ferre X, Moral C, Villalba-Mora E. Design guidelines of mobile apps for older adults: systematic review and thematic analysis. JMIR Mhealth Uhealth 2023;11:e43186 [FREE Full text] [doi: 10.2196/43186] [Medline: 37733401]
34. Tech adoption climbs among older adults. Pew Research Center. 2017. URL: http://www.pewinternet.org/2017/05/17/tech-adoption-climbs-among-older-adults/ [accessed 2024-05-14]
35. Ratcliff CL, Krakow M, Greenberg-Worisek A, Hesse BW. Digital health engagement in the US population: insights from the 2018 health information national trends survey. Am J Public Health 2021;111(7):1348-1351. [doi: 10.2105/AJPH.2021.306282] [Medline: 34014759]
36. Scullin MK, Jones WE, Phenis R, Beevers S, Rosen S, Dinh K, et al. Using smartphone technology to improve prospective memory functioning: a randomized controlled trial. J Am Geriatr Soc 2022;70(2):459-469 [FREE Full text] [doi: 10.1111/jgs.17551] [Medline: 34786698]
37. Stockwell S, Schofield P, Fisher A, Firth J, Jackson SE, Stubbs B, et al. Digital behavior change interventions to promote physical activity and/or reduce sedentary behavior in older adults: a systematic review and meta-analysis. Exp Gerontol 2019;120:68-87. [doi: 10.1016/j.exger.2019.02.020] [Medline: 30836130]
38. Riadi I, Kervin L, Dhillon S, Teo K, Churchill R, Card KG, et al. Digital interventions for depression and anxiety in older adults: a systematic review of randomised controlled trials. Lancet Healthy Longev 2022;3(8):e558-e571 [FREE Full text] [doi: 10.1016/S2666-7568(22)00121-0] [Medline: 36102765]
39. Seifert A, Reinwand DA, Schlomann A. Designing and using digital mental health interventions for older adults: being aware of digital inequality. Front Psychiatry 2019;10:568 [FREE Full text] [doi: 10.3389/fpsyt.2019.00568] [Medline: 31447716]
40. Matthew-Maich N, Harris L, Ploeg J, Markle-Reid M, Valaitis R, Ibrahim S, et al. Designing, implementing, and evaluating mobile health technologies for managing chronic conditions in older adults: a scoping review. JMIR Mhealth Uhealth 2016;4(2):e29 [FREE Full text] [doi: 10.2196/mhealth.5127] [Medline: 27282195]
41. Keränen NS, Kangas M, Immonen M, Similä H, Enwald H, Korpelainen R, et al. Use of information and communication technologies among older people with and without frailty: a population-based survey. J Med Internet Res 2017;19(2):e29 [FREE Full text] [doi: 10.2196/jmir.5507] [Medline: 28196791]
42. Zhang S, Boot WR. Predicting older adults' continued computer use after initial adoption. Innov Aging 2023;7(4):igad029 [FREE Full text] [doi: 10.1093/geroni/igad029] [Medline: 37197443]

Abbreviations

eFHS: electronic Framingham Heart Study
FHS: Framingham Heart Study
MARS: Mobile App Rating Scale
PROMIS: Patient Reported Outcomes Measurement Information System
REDCap: Research Electronic Data Capture

Edited by A Choudhury; submitted 22.01.24; peer-reviewed by D Chong, C Moral; comments to author 29.03.24; revised version received 28.04.24; accepted 01.05.24; published 30.05.24.

Please cite as:
Murabito JM, Faro JM, Zhang Y, DeMalia A, Hamel A, Agyapong N, Liu H, Schramm E, McManus DD, Borrelli B
Smartphone App Designed to Collect Health Information in Older Adults: Usability Study
JMIR Hum Factors 2024;11:e56653
URL: https://humanfactors.jmir.org/2024/1/e56653
doi:10.2196/56653
PMID:38815261

Original Paper

The Influence of Incentive-Based Mobile Fitness Apps on Users' Continuance Intention With Gender Moderation Effects: Quantitative and Qualitative Study

Aaya Faizah[1], SKom; Alifah Fatimah Azzahra Hardian[1], SKom; Rania Devina Nandini[1], SKom; Putu Wuri Handayani[1], Prof Dr; Nabila Cyldea Harahap[1], Dr

[1]Faculty of Computer Science, University of Indonesia, Depok, Indonesia

Corresponding Author:
Putu Wuri Handayani, Prof Dr
Faculty of Computer Science, University of Indonesia
Kampus UI
Depok, 16424
Indonesia
Phone: 62 217863419
Email: Putu.wuri@cs.ui.ac.id

Abstract

Background: A survey conducted by McKinsey & Company reported that, as of May 2022, as many as 26% of Indonesians had recently started to engage actively in physical activity, 32% undertook regular physical activity, and 9% exercised intensely. The Fourth Industrial Revolution has spurred the rapid development of mobile fitness apps (MFAs) used to track people's sports activities. However, public interest in using these apps for any length of time is still relatively low.

Objective: In this study, we aimed to determine the effect of incentives (eg, self-monitoring, social support, platform rewards, and external influence) on the use of MFAs and the moderating effect of gender on users' continuance usage intention.

Methods: The study used a mixed methods approach. Quantitative data were collected through a web-based questionnaire and qualitative data from interviews with 30 respondents. The quantitative data, collected from 379 valid responses, were processed using covariance-based structural equation modeling. The qualitative data were processed using thematic analysis. The MFAs included in this research were those used as sports or physical activity trackers, such as Apple Fitness, Strava, Nike Run Club, and Fita.

Results: The results of the data analysis show that 3 groups of incentives, namely, self-monitoring, platform rewards, and external influence (with the exception of social support), affect the perceived usefulness of these apps. Gender was also shown to moderate user behavior in relation to physical activity. The study showed that women were more likely to be motivated to exercise by social and external factors, while men paid greater attention to the tracking features of the app and to challenges and rewards.

Conclusions: This research contributes to the field of health promotion by providing guidance for MFA developers.

(JMIR Hum Factors 2024;11:e50957) doi:10.2196/50957

KEYWORDS
incentive; fitness; mobile fitness apps; gender; continuance usage intention; Indonesia; mobile phone

Introduction

Background

According to the World Health Organization [1], regular physical activity (PA) is a key factor in the prevention and management of noncommunicable diseases. The Global Status Report on Physical Activity [2] reported that 1.4 billion individuals aged >18 years do not meet the levels of PA recommended to promote and protect health. In 2016, it was reported that, globally, 23% of all men and 32% of all women aged ≥18 years were not sufficiently physically active to stay healthy [1]. This means that approximately 1 in 3 women and 1 in 4 men are not sufficiently active and do not meet the global recommendation of at least 150 minutes of moderate-intensity PA or 75 minutes of high-intensity PA per week [3]. In August 2022, McKinsey & Company released the results of a survey conducted with 1041 Indonesian respondents in which 26% of the respondents stated that they had started to engage actively in personal training, 32% reported that they had been playing sports regularly, and 9% indicated that they had increased the intensity of their sports or fitness activities [4]. These data indicate that the level of Indonesians' interest in, and awareness of, sports, fitness, and personal training is significantly higher than the global average. This conclusion is supported by the increased use of mobile fitness apps (MFAs) in Indonesia, with 29 million users in 2022 [5].

MFAs use several types of incentives, which include self-incentives, peer incentives, and platform incentives [6-9]. In the MFA context, self-incentives involve a self-monitoring (SM) system in which users monitor and track their own behavior [10,11]. Peer incentives are focused on social support (SS), which includes informational, emotional, and material support or the protection provided by fellow users of the app [12]. In the context of MFAs, platform incentives usually take the form of rewards or awards resulting from gamification features [13]. Users who collect a large number of rewards are usually considered to have a higher status on the MFA and feel more satisfied with their use of the app [14,15].

According to Zhu et al [16], very few studies have examined the role played by gender differences in the use of health and fitness apps. Yin et al [17] stated that achievements in sports motivate men more, while social relationships motivate women more. Previous research on MFAs has explored their design [18] and evaluation [19-22], as well as user adoption intentions [23]. In addition, several studies have discussed continuity in the use of MFAs [24-27]. Chiu et al [25] integrated the expectation-confirmation theory (ECT) with the investment model to analyze the continuous use of MFAs. However, research investigating the various types of MFA incentives has been shown to have several limitations [17,26] because the effects of each incentive have mostly been explored separately [28-30]. Per McKinsey & Company's 2022 survey among Indonesian citizens [4], 87% of the respondents intended to continue using their personal training and fitness apps. The market analysis and demographics of this study apply only to MFA users in Indonesia.

Research Question

This research adopted the self-determination theory (SDT) and the ECT. The SDT, as postulated by Ryan and Deci [31], states that there are 3 main psychological needs that drive human behavior: autonomy, relatedness, and competency. If these psychological needs are met, intrinsic motivation will increase and make it easier to maintain certain behaviors [31]. Teixeira et al [32] show that the SDT can be applied to behavioral interventions that relate to exercise or PA. While the SDT has the ability to predict the intensity of a behavior based on the influence of incentive factors [17], the ECT is generally used to predict the continuity of a behavior [25]. The combination of the SDT and the ECT was chosen to analyze the relationship between the incentive factors that affect the use of MFAs and continuity in using them. Thus, the research question is "How do the incentives promoted by MFAs influence users' continuance usage intention (CUI)?" This research can provide guidance for MFA developers by helping them to evaluate their apps.

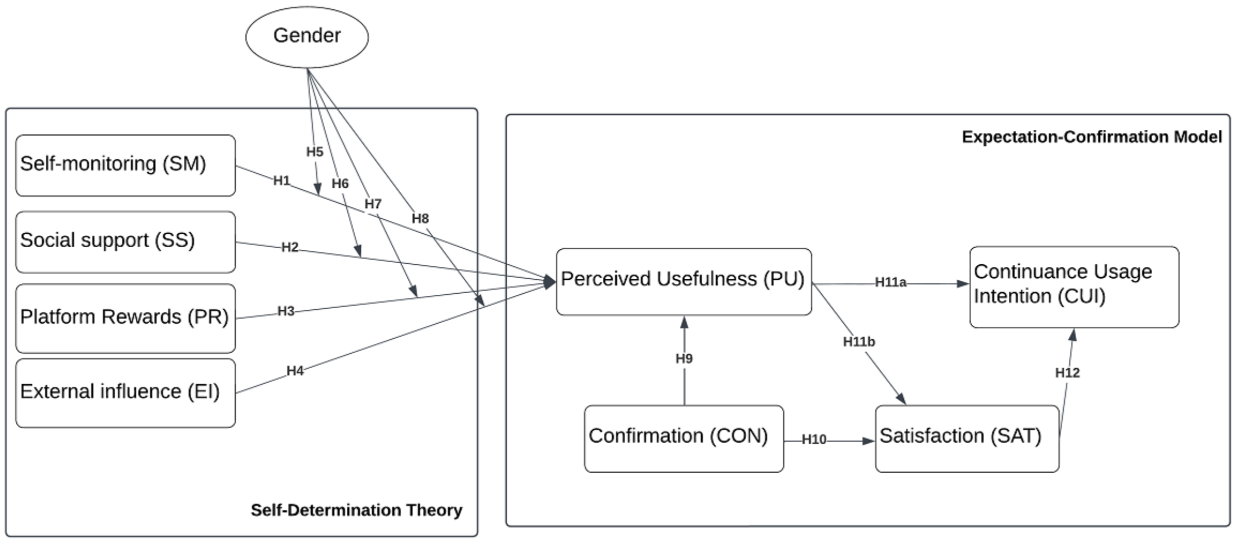
Methods

Research Model

Overview

The model used for this research is based on 2 theories and 1 moderating effect, namely, the SDT and the ECT, with the moderating effect of gender. Significant studies reporting on the use of these 2 theories include those by Yin et al [17], Huang and Ren [26], Chiu et al [25], and Li et al [33]. Yin et al [17] found that incentives are compatible with the SDT in motivating users' PA behaviors. The SDT approach described by Yin et al [17] is the theoretical basis for this research because it analyzes incentives offered by MFAs collectively and uses gender as a moderating variable. The relationship between perceived usefulness (PU) and incentives was also analyzed by Huang and Ren [26]. This research suggests that technology functions in MFAs, such as SM, self-regulation, and goal attainment, have an indirect effect on CUI through PU; for instance, Chiu et al [25] and Li et al [33] found that users' CUI was significantly predicted by ECT. Our research model, which includes 9 variables and 13 hypotheses (described in the following subsections), is presented in Figure 1.

Figure 1. Proposed conceptual model.



The Influence of SM on PU

SM, which is classified as one of the self-incentives in MFAs, includes managing and tracking one's own behavior [17]. These actions enable users to observe their own progress and evaluate their performance against previously set goals [34]. PU refers to the extent to which a person feels that technology can improve their performance of certain tasks [35]. In this study, the task was identified as increasing the user's PA, while, for MFA users, PU implies that using the MFA will enhance their personal training intensity [36,37]. Bhattacherjee [38] argues that when users confirm their initial expectations of the main functionality of a mobile app, they will begin to perceive the app as useful for improving their task performance and thus continue to use it. Huang and Ren [26] measured PU relating to the effectiveness and performance of PA through the use of 4 technological functions of the MFA, one of which is SM. Therefore, we examine the following hypothesis:

- H1: SM has an influence on PU.

The Influence of SS on PU

SS is classified as one of the peer incentives in MFAs [17]. Web-based SS is seen as an important factor affecting the physical and mental health of individuals, such as sports activity and increased well-being [12,39]. Humans have a tendency to behave in ways that are consistent with people in their own social networks, and this can be exploited in the context of mobile health (mHealth) [29]. Chen and Pu [40] conducted research on social incentives by developing the HealthyTogether mobile game, which allows users to participate in PA together and send messages to one another. The authors showed that users significantly increased their PA when using HealthyTogether compared to when they were exercising alone [40]. Edney et al [6] built the Active Team app, which is an MFA with social and gamification functions. The primary outcome of their study was a change in the total daily minutes of moderate to vigorous PA at 3 months, as measured objectively using an accelerometer [6]. Therefore, we propose the following hypothesis:

- H2: SS has an influence on PU.

The Influence of Platform Rewards on PU

The platform rewards include gamification elements, such as badges, points, and leaderboards [17]. The gamification element in MFAs can provide two types of information: (1) the user's PA progress and (2) a comparison of the user's PA with that of other users [41]. From this information, MFA users can observe their progress and experience greater satisfaction as they recognize their own personal training achievements. This leads to higher user competency satisfaction and increased behavioral motivation [13,42]. Yin et al [17] found that platform rewards have a positive relationship with users' PA. This finding is supported by Plangger et al [13] and Huang and Ren [26], who analyzed the effect of the goal-attainment technology function of MFAs, in which users can set their own goals, which are then achieved by undertaking PA. These achievements are then categorized as platform rewards. Huang and Ren [26] also found that this technology function had a positive effect on PU. Therefore, we propose the following hypothesis:

- H3: Platform rewards have a positive influence on PU.

The Influence of External Influence on PU

External influence (EI) is one of the extrinsic motivations identified in the SDT, which means that behavior is motivated through influences that do not depend on internal factors [43]. Huang [28] proposed this variable to explain how PA can be promoted through external factors. One example is companies providing incentives to MFA users as part of their corporate social responsibility initiatives [28]. Several studies discuss EI and PA. One example of EI referred to in this study is the name or *image* of a sponsor of an activity [44]. Low and Pyun [45] explain that sponsorship that gives a good impression to customers or users will produce behavior that tends to be positive. In the context of sporting activities, Huang [28] explains that sponsor characteristics play an important role in participation in a sporting activity. Therefore, because an MFA is a tool that can measure a person's PA, we intend to explore the following hypothesis:

- H4: EI has an influence on PU.

The Influence of Gender on SM and PU

According to Mao et al [7], MFA incentives are not always equally effective for women and men. This is because women and men have different ways of thinking [17]. Yin et al [17] conducted research that assumed that gender would influence the effectiveness of SM incentives, making them more effective for men than for women. This assumption was based on the belief that men generally pay more attention to their own achievements than women [46]. Surprisingly, Yin et al [17] show that gender does not affect the effectiveness of SM in MFAs. This finding relates to the concept of self-regulation, which is strongly driven by self-efficacy [47]. Individuals who decide to use MFAs are generally believed to have high self-efficacy in carrying out PA [17]. Therefore, we plan to test the following hypothesis:

- H5: Gender influences the relationship between SM and PU in MFA users.

The Influence of Gender on SS and PU

With regard to SS, Yin et al [17] explain that SS is one of the factors that most helps to fulfill the relatedness needs described in the SDT. According to Wang et al [9], social ties and commitment are more important for women than for men in shaping their attitudes toward the sharing of information. In considering gender, Yin et al [17] found that women tend to be more influenced by their relatedness needs than men. Women are also believed to be driven more by collective goals, such as pleasure or interpersonal harmony [48,49]. In the context of health apps, Kimbrough et al [50] found that women are usually more affected by environmental conditions and social relationships than men. Thus, we propose the following hypothesis:

- H6: Gender influences the relationship between SS and PU among MFA users.

The Influence of Gender on Platform Rewards and PU

Men tend to focus more on themselves and tend to be more independent than women [46,51]. Men also tend to focus more on completing or achieving individual goals that demonstrate their performance and abilities [46,51]. Relatedly, Vilela and Nelson [52] showed that men tend to be more motivated by their own achievements than women when using information system products. This is due to the general behavioral characteristics of men, who are generally more aggressive, pragmatic, and self-oriented in their behavior compared to women [52]. When specifically applied to incentives and CUI, Yin et al [17] also found an influence between gender and the effectiveness of platform reward incentives. The authors assumed that this is caused by the behavioral characteristics of men, who generally make decisions more rationally and pay greater attention to their own behavior. Thus, we propose the following hypothesis:

- H7: Gender influences the relationship between platform rewards and PU among MFA users.

The Influence of Gender on EI and PU

Sun and Zhang [53] state that women have a higher awareness of the environment than men. Leong et al [54] and Li et al [33] also found that men tend to be less easily influenced by external advice or support. Similarly, Venkatesh et al [55] concluded that women tend to be more influenced by EI, while men are usually less affected by external facilitation in their use of technology. This was confirmed by Weman Josefsson et al [56], who showed that men participate in challenges organized by the community to compete, while women participate for social and autonomy reasons. Hence, we propose the following hypothesis:

- H8: Gender influences the relationship between EI and PU among MFA users.

The Influence of Confirmation of Expectations on PU

Confirmation of expectations refers to the perceived level of conformity between the information system product or service expectation and actual performance [38]. Bhattacherjee [38] explains that PU refers to the individual's perception of the anticipated benefits from the use of IT products or services. The ECT implies that the confirmation of a user's expectations has a positive effect on their perception of the PU of an IT product or service [25,57-59]. According to the cognitive dissonance theory [60], IT users may experience psychological conflict if their initial expectations are not confirmed by their actual use experience [61]. Conversely, if users' initial expectations are confirmed or met, they may display higher investment behavior and reduce their preference for alternative apps [25]. Hsu and Lin [62] state that confirmation of expectations is positively related to the perceived quality of the IT product or service used, with the result that users tend to ignore quality alternatives. Therefore, we propose the following hypothesis:

- H9: Confirmation of expectations has an influence on PU.

The Influence of Confirmation of Expectations on Satisfaction

Chiu et al [25] proposed that confirmation of user expectations affects satisfaction with the app as well as its PU. Satisfaction can be interpreted as an individual's evaluation of their initial experience with a product or service [38]. Chiu et al [25] explain that before downloading an app, users generally have expectations of it, based on detailed information received from the app provider and on ratings and reviews from other users. After using the app, the user gains experience and evaluates the performance of the app based on previously established expectations. In line with the expectation-confirmation model, Chiu et al [25] assume that users' perceptions of postuse benefits and the confirmation of previous expectations determine their satisfaction in using IT products and services. Therefore, we propose the following hypothesis:

- H10: Confirmation of expectations has an influence on satisfaction.

The Influence of PU on CUI

PU refers to the user's perception of the benefits expected from using an IT product or service [61,63]. According to Bhattacherjee [38], expectations based on the user's direct experience have an important role in forming their IT CUI. Chiu et al [25] state that many studies conducted in various contexts [59,64,65] empirically support a positive relationship between PU and CUI. Wu et al [66] show that when users find the mHealth app useful, they show a higher level of satisfaction and tend to use it continuously. Thus, we define the following hypothesis:

- H11a: PU has an influence on CUI.

The Influence of PU on Satisfaction

According to Chiu et al [25], PU also has a strong and positive impact on satisfaction. The authors state that the more benefits users receive from health and fitness apps, the greater their satisfaction [25]. When a user has used an app for an extended period of time, the user will evaluate its performance and form either a confirmation or a disconfirmation of judgment with regard to their expectations [62]. Disconfirmation of expectations affects user satisfaction and creates negative perceptions of the usefulness of MFAs. Conversely, users' positive perceptions of usefulness increase their satisfaction with an app. Therefore, we propose the following hypothesis:

- H11b: PU has an influence on satisfaction.

The Influence of Satisfaction on CUI

Satisfaction can be identified as a significant factor influencing consumer behavior [25]. Bhattacherjee [61] strengthens this definition by explaining that user satisfaction is an important determinant of postadoption behavior relating to IT products and services. In other words, users with higher levels of satisfaction will exhibit greater levels of use of IT products and services than those who are less satisfied [25]. Wu et al [66] confirm that satisfied users are more likely to continue using an app because dissatisfied users can easily switch to other technologies at no additional cost. The relationship between satisfaction and CUI has been identified as one of the strongest relationships in the expectation-confirmation model [63]. Therefore, we propose the following hypothesis:

- H12: Satisfaction has an influence on the CUI of MFA users.

Research Procedure

This study used a mixed methods approach that integrated a quantitative approach, based on a questionnaire, with a qualitative approach, using interviews. The only inclusion criterion for respondents in this study was that they used MFAs. We modified a questionnaire that has been established in previous studies [12,15,16,24,25,28,33,66-70]. Before distributing the questionnaire, a readability test was conducted to validate how easily the questionnaire could be understood by respondents. The readability test was carried out both face-to-face and internet-based, using Google Meet, with 8 people who met the research criterion (ie, they all used MFAs). This readability test was carried out between February 5 and 10, 2023. We then used the results of the readability test to refine the questionnaire.

Once the questionnaire had been refined, we conducted a pilot study from February 20 to 25, 2023, aiming to measure the validity and reliability of the questionnaire by distributing it to 31 selected research respondents. The results of the pilot study were used to check the value of Cronbach α, which, in this pilot study, was 0.832, well over the required value of >0.7.

Research Instruments

The instruments used in this study were a web-based questionnaire and semistructured interview questions. The questionnaire first asked questions regarding the demographics of the respondents, and it then presented statements regarding the research model being tested. Each of the 8 variables exclude the gender variable in the study was assessed by 3 or 4 measurement items, and each indicator was represented by a statement to which participants responded on a Likert scale ranging from 1=*strongly disagree* to 5=*strongly agree*. The questionnaire used in this study is available in Multimedia Appendix 1, and a list of the interview questions is available in Multimedia Appendix 1.

Ethical Considerations

This research was approved by Faculty of Computer Science (approval number S-7/UN2.F11.D1.5/PPM.00.00/2024).

Results

Participant Demographics

We distributed the research questionnaire on the web through various social media platforms such as WhatsApp, Line, Twitter, Instagram, and Telegram. These social media platforms are widely used by Indonesians. The questionnaire distribution was carried out between February 27 and March 20, 2023. Table 1 provides a demographic summary of the respondents. Of the respondents, 75.5% (286/379) were aged between 17 and 25 years, 72.3% (274/379) were women, 25.1% (95/379) were privately employed, and 51.5% (195/379) lived in Greater Jakarta.

Table 1. Respondents' demographics (n=379).

Variable	Respondents, n (%)
Age (years)	
17-25	286 (75.5)
26-35	74 (19.5)
36-45	14 (3.7)
>45	5 (1.3)
Gender	
Woman	274 (72.3)
Man	105 (27.8)
Occupation	
College student	201 (53)
Employee of state-owned enterprise	8 (2.1)
High school student	15 (4)
Privately employed	95 (25.1)
Unemployed	19 (5)
Entrepreneur	23 (6.1)
Housewife	2 (0.5)
Civil servant	3 (0.8)
Other	13 (3.4)
Domicile	
Greater Jakarta	195 (51.5)
Java island	137 (36.1)
Outside of Java island	47 (12.4)

After collecting both the quantitative and qualitative data, we processed the quantitative data using covariance-based structural equation modeling. Using covariance-based structural equation modeling, data processing is carried out in several stages: specification and identification of the research model, estimation of the research model, testing the feasibility of the research model, modification of the research model, and hypothesis testing.

To validate the quantitative data results, we also collected qualitative data by conducting semistructured interviews with 30 respondents. The interviews were conducted both offline and on the web and took 30 to 45 minutes each. The qualitative data analysis was carried out thematically on the basis of the defined hypotheses.

Measurement Model

The factor loading values of all variables and indicators met the Cronbach α standard of >0.7 [71]; thus, the model feasibility test could be carried out. This study yielded average variance extracted values >0.5 as well as Cronbach α and composite reliability values >0.7 [71] (Table 2).

Table 2. Average variance extracted, Cronbach α, and composite reliability values.

Variable	Average variance extracted	Cronbach α	Composite reliability
Self-monitoring	0.968	0.705	0.920
Platform rewards	0.773	0.927	0.872
External influence	0.865	0.816	0.834
Social support	0.638	0.854	0.835
Confirmation of expectations	0.975	0.763	0.885
Perceived usefulness	0.709	0.888	0.848
Satisfaction	0.669	0.889	0.889
Continuance use intention	0.640	0.812	0.842

Structural Model

Next, we tested the structural model with the goodness-of-fit criteria, which included the relative chi-square index, goodness-of-fit index, root-mean-square error of approximation, root mean square residual, normal fit index, comparative fit index, and the Tucker-Lewis Index [71]. The goodness-of-fit values are presented in Table 3, and the R^2 values are shown in Table 4.

Table 3. Goodness-of-fit values.

Goodness-of-fit criteria	Cutoff value	Value	Description
Relative chi-square index	<2	1.956	Good fit
Goodness-of-fit index	≥0.9	0.900	Good fit
Root-mean-square residual	≤0.05	0.048	Good fit
Normal fit index	≥0.9	0.913	Good fit
Comparative fit index	≥0.9	0.955	Good fit
Tucker-Lewis Index	≥0.9	0.948	Good fit
Root-mean-square error of approximation	≤0.08	0.050	Good fit

Table 4. R^2 values.

Variable	R^2	Effect size
Perceived usefulness	0.349	Weak
Satisfaction	0.511	Medium
Continuance use intention	0.714	Strong

Hypotheses Testing

This study used a 2-tailed significance test; thus, the condition for accepting the hypothesis was P<.05 [71]. Table 5 presents the results of hypotheses 1 to 4 and 9 to 12, only one of which (H2) was rejected.

Table 5. Hypotheses testing results.

Hypothesis	Estimate (95% CI)	*P* value	Result
H1: SM[a]→PU[b]	0.319 (0.244 to 0.394)	.001	Accepted
H2: SS[c]→PU	0.060 (−0.019 to 0.143)	.12	Rejected
H3: PR[d]→PU	0.136 (0.044 to 0.219)	.007	Accepted
H4: EI[e]→PU	−0.101 (−0.166 to −0.033)	.006	Accepted
H9: COE[f]→PU	0.323 (0.251 to 0.388)	.001	Accepted
H10: COE→satisfaction	0.541 (0.435 to 0.632)	.002	Accepted
H11a: PU→CUI[g]	0.280 (0.200 to 0.363)	.002	Accepted
H11b: PU→satisfaction	0.218 (0.069 to 0.355)	.003	Accepted
H12: Satisfaction→CUI	0.683 (0.560 to 0.813)	.001	Accepted

[a]SM: self-monitoring.
[b]PU: perceived usefulness.
[c]SS: social support.
[d]PR: platform rewards.
[e]EI: external influence.
[f]COE: confirmation of expectations.
[g]CUI: continuance usage intention.

According to Awang [72], the test for moderation is not significant when the difference in chi-square values between the constrained model and the unconstrained model is <3.84. Table 6 presents a summary of the results of the hypothesis testing using the moderating effect of gender. On the basis of the difference in the chi-square values between the constrained model and the unconstrained model, it can be concluded that all difference values were >3.84 and therefore meet the requirements for calculating the significance of the moderating effect, meaning that H5, H6, H7, and H8 were all accepted.

Table 6. Summary of moderating variable hypothesis testing, with gender as the moderating effect.

Path	Chi-square constrained model	Chi-square unconstrained model	Difference (*df*)	Result
SM[a]→PU[b]	1318.671	1043.151	275.520	H5 accepted
SS[c]→PU	1627.823	1043.151	584.672	H6 accepted
PR[d]→PU	1395.076	1043.151	351.925	H7 accepted
EI[e]→PU	1796.875	1043.151	753.724	H8 accepted

[a]SM: self-monitoring.

[b]PU: perceived usefulness.

[c]SS: social support.

[d]PR: platform rewards.

[e]EI: external influence.

Qualitative Interviews and Validity of the Hypotheses

This research shows that the incentives offered in MFAs in the form of SM (eg, distance walked or run, number of calories expended, time taken, and heart rate) influence users' motivation to undertake PA. The acceptance of H1 is thus in accordance with the findings of Yin et al [17] and Stragier et al [73]. Yin et al [17] state that the user's PA level correlates positively with the amount of SM they do. The majority of interviewees felt that the SM feature provided encouragement for their PA:

> *So I feel happy because I have exercised, more enthusiasm.* [Interviewee 6]

In addition, the interviewees believed that MFAs documented or tracked their progress in PA, which helped them to maintain or even improve their exercise consistency:

> *So that I can compare with previous progress and so that in the future I can look back at my history. Like pace, I also remember what date I did sport.* [Interviewee 9]

An example of a feature that can be implemented is one that displays a summary of the user's performance while exercising, together with visualizations in the form of trends and graphs. Some apps also display comments that describe the user's sports activity performance, based on their activity level. Users can take advantage of these insights to increase their PA levels in their next sports activity.

However, H2 was rejected in this study. H2's rejection aligns with the findings of Sun and Jiang [74] and Kim et al [75]. According to Kim et al [75], social comparison and the user's level of PA are not directly connected. Social comparison here is defined as the relationship between the level of PA and the variable self-efficacy, or a person's belief in their own capabilities [75]. The rejection of H2 indicates that the community or social ecosystem around MFA users does not have a significant impact on motivating the users to exercise. On the basis of the interviews, the SS feature in the app does not have an important effect on PA levels because users do not feel compelled to exercise when using the SS feature:

> *There is no motivation from the engagement side, more from tracking my own progress.* [Interviewee 9]

In addition, nearly one-third of the interviewees (9/30, 30%) admitted that they used the SS feature only to document sports activities that had already been completed.

H3 was accepted in this study. The acceptance of H3 aligns with the studies by Bojd et al [41], Payne et al [42], Plangger et al [13], Goes et al [76], and Hamari and Koivisto [77]. Bojd et al [41] found that the gamification element in MFAs can provide two types of information: (1) the user's PA progress and (2) a comparison of the user's PA with that of other users. Furthermore, when MFA users are able to observe their progress, they feel more satisfied and recognize their own PA competency, which will drive higher user competency satisfaction and behavioral motivation [13,42]. Goes et al [76] and Hamari and Koivisto [77] also highlight the gamification element in MFA, which tracks the user's effort, progress, and achievement of personal goals. According to Goes et al [76], the public nature of user-acquired gamification elements, such as levels, badges, or leaderboards, can generate users' social status on the MFA platform, which encourages social comparison and competitive motivation among users. On the basis of the interviews, MFA users want to take part in challenges (an example of implementing gamification) on the app because they want to obtain limited edition rewards and measure their own capabilities in sports activities:

> *Gamification keeps me motivated and helps me see my activities historically during physical activity based on the badge I have earned.* [Interviewee 21]

Furthermore, the interviewees acknowledged that the rewards they obtain can be used as a benchmark of their capacity in the sports activity against which to build new achievements:

> *I feel happy when I get an achievement because it shows an improvement in my sport. Even though I don't I have specifically targeted certain achievements, but if I can surpass the previous achievements, it means that my sport has improved. The goals that I have set are higher than before.* [Interviewee 17]

Therefore, it would be better if the MFA included challenges that were personalized as well as recommendations that were based on the user's type of sports activity, the user's sports activity goals, and the user's own sports activity history. An example of such a feature could be that, based on the user's history, if they have only managed to run a distance of 3 km, then, to improve their performance, other MFA users could recommend a 4-km challenge.

Huang [28] found that sponsor characteristics play an important role in triggering user behavior. Sponsorship referred to circumstances where the use of a sponsor's product occurred naturally as part of a sponsored event [78]; for example, with an MFA whose function is to promote PA, sponsorship of athletic apparel would be perceived as highly congruent, whereas sponsorship of a cold remedy would reflect low congruence. The H4 finding is in line with the study by Yang et al [79], who stated that the level of involvement of a brand produces a positive association with the brand and strengthens the positive effect of an evaluation impacting one's behavioral intention toward an app. From the interviews, it was found that interviewees were encouraged to take part in a challenge or activity if the activity was associated with the party (public figure, company, etc) that organized it:

> *For a club other than Strava, I think it's cool if you participate, for example, it's like unique. There's definitely a challenge made by Strava every month, so it's not as special as other clubs. The limited edition is more about Heart Month, New Year, and others. I want to take part because it would be a shame if I didn't follow.* [Interviewee 9]

We found that not many interviewees took advantage of EI incentives, but those who did participate focused more on the challenges than on the organizers or the external community. If a user felt capable of taking part in a challenge, they would try to do so:

> *Actually, I see from the challenge, if I feel capable, then I want to join.* [Interviewee 7]

Thus, we argue that it would be better if MFA developers or providers developed challenges for their apps that are created by communities, organizations, and figures with high functional congruence.

With regard to H5, H6, H7, and H8, the results show that, in every case, gender has a moderating effect on the relationship between the variables investigated. This study showed that gender influences the relationship between SM and PU (H5). These results are supported by the studies by Gabriel and Gardner [46] and Sun et al [51], who found that men tend to make decisions based on rationality, while women tend to be more perceptual. According to Gabriel and Gardner [46] and Sun et al [51], men are generally more focused on personal goals that demonstrate their individual performance and abilities, while women are usually less conscious of their own goals and performance. This finding is supported by van Elburg et al [80], who state that men focus more on practical goals and achieving goals when using an mHealth app. We found that our female interviewees usually used the metrics in MFAs for tracking their PA only as monitoring information:

> *I only look at the pulse.* [Interviewee 1]

However, the men usually used these metrics as targets for self-development:

> *To find out whether in sports we have reached the desired target or not. On the other hand, if our sports performance is good, this can also be seen through the information displayed on Apple Watch. Thus, the Apple Watch can be a helpful tool in determining whether our performance has reached the expected level or not.* [Interviewee 12]

Moreover, this study found that gender influenced the extent to which SS incentives affected users' PU (H6). The results of the interviews showed that most female respondents felt more motivated by their social community or by the SS feature provided in the MFA they used. By contrast, the male users used the SS feature, such as sharing their sports activity progress, for personal documentation purposes:

> *Just so you know. Only for review, not to share with other friends.* [Interviewee 28]

Other male respondents stated that this was the case simply because the app posted their activity automatically*:*

> *Because it has to be posted on the Strava application.* [Interviewee 8]

Many male respondents had never used this feature, indicating their lack of interest in the SS feature:

> *I have never tried it.* [Interviewee 18]

However, the female respondents all expressed interest in the SS feature available in MFAs and felt more motivated to exercise due to this feature:

> *I also become motivated to exercise when I see my friends after posting their sports results.* [Interviewee 11]

Some of the female respondents commented that the SS feature of MFAs motivated them to exercise by creating a sense of competition:

> *If I just wake up in the morning and get a notification that my friend has finished exercising, I feel left behind because I just woke up but he has finished exercising. Section it motivates, really.* [Interviewee 9]

Relatedly, Li et al [33] found that women pay greater attention to social relations and are more willing to accept support from those around them. By contrast, Leong et al [54] found that men usually ignore external advice or support due to their sense of independence. These findings are supported by Yin et al [17], who found that SS had a more positive effect on PA in women than in men.

This study also showed that gender influences the relationship between platform rewards and PU (H7). The interviews showed that male respondents were generally more motivated by the challenges, badges, and awards offered by the MFA they were using:

> *Makes me more enthusiastic for the next run, and I use it to keep track of whether I should improve or maintain, for example, I can rank third so I feel I have to improve my performance.* [Interviewee 16]

However, the female respondents usually followed or used this feature only for their own satisfaction and without specific targets or motivations:

> *There is no specific goal to get rewards, but I feel happy and proud of myself if I get them.* [Interviewee 13]

According to Yin et al [17], in the context of PA, men usually pay more attention to meeting their needs for autonomy and competence, such as badges, awards, and so on. This was also demonstrated by Vilela and Nelson [52], who stated that men tend to be more aggressive, pragmatic, and self-oriented. Therefore, they are motivated by the need for achievement when using information system products [52]. Similar findings were identified by Forman et al [81], who showed that the gamification element has a more positive effect on men than on women by arousing their competitive and achievement-oriented motivation. Brandts et al [82] also support this finding and explain that task-based goal setting increases task completion and performance only for men.

This study also showed that gender affected the impact of EI on the user's PU of MFAs (H8). The results of the interviews confirmed that there are 2 main reasons a person will participate in PA supported by the MFA: the match between the organizer of the activity and the user and the match between the user's capabilities and the activity or challenge created. Comparing these 2 reasons, we found that the women were more likely to do something because of a match with the organizers, in contrast to the men, who usually focused more on their own ability to participate in an activity:

> *If Strava doesn't have motivation, if it's a club other than in my opinion, Strava is cool if you join, it's like unique. What Strava makes is there every month, so it's not as special as other clubs to participate on Strava.* [Interviewee 29]

According to Huang [28] and Yang et al [79], the reason female MFA users participate in sports activities is that they experience a *special* feeling because these sports activities are created by a special club. Huang [28] and Yang et al [79] explain that the sponsorship characteristics of a sports activity and high brand involvement play important roles in triggering the behavioral intention of MFA users and their behavior in general. H8 is also supported by the findings of Weman Josefsson et al [56], who explain that men tend to be more influenced by winning rewards than women, who tend to participate more for autonomous and social reasons.

Furthermore, H9 was confirmed in this study. The acceptance of H9 is in accordance with previous research conducted by Bhattacherjee [38], Huang et al [15], Chiu et al [25], Wang et al [9], Cai et al [83], and Wu et al [66]. Wu et al [66] found that PU and user satisfaction are directly influenced by confirmation of expectations, namely, the realization of the expected benefits of using mHealth. This result is supported by Chiu et al [25], who state that PU of the MFA is reflected in the user's enhanced exercise capacity and satisfaction, as evidenced by their increased enjoyment of exercising. Thus, it is to be expected that, after the initial experience, the confirmation level of the user's expectations will have a positive effect on their PU [9,15,38,83]. One of the expectations of a respondent who used an MFA was that they would experience changes and improvements in their PA or exercise, and these expectations were indeed successfully confirmed:

> *Because when I want to download Strava I want to be diligent in exercising, and it is proven that I exercise more often because I can track my sports progress.* [Interviewee 23]

H10 was also accepted by this study, and this result is in accordance with the studies by Bhattacherjee [38], Huang et al [15], Chiu et al [25], Wang et al [9], Cai et al [83], and Wu et al [66]. Wang et al [9] found that confirmation of expectations positively affects user satisfaction with IT products and services. The results of the interviews confirmed that interviewees felt satisfaction when using MFAs:

> *From a user point of view, everything has been fulfilled in my opinion. What I need so far has been achieved.* [Interviewee 24]

> *In my opinion, the features are quite complete, because that's all I really need. The application also provides a reminder if you have passed one day without exercising and automatically arranges for the workout that can be fulfilled the next day to be even tougher.* [Interviewee 10]

This study also showed that PU influences CUI. Acceptance of H11a is in accordance with the studies by Bhattacherjee [38], Huang et al [15], Chiu et al [25], Huang and Ren [26], Wang et al [9], Cai et al [83], Wu et al [66], and Cho et al [24]. Cho et al [24] reported that, in the context of MFAs, perceived benefits were associated with managing health-related information. The interviews confirmed that interviewees would continue using the MFAs if they helped them to be more active in their exercising, and they could track their sports activity progress effectively:

> *I will continue to use it because in my opinion it is also effective and looks simple.* [Interviewee 10]

> *As long as device is connected to the Apple Watch, will still use it. The ability to track different types of exercise separately is one of the advantages of the Apple Watch. This makes me still choose to use the Apple Watch in the future, as long as it meets my sporting needs.* [Interviewee 12]

The study's acceptance of H11b is in accordance with the studies by Bhattacherjee [38], Huang et al [15], Chiu et al [25], Wang et al [9], Cai et al [83], and Wu et al [66]. Cai et al [83] explain that PU is reflected in user satisfaction when exercising using an MFA. The more benefits users obtain from the MFA, the greater their satisfaction [25]. Wang et al [9] also found that satisfaction was a partial mediator between CUI and PU. We found that the level of user satisfaction with an MFA was based not only on its meeting users' sports activity expectations but also on the convenience and effectiveness of the features, the user interface, and the user experience that supported the user's sports activities:

> *I will continue to use Strava, because I am comfortable with Strava.* [Interviewee 14]
>
> *I will continue to use it, because in my opinion it is also effective and the appearance is not a hassle.* [Interviewee 10]
>
> *What makes me satisfied is the user interface, which is easy to use, and the user experience is simple.* [Interviewee 19]

Finally, the effect of satisfaction on CUI was confirmed in this study. The acceptance of H12 is in accordance with the studies by Bhattacherjee [38], Huang et al [15], Chiu et al [25], Wang et al [9], Cai et al [83], and Wu et al [66]. Wu et al [66] found that satisfied users are more likely to continue using an app because dissatisfied users can easily switch to other mHealth technologies. User satisfaction is an important determinant of the postadoption behavior of users of IT products and services [38]. This is supported by Chiu et al [25], who state that user satisfaction with the use of IT products and services is very important for fostering long-term use of IT. The main reason for user satisfaction with an MFA is that the features are complete and meet user needs, with the result that they come to depend on the MFA for their exercise routines:

> *Because I really like it and I have become very dependent on this application for sports. I don't want to exercise if there is no access to this application.* [Interviewee 9]
>
> *This application has fulfilled my daily needs.* [Interviewee 15]

Discussion

Principal Findings

The findings from this study extend previous research by examining the incentive system in MFAs [17,26] and the use of the ECT in the context of mHealth [25,33,38,62,66]. It also expands the understanding of the moderating effect of gender on incentive-based systems [68]. We found that MFAs and the incentives they offer have a strong influence on users' sports activity behaviors and on their intention to continue using the app. The results of this study indicate that the most influential feature of an MFA is the SM incentive feature. MFA users often do not feel like exercising or engaging in PA if the activity is not being tracked by their app. The SM feature was also found to have a greater impact on male users than on female users. This finding regarding gender differs from the results of a study by Yin et al [17], who stated that no gender trend was evident in the effectiveness of the SM feature. Furthermore, in contrast to the study by Yin et al [17], we found that SS had little effect on the PA of MFA users. The results of the qualitative interviews indicate that this is because the social circle of Indonesian MFA users is relatively small, and this small social circle affects the effectiveness of the SS feature.

MFA service providers should evaluate how different app features impact users of different genders to effectively motivate users to keep using their app in the long term. In addition, users feel more satisfied when their expectations regarding the use of an app are met. App developers can increase the PU of their MFA by using the users' social communities (eg, by creating social profile features, group exercises, sporting events organized by recognized organizations or communities, and personalized challenges or awards based on the user's sports activity history). App developers can improve the accuracy of the tracking feature, whether through a smartphone or a smartwatch, with the goal of providing users with more in-depth statistics and data. For user convenience, app providers should also develop tracking features that start automatically.

Limitations

The respondents to both the quantitative and qualitative studies were predominantly aged 17 to 25 years and female; thus, other moderating variables could be considered in a future study. The weak effect size for PU in Table 4 indicates that the differences or relationships between some variables were not significant. This suggests that there are other variables that might influence PU, which were not considered in this study. In future research, another variable that could be considered is PA. This could serve as a metric to determine whether using MFAs with specific incentives increases users' PA [17].

Conclusions

The results of the study show that SM, platform rewards, and EI can all influence the PU of MFAs. However, no relationship was found between SS and the PU of MFAs. Indonesians generally consider MFAs to be useful because these apps allow them to track their sports activities and also offer rewards and awards. The confirmation of a user's initial expectations also affects their perceptions of the usefulness of MFAs. PU and confirmation of expectations also affect user satisfaction with MFAs, which in turn influences the user's desire to continue using the MFA. In addition, gender was shown to influence user behavior when using MFAs. In future research, the scope of EI incentives could be expanded by considering financial reasons for exercising, other people's recommendations, and job demands, among other factors. We suggest considering tangible benefits as additional incentives to determine whether quantifiable benefits, such as assets or money, can increase a person's motivation to exercise.

Acknowledgments

This research was supported by a 2024 internal funding from the Faculty of Computer Science, Universitas Indonesia.

Data Availability

The data sets generated and analyzed during this study are not publicly available due to a lack of authorization to share these data.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study questionnaire and interview questions.
[DOCX File , 21 KB - humanfactors_v11i1e50957_app1.docx]

References

1. Physical activity. World Health Organization. 2022. URL: https://www.who.int/news-room/fact-sheets/detail/physical-activity [accessed 2023-02-09]
2. The global status report on physical activity. World Health Organization. 2022. URL: https://tinyurl.com/mr3hnxnx [accessed 2023-02-09]
3. Global recommendations on physical activity for health. World Health Organization. URL: https://tinyurl.com/ymrcjnmk [accessed 2023-03-10]
4. Consumer sentiment in Indonesia during the coronavirus crisis. McKinsey & Company. 2022. URL: https://tinyurl.com/84bhd2ub [accessed 2023-02-09]
5. Fitness apps- Indonesia. Statista. 2023. URL: https://tinyurl.com/4ary4us8 [accessed 2023-03-09]
6. Edney SM, Olds TS, Ryan JC, Vandelanotte C, Plotnikoff RC, Curtis RG, et al. A social networking and gamified app to increase physical activity: cluster RCT. Am J Prev Med 2020 Feb;58(2):e51-e62 [FREE Full text] [doi: 10.1016/j.amepre.2019.09.009] [Medline: 31959326]
7. Mao X, Zhao X, Liu Y. mHealth app recommendation based on the prediction of suitable behavior change techniques. Decis Support Syst 2020 May;132:113248 [FREE Full text] [doi: 10.1016/j.dss.2020.113248]
8. Tu R, Hsieh P, Feng W. Walking for fun or for “likes”? The impacts of different gamification orientations of fitness apps on consumers’ physical activities. Sport Manag Rev 2021 Feb 03;22(5):682-693 [FREE Full text] [doi: 10.1016/j.smr.2018.10.005]
9. Wang Y, Wang Y, Greene B, Sun L. An analysis and evaluation of quality and behavioral change techniques among physical activity apps in China. Int J Med Inform 2020 Jan;133:104029 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.104029] [Medline: 31759245]
10. Baker RC, Kirschenbaum DS. Self-monitoring may be necessary for successful weight control. Behav Ther 1993;24(3):377-394 [FREE Full text] [doi: 10.1016/s0005-7894(05)80212-6]
11. Boutelle KN, Kirschenbaum DS. Further support for consistent self-monitoring as a vital component of successful weight control. Obes Res 1998 May;6(3):219-224 [FREE Full text] [doi: 10.1002/j.1550-8528.1998.tb00340.x] [Medline: 9618126]
12. Yan L, Tan Y. Good intentions, bad outcomes: the effects of mismatches between social support and health outcomes in an online weight loss community. Prod Oper Manag 2018 Jan 01;27(1):9-27 [FREE Full text] [doi: 10.1111/poms.12793]
13. Plangger K, Campbell C, Robson K, Montecchi M. Little rewards, big changes: using exercise analytics to motivate sustainable changes in physical activity. Inf Manag 2022 Jul;59(5):103216 [FREE Full text] [doi: 10.1016/j.im.2019.103216]
14. Uetake K, Yang N. Inspiration from the “biggest loser”: social interactions in a weight loss program. Mark Sci 2020 May;39(3):487-499 [FREE Full text] [doi: 10.1287/mksc.2018.1116]
15. Huang CK, Chen CD, Liu YT. To stay or not to stay? Discontinuance intention of gamification apps. Inform Technol Peopl 2019 Dec 02;32(6):1423-1445 [FREE Full text] [doi: 10.1108/itp-08-2017-0271]
16. Zhu Y, Wang R, Zeng R, Pu C. Does gender really matter? Exploring determinants behind consumers' intention to use contactless fitness services during the COVID-19 pandemic: a focus on health and fitness apps. Internet Res 2022 Mar 18;33(1):280-307 [FREE Full text] [doi: 10.1108/intr-07-2021-0454]
17. Yin Q, Li L, Yan Z, Guo C. Understanding the effects of self-peer-platform incentives on users' physical activity in mobile fitness apps: the role of gender. Inform Technol Peopl 2021 May 20;35(3):1054-1072 [FREE Full text] [doi: 10.1108/itp-10-2020-0705]
18. Miller AS, Cafazzo JA, Seto E. A game plan: gamification design principles in mHealth applications for chronic disease management. Health Informatics J 2016 Jun;22(2):184-193 [FREE Full text] [doi: 10.1177/1460458214537511] [Medline: 24986104]
19. de Zambotti M, Claudatos S, Inkelis S, Colrain IM, Baker FC. Evaluation of a consumer fitness-tracking device to assess sleep in adults. Chronobiol Int 2015;32(7):1024-1028 [FREE Full text] [doi: 10.3109/07420528.2015.1054395] [Medline: 26158542]

20. Sama PR, Eapen ZJ, Weinfurt KP, Shah BR, Schulman KA. An evaluation of mobile health application tools. JMIR Mhealth Uhealth 2014 May 01;2(2):e19 [FREE Full text] [doi: 10.2196/mhealth.3088] [Medline: 25099179]
21. Willms A, Rhodes RE, Liu S. Effects of mobile-based financial incentive interventions for adults at risk of developing hypertension: feasibility randomized controlled trial. JMIR Form Res 2023 Mar 24;7:e36562 [FREE Full text] [doi: 10.2196/36562] [Medline: 36961486]
22. Bo Y, Liu QB, Tong Y. The effects of adopting mobile health and fitness apps on hospital visits: quasi-experimental study. J Med Internet Res 2023 Jul 28;25:e45681 [FREE Full text] [doi: 10.2196/45681] [Medline: 37505809]
23. Yuan S, Ma W, Kanthawala S, Peng W. Keep using my health apps: discover users' perception of health and fitness apps with the UTAUT2 model. Telemed J E Health 2015 Sep;21(9):735-741 [FREE Full text] [doi: 10.1089/tmj.2014.0148] [Medline: 25919238]
24. Cho H, Chi C, Chiu W. Understanding sustained usage of health and fitness apps: incorporating the technology acceptance model with the investment model. Technol Soc 2020 Nov;63:101429 [FREE Full text] [doi: 10.1016/j.techsoc.2020.101429]
25. Chiu W, Cho H, Chi CG. Consumers' continuance intention to use fitness and health apps: an integration of the expectation–confirmation model and investment model. Inform Technol Peopl 2020 Jun 26;34(3):978-998 [FREE Full text] [doi: 10.1108/itp-09-2019-0463]
26. Huang G, Ren Y. Linking technological functions of fitness mobile apps with continuance usage among Chinese users: moderating role of exercise self-efficacy. Comput Human Behav 2020 Feb;103:151-160 [FREE Full text] [doi: 10.1016/j.chb.2019.09.013]
27. Teng X, Bao Z. Factors affecting users' stickiness of fitness apps: an empirical study based on the S-O-R perspective. Int J Sports Mark Spons 2022 Mar 30;23(4):823-840 [FREE Full text] [doi: 10.1108/ijsms-06-2021-0123]
28. Huang G. Does warm glow promote physical activity? Examining the relative effectiveness of self-benefiting versus other-benefiting incentives in motivating fitness app use by corporate sponsorship programs. Health Commun 2021 Nov;36(13):1719-1730 [FREE Full text] [doi: 10.1080/10410236.2020.1794548] [Medline: 32698621]
29. Pearson E, Prapavessis H, Higgins C, Petrella R, White L, Mitchell M. Adding team-based financial incentives to the Carrot Rewards physical activity app increases daily step count on a population scale: a 24-week matched case control study. Int J Behav Nutr Phys Act 2020 Nov 19;17(1):139-124 [FREE Full text] [doi: 10.1186/s12966-020-01043-1] [Medline: 33208166]
30. Lemola S, Gkiouleka A, Read B, Realo A, Walasek L, Tang NK, et al. Can a 'rewards-for-exercise app' increase physical activity, subjective well-being and sleep quality? An open-label single-arm trial among university staff with low to moderate physical activity levels. BMC Public Health 2021 Apr 23;21(1):782 [FREE Full text] [doi: 10.1186/s12889-021-10794-w] [Medline: 33892665]
31. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78 [FREE Full text] [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
32. Teixeira PJ, Carraça EV, Markland D, Silva MN, Ryan RM. Exercise, physical activity, and self-determination theory: a systematic review. Int J Behav Nutr Phys Act 2012 Jun 22;9:78 [FREE Full text] [doi: 10.1186/1479-5868-9-78] [Medline: 22726453]
33. Li J, Liu X, Ma L, Zhang W. Users' intention to continue using social fitness-tracking apps: expectation confirmation theory and social comparison theory perspective. Inform Health Soc Care 2019 Sep;44(3):298-312 [FREE Full text] [doi: 10.1080/17538157.2018.1434179] [Medline: 29504821]
34. Bandura A. Social cognitive theory of self-regulation. Organ Behav Hum Decis Process 1991 Dec;50(2):248-287. [doi: 10.1016/0749-5978(91)90022-l]
35. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Manag Sci 1989 Aug;35(8):982-1003 [FREE Full text] [doi: 10.1287/mnsc.35.8.982]
36. Barkley JE, Lepp A, Santo A, Glickman E, Dowdell B. The relationship between fitness app use and physical activity behavior is mediated by exercise identity. Comput Human Behav 2020 Jul;108:106313 [FREE Full text] [doi: 10.1016/j.chb.2020.106313]
37. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
38. Bhattacherjee A. Understanding information systems continuance: an expectation-confirmation model. MIS Q 2001 Sep;25(3):351 [FREE Full text] [doi: 10.2307/3250921]
39. Wang Q, Egelandsdal B, Amdam GV, Almli VL, Oostindjer M. Diet and physical activity apps: perceived effectiveness by app users. JMIR Mhealth Uhealth 2016 Apr 07;4(2):e33 [FREE Full text] [doi: 10.2196/mhealth.5114] [Medline: 27056639]
40. Chen Y, Pu P. Healthy together: exploring social incentives for mobile fitness applications. In: Proceedings of the 2nd International Symposium of Chinese CHI. 2014 Presented at: Chinese CHI '14; April 26-27, 2014; Toronto, ON p. 25-34 URL: https://doi.org/10.1145/2592235.2592240 [doi: 10.1145/2592235.2592240]
41. Bojd B, Song X, Tan Y, Yan X. Gamified challenges in online weight-loss communities. SSRN Journal. Preprint posted online April 25, 2018 2018 [FREE Full text] [doi: 10.2139/ssrn.3157331]

42. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
43. Bollók S, Takács J, Kalmár Z, Dobay B. External and internal sport motivations of young adults. Biomed Hum Kinet 2011;3:101-105 [FREE Full text] [doi: 10.2478/V10101-011-0022-5]
44. Koo S, Byon KK, Baker TA. Integrating event image, satisfaction, and behavioral intention: small-scale marathon event. Sport Mark Q 2014;23:59 [FREE Full text]
45. Low XT, Pyun DY. Consumers' perceived functions of and attitude toward corporate sponsors of small-scale amateur sporting events. Event Manag 2016 Jul 25;20(2):227-238 [FREE Full text] [doi: 10.3727/152599516x14643674421771]
46. Gabriel S, Gardner WL. Are there "his" and "hers" types of interdependence? The implications of gender differences in collective versus relational interdependence for affect, behavior, and cognition. J Pers Soc Psychol 1999 Sep;77(3):642-655 [FREE Full text] [doi: 10.1037//0022-3514.77.3.642] [Medline: 10510513]
47. Ryan RM, Williams GC, Heather H, Deci EL. Self-determination theory and physical activity: the dynamics of motivation in development and wellness. Hellenic J Psychol 2009;6(2):107-124 [FREE Full text]
48. Weisskopf WA. The duality of human existence: an essay on psychology and religion by David Bakan. Am J Sociol 1967;73(3):368-369 [FREE Full text] [doi: 10.1086/224499]
49. Spence JT, Helmreich RL. Masculinity and femininity: their psychological dimensions, correlates, and antecedents. Psychol Women Q 1978;4(1):144-146 [FREE Full text] [doi: 10.7560/764439]
50. Kimbrough AM, Guadagno RE, Muscanell NL, Dill J. Gender differences in mediated communication: women connect more than do men. Comput Human Behav 2013 May;29(3):896-900 [FREE Full text] [doi: 10.1016/j.chb.2012.12.005]
51. Sun Y, Lim KH, Jiang C, Peng JZ, Chen X. Do males and females think in the same way? An empirical investigation on the gender differences in web advertising evaluation. Comput Human Behav 2010 Nov;26(6):1614-1624 [FREE Full text] [doi: 10.1016/j.chb.2010.06.009]
52. Vilela AM, Nelson MR. Testing the selectivity hypothesis in cause-related marketing among generation Y: [when] does gender matter for short- and long-term persuasion? J Mark Comm 2013 Oct 11;22(1):18-35 [FREE Full text] [doi: 10.1080/13527266.2013.841272]
53. Sun H, Zhang P. The role of moderating factors in user technology acceptance. Int J Hum Comput Stud 2006 Feb;64(2):53-78 [FREE Full text] [doi: 10.1016/j.ijhcs.2005.04.013]
54. Leong LY, Ooi KB, Chong AY, Lin B. Modeling the stimulators of the behavioral intention to use mobile entertainment: does gender really matter? Comput Human Behav 2013 Sep;29(5):2109-2121 [FREE Full text] [doi: 10.1016/j.chb.2013.04.004]
55. Venkatesh V, Thong JY, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157-178. [doi: 10.2307/41410412]
56. Weman Josefsson K, Johnson U, Lindwall M. Short report: moderations in exercise motivation - gender and age moderates the relations of motivation quality and exercise behavior. Health Psychol Behav Med 2018 Apr 18;6(1):93-103 [FREE Full text] [doi: 10.1080/21642850.2018.1462706] [Medline: 34040823]
57. Albashrawi M, Motiwalla L. When IS success model meets UTAUT in a mobile banking context: a study of subjective and objective system usage. In: Proceedings of the 30th Annual Workshop of the Swedish Artificial Intelligence Society. 2017 Presented at: SAIS '17; May 15-16, 2017; Karlskrona, Sweden URL: https://aisel.aisnet.org/sais2017/1/
58. Susanto A, Chang Y, Ha Y. Determinants of continuance intention to use the smartphone banking services. Ind Manage Data Syst 2016;116:508-525 [FREE Full text] [doi: 10.1108/imds-05-2015-0195]
59. Oghuma AP, Libaque-Saenz CF, Wong SF, Chang Y. An expectation-confirmation model of continuance intention to use mobile instant messaging. Telemat Inform 2016 Feb;33(1):34-47 [FREE Full text] [doi: 10.1016/j.tele.2015.05.006]
60. Festinger L. A Theory of Cognitive Dissonance. Stanford, CA: Stanford University Press; 1957.
61. Bhattacherjee A. An empirical analysis of the antecedents of electronic commerce service continuance. Decis Support Syst 2001 Dec;32(2):201-214 [FREE Full text] [doi: 10.1016/s0167-9236(01)00111-7]
62. Hsu CL, Lin JC. What drives purchase intention for paid mobile apps? – An expectation confirmation model with perceived value. Electron Commer Res Appl 2015 Jan;14(1):46-57 [FREE Full text] [doi: 10.1016/j.elerap.2014.11.003]
63. Ambalov IA. A meta-analysis of IT continuance: an evaluation of the expectation-confirmation model. Telemat Inform 2018 Sep;35(6):1561-1571 [FREE Full text] [doi: 10.1016/j.tele.2018.03.016]
64. Dehghani M, Kim KJ, Dangelico RM. Will smartwatches last? factors contributing to intention to keep using smart wearable technology. Telemat Inform 2018 May;35(2):480-490 [FREE Full text] [doi: 10.1016/j.tele.2018.01.007]
65. Nascimento B, Oliveira T, Tam C. Wearable technology: what explains continuance intention in smartwatches? J Retail Consum Serv 2018 Jul;43:157-169 [FREE Full text] [doi: 10.1016/j.jretconser.2018.03.017]
66. Wu C, Zhou Y, Wang R, Huang S, Yuan Q. Understanding the mechanism between IT identity, IT mindfulness and mobile health technology continuance intention: an extended expectation confirmation model. Technol Forecast Soc Change 2022 Mar;176:121449 [FREE Full text] [doi: 10.1016/j.techfore.2021.121449]

67. Vinnikova A, Lu L, Wei J, Fang G, Yan J. The use of smartphone fitness applications: the role of self-efficacy and self-regulation. Int J Environ Res Public Health 2020 Oct 20;17(20):7639 [FREE Full text] [doi: 10.3390/ijerph17207639] [Medline: 33092090]
68. Sharma P, Sivakumaran B, Marshall R. Impulse buying and variety seeking: a trait-correlates perspective. J Bus Res 2010 Mar;63(3):276-283 [FREE Full text] [doi: 10.1016/j.jbusres.2009.03.013]
69. Sheikh Z, Yezheng L, Islam T, Hameed Z, Khan IU. Impact of social commerce constructs and social support on social commerce intentions. Inform Technol Peopl 2019 Feb 04;32(1):68-93 [FREE Full text] [doi: 10.1108/itp-04-2018-0195]
70. Ng JY, Ntoumanis N, Thøgersen-Ntoumani C, Deci EL, Ryan RM, Duda JL, et al. Self-determination theory applied to health contexts: a meta-analysis. Perspect Psychol Sci 2012 Jul;7(4):325-340 [FREE Full text] [doi: 10.1177/1745691612447309] [Medline: 26168470]
71. Hair Jr JF, Matthews LM, Matthews RL, Sarstedt M. PLS-SEM or CB-SEM: updated guidelines on which method to use. Int J Multivar Data Anal 2017;1(2):107 [FREE Full text] [doi: 10.1504/ijmda.2017.10008574]
72. Awang Z. SEM Made Simple: A Gentle Approach to Learning Structural Equation Modelling. Bangi, Malaysia: MPWS Rich Publication; 2015.
73. Stragier J, Vanden Abeele M, Mechant P, De Marez L. Understanding persistence in the use of online fitness communities: comparing novice and experienced users. Comput Human Behav 2016 Nov;64:34-42 [FREE Full text] [doi: 10.1016/j.chb.2016.06.013]
74. Sun M, Jiang LC. Linking social features of fitness apps with physical activity among Chinese users: evidence from self-reported and self-tracked behavioral data. Inf Process Manage 2022 Nov;59(6):103096 [FREE Full text] [doi: 10.1016/j.ipm.2022.103096]
75. Kim HM, Cho I, Kim M. Gamification aspects of fitness apps: implications of mHealth for physical activities. Int J Hum Comput Interact 2022 May 13;39(10):2076-2089 [FREE Full text] [doi: 10.1080/10447318.2022.2073322]
76. Goes PB, Guo C, Lin M. Do incentive hierarchies induce user effort? Evidence from an online knowledge exchange. Inf Syst Res 2016 Sep;27(3):497-516 [FREE Full text] [doi: 10.1287/isre.2016.0635]
77. Hamari J, Koivisto J. “Working out for likes”: an empirical study on social influence in exercise gamification. Comput Human Behav 2015 Sep;50:333-347 [FREE Full text] [doi: 10.1016/j.chb.2015.04.018]
78. Pappu R, Cornwell TB. Corporate sponsorship as an image platform: understanding the roles of relationship fit and sponsor–sponsee similarity. J Acad Mark Sci 2014 Feb 16;42(5):490-510 [FREE Full text] [doi: 10.1007/s11747-014-0373-x]
79. Yang J, Kanthawala S, Joo E, Kononova A. Can brand sponsorship increase download intention for mHealth apps? The role of issue relevance, brand involvement, and perceived app quality. J Promot Manag 2021 Dec 27;28(6):869-892 [FREE Full text] [doi: 10.1080/10496491.2021.2015513]
80. van Elburg FR, Klaver NS, Nieboer AP, Askari M. Gender differences regarding intention to use mHealth applications in the Dutch elderly population: a cross-sectional study. BMC Geriatr 2022 May 24;22(1):449 [FREE Full text] [doi: 10.1186/s12877-022-03130-3] [Medline: 35610577]
81. Forman EM, Manasse SM, Dallal DH, Crochiere RJ, Berry MP, Butryn ML, et al. Gender differences in the effect of gamification on weight loss during a daily, neurocognitive training program. Transl Behav Med 2021 Apr 26;11(4):1015-1022 [FREE Full text] [doi: 10.1093/tbm/ibaa091] [Medline: 32945860]
82. Brandts J, El Baroudi S, Huber SJ, Rott C. Gender differences in private and public goal setting. J Econ Behav Organ 2021 Dec;192:222-247 [FREE Full text] [doi: 10.1016/j.jebo.2021.09.012]
83. Cai J, Zhao Y, Sun J. Factors influencing fitness app users’ behavior in China. Int J Hum Comput Interact 2021 May 17;38(1):53-63 [FREE Full text] [doi: 10.1080/10447318.2021.1921483]

Abbreviations

CUI: continuance usage intention
ECT: expectation-confirmation theory
EI: external influence
MFA: mobile fitness app
mHealth: mobile health
PA: physical activity
PU: perceived usefulness
SDT: self-determination theory
SM: self-monitoring
SS: social support

Edited by K Adapa; submitted 18.07.23; peer-reviewed by M Mardiana, H Namba; comments to author 27.01.24; revised version received 05.02.24; accepted 29.04.24; published 05.06.24.

Please cite as:
Faizah A, Hardian AFA, Nandini RD, Handayani PW, Harahap NC
The Influence of Incentive-Based Mobile Fitness Apps on Users' Continuance Intention With Gender Moderation Effects: Quantitative and Qualitative Study
JMIR Hum Factors 2024;11:e50957
URL: https://humanfactors.jmir.org/2024/1/e50957
doi:10.2196/50957
PMID:38837199

Original Paper

Using a Smartwatch App to Understand Young Adult Substance Use: Mixed Methods Feasibility Study

Sahiti Kunchay[1], BTech; Ashley N Linden-Carmichael[2], PhD; Saeed Abdullah[1], PhD

[1]College of Information Sciences and Technology, The Pennsylvania State University, University Park, PA, United States

[2]Edna Bennett Pierce Prevention Research Center, The Pennsylvania State University, University Park, PA, United States

Corresponding Author:
Sahiti Kunchay, BTech
College of Information Sciences and Technology
The Pennsylvania State University
Westgate Building
University Park, PA, 16802
United States
Phone: 1 (814) 865 8947
Email: sahiti@psu.edu

Abstract

Background: Young adults in the United States exhibit some of the highest rates of substance use compared to other age groups. Heavy and frequent substance use can be associated with a host of acute and chronic health and mental health concerns. Recent advances in ubiquitous technologies have prompted interest and innovation in using technology-based data collection instruments to understand substance use and associated harms. Existing methods for collecting granular, real-world data primarily rely on the use of smartphones to study and understand substance use in young adults. Wearable devices, such as smartwatches, show significant potential as platforms for data collection in this domain but remain underused.

Objective: This study aims to describe the design and user evaluation of a smartwatch-based data collection app, which uses ecological momentary assessments to examine young adult substance use in daily life.

Methods: This study used a 2-phase iterative design and acceptability evaluation process with young adults (aged 18-25 y) reporting recent alcohol or cannabis use. In phase 1, participants (8/15, 53%) used the data collection app for 14 days on their Apple Watches to report their substance use patterns, social contexts of substance use, and psychosocial risk factors (eg, affect). After this 14-day deployment, the participants completed a user experience survey and a semistructured interview to record their perspectives and experiences of using the app. Formative feedback from this phase informed feature modification and refinement of the app. In phase 2, an additional cohort (7/15, 47%) used the modified app for 14 days and provided feedback through surveys and interviews conducted after the app use period.

Results: Analyses of overall app use patterns indicated high, consistent use of the app, with participants using the app for an average of 11.73 (SD 2.60) days out of 14 days of data collection. Participants reported 67 instances of substance use throughout the study, and our analysis indicates that participants were able to respond to ecological momentary assessment prompts in diverse temporal and situational contexts. Our findings from the user experience survey indicate that participants found the app usable and functional. Comparisons of app use metrics and user evaluation scores indicate that the iterative app design had a measurable and positive impact on users' experience. Qualitative data from the participant interviews highlighted the value of recording substance use patterns, low disruption to daily life, minimal overall burden, preference of platforms (smartphones vs smartwatches), and perspectives relating to privacy and app use in social contexts.

Conclusions: This study demonstrated the acceptability of using a smartwatch-based app to collect intensive, longitudinal substance use data among young adults. The findings document the utility of smartwatches as a novel platform to understand sensitive and often-stigmatized behaviors such as substance use with minimal burden.

(JMIR Hum Factors 2024;11:e50795) doi:10.2196/50795

KEYWORDS

smartwatches; substance use; ecological momentary assessment; mobile health; mHealth; human-centered design; feasibility studies; mobile phone

Introduction

Background

Young adults exhibit some of the highest rates of substance use across all age groups in the United States [1], including alcohol use (50.2% or 17.5 million people), cannabis use (25.9% or 9 million people), vaping nicotine (24% or 8.3 million people), and prescription psychotherapeutic drug use (7.3% or 2.5 million people). Substance use can be associated with significant long-term effects on individuals' health and well-being [2]. As such, there is an urgent need to understand, detect, and mitigate substance use among young adults.

There has been much prior work within the substance use domain in determining various psychological, social, and environmental factors that impact young adults' substance use behaviors. These studies [3-6] highlight the value of collecting mood, affect, situational, and social context data to assess how they affect substance use patterns in this population. In recent years, this domain has shifted from relying on cross-sectional surveys and retrospective data toward using ecological momentary assessments (EMAs) on a daily level to detect relevant within-person trends. With recent advances in ubiquitous technologies and the surge of interest in accessible and affordable health care, there has been an increasing focus on understanding substance use and associated consequences through technology-based solutions. Thus, in the aforementioned studies [3-6], smartphones have been the primary device of choice.

In addition to having adaptable interfaces that support EMAs, smartphones also have extensive sensors that show potential to unobtrusively detect substance use in young adult populations. Prior work in the ubiquitous-computing community has described apps that seek to collect and analyze data to predict drinking episodes. Several studies have investigated the efficacy of inferring alcohol use through a smartphone user's gait [7-9], as well as device use and movement features [10,11]. Smartphone sensors also exhibit potential in detecting cannabis use behaviors from users' gait using accelerometer and gyroscope data [12], as well as from a combination of time features and GPS, accelerometer, SMS text messaging, and smartphone logs [13].

Systems that are capable of capturing behaviors, experiences, and sensor data in real time provide researchers a deeper understanding of the various contexts in which young adults engage in substance use. Although smartphones have been successful in collecting such data and are thus widely used in substance use research, a recent review of EMA protocols determined that compliance for substance use–related EMAs deployed on participants' smartphones was lower than acceptable levels [14]. Hence, there is a need to explore novel interfaces and establish their utility in collecting granular substance use data with high compliance and low perceived burden.

Smartwatches offer a user experience that is distinct from that of a smartphone. The persistent, wearable nature of this device can enable users to observe cues (such as notifications, sounds, and vibrations) and perform quick interactions in diverse situations, such as when the smartphone is out of reach or an inconspicuous use of technology is required to minimize social disruption [15,16]. Moreover, smartwatches offer extensive health-sensing features that allow individuals to track and understand health behaviors. Thus, in recent years, there has been wide adoption of smartwatches: globally, approximately 202 million individuals own smartwatches [17], with 1 in 5 Americans using a smartwatch or fitness tracker [18]. This uptake of smartwatches by consumers has propelled researchers to investigate how smartwatches can be used as instruments of behavioral health studies. In fact, there have been several efforts to investigate whether illnesses and disorders could be recorded or managed through smartwatch-based tools such as those for managing attention-deficit/hyperactivity disorder [19] and posttraumatic stress disorder [20], aiding students with intellectual and developmental disabilities [21], assessing mobility among older adults [22], and managing chronic disorders [23]. In addition, there has also been interesting work in terms of detecting substance use behaviors, such as smoking, using these devices; for example, Skinner et al [24] used the accelerometer and gyroscope data in the Android Wear–based LG G Watch to detect signature hand movements of cigarette smoking.

Smartwatches and fitness trackers have met with resounding success in the health monitoring and self-management market [25,26]. Individuals use these devices to monitor and manage their fitness, sleep, mental health, and menstrual cycles through various apps. *Given their wide adoption for assessing health behaviors, especially by young adults [26], we argue that smartwatches may be well suited to understand substance use trends and patterns in this population*. In fact, Carreiro et al [27] highlight the significant potential yet underuse of wearables in combining detection and interventions for substance use. Importantly, the authors emphasized that wearable-smartphone combinations (such as smartwatches) are especially suitable for understanding and addressing substance use among adolescents and young adults. Recently, several studies pioneered the use of wearable sensors in understanding substance use and associated factors [28-30]. In these studies, participants noted several perceptions that suggested their preference for smartwatch-type interfaces over research-grade sensors for detecting and understanding substance use. Participants noted that these smartwatch interfaces were easy to integrate into their lives, offered various auxiliary features (such as screens, clock faces, and fitness-tracking capabilities), and drew minimal attention from strangers [28,29]. These aspects of smartwatches address many barriers that participants often face while using sensors and wearables in research studies. However, despite their potential and rapid uptake, these devices have rarely been used to assess substance use–related health behaviors in young adults.

Objectives

There is a critical need to better understand and assess substance use behaviors and trends in real-world settings, and this need has so far been addressed by using smartphones for collecting self-report and sensor data. However, the engagement and compliance rates of smartphone apps in this domain are less than ideal, indicating a need to explore the suitability of other interfaces to collect such data. Therefore, this study aims to address this need by assessing the feasibility and acceptability of using smartwatches to collect EMA and sensor data to understand young adult substance use. Our use of smartwatches for this study is motivated by several reasons. First, in recent years, there has been wide adoption of commercially available smartwatches, specifically for health assessments and interventions. Second, smartwatches offer a novel user experience, built-in health sensor data capture, and popularity within young adult populations, thus offering the potential to collect richer, more granular data to understand young adults' substance use with minimal burden. Finally, existing research in the substance use domain suggests that smartwatches may be especially suitable for understanding young adults' substance use behaviors [27].

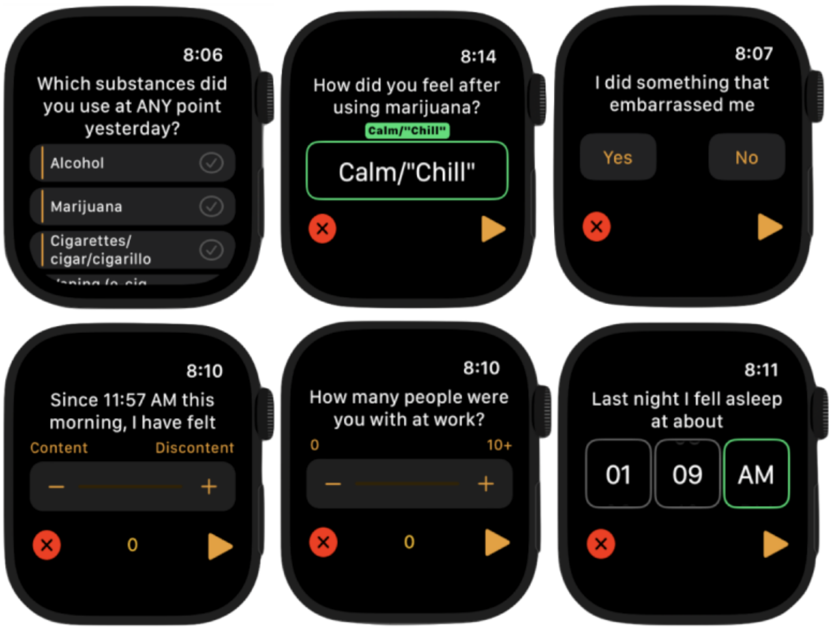
Methods

System Design and Development

Designing apps for smartwatches requires approaches and techniques that vary significantly from those required for typical smartphone app experiences. Smartwatch apps offer a seamless and intuitive experience when they are responsive; involve simple tasks; and make use of features that draw users to the device, such as haptic notifications, glanceable content, intuitive gestures, and a focused core functionality [31].

The primary requirement of the interface concerns ensuring that it provides an experience that results in highly granular and robust data collection, while keeping the perceived burden of interaction low. To address this challenge of high response rates and low study burden, we designed our questionnaire so that each question would take <5 seconds to answer. We expected that keeping the questions concise and interactions intuitive would help maintain low perceived burden and survey fatigue. Examples of these EMA components on the smartwatch (ie, an Apple Watch) are depicted in Figure 1. A companion app on the user's iPhone uploaded all user responses and sensor data to our database.

Figure 1. Screenshots of ecological momentary assessments on an Apple Watch interface depicting the variety of interface elements used to elicit responses from participants.



Iterative Application Design: Phase 1 and Phase 2

Our fundamental approach to app design and development was based on the principles of human-centered design, an approach that heavily incorporates users' experiences and perspectives throughout the design process. It is a nonlinear process that iterates continually between various stages of understanding users, defining the problem domain, generating ideas, prototyping or developing solutions, and testing. This process helps build mobile health (mHealth) systems that are usable, effective, and accessible [32,33].

The creation of the smartwatch app went through continual iterations of user evaluation and development to produce an experience that enables robust data collection while also ensuring that the app is easy to use, minimally invasive, and considerate of users' privacy and security concerns. Thus, we incorporated feedback from participants (8/15, 53%) in phase 1 of the feasibility study, so that participants (7/15, 47%) enrolled in the next phase were able to evaluate a refined app that provided a better user experience. During the initial rounds of testing and development, the app required confirmation from the database for each EMA question, causing a 1- to 2-second delay. This delay generated negative feedback from phase 1 participants. They reported that this delay between questions was frustrating and prompted them to assume that their responses were not recorded. To correct for this delay, we eliminated the step of waiting for the database confirmation before moving to the next screen.

Participants and Procedures

Our methodology for the feasibility study was informed by two main objectives: (1) to fully capture participants' experience using the smartwatch app and their perspectives on its usability; and (2) to collect data that accurately reflect users' lived experiences with substance use, social contexts, affect, behaviors, and experiences. Hence, for this study, we used a mixed methods design with 3 key components: a 14-day in situ data collection period, where participants used the app to answer short EMAs regarding their behaviors and experiences; a poststudy survey that sought to quantitatively capture the usability of the app through various dimensions; and a semistructured interview that sought to capture more nuanced perspectives on participants' experiences with the app.

To be eligible, participants needed to be aged 18 to 25 years, report past-week alcohol or cannabis use, own and use both an iPhone (with iOS version 15 or newer) and an Apple Watch (with watchOS version 8 or newer) to deploy and use the smartwatch app, and be a current student at the local university.

Participants were recruited through convenience sampling, using study flyers posted on the university campus, social media posts, and the university's StudyFinder website. Potential participants were asked to email the study team if they were interested, after which they were sent a link to the screener survey as well as more details about the study. Informed consent to participate in the study was also obtained at this stage.

Eligible participants were immediately directed to a baseline survey in which they provided demographic information, typical substance use behavior, and technological use behaviors. The screener and baseline surveys were collected and managed using REDCap (Research Electronic Data Capture; Vanderbilt University) [34]. After completing the baseline survey, participants were scheduled for a web-based visit with the research staff who explained the research activities, guided them through app installation, and informed them about the compensation structure. Of the 25 eligible participants who completed the baseline survey, 15 (60%) scheduled and attended the web-based visit. After the completion of the 14-day data collection period, participants were requested via email to upload their HealthKit (Apple Inc) data, complete the usability survey, and schedule a second web-based visit for the semistructured interview.

Of the 15 participants who used the app, 12 (80%) completed all research activities. All study activities were conducted virtually between August 2021 and May 2022.

Measures

EMA Data

The types of data we collected from the user through the EMAs related to (1) mood and general affect [35]; (2) experiences of stress; (3) sleep duration; (4) types and amounts of substances used; (5) feelings of intoxication [36,37]; (6) substance use–related consequences [38]; and (7) social context, such as location and social environment. The questions for self-reports explored a wide range of constructs and were sourced from prior research and findings that established their validity and reliability [35-38]. All constructs used in this app were motivated by a wealth of research indicating various associations with substance use [5,39-45]. A full list of all aforementioned EMA items is included in Table S1 in Multimedia Appendix 1 [35-38]).

Participants were sent 5 survey prompts per day at 11 AM, 4 PM, 7 PM, 10 PM, and 1 AM, which were available only for specific time windows or *sessions* every day (11 AM-3 PM, 4 PM-6 PM, 7 PM-9 PM, 10 PM-midnight, and 1 AM-3 AM, respectively). A brief overview of the initial design and development of this app is provided in prior work [46]. For every item, participants had the option of skipping the question if they did not wish to respond.

In the 11 AM session, participants were asked about their experiences and behaviors that occurred at any time on the previous day, and these data were grouped as *prior day* data while analyzing responses. Participants were also asked (in all sessions) about their experiences and behaviors that occurred since their last response, and these data were categorized as *periodic* data during analysis (Table S1 in Multimedia Appendix 1).

Sensor Data

In addition to self-report questionnaires, we also collected sensor data: location (GPS), physical activity, and health data streams. The health data streams serve various purposes: physical exercise, exercise intensity, and the types of exercise are all factors that have significant benefits in reducing substance use, decreasing depression symptoms associated with substance use, and improving the abstinence rate among those using illicit substances [44,45]; sleep has a bidirectional relationship with substance use in young adults, with sleep patterns and duration being significant predictors of cigarette, alcohol, and cannabis use; and the type of substance use is a significant predictor of total sleep duration as well as sleep patterns (eg, weekend oversleep) [43]. Although the limited sample size in this study hinders us from assessing whether these data streams can be effectively leveraged to unobtrusively detect substance use behaviors, the feature is incorporated into the app to examine preliminary associations, as well as for use in future studies with an anticipated larger sample size.

User Experience Evaluation

For the usability survey, we used the System Usability Scale (SUS) [47] to assess the perceived usability of the Apple Watch app, and we used an adapted version of the Mobile Application Rating Scale: User Version (uMARS) [48] and various other items to assess the acceptability of the interface and the EMAs sourced from prior work [49,50]. Both the SUS and uMARS surveys have high reliability and validity and have been extensively used to evaluate digital systems and mHealth systems, respectively.

In the semistructured interview, we queried the participants on whether the app impacted their substance use or substance cravings; whether the app influenced their awareness of substance use patterns; whether they had any concerns about using the app in various social contexts; and whether they had any privacy concerns regarding their substance use data, location data, or HealthKit data. All interviews were conducted via Zoom (Zoom Video Communications, Inc) and were recorded and transcribed using Zoom's live transcription service powered by Otter.ai. The interview script is provided in Textbox S1 in Multimedia Appendix 1.

Analysis

Only deidentified data were used during the analysis of app use data and interview data, blinding the authors to the identity of the participants while reviewing the results.

For our analysis, we focused on analyzing participants' EMA responses, app use patterns, and user perspectives to determine the feasibility and acceptability of the smartwatch app. To understand the effect of the iterative design improvements, we compared various measures between participants from both phases, treating them as separate groups during analysis. These findings are discussed in the *Results* section.

Ethical Considerations

All study activities and methods were approved by The Pennsylvania State University Institutional Review Board (17735) in the northeastern region of the United States in a state in which medical cannabis was legal, but recreational cannabis use was not legal at the time of data collection. A certificate of confidentiality was secured to protect participant responses concerning underage and illegal substance use behavior. All 15 participants provided informed consent before taking part in the study.

Participants were compensated for the study through Amazon gift cards and followed an established structure. Participants were compensated US $5 for completing the baseline survey, up to US $33 for the EMA data collection period, and US $10 for completing the user experience survey and semistructured interview. For the in-the-wild data collection period, participants were compensated US $2 per day if they completed both the 11 AM session and 1 other session during the day, but they were compensated only US $1 if they completed only the 11 AM session. Participants who did not complete the 11 AM session were not compensated for the day. If participants answered even 1 EMA during the 14-day period, they were compensated US $5.

Results

Quantity and Description of EMA Data Set

Participants ranged in age from 20 to 25 (mean 22.20, SD 1.86) years, and all were college students (undergraduate students: 10/15, 67% and graduate students: 5/15, 33%). Two-thirds (10/15, 67%) of the participants identified as female, while one-third (5/15, 33%) identified as male. Of the 15 participants, 5 (33%) identified as Asian, 1 (7%) as Black or African American, and 7 (47%) as White, while 1 (7%) participant preferred not to answer the question about race. Only 1 (7%) of the 15 participants identified as Hispanic or Latinx. Additional participant demographics are reported in Table S2 in Multimedia Appendix 1.

Overall, the 15 participants provided 4796 responses to EMA questions over 210 days. On average, the app collected 320 (SD 151; range 110-652) responses across all participants across all days of the study. Our data consisted of 45 prior-day (collected only at session 1) substance use reports, with a majority of reports mentioning alcohol use (alcohol: n=39, 87%; cannabis: n=12, 27%). We also collected 67 periodic substance use reports, which were reports collected in sessions 1, 2, 3, 4, or 5. Of these 67 periodic substance use reports, a majority included alcohol use, and a small portion included cannabis use, vape (e-cigarette or Juul e-cigarette) use, and cigarette use (reports of periodic alcohol use: n=49, 73%; reports of periodic cannabis use: n=13, 19%). Table 1 details all instances of substance use reported by the participants. Of the 15 participants, 3 (20%) did not report any substance use during the study.

Table 1. App use and substance use reports by participants.

Phase and participant	Total days participated (n=14), n (%)	Total sessions completed (n=70), n (%)	Days compliant (n=14), n (%)	Total EMAs[a] answered, n	Longest consecutive use of app (days; n=14), n (%)	Prior-day substance use reports (n=45), n (%)	Periodic substance use reports (n=67), n (%)
Phase 1							
P1	10 (71)	18 (26)	5 (36)	224	9 (64)	1 (2; alcohol)	1 (1; alcohol)
P2	8 (57)	27 (39)	7 (50)	271	8 (57)	2 (4; alcohol)	2 (3; alcohol)
P3	13 (93)	21 (30)	8 (57)	441	7 (50)	11 (24; alcohol: n=5, 45; cannabis: n=11, 100; vape: n=11, 100)	20 (30; alcohol: n=5, 25; cannabis: n=12, 60; vape: n=18, 90)
P4	13 (93)	25 (36)	6 (43)	259	10 (71)	1 (2; alcohol)	3 (4; alcohol)
P5	7 (50)	12 (17)	5 (36)	110	2 (14)	0 (0)	1 (1; other)
P6	14 (100)	55 (79)	14 (100)	559	14 (100)	1 (2; alcohol)	3 (4; alcohol)
P7	8 (57)	10 (14)	2 (14)	147	7 (50)	3 (7; alcohol)	4 (6; alcohol)
P8	14 (100)	36 (51)	12 (86)	424	14 (100)	4 (9; alcohol)	4 (6; alcohol)
Phase 2							
P9	14 (100)	36 (51)	10 (71)	371	14 (100)	0 (0)	0 (0)
P10	12 (86)	19 (27)	6 (43)	235	10 (71)	2 (4; alcohol: n=2, 100; cannabis: n=1, 50)	2 (3; alcohol: n=2, 100; cannabis: n=1, 50)
P11	14 (100)	28 (40)	13 (93)	380	14 (100)	9 (20; alcohol)	3 (4; alcohol: n=2, 67; cigarettes/cigar/cigarillo: n=1, 33)
P12	12 (86)	22 (31)	8 (57)	219	4 (29)	0 (0)	0 (0)
P13	14 (100)	26 (37)	10 (71)	289	14 (100)	0 (0)	0 (0)
P14	14 (100)	46 (66)	13 (93)	652	14 (100)	10 (22; alcohol)	22 (33; alcohol)
P15	9 (64)	21 (30)	4 (29)	215	9 (94)	1 (2; alcohol)	1 (1; alcohol)

[a]EMA: ecological momentary assessment.

When analyzing the intensity of substance use, we found that participants reported an average consumption of 3.44 (SD 3.09; min=1, max=≥10) alcoholic drinks, with more positive (mean 3.82, SD 2.05; range 0-6) than negative consequences related to alcohol use (mean 0.33, SD 0.66; range 0-3), while participants reporting prior-day cannabis use reported consuming an average of 8.92 (SD 2.23; min=3, max=≥10) hits, with an average of 2.667 (1.370; range 1-6) positive consequences and no negative consequences related to cannabis use.

Periodic substance use reports also included measures that asked participants to describe how they felt after consuming alcohol or cannabis. For alcohol use, the options provided were *buzzed*, *tipsy/happy*, *drunk*, and *wasted*. Most reports of alcohol use described participants feeling *buzzed* (12/28, 43%), followed by feeling *tipsy/happy* (9/28, 32%) and feeling *drunk* (7/28, 25%). For cannabis use, the options provided were *calm/chill*, *relaxed*, *high*, and *stoned*. Most reports of cannabis use described participants feeling *calm/chill* (3/7, 43%) or *high* (2/13, 29%). Cannabis use reports also included the manner in which the substance was consumed. A majority of responses reported cannabis use through pipes (7/13, 54%) or vapes (5/13, 38%).

Participants were also asked about various aspects of their health daily. In session 1, participants were asked about prior-day stress levels and sleep duration. In all sessions, participants were asked about their mood since the last response.

Of the 149 self-reports received for session 1, a total of 148 (99.3%) self-reports contained responses related to stress. In 96 (64.9%) of these 148 self-reports, participants reported that stressful events did not occur. When asked to rate their prior-day stress levels on a scale ranging from 1 to 100, on average, participants reported a stress level of 33.920 (SD 22.181; range 0-90). With respect to sleep, the app collected 147 self-reports, where participants were asked when they went to sleep the prior day and when they woke up on the current day. On average, participants reported 7.290 (SD 1.859; range 0-11.167) hours of sleep. Finally, participants were asked to report their mood through 8 bipolar items, which garnered 3167 self-reports.

Overall, the data collected through the app consisted of a broad range of substance use behaviors and experiences, as well as a variety of health behaviors. This suggests that *participants are willing and able to share substance use data through smartwatches*, along with a variety of measures that have historically been associated with substance use in young adult populations.

App Use

We first examined how regularly participants used the app to answer EMAs. Of the 15 participants, 6 (40%) responded to at least 1 prompt on all 14 days of the study. Most participants (11/15, 73%) responded on ≥10 days. On average, participants provided data on 11.73 (SD 2.60) days out of the 14 days of the study. For participants completing all activities of the study, the average number of days participated was even higher: 12.24 (SD 2.14). Table 1 lists EMA completion details across each participant.

We had 403 sessions with ≥1 EMA response. On average, participants provided data for 26.80 (SD 12.15) sessions. We instructed participants to complete the first session every day along with at least 1 other session. Using these criteria, the overall compliance rate was 59% (8.2/14). On average, participants were compliant for 8.20 (SD 3.67) days out of the 14 days of the data collection period.

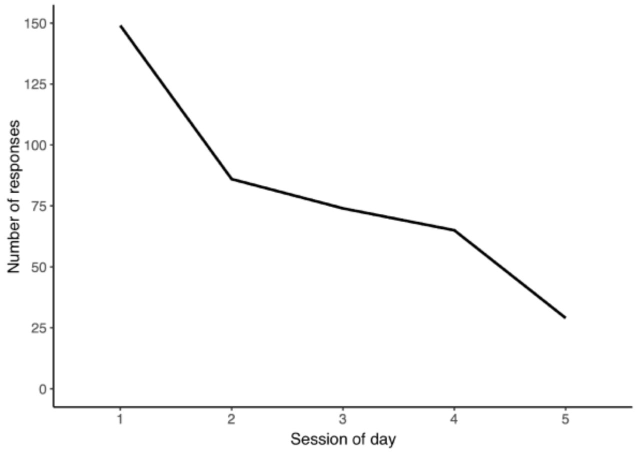
Finally, we also examined consecutive app use—the longest consecutive *streak* of days where participants used the app to provide responses. The longest streak was 14 days: 6 (40%) of the 15 participants used the app every day during the study. The average streak across all participants was 10.00 (SD 3.96) days, indicating sustained engagement with the app for a majority of the study duration.

Contextual Variations in App Use

In this part of our analysis, we wanted to determine whether there were certain times and contexts in which participants were less likely to respond to prompts than others. Toward this effort, we explored how app use patterns varied with time, substance use, and social environments.

In our data set, the response rate varied across sessions (Figure 2). Session 1 (11 AM-3 PM) had the most responses, and app use fell as the day progressed, with the lowest responses being collected during session 5 (1 AM-3 AM). To understand whether the session of day had a significant effect on whether the participant would respond, we used multilevel modeling (using the *lme4* package in R).

Figure 2. Variations in responses by session.



A null model allowed us to calculate the intraclass correlation coefficient (ICC) of whether a participant responded. The ICC was 0.135, which meant that only 13.5% of the variation in responding stemmed from between-person differences, which indicated that a large proportion of the variation arose due to within-person changes. Thus, a random intercept model was created by adding the session of day as a predictor. This model significantly explains more of the variance in participants' responses than the null model and hence is a better fit to the data (χ^2_4=181.6; *P*<.001). Using this model, we found that the session of day had a highly significant effect on whether the participant would respond. The odds and odds ratios calculated using this model indicated that the probability of a participant responding in session 1 was approximately 0.73. Compared to session 1, sessions 2, 3, 4, and 5 were respectively associated with a 76.47%, 82.29%, 85.88%, and 95.68% *decrease* in odds of a participant responding. In other words, *the probability of a participant answering in a particular session decreased significantly across the day*. Model details are described in Table S3 in Multimedia Appendix 1.

A random slope model did not significantly improve the fit of the model (χ^2_{14}=22.9; *P*=.06) and thus was not included for further analysis.

We used a similar method to understand whether the likelihood of reporting substance use varied across the day. The results of our null model calculated an ICC of 0.515, indicating that 51.5% of the variation in reporting substance use stemmed from between-person variances. The results from our model revealed that only session 5 had a significantly higher probability of participants reporting substance use compared to session 1 (estimate=1.70, SE 0.67; *P*=.01). The odds ratio for session 5 indicated that the odds of a participant reporting substance use in session 5 were approximately 5.49 times higher than the odds of a participant reporting substance use in session 1. This model proved to be a significantly better fit to the data than the null model (χ^2_4=11.3; *P*=.02), that is, participants were more likely to report substance use later in the day. Model details are described in Table S4 in Multimedia Appendix 1.

We also explored whether participants were able and likely to respond even when under the influence of substances.

Specifically, we analyzed how participants' responses differed after they reported substance use (compared with reports with no substance use). For this analysis, we used repeated measures correlations to determine within-individual association for paired or repeated measures data using the *rmcorr* package in R. We found no significant moment-level associations of substance use reports with responses in subsequent sessions, that is, whether a participant reported substance use in a specific session had no significant impact on their response to the first (r=0.02, 95% CI −0.08 to 0.13; P=.65), second (r=0.01, 95% CI −0.09 to 0.11; P=.83), third (r=0.03, 95% CI −0.07 to 0.13; P=.56), or fourth (r=−0.03, 95% CI −0.14 to 0.06; P=.46) session after the reported substance use. Random intercept multilevel models confirmed this result: reporting substance use in a specific session was not a significant predictor of whether a participant responded to the first, second, third, or fourth sessions after the session in question. Similarly, we saw no significant associations between social environments (people and places) and participants' likelihood of responding.

To summarize, our findings suggested that *participants were likely to respond to EMA prompts in a variety of social contexts and after consuming substances*. However, we found a time effect, where participants were more likely to respond to prompts earlier in the day.

Differences in Use Patterns Between Design Phases

To investigate whether the improvements made to the smartwatch app had any effect, we compared 5 metrics between phase1 and phase 2 participants with respect to the total number of EMAs answered, the total number of sessions completed, the total number of days participated, the total number of days compliant, and the longest consecutive use of the app. Before running the analysis, we used the Shapiro-Wilk test to check whether the metric values were distributed normally across the phases. The distributions of counts from participants in phase 2 for the total number of days participated (W=0.77; P=.03), the total number of EMAs answered (W=0.74; P=.02), and the longest consecutive use of the app (W=0.77; P=.02) were all significantly nonnormal. Thus, to test for differences between the phases for these 3 variables, we used a nonparametric test, the Wilcoxon rank sum test. For the remaining variables, we used the independent 2-tailed t test (the Welch 2-sample t test) to examine whether the differences were significant.

Our analysis showed that, on average, the total number of days participated among phase 1 participants (mean 10.88, SD 2.95) was lower than that among phase 2 participants (mean 12.71, SD 1.89); however, this difference was not significant (W=17; P=.21; r=−0.32). Similarly, the group means were higher for phase 2 participants compared to those for phase 1 participants in terms of the total number of EMAs answered (phase 1: mean 304.38, SD 155.84; phase 2: mean 337.29, SD 154.78; W=21; P=.86; r=−0.05), the total number of sessions completed (phase 1: mean 25.50, SD 14.55; phase 2: mean 28.29, SD 9.64; t_{12}=−0.44; P=.67), the total number of days compliant (phase 1: mean 7.38, SD 3.93; phase 2: mean 9.14, SD 3.39; t_{13}=−0.94; P=.37), and the longest consecutive use of the app (phase 1: mean 8.88, SD 3.94; phase 2: mean 11.29, SD 3.86; W=17; P=.21; r=−0.33), but none of the differences between the phases were significant.

Although we did not see statistically significant increases in app use metrics after improving the app, the systematically higher engagement in terms of days used, EMAs answered, days compliant, and longest consecutive use suggests that the changes were a step in the right direction.

User Evaluation

For our analysis of the user experience survey deployed after the participants finished their 14-day data collection period, we primarily focused on reporting various measures of usability, describing notable user perceptions, and comparing usability metrics between phase 1 and phase 2 participants to evaluate the effect of app improvements on overall user experience.

SUS Scores

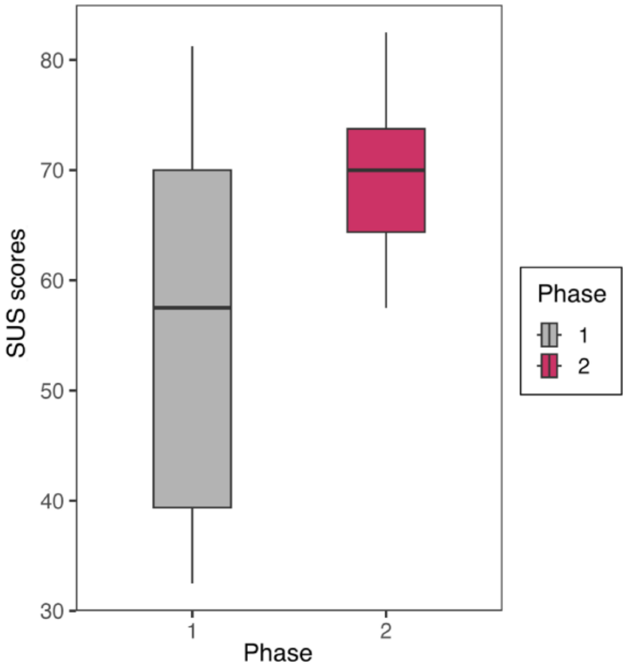
Of the 15 participants who used the app, 12 (80%) completed all research activities. The average SUS score for all 12 participants was 63.54 (SD 18.78). As a comparison point, Bangor et al [51] found that the mean SUS score from 964 usability tests across various interface types was 70. However, a usability study of fitness trackers found that the average SUS score for an Apple Watch interface was 61.36 [52]. While slightly higher than average in terms of smartwatch interface, this score does provide the opportunity to understand pain points within the app. The mean score for each SUS measure is depicted in Table 2.

Table 2. Summary of itemized System Usability Scale (SUS) scores presented overall (combining the results of participants from both phases) and by study phase. Notably, phase 2 participants reported higher mean SUS scores than phase 1 participants, but this difference was not significant.

Measures	SUS items, mean (SD)										Total SUS score, mean (SD)
	1	2	3	4	5	6	7	8	9	10	
Overall	2.83 (1.11)	2.5 (1.17)	3 (1.28)	1.83 (0.94)	3.33 (1.30)	2.67 (1.44)	4.33 (0.49)	3 (1.35)	3.67 (1.07)	1.75 (0.96)	63.54 (18.78)
Phase 1	2.17 (1.17)	3 (1.26)	2.83 (0.98)	1.5 (0.84)	3.33 (1.63)	3.33 (1.63)	4.17 (0.40)	3.33 (1.63)	3.33 (1.21)	1.83 (1.16)	57.08 (22.77)
Phase 2	3.5 (0.55)	2 (0.89)	3.17 (1.60)	2.17 (0.98)	3.33 (1.03)	2 (0.89)	4.5 (0.55)	2.67 (1.03)	4 (0.89)	1.67 (0.82)	70 (12.55)

Before comparing overall mean SUS scores between the phases, we first used the Shapiro-Wilk test to check whether the SUS scores were distributed normally across the phases. The results of this test and an examination of skew and kurtosis values indicated that the SUS scores were distributed normally overall and by phase. Thus, to compare the means, we used the independent *t* test (the Welch 2-sample *t* test). *Although participants in phase 2 reported higher mean SUS scores than those in phase 1, this difference was not statistically significant* (phase 1: mean 57.08, SD 22.77; phase 2: mean 70.0, SD 12.55; t_8=−1.21; *P*=.26). A box plot depicting differences in the SUS scores between the phases is shown in Figure 3.

Figure 3. Box plot of System Usability Scale (SUS) scores by phase of study. Mean SUS scores were higher in phase 2, after we had made changes to the app to correct for delay issues. However, the difference in SUS scores between the phases was not statistically significant.



uMARS Scores

All uMARS items were rated on a scale ranging from 1 to 5. Both the average functionality and aesthetics scores for all 12 participants were 3.88 (SD 0.55) and 3.89 (SD 0.52) respectively, indicating that participants found both measures acceptable (mean ≥3).

In terms of overall mean scores, participants in phase 2 rated app functionality as *good* (mean ≥4), while those in phase 1 rated app functionality as *acceptable*. Participants from both phases rated aesthetics as *acceptable*. Mean functionality scores were higher in phase 2 (phase 1: mean functionality score=3.58, SD 0.61; phase 2: mean functionality score=4.17, SD 0.31; t_7=−2.11; *P*=.07), but mean aesthetics scores were slightly higher in phase 1 (phase 1: mean aesthetics score=3.94, SD 0.71; phase 2: mean aesthetics score=3.83, SD 0.28; t_7=−2.11; *P*=.07).

Notably, the scores for the performance domain of the functionality metric from phase 2 participants exceeded those from phase 1 participants (phase 1: mean performance score=2.33, SD 0.52; phase 2: mean performance score=4.17, SD 0.75; difference in scores between the 2 phases=1.84). Using the Wilcoxon rank sum test, *we found that phase 2 participants rated the app significantly higher on the performance scale than phase 1 participants* (W=1; *P*=.006; *r*=−0.79). There were no significant differences between the phases for the ease of use, navigation, and gestural design domains of uMARS functionality scores. Similarly, none of the aesthetics domains (layout, graphics, and visual appeal) had any significant differences in ratings between the phases.

EMA-Specific Participant Perceptions

Along with ratings of established usability scales, we also asked participants specific survey questions about EMA usability, touching upon the constructs of the ease of use, enjoyment, the speed of answering EMAs, EMA length, interruptibility, and notions of trust and privacy.

Most of the participants (8/12, 67%) either agreed or strongly agreed that the EMAs were easy to fill. However, there were discrepancies between the phases. Most participants in phase 2 (5/6, 83%) agreed or strongly agreed that the EMAs were easy to fill, while 3 (50%) of the 6 participants in phase 1 either disagreed or strongly disagreed with this statement. Similarly, all participants in phase 2 (6/6, 100%) agreed or strongly agreed that they were able to complete the EMAs quickly, while 4 (67%) of the 6 participants in phase 1 disagreed with this statement. Of the 12 participants, 8 (67%) did not think that the EMAs were too long, and among the 4 participants who did think so, a majority (n=3, 75%) were phase 1 participants.

To examine perceived burden and fatigue, participants were asked about the number of days after which they felt tired of answering the EMAs. On average, participants in phase 1 reported fatigue after 6.40 (SD 3.44; range 3-12) days, while participants in phase 2 reported fatigue after 9.17 (SD 3.13;

range 4-12) days. This difference among the phases was not statistically significant (t_8=−1.39; *P*=.20).

Overall, most of the participants felt that the app was acceptable and simple to use. A higher proportion of participants in phase 2 felt so compared to those from phase 1. None of the participant ratings of ease, speed, or fatigue were significantly different across the phases, but the higher ratings in phase 2 suggest that the app changes were a step in the right direction to improve app usability and address critical issues.

Interview Themes

Overview

In this subsection, we present our main findings from our analysis of the semistructured interviews that were conducted after the 14-day data collection period. For the analysis of the interview data, we used inductive thematic analysis to identify common themes using a qualitative interpretivist approach. The primary author conducted the initial analysis and then discussed the themes and codes with the other authors to ensure the validity of the primary findings and to reduce bias.

Overall, participants agreed that their experience using the smartwatch app to answer EMAs was easy, novel, and acceptable, but they also brought up certain key issues with app responsiveness and commented on the suitability of the smartwatch interface for this specific use case.

General App Perceptions

When asked about their overall experience using the app to answer EMAs, most of the participants shared that the app was generally easy to use. A participant recalled that using the app was quickly incorporated into their day, while noting that this was not disruptive to their routine:

> *I mean it kind of turned into, like, an everyday routine where, like I just expected it at certain times and I used to take time out and do it.* [P6]

This sentiment was echoed by another participant:

> *Since it is only like 3 to 5 minutes, I didn't think that's a very disruptive time point, like I could do it in between class or, if I was at dinner [or] lunch and I remembered, I'd typically do it then.* [P4]

Although not disruptive to their daily lives, this participant shared that using the app was different to how they normally used their smartwatch:

> *Disruptive? Not really. I don't normally look at my watch for more than a couple of seconds, so that was a little different, but overall it wasn't really that disruptive.* [P4]

Other participants also noted how completing the EMAs only took a few minutes (generally <2-3 min), unless lagging or responsiveness issues occurred.

Advantages and Challenges of Using Smartwatch Interfaces

Participants presented varied perspectives when it came to the elements of the smartwatch interface; for instance, some of the participants found advantages and preferred the fact that the smartwatch provided a small screen and a personal experience:

> *I think that's the one benefit that the watch did have, is that it's such a small screen that it's hard for anyone to, you know, look at what you're doing, on such a small screen. So the watch definitely had a benefit in kind of, like, protecting your privacy.* [P8]

By contrast, participants also noted that the small screen and the wearable experience presented a hindrance, with a participant sharing the challenges they faced in using the watch to answer EMAs:

> *Well, so for me, just having, just turning my arm and touching my watch is, I don't know if it's a range of motion thing, It's just not the most natural thing to me and so just having to be in this position, looking at my watch, touching stuff, I don't particularly like that.* [P3]

This participant indicated that a bigger screen would provide a more comfortable experience:

> *Just having a larger screen to be able to do everything on, I think it'd be a lot easier.* [P3]

Personal preferences factored greatly into how easy and intuitive participants found various aspects of the app experience. When asked about their perspectives on the various formats in which the questions were presented, such as sliders, checkboxes, and radio buttons, the responses were similarly varied. Some of the participants found all question formats easy to answer:

> *I think all of the formats were very straightforward and in terms of them, like, how they worked, I think they all worked just fine. There was no issue transitioning between the different formats.* [P8]

Others reported issues with the radio button and checkbox formats:

> *I think the multiple select got harder because just, like, being able to see all the options and then be able to click next on an Apple Watch screen [that] is kind of tiny, so in that sense, yes [was a difficult format to answer].* [P7]

Similarly, a participant faced challenges with the slider format:

> *Think the [slider] one, because I think I had to press, if I'm like, you know, perfectly energetic [on the MDMQ] then I had to go all the way plus plus plus plus plus, it was like, a lot of plusses. Other than that, the rest was great.* [P12]

App Responsiveness and Lagging Issues

Phase 1 participants frequently shared their experiences with recurring lag issues, noting that it lengthened app use time and caused disruptions and general frustration:

> *Overall, it was pretty easy and straightforward, but it did start to get frustrating switching between different survey prompts. It would get, like, frozen a lot. So I would click to go to the next prompt, I guess, and it would get frozen, so surveys that were supposed*

> *to take 2 minutes ended up taking upwards of 10 minutes because it would get frozen.* [P8]

> *A lot of the buttons weren't the most responsive, so you had to click them a couple of times before it would actually do anything. And sometimes I had to restart my watch because it just wouldn't have responded.* [P3]

As mentioned in the *Iterative Application Design: Phase 1 and Phase 2* subsubsection in the *Methods* section, we identified that this delay was caused by the data-uploading mechanism, which was corrected for phase 2. As a result, phase 2 participants did not report this frequent lag between questions in their interviews.

Comparisons to iPhone Platform

Several participants believed that having the option to answer the EMAs on both the iPhone and the Apple Watch would offer an easier and more seamless experience and provided the strengths of both devices to support this sentiment. A participant offered some context where such a system would prove useful for them:

> *So I know in the evening, sometimes, especially when I'm just, like, sitting on my couch, laying down, watching TV, I'll take my watch off to charge for the night, but I'll still have my phone with me. So, I'm not gonna get those alerts, if I'm not wearing my watch. So, it's nice to be able to switch, then, to the different interface on my phone, to use that.* [P14]

Another participant shared a similar perspective:

> *I guess, that [having the option to complete surveys on both devices] would be okay, because that way you can at least see the surveys, do on your watch, in case your phone is not in your hand, you still have the watch, you're wearing your watch, you have the option of both.* [P11]

By contrast, a participant shared a scenario where using a smartwatch would prove easier than using a smartphone:

> *Usually you have to open up the phone and then you have to take off your mask [to unlock it using facial recognition]. With the watch, you don't have to do anything, you just, you know, do with the 1 finger, which makes it a lot easier and better.* [P12]

Similarly, another participant noted as follows:

> *I check my watch more than I check my phone. I feel just time wise, and yeah, I feel like it'd be harder to use the phone. Like take my phone out and use it.* [P14]

However, most of the participants agreed that the larger screen size would provide a smoother experience while answering EMAs, with a participant sharing their perspective on how having a bigger screen would benefit their experience:

> *Just cause it's a little bigger, and you can just, like, do it on your phone while walking or something, and like on your watch you can't really do that.* [P10]

Self-Monitoring Substance Use

Several participants shared how using the app provided a valuable self-tracking experience that helped them think about their substance use patterns. Although this was not an intended use of the app, participants found a tangible benefit in keeping track of their substance use to answer the EMAs accurately:

> *It makes you cognizant of your usage, and it makes you cognizant, while looking at the questions, as to what, you know, could be impacted [by substance use].* [P11]

A participant shared how answering the EMAs helped them evaluate their substance use:

> *I think it just forced me to kind of analyze...like, I'd mainly only drink on the weekends, so it made me [think about] how I spend my weekends and how much I was using a substance in a specific time frame. So it made you kind of take a step back and analyze that, which is always, I think, shocking to people, how much or how little they may have been using a substance.* [P8]

Answering frequent EMAs about their substance use helped participants increase their awareness of their substance use patterns and behaviors.

A participant also shared an interesting perspective of how useful they found the self-monitoring aspect of using the app and how they experienced a lack of incentive to track their substance use after the 14-day data collection period ended:

> *I think, just being aware of, like, how many drinks I was consuming. Yeah, because if I don't have to track it, I don't remember how much I drink. So, because I was able to be like oh, like the next window is at 7 o'clock, like, my next notification at 7, like that. I've had to remember that I've had, you know, 2 drinks to put it in that notification.* [P14]

Other participants noted how they already mentally keep track of their substance use, but using the app made them reevaluate their use:

> *It definitely increased my awareness, but I felt like I already knew. If I had work or most of the school days, like, I won't be doing anything like that [substance use], but, more on the weekends. Like, oh, maybe I shouldn't do this tonight, or something like that.* [P7]

Participants used the EMAs to reflectively track their substance use. These interactions augmented their existing self-monitoring practices to periodically and contextually evaluate their substance use behaviors.

Use of the App in Social Settings

All participants reported that they used the app in public and social settings and were comfortable doing so; for instance, a participant shared how their friends felt when they saw the participant using the app:

> *Yeah, like I thought it was totally fine. All my friends knew I was taking, [and] like I didn't care that they*

> *knew. But when I was out at the bars, I was fine taking the surveys, and I don't know if other people knew that I was using my watch, or whatever. But all my friends knew, and they thought it was cool.* [P14]

Another participant also spoke about their use of the app in such settings:

> *Yeah, like, if I got the notification when I was at school, like in class or something, or like walking to class, I would take it then.* [P7]

This indicates that the app is able to effectively collect data in various social settings. This finding is especially meaningful, given the sensitive and often-stigmatized nature of the substance use data that the app collects. The convenience and comfort with which participants are able to share information indicates that using the smartwatch in this way is potentially unintrusive in various social contexts and environments.

Several participants offered insight into how they did not have concerns regarding privacy or security while interacting with the app and shared how the smartwatch platform helped in this aspect:

> *No [I did not feel uncomfortable using the app in public or social settings]. I mean, the watch screen is so small, I don't even think anyone realized what I was doing, that I'm on it.* [P7]

The small screen of the watch ensured that the participants' activities while using the app remained private from their peers and other people in their vicinity and thus helped their perception of the security of their data.

Discussion

App Feasibility and Acceptability

Overall, the app collected 4796 responses to EMA questions from 15 participants over the course of a 2-week-long study. Participants demonstrated high and consistent use of the app, responding on an average of 11.73 (SD 2.60) days and consistently using the app for an average of 10 (SD 3.96) days. Our analysis of app use patterns indicates that participants respond in a variety of contexts: after they consume substances and among different social contexts. The interview data supported these findings: participants were able to quickly incorporate using the app into their daily life and easily provide substance use data, and they were comfortable using the app in diverse social settings. Together, these findings demonstrate that it is indeed feasible to use a smartwatch app to collect substance use data.

With respect to app use, the decrease in participants' responses across the day was an interesting finding. We speculate that the higher response rate in session 1 might be due to the longer availability compared to other sessions (4 h vs 2 h). Participants were also specifically asked to complete session 1 each day and were compensated accordingly. However, our findings also indicate that participants were more likely to report substance use at night, in session 5 (1 AM-3 AM), than in session 1 (11 AM-3 PM), which coincides with substance use patterns among young adults. Together, these results suggest that there are certain time periods that may be better suited to obtaining specific insights into substance use behaviors. Morning and noon may be suitable periods to understand prior-day substance use behaviors, mood, and experiences, while late night might be better suited to understand evening drinking behaviors. As such, there is an opportunity to develop better informed and less burdensome methods for collecting substance use data. Future work should try to replicate our findings regarding the temporal variation of EMA completion rates for substance use.

In terms of user evaluations, the average SUS score for the 12 participants who completed the survey was 63.54 (SD 18.78). Participants in phase 2 reported higher mean SUS scores than those in phase 1 (phase 1: mean 57.08, phase 2: mean 70.00). For context, an SUS score of 70.00 is considered average and acceptable, but it is to be noted that this subjective qualification of SUS scores does not consider smartwatch interfaces. If we factor the interface into our assessment of participants' SUS scores, we can estimate that overall and in phase 2, participants rated the app above average in usability. Furthermore, in terms of mean uMARS scores, participants from phase 2 rated app functionality as good (mean ≥4), while those from phase 1 rated app functionality as acceptable. Although not significant, these findings suggest that the performance improvements to the app had a large and measurable impact on participants' perceptions of usability. Indeed, the improvement also had an impact on app use: on average, the total number of days participated, the total number of EMAs answered, the total number of sessions completed, the total number of days compliant, and the longest consecutive use of the app were all higher among phase 2 participants than among phase 1 participants.

These findings not only establish the user acceptability of smartwatches to collect substance use data but also indicate that app performance, specifically responsiveness to user inputs, is critical for user acceptance. Given the limited computational capability of smartwatches, it is particularly important to aim for responsive design by default. Modifications to improve app responsivity resulted in better perceived usability and user satisfaction, along with systematically higher user evaluation scores and app use metrics. Thus, supporting quick, responsive interactions is a critical consideration when designing EMAs for smartwatches. Researchers and practitioners interested in using these devices as platforms for intensive data collection must focus on efficient, quick, and simple interactions to ensure sustained use as well as acceptable compliance and response rates.

Smartwatches and Substance Use

Our data consisted of 45 prior-day and 67 periodic substance use reports which contain alcohol, cannabis, cigarette/cigar/cigarillo, and e-cigarette/vape use data. Participants were able to share data on a range of variables associated with substance use through the smartwatch.

Furthermore, our interview data highlighted a key benefit that participants found through regularly using the app: tracking and reflecting on their substance use. Using the app to provide substance use data encouraged participants to contemplate on their substance use by requiring them to recollect aspects of their use (when, how much, with whom, etc). Even without a feature that displays the patterns of use, participants noted how the task of recollecting and entering substance use data helped to make them more aware of patterns within their substance use as well as cognizant of the contexts in which they consume various substances. While the benefits of self-monitoring substance use are not limited to the smartwatch interface, it is promising that a smartwatch app is able to successfully promote such experiences.

An aspect of the smartwatch interface that might have helped participants share substance use data confidently is the privacy that it affords through a smaller, more discreet screen. Participants reported how they felt comfortable using the app in social and public settings, saying that the small screen ensured that others in their vicinity would not be able to discern what the participants were doing on their smartwatch. Nevertheless, some participants thought that a larger screen, such as a smartphone screen, might be useful in certain contexts. Participants also noted that some question formats, such as those that require scrolling, are harder to complete on a small screen. Importantly, participants preferred having the option to complete a survey on a smartphone or a smartwatch, depending on what is most convenient at a given time and place. Future studies using smartwatches for health assessments and interventions should ensure that the proposed systems can work comfortably across diverse contexts. One way to accomplish this is by supporting interchangeable use of the app on different devices: smartphones *as well as* smartwatches. Users can then choose which device is most appropriate for their current activity and social environment and use the app correspondingly.

On the whole, our analysis of app use, surveys, and interview data indicate the feasibility and acceptability of using smartwatches in this domain, demonstrating that users are able and willing to use a smartwatch to share substance use data. Participants shared data on a range of substances, experiences, and behaviors and identified aspects of the smartwatch interface that enabled them to do so comfortably. Participants found that comprehensively self-monitoring their substance use through the app was a useful and important feature. Our findings also provide insight into which aspects of the smartwatch interface elicit responses as well as those that do not: while the small screen affords users privacy while relaying sensitive information such as substance use data, it can provide challenges for certain EMA formats and in certain contexts.

Limitations

This study has a number of limitations that are important to discuss, given their potential impact on our findings. First, the study had a small sample size, which we considered to be acceptable, given that the goal of the study was to establish the feasibility and acceptability of a smartwatch-based app for collecting substance use data. However, we acknowledge that the reported findings may not be generalizable to the larger population of young adults who consume substances. Reproducing this study with a larger, more diverse sample can offer a wider perspective on the use of smartwatch-based apps to collect longitudinal, intensive data in this domain. The findings concerning significance should be interpreted with caution, given our small sample size and unstable estimates. Furthermore, participants were already smartwatch owners, which might have had an impact on the perceptions of usability. Thus, understanding the perspectives of novice smartwatch users using the app can help us investigate the effect of novelty on user experience. Finally, most participants in our sample were not binge drinkers or did not exhibit high-intensity substance use. The user experience of the app might differ with participants and circumstances that arise from heavy or hazardous substance use behaviors that are not adequately represented in our sample. Future work focusing on users who exhibit such patterns of substance use can help build more robust systems that cater to a wider range of people who use substances.

Next, we detail limitations associated with our app. Developing data collection apps on the Apple system has the constraint of being platform dependent (limiting the devices on which the apps can be deployed); however, developing for a single ecosystem was the first step in testing the general feasibility of a smartwatch-based data collection app. Implementing the data collection app on multiple ecosystems and running studies with various devices and apps was outside the scope of this study. However, our design and development process focused considerably on creating a reproducible and well-documented codebase so that cross-platform or platform-agnostic implementation can be achieved at a later stage.

Conclusions and Future Work

In recent years, there has been wide adoption of smartwatches for health assessments and interventions. This paper focuses on ascertaining the feasibility and acceptability of using a smartwatch app to collect substance use data from young adults. Our data indicate that it is feasible and acceptable to use smartwatches to collect data about sensitive and stigmatized behaviors, including substance use. On the basis of these findings, we also discuss considerations for future smartwatch apps for health and well-being data collection. These findings have important implications for researchers aiming to leverage smartwatches as an mHealth platform for effective assessments and interventions. In the future, we plan to conduct a larger study, with a randomized between-participants experiment design, to compare app use and user perceptions between smartphones and smartwatches. This future study will help us understand which device results in better compliance, better engagement, and lower perceived burden within the context of substance use data collection. We also intend to use the health sensor data from this larger study to explore whether they can be used to unobtrusively detect substance use or associated behaviors. Finally, we aim to incorporate analyses such as the impact of battery life on app use to gain a nuanced understanding of how smartwatch capabilities impact user experience.

Acknowledgments

The research reported in this paper was supported by the National Institute on Alcohol Abuse and Alcoholism of the National Institutes of Health (K01 AA026854) and a seed grant from the College of Information Sciences and Technology at The Pennsylvania State University.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study materials, participant demographics, and multilevel model details.
[DOCX File , 27 KB - humanfactors_v11i1e50795_app1.docx]

References

1. Substance Abuse and Mental Health Services Administration. Key substance use and mental health indicators in the United States: results from the 2022 national survey on drug use and health. Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. 2023. URL: https://www.samhsa.gov/data/report/2022-nsduh-annual-national-report [accessed 2024-04-09]
2. Substance misuse prevention for young adults. Substance Abuse and Mental Health Services Administration. 2019. URL: https://store.samhsa.gov/product/substance-misuse-prevention-young-adults/pep19-pl-guide-1 [accessed 2023-05-27]
3. Braitman AL, Linden-Carmichael AN, Henson JM. Protective behavioral strategies as a context-specific mediator: a multilevel examination of within- and between-person associations of daily drinking. Exp Clin Psychopharmacol 2017 Jun;25(3):141-155 [FREE Full text] [doi: 10.1037/pha0000111] [Medline: 28240925]
4. Linden-Carmichael AN, Allen HK, Lanza ST. The socio-environmental context of simultaneous alcohol and marijuana use among young adults: examining day-level associations. Drug Alcohol Rev 2021 May 13;40(4):647-657 [FREE Full text] [doi: 10.1111/dar.13213] [Medline: 33188538]
5. Lipperman-Kreda S, Paschall MJ, Robert FS, Morrison CN. Places and social contexts associated with simultaneous use of alcohol, tobacco and marijuana among young adults. Drug Alcohol Rev 2018 Feb 19;37(2):188-195 [FREE Full text] [doi: 10.1111/dar.12537] [Medline: 28422352]
6. Terry-McElrath YM, O'Malley PM, Johnston LD. Alcohol and marijuana use patterns associated with unsafe driving among U.S. high school seniors: high use frequency, concurrent use, and simultaneous use. J Stud Alcohol Drugs 2014 May;75(3):378-389 [FREE Full text] [doi: 10.15288/jsad.2014.75.378] [Medline: 24766749]
7. Arnold Z, Larose D, Agu E. Smartphone inference of alcohol consumption levels from gait. In: Proceedings of the International Conference on Healthcare Informatics. 2015 Presented at: ICHI 2015; October 21-23, 2015; Dallas, TX. [doi: 10.1109/ichi.2015.59]
8. Kao HL, Ho BJ, Lin AC, Chu HH. Phone-based gait analysis to detect alcohol usage. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing. 2012 Presented at: Ubicomp '12; September 5-8, 2012; Pittsburgh, PA. [doi: 10.1145/2370216.2370354]
9. Suffoletto B, Gharani P, Chung T, Karimi H. Using phone sensors and an artificial neural network to detect gait changes during drinking episodes in the natural environment. Gait Posture 2018 Feb;60:116-121 [FREE Full text] [doi: 10.1016/j.gaitpost.2017.11.019] [Medline: 29179052]
10. Bae S, Ferreira D, Suffoletto B, Puyana JC, Kurtz R, Chung T, et al. Detecting drinking episodes in young adults using smartphone-based sensors. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017 Jun 30;1(2):1-36 [FREE Full text] [doi: 10.1145/3090051] [Medline: 35146236]
11. Bae S, Chung T, Ferreira D, Dey AK, Suffoletto B. Mobile phone sensors and supervised machine learning to identify alcohol use events in young adults: implications for just-in-time adaptive interventions. Addict Behav 2018 Aug;83:42-47 [FREE Full text] [doi: 10.1016/j.addbeh.2017.11.039] [Medline: 29217132]
12. Li R, Agu E, Balakrishnan G, Herman D, Abrantes A, Stein M, et al. WeedGait: unobtrusive smartphone sensing of marijuana-induced gait impairment by fusing gait cycle segmentation and neural networks. In: Proceedings of the IEEE Healthcare Innovations and Point of Care Technologies. 2019 Presented at: HI-POCT 2019; November 20-22, 2019; Bethesda, MD. [doi: 10.1109/hi-poct45284.2019.8962787]
13. Bae SW, Chung T, Islam R, Suffoletto B, Du J, Jang S, et al. Mobile phone sensor-based detection of subjective cannabis intoxication in young adults: a feasibility study in real-world settings. Drug Alcohol Depend 2021 Nov 01;228:108972 [FREE Full text] [doi: 10.1016/j.drugalcdep.2021.108972] [Medline: 34530315]
14. Jones A, Remmerswaal D, Verveer I, Robinson E, Franken IH, Wen CK, et al. Compliance with ecological momentary assessment protocols in substance users: a meta-analysis. Addiction 2019 Apr;114(4):609-619 [FREE Full text] [doi: 10.1111/add.14503] [Medline: 30461120]

15. Cecchinato ME, Cox AL, Bird J. Smartwatches: the good, the bad and the ugly? In: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems. 2015 Presented at: CHI EA '15; April 18-23, 2015; Seoul, Republic of Korea. [doi: 10.1145/2702613.2732837]
16. Visuri A, Sarsenbayeva Z, van Berkel N, Goncalves J, Rawassizadeh R, Kostakos V, et al. Quantifying sources and types of smartwatch usage sessions. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI '17; May 6-11, 2017; Denver, CO. [doi: 10.1145/3025453.3025817]
17. Number of users of smartwatches worldwide from 2019 to 2028. Statista. 2023 Nov. URL: https://www.statista.com/forecasts/1314339/worldwideusers-of-smartwatches [accessed 2024-04-09]
18. Vogels EA. About one-in-five Americans use a smart watch or fitness tracker. Pew Research Center. 2020 Jan 09. URL: https://www.pewresearch.org/short-reads/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker/ [accessed 2023-05-27]
19. Dibia V. FOQUS: a smartwatch application for individuals with ADHD and mental health challenges. In: Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility. 2016 Presented at: ASSETS '16; October 23-26, 2016; Reno, NV. [doi: 10.1145/2982142.2982207]
20. Larsen JE, Eskelund K, Christiansen TB. Active self-tracking of subjective experience with a one-button wearable: a case study in military PTSD. arXiv. Preprint posted online on March 9, 2017 [FREE Full text]
21. Zheng H, Motti VG. WeLi: a smartwatch application to assist students with intellectual and developmental disabilities. In: Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility. 2017 Presented at: ASSETS '17; October 20-November 1, 2017; Baltimore, MD. [doi: 10.1145/3132525.3134770]
22. Rouzaud Laborde C, Cenko E, Mardini MT, Nerella S, Kheirkhahan M, Ranka S, et al. Satisfaction, usability, and compliance with the use of smartwatches for ecological momentary assessment of knee osteoarthritis symptoms in older adults: usability study. JMIR Aging 2021 Jul 14;4(3):e24553 [FREE Full text] [doi: 10.2196/24553] [Medline: 34259638]
23. Schickler M, Pryss R, Reichert M, Heinzelmann M, Schobel J, Langguth B, et al. Using wearables in the context of chronic disorders: results of a pre-study. In: Proceedings of the IEEE 29th International Symposium on Computer-Based Medical Systems. 2016 Presented at: CBMS 2016; June 20-24, 2016; Belfast and Dublin, Ireland. [doi: 10.1109/cbms.2016.40]
24. Skinner AL, Stone CJ, Doughty H, Munafò MR. StopWatch: the preliminary evaluation of a smartwatch-based system for passive detection of cigarette smoking. Nicotine Tob Res 2019 Jan 04;21(2):257-261 [FREE Full text] [doi: 10.1093/ntr/nty008] [Medline: 29373720]
25. Dhingra LS, Aminorroaya A, Oikonomou EK, Nargesi AA, Wilson FP, Krumholz HM, et al. Use of wearable devices in individuals with or at risk for cardiovascular disease in the US, 2019 to 2020. JAMA Netw Open 2023 Jun 01;6(6):e2316634 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.16634] [Medline: 37285157]
26. DeMarco J. Nearly 70% of Americans would wear a fitness tracker/smartwatch for discounted health insurance. Value Penguin. 2022 Apr 18. URL: https://www.valuepenguin.com/fitness-tracker-smartwatch-health-survey [accessed 2023-05-27]
27. Carreiro S, Chai PR, Carey J, Lai J, Smelson D, Boyer EW. mHealth for the detection and intervention in adolescent and young adult substance use disorder. Curr Addict Rep 2018 Jun;5(2):110-119 [FREE Full text] [doi: 10.1007/s40429-018-0192-0] [Medline: 30148037]
28. Chapman BP, Lucey E, Boyer EW, Babu KM, Smelson D, Carreiro S. Perceptions on wearable sensor-based interventions for monitoring of opioid therapy: a qualitative study. Front Digit Health 2022 Oct 21;4:969642 [FREE Full text] [doi: 10.3389/fdgth.2022.969642] [Medline: 36339518]
29. Carreiro S, Chintha KK, Shrestha S, Chapman B, Smelson D, Indic P. Wearable sensor-based detection of stress and craving in patients during treatment for substance use disorder: a mixed methods pilot study. Drug Alcohol Depend 2020 Apr 01;209:107929 [FREE Full text] [doi: 10.1016/j.drugalcdep.2020.107929] [Medline: 32193048]
30. Holtyn AF, Bosworth E, Marsch LA, McLeman B, Meier A, Saunders EC, et al. Towards detecting cocaine use using smartwatches in the NIDA clinical trials network: design, rationale, and methodology. Contemp Clin Trials Commun 2019 Jun 04;15:100392 [FREE Full text] [doi: 10.1016/j.conctc.2019.100392] [Medline: 31245651]
31. Designing for watchOS. Apple Inc. URL: https://developer.apple.com/design/human-interface-guidelines/designing-for-watchos [accessed 2023-05-29]
32. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;Suppl:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
33. Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment. World Health Organization. 2016 Dec 12. URL: https://www.who.int/publications/i/item/9789241511766 [accessed 2024-06-10]
34. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
35. Wilhelm P, Schoebi D. Assessing mood in daily life. Eur J Psychol Assessment 2007 Nov 05;23(4). [doi: 10.1027/1015-5759.23.4.258]
36. Cloutier RM, Calhoun BH, Lanza ST, Linden-Carmichael AN. Assessing subjective cannabis effects in daily life with contemporary young adult language. Drug Alcohol Depend 2022 Jan 01;230:109205 [FREE Full text] [doi: 10.1016/j.drugalcdep.2021.109205] [Medline: 34890928]

37. Linden-Carmichael AN, Calhoun BH. Measuring subjective alcohol effects in daily life using contemporary young adult language. Exp Clin Psychopharmacol 2022 Apr;30(2):151-158 [FREE Full text] [doi: 10.1037/pha0000447] [Medline: 33591776]
38. Lee CM, Cronce JM, Baldwin SA, Fairlie AM, Atkins DC, Patrick ME, et al. Psychometric analysis and validity of the daily alcohol-related consequences and evaluations measure for young adults. Psychol Assess 2017 Mar;29(3):253-263 [FREE Full text] [doi: 10.1037/pas0000320] [Medline: 27196690]
39. Rose DN, Bond MJ. Identity, stress and substance abuse among young adults. J Subst Use 2009 Jul 12;13(4):268-282. [doi: 10.1080/14659890801912006]
40. Linden-Carmichael AN, van Doren N, Bray BC, Jackson KM, Lanza ST. Stress and affect as daily risk factors for substance use patterns: an application of latent class analysis for daily diary data. Prev Sci 2022 May;23(4):598-607 [FREE Full text] [doi: 10.1007/s11121-021-01305-9] [Medline: 34716891]
41. Terry-McElrath YM, O'Malley PM, Johnston LD. Simultaneous alcohol and marijuana use among U.S. high school seniors from 1976 to 2011: trends, reasons, and situations. Drug Alcohol Depend 2013 Nov 01;133(1):71-79 [FREE Full text] [doi: 10.1016/j.drugalcdep.2013.05.031] [Medline: 23806871]
42. Gunn RL, Aston ER, Sokolovsky AW, White HR, Jackson KM. Complex cannabis use patterns: associations with cannabis consequences and cannabis use disorder symptomatology. Addict Behav 2020 Jun;105:106329 [FREE Full text] [doi: 10.1016/j.addbeh.2020.106329] [Medline: 32044680]
43. Pasch KE, Latimer LA, Cance JD, Moe SG, Lytle LA. Longitudinal bi-directional relationships between sleep and youth substance use. J Youth Adolesc 2012 Sep;41(9):1184-1196 [FREE Full text] [doi: 10.1007/s10964-012-9784-5] [Medline: 22752878]
44. Prince MA, Collins RL, Wilson SD, Vincent PC. A preliminary test of a brief intervention to lessen young adults' cannabis use: episode-level smartphone data highlights the role of protective behavioral strategies and exercise. Exp Clin Psychopharmacol 2020 Apr;28(2):150-156 [FREE Full text] [doi: 10.1037/pha0000301] [Medline: 31144836]
45. Wang D, Wang Y, Wang Y, Li R, Zhou C. Impact of physical exercise on substance use disorders: a meta-analysis. PLoS One 2014 Oct 16;9(10):e110728 [FREE Full text] [doi: 10.1371/journal.pone.0110728] [Medline: 25330437]
46. Kunchay S, Abdullah S. WatchOver: using Apple watches to assess and predict substance co-use in young adults. In: Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers. 2020 Presented at: UbiComp/ISWC '20 Adjunct; September 12-17, 2020; Virtual Event, Mexico. [doi: 10.1145/3410530.3414373]
47. Brooke J. SUS: a 'quick and dirty' usability scale. In: Usability Evaluation In Industry. Boca Raton, FL: CRC Press; 1996.
48. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
49. Yui BH, Jim WT, Chen M, Hsu JM, Liu CY, Lee TT. Evaluation of computerized physician order entry system-a satisfaction survey in Taiwan. J Med Syst 2012 Dec;36(6):3817-3824. [doi: 10.1007/s10916-012-9854-y] [Medline: 22549624]
50. Wright C, Dietze PM, Agius PA, Kuntsche E, Livingston M, Black OC, et al. Mobile phone-based ecological momentary intervention to reduce young adults' alcohol use in the event: a three-armed randomized controlled trial. JMIR Mhealth Uhealth 2018 Jul 20;6(7):e149 [FREE Full text] [doi: 10.2196/mhealth.9324] [Medline: 30030211]
51. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009 May 01;4(3):114-123.
52. Liang J, Xian D, Liu X, Fu J, Zhang X, Tang B, et al. Usability study of mainstream wearable fitness devices: feature analysis and system usability scale evaluation. JMIR Mhealth Uhealth 2018 Nov 08;6(11):e11066 [FREE Full text] [doi: 10.2196/11066] [Medline: 30409767]

Abbreviations

EMA: ecological momentary assessment
ICC: intraclass correlation coefficient
mHealth: mobile health
REDCap: Research Electronic Data Capture
SUS: System Usability Scale
uMARS: Mobile Application Rating Scale: User Version

Edited by D Gooch; submitted 20.07.23; peer-reviewed by T Fazzino, C Baxter; comments to author 09.02.24; revised version received 14.03.24; accepted 08.04.24; published 20.06.24.

Please cite as:
Kunchay S, Linden-Carmichael AN, Abdullah S
Using a Smartwatch App to Understand Young Adult Substance Use: Mixed Methods Feasibility Study
JMIR Hum Factors 2024;11:e50795
URL: https://humanfactors.jmir.org/2024/1/e50795
doi:10.2196/50795
PMID:

Original Paper

Chatbot for Social Need Screening and Resource Sharing With Vulnerable Families: Iterative Design and Evaluation Study

Emre Sezgin[1], PhD; A Baki Kocaballi[2], PhD; Millie Dolce[1], PhD; Micah Skeens[1], PhD; Lisa Militello[3], PhD; Yungui Huang[1], PhD; Jack Stevens[1], PhD; Alex R Kemper[1], MD

[1]Nationwide Children's Hospital, Columbus, OH, United States

[2]Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, Sydney, Australia

[3]The Ohio State University, Columbus, OH, United States

Corresponding Author:
Emre Sezgin, PhD
Nationwide Children's Hospital
700 Children's dr
Columbus, OH, 43205
United States
Phone: 1 6147223179
Email: emre.sezgin@nationwidechildrens.org

Abstract

Background: Health outcomes are significantly influenced by unmet social needs. Although screening for social needs has become common in health care settings, there is often poor linkage to resources after needs are identified. The structural barriers (eg, staffing, time, and space) to helping address social needs could be overcome by a technology-based solution.

Objective: This study aims to present the design and evaluation of a chatbot, DAPHNE (Dialog-Based Assistant Platform for Healthcare and Needs Ecosystem), which screens for social needs and links patients and families to resources.

Methods: This research used a three-stage study approach: (1) an end-user survey to understand unmet needs and perception toward chatbots, (2) iterative design with interdisciplinary stakeholder groups, and (3) a feasibility and usability assessment. In study 1, a web-based survey was conducted with low-income US resident households (n=201). Following that, in study 2, web-based sessions were held with an interdisciplinary group of stakeholders (n=10) using thematic and content analysis to inform the chatbot's design and development. Finally, in study 3, the assessment on feasibility and usability was completed via a mix of a web-based survey and focus group interviews following scenario-based usability testing with community health workers (family advocates; n=4) and social workers (n=9). We reported descriptive statistics and chi-square test results for the household survey. Content analysis and thematic analysis were used to analyze qualitative data. Usability score was descriptively reported.

Results: Among the survey participants, employed and younger individuals reported a higher likelihood of using a chatbot to address social needs, in contrast to the oldest age group. Regarding designing the chatbot, the stakeholders emphasized the importance of provider-technology collaboration, inclusive conversational design, and user education. The participants found that the chatbot's capabilities met expectations and that the chatbot was easy to use (System Usability Scale score=72/100). However, there were common concerns about the accuracy of suggested resources, electronic health record integration, and trust with a chatbot.

Conclusions: Chatbots can provide personalized feedback for families to identify and meet social needs. Our study highlights the importance of user-centered iterative design and development of chatbots for social needs. Future research should examine the efficacy, cost-effectiveness, and scalability of chatbot interventions to address social needs.

(JMIR Hum Factors 2024;11:e57114) doi:10.2196/57114

KEYWORDS

social determinants of health; social needs; chatbot; conversational agent; primary care; digital health; iterative design; implementation; evaluation; usability; feasibility

Introduction

Background

Unmet social needs (eg, food insecurity, housing insecurity, transportation challenges, and economic instability) are strongly associated with poor health outcomes [1], perpetuating health inequities [2,3] and informing social determinants of health. Children are especially at risk when families face unmet social needs [4,5]. Driven by recent recommendations, there has been a rapid uptake of social need screening [2,6]. Although screening can be relatively straightforward, linkage to resources to address social needs is a major challenge [7,8].

Typically, clinicians provide families who are identified with a social need with a resource sheet. Families are then responsible for follow-up. Most clinics do not have social workers or other staff to help families access services and overcome barriers, such as language or cultural differences, financial constraints, transportation issues, limited internet access, or lack of awareness about available resources. Thus, families are often left to navigate complex social services independently, which can result in significant difficulties in obtaining much-needed assistance and support [9]. This passive provision of information is rarely effective. It is imperative to develop scalable strategies that screen for social needs and effectively link to services.

Digital health technology could improve both screening and resource referral to assist vulnerable populations [10]. Currently, electronic health records (EHRs) help facilitate screening, and patient portals help with bidirectional communication [2]. However, this does not eliminate the need to maintain lists of resources and the need to link individuals to matching resources. Semiautonomous intelligent and conversational digital health technologies, such as chatbots (conversational agents or dialogue systems), can help address these gaps. By using machine learning algorithms and natural language processing, chatbots can deliver personalized feedback and health recommendations to a wide range of users via interactive, user-friendly interfaces that are designed to maintain human conversation [11,12]. The capacity of the technology to reach and assist a large number of users simultaneously offers a cost-effective and efficient method for delivering personalized health services [13,14]. Chatbots have been used for health care communications, including health information seeking, health screening, and health care support, and to improve adherence to recommended care [15-20]. A previous study [21] described a chatbot to screen adults with low and high health literacy for social needs in emergency departments. The authors reported that the performance of the chatbot is comparable to that of traditional screening, and there is a greater interest from lower literacy participants for a chatbot. At a broader scale, chatbots show promise to facilitate social need screening and provide personalized resources to families outside of the traditional clinic setting via speech or text and could improve access [22,23] and further contribute to increased understandability and personalization while addressing social needs [21,24].

The DAPHNE (Dialog-Based Assistant Platform for Healthcare and Needs Ecosystem) chatbot project has been initiated to address unmet social needs via a conversational interface for low-income or resource-limited families, who often have trouble with a complex web of social challenges that include food insecurity, inadequate housing, and financial difficulties [25]. These vulnerable groups typically experience lower incomes, higher rates of unemployment, and diminished access to quality health care services. In the Nationwide Children's Hospital (NCH) primary care clinics, approximately 10% of families are identified to have at least 1 unmet social need [4], with >16% facing food insecurity (based on current data from our ambulatory patient population). This emerging need for social support has been the main motivation of our study. In this paper, we report our findings from the iterative design, prototype development, and evaluation of the DAPHNE chatbot for social need screening and resource referral. In this stage of development, we focus on food insecurity, the most frequently endorsed unmet social need, which has a significant impact on health care costs [26,27].

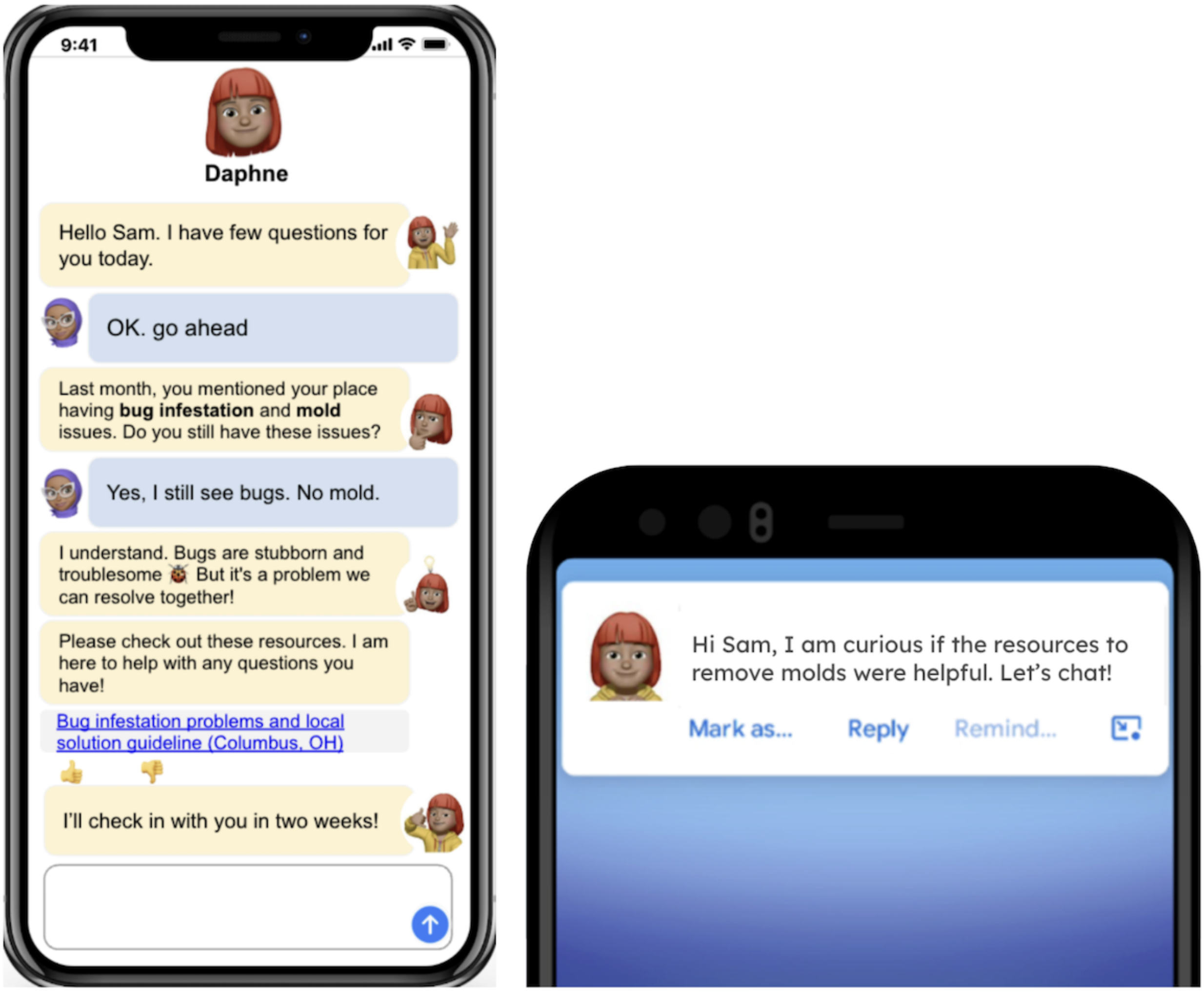
DAPHNE Chatbot

DAPHNE is a web-based application available via a computer or an iOS- or Android-based mobile device over a web browser. Figure 1 presents the initial wire-frame concept. The DAPHNE conversational interface prototype was designed using Adobe XD (Adobe Inc), Expo (650 Industries, Inc), and JavaScript (Oracle Inc) with a secure text-to-speech and speech-to-text service for voice interaction using Amazon Web Services (Amazon.com, Inc). Conversational flow was designed to be rule based. We opted for a rule-based design over pretrained language model or hybrid model at this stage to ensure greater transparency, predictability, and control in system responses, which is crucial for accurately identifying needs and retrieving specified resources.

The architecture, including data storage, conversational intelligence, information search, and referral services, uses Amazon Web Services and Microsoft Azure (Microsoft Corp) backend services. DAPHNE leverages application programming interfaces (APIs) provided by community resource platforms to access resource databases. These platforms, such as FindHelp.org, 211.org, and Cap4Kids.org, provide information about community resources categorized by geographic region. DAPHNE's architecture is designed to be integrated with EHR, enabling the communication of social need screening results to health care providers such as social workers, community health workers, and care teams. Its functionalities are listed in Textbox 1. In the scope of this study, the resource database of DAPHNE was locally created for testing purposes, without leveraging real-time API connection to the community resource platforms. In addition, the prototype was limited to screen 1 social need to reduce complexity during the testing.

Figure 2 outlines the chatbot ecosystem framework. Within the scope of this study, we are focusing on iterative design and the evaluation of engagement using conversational interface (Figure 2A). In the next phases, DAPHNE will have backend cloud services and API connection to enable access to web-based resource databases (Figure 2B) and provider dashboard to track engagement, control content (Figure 2C), and integration to medical records to report back social determinants of health monitoring (Figure 2D).

Figure 1. Initial wire frames and mock-ups.



XSL•FO
RenderX

Textbox 1. DAPHNE (Dialog-Based Assistant Platform for Healthcare and Needs Ecosystem) chatbot functionalities and descriptions.

Profile page

- Users create their account and set up profile details, including name, zip code, family type and size, and income level. The information is to be used for resource-finding queries.

Avatar

- Users can create an avatar to personalize their chatbot experience. For the prototype, we used Apple's Memoji to create an avatar that dynamically reflects emotions [28].

Language selection

- Users can select their preferred language. The prototype included the following languages: Somali, Nepali, and Spanish.

Audio narration

- Users can use the text-to-speech and speech-to-text features to enable audio entry and engagement and listen to the responses.

Multimodal input

- Users can use voice input (using speech to input), assistive buttons with prepopulated responses to select, and text entry with a free-text form to interact with the chatbot.

Social need screening

- DAPHNE uses the following standardized questions [4] to guide the screening process:
  - Food: within the past 6 months, you worried that your food would run out before you had money to buy more.
  - Housing: do you think you are at risk of becoming homeless?
  - Transportation: in the past 12 months, has lack of transportation kept you from medical appointments or from getting medications?
  - Utility: in the past year, has the utility company shut off your service for not paying your bills?

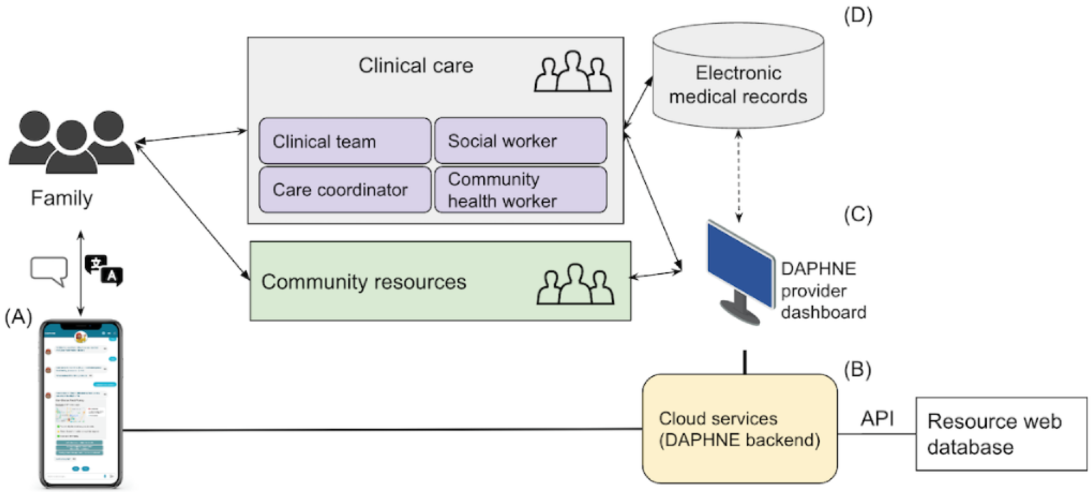
Interactive resource sharing

- DAPHNE can search the resource databases and present matching resources based on user response and ask follow-up questions.

Check-in and reminder notifications

- DAPHNE can send notifications. Scheduled check-in: it can collect information about whether the resource shared was useful. Reminder: it can set and send reminders asking whether the user would like to engage in another time. (Refer to Figure 1 for a reminder notification example.)

Figure 2. Conceptual framework of the chatbot ecosystem: (A) Conversational interface, (B) Backend cloud services and API connection, (C) provider dashboard, and (D) integration to medical records. API: application programming interface; DAPHNE: Dialog-Based Assistant Platform for Healthcare and Needs Ecosystem.



Methods

Study Design and Participants

The research was reported as a three-stage study: (1) understanding family needs and perception toward chatbots, (2) designing the chatbot, and (3) evaluating its feasibility and usability. We used stepwise, user-centered, iterative, and participatory development and improvement processes to ensure the proposed technology meets needs and expectations. Figure 3 presents the design, development, and evaluation stages.

Study 1 aims to understand families' ability to meet social needs and access essential resources and perceptions toward using chatbots to find resources. We conducted a cross-sectional web-based survey. A total of 201 adults in US households participated. Participants were living together with spouses, children, and significant others and self-reported an annual household income of ≤US $29,999 as of August 2023. The participants were recruited through a web-based platform designed for academic and market research, Prolific [29]. We followed a convenience sampling approach, inviting available participants via the survey tool. Survey details are available in Multimedia Appendix 1.

Study 2 focuses on iterative design, which includes the stages of ideation, prototyping, and refinement [30,31]. We held internet-based sessions with an interdisciplinary group of stakeholders. The sessions focused on answering the following research questions to understand design preferences and needs: "What are the pain points in current practices of social need screening and resource sharing?" "Why should we use or not use technology to facilitate this process?" and "How can we design and use a chatbot to connect with families in primary care settings in order to address social needs effectively?" The interdisciplinary stakeholder group (n=10) was formed internally at the NCH, including an epidemiologist (n=1, 10%), a primary care physician (n=1, 10%), a nurse (n=1, 10%), the director of clinical social work services (n=1, 10%), a community health worker (n=1, 10%), a public health scientist (n=1, 10%), an industry partner leader (n=1, 10%), a community partner leader (n=1, 10%), a family advocate (n=1, 10%), and an information system expert (n=1, 10%). Stakeholder group members were recruited within the NCH network (including primary care clinics) in September 2022.

Figure 3. Study process diagram.

Study 3 is a mixed methods evaluation of the prototype. Our evaluation methodology was informed by a feasibility framework [32], technology acceptance model [33], and usability scale—Usability Metric for User Experience–Lite (UMUX-Lite) [34]. We conducted scenario-based usability testing via a focus group interview with community health workers (who are also family advocates as part of the community) and via a web-based survey with social workers to examine the usability and feasibility of DAPHNE (semifunctional prototype) for families and communities. During these sessions, participants interacted with the chatbot to simulate the process of accessing and evaluating social resources (refer to Multimedia Appendix 1 for the scenario and questions). They used the chatbot to enter responses, navigate resource information, and provide feedback on its functionality qualitatively and quantitatively (via UMUX-Lite). Community health workers (n=4) and social workers (n=9) were recruited within the hospital network via email or phone (January to February 2023). Participation was voluntary for all participants.

Ethical Considerations

This study received ethics board approval (NCH institutional review board #00003766). All participants provided informed consent to participate to the study. Participants did not opt out of the study. Collected data and transcripts were deidentified. Participants received compensation for their participation, if permissible. This paper reports an aggregate summary of the data generated during the study, without any identifiable information. Due to privacy and confidentiality reasons, we are not able to share individual data points or transcripts.

Data Collection

In study 1, after participants provided consent, they completed a survey about their experiences. The survey captured their experiences and perceptions regarding the accessibility of social need resources. Questions included items on the awareness of and ability to access community support programs, methods used to obtain resources, and openness to using technological tools such as chatbots for resource assistance. Responses were collected anonymously, and the entire data collection process was structured such that the security and confidentiality of the participants were ensured. In study 2, iterative design sessions consisted of interactive interviews with open discussion guided by the research questions and moderated by a researcher. Wire frames were used to communicate initial design and revised designs of the chatbot (Figure 1). In total, three 1-hour sessions of stakeholder interviews were held between September to December 2022. The research team continuously communicated with the stakeholder group via email to share iterative improvements in the prototype. Throughout the sessions and conversations, stakeholder feedback was captured as conversation notes and observational notes. In study 3, social workers completed a 20-minute web-based scenario-based study to use the chatbot prototype and provide feedback (refer to Multimedia Appendix 1 for scenario and survey details). They responded via a web-based survey tool (REDCap [Research Electronic Data Capture]; Vanderbilt University). Community health workers were invited to a single-session focus group interview at the hospital (approximately 1 hour). The study team introduced the chatbot and its functionalities, shared examples, and provided a scenario-based demonstration. Usability questions and questions about dialogue and conversational design, voice interaction, perceived opportunities, and barriers were verbally discussed, which followed a similar approach to the web-based survey protocol. Data were collected via field notes.

Data Analysis

Study-1 analysis included descriptive statistics to summarize demographic information and responses to survey questions. We compared observed distributions of income, age, and employment status with responses to the questions on the ability to meet social needs, knowledge about community resources, and perception of chatbot use. Then, we conducted chi-square analysis to assess the association and independence of categories. In study 2, we conducted content analysis to inform the chatbot development process [35]. Stakeholder feedback was systematically analyzed by a single researcher to identify emerging themes, patterns, and insights, which were instrumental in understanding stakeholders' needs and expectations. Given the nature of semistructured interviews and scenario-based surveys, study-3 data were analyzed using thematic analysis to synthesize the qualitative data and to understand the meanings and experiences reported in response to open-ended questions and captured during the interviews [36]. The process began with 2 researchers independently conducting initial coding of the data. This coding was primarily inductive, allowing themes to emerge from the data, although a preliminary framework based on existing literature was also considered to guide the analysis. Regular discussions were held to review codes and themes, ensuring consistency and comprehensiveness. Data saturation was assessed to determine when no new themes were emerging, indicating sufficient depth of inquiry. Discrepancies between researchers were resolved through consensus; if consensus could not be reached, a third researcher was consulted to make a final decision, ensuring objectivity and reliability. In addition, we reported the total score of UMUX-Lite, with an expected usability score of ≥60 [34,37]. The thematic analysis and usability results were triangulated to provide a robust understanding of both user satisfaction and deeper user experiences.

Results

Study 1: Family Needs and Perceptions

We surveyed 201 low-income households, each with at least 1 unmet social need, to understand their willingness to use a chatbot for resource assistance. As shown in Table 1, demographic data showed an equal sex split (male: 100/201, 49.8%; female: 100/201, 49.8%; unreported: 1/201, 0.5%). Age distribution skewed toward the 21-40 years range (122/201, 60.7%), and the most reported income level was from US $20,000 to US $29,999 (84/201, 41.8%), followed by <US $10,000 (47/201, 23.4%). Employment status varied, as 24.4% (49/201) of the participants were full-time employees, and 21.9% (44/201) of them were not in paid work. Regarding unmet social needs, 33.3% (67/201) of the participants found it moderately hard to meet. The majority of participants (106/201, 52.7%) were aware of and had used community resources. A substantial portion of participants primarily used the internet for discovering community resources (133/201, 66.2%), and 60.2% (121/201) of the participants were open to using a chatbot for resource finding.

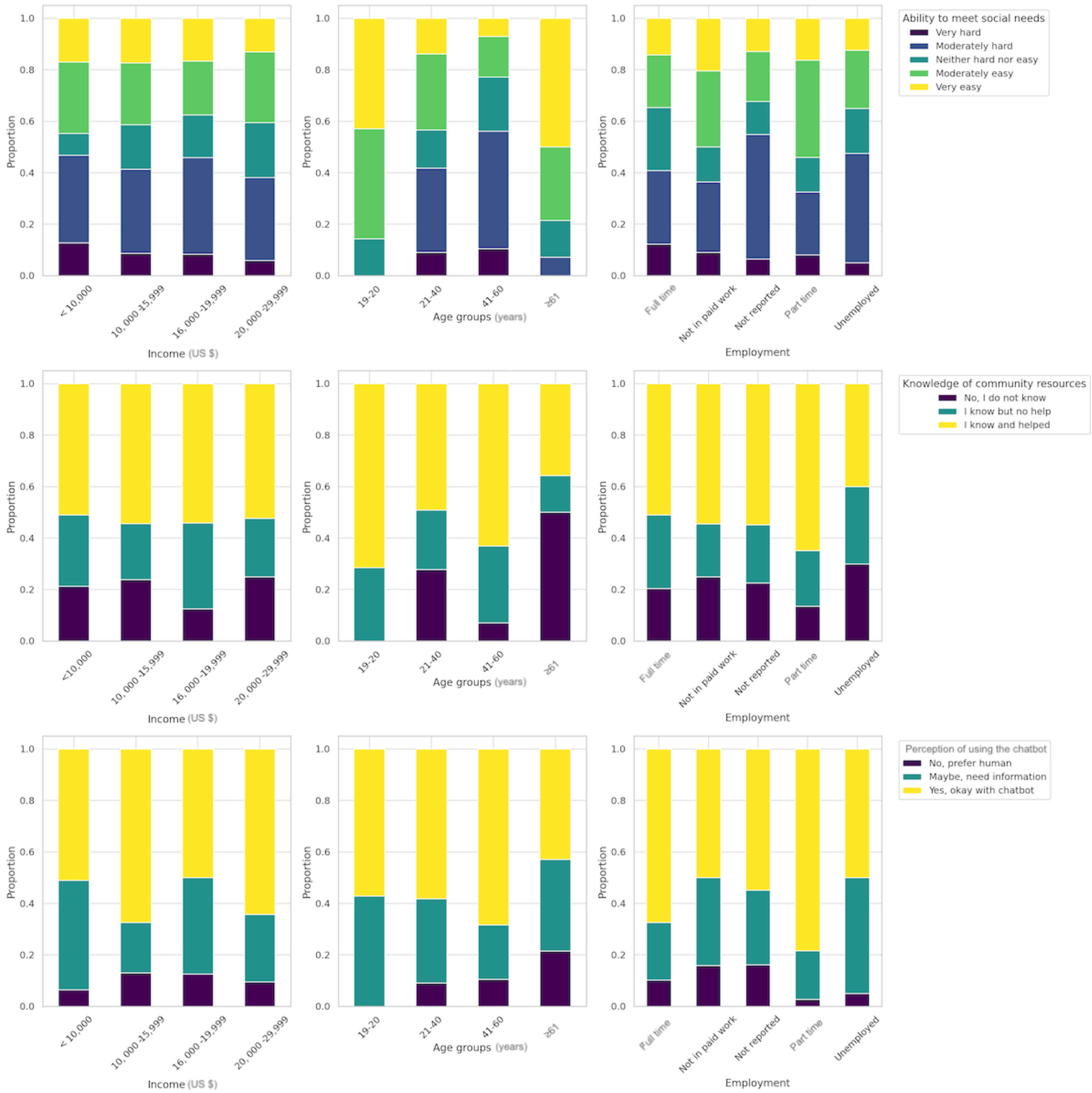
Figure 4 outlines the proportional distribution of the responses. Lower-income groups reported a lower ability to meet social needs, with similar trends observed among middle-aged groups and those who were unemployed. There was an increase in meeting social needs among slightly higher–income earners, younger age groups, and individuals who were fully employed. Knowledge of community resources was lower among individuals who were unemployed, and the oldest age group, those aged ≥61 years, exhibited a lower level of community resource awareness and use of community programs. In terms of chatbot use, while there was a general receptiveness across all income levels, employed and younger individuals, particularly those aged between 21 and 40 years, demonstrated a higher tendency to use chatbot technology. In contrast, the oldest age group demonstrated a greater preference for human interaction over using chatbots.

Table 1. Demographics of web-based survey participants (n=201).

Characteristics	Participants, n (%)
Income range (US $)	
<10,000	47 (23.4)
10,000-15,999	46 (22.9)
16,000-19,999	24 (11.9)
20,000-29,999	84 (41.8)
Sex	
Female	100 (49.8)
Male	100 (49.8)
Not reported	1 (0.5)
Race	
Asian	14 (7)
Black	24 (11.9)
White	146 (72.6)
Other	9 (4.5)
>1	8 (4)
Age group (y)	
19-20	7 (3.5)
21-40	122 (60.7)
41-60	57 (28.4)
≥61	14 (6.9)
Not reported	1 (0.5)
Employment status	
Full time	49 (24.4)
Part time	37 (18.4)
Unemployed (and job seeking)	40 (19.9)
Not in paid work (eg, homemaker, retired, or disabled)	44 (21.9)
Not reported	31 (15.4)
Ability to meet social needs	
Very hard	17 (8.5)
Moderately hard	67 (33.3)
Neither hard nor easy	34 (16.9)
Moderately easy	52 (25.9)
Very easy	31 (15.4)
Knowledge about community resources	
No, I do not know about these programs	45 (22.4)
Yes, but I could not get help	50 (24.9)
Yes, and they helped	106 (52.7)
Methods of finding community resources	
Internet (websites and email)	133 (66.2)
In person (such as at community centers and food banks)	93 (46.3)
Phone calls	61 (30.3)
Mobile apps	27 (13.4)

Characteristics	Participants, n (%)
Other	20 (10)
Perception about using chatbot	
No, I would rather talk to a person	20 (10)
Maybe, I need to know more about how it works	60 (29.9)
Yes, I would be okay with using a chatbot	121 (60.2)

Figure 4. Stacked bar graphs showing the proportional distribution of the survey responses for social needs and perception toward chatbot by income, age group, and employment status.



The chi-square analysis did not yield strong evidence of association of income (χ^2_6=7.9; *P*=.24), age (χ^2_6=6.5; *P*=.37), and ability to meet social needs (χ^2_8=6.2; *P*=.63) with the perception of using a chatbot. However, there is a statistically significant association between knowledge about community resources and chatbot perception (χ^2_4=12.9; *P*=.01). In addition, the relationship between employment and chatbot perception is marginally close to being significant (χ^2_8=15.5; *P*=.051).

Study 2: Iterative Design

The themes were grouped under 3 research questions. Textbox 2 outlines the questions and themes for each question. Themes included common pain points: technology opportunities and challenges and technology considerations, including inclusivity, personalization, and information about accessing resources. On the basis of stakeholder feedback, we improved our chatbot design (Figure 5). We updated the prototype to include chatbot language options, modified language (eg, “What makes it hard to get food?”), and resource education options (ie, eligibility criteria, documentation requirements, and referral guidance). These components were initiated and are under development.

Textbox 2. Questions and themes from stakeholder group sessions.

"What are the pain points in current practices of social need screening and resource sharing?"

- Inadequate or inconsistent screening tools: the tools used for social need screening may not be comprehensive enough for addressing all social needs associated with social determinants of health, resulting in incomplete assessments. In addition, there might be inconsistency in the use of these tools across different settings, leading to variations in the identification and understanding of social needs.
- Limited provider training and awareness: health care providers and other professionals involved in social need screening may lack sufficient training and awareness about how to implement screening, screen for unidentified needs, and address the unmet social needs. This can lead to lower quality of service as well as adversely affect the quality of life.
- Fragmented systems and lack of integration: social need screening and resource-sharing efforts are often fragmented across different divisions, departments, and organizations. This can lead to poor communication and collaboration, creating barriers to the effective identification and provision of resources.
- Insufficient resources and capacity: there may be a lack of adequate resources and capacity to address identified social needs based on the location and resources of institutions (rural vs urban health institution), resulting in unmet needs or long waiting periods for support. This can exacerbate existing disparities and negatively impact health outcomes.
- Stigma and privacy concerns: patients and families may be reluctant to disclose their sensitive information as well as their social needs due to concerns about stigma or privacy. This can prevent the accurate identification of needs and hinder access to appropriate resources.
- Cultural and linguistic barriers: cultural and linguistic differences may negatively impact communication among providers and patients or families, leading to misunderstandings and underestimation of social needs. This can result in the inadequate provision of resources and support.

"Why should we use or not use technology to facilitate this process?"

- Opportunities:
  - Improved efficiency: technology can streamline the screening and resource-sharing process, reducing the time and effort required by both providers and patients or families. Automated systems and digital platforms can facilitate data collection, storage, and retrieval, making it easier to identify and address social needs at scale, especially within low-resource settings.
  - Standardization and consistency: digital tools can help ensure that social need screening is conducted in a standardized and consistent manner across different settings, reducing variations in the identification and understanding of social needs.
  - Personalization and customization: technology can enable more personalized and customized approaches to social need screening and resource sharing, tailoring interventions to the specific needs and preferences of patients and families (which can be beneficial considering cultural appropriateness and language options).
- Challenges:
  - Digital divide: the use of technology may exacerbate existing disparities in access to digital tools, particularly among vulnerable populations. This can result in the further marginalization of those who may be most in need of support yet do not have access to the necessary technology.
  - Privacy and security concerns: storing and sharing sensitive data and private information electronically can raise privacy and security concerns for individuals and institutions. It might be particularly concerning if technology providers have not enforced appropriate safeguards to protect the information.
  - Implementation challenges: introducing a new approach using technology into social need screening and resource-sharing processes may involve significant financial and human resource investments as well as create barriers or burdens related to staff training, infrastructure, and technological compatibility.

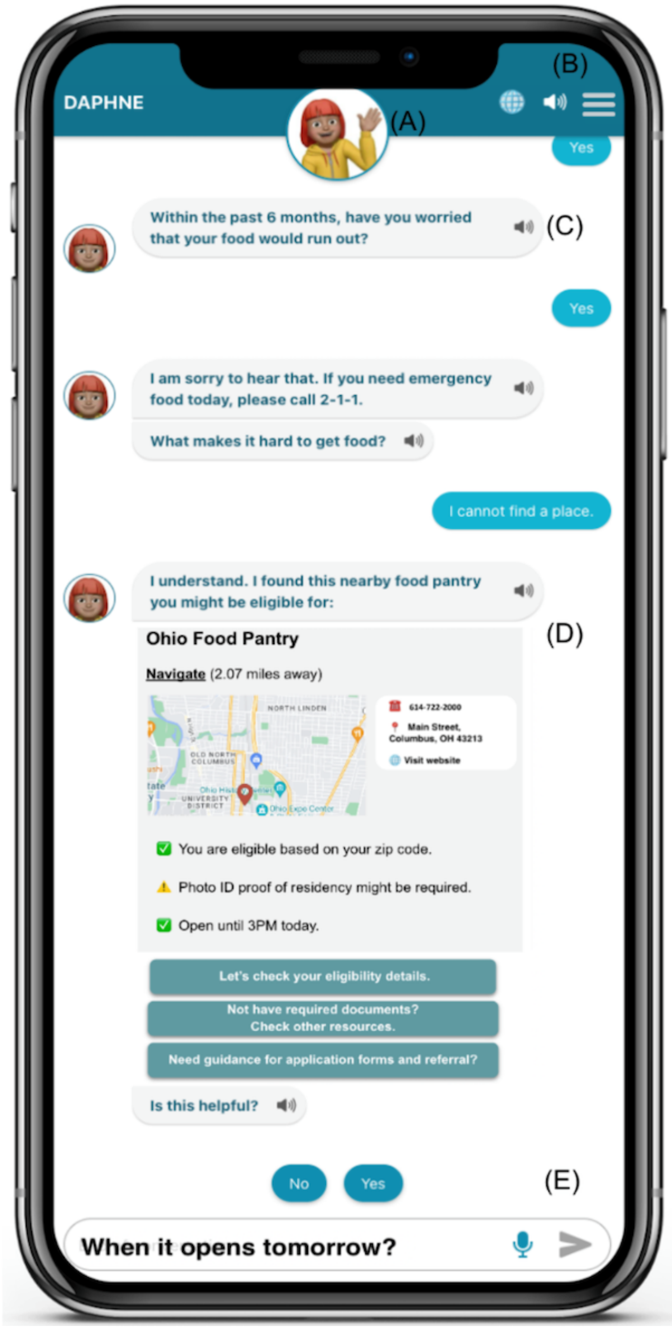
"How can we design and use a chatbot to connect with families in primary care settings in order to address social needs effectively?"

- Leveraging provider and technology collaboration: this means that chatbot and health care providers (eg, primary care team) and community centers can work together to serve families better and more effectively.
- Suggested use case 1: chatbot can be used as a triage follow-up tool for health care providers and community centers to follow-up on patients after their visit to ensure that patient and family needs are met and check whether resources are useful. Thus, the chatbot can timely inform providers to intervene in case of unmet needs as well as help identify invalid or noneligible resources and update their resource list and database accordingly.
- Suggested use case 2: chatbot can be used as a prescreening tool to inform providers and community centers about what needs to be communicated with families, getting them ready for detailed conversation about the resources and how to access them. The chatbot can ease the process of support and patient engagement so that providers can timely serve more families and spare more time to engage as well as for identifying and addressing urgent needs during their conversations.
- Conversational design could be guided to be more inclusive and personal.
- Current screening instruments are not individually relatable or personal, and chatbot can be guided toward more conversational personalized screening, which can eventually inform current screening tools. Reframing dialogues toward positive attitude and social norms are some of the methods discussed.
-

Cultural appropriateness and language barriers could be addressed by chatbot providing language options (eg, Ohio has a high rate of Nepali and Somali refugees with limited English proficiency and requires interpreter services) and culturally guided and appropriate conversations and dialogue flow, accounting in cultural norms and connotation (eg, In some cultures “free resource” still may mean you have to pay back due to cultural expectations or practices).

- Chatbots can help educate families about accessing resources (beyond sharing resources, guiding them on self-referral, how to check eligibility, and how to navigate web-based resources). This may eventually reduce dependency on low-risk or quickly accessible resources by families and patients, reserving time and resources of health institutions and community centers to be spent on patients and families having urgent or complex needs.

Figure 5. Revised semifunctional prototype (web application). (A) customizable avatar; (B) profile and setting menu, language selection, and enable or disable audio narration (text to speech); (C) repeating audio narration for preferred messages; (D) interactive resource screen with navigation and communication details and follow-up questions suggested by stakeholders; and (E) assistive buttons for quick response, text entry feature, and speech-to-text feature for voice interaction.



Study 3: Prototype Evaluation

We collected feedback from 13 participants, including community health workers (n=4, 31%) and social workers (n=9, 69%), within the NCH network (Table 2). Community health workers had <5 years of professional experience (3/4, 75%). They had limited experience with chatbots and conversational agents. They are supporting families through the Connecting Families 4 Success program at NCH, which provides resource linkages to families with identified social needs.

The majority of social workers (6/9, 67%) had 1 to 10 years of experience in their current profession. Most (8/9, 89%) had prior exposure to chatbots. Social workers reported that they serve an average of 165 (SD 229) patients or families monthly. The social workers responded to the usability questions (UMUX-Lite) using a 7-point Likert scale, agreeing that the chatbot's capabilities met expectations toward addressing social needs (average score 5.4, SD 1.1) and that it is easy to use (average score 5.6, SD 1.8). Collectively, they agreed on the usability of chatbot, providing an average score of 72 on the System Usability Scale (calculated using the regression equation developed by Lewis et al [38]).

We analyzed responses from community health workers and social workers together and grouped them under the following 4 overarching themes: user experience (how users perceived the chatbot, its functionality, and its satisfaction and interacted with it), feature preferences (user preferences for additional features in the chatbot), resource concerns (the challenges related to the accuracy, relevance, and timeliness of the resources provided by the chatbot), and perceived disadvantages and challenges (limitations and potential obstacles associated with using the chatbot; Table 3). The themes were informed by the interview and survey questions, focusing on user experience (meeting expectations and ease of use), chatbot dialogue, audio interaction, perceived advantages and disadvantages, and integration to clinical workflow. As outlined in Table 3, our thematic analysis suggests that the use of a chatbot is perceived to be useful for patients, caregivers, and providers and can help with addressing unmet social needs and resource sharing. More specifically, social workers and community health workers appreciated its clear interface but noted the need for detailed options on eligibility and documentation. Opinions on audio narration were mixed, with some valuing it for accessibility and others preferring text for privacy. The need for multilingual support and careful consideration of privacy with EHR integration was also emphasized. Users were concerned about the accuracy and currency of resource information and doubtful of the technology's current capability to update its content without human verification. Major drawbacks included the chatbot's inability to interpret nonverbal cues and complex situations, limited access to necessary technology for some users, and concerns about data privacy and trust.

Table 2. Study-3 participant demographics (n=13).

Demographics	Social worker (n=9), n (%)	Community health worker (n=4), n (%)
Have you ever used chatbots before for any purpose?		
Yes, multiple times	8 (89)	1 (25)
No, never	1 (11)	3 (75)
Age group (y)		
18-24	1 (11)	2 (50)
25-34	3 (33)	1 (25)
35-44	2 (22)	1 (25)
45-54	3 (33)	0 (0)
Experience (y)		
<1	1 (11)	2 (50)
1-5	3 (33)	1 (25)
5-10	3 (33)	1 (25)
10-20	1 (11)	0 (0)
>20	1 (11)	0 (0)

Table 3. Themes from interviews and web-based survey with social workers and community health workers.

Themes	Explanations	Quotes
User experience	• Conversational interface was found to be simple and understandable. It was noted as "easy to use," as it enabled turn-based conversation via natural language through SMS text messaging–like interface with the option of voice interaction. • The simplicity of the dialogue was appreciated, although some suggested adding more options and details about resources, such as expanding options for eligibility, required documents, and forms. Although these components were included during the iterative design process, they were not functionally adapted in the prototype during the testing.	• "[Dialogues] were clear and easy to understand" (Participant #5). • "Dialogue is overall acceptable. I'd be interested in hearing what 'other' options it may generate for people who have food instability due to other reasons" (Participant #8).
Feature preference	• There was mixed feedback on whether audio narration would be preferable. Some participants thought it would be helpful, especially for those who face language barriers or have difficulty reading or writing or physical impairment. However, others believed that families may prefer text or typing in certain cases, such as when they are in public places. • Several participants mentioned the importance of offering multiple language options to engage with a diverse population, including non-English speakers. Current service limitations and inability to follow-up due to language barriers necessitate such alternative support on communicating social needs. • Opinions on integrating chatbot data with EHRs[a] were mixed. Some participants supported integration for better decision-making and follow-up, while others were concerned about privacy, consent, and the potential for surveillance or stigmatization.	• "...voice to text can be sometimes challenging, as with Siri at times and this is more so common for individuals whose first language is not English" (Participant #4). • "I think [chatbot] could [help if integrated to EHR] for basic medical records, but families might try to use this to ask medical questions thinking a doctor might respond" (Participant #3).
Resource concerns	• Participants expressed concerns about the accuracy of resources and how to the resource information is up-to-date. The concerns were also about identifying relevant resources to specific age groups or situations, which may not be available in a single database or web resource. In the current practice, support teams (eg, community centers and social workers) need to reach out and check the availability of resources (eg, food pantry) and eligibility to ensure the resource provider is operational before referring to a patient and family. • Participants emphasized their skepticism behind the need to use technology to ensure that resources are up to date (given the necessity of calling and communicating with the service providers to ensure the resource list is up to date).	• "[resources] might not always be accurate if all the places are not regularly updated in the system" (Participant #9) • "It will also be useful to have more details about the pantry such as open hours of operation" (Participant #6).
Perceived disadvantages and challenges	• Some participants expressed concerns about the chatbot's major communication limitation, which is its inability to assess nonverbal cues and understand nuanced situations; therefore, it may not be able to articulate social need details to guide toward appropriate resources. • Limited access to technology, broadband or Wi-Fi, or mobile devices could pose challenges for some families. • It is noted that some families or patients might be uncomfortable sharing personal information with a chatbot or might not trust the information provided by a chatbot.	• "One disadvantage might be the chatbot not being able to assess nonverbal cues, or other concerns the family might have that can't be typed into a box" (Participant #5). • "[Chatbot] can't share nuanced situations. Can it understand the intersection of different needs?... organizations that have food and housing resources if you are looking for orgs specifically with both" (Participant #4).

[a]EHR: electronic health record.

Discussion

Principal Findings

We conducted a 3-stage study with a multistakeholder group, focusing on the development and evaluation of a social need screening chatbot for families. Engaging stakeholders, including low-income households, health care providers, and family advocates, throughout the need assessment and design and development process helped identify specific needs, preferences, perceptions toward using chatbots and potential barriers to adoption. Our study highlights the importance of a multilayered and user-centered iterative design and the development of chatbots for social needs. Furthermore, it was a promising step forward to develop a chatbot collectively with partners and to serve families effectively via conversational systems [39,40]. It also contributes to the literature [19,21] by providing further evidence on diverse stakeholder perceptions of chatbot use in social need screening and resource sharing.

Our first study highlights that the awareness of community resources is traditionally less among lower-income and unemployed groups, as well as among older households. However, the majority of participants, irrespective of demographics, use the internet to identify community resources. Although this is promising, further investigation is needed on how to leverage current communication technologies to close the gap in digital inclusion, improving the awareness of resources and access [41,42]. Furthermore, a broad acceptance of chatbot technology across all income levels is observed, as younger individuals, notably those aged 21 to 40 years, lead the charge in embracing this digital interaction. This suggests a generational pivot toward using chatbots similarly to prior observations with the acceptance of new technologies [43]. The knowledge and employment status of individuals with social needs are further associated with their perception about chatbots.

Our second study highlights potential implementation areas and improvements for the chatbot to be more engaging and effective. In the current health care ecosystem, chatbots may serve a dual function as follow-up tools and triage systems, as recommended by health care providers and community centers subsequent to family visits, ensuring the effective use of resources and helping meet social needs. In line with this, chatbot data can be used for timely feedback about health care systems about unmet needs as well as to facilitate the updating of resource repositories and databases (eg, user feedback on nonoperational food pantries). When deployed as prescreening instruments, chatbots can enable providers and community centers to be adequately prepared for comprehensive discussions about resource availability and access, thereby streamlining support procedures. Conversational design can be strategically geared toward providing more inclusive and personalized need assessment [44-46]. By enhancing the relational capacities of a chatbot [47] such as showcasing positive attitudes and adhering to social norms (eg, “Others have found assistance through this local agency.”) and implementing behavioral nudges (eg, “Completing this screening will require just a few minutes.”) within the conversational design, its engagement abilities can be bolstered. In addition, a chatbot can address cultural compatibility and linguistic barriers by providing multilingual options and culturally sensitive dialogues [48]. In the context of current practices to adequately address social needs [49], chatbots can act as supportive adjuncts, supplementing and enhancing these efforts. Furthermore, chatbots can play an educational role in assisting families in understanding and accessing resources, steering them through self-referral processes, eligibility assessments, and web-based resource navigation. As a next step, chatbots can be instrumental in augmenting health literacy [50], as they have been well received in addressing social needs among populations with low literacy levels [24].

Building on stakeholder feedback, we implemented improvements in our chatbot prototype. In study 3, our descriptive analysis showed a diverse range of participants in terms of age, experience, and department affiliation, providing a rich perspective on the chatbot’s applicability with various contexts. Participants have generally rated the chatbot’s capabilities and ease of use as average to high. The results of our study indicate that the chatbot designed for addressing social needs is generally well received, with most users finding it easy to use and having a positive user experience overall. This overlaps with the current trajectory of chatbot use in health care, as the capability as well as usefulness of chatbots increases [51,52].

Emerging Opportunities and Barriers

Audio narration emerged as a theme with mixed opinions. While some users believed it could benefit those facing language barriers or with difficulty reading, others felt that text-based communication would be more appropriate in publicly available spaces, which was also noted previously as a common concern about using voice interaction in health information exchange [53].

Resource accuracy and availability were identified as concerns by participants, emphasizing the importance of regularly updating resource information and ensuring that resources are relevant to specific age groups or profiles. To ensure the sustainability and maintenance of resources and accuracy of DAPHNE’s responses, automatic updates of the chatbot resource listings by syncing with APIs provided by established community resource platforms may ensure real-time accuracy. In addition, a user feedback mechanism is in place (regarding whether the resource is helpful) via the chatbot interface, allowing users to report any discrepancies or changes required in the resource information directly through the chatbot interface. Such a human-in-the-loop feedback mechanism is crucial for continuous improvement and helps maintain a high level of trust and reliability in the resources provided [54,55].

Integration with EHRs received mixed feedback, with some users supporting the idea for better decision-making and follow-up. Such implementation is principally viable to support decision-making with a feedback mechanism [56]. Others expressed concerns about privacy and potential stigmatization, which may lead to labeling and internalized negative stereotypes that may reduce disclosing social needs [57]. Therefore, EHR integration and adoption require detailed investigation to reduce barriers and inequality in medical documentation [57,58]. For privacy, data should be transmitted to EHRs via Health Insurance Portability and Accountability Act–compliant services and encrypted, and a governance body should be established for regular oversight [59,60]. In addition, where possible, data anonymization could be implemented to protect patient identity. To address stigma, training health care providers on handling sensitive information respectfully and confidentially is needed, which is crucial for integrating social determinants of health into patient care without bias [61].

Participants raised several disadvantages and challenges related to the chatbot’s ability to assess nonverbal cues and accurately screen needs, limited access to technology, and trust issues when sharing personal information with a nonhuman source. In this regard, using a single modality communication medium (no visual exchange) might limit the chatbot’s ability to process nonverbal cues. Multimodal approaches with chatbots may overcome this limitation in the future [62]. The chatbot’s dependency on technology platforms (computer, Wi-Fi, or smart mobile devices) and data plans may limit access. Although there are existing programs to support broadband access to low-income families (eg, Affordable Connectivity Program by Federal Communications Commission) [63] in the long term, this underscores the importance of strategic consideration of the digital divide and accessibility challenges when designing and implementing chatbots for social needs. In earlier studies, interactive voice response systems and chatbots based on SMS text messaging have been viable alternatives, which might be adapted for social need screening and resource sharing, especially in rural areas [64]. The lack of trust between the chatbot and families might negatively influence its use. Some participants expressed concerns about sharing personal information with a nonhuman actor or not trusting the information provided by the chatbot. Literature has mixed evidence toward trust between humans and chatbot [65], and further research can inform the trust built between families and chatbot.

Expanding With Broader Social Needs

The potential scalability of the DAPHNE chatbot extends beyond its current application in food insecurity. At its current state, its design accommodates the integration of additional social needs by incorporating a flexible, rule-based conversational architecture that can be customized with minimal technical adjustments. For example, the chatbot could be adapted to screen for housing instability or transportation difficulties by updating the dialogue scripts and linking to different resource databases. Moreover, the backend infrastructure, built on cloud services, supports scalability to handle increased user traffic and data volume as the system expands.

Improving Technical Capabilities

Although the current chatbot conversational flow was designed to be rule based, transformer-based large language models and artificial intelligence (AI)–enabled conversational agents are alternative approaches to delegate a variety of tasks [66,67]. These intelligent chatbots include a large range of functionalities that set them apart from their predecessors, such as (1) engaging in discussions across multiple topics; (2) managing multiturn conversations; (3) retaining information from previous conversations; (4) operating in both task-based and non–task-based modes; and, most importantly, (5) collaborating effectively with users. In particular, being able to work together with users, listening to their instructions, and understanding their preferences through naturally occurring conversations open up a wealth of opportunities for both health care providers and families with social needs. Although there are major concerns related to privacy, reliability, and accuracy [68], we can expect the development of hybrid solutions that have the increased conversational competence of chatbots while being constrained by the strict specifications of a rule-based system. In addition, current guidelines and practices for developing skills to engage in a sensitive conversation, such as food insecurity, can be informative for AI-enabled chatbot development [69,70]. For instance, American Hospital Association's guidelines suggesting cultural competency, motivational interviewing, active listening, and empathic inquiry would be valuable input for conversational design and development [71].

Future Research

It is essential to evaluate the long-term effectiveness, scale-up capability, and impact of chatbots for addressing social needs through rigorous and comprehensive evaluation methodologies [72,73]. Our study provides preliminary evidence on the iterative design and evaluation of the chatbot for addressing social needs (focusing on food insecurity screening and resource sharing with text and voice interaction). However, future research should investigate the impact of chatbot interventions on users' health outcomes, quality of life, and access to resources, as well as the cost-effectiveness and scalability of such interventions. While chatbots can play a valuable role in addressing social needs, they are unlikely to replace human service providers entirely. Instead, chatbots can be considered complementary tools that support and enhance existing services by providing timely, personalized, and accessible information and resources. Future research should explore the potential synergies and integration opportunities between chatbots and other digital health interventions, such as telemedicine, mobile health apps, and online support groups, to maximize the overall impact on users' health and well-being.

Limitations

There are several limitations that should be acknowledged while interpreting the study results. First, our participants in studies 2 and 3 consist of stakeholders representing providers more than patients and caregivers, which may potentially skew the feedback toward professional perspectives. Providers may have perspectives or biased interests that differ from those of patients and caregivers, potentially leading to an overemphasis on the functionality and clinical utility of the chatbot. This skew could limit the generalizability of our findings to broader end-user experiences and might overlook key usability challenges faced by less technologically proficient users. Moreover, the diversity and size of our participant sample may not fully represent the broader population, which could limit the generalizability of our findings. Second, our research was conducted in a controlled setting with a single scenario and did not involve any real-world testing and observations. As such, the practical implications of our study remain limited to self-reported and perceived usability, feasibility, and implementation with a limited user experience. Further research in real-world scenarios is required to evaluate the effectiveness and feasibility of the chatbot in addressing social needs. Third, our research focused primarily on qualitative data, thus lacking quantitative information to assess the chatbot's performance with a longitudinal observation. Although we collected preliminary data from households regarding their social needs and perception toward using the chatbot, the effort was limited in terms of demographic diversity and feedback (without chatbot engagement), and it may be subject to self-report bias. Future studies will aim to collect quantitative measures with real-world chatbot use, such as user logs, response accuracy, and user satisfaction rates, to provide a more comprehensive evaluation of the chatbot's performance in addressing social needs.

Conclusions

The study reported the iterative design and evaluation of a chatbot for social need screening and resource identification designed to scale screening and resource sharing for low-resource communities and disadvantaged neighborhoods. Furthermore, it may augment health center services, with low-risk tasks (such as resource finding and sharing) being delegated to the chatbot to scale the services provided. The DAPHNE chatbot has garnered largely favorable responses, providing initial evidence for its practicality and viability. Crucial factors in designing chatbots for social needs involve fostering user confidence, ensuring the precision of resources, and tackling accessibility obstacles. Future studies should investigate the efficacy, cost-efficiency, and expandability of chatbot initiatives, the opportunities provided by conversational AI technologies, and possible collaborations with other established digital health interventions.

Acknowledgments

The authors are thankful to the project advisory board for their valuable feedback and input: Amrik Khalsa, Nicholas Jones, Jeannette Foster, Lori McCullough, Huan Sun, LaToyia Matthews, and Samantha Boch. The authors would like to acknowledge the support of the data science and design and technology teams to the design, development, and testing of the chatbot: Tyler Lieser, Brannon Oiler, Brandon Abbott, Silvio Fernandes, and Amad Hussain. This project was a recipient of phase 1 and 2 prizes of the Health Resources and Services Administration's (HRSA) "Building Bridges to Better Health: A Primary Health Care" Challenge in 2022-2023. Research reported in this publication was partially funded by the HRSA challenge prize.

Conflicts of Interest

ES serves on the editorial board of JMIR Publications.

Multimedia Appendix 1
Survey materials.
[DOCX File , 18 KB - humanfactors_v11i1e57114_app1.docx]

References

1. Hood CM, Gennuso KP, Swain GR, Catlin BB. County health rankings: relationships between determinant factors and health outcomes. Am J Prev Med 2016 Feb;50(2):129-135. [doi: 10.1016/j.amepre.2015.08.024] [Medline: 26526164]
2. Boch S, Keedy H, Chavez L, Dolce M, Chisolm D. An integrative review of social determinants of health screenings used in primary care settings. J Health Care Poor Underserved 2020;31(2):603-622. [doi: 10.1353/hpu.2020.0048] [Medline: 33410796]
3. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Health Care Services, Committee on Integrating Social Needs Care into the Delivery of Health Care to Improve the Nation's Health. Integrating Social Care Into the Delivery of Health Care: Moving Upstream to Improve the Nation's Health. Washington, DC: National Academies Press; 2019.
4. Dolce M, Keedy H, Chavez L, Boch S, Zaim H, Dias B, et al. Implementing an EMR-based health-related social needs screen in a pediatric hospital system. Pediatr Qual Saf 2022 Jan 21;7(1):e512 [FREE Full text] [doi: 10.1097/pq9.0000000000000512] [Medline: 35071955]
5. Tedford NJ, Keating EM, Ou Z, Holsti M, Wallace AS, Robison JA. Social needs screening during pediatric emergency department visits: disparities in unmet social needs. Acad Pediatr 2022;22(8):1318-1327 [FREE Full text] [doi: 10.1016/j.acap.2022.05.002] [Medline: 35537675]
6. LaForge K, Gold R, Cottrell E, Bunce AE, Proser M, Hollombe C, et al. How 6 organizations developed tools and processes for social determinants of health screening in primary care: an overview. J Ambul Care Manage 2018;41(1):2-14 [FREE Full text] [doi: 10.1097/JAC.0000000000000221] [Medline: 28990990]
7. Council on Community Pediatrics, Committee on Nutrition. Promoting food security for all children. Pediatrics 2015 Nov;136(5):e1431-e1438. [doi: 10.1542/peds.2015-3301] [Medline: 26498462]
8. Billioux A, Verlander K, Anthony S, Alley D. Standardized screening for health-related social needs in clinical settings: the accountable health communities screening tool. National Academy of Medicine Perspectives. 2017. URL: https://nam.edu/standardized-screening-for-health-related-social-needs-in-clinical-settings-the-accountable-health-communities-screening-tool/ [accessed 2024-06-21]
9. Vasan A, Kenyon CC, Palakshappa D. Differences in pediatric residents' social needs screening practices across health care settings. Hosp Pediatr 2020 May;10(5):443-446 [FREE Full text] [doi: 10.1542/hpeds.2019-0286] [Medline: 32350005]
10. Arsenijevic J, Tummers L, Bosma N. Adherence to electronic health tools among vulnerable groups: systematic literature review and meta-analysis. J Med Internet Res 2020 Feb 06;22(2):e11613 [FREE Full text] [doi: 10.2196/11613] [Medline: 32027311]
11. Sriwisathiyakun K, Dhamanitayakul C. Enhancing digital literacy with an intelligent conversational agent for senior citizens in Thailand. Educ Inf Technol (Dordr) 2022;27(5):6251-6271 [FREE Full text] [doi: 10.1007/s10639-021-10862-z] [Medline: 35002466]

12. Shumanov M, Johnson L. Making conversations with chatbots more personalized. Comput Hum Behav 2021 Apr;117:106627. [doi: 10.1016/j.chb.2020.106627]
13. Amiri P, Karahanna E. Chatbot use cases in the Covid-19 public health response. J Am Med Inform Assoc 2022 Apr 13;29(5):1000-1010 [FREE Full text] [doi: 10.1093/jamia/ocac014] [Medline: 35137107]
14. Barthelmäs M, Killinger M, Keller J. Using a Telegram chatbot as cost-effective software infrastructure for ambulatory assessment studies with iOS and Android devices. Behav Res Methods 2021 Jun;53(3):1107-1114 [FREE Full text] [doi: 10.3758/s13428-020-01475-4] [Medline: 32989722]
15. Kocaballi AB, Quiroz JC, Rezazadegan D, Berkovsky S, Magrabi F, Coiera E, et al. Responses of conversational agents to health and lifestyle prompts: investigation of appropriateness and presentation structures. J Med Internet Res 2020 Feb 09;22(2):e15823 [FREE Full text] [doi: 10.2196/15823] [Medline: 32039810]
16. Bérubé C, Schachner T, Keller R, Fleisch E, V Wangenheim F, Barata F, et al. Voice-based conversational agents for the prevention and management of chronic and mental health conditions: systematic literature review. J Med Internet Res 2021 Mar 29;23(3):e25933 [FREE Full text] [doi: 10.2196/25933] [Medline: 33658174]
17. Yang S, Lee J, Sezgin E, Bridge J, Lin S. Clinical advice by voice assistants on postpartum depression: cross-sectional investigation using Apple Siri, Amazon Alexa, Google Assistant, and Microsoft Cortana. JMIR Mhealth Uhealth 2021 Jan 11;9(1):e24045 [FREE Full text] [doi: 10.2196/24045] [Medline: 33427680]
18. Miner AS, Milstein A, Schueller S, Hegde R, Mangurian C, Linos E. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. JAMA Intern Med 2016 May 01;176(5):619-625 [FREE Full text] [doi: 10.1001/jamainternmed.2016.0400] [Medline: 26974260]
19. Kocaballi AB, Sezgin E, Clark L, Carroll JM, Huang Y, Huh-Yoo J, et al. Design and evaluation challenges of conversational agents in health care and well-being: selective review study. J Med Internet Res 2022 Nov 15;24(11):e38525 [FREE Full text] [doi: 10.2196/38525] [Medline: 36378515]
20. Chaix B, Bibault JE, Pienkowski A, Delamon G, Guillemassé A, Nectoux P, et al. When chatbots meet patients: one-year prospective study of conversations between patients with breast cancer and a chatbot. JMIR Cancer 2019 May 02;5(1):e12856 [FREE Full text] [doi: 10.2196/12856] [Medline: 31045505]
21. Kocielnik R, Agapie E, Argyle A, Hsieh DT, Yadav K, Taira B, et al. HarborBot: a chatbot for social needs screening. AMIA Annu Symp Proc 2019;2019:552-561 [FREE Full text] [Medline: 32308849]
22. King AC, Campero MI, Sheats JL, Castro Sweet CM, Hauser ME, Garcia D, et al. Effects of counseling by peer human advisors vs computers to increase walking in underserved populations: the COMPASS randomized clinical trial. JAMA Intern Med 2020 Nov 01;180(11):1481-1490 [FREE Full text] [doi: 10.1001/jamainternmed.2020.4143] [Medline: 32986075]
23. Singh B, Olds T, Brinsley J, Dumuid D, Virgara R, Matricciani L, et al. Systematic review and meta-analysis of the effectiveness of chatbots on lifestyle behaviours. NPJ Digit Med 2023 Jun 23;6(1):118 [FREE Full text] [doi: 10.1038/s41746-023-00856-1] [Medline: 37353578]
24. Kocielnik R, Langevin R, George JS, Akenaga S, Wang A, Jones DP, et al. Can I talk to you about your social needs? understanding preference for conversational user interface in health. 2021 Presented at: CUI '21: Proceedings of the 3rd Conference on Conversational User Interfaces; July 27-29, 2021; Bilbao (Online), Spain p. 1-10. [doi: 10.1145/3469595.3469599]
25. Garbarino J, Sherman D. High-risk neighborhoods and high-risk families: the human ecology of child maltreatment. Child Dev 1980 Mar;51(1):188-198. [Medline: 7363733]
26. Gundersen C, Ziliak JP. Food insecurity and health outcomes. Health Aff (Millwood) 2015 Nov;34(11):1830-1839. [doi: 10.1377/hlthaff.2015.0645] [Medline: 26526240]
27. Dillman L, Eichner J, Humienny A, Kinsky S, Liang Q, Kwok EY, et al. The impact of supplemental nutrition assistance program (SNAP) enrollment on health and cost outcomes. NEJM Catal Innov Care Deliv 2023 May 17;4(6) [FREE Full text] [doi: 10.1056/CAT.22.0366]
28. Use Memoji on your iPhone or iPad Pro. Apple Inc. URL: https://support.apple.com/en-us/HT208986 [accessed 2023-05-04]
29. Prolific homepage. Prolific. URL: https://www.prolific.com/ [accessed 2024-06-21]
30. Drain A, Sanders EB. A collaboration system model for planning and evaluating participatory design projects. Int J Design 2019;13(3):39-52 [FREE Full text]
31. Hohenstein J, O'Dell D, Murnane EL, Lu Z, Erickson D, Gay G. Enhancing the usability of an optical reader system to support point-of-care rapid diagnostic testing: an iterative design approach. JMIR Hum Factors 2017 Nov 21;4(4):e29 [FREE Full text] [doi: 10.2196/humanfactors.8621] [Medline: 29162559]
32. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
33. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
34. Lewis JR. Measuring perceived usability: the CSUQ, SUS, and UMUX. Int J Hum Comput Interact 2018 Jan 09;34(12):1148-1156. [doi: 10.1080/10447318.2017.1418805]
35. Krippendorff K. Content Analysis: An Introduction to Its Methodology. Thousand Oaks, CA: SAGE Publications; 1980.

36. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
37. Sauro J. Measuring usability: from the SUS to the UMUX-lite. MeasuringU. 2017 Oct 11. URL: https://measuringu.com/umux-lite/ [accessed 2023-05-08]
38. Lewis JR, Utesch BS, Maher DE. Investigating the correspondence between UMUX-LITE and SUS scores. 2015 Presented at: DUXU 2015: Proceedings of the 4th International Conference on Design, User Experience, and Usability: Design Discourse; August 2-7, 2015; Los Angeles, CA. [doi: 10.1007/978-3-319-20886-2_20]
39. Jovanovic M, Baez M, Casati F. Chatbots as conversational healthcare services. IEEE Internet Comput 2021;25(3):44-51. [doi: 10.1109/mic.2020.3037151]
40. Xu L, Sanders L, Li K, Chow JC. Chatbot for health care and oncology applications using artificial intelligence and machine learning: systematic review. JMIR Cancer 2021 Nov 29;7(4):e27850 [FREE Full text] [doi: 10.2196/27850] [Medline: 34847056]
41. Sieck CJ, Sheon A, Ancker JS, Castek J, Callahan B, Siefer A. Digital inclusion as a social determinant of health. NPJ Digit Med 2021 Mar 17;4(1):52 [FREE Full text] [doi: 10.1038/s41746-021-00413-8] [Medline: 33731887]
42. Sanders CK, Scanlon E. The digital divide is a human rights issue: advancing social inclusion through social work advocacy. J Hum Rights Soc Work 2021;6(2):130-143 [FREE Full text] [doi: 10.1007/s41134-020-00147-9] [Medline: 33758780]
43. Meyers N, Glick AF, Mendelsohn AL, Parker RM, Sanders LM, Wolf MS, et al. Parents' use of technologies for health management: a health literacy perspective. Acad Pediatr 2020;20(1):23-30 [FREE Full text] [doi: 10.1016/j.acap.2019.01.008] [Medline: 30862511]
44. Tudor Car L, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng YL, et al. Conversational agents in health care: scoping review and conceptual analysis. J Med Internet Res 2020 Aug 07;22(8):e17158 [FREE Full text] [doi: 10.2196/17158] [Medline: 32763886]
45. You Y, Gui X. Self-diagnosis through AI-enabled chatbot-based symptom checkers: user experiences and design considerations. AMIA Annu Symp Proc 2021 Jan 25;2020:1354-1363 [FREE Full text] [Medline: 33936512]
46. Kocaballi AB, Berkovsky S, Quiroz JC, Laranjo L, Tong HL, Rezazadegan D, et al. The personalization of conversational agents in health care: systematic review. J Med Internet Res 2019 Nov 07;21(11):e15360 [FREE Full text] [doi: 10.2196/15360] [Medline: 31697237]
47. Zhang J, Oh YJ, Lange P, Yu Z, Fukuoka Y. Artificial intelligence chatbot behavior change model for designing artificial intelligence chatbots to promote physical activity and a healthy diet: viewpoint. J Med Internet Res 2020 Sep 30;22(9):e22845 [FREE Full text] [doi: 10.2196/22845] [Medline: 32996892]
48. Kim H, Koh DY, Lee G, Park JM, Lim YK. Designing personalities of conversational agents. 2019 Presented at: CHI EA '19: Proceedings of the Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, UK. [doi: 10.1145/3290607.3312887]
49. Fischer A, Conigliaro J, Allicock S, Kim EJ. Examination of social determinants of health among patients with limited English proficiency. BMC Res Notes 2021 Aug 05;14(1):299 [FREE Full text] [doi: 10.1186/s13104-021-05720-7] [Medline: 34353369]
50. Mokmin NA, Ibrahim NA. The evaluation of chatbot as a tool for health literacy education among undergraduate students. Educ Inf Technol (Dordr) 2021;26(5):6033-6049 [FREE Full text] [doi: 10.1007/s10639-021-10542-y] [Medline: 34054328]
51. Montenegro JL, da Costa CA, Janssen LP. Evaluating the use of chatbot during pregnancy: a usability study. Healthc Anal 2022 Nov;2:100072 [FREE Full text] [doi: 10.1016/j.health.2022.100072]
52. Sezgin E, D'Arcy S. Editorial: voice technology and conversational agents in health care delivery. Front Public Health 2022 May 30;10:887492 [FREE Full text] [doi: 10.3389/fpubh.2022.887492] [Medline: 35712270]
53. Militello L, Sezgin E, Huang Y, Lin S. Delivering perinatal health information via a voice interactive app (SMILE): mixed methods feasibility study. JMIR Form Res 2021 Mar 01;5(3):e18240 [FREE Full text] [doi: 10.2196/18240] [Medline: 33646136]
54. Wu X, Xiao L, Sun Y, Zhang J, Ma T, He L. A survey of human-in-the-loop for machine learning. Future Gener Comput Syst 2022 Oct;135:364-381 [FREE Full text] [doi: 10.1016/j.future.2022.05.014]
55. Sezgin E. Artificial intelligence in healthcare: complementing, not replacing, doctors and healthcare providers. Digit Health 2023 Jul 02;9:20552076231186520 [FREE Full text] [doi: 10.1177/20552076231186520] [Medline: 37426593]
56. Smith A, Wish J, Spong T, Burns K, Cammarata Hamilton R. Feed to heal: leveraging technology for food insecurity referrals in health care. NEJM Catal Innov Care Deliv 2023 Mar 15;4(4) [FREE Full text] [doi: 10.1056/CAT.22.0373]
57. Silva JM, Durden TE, Hirsch A. Erasing inequality: examining discrepancies between electronic health records and patient narratives to uncover perceived stigma and dismissal in clinical encounters. Soc Sci Med 2023 Apr;323:115837. [doi: 10.1016/j.socscimed.2023.115837] [Medline: 36931038]
58. Gold R, Bunce A, Cowburn S, Dambrun K, Dearing M, Middendorf M, et al. Adoption of social determinants of health EHR tools by community health centers. Ann Fam Med 2018 Sep;16(5):399-407 [FREE Full text] [doi: 10.1370/afm.2275] [Medline: 30201636]
59. Rosenbaum S. Data governance and stewardship: designing data stewardship entities and advancing data access. Health Serv Res 2010 Oct;45(5 Pt 2):1442-1455 [FREE Full text] [doi: 10.1111/j.1475-6773.2010.01140.x] [Medline: 21054365]

60. McKnight R, Franko O. HIPPA compliance with mobile devices among ACGME programs. J Med Syst 2016 May;40(5):129. [doi: 10.1007/s10916-016-0489-2] [Medline: 27079578]
61. Bolster-Foucault C, Ho Mi Fane B, Blair A. Structural determinants of stigma across health and social conditions: a rapid review and conceptual framework to guide future research and intervention. Health Promot Chronic Dis Prev Can 2021 Mar;41(3):85-115 [FREE Full text] [doi: 10.24095/hpcdp.41.3.03] [Medline: 33688694]
62. Lima MR, Wairagkar M, Natarajan N, Vaitheswaran S, Vaidyanathan R. Robotic telemedicine for mental health: a multimodal approach to improve human-robot engagement. Front Robot AI 2021 Mar 18;8:618866 [FREE Full text] [doi: 10.3389/frobt.2021.618866] [Medline: 33816568]
63. Affordable connectivity program. Federal Communications Commission. URL: https://www.fcc.gov/acp [accessed 2024-04-16]
64. Glasgow RE, Knoepke CE, Magid D, Grunwald GK, Glorioso TJ, Waughtal J, et al. The NUDGE trial pragmatic trial to enhance cardiovascular medication adherence: study protocol for a randomized controlled trial. Trials 2021 Aug 11;22(1):528 [FREE Full text] [doi: 10.1186/s13063-021-05453-9] [Medline: 34380527]
65. Rapp A, Curti L, Boldi A. The human side of human-chatbot interaction: a systematic literature review of ten years of research on text-based chatbots. Int J Hum Comput Stud 2021 Jul;151:102630. [doi: 10.1016/j.ijhcs.2021.102630]
66. Johnson D, Goodman R, Patrinely J, Stone C, Zimmerman E, Donald R, et al. Assessing the accuracy and reliability of AI-generated medical responses: an evaluation of the Chat-GPT model. Res Sq 2023 Feb 28:rs.3.rs-2566942 (forthcoming) [FREE Full text] [doi: 10.21203/rs.3.rs-2566942/v1] [Medline: 36909565]
67. Milne-Ives M, de Cock C, Lim E, Shehadeh MH, de Pennington N, Mole G, et al. The effectiveness of artificial intelligence conversational agents in health care: systematic review. J Med Internet Res 2020 Oct 22;22(10):e20346 [FREE Full text] [doi: 10.2196/20346] [Medline: 33090118]
68. Shen Y, Heacock L, Elias J, Hentel KD, Reig B, Shih G, et al. ChatGPT and other large language models are double-edged swords. Radiology 2023 Apr;307(2):e230163. [doi: 10.1148/radiol.230163] [Medline: 36700838]
69. Sellers J. Navigating sensitive topics: from food security to insurance coverage. Oncology Nursing News. 2019 Aug 07. URL: https://www.oncnursingnews.com/view/navigating-sensitive-topics-from-food-security-to-insurance-coverage [accessed 2023-05-07]
70. Butler ED, Morgan AU, Kangovi S. Screening for unmet social needs: patient engagement or alienation? NEJM Catal Innov Care Deliv 2020;1(4). [doi: 10.1056/CAT.19.1037]
71. Screening for social needs: guiding care teams to engage patients. American Hospital Association. URL: https://www.aha.org/toolkitsmethodology/2019-06-05-screening-social-needs-guiding-care-teams-engage-patients [accessed 2023-05-07]
72. Abd-Alrazaq A, Safi Z, Alajlani M, Warren J, Househ M, Denecke K. Technical metrics used to evaluate health care chatbots: scoping review. J Med Internet Res 2020 Jun 05;22(6):e18301 [FREE Full text] [doi: 10.2196/18301] [Medline: 32442157]
73. Denecke K, May R. Developing a technical-oriented taxonomy to define archetypes of conversational agents in health care: literature review and cluster analysis. J Med Internet Res 2023 Jan 30;25:e41583 [FREE Full text] [doi: 10.2196/41583] [Medline: 36716093]

Abbreviations

AI: artificial intelligence
API: application programming interface
DAPHNE: Dialog-Based Assistant Platform for Healthcare and Needs Ecosystem
EHR: electronic health record
NCH: Nationwide Children's Hospital
REDCap: Research Electronic Data Capture
UMUX-Lite: Usability Metric for User Experience–Lite

Edited by Y Gong, A Choudhury; submitted 05.02.24; peer-reviewed by GK Gupta, H Miao, A Hassan; comments to author 29.03.24; revised version received 03.05.24; accepted 24.05.24; published 19.07.24.

Please cite as:
Sezgin E, Kocaballi AB, Dolce M, Skeens M, Militello L, Huang Y, Stevens J, Kemper AR
Chatbot for Social Need Screening and Resource Sharing With Vulnerable Families: Iterative Design and Evaluation Study
JMIR Hum Factors 2024;11:e57114
URL: https://humanfactors.jmir.org/2024/1/e57114
doi:10.2196/57114
PMID:

Original Paper

Effects of a Digital Care Pathway for Multiple Sclerosis: Observational Study

Märt Vesinurm[1], MSc; Anna Maunula[2,3], MD; Päivi Olli[4,5], RN; Paul Lillrank[1], PhD; Petra Ijäs[4,5], MSc, MD, PhD; Paulus Torkki[6], PhD; Laura Mäkitie[4,5], MD, PhD; Sini M Laakso[4,5], MD, PhD

[1]Institute of Healthcare Engineering and Management, Department of Industrial Engineering and Management, Aalto University School of Science, Espoo, Finland

[2]Brain Center, Department of Neurology, Hyvinkää Hospital, Hyvinkää, Finland

[3]Translational Immunology Research Program, University of Helsinki, Helsinki, Finland

[4]Brain Center, Department of Neurology, Helsinki University Hospital, Helsinki, Finland

[5]Department of Clinical Neurosciences, University of Helsinki, Helsinki, Finland

[6]Department of Public Health, Faculty of Medicine, University of Helsinki, Helsinki, Finland

Corresponding Author:
Märt Vesinurm, MSc
Institute of Healthcare Engineering and Management
Department of Industrial Engineering and Management
Aalto University School of Science
Maarintie 8
Espoo, 02150
Finland
Phone: 358 503228815
Email: mart.vesinurm@aalto.fi

Abstract

Background: Helsinki University Hospital has developed a digital care pathway (DCP) for people with multiple sclerosis (MS) to improve the care quality. DCP was designed for especially newly diagnosed patients to support adaptation to a chronic disease.

Objective: This study investigated the MS DCP user behavior and its impact on patient education-mediated changes in health care use, patient-perceived impact of MS on psychological and physical functional health, and patient satisfaction.

Methods: We collected data from the service launch in March 2020 until the end of 2022 (observation period). The number of users, user logins, and their timing and messages sent were collected. The association of the DCP on health care use was studied in a case-control setting in which patients were allowed to freely select whether they wanted to use the service (DCP group n=63) or not (control group n=112). The number of physical and remote appointments either to a doctor, nurse, or other services were considered in addition to emergency department visits and inpatient days. The follow-up time was 1 year (study period). Furthermore, a subgroup of 36 patients was recruited to fill out surveys on net promoter score (NPS) at 3, 6, and 12 months, and their physical and psychological functional health (Multiple Sclerosis Impact Scale) at 0, 3, 6, and 12 months.

Results: During the observation period, a total of 225 patients had the option to use the service, out of whom 79.1% (178/225) logged into the service. On average, a user of the DCP sent 6.8 messages and logged on 7.4 times, with 72.29% (1182/1635) of logins taking place within 1 year of initiating the service. In case-control cohorts, no statistically significant differences between the groups were found for physical doctors' appointments, remote doctors' contacts, physical nurse appointments, remote nurse contacts, emergency department visits, or inpatient days. However, the MS DCP was associated with a 2.05 (SD 0.48) visit increase in other services, within 1 year from diagnosis. In the prospective DCP-cohort, no clinically significant change was observed in the physical functional health between the 0 and 12-month marks, but psychological functional health was improved between 3 and 6 months. Patient satisfaction improved from the NPS index of 21 (favorable) at the 3-month mark to the NPS index of 63 (excellent) at the 12-month mark.

Conclusions: The MS DCP has been used by a majority of the people with MS as a complementary service to regular operations, and we find high satisfaction with the service. Psychological health was enhanced during the use of MS DCP. Our results indicate that DCPs hold great promise for managing chronic conditions such as MS. Future studies should explore the potential of DCPs in different health care settings and patient subgroups.

(JMIR Hum Factors 2024;11:e51872) doi:10.2196/51872

KEYWORDS

digital care pathway; multiple sclerosis; patient satisfaction; outcome; patient reported outcome measures; resource usage; telemedicine; digital care; outpatient clinic; quality of life; quality of care; communication; caregiver; chronic condition; strategy; long-term; patient engagement; digital health service

Introduction

Multiple sclerosis (MS) is a chronic inflammatory disease of the central nervous system with a continuously rising prevalence [1]. According to the Finnish National MS registry, there are currently more than 12,000 patients with MS in Finland. The mean age of diagnosis is around 30 years for relapsing-remitting multiple sclerosis (RRMS), a subtype of disease that accounts for 90% of patients at diagnosis. MS is the third most common cause for disability pension in Finland [2] and the yearly cost of MS to society was estimated to be roughly €50,000 (US $54,400) per patient in a retrospective cross-sectional questionnaire study [3]. People with multiple sclerosis (pwMS) require long-lasting follow-up by a multidisciplinary team of neurologists, trained nurses, and rehabilitation specialists, to prevent disability accumulation.

Receiving a MS diagnosis is a major event often raising thoughts about a threat to career, social and family life, and independence. Social media is full of information on MS intertwining facts, assumptions, and feelings [4]. After receiving an MS diagnosis, it is common to experience severe reactions of stress [5]. Pharmacological treatment options are various for MS and patients would prefer shared decision-making in choosing the right one for them. Successful shared decision-making is not possible without reliable information [6]. A recently published study investigated barriers and adoption strategies for the early use of high-efficacy therapies in MS [7]. Further, 1 identified adoption strategy is providing patients with sufficient and reliable information, enabling shared decision-making, and improving treatment adherence [7]. From a psychological function perspective, a MS diagnosis has also been shown to cause fatigue, which is an important MS symptom related to underperformance in studies and in work [8].

Digital and telehealth services as an add-on to conventional treatment have shown to be useful in chronic illnesses. A Swedish study found that 63.9% of patients with chronic obstructive pulmonary disease and chronic heart failure found a digital platform and structured telephone support useful [9]. Glycemic control, kidney function, and lipid parameters in diabetics [10] were significantly lower for patients in a telemedicine intervention group compared to the control group, although it should be noted that the intervention and control groups were not completely balanced for socioeconomic factors in this study. An intervention of telehealth with a digital platform showed a significant lowering of systolic blood pressure in a patient group that had poorly controlled hypertension at recruitment [11]. Digital services as an option for a face-to-face model of treatment in osteoarthritis have been shown to have only 25% of the cost of the face-to-face model [12].

eHealth interventions for pwMS have shown varying results. Recent meta-analyses have shown telehealth-based rehabilitation to be an efficient intervention for pwMS [13,14]. Outside of rehabilitation, there are fewer studies on the effects of digital health care programs on pwMS. Further, 1 study found that medication adherence increased slightly from 0.85 to 0.87 in pwMS by a pharmacy and chronic disease treatment program [15]. However, the program did not affect health care use or hospital visits during follow-ups. A web-based self-management platform did not affect the use of health care services of pwMS, and at the end of this study, the control group had a better quality of life [16].

Early support for coping with MS and the change in life is mainly provided by national patient organizations. Online services are easy to use anonymously, and in some countries, these services are of high quality such as in the United Kingdom [17]. A few digital patient portals for pwMS have been created [18,19], but the patient satisfaction and patient-reported outcome measures (PROMs) of these have not been studied. A digital well-being app has been shown to improve well-being in patients with chronic conditions [20], among them pwMS. More established are data on the impact of web-based education programs and services, as shown in MS community members using the Understanding Multiple Sclerosis course [21]. Education on pathogenesis, treatment options, and self-care induced a change in lifestyle and self-management in 44.1% (247/560) of users, and the maintenance of the changed health behavior was measurable in two-thirds of them after 6 months. Similarly, in a randomized controlled trial, a significant reduction in fatigue was observed in pwMS after using a web-based self-management program supported by an email contact [22]. To our knowledge, there is only 1 previous digital health service as an add-on to standard care, aimed to give information about MS, give psychological support, and support lifestyle changes after an MS diagnosis [23]. This was found useful by pwMS.

We hypothesized that a digital health care intervention during the first years after diagnosis of RRMS could show an impact on the care of MS both from the perspective of the patient and the organization. We report on the user behavior on the MS digital care pathway (DCP) and adopt a case-control design to evaluate the health care use 1-year post diagnosis of the MS DCP group (n=63) and control group (n=114). In addition, we report on the PROMs on physical and psychological functional health using the Multiple Sclerosis Impact Scale (MSIS-29) questionnaire and patient satisfaction using net promoter score (NPS) of a separately recruited subgroup of users (n=36).

Methods

Patient Process for MS in the Neurology Outpatient Clinic

In Finland, MS is diagnosed solely in the public sector. The neurological outpatient clinic at the Meilahti Hospital covers the publicly organized health care of neurological diseases in Helsinki (0.66 million inhabitants, situation December 31, 2022 [24]). The diagnosis is confirmed with neurological examination, magnetic resonance imaging, and cerebrospinal fluid testing [25].

Follow-up of pwMS is organized as preplanned visits and phone appointments with a neurologist and a nurse. After diagnosis, most patients start immunomodulatory therapy requiring laboratory testing before and during treatment. Usually, a magnetic resonance imaging scan is performed after 6 months of medication initiation followed by a doctor's appointment. In addition, most patients have an appointment with a nurse to discuss in more detail the practicalities of medications and follow-up investigations. Newly diagnosed patients are invited to a group meeting with professionals giving information on the disease where patients have the chance to answer questions and meet peers. These events are held twice a year. During the disease course, patients will have preplanned visits to their neurologist at 6- to 12-month intervals, with additional visits if needed. In addition, patients will contact a nurse if they have new symptoms or questions between the visits. In 2022, Meilahti outpatient clinic had a care relationship with 1115 pwMS.

Contents of the DCP

The MS DCP was developed jointly by the staff and patients of the Helsinki University Hospital (HUS) neurology outpatient clinic in 2018-2020. The idea for the DCP started from the common feedback from pwMS that more support is needed after receiving a diagnosis and that there is health care practitioner-dependent variation in the information given about support services. In addition, pwMS found it difficult to get in touch with the outpatient clinic. The DCP aimed to (1) act as a patient education platform for pwMS, providing up-to-date and reliable information on MS, (2) standardize the information and support for all patients, (3) support the adaptation to a chronic illness of patients newly diagnosed with MS, (4) provide easier contact with the outpatient clinic, and (5) enable the timely use of resources for the patients in most need of health care appointments.

Thus, it was developed mainly with the newly diagnosed pwMS in mind. The DCP operates on the MyPath platform in the Digital Health Village, a digital service provided by all 5 University Hospitals in Finland. The Digital Health Village supports the traditional care paths with digital services and consists currently of 33 hubs, 400 DCPs, and 9 online knowledge centers [26]. The use of DCP is voluntary, does not rule out face-to-face contact at the outpatient clinic, and is free of charge for the patients. The service may be used by a mobile device or computer with an internet connection. The DCP is available 24/7 during the whole care relationship at the outpatient clinic.

The DCP contains information about MS, its treatment, and relapses, and provides support for the adaptation to living with a chronic disease, cessation of smoking, starting physical exercise, and coping at work. The adaptation training course of the DCP provides support to the patient after receiving the diagnosis. The DCP also contains questionnaires on symptoms and functional abilities which are encouraged to be filled in before a doctor's appointment. In addition, the patient has the possibility of sending a message to the health care professional (HCP) at the neurological outpatient clinic, and a video appointment can be launched as well. Patients can store their individual history of MS including time of diagnosis, used medications, social reimbursements, and rehabilitations received on the DCP for their own purposes, for example, used in other contacts with health care. The information content of the DCP is given on videos by professionals and patients sharing their expertise and experiences, on written text and self-assessment questionnaires, and by linking to other reliable websites. The specific contents sections of the DCP are as follows:

1. Welcome to the DCP: information on the DCP and its contents. Both written text and a video (1:06 minutes) with a neurologist introducing the DCP's different sections and how the DCP works.
2. My MS: a questionnaire about the relevant background information on the DCP user. Questions related to health conditions, ability to work, social benefits such as unemployment benefits, relevant health habits, possible rehabilitation, and how long the special reimbursement for MS medications is valid. This is meant to make the doctor's appointment smoother. The questionnaire is not standardized. It was developed for the DCP by the staff and patients of Meilahti Hospital.
3. Symptom questionnaire: a questionnaire about neurological symptoms often associated with MS. For all listed symptoms patients report whether they are asymptomatic, if the symptom is an old one, if an old symptom has become stronger, or if it is a completely new symptom. Patients are also asked to estimate how long they can walk, if they have problems with their medication, and what they would like to talk about with their doctor. The purpose of the questionnaire is for patients to fill it out before doctors' appointments to make the appointments smoother, personalized, and more purposeful. The questionnaire is not standardized. It was developed for the DCP by the staff and patients of Meilahti Hospital.
4. Information on MS: general information on MS. Links to reliable sources of information. A video (3:23 minutes) with two pwMS talking about finding information on the internet about MS, urging to be critical of the found information, what are reliable sources of information, and finding peers on online platforms.
5. Medication: information on medical treatment of MS. Links to reliable sources of information. Information on treatment of relapses with corticosteroid pulse therapy.
6. Adaptation training course: an adaptation training course to support the patient after receiving the diagnosis. The adaptation training course was modified to fill the needs of patients with MS from a template created by a psychologist

for all DCPs at HUS and contains information on the psychological process using the following steps:

- Step 1: shock. The first phase of shock and the different ways individuals will react. There are tools for how to cope at this stage and information on where to seek help from an HCP. A video (1:53 minutes) with two pwMS talking about how to accept the diagnosis.
- Step 2: what has happened? The stage where the individual will start to confront the thought of MS. There are rehearsals to support this stage.
- Step 3: how can I cope with MS? The stage where the individual will start to accept the situation. Encouragement is given to start seeking information on MS.
- Step 4: a part of life: the stage where MS becomes a part of life, and the individual's thoughts are turning to the future. How to take care of one's resources. A video (1:40 minutes) with two pwMS talking about peer support and different forms of peer support.

7. Support for working life: information on the ability to work, and on how to maintain one's ability to work. Links to reliable sources of information are provided.
8. Physical exercise and MS: what the importance of physical exercise is in MS. A video (3:13 minutes) of two pwMS talking about physical exercise and how to find the right sport. A link to another DCP which is developed to help people start exercising is provided.
9. Quit smoking: the negative impact of smoking on MS and information on how to obtain support with quitting.
10. Give feedback: questionnaires on patient satisfaction and response to the DCP. Links to technical and content feedback are provided.
11. Send a message: the patient has the possibility of sending a message to the HCP at the neurological outpatient clinic.

Population and Study Design

All patients under the follow-up for MS in the neurology outpatient clinic at Meilahti Hospital were offered access to the DCP. Data on DCP user behavior were collected on all users of the DCP up to the end of 2022. Data was collected retrospectively and is descriptive in nature.

To evaluate the differences in health care use, a retrospective case-control study was adopted. A case group (n=63) was constructed of the patients who had logged into the DCP at least once and fulfilled the inclusion and exclusion criteria below. A control group (n=112) was constructed of those who had not logged into the DCP but fulfilled the criteria. Data are analyzed with statistical methods.

The inclusion criteria for the health care use part of this study were (1) diagnosed between January 1, 2020, and December 31, 2022, (2) aged ≥18 years, (3) a diagnosis of RRMS, and (4) a recorded Expanded Disability Status Scale (EDSS) score.

The exclusion criteria for this part of this study were (1) a diagnosis of primary-progressive multiple sclerosis (PPMS) and (2) a diagnosis of secondary-progressive MS.

Recently diagnosed pwMS were recruited to a prospective substudy on PROMs from February 1, 2021, to August 30, 2022. The functional impact of MS and patient satisfaction with DCP was assessed, which required filling out surveys. Altogether 118 patients were given access to the MS DCP during the recruitment period, of whom 72 gave initial consent for this study. Of these, 35 were excluded due to a longer time from diagnosis or disease subtype of PPMS, and 1 patient decided to discontinue this study before the end of 12 months. Thus, the final study cohort of PROM consisted of 36 patients. The changes in PROMs are analyzed descriptively.

The inclusion criteria for this part of this study were (1) informed consent, (2) age ≥18 years, (3) ≤3 years from the diagnosis of RRMS at the time of consent, (4) fluent in Finnish language, and (5) experience in using and access to a computer or a smart device.

The exclusion criteria for this part of this study were (1) a diagnosis of PPMS, (2) a diagnosis of secondary-progressive MS, and (3) a severe psychiatric disease (eg, severe depression or a psychotic disease).

It should be noted that while the 3 samples (user behavior, health care use, and PROMs) used in this study are largely overlapping, they are not completely the same. Newly diagnosed patients were emphasized because the principal intended impact of the DCP is focused on newly diagnosed patients.

Measures

User behavior on the DCP was evaluated by descriptive statistics on the number of users, the number of logins, the timing of logins, and messages sent. These outcomes were chosen as simple metrics of how well the DCP is adopted by pwMS and at what point in their patient journey they use the service.

The health care use between the intervention group and the control group was evaluated using 7 simple quantitative measures: number of physical nurse contacts, number of physical doctors' contacts, number of remote nurse contacts, number of remote doctors' contacts, number of other service contacts (including occupational therapy, neuropsychological services, physiotherapy services, rehabilitation services, speech therapy, social worker services, and nutritionist services), number of emergency department (ED) visits, and number of inpatient days. For the first 5, only contacts to the neurological unit were considered as these were assumed to be related to MS. For ED and inpatient days no such filter was included. These measures were selected as the authors believed these can be used to accurately evaluate resource use overall. It should be noted that nonscheduled calls and messages to the MS nurses were not registered in the electronic health record (EHR) as separate contacts and were thus left out of the analysis.

Several factors were considered potential confounders in statistical analyses related to health care use. Previous research has found gender disparities in health care use for pwMS [27]. Age at diagnosis was considered, although previous research has not found a clear association between age at diagnosis and health care use [28]. The EDSS score was considered as a proxy for disabilities, which are known to be associated with higher health care use [29]. Whether or not the pwMS experienced at least one relapse during this study period was considered as a binary variable for disease activity, which is known to be a predictor of health care use [30].

The MSIS-29 [31] was used for measuring the physical and psychological impact of MS during a 1-year follow-up in a prospective substudy cohort. The MSIS-29 self-report questionnaire allows for a detailed evaluation of a patient's functional health and well-being, encompassing physical limitations, emotional well-being, and overall quality of life. The Finnish version of the MSIS-29 has also been found valid and reliable in previous research [32]. The MSIS-29 score was collected from the patients of DCP at the 0-month, 3-month, 6-month, and 12-month marks. The scores were converted into a scale from 0 to 100 for ease of interpretation. A change of 7-8 points for the physical score [33-36] and a change of 4-6 points for the psychological score [35,36] is considered clinically significant. A higher score indicates higher perceived disability.

The NPS was used to assess the satisfaction of the DCP users at 3 key intervals: 3, 6, and 12 months. NPS is based on a simple question of whether the patient or user would recommend the service to others (rated from 0 to 10), and it has been widely adopted from business to health care [37]. NPS enables the comparison of patient satisfaction between different kinds of services and during service development. NPS is calculated as the percentage of promoters (rated 9 or 10) minus the percentage of detractors (rated 0 to 6). For reference, a previous study on a digital solution for MS care considered an NPS of over 0 "good," over 20 "favorable," and over 50 "excellent" [38].

Data Extraction

The quantitative data on user behavior (number and timing of logins, and number of messages) describing the use of the DCP were extracted from Power BI (Microsoft Corp) reports, and the patient information was gathered from EHRs Uranus (Consultants to Government and Industry Incorporated) and Apotti (Epic Systems Corporation) reporting tools. Patient-reported outcomes (MSIS-29 and NPS) were filled electronically in the DCP platform.

Statistical Analysis

Statistical analyses were performed with RStudio software (version 2022.12.0+353; Posit Software, PBC). Demographic data were presented in means and SDs as well as medians with minimum-maximum range and IQR. The Mann-Whitney *U* test was used in comparing the means of continuous variables, not requiring an assumption of normality [39]. Similarly, the medians of continuous variables were compared using the nonparametric Mood median test. A *P* value of <.05 was considered statistically significant.

Multivariable linear regression models [40] were used to estimate the effect of the MS DCP on health care use in the presence of confounders. Further, 3 models with different confounder combinations were used to estimate the effect of the MS DCP on support service use. In addition, models with ED visits, inpatient days, and total elective visits were considered.

Ethical Considerations

This study has been approved by the HUS review board (HUS/2059/2020). The data processing practices followed the European Union Data Protection Directive Rules. This paper conforms with the applicable STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement.

Results

User Behavior on the MS DCP During 2020-2022

The DCP for MS was launched in March 2020 and has since had a steady flow of new users (Figure 1). Overall, as of the end of 2022, a total of 225 patients have had the option to use the service, of which 79.1% (178/225) have logged in to the service. This equals to 16% (178/1115) of all pwMS under follow-up in the Meilahti outpatient clinic. On average, a user of the MS DCP has logged in to the portal 7.4 times. As is evident from Figure 2, most sessions on the DCP take place shortly after the initiation of the access. Over the 2-year observation period starting from the initiation of the service, 18.1% (296/1635) of logins take place during the first month, 72.29% (1182/1635) of logins happen during the first year, and 27.7% (453/1635) during the second year. A minor peak in logins is seen 6 months after the launch of the service, coinciding with a common follow-up visit time. The data on which contents the patients have engaged with in the DCP is not available.

During the observation period, patients using the DCP have sent 1213 messages (average of 6.8 messages/patient) to HCP, usually MS nurses, who have answered the messages roughly in 1 working day. On the other hand, questionnaires on symptoms and functional abilities have been used on average once per patient, even though they were advised to fill it in before every appointment with the doctor.

Figure 1. Cumulative number of patients (circle) and number of new patients (triangle) in the Neurology Outpatient Clinic using the DCP shown from March 2020 until December 2022. DCP: digital care pathway.

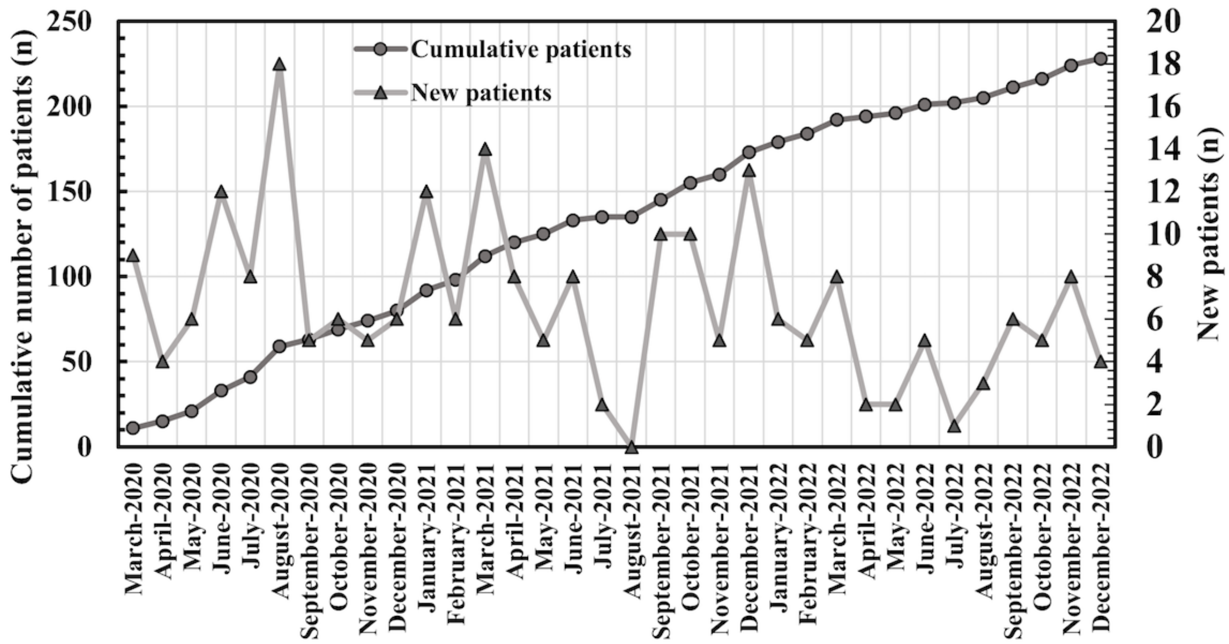
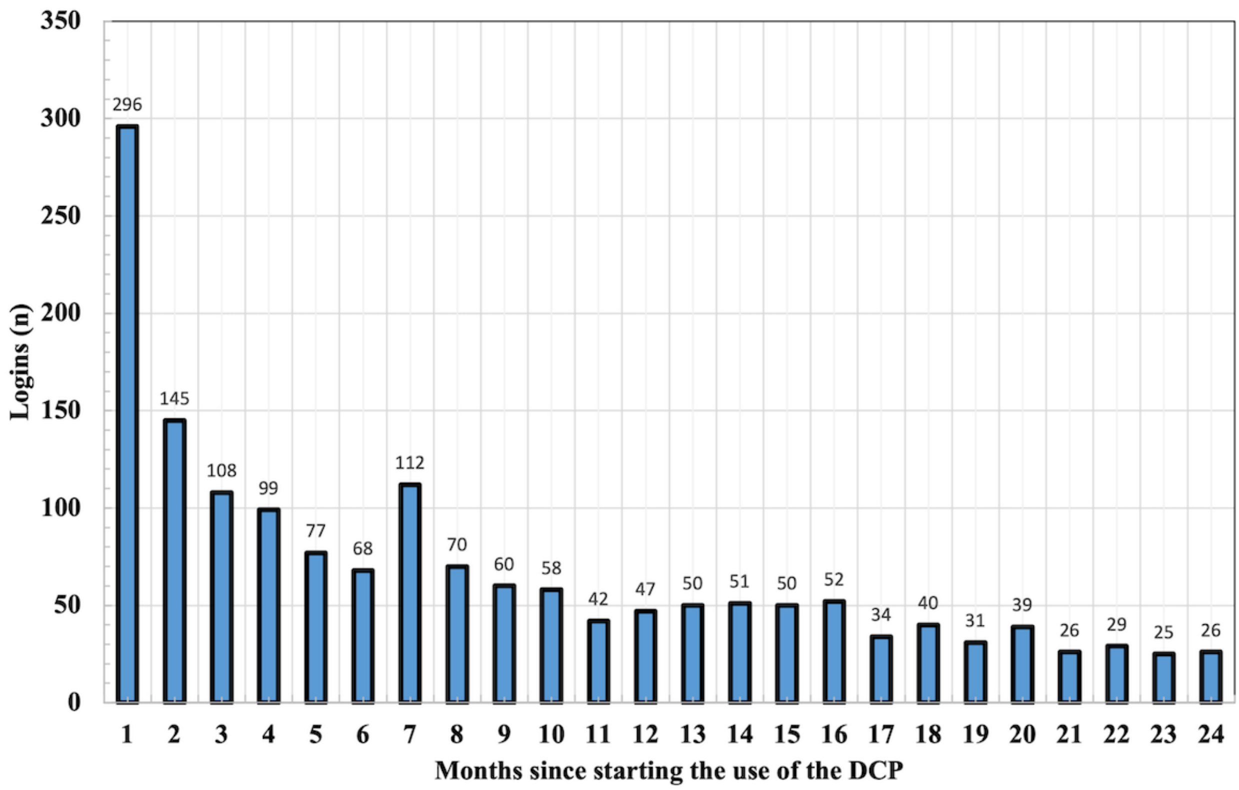


Figure 2. Summative number of logins to the DCP for all pwMS under follow-up in the Neurology Outpatient Clinic. The data are calculated from the time each individual first entered the DCP until 24 months. DCP: digital care pathway; pwMS: people with multiple sclerosis.



Health Care Use

Data on health care use was retrieved from EHRs for a total of 175 patients who fulfilled the inclusion criteria during the observation period. They were separated into an MS DCP user group (n=63) and a control group (n=112). The MS DCP group was made up of the patients who had logged into the DCP at least once. The rest of the patients made up the control group. The groups had similar gender distributions of 73% (46/63) and 70% (79/112) of women, respectively, which well represents the overall female-to-male ratio in the prevalence of MS of roughly 2.35:1 [41]. The mean age was 34.7 (SD 9.83) years in the DCP group and 35.5 (9.76) years in the control group with no significant difference between the two. Similarly, the mean EDSS scores were 1.56 (SD 1.26) and 1.63 (SD 1.19), respectively, with no significant differences. The control group had a higher proportion of pwMS who had relapsed at least once with a total of 13 (N=112, 11.6%) patients, in comparison to only 4 (N=63, 6.4%) in the MS DCP group. This difference, however, was not statistically significant. The characteristics of the MS DCP group and the control group are presented in Table 1.

There were no differences between the MS DCP group and the control group in physical doctor appointments (median 2 visits for both groups), remote doctor appointments (median 8 and 7 contacts, respectively), physical nurse contacts (median 1 visit for both groups), and remote nurse contacts (median 0 visits for both groups; Table 2). Similarly, there was no difference between the groups in emergency care visits (median 0 for both groups) nor inpatient days (median 0 for both groups). A statistically significant difference between the groups was found in other service contacts with the MS DHP group having a median of 1 (minimum-maximum: 0-27) and a mean of 2.57 (SD 4.10) in other service contacts, in contrast to the control group with a median of 0 (minimum-maximum: 0-25) and mean of 0.58 (SD 2.47).

To test for the effects of possible confounders, linear regression models were used (Table 3). The first 3 models tested the association of the DCP and support service use in the presence of confounders. Model 1 considered only the MS DCP, model 2 considered the DCP, gender, and age, and model 3 considered the DCP, gender, age, EDSS score, and relapses. The MS DCP was found to be associated with higher other service use (1.99, 1.95, and 2.05 visits) in all models tested. In addition, linear models were used to confirm no association of the MS DCP with ED visits, inpatient days, and the total number of elective visits (physical or remote, nurses or doctors). No association was found between them.

Table 1. Patient characteristics of the health care use study and PROM[a] cohorts.

Characteristics	MS[b] DCP[c] (n=63)	Control (n=112)	Total (n=175)	PROM subgroup (n=36)
Gender, n (%)				
Female	46 (73)	79 (70.5)	125 (71.4)	24 (66.7)
Male	17 (27)	33 (29.5)	50 (28.6)	12 (33.3)
Age at diagnosis (years)				
Mean (SD)	34.7 (9.83)	35.3 (9.76)	35.1 (9.76)	36.2 (9.9)
18-24, n (%)	11 (17.5)	10 (8.9)	21 (12.0)	3 (8.33)
25-34, n (%)	23 (36.5)	51 (45.5)	74 (42.3)	13 (36.1)
35-44, n (%)	18 (28.6)	33 (29.5)	51 (29.1)	14 (38.9)
45-54, n (%)	9 (14.3)	11 (9.8)	20 (11.4)	4 (11.1)
55 and older, n (%)	2 (3.2)	7 (6.3)	9 (5.1)	2 (5.6)
EDSS[d] score				
Mean (SD)	1.56 (1.26)	1.63 (1.19)	1.60 (1.22)	N/A[e]
Median (minimum-maximum; IQR)	1.5 (0-5.5; 1)	1.75 (0-5; 1)	1.5 (0-5.5; 1)	N/A
Relapse occurred (yes or no)				
Total, n (%)	4 (6.4)	13 (11.6)	17 (9.7)	N/A

[a]PROM: patient-reported outcome measure.

[b]MS: multiple sclerosis.

[c]DCP: digital care pathway

[d]EDSS: expanded disability status scale.

[e]N/A: not applicable.

Table 2. Difference in health care use between the MS[a] DCP[b] group and the control group.

Resource	MS DCP (n=63)	Control (n=112)	*P* value
Physical doctor appointments			
Mean (SD)	1.68 (0.78)	1.64 (0.90)	.75
Median (minimum-maximum; IQR)	2 (0-4; 1)	2 (0-4; 1)	.20
Remote doctor contacts[c]			
Mean (SD)	7.60 (3.33)	8.0 (4.55)	.95
Median (minimum-maximum; IQR)	8 (0-20; 3.5)	7 (0-26; 5)	.05
Physical nurse appointments			
Mean (SD)	1.48 (2.84)	1.16 (2.11)	.90
Median (minimum-maximum; IQR)	1 (0-13; 1)	1 (0-12; 1)	.29
Remote nurse contacts			
Mean (SD)	0.11 (0.41)	0.10 (0.46)	.67
Median (minimum-maximum; IQR)	0 (0-2; 0)	0 (0-4; 0)	.66
Other service contacts[d]			
Mean (SD)	2.57 (4.10)	0.58 (2.47)	<.001
Median (minimum-maximum; IQR)	1 (0-27; 3)	0 (0-25; 0)	<.001
Emergency department visits			
Mean (SD)	0.51 (1.0)	0.69 (2.09)	.88
Median (minimum-maximum; IQR)	0 (0-4; 1)	0 (0-15; 1)	.92
Inpatient days			
Mean (SD)	0.22 (0.71)	0.38 (1.23)	.95
Median (minimum-maximum; IQR)	0 (0-4; 0)	0 (0-6; 0)	.69

[a]MS: multiple sclerosis.

[b]DCP: digital care pathway.

[c]Includes asynchronous contacts (ie, administrative-like work, with no direct contact with the patient).

[d]occupational therapy, neuropsychological services, physiotherapy services, rehabilitation services, speech therapy, social worker services, and nutritionist services.

Table 3. Regression analysis of the effect of the DCP[a] use and confounders on health care use (N=175).

Variables	DV[b]: other service use[c]			DV: emergency department visits	DV: inpatient days	DV: all elective[d]
MS[e] DCP	1.99[f] (0.50)	1.95[f] (0.49)	2.05[f] (0.48)	–0.13 0.28)	–0.12 (0.16)	0.11 (0.81)
Gender: female	N/A[g]	0.39 (0.52)	0.21 (0.51)	0.13 (0.30)	0.18 (0.17)	–0.96 (0.87)
Age at diagnosis (years)	N/A	–0.05[h] (0.02)	–0.07[i] (0.02)	–0.03[h] (0.01)	–0.01 (0.01)	–0.14[f] (0.04)
EDSS[j] score	N/A	N/A	0.63[i] (0.19)	0.13 (0.11)	0.11 (0.07)	0.73[h] (0.33)
Relapses	N/A	N/A	1.21 (0.79)	1.08[h] (0.45)	0.95[f] (0.27)	3.09[h] (1.33)
Constant	0.58 (0.30)	2.20[h] (0.98)	1.71 (0.96)	1.19[h] (0.55)	0.36 (0.33)	15.2[f] (1.63)
Adjusted R^2	0.080	0.10	0.15	0.03	0.08	0.08

[a]DCP: digital care pathway.

[b]DV: dependent variable.

[c]Occupational therapy, neuropsychological services, physiotherapy services, rehabilitation services, speech therapy, social worker services, and nutritionist services.

[d]All elective: sum of remote and physical nurses' and doctors' appointments.

[e]MS: multiple sclerosis.

[f]*P*<.001.

[g]N/A: not applicable.

[h]*P*<.05.

[i]*P*<.01.

[j]EDSS: Expanded Disability Status Scale.

Patient-Reported Outcomes

PROMs were assessed in a cohort of 36 patients (Table 1), with an initial response rate of 97% (35/36). The MSIS-29 looks for clinically important changes in functional health. A change of 7-8 points for the physical score [33-36] and a change of 4-6 points for the psychological score [35,36] is considered clinically significant. There was a large variability between the MSIS-29 scores reported (Tables 4 and 5). Over the course of this study period, there was no clinically significant change in the median physical MSIS-29 scores, but a clinically significant change was observed for the psychological MSIS-29 scores between the 3-month and 6-month points (an improvement from 22.2 to 16.7). The score, however, rose at the 12-month mark. Of the patients who reported MSIS-29 scores for the 0- and 12-month marks, the change in the median physical score was +3.3, and in the psychological score –3.7. The response rate declined during the year-long study period from 97.2% (35/36) to 47.2% (17/36).

An NPS index was calculated at 3 time points: 3, 6, and 12 months after first login to the DCP (Tables 4 and 5). The NPS at the 3-month mark was 21 indicating a "favorable" reception. At the 6-month mark increased to 30 and at the 12-month mark up to 63 passing the "excellent" limit. The response rate dropped from 80.5% (29/36) down to 44.4% (16/36) at 12 months.

Table 4. Patient-reported outcome measures in this study subcohort (N=36).

Measure	Score, median (minimum-maximum; IQR)	Response rate, n (%)
MSIS-29[a] physical score		
0 months	11.7 (0-78.3; 17.5)	35 (97)
3 months	11.7 (0-73.3; 19.2)	31 (86)
6 months	10.8 (0-68.3; 20.8)	26 (72)
12 months	15.0 (0-86.7; 31.7)	17 (47)
MSIS-29 psychological score		
0 months	22.2 (7.4-85.2; 31.5)	35 (97)
3 months	22.2 (0-63.0; 18.95)	31 (86)
6 months	16.7 (3.7-44.4; 17.6)	26 (72)
12 months	18.5 (0-59.3; 22.2)	17 (47)

[a]MSIS-29: Multiple Sclerosis Impact Scale.

Table 5. Patient-reported outcome measures in this study subcohort (N=36).

Measure	Score, n	Response rate (%)
Net promoter score index		
3 months	21	29 (81)
6 months	30	23 (64)
12 months	63	16 (44)

Discussion

Here, we described the contents of MS DCP developed in HUS, user behavior on the DCP, differences in health care use between pwMS using the DCP and those who do not, presented results on changes in perceived impact of MS on both physical and psychological functional health of the MS DCP users, and showed patient satisfaction with the service. We show that the DCP has seen a high adoption rate of 79% (178/225), with most logins taking place within the first year after diagnosis (1182/1635, 72.29%). We find a significant increase in other services use and no other differences in the use of services. There was no clinically significant change in the physical MSIS-29 score and while the psychological score oscillated, a clinically significant change was observed between months 3 and 6. Patient satisfaction as measured with NPS was high and improved during this study period (from 21 “favorable” to 63 “excellent”).

The pwMS in this study were young (Table 1), and age is known to be associated with activity in telemedicine services [42]. Thus, patients with RRMS appear to be an exceptionally well-suited patient group to address through digital health care, as the disease is usually diagnosed at an early age [43] and the intensity of the care required is often highest in the beginning when patients need the most support [44]. The impact of a web-based educative and cognitive training program has been demonstrated in a group of patients with MS in the United States and Germany having depression. Both quality of life and psychological health improved, more so with a therapist supporting the training [45]. In this context, the DCP we describe in this study has most content targeted exactly at the early stages of follow-up.

The results also indicate that those using the DCP find their way into available other (supportive and rehabilitative) services significantly more often than those who do not, with on average 2.05 (SD 0.48) more visits during the year after diagnosis than in the control group. This association stayed strong even when accounting for possible confounders: gender, age, EDSS score, and relapses (Table 3). Assuming the cost-effectiveness of rehabilitative and preventative services, in lieu of overtreatment, this could be assumed to be resource-efficient and a positive development both from the perspective of the health care organization and the patient. However, to demonstrate this, a longer follow-up study would be required to show that pwMS using support services either use less health care or have better health outcomes in the future. We found no change in other measures, which aligns well with the fact that the DCP was intended to be a complementary service rather than replacing some part of the regular patient with the MS pathway of the neurological clinic. Finally, no difference in ED or inpatient days was found. All these findings remained when accounting for confounders (Table 3).

Our findings of the DCP being resource-efficient are in line with previous literature on other chronic diseases from Finland [46,47] and elsewhere [48]. In addition, we assume that our analysis underestimates the resource effects of the DCP because, although the health care use data lacked data on most unscheduled remote nurse contacts, unpublished results of analysis in HUS suggest that the nurse time needed for answering 1 patient phone call is 8 times longer than that needed for answering a message. The MS DCP supports messaging as the means of interaction.

A significant improvement in the psychological score was observed between the third and sixth months. Although this later increased, this is an initial indication that there may have been less psychological burden on the patients, which could be attributed to the DCPs educative role or possible time passing and the patient’s coming to terms with the disease. The response rate reduced a lot from 97% (35/36) at 3 months to 47% (17/36) at 6 months. Although this gives uncertainty to interpretation, there is also a possibility that those pwMS who had the biggest improvement dropped out of this study and stopped logging onto the DCP as it had already done its job and supported them with the start of their MS patient journey. Over the 0-month, 3-month, 6-month, and 12-month marks, there were no significant changes [33,34] in the median MSIS-29 physical scores of the patients with MS DCP indicating that there is at least no adverse effect of using the DCP, which was expected.

Patient satisfaction with care is central to the success of any health care delivery model. We used the NPS as a measure of satisfaction with the DCP. Overall feedback was generally positive, indicating a successful adoption of the DCP, with the final NPS index being 63 “excellent.” Again, however, the response rate dropped during this study period from 81% (29/36) down to 44% (16/36), which together with the increasing trend could be an indication that those who were most satisfied with the service kept coming back to it. The MS DCP has several attributes that can be thought of as contributing to patient satisfaction such as patients having access according to their own needs and returning to the information content whenever there is a need for revision. Previous literature has come to similar conclusions [46].

Our study’s primary strength was the comprehensive examination of the DCP’s impact on multiple dimensions—user behavior measures on the DCP, health care use with a case-control design, and PROMs on the impact of MS on functional health, and patient satisfaction. Specifically, we present data on all users of the DCP and pwMS diagnosed within a defined period as our sample and follow all these pwMS over a period of 1 year. In addition, the PROMs were collected longitudinally at several time points which further strengthens our study. Considering these aspects, this study provided a well-rounded view of the MS DCP. However, there are some significant limitations. First, the cohorts studied in the different parts of this study do not completely overlap and have somewhat different inclusion and exclusion criteria limiting the possibilities of what can be said by combining insights from the groups. Second, it should be noted that although we found a very strong connection between the MS DCP and the use of other services, it is possible that some mediating variable outside of our analysis (ie, the probability of an individual seeking support) could be affecting the results. This fact also introduces some selection bias into our cohort. Third, the section of this study dealing with PROMs had a large reduction in response rates over this study period. While this can be partially explained by a natural attrition rate in longitudinal studies, the sharp decrease in responses may have introduced bias into the results, with those reporting the highest satisfaction levels being more likely to continue engagement with the DCP.

Future studies should investigate whether there are specific patient subgroups, for example, younger patients or more digitally fluent patients that are more likely to engage with and benefit from the MS DCP. Another area for future research is to explore how the cost-effectiveness of the DCP compares to traditional care models in different health care settings, such as other DCP studies from the Finnish context [46].

In conclusion, digital service is a resource-wise way of importing information about MS and support in the early adaptation phase of the disease to pwMS. Our results warrant further research into the long-term effects of DCPs as complementary solutions to existing pathways in MS, and other chronic conditions as well.

Conflicts of Interest

AM discloses the following conflicts of interest: Congress expenses Roche, Merck, AbbVie, Biogen; fees for lecture: Merck. SML discloses the following conflicts of interest: lecture fees Argenx, Biogen, Janssen, Merck, Novartis, Roche; congress expenses Merck, Novartis; advisory fee Argenx, Novartis, Roche, Sanofi, UCB Pharma; investigator for the clinical study Clarion (Merck) and subinvestigator for the clinical study Fenhance (Roche). LM discloses the following conflicts of interest: eCongress fee AbbVie.

References

1. Walton C, King R, Rechtman L, Kaye W, Leray E, Marrie RA, et al. Rising prevalence of multiple sclerosis worldwide: insights from the Atlas of MS, third edition. Mult Scler 2020;26(14):1816-1821 [FREE Full text] [doi: 10.1177/1352458520970841] [Medline: 33174475]
2. Ala-Kauhaluoma M, Laurila H. MS-tautia sairastavat suomalaiset-työssä selviytyminen ja kuntoutus. Finland: Suomen MS-liitto; 2008.
3. Ruutiainen J, Viita A, Hahl J, Sundell J, Nissinen H. Burden of illness in multiple sclerosis (DEFENSE) study: the costs and quality-of-life of Finnish patients with multiple sclerosis. J Med Econ 2016;19(1):21-33 [FREE Full text] [doi: 10.3111/13696998.2015.1086362] [Medline: 26360615]
4. Altunisik E, Firat YE, Kiyak Keceli Y. Content and quality analysis of videos about multiple sclerosis on social media: the case of YouTube. Mult Scler Relat Disord 2022;65:104024. [doi: 10.1016/j.msard.2022.104024] [Medline: 35841755]
5. Chalfant AM, Bryant RA, Fulcher G. Posttraumatic stress disorder following diagnosis of multiple sclerosis. J Trauma Stress 2004;17(5):423-428. [doi: 10.1023/B:JOTS.0000048955.65891.4c] [Medline: 15633921]
6. Kuusisto H, Apila S, Saranto K. Information provision and quality. a pilot study on shared decision-making in multiple sclerosis. Stud Health Technol Inform 2022;295:179-182. [doi: 10.3233/SHTI220691] [Medline: 35773837]
7. Singer BA, Feng J, Chiong-Rivero H. Early use of high-efficacy therapies in multiple sclerosis in the United States: benefits, barriers, and strategies for encouraging adoption. J Neurol 2024;271(6):3116-3130 [FREE Full text] [doi: 10.1007/s00415-024-12305-4] [Medline: 38615277]
8. Chang YT, Kearns PKA, Carson A, Gillespie DC, Meijboom R, Kampaite A, et al. Network analysis characterizes key associations between subjective fatigue and specific depressive symptoms in early relapsing-remitting multiple sclerosis. Mult Scler Relat Disord 2023;69:104429 [FREE Full text] [doi: 10.1016/j.msard.2022.104429] [Medline: 36493562]
9. Barenfeld E, Fuller JM, Wallström S, Fors A, Ali L, Ekman I. Meaningful use of a digital platform and structured telephone support to facilitate remote person-centred care—a mixed-method study on patient perspectives. BMC Health Serv Res 2022;22(1):442 [FREE Full text] [doi: 10.1186/s12913-022-07831-8] [Medline: 35379247]
10. Calikoglu F, Bagdemir E, Celik S, Idiz C, Ozsarı H, Issever H, et al. Telemedicine as a motivational tool to optimize metabolic control in patients with diabetes in Turkey: a prospective, randomized, controlled telediab trial. Telemed e-Health 2023;29(4):518-530. [doi: 10.1089/tmj.2022.0028] [Medline: 36067018]
11. Chen Y, Hung C, Huang C, Lee J, Yu J, Ho Y. The impact of synchronous telehealth services with a digital platform on day-by-day home blood pressure variability in patients with cardiovascular diseases: retrospective cohort study. J Med Internet Res 2022;24(1):e22957 [FREE Full text] [doi: 10.2196/22957] [Medline: 35006089]

12. Ekman B, Nero H, Lohmander LS, Dahlberg LE. Costing analysis of a digital first-line treatment platform for patients with knee and hip osteoarthritis in Sweden. PLoS One 2020;15(8):e0236342 [FREE Full text] [doi: 10.1371/journal.pone.0236342] [Medline: 32785226]
13. Rintala A, Hakala S, Paltamaa J, Heinonen A, Karvanen J, Sjögren T. Effectiveness of technology-based distance physical rehabilitation interventions on physical activity and walking in multiple sclerosis: a systematic review and meta-analysis of randomized controlled trials. Disabil Rehabil 2018;40(4):373-387. [doi: 10.1080/09638288.2016.1260649] [Medline: 27973919]
14. Di Tella S, Pagliari C, Blasi V, Mendozzi L, Rovaris M, Baglio F. Integrated telerehabilitation approach in multiple sclerosis: a systematic review and meta-analysis. J Telemed Telecare 2020;26(7-8):385-399. [doi: 10.1177/1357633X19850381] [Medline: 31132909]
15. Groeneweg M, Forrester SH, Arnold B, Palazzo L, Zhu W, Yoon P, et al. One-year outcomes of an integrated multiple sclerosis disease management program. J Manag Care Spec Pharm 2018;24(5):458-463 [FREE Full text] [doi: 10.18553/jmcp.2018.24.5.458] [Medline: 29694287]
16. Miller DM, Moore SM, Fox RJ, Atreja A, Fu AZ, Lee J, et al. Web-based self-management for patients with multiple sclerosis: a practical, randomized trial. Telemed e-Health 2011;17(1):5-13. [doi: 10.1089/tmj.2010.0133] [Medline: 21214498]
17. Newly diagnosed. MS Society. URL: https://www.mssociety.org.uk/living-with-ms/newly-diagnosed [accessed 2024-06-07]
18. Nittas V, Zecca C, Kamm CP, Kuhle J, Chan A, von Wyl V. Digital health for chronic disease management: an exploratory method to investigating technology adoption potential. PLoS One 2023;18(4):e0284477 [FREE Full text] [doi: 10.1371/journal.pone.0284477] [Medline: 37053272]
19. Van Hecke W, Costers L, Descamps A, Ribbens A, Nagels G, Smeets D, et al. A novel digital care management platform to monitor clinical and subclinical disease activity in multiple sclerosis. Brain Sci 2021;11(9):1171 [FREE Full text] [doi: 10.3390/brainsci11091171] [Medline: 34573193]
20. Parks AC, Williams AL, Kackloudis GM, Stafford JL, Boucher EM, Honomichl RD. The effects of a digital well-being intervention on patients with chronic conditions: observational study. J Med Internet Res 2020;22(1):e16211 [FREE Full text] [doi: 10.2196/16211] [Medline: 31922491]
21. Claflin SB, Campbell J, van der Mei I, Mainsbridge C, Taylor BV. Self-reported behaviour change among multiple sclerosis community members and interested laypeople 6 months following participation in a free online course about multiple sclerosis. Health Promot J Austr 2024;35(1):100-109. [doi: 10.1002/hpja.716] [Medline: 36894322]
22. van Kessel K, Wouldes T, Moss-Morris R. A New Zealand pilot randomized controlled trial of a web-based interactive self-management programme (MSInvigor8) with and without email support for the treatment of multiple sclerosis fatigue. Clin Rehabil 2016;30(5):454-462. [doi: 10.1177/0269215515584800] [Medline: 25952587]
23. Krause N, Riemann-Lorenz K, Rahn AC, Pöttgen J, Köpke S, Meyer B, et al. 'That would have been the perfect thing after diagnosis': development of a digital lifestyle management application in multiple sclerosis. Ther Adv Neurol Disord 2022;15:17562864221118729 [FREE Full text] [doi: 10.1177/17562864221118729] [Medline: 36092248]
24. Väestörakenne. URL: https://stat.fi/tilasto/vaerak [accessed 2024-07-13]
25. Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol 2018;17(2):162-173. [doi: 10.1016/S1474-4422(17)30470-2] [Medline: 29275977]
26. Health and well-being every day. Terveyskylä. URL: https://www.terveyskyla.fi/en [accessed 2024-06-07]
27. Schriefer D, Ness N, Haase R, Ziemssen T. Gender disparities in health resource utilization in patients with relapsing-remitting multiple sclerosis: a prospective longitudinal real-world study with more than 2000 patients. Ther Adv Neurol Disord 2020;13:1756286420960274 [FREE Full text] [doi: 10.1177/1756286420960274] [Medline: 33178335]
28. Mésidor M, Sylvestre M, Marrie RA, Rousseau M. Does age at diagnosis influence the use of health services for multiple sclerosis? Mult Scler Relat Disord 2020;46:102555. [doi: 10.1016/j.msard.2020.102555] [Medline: 33045495]
29. McKay KA, Marrie RA, Fisk JD, Patten SB, Tremlett H. Comorbidities are associated with altered health services use in multiple sclerosis: a prospective cohort study. Neuroepidemiology 2018;51(1-2):1-10. [doi: 10.1159/000488799] [Medline: 29763935]
30. Hawton AJ, Green C. Multiple sclerosis: relapses, resource use, and costs. Eur J Health Econ 2016;17(7):875-884. [doi: 10.1007/s10198-015-0728-3] [Medline: 26438399]
31. Hobart J, Lamping D, Fitzpatrick R, Riazi A, Thompson A. The multiple sclerosis impact scale (MSIS-29): a new patient-based outcome measure. Brain 2001;124(Pt 5):962-973. [doi: 10.1093/brain/124.5.962] [Medline: 11335698]
32. Rosti-Otajärvi E, Hämäläinen P, Wiksten A, Hakkarainen T, Ruutiainen J. Validity and reliability of the finnish version of the multiple sclerosis impact scale-29. Brain Behav 2017;7(7):e00725 [FREE Full text] [doi: 10.1002/brb3.725] [Medline: 28729932]
33. Costelloe L, O'Rourke K, Kearney H, McGuigan C, Gribbin L, Duggan M, et al. The patient knows best: significant change in the physical component of the Multiple Sclerosis Impact Scale (MSIS-29 physical). J Neurol Neurosurg Psychiatry 2007;78(8):841-844 [FREE Full text] [doi: 10.1136/jnnp.2006.105759] [Medline: 17332049]

34. Phillips GA, Wyrwich KW, Guo S, Medori R, Altincatal A, Wagner L, et al. Responder definition of the multiple sclerosis impact scale physical impact subscale for patients with physical worsening. Mult Scler 2014;20(13):1753-1760 [FREE Full text] [doi: 10.1177/1352458514530489] [Medline: 24740371]
35. Smedal T, Johansen H, Myhr K, Strand L. Psychometric properties of a Norwegian version of multiple sclerosis impact scale (MSIS-29). Acta Neurol Scand 2010;122(4):244-251. [doi: 10.1111/j.1600-0404.2009.01298.x] [Medline: 19951267]
36. van der Linden FAH, Kragt JJ, Hobart JC, Klein M, Thompson AJ, van der Ploeg HM, et al. The size of the treatment effect: do patients and proxies agree? BMC Neurol 2009;9:12 [FREE Full text] [doi: 10.1186/1471-2377-9-12] [Medline: 19317921]
37. Krol MW, de Boer D, Delnoij DM, Rademakers JJDJM. The net promoter score—an asset to patient experience surveys? Health Expect 2015;18(6):3099-3109 [FREE Full text] [doi: 10.1111/hex.12297] [Medline: 25345554]
38. Block VJ, Koshal K, Wijangco J, Miller N, Sara N, Henderson K, et al. A closed-loop falls monitoring and prevention app for multiple sclerosis clinical practice: human-centered design of the multiple sclerosis falls insighttrack. JMIR Hum Factors 2024;11:e49331 [FREE Full text] [doi: 10.2196/49331] [Medline: 38206662]
39. Corder GW, Foreman DI. Nonparametric Statistics: A Step-by-Step Approach. New Jersey, USA: John Wiley & Sons; 2014.
40. Katz MH. Multivariable analysis: a primer for readers of medical research. Ann Intern Med 2003;138(8):644-650. [doi: 10.7326/0003-4819-138-8-200304150-00012] [Medline: 12693887]
41. Ahlgren C, Odén A, Lycke J. High nationwide prevalence of multiple sclerosis in Sweden. Mult Scler J 2011;17(8):901-908. [doi: 10.1177/1352458511403794] [Medline: 21459810]
42. Marrie RA, Leung S, Tyry T, Cutter GR, Fox R, Salter A. Use of eHealth and mHealth technology by persons with multiple sclerosis. Mult Scler Relat Disord 2019;27:13-19. [doi: 10.1016/j.msard.2018.09.036] [Medline: 30296732]
43. Compston A, Coles A. Multiple sclerosis. Lancet 2008;372(9648):1502-1517 [FREE Full text] [doi: 10.1016/S0140-6736(08)61620-7] [Medline: 18970977]
44. Dennison L, Yardley L, Devereux A, Moss-Morris R. Experiences of adjusting to early stage multiple sclerosis. J Health Psychol 2011;16(3):478-488. [doi: 10.1177/1359105310384299] [Medline: 21135061]
45. Gold SM, Friede T, Meyer B, Moss-Morris R, Hudson J, Asseyer S, et al. Internet-delivered cognitive behavioural therapy programme to reduce depressive symptoms in patients with multiple sclerosis: a multicentre, randomised, controlled, phase 3 trial. Lancet Digit Health 2023;5(10):e668-e678 [FREE Full text] [doi: 10.1016/S2589-7500(23)00109-7] [Medline: 37775187]
46. Hakanen O, Tolvi M, Torkki P. Cost analysis of face-to-face visits, virtual visits, and a digital care pathway in the treatment of tonsillitis patients. Am J Otolaryngol 2023;44(4):103868 [FREE Full text] [doi: 10.1016/j.amjoto.2023.103868] [Medline: 36996515]
47. Tuomikoski K, Liljamo P, Reponen J, Kanste O. Digihoitopolkujen vaikutukset terveydenhuollon ammattilaisten toimintaprosesseihin erikoissairaanhoidossa. Finn J eHealth eWelfare 2022;14(3):326-338. [doi: 10.23996/fjhw.112648]
48. Voigt I, Benedict M, Susky M, Scheplitz T, Frankowitz S, Kern R, et al. A digital patient portal for patients with multiple sclerosis. Front Neurol 2020;11:400 [FREE Full text] [doi: 10.3389/fneur.2020.00400] [Medline: 32670174]

Abbreviations

DCP: digital care pathway
ED: emergency department
EDSS: Expanded Disability Status Scale
EHR: electronic health record
HCP: health care professional
HUS: Helsinki University Hospital
MS: multiple sclerosis
MSIS-29: Multiple Sclerosis Impact Scale
NPS: net promoter score
PPMS: primary-progressive multiple sclerosis
PROM: patient-reported outcome measure
pwMS: people with multiple sclerosis
RRMS: relapsing-remitting multiple sclerosis
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by A Kushniruk; submitted 15.08.23; peer-reviewed by R Haase, T Ziemssen; comments to author 08.02.24; revised version received 10.06.24; accepted 28.06.24; published 07.08.24.

Please cite as:
Vesinurm M, Maunula A, Olli P, Lillrank P, Ijäs P, Torkki P, Mäkitie L, Laakso SM
Effects of a Digital Care Pathway for Multiple Sclerosis: Observational Study
JMIR Hum Factors 2024;11:e51872
URL: https://humanfactors.jmir.org/2024/1/e51872
doi:10.2196/51872
PMID:

Original Paper

Prediction of Hearing Help Seeking to Design a Recommendation Module of an mHealth Hearing App: Intensive Longitudinal Study of Feature Importance Assessment

Giulia Angonese[1,2]; Mareike Buhl[1,3]; Inka Kuhlmann[1,2]; Birger Kollmeier[1,4,5]; Andrea Hildebrandt[1,2,5]

[1]Cluster of Excellence Hearing4All, Oldenburg, Germany

[2]Psychological Methods and Statistics Lab, Department of Psychology, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany

[3]IHU reConnect, Institut de l'Audition, Fondation Pour l'Audition, Inserm, AP-HP, Institut Pasteur, Université Paris Cité, F-75012 Paris, France

[4]Department of Medical Physics and Acoustics, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany

[5]Research Center Neurosensory Science, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany

Corresponding Author:
Giulia Angonese
Psychological Methods and Statistics Lab
Department of Psychology
Carl von Ossietzky Universität Oldenburg
Ammerländer Heerstraße 114-118
Oldenburg, 26129
Germany
Phone: 49 0441 798 4629
Email: giulia.angonese@uni-oldenburg.de

Abstract

Background: Mobile health (mHealth) solutions can improve the quality, accessibility, and equity of health services, fostering early rehabilitation. For individuals with hearing loss, mHealth apps might be designed to support the decision-making processes in auditory diagnostics and provide treatment recommendations to the user (eg, hearing aid need). For some individuals, such an mHealth app might be the first contact with a hearing diagnostic service and should motivate users with hearing loss to seek professional help in a targeted manner. However, personalizing treatment recommendations is only possible by knowing the individual's profile regarding the outcome of interest.

Objective: This study aims to characterize individuals who are more or less prone to seeking professional help after the repeated use of an app-based hearing test. The goal was to derive relevant hearing-related traits and personality characteristics for personalized treatment recommendations for users of mHealth hearing solutions.

Methods: In total, 185 (n=106, 57.3% female) nonaided older individuals (mean age 63.8, SD 6.6 y) with subjective hearing loss participated in a mobile study. We collected cross-sectional and longitudinal data on a comprehensive set of 83 hearing-related and psychological measures among those previously found to predict hearing help seeking. Readiness to seek help was assessed as the outcome variable at study end and after 2 months. Participants were classified into help seekers and nonseekers using several supervised machine learning algorithms (random forest, naïve Bayes, and support vector machine). The most relevant features for prediction were identified using feature importance analysis.

Results: The algorithms correctly predicted action to seek help at study end in 65.9% (122/185) to 70.3% (130/185) of cases, reaching 74.8% (98/131) classification accuracy at follow-up. Among the most important features for classification beyond hearing performance were the perceived consequences of hearing loss in daily life, attitude toward hearing aids, motivation to seek help, physical health, sensory sensitivity personality trait, neuroticism, and income.

Conclusions: This study contributes to the identification of individual characteristics that predict help seeking in older individuals with self-reported hearing loss. Suggestions are made for their implementation in an individual-profiling algorithm and for deriving targeted recommendations in mHealth hearing apps.

(JMIR Hum Factors 2024;11:e52310) doi:10.2196/52310

KEYWORDS

hearing loss; mobile health; mHealth; older adults; help seeking; mobile study; machine learning; supervised classification; feature importance; profiling; mobile phone

Introduction

Mobile Health Solutions for Hearing Care

Hearing enables individuals to experience their surroundings and communicate with others. Thus, hearing difficulties can have a strong impact on individuals' quality of life. Hearing loss (HL) is one of the most common chronic diseases worldwide and affects 20.3% of the world's population. More than 60% of individuals with HL are aged >50 years, with the principal cause being age-related HL [1]. Untreated hearing difficulties have been associated with lower self-rated health, depression, and anxiety in addition to physical and cognitive decline, dementia, and hospitalization in the older population [2-4]. The primary rehabilitative strategy for individuals with moderate to severe HL is the use of hearing aids (HAs), which increase activity levels, general health, and quality of life [5] and decrease social isolation and depressive symptoms [6] by supporting hearing ability and communication efficacy. Despite the positive effects of hearing rehabilitation, the prevalence of HA use is still limited to approximately 25% of the population with hearing impairment [2,4,7,8]. Moreover, there is an average delay of 9 years between the time a person first acknowledges hearing difficulties and the first contact with a hearing health professional [9].

Developing easily accessible and affordable mobile health (mHealth) solutions in audiology would promote broader and faster access to diagnosis and health services, fostering an early rehabilitation and reducing the impact of HL on the individual. It is estimated that 55% of the global population and 87% of the European population will use internet services on a mobile device by 2030 [10]. Thus, it is evident that mHealth solutions have the potential to significantly impact behaviors in the population. Current studies indicate that the use of tablets and smartphones, including mHealth apps, is steadily increasing among adults aged ≥65 years [11-13], who are even reported to be the fastest-growing population of smartphone users [11]. In the last decades, various mHealth apps have been developed for ear and hearing assessment [14-18] and for HA rehabilitation [19,20], in addition to tele-audiology services [21] and hybrid clinics [22]. The ease of use and accessibility of mHealth solutions can have an impact on self-awareness and recognition of HL and foster knowledge and use of in-person services [19,23,24]. Moreover, mHealth solutions show the potential to promote more equitable health care in low- and middle-income countries where access to health care facilities and professionals is limited [16,25]. Finally, mHealth apps that are quick and easy to use in everyday life have the potential to provide clinicians with important information at both the diagnostic and intervention phases and can help explore and understand daily experiences of the user with HL and facilitate more timely responses [19].

An mHealth hearing app might become the first contact with a hearing diagnostic service and, therefore, should motivate individuals in need to seek professional help in a personalized manner to maximize the impact of early health services on the population. However, even though professional support might be recommended given the hearing test result, many users might still be hesitant to seek help. This is particularly true among the older adult population, in which HL can be slow and gradual and is often considered a natural aspect of aging. In addition, individuals may not be aware of the rehabilitation options and hearing health care services available to them or how to access them [4]. Some users may even choose to ignore indications to seek help for various reasons, such as low awareness of HL or stigma associated with HA use. In addition, individuals high on neuroticism and low on agreeableness might generally distrust external advice. Acquiring more information about users' personal characteristics and generating their individual profiles can inform the creation of targeted recommendations, particularly for those users who need more convincing incentives to take action. These recommendations can aid skeptical individuals in their decision-making process. For example, repeated feedback on daily hearing tests could increase self-awareness of HL, and HA simulators could promote positive expectations regarding HAs. Moreover, information about users' personal characteristics collected via an mHealth hearing app could later assist clinicians in providing personalized counseling. It follows that the assessment of a person with hearing difficulties needs to go beyond the simple quantification of HL and should also take into account individual characteristics that have been shown to influence the readiness of individuals to seek professional help [4,26].

Predictors of Hearing Help Seeking and HA Uptake

Hearing help seeking can be seen as a first crucial step toward the decision to take up an HA [26]. Help seeking takes place after a contemplation stage [3] where a listener in need is initially ambivalent about making changes. Help seekers would then prepare (seek information and plan) and take action [3] toward a change in behavior and attitude, namely, consulting a health care professional about their hearing difficulties [4,27]. Acknowledgment and acceptance of hearing difficulties and their impact on everyday life have been discussed as the most important predictive features of hearing help seeking as well as later HA uptake [5,6,26,28,29]. In the older adult population, HL is frequently perceived as part of the natural aging process, and other health issues are prioritized for treatment [4,30]. Even when HL is identified, individuals might reject the use of an HA due to expected costs, stigma, and negative stereotypes [5,7,28]. However, a positive attitude toward HAs [26,27], high expectations to benefit from them [26,29], and perceived self-efficacy in their daily management [17,18] were shown to promote help seeking and HA uptake. Other relevant covariates that have been identified in the literature are personal attitudes, beliefs, and personality traits. Individuals who are more prone to seeking help and successfully uptake an HA show higher internal locus of control [2,17], self-efficacy [31,32], and agreeableness as well as lower neuroticism and openness [2].

Altogether, such individual characteristics refer to a general self-confidence in the ability to cope with critical situations, good acceptance of others' suggestions and recommendations, and less susceptibility to shame and embarrassment.

Given the wide range of traits and behaviors that have been reported in the literature to be associated with hearing help seeking and HA uptake, predictive models can be developed that take into account the multifaceted nature and association patterns of these traits and behaviors. Machine learning models are built on a large number of predictors simultaneously, usually leading to more accurate predictions than univariate or smaller models that take into account only a limited number of predictors. These models can capture complex and nonlinear relationships between the outcome and its predictors. In addition, when combined with cross-validation (CV) approaches, they draw more robust conclusions and generalize to new data. Currently, the use of machine learning to support health care applications is rapidly growing, particularly in the areas of predictive analytics, diagnosis and treatment, personalized medicine, clinical decision support, and population health management [33]. Machine learning algorithms and feature importance analysis can be used to identify the most relevant predictors of hearing help seeking from the many features hitherto reported in the literature. In an mHealth context, it is crucial to limit the number of features to a small set of key predictors to create a concise and efficient assessment battery.

Rationale and Objectives

This study aimed to identify the most important predictive features for hearing help seeking, planning to design an individual-profiling module for an mHealth hearing app. Such a module will categorize help seekers versus nonseekers and, ultimately, inform the design of targeted or even personalized treatment recommendations. For this purpose, data from a large number of questionnaires and tests covering different hearing-related and psychological characteristics, together with multiple assessments of a hearing screening test, were collected in a longitudinal study that simulated an mHealth hearing app. To target potential users of future apps of this kind, this study was geared toward individuals with subjectively perceived hearing difficulties who had not yet been compensated with HAs. We selected 25 assessment tools based on an extensive literature review of covariates of hearing help seeking and HA uptake. From these, a comprehensive set of 83 features was derived. We used supervised machine learning algorithms to predict the readiness to seek professional help as assessed at the end of the study and after 2 months as self-reports on intention to seek help. Feature importance analysis was used to narrow down the large number of features and identify a small set of key traits to predict hearing help-seeking behavior. On the basis of these results, we aimed to derive suggestions for the implementation of a profiling module as a short and concise assessment battery that can be administered after an audiological test. This could be included in existing or future mHealth apps in a modular manner. Knowledge of a user's propensity to seek help can be used to provide specific recommendations to encourage the use of hearing health care services. Ultimately, our aim was to provide clinicians and mHealth app developers with relevant knowledge about personal characteristics that are helpful in promoting hearing health by encouraging the uptake of hearing health care services and HAs when needed.

The following research questions (RQs) guided our study design and analyses:

1. Which machine learning model can best predict help seeking and categorize individuals into help seekers versus nonseekers? (RQ 1)
2. Which hearing-related and psychological features are most relevant to classify individuals into help seekers versus nonseekers? (RQ 2)
3. How can feature importance measures inform the design of targeted recommendations for users of a future mHealth hearing app? (RQ 3)

Methods

Participants

Adults aged >50 years were recruited between August 2021 and August 2022 through the Ebay minijob announcement web platform and the university intranet and via mailing list services of several German universities' guest audience and senior programs. The inclusion criteria were subjective reports of hearing difficulties in daily life, ownership of and ability to use a smartphone, and good command of the German language. The exclusion criterion was the use of HAs. A total of 192 individuals were enrolled in the study. In total, 3.6% (7/192) of the participants dropped out during the study, resulting in a completion rate of 96.4% (185/192). The final data set included 185 participants—106 (57.3%) female and 79 (42.7%) male (0 diverse)—aged between 47 and 82 years, with a mean age of 63.1 (SD 6.5) years. One participant was aged <50 years (47 years) but was nevertheless included in the final sample given that this value only slightly deviated from the planned age threshold. Of the 185 participants who completed the study, 131 (70.8%) answered the follow-up questionnaire. A descriptive summary of participants' sociodemographic characteristics is provided in Table 1.

Table 1. Main sociodemographic characteristics of the participants (n=185)[a].

Characteristic	Participants, n (%)
Age group (y)	
47-60	70 (37.8)
61-70	83 (44.9)
70-82	32 (17.3)
Sex	
Female	106 (57.3)
Duration of hearing difficulties (y)	
0-1	54 (29.2)
2-5	89 (48.1)
6-10	23 (12.4)
10-21	19 (10.3)
Presence of tinnitus	65 (35.1)
Previous physician consultation for hearing difficulties	89 (48.1)
Presence of visual problems	134 (72.4)
Occupation status	
Employed	67 (36.2)
Monthly income	
<€1500 (US $1611.21)	62 (33.5)
€1500-2500 (US $1611.21-$2685.35)	65 (35.1)
€2500-4000 (US $2685.35-$4296.56)	42 (22.7)
>€4000 (US $4296.56)	16 (8.6)
Residential environment	
Countryside	2 (1.1)
Small town	29 (15.7)
Suburbs	48 (25.9)
City	106 (57.3)
Self-estimated noise level at home	
Low	67 (36.2)
Moderate	113 (61.1)
High	5 (2.7)

[a]These data were acquired during the baseline assessment through a self-report questionnaire.

Procedure

Study Overview

Interested participants contacted the study administrator via email and received extensive written information about the purpose of the project; study design; length of participation and remuneration; possibility to withdraw participation at any time; and data protection, management, and storage. The study design and implementation and data collection, analysis, and storage were conducted in accordance with current literature on ethical considerations in the context of mobile and mHealth apps [34,35]. Security and privacy recommendations were also adhered to. It was clearly stated that a medical diagnostic was not provided in the study. Communication with the participants took place exclusively via email and SMS text message. A pilot study was conducted with a young (aged 23 years), healthy female participant in August 2021 to evaluate the usability and technical functionality of the mobile study.

The web-based study was conducted on the personal mobile phones of the participants to approximate the experience of using an app. Only 3.8% (7/185) of the participants used their computers due to technical difficulties with their smartphones. Data collection was carried out using formr, an open-source web-based application programming interface (API) for the R language that creates automated studies [36]. In formr, different questionnaires and tests (refer to Tables 2 and 3) were linearly chained together as modules of a so-called *run*. A run reproduces the desired design and can be accessed by users through a specific link. The formr software first provides a unique study link to the run, which was shared via email with enrolled participants. Upon accessing this link, participants were assigned a unique visitor session in formr and provided with a second individualized link based on web cookies. The unique visitor session prevented users from providing multiple entries for the same survey. This unique session code enabled the anonymization of the data within formr and, for the duration of the study, was stored in a written coding list, where the participants' names and session codes were recorded. The customer communication platform Twilio (Twilio Inc) [37] was used to send the individualized study link to the participants through daily SMS text message reminders. Automated SMS text message delivery was initiated via the *external link* module in formr, which uses Representational State Transfer API to connect to Twilio. Representational State Transfer API allows a software program (in this case, formr) to expose functionality and data to other programs (Twilio) in a consistent and secure format, ensuring privacy and data protection. With the individualized link received via SMS text message, participants could perform the study on their own smartphone's browser. For the 3.8% (7/185) of participants who completed the study on their computer, the daily SMS text messages were sent to their personal mobile phones as reminders. The individualized link was additionally sent via email at the beginning of the study to allow these participants to access the study via their computer browser.

Table 2. Assessment of hearing-related features (baseline and longitudinal assessment).

Domain and predictor (assessment tool)	Feature for machine learning
Participation and handicap	
Self-reported hearing difficulties (SSQ[a] [38])	• Speech hearing scale • Spatial hearing scale • Qualities of hearing scale
Consequences of hearing loss (HHIE/A[b] [39], 2020)	• Social consequences of hearing loss scale • Emotional consequences of hearing loss scale
Social life participation (Social Network Index [40], 2017, adapted from the Department of Psychology, University of Oldenburg)	• Social network diversity score • Number of people score • Number of nets score
Attitude toward hearing aids	
Hearing aid expectations (ECHO[c] [38])	• Hearing aid expectations (global score)
Hearing aid stigma (ALHQ[d] version 3.0 [41])	• Denial of hearing loss scale • Negative associations scale • Negative coping strategies scale • Manual dexterity and vision scale • Hearing-related esteem scale
Hearing-related personality traits	
Noise sensitivity (WNSS[e] [42], 1997)	• Noise sensitivity (global score)
Hearing habits (SP-HHQ[f] [43])	• Noise annoyance factor • Sound quality factor • Noise sensitivity factor • Unpredictable sounds factor • Openness factor • Warm sounds factor • Environmental sounds factor
Hearing health literacy (HLS-EU-Q16[g] [44], 2015, with 9 additional internally developed items)	• Health literacy (global score)
Hearing performance	
Hearing performance (SRT[h]; DTT[i] [45])	• SRT mean • SRT SD
Hearing feedback type	• Intermediate (percentage of yellow feedback) • Poor (percentage of red feedback)
Others	
Hearing-related sociodemographic data (sociodemographic questionnaire developed for this study by the authors)	• Hearing difficulties—presence • Hearing difficulties—duration (years) • Hearing difficulties—previous consultation with a health professional • Tinnitus—presence • Tinnitus—duration (years) • Tinnitus—previous consultation with a health professional • Motivation to seek professional help before the study • Source of motivation to seek professional help before the study • Motivation to seek professional help after the study • Source of motivation to seek professional help after the study • General attitude toward hearing aids

[a]SSQ: Speech, Spatial, and Qualities of Hearing Scale.

[b]HHIE/A: Hearing Handicap Inventory for the Elderly and Hearing Handicap Inventory for Adults.

[c]ECHO: Expected Consequences of Hearing Aid Ownership.

[d]ALHQ: Attitudes Toward Loss of Hearing Questionnaire.

[e]WNSS: Weinstein Noise Sensitivity Scale.

[f]SP-HHQ: Sound Preference and Hearing Habits Questionnaire.

[g]HLS-EU-Q16: 16-item European Health Literacy Survey Questionnaire.

[h]SRT: speech recognition threshold.

[i]DTT: digit triplet test.

Table 3. Assessment of psychological, general health, and sociodemographic features (baseline and longitudinal assessment).

Domain and predictor (assessment tool)	Feature for machine learning
Personality traits	
Big 5 (NEO-FFI[a] [46])	• Neuroticism scale • Extraversion scale • Openness scale • Agreeableness scale • Conscientiousness scale
Trait anxiety (Geriatric Anxiety Inventory [47], 2016)	• Anxiety (global score)
Trait depression (Geriatric Depression Scale [48], 1986)	• Depression (global score)
Optimism and pessimism (The Optimism-Pessimism Scale - 2 [49], 2012)	• Optimism (global score)
Loneliness (DJG[b] scale [50], 2013)	• Loneliness (global score)
Sensory processing sensitivity (HSPS-G[c] [51])	• Ease of excitation scale • Sensory threshold scale • Esthetic sensitivity scale • Hearing scale
Attitudes and beliefs	
Health locus of control (KKG[d] [52], 1989; Internal-External Control Belief Scale - 4 [53], 2012)	• Internal locus of control scale • Society control scale • External locus of control scale
Attitude toward aging (AAQ[e] [54], 2007)	• Psychosocial scale • Physical scale • Psychological scale
General self-efficacy (GSES[f] [55], 2003)	• General self-efficacy (global score)
Mood	
Affect (daily questionnaire on affect developed for this study by the authors)	• Positive affect pretest mean • Positive affect pretest SD • Negative affect pretest mean • Negative affect pretest SD • Positive affect posttest mean • Positive affect posttest SD • Negative affect posttest mean • Negative affect posttest SD
Stress (PSS[g] [56], 2020)	• Perceived stress (global score)
Cognitive functions	
Figural reasoning (BEFKI[h] [57], 2020)	• Figural reasoning (global score)
Vocabulary (Vocabulary Test [58], 1992)	• Vocabulary (global score)
Digital literacy (Technology Readiness – Short scale [59], 2012)	• Technology commitment (global score)
Others	
General health (SF-12[i] [60])	• Physical health score • Mental health score

Domain and predictor (assessment tool)	Feature for machine learning
General sociodemographic data (sociodemographic questionnaire developed for this study by the authors)	• Age • Sex • Presence of visual problems • Educational degree • Occupation (retired or working) • Weekly working hours • Monthly income • Relationship status • Monthly income of partner • Residential environment • Household size

[a]NEO-FFI: Neuroticism-Extraversion-Openness Five Factor Inventory.

[b]DJG: De Jong Gierveld Loneliness Scale.

[c]HSPS-G: Highly Sensitive Person Scale.

[d]KKG: Kontrollüberzeugungen zu Krankheit und Gesundheit (Control Beliefs about Illness and Health).

[e]AAQ: Attitudes to Aging Questionnaire.

[f]GSES: General Self-Efficacy Scale.

[g]PSS: Perceived Stress Scale.

[h]BEFKI: Berlin Test of Fluid and Crystallized Intelligence.

[i]SF-12: 12-item Short-Form Health Survey.

A detailed list of all assessment tools and their references, as well as the derived features for analysis, is provided in Tables 2 and 3. Each assessment tool was implemented as a *survey* in formr. Most of the surveys included in the study were implemented following the paper-and-pencil version that was retrieved from the literature. The surveys that were developed or adapted specifically for this study can be shared upon request. The items assessed can be inferred from the features provided in Tables 2 and 3. Submission of each survey was possible only after all mandatory questions had been answered. After submission of one survey, the study advanced automatically to the following questionnaire or test planned in the formr run. Users were not allowed to go back and modify their answers after submission.

The total assessment time of 8 hours was distributed across the working days of 3 consecutive weeks, with an overall daily active participation of approximately 30 minutes. The study design is detailed in Multimedia Appendix 1. Participants could select the start date of the study to ensure that the assessment could be easily integrated into their personal schedule. The first week (baseline assessment) included 1 measurement time point per day (requiring approximately 20-30 minutes to complete), which could be performed at any time. On the first day, participants received an email with the study link and their unique participant code. After opening the study link on their browser, each participant was given the possibility to read a summary of the data protection conditions in the first page of the study again. Second, they were asked to provide a telephone number and email address, which were then stored in formr and used for the automatic SMS text message reminders. They would then receive an automatic SMS text message with the individualized link to the study, through which they could begin the assessment. From days 2 to 5, participants received an SMS text message with the link to the study at 7 AM, but they had been previously informed that they could perform the tasks at any time during the day. An email reminder was sent at 7 PM in case participants had not accessed the study link by that time.

The second and third week included 2 measurement time points per day of approximately 15 minutes each. The longitudinal assessments were prompted via SMS text message at 7 AM and 7 PM. Participants were instructed to access the study at their earliest convenience after waking up and before going to bed, thus allowing them to accommodate the study to their daily schedules. In the morning, after clicking on the link they received via SMS text message at 7 AM, each participant was first presented with some questions on baseline mood and sleep quality. They were then required to click on a second link embedded in the following survey page that redirected them to the hearing assessment. Finally, participants were asked again to report on their mood after receiving feedback on their hearing performance. In case the participant forgot to access the link and perform the study, an email reminder was sent at 1 PM. If, after the reminder, the participant still did not take part in the study, the session was established as incomplete. The study administrator had to manually allow the participant to move to the next measurement time point (in this case, the evening assessment) within formr. This same assessment scheme was repeated in the evening. In this case, the SMS text message with the link to the study was sent at 7 PM, and the email reminder was sent at 11 PM.

At the end of the study, participants were asked to provide consent to be contacted after 2 months for a voluntary (and nonremunerated) follow-up questionnaire. Those who provided their consent received an email with a link to a single formr survey that required <5 minutes to complete. The short survey consisted of 2 multiple-choice questions and was completed by 70.8% (131/185) of the participants. Individuals were asked to report again on their action to seek professional help following the feedback received during the study and were asked to indicate whether study participation improved their awareness of hearing difficulties.

Baseline Assessment

Overview

During the baseline assessment, cross-sectional data from a comprehensive set of 25 questionnaires and tests were collected. The questionnaires and tests were distributed on 5 consecutive days to maximize study compliance and avoid priming effects on different questionnaires. The following sections provide a concise summary of the measured predictors (features). We refer to Tables 2 and 3 for a complete list of the assessment tools used in the study. We selected questionnaires and tests that have been previously used in studies investigating their association with hearing help seeking and HA uptake (as cited in the Introduction section and in the Tables 2 and 3). If the tools included in the study had not been previously used in similar literature, we explained our rationale for their selection in the following sections.

Assessment of Hearing-Related Features

First, the assessment included self-reports of participation and perceived handicap, focusing on self-reported hearing difficulties, consequences of HL, and social life participation. In addition, attitudes toward HAs were evaluated using questionnaires on HA expectations and stigma. Hearing-related personality traits were also taken into account. Noise sensitivity was measured as a personality trait, which was shown to be related to affect and neurosensory processing [61]. We further assessed hearing habits aiming to gather more information about sound sensitivity and individuals' sound preference profiles [43]. Finally, hearing health literacy was assessed as well as the ability to search, find, and understand information related to hearing health has been shown to be associated with better self-management of HL [31].

Assessment of Psychological, General Health, and Sociodemographic Features

Personality traits (the Big 5 [62]) were shown to be associated with help seeking and HA use and, therefore, were included in the baseline assessment. Anxiety and depression were also measured given their frequently demonstrated associations with HL [4,6], together with loneliness, which is seen as consequence of untreated HL [3]. We further assessed optimism and sensory processing sensitivity, which refers to an individual's disposition to perceive and process stimuli (including auditory ones) more intensely than the average population [63]. Attitudes and beliefs such as locus of control and self-efficacy were included as well for their association with help seeking and HA use. The belief that HAs are associated with old age and infirmity is often a barrier to HA uptake and use [4]; therefore, attitude toward aging was assessed as well. Perceived stress was measured, too, as high levels of stress that are related to daily life, work, or social situations may boost help-seeking behaviors [3]. General health was assessed given its predictive role for different steps of the HA uptake path [4,6]. For completeness, we also measured cognitive abilities (crystallized and fluid intelligence) despite discordant findings on associations between cognition and HA uptake [4,6,16]. Finally, participants were requested to complete a comprehensive questionnaire on sociodemographics.

Longitudinal Assessment of Hearing and Affect

This microlongitudinal assessment accounts for potential daily fluctuations in hearing performance and affect, which might depend on particular daily events and states. The affect questionnaire included 14 items in line with the circumplex model of affect (Posner et al [64]). A total of 8 items were related to negative affect, and 6 items were related to positive affect [65]. The items are listed in Multimedia Appendix 1. The affect questionnaire was presented before and after the hearing test to assess mood at baseline and after receiving feedback on the hearing test, respectively.

Hearing performance was assessed using the digit triplet test (DTT) [45,66] by Hörzentrum Oldenburg gGmbH. This widely used screening instrument [67] measures speech intelligibility in noise by means of the speech recognition threshold (SRT), which indicates the signal-to-noise ratio (SNR; difference between speech and noise level) at which the participant reaches 50% speech intelligibility. SRT measures obtained using the DTT showed high correlations (r>0.70) with pure tone average measures while being relatively robust against changes in presentation level [68]. Moreover, the DTT has shown to be robust to ambient noise levels outside of audiometric booth environments [68]. Together with its low linguistic and cognitive demands [68], the DTT appears to be suitable for mobile, remote self-test–based screening of hearing abilities in the older population. Smartphone-based DTT has also shown the potential to provide widespread access to hearing screening in low- and middle-income countries and across different socioeconomic strata [69]. After completing the hearing test, participants received feedback on their performance in the form of a traffic light color, where green indicated good performance (SRT<–7.1 dB SNR), yellow indicated intermediate performance (–7.1≥SRT<–5.1 dB SNR), and red reflected poor performance (SRT≥–5.1 dB SNR) [45,70]. The Multimedia Appendix 1 provides detailed information on the hearing test and Multimedia Appendix 2 provides information on its feedback. Participants performed the hearing test with their personal smartphone and headphones. A total of 1.6% (3/185) of the participants used loudspeakers as headphones were not available to them. Calibration of the hardware equipment was not possible due to the remote assessment. However, SRT estimation is relatively robust against changes in presentation level, and no exact calibration is needed [71]. Moreover, the use of different types or qualities of headphones has shown no impact on test reliability [67,69]. Each hearing test began with a signal adjustment trial meant to set the stimulus at approximately 65 dB sound pressure level. The participant was presented with a digit triplet in noise and asked to "adjust the volume to hear both the digits and the noise clearly."

Outcome Measures

Overview

Classification of participants into help seekers and nonseekers was based on self-reports of planned actions to seek professional help for their perceived hearing difficulties and their motivation to seek help. These variables, as retrieved at the end of the study and at follow-up, were chosen as outcome measures for the supervised machine learning (refer to the following sections). This information was also assessed at the beginning of the study. In total, 3 different classifications were considered as outcome measures: action to seek help at study end, action and motivation to seek help at study end, and action to seek help at follow-up. The distribution of participants along the outcome classes considered is summarized in Table 4.

Table 4. Absolute classwise frequencies of observations at the 2-month follow-up across the 3 outcomes considered at the end of the study[a].

	Action at follow-up (n=52, 28.1%), n (%)	No action at follow-up (n=79, 42.7%), n (%)	No follow-up data (n=54, 29.2%), n (%)
Action (n=64, 34.6%)	33 (17.8)	12 (6.5)	19 (10.3)
No action and high motivation (n=47, 25.4%)	12 (6.5)	22 (11.9)	13 (7)
No action and low motivation (n=74, 40%)	7 (3.8)	45 (24.3)	22 (11.9)

[a]The 3 outcomes considered were action to seek help (n=185), action and motivation to seek help (n=185), and action to seek help at follow-up (n=131). In addition, the table provides an overview of those participants who did not complete the follow-up questionnaire.

Action to Seek Help

Help seeking (preparation and action [3]) was assessed at study end using the following question: "Given the feedbacks of this study regarding your hearing performance, have you made an appointment with one of the following physicians or a hearing care professional, or are you planning to do so?" (followed by a list of hearing professionals). This variable was used to create two outcome classes: (1) *action* class (64/185, 34.6%)—participants who were planning to seek professional help in the near future or had already made an appointment—and (2) *no action* class (121/185, 65.4%)—participants not ready to take action who did not plan to consult a hearing health professional in the future.

Action and Motivation to Seek Help

A second outcome measure was taken into account to further differentiate the *no action* class to provide further insights for the design of targeted recommendations in an mHealth hearing app. Information on readiness to take action was combined with the reported motivation to seek help at the end of the study. Motivation was assessed through the following question: "How motivated are you at the moment to seek help regarding your hearing problems?" (1=not motivated at all; 7=very strongly motivated). The answer spectrum was binarized by means of median split to create the following outcome classes: (1) *no action and high motivation* class (47/185, 25.4%)—participants not ready to take action with high motivation who might particularly benefit from personalized and tailored recommendations, (2) *no action and low motivation* class (74/185, 40%)—participants not ready to take action who reported low motivation to seek help, and (3) *action* class (64/185, 34.6%)—participants ready to take action regardless of their motivation level (this class was not further divided with respect to motivation as this would not result in different recommendations. Moreover, data exploration revealed that only 11% (7/64) of the individuals in this category reported low motivation).

Action to Seek Help at Follow-Up

The voluntary follow-up questionnaire (completed by 131/185, 70.8% of the participants) included a question on the intention or action to seek help following the feedback received during the study. The answer range (4 multiple-choices) was binarized to achieve a class allocation comparable to that of the first outcome measure: (1) *action at follow-up* class (52/131, 39.7%)—participants who reported having completed an appointment with a hearing professional, who had an appointment scheduled but not completed, or who were planning to seek help in the near future; and (2) *no action at follow-up* class (79/131, 60.3%)—participants who did not plan to consult a hearing health professional.

Statistical Analysis

Data Preprocessing

Data analysis was performed using the R software (R Foundation for Statistical Computing) [72]. Raw data from all questionnaires were imported from the web-platform formr to the R environment using the dedicated *formr* package [73]. For each questionnaire or test presented at baseline, global scores were computed and considered as features. If both global scores and scale scores were available for a given assessment tool, only the scale scores were retained if considered differentially relevant for the outcome. Hearing test results were sent via email to the investigator from the researchers of Hörzentrum Oldenburg and imported into R as .eml files. The performance feedback category (green, yellow, and red) was additionally extracted and stored for each raw SRT result. Due to the particular implementation of the study in formr, participants could perform the hearing test more than once at each measurement time point. Whenever this happened, only the last SRT result at a given time point was kept for analysis. This led to a removal of 3.9% of the raw SRT results. The longitudinal data on daily mood and hearing performance were summarized into individual means and SDs. The summarized longitudinal data were merged with the cross-sectional data, resulting in a wide-format data frame including 83 features. Tables 2 and 3 provide a complete list of the features considered for analysis.

Completion rate for the baseline questionnaire was 100% (185/185), whereas there were missing data for the longitudinal measures of hearing performance and affect. A complete set of 20 SRT results was collected for 43.8% (81/185) of the participants, whereas at least 15 SRT results were obtained in 95.1% (176/185) of cases. Missing hearing data at a specific time point were considered not available (NA). Where an SRT result was missing, the respective feedback and measures of affect at the posttest time point were established as missing as well (NA). Through visual inspection of the individual SRT distributions, some specific outlier patterns were identified. A total of 7.6% (14/185) of the participants showed a much larger SRT result at the first measurement, which qualified as an outlier following the IQR rule. These large SRT values (indicating poor performance) were considered to be caused by misunderstanding of the hearing test instructions and, therefore, were established as NA. However, the respective feedback category and measures of affect where not established as NA. This is because, despite the unreliable SRT value, participants' mood could still have been affected by the feedback received. With respect to daily affect measures, 1.1% (2/185) of the participants provided no data at measures of posttest affect such that summary measures could not be computed. In these cases, the mean imputation technique was applied—the sample mean and sample variance for negative and positive affect at the posttest time point were imputed to replace missing values.

Machine Learning

Overview

The data were fed into 3 machine learning algorithms for supervised classification. We chose naïve Bayes (NB), random forest (RF), and support vector machine (SVM) among other classifiers to cover a wide range of model complexity (from simple models such as NB to more complex and nonlinear ones such as SVM). The algorithms were implemented in R using the *mlr* package [74] following the approaches described in the work by Rhys [75] and Bischl et al [76]. Given the presence of 3 different outcome measures, the same analysis steps were carried out in parallel for each outcome with a slight difference in the input features included in the analysis. For the first outcome (*action to seek help*), data on motivation at the prestudy time point were kept in the feature space, whereas motivation to seek help at the end of the study was removed. The same applied to the third outcome (*action to seek help at follow-up*). Differently, for the second outcome (*action and motivation to seek help*), all data on motivation at the pre- and poststudy time points were removed from the feature space. Due to the relatively small data set and imbalanced classes, we chose not to split the data into training and test sets but to use CV instead. CV divides the training set into k equally sized parts and considers the kth part as a test set and the $k – 1$ part as a training set at each iteration. Model results are then averaged across all iterations. The implementation details of the 3 algorithms are summarized next.

NB Classifier

This algorithm uses the Bayes rule to predict the probability of an observation belonging to one of the outcome classes given its discriminant function values. Given the prior probability, the likelihood, and the evidence for each observation, the relative posterior probability for each class is computed. The single observation is then assigned to the class with the highest relative posterior probability [75]. The 2 strong assumptions made by NB algorithms are the normal distribution of continuous features (or predictors) and the independence of these features. Model performance will suffer in case of violation of these assumptions [75]. In this implementation, after training the algorithm, repeated 10-fold CV was used to evaluate the model's performance. A stepwise approach was used to select the appropriate number of CV repetitions necessary to achieve accurate and stable performance estimates (50, 100, 150, and 300 CV repetitions).

RF Classifier

Tree-based methods use recursive binary splitting to stratify the features' space in smaller, nonoverlapping regions used for classification. At each iteration of the tree-building process, the algorithm selects among all features the one that best splits the data into 2 branches according to a specific question or rule (node) [75]. The process iterates until a stopping criterion is met and final regions (leaves) are identified. In a classification problem, the mode of the training data within a region is used for prediction—each observation is classified to the majority class within the leaf to which it belongs [77]. Trees are easy to interpret, but they lack predictive power as they tend to overfit the training data. Approaches such as RF can be used to improve prediction accuracy. RF is a nonlinear method that involves the generation of multiple uncorrelated trees from different bootstrapped training sets obtained through sampling with replacement from the original data. The final predicted outcome is retrieved from aggregating the prediction of all built trees and selecting the most frequent or modal prediction [75,77]. This algorithm requires the tuning of a set of hyperparameters that control the learning process and are selected (or tuned) by the algorithm to obtain the best performance. The hyperparameters shown in Textbox 1 were considered.

Textbox 1. Hyperparameters for RF classifier.

Ntree

The number of trees to include. This value is usually fixed at a computationally reasonable value rather than tuned [75]. Ntree was set to 800.

Mtry

The number of features to randomly sample at each time. A popular value is given by $\sqrt{p}$ (where p=the number of predictors) [78]. Different search spaces were explored, with Mtry ranging between 1 and 15.

Nodesize

The minimum number of cases to be included in a leaf. Different search spaces were explored, with Nodesize ranging between 1 and 20.

Tuning the algorithm and finding the best hyperparameter combination requires the definition of an optimization algorithm, or search strategy, and evaluation method. We used grid search with 10-fold CV resampling. To evaluate model performance, nested CV was applied. In this approach, an inner loop tunes the hyperparameters, and an outer loop evaluates a wrapped learner, which comprises the classification task, the learner type (RF), and the hyperparameter tuning process. In this case, a 5-fold CV was applied as an outer resampling strategy.

SVM Classifier

The SVM algorithm iteratively identifies a hyperplane that separates labeled classes also in case of nonlinear data distributions. It does so by adding an extra dimension to the data, which is found through the kernel function, a mathematical transformation of the data. The hyperplane is defined as a surface that has 1 dimension less than the number of variables in the data set. The position of the hyperplane depends on the position of the support vectors, which are training set cases that define the class boundaries [75]. The optimal hyperplane is found by maximizing its margin, which is the region around the hyperplane that touches the fewest training observations. In fact, the distance from a training case to the margin can be viewed as a measure of the correctness of its classification [77]. In case the algorithm needs to separate >2 classes, several models are built and compared to find the one that best predicts new data. SVMs are computationally expensive but tend to perform very well on a variety of tasks conducted on nonlinearly separable classes. In addition, the algorithm has the advantage of making no assumptions on the features' distributional properties [75,77]. Similar to RF, SVM requires hyperparameter tuning. The hyperparameters in Textbox 2 were considered.

Textbox 2. Hyperparameters for SVM classifier.

Kernel

The type of kernel function used to identify the hyperplane. Polynomial, radial, and sigmoid functions were included in the search space [75].

Degree

The shape of the decision boundary (in case of a polynomial kernel). The search space was limited to values from 1 to 5 to avoid the risk of overfitting [75].

Cost

The penalty for having cases fall inside the margin. It is recommended to tune both cost and gamma (refer to the next point) on the logarithmic scale [79], and a popular search space for cost is from 2^{-5} to 2^{15} [80].

Gamma

Influence of each case on the hyperplane. This hyperparameter search space was set to the popular range of 2^{-15} to 2^{3} [80].

Nested CV was used as previously described for RF. An inner resampling loop (with 10-fold CV) was applied for hyperparameter tuning, and an outer loop (with 5-fold CV) was applied for performance evaluation.

Classification Performance Metrics

The algorithms were evaluated in terms of prediction accuracy on the test set, which indicated the overall proportion of cases correctly classified by the model as compared to the observed outcome. However, class imbalance in the sample can negatively impact prediction accuracy, reducing its informativeness as a performance measure [81,82]. The Matthews correlation coefficient (MCC; *yardstick* package [83]) was additionally taken into account to evaluate model performance. This coefficient improves over accuracy measures in case of imbalanced data sets [82,84] and can take values from –1 to 1, with 1 indicating perfect prediction, 0 indicating chance prediction, and –1 indicating inverse prediction. As an additional metric, we computed the confusion matrix (calculateConfusionMatrix(), *mlr* package). Its output provides the absolute number and the proportion of correct model predictions and misclassifications for each outcome class. For binary outcomes, the confusion matrix allows for the estimation of model sensitivity (ie, accurately identifying help seekers) and specificity (ie, accurately identifying nonseekers). In this study, obtaining high specificity was of particular interest in the context of an mHealth hearing app. Indeed, individuals with HL who are not prone to seeking help are the main target population for tailored treatment recommendations and counseling.

Feature Importance

After identifying the best-performing machine learning algorithm, feature importance was considered. Each feature receives a coefficient of importance that indicates its contribution to model performance regardless of the type of relationship (direction and linearity) between the feature and the outcome [85]. In RF, feature importance is model dependent and indicates how much the feature contributes in reducing node impurity. Importance values were retrieved using the function getFeatureImportance() (*mlr* package) applied on the RF model trained using the tuned hyperparameters. These importance results have the advantage of being inherent to the model and closely tied to its performance [86]. Conversely, there are no model-specific importance metrics available for NB and SVM. For these algorithms, the importance value assigned to each feature corresponds to the area under the receiver operating characteristic curve, which is computed from sensitivity and specificity measures [86]. The function varImp() from the *caret* package [87] was applied on the model trained using the function train() (*caret* package) after ensuring comparable performance with that of the same model previously trained on the *mlr* package.

Features with higher importance ranking were considered for inclusion in the profiling algorithm as they represented the most relevant predictors for telling apart help seekers and nonseekers. No statistical criterion exists to determine which importance value threshold should be used to retrieve relevant features. Hence, 3 threshold values were inspected (the first 10, 15, and

20 features in their importance ranking order) and evaluated in terms of predictive accuracy and interpretability. The classification accuracy of these 3 feature sets was assessed on the outcome data obtained at follow-up. For this analysis, the data set was reduced to 131 participants who completed the follow-up questionnaire. The important features were fed into the best-performing machine learning algorithm from the previous step.

Ethical Considerations

The study plan and data management have been approved by the Research Ethics Committee of the Carl von Ossietzky Universität Oldenburg (08.09.2021; EK/2020/020-01). The study supported the autonomy of participants through extensive informed consent, which was given both as a separate written document before enrollment and within the formr survey framework. Debriefing was included, and participants were invited to provide feedback at the end of the study. No direct risks associated with the study design were identified, and privacy risks were accounted for through appropriate data management and data protection concepts for all software and platforms used. Personal information collected during the study was pseudoanonymized using a written coding list stored in a closed locker accessible only to the study administrator. This coding list was destroyed at the end of data collection; therefore, the data have been completely anonymized since. Data collection took place between September 2021 and September 2022. Participants were remunerated with €10 (US $10.74) per hour. A further incentive for study participation was that participants received written feedback on their daily hearing test results.

Results

Machine Learning Classification Performance

Predicting Action to Seek Help

A summary of the model-specific classification accuracy for the first outcome (*action to seek help*) is provided in Table 5. The 3 algorithms showed similar overall performance accuracy estimates on the test set, correctly classifying approximately 65.9% (122/185) to 70.3% (130/185) of the cases in the full data set (n=185). RF was the best-performing algorithm with an accuracy of 70.3% (130/185) and an MCC of 0.28, indicating that the model's prediction improved to approximately 20% over chance. By inspecting the confusion matrix (Figure 1), we observed that RF shows high specificity, correctly classifying 90.9% (110/121) of the cases belonging to the *no action* class. The NB classifier (10-fold CV repeated 50 times) showed the best sensitivity compared with the competing algorithms, with 51% (33/64) of cases in the *action* class being correctly classified. RF and NB were selected for feature importance analyses given the good predictive performance and high specificity of RF as well as the high sensitivity of NB.

Table 5. Model-specific overall performance and class-specific classification accuracy rates for the first outcome, *action to seek help*, measured at study end[a].

Hyperparameters			Overall performance measures		Class-specific classification accuracy	
Model and parameter	Parameter space	Tuned value	Test accuracy	MCC[b]	Action (n=64)	No action (n=121)
NB[c]						
—[d]	—	—	0.66	0.26	0.51	0.75
RF[a]						
Ntree	800	800	0.70	0.28	0.31	0.91
Mtry	(5, 15)	13	—	—	—	—
Nodesize	(1, 5)	1	—	—	—	—
SVM[f]						
Kernel	—	Radial	0.67	0.23	0.39	0.82
Degree	(1, 5)	4	—	—	—	—
Cost	(2^{-5}, 2^{15})	0.01	—	—	—	—
Gamma	(2^{-15}, 2^{3})	69.1	—	—	—	—

[a]Results are based on the full data set (n=185). For random forest and support vector machine, the table additionally shows the hyperparameter search space used in the resampling procedure and the tuned values used for model training and feature importance analysis. The selected naïve Bayes model used 10-fold cross-validation repeated 50 times.

[b]MCC: Matthews correlation coefficient.

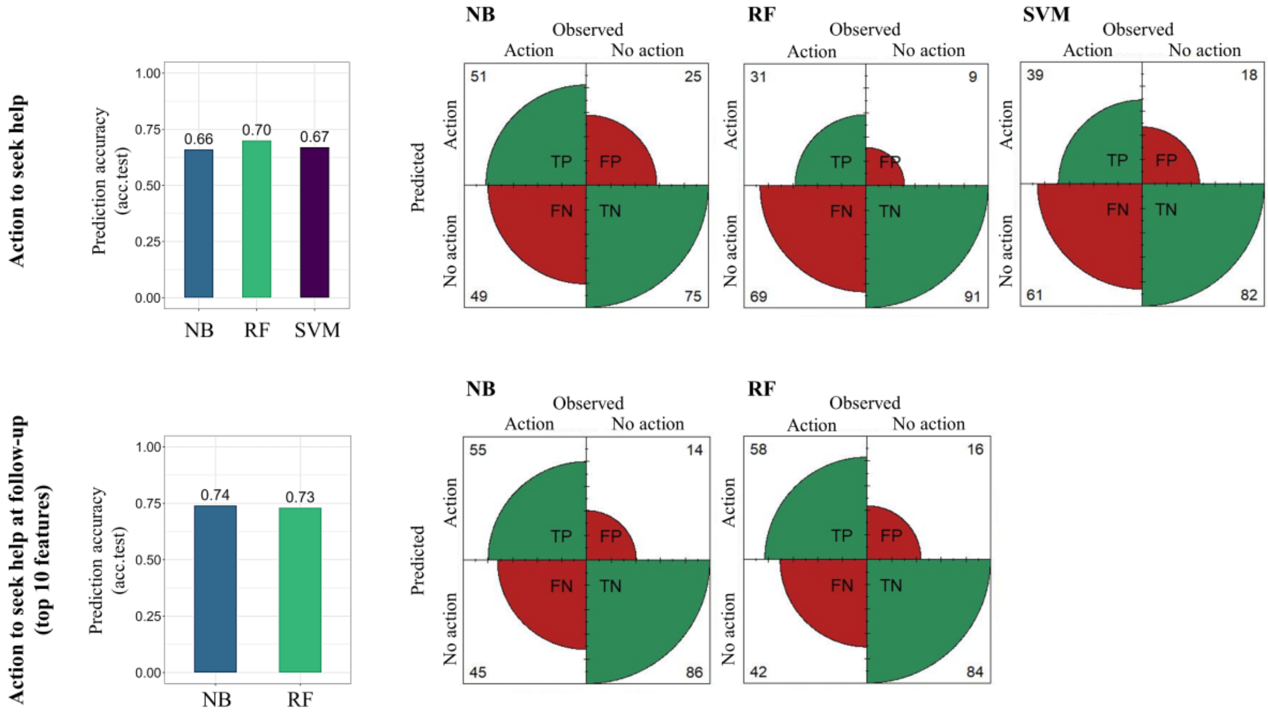
[c]NB: naïve Bayes.

[d]Not applicable.

[e]RF: random forest.

[f]SVM: support vector machine.

Figure 1. Visualization of the model-specific performances (left) and relative confusion matrices (right) for the outcomes action to seek help (first row) measured at study end (n=185) and action to seek help at follow-up (second row) considering the set of top 10 most important features (n=131). Acc.test: prediction accuracy on the test set; FN: false negative; FP: false positive; NB: naïve Bayes; RF: random forest; SVM: support vector machine; TN: true negative; TP: true positive.



Predicting Action and Motivation to Seek Help

Table 6 summarizes the classification performance with respect to the second outcome (*action and motivation to seek help*), which includes 3 classes (refer to the previous section). RF provided the highest accuracy with 55.1% (102/185) and an MCC of 0.30 as compared to NB (10-fold CV repeated 100 times) and SVM. However, the confusion matrix revealed that none of the 3 models was able to adequately tell apart individuals within the no action class who differed with respect to high versus low motivation. All algorithms could only correctly classify 2% (1/47) to 25% (12/47) of cases in the no action and high motivation class. Potentially, an improvement in classification accuracy could be achieved using a larger data set in which the classes are better balanced and with a more reliable and elaborate measure of the participants' motivation to seek help. In view of these limitations, the second outcome was not considered for feature importance analysis.

Table 6. Model-specific overall performance and class-specific classification accuracy rates for the second outcome *action and motivation to seek help* measured at study end[a].

Hyperparameters			Overall performance measures		Class-specific classification accuracy		
Model and parameter	Parameter space	Tuned value	Test accuracy	MCC[b]	Action (n=64)	No action and low motivation (n=74)	No action and high motivation (n=47)
NB[c]							
—[d]	—	—	0.50	0.22	0.47	0.67	0.25
RF[e]							
Ntree	800	—	0.55	0.30	0.55	0.80	0.09
Mtry	(8, 10)	—	—	—	—	—	—
Nodesize	(3, 15)	800	—	—	—	—	—
SVM[f]							
Kernel	—	Sigmoid	0.49	0.20	0.53	0.76	0.02
Degree	(1, 5)	1	—	—	—	—	—
Cost	(2^{-5}, 2^{15})	323	—	—	—	—	—
Gamma	(2^{-15}, 2^{3})	3.05×10^5	—	—	—	—	—

[a]Results are based on the full data set (n=185). For random forest and support vector machine, the table additionally shows the hyperparameter search space used in the resampling procedure and the tuned values used for model training and feature importance analysis. The selected naïve Bayes model used 10-fold cross-validation repeated 100 times.

[b]MCC: Matthews correlation coefficient.

[c]NB: naïve Bayes.

[d]Not applicable.

[e]RF: random forest.

[f]SVM: support vector machine.

Feature Importance

Predicting Action and Motivation to Seek Help at Follow-Up

Feature importance was analyzed based on the RF and NB algorithms predicting the first outcome, *action to seek help* at study end on the full data set (n=185). Each importance value signifies the feature's contribution to the model's performance. However, as detailed in the Methods section, RF and NB models calculate these coefficients differently. Consequently, ranking values were used. In both models, the 83 features were initially ranked in descending order based on their importance values. Features with higher importance rankings were mostly relevant in predicting help-seeking behavior. Three sets of features among the most important ones were taken into account for subsequent analysis: (1) the top 10 features indicated by the 2 models, resulting in a total of 12 best features; (2) the top 15 features indicated by the 2 models, resulting in a total of 19 best features; and (3) the top 20 features indicated by the 2 models, resulting in a total of 28 best features.

Figure 2 shows all 28 features with their importance ranking values originating from the RF and NB models. The specific importance values for each feature are provided in Multimedia Appendix 3. Next, the 3 sets of features (top 10, 15, and 20 features) were evaluated for their predictive performance and classification accuracy on the reduced data set of 131 participants who completed the follow-up questionnaire. NB and RF were trained on the 3 different feature sets for predicting the *action to seek help at follow-up*. The results are summarized in Table 7. They show that all feature sets provided good predictive performance and that the NB algorithm outperformed RF, with an overall accuracy ranging between 73.3% (96/131) and 74.8% (98/131) and an MCC between 0.43 and 0.47. Class-specific classification accuracy was comparable between NB and RF, with the action class correctly classified in 52% (27/52) to 63% (33/52) of the cases and the *no action* class correctly classified in 82% (65/79) to 86% (68/79) of cases. As can be seen from the confusion matrix in Figure 1, sensitivity and specificity measures were comparable for both algorithms predicting action to seek help at follow-up using the set of top 10 most important features. Sensitivity ranged from 55% (NB) to 58% (RF), and specificity ranged from 84% (RF) to 65% (NB).

Figure 2. Ranking of the most important features to predict action to seek help at study end. Importance rankings are shown for the 20 most important features for the 2 models (naïve Bayes [NB] and random forest [RF]) trained on the full data set (n=185), as summarized in Table 5. A total of 28 features are arranged on the y-axis with respect to their average ranking between the 2 models. HA: hearing aid; HL: hearing loss; M: mean.

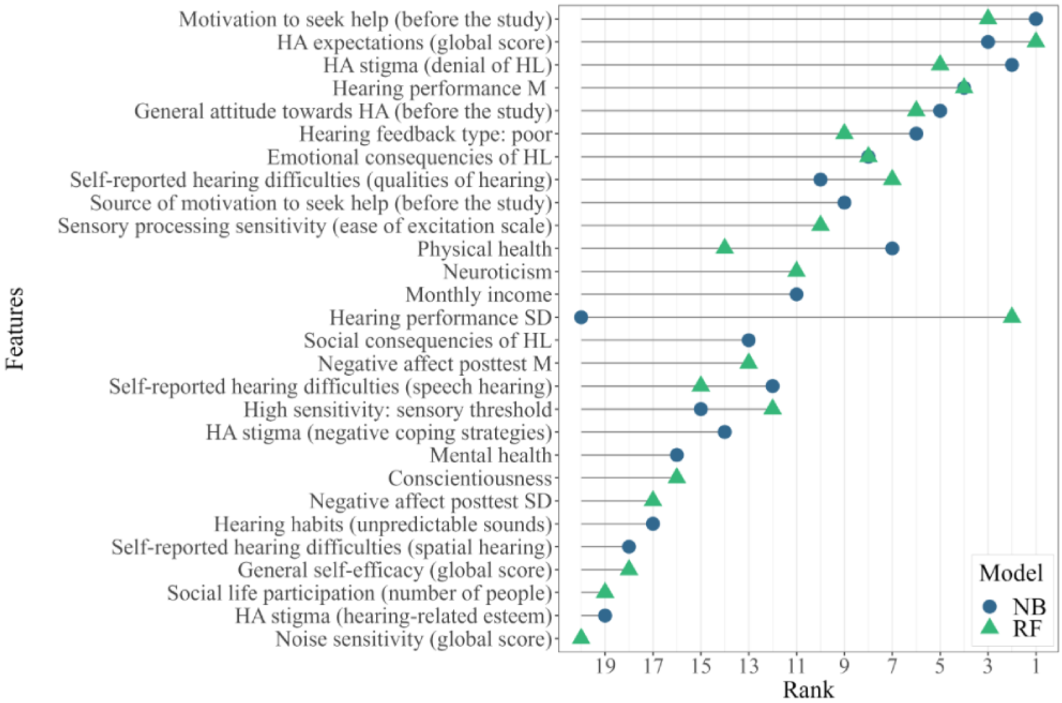


Table 7. Model-specific overall performance and class-specific classification accuracy rates for the outcome *action to seek help at follow-up*[a].

Hyperparameters		Overall performance measures		Class-specific classification accuracy	
Model and parameter	Parameter space	Test accuracy	MCC[b]	Action (n=52)	No action (n=79)
Top 10 features (total=12)					
NB[c]					
—[d]	—	0.74	0.44	0.55	0.86
RF[e]					
Ntree	800	0.73	0.43	0.58	0.84
Mtry	(1, 4)	—	—	—	—
Nodesize	(1, 5)	—	—	—	—
Top 15 features (total=19)					
NB					
—	—	0.73	0.43	0.58	0.83
RF					
Ntree	800	0.73	0.41	0.56	0.84
Mtry	(1, 10)	—	—	—	—
Nodesize	(1, 5)	—	—	—	—
Top 20 features (total=28)					
NB					
—	—	0.75	0.47	0.63	0.83
RF					
Ntree	800	0.70	0.36	0.52	0.82
Mtry	(1, 5)	—	—	—	—
Nodesize	(1, 5)	—	—	—	—

[a]Results are based on the reduced follow-up data set (n=131). The table shows the results for 3 sets of features (10, 15, and 20 most important features for the 2 models). For naïve Bayes, 10-fold cross-validation repeated 50 times was used. For random forest, the table additionally shows the hyperparameter search space used in the resampling procedure. In contrast to the previous results (Tables 5 and 6), the tuned hyperparameters were not retrieved as no importance analysis followed.

[b]MCC: Matthews correlation coefficient.

[c]NB: naïve Bayes.

[d]Not applicable.

[e]RF: random forest.

Important Features

Textbox 3 provides a brief description of 12 features ranked among the most important (previously named as the top 10 features for the 2 models) in the prediction of *action to seek help at follow-up*.

Textbox 3. 12 Features ranked among the most important in the prediction of action to seek help at follow-up.

- *Motivation to seek help* and *source* of this motivation at the beginning of the study
- Individuals' attitude toward and expectations regarding HAs, including *general attitude toward HAs*; *expectations regarding HAs*, as measured using the global score of the Expected Consequences of Hearing Aid Ownership questionnaire [38], which assesses positive and negative expectations regarding HAs, expected services and costs, and assumptions about change in the personal image in case of HA use; and *stigma toward HAs*, as measured using the Denial of Hearing Loss scale of the Attitudes Toward Loss of Hearing Questionnaire [41], which assesses acceptance of HAs and acknowledgment of HL
- *Hearing performance* (mean SRT and its variability as measured using the DTT [45,66]) and *percentage of negative feedback received* (indicating poor performance)
- Perceived consequences of HL, including *emotional consequences of HL*, as measured using the corresponding subscale of the Hearing Handicap Inventory questionnaire [39], and *self-reported hearing difficulties*, as measured using the qualities of hearing subscale of the Speech, Spatial, and Qualities of Hearing Scale [38], which addresses recognition, perceived clarity and naturalness of everyday sounds, and listening effort experienced in different hearing contexts
- *High sensory sensitivity personality*, as assessed through the ease of excitation subscale of the Highly Sensitive Person Scale questionnaire [51], which assesses emotional reactivity to physiological stimuli
- Reported *physical health*, measured using the corresponding items of the 12-item Short-Form Health Survey [60]

As outlined in the Methods section, ranking values up to 10 were arbitrarily chosen to identify the most important features. The following two features had slightly lower importance values (and received a ranking value of 11) but might provide further insights into targeted counseling in an mHealth hearing app: (1) *neuroticism*, which refers to a predisposition to experiencing negative emotions [2] and was assessed through the corresponding items of the Neuroticism-Extraversion-Openness Five-Factor Inventory questionnaire [46]; and (2) *monthly income*, which was categorized using 3 cutoff values (<€1500 [<US $1611.21], €1500-€2500 [US $1611.11-$2685.35], €2500-€4000 [US $2685.35-$4296.56], and >€4000 [>US $4296.56]).

Discussion

Principal Findings

Overview

This study contributes to the identification of individuals' hearing-related, psychological, and general health–related traits that predict the readiness to seek professional help for HL. Cross-sectional and longitudinal data were collected in a comprehensive mobile study. Potential users of a future mHealth hearing app, namely, individuals with subjective hearing difficulties, were classified into help seekers and nonseekers by means of supervised machine learning algorithms. The trait measures used in this study were collected from previous literature investigating health care seeking, particularly in the audiological domain. From these, we derived a comprehensive set of 83 features to be used for prediction and profiling. The 3 algorithms taken into account (NB, RF, and SVM) accurately predicted help-seeking behavior at the end of the study in 65.9% (122/185) to 70.3% (130/185) of cases. In particular, the RF algorithm achieved high specificity, meaning that it was most successful in identifying individuals who might not intend to seek professional help. By selecting a subset of important traits revealed by our empirical feature importance analyses to predict hearing help seeking, the prediction accuracy for action to seek help at the 2-month follow-up reached 74.8% (98/131). This study identified the following features to be most important in the prediction of help-seeking behavior: perceived consequences of HL in daily life, motivation to seek help, attitudes toward HAs, sensory sensitivity, neuroticism, physical health, and income. We conclude that these individual characteristics should be assessed in a profiling module that could complement the main auditory assessment module for hearing screening of existing or future mHealth apps. The degree of HL but, importantly, also its day-to-day variability were among the most important predictors, suggesting the need to perform repeated hearing assessments, which could be prompted by the app at different times of the day. To streamline the implementation of the profiling module in a mobile app, the questionnaires and subscales used to measure these important features should undergo item selection analysis to derive simple and short yet reliable and valid scales. By incorporating a selected machine learning algorithm (RF), the app can profile users into help seekers or nonseekers based on the data collected through this short questionnaire battery. This information would complement the audiological data gathered through existing hearing screening or diagnostic tests, providing an informative user profile. The derived profile would guide the app in selecting the appropriate set of recommendations, optimizing an intervention on help-seeking behavior where needed. The results of this study will also provide suggestions for the design of such targeted treatment recommendations. Ultimately, our aim was to provide clinicians and mHealth app developers with relevant knowledge to promote hearing health by encouraging the uptake of hearing health care services and HAs when needed.

Best Machine Learning Model to Predict Help Seeking and Categorize Individuals Into Help Seekers Versus Nonseekers

In total, 3 machine learning classifiers correctly predicted *action to seek help* at study end in 65.9% (122/185) to 70.3% (130/185) of cases, clearly improving over chance prediction. This is a promising result considering the complexity of the prediction outcome. As discussed previously, several individual factors can influence the decision to seek hearing health care services, and there can be discrepancy among contemplating, planning, and taking concrete action [3]. RF showed the best prediction accuracy and high specificity, whereas NB showed the highest sensitivity. When predicting *action to seek help at follow-up* using the selected important features, the performance of the RF and NB models improved up to 75% despite the smaller data set (n=131). NB showed higher predictive performance for this outcome. Overall, all models exhibited high specificity (ranging from 75% to 86%) and comparatively low sensitivity (31% to 58%). This could be attributed to the fact that the *action* class encompassed individuals who had sought professional help as well as those who were only considering taking action. RF showed high accuracy in identifying the *no action* class both at study end and at follow-up and, therefore, can be considered the best-performing algorithm in this framework. Accurate identification of nonseekers is the most relevant performance outcome in an mHealth app to design targeted recommendations. Indeed, the envisioned profiling algorithm should be a system with high specificity that motivates and promotes help seeking, especially in those cases in which users would not spontaneously take action.

Most Relevant Hearing-Related and Psychological Features to Classify Individuals Into Help Seekers Versus Nonseekers

Hearing performance appears to be one of the most important features to predict help seeking. The association between degree of HL and help seeking, as well as HA uptake, is well established in the literature [4,9,26,28,29]. These results also highlight—to our knowledge, for the first time in the literature—the predictive role of intraindividual fluctuations in hearing performance, emphasizing the need to move beyond the traditional view of hearing as a stable neurosensory process [88]. The implementation of repeated daily measurements of hearing performance provides further insights on the impact of HL on the individual's everyday life. In line with this, feature importance findings emphasize the relevance of self-reports on the consequences of HL. The assessment should consider self-reported listening effort in different contexts as well as perceived handicap and emotional consequences of HL. Indeed, individuals who report greater negative impact of HL in their lives are more prone to seek help and later uptake HAs [9,27]. Individuals' self-awareness of HL can be validated or improved by providing repeated feedback on hearing performance in an mHealth hearing app. As observed at the follow-up survey, 85.5% (112/131) of participants reported increased awareness of their hearing abilities after receiving repeated feedback during the study. Finally, according to these results and previous findings [26,27,29], investigating stigma, attitude, and expectations regarding HAs informs on individuals' readiness to seek help as well as later uptake of an HA. Stigma and negative stereotypes related to HAs may deter individuals from seeking help and can represent a barrier to HA use [7,28].

Audiological factors emerged as the most important features. Nevertheless, other general health and psychological factors were also relevant in the prediction of help seeking. In this study, physical health was an important predictor for help seeking, although the evidence on this relationship is discordant [26]. While people with better self-reported health were more likely to seek help [4,27], HA uptake was predicted by poor self-reported health [6]. Other important factors were related to the personality traits of sensory sensitivity and neuroticism. Individuals characterized by high sensitivity to sensory stimuli [63] and emotionally instable personality traits seem to perceive increased psychological discomfort following HL even in the presence of effective HA treatment [2]. Finally, income emerged as another important predictive feature. This is in line with evidence suggesting that higher socioeconomic status [9], higher income or pension earnings [3,27], and access to financial support [26] promote HA uptake.

Feature Importance Measures Can Inform the Design of Targeted Recommendations for Users of a Future mHealth Hearing App

By assessing and analyzing the aforementioned hearing-related and psychological traits, the algorithm developed in this study aims to profile the user as help seeker or nonseeker. Completing this profile with complementary audiological test results, an informative picture of the user can be derived. Using this profile, clinical experts and intervention app designers could propose different sets of recommendations to assist individuals in their decision-making process in a targeted manner. We propose to first differentiate between profiled help seekers and nonseekers, where the former should receive simple and straightforward recommendations only depending on their hearing status. For nonseekers, there is a need to design more specific and targeted recommendations based on information about the relevant characteristics to predict HA seeking. Users who were profiled as determined help seekers could receive clear and concise guidance on the hearing care they need. Those among them who should uptake a hearing device (given their audiological outcome) could benefit from additional information on available hearing care services and professionals to facilitate faster HA adoption rates. This would facilitate individuals' perceived competence and autonomy, which are important predictors of hearing health–seeking behavior [89]. On the other hand, users with HL who are profiled as nonseekers should receive more elaborate, targeted recommendations to motivate and promote access to hearing care services. Recommendations for nonseekers should be further differentiated and designed depending on their perceived consequences of HL; attitudes toward HAs; sensory sensitivity; neuroticism; and, potentially, income. These recommendations could act as an intervention on modifiable predictive features such as self-recognition of HL and attitude toward HAs. For example, users profiled as nonseekers with good awareness and self-recognition of HL but negative expectations regarding HAs and low income should receive a different set of recommendations than nonseekers with low self-awareness and high neuroticism.

Where HL is detected, the mHealth app could prompt repeated testing on different days and at different times of the day and provide individual feedback on the performance compared to normative data. More detailed feedback on daily hearing performance could improve awareness of the hearing deficit. The app could also inform the user of the risks of an untreated HL and the benefits of early intervention through HAs. Indeed, it has been shown that individuals are more likely to positively change their behavior when provided with actionable and meaningful information on their health status [90]. Where there is a need to promote positive attitudes toward HAs, information could be provided on the wide range of devices available as well as examples of successful peer cases. Knowledge of accessible financial support for HAs by insurance companies could additionally promote HA uptake given the predictive role of income. Furthermore, an implemented HA simulator in an mHealth hearing app could offer possibilities to experience improved listening conditions and promote positive expectations regarding an HA. Elaborate information on HA technologies, such as the benefits of noise control and noise reduction algorithms, could promote help seeking by fostering the knowledge of individuals who are more sensitive to environmental noise (high sensory sensitivity trait). The effectiveness of such recommendations could be further increased through targeting or tailoring communication. Targeted messages are designed for a specific population, whereas tailored communication is individualized to the person and has been shown to be most effective in promoting health behavior change [91]. Indeed, messages that are congruent with the personality traits of the audience are more positively evaluated and persuasive and generate more interest [92]. The predictive role of neuroticism for help-seeking behavior can be considered for efficient communication both in the context of an mHealth hearing app and in clinical counseling. Individuals with a high neuroticism trait are more susceptible to perceived disease [93] and are drawn to action through motives of safety and security [92]. For example, recommendations that target profiled nonseekers with low expectations regarding HAs can be differentiated depending on personality traits. To promote positive HA expectations in individuals with high sensory sensitivity traits, recommendations could focus on the benefits of HAs related to noise control and noise suppression. However, such recommendations may not be effective for people who do not have this high sensitivity trait and who, for example, score high on neuroticism. Instead, they might be more convinced by recommendations that emphasize the risks of an untreated HL and the benefits of an early intervention through HAs.

Limitations and Future Directions

The predictive performance of the machine learning classifiers could be improved in future studies using a larger data set and more balanced classes. Classification accuracy could be further improved by including additional objective measures to complement participant self-reports. Continuous psychophysiological measurements (eg, heart rate variability) could be included as further predictive features. This information could complement the longitudinal assessment of affect and better characterize potential changes in arousal before and after the completion of the auditory measurements. Note that multicollinearity as a potential statistical limitation was ruled out (the correlation plots are available in Multimedia Appendix 4). Future studies might also benefit from a longer follow-up period to properly capture those individuals who took more time to take action to seek help. In this study, measuring help seeking 2 months after the end of the study provided a more valid measure of participants' behavior. For example, of the participants who were categorized as nonseekers at study end, 3.8% (7/185) reported having made an appointment with a hearing professional at follow-up and 6.5% (12/185) were planning to do so in the near future. Another limitation that affects the generalizability of the findings is the specific sample included in this study—older individuals living in Germany, using a smartphone in their daily life, mainly coming from big cities (106/185, 57.3%), and having a monthly income above the national average net salary (€2500 [US $2685.35]) in 31.4% (58/185) of cases. The conclusions about relevant personal factors for predicting hearing help seeking may not be generalizable to people with HL in different socioeconomic situations. For example, socioeconomic factors were found to be the major limiting factor in seeking help for hearing difficulties in a South African periurban community [94]. To address this sociodemographic limitation, future studies may consider alternative recruitment strategies to achieve a more diverse sociodemographic sample. These findings may also not generalize to different age groups. For example, when considering young adults with HL, other personal characteristics may be more important in predicting help seeking.

Looking forward, this study sets milestones for the development and implementation of a short and concise profiling module in an existing or future mobile app for hearing screening linked to targeted recommendations, complementing the audiological assessment in a modular environment. In the context of this mobile study, individuals could provide any type of feedback in an open-question format at the end of the study. Of 185 individuals who completed the study, only 1 (0.005%) participant raised concerns related to usability and the user interface. This participant suggested enhancing the contrast between the fonts and the background and increasing the size of the click buttons. As this feedback was provided toward the end of our data collection period, we were unable to implement this suggestion in our design. Further studies that focus on user interface and usability are necessary for the future implementation of such a module in mHealth solutions that target an older population. Specific implementation strategies should be considered, such as simplicity of design; naturalness of navigation and task flow; clear interface elements; feedback [95,96]; large font sizes; contrasting colors; and clear, consistent, and simple instructions [11]. Perceived ease of use and perceived usefulness should be targeted to promote acceptance and use of mHealth solutions [11].

Conclusions

This research provides initial knowledge regarding a selection of tests and questionnaires that have been shown to predict hearing help seeking in persons with self-reported hearing difficulties. From these, we derived conclusions for the implementation of an individual-profiling algorithm in an mHealth hearing app. This study is innovative in that it considers a comprehensive range of personal characteristics and covariates previously cited in the literature, including 25 assessment tools and 83 features, and narrows them down to identify a short selection of the most important predictors for profiling. Complementing the audiological assessments with such a profiling algorithm will enable an mHealth app to deliver targeted and efficient treatment recommendations depending on relevant individual characteristics. The benefits of such a profiling module might also extend to other functions within an mHealth hearing app. Future studies might explore potential relationships between psychological traits and, among others, HA fitting preferences and endurance in the fine-tuning process toward an optimal aiding solution; openness to try new, elaborate technical solutions; or preference for particular app usability features. We have seen how predictive models that use machine learning algorithms can be used to explore complex association patterns of individual characteristics and behaviors considering multiple predictors simultaneously and drawing robust conclusions through CV approaches. This provides further evidence of the advancement that the use of machine learning algorithms can bring to mHealth technology development [33]. mHealth solutions contribute to the evolution of hearing health care toward predictive, preventive, personalized, and participatory medicine (P4 medicine) [90].

We have seen how individual profiling in an mHealth hearing app can identify nonseekers, acting as a preventive action to reduce the risk of a late intervention for HL. It can also provide clinicians with data-driven insights on the individual health profile of the user for tailored and personalized treatments. Moreover, it can enhance the empowerment and participation of the individual in their own hearing health care, promoting informed decision-making. Indeed, personalization strategies (such as tailored treatment recommendations) increase the effectiveness of mHealth interventions [97]. To conclude, an mHealth hearing app that provides targeted treatment recommendations could facilitate faster access to hearing care services and subsequent earlier intervention where needed to pursue the long-term goal of achieving "hearing for all."

Acknowledgments

This research was funded by the Deutsche Forschungsgemeinschaft (German Research Foundation) under Germany's Excellence Strategy (EXC 2177/1; project ID 390895286). The authors thank all the participants who took part in this study and the universities that helped them with recruitment: Carl von Ossietzky University of Oldenburg, Guest study program; Goethe-University of Frankfurt am Main, University of the 3rd Age; Free University of Berlin, Guest-study-Card Program; University of Cologne, Guest and Senior Citizen Studies; University of Kassel, Guest Students Program; and University of Bielefeld, Studying from 50.

Data Availability

The data sets generated during and analyzed during this study are available from the corresponding authors on reasonable request. Analysis scripts are available in the Zenodo repository [98]. A preprint of the manuscript was published on medRxiv [99] in February 2023 and revised in August 2023. Preliminary results of this research have been presented at symposia and conferences of the Hearing4All Cluster of Excellence, at the Virtual Conference of Computational Audiology 2022, and at the German Society for Psychology Congress 2022. The data will be shared with interested researchers upon request as the data set will be used for further projects and cannot be fully shared at the current time point.

Authors' Contributions

GA, AH, and MB conceptualized the study and were involved in protocol development, study design, and data analysis. IK was involved in study design and data analysis. BK contributed to study conceptualization and obtained funding. GA was responsible for participant recruitment and data collection and wrote the first draft of the manuscript. All authors reviewed and edited the manuscript and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study design overview.
[PDF File (Adobe PDF File), 250 KB - humanfactors_v11i1e52310_app1.pdf]

Multimedia Appendix 2
Hearing test feedback.
[PDF File (Adobe PDF File), 223 KB - humanfactors_v11i1e52310_app2.pdf]

Multimedia Appendix 3
Feature importance values.
[PDF File (Adobe PDF File), 1931 KB - humanfactors_v11i1e52310_app3.pdf]

Multimedia Appendix 4
Feature correlation plots.
[PDF File (Adobe PDF File), 303 KB - humanfactors_v11i1e52310_app4.pdf]

References

1. GBD 2019 Hearing Loss Collaborators. Hearing loss prevalence and years lived with disability, 1990-2019: findings from the Global Burden of Disease Study 2019. Lancet 2021 Mar 13;397(10278):996-1009 [FREE Full text] [doi: 10.1016/S0140-6736(21)00516-X] [Medline: 33714390]
2. Cox RM, Alexander GC, Gray GA. Who wants a hearing aid? Personality profiles of hearing aid seekers. Ear Hear 2005 Feb;26(1):12-26. [doi: 10.1097/00003446-200502000-00002] [Medline: 15692301]
3. Saunders GH, Chisolm TH, Wallhagen MI. Older adults and hearing help-seeking behaviors. Am J Audiol 2012 Dec;21(2):331-337. [doi: 10.1044/1059-0889(2012/12-0028)] [Medline: 23233518]
4. Sawyer CS, Armitage CJ, Munro KJ, Singh G, Dawes PD. Biopsychosocial classification of hearing health seeking in adults aged over 50 years in England. Ear Hear 2020;41(5):1215-1225 [FREE Full text] [doi: 10.1097/AUD.0000000000000839] [Medline: 31985532]
5. Fisher DE, Li CM, Hoffman HJ, Chiu MS, Themann CL, Petersen H, et al. Sex-specific predictors of hearing-aid use in older persons: the age, gene/environment susceptibility - Reykjavik study. Int J Audiol 2015;54(9):634-641 [FREE Full text] [doi: 10.3109/14992027.2015.1024889] [Medline: 25816699]
6. Nixon G, Sarant J, Tomlin D, Dowell R. Hearing aid uptake, benefit, and use: the impact of hearing, cognition, and personal factors. J Speech Lang Hear Res 2021 Feb 17;64(2):651-663. [doi: 10.1044/2020_JSLHR-20-00014] [Medline: 33493400]
7. Southall K, Gagné JP, Jennings MB. Stigma: a negative and a positive influence on help-seeking for adults with acquired hearing loss. Int J Audiol 2010 Nov;49(11):804-814. [doi: 10.3109/14992027.2010.498447] [Medline: 20831460]
8. Yong M, Willink A, McMahon C, McPherson B, Nieman CL, Reed NS, et al. Access to adults' hearing aids: policies and technologies used in eight countries. Bull World Health Organ 2019 Oct 01;97(10):699-710 [FREE Full text] [doi: 10.2471/BLT.18.228676] [Medline: 31656335]
9. Simpson AN, Matthews LJ, Cassarly C, Dubno JR. Time from hearing aid candidacy to hearing aid adoption: a longitudinal cohort study. Ear Hear 2019;40(3):468-476 [FREE Full text] [doi: 10.1097/AUD.0000000000000641] [Medline: 30085938]
10. The mobile economy 2024. GSM Association. URL: https://www.gsma.com/mobileeconomy [accessed 2024-01-25]
11. Hsieh KL, Fanning JT, Rogers WA, Wood TA, Sosnoff JJ. A fall risk mHealth app for older adults: development and usability study. JMIR Aging 2018 Nov 20;1(2):e11569. [doi: 10.2196/11569] [Medline: 31518234]
12. Ahmad NA, Mat Ludin AF, Shahar S, Mohd Noah SA, Mohd Tohit N. Willingness, perceived barriers and motivators in adopting mobile applications for health-related interventions among older adults: a scoping review protocol. BMJ Open 2020 Mar 16;10(3):e033870 [FREE Full text] [doi: 10.1136/bmjopen-2019-033870] [Medline: 32184309]
13. Tong C, Kernoghan A, Lemmon K, Fernandes P, Elliott J, Sacco V, et al. Lessons and reflections from an extended co-design process developing an mHealth app with and for older adults: multiphase, mixed methods study. JMIR Aging 2022 Oct 28;5(4):e39189 [FREE Full text] [doi: 10.2196/39189] [Medline: 36306166]
14. Masalski M, Grysiński T, Kręcicki T. Hearing tests based on biologically calibrated mobile devices: comparison with pure-tone audiometry. JMIR Mhealth Uhealth 2018 Jan 10;6(1):e10 [FREE Full text] [doi: 10.2196/mhealth.7800] [Medline: 29321124]
15. Bright T, Pallawela D. Validated smartphone-based apps for ear and hearing assessments: a review. JMIR Rehabil Assist Technol 2016 Dec 23;3(2):e13 [FREE Full text] [doi: 10.2196/rehab.6074] [Medline: 28582261]
16. Li LY, Wang SY, Wu CJ, Tsai CY, Wu TF, Lin YS. Screening for hearing impairment in older adults by smartphone-based audiometry, self-perception, HHIE screening questionnaire, and free-field voice test: comparative evaluation of the screening accuracy with standard pure-tone audiometry. JMIR Mhealth Uhealth 2020 Oct 27;8(10):e17213 [FREE Full text] [doi: 10.2196/17213] [Medline: 33107828]
17. Lin HY, Chu YC, Lai YH, Cheng HL, Lai F, Cheng YF, et al. A smartphone-based approach to screening for sudden sensorineural hearing loss: cross-sectional validity study. JMIR Mhealth Uhealth 2020 Nov 11;8(11):e23047 [FREE Full text] [doi: 10.2196/23047] [Medline: 33174845]
18. Chen CH, Lin HY, Wang MC, Chu YC, Chang CY, Huang CY, et al. Diagnostic accuracy of smartphone-based audiometry for hearing loss detection: meta-analysis. JMIR Mhealth Uhealth 2021 Sep 10;9(9):e28378 [FREE Full text] [doi: 10.2196/28378] [Medline: 34515644]
19. Paglialonga A, Cleveland Nielsen A, Ingo E, Barr C, Laplante-Lévesque A. eHealth and the hearing aid adult patient journey: a state-of-the-art review. Biomed Eng Online 2018 Jul 31;17(1):101 [FREE Full text] [doi: 10.1186/s12938-018-0531-3] [Medline: 30064497]
20. Frisby C, Eikelboom RH, Mahomed-Asmail F, Kuper H, Moore DR, de Kock T, et al. Mobile health hearing aid acclimatization and support program in low-income communities: feasibility study. JMIR Form Res 2023 Aug 23;7:e46043 [FREE Full text] [doi: 10.2196/46043] [Medline: 37610802]
21. D'Onofrio KL, Zeng FG. Tele-audiology: current state and future directions. Front Digit Health 2021;3:788103 [FREE Full text] [doi: 10.3389/fdgth.2021.788103] [Medline: 35083440]
22. Ratanjee-Vanmali H, Swanepoel DW, Laplante-Lévesque A. Patient uptake, experience, and satisfaction using web-based and face-to-face hearing health services: process evaluation study. J Med Internet Res 2020 Mar 20;22(3):e15875 [FREE Full text] [doi: 10.2196/15875] [Medline: 32196459]

23. Ratanjee-Vanmali H, Swanepoel DW, Laplante-Lévesque A. Characteristics, behaviours and readiness of persons seeking hearing healthcare online. Int J Audiol 2019 Feb;58(2):107-115 [FREE Full text] [doi: 10.1080/14992027.2018.1516895] [Medline: 30289050]
24. Frisby C, Eikelboom R, Mahomed-Asmail F, Kuper H, Swanepoel DW. m-health applications for hearing loss: a scoping review. Telemed J E Health 2022 Aug;28(8):1090-1099. [doi: 10.1089/tmj.2021.0460] [Medline: 34967683]
25. Swanepoel DW. Advancing equitable hearing care: innovations in technology and service delivery. Folia Phoniatr Logop 2023;75(4):201-207. [doi: 10.1159/000530671] [Medline: 37062271]
26. Knoetze M, Manchaiah V, Mothemela B, Swanepoel DW. Factors influencing hearing help-seeking and hearing aid uptake in adults: a systematic review of the past decade. Trends Hear 2023;27:23312165231157255 [FREE Full text] [doi: 10.1177/23312165231157255] [Medline: 36798964]
27. Meyer C, Hickson L, Lovelock K, Lampert M, Khan A. An investigation of factors that influence help-seeking for hearing impairment in older adults. Int J Audiol 2014 Feb;53 Suppl 1:S3-17. [doi: 10.3109/14992027.2013.839888] [Medline: 24405446]
28. Jenstad L, Moon J. Systematic review of barriers and facilitators to hearing aid uptake in older adults. Audiol Res 2011 Mar 23;1(1):e25 [FREE Full text] [doi: 10.4081/audiores.2011.e25] [Medline: 26557310]
29. Knudsen LV, Oberg M, Nielsen C, Naylor G, Kramer SE. Factors influencing help seeking, hearing aid uptake, hearing aid use and satisfaction with hearing aids: a review of the literature. Trends Amplif 2010 Sep;14(3):127-154 [FREE Full text] [doi: 10.1177/1084713810385712] [Medline: 21109549]
30. Schneider JM, Gopinath B, McMahon CM, Britt HC, Harrison CM, Usherwood T, et al. Role of general practitioners in managing age-related hearing loss. Med J Aust 2010 Jan 04;192(1):20-23. [doi: 10.5694/j.1326-5377.2010.tb03395.x] [Medline: 20047543]
31. Convery E, Hickson L, Meyer C, Keidser G. Predictors of hearing loss self-management in older adults. Disabil Rehabil 2019 Aug;41(17):2026-2035. [doi: 10.1080/09638288.2018.1457091] [Medline: 29587551]
32. Meyer C, Hickson L, Fletcher A. Identifying the barriers and facilitators to optimal hearing aid self-efficacy. Int J Audiol 2014 Feb;53 Suppl 1:S28-S37. [doi: 10.3109/14992027.2013.832420] [Medline: 24447235]
33. An Q, Rahman S, Zhou J, Kang JJ. A comprehensive review on machine learning in healthcare industry: classification, restrictions, opportunities and challenges. Sensors (Basel) 2023 Apr 22;23(9):4178 [FREE Full text] [doi: 10.3390/s23094178] [Medline: 37177382]
34. Rooksby J, Asadzadeh P, Morrison A, McCallum C, Gray C, Chalmers M. Implementing ethics for a mobile app deployment. In: Proceedings of the 28th Australian Conference on Computer-Human Interaction. 2016 Presented at: OzCHI '16; November 29-December 2, 2016; Launceston, Australia. [doi: 10.1145/3010915.3010919]
35. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Privacy and security in mobile health apps: a review and recommendations. J Med Syst 2015 Jan;39(1):181. [doi: 10.1007/s10916-014-0181-3] [Medline: 25486895]
36. Arslan RC, Walther MP, Tata CS. formr: a study framework allowing for automated feedback generation and complex longitudinal experience-sampling studies using R. Behav Res Methods 2020 Feb;52(1):376-387 [FREE Full text] [doi: 10.3758/s13428-019-01236-y] [Medline: 30937847]
37. Communication APIs for SMS, voice, email and authentication. Twilio. URL: https://www.twilio.com/ [accessed 2023-01-08]
38. Kiessling J, Grugel L, Meister H, Meis M. Übertragung der Fragebögen SADL , ECHO und SSQ ins Deutsche und deren evaluation. Zeitschrift für Audiologie 2011 Jan;50:6-16 [FREE Full text]
39. Nuesse T, Schlueter A, Lemke U, Holube I. Self-reported hearing handicap in adults aged 55 to 81 years is modulated by hearing abilities, frailty, mental health, and willingness to use hearing aids. Int J Audiol 2021;60(sup2):71-79. [doi: 10.1080/14992027.2020.1858237] [Medline: 33459099]
40. Cohen S, Doyle WJ, Skoner DP, Rabin BS, Gwaltney JMJ. Social ties and susceptibility to the common cold. JAMA 1997 Jun 25;277(24):1940-1944. [Medline: 9200634]
41. Saunders GH, Cienkowski KM, Forsline A, Fausti S. Normative data for the attitudes towards loss of hearing questionnaire. J Am Acad Audiol 2005 Oct;16(9):637-652. [doi: 10.3766/jaaa.16.9.2] [Medline: 16515136]
42. Zimmer K, Ellermeier W. Eine deutsche version der Lärmempfindlichkeitsskala von Weinstein. Zeitschrift für Lärmbekämpfung 1997;44(1997):107-110 [FREE Full text]
43. Meis M, Huber R, Fischer RL, Schulte M, Spilski J, Meister H. Development and psychometric properties of the sound preference and hearing habits questionnaire (SP-HHQ). Int J Audiol 2018 Jun;57(sup3):S118-S129. [doi: 10.1080/14992027.2016.1248795] [Medline: 27875658]
44. Jordan S, Hoebel J. [Health literacy of adults in Germany: findings from the German Health Update (GEDA) study]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2015 Sep;58(9):942-950. [doi: 10.1007/s00103-015-2200-z] [Medline: 26227894]
45. Buschermöhle M, Wagener K, Berg D, Meis M, Kollmeier B. The German digits triplets test (part I): implementations for telephone, internet and mobile devices. Audiol Acoust Für Audiol 2014;54(4):139-145 [FREE Full text]
46. Borkenau P, Ostendorf F. NEO-Fünf-Faktoren-Inventar (NEO-FFI) nach Costa und McCrae : Handanweisung. Göttingen, Germany: Hogrefe Publishing Group; 1993.

47. Gottschling J, Segal DL, Häusele C, Spinath FM, Stoll G. Assessment of anxiety in older adults: translation and psychometric evaluation of the German version of the geriatric anxiety scale (GAS). J Psychopathol Behav Assess 2015 Aug 6;38(1):136-148. [doi: 10.1007/s10862-015-9504-z]
48. Yesavage JA, Sheikh JI. 9/Geriatric Depression Scale (GDS). Clin Gerontol 2008 Oct 25;5(1-2):165-173. [doi: 10.1300/J018v05n01_09]
49. Kemper CJ, Beierlein C, Kovaleva A, Rammstedt B. Eine kurzskala zur messung von optimismus pessimismus: die skala optimismus-pessimismus-2 (SOP2). GESIS Working Papers. 2012. URL: https://www.gesis.org/fileadmin/kurzskalen/working_papers/SOP2_Workingpaper.pdf [accessed 2024-07-03]
50. Tesch-Römer C, Wiest M, Wurm S, Huxhold O. [Loneliness trends in the second half of life: results from the German Ageing Survey (DEAS)]. Z Gerontol Geriatr 2013 Apr;46(3):237-241. [doi: 10.1007/s00391-012-0359-6] [Medline: 22767399]
51. Konrad S, Herzberg PY. Psychometric properties and validation of a German High Sensitive Person Scale (HSPS-G). Eur J Psychol Assess 2019 May;35(3):364-378. [doi: 10.1027/1015-5759/a000411]
52. Lohaus A, Schmitt GM. Kontrollüberzeugungen zu Krankheit und Gesundheit (KKG): testverfahren und testmanual. Hogrefe. 1989. URL: https://pub.uni-bielefeld.de/record/2919685#ama [accessed 2024-07-03]
53. Kovaleva A, Beierlein C, Kemper CJ, Rammstedt B. Eine kurzskala zur messung von kontrollüberzeugung: die akala Internale-Externale-Kontrollüberzeugung-4 (IE-4). GESIS-Working Papers. 2012. URL: https://www.gesis.org/fileadmin/kurzskalen/working_papers/IE4_Workingpaper.pdf [accessed 2024-07-03]
54. Laidlaw K, Power MJ, Schmidt S, WHOQOL-OLD Group. The Attitudes to Ageing Questionnaire (AAQ): development and psychometric properties. Int J Geriatr Psychiatry 2007 Apr;22(4):367-379. [doi: 10.1002/gps.1683] [Medline: 17051535]
55. Schwarzer R, Jerusalem M. SWE. Skala zur Allgemeinen Selbstwirksamkeitserwartung [Verfahrensdokumentation, Autorenbeschreibung und Fragebogen]. Leibniz-Institut für Psychologie (ZPID). 2003. URL: https://www.psycharchives.org/en/item/d041ef33-7cb7-4b29-a92d-29d900fc462b [accessed 2024-07-03]
56. Schneider EE, Schönfelder S, Domke-Wolf M, Wessa M. Measuring stress in clinical and nonclinical subjects using a German adaptation of the Perceived Stress Scale. Int J Clin Health Psychol 2020;20(2):173-181 [FREE Full text] [doi: 10.1016/j.ijchp.2020.03.004] [Medline: 32550857]
57. Schipolowski S, Wilhelm O, Schroeders U. Berliner Test zur Erfassung fluider und kristalliner Intelligenz für die 11. und 12. Jahrgangsstufe. Göttingen, Germany: Hogrefe Publishing; 2020.
58. Schmidt KH, Metzler P. Wortschatztest (WST). Weinheim, Germany: Beltz Test GmbH; 1992.
59. Neyer FJ, Felber J, Gebhardt C. Entwicklung und validierung einer kurzskala zur erfassung von technikbereitschaft. Diagnostica 2012 Apr;58(2):87-99. [doi: 10.1026/0012-1924/a000067]
60. Drixler K, Morfeld M, Glaesmer H, Brähler E, Wirtz MA. [Validation of the Short-Form-Health-Survey-12 (SF-12 Version 2.0) assessing health-related quality of life in a normative German sample]. Z Psychosom Med Psychother 2020 Sep;66(3):272-286. [doi: 10.13109/zptm.2020.66.3.272] [Medline: 32876561]
61. Senese VP, Ruotolo F, Ruggiero G, Iachini T. The Italian version of the Weinstein noise sensitivity scale. Eur J Psychol Assess 2012 Nov;28(2):118-124. [doi: 10.1027/1015-5759/a000099]
62. Goldberg LR. An alternative "description of personality": the Big-Five factor structure. J Pers Soc Psychol 1990;59(6):1216-1229. [doi: 10.1037/0022-3514.59.6.1216]
63. Aron EN, Aron A. Sensory-processing sensitivity and its relation to introversion and emotionality. J Pers Soc Psychol 1997 Aug;73(2):345-368. [doi: 10.1037//0022-3514.73.2.345] [Medline: 9248053]
64. Posner J, Russell JA, Peterson BS. The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. Dev Psychopathol 2005;17(3):715-734 [FREE Full text] [doi: 10.1017/S0954579405050340] [Medline: 16262989]
65. Scott SB, Sliwinski MJ, Zawadzki M, Stawski RS, Kim J, Marcusson-Clavertz D, et al. A coordinated analysis of variance in affect in daily life. Assessment 2020 Dec;27(8):1683-1698. [doi: 10.1177/1073191118799460] [Medline: 30198310]
66. Smits C, Kapteyn TS, Houtgast T. Development and validation of an automatic speech-in-noise screening test by telephone. Int J Audiol 2004 Jan;43(1):15-28. [doi: 10.1080/14992020400050004] [Medline: 14974624]
67. Van den Borre E, Denys S, van Wieringen A, Wouters J. The digit triplet test: a scoping review. Int J Audiol 2021 Dec;60(12):946-963. [doi: 10.1080/14992027.2021.1902579] [Medline: 33840339]
68. Denys S, Hofmann M, van Wieringen A, Wouters J. Improving the efficiency of the digit triplet test using digit scoring with variable adaptive step sizes. Int J Audiol 2019 Oct;58(10):670-677. [doi: 10.1080/14992027.2019.1622042] [Medline: 31187664]
69. Potgieter JM, Swanepoel DW, Myburgh HC, Hopper TC, Smits C. Development and validation of a smartphone-based digits-in-noise hearing test in South African English. Int J Audiol 2015 Jul;55(7):405-411. [doi: 10.3109/14992027.2016.1172269] [Medline: 27121117]
70. Smits C, Merkus P, Houtgast T. How we do it: the Dutch functional hearing-screening tests by telephone and internet. Clin Otolaryngol 2006 Oct;31(5):436-440. [doi: 10.1111/j.1749-4486.2006.01195.x] [Medline: 17014457]

71. Löw T, Bach JH, Buschermöhle M, Bisitz T. Development of new procedures for the Digit Triplets Test as a reliable screening tool. 21. Jahrestagung der Deutschen Gesellschaft für Audiologie. 2018. URL: https://www.dga-ev.com/fileadmin/dga2018/site/data/final/0146.pdf [accessed 2024-02-09]
72. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2019. URL: https://www.r-project.org/ [accessed 2023-01-17]
73. Arslan RC. formr.org survey framework utility R package. Zenodo. 2018 Aug 2. URL: https://zenodo.org/records/1326523 [accessed 2022-12-08]
74. Bischl B, Lang M, Kotthoff L, Schiffner J, Richter J, Studerus E, et al. mlr: machine learning in R. J Mach Learn Res 2017;17(170):1-5 [FREE Full text] [doi: 10.32614/cran.package.mlr]
75. Rhys HI. Machine Learning with R, the Tidyverse, and mlr. Shelter Island, NY: Manning Publications; 2020.
76. Bischl B, Sonabend R, Kotthoff L, Lang M. Applied Machine Learning Using mlr3 in R. Boca Raton, FL: CRC Press; 2023.
77. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning: With Applications in R. New York, NY: Springer; 2013.
78. Probst P, Boulesteix AL, Bischl B. Tunability: importance of hyperparameters of machine learning algorithms. J Mach Learn Res 2019;20:1-32 [FREE Full text]
79. Becker M, Ullmann T, Lang M, Bischl B, Richter J, Binder M. Practical tuning series - tune a support vector machine. mlr-org. 2021 Mar 9. URL: https://mlr-org.com/gallery/optimization/2021-03-09-practical-tuning-series-tune-a-support-vector-machine/index.html [accessed 2022-12-08]
80. Hsu CW, Chang CC, Lin CJ. A practical guide to support vector classification. National Taiwan University. 2016. URL: https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf [accessed 2022-12-08]
81. Ali A, Shamsuddin S, Ralescu A. Classification with class imbalance problem: a review. In: Proceedings of the 10th International Conference on Soft Computing Models in Industrial and Environmental Applications. 2015 Presented at: SOCO 2015; June 15-17, 2015; Burgos, Spain URL: https://www.researchgate.net/publication/288228469
82. Delgado R, Tibau XA. Why Cohen's Kappa should be avoided as performance measure in classification. PLoS One 2019 Sep 26;14(9):e0222916 [FREE Full text] [doi: 10.1371/journal.pone.0222916] [Medline: 31557204]
83. Kuhn M, Vaughan D, Hvitfeldt E. yardstick: tidy characterizations of model performance. The Comprehensive R Archive Network. 2024 Mar 21. URL: https://CRAN.R-project.org/package=yardstick [accessed 2024-07-03]
84. Jurman G, Riccadonna S, Furlanello C. A comparison of MCC and CEN error measures in multi-class prediction. PLoS One 2012;7(8):e41882 [FREE Full text] [doi: 10.1371/journal.pone.0041882] [Medline: 22905111]
85. Saarela M, Jauhiainen S. Comparison of feature importance measures as explanations for classification models. SN Appl Sci 2021 Feb 03;3:272. [doi: 10.1007/s42452-021-04148-9]
86. Kuhn M. The caret package. GitHub. 2019 Mar 27. URL: https://topepo.github.io/caret/index.html [accessed 2022-11-23]
87. Kuhn M, Wing J, Weston S, Williams A, Keefer C, Engelhardt A, R Core Team, et al. caret: classification and regression training. The Comprehensive R Archive Network. 2023 Mar 21. URL: https://CRAN.R-project.org/package=caret [accessed 2024-07-03]
88. Kuhlmann I, Angonese G, Thiel C, Kollmeier B, Hildebrandt A. Are there good days and bad days for hearing? Quantifying day-to-day intraindividual speech perception variability in older and younger adults. J Exp Psychol Hum Percept Perform 2023 Nov;49(11):1377-1394. [doi: 10.1037/xhp0001159] [Medline: 37870818]
89. Ramos MD. Exploring the relationship between planned behavior and self-determination theory on health-seeking behavior among older adults with hearing impairment. Geriatr Nurs 2023;52:1-7. [doi: 10.1016/j.gerinurse.2023.05.001] [Medline: 37187141]
90. Flores M, Glusman G, Brogaard K, Price ND, Hood L. P4 medicine: how systems medicine will transform the healthcare sector and society. Per Med 2013;10(6):565-576 [FREE Full text] [doi: 10.2217/pme.13.57] [Medline: 25342952]
91. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 2007 Jul;133(4):673-693. [doi: 10.1037/0033-2909.133.4.673] [Medline: 17592961]
92. Hirsh JB, Kang SK, Bodenhausen GV. Personalized persuasion: tailoring persuasive appeals to recipients' personality traits. Psychol Sci 2012 Jun;23(6):578-581. [doi: 10.1177/0956797611436349] [Medline: 22547658]
93. Ferguson E. Personality is of central concern to understand health: towards a theoretical model for health psychology. Health Psychol Rev 2013 May;7(Suppl 1):S32-S70 [FREE Full text] [doi: 10.1080/17437199.2010.547985] [Medline: 23772230]
94. Mtimkulu TK, Khoza-Shangase K, Petrocchi-Bartal L. Barriers and facilitators influencing hearing help-seeking behaviors for adults in a peri-urban community in South Africa: a preventive audiology study. Front Public Health 2023 Oct 17;11:1095090 [FREE Full text] [doi: 10.3389/fpubh.2023.1095090] [Medline: 37920587]
95. Wildenbos GA, Peute LW, Jaspers MW. A framework for evaluating mHealth tools for older patients on usability. Stud Health Technol Inform 2015;210:783-787. [Medline: 25991261]
96. Wildenbos GA, Peute L, Jaspers M. Aging barriers influencing mobile health usability for older adults: a literature based framework (MOLD-US). Int J Med Inform 2018 Jun;114:66-75. [doi: 10.1016/j.ijmedinf.2018.03.012] [Medline: 29673606]

97. Rivera-Romero O, Gabarron E, Ropero J, Denecke K. Designing personalised mHealth solutions: an overview. J Biomed Inform 2023 Oct;146:104500 [FREE Full text] [doi: 10.1016/j.jbi.2023.104500] [Medline: 37722446]
98. Angonese G. Machine learning and feature importance for predicting hearing help-seeking. Zenodo. 2023 Feb 14. URL: https://zenodo.org/records/7635920 [accessed 2023-07-09]
99. Angonese G, Buhl M, Kuhlmann I, Kollmeier B, Hildebrandt A. Predicting hearing help-seeking: what features are important for a profiling module within a hearing mHealth application? JMIR Human Factors 2024 May 2:52310 (forthcoming) [FREE Full text] [doi: 10.2196/52310]

Abbreviations

API: application programming interface
CV: cross-validation
HA: hearing aid
HL: hearing loss
MCC: Matthews correlation coefficient
mHealth: mobile health
NA: not available
NB: naïve Bayes
RF: random forest
RQ: research question
SNR: signal-to-noise ratio
SRT: speech recognition threshold
SVM: support vector machine

Edited by D Gooch; submitted 30.08.23; peer-reviewed by Y Chu, M McCullagh; comments to author 02.01.24; revised version received 22.02.24; accepted 02.05.24; published 12.08.24.

Please cite as:
Angonese G, Buhl M, Kuhlmann I, Kollmeier B, Hildebrandt A
Prediction of Hearing Help Seeking to Design a Recommendation Module of an mHealth Hearing App: Intensive Longitudinal Study of Feature Importance Assessment
JMIR Hum Factors 2024;11:e52310
URL: https://humanfactors.jmir.org/2024/1/e52310
doi:10.2196/52310
PMID:39133539

Original Paper

Implementation and Evaluation of a Gait Training Assistant for the Use of Crutches: Usability Study

Milan Anton Wolf[1], MD; Leon Sauerwald[2], MSC; Felix Kosmalla[2], MSC; Florian Daiber[2], PhD; Antonio Krüger[2], PhD; Stefan Landgraeber[1], MD

[1]Department of Ortopedic Surgery, Saarland University Medical Center, Homburg, Germany

[2]German Research Centre for Artificial Intelligence (DFKI), Saarbrücken, Germany

Corresponding Author:
Milan Anton Wolf, MD
Department of Ortopedic Surgery
Saarland University Medical Center
Kirrberger Straße
Homburg, 66421
Germany
Phone: 49 068411624500
Email: milan.wolf@uks.eu

Abstract

Background: Surgical procedures on the lower extremities often require weight-bearing on crutches as part of the rehabilitation process. Orthopedic elective procedures enable patients to learn the correct use of crutches in a controlled preoperative setting. Digital assistance systems can safely circumvent a shortage of skilled staff and any contact restrictions that may be necessary.

Objective: The usability of a newly developed gait training assistant (GTA) for the use of crutches will be evaluated. An intervention group trained to use crutches by the digital trainer will be compared with a control group trained to use crutches conventionally by a physiotherapist.

Methods: As part of the development and implementation of a novel GTA, 14 patients learned to walk with crutches by completing specific exercises while receiving live feedback. Their movements were detected by a depth sensor and evaluated in real time. Specific parameters (step length, synchronous movement, crutch angle, and crutch distance to the feet) were compared with a control group (n=14) trained to use crutches by physiotherapists. The intervention group was also assessed by a physiotherapist. At the end of the study, the patients completed questionnaires to evaluate the usability of the system (Brooke's System Usability Scale score) and patient satisfaction.

Results: All patients trained by the novel GTA were able to use crutches correctly. The intervention group showed significantly better values for crutch angle (mean –6.3°, SD 3.5° vs mean –12.4°, SD 4.5°; $P<.001$) and crutch position (mean 3.3, SD 5.1 cm vs mean –8.5, SD 4.9 cm; $P=.02$). Both groups reported that they felt confident in the use of crutches, were able to follow the instructions, and enjoyed the training. Even though the majority (12/14, 86%) preferred physical therapy over a purely digital approach, most participants enjoyed using the system (13/14, 93%) and were interested in trying out other digital assistants (11/14, 79%). The usability of the GTA was rated above average by the majority (9/14, 64%) of the patients.

Conclusions: The newly designed GTA is a safe method of teaching the use of crutches and is statistically superior to training by a physiotherapist. Even if patients prefer interaction with a physiotherapist over a purely digital approach, digital devices provide a safe and motivating opportunity to learn the essential locomotor skills for rehabilitation.

(JMIR Hum Factors 2024;11:e51898) doi:10.2196/51898

KEYWORDS

telerehabilitation; orthopedics; digital gait trainer; orthopedic; gait; movement; walk; walking; crutch; crutches; sensor; sensors; rehabilitation; usability; digital health; physiotherapy; physical therapy; telehealth; telemedicine; eHealth; virtual; locomotor; locomotion

Introduction

Following accidents or surgical procedures, it is often crucial for the healing process to provide support to the affected body part. In the case of lower limb injuries or surgeries, this can be accomplished by crutches. In the case of elective surgery, patients can receive preoperative instruction in a controlled environment several days prior to the procedure. In Germany, this training is typically provided by specialized physiotherapists.

The use of digital assistive devices has already become ubiquitous in clinical settings and serves to mitigate the strain on hospital personnel, reduce costs, and enhance patient rehabilitation [1]. Despite these advantages, advanced technology applications are currently used in patient with orthopedic care only in the context of studies [2-4].

Walking on crutches is a clearly structured movement sequence. However, the teaching of this skill requires significant staff resources. The projected increase in the number of elective orthopedic surgeries in the future is exacerbating the already persistent shortage of specialized health care personnel [5], which has been further intensified by the COVID-19 pandemic and the associated restrictions on physical contact [6].

The current digital aids for learning to walk with crutches do not provide the comprehensive support that a physiotherapist can offer [7,8]. Some crutches equipped with sensors can monitor proper use, but there are currently no commercially available systems that provide simultaneous guidance and assessment, which puts a strain on the resources of physiotherapists.

This study aims to evaluate the efficacy of a custom-designed gait training assistant (GTA) in teaching the proper use of crutches. The primary objectives of this study are to determine (1) the capability of the digital assistant to impart proper crutch-walking skills using a 3-point gait, (2) patient receptiveness toward the digital walking assistant, and (3) the competence of the digital crutch-walking training assistant in comparison to a human physiotherapist.

Methods

Study Design

A total of 28 patients who underwent elective orthopedic surgery and needed postoperative weight-bearing (3-point gait) on the lower extremity were included. Patients who could not follow the instructions of the GTA, for instance, due to neurological diseases, were excluded.

The patients were randomly assigned to the intervention group (n=14) or control group (n=14). The intervention group learned the proper use of crutches with the GTA. The control group learned the use of crutches with the help of a physiotherapist. Baseline data (age and sex) were collected using a survey. In the intervention group, satisfaction was assessed by a questionnaire specially developed for the study. In addition, the usability of the GTA was evaluated using Brooke's System Usability Scale (SUS) [9]. The GTA and a physiotherapist evaluated the ability of both groups to use the crutches correctly.

Ethical Considerations

A prospective clinical trial was conducted in August 2022 after obtaining patient consent and institutional review board approval from the Saarland Medical Association (318/21). Data was anonymized and no compensation was provided to participants.

Implementation of the GTA

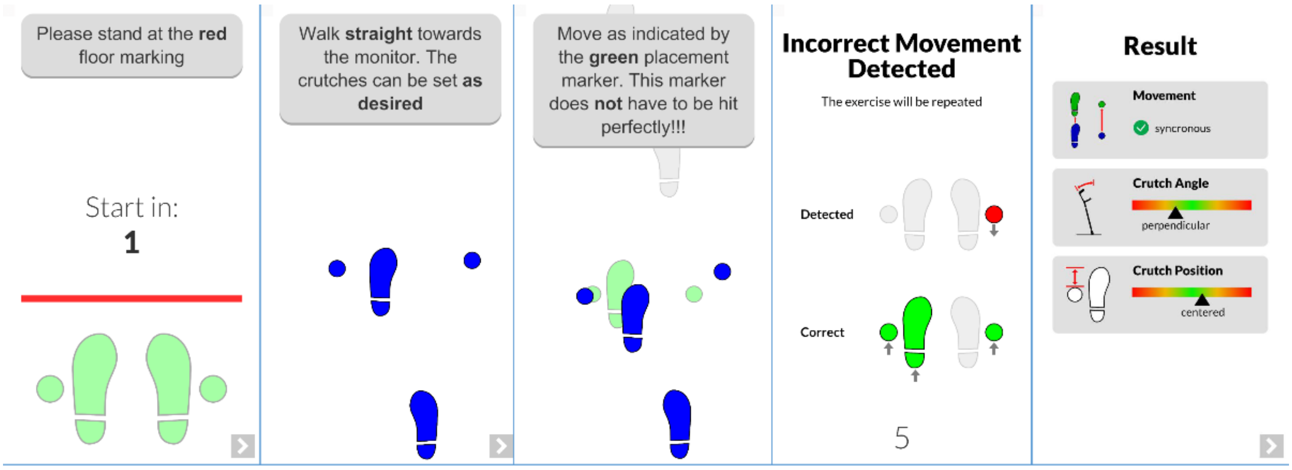
The GTA consists of a screen in portrait mode that gives visual guidance to the user and a 3D camera (Kinect v2 [Microsoft]) that tracks the position of the feet and crutches. While walking toward the screen, the gait pattern is automatically recognized and possible errors are brought to the attention of the user. The screen is also used to give instructions to the user (Figure 1). The user is guided through various exercises that aim to train a specific walking pattern (3-point gait). The software is designed so that each pattern can be described as a sequence of states, where each state specifies the crutch or the foot that needs to be moved. For 3-point gait, for instance, the user needs to move the injured leg, foot, or side and both crutches. Next, the noninjured leg, foot, or side is moved. This pattern is repeated. For simplicity, we focused on 3-point gait and forward movement only during this study, as other walking patterns and backward movement work similarly.

Figure 1. Training with the gait training assistant. The patient uses the crutches while the Kinect camera in front of him analyzes the crutches as well as the feet. The exercise to be completed is displayed with live feedback on the screen above the camera.

For simplicity, 2 infrared markers are attached to each crutch (handle and rubber foot) and each foot of the user to determine the movement of the crutches and feet. These reflective markers are extracted from the 2D infrared image and converted into 3D positions using the Kinect SDK. In this experiment, we focused on the 3-point gait. One step cycle was divided into smaller movements (eg, "Left foot and both crutches moving forward" and "Right foot moving forward") that were derived by the 3D positions of the tracked reflective markers. Besides the necessary compound movements (foot and crutches), individual movements of crutches and feet were also recognized.

As the aim of the system is also to give corrective feedback during the training, the system tries to fit the currently detected movement to the gait pattern. If the system detects a wrong state or a wrong order of movements, it alerts the users and asks them to repeat the exercise (Figure 2).

Figure 2. Excerpt from the instruction display with live feedback from gait training assistant. The first image (left) shows the start screen with the first instruction. The second image represents the instruction to use the crutches. The third image displays the live feedback. The fourth image signals an incorrect movement. The last image (right) provides the analysis of the exercise to the patient.



The 3D positions of the tracked markers are also used to compute a set of metrics while the user is walking. These metrics include crutch angle, crutch position, and the synchronicity of the movement and help to provide feedback to the user after each exercise. The crutch angle describes the tilting of the crutch along the walking path toward the camera when it is placed on the ground. It is computed using the upper and lower markers of the crutch with some additional offset to match the tilting of the crutch cane. For the crutch position, the depth offset between the crutch and the foot is measured after they have moved. As the optimal position is the center of the foot, an offset of 15 cm is subtracted from the tracked foot position. To measure the synchronicity of the movement, the time when the crutch and foot are moving together is divided by the overall moving time for the current state.

Statistical Analysis

The training performance of the GTA was assessed using an ANOVA. The number of correct runs during the analysis phase for each participant (independent variable) was compared between the 2 groups (dependent variable). The 2 groups were compared using a multivariate ANOVA (MANOVA), wherein 2 independent variables (crutch angle and crutch position) were simultaneously used to investigate the differences between the 2 groups. A correlation matrix was used to explore whether there was a correlation between the participants' age and their perceived usability of the system.

Results

Baseline Data

In total, 28 participants were recruited for the primary study. The sample consisted of 13 female and 15 male participants with an age range of 18-73 (mean 41.1, SD 19.1) years. The participants' height ranged from 1.55 m to 1.96 m. Approximately half (15/28, 54%) of the participants had prior experience with crutches, in most cases several years prior to the study. A total of 3 participants had used crutches in the last 12 months, although not for learning a 3-point gait (Table 1).

Table 1. Presentation of the baseline data of the patients.

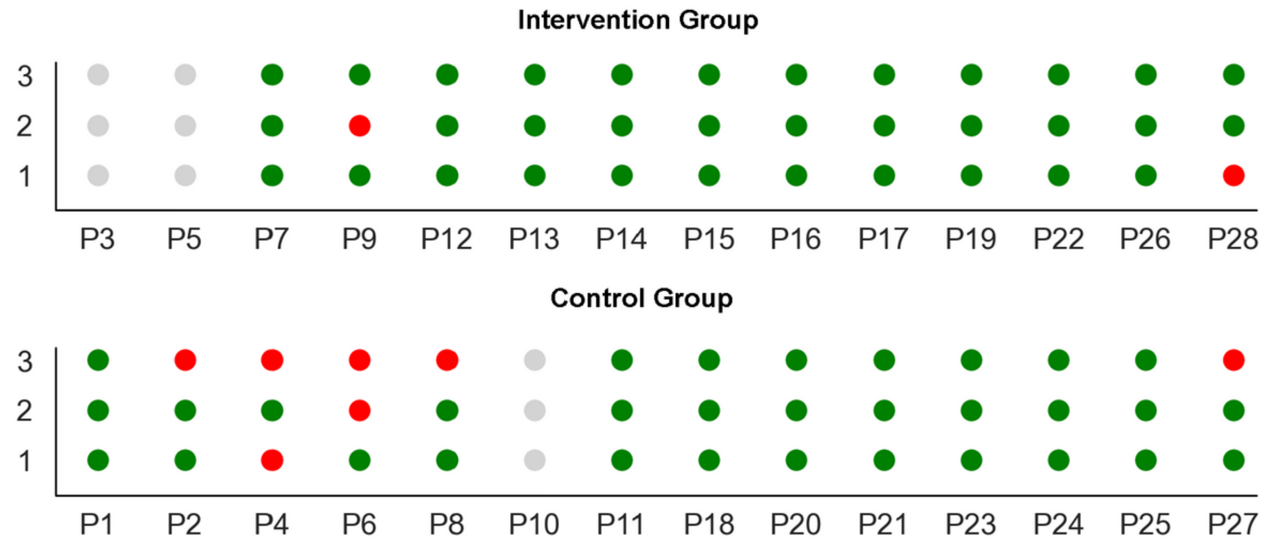
Characteristics	Intervention group (n=14)	Control group (n=14)	Total (n=28)
Sex, n (%)			
Female	6 (43)	7 (50)	13 (46)
Male	8 (57)	7 (50)	15 (54)
Age (years), mean (SD)	44.4 (20.7)	37.8 (17.6)	41.1 (19.1)
Body height (cm), mean (SD)	173.1 (11.6)	175.3 (11.9)	174.2 (11.6)
Used crutches before, n (%)			
Yes	6 (43)	9 (64)	15 (54)
No	8 (57)	5 (36)	13 (46)

A total of 2 participants (P3 and P5; intervention group) could only partially perform the training with the GTA, as after a few exercises, the tracking function of 1 of the crutches was no longer shown in the live position visualization. Subsequent analysis was, therefore, not possible. However, they both filled out all the questionnaires. One other participant (P10; control group) had to be excluded from any analysis including the data recorded by the GTA, as no valid results were computed.

Evaluation of the Training Performance by the GTA

The GTA performance was determined by an analysis of the training sessions. Based on the number of incorrect exercises, the performance was extrapolated. Due to incomplete data collection, P3, P5 (intervention group), and P10 (control group) were excluded from the analysis. In the intervention group, 2 patients were found to have performed 1 of the 3 exercises incorrectly. In the control group, 3 patients performed 1 of the 3 exercises incorrectly and 2 participants performed 2 exercises incorrectly (Figure 3). The differences in the number of correctly performed exercises between the 2 groups were not statistically significant (*P*=.15).

Figure 3. Status for each of the 3 exercises during the analysis: green: success; red: error (incorrect movement); and gray: excluded from the analysis. The upper image shows the results of the intervention group, and the lower image shows the results of the control group.



The parameters selected for comparison of the 2 groups were crutch position and crutch angle. The reason for this is that step length and step speed are individual, patient-specific movements, and the physiotherapists, therefore, had to train synchronous movement depending on the patients' physical condition.

The intervention group showed significantly better values for crutch angle (intervention group vs control group: mean –6.3°, SD 3.5° vs mean –12.4°, SD 4.5°; *P*<.001) and crutch position (intervention group vs control group: mean 3.3, SD 5.1 cm vs mean –8.5, SD 4.9 cm; *P*=.02). A MANOVA confirmed the superiority of the intervention group in the gait metrics (Table 2). Here, group assignment (intervention group and control group) was used as a factor, and crutch angle and crutch position were used as dependent variables. There was homogeneity of the error variances, as assessed by the Levene test (*P*>.05). The homogeneity of covariance was given by the Box test (*P*=.86). All tests revealed a statistical significance for *P*<.05.

Table 2. MANOVA[a] analysis of gait metrics between the intervention group and control group[b].

Test	Value	*F* test (*df*)	*P* value
Pillai trace	0.433	8.39 (2, 22)	.002
Wilks Lambda	0.567	8.39 (2, 22)	.002
Hotelling trace	0.763	8.39 (2, 22)	.002
Roy largest root	0.763	8.39 (2, 22)	.002

[a]MANOVA: multivariate ANOVA.

[b]The analysis showed that the training by the gait training assistant was significantly superior to the training by a physiotherapist.

Evaluation of the GTA Training Performance by a Physiotherapist

The physiotherapist evaluated the ability to walk on crutches based on clinical experience using the parameters of step length, step speed, crutch angle, crutch position, and synchronous movement. Only the intervention group was evaluated, as an assessment of the control groups' training performance by the physiotherapist was estimated to be too biased.

The analysis showed that almost all patients (10/12, 83%) trained on the GTA achieved a perfect result. Only 2 patients in the intervention group (P12 and P13) presented a slight deviation from the ideal score (Table 3).

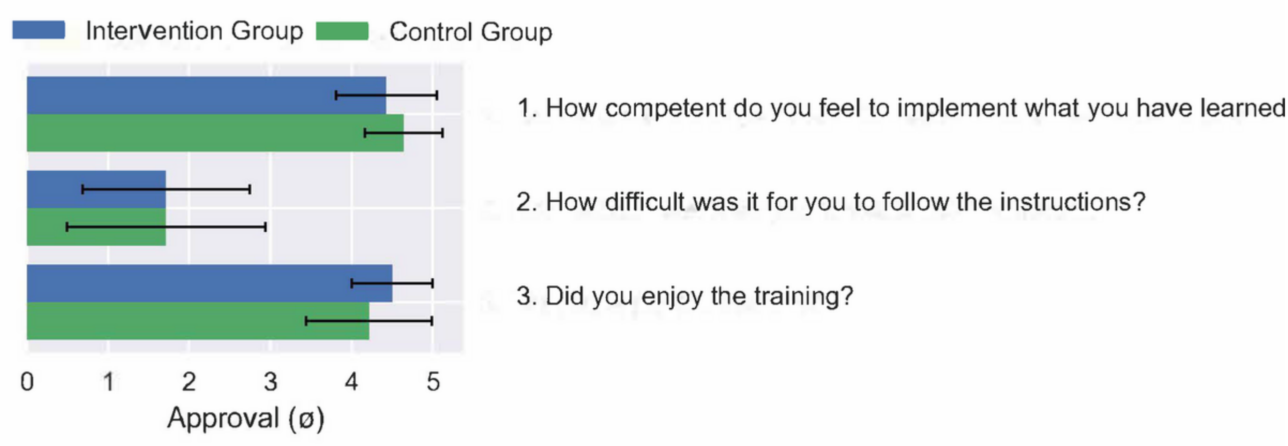
Table 3. Gait training assistant performance of the intervention group, as evaluated by a physiotherapist. Synchronous movement is rated from 1 (not synchronous) to 5 (synchronous). All other metrics are rated from 1 to 5, with 3 as optimal value (eg, rating for step speed: 1=too slow, 3=optimal, and 5=too fast).

Parameters	Patient and rating											
	P7	P9	P12	P13	P14	P15	P16	P17	P19	P22	P26	P28
Step length	3	3	2	3	3	3	3	3	3	3	3	3
Step speed	3	3	2	3	3	3	3	3	3	3	3	3
Crutch angle	3	3	3	3	3	3	3	3	3	3	3	3
Crutch position	3	3	3	3	3	3	3	3	3	3	3	3
Synchronous movement	5	5	5	4	5	5	5	5	5	5	5	5

Training Evaluation

Both groups felt that they were competently trained (intervention group: mean 4.43, SD 0.62; control group: mean 4.64, SD 0.48; Figure 4). All patients reported that they were able to follow the instructions of the physiotherapists and the GTA (intervention group: mean 1.71, SD 1.03; control group: mean 1.71, SD 1.22). The participants in both groups indicated that they enjoyed the training, with the GTA performing slightly better (intervention group: mean 4.5, SD 0.5; control group: mean 4.21, SD 0.77). The differences between the groups were very small and were not statistically significant (MANOVA Wilks Lambda: $F_{3,42}$=0.872; P=.47).

Figure 4. Evaluation of the training and comparison of patients from the control group and Intervention group. Rating from 1 (not at all) to 5 (very). Competence—intervention group: 4.43 (SD 0.62) and control group: 4.64 (SD 0.48). Difficulty—intervention group: 1.71 (SD 1.03) and control group: 1.71 (SD 1.22). Enjoyment—intervention group: 4.5 (SD 0.5) and control group: 4.21 (SD 0.77).



Experience With the GTA

The participants in the intervention group were asked to complete a questionnaire with binary response options (yes or no). All participants indicated that they had learned to use the crutches safely and felt adequately prepared (14/14, 100% agreement), but 43% (6/14) of them indicated a desire for reevaluation by a physiotherapist. Nearly all participants (13/14, 93%) reported that they had understood the instructions provided by the GTA, 86% (12/14) reported that they were able to properly follow the instructions of the GTA, and 79% (11/14) stated that they would like to use a trainer for additional exercises. On the other hand, only 14% (2/14) of the participants wished to have purely digital rehabilitation without the support of a physiotherapist, while 29% (4/14) preferred training with a physiotherapist. Although 93% (13/14) of the participants enjoyed the gamified aspect of the GTA, only 14% (2/14) reported that it was more motivating than working with a physiotherapist (Table 4).

Table 4. Results of the questionnaire given to the patients in the intervention group. The results present the percentage of subjects who agreed with the question (answer: yes).

Question	Approval (n=14), n (%)
1. Have you learned how to safely use forearm crutches?	14 (100)
2. Do you feel prepared using forearm crutches confidently?	14 (100)
3. Would you like to be rechecked by a physical therapist?	6 (43)
4. Did you understand all the instructions of the program?	13 (93)
5. Have you been able to follow all the instructions of the program?	12 (86)
6. Would you like to learn more exercise through an interactive training program?	11 (79)
7. Would you perform a purely digital rehabilitation (ie, without a human physical therapist)?	2 (14)
8. Would you have preferred to learn how to use forearm crutches from a physical therapist?	4 (29)
9. Do you think the interactive trainer is more motivating than a physical therapist?	2 (14)
10. Do you like the playful concept of the interactive trainer?	13 (93)

Assessment of GTA Usability

The usability of the GTA was evaluated using the established SUS [10]. A value of more than 68 can be considered above-average usability. A total of 9 (64%) of the 14 patients rated the usability of the GTA as above average. There was a tendency for younger patients to rate usability higher, but no significance was shown for this observation (Pearson r=–0.381; P=.18).

Discussion

Principal Findings

In the context of this study, a specifically designed GTA was used to teach patients how to use crutches. An RGB-D camera captured the position of the crutches, as well as the position of the study participant in the room. The training program guided the patients through a series of exercises and monitored their progress in real time, providing immediate feedback. This allowed simultaneous training and supervision. The results showed that compared with a control group that received conventional physiotherapy-led crutch training, the intervention group was not inferior and was able to effectively learn proper crutch use by the end of the exercise period. Although patients in the intervention group expressed positive feedback regarding the GTA and perceived it as useful, some remained skeptical about the potential for solely digital rehabilitation. The GTA was rated as having above-average user-friendliness.

Digital interventions are now increasingly integrated into everyday clinical practice for musculoskeletal rehabilitation. In particular, telerehabilitation [11] and mobile health (mHealth) applications [1] are available and prescribed by responsible therapists. At present, purely digital applications that work autonomously and can guide the patient while at the same time recognizing and preventing faulty exercises are not widely available. The programs available to date are mostly used only for the purpose of clinical studies [12]. Digital applications are rarely used in prerehabilitation.

The camera used in this study is the Microsoft Kinect camera released in 2010. Compared with other systems that only use body-worn markers (eg, Vicon [Vicon Motion Systems Ltd] and OptiTrack [NaturalPoint Inc]) for tracking (and are considered the gold standard for accuracy), the Kinect system can achieve similar precision in motion detection [13,14]. The Kinect system has already been embedded in several studies and successfully applied in rehabilitation programs [15,16]. The camera makes it possible to analyze gait without great expense or complex analysis systems. However, the short detection range of the camera is a disadvantage that can only be improved by using other more cost-intensive systems.

The analysis of the correct use of crutches can also be achieved by using sensors in the crutches themselves in addition to external systems such as a camera. Here, acceleration can be measured in addition to pressure. However, the established crutches are used almost exclusively in the measurement of load limits at partial weight-bearing. Tsuda et al [8] tried to overcome the limited detection range of the Kinect system (approximately 4 m) by using crutches with sensors. The disadvantage here is that the freedom of movement gained requires a wired connection to a computer and thus again leads to a disproportionately higher constraint.

In addition to the correct use of the crutches in terms of the movement and positioning of the crutches in the room [17], weight-bearing is important in some cases. Another study investigated forearm crutches with a built-in weight sensor. By using such a system, additional data can be displayed for the patient in real time, and the rehabilitation process can be further improved [18].

According to the available studies, the use of digital training systems is becoming increasingly popular and seems to be effective and safe [7,19,20]. In our study, there were no significant differences between the control and intervention groups. Although this is not proof of equality considering the low number of incorrect exercises, it can be concluded that the assistant can successfully teach a 3-point gait.

The comparison of crutch angle and crutch position reveals that in both cases, the intervention group achieved lower values with statistical significance. In terms of stability, it can be argued that the crutch angle closer to a vertical orientation provides greater contact between the rubber feet and the ground, thus reducing the probability of sliding. Regarding the crutch position, a smaller distance to the center of the foot can have the same effect. From this perspective, training with the digital assistant outperforms training with a physiotherapist. However, it is impractical to compare these metrics regarding a single optimal value. A slightly more tilted crutch can provide almost the same degree of stability, for example, due to the deformation of the rubber foot. Thus, a whole range of optimal values should be considered. This goes hand in hand with the evaluation of metrics by the physiotherapists, who rated all participants with a perfect score, that is, regardless of whether the angle was –14° or –4°. In this regard, both groups performed equally, and we can conclude that the digital assistant can compete with a human physiotherapist.

According to the completed questionnaires, both training approaches (GTA and analog) were similarly acceptable to the participants. Although the majority rejected a purely digital-based rehabilitation, most of them would have preferred training with a physiotherapist and even liked the concept of the assistant. However, the disadvantage of the system mentioned most often was the absence of interpersonal communication. A possible approach in future projects would be the use of artificial intelligence for communication or the use of professionals to monitor performance. Even if this again requires the involvement of staff, in the long term, the use of human resources is significantly lower.

Conclusions

The use of the newly designed GTA is a safe method of learning to use the crutches and is statistically superior to training by a physiotherapist. Even if patients prefer interaction with a physiotherapist over a purely digital approach, digital devices provide a safe and motivating opportunity to learn essential locomotor skills for rehabilitation. Even though the range of the trainers is currently limited, in the future, the use of more advanced hardware can provide a comprehensive physiotherapeutic experience.

Conflicts of Interest

None declared.

References

1. Bäcker HC, Wu CH, Pförringer D, Petersen W, Stöckle U, Braun KF. A review of functional outcomes after the app-based rehabilitation of patients with TKA and THA. J Pers Med 2022;12(8):1342 [FREE Full text] [doi: 10.3390/jpm12081342] [Medline: 36013291]
2. Li CT, Hung GK, Fong KN, Gonzalez PC, Wah SH, Tsang HW. Effects of home-based occupational therapy telerehabilitation via smartphone for outpatients after hip fracture surgery: a feasibility randomised controlled study. J Telemed Telecare 2022;28(4):239-247. [doi: 10.1177/1357633X20932434] [Medline: 32594832]
3. Gilboa Y, Maeir T, Karni S, Eisenberg ME, Liebergall M, Schwartz I, et al. Effectiveness of a tele-rehabilitation intervention to improve performance and reduce morbidity for people post hip fracture—study protocol for a randomized controlled trial. BMC Geriatr 2019;19(1):135 [FREE Full text] [doi: 10.1186/s12877-019-1141-z] [Medline: 31109289]
4. Prieto-Moreno R, Estévez-López F, Molina-Garcia P, Mora-Traverso M, Deschamps K, Claeys K, et al. ActiveHip+: a feasible mHealth system for the recovery of older adults after hip surgery during the COVID-19 pandemic. Digit Health 2022;8:20552076221139694 [FREE Full text] [doi: 10.1177/20552076221139694] [Medline: 36420319]
5. Schäfer A, Brokfeld V. Fachkräftemangel in der rehabilitation am beispiel der rehabilitationskliniken in Nordrhein-Westfalen. Rehabilitation (Stuttg) 2018;57(6):343-345. [doi: 10.1055/a-0791-2374] [Medline: 30577065]
6. De Biase S, Cook L, Skelton DA, Witham M, Hove RT. The COVID-19 rehabilitation pandemic. Age Ageing 2020;49(5):696-700 [FREE Full text] [doi: 10.1093/ageing/afaa118] [Medline: 32470131]
7. Peres B, Azadegan A, Campos PF. Digitally augmenting the physical ground space with timed visual cues for crutch-assisted walking. 2019 Presented at: CHI EA '19: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland, United Kingdom p. 1-6 URL: https://dl.acm.org/doi/proceedings/10.1145/3290607 [doi: 10.7150/jca.29049]
8. Tsuda N, Tarao S, Nomura Y, Kato N. Attending and observing robot for crutch users. 2018 Presented at: HRI '18: Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction; March 5-8, 2018; Chicago, IL p. 259-260 URL: https://dl.acm.org/doi/proceedings/10.1145/3173386 [doi: 10.1145/3173386.3176968]
9. Jordan PW, Thomas B, McClelland IL, Weerdmeester BA. Usability Evaluation in Industry. Milton Park, Oxfordshire, United Kingdom: Taylor & Francis; 1996.
10. PW. Jordan, B. Thomas, IL McClelland, B Weerdmeester. Usability Evaluation In Industry. London, United Kingdom: Taylor and Francis; 1996.
11. LeBrun DG, Martino B, Biehl E, Fisher CM, Gonzalez Della Valle A, Ast MP. Telerehabilitation has similar clinical and patient-reported outcomes compared to traditional rehabilitation following total knee arthroplasty. Knee Surg Sports Traumatol Arthrosc 2022;30(12):4098-4103. [doi: 10.1007/s00167-022-06931-6] [Medline: 35347376]
12. Röhner E, Mayfarth A, Sternitzke C, Layher F, Scheidig A, Groß HM, et al. Mobile robot-based gait training after Total Hip Arthroplasty (THA) improves walking in biomechanical gait analysis. J Clin Med 2021;10(11):2416 [FREE Full text] [doi: 10.3390/jcm10112416] [Medline: 34072524]

13. Hondori HM, Khademi M. A review on technical and clinical impact of Microsoft Kinect on physical therapy and rehabilitation. J Med Eng 2014;2014:846514 [FREE Full text] [doi: 10.1155/2014/846514] [Medline: 27006935]
14. Chang CY, Lange B, Zhang M, Koenig S, Requejo P, Somboon N, et al. Towards pervasive physical rehabilitation using Microsoft Kinect. 2012 Presented at: 2012 6th International Conference on Pervasive Computing Technologies for Healthcare; May 21–24, 2012; San Diego, CA p. 159-162 URL: https://eudl.eu/proceedings/PERVASIVEHEALTH/2012 [doi: 10.4108/icst.pervasivehealth.2012.248714]
15. Peters DM, Fritz SL, Krotish DE. Assessing the reliability and validity of a shorter walk test compared with the 10-meter walk test for measurements of gait speed in healthy, older adults. J Geriatr Phys Ther 2013;36(1):24-30 [FREE Full text] [doi: 10.1519/JPT.0b013e318248e20d] [Medline: 22415358]
16. Hsu JE, Hulet DA, McDonald C, Whitson A, Russ SM, Matsen FA. The contribution of the scapula to active shoulder motion and self-assessed function in three hundred and fifty two patients prior to elective shoulder surgery. Int Orthop 2018;42(11):2645-2651. [doi: 10.1007/s00264-018-4027-3] [Medline: 29987556]
17. Merrett GV, Ettabib MA, Peters C, Hallett G, White NM. Augmenting forearm crutches with wireless sensors for lower limb rehabilitation. Meas Sci Technol 2010;21(12):124008. [doi: 10.1088/0957-0233/21/12/124008]
18. Peres B, Campos PF, Azadegan A. A digitally-augmented ground space with timed visual cues for facilitating forearm crutches' mobility. 2019 Presented at: 17th IFIP TC 13 International Conference; September 2–6, 2019; Paphos, Cyprus p. 184-201 URL: https://link.springer.com/chapter/10.1007/978-3-030-29381-9_12 [doi: 10.1007/978-3-030-29381-9_12]
19. Pedraza-Hueso M, Martín-Calzón S, Díaz-Pernas FJ, Martínez-Zarzuela M. Rehabilitation using kinect-based games and virtual reality. Procedia Comput Sci 2015;75:161-168 [FREE Full text] [doi: 10.1016/j.procs.2015.12.233]
20. Sekhavat YA, Namani MS. Projection-based AR: effective visual feedback in gait rehabilitation. IEEE Trans Human-Mach Syst 2018;48(6):626-636. [doi: 10.1109/thms.2018.2860579]

Abbreviations

GTA: gait training assistant
MANOVA: multivariate ANOVA
mHealth: mobile health
SUS: System Usability Scale

Edited by P Santana-Mancilla; submitted 16.08.23; peer-reviewed by J Ramirez, V Brahmbhatt; comments to author 08.12.23; revised version received 31.12.23; accepted 08.01.24; published 16.08.24.

Please cite as:
Wolf MA, Sauerwald L, Kosmalla F, Daiber F, Krüger A, Landgraeber S
Implementation and Evaluation of a Gait Training Assistant for the Use of Crutches: Usability Study
JMIR Hum Factors 2024;11:e51898
URL: https://humanfactors.jmir.org/2024/1/e51898
doi:10.2196/51898
PMID:

Original Paper

Co-Designing a Smoking Cessation Chatbot: Focus Group Study of End Users and Smoking Cessation Professionals

Hollie Bendotti[1,2], PhD; Sheleigh Lawler[3], PhD; David Ireland[2], PhD; Coral Gartner[3,4], PhD; Henry M Marshall[1,5], PhD

[1]Thoracic Research Centre, Faculty of Medicine, The University of Queensland, Brisbane, Australia

[2]Australia e-Health Research Centre, Commonwealth Scientific and Industrial Research Organisation, Brisbane, Australia

[3]School of Public Health, Faculty of Medicine, The University of Queensland, Brisbane, Australia

[4]National Health and Medical Research Council Centre of Research Excellence on Achieving the Tobacco Endgame, School of Public Health, The University of Queensland, Brisbane, Australia

[5]Department of Thoracic Medicine, The Prince Charles Hospital, Metro North Hospital and Health Service, Brisbane, Australia

Corresponding Author:
Hollie Bendotti, PhD
Thoracic Research Centre
Faculty of Medicine
The University of Queensland
L5 Clinical Sciences Building, The Prince Charles Hospital
627 Rode Road
Brisbane, 4032
Australia
Phone: 61 7 3139 4110
Email: h.bendotti@uq.edu.au

Abstract

Background: Our prototype smoking cessation chatbot, Quin, provides evidence-based, personalized support delivered via a smartphone app to help people quit smoking. We developed Quin using a multiphase program of co-design research, part of which included focus group evaluation of Quin among stakeholders prior to clinical testing.

Objective: This study aimed to gather and compare feedback on the user experience of the Quin prototype from end users and smoking cessation professionals (SCPs) via a beta testing process to inform ongoing chatbot iterations and refinements.

Methods: Following active and passive recruitment, we conducted web-based focus groups with SCPs and end users from Queensland, Australia. Participants tested the app for 1-2 weeks prior to focus group discussion and could also log conversation feedback within the app. Focus groups of SCPs were completed first to review the breadth and accuracy of information, and feedback was prioritized and implemented as major updates using Agile processes prior to end user focus groups. We categorized logged in-app feedback using content analysis and thematically analyzed focus group transcripts.

Results: In total, 6 focus groups were completed between August 2022 and June 2023; 3 for SCPs (n=9 participants) and 3 for end users (n=7 participants). Four SCPs had previously smoked, and most end users currently smoked cigarettes (n=5), and 2 had quit smoking. The mean duration of focus groups was 58 (SD 10.9; range 46-74) minutes. We identified four major themes from focus group feedback: (1) conversation design, (2) functionality, (3) relationality and anthropomorphism, and (4) role as a smoking cessation support tool. In response to SCPs' feedback, we made two major updates to Quin between cohorts: (1) improvements to conversation flow and (2) addition of the "Moments of Crisis" conversation tree. Participant feedback also informed 17 recommendations for future smoking cessation chatbot developments.

Conclusions: Feedback from end users and SCPs highlighted the importance of chatbot functionality, as this underpinned Quin's conversation design and relationality. The ready accessibility of accurate cessation information and impartial support that Quin provided was recognized as a key benefit for end users, the latter of which contributed to a feeling of accountability to the chatbot. Findings will inform the ongoing development of a mature prototype for clinical testing.

(JMIR Hum Factors 2024;11:e56505) doi:10.2196/56505

KEYWORDS

artificial intelligence; chatbot; smoking cessation; behavior change; smoking; mobile health; apps; digital interventions; smartphone; mobile phone

Introduction

In 2019, a total of 1.1 billion people worldwide smoked tobacco regularly, and 13.6% of all deaths were attributable to smoking tobacco [1]. Tobacco's burden on human health is compounded by the chronic relapsing-remitting nature of tobacco dependence [2]; 96%-97% of people are unsuccessful when trying to quit without assistance [3]. Smoking cessation behavioral support increases individual quit attempt success [4] and increases further when pharmacotherapy is used, supporting the physiological and behavioral needs of the individual [5]. However, the scalability and reach of professional interventions such as behavioral counseling are limited due to accessibility [6,7] and individual awareness [8,9]. Digital interventions, such as smartphone apps, have therefore been explored to improve the reach of smoking cessation information and behavioral support. However, there is currently limited evidence to support the routine use of smartphone apps in smoking cessation [10-12], and questions remain surrounding the impact of the level of engagement with an app on its effectiveness.

The use of conversational agents for smoking cessation is an emerging area of research that may, through personalization and tailoring, enhance engagement with digital smoking cessation interventions, thereby increasing their impact. Conversational artificial intelligence (AI), such as chatbots, allows for synchronous communication with users via text and/or audio using natural language processing and machine learning algorithms with a rule-based and/or probabilistic approach [13]. Over time with increasing use and data generation, these interactions may become more natural, responsive, and tailored to the user [13]. Mohr's model of "Supportive Accountability" states that engagement with digital health interventions is promoted with the addition of human support by cultivating a sense of personal accountability to a competent, trustworthy, and caring coach [14]. Chatbots capable of emulating this type of human support via a highly accessible platform such as a smartphone app may bridge the gap between scalable personalized interventions and effective behavioral counseling. Furthermore, the current evidence supporting the effectiveness of conversational AI interventions on smoking cessation outcomes is limited but promising [15-17], and previous studies have found that they are generally an acceptable tool among participants [13,18-20].

Quin is a prototype chatbot that aims to improve access to personalized, evidence-based smoking cessation information and support via an app. Quin has been developed as part of a multiphase program of co-design research by a multidisciplinary team. Quin was designed using a "bottom-up approach," in that we analyzed and applied findings from real-world evidence-based counseling interactions (ie, consumer- and stakeholder-driven research) rather than a top-down approach by applying behavior change theories or frameworks from the outset. A detailed description of the design and development of the Quin prototype has been published elsewhere [21] and a screenshot of Quin's user interface is provided as Figure 1. Our co-design approach is based on a 3-stage model developed by the Good Things Foundation (National Health Service, UK Government), which aims to promote digital inclusion [22]:

- Stage 1: Define user and stakeholder needs and experiences
- Stage 2: Ideas and prototype
- Stage 3: Iterative testing and delivery

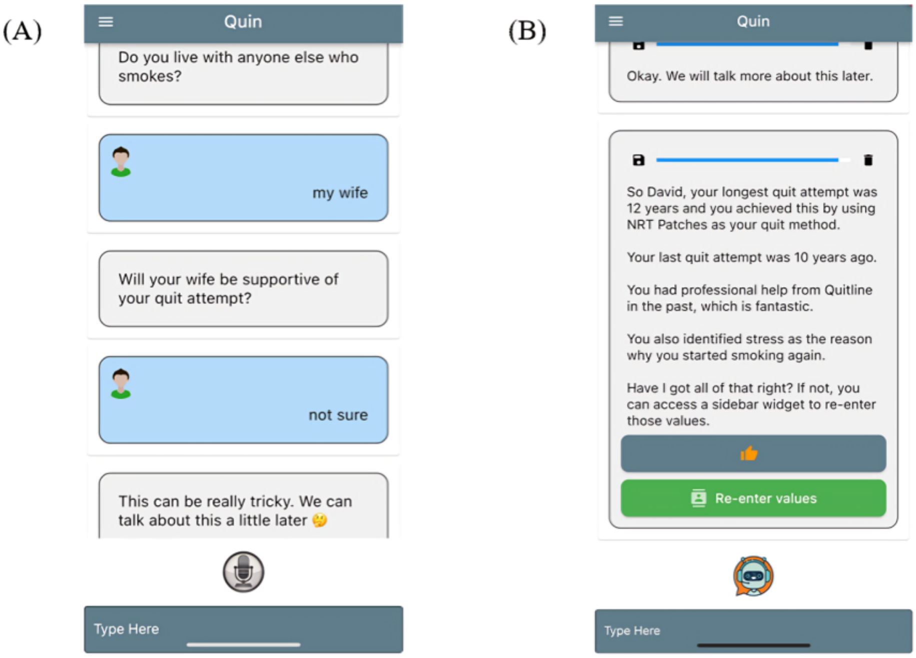
The design foundations of Quin are embedded in stage 1 by understanding and translating user and stakeholder needs and experiences at 2 levels, the platform or user interface (ie, smartphone app) and the content (ie, smoking cessation counseling and education). To understand the user experience of existing mobile smoking cessation (mCessation) apps, we first analyzed unsolicited user reviews of apps to determine important design recommendations across domains of app personalization, relationality, functionality, and credibility [23]. To understand the counselor-patient relationship and interaction, we analyzed real-world Quitline counseling sessions to identify conversation themes and topics, including how topics map to conversation stages and specific statements, questions, and responses from counselors and clients [21, 24]. Quitline counselors are trained to deliver cognitive behavioral therapy strategies and/or motivational interviewing alongside evidence-based smoking cessation information during counseling conversations.

We applied these findings to Quin's conversation design and technical development (stage 2). The types of human-computer interactions with Quin include conversing (via text or speech-to-text), instructing (eg, user selecting options), and responding (eg, notifications or initiating the conversation upon opening) [25]. Quin's dialogue is structured into "initial" and "support" conversations. The initial conversation constructs a personalized "quit plan" from demographic, smoking, and quit histories (Figure 1) to guide a personalized discussion about pharmacotherapy options and behavioral considerations or support. Based on the agreed quit plan, follow-up conversations are scheduled to check in on users' progress and review and to answer questions and troubleshoot issues [21]. Quin was programmed using a collection of natural language processing algorithms within CSIBot, the Commonwealth Scientific and Industrial Research Organisation health chatbot framework. The primary response system is a case-based reasoner, which applies both syntactic matching and sentiment analysis algorithms [21]. Examples of these algorithms are presented in Figure 1. The syntactic matching algorithm uses a radix tree data structure to store and search for a template response that can include wildcard substitutions and extra chatbot operations (eg, log data) [21]. As such, the conversation is guided and tailored by free-text and preset responses which Quin can store and recall to determine subsequent conversation paths.

At all times, we aimed to keep the language and responses very positive and affirming. For example, if the user states that the cost of smoking is a motivator for quitting smoking, Quin reinforces this by saying "Extra cash is a great reward for quitting smoking! Are you saving up for something?" Where applicable within the conversation, Quin also provides links to relevant external resources such as educational videos and Quitline services and generates a "to-do" list based on the user-defined quit plan. For example, if the user has indicated they are interested in using a prescription medication, the list will include "Speak to your doctor about using [medication]." Users also have the option to upload motivational images during the initial conversation, and Quin uses information collected within the conversation to inform a cost analysis to track savings and editable summary of user profile information.

Having developed the prototype Quin (stage 2), we now present the first phase of iterative testing (stage 3), which sought to understand how stakeholders perceive and interact with Quin, a fundamental step for ongoing phases of iterative development and refinement prior to clinical efficacy trials. Our data collection included qualitative inquiry to provide a deeper understanding of the user experience, end user preferences, and priorities, as this is an important aspect of human-centered design. As such, this study aimed to gather feedback on the Quin prototype from end users and smoking cessation professionals (SCPs) via a beta testing process to (1) identify factors that positively and negatively influence the user experience, including general views on chatbot technology for smoking cessation and suggestions for improvement; and (2) compare and contrast subjective feedback from end users and SCPs. Acknowledging the importance of applied consumer-driven research, we will then seek to apply the findings to produce a mature prototype for clinical testing.

Figure 1. (A) Quin user interface; (B) Example of personalization in Quin dialogue.



Methods

This focus group study was completed in Queensland, Australia, and collected qualitative data from focus groups, and feedback logged within the chatbot app.

Ethical Considerations

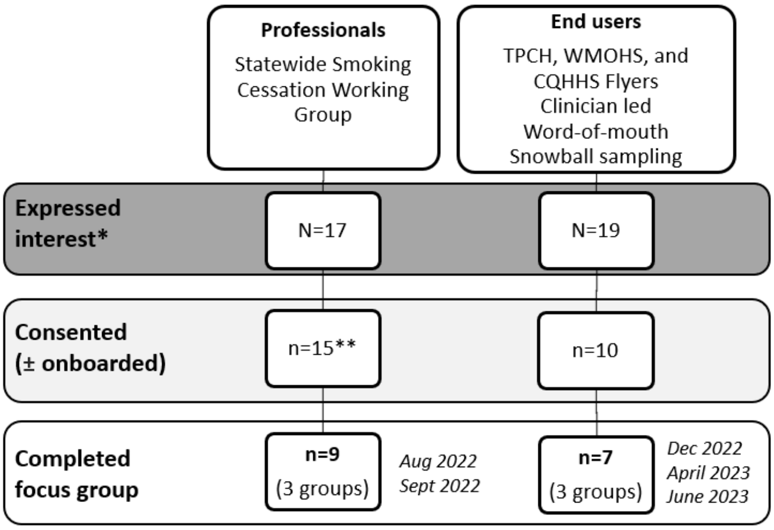
Ethics approval was granted by The Prince Charles Hospital Human Research Ethics Committee (HREC Project ID: 69623). Written informed consent was obtained from all participants before taking part in the study, and they were free to withdraw at any time. End user participants who completed a focus group were given an Aus $40 (US $26.65) gift card. Focus group transcripts were deidentified prior to analysis, and logged in-app feedback was anonymous.

Recruitment and Participants

Multiple methods of active and passive recruitment were used. SCPs from the Queensland State Health Department (Statewide Smoking Cessation Working Group and Quitline) were invited to participate via email distribution lists. End user recruitment was broad and included priority populations for smoking cessation assistance, such as people from regional or rural areas, older people with respiratory conditions or diseases, and people from low-income areas. Targeted recruitment included electronic flyer distribution via a regional public health unit to associated community organizations and physical flyer distribution via clinicians and visual display within respiratory clinics and mental health units at The Prince Charles Hospital and dental clinics at West Moreton Hospital and Health Service. We also recruited end users via word of mouth and snowballing methods. Inclusion criteria for end user participants were people aged 18 years and older who self-identify as currently smoking, either not ready or attempting to quit smoking, or have recently quit smoking (within the previous 12 months); were English-speaking; and owned a smartphone (Apple or Android) and could download apps. Participants who expressed interest were sent an electronic participant information and consent form via the survey platform REDCap (Research Electronic Data Capture; Vanderbilt University) hosted at The University of

Queensland. Following informed consent, participants were sent an electronic survey to collect demographic and smoking and quit history information. The recruitment process is outlined in Figure 2.

Figure 2. Focus group recruitment flow diagram. CQHHS: Central Queensland Hospital and Health Service; TPCH: The Prince Charles Hospital; WMOHS: West Moreton Hospital and Health Service. *Contacted the study team (phone or email) and further information and consent form provided, **9 professionals onboarded to the Quin app.



Beta Testing Procedures and Focus Groups

Consenting participants were onboarded to the Quin app 1 to 2 weeks prior to their scheduled focus group. Onboarding meetings with the lead researcher (HB) or a research assistant were completed via videoconferencing with a summary email of written instructions or only via a summary email if a participant was not available to videoconference. These meetings allowed the research team to assist with downloading the beta version and to provide general testing instructions to maximize the collection of user feedback. Users were encouraged to create multiple demographic profiles to test chatbot responses to different combinations of information and were instructed how to log feedback within the app if anomalies were found (eg, spelling errors and incorrect responses). Logged feedback was stored on a Google Firebase server. Study team members were available to assist with any issues during the testing period.

SCP focus groups were completed first to ensure the accuracy and breadth of smoking cessation information provided within Quin. Following prioritization and incorporation of feedback from SCPs as major updates, end user focus groups were completed. All focus groups were held via videoconferencing. To improve anonymity, participants had the option to leave their camera off and use their first name only (eg, username and in discussion). The focus groups were moderated by the lead investigator (HB), an early career researcher who has formal training and experience in qualitative research as well as expertise in health promotion and public health. Senior researchers SL and/or HMM were present during focus groups to take notes and assist with probing or follow-up questions where applicable. SL is an experienced qualitative researcher with expertise in health psychology, health promotion, and public health. HMM is a thoracic medicine specialist with a bachelor of psychology (honors) and research interests in smoking cessation interventions. At the end of the focus group, upon departure of participants, the investigators remained in the meeting to reflect on the session, feedback, and notes taken. The focus group sessions followed a Human Research Ethics Committee–approved interview guide (Multimedia Appendix 1) based on the categories of feedback within the end user version of the Mobile App Rating Scale [26] but allowed for probing and follow-up questions to expand on discussion. End users were asked exploratory questions regarding smoking and quitting histories and barriers and enablers to smoking cessation as well as experiences, if any, using mCessation apps; SCPs were asked similar questions reframed as their experience of what clients commonly tell them. The moderator ensured all participants were given an opportunity to speak and expand on others' ideas. Focus groups were recorded, transcribed verbatim, and deidentified prior to analysis. All participants had the option to continue testing the app and provide additional feedback after their focus group.

Analysis

Quantitative data from demographic surveys were analyzed descriptively. In-app feedback was downloaded from Google Firebase after completion of all SCP focus groups and after each end -user focus group. This feedback was categorized using content analysis, reviewed, and prioritized for incorporation into Quin by all investigators.

Focus group transcripts were imported into NVivo software (version 12; Lumivero), and analysis was completed in 2 ways. First, after completion of all SCP focus groups, transcripts and notes were reviewed by HB to identify key improvements to be made prior to end user focus groups. These findings were presented, discussed, and agreed upon by our multidisciplinary research team before updates to Quin were made. This approach is reflective of Agile development methods, and we present major changes to Quin using theme, epic, stories, and tasks [27]. Second, following the completion of all focus groups, transcripts were analyzed inductively by HB and SL using thematic analysis [28]. To address the aim of the study, 1 researcher (HB) broadly coded data into categories of positive and negative feedback and suggestions for improvement. Both study team members then generated an initial set of codes within each category, which were reviewed and combined to define and refine themes and subthemes. Disagreements in data interpretation and theme development were able to be resolved via discussion between HB and SL, but a third researcher (HMM) was available to discuss differences in interpretation during thematic analysis. We present the themes and subthemes, including direct quotes, by aforementioned categories and compare and contrast feedback between professionals and end users.

Results

Overview

Three focus groups of 9 SCPs were completed in August (n=4) and September (n=3; n=2) 2022 testing Quin (beta version 21). Three focus groups of 7 end users were completed in December 2022 (n=3) and April (n=2) and June (n=2) 2023, testing Quin beta versions 28, 30, and 31, respectively. The mean duration of focus groups was 58 (SD 10.9; range 46-74) minutes. Participant demographic characteristics are summarized in Table 1. Occupations of SCPs included clinical nurses, occupational therapists, pharmacists, Quitline telephone counselors, and health promotion officers across metropolitan, rural, and Indigenous health services, of which the majority (n=8) delivered smoking cessation counseling and advice regarding nicotine replacement therapy. Most end users currently smoked cigarettes on a daily (n=3) or less than daily basis (n=2), and 2 had quit smoking. Nicotine dependence scores were calculated using the Fagerström Test for Nicotine Dependence [29], but results may be unreliable, as most end users anecdotally indicated they had fully or partially (ie, dual use) transitioned to e-cigarettes (either as a tobacco product or cessation aid) and answered the questions in the context of cigarette smoking.

Table 1. Participant demographics by cohort.

Item	Professionals (n=9)	End users (n=7)
Age (years), n (%)		
25-29	—[a]	2 (29)
30-39	—	2 (29)
40-49	7 (78)	1 (14)
50-59	1 (11)	1 (14)
60-69	1 (11)	—
70+	—	1 (14)
Sex, n (%)		
Female	9 (100)	2 (29)
Male	—	5 (71)
Postcode, n (%)		
Major city	7 (78)	5 (71)
Inner regional	1 (11)	1 (14)
Remote	—	1 (14)
Very remote	1 (11)	—
Indigenous, n (%)		
Aboriginal	1 (11)	—
No	8 (89)	7 (100)
Education, n (%)		
Senior high school	—	2 (29)
Trade certificate	—	2 (29)
Bachelor degree	6 (67)	1 (14)
Master degree	3 (33)	2 (29)
Phone type, n (%)		
Apple	8 (89)	4 (57)
Android	1 (11)	3 (43)
Currently smoke, n (%)		
Daily	—	3 (43)
Less than daily	—	2 (29)
Not at all	9 (100)	2 (29)
Previously smoked, n (%)		
Daily	3 (33)	6 (86)
Less than daily	1 (11)	—
Both daily and less than daily	—	1 (14)
Never smoked	5 (56)	—
Type of tobacco (current)[b], n (%)		
Manufactured cigarettes	—	4 (80)
Roll your own	—	2 (40)
Pipe	—	1 (20)
Heated	—	1 (20)
e-Cigarettes (nonprescription)	—	1 (20)

Item	Professionals (n=9)	End users (n=7)
Type of tobacco (previously smoked)[c], n (%)		
Manufactured cigarettes	2 (50)	1 (50)
Roll-your-own cigarettes	3 (75)	2 (100)
Cigars or cigarillos	—	1 (50)
Age start smoking (years), mean (SD)[d]	16 (4.3)	16 (4.1)
Time to first cigarette[b], n (%)		
Within 5 minutes	—	1 (20)
6 to 30 minutes	—	2 (40)
31 to 60 minutes	—	2 (40)
After 60 minutes	—	—
FTND[e,f], n (%)		
Very low	—	2 (50)
Low	—	2 (50)

[a]Not available.
[b]End users: n=5.
[c]Professionals: n=4 and end users: n=2.
[d]Professionals: n=4 and end users: n=7.
[e]Missing data: n=1 cigarette per day.
[f]FTND: Fagerström Test for Nicotine Dependence.

Logged Feedback

In total, 46 instances of feedback were logged within Quin by 6 SCPs. Feedback related to pattern matching errors (ie, user inputs were misinterpreted or unable to be detected by the chatbot; n=13, 28%); missing conversation trees (ie, topics that require more than 1 response; n=6, 13%), smoking cessation information (n=9, 20%), and responses (ie, information to be programmed; n=8, 17%); logic issues (ie, the decision to enter a conversation tree; n=4, 9%); dialogue content (n=2, 4%); spelling or grammatical errors (n=2, 4%); notifications (n=1, 2%); and user interface (n=1, 2%). In contrast, 4 end users logged 9 instances of feedback related to logic issues (n=2, 22%), pattern matching errors (n=2, 22%), missing responses (n=2, 22%), dialogue content (n=1, 11%), notifications (n=1, 11%), and user interface (n=1, 11%).

Major and Minor Updates to Quin Between Cohorts

Two major updates were made to the Quin prototype between versions 21 and 28 following a review of the focus group and logged feedback from the SCPs, a process illustrated in Figures 3 and 4. First, we reviewed and implemented changes to the conversation flow to improve the visual delivery of information within the app. This included consolidating information into smaller dialogue bubbles, including links to relevant resources, and slowing the delivery of information by including a timer based on the length of text and animated ellipses to indicate the chatbot "typing." Second, we designed and implemented a conversation tree to better handle instances when a user may be craving a cigarette (ie, a "Moment of Crisis"). Quin was programmed to identify more statements that suggest an acute craving and to work sequentially through pharmacotherapy options and behavioral strategies depending on the context of the craving (ie, trigger), of which questions or responses were tailored if the user had created a quit plan. Quin was also programmed to check in with the user to see if the craving had passed and offer positive reinforcement or more options for strategies.

Other minor updates were made between versions to correct errors and improve functionality based on logged feedback, which included spelling errors, connecting programmed but inactive conversation trees, coding missing pattern matching or sentiment analysis, and updates to the chatbot dialogue and smoking cessation information.

Figure 3. Changes to conversation flow using Agile methods.

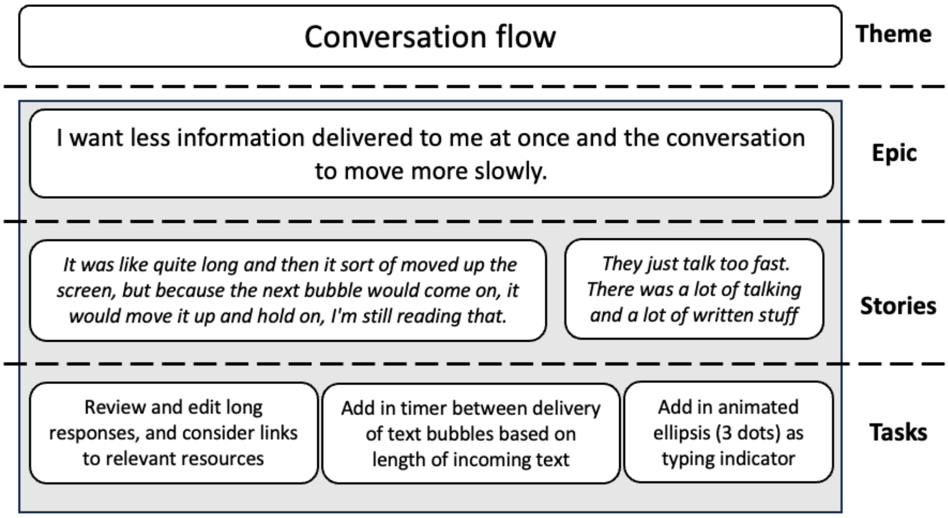
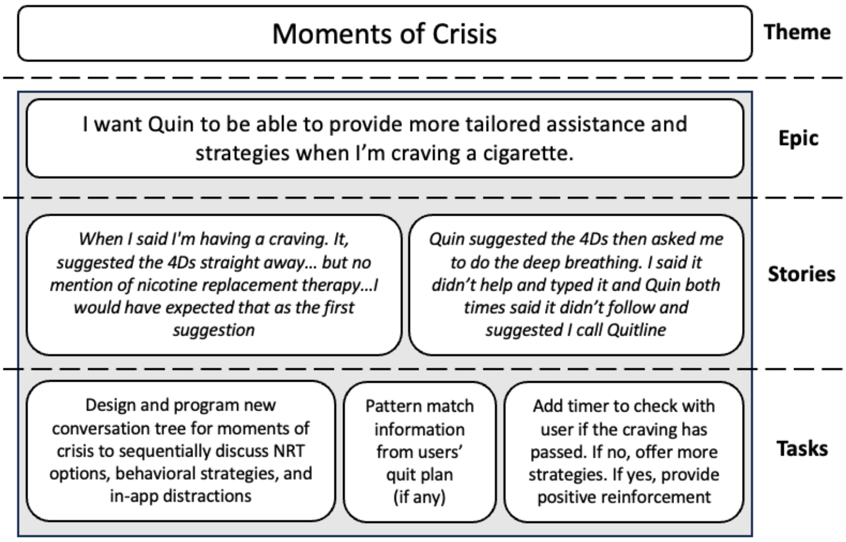


Figure 4. Changes to the conversation in response to acute cigarette cravings using Agile methods. NRT: nicotine replacement therapy.



Feedback Themes

Overview

Four high-level themes encompassing positive and negative user feedback were identified (Figure 5). While these themes are reported separately, there was a crossover between them within individuals' feedback, given the technical connection between the features. We also provide a summary of chatbot design recommendations based on this feedback (Textbox 1).

Figure 5. Positive and negative feedback themes and subthemes. SCP: smoking cessation professional. *End users' feedback only, **SCPs' feedback only.

Textbox 1. Summary of recommendations for smoking cessation chatbot design based on focus group feedback.

Smoking cessation support features

- Provide tailored assistance in moments of acute cravings
- Ability to develop a quit plan based on previous experiences and personal preferences
- Follow-up conversations for user self-reflection on quit plan
- Proactive check-in times and notifications to enhance the sense of accountability
- Accurate educational content and cessation information
- Ability to integrate into or complement existing services (eg, preparation tool and after-hours support)
- Offer distraction and motivational tools within the conversation (eg, photo upload and tracking tools)

Conversation design

- Maximize opportunities for personalization and information gathering or giving
- Speed (delivery) of chatbot responses should emulate human conversation (eg, typing)
- Outsource large amounts of information by linking to other resources
- Include positive reinforcement in responses
- Limit the cyclical nature of conversation trees

Functionality

- Avoid errors in responses and sentiment analysis
- Include safeguards and referral pathways when the chatbot is unable to respond
- Automatically save progress within conversations
- Automatically save, store, and recall user-defined quit plan
- Ability to respond to changing preferences and information

Conversation Design

A key omission from the conversation reported by SCPs was Quin's ability to provide assistance when the user is experiencing a cigarette craving by suggesting rapid-acting nicotine replacement therapy and behavioral strategies. As such, this feedback informed the inclusion of the "Moments of Crisis" conversation tree (Figure 4). The first iteration of this conversation tree was then tested by end users; however, limitations to Quin's ability to respond correctly or appropriately in these instances were also reported.

At times, SCPs found the amount of information being delivered was too much and too fast, which disrupted the visual flow of information as they needed to scroll back through the conversation to read Quin's responses. This feedback informed the previously described updates to the conversation flow (Figure 3). Progression through the conversation, such as when to type a response or use a radio button, was also unclear to some SCPs. After updates were made between cohorts, this issue regarding conversation flow was not raised by end users during open discussion, even when prompted.

Participants across both cohorts found the information provided within Quin to be accurate, relevant, and educational. Some end users reported a humanlike feel to the dialogue content and visual conversation flow and appreciated outsourcing large amounts of information by linking to relevant resources as a more efficient way to learn.

> *It did link to those videos. And one thing that I appreciated was, it was clearly YouTube content that I'd been pushed towards rather than Quin ... I've noticed that Quin, quite cleverly, never has more than a really easily digestible amount of prose that that comes back. And I thought that was a really cool feature.* [End user 1, Focus group 5]

> *Even the little dot dot dots to make you think that it's typing, I think that's really cute. Obviously it needs a moment to process, and that makes you think that you're talking to a human, it really does feel like you're having a chat with someone it's not like for the most part generic robot answers ... for the most part the conversation style with the bot was really good.* [End user 3, Focus group 4]

End users and SCPs also thought the conversation structure emulated that of traditional care, in that users had the ability to form a quit plan alongside receiving smoking cessation information. Furthermore, reminders about aspects of their quit plan within future conversations with Quin were found to be motivating by one end user.

> *I really like this app because there's the information as well as the planning side of things.* [SCP 3, Focus group 2]

One end user noted the chatbot conversation to be quite cyclical and scripted but also acknowledged the resources needed to expand the scope of the technology to improve this. Missed opportunities for giving and gathering information were identified by SCPs including limited prompts to see a general practitioner and asking users about motivations or reasons for previous quit attempts and caffeine and alcohol use. Finally, one SCP perceived the tone of a specific response related to quitting unassisted (ie, cold turkey) to be slightly patronizing and provided suggestions to better promote positive reinforcement. In contrast, no end users commented on the tone of the chatbot dialogue.

> *I found the tone of it was probably a little bit condescending. It was like, "Oh, well, you must have great willpower to do that," and it didn't kind of go on any more to say, "Hey, Great! That's great. Have a try. But remember, we've got lots, lots of options."* [SCP 1, Focus group 2]

Functionality

Issues with chatbot functionality recounted across cohorts were largely related to its inability to respond correctly to some user inputs, which occasionally resulted in the conversation to end abruptly. There were also instances where negative sentiment in users' language was not recognized.

Frustration was expressed by both cohorts when progress was lost during the initial "appointment," as the conversation did not save when they left the app and suggested to include an indication of time required to complete the first conversation. Issues re-engaging with Quin to restart the quit plan discussion and recall quit plan information were also raised.

> *I sort of got halfway through ... when I went back in it didn't sort of pick up where we left off. I found myself having to start again. That was a yeah, a peeve moment for me.* [End user 1, Focus group 5]

End users and SCPs reported a lack of customization for user responses to some questions as well as limited adaptability of chatbot responses to changing quit preferences and information.

> *I've gone in a couple of times since then, typed in questions ... "I'm not having any luck with cold turkey. Can I follow up with something else." And I think I got an exploding robot*, so no response. [SCP 1, Focus group 2] *Represents Quin not being able to provide a response*

Despite the issues highlighted, many participants (largely end users) acknowledged the functionality and potential of Quin relative to its early phase of development and limited resources. Participants from both cohorts reported being able to complete an initial conversation and set up a quit plan. Motivational features prompted by the chatbot, such as the ability to upload images and the cost analysis of money saved per year, were valued by end users.

> *The app is really good in that I can pick up the app, which I have been using a lot, I look at my motivations, so what a picture of myself and my husband ... They're really really motivating as well.* [End user 2, Focus group 5]

Relationality and Anthropomorphism

Both cohorts described limitations to Quin's ability to emulate humanlike characteristics and how this impacted their user experience. Most participants expressed they felt like they were talking to a robot, and this feedback was often linked to functionality issues. Some end users did report they initially felt like they were talking to a human, but errors in chatbot responses impaired this effect, causing frustration, which discouraged them from continuing to use it.

> *End user 2: You think you're talking to someone. And then, when that happened, you're like, "Oh, this is just a bot," and it actually annoys you and makes you go well exit... I'm not talking to this thing. Yeah I had that feeling a couple of times with it.*
>
> *End user 1: Yeah, the same thing. You're talking to it ... You think you're talking to a person, and then ... it answers to something totally irrelevant. Well, I think. "Oh it's a machine I'm talking to."* [Focus group 4]

End users highlighted the benefit of Quin being an impartial proactive support for a quit attempt. Some participants expressed a sense of accountability to the chatbot and appreciated the nonconfrontational nature of the interaction, which in turn limited the fear of judgment. As such, it was a relaxed support tool for first-time quit attempts and a source of positive reinforcement.

> *I think it would be useful for me ... It would keep me honest. Asking me you know "How many of you smokes have you had today? When did you have your last smoke?" Things like that. Yeah, it would keep me honest.* [End user 1, Focus group 4]
>
> *I totally agree with ... End user 1 who said about the chatbot feature being non confronting instead of having to talk to someone. I definitely think that would help. You know, countless people who don't really want to talk about it face-to-face with someone.* [End user 2, Focus group 4]

Role as a Smoking Cessation Support Tool

End users from all focus groups believed the chatbot app was an appropriate and acceptable support tool for smoking cessation. For some, their experience exceeded their expectations, in that they simply enjoyed using the chatbot or they got more cessation support or educational benefit from the technology than anticipated.

> *It was a good tool to bounce back at myself. Which I honestly can say I didn't expect from a bot. It was, it was much more helpful than I thought it would be, even though I'm not actually in the process of quitting smoking. I still found it quite like a quite positive experience to use.* [End user 1, Focus group 6]

Encouragingly, an end user found it provided them the extra support they needed during their current quit attempt. Others felt that Quin was more helpful in comparison to other smoking cessation apps due to the educational content and interactivity with the chatbot. More importantly, one end user also highlighted the benefit of the potential for rapid adaptability of this technology in response to new health challenges and cessation support as a result of industry shifts (eg, vaping prevalence and cessation support).

Ease of access to smoking cessation information and support was seen as a key benefit by both cohorts. From the perspective of SCPs, the chatbot had the potential to integrate into and complement their services as both a client preparation tool and proactive complementary tool alongside professional support including when this support is unavailable (ie, after hours).

> *I think that that is an awesome concept for anything after hours, you know, if you're really really struggling at like three o'clock on a Sunday morning. You can still feel like interacting and getting immediate responses rather than just, you know, reading something that's already there printed. I think you know it's just a good to know that you can have an answer 24/7 essentially, I think that that's really great.* [SCP 2, Focus group 3]

> *The idea of being able to share information from the app with Quitline could be a very beneficial thing ... they've got the background on the client, and it's a really straightforward-based way to start a conversation ... So I think it's a great way to, yeah, it kind of builds rapport. It's strength based ... and possibly cuts down on a bit of time as well, which would be valuable for the individual client as well as the service.* [SCP 3, Focus group 2]

Discussion

Principal Findings

This focus group study, with an embedded beta testing process, aimed to explore SCP and end user feedback on the user experience of a prototype smoking cessation chatbot. We found the qualitative feedback to be similar between cohorts; yet, some feedback subthemes were also exclusive to cohorts, with end users reporting more positive feedback and SCPs reporting more critical feedback. This is likely due to major updates made to Quin between cohorts but may also reflect differing priorities and expectations placed on the chatbot by SCPs and end users, which is a trend we have observed in our previous study of smoking cessation apps [23]. Overall, initial testing feedback on the user experience from both participant cohorts shows promise for Quin as an acceptable smoking cessation support tool; however, aspects of usability require improvement. Our findings build upon and reinforce previous qualitative research on user experiences with conversational agents for smoking cessation [13,18,19,30], and results will be incorporated into future iterations of Quin to produce a mature prototype for clinical testing.

Comparison to Prior Work

Key benefits of Quin highlighted across cohorts included the smoking cessation information being relevant and acceptable and the accessibility of the chatbot as an impartial support tool. Smoking cessation often requires a multifaceted individualized approach [31], delivered over a medium- to long-term period of months to years; expanding the avenues of behavioral support is a step toward overcoming barriers to care. Consulting with SCPs allowed us not only to ensure the accuracy of the information but also to explore the potential interoperability and integration of Quin within their professional services, with many acknowledging it as a preparation tool for both the client and counselor prior to a course of care and support tool outside of service hours. Previous research has found people disclose information to and interact with chatbots similarly to how they would a human [32], and in the context of smoking cessation, that chatbot support was superior to that of family and friends [19]. Participants in our study identified Quin as supportive, nonconfrontational, and nonjudgmental. This finding reflects previous qualitative inquiries, which found that groups of sexual and gender minority young adults [30] and veterans [18] found embodied conversational agents provided nonjudgmental and/or humanlike smoking cessation support. Future first-time clients may therefore feel more comfortable providing smoking and quit histories to Quin, from which a counselor could use this information to build rapport with the client or clients may wish to only engage with Quin due to feelings of judgment or shame commonly experienced by people who smoke.

Our findings are consistent with those reported in a similar qualitative study [19], in that the sense of accountability to the chatbot reported by end users partially supports Mohr's model of "Supportive Accountability" [14]. The model states that engagement with digital health interventions is promoted with the addition of human support by fostering a sense of personal accountability to a legitimate, trustworthy, and caring coach [14]. This qualitative study of 14 people found that engagement with, and feelings of accountability to, a smoking cessation chatbot was attributed to the humanlike features and interaction style of the chatbot alongside users' perceived need for support [19]. However, the legitimacy and trustworthiness [14] of Quin beta versions may have been compromised by functionality issues, which impeded the relationality of the chatbot as a coach. Therefore, improved functionality must be consistent to sustain its legitimacy as a coach.

All participants across cohorts expressed negative feedback regarding the usability and user experience of Quin among aspects of conversation design, functionality, and relationality. Given the technical connection between these features, the functionality of the AI system is at the core of these identified issues. When we consider the usability goals described by Rogers et al [25], the reported inefficiencies in Quin primarily affected the utility (eg, limited tracking features and follow-up prompts) and efficiency (eg, unable to respond correctly). A key design feature of Quin is that the conversation is largely driven by free-text responses via a rule-based system to enhance user engagement and better emulate traditional human support. Previous qualitative research has also found that this approach to smoking cessation chatbot design is preferred by end users, in that responses and questions are tailored and relative to their momentary needs and situations [19]. While Quin's current response system allows for adaptability and greater control over the accuracy of health information being provided, it is limited, in that it requires a defined structure with comprehensive coding of all potential utterances, which may increase the chance of errors, and health information needs to be reviewed and updated on a regular basis. Yet, to rely solely on a predictive model may increase the risk of incorrect information being delivered, which in the context of health behavior change could be problematic or even counterproductive, and limit the degree of personalization in responses [33]. Moreover, rule-based systems require relatively small computational and memory resources and can operate on a mobile device greatly increasing the privacy of user data with minimal operating costs. Most predictive models of suitable performance operate as a server accessible via the internet and require human utterances to be transmitted to the server potentially exposing sensitive data that may be intercepted by a hostile actor. Furthermore, there are known ethical issues and biases associated with the collation and marking of large amounts of training data required for predictive models [34], which can further compound existing inequalities (eg, sex, race, ethnicity, and socioeconomic status). However, the high flexibility of such systems is persuasive, and we believe future technological advances may resolve some of these issues. Therefore, we suggest that a carefully balanced combination of a rule-based system with a more flexible predictive model may improve the adaptability and relationality of a conversational agent intervention as a digital coach.

Contrasting feedback between cohorts following major updates lends support to the Agile development process. Significant changes made to the delivery of information within Quin were validated through end user focus group feedback, with most finding the dialogue and visual conversation flow emulated that of a human and observing specific updates (eg, animated typing ellipses and outsourcing information to links). A similar iterative development process between participant cohorts has also been used in the training of an automated motivational interviewing–based chatbot for unmotivated people who smoke [13]. The traditional path to market for medical devices and pharmaceutical products consists of 4 sequential clinical evaluation phases to determine safety, efficacy, dosing, and effectiveness [35]. In contrast, digital health product development is dynamic and nonlinear due to the rate at which technology evolves and can be modified. Agile development processes allow for greater flexibility and ability to adapt to continuous iterative feedback. Like co-design, Agile methods also prioritize collaboration between developers and end users for effective solutions [36].

Future Directions

Finally, there are potential opportunities for Quin's development and future role in smoking cessation services. Quin's dialogue is continually modifiable, and the benefit of this technology is that it can be easily updated to reflect changing evidence or advice in response to industry shifts and new health challenges and tailored to target specific population groups. The Quin prototype in its current state is the foundation, upon which we plan to iteratively build using our co-design methodology to analyze Quitline cessation support for priority groups (eg, people who are pregnant) and e-cigarette use and to ultimately work backward toward a theory-informed intervention. The accessibility of Quin will allow it to complement and support existing counseling services and primary care. For example, Quin may act as a liaison to initiate cessation support on an on-demand basis after-hours or for the time between referral and initial contact, ideally with interoperability with relevant patient management or health systems to promote continuity of care. Conversely, as previously mentioned, Quin alone as the anonymous coach may be the preferred form of cessation support for people reluctant to access services.

Strengths and Limitations

A strength of this study was the early involvement of key stakeholders in the app development process through an open and informed dialogue. This is an important stage of co-design for digital inclusion [22] and is considered a principle of digital development [37]. Furthermore, the capacity to review and incorporate key feedback between cohort groups allowed for the validation of some updates, evidenced by changes in feedback following major updates to the conversation flow. We also included a variety of end user participants across most age groups, including older age participants who are a key demographic for smoking cessation and mobile health interventions as well as participants from regional and remote locations. However, there are several limitations to this study. First, our participation rate in focus groups was low due to attrition. While we made many efforts to engage with potential participants, some did not respond after contacting the study team to express interest in participating or after consenting or were unable to attend the focus group after being onboarded to the app. As such, our recommendations for smoking cessation chatbot design are limited in their generalizability. Second, current objective nicotine dependence scores among end users were mostly low, which may limit the generalizability of our findings, as people with higher nicotine dependence may have different experiences and needs. This should also be interpreted with caution, as most end users reported in the focus groups that they had transitioned to disposable e-cigarettes, which are known to contain high levels of nicotine, and their use is not captured in the assessment. As such, the true level of nicotine dependence was likely higher among end users. Nevertheless, all end user participants, and some SCPs, had recent previous experience with daily cigarette smoking from a young age and multiple quit attempts. Third, while we encouraged open and honest feedback, a degree of social desirability bias may have influenced subjective feedback.

Conclusions

Feedback from end users and SCPs on the user experience of the Quin prototype highlighted the importance of functionality and its impact on the conversation flow and relationality of chatbot technology. Despite this, participants recognized the benefit of Quin to provide easily accessible information, nonjudgmental support, and positive reinforcement, including how this can complement existing cessation support services. Ongoing development will incorporate these findings to produce a mature prototype for clinical testing, and the design recommendations provided may assist future smoking cessation chatbot developments.

Acknowledgments

The authors would like to thank David Oates for his assistance with onboarding smoking cessation professionals. The authors would also like to thank Jenny Peek, Kalie Green, and Dr Kelly McGowan for their assistance and support with end user recruitment at The Prince Charles Hospital, Central Queensland Hospital and Health Service, and West Moreton Hospital and Health Service, respectively. This work was supported by a New Investigator Grant (NI2021-31) awarded to HB from The Prince Charles Hospital Foundation and by an Investigator Grant (GNT1178331) from the National Health and Medical Research Council awarded to HMM. HB receives a top-up scholarship from the Commonwealth Scientific and Industrial Research Organisation. CG is supported by an Australian Research Council Future Fellowship (FT220100186) and holds funding from National Health and Medical Research Council grants (GNT1198301 and GNT2019252).

Data Availability

The data sets generated and analyzed during this study are not publicly available due to ethical considerations, as the qualitative data are of personal nature and are only available to the study team on approved servers.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Focus group interview guides.
[DOCX File , 30 KB - humanfactors_v11i1e56505_app1.docx]

References

1. GBD 2019 Tobacco Collaborators. Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and attributable disease burden in 204 countries and territories, 1990-2019: a systematic analysis from the Global Burden of Disease Study 2019. Lancet 2021;397(10292):2337-2360 [FREE Full text] [doi: 10.1016/S0140-6736(21)01169-7] [Medline: 34051883]
2. Steinberg MB, Schmelzer AC, Richardson DL, Foulds J. The case for treating tobacco dependence as a chronic disease. Ann Intern Med 2008;148(7):554-556. [doi: 10.7326/0003-4819-148-7-200804010-00012] [Medline: 18378950]
3. Hughes JR, Keely J, Naud S. Shape of the relapse curve and long-term abstinence among untreated smokers. Addiction 2004;99(1):29-38. [doi: 10.1111/j.1360-0443.2004.00540.x] [Medline: 14678060]
4. Lancaster T, Stead LF. Individual behavioural counselling for smoking cessation. Cochrane Database Syst Rev 2017;3(3):CD001292 [FREE Full text] [doi: 10.1002/14651858.CD001292.pub3] [Medline: 28361496]
5. Hartmann-Boyce J, Hong B, Livingstone-Banks J, Wheat H, Fanshawe TR. Additional behavioural support as an adjunct to pharmacotherapy for smoking cessation. Cochrane Database Syst Rev 2019;6(6):CD009670 [FREE Full text] [doi: 10.1002/14651858.CD009670.pub4] [Medline: 31166007]
6. Tall JA, Brew BK, Saurman E, Jones TC. Implementing an anti-smoking program in rural-remote communities: challenges and strategies. Rural Remote Health 2015;15(4):3516 [FREE Full text] [Medline: 26530272]
7. Hutcheson TD, Greiner KA, Ellerbeck EF, Jeffries SK, Mussulman LM, Casey GN. Understanding smoking cessation in rural communities. J Rural Health 2008;24(2):116-124. [doi: 10.1111/j.1748-0361.2008.00147.x] [Medline: 18397444]
8. Roddy E, Antoniak M, Britton J, Molyneux A, Lewis S. Barriers and motivators to gaining access to smoking cessation services amongst deprived smokers—a qualitative study. BMC Health Serv Res 2006;6(1):147 [FREE Full text] [doi: 10.1186/1472-6963-6-147] [Medline: 17087825]
9. van Rossem C, Spigt MG, Kleijsen JRC, Hendricx M, van Schayck CP, Kotz D. Smoking cessation in primary care: exploration of barriers and solutions in current daily practice from the perspective of smokers and healthcare professionals. Eur J Gen Pract 2015;21(2):111-117. [doi: 10.3109/13814788.2014.990881] [Medline: 25649048]
10. Barnett A, Ding H, Hay KE, Yang IA, Bowman RV, Fong KM, et al. The effectiveness of smartphone applications to aid smoking cessation: a meta-analysis. Clinical eHealth 2020;3:69-81. [doi: 10.1016/j.ceh.2020.09.001]
11. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y, Dobson R. Mobile phone text messaging and app-based interventions for smoking cessation. Cochrane Database Syst Rev 2019(10):CD006611. [doi: 10.1002/14651858.cd006611.pub5]
12. Guo YQ, Chen Y, Dabbs AD, Wu Y. The effectiveness of smartphone app-based interventions for assisting smoking cessation: systematic review and meta-analysis. J Med Internet Res 2023;25:e43242 [FREE Full text] [doi: 10.2196/43242] [Medline: 37079352]
13. Almusharraf F, Rose J, Selby P. Engaging unmotivated smokers to move toward quitting: design of motivational interviewing-based chatbot through iterative interactions. J Med Internet Res 2020;22(11):e20251 [FREE Full text] [doi: 10.2196/20251] [Medline: 33141095]
14. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]

15. Bendotti H, Lawler S, Chan GCK, Gartner C, Ireland D, Marshall HM. Conversational artificial intelligence interventions to support smoking cessation: a systematic review and meta-analysis. Digit Health 2023;9:20552076231211634 [FREE Full text] [doi: 10.1177/20552076231211634] [Medline: 37928336]
16. He L, Balaji D, Wiers RW, Antheunis ML, Krahmer E. Effectiveness and acceptability of conversational agents for smoking cessation: a systematic review and meta-analysis. Nicotine Tob Res 2023;25(7):1241-1250 [FREE Full text] [doi: 10.1093/ntr/ntac281] [Medline: 36507916]
17. Whittaker R, Dobson R, Garner K. Chatbots for smoking cessation: scoping review. J Med Internet Res 2022;24(9):e35556 [FREE Full text] [doi: 10.2196/35556] [Medline: 36095295]
18. Abdullah AS, Gaehde S, Bickmore T. A tablet based embodied conversational agent to promote smoking cessation among veterans: a feasibility study. J Epidemiol Glob Health 2018;8(3-4):225-230 [FREE Full text] [doi: 10.2991/j.jegh.2018.08.104] [Medline: 30864768]
19. Alphonse A, Stewart K, Brown J, Perski O. Exploring users' experiences with a quick-response chatbot within a popular smoking cessation smartphone app: semistructured interview study. JMIR Form Res 2022;6(7):e36869 [FREE Full text] [doi: 10.2196/36869] [Medline: 35797093]
20. Karekla M, Savvides SN, Gloster A. An avatar-led intervention promotes smoking cessation in young adults: a pilot randomized clinical trial. Ann Behav Med 2020;54(10):747-760. [doi: 10.1093/abm/kaaa013] [Medline: 32383736]
21. Bendotti H, Ireland D, Lawler S, Oates D, Gartner C, Marshall HM. Introducing Quin: the design and development of a prototype chatbot to support smoking cessation. Nicotine Tob Res 2024;26(5):612-620 [FREE Full text] [doi: 10.1093/ntr/ntad217] [Medline: 37936253]
22. Co-design in health. Goods Things Foundation. 2019. URL: https://digital-health-lab.org/codesign-in-health [accessed 2021-06-09]
23. Bendotti H, Lawler S, Ireland D, Gartner C, Hides L, Marshall HM. What do people want in a smoking cessation app? An analysis of user reviews and app quality. Nicotine Tob Res 2022;24(2):169-177. [doi: 10.1093/ntr/ntab174] [Medline: 34460922]
24. Bendotti H, Lawler S, Gartner C, Ireland D, Marshall HM. Smoking cessation counseling in practice: a qualitative analysis of quitline conversations in Queensland, Australia. Health Educ Behav 2024;51(1):43-53 [FREE Full text] [doi: 10.1177/10901981231206068] [Medline: 37846946]
25. Rogers Y, Sharp H, Preece J. Conceptualizing interaction. In: Interaction Design: Beyond Human-Computer Interaction. 6th Edition. Milton, MA: John Wiley & Sons Inc; 2003:104-137.
26. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
27. Lowe D. Theme, epic, story, task. Scrum & Kanban. 2014. URL: https://scrumandkanban.co.uk/theme-epic-story-task/ [accessed 2023-09-21]
28. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
29. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. Br J Addict 1991;86(9):1119-1127. [doi: 10.1111/j.1360-0443.1991.tb01879.x] [Medline: 1932883]
30. Heffner JL, Watson NL, Serfozo E, Kelly MM, Reilly ED, Kim D, et al. An avatar-led digital smoking cessation program for sexual and gender minority young adults: intervention development and results of a single-arm pilot trial. JMIR Form Res 2021;5(7):e30241 [FREE Full text] [doi: 10.2196/30241] [Medline: 34328430]
31. Stead LF, Koilpillai P, Fanshawe TR, Lancaster T. Combined pharmacotherapy and behavioural interventions for smoking cessation. Cochrane Database Syst Rev 2016;3(3):CD008286 [FREE Full text] [doi: 10.1002/14651858.CD008286.pub3] [Medline: 27009521]
32. Ho A, Hancock J, Miner AS. Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. J Commun 2018;68(4):712-733 [FREE Full text] [doi: 10.1093/joc/jqy026] [Medline: 30100620]
33. Goh YS, Ow Yong JQY, Chee BQH, Kuek JHL, Ho CSH. Machine learning in health promotion and behavioral change: scoping review. J Med Internet Res 2022;24(6):e35831 [FREE Full text] [doi: 10.2196/35831] [Medline: 35653177]
34. Mittermaier M, Raza MM, Kvedar JC. Bias in AI-based models for medical applications: challenges and mitigation strategies. NPJ Digit Med 2023;6(1):113 [FREE Full text] [doi: 10.1038/s41746-023-00858-z] [Medline: 37311802]
35. Pocock SJ. Clinical Trials. 1st Edition. Chichester: Wiley; 2013.
36. Dingsøyr T, Nerur S, Balijepally V, Moe NB. A decade of agile methodologies: towards explaining agile software development. J Syst Softw 2012;85(6):1213-1221. [doi: 10.1016/j.jss.2012.02.033]
37. Design with people. Principles for Digital Development. URL: https://digitalprinciples.org/principles/design-with-people/ [accessed 2023-09-22]

Abbreviations

AI: artificial intelligence
REDCap: Research Electronic Data Capture
SCP: smoking cessation professional

Edited by D Gooch; submitted 17.01.24; peer-reviewed by D Chrimes, U Nair, E Reilly; comments to author 02.04.24; revised version received 27.05.24; accepted 27.06.24; published 19.08.24.

Please cite as:
Bendotti H, Lawler S, Ireland D, Gartner C, Marshall HM
Co-Designing a Smoking Cessation Chatbot: Focus Group Study of End Users and Smoking Cessation Professionals
JMIR Hum Factors 2024;11:e56505
URL: https://humanfactors.jmir.org/2024/1/e56505
doi:10.2196/56505
PMID:

Original Paper

Development of a Digital Health Intervention for the Secondary Prevention of Cardiovascular Disease (INTERCEPT): Co-Design and Usability Testing Study

Irene Gibson[1,2], MA; Lis Neubeck[3], PhD; Marissa Corcoran[4], MSc; Chris Morland[5], BA; Steve Donovan[5]; Jennifer Jones[1], PhD; Caroline Costello[2], MSc; Lisa Hynes[6], PhD; Aisling Harris[6], MSc; Mary Harrahill[7]; Mary Lillis[7], DipAMI; Alison Atrey[2], PhD; Chantal F Ski[8], PhD; Vilius Savickas[9], PhD; Molly Byrne[10], PhD; Andrew W Murphy[11], MD; John William McEvoy[1,2], PhD; David Wood[1,2], MD, MSc; Catriona Jennings[1,2], PhD

[1]School of Medicine, University of Galway, Galway, Ireland
[2]National Institute for Prevention and Cardiovascular Health, Galway, Ireland
[3]School of Health and Social Care, Edinburgh Napier University, Edinburgh, United Kingdom
[4]St James's Hospital, Dublin, Ireland
[5]Citrus Suite, Liverpool, United Kingdom
[6]Croí, West of Ireland Cardiac and Stroke Foundation, Galway, Ireland
[7]Public and Patient Involvement Panel, Croí, West of Ireland Cardiac and Stroke Foundation, Galway, Ireland
[8]School of Nursing and Midwifery, Queen's University Belfast, Belfast, United Kingdom
[9]School of Chemistry, Pharmacy and Pharmacology, University of East Anglia, Norwich, United Kingdom
[10]Health Behaviour Change Research Group, University of Galway, Galway, Ireland
[11]Health Research Board Primary Care Clinical Trials Network Ireland, University of Galway, Galway, Ireland

Corresponding Author:
Irene Gibson, MA
School of Medicine
University of Galway
University Road
Galway, H91 TK33
Ireland
Phone: 353 91 544310
Email: irene.gibson@universityofgalway.ie

Abstract

Background: Secondary prevention is an important strategy to reduce the burden of cardiovascular disease (CVD), a leading cause of death worldwide. Despite the growing evidence for the effectiveness of digital health interventions (DHIs) for the secondary prevention of CVD, the majority are designed with minimal input from target end users, resulting in poor uptake and usage.

Objective: This study aimed to optimize the acceptance and effectiveness of a DHI for the secondary prevention of CVD through co-design, integrating end users' perspectives throughout.

Methods: A theory-driven, person-based approach using co-design was adopted for the development of the DHI, known as INTERCEPT. This involved a 4-phase iterative process using online workshops. In phase 1, a stakeholder team of health care professionals, software developers, and public and patient involvement members was established. Phase 2 involved identification of the guiding principles, content, and design features of the DHI. In phase 3, DHI prototypes were reviewed for clarity of language, ease of navigation, and functionality. To anticipate and interpret DHI usage, phase 4 involved usability testing with participants who had a recent cardiac event (<2 years). To assess the potential impact of usability testing, the System Usability Scale was administered before and after testing. The GUIDED (Guidance for Reporting Intervention Development Studies in Health Research) checklist was used to report the development process.

Results: Five key design principles were identified: simplicity and ease of use, behavioral change through goal setting and self-monitoring, personalization, system credibility, and social support. Usability testing resulted in 64 recommendations for the app, of which 51 were implemented. Improvements in System Usability Scale scores were observed when comparing the results before and after implementing the recommendations (61 vs 83; *P*=.02).

Conclusions: Combining behavior change theory with a person-based, co-design approach facilitated the development of a DHI for the secondary prevention of CVD that optimized responsiveness to end users' needs and preferences, thereby potentially improving future engagement.

(JMIR Hum Factors 2024;11:e63707) doi:10.2196/63707

KEYWORDS

cardiovascular disease; secondary prevention; digital health; intervention development; co-design; usability testing; mobile health; usability; design; conline workshop; social support; behavioral change; self-monitoring

Introduction

Patients with coronary heart disease (CHD) are at high risk of recurrent cardiovascular events [1], with 1 in 5 experiencing a recurrent event in the first year after hospital discharge [2]. To minimize this risk, guidelines recommend early initiation of evidence-based, secondary prevention lifestyle and pharmacological treatments, together with access to a cardiac rehabilitation (CR) program [1,3]. However, in reality, only a minority of patients with CHD achieve optimal risk factor control and less than a half attend CR, usually several weeks or even months after hospital discharge [4]. The promising role of digital health interventions (DHIs), including mobile health (mHealth) apps, in addressing this implementation gap is increasingly recognized [5]. Indeed, the widespread availability of smartphones and mobile devices means that mHealth apps are one of the most common types of DHIs being developed to improve the secondary prevention of CHD [6,7].

mHealth apps have the potential to enhance patient empowerment and self-management [8], with growing evidence to support their use for secondary prevention of cardiovascular disease (CVD) [5]. Recent meta-analysis data show that mHealth apps can improve exercise capacity, physical activity, adherence to medication, and quality of life, as well as reduce hospital readmissions in patients with CHD [9]. However, beyond the research setting, the uptake and usage of health apps is low [6,10,11]. In addition, there is limited understanding of how these apps are developed and what their "active ingredients" include [12,13].

Involving end users in the design and development of DHIs is recognized as essential to maximizing their acceptance, uptake, and effectiveness [5,14]. Despite this, the majority of mHealth apps are designed with minimal input from target end users [13]. For example, in a recent scoping review, the use of co-design methodologies in the development of CVD secondary prevention interventions was reported in only 4 out of 22 studies related to mHealth apps [15]. Furthermore, none of these studies reported using a theoretical framework to guide the development of the intervention, including the use of theory-based behavioral strategies. Although this absence of reporting is not uncommon [16], it impacts on our ability to understand which parts of digital behavior interventions contribute to outcomes [17].

For DHIs to reach their full potential, we need to better understand how they are developed and what the core elements of effective interventions are [13]. This paper aims to describe the development of a multicomponent, complex DHI called "INTERCEPT" to improve secondary prevention in patients with CHD. Developed by the Irish National Institute for Prevention and Cardiovascular Health, INTERCEPT aims to promote self-management and support patients to achieve a healthy lifestyle, manage CVD risk factors, and improve adherence to cardioprotective medications. It includes an mHealth app, which integrates with a web-based health care professional (HCP) portal, a fitness wearable, and a blood pressure monitor. Responding to the need for early initiation of prevention after an index event [1], particularly given delays in patients accessing traditional (CR) programs [18,19], INTERCEPT is designed to be introduced to the patient at the time of their acute hospitalization and before discharge home. In this way, it provides a bridge to CR or an alternative for patients who choose not to join a CR program.

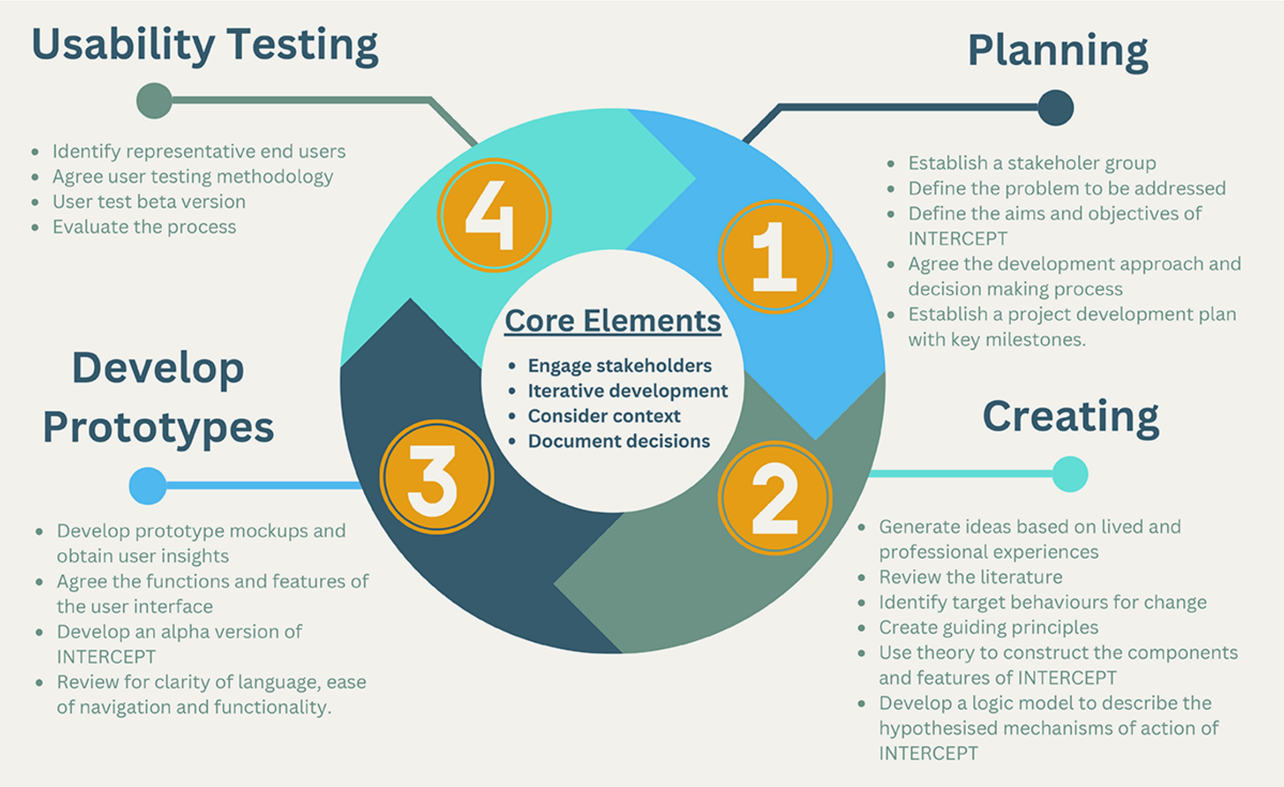
To maximize the potential effectiveness of INTERCEPT, the aim of this study was to adopt a theory- and evidence-based approach to development, integrating end users' (patients and HCPs) perspectives throughout. We use the term "development" to capture the whole process from initial planning to designing and usability testing.

Methods

Overview

We used a theory-driven, person-based approach to the development of INTERCEPT in an iterative co-design process. Our overall approach was guided by the UK Medical Research Council (MRC) framework for developing and evaluating complex interventions [20]. While co-design refers to meaningful end-user engagement across all aspects of intervention development [14], the person-based approach focuses on the psychosocial context of users and the behavioral elements of an intervention [21]. Through a series of qualitative workshops, conducted between May 2021 and December 2022, we adopted a 4-phase approach to development (Figure 1). As development commenced during the COVID-19 pandemic, all workshops were conducted online (Zoom, Zoom Video Communications). This paper is reported in accordance with the GUIDED (Guidance for Reporting Intervention Development Studies in Health Research) checklist (see Multimedia Appendix 1) [22].

Figure 1. Four phases of INTERCEPT development.



Phase 1: Planning

To ensure a diverse and inclusive approach to intervention development, an interdisciplinary stakeholder team was established. This comprised HCPs, software developers, and members from the public and patient involvement (PPI) panel of Croí (Gaelic for heart)—an Irish heart and stroke patient organization. Membership was determined by the individual's relevant expertise in the secondary prevention of CVD, including the design of complex DHIs; health behavior change; software development; and lived experience of CVD. To establish meaningful partnership relationships and enhance engagement, roles and responsibilities were agreed upon from the outset and a process for documenting and making decisions was established [15]. Furthermore, to enable stakeholders to participate to the best of their capabilities, a flexible and supportive approach to development was adopted. For example, workshops were centered around participant availability, and supports to address digital health literacy challenges were made available. Initial meetings focused on project conceptualization and identifying the key needs and challenges to be addressed. This led to the development of the intervention aim and objectives and identification of specific target behaviors for change. These behaviors were related to lifestyle (smoking cessation, making healthy food choices, and increasing physical activity), CVD risk factor management (blood pressure, lipids, and glucose), and adherence to cardioprotective medications. A project plan with key milestones for developing INTERCEPT was established.

Phase 2: Creating

This phase aimed to identify the guiding principles, including the design objectives, content, and features of INTERCEPT. In keeping with the person-based approach, guiding principles help to keep development focused on what is appealing and useful to the intended user, thus helping to maximize its acceptability and effectiveness [21]. INTERCEPT guiding principles were developed by (1) synthesizing the literature [10,23-26], (2) drawing on the stakeholders' clinical and research experiences of developing and implementing DHIs [26-28], and (3) obtaining patient lived experience insights. Informed by these principles, we used social cognitive theory (SCT) [29] and select behavior change techniques (BCTs) from the behavior change taxonomy [30] to help construct the intervention components and features required to achieve the objectives of INTERCEPT. Incorporating behavior change theory in the development of health interventions is known to enhance their effectiveness [20], and SCT is one of the most commonly used theoretical frameworks in secondary prevention DHI studies [12]. This process was facilitated by a series of structured brainstorming workshops with the stakeholder team, where creative idea generation around strategies to address specific target behaviors was encouraged. Contextual insights into the clinical care pathway after a cardiac event were incorporated by involving HCPs and patients in the user journey mapping process. In line with UK MRC guidance, a logic model was developed to articulate the key components of INTERCEPT and its expected mechanisms of action [20].

Phase 3: Developing Prototypes

To help visualize the basic layout and functionality of the app, the software developers initially produced a series of prototype mock-ups of the INTERCEPT digital interface. Through a series of workshops, these mock-ups underwent several iterations, integrating feedback from the entire stakeholder team, until a more refined design solution was reached. This led to the development of a usable first version of the app (alpha prototype), which was pilot-tested among the stakeholder team and reviewed multiple times for clarity of language, ease of navigation, and functionality. At this stage, decisions regarding the INTERCEPT logic (eg, to trigger push notifications), functional requirements (eg, registration process, user login, and safe data storage), integration of data with the HCP portal, and the use of analytics to capture usage patterns were made. Furthermore, to promote user trust and to ensure compliance with general data protection regulations (GDPR), a series of data privacy and security measures were considered in this phase. These included the introduction of encryption and access controls, adhering to data protection standards for data hosting and storage and ensuring data minimization, by developing the app to collect only the necessary personal information. In addition, a privacy policy detailing how personal data are collected, processed, and stored was developed.

INTERCEPT was developed for both iOS (Apple) and Android (Google) mobile phone platforms and was tested multiple times across a range of devices and operating systems to guarantee reliable performance and compatibility. Feedback from the workshops was collated and formulated into design specifications using a custom-designed data extraction spreadsheet. All new design features and content changes were prioritized based on their alignment to the guiding principles and their overall potential to enhance the acceptability and usability of INTERCEPT.

Phase 4: Usability Testing

To further develop INTERCEPT, the finished product (beta version) was subjected to usability testing by a separate group of patients, who were not part of the PPI panel. Usability testing is critical to determining if an intervention is meeting the end-user needs, and, ideally, it should be completed with end users in real-life contexts [31].

End-User Recruitment

Using purposive sampling, participants were recruited through 3 community groups, including Croí, the heart and stroke patient organization; a group representing Travellers (Indigenous ethnic minority individuals); and a farming organization. There was a specific focus on recruiting women, those living in rural isolation, and ethnic minority groups as these individuals are often underrepresented in digital health research [6,12]. Eligibility criteria included participants aged ≥40 years with a recent (≤2 years) diagnosis of CHD. Participants were required to have a smartphone, and family members were included as they play an important role in supporting engagement with technology in the home [32]. To minimize barriers to online engagement, participants were offered training in the use of Zoom technology. Furthermore, as trust in data security and privacy is a frequently reported barrier to DHI engagement [10], participants were informed of the measures taken to ensure that INTERCEPT was compliant with GDPR. We aimed to recruit a sample size of 10-12 participants to test usability, as this has been demonstrated to detect a minimum of 80% of usability problems, which is considered satisfactory for complex intervention testing [33].

Ethical Considerations

Ethical approval for phase 4 was granted by the clinical research ethics committee at Galway University Hospitals (Ref: C.A.2797) on May 11, 2022. Informed consent was obtained from all participants, who were informed of their right to withdraw from the study at any time. Data obtained from participants was handled according to GDPR and all data were anonymous to the study team.

Iterative Co-Design Workshops

Once consent was obtained, participants in phase 4 were provided with a link with instructions to download the app via email, a pedometer (Sportline 340; HRM USA Inc), a blood pressure monitor (UA 651 device; A&D), and a user support manual. The manual incorporated a diary, which participants were encouraged to use to document their experiences of using INTERCEPT [21]. Participants were invited to 5 online workshops over a 12-week period. This was considered sufficient time to allow participants to test the behavioral change elements of the app, such as goal setting and self-monitoring. Furthermore, this allowed for iterative amendments to be made to the app based on usability feedback. We were unable to find guidance for conducting co-design in an online environment; therefore, we adapted best practice principles for facilitation of in-person, co-design workshops, to help optimize engagement [34,35]. This involved presenting the value proposition "sharing your insights will benefit others," having clear workshop objectives with defined roles and expectations, allowing time for participants to share their lived experiences, being supportive around the use of technology, keeping the duration of usability workshops to no longer than 1.5 hours, and offering options for out-of-hours workshops as well as individual sessions. The topic guide (see Multimedia Appendix 2) for the usability workshops was developed with input from the stakeholder team, and participants were encouraged to refer to their diary to aid recollection of their experiences.

The validated System Usability Scale (SUS) was selected to evaluate the impact of usability testing [36]. The SUS assesses components of usability, effectiveness, efficiency, and satisfaction and includes 10 questions with a Likert scale, with values ranging from 0 to 100. Participants in phase 4 were invited to complete the self-administered SUS after an initial 2-week trial of the app and after the last workshop when final modifications to the app were made. To measure eHealth literacy, the validated eHealth Literacy Scale, an 8-item scale, presented as a score out of 40, was utilized [37]. Although there is no fixed cutoff to distinguish between high and low literacy, a higher score reflects a high level of eHealth literacy. Data on baseline characteristics of participants, eHealth literacy, and the SUS were obtained by providing participants with an email link to an online survey that was created using Survey Monkey.

Data Analysis

All quantitative data were analyzed using Stata (version 18; StataCorp). To compare the SUS data between the 2 time points, the nonparametric Mann-Whitney *U* test was used. Patient characteristics, information technology usage, and digital literacy were summarized using descriptive statistics. Qualitative data were analyzed by 2 study team members (IG and CJ) who used a deductive approach to map to 3 key aspects: usability (functionality and ease of navigation), comprehensibility (language), and content.

Results

Participant Characteristics

INTERCEPT followed an iterative cycle of development, involving 29 participants. These included 20 participants from the stakeholder team (4 nurse specialists, 1 physiotherapist, 1 physical activity specialist, 2 dietitians, 1 pharmacist, 2 health psychologists, 2 cardiologists, 2 software developers, and 5 PPI members), who were based in 6 different countries, and a separate group of 9 patients who were recruited for usability testing in phase 4.

INTERCEPT Guiding Principles

A review of the associated literature identified the following key app features: self-monitoring of health behaviors, behavior change motivation, education, personalized content, ease of use, and the ability to integrate with other apps and devices [10,23]. Feedback from the stakeholder team yielded further insights. INTERCEPT should (1) engage and motivate the user, (2) facilitate psychosocial support and contact with an HCP, (3) be credible and evidence based, (4) ensure data privacy, and (5) be introduced early in the recovery journey as patients often experience delays or have limited access to CR. By consolidating findings from the literature with stakeholder team insights, we identified 5 key guiding principles for the development of INTERCEPT. An outline of these principles, comprising of design objectives and intervention features to address these objectives, are presented in Table 1.

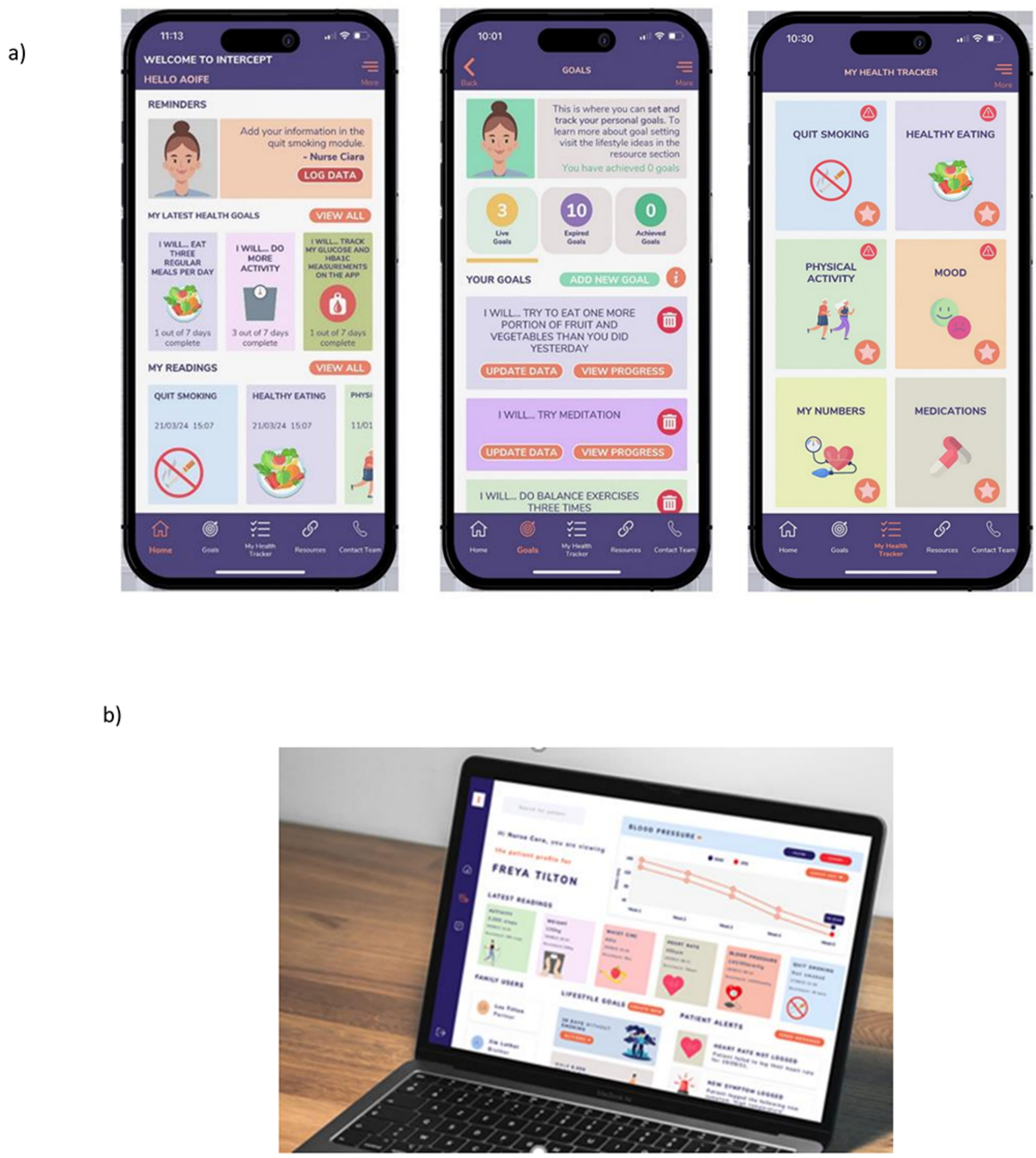
Table 1. INTERCEPT design-guiding principles.

Design objectives	Intervention features
Promote user competence	• Ensure simplicity across layout, language, and navigation procedures • Provide clear guidance on how to use INTERCEPT
Incorporate strategies to motivate and engage users in healthy behaviors	• Support goal setting, self-monitoring, and tracking of behaviors • Acknowledge app usage, and provide tailored feedback on self-reported progress, using rewards where appropriate • Use positive language, promoting the benefits of engaging in health behaviors
Adopt a personalized approach	• Offer choice in relation to how users engage in the app (graded goal setting, turning notification on and off, timing of reminders, and personalized messages)
Promote system credibility	• Be explicit in stating the organizations or people involved in developing INTERCEPT • Use trusted and credible resources, with appropriate links to reputable, noncommercial organizations • Comply with data privacy obligations, ensuring that appropriate data security measures are in place
Social support	• Establish a communication link with health care professionals through integrating the app with a health care professional portal

Developing Prototypes

After a series of workshops (n=7) with the stakeholder team (2 workshops with HCPs and 5 with PPI collaborators), the core components and features of INTERCEPT were agreed upon. In brief, the components included goal setting to motivate and support healthy lifestyle change; a health tracker to support self-monitoring of physical activity, mood, healthy eating, medications, blood pressure, cholesterol, and glucose levels; educational resources to increase knowledge and awareness of healthy lifestyle changes, psychosocial health, and adherence to cardioprotective medications; notifications to motivate and prompt engagement; and a link to an HCP portal to enable remote monitoring and support. A description of these components is provided in the TIDieR (Template for Intervention Description and Replication) checklist (see Multimedia Appendix 3), and screenshots are presented in Figure 2. To operationalize the intervention components and features, 14 BCTs from the taxonomy of BCTs were used. An overview of these BCTs, and how they align to the components and features of INTERCEPT, including the proposed mechanisms of actions and outcomes, is presented in the INTERCEPT logic model (see Multimedia Appendix 4). Examples of how this feedback influenced the end product features are as follows: (1) to promote user competence, additional guidance on goal setting was provided and a hard copy support manual was developed; (2) to motivate and engage users, the value proposition of the app describing its potential benefits was introduced on the home screen; (3) to ensure a personalized approach, a schedule of tailored notification messages mapped to individual usage patterns was developed; and (4) to emphasize psychological health, credible resources and messaging were included.

Figure 2. Screenshots of (A) the home screen, goal setting, and health tracker and (B) the health care professional portal.



Usability Testing

Usability testing was conducted between July and October 2022. In total, 11 individuals were recruited and 9 attended either the workshops (n=6) or semistructured interviews (n=3). Reasons for nonattendance were related to illness and family caring responsibilities. Baseline characteristics of participants, including demographics, CVD diagnosis, technology usage, and digital literacy, are reported in Table 2.

The mean age of participants was 62 years; 67% (6/9) were female, 89% (8/9) had a diagnosis of CVD, and 1 (11%) was a family member. Although the majority (7/9, 78%) of participants reported frequently using apps, the mean eHealth literacy score was 27.6 (SD 6.1). While acknowledging the small sample size, improvements were identified in overall mean SUS scores from 60.8 (SD 23.5) at commencement of usability testing to 82.8 (SD 7.1) after usability testing (*P*=.02; Table 3). Although improvements were observed across all 10 questions, they were significant for questions 2 (I found the app unnecessarily complex; *P*=.04), 3 (I thought the app was easy to use; *P*=.02), 5 (I found the various functions of the app well integrated; *P*=.03), and 9 (I felt confident using the app; *P*=.03).

Table 2. Baseline characteristics of usability testing participants.

Characteristic	Value
Demographics (n=9)	
Gender (female), n (%)	6 (67)
Age (y), mean (SD)	62.2 (11.4)
Enrolled in GMS[a] scheme, n (%)	7 (78)
Geographical region (n=9), n (%)	
Connacht	4 (44)
Leinster	3 (33)
Munster	2 (22)
Urban	3 (33)
Rural	6 (67)
Ethnicity (n=9), n (%)	
White Irish	8 (89)
Irish Traveller	1 (11)
Education level (n=9), n (%)	
Primary education	1 (11)
Secondary education	3 (33)
Technical or vocational	1 (11)
Third level diploma or degree	4 (44)
Cardiovascular diagnosis (self-reported;n=8), n (%)	
Myocardial infarction and percutaneous coronary intervention	3 (33)
Heart failure	2 (22)
Percutaneous coronary intervention	1 (11)
Heart valve disease	1 (11)
Atrial fibrillation	1 (11)
Cardiac rehabilitation program completion (n=8), n (%)	7 (78)
In person	4 (57)
Online	3 (43)
Information technology and digital literacy (n=8)	
Access to internet at home, n (%)	9 (100)
How often do you use apps on your phone? n (%)	
Frequently	7 (78)
Occasionally	1 (11)
Rarely	1 (11)
eHEALS[b] score, mean (SD)	27.6 (6.1)

[a]GMS: General Medical Scheme (refers to the means tested provision of state health care).

[b]eHEALS: eHealth Literacy Scale.

Across the 3 key areas of usability, comprehensibility, and content, participants made 64 suggestions for INTERCEPT, and of these, 51 were implemented. Through stakeholder consensus, reasons for not implementing suggestions were related to practical considerations (budget and time) and misalignment with the guiding principles. A summary of suggestions, across the themes of usability, comprehensibility, and content, were mapped to the intervention components (Multimedia Appendix 5). The following are selected insights. To enhance functionality and minimize the user burden associated with manual data entry, participants identified the importance of integrating data from the fitness wearable and blood pressure monitor with INTERCEPT. However, this change was not feasible to implement immediately, as it required consideration of multiple technical and user experience factors. These included ensuring reliable data transfer and compatibility between INTERCEPT and other devices; privacy and security; and the provision of adequate technical support, including user instructions and manuals to guide setup. Although addressing these factors required additional time and resources, data integration was achieved following usability testing. To motivate the user, participants recommended including rewards for setting and achieving goals and using language to promote more personal ownership, for example, changing "your data" to "my data." In recognizing the acute clinical context of implementation, participants emphasized the importance of including welcoming messages on the home screen to acknowledge the early stages of recovery. To enhance navigation, participants recommended greater integration between components. For example, if a user achieves a low mood score, they should receive prompts with links to the goal setting and resource sections, rather than having to access these components separately.

Table 3. Pre- and postusability testing System Usability Scale scores.

Question	Preusability testing (n=9), mean (SD)	Postusability testing (n=8), mean (SD)	*P* value
1. I think that I would like to use the app frequently	2.7 (1.4)	3.5 (1.1)	.12
2. I found the app unnecessarily complex	2.6 (1.1)	3.6 (0.5)	.04
3. I thought the app was easy to use	2.6 (1.2)	3.8 (0.5)	.02
4. I think that I would need assistance to be able to use this app	2.1 (1.3)	2.6 (0.9)	.39
5. I found the various functions in the app were well integrated	2.3 (1.0)	3.4 (0.7)	.03
6. I thought there was too much inconsistency in the app	2.3 (1.1)	3.1 (0.4)	.05
7. I would imagine that most people would learn to use this app very quickly	2.3 (1.2)	3.3 (1.0)	.08
8. I found the app very cumbersome or awkward to use	2.4 (1.2)	3.4 (0.5)	.06
9. I felt very confident using the app	2.8 (0.7)	3.5 (0.5)	.03
10. I needed to learn a lot of things before I could get going with the app	2.2 (1.6)	3.0 (0.5)	.37
Total score	60.8 (23.5)	82.8 (7.1)	.02

Discussion

Principal Findings

This paper describes the methods used to develop INTERCEPT, a multicomponent DHI, which integrates a smartphone app, HCP portal, fitness wearable, and blood pressure monitor to improve secondary prevention in patients with CHD. By providing a comprehensive and transparent description of methods used, we envision that our reporting will facilitate replication of the design process for future DHIs while also contributing to the growing science of intervention development. For creating effective DHIs, the use of the best combination of approaches to intervention development is required [38]. Accordingly, we used a combination of behavioral change theory, co-design, and the person-based approach to developing INTERCEPT. To our knowledge, this is one of the first studies to adopt this approach for the development of mHealth secondary prevention interventions. The broad range of actions undertaken across the development phases ensured the success of this approach. Many of these actions, for example, planning, designing, and creating, are consistent with guidance from the taxonomy of approaches to developing interventions by O'Cathain et al [39]. This highlights the relevance of using a taxonomy for future CVD DHI development research.

Developing DHIs is increasingly being recognized as a transdisciplinary endeavor, and meaningful stakeholder involvement is critical to successful co-design [15,38]. Despite this, a recent scoping review revealed that establishing meaningful relationships was the least reported process used in co-design studies in CVD secondary prevention [15]. Moreover, PPI is often absent from approaches to intervention development [39]. We addressed this by engaging PPI early in the development process and by using specific strategies, such as agreeing roles and responsibilities to foster meaningful relationships. Given the high level of stakeholder diversity (HCPs, patients, and software developers), the role of facilitators (study team members IG and CJ) was critical to ensuring an authentic understanding of the stakeholders' real-world experiences. Through their backgrounds as CVD nurse specialists, they were able to operate in an empathetic way, while also managing potential power imbalances between members of the stakeholder team. The interdisciplinary expertise of the stakeholder team helped to ensure that the users' needs and preferences, the influence of behavioral theory, best practice CVD prevention guidelines, and technical and practical factors were considered throughout development. Through dynamic, iterative cycles of development, this resulted in the development of a complex intervention, targeting multiple behaviors for change through a personalized user interface, with bidirectional communication via an HCP portal and remote monitoring technology. Consistent with findings from other DHI co-design studies, having the same PPI and HCP group involved throughout enhanced this process, allowing for more in-depth and intensive iteration along the development continuum [35,40].

Although most secondary prevention DHIs focus on outpatient settings, INTERCEPT is designed to be introduced to patients as early as possible after the diagnosis of CHD [41]. Based on recent systematic review and realist synthesis evidence, focusing on early engagement, ensuring a personalized approach, and providing opportunities to connect with an HCP are all features associated with improved participation in telehealth CR and CVD health outcomes [42]. Although HCP support has been identified as one of the most important factors associated with engagement in mHealth DHIs [25], it is important to establish the extent to which it can add further value, without becoming resource intensive [43]. The INTERCEPT portal strives to achieve this balance through protocol-driven remote monitoring and support based on real-time data.

To our knowledge, INTERCEPT is one of the few secondary prevention mHealth DHIs that target multiple behaviors for change. This is particularly important as lifestyle- and medical-related CVD risk factors and their corresponding behaviors are strongly interlinked [1]. Although the guiding principles were key to identifying key components and features to help address these behaviors, using BCTs targeting determinants of SCT helped to operationalize them. Many of the BCTs included in INTERCEPT (eg, self-monitoring of behavior [2.3], information about health consequences [5.1], and goal setting [1.1]) align to recent systematic review evidence for effective BCTs in CR DHIs [12]. By presenting an overview of these BCTs as part of the INTERCEPT logic model, we respond to the increasing calls for explicit reporting of components and expected mechanisms of action of co-designed interventions [15,20].

Similar to findings by Tay et al [44], this paper highlights that it is feasible to develop DHIs in an online environment. Importantly, this enabled us to foster an inclusive approach to intervention development where geography was not a barrier. Indeed, the diverse group of representative end users in the stakeholder team and recruited for usability testing was a key strength of the INTERCEPT development process. For example, among those recruited for usability testing, one-third had never previously attended CR; almost 70% were women; 78% were enrolled in the General Medical Scheme (means tested provision of state health care); 1 individual identified as an Irish Traveller; and eHealth literacy levels were mixed, with scores ranging from 15 to 34 out of 40 (mean score 27.6). Given how cultural and socioeconomic factors including gender and digital literacy influence the acceptability of DHIs, involving diverse populations in their design is essential to optimizing their potential benefits [5]. Outcomes from the usability-testing phase resulted in a number of refinements being made to INTERCEPT. Data from the SUS evaluation suggest that these refinements potentially led to improvements in its usability. This highlights the important value of usability testing in optimizing the user experience and overall quality of the product [35].

Following UK MRC guidance for developing and evaluating complex interventions, our next step in the intervention development process is to evaluate the feasibility of INTERCEPT among a sample of patients with CHD in a real-world clinical setting. By conducting a nonrandomized, mixed methods, feasibility study, the acceptability and usability of INTERCEPT will be assessed. These insights will help to (1) inform further refinement of INTERCEPT and (2) determine the feasibility of a definitive randomized controlled trial. The protocol for this feasibility study is detailed elsewhere [45].

Limitations

The INTERCEPT development process had some limitations. First, the outcomes of our co-design process focus primarily on the development of INTERCEPT. However, there is also a need to report on stakeholder experiences and cost-effectiveness [14,46]. This would help address the paucity of data on the impact of co-design processes on stakeholders, while also informing strategies to optimize participatory approaches for intervention development [47]. Second, co-design is time-consuming, and the evolving nature of technology demands more fast-paced design processes to ensure that DHIs remain relevant and engaging [46]. Although our development timeframe was consistent with in person co-design studies [13], further research is required to ascertain if online approaches can improve efficiencies. Third, usability testing of INTERCEPT required participants to retrospectively report their experiences over a defined period of time, which may have resulted in insights been forgotten or distorted. Including the use of think aloud interviews, whereby participants are asked to give their immediate reactions to every element, may help to address this. However, further guidance on using this approach online is required. Fourth, social desirability bias may have influenced the SUS results; to address this, the survey was self-administered and the results were anonymous to the study team. Finally, because of the low sample size, the SUS results should be interpreted with caution; further research with a larger sample size is warranted to help evaluate the usability testing process.

Conclusions

The proliferation of DHIs including mHealth has enabled the development of new and innovative approaches for the secondary prevention of CVD. However, careful attention to the development of these DHIs is required to increase their effectiveness and uptake in clinical practice. This paper illustrates how combining behavior change theory with a person-based, co-design approach to DHI development is feasible and can result in the successful development of an intervention that responds to end users' needs and preferences, including desired content and features. Additionally, our comprehensive reporting offers guidance to other researchers for developing future DHIs, thus facilitating the translation of evidence into practice.

Acknowledgments

The authors wish to thank the participants who kindly participated in this study; members of the Croí public and patient involvement panel, who advised on all aspects of the development of INTERCEPT (Mary Harrahill, Eilish Kelly, Mary Lillis, Pat McDonald, Terry Small, and Mary Verbruggen); the staff of Croí including Annie Costelloe and Maeve Frawley, who supported with recruitment; the staff of Citrus Suite who supported technical and software development; and Professors Julie Redfern and Robyn Gallagher from the University of Sydney, who provided expert project advice. This research study is supported by the Health Research Board Ireland Collaborative Doctoral Award 2019 (CDA-2019-001); the National Institute for Prevention and Cardiovascular Health, Ireland; the Preventive Cardiology Trust, United Kingdom; and a SOLVE CHD SYNERGY Grant, University of Sydney, Australia. The development of the INTERCEPT software was supported through an unrestricted educational grant by Novartis and NovoNordisk.

Conflicts of Interest

None declared.

Multimedia Appendix 1
GUIDED (Guidance for Reporting Intervention Development Studies in Health Research) checklist.
[DOCX File , 68 KB - humanfactors_v11i1e63707_app1.docx]

Multimedia Appendix 2
Topic guide for usability testing.
[DOCX File , 25 KB - humanfactors_v11i1e63707_app2.docx]

Multimedia Appendix 3
TIDieR (Template for Intervention Description and Replication) checklist.
[DOCX File , 61 KB - humanfactors_v11i1e63707_app3.docx]

Multimedia Appendix 4
INTERCEPT logic model.
[PDF File (Adobe PDF File), 85 KB - humanfactors_v11i1e63707_app4.pdf]

Multimedia Appendix 5
Insights from usability testing.
[DOCX File , 27 KB - humanfactors_v11i1e63707_app5.docx]

References

1. Visseren FLJ, Mach F, Smulders YM, Carballo D, Koskinas KC, Bäck M, et al. 2021 ESC guidelines on cardiovascular disease prevention in clinical practice. Eur Heart J 2021;42(34):3227-3337. [doi: 10.1093/eurheartj/ehab484] [Medline: 34458905]
2. Jernberg T, Hasvold P, Henriksson M, Hjelm H, Thuresson M, Janzon M. Cardiovascular risk in post-myocardial infarction patients: nationwide real world data demonstrate the importance of a long-term perspective. Eur Heart J 2015;36(19):1163-1170. [doi: 10.1093/eurheartj/ehu505] [Medline: 25586123]
3. The BACPR Standards and Core Components for Cardiovascular Prevention and Rehabilitation 2023. London: British Association for Cardiovascular Prevention and Rehabilitation; 2023. URL: https://tinyurl.com/jnj4xu65
4. Kotseva K, de Backer G, de Bacquer D, Rydén L, Hoes A, Grobbee D, et al. Lifestyle and impact on cardiovascular risk factor control in coronary patients across 27 countries: results from the European society of cardiology ESC-EORP EUROASPIRE V registry. Eur J Prev Cardiol 2019;26(8):824-835 [FREE Full text] [doi: 10.1177/2047487318825350] [Medline: 30739508]
5. Tromp J, Jindal D, Redfern J, Bhatt A, Séverin T, Banerjee A, et al. World Heart Federation roadmap for digital health in cardiology. Glob Heart 2022;17(1):61 [FREE Full text] [doi: 10.5334/gh.1141] [Medline: 36051317]
6. Wongvibulsin S, Habeos EE, Huynh PP, Xun H, Shan R, Porosnicu Rodriguez KA, et al. Digital health interventions for cardiac rehabilitation: systematic literature review. J Med Internet Res 2021;23(2):e18773 [FREE Full text] [doi: 10.2196/18773] [Medline: 33555259]
7. Santo K, Redfern J. Digital health innovations to improve cardiovascular disease care. Curr Atheroscler Rep 2020;22(12):71 [FREE Full text] [doi: 10.1007/s11883-020-00889-x] [Medline: 33009975]
8. Cruz-Ramos NA, Alor-Hernández G, Colombo-Mendoza LO, Sánchez-Cervantes JL, Rodríguez-Mazahua L, Guarneros-Nolasco LR. mHealth apps for self-management of cardiovascular diseases: a scoping review. Healthcare (Basel) 2022;10(2):322 [FREE Full text] [doi: 10.3390/healthcare10020322] [Medline: 35206936]
9. Cruz-Cobo C, Bernal-Jiménez M, Vázquez-García R, Santi-Cano MJ. Effectiveness of mHealth interventions in the control of lifestyle and cardiovascular risk factors in patients after a coronary event: systematic review and meta-analysis. JMIR Mhealth Uhealth 2022;10(12):e39593 [FREE Full text] [doi: 10.2196/39593] [Medline: 36459396]

10. Coorey GM, Neubeck L, Mulley J, Redfern J. Effectiveness, acceptability and usefulness of mobile applications for cardiovascular disease self-management: Systematic review with meta-synthesis of quantitative and qualitative data. Eur J Prev Cardiol 2018;25(5):505-521. [doi: 10.1177/2047487317750913] [Medline: 29313363]
11. Amagai S, Pila S, Kaat AJ, Nowinski CJ, Gershon RC. Challenges in participant engagement and retention using mobile health apps: literature review. J Med Internet Res 2022;24(4):e35120 [FREE Full text] [doi: 10.2196/35120] [Medline: 35471414]
12. Kenny E, Coyne R, McEvoy JW, McSharry J, Taylor RS, Byrne M. Behaviour change techniques and intervention characteristics in digital cardiac rehabilitation: a systematic review and meta-analysis of randomised controlled trials. Health Psychol Rev 2024;18(1):189-228 [FREE Full text] [doi: 10.1080/17437199.2023.2185653] [Medline: 36892523]
13. Eyles H, Jull A, Dobson R, Firestone R, Whittaker R, Te Morenga L, et al. Co-design of mHealth delivered interventions: a systematic review to assess key methods and processes. Curr Nutr Rep 2016;5(3):160-167. [doi: 10.1007/s13668-016-0165-7]
14. Slattery P, Saeri AK, Bragge P. Research co-design in health: a rapid overview of reviews. Health Res Policy Syst 2020;18(1):17 [FREE Full text] [doi: 10.1186/s12961-020-0528-9] [Medline: 32046728]
15. Talevski J, Kulnik ST, Jessup RL, Falls R, Cvetanovska N, Beauchamp A. Use of co-design methodology in the development of cardiovascular disease secondary prevention interventions: a scoping review. Health Expect 2023;26(1):16-29 [FREE Full text] [doi: 10.1111/hex.13633] [Medline: 36366855]
16. Buss VH, Varnfield M, Harris M, Barr M. A mobile app for prevention of cardiovascular disease and type 2 diabetes mellitus: development and usability study. JMIR Hum Factors 2022;9(2):e35065 [FREE Full text] [doi: 10.2196/35065] [Medline: 35536603]
17. Michie S, West R, Sheals K, Godinho CA. Evaluating the effectiveness of behavior change techniques in health-related behavior: a scoping review of methods used. Transl Behav Med 2018;8(2):212-224 [FREE Full text] [doi: 10.1093/tbm/ibx019] [Medline: 29381786]
18. Petty-Saphon N. A Needs Assessment for Phase III Cardiac Rehabilitation Services in Ireland. Ireland: Faculty of Public Health Medicine of the Royal College of Physicians of Ireland; 2016.
19. Grace SL, Kotseva K, Whooley MA. Cardiac rehabilitation: under-utilized globally. Curr Cardiol Rep 2021;23(9):118. [doi: 10.1007/s11886-021-01543-x] [Medline: 34269894]
20. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of medical research council guidance. BMJ 2021;374:n2061 [FREE Full text] [doi: 10.1136/bmj.n2061] [Medline: 34593508]
21. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
22. Duncan E, O'Cathain A, Rousseau N, Croot L, Sworn K, Turner KM, et al. Guidance for reporting intervention development studies in health research (GUIDED): an evidence-based consensus study. BMJ Open 2020;10(4):e033516 [FREE Full text] [doi: 10.1136/bmjopen-2019-033516] [Medline: 32273313]
23. Gallagher R, Parker H, Zhang L, Kirkness A, Roach K, Belshaw J, et al. Target audience and preferences related to an Australian coronary heart disease specific mobile app: a mixed methods study. Heart Lung Circ 2020;29(5):696-702. [doi: 10.1016/j.hlc.2019.05.178] [Medline: 31235365]
24. Neubeck L, Lowres N, Benjamin EJ, Freedman SB, Coorey G, Redfern J. The mobile revolution—using smartphone apps to prevent cardiovascular disease. Nat Rev Cardiol 2015;12(6):350-360. [doi: 10.1038/nrcardio.2015.34] [Medline: 25801714]
25. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]
26. Neubeck L, Coorey G, Peiris D, Mulley J, Heeley E, Hersch F, et al. Development of an integrated e-health tool for people with, or at high risk of, cardiovascular disease: the Consumer Navigation of Electronic Cardiovascular Tools (CONNECT) web application. Int J Med Inform 2016;96:24-37. [doi: 10.1016/j.ijmedinf.2016.01.009] [Medline: 26847070]
27. Gibson I, McCrudden Z, Dunne D, Harris A, Hynes L, Murphy E, et al. Harnessing digital health to optimise the delivery of guideline-based cardiac rehabilitation during COVID-19: an observational study. Open Heart 2023;10(1):e002211 [FREE Full text] [doi: 10.1136/openhrt-2022-002211] [Medline: 36927866]
28. Neubeck L, Freedman B, Lowres N, Hyun K, Orchard J, Briffa T, et al. Choice of health options in prevention of cardiovascular events (CHOICE) replication study. Heart Lung Circ 2018;27(12):1406-1414. [doi: 10.1016/j.hlc.2017.09.016] [Medline: 29100841]
29. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
30. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]

31. Daud MH, Yusoff FH, Abdul-Razak S, Baharudin N, Mohamed-Yassin M, Badlishah-Sham SF, et al. Design, development, utility and usability testing of the EMPOWER-SUSTAIN self-management mobile app among primary care physicians and patients with metabolic syndrome. Digit Health 2023;9:20552076231176645 [FREE Full text] [doi: 10.1177/20552076231176645] [Medline: 37312957]
32. Whitelaw S, Pellegrini DM, Mamas MA, Cowie M, van Spall HGC. Barriers and facilitators of the uptake of digital health technology in cardiovascular care: a systematic scoping review. Eur Heart J Digit Health 2021;2(1):62-74 [FREE Full text] [doi: 10.1093/ehjdh/ztab005] [Medline: 34048508]
33. Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. Behav Res Methods Instrum Comput 2003;35(3):379-383. [doi: 10.3758/bf03195514] [Medline: 14587545]
34. McKercher KA. Beyond Sticky Notes: Co-Design for Real: Mindsets, Methods and Movements. Australia: Thorpe-Bowker Identifier Services; 2020.
35. Nesbitt K, Beleigoli A, Du H, Tirimacco R, Clark RA. User experience (UX) design as a co-design methodology: lessons learned during the development of a web-based portal for cardiac rehabilitation. Eur J Cardiovasc Nurs 2022;21(2):178-183. [doi: 10.1093/eurjcn/zvab127] [Medline: 35030261]
36. Brooke J. Usability Evaluation in Industry. London: Taylor & Francis; 1996.
37. Norman CD, Skinner HA. eHEALS: the eHealth literacy scale. J Med Internet Res 2006;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
38. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (integrate, design, assess, and share): a framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. J Med Internet Res 2016;18(12):e317 [FREE Full text] [doi: 10.2196/jmir.5927] [Medline: 27986647]
39. O'Cathain A, Croot L, Sworn K, Duncan E, Rousseau N, Turner K, et al. Taxonomy of approaches to developing interventions to improve health: a systematic methods overview. Pilot Feasibility Stud 2019;5:41 [FREE Full text] [doi: 10.1186/s40814-019-0425-6] [Medline: 30923626]
40. Pelly ML, Fatehi F, Liew D, Verdejo-Garcia A. Digital health secondary prevention using co-design procedures: focus group study with health care providers and patients with myocardial infarction. JMIR Cardio 2023;7:e49892 [FREE Full text] [doi: 10.2196/49892] [Medline: 37902821]
41. Marvel FA, Spaulding EM, Lee MA, Yang WE, Demo R, Ding J, et al. Digital health intervention in acute myocardial infarction. Circ Cardiovasc Qual Outcomes 2021;14(7):e007741 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.121.007741] [Medline: 34261332]
42. Gallegos-Rejas VM, Rawstorn JC, Gallagher R, Mahoney R, Thomas EE. Key features in telehealth-delivered cardiac rehabilitation required to optimize cardiovascular health in coronary heart disease: a systematic review and realist synthesis. Eur Heart J Digit Health 2024;5(3):208-218 [FREE Full text] [doi: 10.1093/ehjdh/ztad080] [Medline: 38774382]
43. Tighe SA, Ball K, Kensing F, Kayser L, Rawstorn JC, Maddison R. Toward a digital platform for the self-management of noncommunicable disease: systematic review of platform-like interventions. J Med Internet Res 2020;22(10):e16774 [FREE Full text] [doi: 10.2196/16774] [Medline: 33112239]
44. Tay BSJ, Edney SM, Brinkworth GD, Cox DN, Wiggins B, Davis A, et al. Co-design of a digital dietary intervention for adults at risk of type 2 diabetes. BMC Public Health 2021;21(1):2071 [FREE Full text] [doi: 10.1186/s12889-021-12102-y] [Medline: 34763701]
45. Gibson I, Jennings C, Neubeck L, Corcoran M, Wood D, Sharif F, et al. Using a digital health intervention "INTERCEPT" to improve secondary prevention in coronary heart disease (CHD) patients: protocol for a mixed methods non-randomised feasibility study. HRB Open Res 2023;6:43 [FREE Full text] [doi: 10.12688/hrbopenres.13781.1] [Medline: 38414839]
46. Malloy J, Partridge SR, Kemper JA, Braakhuis A, Roy R. Co-design of digital health interventions with young people: a scoping review. Digit Health 2023;9:20552076231219117 [FREE Full text] [doi: 10.1177/20552076231219117] [Medline: 38107986]
47. Grindell C, Coates E, Croot L, O'Cathain A. The use of co-production, co-design and co-creation to mobilise knowledge in the management of health conditions: a systematic review. BMC Health Serv Res 2022;22(1):877 [FREE Full text] [doi: 10.1186/s12913-022-08079-y] [Medline: 35799251]

Abbreviations

BCT: behavior change technique
CHD: coronary heart disease
CR: cardiac rehabilitation
CVD: cardiovascular disease
DHI: digital health interventions
GDPR: general data protection regulations
GUIDED: Guidance for Reporting Intervention Development Studies in Health Research
HCP: health care professional
mHealth: mobile health

MRC: Medical Research Council
PPI: public and patient involvement
SCT: social cognitive theory
SUS: System Usability Scale
TIDieR: Template for Intervention Description and Replication

Edited by A Kushniruk, C Jacob; submitted 28.06.24; peer-reviewed by E Harris, A Panahi; comments to author 06.08.24; revised version received 25.08.24; accepted 26.08.24; published 23.10.24.

Please cite as:
Gibson I, Neubeck L, Corcoran M, Morland C, Donovan S, Jones J, Costello C, Hynes L, Harris A, Harrahill M, Lillis M, Atrey A, Ski CF, Savickas V, Byrne M, Murphy AW, McEvoy JW, Wood D, Jennings C
Development of a Digital Health Intervention for the Secondary Prevention of Cardiovascular Disease (INTERCEPT): Co-Design and Usability Testing Study
JMIR Hum Factors 2024;11:e63707
URL: https://humanfactors.jmir.org/2024/1/e63707
doi:10.2196/63707
PMID:39441626

Original Paper

Factors Related to mHealth App Use Among Japanese Workers: Cross-Sectional Survey

Itsuko Ozaki[1], PhD; Mariko Nishijima[2], PhD; Eiji Shibata[3], MD, PhD; Yuri Zako[1], MSN; Chifa Chiang[1], PhD

[1]Graduate School of Nursing, Nagoya City University, Nagoya, Japan
[2]Graduate School of Medicine, Ehime University, Toon, Japan
[3]Yokkaichi Nursing and Medical Care University, Yokkaichi, Japan

Corresponding Author:
Itsuko Ozaki, PhD
Graduate School of Nursing
Nagoya City University
1-4-7 Aoi, Higashi-ku
Nagoya, 461-0004
Japan
Phone: 81 52 982 7343
Email: itsuko@med.nagoya-cu.ac.jp

Abstract

Background: Health care providers can make health guidance more effective by using mobile health technologies such as health apps. Although health care providers need to know who uses health apps, existing studies have yielded inconsistent results.

Objective: The aim of the study was (1) to clarify the prevalence and patterns of health app use to improve health behaviors for preventing lifestyle-related diseases among Japanese workers and (2) to identify the associations among demographic characteristics, health behavior, and internet use and health app use by gender.

Methods: Data were collected from a cross-sectional internet survey in 2023. In total, 2200 participants were included, with an even distribution of men and women in each age group aged 20 to 60 years. The participants were workers with smartphones and reported their gender, age, residence area, marital status, education, employment status, occupation, work pattern, diseases under treatment, health checkups, health guidance, health behaviors, internet use duration, and number of devices used. We asked about current and previous health app use for 1 month. A multivariate logistic regression analysis was conducted by gender.

Results: Of the participants, 472 (21.5%) and 189 (8.6%) were current and previous health app users, respectively. Most current and previous health app users used features that record and track their physical activity and other health behaviors. Health app users—both men and women—were more likely to have health checkups (odds ratio [OR] 1.53, 95% CI 1.12-2.11 and OR 1.51, 95% CI 1.10-2.07, respectively), receive health guidance (OR 2.01, 95% CI 1.47-2.74 and OR 1.86, 95% CI 1.32-2.62, respectively), engage in regular physical activity (OR 2.57, 95% CI 1.91-3.47 and OR 1.94, 95% CI 1.41-2.67, respectively), use the internet for 120-179 minutes per day (OR 1.76, 95% CI 1.13-2.75 and OR 1.70, 95% CI 1.12-2.57, respectively), and were less likely to be older (50-59 years: OR 0.54, 95% CI 0.33-0.88 and OR 0.40, 95% CI 0.25-0.6, respectively, and 60-69 years: OR 0.37, 95% CI 0.22-0.62 and OR 0.47, 95% CI 0.28-0.77, respectively). According to gender, male health app users were more likely to be married (OR 1.69, 95% CI 1.23-2.33) and less likely to work in the security, agriculture, forestry, fishing, manufacturing, or transportation industries (OR 0.62, 95% CI 0.41-0.95). Female health app users were more likely to have a university education or higher (OR 1.55, 95% CI 1.061-2.26), maintain an appropriate body weight (OR 1.52, 95% CI 1.10-2.11), and use 3 or more devices (OR 2.13, 95% CI 1.41-3.23).

Conclusions: Physical activity and health guidance are strong predictors of app use. Health care providers should assess the target populations' preferences for app use based on their characteristics, support their app use, and enhance the effectiveness of health guidance.

(JMIR Hum Factors 2024;11:e54673) doi:10.2196/54673

KEYWORDS

mHealth; mobile health; mobile health apps; prevalence; health promotion; health management; Japanese worker; Japan; cross-sectional survey; disease management; app users; physical activity

Introduction

Internet and communication technologies are developing rapidly worldwide. The proportion of people aged 20-59 years who use the internet in Japan is 95.2-98.4% [1]. The proportion of households with smartphones has risen from 49.5% to 88.6% in the past decade, and individuals are more likely to use a smartphone than a PC when connecting to the internet. Against this background, mobile health (mHealth) technologies have been developed and improved to support behavioral change and disease management. Thousands of health management apps (health apps) are available for smartphones, and their technical features include tracking behavior manually or semiautomatically, app communities, and app reminders [2]. In addition, these apps use behavior change techniques such as goal setting, self-monitoring, and feedback to help users improve their health habits [2]. Interventions involving health apps are increasing and are being evaluated [3-5]. A review of health app interventions reported improved physical functioning, adherence to prescribed medications, ease of symptom evaluation, and reports to care providers for people with chronic conditions [3]. In another review, most studies demonstrated significant improvements in health outcomes [4].

In Japan, one-third of men and one-fifth of women have obesity or overweight [6], leading to chronic diseases that increase medical costs. To prevent chronic diseases, promote health, and extend life expectancy, improving individuals' lifestyles from a young age is important. Consequently, health checkups and guidance are provided to almost all segments of the population in Japan [7]. Under the Industrial Safety and Health Law, employers are required to conduct annual health checkups for full-time employees to prevent the worsening of employee health due to work. Furthermore, employers must endeavor to provide health guidance to employees who need to improve and maintain their health. Additionally, as a strategy to prevent lifestyle-related diseases, medical insurers are obligated to provide a specified health checkup to all insured persons aged between 40 and 74 years under the Act on Ensuring Medical Care for Elderly People. Based on the checkup results, specific health guidance is provided to individuals identified as having a high risk of developing lifestyle-related diseases. Thus, Japan has a unique health care system that aims to promote individuals' efforts to improve their lifestyles based on checkup results. Health care providers, such as physicians, nurse practitioners, and dietitians, are required to provide efficient and effective health guidance. However, workers in Japan often face challenges in receiving health guidance due to time constraints. Health apps have the potential to address this issue by monitoring workers' health-related data and facilitating message exchanges between health care providers and workers. This allows workers to receive health guidance even if they do not frequently meet with health care providers. Given the high rate of smartphone and internet use among the Japanese working population, mHealth can be used to make health guidance more efficient and effective.

Recent studies have examined the effectiveness of using an app for specific health guidance [8,9]. However, health care providers need to know which individuals use health apps to manage their health and lifestyle to effectively provide health guidance in this manner.

Several studies have examined factors associated with the use of health apps. Regarding sociodemographic factors, some studies [10-12] report that women are more likely to use or have health apps than men, while other studies [13-16] report no difference between men and women in terms of having or using health apps. Several studies [10,12,13,15,17] found that younger generations were more likely to have or use health apps, whereas other studies [11,15,16] found no differences between generations. Although Carroll et al [10] reported that individuals with greater education were more likely to have health apps, other studies [11,13-16] found no difference among educational levels. Regarding chronic conditions, Ernsting et al [13] reported that individuals with 1, 2, and 3 or more chronic conditions were more likely to use health apps than those without chronic conditions.

Conversely, Langford et al [17] reported that individuals without a self-reported history of hypertension were more likely to have health apps on their tablets or smartphones than those with a self-reported history of hypertension. Makowsky et al [14] and Robbins et al [18] found no association between health app downloads and chronic conditions. Three studies [10,13,18] that examined health apps and physical activity reported that individuals who maintained physical activity were more likely to use or have health apps than those who did not. On the contrary, Fewings et al [16] found no difference between diet-related app use and physical activity or other health behaviors such as vegetable, fruit, and fast food consumption and smoking status. However, these studies are not easily comparable because of differences in their methods and duration. Consequently, the factors associated with health app use are inconsistent across them. Moreover, few studies have examined health app use and identified its associated factors, which are expected to differ between countries because of variations in the prevalence of smartphones, health care systems, and internet literacy. mHealth technologies are developing rapidly, and attitudes toward and confidence in using health apps may change.

Against this background, we aimed (1) to clarify the prevalence and patterns of health app use to improve health behaviors and thus prevent lifestyle-related diseases among Japanese workers and (2) to identify the associations between demographic characteristics, health behavior, and internet use and health app use by gender. We focused on health apps that have functions for tracking lifestyle, body weight, and medications; searching for health information; and interacting with other users to improve health behaviors and prevent lifestyle-related diseases. Since gender differences exist in the prevalence of lifestyle-related diseases and health behaviors among Japanese citizens [6], gender may influence app use patterns. Therefore, we hypothesized that the factors related to health app use would differ between men and women. We examined which sociodemographic, health behavior, and internet use factors were associated with health app use by gender. Moreover, health care providers should consider clients' preferences for tools that support behavior change during health guidance. For instance, understanding who prefers to use health apps for behavior change during health guidance is useful. Similarly, health care providers can more efficiently introduce health apps to clients by considering factors related to app use. In this study, we specifically focused on the current and previous use of health apps for over a period of more than a month. This is because most smartphones come with preinstalled health apps, and some apps offer a free 30-day trial. Consequently, health app users may use these apps intermittently. For example, they might temporarily stop using them once they can manage their target health behaviors or before the 30-day free trial ends.

Methods

Sample and Procedure

Data for this study were collected from a cross-sectional internet survey conducted by the research company Cross Marketing on February 13-14, 2023. The total number of participants was 2200. We collected data from an equal number of men and women in each age group from 20 to 60 years. Eligible participants were aged 20-69 years, were workers (including part-time work), and had a smartphone. Potential participants were sent an email asking them to complete a preliminary survey to verify their eligibility. Eligible participants were then asked to complete the main survey. The survey was terminated once the target number of participants was reached. Of the 2873 potential participants who completed the preliminary survey, 76.9% (n=2200) completed the primary survey.

All survey procedures were performed by Cross Marketing. Cross Marketing has over 5 million panelists registered across the country. They regularly collect basic information from registered panelists, which was confirmed before conducting the survey. They screened for panelists whose response times were extremely shorter than expected and excluded them. In addition, responses related to demographic characteristics were examined, and panelists with contradictory or inconsistent answers were excluded.

Measures

Demographic Characteristics

We asked the participants about their gender, age, area of residence, marital status, education, employment status, occupation, and work patterns.

Health and Medical Services and Health Behaviors

We asked participants four questions: (1) “Do you have any diseases under treatment?” (yes or no responses), (2) “Do you have annual health checkups or physical examinations?” (yes or no responses), (3) “Have you received health guidance from a doctor, nurse, or dietician?” (3 response options ranging from 0 to ≥2), and (4) “Do you implement the following health behaviors (select all those that you implement): do not smoke, engage in regular physical activity, drink moderately or do not drink, get enough sleep each day, maintain an appropriate weight, and eat breakfast every day?” We reclassified response options into a single dichotomous outcome variable of receiving health guidance (never or more than once).

Internet Use

We asked participants, “Apart from using the internet for work purposes, how much time do you spend on the internet per day? (4 response options ranging from <60 to ≥180 minutes).” Participants were asked to choose all of the devices they used to access the internet on a daily basis (eg, smartphones, computers, tablets, mobile phones (excluding smartphones), internet-enabled television receivers, home game consoles, and wearable devices). We categorized the participants according to the number of devices they used: 1, 2, or 3 or more.

Current and Previous Use of Health Apps

The participants were asked, “Have you ever used health apps to improve health behaviors for chronic disease prevention or weight loss? (Specifically, health apps that have functions for tracking lifestyle, body weight, and medications; searching for health information; and interacting with other users to improve health behavior for weight loss or prevent lifestyle-related diseases).” The response options were “currently using health apps for over 1 month,” “have used health apps in the past for over 1 month,” and “never used health apps (including having downloaded health apps and only used them for less than 1 month).” Participants who are currently using health apps for over 1 month and those who have used health apps in the past for over 1 month were asked the following questions: “What features of the apps do you (or did you) use?” (13 response options, multiple answers), “Do you (or did you) pay to use the health apps?” (yes or no responses), “Do you (or did you) use the apps in conjunction with wearable devices (including those who do not have wearable devices)?” (yes or no responses), and “How motivated are (or were) you to use the apps?” (5 response options, multiple answers). Participants who had never used health apps (including those who downloaded health apps but only used them for less than 1 month) and those who had used health apps in the past for over 1 month but stopped using them were asked why they did not use or stopped using health apps (8 response options with multiple answers).

Statistical Methods

Descriptive statistics were used to describe the participants’ demographic characteristics, health and medical services, health behaviors, internet use, and health app use status. We reclassified the response options into a single dichotomous outcome variable of the use of health apps. We categorized participants who are currently using health apps for over 1 month and those who have used health apps in the past for over 1 month as “users (current and previous),” and those who have never used health apps (including those who had downloaded health apps but only used them for less than 1 month) as “nonusers.” We also used the 2-tailed *t* test and chi-square test to check for differences in these factors according to gender.

We used the chi-square test to clarify the relationship between current or previous use of health apps and age, educational background, occupation, health checkups, health guidance, health behaviors, internet use duration, and number of devices used to access the internet for men and women separately. In addition, binary logistic regression analyses of data from men and women were conducted. The dependent variables were the use of health apps, and the independent variables were age, education, occupation, health checkups, health guidance, health behaviors (smoking, physical activity, drinking, sleep, appropriate weight, breakfast, and snacking), internet use duration, and number of devices used to access the internet. Odds ratios (ORs) and 95% CIs were calculated. Multivariable logistic regression analyses were conducted with all the independent variables included, and adjusted ORs and 95% CIs were calculated. Multicollinearity in multivariable logistic regression models was assessed using the variance inflation factor. Pseudo-R^2 values were also calculated to determine the proportion of variance in the dependent variable that can be explained by the independent variables. Data were analyzed using SPSS Statistics (version 22; IBM Corp).

Ethical Considerations

All the participants were asked to read the information privacy statement and informed consent form, and if they agreed to participate in the survey, they clicked an icon to begin answering the questionnaire. Completion of the internet survey was considered consent to participate in the study. Participants were not asked for any personally identifiable information. There are privacy agreements between Cross Marketing and its panelists. Participants who completed the survey were given points by Cross Marketing as a reward. Ethics approval was granted by the Research Ethics Committee of Nagoya City University, Graduate School of Nursing (ID22042-2).

Results

Participants' Characteristics

The average age of the 2200 participants was 44.7 (SD 13.5; men: mean 44.9, SD 13.6 and women: mean 44.6, SD 13.3) years (Table 1). Most of the participants lived in the Southern Kanto (n=779, 35.4%), Kinki (n=355, 16.1%), and Tokai (n=299, 13.6%) regions, which included metropolitan areas. Nearly half of the participants were unmarried (n=1126, 51.2%) and had an education at the university level or higher (n=1114, 50.6%). Many participants worked full-time (n=1284, 58.4%) and had fixed hours (n=1433, 65.1%). Almost all the participants had no disease under treatment (n=1818, 82.6%). Many participants had undergone health checkups (n=1370, 62.3%) and did not receive health guidance (n=1615, 73.4%). Regarding health behaviors, many participants did not smoke (n=1329, 60.4%), while a few participants did not snack (n=238, 10.8%). One-third of the participants reported using the internet for 60-119 minutes per day, and almost all used smartphones or PCs to do so (n=1972, 89.6% and n=1429, 65%, respectively).

There were statistically significant differences in the participants' demographic characteristics according to gender. More men than women were married, had a university education or higher, worked full-time, and engaged in flexible or discretionary work. The occupation among men was more likely to be security or agriculture, forestry, fishery, production, transportation, and construction, while among women, it was more likely to be office work. In terms of health behaviors, men were more likely to engage in regular physical activity and not snack, while women were more likely not to smoke or maintain an appropriate weight. Men were more likely to spend more time using the internet and to use more devices than women.

Table 1. Demographic characteristics, health-related factors, and internet use according to gender.

	All	Men	Women	*P* value
Age group (years), n (%)				—[a]
20-29	440 (20)	220 (20)	220 (20)	
30-39	440 (20)	220 (20)	220 (20)	
40-49	440 (20)	220 (20)	220 (20)	
50-59	440 (20)	220 (20)	220 (20)	
60-69	440 (20)	220 (20)	220 (20)	
Age (years), mean (SD)	44.7 (13.5)	44.9 (13.6)	44.6 (13.3)	.64[b]
Area of residence, n (%)				.22[c]
Hokkaido	110 (5)	46 (4.2)	64 (5.8)	
Tohoku	128 (5.8)	61 (5.5)	67 (6.1)	
Northern Kanto Koshin	111 (5)	52 (4.7)	59 (5.4)	
Southern Kanto	779 (35.4)	410 (37.3)	369 (33.5)	
Hokuriku	82 (3.7)	40 (3.6)	42 (3.8)	
Tokai	299 (13.6)	146 (13.3)	153 (13.9)	
Kinki	355 (16.1)	181 (16.5)	174 (15.8)	
Chugoku	117 (5.3)	53 (4.8)	64 (5.8)	
Shikoku	59 (2.7)	37 (3.4)	22 (2)	
Kyusyu Okinawa	160 (7.3)	74 (6.7)	86 (7.8)	
Marital status, n (%)				.04[c]
Unmarried	1126 (51.2)	538 (48.9)	588 (53.5)	
Married	1074 (48.8)	562 (51.1)	512 (46.5)	
Education, n (%)				<.001[c]
High school or below	610 (27.7)	295 (26.8)	315 (28.6)	
College or vocational college	476 (21.6)	141 (12.8)	335 (30.5)	
University or higher	1114 (50.6)	664 (60.4)	450 (40.9)	
Employment status, n (%)				<.001[c]
Full-time	1284 (58.4)	767 (69.7)	517 (47)	
Self-employed, business owner	179 (8.1)	110 (10)	69 (6.3)	
Temporary or contract worker, part-time	737 (33.5)	223 (20.3)	514 (46.7)	
Occupation, n (%)				<.001[c]
Management, research, professional	345 (15.7)	283 (25.7)	62 (5.6)	
Medical, education, welfare	337 (15.3)	93 (8.5)	244 (22.2)	
Office	470 (21.4)	135 (12.3)	335 (30.5)	
Sales, marketing, service	566 (25.7)	270 (24.5)	296 (26.9)	
Security, agriculture, forestry, fishery, production, transportation, construction	435 (19.8)	297 (27)	138 (12.5)	
Other	47 (2.1)	22 (2)	25 (2.3)	
Work pattern, n (%)				<.001[c]
Fixed hours	1433 (65.1)	706 (64.2)	727 (66.1)	
Shifts	147 (6.7)	62 (5.6)	85 (7.7)	
Flexible or discretionary work	401 (18.2)	241 (21.9)	160 (14.5)	

	All	Men	Women	*P* value
Other	219 (10)	91 (8.3)	128 (11.6)	
Diseases under treatment, n (%)				.29[c]
No	1818 (82.6)	899 (81.7)	919 (83.5)	
Yes	382 (17.4)	201 (18.3)	181 (16.5)	
Annual medical checkups, n (%)				.57[c]
Yes	1370 (62.3)	692 (62.9)	678 (61.6)	
No	830 (37.7)	408 (37.1)	422 (38.4)	
Health guidance, n (%)				<.001[c]
No	1615 (73.4)	738 (67.1)	877 (79.7)	
More than once	585 (26.6)	362 (32.9)	223 (20.3)	
Health behaviors, n (%)				
Do not smoke	1329 (60.4)	630 (57.3)	699 (63.5)	.003[c]
Engage in regular physical activity	796 (36.2)	451 (41)	345 (31.4)	<.001[c]
Drink alcohol in moderation or do not drink	1008 (45.8)	496 (45.1)	512 (46.5)	.52[c]
Get enough sleep each day	928 (42.2)	454 (41.3)	474 (43.1)	.41[c]
Maintain an appropriate weight	654 (29.7)	294 (26.7)	360 (32.7)	.002[c]
Eat breakfast daily	1093 (49.7)	540 (49.1)	553 (50.3)	.61[c]
Do not eat snacks	238 (10.8)	157 (14.3)	81 (7.4)	<.001[c]
Internet use duration (minutes), n (%)				<.001[c]
<60	528 (24)	221 (20.1)	307 (27.9)	
60-119	712 (32.4)	376 (34.2)	336 (30.5)	
120-179	470 (21.4)	245 (22.3)	225 (20.5)	
≥180	490 (22.3)	258 (23.5)	232 (21.1)	
Devices for internet use (multiple answers), n (%)				
Cell phone	189 (8.6)	100 (9.1)	89 (8.1)	.45[c]
Smartphone	1972 (89.6)	992 (90.2)	980 (89.1)	.40[c]
Tablet	505 (23)	281 (25.5)	224 (20.4)	.004[c]
PC	1429 (65)	805 (73.2)	624 (56.7)	<.001[c]
Wearable device	73 (3.3)	47 (4.3)	26 (2.4)	.02[c]
Home video game consoles with internet access	182 (8.3)	121 (11)	61 (5.5)	<.001[c]
Portable music player with internet access	60 (2.7)	34 (3.1)	26 (2.4)	.36[c]
Other	6 (0.3)	4 (0.4)	2 (0.2)	.69[c]
Number of devices used to access the internet, n (%)				<.001[c]
1	705 (32)	277 (25.2)	428 (38.9)	
2	987 (44.9)	509 (46.3)	478 (43.5)	
≥3	508 (23.1)	314 (28.5)	194 (17.6)	

[a]Not available.

[b]Two-tailed *t* test.

[c]Chi-square test.

Health App Use Status

Of the participants, 472 (21.5%) were current users of health apps for at least 1 month, 189 (8.6%) had used them for at least 1 month in the past, while 1539 (70%) had never used them (this last figure includes those who downloaded health apps but only used them for less than 1 month), with no significant differences according to gender (Table 2). Almost all current and previous health app users used features to track physical activity (n=487, 73.7%) and physical data, such as weight (n=256, 38.7%), sleep status (n=138, 20.9%), and food or diet (n=137, 20.7%). Among current and previous users, 84.4% (n=558) did not pay to use the health app, and 28.6% (n=103) used the health app in conjunction with wearable devices, with significant gender differences. Most users were motivated by their own desire for health management (n=504, 76.2%). Men were more likely to be motivated by referrals during health guidance and offers from their companies or health insurers than women. The reasons for not using or stopping to use health apps were that participants were unfamiliar with or uninterested in health apps and that the apps were cumbersome.

Table 2. Status of health app use according to gender.

	All	Men	Women	*P* value[a]
Use of health apps (n=2200), n (%)				.55
Currently using health apps for at least 1 month	472 (21.5)	239 (21.7)	233 (21.2)	
Used health apps previously for at least 1 month	189 (8.6)	101 (9.2)	88 (8)	
Never used health apps	1539 (70)	760 (69.1)	779 (70.8)	
Participants who reported current and previous use of health apps (n=661)				
What features of the apps do you (or did you) use? (Multiple answers), n (%)				
Tracking physical activity	487 (73.7)	244 (71.8)	243 (75.7)	.29
Tracking food or diet	137 (20.7)	64 (18.8)	73 (22.7)	.25
Tracking alcohol consumption	63 (9.5)	44 (12.9)	19 (5.9)	.002
Tracking sleep status	138 (20.9)	82 (24.1)	56 (17.4)	.04
Tracking physical data (eg, weight)	256 (38.7)	124 (36.5)	132 (41.1)	.23
Tracking medical data (eg, blood pressure and blood glucose level)	87 (13.2)	54 (15.9)	33 (10.3)	.04
Tracking medication	43 (6.5)	22 (6.5)	21 (6.5)	>.99
Automated messages, push notifications, reminders	31 (4.7)	17 (5)	14 (4.4)	.72
Searching for health information	61 (9.2)	29 (8.5)	32 (10)	.59
Earning points and coupons	118 (17.9)	59 (17.4)	59 (18.4)	.76
Sharing information with other users	21 (3.2)	11 (3.2)	10 (3.1)	.55
Competitions and contests	22 (3.3)	14 (4.1)	8 (2.5)	.28
Other	3 (0.5)	0 (0)	3 (0.9)	.11
Do you pay to use health apps? n (%)				.01
Yes	103 (15.6)	66 (19.4)	37 (11.5)	
No	558 (84.4)	274 (80.6)	284 (88.5)	
Do you use the apps in conjunction with wearable devices? n (%)				.001
Yes	189 (28.6)	116 (34.1)	73 (22.7)	
No	472 (71.4)	224 (65.9)	248 (77.3)	
What motivates you to use the app? (Multiple answers), n (%)				
I want to use it for self-health management	504 (76.2)	244 (71.8)	260 (81)	.01
It was referred to me during health guidance	89 (13.5)	66 (19.4)	23 (7.2)	<.001
My company or health insurer provided it	62 (9.4)	43 (12.6)	19 (5.9)	.003
It was recommended by a family member or friend	66 (10)	34 (10)	32 (10)	>.99
Other	27 (4.1)	12 (3.5)	15 (4.7)	.56
Participants who only reported previous use of health apps (n=1728)				
Reasons why they do not use health apps or why they stopped using health apps (Multiple answers), n (%)				
I am unfamiliar with health apps	743 (43)	370 (48.7)	373 (47.9)	>.99
I am not interested in health apps	511 (29.6)	283 (37.2)	228 (29.3)	.003
Health apps are cumbersome to use	496 (28.7)	226 (29.7)	270 (34.7)	.03
Health apps are difficult to use	80 (4.6)	46 (6.1)	34 (4.4)	.17
I am concerned about security, such as personal information leaks	121 (7)	61 (8)	60 (7.7)	.93
Health apps cost a lot	116 (6.7)	57 (7.5)	56 (7.2)	.92
My internet connection is unstable	113 (6.5)	52 (6.8)	61 (7.8)	.44
Other	24 (1.4)	12 (1.6)	12 (1.5)	>.99

[a]Chi-square test.

Relationships Between Current and Previous Health App Use and Demographic Characteristics, Health-Related Factors, and Internet Use

The results of the chi-square test indicated significant differences in terms of marital status (*P*=.001), education (*P*<.001), occupation (*P*=.001), health checkups (*P*<.001), health guidance (*P*<.001), physical activity (*P*<.001), maintaining an appropriate weight (*P*=.01), internet use duration (*P*=.03), and number of devices (*P*<.001) in health app use among men (Multimedia Appendix 1). There were significant differences in terms of age (*P*=.003), education (*P*<.001), health checkups (*P*<.001), health guidance (*P*<.001), physical activity (*P*<.001), drinking in moderation or not drinking (*P*=.04), getting enough sleep (*P*=.01), maintaining an appropriate weight (*P*<.001), eating breakfast (*P*=0.02), internet use duration (*P*=.01), and number of devices (*P*<.001) in health app use among women (Multimedia Appendix 2).

Adjusted ORs for current and previous health app use among men and women were shown in Figures 1 and 2, respectively. The ORs for current and previous health app use were significantly higher for participants who had health checkups (men: OR 1.53, 95% CI 1.12-2.11 and women: OR 1.51, 95% CI 1.10-2.07), received health guidance (men: OR 2.01, 95% CI 1.47-2.74 and women: OR 1.86, 95% CI 1.32-2.62), engaged in physical activity (men: OR 2.57, 95% CI 1.91-3.47 and women: OR 1.94, 95% CI 1.41-2.67), and used the internet for 120-179 minutes per day (men: OR 1.76, 95% CI 1.13-2.75 and women: OR 1.70, 95% CI 1.12-2.57) and significantly lower for those aged 50-59 years (men: OR 0.54, 95% CI 0.33-0.88 and women: OR 0.40, 95% CI 0.25-0.66) and 60-69 years (men: OR 0.37, 95% CI 0.22-0.62 and women: OR 0.47, 95% CI 0.28-0.77) for both men and women. Additionally, the ORs for current and previous health app use were significantly higher for men who were married (OR 1.69, 95% CI 1.23-2.33) while being significantly lower for men who had a job in the security, agriculture, forestry, fishing, production, or transportation industries (OR 0.62, 95% CI 0.41-0.95). For women, the OR for current and previous health app use was significantly higher for those with a university education or higher (OR 1.55, 95% CI 1.061-2.26), those who maintained an appropriate weight (OR 1.52, 95% CI 1.10-2.11), and those who used 3 or more devices (OR 2.13, 95% CI 1.41-3.23). Multimedia Appendix 3 shows the results of the chi-square test and logistic regression model for the data of combined men and women. The variance inflation factor of all the independent variables in the logistic regression models ranged from 1.07 to 1.38. Therefore, multicollinearity was not present in any of the models. The R^2 values were 0.186 for the model of men and 0.172 for the model of women.

Figure 1. Forest plot showing adjusted odds ratios for health app use among men (n=1100).

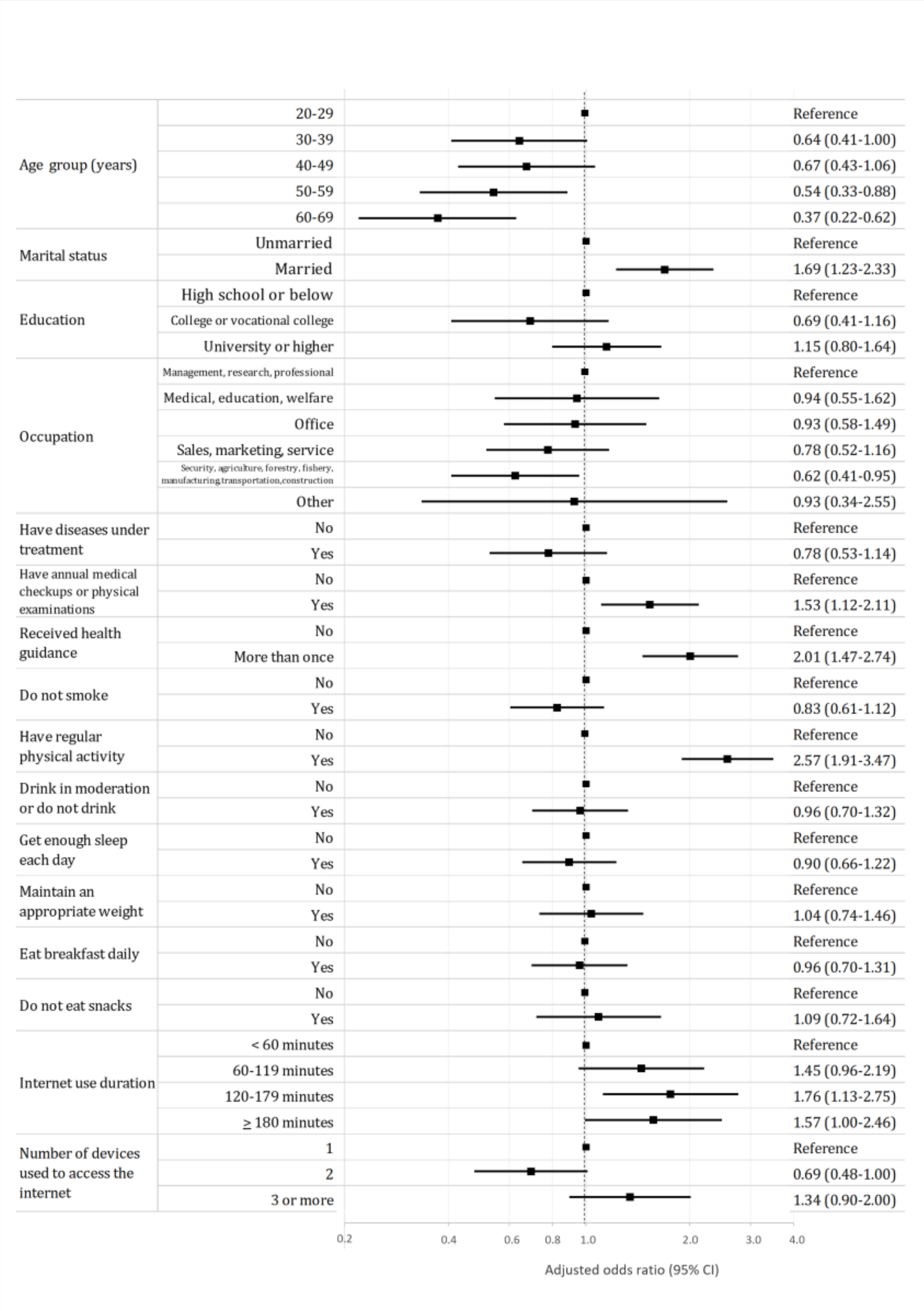
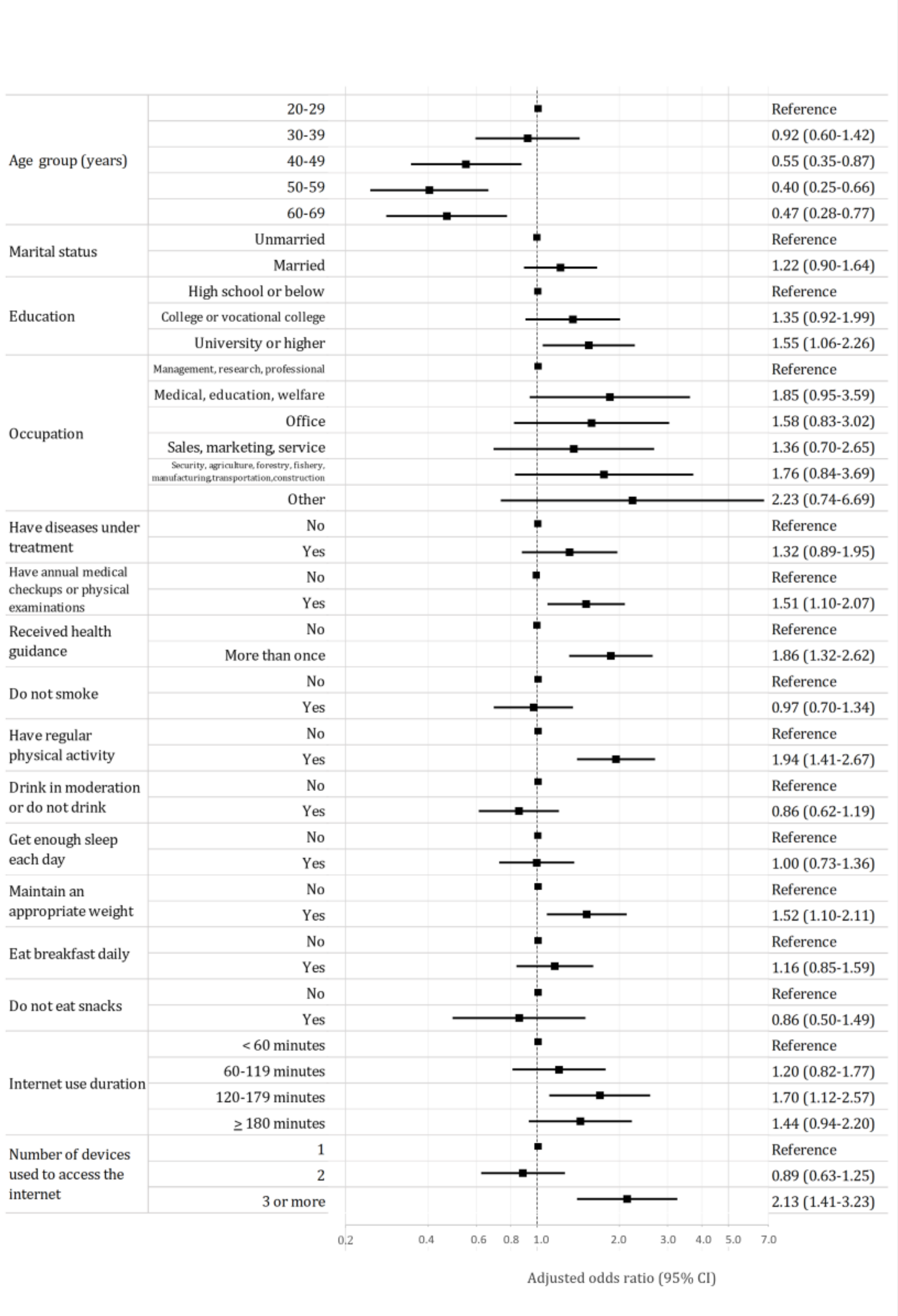


Figure 2. Forest plot showing adjusted odds ratios for health app use among women (n=1100).



Discussion

Principal Findings

In this study, we aimed to map and create an overview of health app use among Japanese workers and factors correlated with health app use by gender. Most current and previous health app users used features that record and track their physical activity and other health behaviors. Health app users were more likely to be younger, receive health guidance, engage in regular physical activity, and spend more time on the internet for both men and women. In terms of gender differences, male health app users were less likely to be married and less likely to work in the security, agriculture, forestry, fishing, production, or transportation industries, whereas female health app users were more likely to have a university education or higher, maintain an appropriate weight, and use 3 or more devices.

The proportions of current and previous health app users were 21.5% (n=472) and 8.6% (n=189), respectively. Ernsting et al [13] reported that the rate of health app use was 20.5% in a survey conducted in Germany in 2015, whereas Xie et al [11] reported that 38.4% of the respondents used health apps in a survey conducted in China between 2016 and 2017. In a survey conducted in Australia in 2020, Fewings et al [16] showed that 25.7% and 37.8% of the respondents were current and previous users of diet-related apps, respectively. These surveys were conducted between 2015 and 2022. Despite the increasing proportion of smartphone use among Japanese citizens [1], the prevalence of health app use among Japanese workers was low compared to the rates reported in these previous studies. However, our findings may have been influenced by the fact that we asked whether participants used health apps continuously for over a month.

The reasons for not using health apps and discontinuing their use were unfamiliarity with the apps and a lack of interest in them. Recording and tracking lifestyle and physical data, which app users in this study reported doing frequently, are useful for self-monitoring and are effective behavior change techniques [19]. Health apps have the potential to be effective tools for improving lifestyles, and their use can be increased by demonstrating their benefits and removing barriers to their use through easy-to-use interfaces.

App users in this study were likelier to have annual health checkups and receive health guidance than those who did not use the apps. A review of health checkups showed that general health checkups are associated with increased chronic disease recognition and treatment, risk factor control, preventive service uptake, and improved patient-reported outcomes [20]. Individuals who undergo a health checkup may reflect on their health status, intend to change their behavior, and consequently use health apps for behavior change. In our study, 13.5% (n=89) of the participants who used a health app reported being introduced to it during health guidance. Similarly, Hogan et al [15] showed that encouragement from health care team members to use an app was strongly associated with app use in a survey study. Individuals who trust the information provided by an app or website are more likely to use the app [16]. Encouragement from a health care provider to use a health app may increase trust in the app and promote its use. However, the features and contents of many apps lack a theoretical basis and safety [21]. Therefore, ensuring quality in the theoretical basis of the app and the safety of the information is crucial [2].

In this study, app users were more likely to engage in physical activity, which aligns with the results of previous studies [10,13,18]. In a qualitative study that explored the perspectives of young adults on apps related to behavior change, Dennison et al [22] showed that the appeal and usefulness of apps depend on whether a user is already motivated to change their lifestyle. Features such as self-monitoring tools and app reminders encourage motivated individuals to change their behavior; app use then helps cement behavior change. However, determining a causal relationship between app use and physical activity engagement based on the results of our cross-sectional survey is challenging, as app users may include individuals with greater motivation and interest in physical activity who are already engaging in physical activity [10,18].

We found no correlation between app use and diseases under treatment, similar to previous studies [14,18]. Individuals who use digital self-tracking tools are motivated to maintain their well-being rather than monitor or mediate medical problems or illnesses, and many monitor their progress in fitness or athletic training [23]. App users may include individuals willing to use the app to improve their health status and maintain their well-being by engaging in physical activity, regardless of whether they are undergoing treatment for any diseases. In this study, women using the app were also likely to maintain an appropriate weight. Weight is an important and easy-to-understand health parameter and is related to aesthetic satisfaction. Female users may intend to use the app to monitor their weight and maintain their well-being.

The app users in this study were likely to spend more time on the internet. Barriers to using health apps are that they are time-consuming and burdensome [22]. Those who spend a lot of time on the internet each day may spend some of that time using a health app and may not face these issues. Conversely, those who spend less than an hour on the internet may find it challenging to use a health app because they do not have enough time to spend on the internet or because they find it burdensome. Meanwhile, those who used the internet for more than 180 minutes did not use health apps significantly. Using the internet for a long time is an unhealthy habit, and these individuals may not be interested in using apps for disease prevention.

Female app users were more likely to use more than 3 devices to access the internet. Oshima et al [24] showed that smartphone ownership was associated with access to health care services, such as looking for health information and receiving necessary medical care. People must search for information about health apps to use suitable apps. In this study, women who used multiple devices may have had higher internet literacy, could easily access information about health apps, and found their favorite apps compared to those who used smartphones only. There was no association between app use and the number of devices among men. App users are uncomfortable with apps operating without user awareness or permission and want awareness of all features and control of all app settings [22]. Therefore, people must be skilled in using devices and technologies for apps. Men were more likely to use multiple devices than women, indicating that the former may be more skilled in using devices and technologies. Thus, the use of multiple devices does not appear to have affected the use of apps by men.

Strengths and Limitations

One strength of this study is that it included participants from diverse backgrounds nationwide. As people may use apps intermittently, we examined both current app use for more than 1 month and previous app use lasting longer than 1 month. We clarified the status of app use and the characteristics of those currently using or who have previously used health apps to improve health behaviors. We found common factors associated with app use in men and women as well as differences. Health care providers can use this information to assess whether the target population is suited to use health apps and how to motivate them to use health apps. For instance, during health checkups and health guidance, both men and women are easily motivated to use health apps for their health management, and health care providers can support them in using health apps. As for women, health care providers need to support them by asking how long they spent on the internet and how many devices they use, and then consider whether they can use health apps. Health care providers can enhance the effectiveness of health guidance by assessing whether a target population prefers to use an app based on their characteristics and by supporting them in using the app to improve their behavior.

Nonetheless, this study has several limitations. First, due to the cross-sectional design, we could not clarify a causal relationship between app use and the mentioned factors. For example, as mentioned earlier, we could not clarify whether participants who engaged in physical activity used the app to manage it or whether app users could manage their physical activity because of using the app. Second, this study involved a web-based survey conducted through a research company. Therefore, the possibility of sample bias cannot be excluded. The proportion of smartphone users among Japanese people aged between 20 and 49 years is more than 90% [1], and the participants in this study were not a specific population compared to the general population. However, people who participate in web-based surveys may be generally interested in internet technology, such as health apps. Therefore, the proportion of current and previous app users may be higher than the actual proportions in the general population.

Meanwhile, the rate of annual health checkups in this study was lower than that of the Japanese population (62.3% vs lowest in 20s: 68.4%; and highest in 50s: 77.4%) [25]. Participants may have been less interested in health behaviors or had difficulty accessing health care services, indicating that the proportion of current and previous app users may be higher in reality. Third, as we inquired regarding previous app use, recall bias is possible. Fourth, we could not clarify various situations of app use. We defined the duration of continuous app use as using an app for more than 1 month and could only clarify current or past app use. However, there are numerous health apps, and their health behavior targets vary. We expect that health app use patterns vary, with participants changing the apps they use in a short time and using multiple apps simultaneously. Future studies are needed to develop a more detailed understanding of app use for health guidance. Finally, although we focused on health-related factors (diseases under treatment, checkups, health guidance, and health behaviors) and internet use (use duration and number of devices), other related factors may influence app use. Future research should clarify the characteristics of individuals who will use a health app when it is offered to a target population and what outcomes will be achieved by using it. In addition, interviews are needed to better understand the kinds of mHealth apps that should be designed for Japanese workers.

Conclusions

This study clarified the status of current and previous use of health apps and related factors to prevent lifestyle-related diseases among Japanese workers. Compared to previous studies, the rate of app use among Japanese workers was low. The reasons for not using apps and discontinuing their use were unfamiliarity and a lack of interest in apps. Therefore, the number of app users can be increased by promoting effective apps for lifestyle improvement and providing easy-to-use and evidence-based apps. Health app users were more likely to be younger, receive health guidance, engage in regular physical activity, and spend more time on the internet for both men and women. Physical activity was strongly associated with app use. Although we could not clarify a causal relationship between app use and engagement in physical activity, app users may be interested and actively engage in physical activity. Health guidance was also strongly associated with app use. The introduction of an app by a health care provider may increase trust in it and promote its use. Health care providers can improve the effectiveness of health guidance by assessing whether a target population prefers to use an app based on their characteristics and by supporting them in using the app to improve their behavior.

Acknowledgments

This work was supported by the Japan Society for the Promotion of Science KAKENHI (grant JP19K11120).

Authors' Contributions

IO cleaned and summarized the dataset, analyzed the data, wrote the first draft, and finalized the manuscript. CC analyzed and interpreted the data and critically revised the draft. All authors contributed to the study design and provided substantial input to the successive drafts.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Relationships between current and previous use of health apps and demographic characteristics, health-related conditions, and internet use among men.

[DOCX File , 41 KB - humanfactors_v11i1e54673_app1.docx]

Multimedia Appendix 2

Relationships between current and previous use of health apps and demographic characteristics, health-related conditions, and internet use among women.

[DOCX File , 41 KB - humanfactors_v11i1e54673_app2.docx]

Multimedia Appendix 3

Relationships between current and previous use of health apps and demographic characteristics, health-related conditions, and internet use among participants.

[DOCX File , 47 KB - humanfactors_v11i1e54673_app3.docx]

References

1. Communications usage trend survey in 2021. Ministry of Internal Affairs and Communications. 2022. URL: https://www.soumu.go.jp/johotsusintokei/tsusin_riyou/data/eng_tsusin_riyou02_2021.pdf [accessed 2022-05-27]
2. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016;13:35 [FREE Full text] [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
3. Lee JA, Choi M, Lee SA, Jiang N. Effective behavioral intervention strategies using mobile health applications for chronic disease management: a systematic review. BMC Med Inform Decis Mak 2018;18(1):12 [FREE Full text] [doi: 10.1186/s12911-018-0591-0] [Medline: 29458358]
4. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
5. Palmer M, Sutherland J, Barnard S, Wynne A, Rezel E, Doel A, et al. The effectiveness of smoking cessation, physical activity/diet and alcohol reduction interventions delivered by mobile phones for the prevention of non-communicable diseases: a systematic review of randomised controlled trials. PLoS One 2018;13(1):e0189801 [FREE Full text] [doi: 10.1371/journal.pone.0189801] [Medline: 29304148]
6. The national health and nutrition survey in Japan. Ministry of Health, Labour and Welfare, Health Service Bureau, Office of Nutrition Health Service Division. 2019. URL: https://www.mhlw.go.jp/content/001066903.pdf [accessed 2020-12-25]
7. OECD reviews of public health: Japan: a healthier tomorrow. Organisation for Economic Co-operation and Development. 2019. URL: https://tinyurl.com/hjdjfbww [accessed 2024-10-03]
8. Kondo M, Okitsu T, Waki K, Yamauchi T, Nangaku M, Ohe K. Effect of information and communication technology-based self-management system DialBeticsLite on treating abdominal obesity in the specific health guidance in Japan: randomized controlled trial. JMIR Form Res 2022;6(3):e33852 [FREE Full text] [doi: 10.2196/33852] [Medline: 35323122]
9. Sakane N, Suganuma A, Domichi M, Sukino S, Abe K, Fujisaki A, et al. The effect of a mHealth app (KENPO-app) for specific health guidance on weight changes in adults with obesity and hypertension: pilot randomized controlled trial. JMIR Mhealth Uhealth 2023;11:e43236 [FREE Full text] [doi: 10.2196/43236] [Medline: 37043287]
10. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017;19(4):e125. [doi: 10.2196/jmir.5604] [Medline: 28428170]
11. Xie Z, Nacioglu A, Or C. Prevalence, demographic correlates, and perceived impacts of mobile health app use amongst Chinese adults: cross-sectional survey study. JMIR Mhealth Uhealth 2018;6(4):e103 [FREE Full text] [doi: 10.2196/mhealth.9002] [Medline: 29699971]
12. Buss VH, Varnfield M, Harris M, Barr M. Mobile health use by older individuals at risk of cardiovascular disease and type 2 diabetes mellitus in an Australian cohort: cross-sectional survey study. JMIR Mhealth Uhealth 2022;10(9):e37343 [FREE Full text] [doi: 10.2196/37343] [Medline: 36069764]
13. Ernsting C, Dombrowski SU, Oedekoven M, Sullivan JLO, Kanzler M, Kuhlmey A, et al. Using smartphones and health apps to change and manage health behaviors: a population-based survey. J Med Internet Res 2017;19(4):e101 [FREE Full text] [doi: 10.2196/jmir.6838] [Medline: 28381394]
14. Makowsky MJ, Jones CA, Davachi S. Prevalence and predictors of health-related internet and digital device use in a sample of South Asian adults in Edmonton, Alberta, Canada: results from a 2014 community-based survey. JMIR Public Health Surveill 2021;7(1):e20671 [FREE Full text] [doi: 10.2196/20671] [Medline: 33416506]
15. Hogan TP, Etingen B, Lipschitz JM, Shimada SL, McMahon N, Bolivar D, et al. Factors associated with self-reported use of web and mobile health apps among US military veterans: cross-sectional survey. JMIR Mhealth Uhealth 2022;10(12):e41767 [FREE Full text] [doi: 10.2196/41767] [Medline: 36583935]
16. Fewings A, Vandelanotte C, Irwin C, Ting C, Williams E, Khalesi S. The use and acceptability of diet-related apps and websites in Australia: cross-sectional study. Digit Health 2022;8:20552076221139091 [FREE Full text] [doi: 10.1177/20552076221139091] [Medline: 36578516]

17. Langford AT, Solid CA, Scott E, Lad M, Maayan E, Williams SK, et al. Mobile phone ownership, health apps, and tablet use in US adults with a self-reported history of hypertension: cross-sectional study. JMIR Mhealth Uhealth 2019;7(1):e12228 [FREE Full text] [doi: 10.2196/12228] [Medline: 31344667]
18. Robbins R, Krebs P, Jagannathan R, Jean-Louis G, Duncan DT. Health app use among US mobile phone users: analysis of trends by chronic disease status. JMIR Mhealth Uhealth 2017;5(12):e197 [FREE Full text] [doi: 10.2196/mhealth.7832] [Medline: 29258981]
19. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
20. Liss DT, Uchida T, Wilkes CL, Radakrishnan A, Linder JA. General health checks in adult primary care: a review. JAMA 2021;325(22):2294-2306. [doi: 10.1001/jama.2021.6524] [Medline: 34100866]
21. Alessa T, Hawley MS, Hock ES, de Witte L. Smartphone apps to support self-management of hypertension: review and content analysis. JMIR Mhealth Uhealth 2019;7(5):e13645 [FREE Full text] [doi: 10.2196/13645] [Medline: 31140434]
22. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
23. Paré G, Leaver C, Bourget C. Diffusion of the digital health self-tracking movement in Canada: results of a national survey. J Med Internet Res 2018;20(5):e177 [FREE Full text] [doi: 10.2196/jmir.9388] [Medline: 29720359]
24. Oshima SM, Tait SD, Thomas SM, Fayanju OM, Ingraham K, Barrett NJ, et al. Association of smartphone ownership and internet use with markers of health literacy and access: cross-sectional survey study of perspectives from project PLACE (population level approaches to cancer elimination). J Med Internet Res 2021;23(6):e24947 [FREE Full text] [doi: 10.2196/24947] [Medline: 34106076]
25. Summary report of comprehensive survey of living conditions. Ministry of Health, Labour and Welfare. 2019. URL: https://tinyurl.com/bdcvbtkx [accessed 2024-09-25]

Abbreviations

mHealth: mobile health
OR: odds ratio

Edited by E Borycki; submitted 18.11.23; peer-reviewed by X Yan, J Zlamal; comments to author 01.04.24; revised version received 25.05.24; accepted 23.09.24; published 25.10.24.

Please cite as:
Ozaki I, Nishijima M, Shibata E, Zako Y, Chiang C
Factors Related to mHealth App Use Among Japanese Workers: Cross-Sectional Survey
JMIR Hum Factors 2024;11:e54673
URL: https://humanfactors.jmir.org/2024/1/e54673
doi:10.2196/54673
PMID:

Original Paper

Using the Person-Based Approach to Co-Create and Optimize an App-Based Intervention to Support Better Sleep for Adolescents in the United Kingdom: Mixed Methods Study

Sarah E Bennett[1], PhD; Milly H Johnston[2], BSc; Georgia Treneman-Evans[2], MSc; James Denison-Day[3], PhD; Anthony Duffy[4], MSc; Amberly Brigden[5], PhD; Paula Kuberka[2], MSc; Nicholas Christoforou[2], LLB; Lee Ritterband[6], PhD; Jewel Koh[1], MSc; Robert Meadows[7], PhD; Doaa Alamoudi[8], PhD; Ian Nabney[8], PhD; Lucy Yardley[1,3], PhD

[1]School of Psychological Science, University of Bristol, Bristol, United Kingdom
[2]Bristol Medical School, University of Bristol, Bristol, United Kingdom
[3]School of Psychology, University of Southampton, Southampton, United Kingdom
[4]Digital Health Circle Lab, School of Interactive Arts & Technology, Simon Fraser University, British Columbia, BC, Canada
[5]School of Engineering, Mathematics and Technology, University of Bristol,, Bristol, United Kingdom
[6]Psychiatry and Neurobehavioral Sciences, University of Virginia Health System, Charlottesville, VA, United States
[7]Department of Sociology, University of Surrey, Guildford, United Kingdom
[8]Department of Computer Science, University of Bristol, Bristol, United Kingdom

Corresponding Author:
Sarah E Bennett, PhD
School of Psychological Science
University of Bristol
12A Priory Rd
Bristol, BS8 1TU
United Kingdom
Phone: 44 07590334234
Email: sarah.bennett@bristol.ac.uk

Abstract

Background: Poor sleep is a common problem in adolescents aged 14 to 18 years. Difficulties with sleep have been found to have a bidirectional link to mental health problems.

Objective: This new research sought to involve young people in the co-creation of a new app, particularly those from underserved communities. The Sleep Solved app uses science-based advice to improve sleep-related behaviors and well-being. The app was developed using the person-based approach, underpinned by the social cognitive theory and the social-ecological model of sleep health.

Methods: Young people (aged 14-18 y) were recruited from across the United Kingdom to contribute to patient and public involvement (PPI) activities. In partnership with our peer researcher (MHJ), we used a multitude of methods to engage with PPI contributors, including web-based workshops, surveys, think-aloud interviews, focus groups, and app beta testing.

Results: A total of 85 young people provided PPI feedback: 54 (64%) young women, 27 (32%) young men, 2 (2%) genderfluid people, 1 (1%) nonbinary person, and 1 (1%) who reported “prefer not to say.” Their levels of deprivation ranged from among the 40% most deprived to the 20% least deprived areas. Most had self-identified sleep problems, ranging from 2 to 3 times per week to >4 times per week. Attitudes toward the app were positive, with praise for its usability and use of science-based yet accessible information. Think-aloud interviews and a focus group identified a range of elements that may influence the use of the app, including the need to pay attention to language choices and readability. User experiences in the form of narrated audio clips were used to normalize sleep problems and provide examples of how the app had helped these users.

Conclusions: Young people were interested in using an app to better support their sleep and mental health. The app was co-created with strong links to theory- and evidence-based sleep hygiene behaviors. Future work to establish the effectiveness of the intervention, perhaps in a randomized controlled trial, would provide support for potential UK-wide rollout.

(JMIR Hum Factors 2024;11:e63341) doi:10.2196/63341

KEYWORDS

behavior change; digital intervention; insomnia; depression; anxiety; sleep; qualitative research; mobile phone

Introduction

Background

Poor sleep is a common issue among adolescents, with significant effects on school performance [1], mood, and behavior [2,3]. Since the 1990s, the amount of sleep per night that adolescents self-report has decreased significantly [4,5]. Poor sleep can impact adolescents' mental health [6]; reduced sleep duration has been associated with loneliness, low mood [7], and depression [8] and can affect adolescents' physical health over time [9].

Adolescence is a period of significant biological and psychosocial changes. These changes include a shift in the sleep-wake cycle, a change in the circadian rhythm [10], and a gradual behavioral shift toward later bedtimes [11-13]. The shift toward later bedtimes is more noticeable on school nights compared to weekends, with later sleep times making it more difficult for adolescents to wake up when they need to [12,13]. Rather than later bedtimes, adolescents have been shown to sleep better when sticking to a regular bedtime routine with consistent sleep and wake times [14].

Sleep hygiene, which can be defined as behavioral practices that promote good sleep, plays a key role in adolescents' sleep quality. Differences in sleep hygiene behavior may partly mediate the relationship between sociodemographic factors and sleep health [15]. Across the wealth of sleep hygiene literature, 3 sleep hygiene behaviors are reported as having a key impact on sleep: avoiding behavioral and cognitive arousal before bedtime, spending less time in bed, and setting a regular wake time [16].

Diet may also play a role in adolescents' sleep [17,18]. The UK National Diet and Nutrition Survey from 2016 to 2019 found that those aged 11 to 18 years had the highest consumption of sugar-sweetened soft drinks and mean saturated fat intake [19]. Diets high in fat and sugar have been associated with shorter sleep duration [20] and significantly higher prevalence of sleep disturbance in adolescents [21] compared to a healthier diet. A high caffeine intake has been associated with sleep problems, including later bedtimes, shorter sleep duration [22,23], difficulty sleeping, and morning tiredness [24]. Dietary influences on sleep may disproportionately affect adolescents from lower-income households. A descriptive analysis of UK survey data from 2005, 2009, and 2014 found that adolescents aged 11 to 15 years from families with lower levels of affluence reported fewer healthy eating behaviors, higher sugary drink consumption, and lower intake of fruits and vegetables [25].

Previous research has indicated that adolescents from more socioeconomically deprived neighborhoods typically experience poorer sleep quality, a shorter sleep duration, and increased daytime sleepiness [26]. They also have lower sleep efficiency, spending more time awake at night [27]. Possible causes may include more chaotic, noisy [28,29], or crowded living environments with lower neighborhood safety, higher levels of crime, or concerns about violence [27,30] when compared to adolescents from higher-income homes [9,31]. Therefore, adolescents from areas of higher socioeconomic deprivation are more likely to experience poorer-quality sleep.

There is a bidirectional relationship between insomnia and mental health difficulties [32-35]. Anxiety may perpetuate and worsen insomnia, with those who experience emotional dysregulation in childhood being more likely to experience anxiety disorders in adolescence [36-38]. As poor sleep during adolescence may be a precursor to mental health problems, targeting poor sleep behaviors early may prevent the development of mental health difficulties, such as anxiety or depression, at a later stage [32].

However, engaging with adolescents at an early stage can be a challenge. There is evidence suggesting that adolescents are less likely to engage with mental health support services due to perceived stigma, embarrassment [39], or a preference for less formal advice and support [40-44]. Instead, smartphones can be a key route to engaging with adolescents; a range of apps to support adolescents with their sleep and well-being have been designed. Delivering Online Zzz's With Empirical Support (DOZE), a 4-week intervention, was developed in co-creation with Canadian adolescents [45]. This intervention includes the ability to track sleep time, average time to fall asleep, and nighttime awakening. A "tips" section provides education to users on how best to make changes to their sleep, such as winding down before bed [45]. Similarly, Sleep Ninja was developed in partnership with Australian adolescents through a series of focus groups focusing on the app design [46,47]. The app features 6 educational sessions, a sleep-tracking function, sleep tips, and reminders to wind down in the evening [46]. Finally, the Sleepio app uses cognitive behavioral therapy for insomnia (CBTi) in 6 web-based, 20-minute sessions [48]. Sleepio has shown promising results in a small population of adolescents (N=39) with mental health problems [48]. However, no apps to date have been co-created with adolescents from the United Kingdom, and none have been co-created with those from disadvantaged backgrounds.

The apps described previously are all intended to provide in-depth support through extended use (4-6 sessions). However, there is some evidence that, for many adults and adolescents, a single session of sleep hygiene advice may be sufficient to improve sleep [49]. Our study aim was to design a highly accessible app suitable for adolescents with lower literacy levels that did not require extensive engagement for users to benefit from sleep hygiene advice and support. The purpose of this app was to serve as an accessible first step of a stepped program of support [50], which would guide adolescents with more persistent and serious sleep problems toward the in-depth support provided by existing apps (specifically, Sleep Healthy Using the Internet [SHUTi] [51,52]).

The Goal of This Study

This paper details the co-creation and optimization of a novel, brief, stepped-care intervention to support adolescents with sleep problems (Table 1) from its initial design phases to co-creation with adolescents from a range of collaborative youth organizations, charities, and educational partners. After outlining the development process, the feasibility of the app and feedback in light of real-world testing are explored. The process was iterative, and contributor-suggested amendments were used to make changes to the intervention as the co-creation process progressed.

Table 1. The 2 stages of Sleep Well.

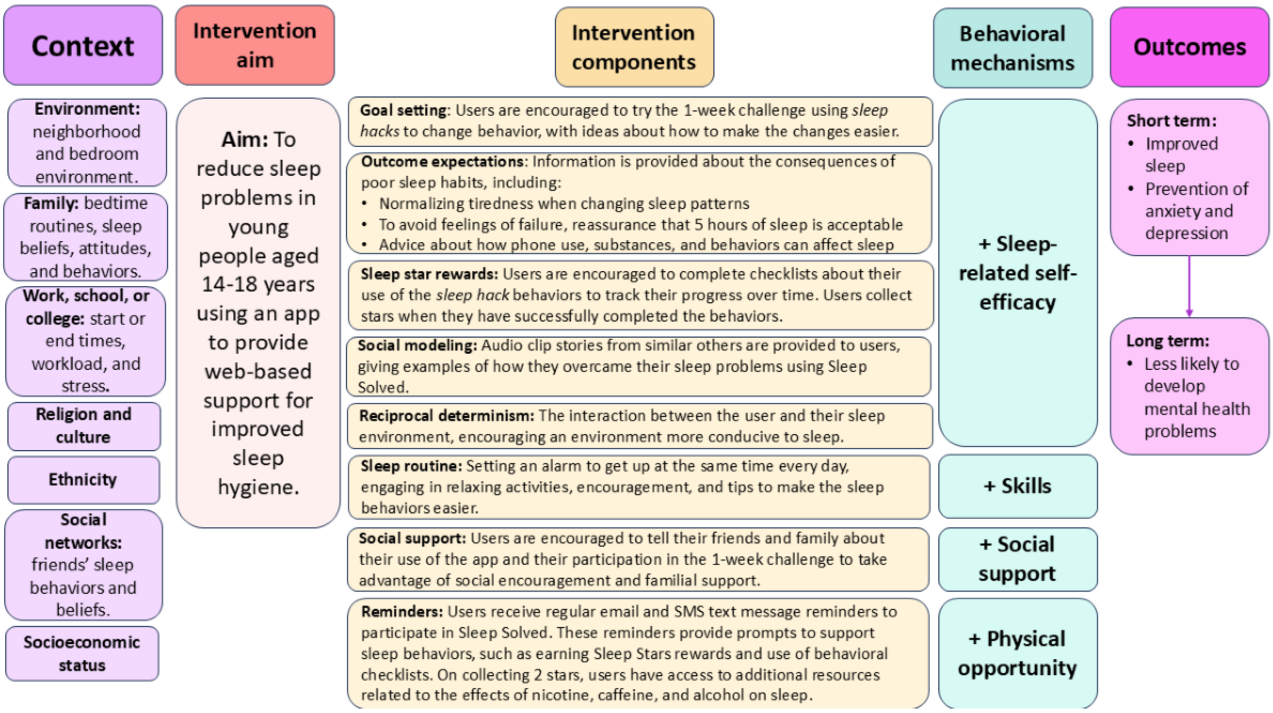
Intervention stage and app	Details	Time when it was offered
Stage 1		
Sleep Solved	A web-based application co-created with adolescents to provide brief (six 5- to 10-min sections) and accessible advice and behavior change support to help improve their sleep. The short Sleep Solved intervention may be sufficient as an intervention for less serious sleep problems.	At sign-up (for iOS and Android users)
Phone Downtime	An Android app created by a University of Bristol PhD student before this study. Phone Downtime allows users to set their desired wake and sleep times to support their sleep schedule while using the Sleep Solved app.	At sign-up (for Android users only)
Stage 2		
SHUTi[a]	A six 20-min–session program to treat insomnia based on internet-delivered cognitive behavioral therapy was adapted from a version of SHUTi shown to be effective for adults in the United States [51,52]. The content was coadapted in this study to better suit adolescents in the United Kingdom.	Bite Back [53] and SHUTi [51] are offered if adolescents score highly on measures of anxiety and depression (RCADS[b]) or insomnia (Insomnia Severity Index) after 6 weeks of access to Sleep Solved.
Bite Back	Developed by researchers in Australia, Bite Back [53] is a web-based 6-session positive psychology program designed to improve well-being and resilience in adolescents.	Bite Back [53] and SHUTi [51] are offered if adolescents score highly on measures of anxiety and depression (RCADS) or insomnia (Insomnia Severity Index) after 6 weeks of access to Sleep Solved.

[a]SHUTi: Sleep Healthy Using the Internet.

[b]RCADS: Revised Child Anxiety and Depression Scale.

The person-based approach was used in the development of the Sleep Solved intervention to provide adolescents with a theoretically informed app that was engaging, clear, and appropriate for their needs [54,55]. This involved initially drawing on clinical theory and evidence to specify the core required intervention elements to support behaviors that would help adolescents sleep better, summarized in a logic model (Figure 1). The social cognitive theory by Bandura [56] was used to underpin the behavioral elements of Sleep Solved, whereas the social-ecological model of sleep health [57] was chosen to represent the multifactorial nature of various moderators on sleep and sleep-related interventions at the individual, social, and societal levels, such as the influence of social networks or family [57]. The logic model features 8 intervention components that make up the key content of the app. How these components relate to social cognitive theory, moderators from the social-ecological model of sleep, and the underpinning behavioral mechanisms is outlined in the following paragraphs.

Figure 1. The Sleep Solved logic model.



We drew on theory and evidence to create guiding principles (Table 2) that were intended to encourage engagement with Sleep Solved. The coproduction work described in this section was then used to co-design the app content and format. Details regarding the virtual co-design of the design aspects of the intervention have been published separately (A Duffy, unpublished data, May 2022).

Table 2. Guiding principles for Sleep Solved.

User characteristics	Design objective	Features to address the objective
• Very low tolerance for engaging with advice that is obvious, dull, not credible, or irrelevant to their sleep context • Low patience for engaging with apps that are dull, uninteresting, unfriendly, patronizing, or misleading or have a confusing user experience and intrusive advertising	• To ensure that all advice is immediately engaging and viewed as interesting, trustworthy, and relevant • To ensure that the app itself is designed to be engaging, friendly, entertaining, and motivating and hold the user's interest and attention, with no advertising	• Present novel content in the most concise and engaging format possible • Simple, short sentences • Good functionality; consistent navigation, theme, gestural design, and user flow • Optimized for mobile use • Provide evidence of credibility and trustworthiness (eg, team expertise and scientific validity of the advice) • Focus on core principles of sleep relevant to a wide range of situations • Personalized using relatable stories from other adolescents with sleep problems • Gamification—users can collect stars to encourage engagement with the app
• Great diversity in needs and preferences for how advice is presented, in accordance with users' social identity and levels of (health) literacy	• To ensure that the advice is appropriate for people with different identities, social contexts, and learning preferences; variety of text, images, and audio clips to reflect different learning styles • Simple language, short sentences, and use of images or icons may be more engaging • Consistent navigation structure; similar graphics, gestural designs, and theme throughout the app for consistency and ease of use • Sleep tips throughout	• Ensure that role models and stories are relevant to diverse contexts and identities • Present advice in a variety of accessible formats • Provide links to a wider range of additional resources or formats (eg, relaxing music, mindfulness, and in-depth resources) • Include a summary of useful hints and tips to improve sleep
• The importance of choice or free will as opposed to paternalistic advice	• Choice architecture—framing advice as an option; emphasizing the benefits of the desired behavior vs the consequences of the undesired behavior	• Avoid telling adolescents what they should do (eg, avoid giving a prescribed amount of sleep that adolescents should be getting) to reduce feelings of anxiety and inadequacy in the target population (aged 14-18 years) • Present facts, actions, and consequences • Frequently asked question pop-ups allowing for more information if required
• A preference for web-, app-, or technology-based interactions over face-to-face	• App use involves interactions with mobile technology rather than face-to-face interactions	• Adolescents using apps may prefer app- and technology-based interactions

Advising adolescents to improve their sleep using a smartphone device known to disrupt sleep may seem counterintuitive, particularly as the nighttime use of smartphones and social media has been linked to disturbed sleep and poor academic achievement [58-61]. However, apps are a key method for engaging adolescents; 81% of adolescents use the internet to seek health-related information [62], whereas digital health interventions to support people with their sleep are flourishing [63]. To avoid too much phone exposure too close to bedtime, users were encouraged to use the app to decide on their preferred sleep behaviors and environment ahead of time, earlier in the evening. Later sections of Sleep Solved remind users that looking at a bright phone screen too close to bedtime could affect their ability to get to sleep [64].

Methods

Phase 1: Initial Intervention Development

Sleep Solved Ideas and Concept Generation

To explore adolescents' views and experiences of sleep and well-being, a web-based focus group was conducted with the Bristol Young People's Advisory Group (YPAG). Patient and public involvement (PPI) contributors were presented with initial ideas for key intervention concepts (eg, a focus on how "science-based" research could help them sleep better and why spending less time in bed could help). Contributors were asked what they thought of these concepts and were invited to share additional ideas for supporting their sleep and well-being.

Sleep Solved Ideas: Name and Design

To involve PPI contributors who preferred to provide written feedback, a Qualtrics survey (Qualtrics) was used to elicit feedback regarding the look and feel of Sleep Solved. PPI contributors were asked for feedback via closed- and open-ended questions and to share ideas about the app's name, logo, color palette, and theme as well as the study's recruitment materials:

- Potential app names: preferred name from 7 options—*Sleep Solved, Snooze It, Science2Sleep, Slumber Master, Rest Pro, Sleep Tracker,* and *Awake to Sleep*
- Color palettes: three options—(1) crimson red, roundel blue, international orange, and black; (2) black, opaline green, deep cream, and pale roundel blue; and (3) Oxford blue, light orange, deep cream, and ruby
- Logos: from a range of 11 possible logo designs
- Design themes: a total of 6 options from other youth-oriented media

New Theme Designs for Sleep Solved

On the basis of the feedback from the first Qualtrics survey, our design partners at PIP Creative designed 2 tailored themes to give the Sleep Solved app a consistent look and feel. PPI contributors were invited via a second Qualtrics survey to rank how much they liked the themes (0-100) and share their thoughts and suggestions regarding the visual design.

New Logo Designs for Sleep Solved

On the basis of the feedback from the first Qualtrics survey, 4 viable app logos were developed by PIP Creative. These logos were presented to PPI contributors via a third Qualtrics survey, and contributors were requested to rank their preferred logos. Comments about the logos were encouraged using open-ended questions.

Sleep Well Design Feedback

To determine the preferred timing of offering the 3 components of Sleep Well (Phone Downtime, Sleep Solved, and SHUTi [51]), 3 different pathway options were presented to adolescents in a fourth Qualtrics survey. For each pathway, PPI contributors were invited to share how likely they were to take part in the Sleep Well study for 6 weeks, giving a rating from 0 ("not at all") to 100 ("a lot") and their reasons why.

PPI Contributors

Adolescents aged 14 to 18 years were recruited to co-design the intervention through partnerships with national educational groups and charities: the Association of Colleges, the E-ACT multiacademy trust, the McPin Foundation, the Bristol YPAG, and Off The Record Bristol. We invited contributors aged between 14 and 18 years who attend secondary school or college and had an Android smartphone.

PPI contributors were recruited from across the United Kingdom. A poster advertising the research was shared within schools and colleges. For recruitment from charities, emails were sent to adolescents within the target age range inviting them to contribute to PPI activities to improve a new sleep app.

PPI contributors could express their interest by scanning a QR code from the poster or following a link within each email. Prospective PPI contributors answered demographic Qualtrics survey questions (age, gender, and ethnicity), how often they had sleep problems, and their preferred method of engagement (one-to-one meeting in person or web-based discussion, web-based group discussions, or written comments). The levels of deprivation were estimated using contributors' postcodes against the 2019 English indices of deprivation data. To ensure that adolescents from ethnically diverse and disadvantaged backgrounds were well represented, we prioritized input from volunteers from ethnic minority groups and more disadvantaged areas [65]. Over the course of a year, PPI contributors engaged in a variety of coproduction activities (Table 2), including identifying behaviors for the app to address; choosing the preferred name, appearance, and other key design elements; providing feedback on how prototype versions of the app could be improved; and trying out the app in real-life use and providing further feedback. PPI contributors shared their views according to their preferred method of contribution (Table 3). PPI contributions and activities are listed in Multimedia Appendix 1.

Table 3. Potential patient and public involvement contributors—preferred method of contribution (N=356).

Expressions of interest (January 2022-March 2022)—"Please select any of the ways you would be happy to talk to us"	Contributors, n (%)
Only written comments	219 (61.5)
Written comments	292 (82)
Web-based group discussion with other adolescents	65 (18.3)
One-to-one web-based meeting (eg, via Zoom)	63 (17.7)

Phase 2: Intervention Optimization

Web-Based Think-Aloud Interviews

We invited contributors to provide detailed feedback on every aspect of the prototype app as it was developed using think-aloud interviews as an opportunity to identify barriers to engagement and feasibility and suggest improvements. PPI contributors met with a researcher via Microsoft Teams (Microsoft Corp) or Zoom (Zoom Video Communications) and provided verbal consent to take part in a one-to-one think-aloud interview. Contributors were given a web address and instructions on how to view Sleep Solved as it would appear on a mobile device. PPI contributors were encouraged to verbalize their immediate thoughts and feelings when interacting with Sleep Solved to provide as much information as possible about their impressions of the app and overall user experience [54]. Contributors were prompted for further information if required—such as "What are you thinking now?" or "What made you choose that option?"—and were encouraged to suggest ways of improving it (see Multimedia Appendix 2 for the topic guide). Interviews were audio recorded and transcribed verbatim. PPI contributors received a £22 (US $28.56) digital voucher for each hour of participation to thank them for their input after taking part in a think-aloud interview. All contributors who provided feedback via a Qualtrics survey were entered into a prize draw to receive a £50 (US $64.92) digital voucher. A versatile multiretailer voucher was carefully chosen by the research team in light of the cost-of-living crisis so that it could be used by contributors to purchase essential items such as food and toiletries. From PPI activity workshops 1 and 2, a total of 4 participants (n=2, 50% male; n=1, 25% female; and n=1, 25% genderfluid) also took part in the think-aloud sessions.

All feedback and suggestions from PPI contributors' think-aloud interviews were collated in a table of changes (Multimedia Appendix 3). This was used directly as a basis for informing intervention changes rather than being subjected to analysis methods. Simple changes (such as clarifying text or rewriting a section that a participant found difficult to understand) would be made directly. Substantial changes were only made if they met the following criteria: (1) the proposed change was suggested by several PPI contributors, (2) the proposed change was likely to have an impact on users' behavior, and (3) the proposed change aligned with the guiding principles for the intervention [54].

Once changes had been made, we obtained additional feedback from participants regarding the updated version of Sleep Solved in further think-aloud interviews [66,67]. The table of changes process continued until no new important suggested changes were raised in feedback from think-aloud interviews [66].

Sleep Solved Real-Life Testing: Think-Aloud Interviews

A selection of our PPI contributors was invited to trial the app for a period of 1 week after taking part in an initial think-aloud interview. After their trial, they were invited for a second think-aloud interview. PPI contributors were encouraged to follow the advice provided and keep a sleep diary for their personal reference during the follow-up interview (Multimedia Appendix 4).

Sleep Solved Real-Life Testing: Web-Based Workshops

The next stage of optimizing Sleep Solved started as soon as the "Must Have" changes were completed. Workshop attendees were contributors from the McPin Foundation's Young People's Network. All contributors were invited to use Sleep Solved for 1 week. Contributors with Android smartphones were also invited to use Phone Downtime alongside Sleep Solved. All were given a sleep diary checklist to complete and provide feedback about their experiences (Multimedia Appendix 5). A group workshop covered contributors' experiences of engaging with every aspect of the app. Contributors were asked about elements of Sleep Solved that they liked or found useful and those that they did not like or found unhelpful or that did not work as expected. Questions explored contributors' own sleep experiences and factors that helped them sleep well. The workshops were documented by group facilitators, with any suggested changes noted in the table of changes. A number of users created and recorded personal accounts of how they had benefited for other app users to listen to.

Ethical Considerations

Ethics approval for the think-aloud interviews was granted by the University of Bristol School of Psychological Science Student Research Ethics Committee (ethics approval code 13084). PPI contributors provided verbal informed consent to take part in a one-to-one think-aloud interview. All data were anonymized, and all identifying information was removed. In compensation for their time, PPI contributors were offered a multiretailer digital voucher of £22 (US $28.56) for each hour of participation after taking part in a think-aloud interview. All contributors who provided feedback via a Qualtrics survey were entered into a prize draw to receive a £50 (US $64.92) digital voucher." (I've included the ethics approval statement that was there previously.)

Results

Overview

An overview of the Sleep Solved content is provided in Textbox 1.

Textbox 1. An overview of the Sleep Solved intervention content.

How can Sleep Solved help me sleep better?

- Introduction to Sleep Solved: types of sleep problems, how sleeping better will help the user, and who made the app
- Explanation of the Phone Downtime app and link to download in Google Play

How can less time in bed help me sleep better?

- 3 kinds of sleep (deep, dreaming, and light)
- 5 to 6 hours and deep and dreaming sleep is what the brain needs. A lot of time in bed leads to additional light sleep—which the brain does not need as much.
- If the user does not sleep well one night, they will sleep better another night.
- Pop-up explaining the “point” of light sleep
- Explanation of why spending time in bed during the day can make it harder to fall asleep at night

Why could sleeping in make me feel bad?

- Explanation of the purpose of cortisol and its role in the sleep-wake cycle
- Graphs to evidence how irregular wake-up times impact cortisol release and how this, in turn, can impact mood
- Pop-up with tips to avoid napping impacting sleep (≤20 min, before 3 PM, and with the curtains open) and an explanation of why

How can I stop worrying in bed?

- Examples of common bedtime worries and explanation that this trains the brain to worry in bed
- Getting up after 20 min if unable to sleep and going back to bed when they feel sleepy
- Tips for calming activities to engage in instead
- Reiteration that, if they do not sleep well one night, they will sleep better another night
- Pop-up—tips on calming the mind before bed
- List of tips for calming the mind in bed

What to do to sleep better?

- The 1-week challenge; users are encouraged to try the 3 “sleep hacks” (sleep hygiene behaviors) for 1 week—getting up at the same time every day, helping their brain calm down before bed, and getting up after 20 min if unable to sleep.
- Pop-ups take the users to previous pages to remind them of the science behind the advice.
- Tips on how to make the “sleep hacks” easier—choosing what they want to do before they go to bed, telling friends and family about the challenge for help and support, and putting their phone out of reach so they will be unable to check it
- Normalization of tiredness and reiteration of napping rules
- Voice recordings (including transcripts) of adolescents explaining how they made the hacks work for them
- Link to personal reminders and diary
- Link to Sleep Healthy Using the Internet

Track my progress and earn stars (added as a result of patient and public involvement feedback)

- Optional reward system—switch to turn weekly challenge on or off
- Each day, the user selects whether they used the sleep hacks, partially used the sleep hacks, or did not use the sleep hacks the night before.
- If they used or partially used the sleep hacks, they earn a blue star.
- If the user collects 5 blue stars in a week, they earn a gold star.

What other things could stop me sleeping well?

- Problem and solution pairs explaining the science behind how various lifestyle factors (eg, caffeine, vaping, and social media) can impact sleep and offering harm reduction solutions

Links to help the brain calm down

- Links to calming sounds, mindfulness and meditation resources

User journey

- The user is presented with a splash screen with an animated Sleep Solved logo.
- Each key element is unlocked in a linear sequence once all pages (besides the optional pop-ups) of the previous element have been viewed.
- Element G—"What other things could stop me sleeping well?"—is unlocked 3 weeks after the user downloads Sleep Solved. Early access to this section is unlocked if the user earns 2 gold stars (see element F).
- The user can navigate back to the main menu at any stage using the hamburger menu.

Phase 1: Intervention Development

Sleep Solved Ideas and Concept Generation

PPI contributors from the Bristol YPAG felt that the 3 ideas presented to them were novel. Many were keen for the app to be "soothing" and "simple," with "noises that help you get to sleep" when feeling worried. These suggestions were implemented through a calming choice of colors; simple gestural app navigation; and the inclusion of "Links to help your brain calm down," such as music and meditation sessions. The "How can I stop worrying in bed?" section gave ideas on calming activities.

Contributors expressed an interest in how phone use, food, and drink can affect sleep. These questions were addressed in the "What other things could stop me sleeping well?" section, which presents problems (how food, drink, and nicotine can affect sleep) with solutions (what users can do instead). How phone use can affect sleep at night was covered in the "How long should I try to get to sleep?" section, where users are advised to "avoid bright lights such as phone screens. Bright light can make your brain think it's time to wake up!"

Sleep Solved Ideas: Name and Design

Results from the first Qualtrics survey helped create the overall look and feel of Sleep Solved. PPI contributors gave feedback on five different elements:

1. Potential app names: of the 7 options, "Sleep Solved" was chosen as the preferred name.
2. Color palettes: a black, opaline green, deep cream, and pale roundel blue color scheme was selected, with an average "like" score of 71% (SD 20.72%). Contributors shared that these colors were "simple," evoking thoughts of "the sea" and "spring awakening."
3. Logos: a bold sans-serif logo in all capital letters was chosen, with an average "like" score of 69% (SD 10.63%). Contributors liked the "bright colours" and readability.
4. Design themes: dark navy was chosen, with cartoonlike graphics in lime green, blue, purple, and orange. Praise was given for the "colourful," "clear" design, which was "easy to find."
5. Sleep Well recruitment materials: changes were made to the poster to make the fonts easier to read and the features of the app clearer.

New Theme Designs for Sleep Solved

On the basis of the feedback from the first Qualtrics survey, 2 potential app themes were designed by PIP Creative. Theme 1 was chosen in the second Qualtrics survey with an average "like" score of 68%. Contributors liked the "sleepy colour theme" and dark colors, which "connotate sleep"; "contrasting colours are cool, it's very appealing." They preferred brightly colored icons as these "draw in my attention." Less praise was given to theme 2, a softer theme with rounded dark purple and navy iconography: "it's not as vibrant and interesting" and "too dull, doesn't draw my attention."

New Logo Designs for Sleep Solved

In total, 4 logos were presented to PPI contributors in the third Qualtrics survey. PPI contributors found a colorful, bold logo made of interconnecting shapes "difficult to read...because of the colour scheme." A bold sans-serif logo was chosen in white and yellow. The bright colors featured in a geometric fan shape next to the logo, making it "better [for] being colourful" but still easy to read.

Sleep Well Design Feedback

The results of the fourth Qualtrics survey are shown in Tables 4 and 5.

Table 4. An overview of the Sleep Solved intervention content.

Route	How likely are you to take part for 6 weeks? (0-100), mean (SD)
1—Phone Downtime for 6 weeks, then access to Sleep Solved for 6 weeks, then access to SHUTi[a]	64.3 (26.1)
2—Phone Downtime and Sleep Solved for 6 weeks, then access to SHUTi	63.3 (25.4)
3—Access to all 3 components right away	55.7 (27.5)

[a]SHUTi: Sleep Healthy Using the Internet.

Table 5. An overview of patient and public involvement contributors' preferred Sleep Solved intervention content (n=23).

	First place		Second place		Third place	
	Values, n (%)	Values, mean (SD)	Values, n (%)	Values, mean (SD)	Values, n (%)	Values, mean (SD)
Route 1	7 (30)	2.09 (0.83)	7 (30)	2.09 (0.83)	9 (39)	2.09 (0.83)
Route 2	7 (30)	1.70 (0.55)	14 (60)	1.70 (0.55)	1 (4)	1.70 (0.55)
Route 3	9 (39)	2.22 (0.93)	2 (8)	2.22 (0.93)	13 (56)	2.22 (0.93)

When asked to rank their preferred options for phasing the steps of the intervention, there was no clear leader. Route 3 was ranked as contributors' first choice 9 times but was ranked in last place by 13 PPI contributors. As route 2 was ranked second by most people (14/23, 60%), this was chosen as a compromise as very few (2/23, 8%) ranked this as their last-placed option compared to route 3 (13/23, 56%). PPI contributors shared that they felt that route 2 was the most time-efficient, logical option:

> *I think [Route 2 is] the best way for people to be able to see how the app is helping their sleep schedule which encourages them to keep using it...then they can determine whether they need the SHUTi app or not.* [PPI contributor; female; aged 17 years; White British ethnicity]

> *Phone Downtime and Sleep Solved don't require much effort so they can be both done at the same time.* [PPI contributor; male; aged 14 years; White British ethnicity]

> *[Route 2] is less time consuming.* [PPI contributor; nonbinary person; aged 18 years; White British ethnicity]

Phase 2: Intervention Optimization

Web-Based Think-Aloud Interviews

Results from the think-aloud interviews were largely positive, with praise for the clear, easily understood content and ease of functionality:

> *I like the way it's broken up into sections.* [JK02]

> *I think one of my favourite parts about it is the explanations behind a lot of it. I like knowing kind of why I'm doing something rather than just saying like this person says it works.* [JK06]

> *"Science based"—makes it more trustworthy, I'm hopeful.* [JK05]

> *It's nice to know, there's various ways that you can improve your sleep, there's not just one way, and that it's okay if one doesn't work for you because there's always another and that you can start off easy and then like, build up.* [JK03]

Nevertheless, the interviews identified numerous small but important ways in which the intervention could be improved. A summary of the think-aloud feedback and the changes implemented can be found in Table 6. Examples from the table of changes can be found in Multimedia Appendix 2. Less positive responses focused on areas in which users found information difficult to understand as it was originally presented or fonts were hard to read.

Table 6. Summary of changes to Sleep Solved as a result of think-aloud interview feedback.

Section of Sleep Solved and problem identified	Changes made
Overall design	
• "The contrast of the light blue text to the white background on highlighted words makes it difficult to read—I have dyslexia. Prefer the white text on a dark background." [JK04] • "The terms 'anxiety' and 'depression' could be seen as quite negative or serious terms." [GT01]	• Light blue font was changed to darker blue font throughout to be easier to read. We retained the use of bold and italic for emphasis. • References to anxiety or depression were changed to "worried" or "low mood" throughout to reduce any negative connotations.
How can Sleep Solved help me sleep better?	
• "Who made this app?": • "Might be good to include more information on why these people are interested in contributing to the app." [JK02] • "This section is unnecessary; I don't care who made the app—I trust if you say it's backed by science." [MJ02]	• The "Who made this app" page was changed to an optional popout: "Who made this app? Tap here." Photos introducing each team member were amended to a brief paragraph of text recognizing the valuable contributions of hundreds of adolescents and providing links to partner charities.
• "Get good marks—isn't clear that you're talking about school. 'Do better in school' would be better." [MJ01] • "Improve your grades would be better." [MJ02]	• References to improving grades were removed as it was felt that this may not be an achievable goal or may not be a priority for all the adolescents using the app.
Why could sleeping in make me feel bad?	
• "The order of the information is hard to understand."[MJ01]	• The text on the final card, "Getting up at the same time every day trains your brain to release cortisol at the right time," was moved to the first slider. • The "if you sleep late your cortisol levels peak late" text was combined with "this may make you feel tired and sleepy when you have to get up earlier" to improve understanding.
• "Is it OK to nap when I'm tired?": • "Why do you have to nap before 3pm?" [GS01 and JK06] • "Why 20 minutes?" [JK06]	• Explanatory text was added to clarify that a 20-min nap would prevent users from sleeping too long or too deeply. • Furthermore, we provided evidence of the science behind this recommendation—that, if users sleep late in the day, their brain will release cortisol on waking, making them feel more awake and making it more difficult to sleep later that night.
How can I stop worrying in bed?	
• "My brain never really stops thinking, maybe tell us how to stop thinking." [JK07]	• Text changed from "thinking" to "worrying": "Lying awake at night, you might start to worry about: how tired you will feel in the morning, all the other worrying things in your life. This trains your brain to worry in bed instead of sleeping!"
What to do to sleep better?	
• "The title, 'What to do to sleep better' does not make it clear that this is the one-week challenge." [MJ01]	• A front page was added stating the following: "This is your one-week challenge! Try these sleep hacks for 1 week!"
Track my progress and earn stars	

• Caffeine—problem:
  • "An exhaustive list of everything that contains caffeine would be helpful. This could be presented as a mind-map to make it more visually appealing." [MJ06]
• Caffeine—graph:
  • "Would be good to indicate the 10hr period somewhere on the graph." [MJ06]
  • "The sun/moon/cloud symbols don't really add anything—they're confusing." [MJ06]
	• A list of all drinks and foods containing caffeine was added as a pop-up section.
	• Changes were made to a graph depicting caffeine levels in the body over time. Consumption of 2 caffeinated drinks was shown over the course of a day, and more accurate timings were added to represent how caffeine levels in the body change over time.

• Food—solution: • "It would be good to have a more extensive list of the things that you can eat." [MJ06]	• More foods were added to the list.

Sleep Solved Real-Life Testing: Think-Aloud Interviews

Overview

Table 7 provides an overview of each PPI contributor's sleep hack use during their 1-week trial of the Sleep Solved app before their second think-aloud interview and the perceived strengths and limitations of each hack.

Table 7. Feedback on Sleep Solved use.

Unique user identifier	Use pattern	Strengths	Limitations
C1	• Hack 1[a]	• Hack 1: going to sleep earlier • Hack 1: feeling less tired and stressed during the day	—[b]
C2	• Hack 1 • Hack 2[c]	• Hack 1: falling asleep more quickly • Hack 2: feeling more awake in the morning • Hack 2: increased productivity and focus • Hack 2: falling asleep faster	• Hack 1: found it difficult to stick to it on weekends • Hack 2: time-consuming during examination periods
C3	• Hack 1 • Hack 2 • Hack 3[d]	• Hack 1: feeling more energized in the morning • Hack 2: reduced worried thoughts and falling asleep faster • Hack 3: falling asleep more easily	• Hack 1: not motivated to maintain the behavior in the long term • Hack 2: did not maintain the behaviors in the long term; enjoys using their phone to watch videos or message friends before bed
C4	• Hack 1 • Hack 2 • AN[e]	• Hack 1: feeling more rested • Hack 2: falling asleep more quickly • AN: falling asleep earlier	• Hack 1: challenging to follow (part-time job with evening shifts) • AN: hard to maintain when very tired and not busy
C5	• Hack 1 • Hack 2 • Hack 3	• Hack 1: falling asleep earlier, napping less during the day, and waking up less during the night	• Hack 1: still sleeping lightly • Hacks 1 and 2: did not improve sleep
C6	• Hack 1 • Hack 2 • Hack 3	• Hack 1; falling asleep earlier, sleeping more deeply, and feeling more alert and refreshed in the mornings • Hack 2: reduced anxiety before bed • Hack 3: falling asleep more quickly	• Hacks 1 and 2: would be hard to maintain when traveling and staying at friends' houses • Hack 3: difficult to maintain during the school holidays
C7	• Hack 2 • Hack 3	• Hack 2: falling asleep earlier and more quickly	• Hack 3: did not improve sleep

[a]Getting up at the same time every day.

[b]No limitations suggested.

[c]Calming your brain.

[d]Less time in bed.

[e]AN: avoiding napping.

Set an Alarm to Get Up at the Same Time Every Day (Hack 1)

PPI contributors reported great variability in their sleep and wake times before trying hack 1, largely due to varying class schedules. Many of those involved reported a notable change in their sleep schedules after trying this hack: falling asleep and rising earlier. However, hack 1 was not feasible for PPI contributors who had to work late-night shifts and, therefore, had a variable sleep schedule. Within the socioecological model, this would correspond to the work, school, and college moderator, which recognizes the impact of variable school, college, and work start times on sleep, particularly in children and adolescents [57,68]. Several PPI contributors described needing motivation to continue the sleep hack on the weekends or during the holidays. This was particularly relevant when taking examinations as they were more likely to try to "catch-up" on sleep on the weekends and were more reluctant to get up at the same time as on weekdays.

Help Your Brain Calm Down Before You Go to Bed: Do Not Do Anything Exciting or Stressful (Hack 2)

Relaxation strategies were perceived as positive by most PPI contributors, finding that this advice helped them reduce their anxious thoughts and fall asleep faster. Calming activities described by contributors included reading, watching television, listening to relaxing music, and ensuring that they stopped homework or revision activities with enough time to "wind down" before bed. Others explained how stopping the use of social media apps before bed had really helped improve their sleep, feeling more "refreshed" on waking. However, a minority found it difficult not to use their phones in the hour before bed, and they continued to message their friends, use social media, and watch videos. In the socioecological model and our logic model, this is represented by the social networks moderator—the friends, acquaintances, and networks that contribute to roles, expectations, beliefs, and behaviors regarding sleep [57].

If You Do Not Get to Sleep After 20 Minutes, Get Up and Do Something Calm Until You Are Sleepy (Hack 3)

Despite initial worries that following this hack would be "impossible," contributors reported finding it easier to fall asleep on returning to bed instead of their usual practice of using their phone or lying awake unable to sleep. Although perceived as initially challenging, after a few days of practice, many could follow the hack, and their sleep latency improved over time. For a minority, this was their favorite sleep hack, and one user described counting to their 20-minute target as so restful that they fell asleep. PPI contributors reported that spending less time in bed helped reduce the overthinking and associated anxiety that they experienced. However, one contributor had concerns that getting up and moving around may disturb other people in their household.

Overall Acceptability

Sleep Solved was deemed acceptable by the PPI contributors. They were appreciative of the range of text formats, simple language, and the use of bullet points to break up information, as well as the use of bright imagery. They described the interactive "slider" navigation elements as "engaging," "fun," and "encouraging." Users praised the science-based content, which provided them with simple yet clear explanations of the reasoning behind the (seemingly contradictory) behaviors to sleep better, such as spending less time in bed. The use of choice architecture (the framing of the advice given as an option, consistent with the guiding principles) was well received. Contributors were given the benefits of the desired behavior compared to the consequences of undesired behaviors. By not telling users which behavior they "should" be doing, many were keen to follow the behavior when the benefits and reasoning for the advice were well explained:

> *I think it explaining that little bit more makes you feel more kind of like it's a good thing, because I'm more aware of what's going on rather than just this is what I'm being told it's going on in my brain, but it makes you feel more kind of knowledgeable in it. It feels very personalized because I can kind of relate to that, that is exactly...how I feel.* [PPI contributor C4; female; aged 18 years; sleep problems 2-3 times per week]

Being given clear reasons for following the suggested behaviors in a particular way was similarly well received by users:

> *[The science-based approach] it's informative in a non-patronising way. If you're patronised, then you're not going to listen...Like if your parents tell you to do something, they're describing [it to] you, but if you know why you should do it, you're more likely to be doing it.* [PPI contributor C1; female; aged 17 years; sleep problems 2-3 times per week]

> *It's a bit humorous as well, which is quite interesting. Rather than...just telling you what to do and what not to do.* [PPI contributor C2; female; aged 16 years; sleep problems 2-3 times per week]

Web-Based Workshop 1: Sleep Solved Real-Life Testing

Feedback from these users was positive. PPI contributors considered the bullet points, visual aids, and graphs to be "helpful for understanding."

Set an Alarm to Get Up at the Same Time Every Day (Hack 1)

This hack was cited as being harder to stick with, particularly on the weekends or during holidays. PPI contributors felt that not having to get up for school or college made their sleep "less disciplined" (M1) as on these days they were able to procrastinate, putting off getting up and out of bed for longer (M2 and M3). However, one contributor liked the fact that, by charging their phone at the other side of the room, they could not be tempted to check on it; they felt that this advice was very helpful (M2).

Help Your Brain Calm Down Before You Go to Bed: Do Not Do Anything Exciting or Stressful (Hack 2)

All PPI contributors liked the concept of taking part in calming activities before bed, sharing that they felt "more refreshed when I wake up" (M3). Stopping exciting or stressful activities such as homework, revising for examinations (M3), or watching Netflix or TikTok before bed (M7) was "really helpful" (M7) and helped them sleep (M2). One PPI contributor shared that this was their examination season, so they were keen to get as much sleep as possible (M1).

If You Do Not Get to Sleep After 20 Minutes, Get Up and Do Something Calm Until You Are Sleepy (Hack 3)

Although this sleep hack was the favorite of some PPI contributors (M7), others worried that they may disturb other people in their family when getting out of bed or moving around (M2), which, when considering the socioecological model, recognizes the impact of moderators such as family, socioeconomic status, and environment on sleep [57]. For some, counting the 20 minutes until it was time to sleep was like "counting sheep," and they found that they fell asleep anyway (M3).

Web-Based Workshop 2: Phone Downtime Testing

Overview

PPI contributors also tested the Phone Downtime app for 1 week, which was an app that was included for Android users of Sleep Solved. Users were encouraged to set the time when they usually went to sleep and woke up. The Phone Downtime app enabled users to tell when they were using their phone during the time they planned to be asleep (during a time window preset by the user).

PPI contributors liked being able to choose their own goals and sleep times and the fact that working days were flexible to accommodate part-time jobs on different days. A new awareness of for just how long they were using their phones at night was welcome as it helped them better schedule their bedtime.

The Need for Interaction

A large proportion of PPI contributors were keen on the addition of a progress chart, journal, or sleep diary. Many expressed a wish to track and measure their sleep behavior changes over time. In its initial prototype stage, adolescents did not feel that there was much to attract them to Sleep Solved once they had completed all the sections. They suggested adding an interactive reward system (eg, where they could collect points and monitor their progress over time).

As a result of contributor feedback, an optional reward system was added to Sleep Solved. Under a new section, "Track my progress and earn stars," users can activate their rewards using a toggle switch to turn their weekly challenge on or off. Each day, a pop-up asks the user whether they (1) used the sleep hacks, (2) partially used the sleep hacks, or (3) did not use the sleep hacks the night before. If they used or partially used the sleep hacks, they earn a blue star. If the user collects 5 blue stars in a week, they earn a gold star. Collecting 2 gold stars enables early access to supplementary content within the app, which provides additional information about the influence of behaviors such as vaping, drinking caffeine, and diet on sleep.

While the need for adolescents to continue to return to the app was not a requirement for effective engagement [69], similar sleep behavior apps have also used elements of gamification and reward systems. These include users tracking their sleep for 3 nights out of 7 to progress to their next "belt" in Sleep Ninja [45], gift cards for users of Sleepio, a digital app to improve sleep behavior using internet-based cognitive behavioral therapy [70], and monetary rewards for children who met their predefined sleep goals [71]. In an exploration of young adults' perspectives on gamification in health apps, rewards such as badges and progress indicators were viewed more positively than self-rewards [72]. Young adults with depression and anxiety have expressed a preference for self-tracking behaviors over time; gamified elements that rewarded progress toward their behavior "goals" increased their motivation to engage with the app and regularly track their behavior [73].

PPI contributors were keen to share their own sleep-related experiences for other adolescents who may use the app in the future. A total of 7 adolescents who participated in the workshops recorded relatable stories related to poor sleep and how Sleep Solved had helped them. A variety of ages, ethnicities, and regional accents were represented following the guiding principle to "ensure that role models and stories were relevant to diverse contexts and identities." These were presented to app users as playable audio clips.

Discussion

Principal Findings

This study used the person-based approach to design and optimize the first step of a 3-stage, app-based intervention aimed at improving sleep-related behaviors in adolescents aged 14 to 18 years. Sleep Solved was considered acceptable, feasible, and easy to use by PPI contributors, including those from underrepresented and diverse backgrounds.

This study highlights that offering varied means of contributing to intervention coproduction is helpful in terms of maximizing the speed of the co-creation process and the diversity of the sample of PPI contributors. It also confirms that, for advice to be engaging for adolescents, it is essential that it is perceived as trustworthy and novel and framed as a choice. A key finding from this study was the importance of offering our PPI contributors varied methods of contribution. This allowed contributors to express feedback in the way in which they felt most comfortable, resulting in diverse and complementary insights.

Web-based surveys were the most popular contribution method and allowed us to purposively recruit adolescents from low-income communities. Web-based surveys facilitated quick feedback on multiple-choice decisions, such as the app name, theme colors, and pathway options for the study. The second most popular contribution method was to take part in a web-based, one-to-one, think-aloud interview. These interviews provided in-depth feedback, allowing us to iteratively optimize the app over a 7-month period.

In comparison to previous research, this study used a 3-stage, app-based intervention, incorporating a co-created first stage followed by CBTi (SHUTi) and mental health support (Bite Back). Levenson et al [74] also designed adolescent sleep interventions with stakeholder input and involvement in a series of 3 focus groups with young adults, adolescents, and health care professionals with experience working with adolescents. These focus groups informed the development of a clinic-based sleep intervention for adolescents aged 13 to 15 years. Similarly, the "Momentum" intervention for young people with insomnia and mental health difficulties was co-created with young people aged 7 to 17 years, with participants encouraged to give feedback on the "look and feel." As advised in the recommendations from the Momentum study, Sleep Well has features that encourage interaction, such as games, videos, interactive elements, quizzes, and rewards.

As with our SHUTi insomnia intervention [51], Palermo et al [75] also used CBTi to develop a 4-session CBTi intervention for adolescents with insomnia and co-occurring mental health conditions. Participants were encouraged to complete sleep diaries during treatment. The results indicated a high level of acceptability and feasibility, with high compliance; 85% (n=34) completed all 6 intervention sessions. Later pilot-testing indicated that the CBTi intervention was associated with improvements in insomnia symptoms, sleep quality, and sleep onset latency.

Focus group and workshop contributors, who were recruited from existing networks such as the Bristol YPAG and the McPin Foundation, were generally highly confident and motivated to engage with the study. They engaged thoroughly with the app over the course of a week, providing important suggestions, and were sufficiently confident to record their personal experiences for inclusion in the app.

Strengths

In line with our guiding principle to ensure that Sleep Solved was perceived as interesting, trustworthy, and novel, our results highlighted that PPI contributors particularly valued the accessible presentation of the science-based advice in Sleep Solved. The perception of the advice as novel was important as adolescents have been shown to have a very low tolerance for behavior change interventions that they perceive as dull or boring [63,64]. The perception of the advice as trustworthy may be a key distinguishing feature of Sleep Solved; in a review of the 76 sleep-related health apps available on Google Play, less than 1 in 3 contained evidence to support their claims [60]. In line with our guiding principle to ensure that advice is framed as a choice, PPI contributors particularly appreciated not feeling "patronized" by the advice in Sleep Solved.

The key strength of this study is that our intervention was coproduced from start to finish, with quality in-depth, varied, and iterative feedback from PPI contributors. The role of the peer researcher (MHJ) in building ongoing, trusting relationships with contributors was a key feature of this success.

Limitations

One major limitation is the representativeness of our PPI contributors. Of those invited to share their feedback, only a small proportion volunteered to contribute to co-design. Therefore, it is likely that our contributors were a more highly motivated subsection of our target demographic, and it is possible that their recommendations may not always represent the views of less motivated adolescents.

In addition, a significantly smaller number of adolescents aged 14 to 15 years volunteered to share their views in think-aloud interviews and group workshops. It is possible that recommendations from those who did not wish to engage with us in more depth may have differed from the suggestions of the adolescents aged 16 to 18 years who helped us develop the app.

Future Work

To foster a more representative PPI sample, forming early relationships with individuals from the relevant community with whom recruitment procedures can be coproduced [69] will likely be beneficial. The formation of these relationships may be made easier by having a more diverse research team [70].

Conclusions

In conclusion, offering multiple methods of providing feedback allows for the collection of holistic feedback from a diverse range of contributors. This feedback resulted in an app viewed as acceptable and engaging by the adolescents who have tried it to date. Further evaluation is needed to determine the feasibility and acceptability of the intervention in a larger group of adolescents, and future research should explore ways of engaging hard-to-reach young people in intervention development research.

Acknowledgments

The authors would like to gratefully acknowledge the support of their partners in the recruitment of patient and public involvement contributors, namely, the Association of Colleges, the E-ACT multiacademy trust, the McPin Foundation, the Bristol Young People's Advisory Group, and Off The Record Bristol. The authors would also like to thankfully acknowledge the work of PIP Creative in the graphic design development of the Sleep Solved intervention. Finally, the authors are very grateful to the funders of this research, The Prudence Trust. LY is a National Institute for Health and Care Research (NIHR) Senior Investigator, and her research program is partly supported by the NIHR Applied Research Collaboration West and NIHR Health Protection Research Unit in Behavioural Science and Evaluation.

Conflicts of Interest

LR reports having equity ownership in BeHealth Solutions, LLC, who originally licensed the Sleep Healthy Using the Internet (SHUTi) program from the University of Virginia. Somryst, a commercial prescription digital therapeutic for insomnia, was developed based on the SHUTi program by Pear Therapeutics, who subsequently sold their license to Nox Health. Nox Health has a royalty agreement with BeHealth Solutions, LLC and the UVA Licensing and Venture Group. LR is a consultant of Nox Health. The terms of this arrangement have been reviewed and approved by the University of Virginia in accordance with its conflict of interest policy.

Multimedia Appendix 1
Patient and public involvement (PPI) activities—engagement activities with our PPI contributors in chronological order.
[DOCX File , 26 KB - humanfactors_v11i1e63341_app1.docx]

Multimedia Appendix 2
An overview of think-aloud questions and prompts asked to patient and public involvement contributors when trialing the prototype Sleep Solved intervention.
[DOCX File , 24 KB - humanfactors_v11i1e63341_app2.docx]

Multimedia Appendix 3
Examples from the table of changes.
[DOCX File , 25 KB - humanfactors_v11i1e63341_app3.docx]

Multimedia Appendix 4
The Sleep Hacks Diary used in think-aloud interview user testing.
[DOCX File , 23 KB - humanfactors_v11i1e63341_app4.docx]

Multimedia Appendix 5
A prototype Sleep Solved app feedback template for patient and public involvement contributors to complete notes and feedback about their experiences.
[DOCX File , 22 KB - humanfactors_v11i1e63341_app5.docx]

References

1. Saxvig IW, Pallesen SI, Wilhelmsen-Langeland A, Molde H, Bjorvatn B. Prevalence and correlates of delayed sleep phase in high school students. Sleep Med 2012 Feb;13(2):193-199. [doi: 10.1016/j.sleep.2011.10.024] [Medline: 22153780]
2. Booth SA, Carskadon MA, Young R, Short MA. Sleep duration and mood in adolescents: an experimental study. Sleep 2021 May 14;44(5):zsaa253. [doi: 10.1093/sleep/zsaa253] [Medline: 33245773]
3. Pasch KE, Laska MN, Lytle LA, Moe SG. Adolescent sleep, risk behaviors, and depressive symptoms: are they linked? Am J Health Behav 2010;34(2):237-248 [FREE Full text] [doi: 10.5993/ajhb.34.2.11] [Medline: 19814603]
4. Keyes KM, Maslowsky J, Hamilton A, Schulenberg J. The great sleep recession: changes in sleep duration among US adolescents, 1991-2012. Pediatrics 2015 Mar;135(3):460-468 [FREE Full text] [doi: 10.1542/peds.2014-2707] [Medline: 25687142]
5. Liu D, Santhanam R, Webster J. Toward meaningful engagement: a framework for design and research of gamified information systems. MIS Q 2017 Apr 4;41(4):1011-1034 [FREE Full text] [doi: 10.25300/misq/2017/41.4.01]
6. Gradisar M, Kahn M, Micic G, Short M, Reynolds C, Orchard F, et al. Sleep's role in the development and resolution of adolescent depression. Nat Rev Psychol 2022;1(9):512-523 [FREE Full text] [doi: 10.1038/s44159-022-00074-8] [Medline: 35754789]
7. Palmer CA, Oosterhoff B, Massey A, Bawden H. Daily associations between adolescent sleep and socioemotional experiences during an ongoing stressor. J Adolesc Health 2022 Jun;70(6):970-977 [FREE Full text] [doi: 10.1016/j.jadohealth.2022.01.127] [Medline: 35256239]
8. Yang Y, Liu XM, Liu ZZ, Tein JY, Jia CX. Life stress, insomnia, and anxiety/depressive symptoms in adolescents: a three-wave longitudinal study. J Affect Disord 2023 Feb 01;322:91-98. [doi: 10.1016/j.jad.2022.11.002] [Medline: 36372126]
9. Tucker CJ, Sharp EH, Van Gundy KT, Rebellon C. Household chaos, hostile parenting, and adolescents' well-being two years later. J Child Fam Stud 2018 Jul 19;27(11):3701-3708. [doi: 10.1007/s10826-018-1198-x]
10. Darchia N, Cervena K. The journey through the world of adolescent sleep. Rev Neurosci 2014;25(4):585-604. [doi: 10.1515/revneuro-2013-0065] [Medline: 24717334]
11. Galland BC, Taylor BJ, Elder DE, Herbison P. Normal sleep patterns in infants and children: a systematic review of observational studies. Sleep Med Rev 2012 Jun;16(3):213-222. [doi: 10.1016/j.smrv.2011.06.001] [Medline: 21784676]
12. Crowley SJ, Acebo C, Carskadon MA. Sleep, circadian rhythms, and delayed phase in adolescence. Sleep Med 2007 Sep;8(6):602-612. [doi: 10.1016/j.sleep.2006.12.002] [Medline: 17383934]
13. Wolfson AR, Carskadon MA. Sleep schedules and daytime functioning in adolescents. Child Dev 1998 Aug;69(4):875-887. [Medline: 9768476]
14. Buxton OM, Chang AM, Spilsbury JC, Bos T, Emsellem H, Knutson KL. Sleep in the modern family: protective family routines for child and adolescent sleep. Sleep Health 2015 May 01;1(1):15-27 [FREE Full text] [doi: 10.1016/j.sleh.2014.12.002] [Medline: 26779564]
15. Glasbeek MP, Inhulsen MB, Busch V, van Stralen MM. Sleep reduction in adolescents: socio-demographic factors and the mediating role of sleep hygiene practices. Sleep Epidemiol 2022 Dec;2:100024 [FREE Full text] [doi: 10.1016/j.sleepe.2022.100024]
16. Stepanski EJ, Wyatt JK. Use of sleep hygiene in the treatment of insomnia. Sleep Med Rev 2003 Jun;7(3):215-225. [doi: 10.1053/smrv.2001.0246] [Medline: 12927121]
17. Weiss A, Xu F, Storfer-Isser A, Thomas A, Ievers-Landis CE, Redline S. The association of sleep duration with adolescents' fat and carbohydrate consumption. Sleep 2010 Sep;33(9):1201-1209 [FREE Full text] [doi: 10.1093/sleep/33.9.1201] [Medline: 20857867]
18. Alahmary SA, Alduhaylib SA, Alkawii HA, Olwani MM, Shablan RA, Ayoub HM, et al. Relationship between added sugar intake and sleep quality among university students: a cross-sectional study. Am J Lifestyle Med 2022;16(1):122-129 [FREE Full text] [doi: 10.1177/1559827619870476] [Medline: 35185434]
19. National Diet and Nutrition survey rolling programme years 9 to 11 (2016/2017 to 2018/2019): a survey carried out on behalf of Public Health England and the Food Standards Agency. Public Health England. 2022. URL: https://assets.publishing.service.gov.uk/media/5fd23324e90e07662b09d91a/NDNS_UK_Y9-11_report.pdf [accessed 2024-04-29]

20. Nuss T, Morley B, Scully M, Wakefield M. Energy drink consumption among Australian adolescents associated with a cluster of unhealthy dietary behaviours and short sleep duration. Nutr J 2021 Jul 05;20(1):64 [FREE Full text] [doi: 10.1186/s12937-021-00719-z] [Medline: 34225738]
21. Otsuka Y, Kaneita Y, Itani O, Osaki Y, Higuchi S, Kanda H, et al. Association between unhealthy dietary behaviors and sleep disturbances among Japanese adolescents: a nationwide representative survey. Sleep Biol Rhythms 2018 Oct 17;17(1):93-102. [doi: 10.1007/s41105-018-0193-3]
22. Aepli A, Kurth S, Tesler N, Jenni OG, Huber R. Caffeine consuming children and adolescents show altered sleep behavior and deep sleep. Brain Sci 2015 Oct 15;5(4):441-455 [FREE Full text] [doi: 10.3390/brainsci5040441] [Medline: 26501326]
23. Pollak CP, Bright D. Caffeine consumption and weekly sleep patterns in US seventh-, eighth-, and ninth-graders. Pediatrics 2003 Jan;111(1):42-46. [doi: 10.1542/peds.111.1.42] [Medline: 12509552]
24. Orbeta RL, Overpeck MD, Ramcharran D, Kogan MD, Ledsky R. High caffeine intake in adolescents: associations with difficulty sleeping and feeling tired in the morning. J Adolesc Health 2006 Apr;38(4):451-453. [doi: 10.1016/j.jadohealth.2005.05.014] [Medline: 16549311]
25. Simon A, Owen C, O'Connell R, Brooks F. Changing trends in young people's food behaviour and wellbeing in England in relation to family affluence between 2005 and 2014. J Youth Stud 2017 Nov 21;21(5):687-700. [doi: 10.1080/13676261.2017.1406599]
26. Felden É, Leite CR, Rebelatto CF, Andrade RD, Beltrame TS. Sono em adolescentes de diferentes níveis socioeconômicos: revisão sistemática. Rev Paul Pediatr 2015 Dec;33(4):467-473 [FREE Full text] [doi: 10.1016/j.rpped.2015.01.011] [Medline: 26298657]
27. Zeringue MM, Saini EK, Fuller-Rowell TE, Hinnant JB, El-Sheikh M. Neighborhood environment and adolescent sleep: the role of family socioeconomic status. Sleep Med 2023 Sep;109:40-49 [FREE Full text] [doi: 10.1016/j.sleep.2023.06.014] [Medline: 37413781]
28. Mayne SL, Morales KH, Williamson AA, Grant SF, Fiks AG, Basner M, et al. Associations of the residential built environment with adolescent sleep outcomes. Sleep 2021 Jun 11;44(6):zsaa276 [FREE Full text] [doi: 10.1093/sleep/zsaa276] [Medline: 33507268]
29. Sheehan C, Powers D, Margerison-Zilko C, McDevitt T, Cubbin C. Historical neighborhood poverty trajectories and child sleep. Sleep Health 2018 Apr;4(2):127-134 [FREE Full text] [doi: 10.1016/j.sleh.2017.12.005] [Medline: 29555124]
30. Philbrook LE, Buckhalt JA, El-Sheikh M. Community violence concerns and adolescent sleep: physiological regulation and race as moderators. J Sleep Res 2020 Jun 30;29(3):e12897 [FREE Full text] [doi: 10.1111/jsr.12897] [Medline: 31362331]
31. Evans GW, Gonnella C, Marcynyszyn LA, Gentile L, Salpekar N. The role of chaos in poverty and children's socioemotional adjustment. Psychol Sci 2005 Jul 01;16(7):560-565. [doi: 10.1111/j.0956-7976.2005.01575.x] [Medline: 16008790]
32. Narmandakh A, Roest AM, Jonge PD, Oldehinkel AJ. The bidirectional association between sleep problems and anxiety symptoms in adolescents: a TRAILS report. Sleep Med 2020 Mar;67:39-46. [doi: 10.1016/j.sleep.2019.10.018] [Medline: 31887607]
33. Yasugaki S, Okamura H, Kaneko A, Hayashi Y. Bidirectional relationship between sleep and depression. Neurosci Res 2023 Apr 26:S0168-0102(23)00087-1 [FREE Full text] [doi: 10.1016/j.neures.2023.04.006] [Medline: 37116584]
34. Alvaro PK, Roberts RM, Harris JK, Bruni O. The direction of the relationship between symptoms of insomnia and psychiatric disorders in adolescents. J Affect Disord 2017 Jan 01;207:167-174. [doi: 10.1016/j.jad.2016.08.032] [Medline: 27723540]
35. Jansson-Fröjmark M, Lindblom K. A bidirectional relationship between anxiety and depression, and insomnia? A prospective study in the general population. J Psychosom Res 2008 Apr;64(4):443-449 [FREE Full text] [doi: 10.1016/j.jpsychores.2007.10.016] [Medline: 18374745]
36. Haugland BS, Hysing M, Baste V, Wergeland GJ, Rapee RM, Hoffart A, et al. Sleep duration and insomnia in adolescents seeking treatment for anxiety in primary health care. Front Psychol 2021 Mar 24;12:638879 [FREE Full text] [doi: 10.3389/fpsyg.2021.638879] [Medline: 33841272]
37. Pesonen AK, Martikainen S, Heinonen K, Wehkalampi K, Lahti J, Kajantie E, et al. Continuity and change in poor sleep from childhood to early adolescence. Sleep 2014 Feb 01;37(2):289-297 [FREE Full text] [doi: 10.5665/sleep.3400] [Medline: 24497657]
38. McMakin DL, Alfano CA. Sleep and anxiety in late childhood and early adolescence. Curr Opin Psychiatry 2015 Nov;28(6):483-489 [FREE Full text] [doi: 10.1097/YCO.0000000000000204] [Medline: 26382163]
39. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
40. Lindsey MA, Joe S, Nebbitt V. Family matters: the role of mental health stigma and social support on depressive symptoms and subsequent help seeking among African American boys. J Black Psychol 2010 Nov 01;36(4):458-482 [FREE Full text] [doi: 10.1177/0095798409355796] [Medline: 20953336]
41. Gould MS, Greenberg T, Munfakh JL, Kleinman M, Lubell K. Teenagers' attitudes about seeking help from telephone crisis services (hotlines). Suicide Life Threat Behav 2006 Dec 07;36(6):601-613. [doi: 10.1521/suli.2006.36.6.601] [Medline: 17250466]

42. Boyd C, Hayes L, Nurse S, Aisbett DL, Francis K, Newnham K, et al. Preferences and intention of rural adolescents toward seeking help for mental health problems. Rural Remote Health 2011;11(1):1582 [FREE Full text] [Medline: 21319934]
43. Radez J, Reardon T, Creswell C, Lawrence PJ, Evdoka-Burton G, Waite P. Why do children and adolescents (not) seek and access professional help for their mental health problems? A systematic review of quantitative and qualitative studies. Eur Child Adolesc Psychiatry 2021 Feb 21;30(2):183-211 [FREE Full text] [doi: 10.1007/s00787-019-01469-4] [Medline: 31965309]
44. Muthupalaniappen L, Omar J, Omar K, Iryani T, Hamid SN. Emotional and behavioral problems among adolescent smokers and their help-seeking behavior. Southeast Asian J Trop Med Public Health 2012 Sep;43(5):1273-1279. [Medline: 23431837]
45. Carmona NE, Usyatynsky A, Kutana S, Corkum P, Henderson J, McShane K, et al. A transdiagnostic self-management web-based app for sleep disturbance in adolescents and young adults: feasibility and acceptability study. JMIR Form Res 2021 Nov 01;5(11):e25392 [FREE Full text] [doi: 10.2196/25392] [Medline: 34723820]
46. Werner-Seidler A, Wong Q, Johnston L, O'Dea B, Torok M, Christensen H. Pilot evaluation of the Sleep Ninja: a smartphone application for adolescent insomnia symptoms. BMJ Open 2019 May 27;9(5):e026502 [FREE Full text] [doi: 10.1136/bmjopen-2018-026502] [Medline: 31133584]
47. Werner-Seidler A, O'Dea B, Shand F, Johnston L, Frayne A, Fogarty AS, et al. A smartphone app for adolescents with sleep disturbance: development of the Sleep Ninja. JMIR Ment Health 2017 Jul 28;4(3):e28 [FREE Full text] [doi: 10.2196/mental.7614] [Medline: 28754651]
48. Cliffe B, Croker A, Denne M, Smith J, Stallard P. Digital cognitive behavioral therapy for insomnia for adolescents with mental health problems: feasibility open trial. JMIR Ment Health 2020 Mar 03;7(3):e14842 [FREE Full text] [doi: 10.2196/14842] [Medline: 32134720]
49. Schleider JL, Weisz JR. Little treatments, promising effects? Meta-analysis of single-session interventions for youth psychiatric problems. J Am Acad Child Adolesc Psychiatry 2017 Feb;56(2):107-115. [doi: 10.1016/j.jaac.2016.11.007] [Medline: 28117056]
50. Kruzan K, Ng A, Stiles-Shields C, Lattie EG, Mohr DC, Reddy M. The perceived utility of smartphone and wearable sensor data in digital self-tracking technologies for mental health. Proc SIGCHI Conf Hum Factor Comput Syst 2023 Apr;2023:88 [FREE Full text] [doi: 10.1145/3544548.3581209] [Medline: 38873656]
51. Ritterband LM, Thorndike FP, Gonder-Frederick LA, Magee JC, Bailey ET, Saylor DK, et al. Efficacy of an internet-based behavioral intervention for adults with insomnia. Arch Gen Psychiatry 2009 Jul;66(7):692-698 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2009.66] [Medline: 19581560]
52. Shaffer KM, Smith JG, Glazer JV, Camacho F, Chow PI, Mattos M, et al. Effects of an internet-delivered insomnia intervention for older adults: a secondary analysis on symptoms of depression and anxiety. J Behav Med 2022 Oct;45(5):728-738 [FREE Full text] [doi: 10.1007/s10865-022-00353-y] [Medline: 35932397]
53. Manicavasagar V, Horswood D, Burckhardt R, Lum A, Hadzi-Pavlovic D, Parker G. Feasibility and effectiveness of a web-based positive psychology program for youth mental health: randomized controlled trial. J Med Internet Res 2014 Jun 04;16(6):e140 [FREE Full text] [doi: 10.2196/jmir.3176] [Medline: 24901900]
54. Yardley L, Muller I, Arden-Close E, Ainsworth B. The person-based approach to enhancing the acceptability and feasibility of interventions. Pilot Feasibility Stud 2015 Oct 26;1(1):37 [FREE Full text] [doi: 10.1186/s40814-015-0033-z] [Medline: 27965815]
55. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
56. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Adv Behav Res Ther 1978 Jan;1(4):139-161 [FREE Full text] [doi: 10.1016/0146-6402(78)90002-4]
57. Grandner MA. Social-ecological model of sleep health. In: Grandner MA, editor. Sleep and Health. New York, NY: Academic Press; 2019:45-53.
58. Zhang SC, Yang R, Li DL, Wan YH, Tao FB, Fang J. Association of health literacy and sleep problems with mental health of Chinese students in combined junior and senior high school. PLoS One 2019 Jun 7;14(6):e0217685 [FREE Full text] [doi: 10.1371/journal.pone.0217685] [Medline: 31173621]
59. Evers K, Chen S, Rothmann S, Dhir A, Pallesen S. Investigating the relation among disturbed sleep due to social media use, school burnout, and academic performance. J Adolesc 2020 Oct;84:156-164 [FREE Full text] [doi: 10.1016/j.adolescence.2020.08.011] [Medline: 32920331]
60. Vik FN, Nilsen T, Øverby NC. Associations between sleep deficit and academic achievement - triangulation across time and subject domains among students and teachers in TIMSS in Norway. BMC Public Health 2022 Sep 21;22(1):1790 [FREE Full text] [doi: 10.1186/s12889-022-14161-1] [Medline: 36131267]
61. Dewi RK, Efendi F, Has EM, Gunawan J. Adolescents' smartphone use at night, sleep disturbance and depressive symptoms. Int J Adolesc Med Health 2018 Nov 17;33(2):10.1515/ijamh-2018-0095. [doi: 10.1515/ijamh-2018-0095] [Medline: 30447141]

62. Ghaddar SF, Valerio MA, Garcia CM, Hansen L. Adolescent health literacy: the importance of credible sources for online health information. J Sch Health 2012 Jan;82(1):28-36 [FREE Full text] [doi: 10.1111/j.1746-1561.2011.00664.x] [Medline: 22142172]
63. Duffy A, Christie GJ, Moreno S. The challenges toward real-world implementation of digital health design approaches: narrative review. JMIR Hum Factors 2022 Sep 09;9(3):e35693 [FREE Full text] [doi: 10.2196/35693] [Medline: 36083628]
64. Potter GD, Skene DJ, Arendt J, Cade JE, Grant PJ, Hardie LJ. Circadian rhythm and sleep disruption: causes, metabolic consequences, and countermeasures. Endocr Rev 2016 Dec;37(6):584-608 [FREE Full text] [doi: 10.1210/er.2016-1083] [Medline: 27763782]
65. Bradbury K, Watts S, Arden-Close E, Yardley L, Lewith G. Developing digital interventions: a methodological guide. Evid Based Complement Alternat Med 2014;2014:561320 [FREE Full text] [doi: 10.1155/2014/561320] [Medline: 24648848]
66. Bradbury K, Morton K, Band R, van Woezik A, Grist R, McManus RJ, et al. Using the person-based approach to optimise a digital intervention for the management of hypertension. PLoS One 2018;13(5):e0196868 [FREE Full text] [doi: 10.1371/journal.pone.0196868] [Medline: 29723262]
67. Essery R, Pollet S, Smith KA, Mowbray F, Slodkowska-Barabasz J, Denison-Day J, et al. Planning and optimising a digital intervention to protect older adults' cognitive health. Pilot Feasibility Stud 2021 Aug 18;7(1):158 [FREE Full text] [doi: 10.1186/s40814-021-00884-2] [Medline: 34407886]
68. Hui SK, Grandner MA. Trouble sleeping associated with lower work performance and greater health care costs: longitudinal data from Kansas State Employee Wellness Program. J Occup Environ Med 2015 Oct;57(10):1031-1038 [FREE Full text] [doi: 10.1097/JOM.0000000000000534] [Medline: 26461857]
69. Condon L, Bedford H, Ireland L, Kerr S, Mytton J, Richardson Z, et al. Engaging Gypsy, Roma, and Traveller communities in research: maximizing opportunities and overcoming challenges. Qual Health Res 2019 Jul;29(9):1324-1333 [FREE Full text] [doi: 10.1177/1049732318813558] [Medline: 30600758]
70. Hilliard F, Goldstein E, Nervik K, Croes K, Ossorio PN, Zgierska AE. Voices of women with lived experience of substance use during pregnancy: a qualitative study of motivators and barriers to recruitment and retention in research. Fam Community Health 2023;46(1):1-12 [FREE Full text] [doi: 10.1097/FCH.0000000000000349] [Medline: 36383229]
71. Mitchell JA, Morales KH, Williamson AA, Huffnagle N, Eck C, Jawahar A, et al. Engineering a mobile platform to promote sleep in the pediatric primary care setting. Sleep Adv 2021;2(1):zpab006 [FREE Full text] [doi: 10.1093/sleepadvances/zpab006] [Medline: 33981997]
72. Nour MM, Rouf AS, Allman-Farinelli M. Exploring young adult perspectives on the use of gamification and social media in a smartphone platform for improving vegetable intake. Appetite 2018 Jan 01;120:547-556. [doi: 10.1016/j.appet.2017.10.016] [Medline: 29032184]
73. Beltzer ML, Meyerhoff J, Popowski SA, Mohr DC, Kornfield R. Mental health self-tracking preferences of young adults with depression and anxiety not engaged in treatment: qualitative analysis. JMIR Form Res 2023 Oct 06;7:e48152 [FREE Full text] [doi: 10.2196/48152] [Medline: 37801349]
74. Levenson JC, Ford HA, Reyes ZM, Mukundan A, Patel G, Bahary S, et al. Designing adolescent sleep interventions with stakeholder input. Sleep Health 2021 Oct;7(5):581-587 [FREE Full text] [doi: 10.1016/j.sleh.2021.07.007] [Medline: 34474988]
75. Palermo TM, Bromberg MH, Beals-Erickson SE, Law EF, Durkin L, Noel M, et al. Development and initial feasibility testing of brief cognitive-behavioral therapy for insomnia in adolescents with comorbid conditions. Clin Pract Pediatr Psychol 2016 Jun;4(2):214-226. [doi: 10.1037/cpp0000140]

Abbreviations

CBTi: cognitive behavioral therapy for insomnia
DOZE: Delivering Online Zzz’s With Empirical Support
PPI: patient and public involvement
YPAG: Young People’s Advisory Group

Edited by C Jacob; submitted 24.06.24; peer-reviewed by K Coteur; comments to author 29.08.24; revised version received 17.10.24; accepted 21.10.24; published 31.10.24.

Please cite as:
Bennett SE, Johnston MH, Treneman-Evans G, Denison-Day J, Duffy A, Brigden A, Kuberka P, Christoforou N, Ritterband L, Koh J, Meadows R, Alamoudi D, Nabney I, Yardley L
Using the Person-Based Approach to Co-Create and Optimize an App-Based Intervention to Support Better Sleep for Adolescents in the United Kingdom: Mixed Methods Study
JMIR Hum Factors 2024;11:e63341
URL: https://humanfactors.jmir.org/2024/1/e63341
doi:10.2196/63341
PMID:

Original Paper

A New Research Model for Artificial Intelligence–Based Well-Being Chatbot Engagement: Survey Study

Yanrong Yang[1*], MSc; Jorge Tavares[1*], PhD; Tiago Oliveira[1*], PhD

NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Lisboa, Portugal

[*]all authors contributed equally

Corresponding Author:
Yanrong Yang, MSc
NOVA Information Management School (NOVA IMS)
Universidade Nova de Lisboa
Campus de Campolide
Lisboa, 1070-312
Portugal
Phone: 351 213 828 610
Email: 20220021@NOVAIMS.UNL.PT

Abstract

Background: Artificial intelligence (AI)–based chatbots have emerged as potential tools to assist individuals in reducing anxiety and supporting well-being.

Objective: This study aimed to identify the factors that impact individuals' intention to engage and their engagement behavior with AI-based well-being chatbots by using a novel research model to enhance service levels, thereby improving user experience and mental health intervention effectiveness.

Methods: We conducted a web-based questionnaire survey of adult users of well-being chatbots in China via social media. Our survey collected demographic data, as well as a range of measures to assess relevant theoretical factors. Finally, 256 valid responses were obtained. The newly applied model was validated through the partial least squares structural equation modeling approach.

Results: The model explained 62.8% (R^2) of the variance in intention to engage and 74% (R^2) of the variance in engagement behavior. Affect (β=.201; *P*=.002), social factors (β=.184; *P*=.007), and compatibility (β=.149; *P*=.03) were statistically significant for the intention to engage. Habit (β=.154; *P*=.01), trust (β=.253; *P*<.001), and intention to engage (β=.464; *P*<.001) were statistically significant for engagement behavior.

Conclusions: The new extended model provides a theoretical basis for studying users' AI-based chatbot engagement behavior. This study highlights practical points for developers of AI-based well-being chatbots. It also highlights the importance of AI-based well-being chatbots to create an emotional connection with the users.

(JMIR Hum Factors 2024;11:e59908) doi:10.2196/59908

KEYWORDS

artificial intelligence–based chatbot; AI-based chatbot; mental well-being; intention to engage; engagement behavior; theoretical models; mobile phone

Introduction

Overview

According to the World Health Organization (2019), >80% of people worldwide face challenges in accessing mental health services [1]. This lack of access can be attributed to various factors, such as inadequate attention to health care, limited availability of medical resources, and the inability to afford the high costs of treatment [1,2]. Accessibility and scalability of mental health services need to be addressed [3].

An artificial intelligence (AI)–based well-being chatbot can engage in conversations with humans in a relatively natural manner, offering companionship, emotional support, and guidance for emotional well-being [4]. Therapeutic well-being chatbots work by simulating how a mental health professional would treat a user [5], and companionship well-being chatbots facilitate or develop a social relationship with the user through chatting to alleviate and channel negative emotions, such as loneliness and irritability [6,7]. These new digital interventions provide considerable relief to individuals who need psychotherapeutic help but are plagued by a lack of time, space, or resources to access it [8-10]. Individuals who have interacted with these chatbots have expressed satisfaction with their experiences and have shown a positive attitude toward the future development of this technology [11]. These chatbots allow users to discuss private topics anonymously, effectively avoiding any feelings of shyness that may arise [12]. Well-being chatbots have also been used by professionals as an effective complementary tool to traditional face-to-face therapy [13,14]. In addition, they contribute positively to the dissemination of mental health knowledge and the promotion of healthy behaviors [11,15].

A growing number of research findings support the idea that digital mental health interventions, for instance, well-being chatbots, reduce the risk of chronic diseases by improving patients' psychosocial well-being and promoting other health behaviors [16-18]. They can help users overcome barriers to mental health support, and users can anonymously accept help from chatbots [19-21]. Scholars have taken notice of this phenomenon, and chatbot effectiveness, software design and development, use, and user satisfaction are being emphasized [22,23]. However, the problem of low engagement and high dropout rates between users and chatbots have not been prioritized, particularly in studying engagement behaviors through theoretical models. This will severely influence the user experience and effectiveness [3,8]. Exploring the factors influencing users' engagement behavior with well-being chatbots is critical to comprehend and refine this association, to serve users better [8,24].

This study aimed to investigate user intention to engage with well-being chatbots and engagement behavior by developing a new theoretical model that combines the theory of interpersonal behavior (TIB), diffusion of innovation (DOI), and trust. The goal is to understand the relationships among various factors and analyze their impact on the intention to engage and engagement behavior. We gathered data through a web-based survey to examine this model and identify the relationships between different factors. This research contributes to expanding the existing knowledge on theoretical models, particularly in the context of a human-centered digital mental health intervention. In addition, it will assist in designing, developing, and improving user-centered well-being chatbots; alleviating the problem of mental health medical resources; and helping to improve the overall well-being of the population. We have two research questions related to the objective of this study: (1) What factors influence users engaged with AI-based well-being chatbots? (2) How could AI-based well-being chatbot service be improved using the results of this study to improve users' engagement and experience?

Theoretical Background Rationale

Published studies about adoption of AI-based well-being chatbots tend to focus on either emotional or technical components of this technology but not on a more integrated approach to study this new technology [25-30]. Particularly in digital health adoption, the most used theories, the technology acceptance model and unified theory of acceptance and use of technology, mostly focus on general technology adoption drivers [31,32]. Explaining the interaction of AI-based well-being chatbots with users goes beyond a simple technical interaction, it has been documented that they can create a psychological connection, like a friendship [33,34]. Therefore, we use the TIB, specifically its affect construct, to understand the relationship between a user and an AI-based well-being chatbot [29,30]. AI-based well-being chatbots are innovative technologies in the field of mental health care and personal well-being, and the application of DOI theory is beneficial for studying the factors that contribute to the adoption of AI-based well-being chatbots [35]. Trust is a key factor, particularly when dealing with personal and sensitive data [36,37], like the sharing process between the user and AI-based chatbots when it concerns mental health and personal well-being [38]. Without trust in the treatment intervention, the expected health outcomes between both parties may not be achieved [39]. The study brings these theories together through a new approach that combines relevant phycological factors for the adoption of AI-based well-being chatbots, which can be measured with the TIB and trust theory and the technical and innovation component of this new technology, which can be measured with the DOI theory.

Engagement Behavior in Digital Mental Health Intervention

Mental well-being is an increasingly important health topic of public concern. AI-based chatbots empower mental health and well-being through AI technology to provide emotional support to human beings [40]. AI-based chatbots enable user interaction based on text or voice support and complete corresponding tasks, recognizing users' emotions and providing solutions [41]. The services of AI-based chatbots for mental well-being as a new digital mental health intervention to users are evolving, and it is crucial to study users' engagement behavior.

Engagement is a multidimensional concept that includes not only the formation of interest or adherence to a predefined plan, but also the development of trust, integration, and ongoing participation [42]. In this study, engagement behaviors are defined as the behaviors of users interacting with a well-being chatbot. The well-being chatbot serves as a new type of digital health intervention that provides users with mental health self-management and psychotherapy services [42]. Users' engagement is an important factor, influencing the effectiveness of mental health interventions [43]. Research has shown that high engagement is associated with high intervention effectiveness [29,30,44,45]. In mental health treatment, participation in ongoing treatment is necessary for recovery [46,47]. A study of a digital mental health intervention found that >70% of users failed to complete all treatment modules and >50% withdrew before completing all treatment modules in general [29,48]. An analysis of mental health applications use showed that the average 15-day retention rate was only 3.9% [49]. Another study showed that mobile apps that emphasized user participation in design increased the effectiveness of interventions for depression and anxiety [50]. In a meta-analysis study of the impact of digital mental health engagement on mental health outcomes, users with higher levels of access showed substantial or moderate improvements in postintervention mental health outcomes [29]. This study explores the relationship between factors around engagement behaviors. This will help to uncover the insight of users' willingness to engage and their engagement behaviors and improve the design capabilities and services of well-being chatbot.

TIB Theory

TIB was developed by Triandis [51] in 1977. It is similar to the ABC (attitude-behavior-context) model by Stern [52], combining internal and external factors, including affect, social factors, perceived consequences, habit, and facilitating conditions, to understand intended behavior [53]. In the context of engagement with well-being chatbot research, TIB is a well-suited theoretical model because well-being chatbots operate in a way similar to social software, where communication with users is accomplished through text dialogue and voice dialogue [54]. Users communicate trial experiences and results, and even recommend an AI-based chatbot to others [55]. The affect factor can seriously impact an individual's willingness to communicate [56]. People will recommend their favorite products to each other, and this recommendation behavior will influence the individual's intention [57]. Individuals past communication habit of using mobile apps will influence their willingness to use them [58]. If individuals frequently use instant messaging apps, they will be accustomed to this online communication method. TIB contains the above 3 critical aspects known as affect, habit, and social factors. Therefore, TIB is chosen as a theoretical basis for our model.

DOI Theory

DOI describes the process by which people embrace new ideas, use new products, and engage in new practices [59]. In general, only a few people have an attitude of developmental acceptance of new ideas and are willing to try them out and embrace them in the initial stages. As these people propagate them, gradually, more people begin to embrace them; the innovative idea or product thus diffuses through the population and eventually reaches saturation [60].

DOI was proposed by Rogers [59] in 1995, it helps us to understand the characteristics of an innovation and what attracts users to it. According to Rogers' research, 5 key features influence the adoption of an innovation: relative advantages, complexity, compatibility, observability, and trialability. Well-being chatbots are an innovative technology that has emerged in recent years, but they are not widespread in daily life. As their contribution to the mental health field, studying their dissemination among people leads to its understanding and acceptance by more people can contribute to human health and well-being. Therefore, extending the TIB model by adopting the properties of DOI is crucial. Among the 5 characteristics, observability can be considered equivalent to the combined effect of demonstrability and visibility [61]. Visibility was not used in this study because AI-based chatbot engagement was treated as a personal experience. Still, results demonstrability was used in our research model. Trialability was also not adopted because there was no evidence of whether the user had trialed a chatbot.

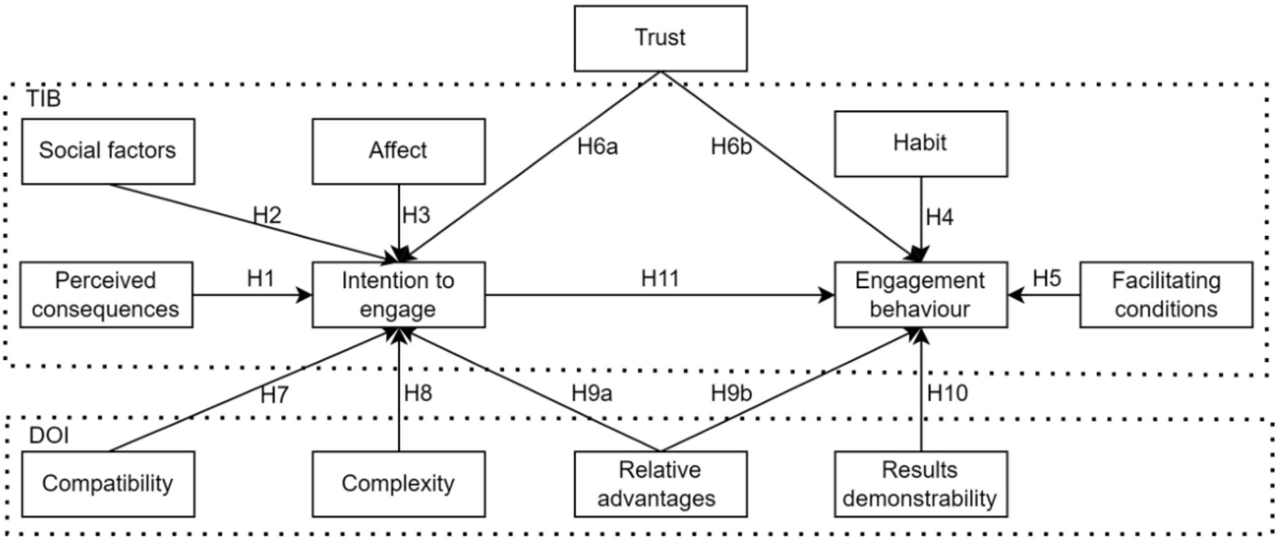
Trust

Trust is defined as the willingness of one party to accept the actions of another party, irrespective of the latter's ability to control them [62]. The trust placed in machines is determined as the willingness of users to accept the information generated by machines and to adhere to their recommendations [63]. This indicates that one party intents to form a relatively secure attachment to the other, despite the potential for negative outcomes [64]. This represents a psychological mechanism that can reduce uncertainty and increase the likelihood of successful interaction with other entities within the environment [65]. Trust is a prerequisite for any social interaction and is instrumental in fostering collaboration and cooperation between individuals [66]. It serves as a key factor in successful transactions and establishing long-term relationships [67]. In the field of mental health, trust is of paramount importance in the relationship between patients and health care professionals [68,69]. The interactive behavior of users with chatbots for health purposes is analogous to that of patients with their doctors; the establishment of collaborative and cooperative relationships based on trust is conducive to the achievement of health objectives [70,71]. Therefore, trust represents a crucial element in the investigation of engagement with chatbots.

Research Model and Hypotheses

Following the theoretical rationale, TIB, DOI, and trust were combined to support the understanding of users' intention to engage with AI-based well-being chatbots and engagement behavior in our research, as shown in Figure 1.

Figure 1. Research model. DOI: diffusion of innovation; H: hypothesis; TIB: theory of interpersonal behavior.



Development of Hypotheses

Perceived consequences are the positive or negative results of an action after it has occurred and the possibility of the outcome occurring [72]. When the perceived consequences are positive, the individual will be prompted to engage with the behavior to achieve advantages; when the perceived consequences are negative, the individual's willingness to engage with the behavior will be reduced [53]. It had been validated to have a relevant effect on behavioral intention [53,73]. Well-being chatbots offer mental companionship and emotional support, contributing positively to users' emotional and mental well-being [9]. Therefore, we assume that perceived consequences will be positively related to the intention to engage with well-being chatbots (hypothesis 1).

Social factors are related to the extent to which people are influenced by others who are significant to them [51,73]. Individuals in a group or those observed by a group will comply with some of the unwritten rules within the group, and the likelihood that an individual will act in accordance with the group's demands increases under the pressure of the group [51,74,75]. Social factors have been shown in several studies to positively influence individual behavioral intention [73,76-78]. In health care, the influence of social factors was examined and affirmed from multiple perspectives in a study of clinicians' adoption of mHealth (mobile health) tools [79]. In chatbots that provide services in a social context, users decisions are influenced by perceptions of how those around them use these services [78]. Regarding the context of a well-being chatbot engagement, we hypothesize that social factors will be positively related to the intention to engage with well-being chatbots (hypothesis 2).

Affect is used to describe the mental representation of internal bodily sensations associated with emotions, behaviors, or personality tendencies [80]. It is the purely emotional part of an individual's attitude and contains positive or negative emotions, for instance excitement, joy, depression, and displeasure [81,82]. Affect has been shown to have an influence on behavioral intention in studies on information technology applications [76,83,84]. In the context of AI-based well-being chatbot engagement research, we assume that affect will positively influence individuals' intention to engage with AI-based well-being chatbots. Thus, we propose that affect will be positively related to the intention to engage with well-being chatbots (hypothesis 3).

Habit is a learned behavior, an automatic response to a steady stream of contextual cues [85], and it is regarded as a major influence on behavior [86]. A study has shown that an individual's habits can predict future behavior to some extent [87]. Because the popularity of the internet as well as smartphones and the effectiveness of using digital interventions for health behaviors have been proven [51,53,54], this has caused health care apps to gradually become a way to optimize people's daily health care behavioral habits [88]. As the well-being chatbot serves as a health care information system, we assume that habit will be positively related to engagement behavior (hypothesis 4).

"Facilitating conditions" is a term that refers to objective elements in the environment that enable the easy execution of behavior [51]. In the IT context, it is defined as the resources necessary to support the use of a system, such as access to the internet or a smartphone [31]. Facilitating conditions have been identified as a key factor which influenced individuals' behavior related to engagement [83]. Thus, we hypothesize that facilitating conditions will be positively related to engagement behavior (hypothesis 5).

Trust has been recognized as one of the critical factors in human-robot interaction research [89,90]. Users' trust in AI-based chatbots is based on the AI-based chatbot's performance and services being dependable, trustworthy, and being able to assist in achieving the user's intended purpose [37]. Developing and nurturing trust in the psychotherapy process to establish a good therapeutic relationship through engagement and ultimately effective treatment is crucial [91,92]. Meanwhile, trust was identified to have a major influence on the intention to act on eHealth websites [93]. Accordingly, trust influences users' willingness to intent and engage with AI-based well-being chatbots [91-93]. So, we assume that trust positively influences an individual's intention to engage AI-based well-being chatbots (hypothesis 6a) and trust positively influences users' engagement behavior (hypothesis 6b).

Compatibility refers to the extent to which the innovation matches the existing values and beliefs, previous experiences, and demands of potential users [59,94]. It provides a good indicator of how extensively an innovation complies with potential users' lifestyles, needs, and preferences [60]. In previous research, compatibility was identified as one that influenced the intention to behavior [95]. Well-being chatbots meet the real-time needs of users [96-98], and chatbot mobile apps match the habits of smartphone users [99]. In this research, we assume that compatibility will be positively related to the intention to engage with AI-based well-being chatbots (hypothesis 7).

Complexity is a measure of how difficult it is to understand and use an innovation [59]. It is a systematic form that is associated with almost all aspects of health care [100]. Complexity has been proven to have an impact on digital technology in health and well-being apps [101]. In another study on health care chatbots, complexity had a strong impact on the ability of chatbots to successfully provide health information and adoption behavior [100,102,103]. In this research, we assume that low complexity will be positively related to the intention to engage with well-being chatbots (hypothesis 8).

Relative advantages is a term that refers to the degree to which an innovation is better than the object it replaces [94]. Innovation with greater relative advantage is beneficial for its diffusion [104]. It has been shown that an innovation will not be used if potential users believe that there is no comparative advantage in the adoption of the innovation over its earlier counterparts [105]. AI-based well-being chatbots are more empathetic than their earlier counterparts and even have memory functions, these advantages motivate users to interact and engage with them more [96,106-110]. Thus, we assume that relative advantages will be positively related to the intention to engage with AI-based well-being chatbots (hypothesis 9a) and relative advantages will be positively related to engagement behavior (hypothesis 9b).

Results demonstrability is the degree to which innovative results are presented and disseminated [61]. Innovations will be more adopted if they generate demonstrably positive results; if the converse is the case, the chances of the innovation being adopted become lower [94]. Studies have shown that results demonstrability is a potential predictor of behavioral adoption [111]. AI-based well-being chatbots can serve users as an mHealth app. Thus, we assume that results demonstrability will be positively related to the intention to engage with AI-based well-being chatbot engagement (hypothesis 10).

Intention to engage in a behavior is the most direct determinant of an individual's behavior [112]. Exploring the relationship between intention to engage and engagement behavior helps to improve user experience and interaction effectiveness [9,113]. This also has a positive effect on the design of well-being chatbots in terms of enhancing user engagement [114,115]. Therefore, the intention to engage influences engagement and is an important factor in the study of user engagement with well-being chatbots. We assume that intention to engage with AI-based well-being chatbots will be positively related to engagement behavior (hypothesis 11).

Age, gender, education and chronic disease status were implemented in the research model as control variables [116].

Methods

Ethical Considerations

Approval was obtained from the NOVA Information Management School Ethics Committee, NOVA University of Lisbon (INFSYS2023-5-257970). The procedures used in this study adhere to the tenets of the Declaration of Helsinki. All participants were aged at least 18 years, and informed consent was obtained from them. All data were collected anonymously, and participants were not compensated.

Data Collection and Sample

The questionnaire was developed in English on the Qualtrics platform. The survey was designed per the guidelines and the Checklist for Reporting Results of Internet E-Surveys (CHERRIES), which is presented in Multimedia Appendix 1 [117]. We explained to participants that participation was voluntary, and their data would be collected anonymously. We took measures to ensure that participants clearly understood what a well-being chatbot entails by introducing its concept and benefits at the start of the survey. Meanwhile, we described the functionality and use of an AI-based well-being chatbot. We engaged 2 experts and 2 colleagues to review and evaluate the questions to ensure that the topics were clear, relevant to the subject matter, and easy to understand. Once the questionnaire was finalized, a translator translated the questions into Chinese. Then, another translator was responsible for doing a back-translation and comparing it with the original English version to ensure accuracy [118]. Then, 40 participants were selected for pretesting to validate the questions' understandability and the survey scale. No issues were reported that could indicate that the survey items were unreliable. Action was taken to prevent potential issues with single source and common source bias. The questionnaire was placed in 3 different web platforms to ensure the maximum coverage and avoid a single-source bias [117,119,120].

We distributed the survey on 3 popular social media mobile apps: WeChat, Weibo, and Douban. WeChat is China's most popular social media network, with 1.3 billion active users in 2022 [121]. Weibo is China's second-largest social platform after WeChat, with 582 million active users at the end of the first quarter of 2022 [122]. Douban is an interest-oriented social network community with 75 million users as of 2020 [123,124]. Publishing the questionnaire across the 3 social media platforms will ensure fair data collection.

The framework's independent and dependent variable items were collected in a single questionnaire. We assessed if there was a clear understanding of what was being measured by the constructs, to avoid the risk of common source bias [119,120]. The aim was for the respondents to avoid using the same mental process or heuristics when replying to questions about different constructs [119,120,125]. The assessment of our pilot survey was that there was no reason for concern. For added precaution, additional features were incorporated in the final survey to enable psychological separation. While designing a survey, psychological separation should ensure that the measures of the different constructs are unrelated [120]. Different instructions for different sections of the survey were provided, and the sections of the survey that measure different constructs were physically separated [120].

Finally, 256 valid replies from well-being chatbot users were collected from May to October 2023. The web-based survey did not impose any restrictions on participants other being an adult aged ≥18 years.

Measurement

The scales of all the variables in this study were produced concerning the relevant literature. Minor modifications were carried out according to the characteristics of AI-based chatbots. We used a 7-point scale to assess the variables from 1="strongly disagree" to 7="strongly agree." The questionnaire with the measurement items and references for each variable are provided in Multimedia Appendix 2.

Data Analysis

The data were analyzed using the partial least squares structural equation modeling (PLS-SEM) approach using Smart-PLS (version 4.0) [126], which is suitable for analyzing and predicting complex models and nonnormally distributed data. PLS-SEM can also handle models that include both reflective and formative variables [127].

Reflective and formative construct measurements were included in the research. In reflective measurement models, causality flows from the underlying construct to the indicator. In contrast, in formative measurement models, causality flows in the opposite direction, from indicators to constructs [128]. Reflective constructs measure entities with a series of positively correlated items [129,130]. In contrast, the formative construct is a singular construct which is constituted by the aggregation of multiple indicators without any a priori assumptions regarding the interrelationships between these elements [129,130]. Reflective and formative measurement models should be evaluated separately [128].

Results

Sample Characteristics

Of the 256 valid samples, all had experience in using AI-based chatbot for mental health care. The participants' average age was 30.9 years, and 55.9% of participants were younger than 30 years. The average age in other studies in China with the same scope as this study has ranged between 21 and 34.8 years [25-28]. The high proportion of young women was also present in demographic data from other studies, particularly studies on health technology adoption behaviors [109,131,132]. A recent Chinese study from 2023 showed that 77% of users of digital mental health technologies in China were female [28], which aligns with our study participants' demographics. Approximately 91% of the participants held higher education degrees, which is more prevalent in innovation technology adoption studies [133,134]. However, the number of participants with chronic diseases was close to that of those without any disease, which is also reflected in the results of previous studies, which have found that chronic diseases have an impact on health applications [116]. The sample characteristics are shown in Table 1.

Table 1. Demographic data (n=256).

Characteristics	Participants, n (%)
Age (years)	
18-29	143 (56)
30-44	84 (33)
45-59	13 (5)
≥60	16 (6)
Gender	
Women	187 (73)
Men	69 (27)
University education	
Degree	234 (91)
No degree	22 (9)
Chronic disease status	
Yes	111 (43)
No	145 (57)

Measurement Model

Formative and reflective constructs were included in our model. They were measured separately. First, for reflective constructs, the construct items' reliability was assessed by computing the value of each item. The loading values of all reflective construct items were above the threshold of 0.7, and they were accepted [135] and are listed in Multimedia Appendix 3. Then, we applied the Cronbach α reliability coefficient and composite reliability (CR) to measure their internal consistency. All Cronbach α and CR scores were above 0.7, and the model was proven to have good reliability [135]. Meanwhile, we examined convergence validity by assessing the average variance extracted (AVE); the value of AVE for each construct was >0.5 [135]. All detailed indicators for mean, SD, Cronbach α, CR, and AVE are shown in Table 2. We used the heterotrait-monotrait (HTMT) ratio as the main criterion to assess discriminant validity, following the latest guidelines recommendation [135-137]. The HTMT values were below the threshold value of 0.90 [135], thus confirming discriminant validity. The results are shown in Table 2. In addition, cross-loadings and the Fornell-Larcker criterion were also evaluated for discriminant validity, and the results also confirm discriminant validity (Multimedia Appendices 3 and 4).

Table 2. Indicators of reflective constructs.

Construct	Values, mean (SD)	Cronbach α	CR[a]	AVE[b]	HTMT[c] values										
					PC[d]	Affect	Habit	FC[e]	Trust	Compatibility	Complexity	RA[f]	RD[g]	ITE[h]	EB[i]
PC	4.726 (1.306)	.884	.885	.742	—[j]	—	—	—	—	—	—	—	—	—	—
Affect	4.638 (1.533)	.939	.942	.890	.641	—	—	—	—	—	—	—	—	—	—
Habit	4.333 (1.329)	.822	.828	.651	.720	.695	—	—	—	—	—	—	—	—	—
FC	4.770 (1.269)	.846	.846	.684	.684	.602	.573	—	—	—	—	—	—	—	—
Trust	4.461 (1.280)	.893	.896	.703	.839	.702	.768	.614	—	—	—	—	—	—	—
Compatibility	4.508 (1.344)	.851	.853	.771	.725	.746	.732	.666	.767	—	—	—	—	—	—
Complexity	4.921 (1.292)	.896	.900	.763	.690	.679	.648	.811	.697	.791	—	—	—	—	—
RA	4.663 (1.242)	.849	.855	.689	.816	.707	.768	.676	.842	.805	.752	—	—	—	—
RD	4.620 (1.362)	.856	.858	.777	.691	.659	.659	.685	.753	.843	.837	.836	—	—	—
ITE	4.575 (1.415)	.894	.894	.825	.728	.726	.776	.771	.759	.766	.701	.766	.720	—	—
EB	4.567 (1.362)	.896	.896	.763	.779	.723	.803	.653	.827	.758	.662	.760	.736	.897	—

[a]CR: composite reliability.

[b]AVE: average variance extracted.

[c]HTMT: heterotrait-monotrait.

[d]PC: perceived consequences.

[e]FC: facilitating conditions.

[f]RA: relative advantages.

[g]RD: results demonstrability.

[h]ITE: intention to engage.

[i]EB: engagement behavior.

[j]Not applicable.

Social factors (SF) were measured as a formative construct in our research model [53]. We assessed the collinearity among indicators of the formative construct by calculating the variance inflation factor (VIF). The VIF values (Table 3) were below the cutoff value of 5 [135], which meant that there was no collinearity. Finally, we implemented a bootstrapping approach with 5000 resamples for identifying the statistical significance of each path. Social factor (SF) 1 and SF2 present statistically significant outer weights; SF3, SF4 and SF5 did not present statistically significant outer weights. Thus, we verified the SF3, SF4 and SF5 outer loading values, which were all >0.5 [135]. This aspect means that all SF items were relevant.

Table 3. Indicators of formative construct.

SF[a]	VIF[b]	Outer weights	*P* values (outer weights)	Outer loadings	*P* values (outer loadings)
SF1	2.386	.224	.04	.826	<.001
SF2	2.411	.421	<.001	.881	<.001
SF3	2.441	.253	.07	.834	<.001
SF4	2.841	.111	.37	.781	<.001
SF5	2.520	.198	.13	.738	<.001

[a]SF: social factor.

[b]VIF: variance inflation factor.

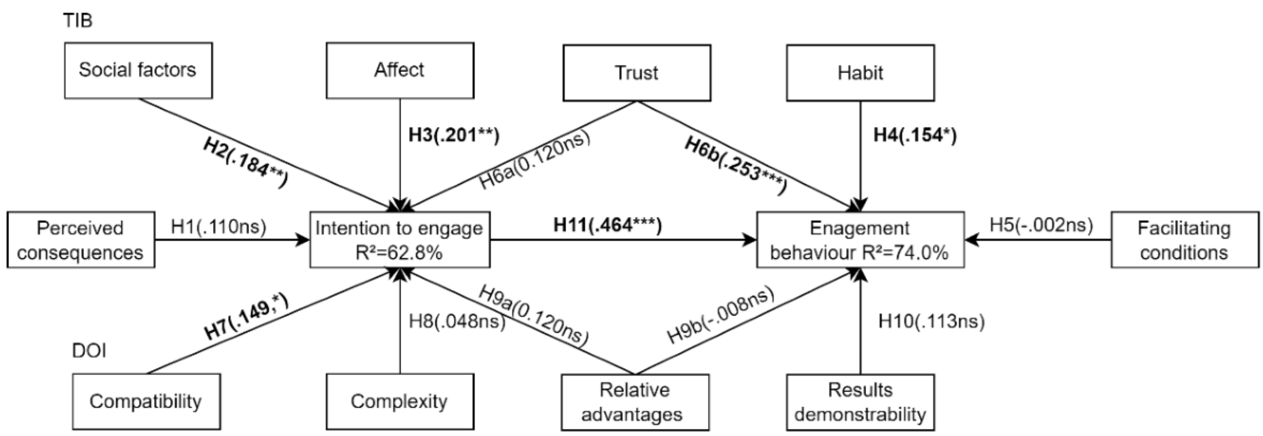
In addition, we used Harman 1-factor test method to probe for common method variance (CMV).

The total variance extracted by 1 factor was less than the recommended threshold of 50%. Hence, this data should not present any problem with CMV [120]. Afterward, the marker variable technique was adopted to assess the CMV, and an unrelated construct was defined as a marker variable to determine the relationship between it and each construct in the research model [125]. We obtained 0.055 (5.5%) as the maximum shared variance with other variables. Therefore, the value can be considered low [138]. After verification by 2 methods, it was concluded that the influence of CMV can be excluded from this study [120,125].

Structural Model

The structural model explained 62.8% variance in intention to engage, and 74% of variance in engagement behavior. Both R^2 are regarded as high by the literature [135]. High R^2 values indicate that our key target variables can be well predicted via the PLS path model [135]. Figure 2 shows the structural model results and identifies which latent variables are statistically significant.

Figure 2. Structural model results. DOI: diffusion of innovation; H: hypothesis; ns: nonsignificant; TIB: theory of interpersonal behavior. **P*<.05; ***P*<.01; ****P*<.001.



Regarding the intention to engage, affect (β=.201; *P*=.002), SF (β=.184; *P*=.007), and compatibility (β=.149; *P*=.03) were statistically significant. Hypotheses 2, 3, and 7 were supported. Perceived consequences (β=.110; *P*=.12), trust (β=.120; *P*=.11), complexity (β=.048; *P*=.47), and relative advantages (β=.120; *P*=.16) were not statistically significant; hence hypotheses 1, 6a, 8, and 9a were not supported.

About engagement behavior, intention to engage (β=.464, *P*<.001), habit (β=.154; *P*=.01) and trust (β=.253, *P*<.001) were statistically significant, and facilitating conditions (β=−.002; *P*=.98), relative advantages (β=−.008; *P*=.91), and results demonstrability (β=.113; *P*=.10) were not statistically significant. Hypotheses 11, 6b, and 4 were supported, and hypotheses 5, 9b, and 10 were rejected, as shown in Table 4.

Table 4. Hypothesized path analysis.

Hypothesis	Path	β	*P* values	Supported
H1	Perceived consequences → intention to engage	.110	.12	No
H2	Social factors → intention to engage	.184	.007	Yes
H3	Affect → intention to engage	.201	.002	Yes
H4	Habit → engagement behavior	.154	.01	Yes
H5	Facilitating conditions → engagement behavior	−.002	.98	No
H6a	Trust → intention to engage	.120	.11	No
H6b	Trust → engagement behavior	.253	<.001	Yes
H7	Compatibility → intention to engage	.149	.03	Yes
H8	Complexity → intention to engage	.048	.47	No
H9a	Relative advantages → intention to engage	.120	.16	No
H9b	Relative advantages → engagement behavior	−.008	.91	No
H10	Results demonstrability → engagement behavior	.113	.10	No
H11	Intention to engage → engagement behavior	.464	<.001	Yes

The PLSpredict algorithm was used to assess the framework predictive power. The method uses training and hold out samples to generate and evaluate predictions from PLS path model estimations [139,140]. The guideline recommendation was followed, and the number of folds was set to 10 [140]. This approach was done because it is possible to achieve a statistical power of 80% to detect minimum R^2 values of 0.1 in the endogenous constructs in the structural model for a significance level of 1% [137,140]. The first parameter to be evaluated was the $Q^2_{predict}$ of the indicators concerning our endogenous variables that was above 0, showing that the model demonstrates predictive power [140]. To evaluate the predictive magnitude, we compared the PLS-SEM study model with the naive linear regression model (LM) to see if it can outperform the LM benchmark [140]. Because the prediction errors distribution was considerably asymmetrical, with high kurtosis values (>1) [135], the mean absolute error was the more appropriate prediction statistic [140].

The PLS-SEM analysis yielded lower prediction errors for most of the dependent variables' indicators, as seen in Table 5; this indicates a medium predictive power for the study model [140]. When complex models are used, involving several theories, such as those explaining human behavior, R^2 values higher than 0.5 can be regarded as substantial [135,137]. The model in this research study is complex, and achieving medium predictive power is challenging in such models [137,140]. Given that this research model shows a substantial R^2 for a complex model and medium predictive power, it provides confidence in its use for real-world applications [135,137,140].

Table 5. Prediction summary.

Indicators	$Q^2_{predict}$	PLS_SEM_RMSE[a]	PLS_SEM_MAE[b]	LM_RMSE[c]	LM_MAE[d]	Kurtosis	Skewness
EB1[e]	.431	1.165	.844	1.149	.810	2.029	−.542
EB2	.461	1.157	*.863* [f]	1.196	.886	.673	−.048
EB3	.490	1.114	*.811*	1.230	.913	1.550	−.152
EB4	.557	1.057	*.776*	1.157	.831	2.030	−.413
ITE1[g]	.485	1.128	.835	1.071	.785	2.374	−.461
ITE2	.472	1.135	*.814*	1.167	.869	2.782	−.740
ITE3	.452	1.156	*.861*	1.189	.884	1.862	−.627

[a]PLS_SEM_RMSE: partial least squares structural equation modeling root mean squared error.

[b]PLS_SEM_MAE: partial least squares structural equation modeling mean absolute error.

[c]LM_RMSE: linear regression model root mean squared error.

[d]LM_MAE: linear regression model mean absolute error.

[e]EB: engagement behavior.

[f]Most relevant errors to define the model predictive power are highlighted in italics.

[g]ITE: intention to engage.

Discussion

Principal Findings

The TIB was considered in this study, and it was extended with the DOI theory and trust to explore in depth the factors influencing users' engagement behavior with AI-based chatbots for well-being. The new extended model was well explained, with R^2 values of 62.8% for the variance in intention to engage and 74% for the variance in engagement behavior. RQ1 was answered in this study. AI-based well-being chatbots are gradually beginning to play an active role in the field of mental health. In this context, this study obtained important results, including affect, habit, SF, trust, compatibility and intention to engage as determinants influencing users' engagement behavior. By extending the TIB to include important research variables in the model, a foundation was laid for future, related theoretical research and practice. The study's results affirmed the significance of affect on users' intention to engage. Our measurement of user affect contains both positive and negative emotions [141]. From the user's perspective, there is a willingness to engage with the chatbot. It indicates that users have a positive impression of the well-being chatbot's service and that users are inclined to deal with their emotions by engaging with a well-being chatbot. Users are beginning to be comfortable with the service as a mental health intervention that is available anytime, anywhere, without an appointment [142]. From the perspective of well-being chatbot characteristics, empathetic chatbots can understand users' emotions and provide professional psychological counseling or companionship [96]; for instance, suggesting meditation, outdoor activities, or socializing with friends [7,143,144]. This proactive intervention on users' emotions promotes users' willingness to participate with sustained engagement behaviors.

Habit has been extensively studied in previous research on health app use and engagement behavior [145]. In the current information age, smartphone use has become ingrained in people's daily lives. Engaging with mobile apps for specific purposes has become a habit for many individuals [146]. This study confirmed that habits were an essential factor influencing users' engagement behavior, aligning with other studies' findings [147]. In addition, the significant influence of SF indicated that the opinions of friends, family, and medical professionals substantially impacted individuals' intention to engage with well-being chatbots. This finding is consistent with previous studies [78].

Furthermore, as AI-based well-being chatbots represent an innovative technology in a digital mental health intervention, proactively exploring methods for promoting their adoption is worthwhile. This research model incorporates features from the DOI framework to extend the TIB model. Among these features, compatibility was found to impact the intention to engage significantly, while the other 3 aspects did not demonstrate statistical significance. Because chatbots work like any other social software, the experience is consistent with past experiences [148,149]. They can be used easily and without extra effort, making users more willing to use them, which is also consistent with previous research [150].

Trust, habit, and intention to engage were statistically significant for chatbot engagement. They are both internal (subjective) elements of individuals [151,152], whereas other factors had no statistically significant influence, such as facilitating conditions, relative advantages, and results demonstrability. It indicated that the user's subjective sense of experience played a decisive role in engagement compared with other factors [153,154]. This suggests that a lack of user-centeredness in product design or a lack of information about mental health services in terms of content that meets users' needs could directly reduce user engagement [155]. Consistent with the findings of this study, it further underlines the importance of user-centeredness.

Potential explanations for the constructs that were nonsignificant are also addressed. The facilitating conditions hypothesis was not supported, aligning with previous research [32,156]. When engaging with an AI-based well-being chatbot, users' ability to access smartphones, computers, or the internet, as well as their knowledge of how to use them, did not become barriers—most likely because our respondents were young and highly educated [157]. It could have been expected that complexity would be statistically significant; nonetheless, other studies of new technologies in health care, when complexity was evaluated as part of DOI, also obtained nonsignificant results [13,91,94,157-160]. A possible explanation, also supported by the literature, is that early adopters of new technologies, as in the case of our study, have a higher cognitive ability and are more accustomed to managing complexity, so they do not perceive complexity as an obstacle to using new technologies, including AI-based well-being chatbots [13,91,94,157,160]. Some research indicates that relevant advantages could influence users' willingness to adopt new technologies in the initial implementation phase [161,162]. However, not all studies support this. The nonconfirmation of the relative advantages hypothesis in our research suggests that individuals focused more on the experiences and value derived from chatbots' services rather than the direct benefits of outcomes [163]. The emotional component—affect—is much more relevant than technical factors, minimizing their impact on the model. Among younger people with higher level of education, perceived consequences show different influences on behavioral intention [164]. In our study, perceived consequences had a nonsignificant impact on the intention to engage with a well-being chatbot, suggesting that individuals become more focused on their engaging experience rather than on the positive or negative consequences.

Age, gender, education, and chronic disease status as control variables were not statistically significant for the 2 dependent variables (intention to engage and engagement behavior) in the model. Our study respondents were young, were mostly highly educated, included a high proportion of women, and had a ratio of respondents with chronic diseases aligned with the literature [25-28,116]. In addition, early adopters exhibit behavior toward a technology that differs from those who adopt it later [94,157,160]. AI-based well-being chatbots are a new technology, currently being used by early adopters with no significant heterogeneity, which, at this early stage of implementation, does not contribute to significant results for the control variables.

Theoretical Implications

First, we provided an integrated perspective of TIB, DOI and trust to uncover the critical factors influencing chatbot intention to engage and engagement and how these factors influence individual decision-making. In previous studies, the technology acceptance model and unified theory of acceptance and use of technology were considered the most used and integrated information systems theories, and there were few relevant adoptions of TIB [165]. However, we believed that affect and habit in TIB were most appropriate for explaining AI-based well-being chatbot use behavior based on its characteristics. The new integrated model provides important theoretical support for future research on AI-based chatbots and other products.

Trust is among the most important factors determining human intention to adopt smart technology products [166]. Adding trust to the research model explained users' intention to engage and engagement behavior in a multidimensional way. The results of this study provided empirical evidence for the completeness of future AI information system application models. Nowadays, AI technology is being increasingly applied in diverse information systems and the autonomy of information systems has been enhanced. Theoretically validating the impact of trust on AI products is imperative.

Practical Implications

The results of our study model can provide guidance on better implementing AI-based well-being chatbots to increase their adoption. The triangulation between users, companies that develop AI-based well-being chatbots, and mental health practitioners is highly relevant and should be considered during the system's development and real-world use. In our model, affect had a statistically significant impact on the intention to engage with AI-based well-being chatbots. It is crucial for designers and developers to recognize the influence of users' affect on their willingness to engage with AI-based well-being chatbots. Affect describes a wide range of feelings, both positive and negative [72]. Developers should address the needs of users in different emotional states during the design and development process while simultaneously enhancing the system's ability to recognize emotions and build emotional connections between chatbots and users [167,168]. Because affect is a complex dimension that transcends the purely IT aspects of the chatbot, developers should seek the support of mental health practitioners from the early stages of these systems' development. Habit is also a statistically significant construct, so providing good support services should be a key area of focus for AI-based well-being chatbot companies, as they contribute to user experience and help maintain users' habits of continuous engagement with the chatbot. According to our model results, SF also contribute to the intention to engage AI-based well-being chatbots. It suggests that individuals relevant to users can influence the adoption of AI-based well-being chatbots [51]. This indicates that mental health practitioners may influence users' intention to engage with AI-based well-being chatbots. For companies developing AI-based well-being chatbots, it is important to engage with mental health practitioners. Developers could also leverage social media to promote their AI-based well-being chatbots, considering the relevance of SF.

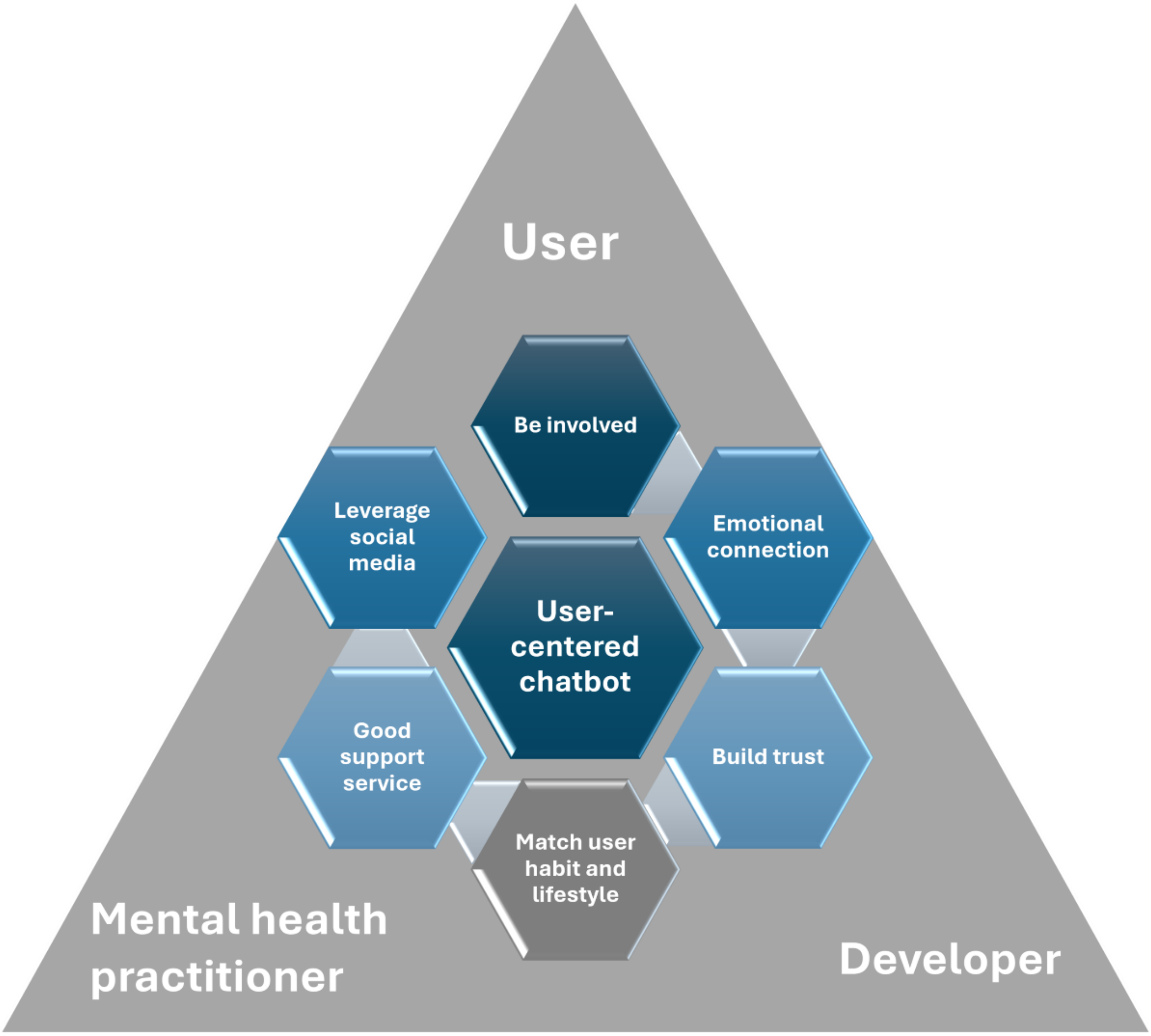
Trust was another statistically significant construct in the model. Establishing user trust is pivotal for driving participation behavior, and devising strategies to cultivate user trust requires careful consideration. They should consider introducing mental health practitioners into the system development process [169]. The incorporation of mental health practitioners enhances the efficacy and competitive advantage of chatbots, while reinforcing user trust and fostering their intention to engage. In some current practices, mental health experts have already been involved in the chatbot development process, with positive outcomes [170].

Another relevant construct in the model is compatibility. It is recommended that users be included in the entire life cycle of the AI-based well-being chatbot, as this allows a comprehensive understanding of users' habits and lifestyles, thereby facilitating product compatibility [8,171]. By communicating with users in greater depth, it is possible to gain insights into their interactions with the chatbot and to establish trust. The emotional connection between AI-based well-being chatbots and their users represents the core value of this technology, requiring designers and developers to prioritize a user-centered approach in their work [172].

Although this research did not cover older people or those with special needs, future developments specifically targeting these groups should account for the fact that they may not have access to certain resources or may have lower digital literacy [173]. In these cases, the constructs of complexity and facilitating conditions, which were nonsignificant in this research, should play a critical role. Less complex systems and providing the right resources should increase the adoption of AI-based well-being chatbots within these groups.

The involvement of mental health practitioners and users, integrating and applying their feedback throughout the AI-based well-being chatbot's development, is essential for successfully implementing the chatbot's entire life cycle. It is crucial to use AI-based well-being chatbots for therapeutic purposes under the guidance of qualified professionals to prevent misuse, which could potentially result in risky behaviors [174,175]. Figure 3 provides a graphical representation of the suggested practical implications. RQ2 about improving user experience and engagement from a practical point of view, leading to enhanced service levels was answered.

Figure 3. Suggested practical implications.



Limitations and Future Research

Our study used a convenience sample for which an online, multiplatform collection approach with a large coverage of the Chinese population was used to prevent single-source bias [120]. Still, the approach was not entirely random because we only posted the questionnaire, making it available, and we did not send messages to all platform users. In addition, the data source was from only one country. Future studies could use probabilistic sampling approaches. Access to large databases of AI-based well-being chatbot users could support studies with random sampling. Future research should expand the data sources to include participants from different countries and ethnicities, as well as special target populations (eg, people diagnosed with depression, older people). A multicountry approach as the next step may be used to evaluate if the findings are generalizable. Future studies could also collect the independent and dependent variables in different moments in time to reduce the probability of common-source bias [120]. Collection of real use data from the AI-based well-being chatbots could also be an advantage in future studies. Specific criteria for measuring engagement, such as the number of minutes of participation in using the chatbot, the number of logins, and the number of completed modules were not counted [176]. A future study may also explore engagement with AI-based well-being chatbots from mental health practitioners' perspective, as well as conduct qualitative research or quantitative research.

Conclusions

AI-based well-being chatbots provide users with emotional support to help alleviate conditions such as loneliness and anxiety. They are an effective solution to the lack of resources for mental health care. Exploring the factors affecting their use carries great significance. This paper extended past models by using DOI and trust theory, based on TIB. It proposed an integrated model that effectively explained the factors affecting individuals' acceptance and engagement with AI-based well-being chatbots. Among them, affect, habit, and trust play vital roles. The important theoretical role of the TIB model in the context of chatbots was validated. In addition, recommendations for the design of well-being chatbots were presented. For example, human-centered design concepts, attention to ethical issues, and building trust through characterization have important practical implications.

Acknowledgments

This work was supported by national funds through Fundação para a Ciência e a Tecnologia, under the project UIDB/04152/2020, Centro de Investigação em Gestão de Informação (MagIC Information Management Research Center/NOVA Information Management School).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Checklist for Reporting Results of Internet E-Surveys (CHERRIES) checklist.

[PDF File (Adobe PDF File), 117 KB - humanfactors_v11i1e59908_app1.pdf]

Multimedia Appendix 2

Study questionnaire.

[PDF File (Adobe PDF File), 100 KB - humanfactors_v11i1e59908_app2.pdf]

Multimedia Appendix 3

Outer loadings and cross-loadings.

[PDF File (Adobe PDF File), 174 KB - humanfactors_v11i1e59908_app3.pdf]

Multimedia Appendix 4

Fornell-Larcker criterion.

[PDF File (Adobe PDF File), 158 KB - humanfactors_v11i1e59908_app4.pdf]

References

1. The WHO special initiative for mental health (2019-2023) : universal health coverage for mental health. World Health Organization. 2019. URL: https://iris.who.int/handle/10665/310981 [accessed 2023-06-16]
2. Mojtabai R, Olfson M, Sampson NA, Jin R, Druss B, Wang PS, et al. Barriers to mental health treatment: results from the National Comorbidity Survey replication. Psychol Med 2011 Aug;41(8):1751-1761 [FREE Full text] [doi: 10.1017/S0033291710002291] [Medline: 21134315]
3. Jabir AI, Lin X, Martinengo L, Sharp G, Theng YL, Tudor Car L. Attrition in conversational agent-delivered mental health interventions: systematic review and meta-analysis. J Med Internet Res 2024 Feb 27;26:e48168 [FREE Full text] [doi: 10.2196/48168] [Medline: 38412023]
4. MacNeill AL, MacNeill L, Doucet S, Luke A. Professional representation of conversational agents for health care: a scoping review protocol. JBI Evid Synth 2022 Feb;20(2):666-673. [doi: 10.11124/JBIES-20-00589] [Medline: 34374689]
5. Gray-Burrows KA, Willis TA, Foy R, Rathfelder M, Bland P, Chin A, et al. Role of patient and public involvement in implementation research: a consensus study. BMJ Qual Saf 2018 Oct;27(10):858-864 [FREE Full text] [doi: 10.1136/bmjqs-2017-006954] [Medline: 29666310]
6. Brandtzaeg PB, Skjuve M, Følstad A. My AI friend: how users of a social chatbot understand their human–AI friendship. Hum Commun Res 2022 Jul;48(3):404-429. [doi: 10.1093/hcr/hqac008]
7. Xie T, Pentina I, Hancock T. Friend, mentor, lover: does chatbot engagement lead to psychological dependence? J Serv Manag 2023;34(4):806-828. [doi: 10.1108/JOSM-02-2022-0072]
8. Sadasivan C, Cruz C, Dolgoy N, Hyde A, Campbell S, McNeely M, et al. Examining patient engagement in chatbot development approaches for healthy lifestyle and mental wellness interventions: scoping review. J Particip Med 2023 May 22;15:e45772 [FREE Full text] [doi: 10.2196/45772] [Medline: 37213199]
9. Kettle L, Lee YC. User experiences of well-being chatbots. Hum Factors 2024 Jun;66(6):1703-1723. [doi: 10.1177/00187208231162453] [Medline: 36916743]
10. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry 2019 Jul;64(7):456-464 [FREE Full text] [doi: 10.1177/0706743719828977] [Medline: 30897957]
11. Ahmed A, Ali N, Aziz S, Abd-alrazaq AA, Hassan A, Khalifa M, et al. A review of mobile chatbot apps for anxiety and depression and their self-care features. Comput Methods Programs Biomed Update 2021;1:100012. [doi: 10.1016/j.cmpbup.2021.100012]

12. Nelekar S, Abdulrahman A, Gupta M, Richards D. Effectiveness of embodied conversational agents for managing academic stress at an Indian University (ARU) during COVID-19. Br J Educ Technol 2021 Dec 20;53(3):491-511. [doi: 10.1111/bjet.13174]
13. Baines R, Bradwell H, Edwards K, Stevens S, Prime S, Tredinnick-Rowe J, et al. Meaningful patient and public involvement in digital health innovation, implementation and evaluation: a systematic review. Health Expect 2022 Aug;25(4):1232-1245 [FREE Full text] [doi: 10.1111/hex.13506] [Medline: 35526274]
14. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
15. Peuters C, Maenhout L, Cardon G, De Paepe A, DeSmet A, Lauwerier E, et al. A mobile healthy lifestyle intervention to promote mental health in adolescence: a mixed-methods evaluation. BMC Public Health 2024 Jan 02;24(1):44 [FREE Full text] [doi: 10.1186/s12889-023-17260-9] [Medline: 38166797]
16. Chatterjee A, Prinz A, Gerdes M, Martinez S. Digital interventions on healthy lifestyle management: systematic review. J Med Internet Res 2021 Nov 17;23(11):e26931 [FREE Full text] [doi: 10.2196/26931] [Medline: 34787575]
17. Widmer RJ, Collins NM, Collins CS, West CP, Lerman LO, Lerman A. Digital health interventions for the prevention of cardiovascular disease: a systematic review and meta-analysis. Mayo Clin Proc 2015 Apr;90(4):469-480 [FREE Full text] [doi: 10.1016/j.mayocp.2014.12.026] [Medline: 25841251]
18. Fadhil A, Gabrielli S. Addressing challenges in promoting healthy lifestyles: the al-chatbot approach. In: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2017 Presented at: PervasiveHealth '17; May 23-26, 2017; Barcelona, Spain. [doi: 10.1145/3154862.3154914]
19. Escriva Boulley G, Leroy T, Bernetière C, Paquienseguy F, Desfriches-Doria O, Préau M. Digital health interventions to help living with cancer: a systematic review of participants' engagement and psychosocial effects. Psychooncology 2018 Dec;27(12):2677-2686. [doi: 10.1002/pon.4867] [Medline: 30152074]
20. Haque MD, Rubya S. An overview of chatbot-based mobile mental health apps: insights from app description and user reviews. JMIR Mhealth Uhealth 2023 May 22;11:e44838 [FREE Full text] [doi: 10.2196/44838] [Medline: 37213181]
21. Belen-Saglam R, Nurse JR, Hodges D. An investigation into the sensitivity of personal information and implications for disclosure: a UK perspective. Front Comput Sci 2022 Jun 30;4:1-22. [doi: 10.3389/fcomp.2022.908245]
22. Guerreiro MP, Angelini L, Rafael Henriques H, El Kamali M, Baixinho C, Balsa J, et al. Conversational agents for health and well-being across the life course: protocol for an evidence map. JMIR Res Protoc 2021 Sep 17;10(9):e26680 [FREE Full text] [doi: 10.2196/26680] [Medline: 34533460]
23. Diederich S, Brendel AB, Morana S, Kolbe L. On the design of and interaction with conversational agents: an organizing and assessing review of human-computer interaction research. J Assoc Inform Syst 2022;23(1):96-138. [doi: 10.17705/1jais.00724]
24. Al-Shafei M. Navigating human-chatbot interactions: an investigation into factors influencing user satisfaction and engagement. Int J Hum Comput Interact 2024 Jan 12:1-18. [doi: 10.1080/10447318.2023.2301252]
25. Qin Z, Ng S, Wu W, Zhang S. What Chinese women seek in mental health apps: insights from analyzing Xiaohongshu user posts during the COVID-19 pandemic. Healthcare (Basel) 2024 Jun 28;12(13):1297 [FREE Full text] [doi: 10.3390/healthcare12131297] [Medline: 38998832]
26. Shan Y, Ji M, Xie W, Zhang X, Qian X, Li R, et al. Use of health care chatbots among young people in China during the Omicron wave of COVID-19: evaluation of the user experience of and satisfaction with the technology. JMIR Hum Factors 2022 Jun 09;9(2):e36831 [FREE Full text] [doi: 10.2196/36831] [Medline: 35576058]
27. Aziz M, Erbad A, Almourad MB, Altuwairiqi M, McAlaney J, Ali R. Did usage of mental health apps change during COVID-19? A comparative study based on an objective recording of usage data and demographics. Life (Basel) 2022 Aug 19;12(8):1266 [FREE Full text] [doi: 10.3390/life12081266] [Medline: 36013444]
28. Zhang X, Lewis S, Chen X, Berry N, Bucci S. Technology use and attitudes towards digital mental health in people with severe mental health problems: a survey study in China. Front Psychiatry 2023;14:1261795 [FREE Full text] [doi: 10.3389/fpsyt.2023.1261795] [Medline: 38076680]
29. Gan DZ, McGillivray L, Han J, Christensen H, Torok M. Effect of engagement with digital interventions on mental health outcomes: a systematic review and meta-analysis. Front Digit Health 2021;3:764079 [FREE Full text] [doi: 10.3389/fdgth.2021.764079] [Medline: 34806079]
30. Neimeyer RA, Kazantzis N, Kassler DM, Baker KD, Fletcher R. Group cognitive behavioural therapy for depression outcomes predicted by willingness to engage in homework, compliance with homework, and cognitive restructuring skill acquisition. Cogn Behav Ther 2008;37(4):199-215. [doi: 10.1080/16506070801981240] [Medline: 18608311]
31. Venkatesh V, Thong JY, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157-178. [doi: 10.2307/41410412]
32. Tavares J, Oliveira T. Electronic health record patient portal adoption by health care consumers: an acceptance model and survey. J Med Internet Res 2016 Mar 02;18(3):e49 [FREE Full text] [doi: 10.2196/jmir.5069] [Medline: 26935646]
33. Pentina I, Hancock T, Xie T. Exploring relationship development with social chatbots: a mixed-method study of replika. Comput Human Behav 2023 Mar;140:107600. [doi: 10.1016/j.chb.2022.107600]

34. Zhu Y, Wang R, Pu C. "I am chatbot, your virtual mental health adviser." What drives citizens' satisfaction and continuance intention toward mental health chatbots during the COVID-19 pandemic? An empirical study in China. Digit Health 2022;8:20552076221090031 [FREE Full text] [doi: 10.1177/20552076221090031] [Medline: 35381977]
35. Reinhardt R, Hietschold N, Spyridonidis D. Adoption and diffusion of innovations in health care. In: Gurtner S, Soyez K, editors. Challenges and Opportunities in Health Care Management. Cham, Switzerland: Springer; 2015:211-221.
36. Saglam RB, Nurse JR, Hodges D. Privacy concerns in chatbot interactions: when to trust and when to worry. In: Proceedings of the 23rd HCI International Conference. 2021 Presented at: HCII 2021; July 24-29, 2021; Virtual Event. [doi: 10.1007/978-3-030-78642-7_53]
37. Liu K, Tao D. The roles of trust, personalization, loss of privacy, and anthropomorphism in public acceptance of smart healthcare services. Comput Human Behav 2022 Feb;127:107026. [doi: 10.1016/j.chb.2021.107026]
38. Zhao S, Cao Y, Cao H, Liu K, Lv X, Zhang J, et al. Chinese version of the mHealth app usability questionnaire: cross-cultural adaptation and validation. Front Psychol 2022;13:813309 [FREE Full text] [doi: 10.3389/fpsyg.2022.813309] [Medline: 35185732]
39. Wilkins CH. Effective engagement requires trust and being trustworthy. Med Care 2018 Oct;56 Suppl 10 Suppl 1(10 Suppl 1):S6-S8 [FREE Full text] [doi: 10.1097/MLR.0000000000000953] [Medline: 30015725]
40. Jiménez FR, Voss KE. An alternative approach to the measurement of emotional attachment. Psychol Market 2014 Apr 10;31(5):360-370. [doi: 10.1002/mar.20700]
41. Grové C. Co-developing a mental health and wellbeing chatbot with and for young people. Front Psychiatry 2020;11:606041 [FREE Full text] [doi: 10.3389/fpsyt.2020.606041] [Medline: 33597898]
42. Andersson G, Titov N. Advantages and limitations of internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11 [FREE Full text] [doi: 10.1002/wps.20083] [Medline: 24497236]
43. Bijkerk LE, Oenema A, Geschwind N, Spigt M. Measuring engagement with mental health and behavior change interventions: an integrative review of methods and instruments. Int J Behav Med 2023 Apr;30(2):155-166 [FREE Full text] [doi: 10.1007/s12529-022-10086-6] [Medline: 35578099]
44. Glenn D, Golinelli D, Rose RD, Roy-Byrne P, Stein MB, Sullivan G, et al. Who gets the most out of cognitive behavioral therapy for anxiety disorders? The role of treatment dose and patient engagement. J Consult Clin Psychol 2013 Aug;81(4):639-649 [FREE Full text] [doi: 10.1037/a0033403] [Medline: 23750465]
45. Hundt NE, Amspoker AB, Kraus-Schuman C, Cully JA, Rhoades H, Kunik ME, et al. Predictors of CBT outcome in older adults with GAD. J Anxiety Disord 2014 Dec;28(8):845-850 [FREE Full text] [doi: 10.1016/j.janxdis.2014.09.012] [Medline: 25445074]
46. Becker MP, Christensen BK, Cunningham CE, Furimsky I, Rimas H, Wilson F, et al. Preferences for early intervention mental health services: a discrete-choice conjoint experiment. Psychiatr Serv 2016 Feb;67(2):184-191. [doi: 10.1176/appi.ps.201400306] [Medline: 26369880]
47. Dixon LB, Holoshitz Y, Nossel I. Treatment engagement of individuals experiencing mental illness: review and update. World Psychiatry 2016 Feb;15(1):13-20 [FREE Full text] [doi: 10.1002/wps.20306] [Medline: 26833597]
48. Karyotaki E, Kleiboer A, Smit F, Turner DT, Pastor AM, Andersson G, et al. Predictors of treatment dropout in self-guided web-based interventions for depression: an 'individual patient data' meta-analysis. Psychol Med 2015 Oct;45(13):2717-2726 [FREE Full text] [doi: 10.1017/S0033291715000665] [Medline: 25881626]
49. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
50. Wu A, Scult MA, Barnes ED, Betancourt JA, Falk A, Gunning FM. Smartphone apps for depression and anxiety: a systematic review and meta-analysis of techniques to increase engagement. NPJ Digit Med 2021 Feb 11;4(1):20 [FREE Full text] [doi: 10.1038/s41746-021-00386-8] [Medline: 33574573]
51. Triandis HC. Interpersonal Behavior. Pacific Grove, CA: Brooks/Cole Publishing Company; 1977.
52. Stern PC. New environmental theories: toward a coherent theory of environmentally significant behavior. J Soc Issues 2002 Dec 17;56(3):407-424. [doi: 10.1111/0022-4537.00175]
53. Pee LG, Woon IM, Kankanhalli A. Explaining non-work-related computing in the workplace: a comparison of alternative models. Inf Manag 2008 Mar;45(2):120-130. [doi: 10.1016/j.im.2008.01.004]
54. Singh B, Olds T, Brinsley J, Dumuid D, Virgara R, Matricciani L, et al. Systematic review and meta-analysis of the effectiveness of chatbots on lifestyle behaviours. NPJ Digit Med 2023 Jun 23;6(1):118 [FREE Full text] [doi: 10.1038/s41746-023-00856-1] [Medline: 37353578]
55. Park D, Kim H. Determinants of intentions to use digital mental healthcare content among university students, faculty, and staff: motivation, perceived usefulness, perceived ease of use, and parasocial interaction with AI chatbot. Sustainability 2023 Jan 03;15(1):872. [doi: 10.3390/su15010872]
56. Zarrinabadi N, Lou NM, Darvishnezhad Z. To praise or not to praise? Examining the effects of ability vs. effort praise on speaking anxiety and willingness to communicate in EFL classrooms. Innov Lang Learn Teach 2021 Jun 17;17(1):88-101. [doi: 10.1080/17501229.2021.1938079]

57. Flavián C, Akdim K, Casaló LV. Effects of voice assistant recommendations on consumer behavior. Psychol Market 2022 Nov 30;40(2):328-346. [doi: 10.1002/mar.21765]
58. Yang R, Wibowo S, Mubarak S. An investigation into domestic violence victims' adoption of chatbots for help-seeking: based on the UTAUT2 and health belief models. In: Proceedings of the PACIS 2023. 2023 Presented at: PACIS 2023; July 8-12, 2023; Nanchang, China.
59. Rogers EM. Diffusion of Innovations, Fourth Edition. New York, NY: Free Press; 2010.
60. Rogers EM. Diffusion of innovations: modifications of a model for telecommunications. In: Stoetzer MW, Mahler A, editors. Die Diffusion von Innovationen in der Telekommunikation. Berlin, Heidelberg: Springer; 1995.
61. Moore GC, Benbasat I. Development of an instrument to measure the perceptions of adopting an information technology innovation. Inf Syst Res 1991 Sep;2(3):192-222. [doi: 10.1287/isre.2.3.192]
62. Mayer RC, Davis JH, Schoorman FD. An integrative model of organizational trust. Acad Manag Rev 1995 Jul 01;20(3):709-734. [doi: 10.2307/258792]
63. Hancock PA, Billings DR, Schaefer KE, Chen JY, de Visser EJ, Parasuraman R. A meta-analysis of factors affecting trust in human-robot interaction. Hum Factors 2011 Oct;53(5):517-527. [doi: 10.1177/0018720811417254] [Medline: 22046724]
64. McKnight DH, Chervany NL. Trust and distrust definitions: one bite at a time. In: Falcone R, Singh M, Tan YH, editors. Trust in Cyber-societies. Berlin, Germany: Springer; 2001.
65. Lukyanenko R, Maass W, Storey VC. Trust in artificial intelligence: from a Foundational Trust Framework to emerging research opportunities. Electron Markets 2022 Nov 28;32(4):1993-2020. [doi: 10.1007/s12525-022-00605-4]
66. Misztal B. Trust in Modern Societies: The Search for the Bases of Social Order. Cambridge, UK: Polity Press; 2013.
67. Koehn D. Should we trust in trust? Am Bus Law J 2007 Aug 27;34(2):183-204. [doi: 10.1111/j.1744-1714.1996.tb00695.x]
68. Du L, Xu J, Chen X, Zhu X, Zhang Y, Wu R, et al. Rebuild doctor-patient trust in medical service delivery in China. Sci Rep 2020 Dec 15;10(1):21956 [FREE Full text] [doi: 10.1038/s41598-020-78921-y] [Medline: 33319841]
69. Chipidza FE, Wallwork RS, Stern TA. Impact of the doctor-patient relationship. Prim Care Companion CNS Disord 2015;17(5):10.4088/PCC.15f01840 [FREE Full text] [doi: 10.4088/PCC.15f01840] [Medline: 26835164]
70. Seitz L, Bekmeier-Feuerhahn S, Gohil K. Can we trust a chatbot like a physician? A qualitative study on understanding the emergence of trust toward diagnostic chatbots. Int J Hum Comput Stud 2022 Sep;165:102848. [doi: 10.1016/j.ijhcs.2022.102848]
71. Powell J. Trust me, I'm a chatbot: how artificial intelligence in health care fails the Turing test. J Med Internet Res 2019 Oct 28;21(10):e16222 [FREE Full text] [doi: 10.2196/16222] [Medline: 31661083]
72. Triandis HC. Values, attitudes, and interpersonal behavior. Nebr Symp Motiv 1980;27:195-259. [Medline: 7242748]
73. Woon IM, Pee LG. Behavioral factors affecting internet abuse in the workplace – an empirical investigation. In: Proceedings of the SIGHCI 2004. 2004 Presented at: SIGHCI 2004; December 10-11, 2004; Washington, DC URL: https://aisel.aisnet.org/sighci2004/5/
74. Madden TJ, Ellen PS, Ajzen I. A comparison of the theory of planned behavior and the theory of reasoned action. Pers Soc Psychol Bull 1992 Feb 01;18(1):3-9. [doi: 10.1177/0146167292181001]
75. Ajzen I. From intentions to actions: a theory of planned behavior. In: Kuhl J, Beckmann J, editors. Action Control: From Cognition to Behavior. Berlin, Heidelberg: Springer-Verlag; 1985:11-39.
76. Chang MK, Cheung W. Determinants of the intention to use internet/WWW at work: a confirmatory study. Inf Manag 2001 Nov;39(1):1-14. [doi: 10.1016/s0378-7206(01)00075-1]
77. Fanta GB, Pretorius L. Sociotechnical factors of sustainable digital health systems: a system dynamics model. Health Policy Technol 2023 Mar;12(1):100729. [doi: 10.1016/j.hlpt.2023.100729]
78. Haoyue LL, Cho H. Factors influencing intention to engage in human–chatbot interaction: examining user perceptions and context culture orientation. Univ Access Inf Soc 2024 Jan 29. [doi: 10.1007/s10209-023-01087-7]
79. Jacob C, Sanchez-Vazquez A, Ivory C. Social, organizational, and technological factors impacting clinicians' adoption of mobile health tools: systematic literature review. JMIR Mhealth Uhealth 2020 Feb 20;8(2):e15935 [FREE Full text] [doi: 10.2196/15935] [Medline: 32130167]
80. Barrett LF, Bliss-Moreau E. Affect as a psychological primitive. Adv Exp Soc Psychol 2009;41:167-218 [FREE Full text] [doi: 10.1016/S0065-2601(08)00404-8] [Medline: 20552040]
81. Alexander R, Aragón OR, Bookwala J, Cherbuin N, Gatt JM, Kahrilas IJ, et al. The neuroscience of positive emotions and affect: implications for cultivating happiness and wellbeing. Neurosci Biobehav Rev 2021 Feb;121:220-249 [FREE Full text] [doi: 10.1016/j.neubiorev.2020.12.002] [Medline: 33307046]
82. Diener E, Suh EM, Lucas RE, Smith HL. Subjective well-being: three decades of progress. Psychol Bull 1999 Mar;125(2):276-302. [doi: 10.1037/0033-2909.125.2.276]
83. Chin WY, Chua HN. Using the theory of interpersonal behavior to predict information security policy compliance. In: Proceedings of the Eighth International Conference on eDemocracy & eGovernment. 2021 Presented at: ICEDEG 2021; July 28-30, 2021; Quito, Ecuador. [doi: 10.1109/icedeg52154.2021.9530849]
84. Moody GD, Siponen M. Using the theory of interpersonal behavior to explain non-work-related personal use of the internet at work. Inf Manag 2013 Sep;50(6):322-335. [doi: 10.1016/j.im.2013.04.005]

85. Gardner B, Abraham C, Lally P, de Bruijn GJ. Towards parsimony in habit measurement: testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. Int J Behav Nutr Phys Act 2012 Aug 30;9:102 [FREE Full text] [doi: 10.1186/1479-5868-9-102] [Medline: 22935297]
86. Sheeran P, Gollwitzer PM, Bargh JA. Nonconscious processes and health. Health Psychol 2013 May;32(5):460-473. [doi: 10.1037/a0029203] [Medline: 22888816]
87. Orbell S, Hodgkins S, Sheeran P. Implementation intentions and the theory of planned behavior. Pers Soc Psychol Bull 1997 Sep;23(9):945-954. [doi: 10.1177/0146167297239004] [Medline: 29506445]
88. Badawy SM, Thompson AA, Kuhns LM. Medication adherence and technology-based interventions for adolescents with chronic health conditions: a few key considerations. JMIR Mhealth Uhealth 2017 Dec 22;5(12):e202 [FREE Full text] [doi: 10.2196/mhealth.8310] [Medline: 29273573]
89. Lee JD, See KA. Trust in automation: designing for appropriate reliance. Hum Factors 2004;46(1):50-80. [doi: 10.1518/hfes.46.1.50_30392] [Medline: 15151155]
90. Montague E. Validation of a trust in medical technology instrument. Appl Ergon 2010 Oct;41(6):812-821 [FREE Full text] [doi: 10.1016/j.apergo.2010.01.009] [Medline: 20189163]
91. Brown JE, Halpern J. AI chatbots cannot replace human interactions in the pursuit of more inclusive mental healthcare. SSM Mental Health 2021 Dec;1:100017. [doi: 10.1016/j.ssmmh.2021.100017]
92. D'Alfonso S. AI in mental health. Curr Opin Psychol 2020 Dec;36:112-117. [doi: 10.1016/j.copsyc.2020.04.005] [Medline: 32604065]
93. Sillence E, Blythe JM, Briggs P, Moss M. A revised model of trust in internet-based health information and advice: cross-sectional questionnaire study. J Med Internet Res 2019 Nov 11;21(11):e11125 [FREE Full text] [doi: 10.2196/11125] [Medline: 31710297]
94. Scott SD, Plotnikoff RC, Karunamuni N, Bize R, Rodgers W. Factors influencing the adoption of an innovation: an examination of the uptake of the Canadian Heart Health Kit (HHK). Implement Sci 2008 Oct 02;3:41 [FREE Full text] [doi: 10.1186/1748-5908-3-41] [Medline: 18831766]
95. Hofstede G, Bond MH. The Confucius connection: from cultural roots to economic growth. Organ Dyn 1988 Mar;16(4):5-21. [doi: 10.1016/0090-2616(88)90009-5]
96. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. JMIR Mhealth Uhealth 2018 Nov 23;6(11):e12106 [FREE Full text] [doi: 10.2196/12106] [Medline: 30470676]
97. Ortega‐Ochoa E, Arguedas M, Daradoumis T. Empathic pedagogical conversational agents: a systematic literature review. Brit J Educ Technol 2023 Dec 06;55(3):886-909. [doi: 10.1111/bjet.13413]
98. Zhang X, Shah J, Han M. ChatGPT for fast learning of positive energy district (PED): a trial testing and comparison with expert discussion results. Buildings 2023 May 27;13(6):1392. [doi: 10.3390/buildings13061392]
99. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119 [FREE Full text] [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
100. Nason R. Challenges of implementing complexity in healthcare. Healthc Manage Forum 2023 Nov;36(6):368-372 [FREE Full text] [doi: 10.1177/08404704231191956] [Medline: 37544743]
101. Buckingham SA, Walker T, Morrissey K, Smartline project team. The feasibility and acceptability of digital technology for health and wellbeing in social housing residents in Cornwall: a qualitative scoping study. Digit Health 2022;8:20552076221074124 [FREE Full text] [doi: 10.1177/20552076221074124] [Medline: 35096410]
102. Wang X, Luo R, Liu Y, Chen P, Tao Y, He Y. Revealing the complexity of users' intention to adopt healthcare chatbots: a mixed-method analysis of antecedent condition configurations. Inf Process Manag 2023 Sep;60(5):103444. [doi: 10.1016/j.ipm.2023.103444]
103. Biro J, Linder C, Neyens D. The effects of a health care chatbot's complexity and persona on user trust, perceived usability, and effectiveness: mixed methods study. JMIR Hum Factors 2023 Feb 01;10:e41017 [FREE Full text] [doi: 10.2196/41017] [Medline: 36724004]
104. Vagnani G, Volpe L. Innovation attributes and managers' decisions about the adoption of innovations in organizations: a meta-analytical review. Int J Innov Studies 2017 Nov;1(2):107-133. [doi: 10.1016/j.ijis.2017.10.001]
105. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629 [FREE Full text] [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
106. Zhou L, Gao J, Li D, Shum HY. The design and implementation of Xiaoice, an empathetic social chatbot. Comput Linguist 2020 Mar 01;46(1):53-93. [doi: 10.1162/coli_a_00368]
107. Ghandeharioun A, McDuff D, Czerwinski M, Rowan K. EMMA: an emotion-aware wellbeing chatbot. In: Procedings of the 8th International Conference on Affective Computing and Intelligent Interaction. 2019 Presented at: ACII 2019; September 3-6, 2019; Cambridge, UK. [doi: 10.1109/acii.2019.8925455]

108. Curtis RG, Bartel B, Ferguson T, Blake HT, Northcott C, Virgara R, et al. Improving user experience of virtual health assistants: scoping review. J Med Internet Res 2021 Dec 21;23(12):e31737 [FREE Full text] [doi: 10.2196/31737] [Medline: 34931997]
109. Ben-Zeev D, Chander A, Tauscher J, Buck B, Nepal S, Campbell A, et al. A smartphone intervention for people with serious mental illness: fully remote randomized controlled trial of CORE. J Med Internet Res 2021 Nov 12;23(11):e29201 [FREE Full text] [doi: 10.2196/29201] [Medline: 34766913]
110. Lee M, Ackermans S, Van As N, Chang H, Lucas E, IJsselsteijn W. Caring for Vincent: a chatbot for self-compassion. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19; May 4-9, 2019; Glasgow, Scotland. [doi: 10.1145/3290605.3300932]
111. Menzli LJ, Smirani LK, Boulahia JA, Hadjouni M. Investigation of open educational resources adoption in higher education using Rogers' diffusion of innovation theory. Heliyon 2022 Jul 11;8(7):e09885. [doi: 10.1016/j.heliyon.2022.e09885] [Medline: 35874064]
112. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
113. Entenberg GA, Dosovitsky G, Aghakhani S, Mostovoy K, Carre N, Marshall Z, et al. User experience with a parenting chatbot micro intervention. Front Digit Health 2022 Jan 11;4:989022 [FREE Full text] [doi: 10.3389/fdgth.2022.989022] [Medline: 36714612]
114. Agapie E, Chang K, Patrachari S, Neary M, Schueller SM. Understanding mental health apps for youth: focus group study with Latinx youth. JMIR Form Res 2022 Oct 18;6(10):e40726. [doi: 10.2196/40726] [Medline: 36256835]
115. Jakob R, Harperink S, Rudolf AM, Fleisch E, Haug S, Mair JL, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371 [FREE Full text] [doi: 10.2196/35371] [Medline: 35612886]
116. Kontos E, Blake KD, Chou WY, Prestin A. Predictors of eHealth usage: insights on the digital divide from the health information national trends survey 2012. J Med Internet Res 2014 Jul 16;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
117. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
118. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. Value Health 2005;8(2):94-104 [FREE Full text] [doi: 10.1111/j.1524-4733.2005.04054.x] [Medline: 15804318]
119. Fielding NG, Lee RM, Blank G. The SAGE Handbook of Online Research Methods. Thousand Oaks, CA: SAGE Publications Ltd; 2017.
120. Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol 2003 Oct;88(5):879-903. [doi: 10.1037/0021-9010.88.5.879] [Medline: 14516251]
121. Liang X, Yan M, Li H, Deng Z, Lu Y, Lu P, et al. WeChat official accounts' posts on medication use of 251 community healthcare centers in Shanghai, China: content analysis and quality assessment. Front Med (Lausanne) 2023 Jun 12;10:1155428 [FREE Full text] [doi: 10.3389/fmed.2023.1155428] [Medline: 37378298]
122. Zeng L, Li RY, Zeng H. Weibo users and Academia's foci on tourism safety: implications from institutional differences and digital divide. Heliyon 2022 Dec 10;9(3):e12306 [FREE Full text] [doi: 10.1016/j.heliyon.2022.e12306] [Medline: 36923855]
123. Bao Z, Zhu Y. Understanding online reviews adoption in social network communities: an extension of the information adoption model. Inf Technol People 2023 Sep (forthcoming). [doi: 10.1108/ITP-03-2022-0158]
124. Chou Y, Lin W. Blockbuster or flop? Effects of social media on the Chinese film market. SSRN J 2023 Jul 25:1-41. [doi: 10.2139/ssrn.4514856]
125. Lindell MK, Whitney DJ. Accounting for common method variance in cross-sectional research designs. J Appl Psychol 2001 Feb;86(1):114-121. [doi: 10.1037/0021-9010.86.1.114] [Medline: 11302223]
126. Hair JJ, Sarstedt M, Hopkins L, Kuppelwieser V. Partial least squares structural equation modeling (PLS-SEM): an emerging tool for business research. Eur Bus Rev 2014 Feb;26(2):106-121. [doi: 10.1108/EBR-10-2013-0128]
127. Ringle CM, Sarstedt M, Straub DW. Editor's comments: a critical look at the use of PLS-SEM in "MIS Quarterly". MIS Q 2012 Mar;36(1):iii-xiv. [doi: 10.2307/41410402]
128. Freeze R, Raschke RL. An assessment of formative and reflective constructs in IS research. In: Proceedings of the Fifteenth European Conference on Information Systems. 2007 Presented at: ECIS 2007; June 7-9, 2007; St. Gallen, Switzerland.
129. Bollen K, Lennox R. Conventional wisdom on measurement: a structural equation perspective. Psychol Bull 1991;110(2):305-314. [doi: 10.1037//0033-2909.110.2.305]
130. Edwards JR, Bagozzi RP. On the nature and direction of relationships between constructs and measures. Psychol Methods 2000 Jun;5(2):155-174. [doi: 10.1037/1082-989x.5.2.155] [Medline: 10937327]

131. Carl E, Witcraft SM, Kauffman BY, Gillespie EM, Becker ES, Cuijpers P, et al. Psychological and pharmacological treatments for generalized anxiety disorder (GAD): a meta-analysis of randomized controlled trials. Cogn Behav Ther 2020 Jan;49(1):1-21. [doi: 10.1080/16506073.2018.1560358] [Medline: 30760112]
132. Seegan PL, Miller MJ, Heliste JL, Fathi L, McGuire JF. Efficacy of stand-alone digital mental health applications for anxiety and depression: a meta-analysis of randomized controlled trials. J Psychiatr Res 2023 Aug;164:171-183. [doi: 10.1016/j.jpsychires.2023.06.019] [Medline: 37352813]
133. Al-Somali SA, Gholami R, Clegg B. An investigation into the acceptance of online banking in Saudi Arabia. Technovation 2009 Feb;29(2):130-141. [doi: 10.1016/j.technovation.2008.07.004]
134. Viana Pereira F, Tavares J, Oliveira T. Adoption of video consultations during the COVID-19 pandemic. Internet Interv 2023 Mar;31:100602 [FREE Full text] [doi: 10.1016/j.invent.2023.100602] [Medline: 36694630]
135. Hair JFJ, Hult GT, Ringle CM, Sarstedt M. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). Thousand Oaks, CA: SAGE Publications; 2014.
136. Henseler J, Ringle CM, Sarstedt M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. J Acad Mark Sci 2014 Aug 22;43:115-135. [doi: 10.1007/s11747-014-0403-8]
137. Hair JF, Risher JJ, Sarstedt M, Ringle CM. When to use and how to report the results of PLS-SEM. Eur Bus Rev 2019;31(1):2-24. [doi: 10.1108/ebr-11-2018-0203]
138. Johnson RE, Rosen CC, Djurdjevic E. Assessing the impact of common method variance on higher order multidimensional constructs. J Appl Psychol 2011 Jul;96(4):744-761. [doi: 10.1037/a0021504] [Medline: 21142343]
139. Ringle CM, Wende S, Becker JM. SmartPLS 4. SmartPLS GmbH. URL: http://www.smartpls.com [accessed 2024-10-13]
140. Shmueli G, Sarstedt M, Hair JF, Cheah JH, Ting H, Vaithilingam S, et al. Predictive model assessment in PLS-SEM: guidelines for using PLSpredict. Eur J Mark 2019;53(11):2322-2347. [doi: 10.1108/ejm-02-2019-0189]
141. Bradburn NM, Noll CE. The Structure of Psychological Well-Being. New York, NY: Aldine Publishing Company; 1969.
142. Wilson L, Marasoiu M. The development and use of chatbots in public health: scoping review. JMIR Hum Factors 2022 Oct 05;9(4):e35882 [FREE Full text] [doi: 10.2196/35882] [Medline: 36197708]
143. McKay FH, Wright A, Shill J, Stephens H, Uccellini M. Using health and well-being apps for behavior change: a systematic search and rating of apps. JMIR Mhealth Uhealth 2019 Jul 04;7(7):e11926 [FREE Full text] [doi: 10.2196/11926] [Medline: 31274112]
144. Su Z, Figueiredo MC, Jo J, Zheng K, Chen Y. Analyzing description, user understanding and expectations of AI in mobile health applications. AMIA Annu Symp Proc 2021 Jan 25;2020:1170-1179 [FREE Full text] [Medline: 33936493]
145. Huang CY, Yang MC, Huang MC. An empirical study on factors influencing consumer adoption intention of an AI-powered chatbot for health and weight management. Int J Performability Eng 2021;17(5):422-432. [doi: 10.23940/ijpe.21.05.p2.422432]
146. Hsiao CH, Chang JJ, Tang KY. Exploring the influential factors in continuance usage of mobile social apps: satisfaction, habit, and customer value perspectives. Telemat Inform 2016 May;33(2):342-355. [doi: 10.1016/j.tele.2015.08.014]
147. Bhattacherjee A. Understanding information systems continuance: an expectation-confirmation model. MIS Q 2001 Sep;25(3):351-370. [doi: 10.2307/3250921]
148. Okey OD, Udo EU, Rosa RL, Rodríguez DZ, Kleinschmidt JH. Investigating ChatGPT and cybersecurity: a perspective on topic modeling and sentiment analysis. Comput Secur 2023 Dec;135:103476. [doi: 10.1016/j.cose.2023.103476]
149. Ogilvie L, Prescott J, Carson J. The use of chatbots as supportive agents for people seeking help with substance use disorder: a systematic review. Eur Addict Res 2022;28(6):405-418 [FREE Full text] [doi: 10.1159/000525959] [Medline: 36041418]
150. Rahimi B, Nadri H, Lotfnezhad Afshar H, Timpka T. A systematic review of the technology acceptance model in health informatics. Appl Clin Inform 2018 Jul;9(3):604-634 [FREE Full text] [doi: 10.1055/s-0038-1668091] [Medline: 30112741]
151. Spitzberg BH. Issues in the development of a theory of interpersonal competence in the intercultural context. Int J Intercult Relat 1989;13(3):241-268. [doi: 10.1016/0147-1767(89)90012-6]
152. Ibrahim A, Knox K, Rundle-Thiele S, Arli D. Segmenting a water use market: theory of interpersonal behavior insights. Soc Mark Q 2017 Nov 19;24(1):3-17. [doi: 10.1177/1524500417741277]
153. Kuhail MA, Thomas J, Alramlawi S, Shah SJ, Thornquist E. Interacting with a chatbot-based advising system: understanding the effect of chatbot personality and user gender on behavior. Informatics 2022 Oct 10;9(4):81. [doi: 10.3390/informatics9040081]
154. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
155. Vaghefi I, Tulu B. The continued use of mobile health apps: insights from a longitudinal study. JMIR Mhealth Uhealth 2019 Aug 29;7(8):e12983 [FREE Full text] [doi: 10.2196/12983] [Medline: 31469081]
156. Yuan S, Ma W, Kanthawala S, Peng W. Keep using my health apps: discover users' perception of health and fitness apps with the UTAUT2 model. Telemed J E Health 2015 Sep;21(9):735-741. [doi: 10.1089/tmj.2014.0148] [Medline: 25919238]
157. Tavares J, Oliveira T. New integrated model approach to understand the factors that drive electronic health record portal adoption: cross-sectional national survey. J Med Internet Res 2018 Nov 19;20(11):e11032 [FREE Full text] [doi: 10.2196/11032] [Medline: 30455169]

158. Yi MY, Jackson JD, Park JS, Probst JC. Understanding information technology acceptance by individual professionals: toward an integrative view. Inf Manag 2006 Apr;43(3):350-363. [doi: 10.1016/j.im.2005.08.006]
159. Oliveira T, Thomas M, Baptista G, Campos F. Mobile payment: understanding the determinants of customer adoption and intention to recommend the technology. Comput Hum Behav 2016 Aug;61:404-414. [doi: 10.1016/j.chb.2016.03.030]
160. Rogers EM. Diffusion of Innovations. Washington, DC: Free Press; 2003.
161. Min S, So KK, Jeong M. Consumer adoption of the Uber mobile application: insights from diffusion of innovation theory and technology acceptance model. J Travel Tour Mark 2018 Sep 04;36(7):770-783. [doi: 10.1080/10548408.2018.1507866]
162. Al-Rahmi WM, Yahaya N, Alamri MM, Alyoussef IY, Al-Rahmi AM, Kamin YB. Integrating innovation diffusion theory with technology acceptance model: supporting students' attitude towards using a massive open online courses (MOOCs) systems. Interact Learn Environ 2019 Jun 20;29(8):1380-1392. [doi: 10.1080/10494820.2019.1629599]
163. Zhang X, Yu P, Yan J, Ton A M Spil I. Using diffusion of innovation theory to understand the factors impacting patient acceptance and use of consumer e-health innovations: a case study in a primary care clinic. BMC Health Serv Res 2015 Feb 21;15:71 [FREE Full text] [doi: 10.1186/s12913-015-0726-2] [Medline: 25885110]
164. Li H, Novack DH, Duke P, Gracely E, Cestone C, Davis T. Predictors of medical students' ethical decision-making: a pilot study using the Theory of Interpersonal Behavior. Patient Educ Couns 2020 May 23:S0738-3991(20)30292-5 (forthcoming). [doi: 10.1016/j.pec.2020.05.026] [Medline: 32540095]
165. Jacob C, Sanchez-Vazquez A, Ivory C. Understanding clinicians' adoption of mobile health tools: a qualitative review of the most used frameworks. JMIR Mhealth Uhealth 2020 Jul 06;8(7):e18072 [FREE Full text] [doi: 10.2196/18072] [Medline: 32442132]
166. Kaur D, Uslu S, Rittichier KJ, Durresi A. Trustworthy artificial intelligence: a review. ACM Comput Surv 2023 Mar 31;55(2):1-38 [FREE Full text] [doi: 10.1145/3491209]
167. Thieme A, Belgrave D, Doherty G. Machine learning in mental health: a systematic review of the HCI literature to support the development of effective and implementable ML systems. ACM Trans Comput Hum Interact 2020 Aug 17;27(5):1-53. [doi: 10.1145/3398069]
168. Gao J, Yan D, Dong M. Black-box adversarial attacks through speech distortion for speech emotion recognition. EURASIP J Audio Speech Music Process 2022 Aug 17;2022. [doi: 10.1186/s13636-022-00254-7]
169. Coghlan S, Leins K, Sheldrick S, Cheong M, Gooding P, D'Alfonso S. To chat or bot to chat: ethical issues with using chatbots in mental health. Digit Health 2023;9:20552076231183542 [FREE Full text] [doi: 10.1177/20552076231183542] [Medline: 37377565]
170. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
171. Maenhout L, Peuters C, Cardon G, Compernolle S, Crombez G, DeSmet A. Participatory development and pilot testing of an adolescent health promotion chatbot. Front Public Health 2021 Nov 11;9:724779 [FREE Full text] [doi: 10.3389/fpubh.2021.724779] [Medline: 34858919]
172. Koulouri T, Macredie RD, Olakitan D. Chatbots to support young adults' mental health: an exploratory study of acceptability. ACM Trans Interact Intell Syst 2022 Jul 20;12(2):1-39. [doi: 10.1145/3485874]
173. Frishammar J, Essén A, Bergström F, Ekman T. Digital health platforms for the elderly? Key adoption and usage barriers and ways to address them. Technol Forecast Soc Change 2023 Apr;189:122319. [doi: 10.1016/j.techfore.2023.122319]
174. Abd-Alrazaq AA, Rababeh A, Alajlani M, Bewick BM, Househ M. Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. J Med Internet Res 2020 Jul 13;22(7):e16021 [FREE Full text] [doi: 10.2196/16021] [Medline: 32673216]
175. Pandey S, Sharma S, Wazir S. Mental healthcare chatbot based on natural language processing and deep learning approaches: Ted the therapist. Int J Inf Technol 2022 Jun 02;14(7):3757-3766. [doi: 10.1007/s41870-022-00999-6]
176. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]

Abbreviations

ABC: attitude-behavior-context
AI: artificial intelligence
AVE: average variance extracted
CR: composite reliability
DOI: diffusion of innovation
HTMT: heterotrait-monotrait
ITE: intention to engage
mHealth: mobile health
PLS-SEM: partial least squares structural equation modeling

SF: social factors
TIB: theory of interpersonal behavior
VIF: variance inflation factor

Edited by A Kushniruk, E Borycki; submitted 27.04.24; peer-reviewed by CL Gan, Y Shahsavar; comments to author 07.07.24; revised version received 25.08.24; accepted 13.09.24; published 11.11.24.

Please cite as:
Yang Y, Tavares J, Oliveira T
A New Research Model for Artificial Intelligence–Based Well-Being Chatbot Engagement: Survey Study
JMIR Hum Factors 2024;11:e59908
URL: https://humanfactors.jmir.org/2024/1/e59908
doi:10.2196/59908
PMID:

Original Paper

The Doctors, Their Patients, and the Symptom Checker App: Qualitative Interview Study With General Practitioners in Germany

Christine Preiser[1], MA, Dr phil; Natalia Radionova[1], MA; Eylem Ög[1], MA; Roland Koch[2], Dr med; Malte Klemmt[3], Dr phil; Regina Müller[4], Dr rer med; Robert Ranisch[5], Prof Dr; Stefanie Joos[2], Prof Dr; Monika A Rieger[1], Prof Dr

[1]Institute of Occupational and Social Medicine and Health Services Research, University Hospital Tübingen, Tübingen, Germany

[2]Institute for General Practice and Interprofessional Care, University Hospital Tübingen, Tübingen, Germany

[3]Institute for General Practice and Palliative Care, Hannover Medical School, Hannover, Germany

[4]Institute of Philosophy, University Bremen, Bremen, Germany

[5]Faculty of Health Sciences Brandenburg, University of Potsdam, Potsdam, Germany

Corresponding Author:
Christine Preiser, MA, Dr phil
Institute of Occupational and Social Medicine and Health Services Research
University Hospital Tübingen
Wilhelmstr. 27
Tübingen, 72074
Germany
Phone: 49 70712984361
Email: christine.preiser@med.uni-tuebingen.de

Abstract

Background: Symptom checkers are designed for laypeople and promise to provide a preliminary diagnosis, a sense of urgency, and a suggested course of action.

Objective: We used the international symptom checker app (SCA) Ada App as an example to answer the following question: How do general practitioners (GPs) experience the SCA in relation to the macro, meso, and micro level of their daily work, and how does this interact with work-related psychosocial resources and demands?

Methods: We conducted 8 semistructured interviews with GPs in Germany between December 2020 and February 2022. We analyzed the data using the integrative basic method, an interpretative-reconstructive method, to identify core themes and modes of thematization.

Results: Although most GPs in this study were open to digitization in health care and their practice, only one was familiar with the SCA. GPs considered the SCA as part of the "unorganized stage" of patients' searching about their conditions. Some preferred it to popular search engines. They considered it relevant to their work as soon as the SCA would influence patients' decisions to see a doctor. Some wanted to see the results of the SCA in advance in order to decide on the patient's next steps. GPs described the diagnostic process as guided by shared decision-making, with the GP taking the lead and the patient deciding. They saw diagnosis as an act of making sense of data, which the SCA would not be able to do, despite the huge amounts of data.

Conclusions: GPs took a techno-pragmatic view of SCA. They operate in a health care system of increasing scarcity. They saw the SCA as a potential work-related resource if it helped them to reduce administrative tasks and unnecessary patient contacts. The SCA was seen as a potential work-related demand if it increased workload, for example, if it increased patients' anxiety, was too risk-averse, or made patients more insistent on their own opinions.

(JMIR Hum Factors 2024;11:e57360) doi:10.2196/57360

KEYWORDS

symptom checker app; qualitative interviews; general practice; perceived work-related psychosocial stress; job satisfaction; professional identity; medical diagnosis

Introduction

The right to diagnose is reserved for physicians and is a core element of professional authority [1]. Finding the proper diagnosis as part of GPs' expertise has been reported to contribute to physicians' perceived job satisfaction [2]. Patients also play an important role in the process. They first assess their symptoms, schedule an appointment, present their symptoms, and react to the diagnosis [3,4].

But what happens when technologies enter this process? Here we explore this with the example of the perspectives of general practitioners (GPs) on the Ada App, an internationally available symptom checker app (SCA).

Since the 1950s, physicians and computer scientists have been exploring how computers can be used to assist and improve the diagnostic process [5,6]. Part of the technologies are primarily designed to assist trained physicians or other health care professionals. Other technologies, such as symptom checkers, are designed for laypersons. A plethora of different apps exist [7]. The wording about what browser- and app-based symptom checkers actually deliver is rather heterogeneous, as the debate is ongoing [8]. For the purpose of this study, we use an operational definition: symptom checkers provide lay users with preliminary diagnoses, give a first sense of urgency, and suggest a course of action [9]. As such, we understand symptom checkers as "sociocultural artifacts" that are "(nonhuman) participants in networks of meaning and power relations" [10].

Symptom checkers are discussed ambivalently in the literature. Some expect that symptom checkers could outperform physicians' diagnostic capabilities [11], increase anxieties in patients [12], disrupt the doctor-patient relationship [10,13-15], or cause overuse of the health care system [16,17]. Further perspectives suggest that physicians' expertise and authority remain unchallenged by symptom checkers [10,18,19], patients could be empowered [12,18], and symptom checkers might have the potential to reduce physicians' workload and relieve an overburdened health care system [15,20,21]. Therefore, despite physicians not being the primary target group, symptom checkers might have far-reaching impacts on their work in terms of work content and work organization. These two dimensions are central to established models of perceived work-related stress [22-25].

As GPs are the primary access point to the health care system for patients [26], they might be particularly affected by laypersons' respectively patients' use of symptom checkers and the aforementioned impacts such as the use of the health care system or changes in the doctor-patient relationship. To date, little is known about the lived experiences and perspectives of GPs in regard to SCAs. A survey among GPs in the United States showed that only 30% had first-hand experiences with chatbots in health care [27]. One survey from Finland analyzed the experiences of health care professionals with a symptom checker at occupational health clinics and found that symptom checkers were of limited relevance in their daily work [15]. In both studies, the attitudes of the respondents mirrored the aforementioned discourse. While the studies give a first overview of relevant topics in the field, there is no in-depth analysis of how GPs experience SCAs as nonhuman participants in their daily work and how they see SCAs in relation to their own "apostolic function" [4], that is, their expectations on patients, their illness concepts, and personal and professional values and attitudes. Depending on how GPs disintegrate or integrate SCAs in their "apostolic function," this, too, might have further impacts in regard to job satisfaction and perceived work-related stress.

This study is embedded in the multidisciplinary joint project CHECK.APP, in which the ethical, legal, and social implications of SCAs in general practice in Germany are analyzed using the example of the Ada App [28]. The Ada App [29] had originally been conceptualized as a clinical decision support system but was later redesigned for laypersons. Laypersons enter their symptoms guided by the questions in the app and will be presented with a ranking of several potential diseases in combination with a sense of urgency and potential courses of action. It appeared to be a promising case for a wider phenomenon. Its advertising campaign #tellada won the German brand award in 2019 and included slogans such as "How are you? Be honest" [30,31]. At the same time, the actual number of users remained unclear. A health insurance had planned to implement the app, but it canceled its plans due to data privacy concerns [32-34]. As such, the app represents the aforementioned ambivalence around SCAs. In this study, we take a technology-in-practice approach [35] and will focus on GPs' lived experiences with the SCA and their perspectives on it. From an occupational health perspective, we are particularly interested in the following research questions: (1) How do GPs experience the SCA in relation to the macro, meso, and micro levels of their daily work? (2) How does this relate to their professional identity, job satisfaction, and psychosocial resources and demands?

Methods

Study Design

In a scoping review on the impacts of laypersons' use of SCAs on physicians in primary care, we showed that while some publications include the perspectives and voices of physicians, they primarily portray expectations rather than lived experiences [36]. In this empirical study, we aimed to fill this gap by exploring the experiences of GPs in Germany through the in-depth reconstructive analysis of semistructured qualitative interviews. The interview study was embedded in the joint project CHECK.APP which integrates the perspectives of users, experts, and GPs and works with a mixed methods research design, including a scoping review on ethical, legal, and social aspects [8], a survey among the wider German population, a diary and interview study with app users, qualitative interviews with GPs, and qualitative interviews with experts with regards to SCAs in general and the German health care system in particular [28].

Sample

We used several common recruitment strategies: as we know from previous studies, it is almost impossible to contact GPs directly by telephone, as practice assistants act as gatekeepers [37,38]. We contacted GPs by letter, email, or fax, depending on the information available through internet searches, and sent follow-up messages after 2-4 weeks. In addition, one of the GPs in the collaborative project (RK), who was not directly involved in this part of the study, contacted GPs through his institute's practice network. Our final sample included 8 GPs from different regions of Germany, representing different theoretically derived dimensions of our intended sample (Table 1).

Table 1. Sample of the study.

Relative dimensions and feature	Value (N=8), n
GPs[a]	
Age (in years)	
<45	3
45-55	2
55+	3
Gender	
Women	3
Men	5
Diverse	—[b]
Race	
Black	—
Person of color	—
White	8
Migrant	
Yes	1
No	7
Experiences with SCA[c]	
Yes	1
No	7
Use of digital tools	
No indicators	2
Indicators such as browser- or app-based system to book appointments, telemedical appointments, focus of the practice, and profile in social media	6
Practice	
Location	
Rural	3
Suburban	3
City	2
Structure of the practice	
Single practice	3
Joint practice	5
Medical center	—
Size of the practice	
<5 employees	3
>5 employees	5

[a]GP: general practitioner.

[b]Indicate the dimensions that, despite our efforts, were not represented in our sample.

[c]SCA: symptom checker app.

Data Collection

We developed the interview guide in several rounds, translating our research interests into final interview questions [39] (Multimedia Appendix 1). Two researchers (NR and CP) and one research assistant (EÖ) each conducted at least one of the interviews. Each interviewer obtained verbal and written consent for their interview. All 3 interviewers were trained sociologists with 3 (EÖ), 6 (NR), and 10 (CP) years of professional experience in health services research and occupational medicine. The interviews were conducted between December 2020 and February 2022 using the video conferencing tool VidyoConnect (version 21.6.3.17468; Vidyo Inc), which is provided by the University Hospital of Tübingen. The interviews lasted between 30 and 63 minutes and were recorded using an external audio recorder. The videos were not recorded. The files were transcribed by a certified office and pseudonymized by NR.

Data Analysis

The interviews were analyzed by NR, EÖ, and CP using the integrative basic method [40]. This reconstructive-interpretive method allowed the reconstruction of manifest and latent meanings by analyzing semantics, syntax, and metaphors. We chose agency and positioning as analytical approaches [40] because these proved particularly promising for understanding how GPs see and navigate themselves and their agency in a network of potentially conflicting participants and interests. We used the "Risk assessment of work-related psychological stress" of the Joint German Occupational Safety and Health Strategy (Gemeinsame Deutsche Arbeitsschutzstrategie) which operationalizes established models of work-related perceived stress [41] as a core sensitizing concept [42] for our analysis. It defines work content, social relations, work organization, work environment, and new forms of work as dimensions of work-related stress.

All interviews were interpreted line by line in order to identify the main motives and modes of thematization. Analytical case protocols were written for each interview. Interviews were continuously compared with each other. For the purpose of quality assurance, the analyses were discussed with MAR and researchers of the joint project [40]. Furthermore, we conducted a member check [43] with study participants and experts on SCAs in Germany in April 2022. The reporting of this study follows the Standards for Reporting Qualitative Research guidelines [44].

Ethical Considerations

The ethics committee of the Medical Faculty and University Hospital of Tübingen has approved the study (464/2020BO). All study participants were informed and gave consent verbally and in written form.

Results

Overview

The results reflect the perspectives of GPs and are presented along the main stages of the diagnosis process: the unorganized stage, the patient's decision to see a GP, and the shared process of exploring the patient's condition.

Doctor, Have You Heard of Ada?

The GPs in this study were mostly receptive to digitalization in their area of practice and used digital tools to varying degrees to manage patient volumes and streamline workflows. These included tools that helped to free up the telephone line, tools that helped with documentation, and tools that enabled easy and direct digital communication between patients and the practice. Several GPs made it clear that the digital tools were not suitable for all patient groups but helped to relieve capacity for those patients who did not use digital tools (eg, ensuring telephone availability). Although the GPs in this study were rather receptive to digital tools, only one GP was aware of symptom checkers in general and the introduced SCA in particular. This GP used it as an additional interlocutor:

> *GP 3: (2) So I personally use it for patients with … rare symptoms or unusual laboratory constellations ... too. So then ... I ask Ada, ... simply to get differential diagnoses again and then think about it: Could any of these differential diagnoses be correct?*
>
> *Interviewer: During the consultation?*
>
> *GP 3: No, in the evening on the couch (laughs).*
>
> *Interviewer: (laughs) So you don't finish work, but still google about Ada in the evening ... another symptom?*
>
> *GP 3: Exactly. So that's ... I usually already have an idea, and if it's a more complicated case, I also take a blood sample. And the blood values ... I go through them in the evening anyway. And if I then somehow come to a standstill or think maybe I've forgotten something in the differential diagnosis, then I often use Ada and ... see if it gives me new ideas, new impulses, yes.* [female, <45 years old, rural area]

This means that our results reflect expectations rather than experiences of the specific SCA and SCAs in general. However, the differences between GPs who were familiar with the SCA and those who were not were mainly in aspects of the app's practicality, not in perspectives on the app or their work.

The SCA in the Black Box of the "Unorganized Stage"

GPs normalized patients' desire to explore their condition and considered it helpful if patients had already thought about their symptoms, as this might help some patients to accept the outcome of the consultation.

> *So there is the example of the well-informed, intelligent patient who has already obtained preliminary information, which may not always be correct, but which sometimes makes it more difficult to find a diagnosis because it is unfiltered information. But I would say that, looking at all patients, it is easier if the patients have been informed in advance. You often have to revise patients' misjudgements, but they have already considered the issue in more detail. That is on average ... It certainly always depends on the type of patient you have in the*

practice, but ultimately it helps rather than harms if the patients are preinformed. [GP 8, male, 55+ years old, big city]

At best, the SCA could help patients reflect on their condition in preparation for a consultation with their GP, or reduce unnecessary patient anxiety. However, patients may include irrelevant information or omit relevant information, leading to misleading results. GPs portrayed patients as laypeople who often focus on and present subordinate aspects, use the wrong search terms when searching the internet, or are unable to assess the quality of information and follow the most appropriate course of action. According to GPs, internet searches in particular can increase patients' uncertainty and anxiety, as they may experience a flood of (negative) information and focus on the most serious potential outcomes. Some GPs preferred the SCA to internet searches because the information might be more evidence-based and focused. As such, GPs did not consider the SCA to be suitable for all patients, but only for those with eHealth literacy, general health literacy, and anxious patients, for whom the SCA might be the lesser of two evils compared to internet searches.

(...) then the app might come up with an initial result that is not quite as bad as a search engine. I don't know whether this will reassure patients, because someone who is worried about their health from home doesn't know whether they'll trust the app's result, and in the end the app has to present the result, so if in doubt, they'll go to their doctor. If one or two people with a cold don't rush straight to the GP's surgery, then perhaps there could be a marginal relief effect for GP surgeries... [GP 7, male, 45-55 years old, suburban area]

However, GPs mostly portrayed patients' path to information as a black box for GPs, as patients did not necessarily disclose their sources of information. It became clear that GPs also rarely actively asked patients where they got their information from, as they considered the source of information to be secondary once patients were in the consultation room. GPs therefore focused on what they could control: direct contact and discussion with patients. In addition, GPs emphasized that asking about the source of information would in most cases take up scarce time without adding value to their work. They felt it was more important to probe the patient's understanding of their condition ("disease model" [German: Krankheitsmodell]).

(Better) Too Early, (Than) Too Late—The SCA and Patients' Timing

GPs expressed that they were dependent on patients' decisions about when to see a doctor. They expected patients to go to the GP at the right time, but in their narratives, patients often went either too early or too late. This reasoning was embedded in the GPs' understanding of health care as a system with limited resources, with the scarcest resource being health professionals' time. In all interviews, GPs referred to the context of their working conditions, which may also shape their expectations of SCA. All GPs reported an intense workload with a high number of patient contacts. The workload has been increased by the COVID-19 pandemic. When in doubt, GPs preferred patients to seek help too early to avoid avoidable suffering. For them, the most important question about the SCA was: would it encourage patients to seek help at the right time? If it was too risk-averse, it would send patients too early and lead to oversupply, sabotaging the GP's mandate.

So, the back pain doesn't have to be something bad, you're doing this and this and this and we'll talk again in a fortnight and you'll tell me how it's gone then. So a watchful waiting approach. Whereas so-called red flags in general medicine, that is, obviously highly conspicuous and potentially dangerous symptom indications, lead to immediate consequences. This is how we handle evidence-based general medicine, at least in my practice. If the Ada app now turns everything into red flags ... it not only destroys our health care system but also unsettles patients and does the opposite of what is also very important in general practice, namely the prevention of overdiagnosis. For a variety of reasons, of course also for reasons of cost, the economic costs, but also (pauses) to ... yes, to keep the feeling of illness away from patients. [GP 5, male, 45-55 years old, rural area]

Conversely, if the SCA was not risk-averse enough, it would lead to underuse for patients. GPs attributed the potential for both to the SCA. The SCA could give wrong results and exacerbate the challenge of resource scarcity, or it could give patients the right sense of urgency and self-care instructions for simple cases. However, GPs felt that this risk or potential of the SCA would ultimately depend on the competence or personality of the user, not the technology itself.

Some GPs were considering how to integrate the SCA into their workflow, particularly with regard to practicalities that might have the potential to reduce unnecessary patient contact and thus workload. Some imagined that the results could be made available to GPs in advance.

So ultimately it's just a decision: He has to come or he doesn't have to come, or I have to visit him or not visit him. Those are the two things that are ultimately at the end of the decision-making tree. Yes. And I have to make the decision in the end when I have looked at it and realize: OK, it doesn't look very good somehow. Then you just have to say: You have to come. Or you just have to say: I'm going there. So, that ... (2) But you don't have to ask about all this previous history because it's already done. And, well, you have to be able to rely on it, that's the crucial thing (laughs). [GP 4, male, 55+ years old, suburban area]

In this way, the SCA would document the initial history and become a tool for communication between the patient and the GP, while the GP would ultimately decide how to proceed.

Finding Diagnosis—Humble Paternalism and the Art of Sense-Making

GPs described finding a diagnosis as a process that could involve several steps, patient contacts, and time (watchful waiting, see also the response of GP 5 above). From the GPs' perspective, both GPs and patients had a common goal: to find out what the patient had and to take the necessary measures. GPs and patients also shared the challenge of limited knowledge (see also the response of GP 3). However, from the GPs' perspective, GPs as medical experts and patients as medical laypeople had different knowledge limitations and different roles in the process. In the interviews, GPs positioned themselves as experts who, as such, were better able to search for information and assess information quality (including internet searches in medical databases and popular search engines), given their existing broad medical knowledge and years of professional experience. GPs positioned themselves as the medical authority who knew better, but who also had an ethical and legal responsibility to do their best to reach the right conclusion and treatment. GPs expected trust from their patients and offered skepticism in return. In the diagnostic process, they expected patients to share (information relevant to the GP) but not to overshare (information irrelevant to the GP) and thus to contribute smoothly to the GP's work.

> *So my favorite patient is the one who describes their complaints and not immediately their interpretation: "I read on Google, and that fits together." (laughs) "No, stop, (laughs) that's what I do (laughs). I would need your complaints (2) and not your interpretation of your complaints, please leave that to me." Hm, well, I would say about 20% of patients already prepare information from the internet. But to be honest, that bothers me more than it helps.* [GP 6, female, 55+ years old, big city]

The GPs in this study reacted negatively to terms such as "self-diagnosis," which we used in the interview guide in reference to common wording in the literature on SCA. Our interviewees rarely used the term "diagnosis" themselves. Rather, they spoke of "ideas," "assessments," "interpretations," "perceptions," "categorizations," etc in relation to SCA, but also in relation to their work. GPs considered it potentially helpful for patients to gain an initial understanding of their case, as long as they remained open to the GPs' guidance. They therefore describe the process of exploration as one of guided shared decision-making, with the GP guiding the patient and knowing best, but ultimately knowing that the patient will decide their own direction.

When it came to the SCA in this process, the GPs drew clear boundaries.

> *But yes, I actually see myself more as a symptom checker myself, so (laughs) people come to me and tell me their symptoms and I'm the one who helps them categorize them. (...) I think I can do that better than any app (laughs).* [GP 1, female, <45 years old, rural area]

The GPs ascribed to the SCA the potential to act as a nonhuman actor. The SCA could help guide the process of exploration—but it could also be an uncontrollable element that "spits out diagnoses," as one GP stated. GPs problematized the SCA if it disrupted their workflow and the doctor-patient relationship. GPs were concerned that the SCA could create avoidable extra work over which GPs had no control if GPs had to deal with the impact of the app on patients before they could focus on their work. The SCA was seen as an intruder if it cemented patients' insistence on their ideas and lay diagnoses, thus causing extra work:

> *So I think that if the app ... for example, really just says: "Go to your GP at short notice, he should clarify this," I have no problem with that at all, I think that's a great thing. Then I don't have to do a lot of educational work, I can look at the patient and then make a decision and discuss it with them. However, if the app now throws specific diagnoses into the room, and that's what it does here, i.e. infection with the bacterium Clostridium difficile, then it could be that the app leads to more work for me because then I first have to work through the app's diagnoses and reassure the patient, and above all I have to justify why I think this is not the right thing to do, even though the app suggests it. (...) I would find that unfortunate.* [GP 2, male, <45 years old, suburban area]

The GPs positioned themselves as the real intelligence against the artificial intelligence (AI) behind the SCA. From their point of view, the SCA could not offer anything that a human could not: the SCA had no empathy and could not offer physical examinations. GPs could send patients to a specialist when they reached the limits of their expertise, while the app would not send users to another, more appropriate app. They made it clear that the diagnostic process is more complex than a compilation of symptoms. From the GPs' perspective, diagnosing and finding the line between ill and not ill was a process of making sense of information. The SCA could collect potentially relevant information, and at best, organize it, but would not be able to make sense of it. GPs presented their view as limited but more objective, neutral, and less biased than that of the SCA, whose results were seen as one-dimensional. Ultimately, GPs concluded:

> *That depends on what the take-home message is for the patient at the end of the AI utilization, as I just said. If the take-home message is: You have complaints that (3) ... should result in a doctor's consultation within the next week - then I think that's fine. (00:30:55) If it's different, (...), then I would say to the patient: OK, now ... we're starting from scratch, please don't tell me anything about Ada, but tell me everything you told Ada again because I'm not going to let an AI take my decision-making away from me. If I go through a PHQ-9 questionnaire ... with my patient who has depression, then the questionnaire doesn't make the diagnosis, I do. But it can contribute to the validation of the diagnosis, in addition to my medical skills. And that's why AI can be part of the doctor-patient relationship, but in my view, it can't replace it - at least in the GP sector, where it's all about relationships.* [GP 5, male, 45-55 years old, rural area]

Discussion

SCAs as a “Proximate Future”

This study is one of the first empirical studies of GPs’ perceptions of SCA in their daily work. One of our main findings is that SCAs seem to be much less relevant in the current daily work of GPs in Germany than we had initially expected from the scientific literature and public discourse on symptom checkers. There may be several reasons for this. As we found out in the joint project, the use of symptom checkers is not widespread in the German population, that is, only a small proportion of patients actually use symptom checkers [45]. Furthermore, patients have been shown to be reluctant to share their source of information with physicians for fear of criticism from their physicians and disruption of the physician-patient relationship [46], so even if patients are using the particular SCA or other symptom checkers, they are unlikely to tell their GPs. In addition, GPs in this study reported that they did not ask patients about potential sources of information and did not differentiate between digital sources, but used the umbrella term “internet searches” for all types of digital tools. GPs might consequently have experiences with patients who use symptom checkers but might not be aware of this experience. To date, symptom checkers are rather a “proximate future” in German health care, a yet unachieved future envisioned by tech companies or other stakeholders as to be “just around the corner” and to be about to solve pressing issues of the present [47]. Centering “proximate futures” and the associated techno-utopian or techno-dystopian visions tend to distract from the unresolved issues of the present [47]. For us, then, the question is what we learn about the present work of GPs through the lens of GPs’ perceptions of symptom checkers.

GPs’ Perception of the SCA in Their Daily Work

GPs used language that presented the SCA and symptom checkers as “(nonhuman) participants in networks of meaning and power relations” [10]. Our results address SCA in GPs’ perspectives in relation to the macro, meso, and micro levels of their work.

At the macro level of the German health care system, we are to date not only facing a demographic change in the German population and among GPs (in 2022, about 36.6% of the GPs in Germany were 60 years or older [48]) but also a shortage of GPs due to lack of young practitioners, physicians’ wish for more part-time work, and an uneven distribution between urban and rural, as well as high- and low-income areas [49,50]. This is prognosed to intensify within the next decade. Some authors envision symptom checkers or other AI- or algorithm-based technologies as promising tools in a health care system of increased scarcity [17,51]. At present, symptom checkers are hardly used among the population in Germany [45,52] and—as shown by our data—are hardly known among GPs. GPs were critical about the future potentials of symptom checkers as useful participants in the German health care system if symptom checkers were too risk-adverse or not risk-adverse enough and led patients to the medically wrong time within the health care system. Patients’ individual decisions to follow the suggestions of the SCA might thus have impacts on the macro level and lead to over- or undersupply. Nevertheless, patients’ current decision-making process is more complex than following the results of a symptom checker [52,53].

At the meso level, GPs considered the SCA as a nonhuman participant on the organizational level of the practice and the interpersonal level of the physician-patient relationship. Digitization in the German health care system is a fragmented and slow process [54,55]. On an organizational level, GPs work in an environment that is characterized by a patchwork of technologies that lack interoperability. They perceived symptom checkers to feed into this pattern instead of improving it. Some GPs currently use digital tools to streamline patient contacts and administrative tasks. GPs welcomed the SCA as a potential future technology if it facilitated documentation and administration or helped to reduce unnecessary patient visits. In regards to the physician-patient relationship, GPs located the SCA in the black box of the patient’s search as part of the patient's “unorganized stage” [4] and “prediagnosis work” of patients [3], where it might complicate but not fundamentally damage the GPs’ and patients’ shared process of exploration. The SCA was attributed to the larger phenomenon of internet searches and “Dr Google” [56] which does not create a new phenomenon, but at worst reinforces an already existing one, namely patients who distrust their own health literacy, anxious patients, or patients who insist on their own assessment instead of trusting that of the GP. However, as the SCA is hardly present in the GPs’ daily work, it is not perceived as having a real impact on the meso level.

On the micro level, GPs positioned the SCA in relation to themselves. GPs in this study reacted negatively to terms such as “self-diagnosis,” often used in the popular and scientific literature on symptom checkers [16,46,57], and rejected the idea that the SCA could provide a proper diagnosis. They saw the SCA as lacking the holistic view, empathy, accumulated experience, and flexibility of human physicians ([58]). Similar to other studies, GPs saw diagnosis as a process of making sense of data and information and understood the SCA as monodimensional and static, full of correct information but unable to make full sense of the data (cf. [58]). They saw themselves as real, complex, and flexible, adaptive intelligence, acknowledging their own biases and knowledge limitations, compared to an artificial, schematic thinking intelligence ([10,20,58]). They therefore framed medicine as science and art, and technology as data minus art in the present and the future. As such, the SCA would not touch the GP’s professional expertise. This perspective also resonates with the current quality of symptom checkers. Databases in medicine might grow, but are biased and reinforce inequalities, and health data is of economic and political interest [59-62]. When it comes to SCAs, currently, the diagnostic accuracy of symptom checkers does not match the vision [63-65] even after decades of research [6,7,66], but they have the potential to provide an initial sense of urgency [17,63,64].

The SCA in Relation to Perceived Work-Related Resources and Demands

From an occupational health perspective, we were particularly interested in what our findings mean in the context of professional identity, subjective job satisfaction, work content, work organization, social relationships, and work-related psychosocial resources and demands [67]. The focus on the SCA in the following should not obscure the fact that the causes of the main work-related stressors for GPs can be found and should be addressed at the macro level [68,69].

Studies have repeatedly shown that GPs in Germany have a high workload, long working hours, and a higher prevalence of burnout [70,71]. The GPs in this study are no exception to this and faced an even increased workload due to the COVID-19 pandemic. Lack of enough time is a key stressor for GPs. As a consequence, their perspective on both patients and the SCA alike is structured by the question: Who or what will cost time for what? GPs see some aspects of the SCA as a potential resource in their work. Similar to other publications, the SCA is seen as a better source of health information for patients using internet searches than, for example, popular search engines or digital encyclopedias ([12,46]). The SCA is seen as potentially increasing work demands if it increases patients' anxiety, is too risk-averse and sends patients to the doctor unnecessarily or makes patients more insistent on their view, that is, if the effects of the SCA add to the already high workload or interfere with the doctor-patient relationship as an important part of GPs' work. GPs also see the SCA as a potential resource if it "streamlines access to physicians" [58], for example, if they can see and work with the results and inform patients whether or not they should come to the practice ([17]). The latter aspect has implications for the organization of work of GPs and possibly the practice team, for example, the question of when and how to integrate the results of the SCA into the workflow [72]. A study of a symptom checker embedded in an occupational health clinic shows that physicians do not integrate the results into their workflow [15]. As implementation theories and patient-reported outcome studies have shown, the additional information will only be used by health care professionals if it is meaningfully embedded into workflows [73-75]. In another recent study, the use of an AI-based chatbot did not reduce the workload of GPs or the duration of patient visits [76]. Furthermore, if the integration of SCAs leads to an even higher density of patient-related decisions to be made by the GP or the practice team, or if they are dealt with during breaks, the potential resource could become a demand.

For the GPs in this study, similar to other studies, patient work and especially the doctor-patient relationship is an important resource in terms of their subjective job satisfaction [68], as well as finding the proper diagnosis [2]. It has been shown that administrative tasks are the least favored tasks and inhibit job satisfaction [2,77]. Anything that threatens job satisfaction is viewed critically by GPs and—from an occupational health point of view—has to be seen as critical in terms of work-related psychosocial stress, even more so if it increases an already high workload. The SCAs were seen as a potential resource if they allowed GPs to focus on patient contact, thus reducing unfavored work tasks and allowing more time for favored work content. Some authors envision the physician-patient relationship to transform fundamentally through patients' use of SCAs [78]. GPs in this study appear to be protective of the physician-patient relationship. This study shows that GPs currently see the SCA as another means of diversifying patients' access to knowledge, rather than as a challenge to the doctor-patient relationship or their professional identity as medical experts. Treating the source of information as secondary and focusing on direct interaction with the patient can also be read as a resource for maintaining control in a network of increasing numbers of nonhuman participants. In this way, GPs also protect important pillars of their job satisfaction: the physician-patient relationship, their medical and interpersonal expertise, and the lead in the diagnostic process.

Strengths and Limitations

This study focuses on the perspective of GPs, although practice assistants also play an important role in general practice [79]. Initially, we had aimed to conduct 10 interviews with GPs that represent heterogeneous lived experiences in general practice [28]. Finding interview partners proved to be a challenge. GPs are known to have very high workloads and long working hours, so it is generally difficult to involve them in research studies [37,38,80]. The situation was exacerbated by the fact that we were looking for interview partners at the peak of the COVID-19 pandemic due to the funding period. We used various common strategies to reach out to GPs and eventually succeeded in finding 8 interview partners. The interviews were conducted between December 2020 and February 2022. We do not see striking differences between earlier and later interviews, probably also due to the fact that no COVID-19–related symptom checkers were used in Germany which might have introduced symptom checkers to a wider population. The sample only includes White GPs despite attempts to reach GPs of color and Black GPs. In addition to the aforementioned challenges, we assume unfitting sampling strategies and too homogeneous professional networks of the researchers as further reasons. It remains unclear, which impact race would have had on the perceptions of and experiences with the SCA. Within the given sample, the interview partners represent a variety of contexts and experiences and the data are rich enough to include shared patterns and conflicting perspectives [81]. Our results create resonance [40] with existing literature but also expand it. On the one hand, in terms of the issues raised by GPs in this study, we see strong parallels with the discourse on SCAs that we developed in our corresponding scoping review [36]. On the other hand, our data suggest that neither techno-utopian nor techno-dystopian visions of the literature are a dominant perceived reality in general practice in Germany. Instead, practitioners operating in a health care system of increasing scarcity display an attitude of techno-pragmatism in their daily work.

Conclusions

Some of the current scientific literature on symptom checkers presents rather techno-utopian visions of symptom checkers as a means to make health care more humane by supporting health care professionals and patients [58,82], democratizing access to knowledge and expertise [12,18], and relieving a burdened health care system [15,20,21]. Others emphasize techno-dystopian visions of a technology that outperforms or replaces humans [11,83] and becomes an uncontrollable participant in health care systems [16]. Our results show that symptom checkers are a "proximate future" [47] rather than a lived experience among GPs in Germany. The German Federal Association of Statutory Health Insurance Physicians established the so-called Patient-Navi [84], which shows similarities to symptom checkers, and some health insurance agents are working on their own symptom checkers. Our results help to understand the context in which these technologies might enter and to identify possible long-term effects in the future. Given the fact that the main challenges for GPs and patients can be found on the macro level of the health care system, this study also highlights that singular technological solutions do not solve complex societal issues, but can at best be one means in a plethora of means taken.

Acknowledgments

We thank all interview partners for their time and perspectives. The project was fully funded by the German Federal Ministry of Education and Research (grant 01GP1907A). The work of the Institute of Occupational and Social Medicine and Health Services Research Tübingen was supported by an unrestricted grant from the employers' association of the metal and electric industry Baden-Württemberg (Südwestmetall). The funding bodies had no role in the design of the study phases, during its execution, in the analyses of the data, or in the decision to submit any results. We acknowledge the support from the Open Access Publishing Fund of the University of Tübingen.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.
[DOCX File , 23 KB - humanfactors_v11i1e57360_app1.docx]

References

1. de Swaan A. The reluctant imperialism of the medical profession. Soc Sci Med 1989;28(11):1165-1170. [doi: 10.1016/0277-9536(89)90009-9] [Medline: 2660273]
2. Behmann M, Schmiemann G, Lingner H, Kühne F, Hummers-Pradier E, Schneider N. Job satisfaction among primary care physicians: results of a survey. Dtsch Arztebl Int 2012;101(11):193-200 [FREE Full text] [doi: 10.3238/arztebl.2012.0193]
3. Jutel AGG. Putting a Name to It: Diagnosis in Contemporary Society. Baltimore, Maryland: Johns Hopkins University Press; 2014.
4. Balint M. The doctor, his patient, and the illness. Lancet 1955;268(6866):683-688. [doi: 10.1016/s0140-6736(55)91061-8] [Medline: 14354967]
5. Engle RL, Davis BJ. Present concepts of the meaning and limitations of medical diagnosis. Arch Intern Med 1963;112:512-519. [doi: 10.1001/archinte.1963.03860040108009] [Medline: 14051369]
6. Engle RL. Attempts to use computers as diagnostic aids in medical decision making: a thirty-year experience. Perspect Biol Med 1992;35(2):207-219. [doi: 10.1353/pbm.1992.0011] [Medline: 1557299]
7. Pairon A, Philips H, Verhoeven V. A scoping review on the use and usefulness of online symptom checkers and triage systems: how to proceed? Front Med (Lausanne) 2022;9:1040926 [FREE Full text] [doi: 10.3389/fmed.2022.1040926] [Medline: 36687416]
8. Müller R, Klemmt M, Ehni HJ, Henking T, Kuhnmünch A, Preiser C, et al. Ethical, legal, and social aspects of symptom checker applications: a scoping review. Med Health Care Philos 2022;25(4):737-755 [FREE Full text] [doi: 10.1007/s11019-022-10114-y] [Medline: 36181620]
9. Chambers D, Cantrell AJ, Johnson M, Preston L, Baxter SK, Booth A, et al. Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review. BMJ Open 2019;9(8):e027743 [FREE Full text] [doi: 10.1136/bmjopen-2018-027743] [Medline: 31375610]
10. Lupton D, Jutel A. 'It's like having a physician in your pocket!' a critical analysis of self-diagnosis smartphone apps. Soc Sci Med 2015;133:128-135. [doi: 10.1016/j.socscimed.2015.04.004] [Medline: 25864149]
11. Gilbert S, Mehl A, Baluch A, Cawley C, Challiner J, Fraser H, et al. How accurate are digital symptom assessment apps for suggesting conditions and urgency advice? A clinical vignettes comparison to GPs. BMJ Open 2020;10(12):e040269 [FREE Full text] [doi: 10.1136/bmjopen-2020-040269] [Medline: 33328258]
12. Kuhn S, Jungmann SM, Jungmann F. Künstliche intelligenz für Ärzte und Patienten: "Googeln" war gestern. Dtsch Arztebl 2018;115(26):A1262-A1266.
13. Jutel A, Lupton D. Digitizing diagnosis: a review of mobile applications in the diagnostic process. Diagnosis (Berl) 2015;2(2):89-96 [FREE Full text] [doi: 10.1515/dx-2014-0068] [Medline: 29540025]
14. Denecke K, Tschanz M, Dorner TL, May R. Intelligent conversational agents in healthcare: hype or hope? Stud Health Technol Inform 2019;259:77-84. [Medline: 30923277]

15. Kujala S, Hörhammer I, Hänninen-Ervasti R, Heponiemi T. Health professionals' experiences of the benefits and challenges of online symptom checkers. Stud Health Technol Inform 2020;270:966-970. [doi: 10.3233/SHTI200305] [Medline: 32570525]
16. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ 2015;351:h3480 [FREE Full text] [doi: 10.1136/bmj.h3480] [Medline: 26157077]
17. Altendorf M, Luckmann M, Lägel R, Amelung VE. Use of symptom checkers to increase care efficiency—competent patients and better decisions? Monitor Versorgungsforschung 2023:66-70. [doi: 10.24945/MVF.02.23.1866-0533.2487]
18. Müschenich M, Wamprecht L. Health 4.0—how are we doing tomorrow? Bundesgesundheitsbl 2018;61(3):334-339. [doi: 10.1007/s00103-018-2702-6] [Medline: 29411045]
19. Miles O, West R, Nadarzynski T. Health chatbots acceptability moderated by perceived stigma and severity: a cross-sectional survey. Digit Health 2021;7:20552076211063012 [FREE Full text] [doi: 10.1177/20552076211063012] [Medline: 34917391]
20. Summerton N, Cansdale M. Artificial intelligence and diagnosis in general practice. Br J Gen Pract 2019;69(684):324-325 [FREE Full text] [doi: 10.3399/bjgp19X704165] [Medline: 31249070]
21. Gottliebsen K, Petersson G. Limited evidence of benefits of patient operated intelligent primary care triage tools: findings of a literature review. BMJ Health Care Inform 2020;27(1):e100114 [FREE Full text] [doi: 10.1136/bmjhci-2019-100114] [Medline: 32385041]
22. Bakker AB, Demerouti E, de Boer E, Schaufeli WB. Job demands and job resources as predictors of absence duration and frequency. J Vocat Behav 2003;62(2):341-356. [doi: 10.1016/s0001-8791(02)00030-1]
23. Karasek R, Theorell T. Healthy Work: Stress, Productivity, and the Reconstruction of Working Life. New York: Basic Books, Inc; 1990.
24. Johnson JV, Hall EM. Job strain, work place social support, and cardiovascular disease: a cross-sectional study of a random sample of the Swedish working population. Am J Public Health 1988;78(10):1336-1342. [doi: 10.2105/ajph.78.10.1336] [Medline: 3421392]
25. Siegrist J. Adverse health consequences of continued effort-reward imbalance at work. Int J Psychol 1992;27(3-4):599. [doi: 10.1037//1076-8998.1.1.27]
26. Döring A, Paul F. The German healthcare system. EPMA J 2010;1(4):535-547 [FREE Full text] [doi: 10.1007/s13167-010-0060-z] [Medline: 23199108]
27. Palanica A, Flaschner P, Thommandram A, Li M, Fossat Y. Physicians' perceptions of chatbots in health care: cross-sectional web-based survey. J Med Internet Res 2019;21(4):e12887 [FREE Full text] [doi: 10.2196/12887] [Medline: 30950796]
28. Wetzel A, Koch R, Preiser C, Müller R, Klemmt M, Ranisch R, et al. Ethical, legal, and social implications of symptom checker apps in primary health care (CHECK.APP): protocol for an interdisciplinary mixed methods study. JMIR Res Protoc 2022;11(5):e34026 [FREE Full text] [doi: 10.2196/34026] [Medline: 35576570]
29. Ada. GmbH AH. 2020. URL: https://ada.com/ [accessed 2024-07-29]
30. GmbH RfFM. German Innovation Award: Winners 2019. 2019. URL: https://www.german-innovation-award.de/en/winners/preis/gewinner/ada/ [accessed 2023-11-23]
31. GmbH RfFS. German Brand Award: Winners 2019. 2019. URL: https://www.german-brand-award.com/preistraeger/galerie/detail/26114-tellada.html [accessed 2023-11-23]
32. Tremmel S, Gieselmann H, Eikenberg R. Massive Datenschutzmängel in der Gesundheits-App Ada. 2019. URL: https://www.heise.de/news/Massive-Datenschutzmaengel-in-der-Gesundheits-App-Ada-4549354.html [accessed 2023-11-30]
33. ÄrzteZeitung. TK legt Kooperation mit Ada Health auf Eis. 2019. URL: https://www.aerztezeitung.de/Wirtschaft/TK-legt-Kooperation-mit-Ada-Health-auf-Eis-404721.html [accessed 2023-11-30]
34. Sauer J. Hausärzte-Chef erteilt Diagnose-App Ada klare Absage. 2022. URL: https://www.hausarzt.digital/politik/hausaerzteverband/bundesverband/weigeldt-erteilt-diagnose-app-ada-klare-absage-42908.html [accessed 2022-05-22]
35. Timmermans S, Berg M. The practice of medical technology. Sociol Health Illness 2003;25:97-114 [FREE Full text] [doi: 10.1111/1467-9566.00342] [Medline: 14498932]
36. Radionova N, Ög E, Wetzel A, Rieger MA, Preiser C. Impacts of symptom checkers for laypersons' self-diagnosis on physicians in primary care: scoping review. J Med Internet Res 2023;25:e39219 [FREE Full text] [doi: 10.2196/39219] [Medline: 37247214]
37. Güthlin C, Beyer M, Erler A, Gensichen J, Hoffmann B, Mergenthal K. Recruitment of family practitioners for research: experiences from five studies. Z Allg Med 2012;88(4):173-181. [doi: 10.3238/zfa.2012.0173-0181]
38. Hummers-Pradier E, Scheidt-Nave C, Martin H, Heinemann S, Kochen MM, Himmel W. Simply no time? Barriers to GPs' participation in primary health care research. Fam Pract 2008;25(2):105-112. [doi: 10.1093/fampra/cmn015] [Medline: 18417465]
39. Helfferich C. Die Qualität qualitativer Daten: Manual für die Durchführung qualitativer Interviews. Wiesbaden: VS Verlaf für Sozialwissenschaften; 2011.
40. Kruse J. Qualitative Interviewforschungin integrativer Ansatz. Weinheim, Basel: Juventa Verlag GmbH; 2015.
41. Gemeinsame Deutsche Arbeitsschutzstrategie. Consideration of psychosocial factors in risk assessment: recommendations for implementation in business practice. Gemeinsame Deutsche Arbeitsschutzstrategie. 2022. URL: https://www.gda-portal.de/EN/Download/pdf/Psychosocial-factors-in-risk-assessment.pdf?__blob=publicationFile&v=1 [accessed 2024-05-24]

42. Blumer H. What is wrong with social theory? Am Sociol Rev 1954;19(1):3-10. [doi: 10.2307/2088165]
43. Birt L, Scott S, Cavers D, Campbell C, Walter F. Member checking: a tool to enhance trustworthiness or merely a nod to validation? Qual Health Res 2016;26(13):1802-1811. [doi: 10.1177/1049732316654870] [Medline: 27340178]
44. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
45. Wetzel AJ, Koch R, Koch N, Klemmt M, Müller R, Preiser C, et al. 'Better see a doctor?' status quo of symptom checker apps in Germany: a cross-sectional survey with a mixed-methods design (CHECK.APP). Digit Health 2024;10:20552076241231555 [FREE Full text] [doi: 10.1177/20552076241231555] [Medline: 38434790]
46. Farnood A, Johnston B, Mair FS. A mixed methods systematic review of the effects of patient online self-diagnosing in the 'smart-phone society' on the healthcare professional-patient relationship and medical authority. BMC Med Inform Decis Making 2020;20(1):253 [FREE Full text] [doi: 10.1186/s12911-020-01243-6] [Medline: 33023577]
47. Dourish P, Bell G. Divining a Digital Future: Mess and Mythology in Ubiquitous Computing. Cambridge, Massachusetts: The MIT Press; 2011.
48. Bundesvereinigung K. Gesundheitsdaten: Niedergelassene Ärztinnen und Ärzte werden immer älter. 2023. URL: https://gesundheitsdaten.kbv.de/cms/html/16397.php [accessed 2023-11-30]
49. Bundesvereinigung K. Ärztemangel. URL: https://www.kbv.de/html/themen_1076.php [accessed 2023-11-30]
50. Bundesvereinigung K. Arztzeit-Mangel: Warum immer weniger Zeit für die Patientenversorgung da ist. URL: https://www.kbv.de/html/themen_38343.php [accessed 2023-11-30]
51. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mhealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
52. Koch R, Steffen MT, Wetzel AJ, Preiser C, Klemmt M, Ehni H, et al. An interface between ehealth, health literacy and health-related behavior: qualitative interview study on laypersons experiences with a mobile symptom checker app. JMIR Preprints Preprint posted online on May 17, 2024. [doi: 10.2196/preprints.60647]
53. Ubel PA. Beyond costs and benefits: understanding how patients make health care decisions. Oncologist 2010;15 Suppl 1:5-10 [FREE Full text] [doi: 10.1634/theoncologist.2010-S1-5] [Medline: 20237210]
54. Nohl-Deryk P, Brinkmann JK, Gerlach FM, Schreyögg J, Achelrod D. Barriers to digitalisation of healthcare in Germany: a survey of experts. Gesundheitswesen 2018;80(11):939-945. [doi: 10.1055/s-0043-121010] [Medline: 29301149]
55. Caumanns J. For discussion: the state of digitization of the German healthcare system. Z Evid Fortbild Qual Gesundhwes 2019;143:22-29. [doi: 10.1016/j.zefq.2019.04.002] [Medline: 31204267]
56. Lee K, Hoti K, Hughes JD, Emmerton LM. Consumer use of "Dr google": a survey on health information-seeking behaviors and navigational needs. J Med Internet Res 2015;17(12):e288 [FREE Full text] [doi: 10.2196/jmir.4345] [Medline: 26715363]
57. Aboueid S, Liu RH, Desta BN, Chaurasia A, Ebrahim S. The use of artificially intelligent self-diagnosing digital platforms by the general public: scoping review. JMIR Med Inform 2019;7(2):e13445 [FREE Full text] [doi: 10.2196/13445] [Medline: 31042151]
58. Blease C, Kaptchuk TJ, Bernstein MH, Mandl KD, Halamka JD, DesRoches CM. Artificial intelligence and the future of primary care: exploratory pualitative study of UK general practitioners' views. J Med Internet Res 2019;21(3):e12802 [FREE Full text] [doi: 10.2196/12802] [Medline: 30892270]
59. Broussard M. More Than a Glitch: Confronting Race, Gender, and Ability Bias in Tech. Cambridge, Massachusetts: MIT Press; 2023.
60. Noble S. Algorithms of Oppression: How Search Engines Reinforce Racism. New York: New York University Press; 2018.
61. Benjamin R, editor. Captivating Technology: Race, Carceral Technoscience, and Liberatory Imagination in Everyday Life. Durham: Duke University Press; 2019.
62. Cheung S. Disambiguating the benefits and risks from public health data in the digital economy. Big Data Soc 2020;7(1):205395172093392. [doi: 10.1177/2053951720933924]
63. Schmieding ML, Mörgeli R, Schmieding MAL, Feufel MA, Balzer F. Benchmarking triage capability of symptom checkers against that of medical laypersons: survey study. J Med Internet Res 2021;23(3):e24475 [FREE Full text] [doi: 10.2196/24475] [Medline: 33688845]
64. Hennemann S, Kuhn S, Witthöft M, Jungmann SM. Diagnostic performance of an app-based symptom checker in mental disorders: comparative study in psychotherapy outpatients. JMIR Ment Health 2022;9(1):e32832 [FREE Full text] [doi: 10.2196/32832] [Medline: 35099395]
65. Nateqi J, Lin S, Krobath H, Gruarin S, Lutz T, Dvorak T, et al. From symptom to diagnosis-symptom checkers re-evaluated : are symptom checkers finally sufficient and accurate to use? An update from the ENT perspective. HNO 2019;67(5):334-342. [doi: 10.1007/s00106-019-0666-y] [Medline: 30993374]
66. Wallace W, Chan C, Chidambaram S, Hanna L, Iqbal FM, Acharya A, et al. The diagnostic and triage accuracy of digital and online symptom checker tools: a systematic review. NPJ Digital Med 2022;5(1):118 [FREE Full text] [doi: 10.1038/s41746-022-00667-w] [Medline: 35977992]

67. Arbeitsschutzstrategie G. Consideration of Psychosocial Factors in Risk Assessment: Recommendations for Implementation in Business Practice. Gemeinsame Deutsche: Arbeitsprogramm Psyche; 2022.
68. Gündel H, Born M, Drews A, Mulfinger N, Junne F, Müller A. Gesundheit von Krankenhauspersonal: Kaum Spielräume für Verbesserungen. Dtsch Arztebl 2020;117(47):A-2281-B-1927.
69. West CP, Dyrbye LN, Shanafelt TD. Physician burnout: contributors, consequences and solutions. J Intern Med 2018;283(6):516-529 [FREE Full text] [doi: 10.1111/joim.12752] [Medline: 29505159]
70. Viehmann A, Kersting C, Thielmann A, Weltermann B. Prevalence of chronic stress in general practitioners and practice assistants: personal, practice and regional characteristics. PLoS One 2017;12(5):e0176658 [FREE Full text] [doi: 10.1371/journal.pone.0176658] [Medline: 28489939]
71. Dreher A, Theune M, Kersting C, Geiser F, Weltermann B. Prevalence of burnout among German general practitioners: comparison of physicians working in solo and group practices. PLoS One 2019;14(2):e0211223 [FREE Full text] [doi: 10.1371/journal.pone.0211223] [Medline: 30726284]
72. Kujala S, Hörhammer I. Health care professionals’ experiences of web-based symptom checkers for triage: cross-sectional survey study. J Med Internet Res 2022;24(5):e33505. [doi: 10.2196/33505]
73. Grol R, Wensing M, Eccles M, Davis D. Improving Patient Care: The Implementation of Change in Health Care. Chichester: Wiley Blackwell; 2013.
74. Antunes B, Harding R, Higginson IJ, EUROIMPACT. Implementing patient-reported outcome measures in palliative care clinical practice: a systematic review of facilitators and barriers. Palliative Med 2014;28(2):158-175 [FREE Full text] [doi: 10.1177/0269216313491619] [Medline: 23801463]
75. Maguire R, McCann L, Miller M, Kearney N. Nurse's perceptions and experiences of using of a mobile-phone-based advanced symptom management system (ASyMS) to monitor and manage chemotherapy-related toxicity. Eur J Oncol Nurs 2008;12(4):380-386. [doi: 10.1016/j.ejon.2008.04.007] [Medline: 18539527]
76. Mahlknecht A, Engl A, Piccoliori G, Wiedermann CJ. Supporting primary care through symptom checking artificial intelligence: a study of patient and physician attitudes in Italian general practice. BMC Prim Care 2023;24(1):174 [FREE Full text] [doi: 10.1186/s12875-023-02143-0] [Medline: 37661285]
77. Ochmann R, an der Heiden I, Bernhard J. IHP-Survey Deutschland 2022—Befragung der Hausärztinnen und Hausärzte in Deutschland: Bericht für das Bundesministerium für Gesundheit. Berlin: iGES Institut; 2022.
78. Fiske A, Buyx A, Prainsack B. The double-edged sword of digital self-care: physician perspectives from Northern Germany. Soc Sci Med 2020;260:113174 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113174] [Medline: 32659512]
79. Litchfield I, Burrows M, Gale N, Greenfield S. Understanding the invisible workforce: lessons for general practice from a survey of receptionists. BMC Primary Care 2022;23(1):230 [FREE Full text] [doi: 10.1186/s12875-022-01842-4] [Medline: 36085008]
80. Poß-Doering R, Kunz A, Pohlmann S, Szecsenyi J. Engaging family physicians in a research project to pilot an electronic health record. Z Allg Med 2019;95:515-519. [doi: 10.3238/zfa.2019.0515-0519]
81. Malterud K, Siersma VD, Guassora AD. Sample size in pualitative interview studies: guided by information power. Qual Health Res 2016;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
82. Fogel AL, Kvedar JC. Artificial intelligence powers digital medicine. NPJ Digital Med 2018;1:5 [FREE Full text] [doi: 10.1038/s41746-017-0012-2] [Medline: 31304291]
83. Wattanapisit A, Teo CH, Wattanapisit S, Teoh E, Woo WJ, Ng CJ. Can mobile health apps replace GPs? A scoping review of comparisons between mobile apps and GP tasks. BMC Med Inf Decis Making 2020;20(1):5 [FREE Full text] [doi: 10.1186/s12911-019-1016-4] [Medline: 31906985]
84. Versorgung ZK. Neues Angebot für Patient:innen zur Medizinischen Ersteinschätzung Online. 2021. URL: https://www.zi.de/das-zi/medien/medieninformationen-und-statements/detailansicht/7-dezember-2021 [accessed 2023-11-30]

Abbreviations

AI: artificial intelligence
GP: general practitioner
SCA: symptom checker app

Edited by Y Gong; submitted 14.02.24; peer-reviewed by K Savolainen, F Balzer; comments to author 07.06.24; revised version received 01.08.24; accepted 22.08.24; published 18.11.24.

Please cite as:
Preiser C, Radionova N, Ög E, Koch R, Klemmt M, Müller R, Ranisch R, Joos S, Rieger MA
The Doctors, Their Patients, and the Symptom Checker App: Qualitative Interview Study With General Practitioners in Germany
JMIR Hum Factors 2024;11:e57360
URL: https://humanfactors.jmir.org/2024/1/e57360
doi:10.2196/57360
PMID:

Original Paper

Mobile App for Improving the Mental Health of Youth in Out-of-Home Care: Development Study Using an Intervention Mapping Approach

Jinyoung Park[1], PhD; Jungeun Lee[2], PhD; Dabok Noh[3], PhD

[1]College of Nursing, Yonsei University, Seoul, Republic of Korea

[2]College of Social Science, Dankook University, Gyeonggi-do, Republic of Korea

[3]College of Nursing, Eulji University, Gyeonggi-do, Republic of Korea

Corresponding Author:
Dabok Noh, PhD
College of Nursing
Eulji University
553 Sanseong-daero, Sujeong-gu, Seongnam-si
Gyeonggi-do, 13135
Republic of Korea
Phone: 82 31 740 7415
Fax: 82 31 740 7359
Email: daboknoh@eulji.ac.kr

Abstract

Background: Youth in out-of-home care encounter substantial mental health challenges because of the absence of stable family and social support systems. Their vulnerability is heightened by trauma, neglect, and abuse. They struggle, especially when transitioning to independent living, coping with loneliness, anxiety, and pressure.

Objective: This study aimed to develop a mobile app with high accessibility and long-term continuous effects to support independent living and improve mental health among youth in out-of-home care. The approach used was the systematic and step-by-step intervention mapping (IM) framework.

Methods: The program was created using the IM framework and had 6 steps. Drawing from data from individual and focus group interviews and literature reviews, we developed a logical model of the problem. We established program outcomes and objectives, defining performance objectives and variable determinants. We identified theoretical and evidence-based methods that influence determinants. The app design integrated these methods into practical applications, allowing for the creation of self-management and emotional support tools. The development process included ongoing discussions between app designers and the research team to ensure that user needs and preferences were addressed.

Results: Individual interviews and focus group discussions revealed challenges in managing daily routines and regulating emotions. The program design was based on the transtheoretical model, social cognitive theory, and elaboration likelihood model. Key features included goal setting, structured routines, emotion recognition flashcards, character models demonstrating emotion regulation strategies, verbal persuasion, and self-monitoring tools to support habit formation and emotion regulation. An implementation plan was developed to facilitate the app's adoption, execution, and maintenance, while an evaluation plan was established, including app usage analytics, user logs, and feedback surveys. A randomized controlled trial will be conducted to assess the app's impact on mental health outcomes, focusing on reducing anxiety and depressive symptoms, improving emotion regulation, and enhancing daily living skills.

Conclusions: The IM framework was beneficial in developing a mobile app to enhance the mental health of youth in out-of-home care. The study produced a program grounded in theory and evidence that caters to the needs of these individuals. Further research should aim to verify the app's effectiveness in real-world settings and refine it continuously based on user input.

(JMIR Hum Factors 2024;11:e64681) doi:10.2196/64681

KEYWORDS

out-of-home youth; mental health intervention; mobile app; intervention mapping; youth; mental health; mHealth; mobile health; app; interview; need; focus group; emotion; emotional; young adult; independent living; emotional support; tool; emotion regulation; user; app usage

Introduction

In South Korea, "out-of-home youth" refers to individuals aged 9-24 years who have been separated from their guardians due to family conflicts, abuse, violence, neglect, family dissolution, or elopement and thus require social protection and support [1].

Youth in out-of-home care settings, such as child welfare facilities, orphanages, group homes, and youth shelters, are vulnerable to mental health issues due to the absence of stable family and social support systems [2,3]. This vulnerability is worsened by trauma, neglect, and abuse, which are prevalent among these youth [4-6]. As a result, they face long-term mental health challenges, and recent evidence reports that individuals who have lived in out-of-home care experience higher mortality rates in adulthood, largely due to increased rates of self-harm, accidents, and other mental health and behavioral factors [7].

These youth are especially vulnerable when transitioning from out-of-home care to living independently [4,5]. They often fail to make realistic and specific plans for independence, resulting in anxiety and pressure [8]. When individuals leave shelters, they face a sudden lack of support, leading to economic challenges such as housing instability, inadequate job opportunities, and financial hardship [9,10]. During the transition from youth to adulthood, individuals often face various crises, and the lack of a social support system can intensify feelings of isolation, loneliness, and despair, ultimately impacting their mental health [4,5,8]. Similarly, unaccompanied youth who are homeless experience high rates of obesity and poor diet quality, with many having deficiencies in essential nutrients. This exacerbates their physical and mental vulnerabilities, contributing to long-term health issues [11].

Various interventions have been proposed to address the mental health challenges of out-of-home care youth [12]. Some studies have attempted interventions for mental health, such as art therapy [13], cognitive behavior therapy [14,15], and family therapy [16]. These interventions predominantly rely on face-to-face counseling. Furthermore, a systematic review of interventions for care-experienced children and young people found that over half of the reviewed studies were conducted in the United States, raising concerns about regional bias and the generalizability of these interventions to other contexts [17].

In addition, according to a meta-analysis by Trubey et al [12], interventions aimed at improving the mental health of children and young people in out-of-home care showed significant positive effects on depression, anxiety, and social functioning difficulties in the short term (0-6 months); however, there was no evidence of effectiveness for long-term outcomes (>6 months) [12]. Furthermore, youth in out-of-home care often lose contact with the facilities when they become independent, leading to inconsistent management and heightened vulnerability [4,5]. Hence, youth living independently should receive ongoing and structured assistance.

The intervention mapping (IM) framework is effective in designing systematic, evidence-based health interventions [18]. For example, Svendsen et al [19] developed a self-management app for low back pain, improving adherence and outcomes through tailored advice. Hadjiconstantinou et al [20] addressed psychological and behavioral factors in their type 2 diabetes management program, and Wong et al [21] enhanced recovery and autonomy with a stroke self-management intervention. These studies demonstrate the use of IM in designing various health interventions.

In this study, we created a mobile app to enhance independent living, accessibility, and long-term mental health benefits for youth in out-of-home care based on the IM approach.

Methods

Overview

A mobile app was developed from October 2023 to June 2024 to enhance the mental health of youth in out-of-home care. The app was created using the IM approach, which is a systematic method that integrates theoretical and empirical evidence to create impactful health promotion programs tailored to the needs of youth in out-of-home care [18]. It involves engaging stakeholders throughout the program development process to ensure that their perspectives and needs are considered [22]. It consists of six steps: (1) a logical model of the problem, (2) program outcomes and objectives (logical model of change), (3) program design, (4) program production, (5) program implementation plan, and (6) evaluation [18].

Step 1: Logical Model of the Problem

Establish and Work With a Planning Group

Initially, we established a multidisciplinary planning group to incorporate expertise from various domains, with a particular focus on mental health improvement, youth care, and social welfare. The group consisted of a psychiatric and mental health nurse with a PhD, a PhD graduate in child and family studies, 2 social workers serving as directors at youth shelters, and a social worker employed at a youth shelter. The planning group worked closely with an app development team, which included an app developer, a designer, and a lead app developer who also served as the server and web developer, to ensure that the digital platform was appropriately tailored to meet the needs of youth in out-of-home care. To gather feedback directly from the target population for this intervention without exposing them to the pressures or responsibilities of participating in a formal planning group, we decided not to include youth in the planning group. Instead, we conducted individual interviews to assess their needs.

Needs Assessment

We conducted a needs assessment to analyze respondents' mental health and quality of life issues, as well as their underlying causes. The methods used for this assessment included conducting a literature review, holding individual interviews with youth living in youth shelters, and organizing focus group interviews with shelter workers.

The target population for this intervention was youth aged 18-25 years who receive out-of-home care at youth shelters and are preparing for independent living after leaving such care. This focus was motivated by two main reasons. First, previous research indicated that youth in out-of-home care are vulnerable to mental and physical health issues due to the absence of stable family and social support systems [2,3,7]. Second, youth in out-of-home care in South Korea must transition to independent living between the ages of 18 and 25 years. Thus, our target population consisted of individuals aged 18 to 25 years who are in the process of transitioning to independence.

Individual Interviews With Youth

We conducted individual interviews with 5 youths residing in 2 youth shelters. The participants were chosen purposefully to include those who could share their experiences living away from family and preparing for independence. The interview questions were designed to gather insights into the participants' experiences with mental health difficulties and their specific needs, which were crucial in structuring the intervention. The participants were asked questions such as "Have you ever experienced psychological or emotional difficulties?" and "Have you sought help during tough times?" In addition to these, they were asked about concerns related to independent living, including "What is your biggest worry as you prepare to live independently?" and "What challenges do you expect to face in your daily life after leaving the youth shelter?" as well as "What emotional or psychological challenges do you anticipate after becoming independent?" To explore their perspectives on digital interventions, participants were asked, "If there was a mental health program available through a mobile app, what kind of support would you want?" Finally, they were asked, "What advice would you like to give to the program developers?" and "Is there anything else you would like to add?" The interviews, conducted in the counseling rooms of the shelters, lasted around 1 hour.

Focus Group Interviews With Shelter Workers

We also conducted 2 focus group interviews with 10 shelter workers from 2 youth shelters. These workers were selected based on their experience of working at youth shelters for 1 year or more and their deep understanding of the experiences of out-of-home youth. The interviews focused on discussing the mental health challenges faced by these youth, instances where assistance was sought, and the need for intervention programs. To begin the interviews, the participants were asked to introduce themselves by sharing how long they have worked in their roles and a brief overview of their responsibilities. We then explored perspectives on the mental health status of youth residing in youth shelters, asking, "What do you think about the mental health condition of these youths?" As the conversation progressed, we delved into specific experiences while working in youth shelters: "Have you observed any cases of mental health difficulties among these youths?" Thereafter, we discussed experiences in assisting youth in need: "Have you encountered any instances where a youth sought help with mental health issues?" To further understand their insights, we asked, "In your opinion, what do you think is beneficial for the mental health of youths?" We also discussed the potential of mobile applications in this context, asking, "If a mental health promotion program were to be delivered through a mobile app, in what ways do you think it could be helpful?" To conclude the interviews, we asked, "What advice would you like to give to the program developers?" and "Is there anything else you would like to add?" Each interview lasted about 1 hour and was conducted in the meeting rooms of the two youth shelters.

All interviews were audio-recorded and subsequently transcribed. The interviews were exploratory, focusing on gathering real-world needs and collaborating with stakeholders. Insights obtained from the open-ended, semistructured interviews were integrated iteratively, and any emergent trends were reviewed with the planning group.

Finally, we outlined the intervention's context, which includes the population, community, and setting. In the final step of Phase 1, we defined the program's goals.

Step 2: Program Outcomes and Objectives: Logical Model of Change

In the second step, we identified the expected outcomes from the results of Step 1. We defined the performance objectives for these outcomes and chose their determinants. A matrix was created to outline the change objectives that would impact the determinants and help achieve the performance objectives.

Step 3: Program Design

In Step 3, we conceptualized and designed the program, selecting theory- and evidence-based change methods to achieve the change objectives determined in Step 2. Theoretically, grounded change methods were translated into practical applications, with behavior change techniques (BCTs) [23] systematically applied. This phase focused on ensuring that each component of the program was based on evidence-based methods and tailored to meet the specific needs of youth in out-of-home care. Key tasks included identifying and selecting appropriate change methods based on established theories like the transtheoretical model (TTM), social cognitive theory (SCT), and elaboration likelihood model (ELM).

The TTM was chosen for its effectiveness in promoting behavior change through consciousness-raising and reinforcement management. Liu et al [24] demonstrated that TTM-based interventions help individuals understand the benefits of behavior change and maintain positive habits. By applying TTM, we aimed to support users in recognizing the value of structuring their daily routines and encouraging long-term adherence to healthy behaviors.

The SCT was incorporated for its focus on self-efficacy and observational learning. The theory postulates that internal processes, such as goal setting and self-efficacy, lead to behavioral outcomes [25], which informed our decision to apply SCT. This model supports users in setting goals, building self-efficacy, and learning emotion regulation strategies by observing others.

The ELM was selected to enhance emotional recognition through repeated exposure to emotional stimuli. Petty and Cacioppo [26] emphasized that repeated messaging can significantly influence attitudes and behavior, making ELM a suitable model for improving users' emotional awareness and regulation. This theory helps explain how consistent exposure to emotional content can lead to deeper emotional understanding and behavior change.

These theoretical methods were then translated into practical applications by designing specific features within the mobile app to deliver the selected change methods effectively. We prioritized a user-centered design approach, as recommended by Johnson et al [27] and Bakker et al [28], for enhancing the app's usage and effectiveness. The practical applications were developed to be user-friendly and engaging, allowing users to easily interact with the app and incorporate desired behaviors into their daily routines. The aim was to create a structured theoretical framework to help youth in out-of-home care manage their daily lives independently and improve their emotion regulation abilities.

Step 4: Program Production

Each component of the intervention app was developed systematically by integrating theoretical foundations, designing the user interface, and implementing functionalities. This step aimed to ensure the app's practical applicability and functionality. Educational content, messages, and images depicting characters and emotions were also developed in a structured manner. An iterative development process was employed, working with the app development team to check and revise the prototype. Regular meetings were held with the lead app developer to communicate requirements, review screen designs, and discuss necessary modifications. These discussions were essential to consider and finalize the design modifications and ensure that the app met all the requirements.

Step 5: Program Implementation Plan

The fifth step involved creating a comprehensive program implementation plan. This plan included strategies to ensure that the mobile app, which was designed to improve the mental health of youth in out-of-home care, was effectively adopted, used, and maintained over time.

Step 6: Evaluation Plan

The sixth step was to develop a plan to evaluate if the app had achieved its goals and objectives. This plan focused on conducting a thorough assessment of the app's effectiveness, covering both process and outcome evaluations.

Ethical Considerations

This study was approved by the Eulji University Institutional Review Board (approval EU23-43). Participants in the needs assessment interviews received detailed information about the study, and written consent was obtained before their participation. They were assured of their right to withdraw consent verbally at any stage of the interview, even after it had started, without facing any negative consequences. All collected data were anonymized and securely stored in encrypted files, accessible only to the research team. As an expression of gratitude, participants were offered a gift card worth 50,000 Korean won (approximately $35 USD).

Results

Step 1: Logical Model of the Problem

Our literature review revealed that young people living in youth shelters often experience a high prevalence of mental health issues, like anxiety, depression, suicidal thoughts, and a poor quality of life [4]. After leaving youth shelters, they often experience isolation, loneliness, despair, and suicidal thoughts as they navigate the challenges of independent living [7].

Individual interviews were conducted with 5 adolescents aged 18-23 years who had been living in shelters for 5 months to 5 years. Many of the participants appeared to be either overweight or obese. Participants have experienced feelings of loneliness, depression, and, in some cases, suicidal thoughts. When facing difficult times, they sought support primarily from their romantic partners, peers, and shelter staff. When asked about their biggest concerns regarding independent living, financial stability was the primary worry, especially related to securing housing and managing living expenses after leaving the shelter. Day-to-day challenges, such as waking up on time and preparing meals, were significant concerns for many due to their limited experience in these areas. They described their current unstructured daily lifestyle—waking up late, spending most of their time lethargically in the shelter, and frequently looking at their smartphones. This reflects a lack of knowledge about healthy daily routines and the absence of motivation to maintain them (lack of knowledge about the advantages of daily routines, lack of motivation to structure daily routines). Psychologically, they expressed fears of dealing with loneliness after becoming independent. They mentioned that they would appreciate having a mobile app they could use daily, especially one that included visual elements like characters to make the experience more engaging.

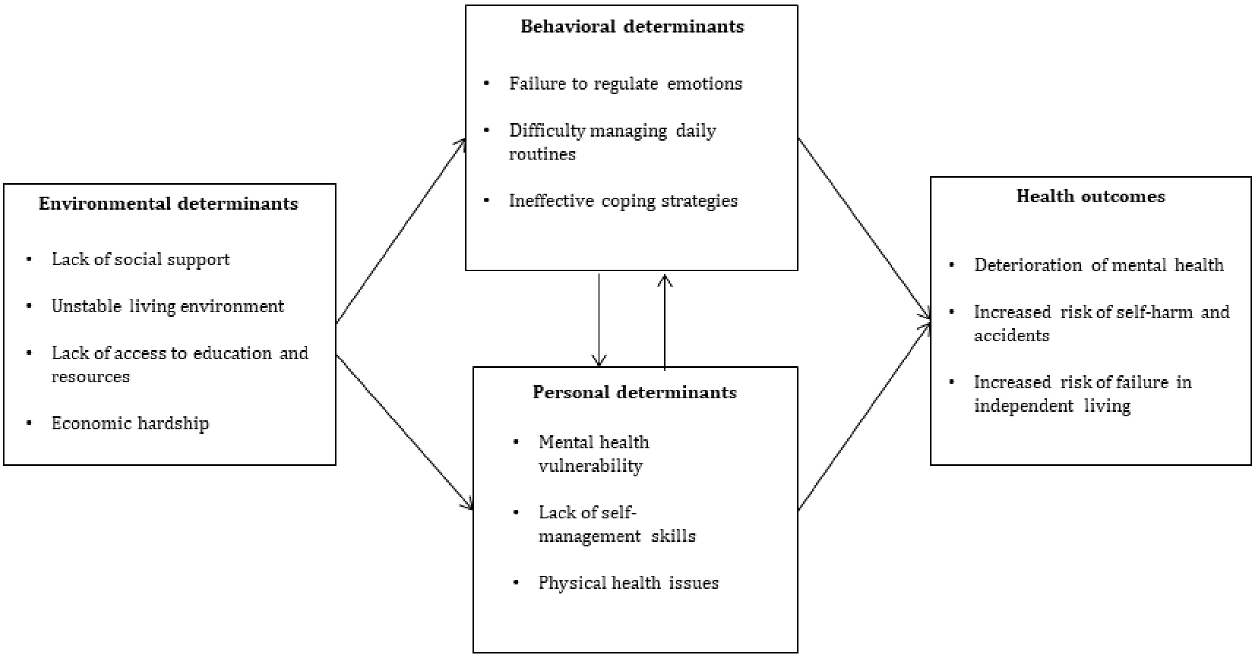
Focus group interviews were conducted with shelter staff between the ages of 29 and 59 years, who had 1 to 9 years and 11 months of work experience. They specialized in counseling psychology, social welfare, youth education, counseling, and youth studies. The staff explained that the mental health of adolescents entering shelters recently was worse compared to those in the past, with many experiencing higher levels of depression and anxiety. Most adolescents spend the majority of their day on their phones, engaging in social media, calls, YouTube videos, and games. During their days off, they often wake up around 1-2 PM, have meals at the shelter, socialize with friends, and return just before bedtime. This indicates a lack of motivation to adopt healthy behaviors, such as exercise or balanced meals, and the absence of habits that support a healthy lifestyle (lack of knowledge about the importance of healthy behaviors, lack of habit in maintaining healthy behaviors).

The staff observed that some adolescents at the shelter have difficulties with emotional regulation, including depression and anxiety, and actively seek counseling. At times, some adolescents stay up past bedtime due to self-harm urges and turn to teachers for assistance. They also discussed how adolescents struggle to regulate their emotions even in the presence of peers and teachers. Furthermore, when these adolescents transition to living independently, they experience loneliness and depression, prompting them to seek support. This is connected to the lack of strategies for emotional regulation and the absence of motivation to implement those strategies consistently (lack of knowledge about emotional regulation strategies, lack of motivation to regulate emotions).

The staff mentioned that some adolescents struggle to manage their daily lives and emotions even after becoming independent. In some cases, they choose to return to the shelter. In addition, some adolescents do not maintain contact after leaving. Therefore, the staff suggested that a mobile app could assist youth in managing their daily lives and emotions postindependence. The program will help youth develop the knowledge and habits necessary for managing their emotions, as well as motivate them to consistently track and regulate their emotional states (lack of habit in tracking emotions, lack of acceptance of one's emotions).

Out-of-home youth living in shelters encounter challenges due to a lack of support from family and caregivers, resulting in feelings of anxiety, depression, and a tendency to engage in self-injury. In their transition to independent living, many youth face loneliness, leading to feelings of powerlessness, difficulties in daily life management (lack of structured routines, poor habit formation), and struggles with emotional regulation (lack of emotional regulation strategies, lack of motivation). The proposed program aims to empower youth in out-of-home care to address these challenges by effectively managing their daily lives, regulating their emotions, and improving their stability and emotional well-being (Figure 1).

Figure 1. Logical model of the problem.



Step 2: Program Outcomes and Objectives: Logical Model of Change

Based on the findings of the initial phase, the expected program outcomes were: (1) young people managing their daily lives independently and (2) young people enhancing their emotion regulation. Specific performance objectives for independently managing daily lives included structuring and maintaining daily routines, as well as developing healthy habits. In addition, performance objectives for enhancing emotion regulation in young people involved in expressing their emotions and implementing strategies for emotion regulation.

There is evidence that knowledge influences attitude, leading to behavioral changes. Habit formation is crucial for lasting behavioral change, especially when dealing with changing attitudes and temptations [29]. In this study, knowledge, attitude, and habits were chosen as determinants. For out-of-home youth with little family and adult support, forming habits is a key factor in being able to manage daily life and emotions independently. The change objectives are outlined for each of the four performance objectives in the three determinant areas (Table 1).

Table 1. Matrix of change objectives.

Outcome and PO[a]	Determinant		
	Knowledge or cognition	Attitude or motivation	Habit
Young people manage their daily lives independently			
PO 1. Young people structure and maintain their daily routines (e.g., time management, scheduling)	K.1. Identify the advantages of structuring daily routines	A.1. Recognize the advantages of structuring daily routines	• H.1.a. Monitor the progress of achievement of daily routines • H.1.b. Stick to routines to turn them into habits
PO 2. Young people adapt healthy behaviors (e.g., regular exercise, balanced diet)	K.2. Identify the importance of healthy behaviors	A.2. Recognize the importance of healthy behaviors	• H.2.a. Monitor the progress of achievement of healthy behaviors • H.2.b. Stick to healthy behaviors to turn them into habits
Young people enhance their emotion regulation abilities			
PO 3. Young people express their emotions	K.3. Identify and name a range of emotions	A.3. Express an accepting attitude towards one's own emotions	• H.3.a. Incorporate emotional expression into daily routines • H.3.b. Track one's own emotions
PO 4. Young people implement strategies to regulate emotions	K.4. Identify the strategies to regulate one's own emotions	A.4. Express an attitude of openness to regulate emotions	• H.4. Incorporate emotion regulation into daily routines

[a]PO: performance objective.

Step 3: Program Design

The theory-based program design and practical applications to achieve change objectives are detailed in Table 2. Core components of the program include structuring and maintaining routines, and regulating emotions delivered by a smartphone app. The design was based on behavioral change theories to provide users with tools to manage their daily lives and emotions. BCTs [23] were applied to operationalize the theoretical methods and provide structured guidance to users.

Table 2. Explanations of changeable determinants, theory- and evidence-based change methods, and practical applications.

Determinant, change objective, and theory	Method	BCTs[a]	Practical application	Text description
Knowledge or cognition				
• K.1. Identify the advantages of structuring daily routines • K.2. Identify the importance of healthy behaviors				
Transtheoretical model	Consciousness raising	4.1 Instruction on how to perform a behavior	Intro session with educational messages on the app	The intro session provides users with educational messages explaining the importance of structuring daily routines and maintaining healthy behaviors.
• K.3. Identify and name a range of emotions				
Elaboration likelihood model	Message repetition	4.1 Instruction on how to perform a behavior	Emotion flashcards	Flashcards repeatedly show various emotions to help users recognize and label their emotions more accurately.
• K.4. Identify strategies to regulate one's own emotions				
Social cognitive theory	Observational learning	2.1 Others monitoring with awareness	Observing a character model regulating their emotions	Users learn emotion regulation strategies by observing how character models handle their emotions.
Attitude or motivation				
• A.1. Recognize the advantages of structured daily routines • A.2. Recognize the importance of healthy behaviors				
Social cognitive theory	Goal setting	1.1 Goal setting (behavior)	Setting daily routines, including healthy behaviors	Users set personal goals for structured routines and healthy behaviors, such as regular exercise and a balanced diet, to develop positive habits
• A.3. Express an accepting attitude towards one's own emotions				
Elaboration likelihood model	Message repetition	5.4 Self-assessment of affective consequences	Selecting an emotion flashcard and choosing the corresponding word for the emotion	The flashcards and repeated exposure help users build a positive attitude toward acknowledging their emotions.
Social cognitive theory	Observational learning	3.3 Social support (emotional)	Character showing empathy toward their emotions	Character models in the app show empathy towards users' emotions, helping to build a supportive and motivating environment for emotion regulation.
• A.4. Express an attitude of openness to regulate emotions				
Social cognitive theory	Verbal persuasion	15.1 Verbal persuasion to boost self-efficacy	Offering positive reinforcement through verbal encouragement	Verbal encouragement, such as "you can do it," boosts users' confidence in their ability to regulate emotions and motivates them to make initial efforts.
Habit				
• H.1.a. Monitor the progress of achievement of daily routines • H.1.b. Stick to routines to turn them into habits • H.2.a. Monitor the progress of achievement of healthy behaviors • H.2.b. Stick to healthy behaviors to turn them into habits				
Social cognitive theory	Self-monitoring	2.3 Self-monitoring of behavior	Tracking daily achievements using a calendar function	Users track daily routines using a calendar that displays progress from the start of the month to the present day. Progress is visualized with a cake filling piece by piece on the daily screen.

Determinant, change objective, and theory	Method	BCTs[a]	Practical application	Text description
Social cognitive theory	Feedback	2.2 Feedback on behavior	Informing and displaying routine accomplishments	The program provides feedback by completing the cake visual when all daily routines are achieved, motivating users to maintain routines.
Transtheoretical model	Reinforcement management	10.2 Material reward	Rewards for maintaining routines; points for achievements	Users earn points for maintaining routines and healthy behaviors, reinforcing positive habit formation.
• H.3.a. Incorporate emotional expression into daily routines • H.3.b. Track one's own emotions				
Social cognitive theory	Feedback	2.2 Feedback on behavior	Notifications to encourage the expression of emotions	If users do not log their emotions, the app sends a pop-up notification reminding them to record their emotions as part of their daily routine.
Social cognitive theory	Self-monitoring	2.4 Self-monitoring of the outcome of behavior	Visualizing weekly emotions in a graph	Users can monitor their emotions through visual graphs that summarize their emotional states over time, encouraging self-reflection and awareness.
• H.4. Incorporate regulating emotions into daily routines				
Transtheoretical model	Reflection	15.4 Self-talk	Keeping an emotion diary	Users keep an emotion diary to reflect on their emotional states, helping to improve emotional regulation.
Transtheoretical model	Reinforcement management	10.2 Material reward	Points for completing each routine task and diary entry	Users earn points for completing routine tasks and maintaining an emotion diary, encouraging consistency in emotional regulation.

[a]BCT: behavior change technique.

Consciousness-raising techniques from the TTM helped users understand the benefits of organizing routines and fostering healthy behaviors [30]. Emotion recognition was enhanced using the ELM through repeated exposure to emotion flashcards [26]. Observational learning from the SCT was applied by character models demonstrating emotion regulation [25].

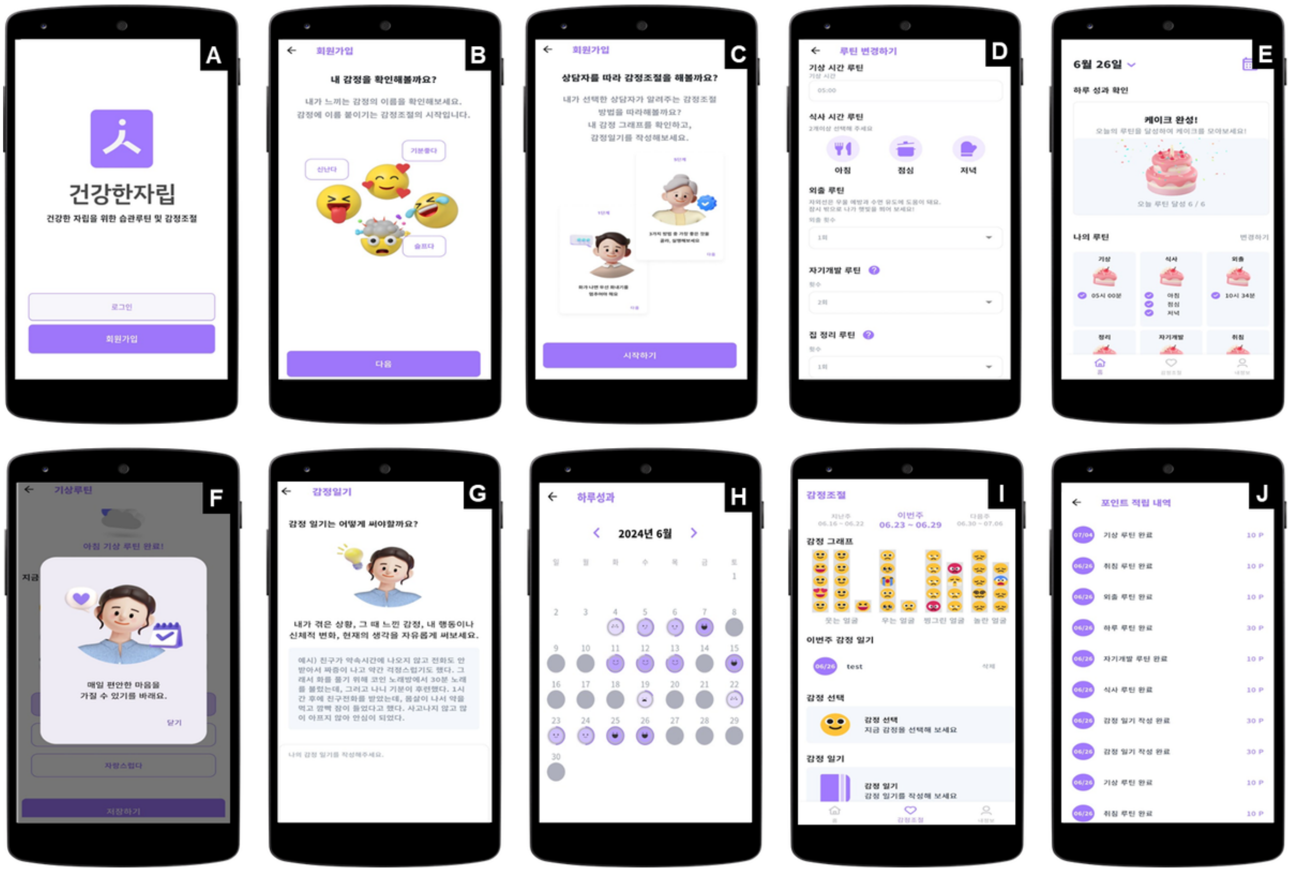
To further enhance motivation, the SCT's goal setting allowed users to establish routines and healthy behaviors independently, while verbal persuasion and supportive character models built confidence in emotion regulation. Self-monitoring and reinforcement strategies from the SCT and TTM supported habit formation and emotional regulation [31,32].

These theoretical methods were operationalized into practical strategies through the application of BCTs [23], including goal setting, self-monitoring, and reinforcement, which facilitated users in adopting and sustaining behavior changes in their daily routines.

Step 4: Program Production

The fourth step involved developing the specific components of the program to give it a viable form. To create an app for youth in out-of-home care, the program was designed by incorporating information gathered in Steps 1 to 3. Prototypes were utilized in designing the app's structure and main features, focusing on making the interface intuitive and visually appealing for easy use by the youth. Figure 2 offers an overview of the mobile app's user interface, showcasing the main components and navigation elements. It presents the home screen comprehensively, including the dashboard, navigation menu, and key interactive features like touchable buttons and icons.

Figure 2. Overview of the app interface. A: Login screen. B, C: Introduction to the app's purpose and features for managing routines and regulating emotions. D: Screen for setting and modifying daily routines, such as wake-up time, bedtime, meal frequency, and personal development activities. E: Visual representation of daily achievements using a cake image that fills as routines are completed. F: Feedback section where users select their feelings after completing routines and receive messages from a counselor. G: Emotion diary for recording and managing feelings. H: Calendar that shows monthly progress. I: Graph that visualizes emotional changes over time. J: Points section where users can view and track points earned for completing routines.



Once users sign in and provide their information, they can use the app after obtaining approval from the administrator. Approved users receive guidance on using the app, including details regarding the importance of daily routines and healthy behaviors. Users then select a character to represent themselves and another character to act as their counselor. Subsequently, they set their daily routines, which include wake-up time, bedtime, number of outings per day, number of meals per day (at least 2), personal development activities (eg, reading, exercising, and studying), and household chores (eg, cleaning and washing dishes). The default home screen displays the current date and provides visual feedback, showing a cake image that is filled in based on routine completion. In addition, the calendar function allows users to visually track their daily achievements and serves as a self-monitoring tool to easily observe their progress.

During the day, users can input whether they completed their routines and the time spent doing so. They can then select their emotions from a list of 12 emoticons, along with appropriate verbal expressions. The counselor's character provides messages based on the selected emotions to create an emotionally supportive environment. For instance, if a user chooses "anxious," they receive advice such as, "Something is making you anxious. Consider the causes and prioritize tasks to write down. Talking or writing about your problems can help calm your negative emotions." This can aid users in recognizing and managing their emotions.

The Emotion Regulation tab shows facial images gathered over a week in a graph, allowing users to visually monitor their emotional changes. In addition, users can write emotion diaries to document and analyze their feelings, aiding in improved understanding and management of emotions.

In the My Info tab, users can view basic information such as their characters, accumulated points, and set routines. They can also check and modify their routines at any time. The Settings tab offers options to change passwords, set notifications, and log out or reset accounts.

The development of the mobile intervention app followed a structured process that included defining requirements, screen planning, development, testing, and deployment. The development environment consisted of both a web-based system for data and user management and a mobile application. The mobile application was developed as a hybrid app designed for both Android and iOS platforms to maximize accessibility for users. The app was registered and made available through the Google Play Store and Apple App Store. User data was securely stored in the MySQL database, and the server was hosted on Amazon Web Services to ensure robust data protection. Security measures were carefully implemented to protect user information. User authentication was managed using JSON

Web Token, providing a secure method for validating user sessions. All data transmissions were encrypted using SSL/TLS through the HTTPS protocol, ensuring that sensitive information remained protected during communication between the app and the server. In addition, server access control was strictly managed to further enhance security.

Step 5: Program Implementation Plan

The fifth step is a detailed plan on how we envision the adoption, implementation, and long-term maintenance of the app within shelter settings. The program uses outcomes (adoption and implementation) that were developed with specific performance objectives for both the shelter staff and youth. Using normalization process theory [33], we identified changeable determinants that could influence the successful integration of the app into daily routines. These determinants were crossed with performance objectives to establish clear change objectives. The theoretical foundation for our strategies is based on BCTs [23], which informed the methods for fostering app adoption and ensuring its sustained use. Practical strategies include training sessions for staff, peer ambassador programs, and motivational tools for the youth, which were all designed to encourage continuous engagement with the app. Table 3 outlines the detailed performance objectives, change objectives, and corresponding strategies for both adoption and implementation.

Table 3. Plan for program adoption and implementation.

Program use outcomes, performance objectives, and determinants (NPT[a])	Change objectives	BCTs[b]	Practical strategies
Adoption outcome			
Shelter staff adopt the app			
Coherence: Staff understand the benefits of the app	Staff will develop an understanding of how the app supports mental health management	4.1 Instruction on how to perform a behavior	Provide informational sessions and materials on the app's benefits for mental health.
Staff encourage youth to use the app regularly			
Cognitive participation: Staff engage in the program	Staff will actively promote app use to youth in their daily routines	1.8 Behavioral contract	Staff will commit to a plan to introduce the app to youth and track usage
Collective action: Staff train youth on app use	Staff will train youth to use the app and integrate it into daily life	4.1 Instruction on how to perform a behavior	Staff will provide step-by-step training on app functionality and its benefits.
Implementation outcome			
Youth use the app daily			
Coherence: Youth understand how the app helps manage emotions	Youth will understand the purpose of the app and its features	4.1 Instruction on how to perform a behavior	Educational sessions and motivational posters will explain the app's purpose.
Cognitive participation: Youth engage in regular use	Youth will commit to using the app daily to track emotions and routines	8.1 Behavioral practice or rehearsal	Youth will be encouraged to practice app use through daily emotional logging and tracking routines.
Reflexive monitoring: Youth track progress and give feedback	Youth will evaluate how the app helps them manage their daily routines	2.3 Self-monitoring of behavior	Youth will log daily routines and emotions; feedback will be collected for improvement.
Sustainability			
Maintain app usage long-term			
Collective action: Peer ambassadors promote the app	Ambassadors will share success stories to encourage long-term app use	6.3 Information about others' approval	Peer ambassadors will share their experiences and offer support through group sessions.
Reflexive monitoring: Continuous improvement	Youth will provide feedback for app improvements	1.5 Review of behavior goals	Regular feedback sessions will be held to improve app functionality and user experience.

[a]NPT: normalization process theory.

[b]BCT: behavior change technique.

Step 6: Evaluation Plan

The sixth step involved creating a detailed and comprehensive plan to evaluate the designed mobile app. This evaluation plan aims to systematically analyze the implementation process and effectiveness and identify areas for improvement, if necessary.

Process Evaluation

Process evaluation confirms whether the app is being implemented as planned and assesses the fidelity of its implementation. Key indicators include the number of youths using the app, frequency of app usage, levels of user engagement, and the frequency with which users use the key features of the app. These data will be collected through various methods, including app usage analytics software, user logs, and feedback surveys from youth residing in shelters. Real-time data collection will be implemented to analyze usage patterns and engagement levels. User logs will track individual user activities to help identify the most and least used features. Regular feedback surveys will assess subjective user experiences and satisfaction, covering aspects such as ease of use, usefulness of features, and overall satisfaction. In addition, in-depth interviews will be conducted to gain a more detailed understanding of user experiences, focusing on the strengths, weaknesses, and areas for improvement of the app. This comprehensive data collection will help determine if the app is being implemented as planned, identify any unforeseen issues, and understand the most beneficial features to users.

Effectiveness Evaluation

The program's effectiveness will be assessed through a randomized controlled trial. Participants will be assigned randomly to either the experimental or control group using computer-generated random allocation. The experimental group will receive the app intervention, while the control group will receive standard care provided by the shelters during the trial. In consideration of ethical concerns, the control group will be given access to the app after the intervention period has ended, ensuring that all participants can benefit from the app. The evaluation will focus on the app's effects on youth's mental health, daily living skills, and emotion regulation. Key measures will include decreases in anxiety and depressive symptoms, better emotion regulation, and improved daily living skills. Data collection methods will include pre- and post-surveys, standardized mental health assessment tools (such as the Beck Depression Inventory and Generalized Anxiety Disorder Scale), and user interviews.

Quantitative data analysis will be conducted using IBM SPSS (version 26). The 1-way, repeated-measures, multivariate ANOVA will be used to assess the intervention effects. Qualitative data from interviews and open-ended survey responses will undergo thematic analysis to gain deep insights into user experience and app effectiveness. This thorough evaluation plan aims to systematically evaluate the app's implementation process and outcomes, aiming to enhance youth's mental health. Our goal is to optimize the app's effectiveness and promptly make any necessary improvements.

Discussion

Principal Results and Comparisons With Previous Work

A mobile app was developed in this study to enhance the mental health of youth in out-of-home care using the IM approach. The app was created through a systematic and step-by-step development process.

Earlier studies have used the IM approach to create successful health promotion programs for different health concerns [19-21]. Nonetheless, IM-based mental health promotion initiatives aimed at youth residing in shelters are scarce. Furthermore, programs developed in previous studies frequently depend on brief interventions centered on in-person counseling, posing challenges in maintaining lasting impacts [12,17]. In South Korea, mental health apps such as MindLink [34] and Mind Café provide early intervention and emotional support to youth, but they are not specifically designed for youth residing in shelters. MindLink, for example, focuses on psychosis intervention, while Mind Café provides short-term emotional support through counseling, but it may be limited due to cost-related barriers to sustained access. These apps are valuable for their specific purposes but lack tailored features to address the unique needs of youth in out-of-home care. The mobile app developed in this study was created to overcome these limitations by providing accessible emotional support and daily management functions to youth. Emotion flashcards and self-monitoring tools assist users in recognizing and regulating their emotions independently, helping them cultivate their ability to self-manage even after leaving the shelter.

Our program was designed after conducting a comprehensive needs assessment of youth living in shelters. A previous study has emphasized the importance of a thorough needs assessment to improve program effectiveness [35]. This study investigated the behavioral and environmental factors influencing mental health issues by reviewing literature and interviewing youth in out-of-home care and shelter staff with the aim of providing tailored interventions.

We designed a program using theories like the TTM, SCT, and ELM to improve the scientific validity of the program's mental health promotion program. By integrating these theories, our program identified knowledge, attitudes, and habit formation as key determinants, applying each theory's methods to enhance the app's effectiveness. Previous studies that incorporated these theories have shown positive outcomes.

We prioritized a user-centered design approach to enhance the app's usage and effectiveness. To increase youth's engagement and persistence, we implemented various strategies. For instance, users have the option to personalize their experience by selecting both their own character and their counselor's character. The app's user-friendly design, combined with a visual reward system for completing daily tasks, sparks interest and boosts engagement, encouraging continuous use. Users can establish their daily routines and receive visual cues indicating progress, such as a cake image that fills up as they complete tasks, providing a sense of achievement. When users choose an emotion from a list, the counselor responds with appropriate messages to foster emotional support. For example, if a user selects "anxious," the counselor's character might suggest, "Identify what's causing your anxiety and focus on tackling simple tasks first. Acknowledging and addressing your problems can help alleviate negative feelings." This feature enhances emotional support and equips users with tools to manage their emotions effectively. The Emotion Regulation section displays facial expressions accumulated over a week, allowing users to monitor changes in their emotions. By keeping emotion diaries, users can record and reflect on their feelings, which enhances their emotional awareness and regulation skills.

Limitations

The interviews conducted during the needs assessment of this study's development process focused on gathering real-world needs and collaborating with stakeholders. Therefore, insights obtained from these interviews were integrated iteratively, and any emergent trends were reviewed with the planning group. However, using formal analysis frameworks in future projects could enhance rigor and reliability. This study focused on developing a mobile app but did not confirm its effectiveness through application and evaluation. This limitation hinders the assessment of the app's effectiveness and user satisfaction. Future research should validate the app's long-term effectiveness in real-world scenarios. The study specifically targeted youth living in shelters in a specific region, limiting the generalizability of the findings to youth in other regions and cultural contexts. Further research is necessary to evaluate the app's use among youth from various regions and cultural backgrounds.

Conclusions

This study outlines the development process of a mobile app aimed at enhancing the mental health of youth in out-of-home care using the IM approach. The app was created to assist youth in managing their daily lives and regulating their emotions independently, offering high accessibility and continuous support. The systematic development process and theoretical foundations for designing the app's structure and features are described, along with plans for evaluating its process and effectiveness.

Although real-world use of the app was not evaluated in this study, the systematic approach and substantial stakeholder involvement in the development phase indicate its potential effectiveness. Future research should focus on assessing the app's real-world usage and incorporating user feedback for continuous improvement. Additional studies are anticipated to establish a strong foundation for developing and implementing digital health interventions that can significantly improve the mental health and well-being of youth in out-of-home care.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT; NRF-2022R1F1A1063208).

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Authors' Contributions

DN, JL, and JP contributed to conceptualization. DN and JL contributed to the methodology. DN and JP contributed to writing—original draft preparation. DN and JP contributed to writing—review and editing. DN performed supervision. DN handled project administration and funding acquisition. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

References

1. Youth Welfare Support Act, Article 2 Clause 5. Korea Legislation Research Institute. 2020. URL: https://elaw.klri.re.kr/kor_mobile/viewer.do?hseq=62168&type=sogan&key=11 [accessed 2024-10-20]
2. Gewirtz O'Brien JR, Edinburgh LD, Barnes AJ, McRee A. Mental health outcomes among homeless, runaway, and stably housed youth. Pediatrics 2020;145(4):e20192674. [doi: 10.1542/peds.2019-2674] [Medline: 32152134]
3. Kim S, Chun J. Aggressive behaviors among Korean children in out-of-home care: the role of placement characteristics. Children and Youth Services Review 2016 Jul;66:56-61. [doi: 10.1016/j.childyouth.2016.04.023]
4. Kim E. Independent living experiences of run away female-youth discharged from shelter. Korean Journal of Youth Studies 2019;26(1):55-77. [doi: 10.21509/KJYS.2019.01.26.1.55]
5. Rahamim A, Mendes P. Mental health supports and young people transitioning from out-of-home care in Victoria. Children Australia 2015;41(1):59-68. [doi: 10.1017/cha.2015.35]
6. Kolko DJ, Hurlburt MS, Zhang J, Barth RP, Leslie LK, Burns BJ. Posttraumatic stress symptoms in children and adolescents referred for child welfare investigation. a national sample of in-home and out-of-home care. Child Maltreat 2010 Feb;15(1):48-63. [doi: 10.1177/1077559509337892] [Medline: 19564628]
7. Murray ET, Lacey R, Maughan B, Sacker A. Association of childhood out-of-home care status with all-cause mortality up to 42-years later: Office of National Statistics Longitudinal Study. BMC Public Health 2020 May 20;20(1):735 [FREE Full text] [doi: 10.1186/s12889-020-08867-3] [Medline: 32434479]

8. Jeong MJ, Kim SJ. A study on readiness for independent living of youth in shelter. J Soc Sci Res 2016 Apr;42(1):177-207. [doi: 10.15820/khjss.2016.42.1.008]
9. Stubbs A, Baidawi S, Mendes P. Young people transitioning from out-of-home care: their experience of informal support. A scoping review. Child Youth Serv Rev 2023 Jan;144:106735. [doi: 10.1016/j.childyouth.2022.106735]
10. Nho CR, Park EH, McCarthy ML. Case studies of successful transition from out-of-home placement to young adulthood in Korea. Child Youth Serv Rev 2017 Aug;79:315-324. [doi: 10.1016/j.childyouth.2017.06.035]
11. Hatsu I, Gunther C, Hade E, Vandergriff S, Slesnick N, Williams R, et al. Unaccompanied homeless youth have extremely poor diet quality and nutritional status. Int J Adolesc Youth 2019;24(3):319-332 [FREE Full text] [doi: 10.1080/02673843.2018.1538885] [Medline: 31485095]
12. Trubey R, Evans R, McDonald S, Noyes J, Robling M, Willis S, et al. Effectiveness of mental health and wellbeing interventions for children and young people in foster, kinship, and residential care: systematic review and meta-analysis. Trauma Violence Abuse 2024 Oct;25(4):2829-2844 [FREE Full text] [doi: 10.1177/15248380241227987] [Medline: 38362816]
13. Brillantes-Evangelista G. An evaluation of visual arts and poetry as therapeutic interventions with abused adolescents. Arts Psychother 2013 Feb;40(1):71-84. [doi: 10.1016/j.aip.2012.11.005]
14. Hyun MS, Chung HI, Lee YJ. The effect of cognitive-behavioral group therapy on the self-esteem, depression, and self-efficacy of runaway adolescents in a shelter in South Korea. Appl Nurs Res 2005 Aug;18(3):160-166. [doi: 10.1016/j.apnr.2004.07.006] [Medline: 16106334]
15. Zhang J, Wu Q, Slesnick N. Social problem-solving and suicidal ideation among homeless youth receiving a cognitive therapy intervention: a moderated mediation analysis. Behav Ther 2021 May;52(3):552-563 [FREE Full text] [doi: 10.1016/j.beth.2020.07.005] [Medline: 33990233]
16. Slesnick N, Erdem G, Bartle-Haring S, Brigham GS. Intervention with substance-abusing runaway adolescents and their families: results of a randomized clinical trial. J Consult Clin Psychol 2013 Aug;81(4):600-614 [FREE Full text] [doi: 10.1037/a0033463] [Medline: 23895088]
17. Evans R, MacDonald S, Trubey R, Noyes J, Robling M, Willis S, et al. Interventions targeting the mental health and wellbeing of care-experienced children and young people in higher-income countries: Evidence map and systematic review. Syst Rev 2023 Jul 01;12(1):111 [FREE Full text] [doi: 10.1186/s13643-023-02260-y] [Medline: 37393358]
18. Bartholomew Eldredge LK. Planning Health Promotion Programs: An Intervention Mapping Approach. 4th edition. Hoboken, NJ: Wiley; 2016.
19. Svendsen MJ, Sandal LF, Kjær P, Nicholl BI, Cooper K, Mair F, et al. Using intervention mapping to develop a decision support system-based smartphone app (selfBACK) to support self-management of nonspecific low back pain: development and usability study. J Med Internet Res 2022 Jan 24;24(1):e26555 [FREE Full text] [doi: 10.2196/26555] [Medline: 35072645]
20. Hadjiconstantinou M, Schreder S, Brough C, Northern A, Stribling B, Khunti K, et al. Using intervention mapping to develop a digital self-management program for people with type 2 diabetes: tutorial on MyDESMOND. J Med Internet Res 2020 May 11;22(5):e17316 [FREE Full text] [doi: 10.2196/17316] [Medline: 32391797]
21. Wong AWK, Fong MWM, Munsell EGS, Metts CL, Lee SI, Nicol GE, et al. Using intervention mapping and behavior change techniques to develop a digital intervention for self-management in stroke: development study. JMIR Hum Factors 2023 Jul 24;10:e45099 [FREE Full text] [doi: 10.2196/45099] [Medline: 37486748]
22. Fernandez ME, Ruiter RAC, Markham CM, Kok G. Intervention mapping: theory- and evidence-based health promotion program planning: perspective and examples. Front Public Health 2019;7:209 [FREE Full text] [doi: 10.3389/fpubh.2019.00209] [Medline: 31475126]
23. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
24. Li X, Yang S, Wang Y, Yang B, Zhang J. Effects of a transtheoretical model - based intervention and motivational interviewing on the management of depression in hospitalized patients with coronary heart disease: a randomized controlled trial. BMC Public Health 2020 Mar 30;20(1):420 [FREE Full text] [doi: 10.1186/s12889-020-08568-x] [Medline: 32228532]
25. Bandura A. Social cognitive theory of self-regulation. Organ Behav Hum Decis Process 1991;50(2):248-287. [doi: 10.1016/0749-5978(91)90022-L]
26. Petty RE, Cacioppo JT. The elaboration likelihood model of persuasion. Adv Exp Soc Psychol 1986;19:123-205. [doi: 10.1016/S0065-2601(08)60214-2]
27. Johnson RW, White BK, Gucciardi DF, Gibson N, Williams SA. Intervention mapping of a gamified therapy prescription app for children with disabilities: user-centered design approach. JMIR Pediatr Parent 2022 Aug 09;5(3):e34588 [FREE Full text] [doi: 10.2196/34588] [Medline: 35943782]
28. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 01;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]

29. Verplanken B, Orbell S. Attitudes, habits, and behavior change. Annu Rev Psychol 2022 Jan 04;73:327-352. [doi: 10.1146/annurev-psych-020821-011744] [Medline: 34587780]
30. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. J Consult Clin Psychol 1983 Jun;51(3):390-395. [doi: 10.1037//0022-006x.51.3.390] [Medline: 6863699]
31. Schunk DH, DiBenedetto MK. Motivation and social cognitive theory. Contemp Educs Pchol 2020 Jan;60:101832. [doi: 10.1016/j.cedpsych.2019.101832]
32. Scott L, Andrewes T. Using the transtheoretical model of behaviour change to analyse the impact of stopping exercise: a reflection. Br J Nurs 2021 Nov 11;30(20):1203-1205. [doi: 10.12968/bjon.2021.30.20.1203] [Medline: 34761974]
33. May CR, Mair F, Finch T, MacFarlane A, Dowrick C, Treweek S, et al. Development of a theory of implementation and integration: Normalization Process Theory. Implement Sci 2009 May 21;4:29 [FREE Full text] [doi: 10.1186/1748-5908-4-29] [Medline: 19460163]
34. Kim SW, Lee GY, Yu HY, Jung EI, Lee JY, Kim SY, et al. Development and feasibility of smartphone application for cognitive-behavioural case management of individuals with early psychosis. Early Interv Psychiatry 2018 Dec;12(6):1087-1093. [doi: 10.1111/eip.12418] [Medline: 28516480]
35. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]

Abbreviations

BCT: behavior change technique
ELM: elaboration likelihood model
IM: intervention mapping
SCT: social cognitive theory
TTM: transtheoretical model

Edited by C Jacob; submitted 30.07.24; peer-reviewed by A Frank; comments to author 29.08.24; revised version received 20.10.24; accepted 06.11.24; published 21.11.24.

Please cite as:
Park J, Lee J, Noh D
Mobile App for Improving the Mental Health of Youth in Out-of-Home Care: Development Study Using an Intervention Mapping Approach
JMIR Hum Factors 2024;11:e64681
URL: https://humanfactors.jmir.org/2024/1/e64681
doi:10.2196/64681
PMID:

Original Paper

Mobility-Based Smartphone Digital Phenotypes for Unobtrusively Capturing Everyday Cognition, Mood, and Community Life-Space in Older Adults: Feasibility, Acceptability, and Preliminary Validity Study

Katherine Hackett[1], PhD; Shiyun Xu[2], BS; Moira McKniff[1], BA; Lido Paglia[3], BS; Ian Barnett[2], PhD; Tania Giovannetti[1], PhD

[1]Department of Psychology and Neuroscience, Temple University, Philadelphia, PA, United States

[2]Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, PA, United States

[3]Information Technology, College of Science & Technology, Temple University, Philadelphia, PA, United States

Corresponding Author:
Tania Giovannetti, PhD
Department of Psychology and Neuroscience
Temple University
1701 N 13th St
Philadelphia, PA, 19122
United States
Phone: 1 484 843 1321
Email: tania.giovannetti@temple.edu

Abstract

Background: Current methods of monitoring cognition in older adults are insufficient to address the growing burden of Alzheimer disease and related dementias (AD/ADRD). New approaches that are sensitive, scalable, objective, and reflective of meaningful functional outcomes are direly needed. Mobility trajectories and geospatial life space patterns reflect many aspects of cognitive and functional integrity and may be useful proxies of age-related cognitive decline.

Objective: We investigated the feasibility, acceptability, and preliminary validity of a 1-month smartphone digital phenotyping protocol to infer everyday cognition, function, and mood in older adults from passively obtained GPS data. We also sought to clarify intrinsic and extrinsic factors associated with mobility phenotypes for consideration in future studies.

Methods: Overall, 37 adults aged between 63 and 85 years with healthy cognition (n=31, 84%), mild cognitive impairment (n=5, 13%), and mild dementia (n=1, 3%) used an open-source smartphone app (mindLAMP) to unobtrusively capture GPS trajectories for 4 weeks. GPS data were processed into interpretable features across categories of activity, inactivity, routine, and location diversity. Monthly average and day-to-day intraindividual variability (IIV) metrics were calculated for each feature to test a priori hypotheses from a neuropsychological framework. Validation measures collected at baseline were compared against monthly GPS features to examine construct validity. Feasibility and acceptability outcomes included retention, comprehension of study procedures, technical difficulties, and satisfaction ratings at debriefing.

Results: All (37/37, 100%) participants completed the 4-week monitoring period without major technical adverse events, 100% (37/37) reported satisfaction with the explanation of study procedures, and 97% (36/37) reported no feelings of discomfort. Participants' scores on the comprehension of consent quiz were 97% on average and associated with education and race. Technical issues requiring troubleshooting were infrequent, though 41% (15/37) reported battery drain. Moderate to strong correlations (r≥0.3) were identified between GPS features and validators. Specifically, individuals with greater activity and more location diversity demonstrated better cognition, less functional impairment, less depression, more community participation, and more geospatial life space on objective and subjective validation measures. Contrary to predictions, greater IIV and less routine in mobility habits were also associated with positive outcomes. Many demographic and technology-related factors were not associated with GPS features; however, income, being a native English speaker, season of study participation, and occupational status were related to GPS features.

Conclusions: Theoretically informed digital phenotypes of mobility are feasibly captured from older adults' personal smartphones and relate to clinically meaningful measures including cognitive test performance, reported functional decline, mood, and community activity. Future studies should consider the impact of intrinsic and extrinsic factors when interpreting mobility phenotypes. Overall, smartphone digital phenotyping is a promising method to unobtrusively capture relevant risk and resilience factors in the context of aging and AD/ADRD and should continue to be investigated in large, diverse samples.

(JMIR Hum Factors 2024;11:e59974) doi:10.2196/59974

KEYWORDS

digital phenotyping; digital biomarkers; monitoring; mHealth; cognition; mobility; life space; depression; location data; Alzheimer disease; aging; mobile phone

Introduction

Background

Alzheimer disease and related dementias (AD/ADRD) place immense pressure on our health care system. They represent a global issue that is worsening and will be exacerbated by insufficient disease screening methods and a lack of ecologically valid outcome measures for clinical trials [1]. New and innovative approaches for early detection and monitoring are direly needed to address this global crisis. In this paper, we present results from a proof-of-concept study demonstrating the promise of smartphone digital phenotyping to capture clinically relevant risk factors and outcomes in the context of aging and AD/ADRD.

Decades of clinical trial research indicate that early intervention will be critical for effective AD/ADRD treatment [2-5]. Biomarker testing (eg, positron emission tomography neuroimaging and cerebrospinal fluid), traditional neuropsychological evaluation, and clinician- or informant-rated assessments have historically been the gold-standard methods for detection and diagnosis. However, these methods present drawbacks, including limited accessibility (eg, proximity and cost), scalability (ie, logistical constraints for wide-scale clinical implementation), prognostic value (ie, inconsistent correspondence with clinical progression), and ecological validity (ie, poor representation of real-world functioning) [6-10]. Traditional neuropsychological measures are also affected by sociocultural factors, such as educational quality, socioeconomic status, native language status, and acculturation, rendering them inappropriate for the increasingly diverse global population [11-20]. Subtle difficulties in complex everyday tasks signal early decline and are critical to assess; however, standard functional assessments are limited by recall bias, availability of an informant reporter, bias due to attributes of the informant, outdated items, and poor sensitivity [21-27]. Patterns and trajectories of cognitive and functional decline are also extremely heterogeneous and person-specific [28-30] and are thus difficult to assess using a one-time standardized assessment measure. Taken together, health care systems are ill-equipped to screen for early signs of cognitive impairment at scale, as evidenced by recent estimates that 92% of cases of mild cognitive impairment (MCI) remain undiagnosed [31-34]. In addition to poor screening methods, traditional clinical trial outcome measures are not sufficiently meaningful and precise to demonstrate therapeutic benefit at early stages [4,25,35-40]. Thus, a growing priority in the field of AD/ADRD is to develop and implement sensitive and functionally meaningful screening and outcome measures as new therapies are evaluated earlier in the disease course [41].

New digital assessment methods have great potential for efficient, accessible, sensitive, and objective assessment of early cognitive and functional changes reflecting risk for AD/ADRD [6,7,42-46]. Digital tools can capture microlevel behavioral data with increased sensitivity and reduced sample size requirements compared with traditional paper-and-pencil neuropsychological measures and functional scales [47]. Gathering this information at home can address accessibility limitations for those in rural environments or who face hardships traveling to and from a clinical site, and may provide a more reliable and ecologically valid representation of real-world functioning compared to traditional evaluation at a single time point in a highly controlled setting [48-50]. As global technology use and affordability of personal devices continue to rise [51,52], new technologies can potentially address crucial gaps in the scalability and accessibility of current methods and counteract higher rates of missed diagnosis among populations experiencing low socioeconomic status [34,53-57].

Digital phenotyping is one innovative approach that uses the “moment-by-moment quantification of the individual-level human phenotype in situ” based on interactions with technology, including smartphones and smart home devices, to capture social and behavioral data passively, continuously, and with minimal interference [58-60]. It collects high-frequency, fine-grained data reflecting everyday behaviors “in the wild” without relying on user engagement, subjective reports, or burdensome procedures. Preliminary support has been demonstrated in psychiatry studies leveraging a host of sensors (eg, GPS, accelerometer, Wi-Fi or Bluetooth signals, ambient sound and light, app use, call and text message metadata, and keystroke dynamics) and imputed behavioral features (eg, time spent at home, sleep cycles, level of socialization, and routine or anomalies) to predict clinical outcomes, including depression and bipolar disorder symptoms, suicide risk, psychosis relapse, and depression treatment response [61-70]. In the context of neurodegenerative disorders, several sensors—particularly keystroke dynamics and phone or battery use metrics—have shown associations with neuropsychological test performance, diagnostic severity, and even gray matter volume in clinical cohorts, including those with MCI, Alzheimer disease (AD), frontotemporal dementia, and multiple sclerosis [71-77].

Basic questions of feasibility, acceptability, and ethical considerations related to data privacy are important to weigh when considering the highly sensitive nature of digital phenotyping data in vulnerable populations [78-80]. Few studies have proactively addressed these questions in the context of older adults with cognitive decline [65]. Furthermore, many existing studies have used exploratory approaches without a priori hypotheses [81]. As described by Hackett and Giovannetti [7] and Leaning et al [66], some of the many interpretive and logistical challenges of digital phenotyping can be mitigated with conceptual models and clinically informed features to provide context to results and improve reproducibility.

In this paper, we present findings from a proof-of-concept study evaluating a smartphone digital phenotyping protocol to assess cognition, everyday function, and mood in a cohort of older adults with and without cognitive decline. Here, we focus on smartphone-derived GPS data as the digital phenotyping sensor of interest. Our study design and analytic approach were informed by a conceptual framework proposed by Hackett and Giovannetti [7] based on established trends in cognitive neuroscience, neuropsychology, neurology, and computer science literature [7]. The conceptual framework (ie, the Variability in Everyday Behavior [VIBE] model) posits that pathological cognitive decline is accompanied by a reduction in everyday activities, worsening mood, and lower scores on standardized neuropsychological measures. These declining mean-level trends occur alongside increases in intraindividual variability (IIV) on measures of cognition and everyday function as individuals become more inefficient and work to compensate for underlying disease progression [7]. Trends of decreasing *levels* of everyday activity and increasing *variability* in everyday activities may be indexed by passively obtained smartphone data such as GPS trajectories, hence the focus of this study.

Objectives

The primary aims of this study were twofold: (1) to examine the feasibility and acceptability of a digital phenotyping protocol among older adult smartphone users, and (2) to examine associations between passively obtained GPS movement trajectories collected over a 1-month study period and traditional validated measures of cognition, everyday functioning, mood, and mobility habits collected at baseline. Data were collected using the Learn, Assess, Manage, and Prevent (LAMP) platform, an open-source platform for research and clinical use, via the mindLAMP app [82-84]. Data for the second aim were analyzed and interpreted according to the following a priori hypotheses based on our conceptual framework (ie, the VIBE model), shown in Textbox 1.

Textbox 1. Hypotheses informed by the Variability in Everyday Behavior model for individuals along the continuum from healthy cognition to MCI.

Cognition and function

Average GPS metrics of activity will show a positive linear relation with measures of cognition and function, whereas GPS intraindividual variability (IIV) for all categories will show a negative relation with cognition and function.

Mood

Average GPS metrics of activity will show a negative linear relation with depression (ie, greater overall mobility will be associated with less depression), whereas GPS IIV for all categories will show a positive relation with depression (ie, greater mobility variability will be associated with more depression).

Mobility

Average GPS metrics of activity will be positively associated with objective and subjective measures of gait speed, life space, and community participation.

The study also included two exploratory aims: (1) to examine relations between GPS features and participant intrinsic and extrinsic factors (eg, sociodemographic and contextual) to inform the selection of covariates or moderating variables in future studies and (2) to explore whether patterns of mobility—rather than absolute amounts of mobility—also relate to validators. Overall, our results provide preliminary support for the feasibility, acceptability, and validity of digital phenotyping in older adults, along with key insights that can be used to inform future studies.

Methods

Recruitment

Participants aged >60 years with healthy cognition, diagnoses of MCI, or mild AD were recruited from specialty dementia clinics and the community within the Philadelphia region, beginning in January 2022. Recruitment also involved contacting previous participants of other research studies within our laboratory, consistently attending community outreach events to establish trust and familiarity with our research team, and providing educational presentations on topics related to cognitive health and aging at local community centers. Individuals who expressed interest in participating in our study were contacted by a member of the study team to schedule a study appointment. During the scheduling call, the team member reviewed basic eligibility criteria (eg, age, use of a smartphone, and availability of a study partner), and participants who met criteria were scheduled for an initial session. Inclusion and exclusion criteria were reviewed again in detail at the start of the baseline session to ensure eligibility. General inclusion criteria for all participants were (1) aged ≥60 years, (2) fluent in English, (3) existing smartphone user (iOS or Android; meeting minimum software version compatibility) for at least 1 year before joining the study, (4) Wi-Fi connectivity at home, (5) phone use on a daily basis; (6) no plans to purchase or switch to a new smartphone over the next 4 weeks, and (7) availability of an informant reporter who has knowledge of the participant's daily functioning. Exclusion criteria were (1) a history of severe psychiatric or nervous system disorders (other than dementia), (2) current metabolic or systemic disorders, (3) severe sensory or motor deficits precluding smartphone use, (4) intellectual disability, and (5) scheduled surgery or major travel over the 4-week study period. Participants with self- or clinician-reported diagnoses of healthy cognition, MCI, or mild AD completed comprehensive neuropsychological testing during a baseline visit (as described in the subsequent sections), and Jak/Bondi neuropsychological actuarial criteria were used to confirm diagnostic group membership [85]. Follow-up consensus diagnosis was used to account for atypical clinical factors that may impact the accuracy of actuarial diagnosis (eg, English as a second language and co-occurring mood or psychiatric concerns).

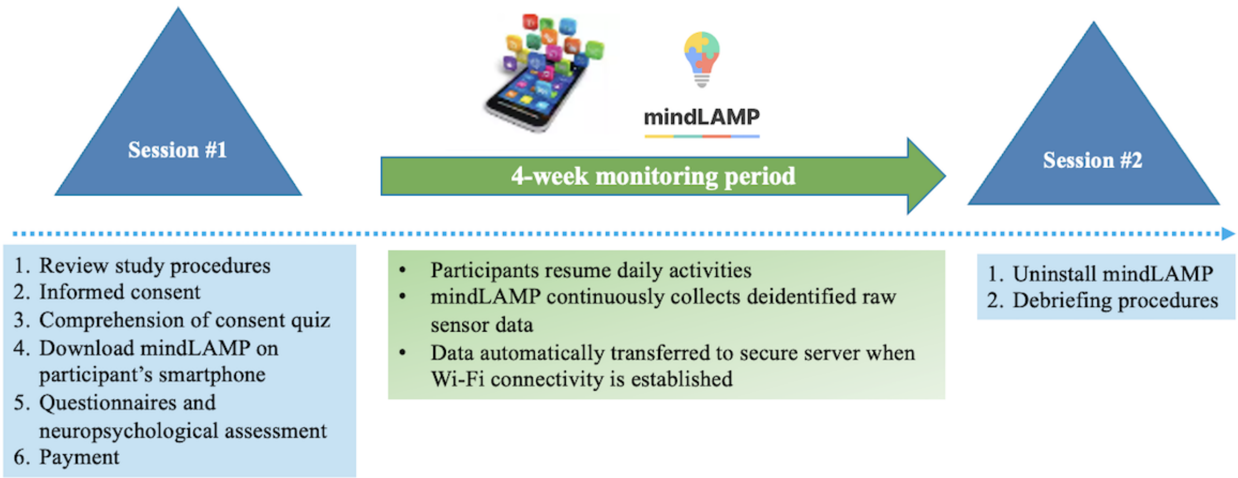
Study informants were also recruited for each participant to corroborate responses on the self-reported functional decline validation measure, to confirm other data pertinent to clinical history, and to provide assistance with potential technical difficulties—particularly for participants with MCI or dementia. General inclusion and exclusion criteria for all informants included (1) aged ≥18 years; (2) fluent in English; (3) cognitively healthy with no diagnosis of dementia, MCI, or other neurological and psychiatric disorder; (4) available and willing to complete study questionnaires in person, by phone, or by web; (5) having at least weekly contact with the participant; and (6) reports that they are knowledgeable of the participant's daily functioning and smartphone use.

Study Procedures

Study Timeline

Participants meeting eligibility criteria were scheduled for an in-person visit (session 1) and were enrolled in the study for approximately 4 weeks. As outlined in Figure 1, participants and informants completed 2 study visits, separated by 4 weeks. Session 1 lasted 2 to 4 hours and included a detailed review of study procedures, informed consent, comprehension of consent quiz, cognitive testing, questionnaires, and configuration of the study app (mindLAMP) on participants' personal smartphones (ie, downloading mindLAMP, logging in using participant-specific secure credentials, and enabling continuous location services). At the completion of session 1, participants were instructed to resume their daily routines for 4 weeks and were compensated US $50 upfront for their participation. They were asked not to enter low battery or airplane mode and not to log out of the mindLAMP app, which would impact GPS data quality. During the 4-week study period, mindLAMP passively and securely collected GPS data without user engagement. At the end of the study period, participants completed session 2 (in-person or remotely), which consisted of debriefing questionnaires. The mindLAMP app was deleted from participants' smartphones at session 2, halting data collection.

Figure 1. Study timeline.



Privacy and Security Safeguards

The LAMP platform was used for the collection of digital phenotyping (GPS) data via the mindLAMP smartphone app, which is available for free on the Apple App Store and Google Play Store. LAMP is an open-source platform for research and clinical use developed by the Division of Digital Psychiatry at Beth Israel Deaconess Medical Center [83]. It uses industry-standard encryption protocols to render information collected from smartphones unidentifiable and transmits data to a secure cloud database via Health Insurance Portability and Accountability Act (HIPAA)–compliant methods whenever Wi-Fi connectivity is established. Participants logged in to mindLAMP via a randomly generated 10-digit user ID (UID) generated within the LAMP platform by the study coordinator, which was only used for the collection of digital phenotyping data. Therefore, no personally identifiable information is associated with the data collected by mindLAMP. Participants' 6-digit study ID, used for all other clinical data collected within the laboratory at Temple University, is linked to their mindLAMP UID on a university-approved secure research database only (REDCap [Research Electronic Data Capture; Vanderbilt University]).

The aforementioned security and privacy information was thoroughly reviewed with participants during informed consent at session 1. This process included a review of written and visual handouts depicting privacy safeguards, examples of the scope of data collected, and results of other published studies that used mindLAMP [86]. Participants could choose to end study participation and have their study data deleted at any time. After reviewing the consent form, a 10-item comprehension of consent quiz with yes or no response options was administered to ensure participants completely understood the information outlined in the consent form. This quiz covered details including the purpose of mindLAMP (eg, "This study requires downloading the mindLAMP app which collects information from my smartphone sensors"), possible risks or benefits to study participation, such as potential battery drain (eg, "It is possible that I will notice a reduction in my phone's battery life while mindLAMP is running on my phone"), and data security and encryption methods (eg, "The mindLAMP app uses a secure encryption system called 'hashing' to make all information that it collects unidentifiable and untraceable"). Multimedia Appendix 1 presents the complete comprehension of consent quiz. Incorrect items were further reviewed until comprehension was established. These procedures were informed by the Digital Health Checklist and other materials from the Research Center for Optimal Digital Ethics health team, which encourages digital health researchers to proactively identify gaps in the communication of study risks, benefits, and privacy and security details [78,79].

Backend Technical Implementation

Data collection and storage were supported by a self-deployed version of the open-source LAMP Platform. A Temple University-approved secure cloud server (1-TB capacity) was purchased and configured with the LAMP application programming interface to enable research participants to connect to the LAMP platform and access the mindLAMP app. Study data were stored in our instance of the mindLAMP database (ie, our copy of the LAMP platform located on a study-specific cloud server) via CouchDB, ensuring a standardized data format consistent with other studies using the LAMP platform [84]. The functionality of the mindLAMP app itself is continually maintained by the team at mindLAMP at the Beth Israel Deaconess Division of Digital Psychiatry. Ongoing security monitoring and backup of study data were maintained by Temple University IT. To monitor unexpected periods of missing data due to potential technical issues (ie, phone powered off, mindLAMP logged out, or permissions reset), we created a code to automate an email alert to the study team (Cronjob) when there were >3 days of missing sensor data. In these instances, the email message included the UID and the corresponding missing sensor, and a member of the study team promptly reached out to the participant to troubleshoot.

Measures

Feasibility and Acceptability

To assess feasibility, we tracked the number of participants who completed the 4-week study period after providing consent and completing session 1 and those who requested to withdraw for any reason. Feasibility was also operationalized by performance on the comprehension of consent quiz, which demonstrates participants' ability to comprehend complex technical information specific to digital phenotyping studies. To gauge acceptability of the informed consent and overall study procedures, we administered a debriefing survey at session 2 after participants had completed the full study period. Participants were asked to rate their level of satisfaction with the explanation of the study procedures at session 1 on a scale of 4 (very satisfied—all components of the study were clearly explained) to 1 (very unsatisfied—all components of the study were poorly explained).

Participants were also asked if they experienced any major difficulties with their phones, if they experienced any feelings of discomfort or paranoia due to the study app running on their smartphones, or if there were any major changes in how they used their smartphones during the study period (yes or no). If they answered "yes" to changes in smartphone use, participants selected all applicable options, including "I used my phone less/more overall," "I charged my phone less/more," "I carried my phone with me less/more," and "other." Finally, troubleshooting contacts between study staff and participants were tracked and reported as part of feasibility findings.

Validators

Overview

Validation measures administered at session 1 included neuropsychological tests used widely in the clinical diagnosis of MCI and dementia; self- and informant-report measures of cognitive decline and everyday functioning; questionnaires pertaining to mood; and measures of gait speed, geospatial life space, and community participation. These measures have demonstrated strong psychometric properties and were therefore used together as validation comparisons against digital phenotyping data. More details are provided in the subsequent sections.

Neuropsychological Tests

The neuropsychological test battery included the Hopkins Reading Test [87] as an estimate of premorbid intellectual ability (IQ), the Mini-Mental State Examination [88] as a global cognitive screener, and tests of attention (Trail Making Test-Part A [89] and Wechsler Memory Scale—Revised Digit Span Forward [90]), processing speed (Salthouse Letter and Pattern Comparison [91]), executive function (Trail Making Test-Part B [89] and Wechsler Memory Scale—Revised Digit Span Backward [90]), episodic memory (Hopkins Verbal Learning Test—Revised Delayed Recall [92] and Brief Visuospatial Memory Test-Revised Delayed Recall [93]), and language (Animal Fluency [94] and Boston Naming Test 30-item version [95]). Raw scores of each test were transformed into demographically corrected *t* scores using the Calibrated Neuropsychological Normative System [94], adjusting for age, sex, education, and estimated premorbid IQ (ie, Hopkins Reading Test score), which enabled more accurate estimation of cognitive ability within our diverse sample. An average *t* score was computed for the 2 tests within each cognitive domain to generate composite scores for attention, processing speed, executive function, delayed memory recall, and language abilities to streamline presentation of results.

Reported Cognitive Decline and Everyday Functioning

Self-reported cognitive decline was collected using the Everyday Cognition Scale-Short Form [96]. Self- or informant-reported everyday functioning was captured with the Functional Activities Questionnaire (FAQ [97]). Self-reported FAQ was used for participants with healthy cognition, and informant-reported FAQ was used for participants with MCI or dementia. Higher scores on these measures indicate more cognitive decline and more functional impairment, respectively.

Mood

Mood symptoms were indexed using the 15-item Geriatric Depression Scale (GDS [98]), a widely used self-report measure of depressive symptoms among older adults that requires participants to indicate whether they experience a list of common depression symptoms in a "yes/no" format. Raw scores were transformed into demographically corrected *t* scores using the Calibrated Neuropsychological Normative System as mentioned earlier. Higher *t* scores indicate higher levels of depression.

Gait Speed, Life Space, and Community Participation

The Timed Up and Go Test (TUG) was administered at session 1 as an objective measure of gait speed. This task measures the time it takes to rise from a chair, walk 10 feet, turn, walk back to the chair, and sit down. It is widely used to examine balance, functional mobility, gait speed, and fall risk in older adults [99,100]. Participants also completed the University of Alabama at Birmingham Life-Space Assessment (LSA), a self-report measure of mobility for community-dwelling older adults [101]. It captures the level of independence and spatial extent of a person's life over the preceding month and has shown strong associations with mobility within the home and community and with performance of activities of daily living [102]. Constricted life space has also been associated with risk for MCI and dementia [103,104]. The Australian Community Participation Questionnaire (ACPQ) 15-item version was administered as an additional measure of concurrent validity and assesses the extent to which someone engages in a range of community activities. Subscales include contact with immediate household, extended family, friends, and neighbors; participating in organized community activities; taking an active interest in current affairs; and religious observance [105]. An index of breadth of participation across the 7 domains was derived using a mean-split procedure for each domain, followed by summing these scores to generate an overall index ranging from 1 to 7 (as described by Brett et al [106]).

Other Participant Features

Demographic

Demographic data included participants' self-reported biological sex assigned at birth, age, race, ethnicity, current living status (alone or with others), current occupational status, educational attainment, and other information related to socioeconomic status (eg, highest household annual income).

Technology Use

Participants completed a 6-item Habitual Smartphone Behavior subscale [107] to assess smartphone use patterns, providing responses ranging from "strongly agree" to "strongly disagree" to questions such as "Smartphone usage is part of my daily routines." We also asked participants, "Do you usually have your phone with you when you leave home?" (smartphone portability), to which they could reply, "Yes- I almost never leave my house without my phone," "In between—I leave my house without my phone about half the time," or "No, I often leave my house without my phone." The Mobile Device Proficiency Questionnaire was administered as a measure of digital literacy [108]. Participants also indicated their smartphone operating system (Android vs iOS).

Seasonal and COVID-19 Factors

Dates of study participation were collected and coded as winter, fall, spring, and summer to explore potential seasonal effects on mobility habits. Because study participation took place during the COVID-19 pandemic for some participants, we asked about the impact of the COVID-19 pandemic on social participation, routines, and mobility behaviors at the time of study participation. Participants were asked, "On a scale of 1-5, how isolated or cut off from family and friends are you feeling due to limited/canceled social gatherings resulting from COVID-19?" "On a scale of 1-5, how disruptive has the COVID-19 pandemic been to your daily routines and activities?" and "On a scale of very much limited to very much expanded, how much has the COVID-19 pandemic changed your mobility/your movements outside of the home?"

Self-Reported Health Changes During the Study Period

At the end of the study period during session 2, participants reported whether there were any changes in their overall health during the study period by responding to a single question on the debriefing survey. Response options included (1) yes, significant change; (2) yes, a little change; or (3) no change. If they endorsed any change in health, they were given the option to elaborate.

Digital Phenotyping (mindLAMP)

Though the mindLAMP app enables the collection of a wide array of deidentified passive and active data, this study focused on passively obtained GPS data. mindLAMP was configured to continuously record the device's GPS coordinates at a maximum frequency of 1 Hz. Raw data outputs include latitude, longitude, altitude, and the coordinates' estimated accuracy. At study completion, these raw data were extracted from the study server and processed into daily summary features (Table 1) using a publicly available R script developed by Barnett and Onnela [109,110]. GPS data from smartphone devices are prone to large amounts of missing data [111]; therefore, advanced multiple imputation methods based on weighted resampling of the observed data were used to account for missingness before feature calculation. These imputation methods are automatically incorporated within the aforementioned processing script and are described in detail in the study by Barnett and Onnela [109].

Table 1. Daily GPS features generated from mindLAMP raw data.

Category and feature	Feature description
Activity	
Distance traveled	The sum of all flight lengths that day (m)
Radius of gyration	Average distance (m) a person is from their center of mass (average position) on a given day
Maximum diameter	Maximum diameter (m); longest pairwise distance between any 2 pause locations that occur that day
Maximum distance home	Maximum distance from home (m); distance between home and farthest pause location from home that day
Average flight length	The average length of all flights that day (m)
Average flight duration	The average duration of all flights that day (s)
Inactivity	
Home time	Time spent at home (min); amount of time spent that day within 200 m of home (the significant location with the largest total amount of time between 9 PM and 6 AM throughout the study period).
Probability paused	Fraction of time a person is stationary (paused) during a day, relative to time spent mobile or in flight.
Routine	
Circadian routine	Physical circadian routine; the fraction of time a person is in the same place (within a 200 m radius) at the same time of each day throughout the study period. Ranges from 0=completely different routine to 1=identical routine.
Weekend circadian routine	Physical circadian routine weekend or weekday stratified; similar to circadian routine except comparisons are stratified by grouping together weekends and grouping together weekdays. Higher scores reflect greater overall routine.
Location diversity	
Significant location entropy	Location entropy across a person's significant locations for the day; large values indicate spreading time out across many different locations fairly evenly for that day; small values indicate a concentration at a few significant locations.
Significant locations visited	The number of significant locations a person is within 200 m of that day. Determined using K-means on the set of all pause locations with a minimum duration of 10 minutes (longer pauses given additional weight and no 2 cluster centers within 400 m of one another).
Other	
Minutes missing	The number of minutes of missing data preimputation in a person's GPS mobility trace that day (min)

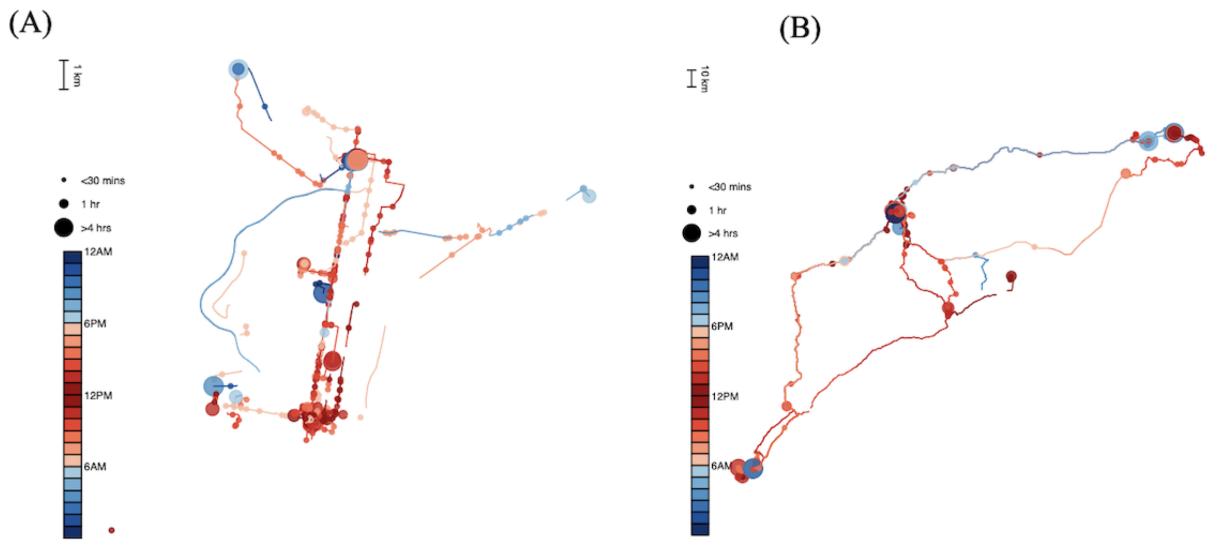
Supplementary materials from Barnett and Onnela [109] give a full description of GPS features and the methods used to calculate each. A flight is defined as a segment of linear movement; pauses are periods when a person does not move; and curved movement is approximated by multiple sequential flights.

Statistical Analyses

Preliminary Processing

The individual mobility traces derived from GPS data were visually inspected by examining each participant's deidentified mobility plots averaged across the study period (Figure 2). These plots are generated automatically through our processing script and enable visualization of overall mobility habits, amount of time spent at various locations, and time of day. Following visual inspection, daily mobility features (Table 1) were collapsed into monthly overall average and monthly (day-to-day) individual SD (iSD) for each participant to generate estimates of mean mobility and mobility IIV across the study period. For example, 30 days of daily distance traveled estimates for 1 participant were reduced to an average distance traveled per day, and an iSD of daily distance traveled across the 30-day study period. This approach enabled the examination of a priori hypotheses as an important first step in the validation of GPS data. All variables (GPS and validation measures) with highly skewed distributions were transformed using log(x+c) transformation to reduce the influence of outliers and make parametric analyses more robust. Pearson correlation analyses were used to explore relations among individual GPS features.

Figure 2. Example mobility plots depicting overall trajectories over 4 weeks for 2 participants. Solid lines depict flights from one location to another; circles indicate periods spent stationary, with larger sizes reflecting more time spent stationary at that location, color-coded by the time of day. (A) A 67-year-old woman living in northern Philadelphia. (B) A 70-year-old woman living in a suburb of Philadelphia.



Other preprocessing steps were completed in individual cases. Specifically, some participants had unanticipated travel during their scheduled study period (which was an exclusion criterion for validation purposes). In these cases, we extended their study duration and excluded days spent traveling from the raw data before feature extraction. In total, 3 (8%) of the 37 participants reported major unanticipated health changes during the study period (1 (3%) reported eye surgery, 1 (3%) reported a major fall with loss of consciousness, and 1 (3%) reported a fall with head trauma and COVID-19 infection) and were excluded from preliminary validity analyses in this study, because major health events during passive data collection would confound relations with baseline validators.

Primary Analyses

Descriptive statistics were first used to examine feasibility and acceptability outcomes across the entire sample (N=37; aim 1). Next, Pearson or Spearman correlation analyses were used to examine the relations between the mean (monthly average) and IIV (monthly iSD) of each mobility feature against validators collected at session 1 (ie, cognition, everyday function, mood, and mobility; aim 2). All validity analyses were conducted across the entire sample after excluding 3 participants with major health events (final n for aim 2 validation analyses=34). Given the small sample size, we interpreted correlation coefficients with r>0.3 as meaningful, regardless of statistical significance. Correlations surviving Bonferroni correction of P<.0038 (α corresponding to .05/13 GPS features) are noted to account for multiple comparisons. As mentioned earlier, we tested several a priori hypotheses based on our proposed conceptual framework (ie, the VIBE model, see Textbox 1). Specifically, we predicted that GPS mean-level activity metrics would show a positive linear relation with global cognition, function, and mood, whereas GPS IIV metrics would show a negative relation with cognition, function, and mood. Hypotheses from the original VIBE model were adjusted to consider trends along the continuum from healthy cognition to MCI only (excluding trends from MCI to dementia), given this study sample only included 1 participant with dementia. We also predicted a positive linear relation between GPS mean-level activity metrics and objective and subjective measures of gait speed, life space, and community participation that would reflect concurrent validity.

Exploratory Analyses

A series of exploratory analyses examined relations between GPS metrics and other participant factors such as demographics, technology habits, and smartphone type, as well as environmental factors (eg, COVID-19 and seasonal impact). Our goal was to better understand what intrinsic and extrinsic factors are associated with GPS trajectories in older adults so that future studies can consider these variables as potential covariates or moderators. This is important given the high heterogeneity in individual features and incidental factors that may impact the generalizability of between-group differences in GPS trajectories. Spearman correlation analyses were used for continuous, dichotomous, and ordinal variables, and 1-way ANOVA was used for categorical variables with a Welch test for unequal group variances.

To explore the utility of a GPS composite score, a Gaussian mixture modeling (GMM) approach was used to generate a nuanced yet singular representation of mobility trajectories. For each participant's set of daily GPS features (Table 1), we fit a GMM with k=3 mixture distributions to allow for flexible modeling with consideration of differences in the distribution for each person due to various factors (eg, weekend and weekday differences). The choice of k=3 was determined to maximize clustering complexity subject to our sample size limitations to avoid overfitting. This method also allowed us to reduce the full set of GPS features to a single dimension. Next, we created a distance matrix between each pair of participants in the sample by calculating the integral of the squared distance between each GMM density (the larger the distance, the more different the pair's overall mobility patterns). After calculating the distance matrix, we used multidimensional scaling (MDS) to extract a 1D representation of this distance matrix (hereafter termed "MDS1"), akin to a principal component [112]. This MDS1 metric is a relative measure that represents the similarity of overall mobility patterns across participants. For example, if 2 individuals have similar MDS1 values, their overall mobility patterns are similar, whereas 2 individuals with very different MDS1 values demonstrate different mobility patterns. The MDS1 variable was used in exploratory correlation analyses to identify whether patterns of mobility relate to validators, in contrast to total amounts of variability in individual mobility features.

Ethical Considerations

All aspects of the study protocol received ethics approval from the institutional review board at Temple University (protocol number 27013). As described earlier, all participants provided informed consent, were given the option to opt-out of study participation at any time, and were compensated for their time and effort. Study data were deidentified according to privacy procedures outlined earlier. Participants were compensated US $50 upfront for their participation.

Results

Participant Characteristics

A total of 37 individuals participated in our study between April 2022 and January 2024. The full sample was included in the analysis of feasibility and acceptability outcomes, whereas a subset (n=34, 92%) was included in preliminary validity analyses after excluding those who experienced major unexpected health changes. The 3 (8%) participants excluded due to major health events were aged on average 70.9 (SD 6.9) years, 67% (2/3) were female, 100% (3/3) identified as non-Hispanic White, and they completed on average 16.9 (SD 1.8) years of education. In addition, 2 (67%) of the 3 participants were Android users, and all had healthy cognition.

Of the validation subset (n=34), participation was distributed fairly evenly across all 4 seasons: 6 (18%) in the winter, 10 (29%) in the fall, 9 (26%) in the spring, and 9 (26%) in the summer. Participants' age ranged from 63 to 85 (mean 71.6, SD 5.5) years and they were on average highly educated (mean 16.4, SD 2.7 years; range 10-20 years). A majority of participants identified as female (23/34, 68%) and non-Hispanic or Latinx (33/34, 97%). A total of 56% (19/34) participants identified as White, 35% (12/34) as Black or African American, and 6% (2/34) as Asian. Most participants lived with others (23/34, 68%) and were retired (27/34, 79%). The majority (26/34, 76%) of participants were iPhone users, whereas 8 (24%) were Android users. Most participants met diagnostic criteria for healthy cognition (n=28, 82%) with a minority meeting criteria for MCI (n=5, 15%) or mild dementia (n=1, 3%). Scores on the Mini-Mental State Examination ranged from 24 to 30 (mean 28.3, SD 1.3), and according to the FAQ, participants on average experienced minimal difficulties with everyday functioning (mean 1.5, SD 2). Average responses on the GDS revealed low levels of self-reported depression (mean 1.5, SD 1.5) [113]. Participant demographic characteristics are detailed in Table 2, and scores on validation measures of cognition, everyday functioning, mood, and mobility are outlined in Table 3.

Table 2. Participant demographics. Includes data from subset (n=34) included in validation analyses.

Demographics	Values
Age (y), mean (SD)	71.6 (5.5)
Education (y), mean (SD)	16.4 (2.7)
Sex, n (%)	
Male	11 (32)
Female	23 (68)
Race, n (%)	
Asian	2 (6)
Black or African American	12 (35)
White	19 (56)
Not reported	1 (3)
English as a second language, n (%)	4 (12)
Ethnicity, n (%)	
Hispanic or Latinx	1 (3)
Not Hispanic or Latinx	33 (97)
Living status, n (%)	
Live alone	11 (32)
Live with others	23 (68)
Current occupational status, n (%)	
Full-time employee, volunteer, or student	3 (9)
Part-time employee, volunteer, or student	4 (12)
Retired	27 (79)
Highest annual household income (US $), n (%)	
<30,0000	2 (6)
30,000-49,000	4 (12)
50,000-69,000	3 (9)
70,000-89,000	5 (15)
90,000-99,000	3 (9)
100,000-149,000	6 (18)
≥150,000	8 (24)
Prefer not to answer	3 (9)
Phone type, n (%)	
iPhone	26 (76)
Android	8 (24)
Smartphone portability [a] , n (%)	
Yes	32 (94)
Half the time	2 (6)
No	0 (0)
Consensus diagnosis, n (%)	
Healthy cognition	28 (82)
MCI[b]	5 (15)
Mild dementia	1 (3)

[a]Smartphone portability=usually has smartphone when leaves home (no=1, in between=2, and yes=3).
[b]MCI: mild cognitive impairment.

Table 3. Participant baseline validation measures. Includes data from subset (n=34) included in validation analyses.

Validation measures	Scores, mean (SD; range)
Neuropsychological test (composite *t* scores)	
Global cognition (MMSE[a])	52.4 (8.1; 37-67)
Attention	52.1 (5.8; 38-66)
Processing speed	54.5 (7.9; 36-74)
Executive function	50.9 (7.3; 38-66)
Memory	47.4 (11.1; 20-75)
Language	50 (9.3; 28-72)
Self- and informant-reported functioning	
Functional Activities Questionnaire	1.5 (2; 0-6)
Everyday Cognition Scale-Short Form	1.3 (0.2; 1.0-1.8)
Mood	
Geriatric Depression Scale (raw score)	1.5 (1.5; 0-5)
Geriatric Depression Scale (*t* score)	52.2 (10.4; 38-76)
Mobility	
Timed up and Go Test (s)	11 (4; 6-29)
Life-Space Assessment	78.5 (20.1; 36-114)
Australian Community Participation Questionnaire	3.8 (1.8; 1-7)

[a]MMSE: Mini-Mental State Examination.

Feasibility and Acceptability (Primary Aim 1)

All 37 (100%) participants who began the study completed the 4-week monitoring period and session 2 without requesting to withdraw. Participants scored on average 97% (SD 5.7%) on the comprehension of consent quiz. In total, 2 (5%) participants had an initial score of 80% (8/10 questions correct), 7 (19%) scored 90%, and 28 (76%) scored 100% on their first attempt. The most frequently incorrect item was “Using the mindLAMP app will help improve my cognitive functioning,” to which 5 (13%) participants answered “yes.” Better performance on the comprehension of consent quiz was associated with higher education (r_s=0.65; P<.001) and differed across racial groups ($F_{2,33}$=8.4; η^2=0.34; P=.001), with better performance among White participants versus Black participants according to post hoc comparisons (P=.02). When including education as a covariate, group differences for race remained statistically significant, though a lower effect size was noted ($F_{2,32}$=3.3; η^2=0.17; P=.049). Performance on the quiz was not associated with age, English as a second language, or cognitive status (all *P* values >.05).

Satisfaction ratings on the debriefing questionnaire at session 2 (ie, responses to the question “How satisfied are you with the study team’s explanation of this study? Did the study team accurately convey what it would be like to participate in this study during the consent process at your first study visit?”) revealed high levels of satisfaction. Specifically, 84% (31/37) of participants reported they were “very satisfied,” and 16% (6/37) reported they were “satisfied.”

Regarding new issues with or changes to phone use, 92% (34/37) of participants reported they did not experience any new problems using their smartphone during the study period. However, 1 (3%) participant experienced technical issues that were determined to be unrelated to the study application, 1 (3%) reported that some of their text messages were disrupted (unrelated to the study app as we did not collect information from text messages), and 1 (3%) reported that their phone was “a little slow and lack of charge.” In total, 97% (36/37) of participants reported they did not experience any feelings of being uncomfortable, suspicious, or paranoid due to the study app running on their smartphone. Although participants were instructed to go about their daily lives and smartphone use as they normally would, 46% (17/37) of participants reported there were major changes in how they used their smartphone during the study period. Specifically, 3% (n=1) used their phone less overall, 41% (n=15) charged their phone more, 5% (n=2) carried their phone less, and 14% (n=5) carried their phone more.

Troubleshooting contacts related to missing GPS data were infrequent. During the entire study period across 37 participants, only 6 incidents were logged affecting 5 (14%) unique participants, with causes including (1) Android phone went into “safe mode,” (2) location permissions reset from “always” to “only while using app,” (3) outdated version of mindLAMP installed on phone, (4) low battery mode enabled, and (5) mindLAMP app was accidentally deleted. All incidents were promptly resolved by the study coordinator remotely by speaking with the participant over the phone to guide them to either reconfigure their phone settings or re-download and login to mindLAMP.

Validity Analyses (Primary Aim 2)

Preliminary Analyses

Examining the untransformed GPS data revealed that on average each day, participants spent 1074 (SD 192) minutes at home (ie, about 18 hours), traveled 42,676 (SD 34,694) m, spent time at 1.56 (SD 0.54) unique locations, and had 423 (SD 352) minutes of missing GPS data (approximately 29% of the day, which reflects relatively high data quality and frequency relative to other smartphone digital phenotyping studies using interval sampling approaches [111]). Correlations among GPS features revealed strong associations among features reflecting activity (distance traveled, radius of gyration, maximum diameter, and maximum distance from home), which were negatively associated with features reflecting inactivity (time spent at home and stationary time) and the 2 indices of physical circadian routine. Significant locations visited and significant location entropy were intercorrelated, suggesting a distinct construct related to location diversity. These associations together support the conceptual GPS feature categories outlined in Table 1 and are used throughout to streamline the presentation of results. Tables S1 and S2 in Multimedia Appendix 2 present descriptive statistics of GPS data and intercorrelations among all GPS features.

Relations Between Average Mobility Features and Measures of Cognition, Mood, and Everyday Function

We predicted significant relations between monthly average GPS activity features and baseline neuropsychological measures, mood, and everyday function, such that greater overall mobility would be associated with better performance on neuropsychological tests, less depression, and less reported cognitive and functional decline. Most correlations were in the predicted direction (Table 4; results for individual neuropsychological tests and function questionnaires are reported in Table S3 in Multimedia Appendix 2).

Correlations with neuropsychological composites and individual tests showed that greater GPS activity was associated with better scores on the language composite (Table 4) and Digit Span Forward test (Table S3 in Multimedia Appendix 2). GPS measures of inactivity and physical circadian routine were associated with lower language composite and individual test scores. Of note, correlations between greater physical circadian routine and lower language scores were the only relations to survive correction for multiple comparisons ($P<.0038$; Table 4; Table S3 in Multimedia Appendix 2). Associations between GPS activity features and the memory composite were not significant and weak but were in the opposite direction of most other neuropsychological composites. Analyses of individual neuropsychological tests showed greater average flight length was associated with worse delayed verbal recall ($r=–0.38$, $P=.03$; Table S3 in Multimedia Appendix 2), but the correlation coefficient was not statistically significant after correction for multiple comparisons.

Regarding mood, more inactivity (eg, more home time), greater routine, and less location diversity were associated with greater depression symptoms (Table 4). Less location diversity was also associated with greater reported functional impairment (FAQ; $r=–0.36$, $P=.04$; Table S3 in Multimedia Appendix 2). However, the associations between mobility features and self-reported mood and function did not survive correction for multiple comparisons. Correlations between GPS features and self-reported cognitive decline on the Everyday Cognition Scale-Short Form were weak and not statistically significant (Table S3 in Multimedia Appendix 2).

Table 4. Bivariate Pearson correlations between monthly average GPS features and neuropsychological measures (*t* scores)[a].

Category and GPS monthly average feature	Global cognition (MMSE[b])	Attention	Processing speed	Executive function	Memory	Language	GDS[c]
Activity							
Distance traveled	0.15	0.33[d]	–0.01	0.07	–0.17	0.33[d]	–0.24
Radius of gyration	0.21	0.25	–0.03	0.11	–0.10	0.34[d]	–0.21
Maximum diameter	0.19	0.33[d]	0.01	0.10	–0.12	0.35[d,e]	–0.20
Maximum distance home	0.22	0.27	0	0.08	–0.10	0.43[d,e]	–0.19
Average flight length	0.22	0.26	–0.05	0.03	–0.27	0.38[d,e]	–0.33[d]
Average flight duration	–0.05	0.15	–0.10	0.07	–0.01	0.16	0.01
Inactivity							
Home time	–0.08	–0.11	0.22	0.16	0.17	–0.45[d,f]	0.38[d,e]
Probability paused	–0.12	–0.28	0.13	–0.02	0.11	–0.22	0.29
Routine							
Circadian routine	–0.07	–0.11	0.16	0.16	0.20	*–0.51*[d,f,g]	0.40[d,e]
Weekend circadian routine	–0.08	–0.12	0.14	0.15	0.19	*–0.53*[d,f,g]	0.38[d,e]
Location diversity							
Significant location entropy	0.04	0.20	–0.09	–0.06	–0.25	0.23	–0.36[d,e]
Significant locations visited	–0.20	0.26	–0.02	–0.21	–0.14	0.07	–0.35[d,e]
Other							
Minutes missing	0.28	0.47[d,f]	0.08	–0.01	–0.24	0.37[d,e]	–0.39[d,e]

[a]Data represent effect size as measured by bivariate Pearson correlation coefficients (*r*), whereby 0.10, 0.30, and 0.50 represent small, moderate, or large effects, respectively. All neuropsychological measures reflect *t* scores corrected for age, sex, education, and estimated premorbid IQ.

[b]MMSE: Mini-Mental State Examination.

[c]GDS: 15-item Geriatric Depression Scale *t* score.

[d]Moderate to large effect size.

[e]*P*<.05 (2-tailed).

[f]*P*<.01 (2-tailed).

[g]Italics indicate correlation coefficients surviving Bonferroni correction (*P*<.0038).

Relations Between Average GPS Features and Measures of Gait Speed, Life Space, and Community Participation

Contrary to our hypotheses, performance on the TUG measure of gait speed was not associated with any of the average GPS features (Table 5). By contrast, self-reported measures of geospatial life space (LSA), and community participation (ACPQ) were associated with many GPS features. Overall, greater GPS activity was associated with more self-reported life space and more community participation, whereas greater inactivity and physical circadian routine were associated with less geospatial life space and less community participation. Correlations between greater radius of gyration, less physical circadian routine, and greater geospatial life space survived correction for multiple comparisons (*P*<.0038; Table 5). Relations with the ACPQ were driven by the domains of extended family and friends (eg, participants who reported they tend to visit friends more often demonstrated significantly higher levels of GPS activity and lower physical circadian routine indices; *P*<.0038; Table S5 in Multimedia Appendix 2).

Table 5. Bivariate Spearman correlations between average GPS features and measures of gait speed, life space, and community participation[a].

GPS monthly average feature	TUG[b]	LSA[c]	ACPQ[d]
Activity			
Distance traveled			
r	–0.16	0.40[e]	0.34[e]
P value	.37	.02	.048
Radius of gyration			
r	–0.17	*0.52* [e,f]	0.39[e]
P value	.35	.002	.02
Maximum diameter			
r	–0.18	0.44[e]	0.40[e]
P value	.31	.009	.02
Maximum distance home			
r	–0.11	0.44[e]	0.42[e]
P value	.54	.008	.01
Average flight length			
r	–0.05	0.23	0.29
P value	.77	.18	.10
Average flight duration			
r	0.01	0.34[e]	–0.09
P value	.94	.052	.62
Inactivity			
Home time			
r	0.12	–0.42[e]	–0.39[e]
P value	.49	.01	.02
Probability paused			
r	0.13	–0.48[e]	–0.24
P value	.46	.004	.18
Routine			
Circadian routine			
r	0.12	–0.45[e]	–0.45[e]
P value	.50	.007	.008
Weekend circadian routine			
r	0.18	*–0.50*[e,f]	–0.48[e]
P value	.32	.002	.004
Location diversity			
Significant location entropy			
r	–0.15	0.34[e]	0.30[e]
P value	.40	.052	.08
Significant locations visited			
r	–0.18	0.29	0.21
P value	.31	.09	.23

GPS monthly average feature	TUG[b]	LSA[c]	ACPQ[d]
Other			
Minutes missing			
r	0.17	0.16	0.32[e]
P value	.34	.35	.07

[a]*r* represents effect size as measured by bivariate Spearman correlation coefficient, whereby 0.10, 0.30, and 0.50 represent small, moderate, or large effects, respectively.

[b]TUG: Timed Up and Go Test.

[c]LSA: Life-Space Assessment.

[d]ACPQ: Australian Community Participation Questionnaire.

[e]Moderate to large effect size.

[f]Italics indicate correlation coefficients surviving Bonferroni correction (*P*<.0038).

Relations Between GPS Variability and Measures of Cognition, Mood, and Everyday Function

We predicted a negative relation between day-to-day variability in GPS features and baseline measures of cognition, mood, and everyday function, such that greater IIV would be associated with lower neuropsychological test scores, more depression, and more reported cognitive and functional decline. Contrary to our hypotheses, we saw that higher variability in most GPS features (ie, greater day-to-day iSD in mobility habits) was associated with *better* scores on neuropsychological measures of attention and language, as seen in Table 6 and in Tables S6 and S7 in Multimedia Appendix 2. Correlations between greater IIV in home time and physical circadian routine and higher language composite scores survived correction for multiple comparisons (*P*<.0038; Table 6). Again, relations with memory were in the opposite direction, such that higher GPS IIV in location entropy was associated with a worse memory composite. Analyses with individual tests showed significant associations between higher GPS IIV in average flight length and worse verbal memory and higher GPS IIV in location entropy and worse visual memory (*P*<.0038; Table S7 in Multimedia Appendix 2). Regarding relations with mood, greater IIV in location diversity was associated with less depression, though results did not survive correction for multiple comparisons. This was contrary to expectations but consistent with the aforementioned results suggesting greater variability in mobility habits is overall beneficial. GPS IIV was not significantly associated with self-reported cognitive decline or functional impairment, though results are directionally consistent, such that greater IIV was weakly related to less reported functional decline (Table S7 in Multimedia Appendix 2).

Table 6. Bivariate Pearson correlations between monthly IIV GPS features and neuropsychological measures (*t* scores)[a].

Category and GPS monthly IIV[b] feature	Global cognition (MMSE[c])	Attention	Processing speed	Executive function	Memory	Language	GDS[d]
Activity							
Distance traveled	0.22	0.24	0.01	0.15	–0.08	0.37[e,f]	–0.14
Radius of gyration	0.27	0.15	0	0.12	0.05	0.33[e]	–0.07
Maximum diameter	0.27	0.20	0.06	0.13	0.03	0.37[e,f]	–0.06
Maximum distance home	0.27	0.16	0.03	0.12	0.04	0.40[e,f]	–0.06
Average flight length	0.27	0.16	0.04	0.18	–0.32[e]	0.35[e,f]	–0.28
Average flight duration	0	0.06	–0.08	0.10	–0.03	0.18	0.02
Inactivity							
Home time	0.23	0.15	–0.13	0	–0.18	*0.64* [e,g,h]	–0.31[e]
Probability paused	0.17	0.27	–0.10	0.13	–0.09	0.29	–0.10
Routine							
Circadian routine	0.15	0.17	0.10	0.07	–0.02	*0.59* [e,g,h]	–0.20
Weekend circadian routine	0.16	0.13	–0.03	0.04	–0.02	*0.55* [e,g,h]	–0.28
Location diversity							
Significant location entropy	0.01	0.25	0.06	–0.09	–0.45[e,g]	0.35[e,f]	–0.47[e,g]
Significant locations visited	–0.21	0.11	–0.07	–0.33[e]	–0.13	0.32[e]	–0.36[e,f]
Other							
Minutes missing	0.29	0.06	–0.13	0.03	0	0.41[e,f]	–0.22

[a]Data represent effect size as measured by bivariate Pearson correlation coefficients (*r*), whereby 0.10, 0.30, and 0.50 represent small, moderate, or large effects, respectively. All neuropsychological measures reflect *t* scores corrected for age, sex, education, and estimated premorbid IQ. Table S6 in Multimedia Appendix 2 gives exact *P* values.

[b]IIV: intraindividual variability.

[c]MMSE: Mini-Mental State Examination.

[d]GDS: 15-item Geriatric Depression Scale *t* score.

[e]Moderate to large effect size.

[f]*P*<.05 (2-tailed).

[g]*P*<.01 (2-tailed).

[h]Italics indicate correlation coefficients surviving Bonferroni correction (*P*<.0038).

Exploratory Analyses

Relations Between GPS Metrics and Participant Intrinsic and Extrinsic Factors

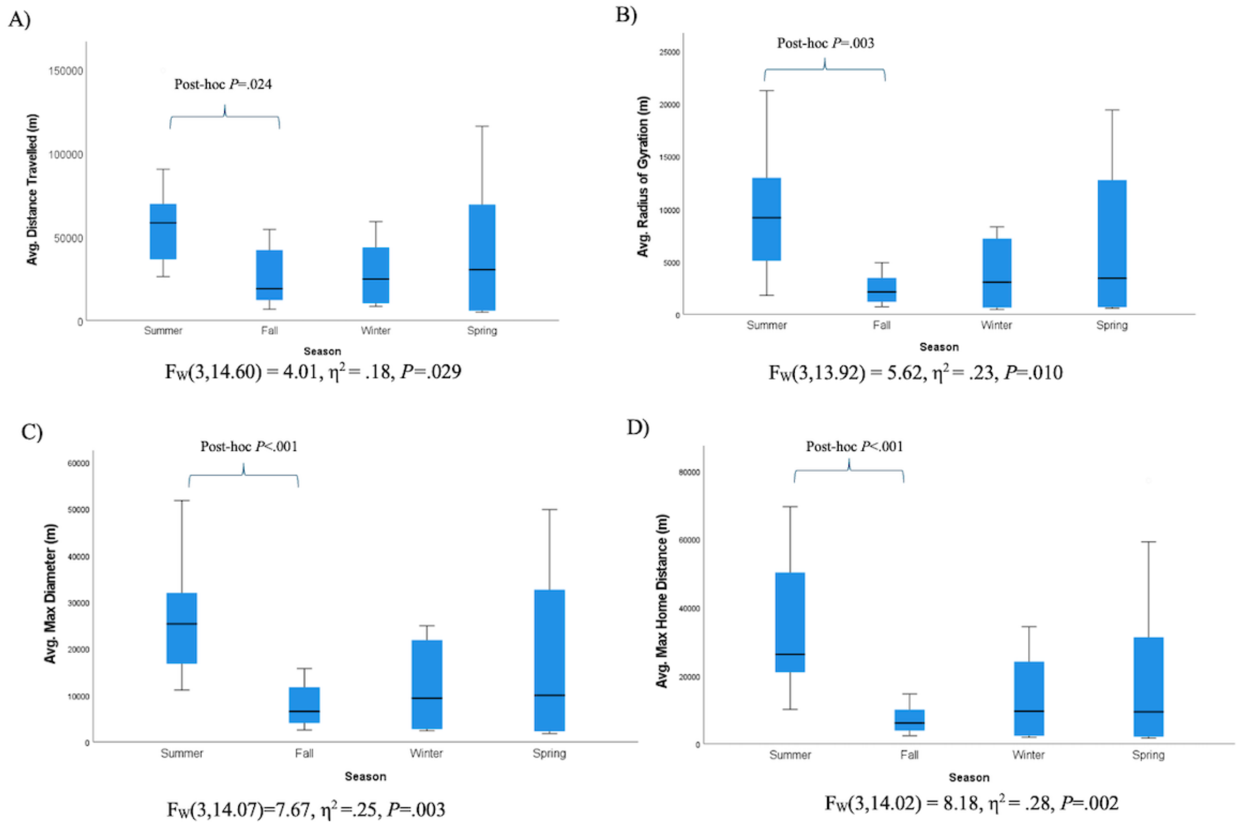
To inform future studies in the selection of covariates or moderating variables, we conducted exploratory correlations between GPS average and IIV features and various participant, environmental, and contextual factors (Tables S8 and S9 in Multimedia Appendix 2). Average GPS features were unrelated to several sociodemographic factors, including age, sex, and cohabitation status (all *P* values >.05). Higher education was associated with less location diversity only (*r*=–0.35; *P*=.04) but was not significant after correction for multiple comparison. By contrast, lifetime annual income and native language status appeared to be more relevant. Overall, higher lifetime income and native English language status were associated with greater levels of GPS activity, less routine, and greater location diversity (0.36≤|*r*|≤0.55). The association between higher lifetime income and greater GPS activity measures (radius of gyration and maximum distance from home) remained significant after correction for multiple comparisons (*P*<.0038; Table S8 in Multimedia Appendix 2).

Technology factors were largely unrelated to average GPS features. Phone type (iPhone vs Android) was unrelated to all features except for the overall number of minutes missing; iPhone users had significantly less missing GPS data than Android users, and the relation remained significant after correction for multiple comparisons (*r*=–0.64; *P*<.0038). More habitual smartphone use was associated with less location diversity only (*r*=–0.40; *P*=.02), though this relation did not survive correction for multiple comparisons. Whether or not participants typically carry their phone when leaving home (smartphone portability) was unrelated to all average GPS features, as was digital literacy as measured by the Mobile Device Proficiency Questionnaire (all P values >.05). Regarding the impact of COVID-19, individuals who reported more COVID-19–related isolation (ie, due to limited or canceled social gatherings) demonstrated significantly less location diversity (r=–0.51; P<.0038), whereas those who reported a greater expansion in mobility habits due to COVID-19 demonstrated more location diversity (r=0.34; P=.047), though this latter relation was not significant after correction for multiple comparisons.

One-way ANOVA of average GPS features was used to explore group differences across categorical demographic variables (race, occupational status, and season), with the Welch test for unequal group variances. Group differences for race (Black, Asian, and White) were observed only on the minutes missing feature ($F_{2,30}$=6.32; η^2=0.30; P=.005), with Black participants having greater amounts of missing GPS data than White and Asian participants according to post hoc comparisons (P=.003 and P=.02, respectively). A follow-up chi-square test was performed to examine the relationship between race and phone type given prior findings that Android users have greater missing data than iPhones [114]. The relation was significant, such that a greater proportion of Black participants owned Androids, χ^2_2=6.9, P=.03 (N=33), suggesting that group differences in missing data could be related to phone type.

No group differences in average GPS features were observed according to current occupational status (eg, retired and working full vs part-time). By contrast, several activity features were significantly different across the winter, fall, spring, and summer seasons ($4.01 \leq F_{3,30} \leq 8.8$; $0.18 \leq \eta^2 \leq 0.28$; all P values <.05). According to post hoc comparisons, these differences were driven by significantly greater GPS activity in the summer versus fall months (Figure 3).

Figure 3. One-way ANOVA of GPS activity features by season.

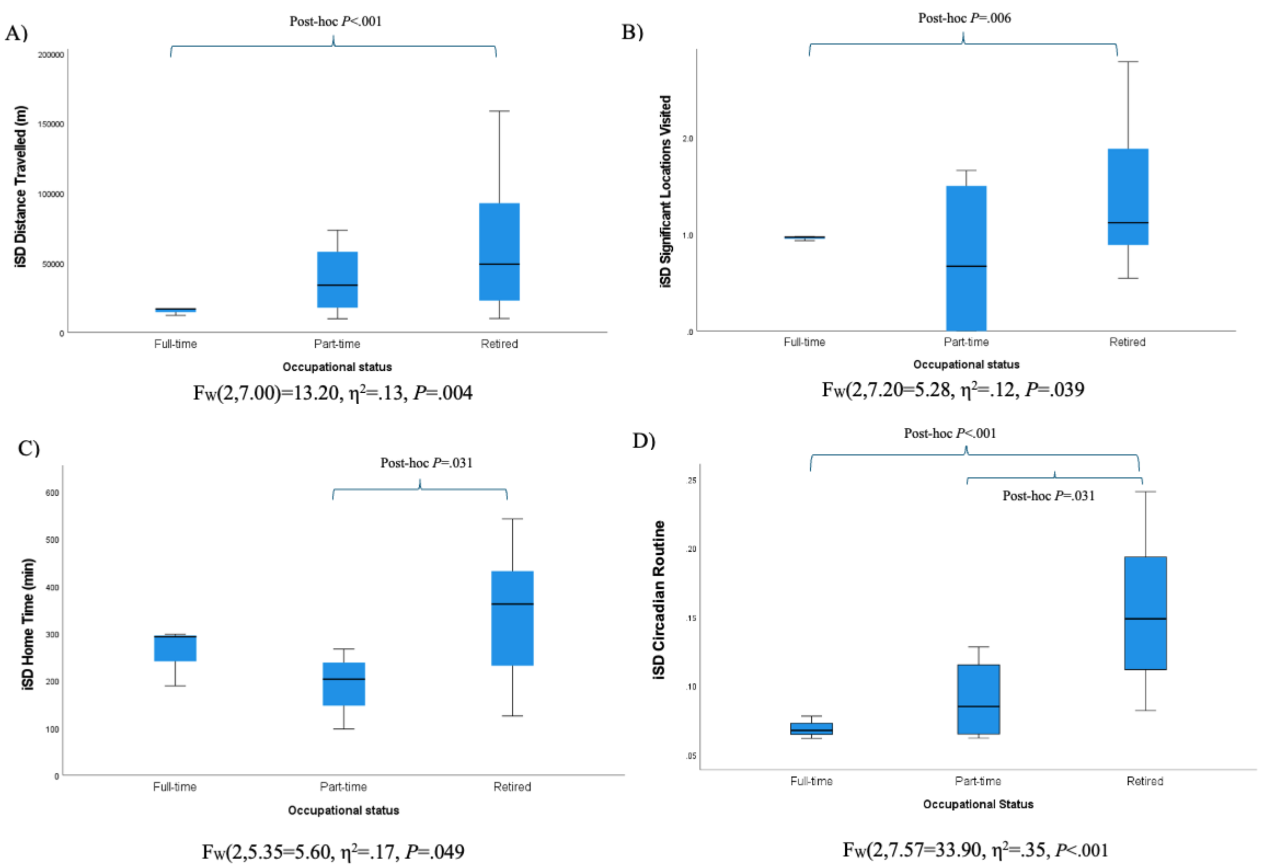


Spearman correlations between GPS IIV features and the participant factors assessed earlier are detailed in Table S9 in Multimedia Appendix 2. Similar to the results with average GPS metrics, relevant sociodemographic factors appeared to be income and native language status. Higher lifetime income and native English language status were associated with more variability in activity, routine, and location diversity features ($0.35 \leq |r| \leq 0.56$), with associations between higher income and greater IIV in activity (radius of gyration, maximum diameter, and maximum distance from home) surviving correction for multiple comparisons (P<.0038).

Phone type was more prominently related to GPS IIV measures than GPS average measures; iPhone users showed greater GPS IIV than Android users on several GPS features ($0.34 \leq r \leq 0.52$), though only greater GPS IIV in minutes missing survived correction for multiple comparisons (P<.0038). Of note, iPhone ownership was also positively correlated with income (r=0.49; P=.005), making it difficult to determine whether higher levels of IIV among iPhone users relate to differences in phone type or to behavioral differences in lifestyle afforded by higher income. Regarding other technology factors, habitual smartphone use and smartphone portability were unrelated to GPS IIV metrics, whereas digital literacy was related to more variability in routine only (r=0.34; P=.047)—though this relation did not survive correction for multiple comparisons. Individuals who reported more COVID-19–related social isolation had significantly lower IIV in location diversity (ie, more day-to-day consistency in the number of locations visited; r=–0.55; P<.0038), whereas those who reported a greater expansion in mobility habits due to COVID-19 had greater IIV in location diversity (r=0.37; P=.03), though this latter relation was not significant after correction for multiple comparisons.

One-way ANOVA of GPS IIV features across categorical demographic variables revealed a significant difference across racial groups on only 1 measure—IIV for flight length ($F_{W2,18.63}$=11.96; η^2=0.12; P<.001). This group difference was driven by lower IIV in Asian participants compared with both Black participants and White participants according to post hoc tests. Effects of current occupational status were observed on several IIV metrics (5.28≤$F_{2,31}$≤33.90; 0.12≤η^2≤0.35; all P values <.05), such that retired participants demonstrated significantly greater IIV in mobility habits compared to those working full- or part-time (Figure 4). Consistent with the results of GPS average metrics, several GPS IIV metrics differed according to the season of study participation, with significantly more IIV in activity features observed during summer months versus the fall and spring months (2.41≤$F_{3,30}$≤3.78; 0.19≤η^2≤0.27; all P values <.05).

Figure 4. One-way ANOVA of GPS intraindividual variability features by occupational status.



Relative Mobility Patterns

A GMM, MDS approach was used to derive a relative mobility feature (MDS1), including all GPS features simultaneously. This allowed us to explore how similarity in mobility patterns, as reflected through a 1D value, related to validators of cognition, function, and mood. Relative mobility patterns were moderately associated with a cognitive measure of language (|r|=0.36; P=.04) and a global cognitive screener (|r|=0.31; P=.08; Table S10 in Multimedia Appendix 2). Relative mobility patterns were also associated with depression (|r|=0.39; P=.02) and community participation (|r|=0.51; P=.002). Of note, only the relation between MDS1 and community participation (ACPQ) survived correction for multiple comparisons (P<.0029). While the MDS1 feature represents a linear contrast combining multiple individual GPS features that excel at differentiating mobility patterns across the sample, this comes at the cost of interpretability of the associations related to MDS1 (eg, positive vs negative correlation coefficients are not meaningful); however, these results provide additional converging evidence that individuals with similar mobility profiles may have similar levels of underlying cognitive ability, depression, and community participation.

Discussion

Principal Findings

This study investigated a 4-week smartphone digital phenotyping protocol as a novel method to assess everyday cognition, function, and mood in a cohort of 37 older adults. Our preliminary, proof-of-concept results suggest that theoretically informed digital phenotypes of mobility are feasibly captured from older adults' personal smartphones and are associated with clinically relevant data pertinent to cognitive aging and AD/ADRD. Findings and implications provide key insights to inform the design and interpretation of future studies using this method in larger, more diverse cohorts.

One of our primary aims was to examine the feasibility and acceptability of smartphone digital phenotyping among older adults. All participants completed the 4-week monitoring period without dropping out, and 97% (36/37) reported no feelings of discomfort during debriefing procedures. By contrast, almost half (17/37, 46%) reported changes in how they used their smartphone, with 41% (15/37) endorsing charging their phone more frequently. Battery drain was a communicated risk and is common in high-frequency continuous data collection, including GPS [111]. Increased phone charging behavior may limit the naturalistic aspect of this approach and will be important to address in future designs—particularly if battery power is considered as a digital biomarker in and of itself. Future studies should explore whether lower sampling frequencies are sufficient, as this can be adjusted via the LAMP platform and would lead to less battery drain. It is also likely that advances in smartphone battery life span will ultimately circumvent this issue; in the interim, participants may be provided with portable batteries for daily outings.

We were also interested in how participants would respond to and comprehend complex technical details of this study. Participants demonstrated good comprehension of study procedures, as demonstrated by an average score of 97% on the comprehension of consent quiz, and all reported satisfaction with the study team's initial review of study procedures. These findings are encouraging given the dearth of studies investigating older adults' attitudes and concerns about passive sensing technologies [78,115] and suggest that our upfront efforts to enhance privacy, security, transparency, and comprehension were effective [79]. However, despite high overall comprehension, we observed that higher education and identifying as White versus Black were associated with better performance on the comprehension of consent quiz—suggesting the language used in our materials may not be culture-free and should be revised using co-design or focus group approaches with increasingly diverse perspectives. Importantly, the most frequently incorrect item on the quiz pertained to potential benefits of study participation ("Using the mindLAMP app will help improve my cognitive functioning," yes or no). Clear communication of potential benefits in research is a core ethical requirement, and future study materials should clarify expectations for potential benefits to facilitate trust, particularly among historically marginalized groups [116-119].

Regarding our second primary aim of establishing preliminary validity, we observed converging support that unobtrusively obtained movement trajectories from smartphones are related to established clinically relevant variables, including cognition, function, and mood. This is not surprising but is highly encouraging. Movement trajectories reflect many facets of everyday cognition, including the ability and motivation to travel outside the home, the degree of someone's spatial routine, number of unique locations someone can visit, and the total amount and duration of movement—behaviors that require abundant cognitive and psychological resources. Here, individuals with better cognition, less functional impairment, and less depression did, in fact, demonstrate greater overall mobility according to several GPS features. Specifically, they traveled farther, spent less time at home, and had greater diversity in the locations they visited. Although these associations did not remain statistically significant after correction for multiple comparisons, these moderate-level effects in a relatively small sample are encouraging as they are consistent with our hypotheses and with prior studies showing greater physical activity, less time at home, greater life space (the extent of movement through the environment during daily functioning), and engagement in a variety of activities (environmental complexity) are associated with better cognition, less depression, and reduced risk for MCI and dementia or AD/ADRD [66,103,104,120-128].

Inconsistent with our hypotheses, individuals with better cognition demonstrated significantly *greater* day-to-day variability in GPS features and *less* physical circadian routine, and these associations did survive correction for multiple comparisons. We also saw that greater mobility variability and less routine were associated with less depression, although associations did not survive correction for multiple comparisons. Regarding depression, it is conceptually reasonable that more varied mobility habits could be protective against depression, and this has been demonstrated in at least 1 study of younger adults where lower location diversity was associated with more depression symptoms [129]. With respect to the significant negative associations between cognition and routine, it is possible that the older adults in our cohort with more cognitive difficulties intentionally engaged in more predictable and less demanding daily activities to compensate for underlying mild difficulties, leading to more consistent physical circadian routines and lower day-to-day variability in mobility habits. This is a pattern identified in the literature and is often recommended as an intervention in clinical practice [130-134].

The VIBE model predictions regarding variability were informed by observations that individuals with cognitive impairment demonstrate increased IIV compared with healthy controls while performing single, standardized tasks in the clinic or laboratory where task demands are the same for everyone. Greater IIV on constrained tasks reflects an inability to maintain consistent levels of performance [7,135,136]. This study instead involved unconstrained mobility habits, which individuals may modify to compensate for cognitive difficulties. Therefore, the observation of GPS IIV as indicative of positive rather than negative outcomes may be specific to unconstrained geospatial routines or to our relatively functionally healthy sample. We may still observe that greater IIV associates with worse outcomes when considering more fine-grained digital biomarkers (eg, diurnal phone use patterns and accelerometer-based sleep estimates) that are relatively more constrained, as has been identified in other studies examining IIV in gait speed, medication-taking routine, and computer use [137-139]. We may also see that greater IIV in broad everyday behaviors is a marker of resilience early on but becomes maladaptive in later stages of neurocognitive decline. Larger samples with greater heterogeneity in cognitive ability are needed to answer this question, in addition to longitudinal designs that would enable monitoring of within-person IIV trends over time. In the meantime, our relatively cognitively healthy sample provides insights into how variability behaves in early stages and unconstrained settings. Consistent with past reports of significant task and timescale effects on IIV [140], IIV in continuous mobility trajectories may be mechanistically distinct from IIV in constrained settings.

Concurrent validity was supported by strong and significant associations between mean-level GPS features and baseline scores on the LSA and the ACPQ—self-report measures of geospatial life space and community activity participation, respectively. These 2 constructs are highly relevant in the context of aging and AD/ADRD and represent key measures of risk. The LSA measures the extent, frequency, and independence of movement within and outside the home. Constricted life space has been associated with increased risk of AD, MCI, and cognitive decline in racially diverse groups and may represent an early functional marker of prodromal decline as individuals compensate for early subtle difficulties and limit their range of movement or activity complexity [103,104,130]. Social engagement—particularly leaving the home to visit friends—protects against social isolation, promotes cognitive reserve, and represents a complex activity requiring cognitive flexibility [141-145]. Thus, the ability to unobtrusively and longitudinally measure these key risk and resilience factors without the burden or bias of self- or informant-report represents a noteworthy application for smartphone-derived mobility trajectories. Minimal associations were identified between GPS features and the TUG measure of gait speed. It is possible that low heterogeneity in TUG scores or small variations in the administration of the TUG played a role. It is also likely that gait speed and coordination are lower-level features of mobility that are independent of broader mobility habits, at least within this sample of functionally independent older adults.

We examined many individual and contextual influences on mobility features to inform future studies in selecting covariates or moderating factors. This is a critical open question in the field of digital phenotyping [66], and our preliminary results provide important insights into which sociodemographic, contextual, and technological factors should be considered when interpreting these data. Age, education, sex, race, and cohabitation status appeared to be minimally associated with most mean-level and variability metrics, providing partial support for the objectivity of digital phenotyping features. Nonetheless, other sociodemographic factors, including higher lifetime income and native English language status, were moderately associated with several mobility metrics that appear to be advantageous (ie, greater activity and greater IIV). This could reflect an association between social determinants of health and access to transportation, opportunities for socialization outside of the home, or an ability to engage in spontaneous activities—and may suggest mechanisms for income and acculturation as resilience factors.

Other extrinsic factors with relatively clear and interpretable effects included season, COVID-19–related effects, and occupational status. Participants demonstrated greater activity, less home time, and more mobility variability in the summer months, which is reasonable given the increased leisure activities that typically occur in the summer. Therefore, seasonal differences are relevant if interpreting data in pre-post designs, suggesting investigators should aim to control for season or re-evaluate during the same season if possible. Participants reporting more COVID-19–related isolation visited fewer locations, whereas those reporting more COVID-19–related mobility expansion visited more. Individuals who were retired demonstrated more variable mobility habits than those working full or part time, which aligns with differential degrees of consistency depending on occupational status. In addition to shedding light on which factors are relevant when interpreting digital phenotyping data, these associations provide additional validation for the mobility features in this study.

Phone type was unrelated to all mean-level GPS metrics except for the number of minutes of missing data, which was significantly lower for iPhones. This finding is consistent with a previous study by Kiang et al [114] that identified lower rates of missing GPS data among iOS users and overall bodes well for the generalizability of mean-level metrics across different phone types and operating systems. Missing data were also higher among Black participants, which could be due to a higher proportion of Android ownership among Black participants in our study. Thus, controlling for phone type may be important when interpreting the minutes of missing data feature. iPhone users also demonstrated more variability in most GPS features, although only the missing data feature remained statistically significant after correction for multiple comparisons. Given iPhone ownership was positively associated with income, it is difficult to interpret whether increased variability among iPhone users was related to operating system factors, to aspects of resilience associated with income, or whether this reflects a spurious finding. Future studies with larger sample sizes should work to clarify these questions.

In considering the neuropsychological correlates of our GPS features, relations with cognition were strongest and most consistent for measures of language. In fact, relations between GPS features and the language composite were the only ones to survive correction for multiple comparisons (ie, less overall physical circadian routine, greater IIV in circadian routine, and greater IIV in home time were all strongly [$|r|$>0.5] associated with the language composite at P<.0038). This was somewhat unexpected given language abilities are typically not as critical to the completion of everyday activities compared to domains such as executive functioning [146,147]. Nonetheless, intact language functioning (specifically semantic access and retrieval) relies on left temporal and prefrontal integrity and connectivity, which are highly relevant neuroanatomical regions in the pathogenesis of AD. Interestingly, relations between GPS features and measures of memory were in the opposite direction compared to other cognitive domains; for example, more day-to-day variability in average flight length was associated with worse delayed memory performance. This finding may reflect inefficient planning and daily errors resulting in occasional backtracking when visiting common locations due to forgotten items. In general, differential relations between

GPS features and measures of memory and language—2 cognitive domains implicated in AD pathology—suggest relations between GPS features and cognitive abilities may not be as straightforward as our model predicted. Future studies should continue to investigate whether mobility phenotypes are uniquely related to specific neuropsychological and neuroanatomical correlates, rather than focus on a global cognitive composite. This may involve developing even more fine-grained GPS features to capture distinct functional difficulties in everyday life (eg, backtracking and repetitive motions).

Limitations and Strengths

Our study has several limitations that are worth noting. Our relatively small sample size of 37 individuals coupled with a relatively large number of analyses limits generalizability; thus, findings should be considered preliminary and proof-of-concept. Digital phenotyping is a new research area with relatively few established standards, yet preliminary validation of digital phenotypes has been reported in samples of under 50 participants [61,73,148,149] for ≤30 days of data collection [73,150-152]. Another major limitation is the restricted diversity of our cohort in terms of cognition and diagnostic severity (which precluded investigation of diagnostic group differences), demographics (ethnicity, education, and socioeconomic status), and severity of depressive symptoms. Furthermore, all participants lived within driving distance of Temple University and therefore reflected an urban and suburban cohort. In addition, individual data-processing exclusions were required to account for unanticipated travel and health events. Given our aim of establishing preliminary validity, there was a need to ensure participants met strict inclusion criteria, which included relative stability in their health and physical location. This level of control and oversight may be unrealistic in larger, longitudinal studies and may need to be replaced by advanced statistical methods in the future [153]. Finally, it is worth mentioning that mobility traces from smartphone GPS sensors are a proxy for actual movement trajectories and depend upon the participant having their smartphone on their body, requiring multiple imputation and inference for missing GPS data before feature calculation during preprocessing [109]. It is also critical that participants are the unique users of their smartphones, as others have mentioned issues with shared devices [154,155]. Sensors worn on the body may provide more accurate and reliable measures of mobility patterns but present other drawbacks (eg, may be perceived as more intrusive by participants, may disrupt existing habits, and are typically costly).

This study has several strengths. Our cohort was well characterized compared to many prior digital phenotyping studies, including a comprehensive neuropsychological battery with 10 individual neurocognitive tests, objective and subjective validation measures, and application of combined actuarial and consensus diagnosis criteria to accurately classify participants. Our study design and hypotheses were theoretically informed by a conceptual model, which improves the interpretability and replicability of our findings. Toward this aim, we examined a range of interpretable mobility features (ie, 13 monthly average features and 13 monthly iSD features) with correction for multiple comparisons before reducing features into a principal component (MDS1). Results from the singular MDS1 component provided additional evidence that individuals with similar cognitive, functional, and mood profiles may demonstrate similar mobility profiles. Although many GPS features were intercorrelated, the presence of differential and unique correlations suggests individual features may be useful in clarifying specific behaviors and should be preserved as we continue to learn more about what these features signal. In addition, relative mobility profiles similar to the singular MDS1 component may be useful in identifying and classifying individuals with similar levels of underlying community participation and functional resilience. With regard to our technical protocol, the use of an open-source platform and a publicly available data processing script facilitates replicability, which is key to ongoing validation efforts [156]. Attempts at minimizing missing data using an automated checking script represent another strength and should be incorporated in future studies given the impact of missing data on subsequent findings [66]. Finally, our device-agnostic approach leveraging personally owned smartphones versus a study-issued device represents a notable strength as it promotes the naturalistic, unobtrusive nature of this method and affords increased scalability.

Future Directions

Future studies with larger and more diverse cohorts will be critical to replicate the present findings and address open questions. Given the lack of accessible and unbiased diagnostics available to individuals from low-income and marginalized groups [53-55,157], increased diversity in terms of race, ethnicity, educational attainment, and socioeconomic status is a priority for subsequent studies. Validating this method in more clinically heterogeneous samples will be critical to evaluate its ability to distinguish between age-related cognitive decline, neurodegenerative decline, and other medical or psychiatric conditions, and between various clinical and biological subtypes of dementia and ADRD. Both activity-level and IIV metrics should be evaluated, as variability appears to be fundamentally distinct from mean-level metrics and relates to AD risk [158]. Next steps will also investigate relations between validation measures and other sensors (accelerometer, device state, and steps) to extend the present findings and test our theoretical framework in behavioral features other than mobility trajectories. Machine learning or cluster approaches integrating multiple sensor streams may be helpful in determining clinically useful digital phenotypes, thereby reducing the analytic burden and narrowing the focus on clinically relevant, nonredundant features. Additional open questions include the test-retest reliability of digital phenotypes; the incremental utility of ecological momentary assessment, which can provide context to passive data [7,49,159]; determining the minimum necessary sensor sampling rate and duration to reduce battery drain; and evaluating within-day variability, diurnal patterns, and time of day effects on GPS features [26,49,75,138,160].

Because there is no ground truth for the application and interpretation of digital phenotyping data in aging and ADRD populations, additional studies are needed so that results can be compared across samples and insights can be consolidated to inform gold-standard approaches and normative data. As the field continues to evolve, it is likely that longitudinal monitoring of within-person fluctuations will be the ideal approach to capture individually relevant changes rather than population- or group-based approaches [7,63,156,161]. This is particularly relevant in the context of first-line primary care screening and ADRD clinical trials, where scalable risk detection, trial screening, and treatment response measures are direly needed. However, with longitudinal designs and increased sample sizes, it will be important to consider whether the data storage, processing, and interpretation burdens justify breaking from the status quo. Workflows that enhance the efficiency and scalability of this method will be critical, not just for research participants and patients but also for those collecting and interpreting the data [7,80,111].

Conclusions

As the population of older adults continues to rise, efforts to identify new tools to detect risk for future cognitive decline and measure treatment response are critical, particularly as new pharmacological interventions gain approval. Current gold standard methods are costly, burdensome, not widely accessible, and not part of routine clinical care. Measures that are sensitive to cognitive decline that can be passively and affordably integrated into everyday life without burden, disruption, or self-report bias are needed to serve as a first-line approach. Our study demonstrates that unobtrusively obtained GPS movement trajectories from personal smartphones may in the future be one such first-line approach, enabling clinicians and researchers to efficiently assess cognitive status, mood, and dementia risk on a broader scale. Individuals with at-risk data profiles may ultimately be referred for more comprehensive evaluation and directed toward appropriate intervention or research settings, leading to cost savings, reduced burden, and faster access to care. Much work remains to be done before we determine how smartphone digital phenotyping can be integrated into our current health care system; however, preliminary results suggest that it is a worthwhile endeavor and serve to inform follow-up studies that are necessary to answer important, outstanding questions.

Acknowledgments

The research reported in this publication was supported by the National Institute on Aging (awards F31AG069444 and T32AG066598 [KH]; R01AG062503 and R21AG066771 [TG]; R01MH116884 [IB]). We thank Drs. Deborah Drabick, David Smith, Chiu Tan, and Lauren Alloy for their valuable contributions to this project throughout their role on KH’s dissertation committee. We would also like to acknowledge Emma Pinsky, Sophia Holmqvist, and Marina Kaplan for assistance with data collection and entry. Please note that KH's current affiliation and source of her T32 funding is the Department of Medicine, Division of General Internal Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, United States.

Authors' Contributions

KH, TG, and IB contributed to conceptualization. IB, SX, LP, KH, and TG contributed to the methodology. KH and MM contributed to data collection and implementation. KH, SX, IB, and TG contributed to the formal analysis. KH and TG contributed to writing the original draft. KH, SX, MM, LP, IB, and TG contributed to writing, reviewing, and editing the manuscript. TG contributed to project administration and funding acquisition. All authors read and agreed to the version of the manuscript intended for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Comprehension of consent quiz.

[PNG File , 321 KB - humanfactors_v11i1e59974_app1.png]

Multimedia Appendix 2

Supplementary data and analyses.

[XLS File (Microsoft Excel File), 217 KB - humanfactors_v11i1e59974_app2.xls]

References

1. Branca C, Oddo S. Paving the way for new clinical trials for Alzheimer's disease. Biol Psychiatry 2017 Jan 15;81(2):88-89. [doi: 10.1016/j.biopsych.2016.10.016] [Medline: 27938878]
2. Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011 May;7(3):280-292. [doi: 10.1016/j.jalz.2011.03.003] [Medline: 21514248]

3. Sperling RA, Rentz DM, Johnson KA, Karlawish J, Donohue M, Salmon DP, et al. The A4 study: stopping AD before symptoms begin? Sci Transl Med 2014 Mar 19;6(228):228fs13 [FREE Full text] [doi: 10.1126/scitranslmed.3007941] [Medline: 24648338]
4. Petersen RC, Aisen PS, Andrews JS, Atri A, Matthews BR, Rentz DM, et al. Expectations and clinical meaningfulness of randomized controlled trials. Alzheimers Dement 2023 Jun 07;19(6):2730-2736 [FREE Full text] [doi: 10.1002/alz.12959] [Medline: 36748826]
5. Cummings J, Apostolova L, Rabinovici GD, Atri A, Aisen P, Greenberg S, et al. Lecanemab: appropriate use recommendations. J Prev Alzheimers Dis 2023;10(3):362-377 [FREE Full text] [doi: 10.14283/jpad.2023.30] [Medline: 37357276]
6. Giovannetti T, Hackett K, Tassoni MB, Mis R, Simone SM. Assessing and predicting everyday function. In: Brown GG, Crosson B, Haaland KY, King TZ, editors. APA Handbook of Neuropsychology. New York, NY: American Psychological Association; 2022:275-292.
7. Hackett K, Giovannetti T. Capturing cognitive aging in vivo: application of a neuropsychological framework for emerging digital tools. JMIR Aging 2022 Sep 07;5(3):e38130 [FREE Full text] [doi: 10.2196/38130] [Medline: 36069747]
8. Jutten RJ, Papp KV, Hendrix S, Ellison N, Langbaum JB, Donohue MC, et al. Why a clinical trial is as good as its outcome measure: a framework for the selection and use of cognitive outcome measures for clinical trials of Alzheimer's disease. Alzheimers Dement 2023 Feb 10;19(2):708-720 [FREE Full text] [doi: 10.1002/alz.12773] [Medline: 36086926]
9. Rabin LA, Paolillo E, Barr WB. Stability in test-usage practices of clinical neuropsychologists in the United States and Canada over a 10-year period: a follow-up survey of INS and NAN members. Arch Clin Neuropsychol 2016 May 16;31(3):206-230. [doi: 10.1093/arclin/acw007] [Medline: 26984127]
10. Teunissen CE, Verberk IM, Thijssen EH, Vermunt L, Hansson O, Zetterberg H, et al. Blood-based biomarkers for Alzheimer's disease: towards clinical implementation. Lancet Neurol 2022 Jan;21(1):66-77. [doi: 10.1016/S1474-4422(21)00361-6] [Medline: 34838239]
11. Brickman AM, Cabo R, Manly JJ. Ethical issues in cross-cultural neuropsychology. Appl Neuropsychol 2006 Jun;13(2):91-100. [doi: 10.1207/s15324826an1302_4] [Medline: 17009882]
12. Byrd DA, Rivera-Mindt MG. Neuropsychology's race problem does not begin or end with demographically adjusted norms. Nat Rev Neurol 2022 Jan 11;18(3):125-126. [doi: 10.1038/S41582-021-00607-4]
13. Chan KC, Barnes LL, Saykin AJ, Sano M, Au R, Craft S, et al. Racial - ethnic differences in baseline and longitudinal change in neuropsychological test scores in the NACC Uniform Data Set 3.0. Alzheimers Dement 2021 Dec 31;17(S6):e054653. [doi: 10.1002/alz.054653]
14. Fyffe DC, Mukherjee S, Barnes LL, Manly JJ, Bennett DA, Crane PK. Explaining differences in episodic memory performance among older African Americans and Whites: the roles of factors related to cognitive reserve and test bias. J Int Neuropsychol Soc 2011 Jul 06;17(4):625-638 [FREE Full text] [doi: 10.1017/S1355617711000476] [Medline: 23131601]
15. Heaton R, Miller S, Taylor MJ, Grant I. Revised comprehensive norms for an expanded halstead-reitan battery: demographically adjusted neuropsychological norms for African American and Caucasian adults. PAR, Inc. 2004. URL: https://www.parinc.com/products/RCNAAC [accessed 2024-04-29]
16. Manly JJ, Jacobs DM, Touradji P, Small SA, Stern Y. Reading level attenuates differences in neuropsychological test performance between African American and White elders. J Int Neuropsychol Soc 2002 Mar 16;8(3):341-348. [doi: 10.1017/s1355617702813157] [Medline: 11939693]
17. Manly JJ, Byrd DA, Touradji P, Stern Y. Acculturation, reading level, and neuropsychological test performance among African American elders. Appl Neuropsychol 2004 Mar;11(1):37-46. [doi: 10.1207/s15324826an1101_5] [Medline: 15471745]
18. Postal K. President's annual state of the academy report. Clin Neuropsychol 2018 Jan 22;32(1):1-9. [doi: 10.1080/13854046.2017.1406993] [Medline: 29271311]
19. Sachs BC, Steenland K, Zhao L, Hughes TM, Weintraub S, Dodge HH, et al. Expanded demographic norms for version 3 of the Alzheimer Disease Centers' neuropsychological test battery in the uniform data set. Alzheimer Dis Assoc Disord 2020;34(3):191-197 [FREE Full text] [doi: 10.1097/WAD.0000000000000388] [Medline: 32483017]
20. Tan YW, Burgess GH, Green RJ. The effects of acculturation on neuropsychological test performance: a systematic literature review. Clin Neuropsychol 2021 Apr 30;35(3):541-571. [doi: 10.1080/13854046.2020.1714740] [Medline: 31996089]
21. Benge JF, Aguirre A, Scullin MK, Kiselica A, Hilsabeck RC, Paydarfar D, et al. Digital methods for performing daily tasks among older adults: an initial report of frequency of use and perceived utility. Exp Aging Res 2024 Feb 05;50(2):133-154. [doi: 10.1080/0361073X.2023.2172950] [Medline: 36739553]
22. Graves LV, Beacham D, Sawitzky N, Lozano J, Weberg M, Reyes K, et al. Informant sex/gender moderates associations between reported functioning and memory performance in Mexican American adults. Clin Neuropsychol 2024 Apr 12;38(3):715-737. [doi: 10.1080/13854046.2023.2246669] [Medline: 37571992]
23. Graves LV, Hamill S, Larry M, Williams D. Informant characteristics influence reports of participant functioning and their associations with neuropsychological performance in non-Hispanic black adults. Arch Clin Neuropsychol 2023 Oct 20;38(7):1054-1067 [FREE Full text] [doi: 10.1093/arclin/acad022] [Medline: 36881929]

24. Hackett K, Mis R, Drabick DA, Giovannetti T. Informant reporting in mild cognitive impairment: sources of discrepancy on the functional activities questionnaire. J Int Neuropsychol Soc 2020 May 22;26(5):503-514 [FREE Full text] [doi: 10.1017/S1355617719001449] [Medline: 31964443]
25. Siemers E, Holdridge KC, Sundell KL, Liu-Seifert H. Function and clinical meaningfulness of treatments for mild Alzheimer's disease. Alzheimers Dement (Amst) 2016 Mar 09;2(1):105-112 [FREE Full text] [doi: 10.1016/j.dadm.2016.02.006] [Medline: 27239541]
26. Wu CY, Dodge HH, Gothard S, Mattek N, Wright K, Barnes LL, et al. Unobtrusive sensing technology detects ecologically valid spatiotemporal patterns of daily routines distinctive to persons with mild cognitive impairment. J Gerontol A Biol Sci Med Sci 2022 Oct 06;77(10):2077-2084 [FREE Full text] [doi: 10.1093/gerona/glab293] [Medline: 34608939]
27. Gomar JJ, Bobes-Bascaran MT, Conejero-Goldberg C, Davies P, Goldberg TE, Alzheimer's Disease Neuroimaging Initiative. Utility of combinations of biomarkers, cognitive markers, and risk factors to predict conversion from mild cognitive impairment to Alzheimer disease in patients in the Alzheimer's disease neuroimaging initiative. Arch Gen Psychiatry 2011 Sep 05;68(9):961-969. [doi: 10.1001/archgenpsychiatry.2011.96] [Medline: 21893661]
28. Mungas D, Beckett L, Harvey D, Farias ST, Reed B, Carmichael O, et al. Heterogeneity of cognitive trajectories in diverse older persons. Psychol Aging 2010 Sep;25(3):606-619 [FREE Full text] [doi: 10.1037/a0019502] [Medline: 20677882]
29. Wang X, Ye T, Zhou W, Zhang J, Alzheimer's Disease Neuroimaging Initiative. Uncovering heterogeneous cognitive trajectories in mild cognitive impairment: a data-driven approach. Alzheimers Res Ther 2023 Mar 20;15(1):57 [FREE Full text] [doi: 10.1186/s13195-023-01205-w] [Medline: 36941651]
30. Haaksma ML, Rizzuto D, Leoutsakos JM, Marengoni A, Tan EC, Olde Rikkert MG, et al. Predicting cognitive and functional trajectories in people with late-onset dementia: 2 population-based studies. J Am Med Dir Assoc 2019 Nov;20(11):1444-1450 [FREE Full text] [doi: 10.1016/j.jamda.2019.03.025] [Medline: 31109912]
31. Perneczky R, Jessen F, Grimmer T, Levin J, Flöel A, Peters O, et al. Anti-amyloid antibody therapies in Alzheimer's disease. Brain 2023 Mar 01;146(3):842-849. [doi: 10.1093/brain/awad005] [Medline: 36655336]
32. Federman AD, Becker JH, Mindt MR, Cho D, Curtis L, Wisnivesky J. Rates of undiagnosed cognitive impairment and performance on the Montreal Cognitive Assessment among older adults in primary care. J Gen Intern Med 2023 Aug 22;38(11):2511-2518. [doi: 10.1007/s11606-023-08102-w] [Medline: 36814049]
33. Liu Y, Jun H, Becker A, Wallick C, Mattke S. Detection rates of mild cognitive impairment in primary care for the United States Medicare population. J Prev Alzheimers Dis 2024;11(1):7-12. [doi: 10.14283/jpad.2023.131] [Medline: 38230712]
34. Mattke S, Jun H, Chen E, Liu Y, Becker A, Wallick C. Expected and diagnosed rates of mild cognitive impairment and dementia in the U.S. Medicare population: observational analysis. Alzheimers Res Ther 2023 Jul 22;15(1):128 [FREE Full text] [doi: 10.1186/s13195-023-01272-z] [Medline: 37481563]
35. DiBenedetti DB, Slota C, Wronski SL, Vradenburg G, Comer M, Callahan LF, et al. Assessing what matters most to patients with or at risk for Alzheimer's and care partners: a qualitative study evaluating symptoms, impacts, and outcomes. Alzheimers Res Ther 2020 Jul 30;12(1):90 [FREE Full text] [doi: 10.1186/s13195-020-00659-6] [Medline: 32731886]
36. Andrews JS, Desai U, Kirson NY, Zichlin ML, Ball DE, Matthews BR. Disease severity and minimal clinically important differences in clinical outcome assessments for Alzheimer's disease clinical trials. Alzheimers Dement (N Y) 2019 Aug 02;5(1):354-363 [FREE Full text] [doi: 10.1016/j.trci.2019.06.005] [Medline: 31417957]
37. Lansdall CJ, McDougall F, Butler LM, Delmar P, Pross N, Qin S, et al. Establishing clinically meaningful change on outcome assessments frequently used in trials of mild cognitive impairment due to Alzheimer's disease. J Prev Alzheimers Dis 2023;10(1):9-18. [doi: 10.14283/jpad.2022.102] [Medline: 36641605]
38. Evans S, McRae-McKee K, Wong MM, Hadjichrysanthou C, De Wolf F, Anderson R. The importance of endpoint selection: how effective does a drug need to be for success in a clinical trial of a possible Alzheimer's disease treatment? Eur J Epidemiol 2018 Jul 23;33(7):635-644 [FREE Full text] [doi: 10.1007/s10654-018-0381-0] [Medline: 29572656]
39. Posner H, Curiel R, Edgar C, Hendrix S, Liu E, Loewenstein DA, et al. Outcomes assessment in clinical trials of Alzheimer's disease and its precursors: readying for short-term and long-term clinical trial needs. Innov Clin Neurosci 2017;14(1-2):22-29 [FREE Full text] [Medline: 28386518]
40. Tochel C, Smith M, Baldwin H, Gustavsson A, Ly A, Bexelius C, ROADMAP consortium. What outcomes are important to patients with mild cognitive impairment or Alzheimer's disease, their caregivers, and health-care professionals? A systematic review. Alzheimers Dement (Amst) 2019 Dec 09;11(1):231-247 [FREE Full text] [doi: 10.1016/j.dadm.2018.12.003] [Medline: 30906845]
41. Cummings J, Lee G, Nahed P, Kambar ME, Zhong K, Fonseca J, et al. Alzheimer's disease drug development pipeline: 2022. Alzheimers Dement (N Y) 2022 May 04;8(1):e12295 [FREE Full text] [doi: 10.1002/trc2.12295] [Medline: 35516416]
42. Kourtis LC, Regele OB, Wright JM, Jones GB. Digital biomarkers for Alzheimer's disease: the mobile/ wearable devices opportunity. NPJ Digit Med 2019;2 [FREE Full text] [doi: 10.1038/s41746-019-0084-2] [Medline: 31119198]
43. Beattie ZT, Gothard S, Barnes LL, Calloway S, Gonzales MM, Jones BD, et al. The collaboration and harmonization of digital biomarkers for Alzheimer's research. Alzheimers Dement 2022 Dec 20;18(S2):e067984. [doi: 10.1002/alz.067984]
44. Papp KV, Samaroo A, Chou HC, Buckley R, Schneider OR, Hsieh S, et al. Unsupervised mobile cognitive testing for use in preclinical Alzheimer's disease. Alzheimers Dement (Amst) 2021 Sep 30;13(1):e12243 [FREE Full text] [doi: 10.1002/dad2.12243] [Medline: 34621977]

45. Weizenbaum EL, Soberanes D, Hsieh S, Molinare CP, Buckley RF, Betensky RA, et al. Capturing learning curves with the multiday Boston Remote Assessment of Neurocognitive Health (BRANCH): feasibility, reliability, and validity. Neuropsychology 2024 Feb;38(2):198-210. [doi: 10.1037/neu0000933] [Medline: 37971862]
46. Popp Z, Low S, Igwe A, Rahman MS, Kim M, Khan R, et al. Shifting from active to passive monitoring of Alzheimer disease: the state of the research. J Am Heart Assoc 2024 Jan 16;13(2):e031247 [FREE Full text] [doi: 10.1161/JAHA.123.031247] [Medline: 38226518]
47. Dodge HH, Zhu J, Mattek NC, Austin D, Kornfeld J, Kaye JA. Use of high-frequency in-home monitoring data may reduce sample sizes needed in clinical trials. PLoS One 2015 Sep 17;10(9):e0138095 [FREE Full text] [doi: 10.1371/journal.pone.0138095] [Medline: 26379170]
48. Sliwinski MJ, Mogle JA, Hyun J, Munoz E, Smyth JM, Lipton RB. Reliability and validity of ambulatory cognitive assessments. Assessment 2018 Jan;25(1):14-30 [FREE Full text] [doi: 10.1177/1073191116643164] [Medline: 27084835]
49. Weizenbaum E, Torous J, Fulford D. Cognition in context: understanding the everyday predictors of cognitive performance in a new era of measurement. JMIR Mhealth Uhealth 2020 Jul 23;8(7):e14328 [FREE Full text] [doi: 10.2196/14328] [Medline: 32706680]
50. Nicosia J, Aschenbrenner AJ, Balota DA, Sliwinski MJ, Tahan M, Adams S, et al. Reliability, validity, and feasibility of a smartphone‐based cognitive assessment for preclinical Alzheimer disease. Alzheimers Dement 2023 Jun 16;19(S4):e063363. [doi: 10.1002/alz.063363]
51. Anderson M, Perrin A. Tech adoption climbs among older adults. Pew Research Center. 2017. URL: https://www.pewresearch.org/internet/2017/05/17/tech-adoption-climbs-among-older-adults/ [accessed 2024-04-29]
52. Mobile fact sheet. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2024-04-29]
53. Holm E, Jacobsen KK, de Lony TB, Lembeck M, Pedersen H, Andersson C, et al. Frequency of missed or delayed diagnosis in dementia is associated with neighborhood socioeconomic status. Alzheimers Dement (N Y) 2022 Mar 24;8(1):e12271 [FREE Full text] [doi: 10.1002/trc2.12271] [Medline: 35356741]
54. Ferri CP, Jacob KS. Dementia in low-income and middle-income countries: different realities mandate tailored solutions. PLoS Med 2017 Mar 28;14(3):e1002271 [FREE Full text] [doi: 10.1371/journal.pmed.1002271] [Medline: 28350797]
55. Connolly A, Gaehl E, Martin H, Morris J, Purandare N. Underdiagnosis of dementia in primary care: variations in the observed prevalence and comparisons to the expected prevalence. Aging Ment Health 2011 Nov;15(8):978-984. [doi: 10.1080/13607863.2011.596805] [Medline: 21777080]
56. Lin PJ, Daly AT, Olchanski N, Cohen JT, Neumann PJ, Faul JD, et al. Dementia diagnosis disparities by race and ethnicity. Med Care 2021 Aug 01;59(8):679-686 [FREE Full text] [doi: 10.1097/MLR.0000000000001577] [Medline: 34091580]
57. Gianattasio KZ, Prather C, Glymour MM, Ciarleglio A, Power MC. Racial disparities and temporal trends in dementia misdiagnosis risk in the United States. Alzheimers Dement (N Y) 2019 Jan;5(1):891-898 [FREE Full text] [doi: 10.1016/j.trci.2019.11.008] [Medline: 31890853]
58. Insel TR. Digital phenotyping: technology for a new science of behavior. JAMA 2017 Oct 03;318(13):1215-1216. [doi: 10.1001/jama.2017.11295] [Medline: 28973224]
59. Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. Nat Biotechnol 2015 May;33(5):462-463. [doi: 10.1038/nbt.3223] [Medline: 25965751]
60. Torous J, Kiang MV, Lorme J, Onnela JP. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. JMIR Ment Health 2016 May 05;3(2):e16 [FREE Full text] [doi: 10.2196/mental.5165] [Medline: 27150677]
61. Barnett I, Torous J, Staples P, Sandoval L, Keshavan M, Onnela JP. Relapse prediction in schizophrenia through digital phenotyping: a pilot study. Neuropsychopharmacology 2018 Jul;43(8):1660-1666 [FREE Full text] [doi: 10.1038/s41386-018-0030-z] [Medline: 29511333]
62. Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT. Next-generation psychiatric assessment: using smartphone sensors to monitor behavior and mental health. Psychiatr Rehabil J 2015 Sep;38(3):218-226 [FREE Full text] [doi: 10.1037/prj0000130] [Medline: 25844912]
63. Cohen A, Naslund JA, Chang S, Nagendra S, Bhan A, Rozatkar A, et al. Relapse prediction in schizophrenia with smartphone digital phenotyping during COVID-19: a prospective, three-site, two-country, longitudinal study. Schizophrenia (Heidelb) 2023 Jan 27;9(1):6 [FREE Full text] [doi: 10.1038/s41537-023-00332-5] [Medline: 36707524]
64. Kelkar RS, Currey D, Nagendra S, Mehta UM, Sreeraj VS, Torous J, et al. Utility of smartphone-based digital phenotyping biomarkers in assessing treatment response to transcranial magnetic stimulation in depression: proof-of-concept study. JMIR Form Res 2023 Sep 01;7:e40197 [FREE Full text] [doi: 10.2196/40197] [Medline: 37656496]
65. Lane E, D'Arcey J, Kidd S, Onyeaka H, Alon N, Joshi D, et al. Digital phenotyping in adults with schizophrenia: a narrative review. Curr Psychiatry Rep 2023 Oct 20;25(11):699-706. [doi: 10.1007/S11920-023-01467-Z]
66. Leaning IE, Ikani N, Savage HS, Leow A, Beckmann C, Ruhé HG, et al. From smartphone data to clinically relevant predictions: a systematic review of digital phenotyping methods in depression. Neurosci Biobehav Rev 2024 Mar;158:105541 [FREE Full text] [doi: 10.1016/j.neubiorev.2024.105541] [Medline: 38215802]

67. Ren B, Xia CH, Gehrman P, Barnett I, Satterthwaite T. Measuring daily activity rhythms in young adults at risk of affective instability using passively collected smartphone data: observational study. JMIR Form Res 2022 Sep 14;6(9):e33890 [FREE Full text] [doi: 10.2196/33890] [Medline: 36103225]
68. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. J Med Internet Res 2015 Jul 15;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
69. Vesel C, Rashidisabet H, Zulueta J, Stange JP, Duffecy J, Hussain F, et al. Effects of mood and aging on keystroke dynamics metadata and their diurnal patterns in a large open-science sample: a BiAffect iOS study. J Am Med Inform Assoc 2020 Jul 01;27(7):1007-1018 [FREE Full text] [doi: 10.1093/jamia/ocaa057] [Medline: 32467973]
70. Zulueta J, Piscitello A, Rasic M, Easter R, Babu P, Langenecker SA, et al. Predicting mood disturbance severity with mobile phone keystroke metadata: a BiAffect digital phenotyping study. J Med Internet Res 2018 Jul 20;20(7):e241 [FREE Full text] [doi: 10.2196/jmir.9775] [Medline: 30030209]
71. Chen MH, Leow A, Ross MK, DeLuca J, Chiaravalloti N, Costa SL, et al. Associations between smartphone keystroke dynamics and cognition in MS. Digit Health 2022 Dec 05;8:20552076221143234 [FREE Full text] [doi: 10.1177/20552076221143234] [Medline: 36506490]
72. Chen R, Jankovic F, Marinsek N. Developing measures of cognitive impairment in the real world from consumer-grade multimodal sensor streams. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019 Presented at: KDD '19; August 4-8, 2019; Anchorage, AK p. 2145-2155 URL: https://dl.acm.org/doi/10.1145/3292500.3330690 [doi: 10.1145/3292500.3330690]
73. Dagum P. Digital biomarkers of cognitive function. NPJ Digit Med 2018 Mar 28;1(1):10 [FREE Full text] [doi: 10.1038/s41746-018-0018-4] [Medline: 31304295]
74. Gordon M, Gatys L, Guestrin C, Bigham JP, Trister A, Patel K. App usage predicts cognitive ability in older adults. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19; May 4-9, 2019; Glasgow, Scotland p. 168 URL: https://dl.acm.org/doi/fullHtml/10.1145/3290605.3300398 [doi: 10.1145/3290605.3300398]
75. Ning E, Cladek AT, Ross MK, Kabir S, Barve A, Kennelly E, et al. Smartphone-derived virtual keyboard dynamics coupled with accelerometer data as a window into understanding brain health: smartphone keyboard and accelerometer as window into brain health. Proc SIGCHI Conf Hum Factor Comput Syst 2023 Apr;2023:1-15 [FREE Full text] [doi: 10.1145/3544548.3580906] [Medline: 38115842]
76. Ross MK, Demos AP, Zulueta J, Piscitello A, Langenecker SA, McInnis M, et al. Naturalistic smartphone keyboard typing reflects processing speed and executive function. Brain Behav 2021 Nov 06;11(11):e2363 [FREE Full text] [doi: 10.1002/brb3.2363] [Medline: 34612605]
77. Paolillo EW, Casaletto KB, Clark AL, Taylor JC, Heuer HW, Wise AB, ALLFTD Consortium. Examining associations between smartphone use and clinical severity in frontotemporal dementia: proof-of-concept study. JMIR Aging 2024 Jun 26;7:e52831 [FREE Full text] [doi: 10.2196/52831] [Medline: 38922667]
78. Bartlett Ellis R, Wright J, Miller LS, Jake-Schoffman D, Hekler EB, Goldstein CM, et al. Lessons learned: beta-testing the digital health checklist for researchers prompts a call to action by behavioral scientists. J Med Internet Res 2021 Dec 22;23(12):e25414 [FREE Full text] [doi: 10.2196/25414] [Medline: 34941548]
79. Nebeker C, Bartlett Ellis RJ, Torous J. Development of a decision-making checklist tool to support technology selection in digital health research. Transl Behav Med 2020 Oct 08;10(4):1004-1015 [FREE Full text] [doi: 10.1093/tbm/ibz074] [Medline: 31120511]
80. Huckvale K, Venkatesh S, Christensen H. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. NPJ Digit Med 2019;2:88 [FREE Full text] [doi: 10.1038/s41746-019-0166-1] [Medline: 31508498]
81. Schmidt S, D'Alfonso S. Clinician perspectives on how digital phenotyping can inform client treatment. Acta Psychol (Amst) 2023 May;235:103886 [FREE Full text] [doi: 10.1016/j.actpsy.2023.103886] [Medline: 36921359]
82. Vaidyam A, Halamka J, Torous J. Actionable digital phenotyping: a framework for the delivery of just-in-time and longitudinal interventions in clinical healthcare. Mhealth 2019 Aug;5:25 [FREE Full text] [doi: 10.21037/mhealth.2019.07.04] [Medline: 31559270]
83. Torous J, Wisniewski H, Bird B, Carpenter E, David G, Elejalde E, et al. Creating a digital health smartphone app and digital phenotyping platform for mental health and diverse healthcare needs: an interdisciplinary and collaborative approach. J Technol Behav Sci 2019 Apr 27;4(2):73-85 [FREE Full text] [doi: 10.1007/S41347-019-00095-W]
84. Vaidyam A, Halamka J, Torous J. Enabling research and clinical use of patient-generated health data (the mindLAMP Platform): digital phenotyping study. JMIR Mhealth Uhealth 2022 Jan 07;10(1):e30557 [FREE Full text] [doi: 10.2196/30557] [Medline: 34994710]
85. Jak AJ, Bondi MW, Delano-Wood L, Wierenga C, Corey-Bloom J, Salmon DP, et al. Quantification of five neuropsychological approaches to defining mild cognitive impairment. Am J Geriatr Psychiatry 2009 May;17(5):368-375. [doi: 10.1097/jgp.0b013e31819431d5]
86. Melcher J, Hays R, Torous J. Digital phenotyping for mental health of college students: a clinical review. Evid Based Ment Health 2020 Nov;23(4):161-166 [FREE Full text] [doi: 10.1136/ebmental-2020-300180] [Medline: 32998937]

87. Schretlen DJ, Winicki JM, Meyer SM, Testa SM, Pearlson GD, Gordon B. Development, psychometric properties, and validity of the hopkins adult reading test (HART). Clin Neuropsychol 2009 Aug;23(6):926-943. [doi: 10.1080/13854040802603684] [Medline: 19191072]
88. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
89. Reitan RM. Validity of the trail making test as an indicator of organic brain damage. Percept Mot Skills 1958 Dec 01;8(3):271-276. [doi: 10.2466/pms.1958.8.3.271]
90. Wechsler D. Wechsler Memory Scale-Revised: Manual. San Antonio, TX: Psychological Corporation; 1987.
91. Salthouse TA. Mediation of adult age differences in cognition by reductions in working memory and speed of processing. Psychol Sci 1991 May 01;2(3):179-183. [doi: 10.1111/j.1467-9280.1991.tb00127.x]
92. Brandt J, Benedict RH. Hopkins Verbal Learning Test-- Revised Professional Manual. New York, NY: Psychological Assessment Resources; 2001.
93. Benedict RH, Schretlen D, Groninger L, Dobraski M, Shpritz B. Revision of the brief visuospatial memory test: studies of normal performance, reliability, and validity. Psychol Assess 1996 Jun;8(2):145-153. [doi: 10.1037/1040-3590.8.2.145]
94. Schretlen DJ, Testa SM, Pearlson GD. Calibrated neuropsychological normative system professional manual. PAA. 2022. URL: https://paa.com.au/product/cnns/ [accessed 2024-04-29]
95. Goodglass H, Kaplan E. Boston diagnostic aphasia examination. In: Kreutzer JS, DeLuca J, Caplan B, editors. Encyclopedia of Clinical Neuropsychology. Cham, Switzerland: Springer; 2022:428-430.
96. Tomaszewski Farias S, Mungas D, Harvey DJ, Simmons A, Reed BR, Decarli C. The measurement of everyday cognition: development and validation of a short form of the Everyday Cognition scales. Alzheimers Dement 2011 Nov;7(6):593-601 [FREE Full text] [doi: 10.1016/j.jalz.2011.02.007] [Medline: 22055976]
97. Pfeffer RI, Kurosaki TT, Harrah CH, Chance JM, Filos S. Measurement of functional activities in older adults in the community. J Gerontol 1982 May 01;37(3):323-329. [doi: 10.1093/geronj/37.3.323] [Medline: 7069156]
98. Yesavage JA. Geriatric depression scale. Psychopharmacol Bull 1988;24(4):709-711. [doi: 10.1037/T00930-000]
99. Bohannon RW. Reference values for the timed up and go test: a descriptive meta-analysis. J Geriatr Phys Ther 2006;29(2):64-68. [doi: 10.1519/00139143-200608000-00004] [Medline: 16914068]
100. Podsiadlo D, Richardson S. The timed "Up and Go": a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc 1991 Feb 27;39(2):142-148. [doi: 10.1111/j.1532-5415.1991.tb01616.x] [Medline: 1991946]
101. Baker PS, Bodner EV, Allman RM. Measuring life-space mobility in community-dwelling older adults. J Am Geriatr Soc 2003 Nov 24;51(11):1610-1614. [doi: 10.1046/j.1532-5415.2003.51512.x] [Medline: 14687391]
102. Peel C, Sawyer Baker P, Roth DL, Brown CJ, Brodner EV, Allman RM. Assessing mobility in older adults: the UAB study of aging life-space assessment. Phys Ther 2005 Oct;85(10):1008-1119. [Medline: 16180950]
103. Crowe M, Andel R, Wadley VG, Okonkwo OC, Sawyer P, Allman RM. Life-space and cognitive decline in a community-based sample of African American and Caucasian older adults. J Gerontol A Biol Sci Med Sci 2008 Nov 01;63(11):1241-1245 [FREE Full text] [doi: 10.1093/gerona/63.11.1241] [Medline: 19038840]
104. James BD, Boyle PA, Buchman AS, Barnes LL, Bennett DA. Life space and risk of Alzheimer disease, mild cognitive impairment, and cognitive decline in old age. Am J Geriatr Psychiatry 2011 Nov;19(11):961-969. [doi: 10.1097/jgp.0b013e318211c219]
105. Berry HL, Rodgers B, Dear KB. Preliminary development and validation of an Australian community participation questionnaire: types of participation and associations with distress in a coastal community. Soc Sci Med 2007 Apr;64(8):1719-1737. [doi: 10.1016/j.socscimed.2006.12.009] [Medline: 17241727]
106. Brett L, Georgiou A, Jorgensen M, Siette J, Scott G, Gow E, et al. Ageing well: evaluation of social participation and quality of life tools to enhance community aged care (study protocol). BMC Geriatr 2019 Mar 12;19(1):78 [FREE Full text] [doi: 10.1186/s12877-019-1094-2] [Medline: 30871472]
107. van Deursen AJ, Bolle CL, Hegner SM, Kommers PA. Modeling habitual and addictive smartphone behavior: the role of smartphone usage types, emotional intelligence, social stress, self-regulation, age, and gender. Comput Human Behav 2015 Apr;45:411-420. [doi: 10.1016/j.chb.2014.12.039]
108. Roque NA, Boot WR. A new tool for assessing mobile device proficiency in older adults: the mobile device proficiency questionnaire. J Appl Gerontol 2018 Feb 11;37(2):131-156 [FREE Full text] [doi: 10.1177/0733464816642582] [Medline: 27255686]
109. Barnett I, Onnela JP. Inferring mobility measures from GPS traces with missing data. Biostatistics 2020 Apr 01;21(2):e98-112 [FREE Full text] [doi: 10.1093/biostatistics/kxy059] [Medline: 30371736]
110. Ian B. Ian Barnett: profile. GitHub. URL: https://github.com/owianbarnett [accessed 2024-04-29]
111. Onnela JP. Opportunities and challenges in the collection and analysis of digital phenotyping data. Neuropsychopharmacol 2020 Jul 17;46(1):45-54. [doi: 10.1038/S41386-020-0771-3]
112. Torgerson WS. Multidimensional scaling: I. Theory and method. Psychometrika 1952 Dec;17(4):401-419. [doi: 10.1007/bf02288916]

113. Almeida OP, Almeida SA. Short versions of the geriatric depression scale: a study of their validity for the diagnosis of a major depressive episode according to ICD-10 and DSM-IV. Int J Geriatr Psychiatry 1999 Oct;14(10):858-865. [doi: 10.1002/(sici)1099-1166(199910)14:10<858::aid-gps35>3.0.co;2-8] [Medline: 10521885]
114. Kiang MV, Chen JT, Krieger N, Buckee CO, Alexander MJ, Baker JT, et al. Sociodemographic characteristics of missing data in digital phenotyping. Sci Rep 2021 Jul 29;11(1):15408 [FREE Full text] [doi: 10.1038/s41598-021-94516-7] [Medline: 34326370]
115. Paolillo EW, Casaletto KB, Clark AL, Taylor JC, Heuer HW, Wise AB, et al. Passively monitored smartphone battery percentage relates to disease severity in frontotemporal dementia: a proof‐of‐concept study. Alzheimers Dement 2023 Dec 25;19(S18):e079178. [doi: 10.1002/alz.079178]
116. Eliacin J, Polsinelli AJ, Epperson F, Gao S, Van Heiden S, Westmoreland G, et al. Barriers and facilitators to participating in Alzheimer's disease biomarker research in black and white older adults. Alzheimers Dement (N Y) 2023 Jun 05;9(2):e12399 [FREE Full text] [doi: 10.1002/trc2.12399] [Medline: 37287470]
117. Indorewalla KK, O'Connor MK, Budson AE, Guess DiTerlizzi C, Jackson J. Modifiable barriers for recruitment and retention of older adults participants from underrepresented minorities in Alzheimer's disease research. J Alzheimers Dis 2021;80(3):927-940. [doi: 10.3233/JAD-201081] [Medline: 33612540]
118. Minkler M, Chang C. Community-based participatory research: a promising approach for studying and addressing immigrant health. In: Schenker MB, Castañeda X, Rodriguez-Lainz A, editors. Migration and Health: A Research Methods Handbook. Los Angeles, CA: University of California Press; 2014:361-376.
119. Neffa-Creech D, Aggarwal R, Stowell C, Menon U, George S, Plant A, et al. Understanding barriers and facilitators to signing up for a mobile-responsive registry to recruit healthy volunteers and members of underrepresented communities for Alzheimer's disease prevention studies. J Prev Alzheimers Dis 2023;10(4):865-874. [doi: 10.14283/jpad.2023.67] [Medline: 37874109]
120. Bangen KJ, Calcetas AT, Thomas KR, Wierenga C, Smith CN, Bordyug M, et al. Greater accelerometer-measured physical activity is associated with better cognition and cerebrovascular health in older adults. J Int Neuropsychol Soc 2023 Nov 15;29(9):859-869 [FREE Full text] [doi: 10.1017/S1355617723000140] [Medline: 36789631]
121. Buchman AS, Boyle PA, Wilson RS, Tang Y, Bennett DA. Frailty is associated with incident Alzheimer's disease and cognitive decline in the elderly. Psychosom Med 2007 Jun;69(5):483-489. [doi: 10.1097/psy.0b013e318068de1d] [Medline: 17556640]
122. Buchman AS, Boyle PA, Yu L, Shah RC, Wilson RS, Bennett DA. Total daily physical activity and the risk of AD and cognitive decline in older adults. Neurology 2012 Apr 24;78(17):1323-1329. [doi: 10.1212/wnl.0b013e3182535d35]
123. Casaletto KB, Rentería MA, Pa J, Tom SE, Harrati A, Armstrong NM, et al. Late-life physical and cognitive activities independently contribute to brain and cognitive resilience. J Alzheimers Dis 2020;74(1):363-376 [FREE Full text] [doi: 10.3233/JAD-191114] [Medline: 32039854]
124. Gallaway PJ, Miyake H, Buchowski MS, Shimada M, Yoshitake Y, Kim A, et al. Physical activity: a viable way to reduce the risks of mild cognitive impairment, Alzheimer’s disease, and vascular dementia in older adults. Brain Sci 2017 Feb 20;7(2):22 [FREE Full text] [doi: 10.3390/brainsci7020022] [Medline: 28230730]
125. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet commission. Lancet 2020 Aug;396(10248):413-446. [doi: 10.1016/s0140-6736(20)30367-6]
126. Petersen J, Austin D, Mattek N, Kaye J. Time out-of-home and cognitive, physical, and emotional wellbeing of older adults: a longitudinal mixed effects model. PLoS One 2015 Oct 5;10(10):e0139643 [FREE Full text] [doi: 10.1371/journal.pone.0139643] [Medline: 26437228]
127. Scarmeas N, Levy G, Tang MX, Manly J, Stern Y. Influence of leisure activity on the incidence of Alzheimer's disease. Neurology 2001 Dec 26;57(12):2236-2242 [FREE Full text] [doi: 10.1212/wnl.57.12.2236] [Medline: 11756603]
128. Wang L, Larson EB, Bowen JD, van Belle G. Performance-based physical function and future dementia in older people. Arch Intern Med 2006 May 22;166(10):1115-1120. [doi: 10.1001/archinte.166.10.1115] [Medline: 16717174]
129. Rykov Y, Thach TQ, Bojic I, Christopoulos G, Car J. Digital biomarkers for depression screening with wearable devices: cross-sectional study with machine learning modeling. JMIR Mhealth Uhealth 2021 Oct 25;9(10):e24872 [FREE Full text] [doi: 10.2196/24872] [Medline: 34694233]
130. Bergua V, Bouisson J, Dartigues JF, Swendsen J, Fabrigoule C, Pérès K, et al. Restriction in instrumental activities of daily living in older persons: association with preferences for routines and psychological vulnerability. Int J Aging Hum Dev 2013 Nov 26;77(4):309-329. [doi: 10.2190/ag.77.4.c]
131. Insel KC, Einstein GO, Morrow DG, Koerner KM, Hepworth JT. Multifaceted prospective memory intervention to improve medication adherence. J Am Geriatr Soc 2016 Mar;64(3):561-568 [FREE Full text] [doi: 10.1111/jgs.14032] [Medline: 27000329]
132. Suchy Y, Gereau Mora M, DesRuisseaux LA, Brothers SL. It's complicated: executive functioning moderates impacts of daily busyness on everyday functioning in community-dwelling older adults. J Int Neuropsychol Soc 2023 Nov 14;29(9):850-858. [doi: 10.1017/S1355617723000048] [Medline: 37057862]

133. Tomaszewski Farias S, Schmitter-Edgecombe M, Weakley A, Harvey D, Denny KG, Barba C, et al. Compensation strategies in older adults: association with cognition and everyday function. Am J Alzheimers Dis Other Demen 2018 May 23;33(3):184-191 [FREE Full text] [doi: 10.1177/1533317517753361] [Medline: 29357670]
134. Tassoni MB, Drabick DA, Giovannetti T. The frequency of self-reported memory failures is influenced by everyday context across the lifespan: implications for neuropsychology research and clinical practice. Clin Neuropsychol 2023 Aug 24;37(6):1115-1135. [doi: 10.1080/13854046.2022.2112297] [Medline: 36000515]
135. Bielak AA, Hultsch DF, Strauss E, Macdonald SW, Hunter MA. Intraindividual variability in reaction time predicts cognitive outcomes 5 years later. Neuropsychology 2010 Nov;24(6):731-741. [doi: 10.1037/a0019802] [Medline: 20853957]
136. Bielak AA, Hultsch DF, Strauss E, MacDonald SW, Hunter MA. Intraindividual variability is related to cognitive change in older adults: evidence for within-person coupling. Psychol Aging 2010 Sep;25(3):575-586. [doi: 10.1037/a0019503] [Medline: 20853965]
137. Dodge HH, Mattek NC, Austin D, Hayes TL, Kaye JA. In-home walking speeds and variability trajectories associated with mild cognitive impairment. Neurology 2012 Jun 12;78(24):1946-1952. [doi: 10.1212/wnl.0b013e318259e1de]
138. Hayes TL, Abendroth F, Adami A, Pavel M, Zitzelberger TA, Kaye JA. Unobtrusive assessment of activity patterns associated with mild cognitive impairment. Alzheimers Dement 2008 Nov;4(6):395-405 [FREE Full text] [doi: 10.1016/j.jalz.2008.07.004] [Medline: 19012864]
139. Kaye J, Mattek N, Dodge HH, Campbell I, Hayes T, Austin D, et al. Unobtrusive measurement of daily computer use to detect mild cognitive impairment. Alzheimers Dement 2014 Jan;10(1):10-17 [FREE Full text] [doi: 10.1016/j.jalz.2013.01.011] [Medline: 23688576]
140. Jutten RJ, Amariglio RE, Maruff P, Properzi MJ, Rentz DM, Johnson KA, et al. Increased intraindividual variability in reaction time performance is associated with emerging cognitive decline in cognitively unimpaired adults. Neuropsychology 2024 Feb;38(2):184-197. [doi: 10.1037/neu0000928] [Medline: 37971861]
141. Biddle KD, d'Oleire Uquillas F, Jacobs HI, Zide B, Kirn DR, Rentz DM, et al. Social engagement and amyloid-β-related cognitive decline in cognitively normal older adults. Am J Geriatr Psychiatry 2019 Nov;27(11):1247-1256 [FREE Full text] [doi: 10.1016/j.jagp.2019.05.005] [Medline: 31248770]
142. Fratiglioni L, Wang HX, Ericsson K, Maytan M, Winblad B. Influence of social network on occurrence of dementia: a community-based longitudinal study. Lancet 2000 Apr 15;355(9212):1315-1319. [doi: 10.1016/S0140-6736(00)02113-9] [Medline: 10776744]
143. Lee M, Whitsel E, Avery C, Hughes TM, Griswold ME, Sedaghat S, et al. Variation in population attributable fraction of dementia associated with potentially modifiable risk factors by race and ethnicity in the US. JAMA Netw Open 2022 Jul 01;5(7):e2219672 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.19672] [Medline: 35793088]
144. Penninkilampi R, Casey AN, Singh MF, Brodaty H. The association between social engagement, loneliness, and risk of dementia: a systematic review and meta-analysis. J Alzheimers Dis 2018;66(4):1619-1633. [doi: 10.3233/JAD-180439] [Medline: 30452410]
145. Ren Y, Savadlou A, Park S, Siska P, Epp JR, Sargin D. The impact of loneliness and social isolation on the development of cognitive decline and Alzheimer's disease. Front Neuroendocrinol 2023 Apr;69:101061 [FREE Full text] [doi: 10.1016/j.yfrne.2023.101061] [Medline: 36758770]
146. Mcalister C, Schmitter-Edgecombe M, Lamb R. Examination of variables that may affect the relationship between cognition and functional status in individuals with mild cognitive impairment: a meta-analysis. Arch Clin Neuropsychol 2016 Mar 06;31(2):123-147 [FREE Full text] [doi: 10.1093/arclin/acv089] [Medline: 26743326]
147. Royall DR, Lauterbach EC, Kaufer D, Malloy P, Coburn KL, Black KJ, Committee on Research of the American Neuropsychiatric Association. The cognitive correlates of functional status: a review from the Committee on Research of the American Neuropsychiatric Association. J Neuropsychiatry Clin Neurosci 2007 Jul;19(3):249-265. [doi: 10.1176/jnp.2007.19.3.249] [Medline: 17827410]
148. Sverdlov O, Curcic J, Hannesdottir K, Gou L, De Luca V, Ambrosetti F, et al. A study of novel exploratory tools, digital technologies, and central nervous system biomarkers to characterize unipolar depression. Front Psychiatry 2021 May 6;12:640741 [FREE Full text] [doi: 10.3389/fpsyt.2021.640741] [Medline: 34025472]
149. Torous J, Staples P, Barnett I, Sandoval LR, Keshavan M, Onnela JP. Characterizing the clinical relevance of digital phenotyping data quality with applications to a cohort with schizophrenia. NPJ Digit Med 2018;1:15 [FREE Full text] [doi: 10.1038/s41746-018-0022-8] [Medline: 31304300]
150. Murnane EL, Abdullah S, Matthews M, Kay M, Kientz JA, Choudhury T, et al. Mobile manifestations of alertness: connecting biological rhythms with patterns of smartphone app use. In: Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services. 2016 Sep Presented at: MobileHCI '16; September 6-9, 2016; Florence, Italy p. 465-477 URL: https://dl.acm.org/doi/10.1145/2935334.2935383 [doi: 10.1145/2935334.293538]
151. Quisel T, Kale DC, Foschini L. Intra-day activity better predicts chronic conditions. arXiv. Preprint posted online December 4, 2016 [FREE Full text]
152. Stachl C, Au Q, Schoedel R, Gosling SD, Harari GM, Buschek D, et al. Predicting personality from patterns of behavior collected with smartphones. Proc Natl Acad Sci U S A 2020 Jul 28;117(30):17680-17687 [FREE Full text] [doi: 10.1073/pnas.1920484117] [Medline: 32665436]

153. Barnett I, Torous J, Staples P, Keshavan M, Onnela JP. Beyond smartphones and sensors: choosing appropriate statistical methods for the analysis of longitudinal data. J Am Med Inform Assoc 2018 Dec 01;25(12):1669-1674 [FREE Full text] [doi: 10.1093/jamia/ocy121] [Medline: 30272176]
154. Zawada SJ, Ganjizadeh A, Hagen CE, Demaerschalk BM, Erickson BJ. Feasibility of observing cerebrovascular disease phenotypes with smartphone monitoring: study design considerations for real-world studies. Sensors 2024 Jun 02;24(11):3595. [doi: 10.3390/S24113595]
155. Johnson SA, Karas M, Burke KM, Straczkiewicz M, Scheier ZA, Clark AP, et al. Wearable device and smartphone data quantify ALS progression and may provide novel outcome measures. NPJ Digit Med 2023 Mar 06;6(1):34 [FREE Full text] [doi: 10.1038/s41746-023-00778-y] [Medline: 36879025]
156. Langholm C, Kowatsch T, Bucci S, Cipriani A, Torous J. Exploring the potential of apple SensorKit and digital phenotyping data as new digital biomarkers for mental health research. Digit Biomark 2023;7(1):104-114 [FREE Full text] [doi: 10.1159/000530698] [Medline: 37901364]
157. GBD 2019 Dementia Forecasting Collaborators. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. Lancet Public Health 2022 Feb;7(2):e105-e125 [FREE Full text] [doi: 10.1016/S2468-2667(21)00249-8] [Medline: 34998485]
158. Aschenbrenner AJ, Hassenstab J, Morris JC, Cruchaga C, Jackson JJ. Relationships between hourly cognitive variability and risk of Alzheimer's disease revealed with mixed-effects location scale models. Neuropsychology 2024 Jan;38(1):69-80. [doi: 10.1037/neu0000905] [Medline: 37079810]
159. Moore RC, Kaufmann CN, Rooney AS, Moore DJ, Eyler LT, Granholm E, et al. Feasibility and acceptability of ecological momentary assessment of daily functioning among older adults with HIV. Am J Geriatr Psychiatry 2017 Aug;25(8):829-840 [FREE Full text] [doi: 10.1016/j.jagp.2016.11.019] [Medline: 28065496]
160. Schmidt C, Collette F, Cajochen C, Peigneux P. A time to think: circadian rhythms in human cognition. Cogn Neuropsychol 2007 Oct;24(7):755-789. [doi: 10.1080/02643290701754158] [Medline: 18066734]
161. Wisniewski H, Henson P, Torous J. Using a smartphone app to identify clinically relevant behavior trends symptom report, cognition scores, and exercise levels: a case series. Front Psychiatry 2019;10:652 [FREE Full text] [doi: 10.3389/fpsyt.2019.00652] [Medline: 31607960]

Abbreviations

ACPQ: Australian Community Participation Questionnaire
AD/ADRD: Alzheimer disease and related dementias
AD: Alzheimer disease
FAQ: Functional Activities Questionnaire
GMM: Gaussian mixture modeling
HIPAA: Health Insurance Portability and Accountability Act
IIV: intraindividual variability
iSD: individual SD
LAMP: Learn, Assess, Manage, and Prevent
LSA: Life-Space Assessment
MCI: Mild Cognitive Impairment
MDS: multidimensional scaling
REDCap: Research Electronic Data Capture
TUG: Timed Up and Go Test
VIBE: Variability in Everyday Behavior

Edited by A Kushniruk, E Borycki; submitted 29.04.24; peer-reviewed by E Weizenbaum, I Leaning, S Zawada; comments to author 05.07.24; revised version received 29.08.24; accepted 30.09.24; published 22.11.24.

Please cite as:
Hackett K, Xu S, McKniff M, Paglia L, Barnett I, Giovannetti T
Mobility-Based Smartphone Digital Phenotypes for Unobtrusively Capturing Everyday Cognition, Mood, and Community Life-Space in Older Adults: Feasibility, Acceptability, and Preliminary Validity Study
JMIR Hum Factors 2024;11:e59974
URL: https://humanfactors.jmir.org/2024/1/e59974
doi:10.2196/59974
PMID:

Original Paper

Investigating the Acceptability of an Interactive Television Intervention Promoting Social Links Among Older Adults Living at Home and in Care Institutions: Qualitative Interview and Questionnaire Study

Bérangère Naudé[1,2], PhD; Anne-Sophie Rigaud[1,2,3], Prof Dr; Maribel Pino[1,2,3], PhD

[1]Faculté de Médecine, Université Paris Cité, Maladie d’Alzheimer, Paris, France
[2]Broca Living Lab, CEN STIMCO, Paris, France
[3]Service Gériatrie 1&2, Centre Mémoire de Ressources et Recherches Ile de France, AP-HP, Hôpital Broca, Paris, France

Corresponding Author:
Bérangère Naudé, PhD
Broca Living Lab
CEN STIMCO
54 rue Pascal
Paris, 75013
France
Phone: 33 1 44 08 33 60
Email: berangere.naude@etu.u-paris.fr

Abstract

Background: When older adults (OAs) can no longer live independently at home, they have the option to choose from various types of geriatric care institutions, such as residential facilities or nursing homes. For several years now, thanks to the development of interactive television (iTV), social link functions have been accessible directly on televisions, tools that are already integrated into residents’ rooms. The acceptance of technologies specifically targeting older users, as well as iTV, has been widely documented in the literature, incorporating factors from the innovation resistance model.

Objective: This research aims to enrich the acceptance of existing models of innovation by OAs living in different settings.

Methods: User tests were carried out to evaluate OAs’ experiences with iTV and identify the factors involved in its acceptance. A total of 32 OAs living at home, in nursing homes, or in residential facilities in France were interviewed between November 2022 and June 2023. iTV acceptance was examined using an interview grid based on the technology acceptance model and included the following factors: intention to use, perceived usefulness, perceived ease of use, user resistance, anxiety, facilitating conditions, and user characteristics.

Results: The deductive qualitative analysis based on the technology acceptance model helped to identify 33 concepts.

Conclusions: This study has contributed to the literature on the acceptance of iTV by OAs living at home and in geriatric institutions, particularly by enriching existing models and proposing new avenues for reflection.

(JMIR Hum Factors 2024;11:e58051) doi:10.2196/58051

KEYWORDS

interactive television; iTV; acceptance; older adults; nursing home; residential facility; technology acceptance model; TAM; mobile phone

Introduction

Background

Older adults (OAs) experience perceptual, physiological, and cognitive changes with aging [1,2], which may affect their daily independence. When it is no longer possible for them to remain at home (due to, eg, the presence of cognitive problems, functional problems, or a combination of both cognitive and functional problems) [3], OAs may choose between different types of accommodation, including residential facilities (RFs) or nursing homes (NHs). In France, RFs represent a step between home and institutional care [4]. For several years now, and especially since the COVID-19 pandemic, NHs and RFs have been equipped with information and communication technologies (ICTs; eg, tablets, computers, and smartphones) to promote a feeling of connection between residents and their relatives [5,6]. However, access is not the same as accessibility. In fact, OAs appear to experience stress when using technology, also known as technological anxiety, which is described as (free translation) "a negative psychological state associated with the use or fear of having to use ICT. This experience produces feelings of anxiety, mental fatigue, skepticism and inefficiency" [7]. OAs who are anxious and lack confidence in their own abilities also show greater resistance to gerontechnologies [8-10] and an increased need for support.

With the development of interactive television (iTV), social networking functions (eg, social networks, video calls, and text messages) are now accessible directly on the television screen [11,12]. The appeal of this new technology, in contrast to other ICTs, is rooted in its familiarity. Given that television systems are already a fundamental component of OAs' daily routines (ie, their main source of information and entertainment) [12], iTV may offer a less anxiety-inducing alternative to traditional ICTs [12-16]. Although iTV seems to be able to facilitate access to social link functionalities within geriatric institutions, its long-term use depends on several other factors, including organizational (eg, staff turnover and additional workload), human (eg, health, self-efficacy, and technological and social habits), ethical (eg, privacy and agism), and technological factors (eg, technical and ergonomic problems with iTV and nonexistent or inadequate training) [17-19]. The literature on the acceptance of technologies by OAs is often stereotypical [20], considering the older user to be resistant to any form of innovation and lacking motivation to use a technology [21]. However, some authors [21] challenged these studies, explaining that while OAs value their independence, privacy, and social interaction, products designed specifically for this population focus mainly on safety and assistance aspects. In the technology paradoxes framework [22], consumers face paradoxes while using technology, creating negative emotions, such as anxiety and stress. To combat these emotions, consumers tend to adopt various pre- or postacquisition coping strategies to avoid or confront the technology. Applying this framework to older consumers, Wilson-Nash and Tinson [23] found that most of the strategies adopted by OAs are confrontational, showing that despite previous assumptions of lack of dynamism or willingness [24,25], OAs are keen to master the technology. There is a wealth of literature on the factors influencing the adoption of technological innovations, and several explanatory models have been developed. The 2 main ones are the technology acceptance model (TAM) proposed by Davis [26] and the unified theory of acceptability and use of technology (UTAUT) proposed by Venkatesh et al [27].

Models of Technology Acceptance

According to the TAM, the acceptance of technology, that is, the extent to which a product will be used or not, depends on 2 main factors: perceived usefulness (PU) and perceived ease of use (PEOU) [26]. Subsequently, other extended versions of the TAM emerged, culminating in the proposal of an integrative model named UTAUT [27]. Venkatesh et al [27] differentiate between the determining factors, that is, factors influencing the intention to use (IU) or the use of technologies, and the moderating factors, that is, factors influencing the determining factors. These include expected performance; expected effort; social influence; and the presence of facilitating conditions (FCs), such as documentation. Although these models have been used in a variety of contexts, their original field of application is the workplace. Venkatesh et al [28] then proposed a second version of the UTAUT (UTAUT 2), which is better adapted to the context of consumer use. Three other factors have been added to the existing determining factors: hedonic motivation, monetary value, and habit. Finally, these same authors proposed ways of developing the model, such as integrating contextual factors (eg, organization and physical environment) and individual factors (eg, user and technology characteristics) [29]. At the same time, Chen and Chan [8] have been looking at the acceptance of technologies by OAs and have developed the Senior TAM. This model incorporates factors from both TAM and UTAUT, such as PU, PEOU, attitude to use, FCs, self-efficacy, and anxiety about gerontechnologies. The authors also added age-related factors such as perceived health, cognitive ability, attitude to aging, satisfaction with life, social relationships, and level of physical functioning. Age, gender, level of education, and economic status were also considered as control variables.

Innovation Resistance

The adoption of a technological innovation may also depend on the user's degree of resistance to change, that is, to the maintenance of the status against the pressure of change [30]. Any disruption to users' routines can alter their psychological equilibrium, which then needs to be adapted to reduce resistance [31,32]. An innovation can only be adopted once the initial resistance has been overcome by users [33]. Several authors have attempted to identify the factors influencing this resistance, such as the practice habit, and the perceived risk of the innovation [34] or the innovation, propagation mechanism, and user characteristics (UCs) [33]. According to Joseph [35], 3 types of factors seem to influence resistance: functional, psychological, and informational factors. Finally, user resistance (UR) seems to depend on several multidimensional factors.

The Acceptance Model Used in This Study

In their study on the acceptance of iTV, Im et al [36] proposed an integrative model combining the models of acceptance (TAM [26]) and resistance to innovation [33]. They aimed to go beyond previous theoretical frameworks to better explain the dynamic nature of adoption [37]. According to these authors, the factors influencing iTV adoption are PEOU, PU, UCs (eg, previous experience with innovations and self-efficacy), resistance to innovation, and IU. However, this model does not consider the characteristics of OAs (eg, cognitive and physical disorders, lack of digital literacy, technological anxiety, and low sense of self-efficacy) [2,38,39], nor does it consider the specific characteristics of geriatric residents [40].

This paper presents a qualitative study investigating the factors that might influence the acceptance of the iTV system by OAs living at home, in RFs, and in NHs. The first step was then to create an acceptance interview grid addressing both the specific features of the technology and those of the population studied.

Therefore, this study aimed to enrich existing models of acceptance by suggesting new lines of thought on the determinants of acceptance proposed in these models.

Methods

Overview

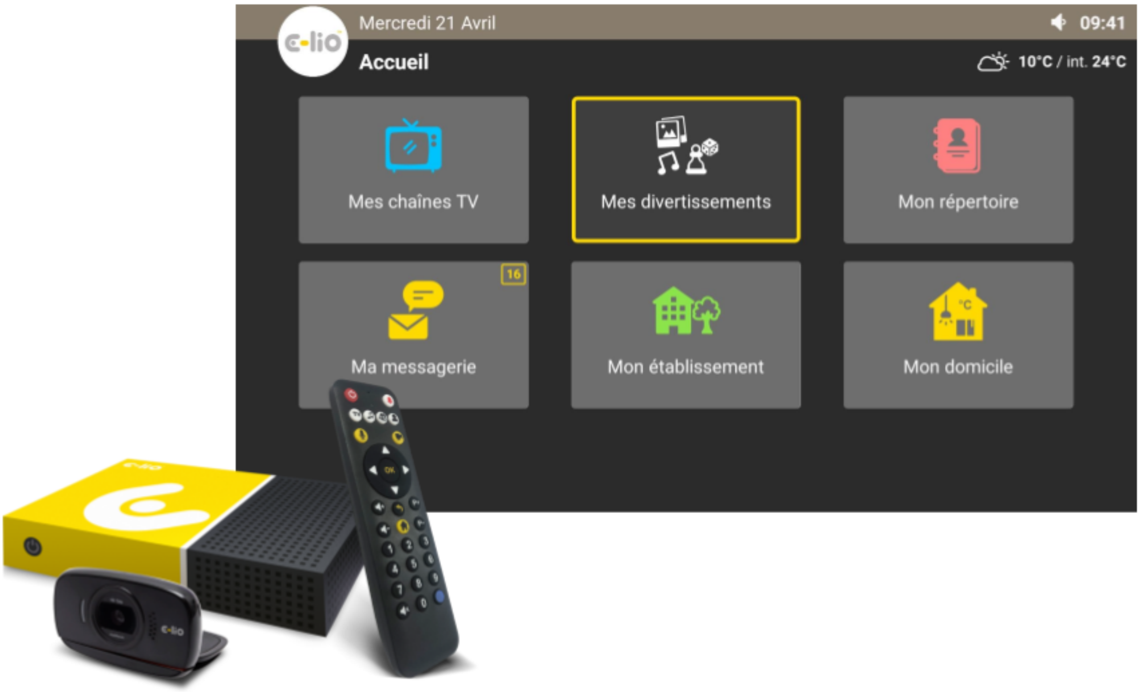
The method used was borrowed from the user-centered design and the qualitative and quantitative methods of user testing [41]. The aim was to place participants in real-life conditions based on pre-established iTV use scenarios. By combining observation and interviews with the participants, the researcher was able to identify the difficulties encountered when using the technology. As part of this study, the participants could receive cues if they were stuck, asked for help, or made a mistake.

Materials

The iTV

The iTV selected for the user tests (e-lioTV) is a system developed by Technosens [42] and installed in geriatric institutions since 2011. A total of 53 NHs and RFs in France are equipped with this iTV, which is installed directly in residents' bedrooms. This iTV consists of a box connected to a television screen and offers communication (eg, text messages, photos, and video calls); entertainment (eg, television channels and radio); and information services (eg, news from the institution; Figure 1). All these functions can be accessed via a single remote control.

Figure 1. The e-lioTV system and the main menu interface: the box, camera, and remote control.



Questionnaires and Interview Grid

To identify the different profiles of participants, we administered a sociodemographic and a technological habit (eg, frequency of use of technologies) questionnaire. A semistructured interview to examine iTV acceptance was conducted at the end of the user tests. An interview guide (Table 1) was constructed based on the literature on the acceptance of technologies [26,43,44], and in particular iTVs [36], among OAs [8]. The final extended TAM (e-TAM) consists of the following seven dimensions: (1) IU; (2) PU; (3) PEOU; (4) UR; (5) anxiety; (6) FCs (eg, peer support and self-efficacy); and (7) UCs (eg, previous experience and interest in technology, social influence, and perceived health). Participants were asked to give their opinion on statements using a 5-point Likert scale ("1" indicating "strongly disagree," "2" indicating "disagree," "3" indicating "neutral," "4" indicating "agree," and "5" indicating "strongly agree"). We then asked them to explain their responses.

Table 1. The interview guide based on the extended technology acceptance model.

Dimensions and affirmations		References
Perceived utility	• "E-lioTV is practical for me." • "E-lioTV provides me with various useful functions and services."	[36]
Perceived ease of use	• "It takes much time and effort to understand and use e-lioTV." • "I found e-lioTV easy to use."	[26,36]
User resistance	• "Using e-lioTV is burdensome." • "E-lioTV is not for me."	[36]
Anxiety	• "I'm afraid of using e-lioTV." • "I'm afraid that if I press the wrong button on the remote, I might break something on e-lioTV."	[43]
Intention to use	• "I want to have e-lioTV in my home." • "I would recommend others using e-lioTV."	[36]
Facilitating conditions	• "In general, if I have a problem with a device, I ask my family or care staff for help." • "It is easy for me to understand and use new devices."	[36,44]
User characteristics	• "The people I care about push me to use messages and video calls." • "In general, I'm curious to learn how to use a new device."	[44,45]

Participants and Recruitment

Participants of the study lived in geriatric institutions (eg, NH and RF) or at home and were not e-lioTV users. Participants living at home were recruited in Paris, while others were recruited in the Saint-Etienne and Grenoble regions. The inclusion criteria were that the participants should be (1) aged >60 years, (2) literate, and (3) able to use a television remote control by themselves. Exclusion criteria were (1) having a Mini-Mental State Examination score of <16 [46] and (2) having visual, hearing, or motor problems that prevented them from using a television remote control, unless they could be compensated for by a technical aid. To this end, for NH and RF residents, professionals were asked to draw up a preliminary list of willing residents who met the above criteria.

In total, 38 participants were approached (RF: n=10, 26%; NH: n=12, 32%; and home: n=16, 42%) between November 2022 and June 2023. Before each scenario, the researcher asked whether the participant wished to continue. Of 38 participants, 5 (13%) chose to stop before the end of the test due to increasing fatigue and 1 (3%) person was excluded because he was illiterate. A total of 32 participants were included, including 23 (72%) women and 9 men (28%). The demographic characteristics of the 3 groups are presented in Table 2. These data seemed in line with the literature, with participants living in NH belonging almost entirely to the category of those aged >80 years [4]. In this study, it is interesting to note that almost all the participants living at home had a higher education degree (postgraduate degree) compared with the institutionalized participants.

Table 2. Sociodemographic characteristics of participants (N=32).

	Gender, n (%)		Age (y), n (%)			Education, n (%)	
	Women (n=23)	Men (n=9)	60 to 69 (n=5)	70 to 79 (n=11)	>80 (n=16)	Undergraduate (n=17)	Postgraduate (n=15)
NH[a] (n=7)	4 (17)	3 (33)	0 (0)	1 (9)	6 (38)	7 (41)	0 (0)
RF[b] (n=9)	7 (30)	2 (22)	3 (60)	1 (9)	5 (31)	8 (47)	1 (7)
Home (n=16)	12 (52)	4 (44)	2 (40)	9 (82)	5 (31)	2 (12)	14 (93)

[a]NH: nursing home.

[b]RF: residential facility.

To explore participants' technological expertise, we assessed the frequency of mobile phone use by functionality (eg, calling, answering a call, sending a text message, and checking a text message) and the overall frequency of computer and tablet use. We then categorized their expertise with the technology as follows: a person is considered an "expert" if they use the feature or technology "several times a day" or "every day or almost every day," and a person is considered a "nonexpert" if they do not use the feature or technology or if they use it "every week" or "every month." According to Table 3, most participants (RF: 9/9, 100%; NH: 6/7, 86%) in the RF group and NH group were not experts in the use of computers, unlike participants in the home group. In addition, some of the participants in the RF group (make a call: 5/9, 55%; send a text message: 4/9, 44%) used their mobile phones more frequently than those in the NH group (make a call: 1/7, 14%; send a text message: 0/7, 0%). It is interesting to note that among the smartphone users (NH: n=0, 0%; RF: n=6, 67%; and home: n=15, 94%), those living in RF tended to restrict the use of their mobile phones to calling and messaging.

Table 3. Technological expertise of participants (N=32).

Level of expertise	Mobile phone, n (%)				Computer, n (%)	Tablet, n (%)
	Make a call	Answer a call	Send a text message	Receive a text message		
NH[a] (n=7), n (%)						
Expert	1 (14)	2 (29)	0 (0)	0 (0)	1 (14)	0 (0)
Nonexpert	6 (86)	5 (71)	7 (100)	7 (100)	6 (86)	7 (100)
RF[b] (n=9), n (%)						
Expert	5 (56)	8 (89)	4 (44)	5 (56)	0 (0)	1 (11)
Nonexpert	4 (44)	1 (11)	5 (56)	4 (44)	9 (100)	8 (89)
Home (n=16), n (%)						
Expert	14 (88)	15 (94)	15 (94)	15 (94)	16 (100)	5 (31)
Nonexpert	2 (12)	1 (6)	1 (6)	1 (6)	0 (0)	11 (69)
Total (N=32), n (%)						
Expert	20 (62)	25 (78)	19 (59)	20 (62)	17 (53)	6 (19)
Nonexpert	12 (38)	7 (22)	13 (41)	12 (38)	15 (47)	26 (81)

[a]NH: nursing home.

[b]RF: residential facility.

Procedure

The tests were carried out individually and systematically by 2 people (BN and a colleague). While the first author conducted each part of the procedure, the other researcher was invited to observe and take notes. Informed consent was obtained on the day of the appointment after the participant had been reminded of the objectives of the research. Before starting the scenarios on the iTV, the overall cognitive level of each participant was assessed, either by the institution's psychologist or by the researcher (first author), using the Mini-Mental State Examination. Then, participants were asked to complete a sociodemographic questionnaire and to share their technology use habits. After a brief presentation of the principle and overall functioning of the iTV (eg, the different components of the tool, such as the box, camera, and remote control, and the main features), participants' first impressions were gathered. The iTV was used through 3 scenarios of increasing difficulty, each with more steps than the last (eg, answering a video call, checking a text message, and making a video call). The researcher then collected participants' opinions or recommendations on the iTV throughout the scenarios. Once the 3 scenarios had been completed, the semistructured interview, based on the e-TAM, was conducted. If necessary, the researcher asked participants to develop some answers. Finally, and at the participant's request, the researcher could also provide an additional explanation of the iTV's marketing status and target audience, as well as the rest of its functionalities not used in this test (eg, radio, news from the institution, and games). Participants were then invited to react on their behalf, as well as on behalf of someone close to them.

Ethical Considerations

Ethics approval for this study was obtained from the Université Paris Cité Research and Ethics Committee in November 2021 (approval number 00012021-91). Validation of the data management procedures and related compliance with the General Data Protection Regulation was obtained from the data protection office and registered in the general register of Greater Paris University Hospitals (Assistance Publique-Hôpitaux de Paris) in February 2022 (20220228123925). The collected data were anonymized in such a way that identification of individuals or sources of information is not possible. Participation in the study was voluntary and subject to the signing of an informed consent form. Participants were informed of their right to withdraw from the research project (resulting in the deletion of their data) and their right to access and rectify information concerning them.

Qualitative Analysis

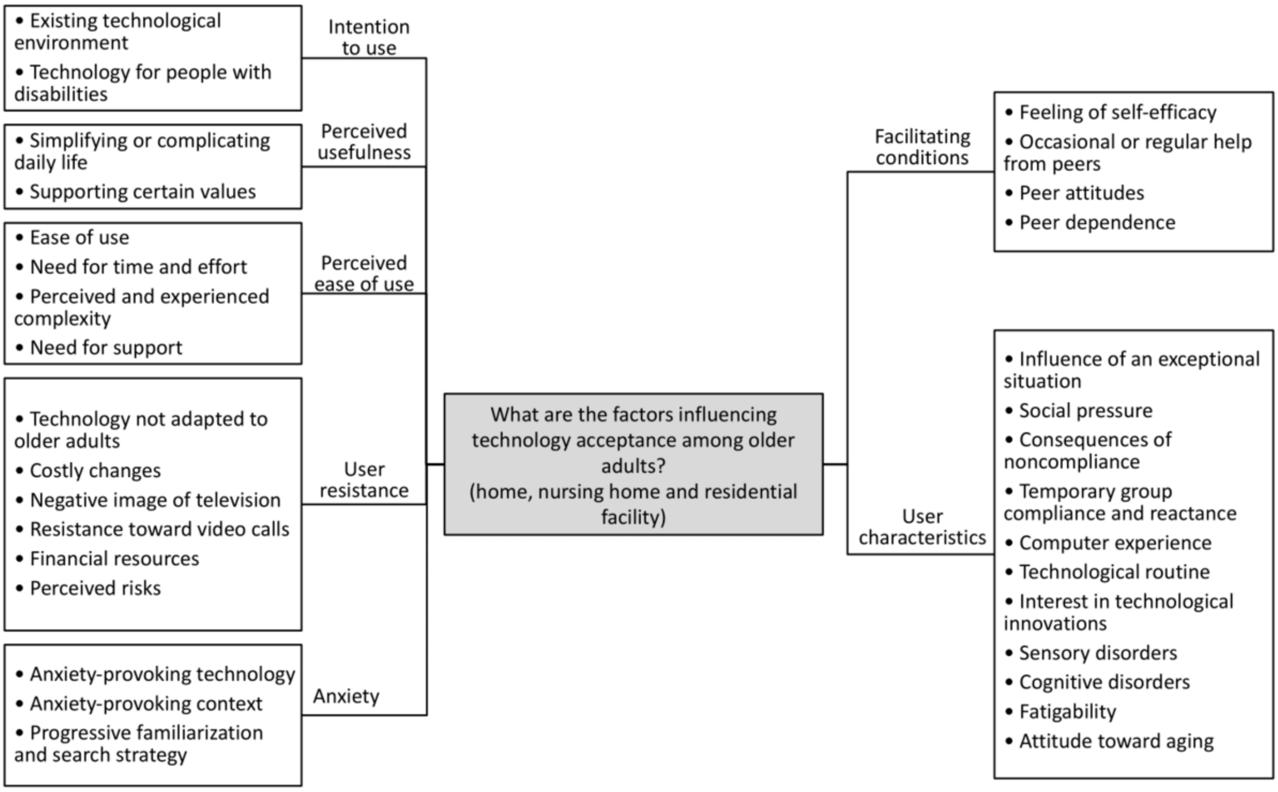
Two authors participated in the qualitative analysis (BN and ASR). Deductive thematic analysis was carried out based on the transcribed user tests. The aim was to identify the factors that could influence performance on the iTV and, therefore, its acceptance, based on the e-TAM factors. To this end, we took inspiration from the Qualitative Analysis Guide of Leuven [47] following three preliminary steps before coding the interviews: (1) highlighting important elements and writing a mini report on the participant's characteristics, (2) designing conceptual schemes, and (3) validity testing. Once the conceptual schemes had been tested and validated, we were able to draw up a list of concepts used to code the transcripts. These concepts were classified according to the e-TAM factors, allowing us to develop some of them further or even to create new ones. An overall conceptual framework for our transcripts, in the form of a mind map, was thus created. Finally, each concept was described, including a short summary of the points raised in the interviews and a few key verbatims.

Results

Overview

According to the e-TAM, iTV acceptance depended on 7 factors: IU, PU, PEOU, UR to iTV, anxiety when using iTV, FCs, and UCs. A summary of the 33 concepts found during the deductive qualitative analysis is presented in Figure 2. Figures 3-9 show the number of participants who rated 1, 2, 3, 4, or 5 on the Likert scale. Different colors and symbols are used to represent the 3 living environments.

Figure 2. Mind map of themes and subthemes from the deductive qualitative analysis based on the extended technology acceptance model.

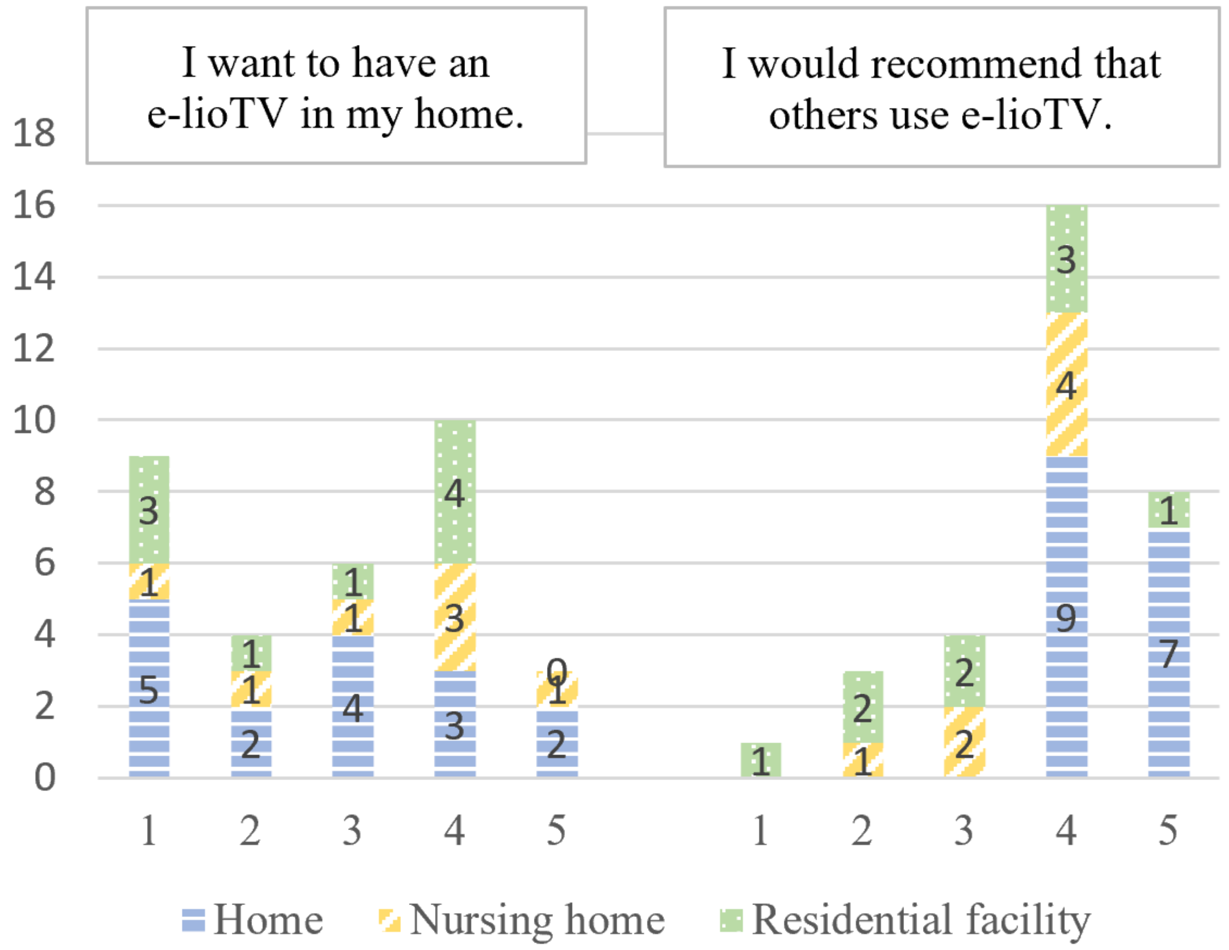


IU Dimension

Overview

Although they seemed to perceive an interest in iTV in general, participants were equally divided between those who would like to have an iTV at home (13/32, 41%) and those who would not (13/32, 41%). However, all participants living at home would still recommend iTV to their friends and family (Figure 3).

Figure 3. Intended use of interactive television by older adults living at home, in residential facilities, and in nursing homes (1: strongly disagree; 5: strongly agree).



Existing Technological Environment

Many participants (30/32, 94%) already owned several technologies, some for several years, and had well-established communication and information habits. Some of them (6/32, 19%) saw no point in using an iTV as a complement. The technology was perceived as a gadget with no added value compared to existing technology (fewer functions or accessibility options, such as voice):

> *Me personally? No, no, I won't use it....Because I have other things, things that react to my voice, things that are much more responsive.* [Participant 2, home group]

There was a multiplicity of everyday technologies:

> *I have enough machines to do it.* [Participant 13, home group]

It is interesting to note that this barrier only concerned OAs living at home, that is, in this study, the participants who owned the most technology.

Technology for People With Disabilities

E-lioTV could be seen as a technology for people who are losing their independence with temporary or permanent cognitive or mobility impairments. In fact, one of the advantages of the iTV identified was the centralization of functions on a technology that is already used on a daily basis and located in a single room, often the most frequented in the home. In the event of a call, this could reduce the risk of falling (getting up in a hurry if the phone rings):

> *It's certain that if one day I'm a lot less mobile, and if I'm less able to flit from one thing to another, I'll use this device.* [Participant 2, home group]

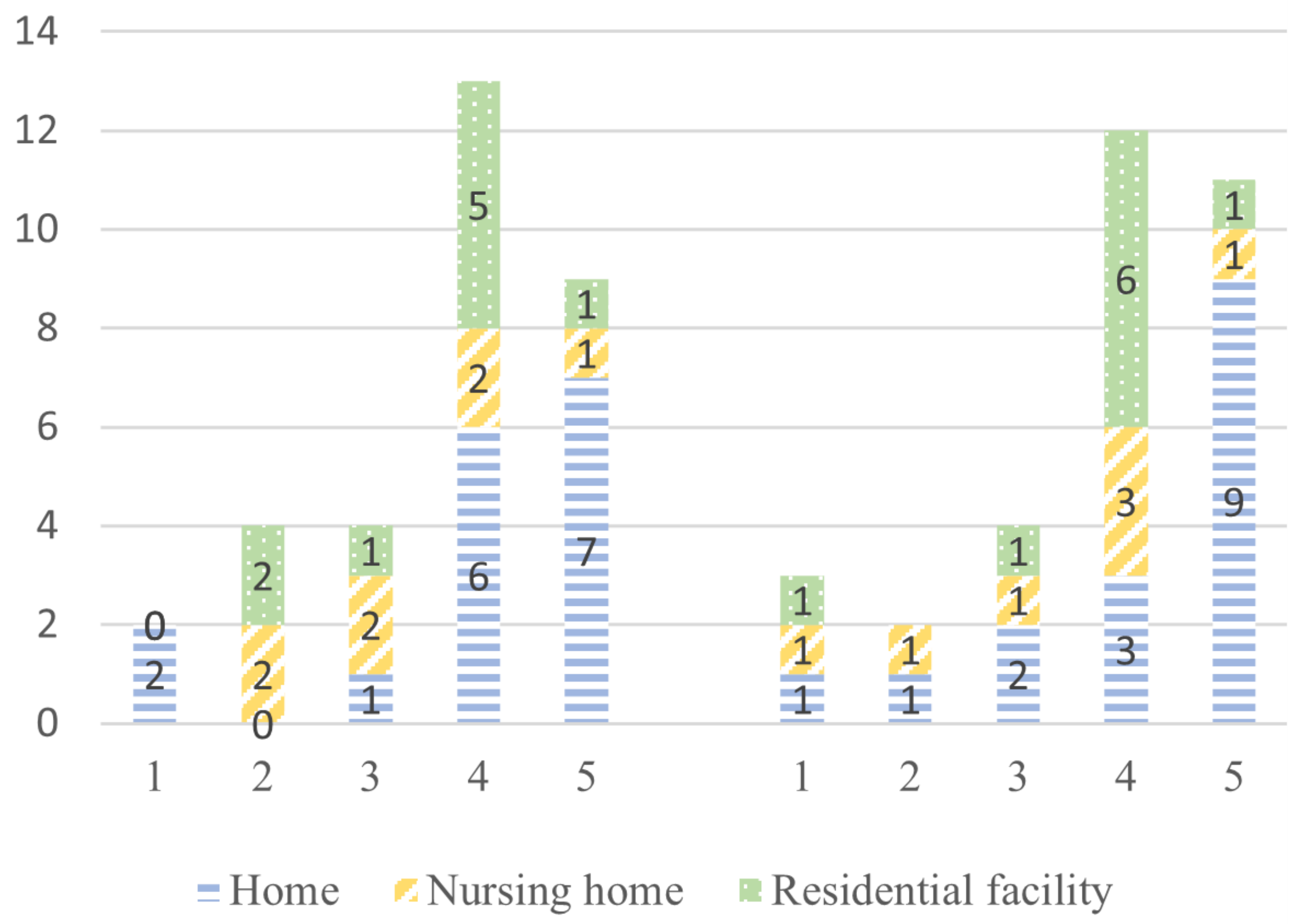
However, this comment was only made by OAs living at home (7/32, 22%), and if we observe the left side of Figure 3, no link can be made between living settings and the IU.

PU Dimension

Overview

The iTV seemed practical for most participants living at home (13/16, 81%) and in RF (6/9, 67%), with the services offered perceived as useful (Figure 4). However, the opinions of the participants living in NHs seemed to be more nuanced.

Figure 4. Perceived usefulness of interactive television by older adults living at home, in residential facilities, and in nursing homes (1: strongly disagree; 5: strongly agree).



Simplifying or Complicating Daily Life

Most participants (22/32, 69%) had a very positive attitude toward iTV and were satisfied with the principle of having a television screen that brings together several functions (23/32, 72%). According to 1 participant, this principle is "very clever" (Participant 3, home group) and avoids the multiplication of technologies in the home (10/32, 31%). Using a television system also means that functionalities such as games can be accessed on a larger screen. Furthermore, it is easier to position oneself in front of a camera on a television screen than on a smartphone. However, some participants (4/32, 12%) seemed to consider technological innovations as an unnecessary complication of everyday life. For example, even before the operation of the iTV was explained to them, some participants showed no interest in discovering this technology:

> *I don't need it because I just want to keep watching the news and that's all...I'm starting to say: "Whoa, let's just stay as we are, let's not complicate our lives.*" [Participant 6, NH group]

The value of technology can also be linked to a person's lifestyle and job. One former farmer, who had always lived in a rural environment, had never understood the usefulness of a computer:

> *And then, in the countryside, it's not really that useful. The main thing is to have the essentials.* [Participant 14, NH group]

Supporting Certain Values

According to some participants (12/32, 37%), the iTV's features were relevant because they supported values that were important to them, including connection with the outside world; the iTV could facilitate interaction with loved ones and provide people with the opportunity to take part in moments of life through photos and video calls:

> *Oh yes, indeed, you can travel that way.* [Participant 7, NH group]

Autonomy and health was also important; other participants appreciated the reflection involved when using the iTV, as well as the games:

> *That way, he can do cognitive remediation on his own.* [Participant 3, home group]

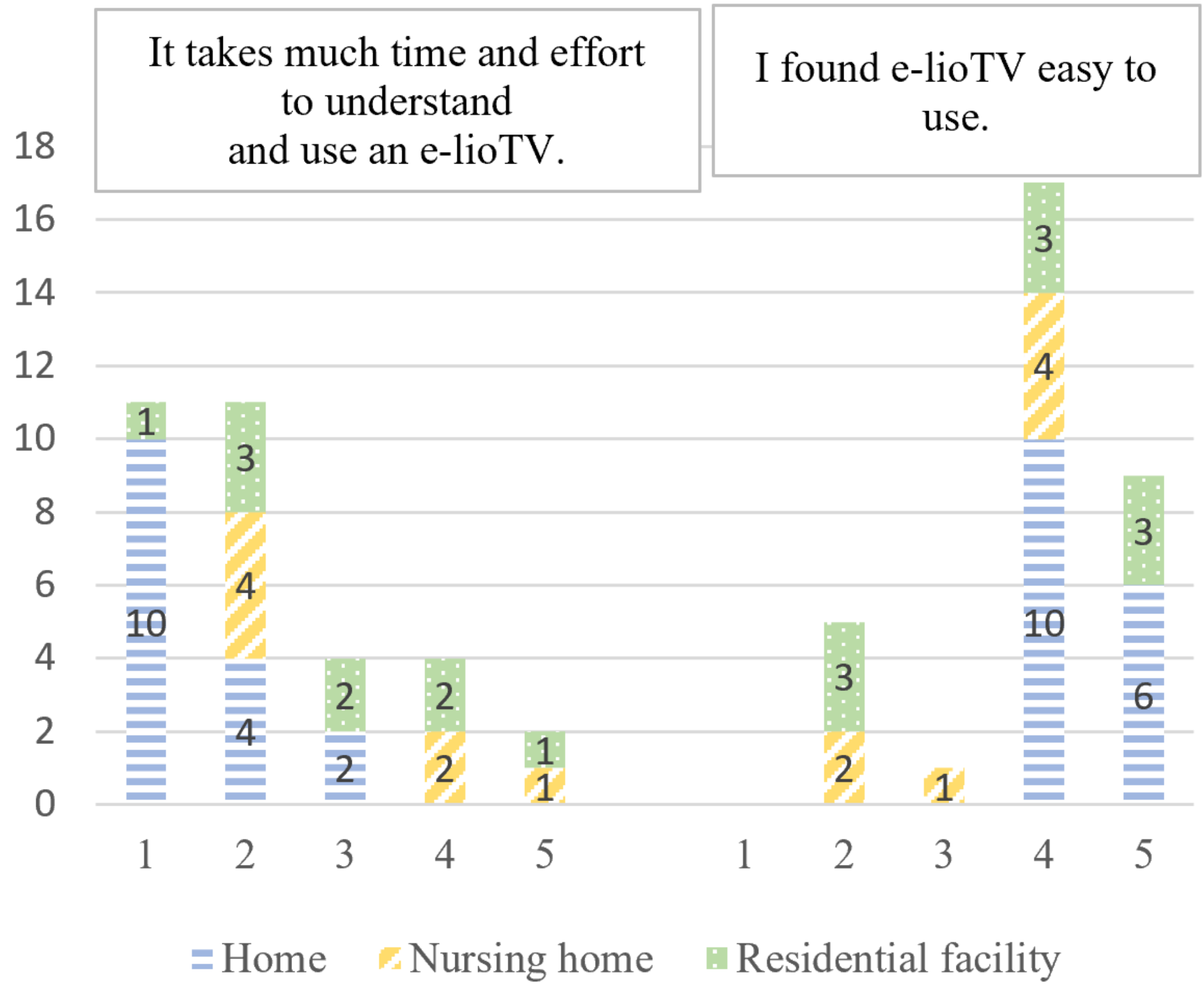
Video calls could also help to reassure family and friends.

PEOU Dimension

Overview

Most participants found the iTV easy to use (NH: 4/7, 57%; RF: 6/9, 67%; and home: 16/16, 100%). However, a few participants living in RFs (3/9, 33%) and NHs (3/7, 43%) reported that they needed a little more time to learn how to use it (Figure 5).

Figure 5. Perceived ease of use of interactive television by older adults living at home, in residential facilities, and in nursing homes (1: strongly disagree; 5: strongly agree).



Ease of Use

The iTV seemed easy to use for most participants (26/32, 81%), even for those with no computer skills or with mild cognitive impairment. Several participants (14/32, 44%) found that the remote control had few new buttons, making it easy to identify the most important ones. The icons also seemed logical and easy to understand. The information inserted on the television screen served as a memory aid; instead of spending time looking for the solution, they felt that the system was doing the work for them by pointing out the right buttons, something that is all the more relevant for OAs:

> *Well yes, everything has been spoon-fed, you know.* [Participant 1, home group]

> *It's visual, but maybe you need that more when you're older too.* [Participant 14, home group]

This guidance greatly reduced the amount of information to be memorized, facilitating familiarization with the system and success during the first moments of interaction, an important phase in forming a good first impression. Finally, the participants had the feeling that they were being supported:

> *It's pretty simple, isn't it, because you're really being taken by the hand from the beginning to the end.* [Participant 1, home group]

Finally, 1 participant was used to a certain complexity on a daily basis (2 remote controls) and seemed pleasantly surprised to be able to do so many things with just 1 remote control.

Need for Time and Effort

Although the iTV was considered rather easy to use, some participants (17/32, 53%) said that they needed time to get used to it. This time was needed to develop automatisms and reflexes and to become familiar with the buttons and the iTV's operating principle, that is, looking at the screen to find the right button:

> *It requires some adaptation, and adaptation takes a little longer now.* [Participant 11, home group]

However, although learning and problem-solving take longer with age, participants were not bothered, also because they had more time to spend on this than young adults:

> *It took me a little while to find it, but does it need to be immediate? I'm not sure.* [Participant 8, home group]

Regarding the need for effort, several participants (6/32, 19%) emphasized the cognitive resources required to handle the iTV. Using this tool seemed to have a cost in terms of attention and mental flexibility to make the link between the guidance on the screen and the buttons on the remote control, select the relevant information on the screen, and remember the actions of certain buttons:

> *I was doing gymnastics between looking at the screen, looking at the remote control.* [Participant 6, NH group]

Perceived and Experienced Complexity

Some participants (9/32, 28%) experienced difficulties in learning how to use the iTV, such as recognizing the icons on the remote control, pressing the buttons gently, or identifying the actions to be performed on the television screen. One participant stressed the need to simplify the technology as much as possible to avoid discouraging users:

> *It can be very discouraging for someone who doesn't understand it very well, and maybe not curious enough to insist either.* [Participant 14, home group]

Indeed, when faced with several failures, 1 participant was tempted to give up:

> *Pfff. What am I going to do? I can't do anything. I'm stuck.* [Participant 15, RF group]

Beyond the experienced complexity, the perceived complexity could influence the way the iTV was approached. Indeed, even before using the device for the first time, participants had already formed an initial impression based on their personal experience with their television. Some participants (4/32, 12%) perceived the remote control as complex (those who were using a remote control with fewer buttons on a daily basis), while others imagined that the iTV would be simple to use, similar to an ordinary television. But this first impression could be misleading, particularly because of the difficulty of applying knowledge of other technologies to this device:

> *It looks very simple, but in the end, you have to find out what it means on the remote control.* [Participant 14, home group]
>
> *I don't know why...I end up thinking about how to use a remote control.* [Participant 14, home group]

Need for Support

Most participants (14/32, 44%) said that they needed extra support or even guidance to be able to use the iTV. The fact that there were no instructions and that they had to manage on their own from the start seemed to disconcert some of them:

> *We're letting people get started without giving them a user manual?* [Participant 2, home group]

One participant highlighted the necessity for OAs to record all details in writing. A summary note of the remote control with the actions associated with each button might be enough, particularly in the early stages of the interaction. However, 1 participant was reluctant to have a user manual, considering it too complicated in general. Another participant preferred to have support from someone, particularly to reassure her during the first moments of interaction:

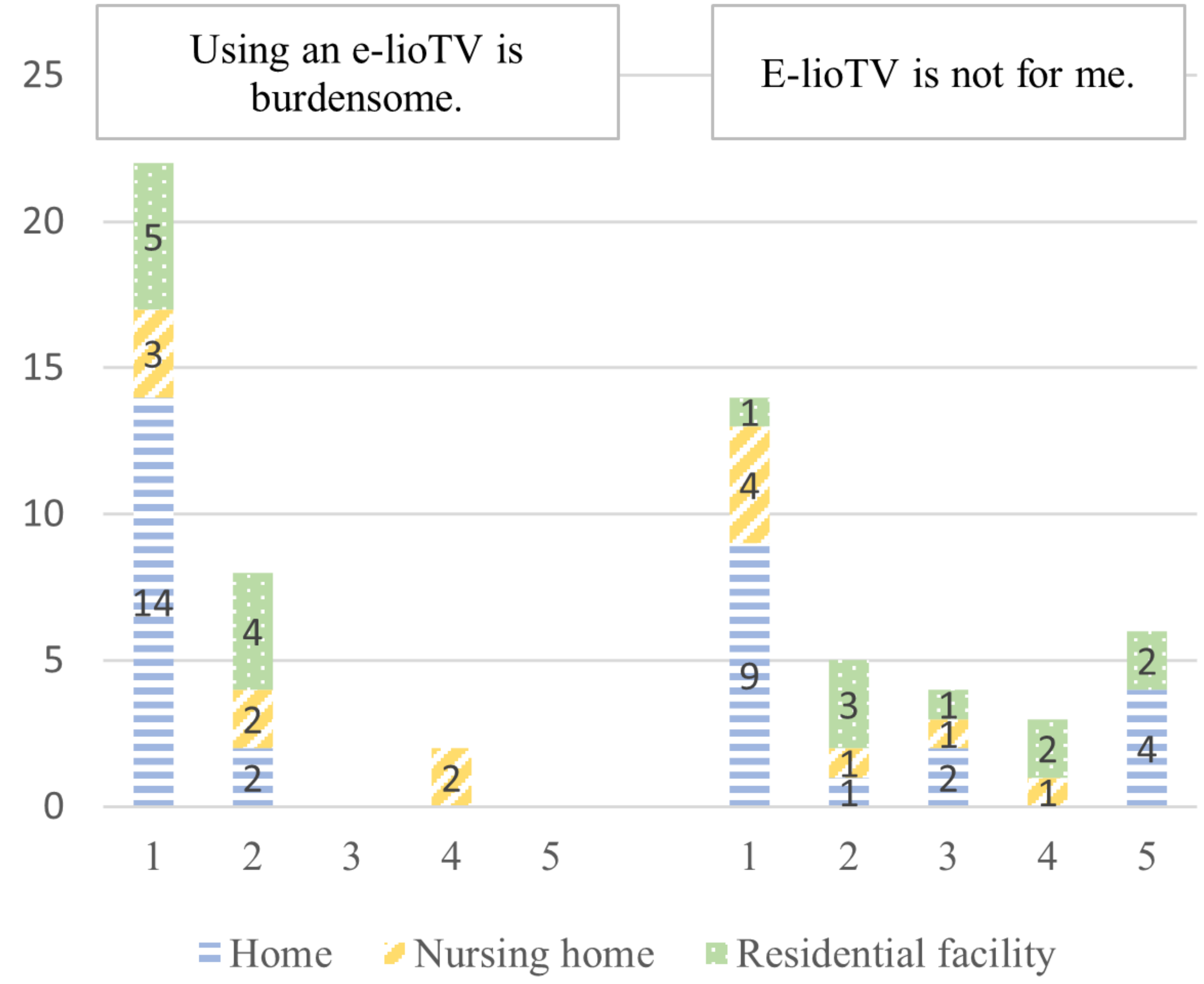
> *Well, at least.... I don't know, one or two days to fully understand the buttons.* [Participant 15, RF group]

UR Toward the iTV

Overview

Although participants did not seem to find the iTV use burdensome (30/32, 94%), almost half of them (13/32, 41%) felt that iTV was not adapted to their needs and uses or did not give an opinion (Figure 6).

Figure 6. User resistance toward interactive television by older adults living at home, in residential facilities, and in nursing homes (1: strongly disagree; 5: strongly agree).



Technology Not Adapted to OAs

Some participants (7/32, 22%) immediately thought that the iTV was not suited to them or their NH neighbors for various reasons, including health conditions, with a participant stating, “We’re all a bit lost, I don’t think it’s really suited to our home” (participant 6, NH group). Other reasons included poor knowledge of technology use and age. This last point can be seen as self-agism. One participant seemed outraged by this stereotype:

> *[This television is not for me] Why? Because I’m stupid? And I’m old? [Laughs].* [Participant 7, NH group]

Finally, although participants found the functionalities rather useful, they almost systematically considered that the iTV was made for someone else, seeing no point in changing their habits.

Costly Changes

Some participants (4/32, 12%) had a negative attitude toward application updates or the switch to smart televisions, showing annoyance or even stress when using them. OAs found it difficult to adapt to change, with each new feature requiring an effort to understand and learn, which became even more demanding with age (eg, slowness and need for repetition):

> *That’s what’s annoying, it’s the constant updates that turn everything upside down. So, you make an effort to keep up with it, you start to assimilate it more or less and that’s it, off you go again.* [Participant 12, home group]

So, although the television remote control is used on a daily basis, most participants (5/32, 16%) were not used to using other buttons apart from the television channels to keep control of the situation. Therefore, using an iTV necessarily requires a learning curve. Another explanation for this annoyance could be the imbalance between the cost and benefit of the change, as it was not necessarily perceived as an improvement, especially if it made other functions less accessible. One participant showed particular resistance, mainly because using iTV made it harder for her to access television channels.

Negative Image of Television

A few participants (5/32, 16%) said that they disliked television in general, preferring to use another technology (eg, computer) or do another activity. The television system is generally perceived as a technology that isolates people and encourages immobility. In addition, 2 participants (6%) found the programs they watched uninteresting. In this case, the television system was confused with the functionality of the television channels, affecting the overall iTV experience:

> *But I'll tell you, honestly... logically I don't like TV.* [Participant 15, RF group]

Resistance Toward Video Calls

A few participants (8/32, 25%) did not like making video calls because they were either uncomfortable with seeing themselves on a screen (7/32, 22%) or perceived video calls as an intrusion into people's privacy (1/32, 3%).

Financial Resources

Several participants (6/32, 19%) quickly raised the issue of the iTV's price. Aside from the usefulness of the system, this aspect was quickly considered by the participants when they were planning to use the system:

> *What bothers me is the money, to change television.* [Participant 14, NH group]

Moreover, as soon as the question of a subscription was raised, the participants seemed reluctant.

Perceived Risks

The participants (10/32, 31%) showed a certain reluctance toward technology in general or toward the iTV in particular because risks, such as addiction and social barriers (6/32, 19%). Several participants felt that the use of technology represented a social barrier, intruding between people and making communication difficult:

> *But it's frightening how many people have their smartphones and don't look at the person next to them. It's crazy.* [Participant 1, home group]

Also, 2 participants (6%) showed contempt or annoyance toward people who spend a lot of time on their phones. One participant feared becoming an addict and refused to learn how to use a digital tool:

> *There are dangers there, I have to be careful.* [Participant 8, NH group]

Finally, 1 participant feared that iTV would be used as an excuse to reduce the number of carers or reduce the number of family visits.

Another barrier was excessive solicitation (4/32, 12%); some participants showed a certain resistance to ICTs (eg, social networks and iTV) because of oversolicitation. According to 1 participant, receiving text messages every day was inconceivable. Indeed, this form of remote communication is relatively recent:

> *No, no, no, once in a while! If people send me messages every day, it's going to explode.* [Participant 1, RF group]

Oversolicitation could then lead to annoyance or even refusal to use the technology:

> *I don't want that but that's the flaw, it's the instantaneity of today's world.* [Participant 12, home group]

However, 1 participant mentioned the possibility of an evolution in this resistance with the evolution in the importance of certain values with age, that is, favoring social contact over entertainment:

> *Human contact is even more important at that age than seeing a documentary on which she falls asleep, eh.* [Participant 12, home group]

Another barrier was a lack of control over data (4/32, 12%); 2 (6%) participants mentioned the dangers of the internet and the importance of data protection. One of them was very reluctant to go online because someone she knew had been scammed. Generally speaking, the lack of control over technology seemed to bother the participants:

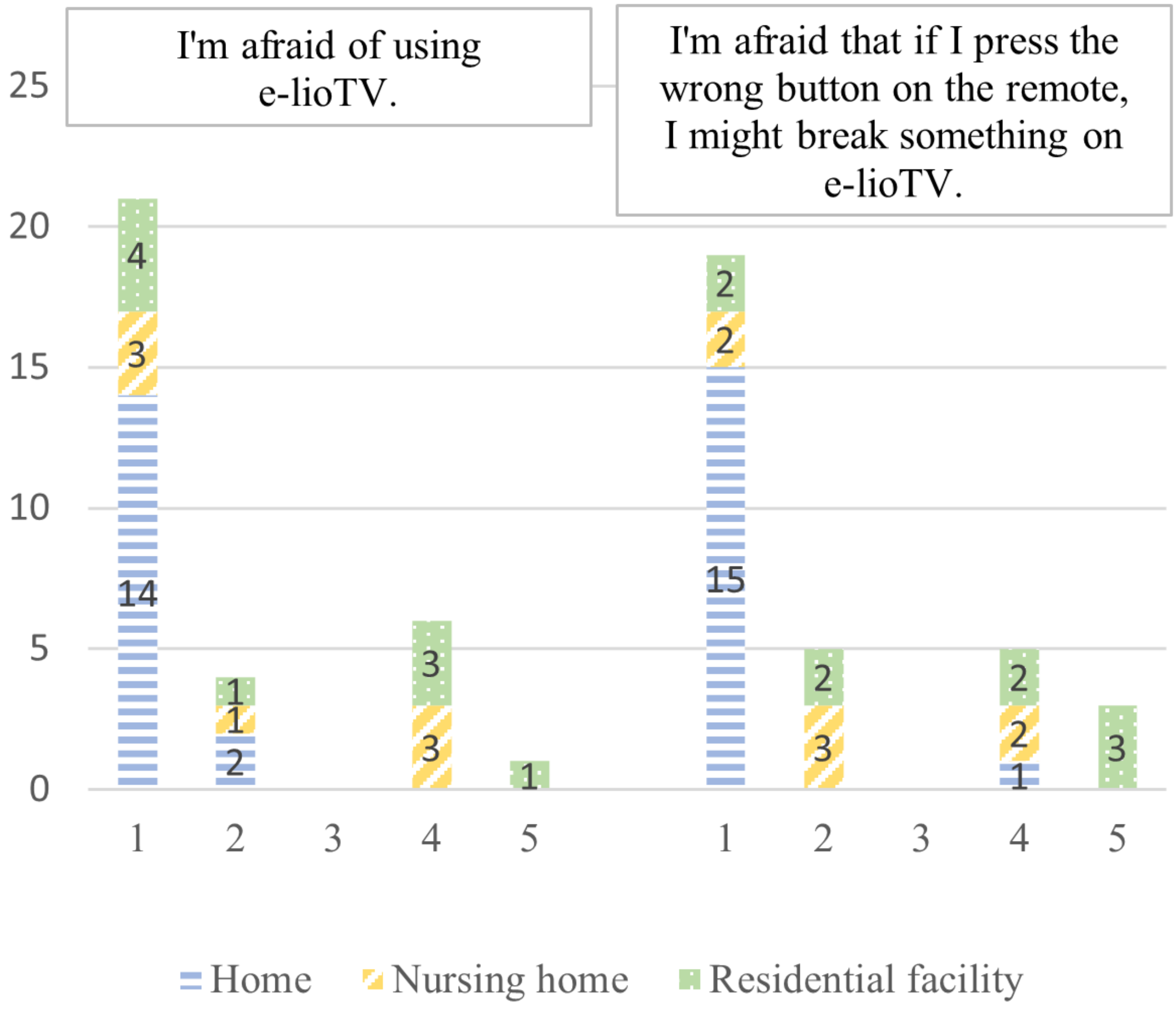
> *And can we deactivate this action? That's the problem.* [Participant 12, home group]

Anxiety Toward iTV Use

Overview

The majority reported no apprehension about using (25/32, 78%) or even making mistakes on the iTV (24/32, 75%; Figure 7). The few participants (7/32, 22%) who reported some anxieties were those who lived in RFs or NHs.

Figure 7. Anxiety while using interactive television by older adults living at home, in residential facilities, and in nursing homes (1: strongly disagree; 5: strongly agree).



Anxiety-Provoking Technology

Several participants (14/32, 44%) reported feeling anxious when using their device: fear of making irreparable mistakes and getting stuck, having lost their bearings:

> *I'm not going to look too hard when it's set, because I'm always afraid of going wrong anyway.* [Participant 14, home group]

Some participants (4/32, 12%) felt this apprehension every time they used their technology, while others (10/32, 31%) were especially stressed during the learning phase:

> *Computers really put me off because I can't get the hand of it...Oh yes, I screw up... I lose things... Pfff.* [Participant 11, home group]

This apprehension was not the same depending on the technology: some participants (3/32, 9%) felt more comfortable with their computer than with their smartphone, with which they were less used to interacting, while others (2/32, 6%) apprehended using the television system:

> *I hope there's no remote control.* [Participant 3, home group]

The remote controls seemed to generate stress because of their diversity (ie, each brand of television has a different remote control, as a participant stated: "Look, I don't really like remote controls, but because they change all the time" [Participant 5, home group]); a large number of buttons; or the presence of generic buttons (eg, colored buttons with no symbols on it and arrows). During the test, 1 participant did not feel confident because of the workload required by the iTV functioning (ie, linking the information on the screen with that on the remote control). These anxious OAs did not necessarily show resistance, but they saw the iTV as inevitably more complicated to use, making it difficult to apply their digital literacy:

> *I do it all the time. I was looking for something more complicated.* [Participant 11, RF group]

To cope with this technological anxiety, some participants (4/32, 12%) adopted coping strategies such as neglect (eg, a participant preferred having someone operate her TV for her) or avoidance (eg, ignoring certain buttons or functions on their television and remote control), stated as follows:

> *Oh no, no, I just do the channel and that's it. No, I haven't explored it, and I don't even know what these are [arrow buttons].* [Participant 7, NH group]

Anxiety-Provoking Context

One participant mentioned the influence of the context in which the technology was used on the stress felt at the time of use:

> *For train tickets for example, I don't know... there's something about it, there's the urgency and the stakes.* [Participant 5, home group)]

According to her, what was most stressful was the stakes behind the successful use of the technology, as well as the potential consequences if it failed (eg, automated teller machine and train tickets). In the case of iTV, the ringtone used to notify the user of a call could also be stressful, with the user rushing to find the button to pick up the call (2/32, 6%).

Progressive Familiarization and Search Strategy

Some participants (11/32, 34%) were not afraid of pressing the wrong button and dared to explore. This confidence may have come from their own experience of computers or from their confidence in their ability to solve problems: they were aware that it was always possible to go back or ask for help:

> *Well, sometimes the experience consists of going back.* [Participant 2, home group]

> *You can go wrong at first, but you can go back, you can fix it.* [Participant 10, RF group]

The participants' level of apprehension could then influence their strategies for finding solutions. If they found themselves stuck, some (5/32, 16%) opted for testing unknown or unused buttons, while others preferred to test randomly, seeing the mistake as an opportunity to learn something else. It is interesting to note that this confidence was acquired gradually, as OAs learned how the technologies worked and overcame their initial reluctance:

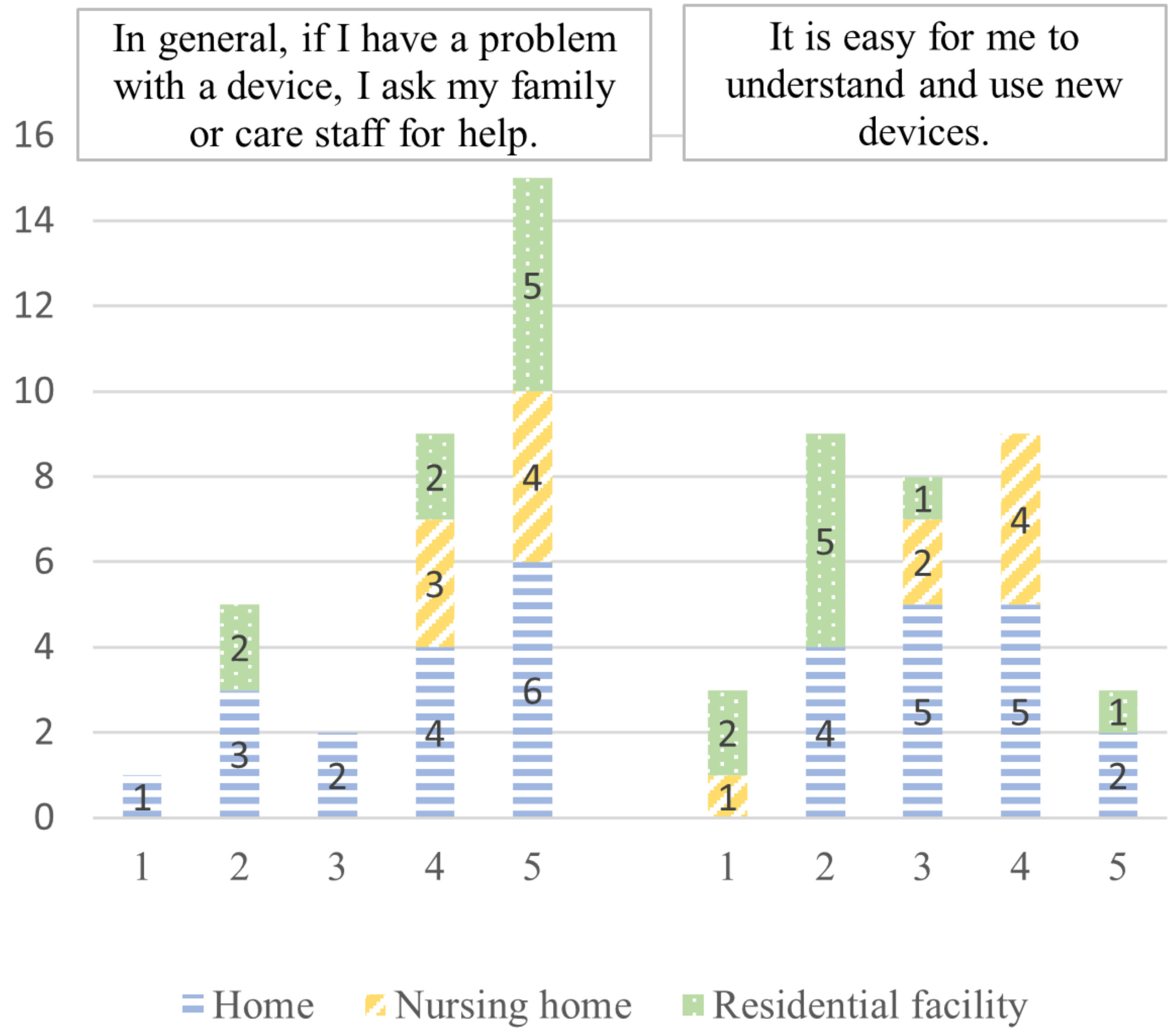
> *Now it's easy, because I've learned, but before, no, it was a real barrier.* [Participant 5, home group]

FC Dimension

Overview

Most participants (24/32, 75%) considered having support from their families or care staff (those living in NH or RF; Figure 8). The few respondents who did not call on family and friends in case of problems were those who lived at home (4/16, 25%) and in RFs (2/9, 22%). However, when it came to their ability to learn how to use a new device, the participants were more divided: of 32 participants, 8 (25%) were undecided, 12 (38%) were fairly confident, and 12 (38%) were not. Interestingly, the participants who were least confident about using a new device were those living in RFs.

Figure 8. Facilitating conditions for the use of interactive television by older adults living at home, in residential facilities, and in nursing homes (1: strongly disagree; 5: strongly agree).



Feeling of Self-Efficacy

As shown on the right of Figure 8, participants' feelings of self-efficacy varied considerably. Some participants (12/32, 37%) were confident in their ability to use and solve problems on the iTV, as 3 (9%) of them pointed out the role of expert that they assumed among their friends and family. Therefore, autonomy in finding solutions (eg, consulting YouTube [Google LLC] tutorials and online forums) could be an indicator of a feeling of personal efficacy. It is interesting to note that the feeling of anxiety depended on the technology and remained variable with time. For example, a participant felt confident using the iTV but doubted their ability to use a tablet. Even before they started the test, other participants (12/32, 37%) were skeptical about their ability to use the iTV independently, as a participant stated, "I'll need help with this" (Participant 2, RF group), and seemed to consider its use as very difficult or impossible:

> *I was looking for something more complicated.* [Participant 11, RF group]

One participant even questioned their place in this test, stating "I'm wasting your time, aren't I?" [Participant 7, NH group]. This lack of confidence in their abilities could have influenced their first experience with the iTV in several ways. As participants used the iTV, mistakes could be more prominent than successes, as a participant stated, "Well, that's a good start!" [Participant 9, RF group], and they tended to blame themselves (eg, slowness, lack of thinking, and impatience and haste) for the slightest problem they encountered, never questioning the design or functioning of the technology:

> *Well, maybe it's because I'm not paying enough attention.* [Participant 6, home group]

The tests revealed the difference between participants who were actively looking for a solution and those who gave up at the first sign of difficulty and preferred to rely on the researcher's help. Finally, a bias identified among some participants (4/32, 12%) was the supposed comparison with other people, younger or of the same age, making the situation even more frustrating:

> *I'm sure someone younger would have understood straight away.* [Participant 14, home group]

Occasional or Regular Help From Peers

In most cases (20/32, 62%), peers provided occasional assistance, in particular, to help participants in the event of a problem, to support them as they familiarize themselves with a new technology or feature, or to explore other functionalities:

> *My son told me "there's a yellow line there." He says you have to press really hard.* [Participant 14, NH group]

Training people in a new technology or function sometimes meant reassuring them during the first moments of interaction. On a few occasions, participants' relatives were able to act as advisers, pointing out the following additional features that could meet their needs:

- Audio message to address difficulty in writing SMS text messages
- Subtitle option to address difficulty in understanding a television program
- Pause function on television to help with urinary problems

Finally, other participants could also benefit from regular help with administrative formalities. In this case, the relatives provided more than just help as they were doing things directly for the participants. Help often came from younger family members and occasionally from friends or digital advisers via local councils. For participants living in institutions, the care staff were able to absorb some of the family's support role, especially when the family was not nearby. This feeling of trust between the resident and the professionals could facilitate the adoption of a new tool.

Peer Attitudes

Some participants (4/32, 12%) complained about the annoyance and impatience of their relatives:

> *My nephew often reproaches me: "You're not watching! It's indicated!"* [Participant 16, RF group]

Moreover, peers may have lost interest when faced with the participant's resistance or may not be available to help if problems arise:

> *My family... they'll tell me "I don't have the time!"* [Participant 15, home group]

However, as 1 participant pointed out, the trainer's attitude can be passed on to the learner and, therefore, influence their attitude toward the technology.

Peer Dependence

Several participants (10/32, 31%) depended entirely on their relatives for certain tasks (eg, video calls and tax declarations), which could accentuate the gap between their current skills and modern technological requirements:

> *For example, I have taxes, but my son does them, so I don't even know how to do them myself.* [Participant 11, home group]

Depending on several people can constitute a barrier and complicate the day-to-day use of technology. For example, 1 participant was very frustrated and angry at having "lost" her computer when she moved into an institution, and another reported his dependence on his wife for the tablet they share. This dependence could be induced by peers or desired by the participants themselves. In the first case, peers might think that they were helping by sparing them the difficulties associated with the technology. This could be frustrating for 1 participant, who felt she was capable of managing her own accounts:

> *Because he's afraid I don't know how to manage my account, so he checks!...No, because I don't need him, I do it myself.* [Participant 16, RF group]

In the second case, some participants tended to wait to be contacted or for someone to do it for them: although they had taken part in the test and shown their ability to use remote control, some participants living in NH (2/7, 29%) preferred to delegate its use to care staff, mainly because of the anxiety-provoking technology:

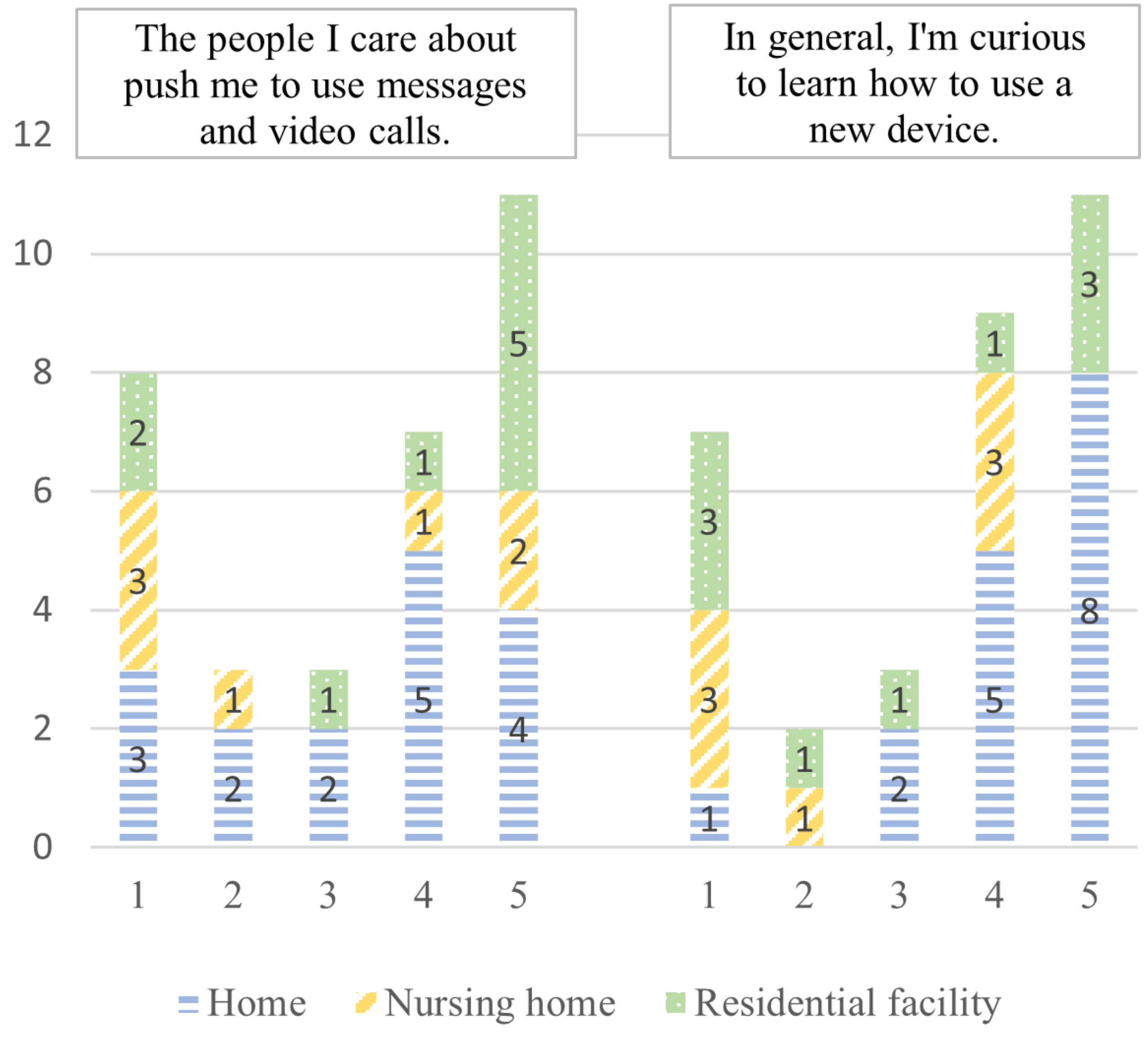
> *When they come, if the TV is on or off, they put it on for me or turn it off. I mean, I know how to turn it off, but still... I don't feel safe using it.* [Participant 14, NH group]

UC Dimension

Overview

The participants from the 3 settings appeared to be divided into 2 groups in terms of the perceived social pressure and their curiosity about technology (Figure 9). In contrast to participants living in RFs who perceived a strong influence from their relatives (6/9, 78%), participants living at home and in NHs seemed more divided.

Figure 9. Participants' characteristics (1: strongly disagree; 5: strongly agree).



Finally, participants living at home seemed to be more curious about the idea of discovering a new technology (13/16, 81%) compared to participants living in RFs and NHs, who seemed more nuanced.

Influence of an Exceptional Situation

In response to exceptional situations (eg, a relative moving abroad and the COVID-19 pandemic), peers may have adopted different behaviors in an attempt to maintain cohesion within the group of friends or family. For example, during the COVID-19 pandemic, several participants (11/32, 34%) mentioned using video calls to stay in touch with their loved ones and to take part in activities (eg, a multimedia café organized by a laboratory). As a result of these exceptional situations, certain functionalities such as video calls or text messages may have become the norm in terms of communication within the family, within a group of friends, or even in the professional environment (eg, association). The iTV could then have more added value since the COVID-19 pandemic and confinements, with the more widespread use of video calls:

> *If she'd had that [iTV with video calls], it would have made our lives a lot easier because we could have seen her. And it would have been nice for her, you know.* [Participant 5, home group]

Social Pressure

Both private and professional peers could have a direct influence on the use of a feature or technology by encouraging the use of particular apps (eg, Google Maps [Google LLC] to find the way home and Skype [Microsoft Corp] for remote meetings; 3/32, 9%). Relatives could also offer participants a technology (eg, a tablet, smartphone, or a connected watch), first perceiving it as a gadget and then, after a period of familiarization, discovering its usefulness in everyday life:

> *No, no, it [connected watch] was offered to me because... I said it was a gadget, but when I saw the applications…* [Participant 2, home group]

Furthermore, some participants (4/32, 12%) felt forced to use several apps depending on the habits of the person they were talking to (eg, Messenger [Meta Platforms], WhatsApp [Meta Platforms], and Signal [Signal Technology Foundation]):

> *I receive messages from people who each have the good idea of having a different application.* [Participant 2, home group]

They also felt oversolicited by their peers: depending on the generation, it is not usual to receive text messages every day or several times a day. However, this influence did not necessarily enable participants to become more skilled at using the features. In the case of video calls, 1 participant (3%) simply received the calls without really understanding how they worked:

> *He calls my sister, the kids, and we all see each other on the mobile. But I'm not the one doing it, I don't know how.* [Participant 11, home group]

Finally, 2 (6%) participants pointed out their own influence on their mode of communication:

> *It would be the other way round...It's easier for me to read than to listen. So, I prefer to have messages...* [Participant 3, home group]

They pointed out the frequency of exchanges with their loved ones:

> *And then I solicit them too, to get news.* [Participant 7, NH group]

Consequences of Noncompliance

Whether or not participants are resistant to ICTs (eg, video calls messages), they sometimes represent the only way for them to keep in touch with their loved ones living abroad (5/32, 16%):

> *I have children who live in Japan, so I only see them that way.* [Participant 8, home group]

For them, resisting these innovations would mean isolation and rejection by their peers. One participant felt excluded from his family following his refusal to use several technologies, and another felt ashamed of being out of step with her RF neighbors:

> *I told them I'm hopeless, I'm old-fashioned! So, they laughed because it can seem completely absurd nowadays.* [Participant 9, RF group]

Finally, 1 participant saw no alternative but to submit to the group norms:

> *Well, it's not that they [the peers] push me, it's mandatory.* [Participant 5, home group]

Temporary Group Compliance and Reactance

The adoption of certain modes of communication (eg, video calls) may have been only temporary, which is the time needed to overcome the exceptional situation that had prompted relatives to adopt them in the first place (eg, the COVID-19 pandemic and relatives moving abroad). Thus, the social pressure present at a given moment is not necessarily effective with time, especially if the participants were initially resistant to change or video calls (7/32, 22%):

> *Well, I'm the one who doesn't want to....We used to do it during the COVID with the young persons who couldn't come and see us. Otherwise, I don't like it, I don't like being seen, I don't think it's... for me anymore.* [Participant 11, RF group]

Although 1 participant communicated a lot via text messages, she seemed annoyed by the regular solicitations from her peers, thus provoking psychological reactance:

> *[People who are important to me push me to use messages and video calls] Yes, that's true, but I don't reply [laughs].* [Participant 7, NH group]

Computer Experience

Most participants (18/32, 56%) saw their previous experience with technology (eg, smartphone, computer, or tablet) as facilitating, if not essential, for using iTV. This technology seemed to partially use knowledge already acquired through other technologies (eg, icons and navigation), making it easier for people who already had certain reflexes and rudimentary knowledge to learn:

> *Yes, there are lots of gestures now that are new, that have been learned and that correspond.* [Participant 9, home group]

Certain symbols that were once incomprehensible (eg, menu and contact) have come to mean something, thanks to the use of technology in everyday life, and therefore, familiarity with technical vocabulary and overall operation:

> *But finally, I think it's an icon [contact] that has become universal now, for everyone who uses a smartphone.* [Participant 9, home group]

However, some participants (3/32, 9%) stressed that it was not necessary to have highly developed digital literacy to use iTV:

> *It's quite accessible to people who don't have any skills, well, who don't have a... how should I say... a culture.* [Participant 16, home group]

Sometimes, certain prior knowledge could even hinder the use of iTV, as in the case of 1 participant who had learned to press hard on the remote control buttons, thus causing several actions on the iTV instead of 1 action.

Some participants (2/32, 6%) pointed out the late access their generation had to the internet. Indeed, several participants (3/32, 9%) had acquired more or less advanced computer skills or even automatic skills with the arrival of computers in the workplace. However, the training received and the use of digital technology (whether at work or not) often date back several decades. So, with the rapid development of technology, 1 participant said she felt out of date:

> *Let's just say it's changed so much since then, it's evolved so much that I'm sure I'm... [out of date].* [Participant 16, RF group]

However, 2 (6%) participants thought that the digital literacy of OAs would evolve rapidly in the future:

> *But I think that use will increase because, in NH, you are dealing with an age group that is of my generation and I belong to a generation that has had very little access to the internet.* [Participant 8, home group]

Technological Routine

In many cases (11/32, 34%), using the computer, tablet, or television was part of a well-established routine:

> *I've had this habit for a very, very, very long time.* [Participant 4, home group]

This routinized use could then lead to various events, such as the creation of habits that are difficult to disrupt. In addition, this routine was all the more difficult to change, and it was hard to establish with peers. Other problems included the development of automatisms, making it harder to transfer knowledge to other technologies:

> *I'm so used to [typing] 3, that's it. But that's to find the right button [on the remote control].* [Participant 6, NH group]

Another problems was the neglect of buttons not used on a daily basis:

> *As I only ever use the same buttons, there may be things I have that I don't even look at.* [H-14]

Interest in Technological Innovations

Participants (14/32, 44%) seemed curious to discover the iTV functionalities and found it fun to try to figure out how the iTV worked, perceiving mistakes as challenges to be overcome.

Some participants (3/32, 9%) were rather curious about new technologies in general (eg, robots and computers), being interested in them since the early days of computing. The time and effort invested at the time would have made it easier for them to learn how to use a new technology:

> *So I bought some books, I didn't understand anything... And I read and reread the same thing over and over again, and then that was it, I was able to design programs. That may also explain why, perhaps more quickly than others, I quickly understand how to use the software.* [Participant 2, home group]

Two (6%) participants emphasized the importance of initial motivation in the learning process, noting that imagining the potential benefits of technology can stimulate its exploration:

> *It can be exciting, in the sense that it's great, I'm going to have something better and everything.* [Participant 4, home group]

Although curiosity seems crucial when learning to use the iTV, it may not be enough due to the apprehension of making irreparable mistakes or the lack of PU:

> *But I don't see what it's for, what it can lead to. Yes, I see it as a game.* [Participant 10, RF group]

Sensory Disorders

With advancing age, sensory problems (9/32, 28%; eg, visual, auditory, and touch) could appear and develop with time, limiting the use of certain technologies, such as smartphones (eg, difficulty typing on a small keyboard) and television (eg, visual fatigue). Participants who wore glasses to improve near vision generally did not wear them to watch television (11/32, 34%). In the case of iTV, these participants then had difficulty recognizing certain symbols on the screen and the remote control:

> *Oh maybe with glasses I'd see better.* [Participant 10, home group]

In terms of hearing, 1 participant with hearing aids felt the sound of the iTV was not loud enough: on a daily basis, he used headphones plugged directly into his television. With regard to touch problems, some participants (5/32, 16%) found it difficult to press certain buttons correctly (eg, arrows and back buttons) because of their design (ie, size and shape of the button) or their location (ie, in the middle of 2 buttons). One participant was convinced that she pressed the right button when, in fact, she pressed the wrong one:

> *That's what I've pressed now, isn't it? [Researcher] You pressed just below it.* [Participant 14, NH group]

Cognitive Disorders

With age, certain cognitive disorders can affect memory, information processing speed, and mental flexibility. During the test, some participants (15/32, 47%) could not remember and frequently asked for the function of buttons they had already used (eg, camera button). Performing the action once seemed insufficient for the information to be retained, which underlines the importance of reminding people on the screen of the buttons to use. Two participants (6%) complained that they were slow in thinking, learning, or even adapting to a new way of functioning:

> *You have to adapt, and it takes a bit longer to adapt now...You're not as fast and your brain doesn't work as quickly as it did a few years ago.* [Participant 11, home group]

Another manifestation of cognitive impairments was the false recognition of certain actions and buttons. Some participants (5/32, 16%) tried to remember a button they had supposedly already used instead of scanning the remote control again to identify the right button. Because of the difficulties mentioned above, 1 participant living at home could not imagine the iTV being used by people with cognitive impairments. Two (6%) others living in NHs felt that they had been in decline for some time, partly due to various events (eg, retirement, the COVID-19 pandemic, and heatwaves) also affecting the residents around them. Thus, 1 participant no longer considered herself "adapted" [Participant 6, NH group] to new technologies such as iTV, while another complained that her life had "shrunk" [Participant 7, NH group], giving her the feeling she "no longer had time to do anything" [Participant 7, NH group]. Finally, participants conscious of their cognitive decline and the impact on their performance questioned and blamed themselves more often for the difficulties encountered during the test:

> *Yes, but maybe if our brains were a bit more developed, we wouldn't do anything stupid either.* [Participant 16, RF group]

Fatigability

As people age, they may become more fatigued, impacting the time spent on certain activities. Because of the effort required to discover the iTV (eg, gymnastics between the screen and the remote control, searching for information, and the number of steps in the scenario), some participants (8/32, 25%) apparently reached cognitive saturation, with increasing difficulty in maintaining their attention, refocusing after a distraction, or even blocking on a task:

> *The trick is to pay attention, you can't do random things. You have to try and think, which isn't always easy at our age.* [Participant 10, RF group]

The fatigue accumulated throughout the test may have hampered the use of the iTV and led to frustration if the participant got stuck on a task already completed:

> *My brain doesn't want to work anymore.* [Participant 4, RF group]

For example, 1 participant gradually lost sight of the test objective, constantly wanting to return to the TV channels instead of following the instructions. By making repeated mistakes, 2 (6%) participants admitted confusing good and bad learning.

Attitude Toward Aging

Among the participants living at home, some seemed to perceive the onset of cognitive problems as imminent and inevitable. Beyond the participants' current state of health, the image they had of aging and their perception of the time they had left in good health, or even to live, could influence their commitment to iTV, as well as its acceptability. One participant seemed impatient, declaring that he had no more time to lose with technologies that were complex to learn or that didn't meet his needs:

> *This thing is a tool, so it has to provide me with the services I need quickly and immediately. I've got no time to lose in my life, I've got 89 anyway [laughs].* [Participant 9, home group]

Others did not see the point of learning to use the iTV, believing that they didn't have much time left to enjoy it:

> *I'm in my 91s this year. So maybe I won't get much out of it.* [Participant 14, NH group]

Discussion

Principal Findings

The main aim of this research was to contribute to the enrichment of existing models of acceptance of innovations by OAs. This study focused on the acceptability of an iTV system by 32 people living in different settings (eg, home, NH, and RF). A deductive qualitative analysis based on e-TAM identified 33 concepts, each related to the following themes: IU, PU, PEOU, UR, anxiety about the iTV, FCs, and UCs.

Advantages of the Methodology Used

This qualitative research complements a previous study, currently being submitted, aimed at assessing the perceived usability and ability of OAs to use the iTV. The data extracted from these tests were self-reported and subject to certain biases. For example, some authors have shown the influence of stereotypes on the perceived ability to innovate [48], while others have discovered that OAs who reported using smartphones were in fact only using the basic functions of the phone (ie, making calls and sending SMS text messages) [49]. Therefore, using a purely quantitative study does not allow for a comprehensive understanding of the mechanisms behind the acceptance of technological innovations.

The deductive qualitative analysis used in this study is based on the Qualitative Analysis Guide by Leuven [47]. One of the advantages of this analysis method is the combination of approaches used. The transcripts were first analyzed separately, using a case-oriented narrative approach. This approach ensured that the interaction between the ideas from each transcript and the specific characteristics of the participants was not overlooked, particularly with the help of conceptual diagrams. Then, by cross-referencing all these ideas, more global concepts were developed. It was only at this point in the analysis that all the data were interpreted, allowing themes to be developed by drawing on the richness of the data and doing justice to the complexity of the experiences of the participants [50].

Technology Paradoxes

This qualitative analysis has highlighted the different behaviors of OAs when they are confronted with the use of technological innovation. Participants (20/32, 62%) were both very curious about the functioning of iTV and anxious about having to use a television remote control. Some (15/32, 47%) said that they were already beyond the stage of being apprehensive about using technology, with 1 participant mocking these anxious people:

> *You'd think we were going to bring out the atomic bomb! "No, no, no, I don't want to touch it! It's going to be dangerous!"* [Participant 4, home group]

It is interesting to note that these behaviors seemed to be influenced by different factors, such as the level of digital literacy or place of residence. In our study, the participants who lived in institutions (16/32, 50%) were also the ones who were the least familiar with the use of smartphones or computers. These results are in line with the literature on the use of technology by OAs [51-53] and highlight the nuances of older users' experience when confronted with technological innovations.

In addition to current models of technology acceptance (eg, TAM and UTAUT), the authors have proposed a framework of technology paradoxes [22,23]. This model states that users, when using a technology, are subject to a certain number of paradoxes that create strong emotions, such as anxiety and stress. Wilson-Nash and Tinson [23] identified 3 types of paradox: functional, social, and psychological.

Among the functional paradoxes, the paradox of chaos versus control was the origin of several situations reported by participants and classified in the category of resistance. In the case of the participant complaining about the addition of functions on her television, she experienced the transition from a situation of control (ie, watching television channels) to a chaotic situation (ie, YouTube starting at the same time as the television), resulting in frustration and the stress of losing one's bearings.

In order to deal with the emotions generated by paradoxes, users may adopt various coping strategies before or after acquiring technology to avoid them or confront them. Wilson-Nash and Tinson [23] have identified 5 coping strategies used by OAs: neglect, partnership, control, adjustment, and acceptance. The partnership strategy seems to have been used by some participants in this study (3/32, 9%) via a mechanism of humanization of the technology. The participants tended to consider the indications displayed on the left of the television screen as instructions to follow or even tips given by the iTV. The participants spoke of the "interrelationship" and mutual assistance that existed between them as users interacting with the remote control and the television screen:

> *View the photos, I press OK? Because they tell me to.* [Participant 11, home group]

Therefore, the personification of iTV observed by the researcher could prove to be a strategy for confronting the technology, enabling participants to overcome their anxiety when using the iTV. It would be interesting to investigate the humanization of technology further and to test whether reinforcing this feeling (eg, the presence of a virtual agent on the iTV) could influence the perceived usability and acceptability of the iTV.

Technologies are often perceived as isolating people from others, as evidenced by the risks perceived by the participants in this study. However, almost all participants (24/32, 75%) also admitted to calling on certain members of their family or friends to resolve problems encountered with their technology. Thus, mastering the technology, which is another adaptation strategy, also includes a social dimension.

The acceptance strategy put forward by Wilson-Nash and Tinson [23] states that OAs, because of their life experience and time perspective, have a greater tolerance of technology faults. In this study, the difficulties encountered and the time required to understand how the iTV worked did not seem to have a major influence on its perceived usability. Therefore, error tolerance may be a differentiating factor between generations, as may apprehension when learning how to use an iTV.

Sense of Self-Efficacy and Performance

The feeling of self-efficacy refers to the evaluation of one's personal abilities, not one's personal worth [39]. Belief in one's abilities can influence the level of effort put into using and learning the technology, as well as thoughts and emotional reactions when using the technology [54]. Thus, the fact of not believing in one's abilities, or even failing several times, does not normally influence one's self-esteem but, on the contrary, can greatly harm motivation and performance. In this study, by blaming themselves for the difficulties they encountered, the participants tended to feel ashamed of not succeeding and even devalued themselves:

> *So turning it off, I don't know where it is. That's what I told you, eh, I'm ashamed but... [laughs].* [Participant 9, RF group]

In this case, the perceived inability to use an iTV seemed to influence self-esteem. Furthermore, Arning and Ziefle [55] stated that when a device presents technical difficulties, OAs experience a loss of technical confidence and begin to question their own value. These elements refer to the paradox of competence versus incompetence proposed by Wilson-Nash and Tinson [23], where users can feel both a sense of accomplishment when using technology successfully and a sense of self-doubt when encountering difficulties.

Authors have also found that OAs who consider themselves targeted by a stereotype tend to reduce their capacity for innovation (ie, their ability to mobilize their resources) to avoid appearing incompetent [48]. Therefore, agist stereotypes may indirectly influence older users' feelings of efficacy.

Another factor that can influence OAs' sense of self-efficacy is institutionalization, which is often associated with an increase in the degree of dependency of OAs. An institutionalization syndrome is characterized by apathy, indifference, reduced cognitive ability, difficulty in expressing feelings, and loss of autonomy [56]. The lives of residents in geriatric institutions are routinized to facilitate the organization of care, implying that residents may lose control over their activities. However, residents who are given control over their daily lives are more socially active, more involved in activities, happier, healthier, and live longer than those who are kept staff-dependent [57,58]. Nevertheless, beyond the possibility of exercising control, the desire for personal control tends to diminish with age, as does self-esteem and the belief in efficacy [39]. One participant in this study, living in an NH, relied heavily on the care staff to use her television. Although she took part in the test and showed that she was capable of using television, she preferred to delegate its use on a daily basis, in particular, because it made her anxious. Other levers exist, such as the presence of a role model (ie, a person with whom we can identify [age and sex] and who is able to manage) or resorting to social persuasion (ie, support and encouragement from carers and care staff) [39].

The Influence of Motivation on the Acceptability of ICTs

Several participants (3/32, 9%) stressed the influence of motivation in the acceptance of a technology or feature. According to the literature, there is an essential distinction between voluntary behavior and behavior resulting from external pressure or control [59]. Intrinsically motivated behaviors are performed because of interest and to satisfy innate psychological needs for competence and autonomy, whereas extrinsically motivated behaviors are often carried out for external and instrumental reasons.

In the literature, enjoyment has often been used to represent this notion of intrinsic motivation [60]. Therefore, the enjoyment experienced when using a technology could reduce the perception of the effort made during its use [61] and thus be a predictor of the IU [60]. In this study, some participants (3/32, 9%) expressed curiosity about the iTV and emphasized the playful nature of its handling, taking a certain enjoyment in discovering its functioning. What was categorized as an interest in technological innovations could, therefore, represent a source of intrinsic motivation for the participants. This study appears to align with existing literature, although there are certain limitations to the concept of enjoyment. Indeed, while the use of the iTV could be enjoyable, some participants (13/32, 41%) saw no point in using it on a daily basis. Therefore, PU remains an essential factor in the long-term acceptance of technology.

Another aspect raised in this study is the extrinsic motivation to use a new technology. Some participants (4/32, 12%) mentioned a certain pressure to learn how to use a computer, for example, with the digitization of administrative services (eg, train e-tickets and tax). Indeed, 1 participant said that under normal circumstances, she would never have explored a new technology on her own. However, although this digitalization is pushing OAs to acquire digital knowledge, this learning process remains costly. The obligation, combined with the difficulty of increasing skills and knowledge, seemed to irritate these participants. Indeed, to avoid constantly disturbing her family and friends, 1 participant put a certain amount of pressure on herself to solve her problems on her own:

> *They seem to say that we all have to learn and manage. But it's not... it's not easy, is it?* [Participant 11, home group]

Limitations

This study had several limitations, such as the small number of user tests conducted. The small sample in the study prevented us from generalizing the results to OAs living at home, in NH, and in RF.

Another limitation was the recruitment method for OAs living at home. Owing to technical and logistical constraints, user tests with these participants were conducted solely in the Paris region, while tests with OAs in NHs and RFs primarily took place in rural areas of the Grenoble region.

The final limitation of the study was the influence of experimenter bias, which could either negatively impact or inadvertently enhance participants' acceptance of the iTV. Throughout the scenarios, the researcher was able to provide clues on how to use the iTV, if the participants asked for them, if they were blocked, or if they made a mistake. The presence of the researcher in this supportive role could then help the participants to overcome the problems encountered more easily, thereby reducing the perceived difficulty of use and anxiety:

> *[I felt confident when I used e-lioTV] Yes, but because you were there, sweetheart!... Otherwise, it would have gone out of the window.* [Participant 4, RF group]

However, the presence of the researcher could also increase anxiety, placing the participants in a situation of evaluation:

> *It's the setting in a way. It's an exercise, I have to pass, so there's stress.* [Participant 9, home group]
>
> *And I don't know... You're destabilizing me [laughs].* [Participant 7, NH group]

Future Work

This research was inspired by the user-testing method, which is rooted in user-centered design. This method enabled us to observe the behavior of OAs, as they familiarized themselves with the iTV, and to identify the factors influencing its use. However, this evaluation was a time evaluation and did not allow us to monitor and understand changes in acceptance with time. A future study could focus on the long-term use of the iTV, specifically examining how the influence of the factors identified in this research evolves with time. It would be valuable to investigate whether these factors continue to affect acceptance of the iTV after several months of use. Moreover, the e-TAM developed in this study was based on different models from the literature (eg, TAM and Senior TAM). Another perspective could be to incorporate concepts from user experience research to propose a unified model.

Recommendations

Overview

Conducting user tests with OAs living in different living environments requires techniques to be personalized and adapted, in particular, to compensate for their frailty and low levels of self-efficacy and digital literacy [62]. The following recommendations seek to share best practices for implementing user-centered design methods with OAs. They are intended for researchers and companies aiming to develop user-friendly technologies tailored to OAs.

Prioritize the Use of Codiscovery Learning

The think-aloud method requires users to express their thoughts, feelings, and actions while interacting with a tool. During the user tests, we observed that OAs struggled to verbalize their experiences while using the iTV, often forgetting to do so altogether. This is consistent with findings in the literature indicating that concurrent think-aloud practice induces greater cognitive load and stress in older participants, thus negatively affecting their performance [63-65]. Therefore, researchers should prioritize codiscovery learning, where users collaborate in teams. This approach encourages them to verbalize their thought processes while interacting with one another to complete tasks [64].

Account for Desirability Bias in Researcher-Participant Interactions

During user tests, participants often feel that they are being evaluated, leading them to adopt behaviors that present them in a positive light. This social desirability bias can result in excessively positive feedback about the device, even in the face of clear usability limitations. Among the various tactics adopted by participants, some may portray themselves as weak or dependent to elicit assistance, for example: "So, uh, I don't know how to do this... As I said, I'm not good at this!" (Participant 9, RF group). In such cases, researchers can respond to each question with another question and reassure the participant that they are on the right track [62]. This technique, integrated into the protocol of this study, enabled some participants with lower self-efficacy to successfully complete each scenario.

Conclusions

This study aimed to describe in detail the factors influencing iTV acceptance among OAs living in different settings (home, NH, and RF), using the example of e-lioTV. A total of 32 OAs used the communication functionalities of the iTV system (ie, messaging and video calling) and shared their opinions about the learning process, daily use, and iTV adoption. On the basis of the e-TAM built on the technological acceptance literature, 33 concepts were identified among the 7 determinants of iTV use (IU, PEOU, PU, UR, anxiety toward iTV use, FCs, and UCs). No new determinant was found during the qualitative analysis, even though some factors (eg, FCs and UCs) were completed and extended to consider all the nuances of OAs' experience when interacting with an iTV system.

To conclude, the iTV acceptance seemed to be context, technology, and characteristic dependent. The participants seemed to agree with recommending the iTV to others; they also found using the iTV relatively easy and pleasant, and thus, they felt confident when using it. Almost all participants considered having support from their family and friends to learn or use a new technology. However, the participants had a divided opinion on iTV IU, the feeling of self-efficacy, and social pressure.

It is also interesting to point out the dimensions that generated the most disagreement between the institutionalized participants and those living at home. While participants living at home seemed to be quite consistent on some dimensions, the institutionalized participants shared more nuanced opinions about the usefulness of the iTV, the effort required to learn how to use it, the resistance to iTV, and the apprehension about pressing the wrong button. Finally, persons living at home were more curious about and interested in technological innovations than other participants.

Acknowledgments

This research was funded by Technosens (Grenoble), Association Nationale de la Recherche et de la Technologie (grant 2021/0344), and the article processing charge was funded by Assistance Publique-Hôpitaux de Paris. This paper was part of the doctor of philosophy thesis (Convention Industrielle de Formation par la Recherche) of BN, which aimed to study the development and use of an iTV system in older person care institutions. The authors would like to sincerely thank the Technosens team, especially Cindy Chassang for her scientific supervision and Frédérick Lorent, Mélissa Neyrand, Yoann Roche, and Cyril Thevenet for their support during the evaluation stage. Finally, the authors would like to thank the older person care institutions for agreeing to welcome the user tests and all participants for their contribution to this study.

Authors' Contributions

The study was conceptualized and formally analyzed by BN and ASR. BN was involved in the investigation and project administration of the study. BN, ASR, and MP were involved in the study methodology, writing the original draft, and reviewing and editing it. All authors have read and agreed to the published version of the manuscript. The funders had no role in the design of the study; the data collection, analyses, or interpretation; the writing of the manuscript; or the decision to publish the results.

Conflicts of Interest

None declared.

References

1. Smith DB, Meshkati N, Robertson MM. The older driver and passenger. In: Peacock B, Karwowski W, editors. Automotive Ergonomics. London, UK: Taylor & Francis; 1993.
2. Matthews G, Roy Davies D, Stammers RB, Westerman SJ. Human Performance: Cognition, Stress, and Individual Differences. London, UK: Psychology Press; 2000.
3. Luppa M, Luck T, Weyerer S, König HH, Brähler E, Riedel-Heller SG. Prediction of institutionalization in the elderly. A systematic review. Age Ageing 2010 Jan 23;39(1):31-38. [doi: 10.1093/ageing/afp202] [Medline: 19934075]
4. Boucaud-Maitre D, Meillon C, Letenneur L, Villeneuve R, Dartigues JF, Amieva H, et al. Health trajectories of elderly living in French senior housing: a longitudinal perspective. Sci Rep 2023 Apr 04;13(1):5471 [FREE Full text] [doi: 10.1038/s41598-023-32429-3] [Medline: 37015961]
5. Schuster AM, Cotten SR. COVID-19's influence on information and communication technologies in long-term care: results from a web-based survey with long-term care administrators. JMIR Aging 2022 Jan 12;5(1):e32442 [FREE Full text] [doi: 10.2196/32442] [Medline: 34878989]
6. Naudé B, Rigaud AS, Pino M. Video calls for older adults: a narrative review of experiments involving older adults in elderly care institutions. Front Public Health 2021 Jan 14;9:751150 [FREE Full text] [doi: 10.3389/fpubh.2021.751150] [Medline: 35096731]
7. Fuhrer C. Diminuer le technostress pour que s'exprime la capacité d'absorption de l'utilisateur? La Revue des Sciences de Gestion 2021 Mar;309-310:57-71. [doi: 10.3917/rsg.309.0058]
8. Chen K, Chan AH. Gerontechnology acceptance by elderly Hong Kong Chinese: a senior technology acceptance model (STAM). Ergonomics 2014 Mar 24;57(5):635-652. [doi: 10.1080/00140139.2014.895855] [Medline: 24655221]

9. Jarvis MA, Sartorius B, Chipps J. Technology acceptance of older persons living in residential care. Inf Dev 2019 Jun 13;36(3):339-353. [doi: 10.1177/0266666919854164]
10. Wilson G, Gates JR, Vijaykumar S, Morgan DJ. Understanding older adults' use of social technology and the factors influencing use. Ageing Soc 2021 Apr 20;43(1):222-245 [FREE Full text] [doi: 10.1017/s0144686x21000490]
11. Ahmed BS, Bures M. Testing of smart TV applications: key ingredients, challenges and proposed solutions. arXiv Preprint posted online on March 14, 2019 [FREE Full text] [doi: 10.1007/978-3-030-02686-8_20]
12. Epelde G, Valencia X, Carrasco E, Posada J, Abascal J, Diaz-Orueta U, et al. Providing universally accessible interactive services through TV sets: implementation and validation with elderly users. Multimed Tools Appl 2011 Dec 22;67(2):497-528. [doi: 10.1007/s11042-011-0949-0]
13. Coelho J, Guerreiro T, Duarte C. Designing TV interaction for the elderly – a case study of the design for all approach. In: Biswas P, Duarte C, Langdon P, Almeida L, Jung C, editors. A Multimodal End-2-End Approach to Accessible Computing. London, UK: Springer; 2013.
14. Mitchell V, Nicolle C, Maguire M, Boyle H. Web-based interactive TV services for older users. Gerontechnology 2007;6(1):20-32. [doi: 10.4017/gt.2007.06.01.003.00]
15. Santana-Mancilla PC, Anido-Rifón LE. The technology acceptance of a TV platform for the elderly living alone or in public nursing homes. Int J Environ Res Public Health 2017 Jun 08;14(6):617 [FREE Full text] [doi: 10.3390/ijerph14060617] [Medline: 28594386]
16. Zamir S, Hennessy CH, Taylor AH, Jones RB. Video-calls to reduce loneliness and social isolation within care environments for older people: an implementation study using collaborative action research. BMC Geriatr 2018 Mar 02;18(1):62 [FREE Full text] [doi: 10.1186/s12877-018-0746-y] [Medline: 29499659]
17. Naudé B, Rigaud AS, Kamali L, Pino M. Barriers and enablers for the use of digital interactive television in nursing home settings: an interview case study with older adults and professionals. Int J Environ Res Public Health 2023 Jan 18;20(3):1813 [FREE Full text] [doi: 10.3390/ijerph20031813] [Medline: 36767181]
18. Zamir S, Hennessy C, Taylor A, Jones R. Intergroup 'Skype' quiz sessions in care homes to reduce loneliness and social isolation in older people. Geriatrics (Basel) 2020 Nov 11;5(4):90 [FREE Full text] [doi: 10.3390/geriatrics5040090] [Medline: 33187242]
19. Tsai HH, Tsai YF. Changes in depressive symptoms, social support, and loneliness over 1 year after a minimum 3-month videoconference program for older nursing home residents. J Med Internet Res 2011 Nov 15;13(4):e93 [FREE Full text] [doi: 10.2196/jmir.1678] [Medline: 22086660]
20. Niemelä‐Nyrhinen J. Baby boom consumers and technology: shooting down stereotypes. J Consum Mark 2007 Aug 07;24(5):305-312. [doi: 10.1108/07363760710773120]
21. Lee C, Coughlin JF. PERSPECTIVE: older adults' adoption of technology: an integrated approach to identifying determinants and barriers. J Prod Innov Manag 2014 Jun 03;32(5):747-759. [doi: 10.1111/jpim.12176]
22. Mick DG, Fournier S. Paradoxes of technology: consumer cognizance, emotions, and coping strategies. J Consum Res 1998;25(2):123-143. [doi: 10.1086/209531]
23. Wilson-Nash C, Tinson J. 'I am the master of my fate': digital technology paradoxes and the coping strategies of older consumers. J Mark Manag 2021 Jul 09;38(3-4):248-278. [doi: 10.1080/0267257x.2021.1945662]
24. McCreadie C, Tinker A. The acceptability of assistive technology to older people. Ageing Soc 2005 Jan 10;25(01):91-110. [doi: 10.1017/S0144686X0400248X]
25. Peacock SE, Künemund H. Senior citizens and Internet technology: reasons and correlates of access versus non-access in a European comparative perspective. Eur J Ageing 2007 Dec 26;4(4):191-200 [FREE Full text] [doi: 10.1007/s10433-007-0067-z] [Medline: 28794788]
26. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
27. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
28. Venkatesh V, Thong JY, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157-178. [doi: 10.2307/41410412]
29. Venkatesh V, Thong JY, Xu X. Unified theory of acceptance and use of technology: a synthesis and the road ahead. J Assoc Inf Syst 2016 May;17(5):328-376. [doi: 10.17705/1jais.00428]
30. Zaltman G, Wallendorf M. Consumer Behavior: Basic Findings and Management Implications. Hoboken, NJ: John Wiley & Sons; 1979.
31. Heider F. The Psychology of Interpersonal Relations. Hillsdale, NJ: Lawrence Erlbaum Associates; 1958.
32. Osgood CE, Tannenbaum PH. The principle of congruity in the prediction of attitude change. Psychol Rev 1955;62(1):42-55. [doi: 10.1037/h0048153]
33. Ram S. A model of innovation resistance. Adv Consum Res 1987;14(1):208-212 [FREE Full text]
34. Psychology of innovation resistance: the less developed concept (LDC) in diffusion research. Jagdish Sheth. 1981 Jan 1. URL: https://www.jagsheth.com/marketing-research/psychology-of-innovation-resistance-the-less-developed-concept-ldc-in-diffusion-research/ [accessed 2024-01-19]

35. Joseph RC. Individual resistance to IT innovations. Commun ACM 2010 Apr;53(4):144-146. [doi: 10.1145/1721654.1721693]
36. Im H, Jung J, Kim Y, Shin DH. Factors affecting resistance and intention to use the smart TV. J Media Bus Stud 2015 Mar 16;11(3):23-42. [doi: 10.1080/16522354.2014.11073583]
37. Lee S. An integrated adoption model for e-books in a mobile environment: evidence from South Korea. Telemat Inform 2013 May;30(2):165-176. [doi: 10.1016/j.tele.2012.01.006]
38. Wang CH, Wu CL. Bridging the digital divide: the smart TV as a platform for digital literacy among the elderly. Behav Inform Technol 2021 Jun 21;41(12):2546-2559. [doi: 10.1080/0144929x.2021.1934732]
39. Bandura A. Self Efficacy: The Exercise of Control. New York, NY: Worth Publishers; 1997.
40. Cotten SR, Yost EA, Berkowsky RW, Winstead V, Anderson WA. Designing Technology Training for Older Adults in Continuing Care Retirement Communities. Boca Raton, FL: CRC Press; 2016.
41. Abras C, Maloney-Krichmar D, Preece J. User-centered design. In: Bainbridge W, editor. Encyclopedia of Human-Computer Interaction. Thousand Oaks, CA: Sage Publications; 2004.
42. Digital tools for seniors for care facility administrators. E-lio. URL: https://www.e-lio.fr/?lang=en [accessed 2022-09-22]
43. George Saadé R, Kira D. The emotional state of technology acceptance. Issues Informing Sci Inf Technol 2006;3:529-539. [doi: 10.28945/913]
44. Yu-Huei C, Ja-Shen C, Ming-Chao W. Why do older adults use wearable devices: a case study adopting the senior technology acceptance model (STAM). In: Proceedings of the Portland International Conference on Management of Engineering and Technology. 2019 Presented at: PICMET 2019; August 25-29, 2019; Portland, OR. [doi: 10.23919/picmet.2019.8893767]
45. Guner H, Acarturk C. The use and acceptance of ICT by senior citizens: a comparison of technology acceptance model (TAM) for elderly and young adults. Univ Access Inf Soc 2018 Dec 15;19(2):311-330. [doi: 10.1007/s10209-018-0642-4]
46. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
47. Dierckx de Casterlé B, Gastmans C, Bryon E, Denier Y. QUAGOL: a guide for qualitative data analysis. Int J Nurs Stud 2012 Mar;49(3):360-371 [FREE Full text] [doi: 10.1016/j.ijnurstu.2011.09.012] [Medline: 21996649]
48. Bae H, Jo SH, Lee E. Why do older consumers avoid innovative products and services? J Serv Mark 2020 Jul 03;35(1):41-53. [doi: 10.1108/jsm-10-2019-0408]
49. Petrovčič A, Slavec A, Dolničar V. The ten shades of silver: segmentation of older adults in the mobile phone market. Int J Hum Comput Interact 2017 Nov 13;34(9):845-860. [doi: 10.1080/10447318.2017.1399328]
50. Ayres L, Kavanaugh K, Knafl KA. Within-case and across-case approaches to qualitative data analysis. Qual Health Res 2003 Jul 01;13(6):871-883. [doi: 10.1177/1049732303013006008] [Medline: 12891720]
51. Mansvelt J, Elms J, Dodds S. Connecting meanings of ageing, consumption, and information and communication technologies through practice. Geogr Res 2020 May 29;58(3):289-299. [doi: 10.1111/1745-5871.12392]
52. Kim MJ, Preis MW. Why seniors use mobile devices: applying an extended model of goal-directed behavior. J Travel Tour Mark 2015 Aug 24;33(3):404-423. [doi: 10.1080/10548408.2015.1064058]
53. Chiu CJ, Hu YH, Lin DC, Chang FY, Chang CS, Lai CF. The attitudes, impact, and learning needs of older adults using apps on touchscreen mobile devices: results from a pilot study. Comput Hum Behav 2016 Oct;63:189-197. [doi: 10.1016/j.chb.2016.05.020]
54. Alvseike B. Feasibility of the iPad as a hub for smart house technology in the elderly; effects of cognition, self-efficacy, and technology experience. J Multidiscip Healthc 2012 Nov;5:299-306 [FREE Full text] [doi: 10.2147/jmdh.s35344]
55. Arning K, Ziefle M. Understanding age differences in PDA acceptance and performance. Comput Hum Behav 2007 Nov;23(6):2904-2927. [doi: 10.1016/j.chb.2006.06.005]
56. Cobo CM. The influence of institutionalization on the perception of autonomy and quality of life in old people. Rev Esc Enferm USP 2014 Dec;48(6):1013-1019. [doi: 10.1590/s0080-623420140000700008]
57. Langer EJ, Rodin J. The effects of choice and enhanced personal responsibility for the aged: a field experiment in an institutional setting. J Pers Soc Psychol 1976 Aug;34(2):191-198. [doi: 10.1037//0022-3514.34.2.191] [Medline: 1011073]
58. Rodin J, Langer EJ. Long-term effects of a control-relevant intervention with the institutionalized aged. J Pers Soc Psychol 1977 Dec;35(12):897-902. [doi: 10.1037//0022-3514.35.12.897] [Medline: 592095]
59. Ryan RM, Deci EL. Intrinsic and extrinsic motivations: classic definitions and new directions. Contemp Educ Psychol 2000 Jan;25(1):54-67. [doi: 10.1006/ceps.1999.1020] [Medline: 10620381]
60. Zhang S, Zhao J, Tan W. Extending TAM for online learning systems: an intrinsic motivation perspective. Tinshhua Sci Technol 2008 Jun;13(3):312-317. [doi: 10.1016/s1007-0214(08)70050-6]
61. Venkatesh V, Speier C, Morris MG. User acceptance enablers in individual decision making about technology: toward an integrated model. Decis Sci 2007 Jun 07;33(2):297-316. [doi: 10.1111/j.1540-5915.2002.tb01646.x]
62. Franz R, Neves BB. Usability is ageless: conducting usability tests with older adults. In: Neves B, Vetere F, editors. Ageing and Digital Technology. Singapore, Singapore: Springer; 2019.
63. Dickinson A, Arnott J, Prior S. Methods for human – computer interaction research with older people. Behav Inf Technol 2007 Jul;26(4):343-352. [doi: 10.1080/01449290601176948]

64. Barbosa Neves B, Franz R, Judges R, Beermann C, Baecker R. Can digital technology enhance social connectedness among older adults? A feasibility study. J Appl Gerontol 2019 Jan 22;38(1):49-72. [doi: 10.1177/0733464817741369] [Medline: 29166818]
65. van den Haak MJ, de Jong MD, Schellens PJ. Evaluation of an informational web site: three variants of the think-aloud method compared. Tech Commun 2007;54(1):58-71.

Abbreviations

e-TAM: extended technology acceptance model
FC: facilitating condition
ICT: information and communication technology
iTV: interactive television
IU: intention to use
NH: nursing home
OA: older adult
PEOU: perceived ease of use
PU: perceived usefulness
RF: residential facility
TAM: technology acceptance model
UC: user characteristic
UR: user resistance
UTAUT: unified theory of acceptability and use of technology

Edited by Y Gong; submitted 04.03.24; peer-reviewed by P Iranmanesh, R Kaur, K Mouloudj; comments to author 15.08.24; revised version received 25.09.24; accepted 23.10.24; published 02.12.24.

Please cite as:
Naudé B, Rigaud AS, Pino M
Investigating the Acceptability of an Interactive Television Intervention Promoting Social Links Among Older Adults Living at Home and in Care Institutions: Qualitative Interview and Questionnaire Study
JMIR Hum Factors 2024;11:e58051
URL: https://humanfactors.jmir.org/2024/1/e58051
doi:10.2196/58051
PMID:

Users' Perceived Service Quality of National Telemedicine Services During the COVID-19 Pandemic in Bangladesh: Cross-Sectional Study

Fatema Khatun[1], MBBS, MSc, PhD; Novel Chandra Das[1], MSc; Md Rakibul Hoque[2], BBA, MBA, MSc, PhD; Kazi Nazmus Saqeeb[3], MBBS; Monjur Rahman[4], MSc; Kyung Ryul Park[5], BA, BSc, MA, PhD; Sabrina Rasheed[1], BSc, PGDip, PhD; Daniel D Reidpath[6], BA, DipEd, PhD

[1]Health Systems and Population Studies Division, International Centre for Diarrheal Disease Research, Bangladesh (icddr,b), 68, Shaheed Tajuddin Ahmed Sharani, Mohakhali, Dhaka, Bangladesh

[2]Department of Management Information Systems, University of Dhaka, Dhaka, Bangladesh

[3]Nutrition Research Division, International Centre for Diarrheal Disease Research, Bangladesh (icddr,b), Dhaka, BangladeshBangladesh

[4]Maternal and Child Health Division, International Centre for Diarrheal Disease Research, Bangladesh (icddr,b), Dhaka, BangladeshBangladesh

[5]Graduate School of Science and Technology Policy, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea

[6]The Institute for Global Health and Development, Queen Margaret University, Edinburgh, United Kingdom

Corresponding Author:
Fatema Khatun, MBBS, MSc, PhD
Health Systems and Population Studies Division, International Centre for Diarrheal Disease Research, Bangladesh (icddr,b), 68, Shaheed Tajuddin Ahmed Sharani, Mohakhali, Dhaka, Bangladesh

Abstract

Background: COVID-19 created an opportunity for using teleconsultation as an alternative way of accessing expert medical advice. Bangladesh has seen a 20-fold increase in the use of teleconsultation during the pandemic.

Objective: The aim of our study was to assess the influence of service quality and user satisfaction on the intention to use teleconsultation in the future among users of national teleconsultation services during the pandemic.

Methods: A cross-sectional survey was conducted in 2020 among users of the national teleconsultation service—Shastho Batayon for acute respiratory infection. A validated mobile health service quality model based on structural equation modeling and confirmatory factor analysis was used to analyze the data with SmartPLS (version 3.0).

Results: Among the 2097 study participants, 1646 (78.5%) were male, 1416 (67.5%) were aged 18 - 39 years, 1588 (75.7%) were urban residents, 1348 (64.2%) had more than 10 years of schooling, and 1657 (79%) were from middle-income households. From a consumer perspective, the quality of the service platform (β=.946), service interaction (β=.974), and outcome (β=.955) contributed to service quality. Service quality was positively associated with user satisfaction (β=.327; *P*<.001) and intention to use teleconsultation services (β=.102; *P*<.001). User satisfaction was positively associated with the intention to use teleconsultation services (β=.311; *P*<.001).

Conclusions: The increase in the use of teleconsultation during the pandemic indicated that such services were potentially used for emergencies. However, the future use of teleconsultation will be dependent on the quality of service and user satisfaction. Our findings are relevant for low-income contexts where teleconsultation services are used to address gaps in service delivery.

(JMIR Hum Factors 2024;11:e46566) doi:10.2196/46566

KEYWORDS

telemedicine; COVID-19; LMIC; low- and middle-income countries; Bangladesh; service quality; user satisfaction; structural equation modeling; digital health

Introduction

Background

Globally, 58% of countries have adopted telemedicine to address the disruption of health care services during the COVID-19 pandemic [1]. Telemedicine facilitates the delivery of remote health care services at low cost [2] and has shown to enable quick and equitable access to health care for patients in hard-to-reach areas [3]. Telemedicine has been effectively used to provide health services during previous SARS-CoV-2 and Middle East respiratory syndrome outbreaks [4], and during the COVID-19 pandemic, the use of telemedicine has increased globally to minimize the risks of viral transmission [5,6].

Researchers around the world reported widespread use of telemedicine in many low- and middle-income countries (LMICs) where the provision of protective gear, vaccines, and hospital facilities was inadequate during the pandemic [7-13].

In Bangladesh, there was severe disruption of primary health care provision during the early stage of the pandemic, as health systems responded to lockdown measures, deaths among health care providers, and the need to mitigate and limit the spread of COVID-19 [14]. One of the response measures by the government of Bangladesh was to make the existing national teleconsultation platform “Shastho Batayon 16263” toll-free [15]. As travel restriction was in place and access to face-to-face consultation was limited, calls to the national teleconsultation platform increased 20-fold during the pandemic [16]. However, there were concerns about the negative impact of the sudden increase in demand for teleconsultation services on the existing infrastructure and human resources of Shastho Batayon (SB) [17] and whether any compromise in terms of service quality of the platform would negatively impact patient satisfaction and intention to avail teleconsultation services in the future [18].

Although the promises are offered by telemedicine in low-resource settings, ensuring the quality of telemedicine services is a major challenge [19]. Despite the challenges of quality, however, there are benefits of using telemedicine during a pandemic such as COVID-19 [20], and it is likely that many patients and providers will continue to use this service in the postpandemic period [21]. Hence, it is important to understand how the service quality of telemedicine affects patient satisfaction and intention to use the service in the future [22].

The World Health Organization’s guiding principles for implementing telemedicine services during COVID-19 listed “user satisfaction” as one of the key components for evaluation [23]. In health care, service quality is a major indicator of patient satisfaction [24-27]. Satisfaction is associated with increased user retention and intention to use the service in the future [24,25,28].

There are a number of theories and models that address quality dimensions of information and communication technology–based services. A systematic review of the service quality models identified 30 models, which described that the most common dimensions were tangibles, reliability, responsiveness, empathy, and assurance [29]. In a developing country context, the mobile health (mHealth) service quality model tested by Akter et al [30] measures the association between service quality and satisfaction on the intention to use the service in the future. This conceptual model describes platform quality, interaction quality, and outcome quality as primary dimensions of service quality for mHealth and proposes a direct association between service quality and patient satisfaction and an indirect association between service quality and intention to continue to use the service through patient satisfaction [30]. The conceptual model has been previously validated for Bangladesh. Using the model for mHealth service quality, we aimed to assess the influence of service quality and satisfaction on the intention to use SB services in the future among the users who availed the service during the pandemic. The insights from this study will significantly contribute to the sustainability of national teleconsultation services during future health emergency situations and beyond.

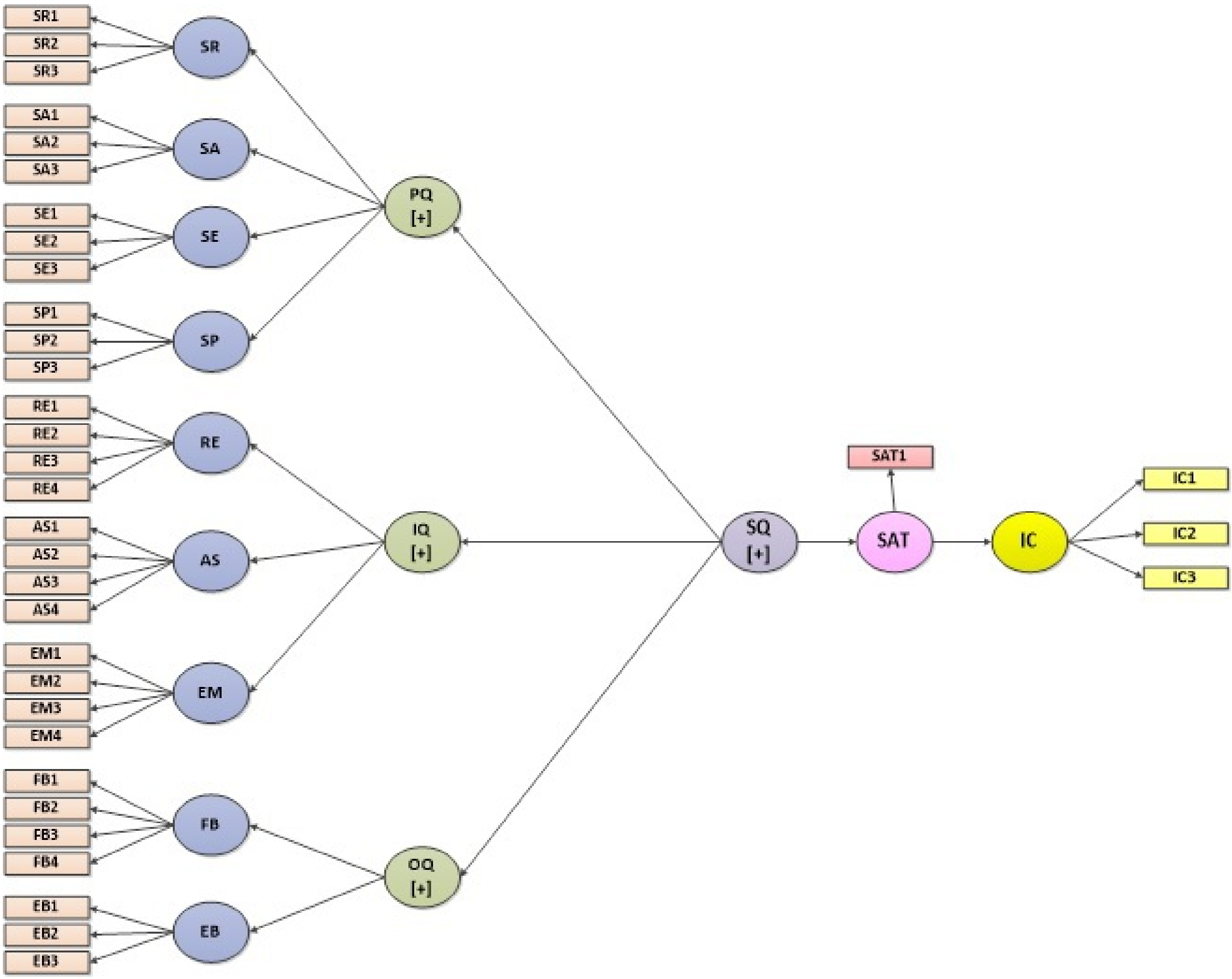
Theoretical Framework

For this study, we used the mHealth service quality model [30] to design the study. A number of empirical studies have used the mHealth service quality model to describe the association between service quality and users’ intention to continue using eHealth services [24,30-32], but the model by Akter et al [30] is chosen, as it was used to teleconsultation in the context of LMICs. The conceptual framework and the hypotheses are detailed in the following subsections.

Conceptual Model: A Hierarchical, Multidimensional Service Quality Model

The conceptual model (Figure 1) of service quality used in this study postulates a direct relationship between service quality and satisfaction and that service quality indirectly works on intention to use, mediated by satisfaction [33-36]. In the service quality model, service quality comprises 3 primary dimensions (platform quality, interaction quality, and outcome quality) and 9 subdimensions. The dimension platform quality has systems reliability, systems availability, systems efficiency, and systems privacy as subdimensions; the interaction quality dimension has responsiveness, assurance, and empathy as subdimensions; and the outcome quality dimension has functional benefit and emotional benefit as subdimensions.

Figure 1. Conceptual framework of service quality conceptual model for national teleconsultation services users during the COVID-19 pandemic in Bangladesh, October-December 2020 (N=2097). AS: assurance; EB: emotional benefit; EM: empathy; FB: functional benefit; IC: intension; IQ: interaction quality; OQ: outcome quality; PQ: platform quality; RE: responsiveness; SA: systems availability; SAT: satisfaction; SE: systems efficiency; SP: systems privacy; SQ: service quality; SR: systems reliability.



Definition of the Constructs

The definitions of constructs such as platform quality, interaction quality, outcome quality, user satisfaction, and intention to continue are detailed in Multimedia Appendix 1.

Hypotheses

In this study, we tested 3 hypotheses:

- Hypothesis 1: SB service quality has a positive impact on user satisfaction.
- Hypothesis 2: User satisfaction with SB is positively associated with the intention to use the services in the future.
- Hypothesis 3: SB service quality is positively associated with the intention to use the service in the future.

Methods

Study Settings and Participants: SB

The Ministry of Family and Welfare of Bangladesh, in an effort to facilitate access to health services, runs a national health helpline with SB through its Management Information System division. The service is available 24/7, and callers can consult doctors and get limited prescriptions and referrals. During the COVID-19 pandemic, the SB number was made toll-free.

We conducted this study among the users of SB from October 19, 2020, to December 31, 2020. Our study participants were those who called SB with their flu-like syndromes and were provided a provisional diagnosis of suspected or confirmed COVID-19. We approached the participants from the random list collected from the Management Information System database who sought teleconsultation services from any geographical area of Bangladesh and were interested to join the study, and those participants were enrolled as study participants. We were provided with a total list of 7400 people, with batches of 100 participants sent out each day from SB, and finally, 2097 participants were enrolled in the study. All the interviews were conducted over the phone.

Sample Size

To use structural equation modeling (SEM), the number of constructs (latent variables), number of observed variables, and effect size of the loading coefficient with a 95% confidence level and 80% power were considered with a view to calculating the required sample size through Enrico multivariate software

[37]. To identify the effect size of the loading coefficient of 0.15, 35 observed variables, and 15 latent variables with 95% confidence level and 80% power, the minimum required sample size was 378. Thus, the study fulfills the requirement of the necessary sample size to apply SEM by collecting information from 2097 participants.

Data Collection

Within 2 weeks of seeking care from SB by the selected caller, we administered this satisfaction survey about their call to SB. We collected a list of 100 callers (randomly selected) every day based on our selection criteria (suspected or confirmed COVID-19 cases) from October 19, 2020, to December 31, 2020.

The list was then divided and distributed among the data collectors for conducting a telephone survey. Data were collected using a questionnaire previously validated in Bangladesh [32]. All the data were collected through a web-based Android-based data collection app or tool by trained data collectors. In total, 5% of the interviews were randomly rechecked by the supervisor to ensure the data quality and validity of the interviews.

Ethical Considerations

The study was approved by the institutional review board of the International Centre for Diarrheal Disease Research, Bangladesh (protocol: PR#19091). Informed verbal consent was taken from all the study participants over the phone. In total, 31 study participants' age was below 18 years so we took informed assent from their guardians over the phone to participate in the study. The privacy, anonymity, and confidentiality of the research data and information were strictly maintained.

Variables

The questionnaire included questions on user demographics (age, sex, socioeconomic status, and place of residence). To understand users' perception of service quality, the questionnaire included questions related to systems reliability, availability, efficiency, privacy, responsiveness, assurance, empathy, functional, emotional benefit, and satisfaction along with the intention to continue using the service in the future. A 5-point Likert scale with a range from 1=strongly disagree to 5=strongly agree was used to measure the responses. The statements used in the questionnaire were previously validated in Bangladesh [32]. The questionnaire is available in Multimedia Appendix 2.

Normality Check

The assessment of data normality was conducted through the estimation of skewness and kurtosis for each measurement item. The skewness values ranged from −0.319 to 0.069, while kurtosis values ranged from −0.543 to 0.671, where values within the range of ±2 for skewness and ±5 for kurtosis are deemed acceptable indicators of normal distribution [38], our dataset exhibited no substantive issues with skewness or kurtosis.

Data Analysis

From the list of 7400 individual users of SB, we were able to reach only 2163 due to incorrect numbers, switched off phones, unavailable networks, use of neighbors' or relatives' phone numbers for SB services, and because callers did not receive our calls. The initial response rate was 29.2%. Of those we were able to reach, 3.9% (n=66) of respondents refused to participate in the study so our final sample was 2097.

SmartPLS (version 3.0) was used to analyze the data. Component-based SEM or partial least squares (PLS) path modeling was used to understand the direct effect between service quality and satisfaction and satisfaction and intention and the indirect effect between service quality and intention. Confirmatory factor analysis (CFA) was used to test the hypotheses. First-order CFA was used to test the validity and reliability of measurement items of the constructs. Second- and third-order CFA were used in the measurement model because there are latent variables that were constructed as dimensions of another variable. To assess the second-order reflective model of service quality, this study used PLS Graph (version 3.0; SmartPLS GmbH) [39] and PLS path modeling with a path weighing scheme for the inside approximation [40-42].

Results

Sociodemographic Characteristics of Study Participants

Among the 2097 study participants, the majority were from urban areas (n=1588, 75.7%), had more than 10 years of schooling (n=1348, 64.3%), and were from middle-income households (n=1657, 79%; Table 1). Most of the service recipients were male (n=1646, 78.5%), 67.5% (n=1416) were between 18 and 39 years of age, 63.2% (n=1325) were married, and 37.3% (n=782) had more than 4 members in their household (Table 1).

Table . Sociodemographic characteristics of the study participants of Shastho Batayon teleconsultation services users during the COVID-19 pandemic in Bangladesh, October-December 2020 (N=2097).

Characteristics		Values, n (%)
Age (years)		
	<18	31 (1.5)
	18 - 39	1416 (67.5)
	40 - 59	579 (27.6)
	≥60	71 (3.4)
Sex		
	Male	1646 (78.5)
	Female	451 (21.5)
Residence		
	Village	290 (13.8)
	Semiurban	219 (10.4)
	Urban	1588 (75.7)
Education (years)		
	≤5	331 (15.8)
	6 - 10	418 (19.9)
	>10	1348 (64.3)
Marital status		
	Unmarried	732 (34.5)
	Married	1325 (63.2)
	Widowed	37 (1.7)
	Divorced	3 (0.1)
Occupation		
	Business	263 (12.5)
	Unskilled labor	18 (0.86)
	Skilled labor	30 (1.4)
	Service holder	1035 (49.3)
	Agriculture worker	33 (1.6)
	Student	252 (12)
	Housewife	255 (12.2)
	Unemployed	118 (5.6)
	Others	93 (4.4)
Family member		
	≤4	1315 (62.7)
	>4	782 (37.3)
Monthly household expenditure (US $)		
	>500	182 (8.7)
	50 - 500	1657 (79)
	<50	258 (12.3)

Assessment of the Measurement Model

The study assesses the psychometric properties of the first-order measurement model by examining reliability, convergent validity (Table 2), and discriminant validity (Table 3). Cronbach α, composite reliabilities, and average variance extracted from the data exceeded the cutoff values of 0.5, 0.7, and 0.5, respectively (Table 2), which indicate adequate scale reliability and validity [40,43]. The model was considered satisfactory in terms of reliability, convergent validity, and discriminant validity. The result showed that the composite reliabilities and average variance extracted from the second- and third-order scales were greater than 0.8 and 0.5, respectively, which indicated that the higher-order measures were reliable (Table 2).

Table . Confirmatory factor analysis and psychometric properties of the hierarchical service quality scale[a].

Factor and items	Loadings	Cronbach α	First-order constructs		Second-order constructs			Third-order constructs		
			CR[b]	AVE[c]	Constructs	CR	AVE	Constructs	CR	AVE
Systems reliability (SR)		0.950	0.969	0.911	Platform quality	0.956	0.647	Service quality	0.985	0.685
SR1	0.942									
SR2	0.971									
SR3	0.951									
Systems availability (SA)		0.872	0.922	0.797	—[d]	—	—	—	—	—
SA1	0.908									
SA2	0.927									
SA3	0.841									
Systems efficiency (SE)		0.90	0.938	0.834	—	—	—	—	—	—
SE1	0.901									
SE2	0.932									
SE3	0.906									
Systems privacy (SP)		0.923	0.951	0.867	—	—	—	—	—	—
SP1	0.946									
SP2	0.943									
SP3	0.903									
Responsiveness (RE)		0.920	0.943	0.806	Interaction quality	0.975	0.768	—	—	—
RE1	0.901									
RE2	0.909									
RE3	0.892									
RE4	0.890									
Assurance (AS)		0.936	0.954	0.839	—	—	—	—	—	—
AS1	0.903									
AS2	0.916									
AS3	0.926									
AS4	0.918									
Empathy (EM)		0.952	0.966	0.875	—	—	—	—	—	—
EM1	0.946									
EM2	0.957									
EM3	0.940									
EM4	0.899									
Functional benefit (FB)		0.953	0.966	0.876	Outcome quality	0.979	0.869	—	—	—
FB1	0.926									
FB2	0.933									
FB3	0.942									
FB4	0.943									
Emotional benefit (EB)		0.961	0.975	0.928	—	—	—	—	—	—

Factor and items		Loadings	Cronbach α	First-order constructs		Second-order constructs			Third-order constructs		
				CR[b]	AVE[c]	Constructs	CR	AVE	Constructs	CR	AVE
	EB1	0.958									
	EB2	0.965									
	EB3	0.966									
Satisfaction	—	1	—	1	1	—	—	—	—	—	—
Intention to continue (IC)			—	0.951	0.866	—	—	—	—	—	—
	IC1	0.935									
	IC2	0.934									
	IC3	0.922									

[a]Convergent validity: loadings >0.70. Scale reliability: CR>0.80, AVE>0.50.
[b]CR: composite reliability.
[c]AVE: average variance extracted.
[d]Not applicable.

Table . Correlation of first-order constructs[a].

	AS[b]	EB[c]	EM[d]	FB[e]	IC[f]	RE[g]	SA[h]	SAT[i]	SE[j]	SP[k]	SR[l]
AS	0.916	—[m]	—	—	—	—	—	—	—	—	—
EB	0.852	0.963	—	—	—	—	—	—	—	—	—
EM	0.899	0.858	0.936	—	—	—	—	—	—	—	—
FB	0.895	0.935	0.907	0.936	—	—	—	—	—	—	—
IC	0.761	0.824	0.753	0.795	0.930	—	—	—	—	—	—
RE	0.860	0.793	0.856	0.834	0.709	0.898	—	—	—	—	—
SA	0.637	0.635	0.628	0.657	0.562	0.697	0.893	—	—	—	—
SAT	0.324	0.308	0.323	0.333	0.311	0.269	0.173	1.000	—	—	—
SE	0.817	0.786	0.795	0.827	0.712	0.786	0.669	0.304	0.913	—	—
SP	0.682	0.621	0.670	0.682	0.567	0.662	0.568	0.230	0.755	0.931	—
SR	0.773	0.783	0.766	0.798	0.689	0.751	0.659	0.301	0.801	0.603	0.955

[a]Discriminant validity: square root of average variance extracted on the diagonal>correlation coefficients.
[b]AS: assurance.
[c]EB: emotional benefit.
[d]EM: empathy.
[e]FB: functional benefit.
[f]IC: intension.
[g]RE: responsiveness.
[h]SA: systems availability.
[i]SAT: satisfaction.
[j]SE: systems efficiency.
[k]SP: systems privacy.
[l]SR: systems reliability.
[m]Not applicable.

Structural Equation Modeling

Our results showed that the third-order construct, service quality, was associated with the second-order constructs—platform quality (β=.946), interaction quality (β=.974), and outcome quality (β=.955), which explained 90%, 95%, and 91% of the overall quality variance, respectively (Figure 2). The second-order construct, outcome quality, was significantly associated with their first-order constructs—functional benefit (β=.987) and emotional benefit (β=.979; Figure 2). All the path coefficients from service quality to second-order and third-order components were significant at *P*<.001 (Figure 3). Therefore, we found that the 31 items, grouped into 9 factors, can be used to measure the overall service quality of SB teleconsultation services.

Figure 2. Path coefficients of the research model for national teleconsultation services users during the COVID-19 pandemic in Bangladesh, October-December 2020 (N=2097). AS: assurance; EB: emotional benefit; EM: empathy; FB: functional benefit; IC: intension; IQ: interaction quality; OQ: outcome quality; PQ: platform quality; RE: responsiveness; SA: systems availability; SAT: satisfaction; SE: systems efficiency; SP: systems privacy; SQ: service quality; SR: systems reliability.

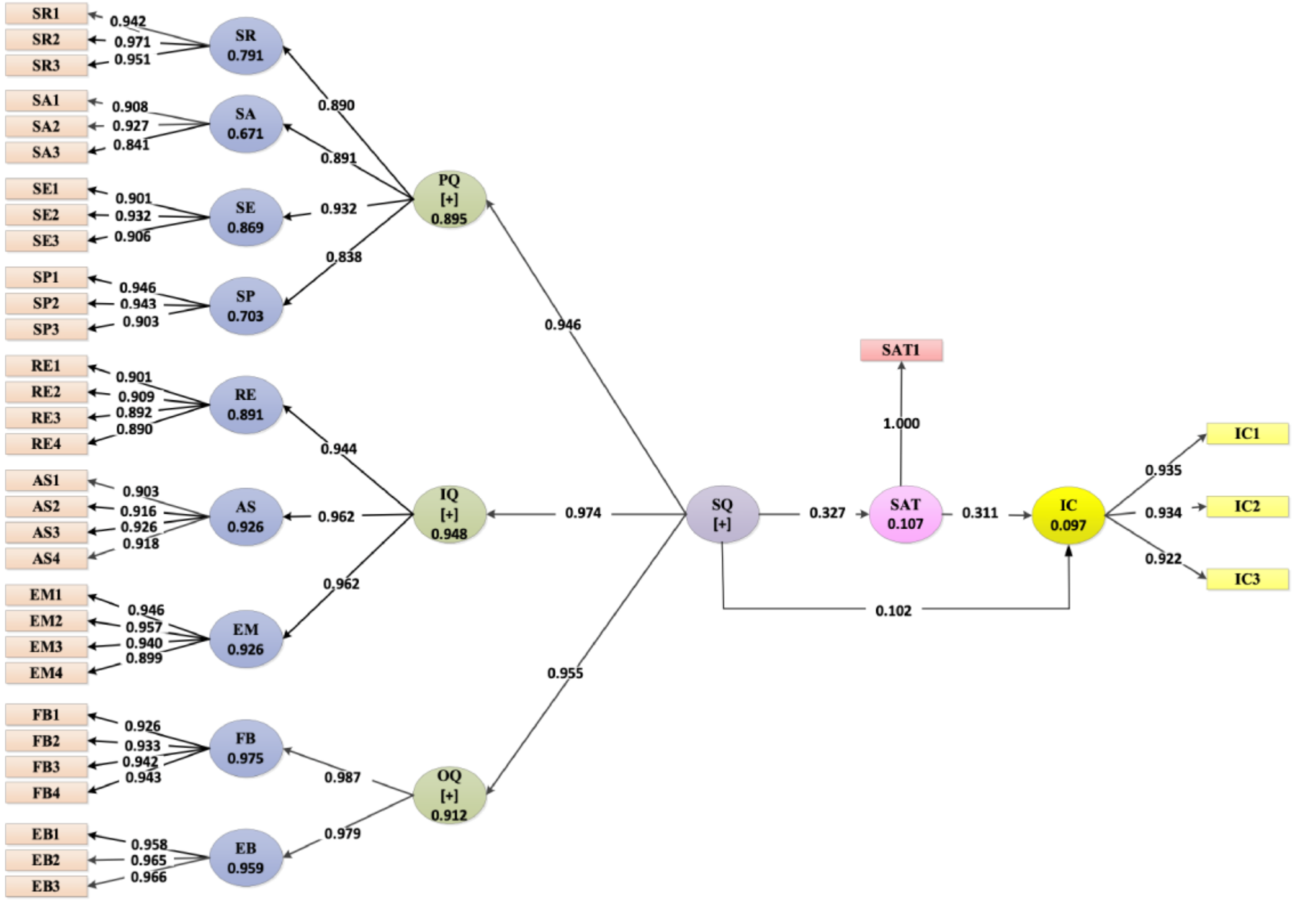
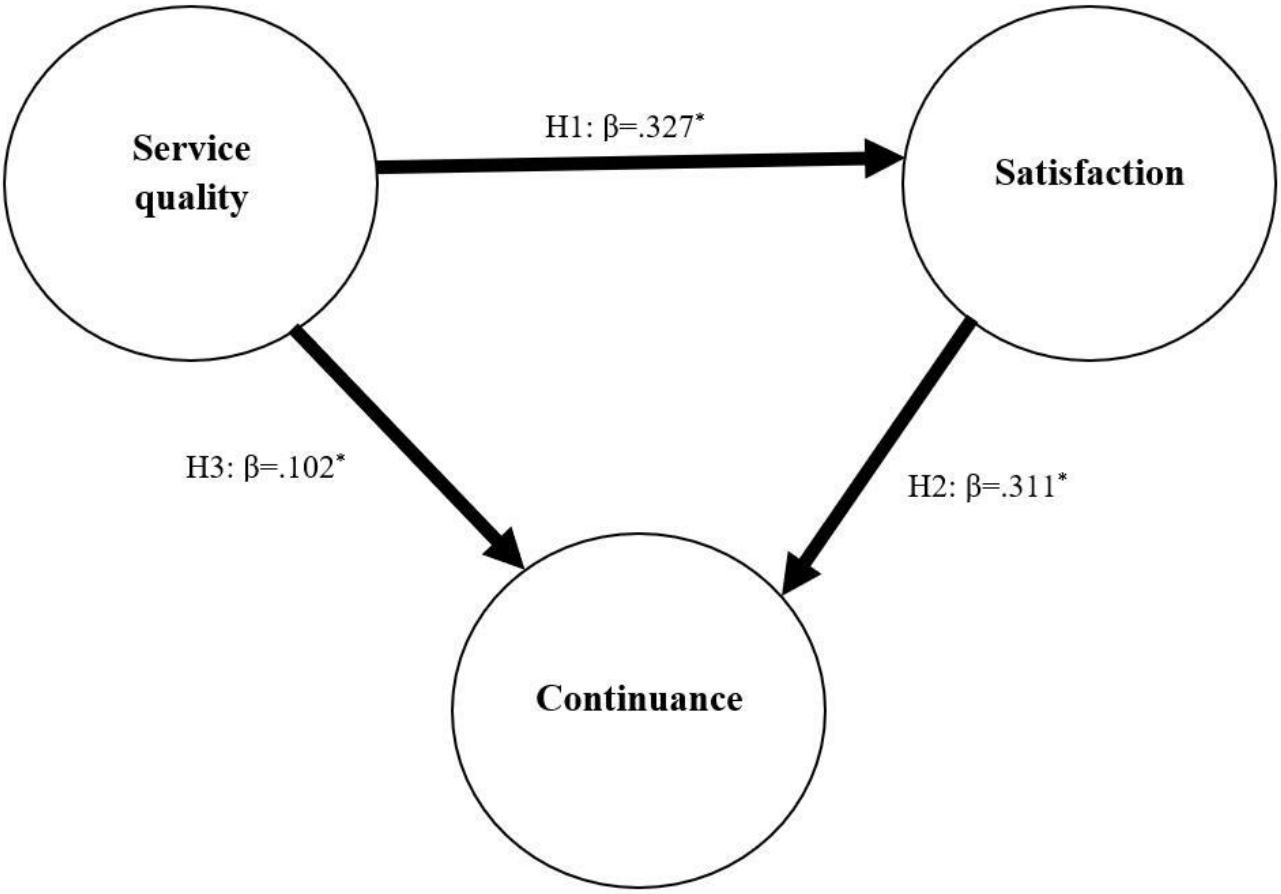


Figure 3. Testing hypotheses regarding the association between Shastho Batayon service quality, user satisfaction, and continuance intention, October-December 2020 (N=2097). H: hypothesis. *P<.001.



Hypotheses Testing

- Hypothesis 1: The β from service quality to satisfaction was .327, indicating that SB service quality was positively associated with user satisfaction.
- Hypothesis 2: The β from satisfaction to intention to continue use was .311, indicating that user satisfaction with SB was positively associated with the intention to use the services in the future.
- Hypothesis 3: The β from service quality to continuance intention was .102, indicating that SB service quality was associated with the intention to use the service in the future.

Discussion

Principal Findings

Our study describes the use of toll-free national teleconsultation services (SB) during the COVID-19 pandemic using a service quality model validated in Bangladesh [16]. In our analysis, we assessed the impact of service quality on user satisfaction and intention to use SB services in the future. During the pandemic, with the rapid increase in the number of calls to SB, the service provision had to be scaled up rapidly to cater to the needs of the population. However, our study indicates that teleconsultation services such as SB have to pay attention to service quality if they want the service recipients to continue to avail the services beyond the pandemic. In the context of Bangladesh, researchers have described [44,45] factors such as usefulness, perceived reliability, price value, technology anxiety, expectations, performance, disconfirmation, and enjoyment influencing the adoption, satisfaction, and continuance of mHealth and telemedicine services in Bangladesh [44,46-49], although in most cases, the studies did not use validated scales for service quality. In most cases, the mHealth interventions studied were small-scale or were being piloted among both rural and urban populations. Only 1 study was conducted to conceptualize and validate service quality scales [31]. To our knowledge, this is the first study to assess the association between service quality and both user satisfaction and intention to use the service in the future for national teleconsultation service.

The service quality of SB was positively associated with user satisfaction. This finding emphasizes the importance of looking at the SB service from the viewpoint of “systems as service” as described by earlier researchers [30]. In terms of the 3 subdomains of service quality, such as platform quality, interaction quality, and outcome quality, SB services were perceived to provide the needed health care during the COVID-19 pandemic. Although in other studies, the price of the service was an important determinant of user satisfaction along with service quality [50], calls to SB were made toll-free during the pandemic period, which could have contributed to both increase in the number of calls [51] and user satisfaction. It is important that the insights related to the rapid scaling up of the SB services while maintaining service quality during emergencies are systematically recorded and included in the standard operating procedures in the future.

We found that service quality and user satisfaction were important determinants of the intention to use SB services in the future. Similar findings about the association between telemedicine service quality and user satisfaction on intention to use the services have been reported from both developed and developing countries [52,53]. Whether users will continue to use SB services for their future health care needs after the COVID-19 pandemic will depend on both service quality and user satisfaction. A sound technological platform is a necessity for providing reliable teleconsultation services. However, service quality also depends on the quality of interaction between the service providers and the users. Interaction quality could be improved by adequately training the service providers to respond to users' needs with empathy and assurance [31]. Another issue to consider is the specific nature of the COVID-19 pandemic, where normal health care delivery was disrupted [54], and it was not easy to reach trained medical professionals [55]. During the pandemic, a protocolized management of COVID-19 was approved nationally [56] and was implemented through SB. However, in normal circumstances, people might prefer reaching out to trusted health facilities and doctors, as trust between provider and patient has been shown to be a crucial component of adherence to treatment [57,58]. In Bangladesh, previous studies have shown that preference for trusted health care professionals and lack of trust on telemedicine doctors created a significant barrier in the use of teleconsultation.

Though telemedicine services have improved and technological development has risen in recent years, SB services are only limited to teleconsultation and e-prescription [51]. In addition, prescriptions provided by telemedicine are limited to adhere with health regulations [59]. Despite its limitations, SB has proven to be an important source of health information during emergency situations where normal health care delivery is disrupted [60]. The low-cost access to professional medical advice provided has the potential to bolster health care provision in the future. It is important, therefore, that the SB services are monitored to ensure the quality of service provision and address the gaps. It is also important to think through how best to optimally use SB services to enhance health care access and use during nonpandemic times. In this regard, it is important to address the existing digital divide that is apparent in the use of SB. Most of the callers of SB were better educated, young, male, and urban residents. It will be important to address the barriers faced by population groups that do not use SB services.

Overall, the findings of our study suggest that decision makers should consider "service quality" as an important strategic objective to ensure positive satisfaction and continuance intentions. Good-quality teleconsultation services can help health providers enhance health care coverage during pandemics. Policy makers can develop a tailored regulatory framework to ensure the quality of SB services. In this regard, investments in robust telecommunication infrastructure, especially in rural and hard-to-reach areas, are important to enhance technology accessibility and connectivity. Although teleconsultation services such as SB cannot be an alternative to regular health care, there are important lessons to be learnt from SB. The SB physicians were trained thoroughly in dealing with patients with professionalism and empathy, which led to increased patient satisfaction. As patient satisfaction was an important determinant of intention to use the service in the future, there should be constant efforts made to train health care providers to prioritize patient-centered care, incorporate patient feedback, and ensure user-friendly and accessible telemedicine platforms. Finally, a commitment to research and development, including resource allocation for studies on telemedicine's impact on user satisfaction, needs to be made to foster innovation and improvement of teleconsultation services.

The strength of the study is the use of a validated service quality model to understand the impact of service quality on patient satisfaction and intention to continue using the service [30]. Another strength of this study is its focus on pandemic health care needs. The study had a few limitations: first, the context of the study is a single provider, and the study applies to the pandemic context. The findings regarding SB's acceptability may not be generalizable to a nonpandemic context. In the study, we were only able to measure the intention to continue to use the SB services rather than the actual continuance. In the future, longitudinal studies could be conducted to follow up and measure service quality and satisfaction in relationship to the actual continued use of the application.

Conclusions

This study used a validated framework to evaluate national teleconsultation services in Bangladesh during the COVID-19 pandemic. The findings of this study imply that service quality in health care is an important factor to improve user satisfaction. This study is significant in the context of Bangladesh and other LMICs, where there are human resource constraints in health care, which can be addressed through teleconsultation services [61]. The key to improving service quality and users' satisfaction is a combination of an enabling environment and infrastructure that includes a robust platform, trained workforce, data privacy and confidentiality, and proper policies and legislations [30,32]. Additionally, our study findings imply that any teleconsultation service must have a guideline for service providers to ensure that good quality service is provided through teleconsultation. While the advances in telemedicine hold immense promise for improving health services, optimal benefits can be availed if service quality is adequately evaluated and monitored.

Acknowledgments

The authors are thankful to all the participants who took part in the study. The authors thank the Synesis IT team and the Management Information System team of the Directorate General of Health Services. The authors would like to thank Dr Sifat Parveen Sheikh for all her support in the project. The Swedish International Development Cooperation Agency and Global Affairs Canada gratified the authors by providing funds for the project. The authors also thank the governments of Bangladesh, Canada, Sweden, and the United Kingdom for providing core or unrestricted support.

Data Availability

The dataset is not publicly available due to the data privacy policy of the organization. Our institutional review board has required that the personal information of the participants is not disclosed. Data related to this manuscript are available upon request to those who meet the criteria of the organization's policy from Ms Shiblee Sayeed (shiblee_s@icddrb.org) at the research administration of the International Centre for Diarrheal Disease Research, Bangladesh.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Definition of the constructs.
[DOCX File, 18 KB - humanfactors_v11i1e46566_app1.docx]

Multimedia Appendix 2
Questionnaire for teleconsultation services user.
[DOCX File, 34 KB - humanfactors_v11i1e46566_app2.docx]

References

1. COVID-19 significantly impacts health services for noncommunicable diseases. World Health Organization. URL: https://www.who.int/news/item/01-06-2020-covid-19-significantly-impacts-health-services-for-noncommunicable-diseases [accessed 2024-12-11]
2. Craig J, Patterson V. Introduction to the practice of telemedicine. J Telemed Telecare 2005;11(1):3-9. [doi: 10.1177/1357633X0501100102] [Medline: 15829036]
3. Howitt P, Darzi A, Yang GZ, et al. Technologies for global health. Lancet 2012 Aug 4;380(9840):507-535. [doi: 10.1016/S0140-6736(12)61127-1] [Medline: 22857974]
4. Ohannessian R, Duong TA, Odone A. Global telemedicine implementation and integration within health systems to fight the COVID-19 pandemic: a call to action. JMIR Public Health Surveill 2020 Apr 2;6(2):e18810. [doi: 10.2196/18810] [Medline: 32238336]
5. Ramaswamy A, Yu M, Drangsholt S, et al. Patient satisfaction with telemedicine during the COVID-19 pandemic: retrospective cohort study. J Med Internet Res 2020 Sep 9;22(9):e20786. [doi: 10.2196/20786] [Medline: 32810841]
6. Akseer N, Kandru G, Keats EC, Bhutta ZA. COVID-19 pandemic and mitigation strategies: implications for maternal and child health and nutrition. Am J Clin Nutr 2020 Aug 1;112(2):251-256. [doi: 10.1093/ajcn/nqaa171] [Medline: 32559276]
7. Kaye AD, Okeagu CN, Pham AD, et al. Economic impact of COVID-19 pandemic on healthcare facilities and systems: international perspectives. Best Pract Res Clin Anaesthesiol 2021 Oct;35(3):293-306. [doi: 10.1016/j.bpa.2020.11.009] [Medline: 34511220]
8. Haldane V, Zhang Z, Abbas RF, et al. National primary care responses to COVID-19: a rapid review of the literature. BMJ Open 2020 Dec 8;10(12):e041622. [doi: 10.1136/bmjopen-2020-041622] [Medline: 33293398]
9. Humayun A, Shahabuddin S, Afzal S, Malik AA, Atique S, Iqbal U. Healthcare strategies and initiatives about COVID19 in Pakistan: telemedicine a way to look forward. Comput Methods Programs Biomed Update 2021;1:100008. [doi: 10.1016/j.cmpbup.2021.100008] [Medline: 34337588]
10. Peiris D, Sharma M, Praveen D, Bitton A, Bresick G, Coffman M, et al. Strengthening primary health care in the COVID-19 era: a review of best practices to inform health system responses in low- and middle-income countries. WHO South East Asia J Public Health 2021;10(3):6. [doi: 10.4103/2224-3151.309867]
11. Nit B, Kobashi Y, Vory S, et al. The introduction of telemedicine is required immediately in Cambodia: barriers and lessons from COVID-19. J Glob Health 2021 Mar 27;11:03047. [doi: 10.7189/jogh.11.03047] [Medline: 33828832]
12. Chowdhury SR, Sunna TC, Ahmed S. Telemedicine is an important aspect of healthcare services amid COVID‐19 outbreak: its barriers in Bangladesh and strategies to overcome. Int J Health Plann Manage 2021 Jan;36(1):4-12. [doi: 10.1002/hpm.3064]
13. Malhotra N, Sakthivel P, Gupta N, Nischal N, Ish P. Telemedicine: a new normal in COVID era; perspective from a developing nation. Postgrad Med J 2022 Mar 1;98(e2):e79-e80. [doi: 10.1136/postgradmedj-2020-138742]
14. Bangladesh health system review. World Health Organization. 2015. URL: https://iris.who.int/bitstream/handle/10665/208214/9789290617051_eng.pdf [accessed 2024-11-29]
15. Khan MAH, Cruz VDO, Azad AK. Bangladesh’s digital health journey: reflections on a decade of quiet revolution. WHO South East Asia J Public Health 2019 Sep;8(2):71-76. [doi: 10.4103/2224-3151.264849] [Medline: 31441440]
16. Ahmed R. Bangladesh fares poorly in use of technology to fight COVID-19. Prothomalo. 2020. URL: https://en.prothomalo.com/bangladesh/digital-bangladesh-fares-poorly-in-use-of-technology-to-fight-covid-19 [accessed 2024-11-14]

17. Biswas RK, Huq S, Afiaz A, Khan HTA. A systematic assessment on COVID-19 preparedness and transition strategy in Bangladesh. J Eval Clin Pract 2020 Dec;26(6):1599-1611. [doi: 10.1111/jep.13467] [Medline: 32820856]
18. Bali S. Barriers to development of telemedicine in developing countries. In: Telehealth: IntechOpen; 2019. [doi: 10.5772/intechopen.81723]
19. Hong Z, Li N, Li D, et al. Telemedicine during the COVID-19 pandemic: experiences from Western China. J Med Internet Res 2020 May 8;22(5):e19577. [doi: 10.2196/19577] [Medline: 32349962]
20. Hoque R, Sorwar G. Understanding factors influencing the adoption of mHealth by the elderly: an extension of the UTAUT model. Int J Med Inform 2017 May;101:75-84. [doi: 10.1016/j.ijmedinf.2017.02.002] [Medline: 28347450]
21. Abdel-Wahab M, Rosenblatt E, Prajogi B, Zubizarretta E, Mikhail M. Opportunities in telemedicine, lessons learned after COVID-19 and the way into the future. Int J Radiat Oncol Biol Phys 2020 Oct 1;108(2):438-443. [doi: 10.1016/j.ijrobp.2020.07.006] [Medline: 32890528]
22. Duggirala M, Rajendran C, Anantharaman RN. Patient ‐ perceived dimensions of total quality service in healthcare. Benchmarking 2008 Aug 29;15(5):560-583. [doi: 10.1108/14635770810903150]
23. Implementing telemedicine services during COVID-19: guiding principles and considerations for a stepwise approach. World Health Organization. 2020. URL: https://www.who.int/publications/i/item/WPR-DSE-2020-032 [accessed 2024-11-29]
24. Oppong E, Hinson RE, Adeola O, Muritala O, Kosiba JP. The effect of mobile health service quality on user satisfaction and continual usage. Total Qual Manag Bus Excell 2021 Jan 2;32(1-2):177-198. [doi: 10.1080/14783363.2018.1541734]
25. Tweena SN, Oriade A, Mahdi M. An empirical investigation of the impact of service quality on satisfaction, usage behaviors, and quality of life in m-health service context: a case of Bangladesh. Glob Conf on Serv Mgmt 2021;2:374. [doi: 10.5038/2641-5062-v2]
26. Lee WI, Figueredo NM. Exploring the perspective of service quality in mHealth services during the COVID-19 pandemic. Int J Econ Manag Eng 2021;15(5):562-566.
27. Khatun F, Hanifi SMA, Iqbal M, et al. Prospects of mHealth services in Bangladesh: recent evidence from Chakaria. PLoS One 2014;9(11):e111413. [doi: 10.1371/journal.pone.0111413] [Medline: 25375255]
28. Sowon K, Chigona W. Trust in mhealth: how do maternal health clients accept and use mhealth interventions? Presented at: SAICSIT ’20: Conference of the South African Institute of Computer Scientists and Information Technologists 2020; Sep 14-16, 2020; Cape Town, South Africa. [doi: 10.1145/3410886.3410895]
29. Preaux J, Casadesús M, Bernardo M. A conceptual model to evaluate service quality of direct-to-consumer telemedicine consultation from patient perspective. Telemed J E Health 2023 Feb;29(2):156-171. [doi: 10.1089/tmj.2022.0089] [Medline: 35771956]
30. Akter S, D’Ambra J, Ray P. Service quality of mHealth platforms: development and validation of a hierarchical model using PLS. Electron Markets 2010 Dec;20(3-4):209-227. [doi: 10.1007/s12525-010-0043-x]
31. Akter S, D’Ambra J, Ray P. Development and validation of an instrument to measure user perceived service quality of mHealth. Inf Manag 2013 Jun;50(4):181-195. [doi: 10.1016/j.im.2013.03.001]
32. Akter S, D’Ambra J, Ray P. User perceived service quality of mhealth services in developing countries. Presented at: ECIS 2010 Proceedings; Jun 7-9, 2010; Pretoria, South Africa.
33. Adam Mahmood MO, Burn JM, Gemoets LA, Jacquez C. Variables affecting information technology end-user satisfaction: a meta-analysis of the empirical literature. Int J Hum Comput Stud 2000 Apr;52(4):751-771. [doi: 10.1006/ijhc.1999.0353]
34. Zviran M, Erlich Z. Measuring IS user satisfaction: review and implications. Commun Assoc Inf Syst 2003;12(1):5. [doi: 10.17705/1CAIS.01205]
35. Cronin JJJ, Taylor SA. Measuring service quality: a reexamination and extension. J Mark 1992 Jul;56(3):55-68. [doi: 10.1177/002224299205600304]
36. Dabholkar PA, Shepherd CD, Thorpe DI. A comprehensive framework for service quality: an investigation of critical conceptual and measurement issues through a longitudinal study. J Retailing 2000 Jun;76(2):139-173. [doi: 10.1016/S0022-4359(00)00029-4]
37. A-priori sample size calculator for structural equation models. DanielSoper. URL: https://www.danielsoper.com/statcalc/calculator.aspx?id=89 [accessed 2024-12-11]
38. Bentler P. EQS Structural Equations Model Program Manual: Multivariate Software; 2006.
39. Chin DN. Empirical evaluation of user models and user-adapted systems. User Model User-Adapt Interact 2001;11(1/2):181-194. [doi: 10.1023/A:1011127315884]
40. Chin WW. The partial least squares approach to structural equation modeling. In: Modern Methods for Business Research: Lawrence Erlbaum Associates; 1998:295-336.
41. Tenenhaus M, Vinzi VE, Chatelin YM, Lauro C. PLS path modeling. Comput Stat Data Anal 2005 Jan;48(1):159-205. [doi: 10.1016/j.csda.2004.03.005]
42. Wetzels M, Odekerken G, van Oppen C. Using PLS path modeling for assessing hierarchical construct models: guidelines and empirical illustration. MIS Q 2009;33(1):177-195. [doi: 10.2307/20650284]
43. Fornell C, Larcker DF. Structural equation models with unobservable variables and measurement error: algebra and statistics. The University of Michigan Library. 1981. URL: https://deepblue.lib.umich.edu/bitstream/handle/2027.42/35622/b1378752.0001.001.pdf?sequence=2&amp

44. Alam MZ, Khanam L. Understanding the determinants of adoption of mHealth services among older women's perspective in Bangladesh. Int J Pharm Healthc Mark 2023 Mar 22;17(1):132-152. [doi: 10.1108/IJPHM-05-2021-0055]
45. Khatun F, Heywood AE, Ray PK, Hanifi SMA, Bhuiya A, Liaw ST. Determinants of readiness to adopt mHealth in a rural community of Bangladesh. Int J Med Inform 2015 Oct;84(10):847-856. [doi: 10.1016/j.ijmedinf.2015.06.008] [Medline: 26194141]
46. Amin R, Hossain MA, Uddin MM, Jony MTI, Kim M. Stimuli influencing engagement, satisfaction, and intention to use telemedicine services: an integrative model. Healthcare (Basel) 2022;10(7):1327. [doi: 10.3390/healthcare10071327]
47. Khatun F, Heywood AE, Ray PK, Bhuiya A, Liaw ST. Community readiness for adopting mHealth in rural Bangladesh: a qualitative exploration. Int J Med Inform 2016 Sep;93:49-56. [doi: 10.1016/j.ijmedinf.2016.05.010] [Medline: 27435947]
48. Uzir MUH, Al Halbusi H, Lim R, et al. Applied artificial intelligence and user satisfaction: smartwatch usage for healthcare in Bangladesh during COVID-19. Technol Soc 2021 Nov;67:101780. [doi: 10.1016/j.techsoc.2021.101780]
49. Zobair KM, Sanzogni L, Houghton L, Islam MZ. Forecasting care seekers satisfaction with telemedicine using machine learning and structural equation modeling. PLoS ONE 2021;16(9):e0257300. [doi: 10.1371/journal.pone.0257300]
50. Birkmeyer S, Wirtz BW, Langer PF. Determinants of mHealth success: an empirical investigation of the user perspective. Int J Inf Manage 2021 Aug;59:102351. [doi: 10.1016/j.ijinfomgt.2021.102351] [Medline: 36540093]
51. Khatun F, Ahmed NU, Rahman H, et al. The promise of teleconsultation in the era of pandemic: a case from Bangladesh. Telemed J E Health 2023 Apr;29(4):602-606. [doi: 10.1089/tmj.2021.0529] [Medline: 35861775]
52. Wang J, Cao Y. Factors influencing continuous intention to use telemedicine after the COVID-19 pandemic in China: an extended technology acceptance model. Open J Soc Sci 2022;10(12):344-359. [doi: 10.4236/jss.2022.1012023]
53. Amin R, Hossain MA, Uddin MM, Jony MTI, Kim M. Stimuli influencing engagement, satisfaction, and intention to use telemedicine services: an integrative model. Healthcare (Basel) 2022 Jul 18;10(7):1327. [doi: 10.3390/healthcare10071327]
54. Rahman M, Rahman A, Siddique KB, Islam R, Ur Rahim MM. Impact of COVID-19 on essential health care of rural people in Northern Bangladesh: a cross-sectional study. Int J Transl Med Res Public Health 2023;7(1):1-10. [doi: 10.21106/ijtmrph.410]
55. Razu SR, Yasmin T, Arif TB, et al. Challenges faced by healthcare professionals during the COVID-19 pandemic: a qualitative inquiry from Bangladesh. Front Public Health 2021;9:647315. [doi: 10.3389/fpubh.2021.647315]
56. Guidelines and rules [Article in Bangla]. Directorate General of Health Services (DGHS). URL: https://old.dghs.gov.bd/index.php/bd/publication/guideline [accessed 2024-12-11]
57. Mohiuddin AK. An extensive review of patient satisfaction with healthcare services in Bangladesh. Patient Exp J 2020;7(2):59-71. [doi: 10.35680/2372-0247.1415]
58. Haque MM. Patient-doctor trust at local healthcare centers in rural bangladesh. In: Policy Response, Local Service Delivery, and Governance in Bangladesh: Springer; 2021:233-257. [doi: 10.1007/978-3-030-66018-5_10]
59. Telemedicine guidelines. Bangladesh Medical & Dental Council. 2020. URL: https://www.bmdc.org.bd/docs/BMDC_Telemedicine_Guidelines_July2020.pdf [accessed 2024-11-14]
60. Uddin Ahmed N, Rahman M, Shekhor Roy S, et al. Impact of telehealth services through "Shastho Batayon 16263" for tackling the COVID-19 pandemic in Bangladesh. East J Healthc 2022;2(1):13-17. [doi: 10.31557/ejhc.2022.2.1.13-17]
61. Motamarri S, Akter S, Ray P, Tseng CL. mHealth: a better alternative for healthcare in developing countries. Presented at: PACIS-2012 Proceedings; Jul 11-15, 2012; Ho Chi Minh City, Vietnam.

Abbreviations

CFA: confirmatory factor analysis
LMIC: low- and middle-income country
mHealth: mobile health
PLS: partial least squares
SEM: structural equation modeling

Edited by C Jacob; submitted 25.02.23; peer-reviewed by H Liliana, R Walczak; revised version received 22.08.24; accepted 26.08.24; published 23.12.24.

Please cite as:
Khatun F, Das NC, Hoque MR, Saqeeb KN, Rahman M, Park KR, Rasheed S, Reidpath DD
Users' Perceived Service Quality of National Telemedicine Services During the COVID-19 Pandemic in Bangladesh: Cross-Sectional Study
JMIR Hum Factors 2024;11:e46566
URL: https://humanfactors.jmir.org/2024/1/e46566
doi:10.2196/46566

Original Paper

Perspectives on the COVID-19 Vaccination Rollout in 17 Countries: Reflexive Thematic and Frequency Analysis Based on the Strengths, Weaknesses, Opportunities, and Threats (SWOT) Framework

Vanja Kopilaš[1*], PhD; Khrystyna Nasadiuk[2*], MD, PhD; Lucia Martinelli[3], PhD; Lenka Lhotska[4,5], PhD; Zoran Todorovic[6,7], MD, PhD; Matjaz Vidmar[8,9], PhD; Helena Machado[10], PhD; Anna Lydia Svalastog[11,12], PhD; Srećko Gajović[13], MD, PhD

[1]Department of Psychology, Faculty of Croatian Studies, University of Zagreb, Zagreb, Croatia
[2]Department of Biochemistry, Danylo Halytsky Lviv National Medical University, Lviv, Ukraine
[3]Science in Society Program, MUSE-Science Museum, Trento, Italy
[4]Czech Institute of Informatics, Robotics and Cybernetics, Czech Technical University in Prague, Prague, Czech Republic
[5]Faculty of Biomedical Engineering, Czech Technical University in Prague, Prague, Czech Republic
[6]University Hospital Medical Center "Bežanijska kosa", Belgrade
[7]Department of Pharmacology, Clinical Pharmacology and Toxicology, Faculty of Medicine, University of Belgrade, Belgrade
[8]Institute for the Study of Science, Technology and Innovation, The University of Edinburgh, Edinburgh, United Kingdom
[9]School of Engineering, The University of Edinburgh, Edinburgh, United Kingdom
[10]Institute for Social Sciences, University of Minho, Braga, Portugal
[11]Østfold University College, Halden, Norway
[12]Centre for Research Ethics and Bioethics, Uppsala University, Uppsala, Sweden
[13]BIMIS-Biomedical Research Center Šalata, University of Zagreb School of Medicine, Zagreb, Croatia
[*]these authors contributed equally

Corresponding Author:
Srećko Gajović, MD, PhD
BIMIS-Biomedical Research Center Šalata
University of Zagreb School of Medicine
Šalata 4
Zagreb, 10000
Croatia
Phone: 385 1 4566 948
Email: srecko.gajovic@hiim.hr

Abstract

Background: As the SARS-CoV-2 virus created a global pandemic and rapidly became an imminent threat to the health and lives of people worldwide, the need for a vaccine and its quick distribution among the population was evident. Due to the urgency, and on the back of international collaboration, vaccines were developed rapidly. However, vaccination rollouts showed different success rates in different countries and some also led to increased vaccine hesitancy.

Objective: The aim of this study was to identify the role of information sharing and context sensitivity in various vaccination programs throughout the initial COVID-19 vaccination rollout in different countries. Moreover, we aimed to identify factors in national vaccination programs related to COVID-19 vaccine hesitancy, safety, and effectiveness. Toward this end, multidisciplinary and multinational opinions from members of the Navigating Knowledge Landscape (NKL) network were analyzed.

Methods: From May to July 2021, 25 completed questionnaires from 27 NKL network members were collected. These contributors were from 17 different countries. The responses reflected the contributors' subjective viewpoints on the status and details of the COVID-19 vaccination rollout in their countries. Contributors were asked to identify strengths, weaknesses, opportunities, and threats (ie, SWOT) of the respective vaccination programs. The responses were analyzed using reflexive thematic analysis, followed by frequency analysis of identified themes according to the represented countries.

Results: The perspectives of NKL network members showed a link between organizational elements of the vaccination rollout and the accompanying societal response, both of which were related to strengths and weaknesses of the process. External sociocultural variables, improved public communication around vaccination-related issues, ethical controversies, and the spread of disinformation were the dominant themes related to opportunities and challenges. In the SWOT 2×2 matrix, *Availability* and *Barriers* emerged as internal categories, whereas *Transparent communication and promotion* and *Societal divide* emerged as key external categories.

Conclusions: Inventory of themes and categories inspired by elements of the SWOT framework provides an informative multidisciplinary perspective for effective implementation of public health strategies in the battle against COVID-19 or any future pandemics of a similar nature.

(JMIR Hum Factors 2024;11:e44258) doi:10.2196/44258

KEYWORDS

SARS-CoV-2 virus; COVID-19 vaccination; pandemic; hesitancy; safety; vaccination; COVID-19; tool; implementation; vaccine hesitancy; effectiveness; sociocultural; communication; disinformation

Introduction

On March 11, 2020, the World Health Organization declared the COVID-19 pandemic [1]. COVID-19 first appeared in December 2019 in Wuhan, China. This disease, caused by the SARS-CoV-2 virus, led to an unprecedented challenge for health institutions that required most countries to integrate their efforts to globally mitigate the spread of the disease [2-4].

Various policies to control the spread of the virus have been adopted in different countries. Some of them were drastic, such as national lockdowns, as well as initiating the widespread use of individual protection devices and means [5]. The individual protective measures included recommending frequent hand washing and application of sanitizers, maintaining social distance, and mandatory wearing of face masks or respirators. However, even a simple measure of covering the face proved to have psychological, cultural, religious, and behavioral implications at both the individual and communal levels [6]. Moreover, the policies aimed to stop the spread of the virus impacted the psychological well-being of the population [7]. These implications should be considered in public campaign strategies aimed at achieving effective public consent toward the adoption of protective measures.

The publication of the genetic sequence of SARS-CoV-2 on January 11, 2020, resulted in the explosion of comprehensive studies on the virus and stirred global research and development activity to develop vaccines against the virus [8]. To accelerate this work, next-generation vaccine technology platforms have been deployed and the first COVID-19 vaccine candidate entered human clinical testing as early as March 16, 2020. In December 2020—in record time and following collaborative efforts of the global scientific community, pharmaceutical companies, and governments—several types and brands of vaccines based on different technologies and mechanisms of action became available for mass deployment [9].

The importance of mass vaccination has been established in the context of previous epidemics such as in the case of smallpox eradication and the incidence reduction of measles and polio [10]. The goal of mass vaccination programs is to interrupt person-to-person disease transmission by surrounding infected people with a high proportion of vaccinated individuals who have developed protective antibodies against the infection (ie, reaching "herd immunity") [11]. Public health experts have prioritized increased vaccination delivery with the hope to resume socioeconomic activities [12]. According to one study, to reduce the number of confirmed COVID-19 cases and deaths, it was estimated that, on average, the administration of 80 vaccine doses per 100 people was necessary [13]. However, the efficacy of vaccination is challenged by an increasing number of mutated strains, clinically proven possibilities of reinfection, and globally uneven rates of vaccination [14,15].

The vaccination process depends on various societal factors such as vaccine hesitancy, vaccine refusal, practical aspects of its application, and uneven/unequal vaccine rollout [16]. Even prior to the COVID-19 pandemic, the World Health Organization identified vaccine hesitancy as one of the top 10 global threats to public health [17]. After the appearance of COVID-19, this issue has gained a completely different dimension, and several people showed different degrees of vaccination acceptance from total refusal to hesitation, including health workers [18]. Levels of COVID-19 vaccine acceptance and obstacles to its rollout are country- and context-dependent [19]. Research has shown that most people are neither absolutely for nor against COVID-19 vaccines [20,21]. Hence, to begin to understand vaccine acceptance, it is important to gain insight into the reasons behind individual and collective decision-making [22].

SARS-CoV-2 is a novel virus, and various questions about dealing with this threat by mass vaccination emerged during the pandemic, including the efficacy of vaccines, the duration of the vaccine's effect, and the impact of new virus variants [23]. The rapid vaccine development raised questions regarding safety, availability, and efficacy [24]. This is not surprising given the fact that vaccine development usually takes 10-15 years, whereas COVID-19 vaccines were developed in less than 1 year [25]. In addition, there are various factors that can increase disease spread and mortality rates that seem incoherent with the proposal for a uniform global vaccination rollout. The mortality rates were lower in countries investing more in the health system and vice versa [15,26]. Research from the United States showed that prosperous states with a higher population of older people and a higher number of physicians had a lower rate of vaccine hesitancy compared to that of other states [12].

The availability of vaccines, both in terms of the number of doses and equal distribution, has been an important issue within various countries, involving technical as well as socioeconomic aspects [27]. Timing is also very important since seasonal and environmental factors play important roles in the reduction of COVID-19 symptomatology [15,28]. Due to the numerous factors involved, interdisciplinary collaboration appears to be an appropriate solution to tackle vaccine hesitancy [29].

To facilitate a discussion on a successful vaccination rollout process, in this study, an analysis inspired by the strengths, weaknesses, opportunities, and threats (SWOT) framework was performed to explore perceptions and establish an informative perspective of the vaccination campaigns in 17 countries during the first phase of the mass vaccination programs in the first half of 2021. To facilitate this research, the scholars from the interdisciplinary Navigating Knowledge Landscape (NKL) research network were surveyed between May and July 2021. They were asked to provide information and their own opinions about the vaccination rollout programs in their respective countries. The participating scholars belonged to different disciplines, creating a specific combination of sociological and cultural analytical competences merged with medical and public health expertise. The aim of this interdisciplinary and transnational analysis was to better understand how information-sharing practices and social context were intertwined to coproduce public opinion on vaccination as a response to the COVID-19 pandemic.

SWOT, as a strategic planning framework, is usually used in evaluation of an organization, plan, project, or business activity. It is therefore a significant tool for situation analysis that helps managers identify organizational and environmental factors affecting performance and operations [30]. The framework can be used to identify favorable and unfavorable factors and conditions, solve current problems in a targeted manner, recognize the challenges and obstacles faced, and formulate strategic plans to guide scientific decisions [30-33]. The SWOT framework strives to offer a comprehensive, systematic, and accurate description of the scenario in which a topic is located [34]. SWOT analysis has two dimensions: internal and external. The internal dimension includes organizational factors focusing on strengths and weaknesses, whereas the external dimension includes environmental factors, namely opportunities and threats [30]. Since SWOT analysis is primarily used in organizational studies, our goal was to use its elements as a conceptual and narrative analysis tool where focus was placed on the intertwining viewpoints of social, political, and public health practices. A similar approach has already been applied as a research method in which aspects of the SWOT framework were used to yield more precise and organized data [35]. However, to date, a SWOT-based analysis of the COVID-19 vaccination campaign has only been reported for India and Zimbabwe [36-38]. Therefore, with this study, we aimed to offer a new transdisciplinary and multinational viewpoint of the vaccination process.

Methods

Study Design

The data set included 25 contributions from 27 members of the NKL research network, collected from May to July 2021. These members contributed their viewpoints through a questionnaire aimed at mapping, in a representative manner, the rollout of the vaccination campaigns against SARS-CoV-2 during the early stages when vaccines were available to the general public.

All contributions were collected in a public data set, which is available with open access in Mendeley Data [39].

Study Sample

The 27 contributors were from 17 different countries: Australia, Austria, Croatia, Czech Republic, Germany, Hungary, Italy, Norway, Portugal, Romania, Serbia, Slovenia, South Korea, Sweden, Turkey, Ukraine, and the United Kingdom (including England and Scotland). Three contributions from the same country were received from Slovenia, Sweden, and Portugal; two from Croatia and the United Kingdom; and one contribution from each of the rest of the countries. Two contributions were coauthored (from Australia and Sweden). The contributors come from different academic backgrounds, but most of them are experts in the fields of life sciences, sociology, philosophy, and medicine. However, it is important to note that the contributors were expressing their own opinions and perceptions.

Measures

The questionnaire contained three parts asking about the status and details of COVID-19 vaccination in the respondent’s country. Contributors were asked to return a short-text (ie, narrative) answer of 200-300 words per part. In this study, we focused only on the SWOT-related aspect of the responses (ie, Part 1) and the responses to the other parts of the questionnaire (Parts 2 and 3) were considered only to identify the eventual contribution to the SWOT-inspired analysis. SWOT elements were selected among the entire response text during the analysis process. The specific questions are presented in Textbox 1.

Textbox 1. Questionnaire items.

- Part 1: The national vaccination program

Describe the COVID-19 vaccination program in your country: what were its strengths, main weaknesses, opportunities to improve it, and threats to its success?

- Part 2: Public discourse and ethics

How would you describe public responses to your country's vaccination program? What is your impression on the various collective attitudes toward the vaccination program in your country? Were there any ethical issues or concerns around the vaccine program in your country?

- Part 3: Personal experience

What is your personal experience, opinion, or attitude regarding COVID-19 vaccination?

Ethical Approval

Ethical approval for this study was obtained from The University of Edinburgh, Scotland, United Kingdom.

Data Analysis Procedure

To fulfill the study's aims and obtain results, reflexive thematic analysis [40] and descriptive statistics (frequency analysis) of the themes were performed. This method is considered appropriate for exploratory research such as our study. Moreover, flexibility of the thematic analysis and opportunity for theme development seemed a great fit and application for our data set [40,41]. The open-ended questions allowed for formulating responses that enabled the respondents to frame the description of the vaccination process in their countries according to their own personal views.

For the purposes of reflexive thematic analysis, we divided the responses into four categories according to the elements of the SWOT framework. The subcategories of each category were identified and a list of the themes for each SWOT element was established. In a subsequent step, we analyzed the data for patterns and recurring topics. We looked for country-specific differences and similarities in regulations and practices. In addition, close attention was paid to how the experts made sense of their experiences with the vaccination process and how the issues addressed were expressed. In presentation of the research results, focus was placed on themes identified throughout the reflexive thematic analysis. The results were then contextualized based on the existing literature.

Following that, frequency analysis of the identified themes was performed in relation to the corresponding countries. In the case of multiple contributions from the same country referring to the same theme or subtheme, only one data point was counted. The obtained results are presented in the form of tables and graphs.

Results

Thematic Analysis

Overview of Themes

Reflexive thematic analysis of collected contributions was performed independently by two researchers (VK, KN). Through the process of the reflexive thematic analysis [40], the numbers of themes respectively belonging to the elements of strengths, weaknesses, opportunities, and threats were established (Figure 1).

Figure 1. Graphical presentation of the established themes within each of the strengths, weaknesses, opportunities, and threats (SWOT) elements.

Thematic analysis of the vaccination process yielded a nearly even distribution of the four SWOT elements across all included countries and contributors, with 7 themes identified for strengths, 5 themes identified for weaknesses, 6 themes identified for opportunities, and 7 themes identified for threats. In total, analysis of the SWOT elements covered 25 different themes.

The contributors shared their subjective perceptions of the effectiveness of the vaccination campaigns in their countries, which ranged from claims of success to voices of criticism. The United Kingdom was the first country in the world to start the COVID-19 vaccination program in December 2020. Shortly afterward, the vaccine rollout was launched in the United States and the countries of the European Union, albeit with some delay (3 months) in Ukraine. In many countries represented in this study, the vaccination rollout started with some constraints, poor planning/management, and delays with vaccines delivery, but improved over time. In Portugal, an efficient organization of the vaccination process was achieved with the change of the Head of the Vaccination Task Force. In the countries of the European Union, the vaccination process was coordinated with that of other member countries (Croatia), although this collaboration was not always perceived as efficient, as pointed out by a contributor from Sweden.

A successful vaccination program was achieved in the United Kingdom, being respectively described as "overall…a large success" and "an overwhelming success" [39]. The contributions from Portugal and Serbia highly rated the results of the vaccination programs in their respective countries in relation to the high vaccination rate and being ahead of plans/schedule. A relatively successful vaccination process was also reported in Croatia, Hungary, Italy, and Norway. Efficient implementation was noted in Turkey, and an active vaccination process was described in South Korea with major public facilities offering vaccinators discounts or exemptions from paying admission or usage fees. For some other countries, the collected contributions reported low vaccination coverage in the survey period (May to July 2021), including Australia, Romania, and Ukraine. The respondent in Romania specifically reported low coverage for high-risk groups and people over 65 years old. Moreover, very low coverage of rural areas occurred due to lack of local community involvement, especially mayors, policy makers, and family doctors, with some of the latter refusing to dispense vaccines.

Slow rollout of the vaccines was noted in Australia, Austria, and Germany. In Australia, the delayed vaccine rollout has been described as a "vaccine stroll out," as by July 2021, only 6% of the Australian population had been vaccinated [39]. Moreover, some individuals in priority groups such as older people or those with disabilities living in long-term care homes were still waiting for their second or even their first dose of the vaccine. In addition, in some countries, the vaccination points were hard to access in remote, rural areas (Australia).

If we are to judge vaccination rollout success by looking at the percentage of people who had received at least one dose of the vaccine during the time period corresponding to our data collection, the most successful country in our sample was the United Kingdom, with approximately 70% of the population receiving at least one dose (Multimedia Appendix 1) [42]. The lowest percentage was reported in Ukraine, where only approximately 8% of people had received a single vaccine dose [42].

Strengths

The primary themes related to strengths included (1) societal discussion/consensus on priorities to get the vaccine, (2) defined vaccination strategy/plan, (3) vaccine availability, (4) positive attitudes toward vaccines and the vaccination process, (5) practical aspects of the vaccination solved (eg, medical personnel satisfied, sites easy to access, fast process, no long queues), (6) well-designed public communication campaign on the vaccination process, and (7) flexibility to provide vaccines.

High availability of vaccines was reported in Hungary, Italy, Sweden, and the United Kingdom. Following the controversies around the possible side effects of AstraZeneca vaccines, stocks of the European Union–approved vaccines were excessive in Slovenia. The wide availability of vaccines to whoever wanted them was considered a strength of the vaccination campaigns. In Romania, free vaccination has been offered to everybody who wanted one, including those with Romanian or European citizenship. In Portugal, vaccination was available independent of legal status, including to undocumented migrants. Free vaccination was also offered in Serbia to people from abroad, primarily citizens of neighboring countries, with no restrictions.

Medical workers played a key role in achieving successful vaccination campaigns. Family doctors contributed to the success of the vaccine rollout in Croatia and a helpful approach was reported by the medical staff of the Czech Republic. For Portugal, strong commitment of health care professionals and communication initiatives of the medical doctors to clarify doubts related to the vaccine's side effects were noted.

Transparent planning and strategies, as well as prioritization of people with higher infection risk or greater vulnerability, were commonly reported strengths of the vaccination programs. In most countries, the prioritization was perceived as fair, although in some countries controversial cases of people from nonpriority groups being vaccinated early also occurred (Portugal, Slovenia). The priority groups in most countries included older adults, those with underlying health conditions, and workers exposed to a high infection risk. By contrast, vaccination of health care professionals has not been prioritized in Sweden. In all countries, the vaccine was provided free of charge, dispensed on a voluntary basis; however, mandatory vaccination was reported for medical workers in Italy and South Korea and for people in high-risk jobs in Australia. Moreover, an easy registration process, owing to easy-to-access platforms such as apps, web pages, or via the phone, was described for Turkey and Ukraine. Automatic enrollment based on medical records via general practitioners (eg, family doctors) was available in the United Kingdom. An efficient registration process in the Czech Republic was also claimed as a strength.

Weaknesses

The primary themes related to weaknesses were as follows: (1) social divide due to the vaccine distribution and side effects,

(2) unclear vaccination strategy/plan, (3) lack of vaccines, (4) negative attitudes or hesitancy toward vaccines and the vaccination process, and (5) barriers to access vaccines.

Lack and shortage of vaccines were emphasized in Ukraine and Turkey, as well as at the beginning of the vaccination rollout in some other countries, where delayed deliveries were also reported. Delayed rollout to the remote Aboriginal communities was noted in Australia. The registration process was essential in achieving an effective rollout of the vaccines. Poor functioning of the distribution organization was highlighted by many participants from different countries, especially in the early stages of vaccination programs. Getting a vaccination appointment was rated as difficult in Sweden.

Trust was pointed out as an important issue in several contributions. A low level of trust in the medical science (Croatia), in the effectiveness and safety of the vaccines (Ukraine, Romania, Slovenia), and in the official authorities (Slovenia) were reported. An overall high level of skepticism in society at large was observed (Germany). Lack of enthusiasm and willingness to be vaccinated or vaccine hesitancy were widely reported (Austria, Norway, Romania, Slovenia, Ukraine). Despite the very successful vaccination process in Serbia, only a small percentage of younger people and health workers were vaccinated in the country. High hesitation among young people was also reported in Slovenia. In contrast, in Ukraine, young people were rather eager to be vaccinated, despite the high level of general hesitancy noted in the country. In Australia, vaccine hesitancy was exacerbated by the risks of side effects reported for the AstraZeneca vaccine.

Lack of clear and coherent communication on how to receive the vaccination was considered an important barrier to access in Slovenia. The need for suitable and unequivocal guidelines about vaccination was stressed in Italy, as constant changes have confused the population and discouraged vaccination, while different rules in different parts of the country and frequent regulation changes were noted to have discouraged vaccination in Germany. Unavailability of scientific information to foreigners/migrants, especially for those not fluent in the local language (eg, the home workers caring for the older population) was stressed for Italy. The digitally based vaccination approach was considered an important barrier to those not having adequate digital abilities. In Sweden, despite having one of the highest internet coverage rates in the world, people living in socially disadvantaged areas, including asylum seekers and migrants, and older adults or people with cognitive impairment who did not master the digital skills required were at risk to be excluded from accessing important information. A low level of digitalization was also mentioned as an obstacle to vaccination success in Romania.

Opportunities

The primary themes related to opportunities were as follows: (1) adding more flexibility; (2) increasing vaccine availability and multiple options for registration; (3) active outreach to marginalized groups, vulnerable citizens, refugees, and ethnic minorities; (4) improving the role of the media (better communication) and national awareness campaigns; (5) information sharing about the usefulness of the vaccination process; and (6) provisions for vaccinated individuals.

The freedom to choose to make an appointment for vaccination, no matter where people were registered (Sweden), and adding more flexibility to accessing vaccination (Croatia) were considered among the opportunities to improve the vaccination rollout.

To motivate people to be vaccinated, financial support (approximately US $30) was offered in Serbia. Vaccination coupons or exemptions from admission or usage fees of public facilities (approximately US $900) were introduced by a National Vaccine Injury Compensation Program in South Korea. In addition, this country also allowed a one-day “vaccination leave” from work to be taken the day after receiving the vaccine, along with an additional one-day leave in the case of experiencing some subsequent side effects [39]. In Ukraine, in the unlikely case that vaccination would cause disability or death, a compensation allowance (approximately US $21,000-27,000) was promised by the government.

Among the opportunities to improve the vaccination process, the freedom to choose among the available vaccine brands/types was recognized as a good strategy to counteract the arising doubts about a certain brand of vaccine (Slovenia, Ukraine). The choice of vaccine brand was also available in Turkey and in Serbia, contributing to successful vaccination campaigns. Finally, a more responsible role of media was mentioned as an opportunity to improve people’s attitude toward vaccination, as pointed out for Croatia and Ukraine. Moreover, in various countries, the respondents suggested that improvement of communication strategies and specific information programs might be crucial to reach vaccine-hesitant citizens and facilitate the vaccine rollout.

In addition to traditional media, social media were noted to play a role. Social media influencers were identified to positively contribute to motivating people to be vaccinated, producing a “crowd effect,” as reported for Croatia and Ukraine, where public figures, such as the President and the health minister gave declarations through the media. Vivid promotions in favor of vaccination by persuasive political and medical discourses, accompanied by enthusiastic argumentations in favor of science and against conspiracy theories and vaccination skepticism, were described for Slovenia.

Threats

The primary themes related to threats were as follows: (1) appearance of new virus strains/lower efficacy of the vaccines, (2) unforeseen side effects of the vaccines, (3) spreading disinformation, (4) ethical controversies, (5) legal controversies, (6) religious controversies, and (7) a change in the behavior of vaccinated individuals that facilitates the spread of infection.

Low trust in the efficacy and safety of the vaccines (Romania, Slovenia, Ukraine); a negative influencing role played by some media communications, especially when stressing the side effects (Serbia, Sweden), and alleged corruption related to the vaccine prioritization (Slovenia) were regarded as relevant threats to be considered for achieving successful mass vaccination campaigns. Insufficient information, disinformation, or misinformation in the media and on the internet were reported for the Czech Republic, Sweden, and Romania, while development of conspiracy theories about vaccines was pointed out for Slovenia and Ukraine. Disputable communication from the government regarding vaccines and other public health measures such as lockdowns was described for Germany. Lack of adequate public communication strategies was also noted in Slovenia. Failures in communication with people from different cultural groups were reported in Australia.

Ethical concerns associated with the use of leftover doses were pointed out by respondents from Sweden and Portugal, referring to a lack of planning for how to handle leftover vaccines that could not be administered the next day or to the overall mismanagement of vaccine administration. In contrast, the opportunity to get a leftover dose was marked as a strength at the beginning of the vaccination campaign in Ukraine, where this was the only option to be vaccinated for those in nonpriority groups. Confusing messages from religious leaders and local community priests were reported in Romania. Concerns of disobeying the Islamic conduct codes raised by vaccination opponents was described for Turkey, as during the month of Ramadan fears were prompted that vaccination during the fasting period was not acceptable.

Frequency Analysis

Overview

To explore the distribution of the responses by countries, frequency analysis was performed (Figures 2-5). Responses reporting a certain theme are marked in the figures in green color and assigned a value of 1, whereas those that did not mention the theme are marked with light yellow and assigned a value of 0. The total score corresponds to the sum of values of all related responses. Additionally, the average percentage of responses distributed for each element and theme was calculated (Multimedia Appendices 2-5).

Figure 2. Overview of the opinions covering strengths-related themes by country. Green indicates presence of a theme (assigned a value of 1) and yellow indicates absence of the theme (assigned a value of 0). The total score corresponds to the sum of values of all related responses.

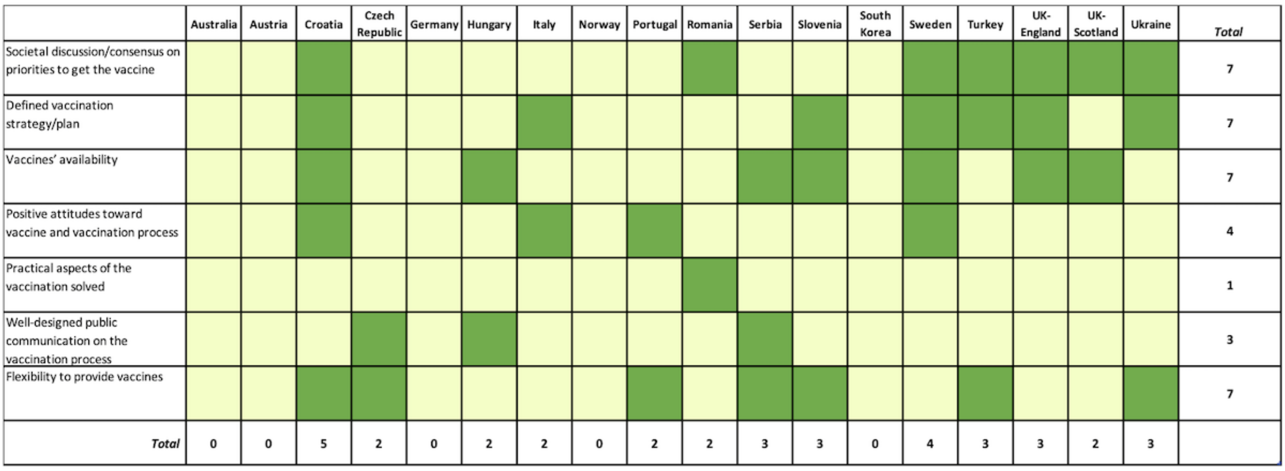


Figure 3. Overview of the opinions covering weaknesses-related themes by country. Green indicates presence of a theme (assigned a value of 1) and yellow indicates absence of the theme (assigned a value of 0). The total score corresponds to the sum of values of all related responses.

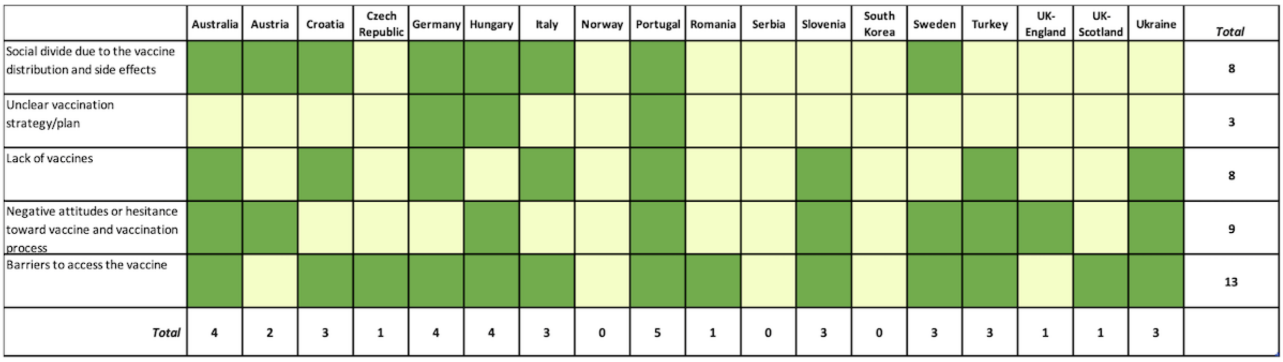


Figure 4. Overview of the opinions covering opportunities-related themes by country. Green indicates presence of a theme (assigned a value of 1) and yellow indicates absence of the theme (assigned a value of 0). The total score corresponds to the sum of values of all related responses.

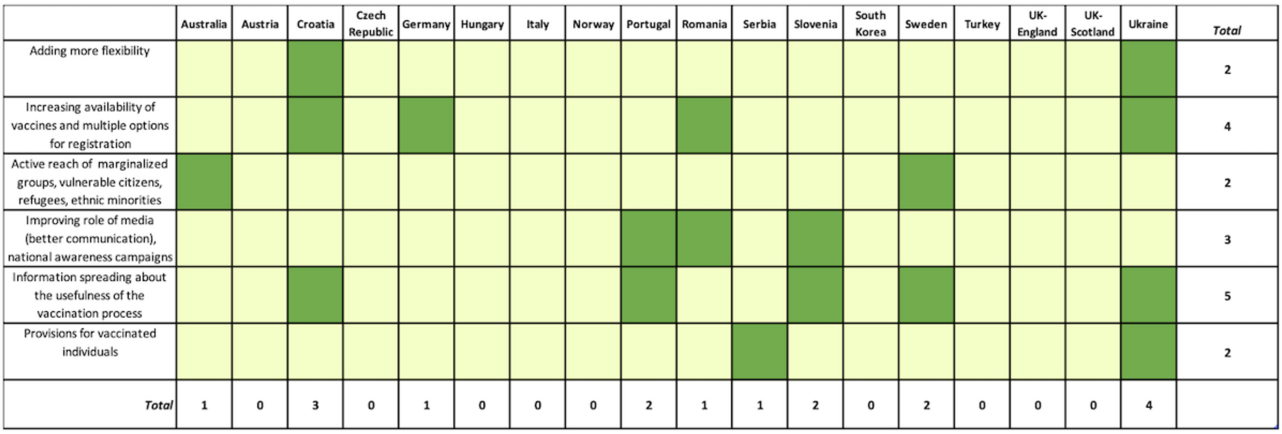
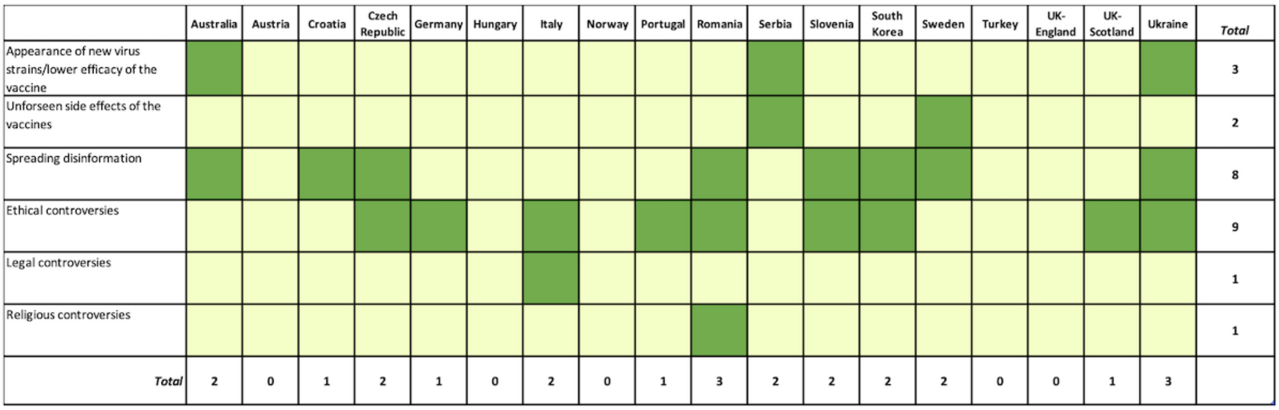


Figure 5. Overview of the opinions covering threats-related themes by country. Green indicates presence of a theme (assigned a value of 1) and yellow indicates absence of the theme (assigned a value of 0). The total score corresponds to the sum of values of all related responses.



Strengths

Three themes dominated the analysis of strengths, each being covered in 7 reports: societal discussion/consensus on priorities to get the vaccine, defined vaccination strategy/plan, vaccines' availability, and flexibility to provide vaccines.

The strengths theme "societal discussion/consensus on priorities to get the vaccine" was mentioned in Croatia, Romania, Sweden, Turkey, UK Scotland, UK England, and Ukraine. "Defined vaccination strategy/plan" was reported in Croatia, Italy, Slovenia, Sweden, Turkey, UK (England), and Ukraine. "Wide vaccines availability" was indicated in Croatia, Hungary, Serbia, Slovenia, Sweden, and the United Kingdom. "Positive attitudes of the society toward vaccines and vaccination process" were reported for Croatia, Italy, Portugal, and Sweden. Logistic aspects of the vaccination being solved (including satisfaction of the medical personnel, sites easy to access for registration, fast process, no waiting in line) were noted for Romania, while well-designed public communication on the vaccination process was described for the Czech Republic, Hungary, and Serbia. Flexibility to provide vaccines was highlighted as a potential strength in the contributions from Croatia, Czech Republic, Portugal, Serbia, Slovenia, Turkey, and Ukraine. Relatively even distribution was identified across strengths categories with the exception of practical aspects of the vaccination solved that was reported by only one contributor.

Weaknesses

Most frequently reported weaknesses were barriers to access the vaccination (13 reports) and negative attitudes or hesitancy toward vaccines and the vaccination process (9 reports).

Social divide due to the vaccine distribution and side effects were considered weaknesses in Australia, Austria, Croatia, Germany, Hungary, Italy, Portugal, and Sweden. Unclear vaccination strategy/plan was described in Germany, Hungary, and Portugal. Lack of vaccines was noted in Australia, Croatia, Germany, Italy, Portugal, Slovenia, Turkey, and Ukraine (note that the questionnaire addressed these issues only related to the first 6 months of the vaccination campaigns). Negative attitudes or hesitancy toward vaccines and the vaccination process were described in Australia, Austria, Hungary, Portugal, Slovenia, Sweden, Turkey, UK England, and Ukraine, while barriers to access vaccination, including problems with prioritization, registration, and unfair/nontransparent distribution of the vaccines were noted in Australia, Croatia, Germany, Hungary, Italy, Czech Republic, Portugal, Romania, Slovenia, Sweden, Turkey, UK Scotland, and Ukraine. In addition, 72.2% of the contributions reported barriers to access vaccines as a weakness. Conversely, only 16.6% of our sample reported an unclear vaccination strategy/plan as weakness.

Opportunities

The frequencies of the selected opportunities to improve the vaccination process were rather low including a maximum of 5 countries. Adding more flexibility to the vaccination process was mentioned in Croatia and Ukraine; increasing availability of the vaccines and multiple options for registration were mentioned in Croatia, Germany, Romania, and Ukraine; active reach of marginalized groups, vulnerable citizens, refugees, and ethnic minorities was mentioned in Australia and Sweden; improving the role of the media (better communication) and national awareness campaigns were indicated in Portugal, Romania, and Slovenia; information spreading about the usefulness of the vaccination process was highlighted in Croatia, Portugal, Slovenia, Sweden, and Ukraine; and monetary provisions for vaccinated individuals were mentioned in Serbia and Ukraine. Contributors did not report opportunities in large numbers. The highest percentage (27.7%) of responses related to the opportunities-related themes was attributed to spreading information about the usefulness of the vaccination process.

Threats

Concerning the possible threats to a vaccination campaign's success, the contributors from Australia, Serbia, and Ukraine remarked the possible appearance of new virus strains and lower efficacy of the vaccine; unforeseen side effects were noted as possible threats in the contributions from Serbia and Sweden; spreading disinformation were noted or could be concluded from the abstracts from Australia, Croatia, Czech Republic, Romania, Slovenia, South Korea, Sweden, and Ukraine. Other possible threats to vaccination success mentioned were ethical (Czech Republic, Germany, Italy, Portugal, Romania, Slovenia, South Korea, UK Scotland, and Ukraine), along with legal (Italy) and religious controversies (Romania). Ethical controversies (50%) and spreading information (44.4%) were the most highly represented threats-related themes.

In this study, the mainly acknowledged threat feature for achieving successful vaccination campaigns reported by the respondents was related to the likely occurrence of viral mutations, resulting in new virus strains with the ability to escape the immunizing effects of the present available vaccines. This fact has been pointed out as a relevant source of uncertainties and doubts about the vaccines' effectiveness as well as about their overall reliability and utility.

Discussion

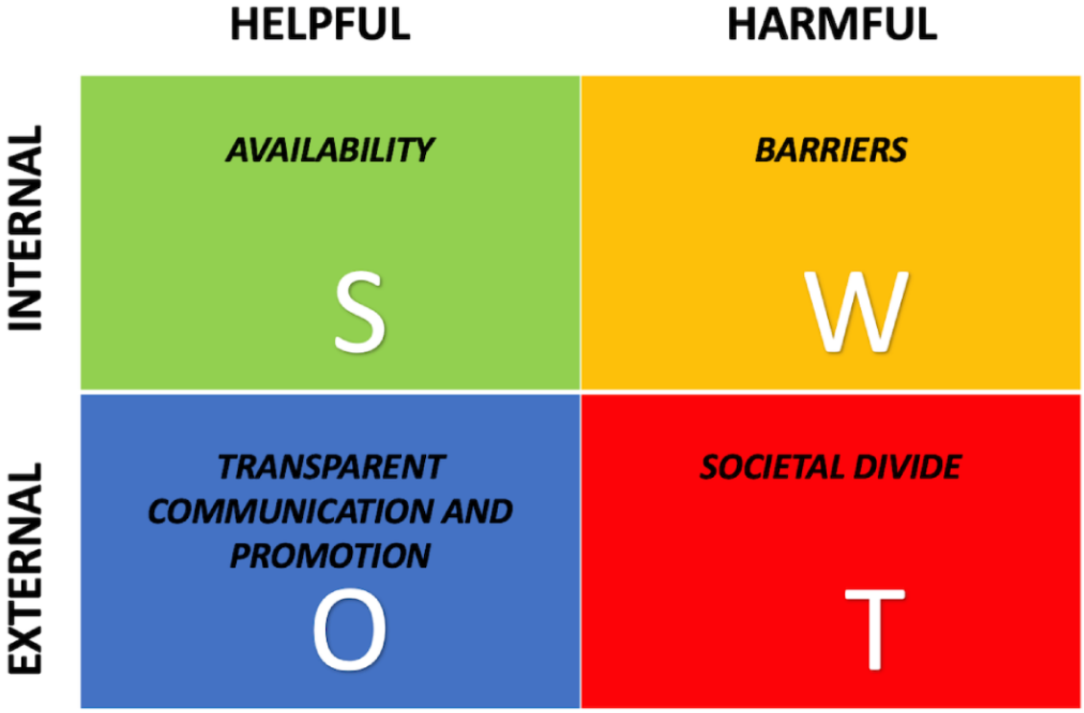
Principal Findings

Since its introduction in the 18th century to the present day, vaccination has been one of the most effective tools in the battle against infectious diseases [10,43]. Owing to the high efficacy of vaccines, the public health burden of infectious diseases has been significantly reduced throughout the years [10]. However, despite their proven track record, the phenomenon known as "vaccine hesitancy" has been around almost as long as vaccination itself. This reluctance to accept an injection of an "unknown substance" into the body is exasperated by a need to vaccinate a large number of healthy individuals, including in the case of COVID-19 [44].

This study, based on an analysis of interdisciplinary experts' viewpoints in 17 different countries inspired by the SWOT framework, allowed us to identify 25 themes distributed across the four SWOT elements. To our knowledge, this is the first study to analyze and compare the vaccine rollout process in various different countries. The frequency of the appearance of these themes and their distribution across the countries allowed us to select those that stand out. As the contributions were inspired by the SWOT framework, the presented analysis could be easily synthesized into the four main overarching SWOT categories (Figure 6). With respect to the strengths of the vaccination process, the identified seven themes correspond to a single category referred as *Availability*, being the major strength of the successful vaccination program. When weaknesses were described, the five themes identified could be best described by a single category termed *Barriers*, which were either not recognized or not addressed by the vaccination programs. The external aspects of opportunities described via the six themes identified fit under category *Transparent communication and promotion*, which allows other societal forces to contribute to the vaccination process. Finally, the seven themes describing threats correspond to the *Societal divide* category, where a polarized society has the potential to spoil even well-thought-out initiatives.

We believe that these categories offer the best representation of the most frequently reported themes in each of the SWOT elements. However, due to the intertwining factors present in the vaccination rollout process, it is important to not look at this distribution as a binary (presence/absence) phenomenon. This is particularly relevant when splitting the identified themes into "internal" and "external" categories. In the current SWOT 2×2 matrix (Figure 6), *Availability* and *Barriers* are labeled as internal categories, whereas *Transparent communication and promotion* and *Societal divide* are suggested as external categories [45]. However, within the *Societal divide* category labeled as a threat, there are ethical, religious, and legal controversies reported as important themes. Therefore, one cannot classify a controversy per se as a threat, as controversies can serve just as much as a source of debate with the potential to improve the vaccination process.

Figure 6. Synthesized themes under the strengths (S), weaknesses (W), opportunities (O), and threats (T) framework.



The specific time window when the study was performed corresponds to a relatively early phase of the vaccination process (on average half of the population had been vaccinated in the analyzed countries). This leads to a very specific bias in the submitted data: urgency to tackle an important and pressing issue. We reiterate that our study analyzed the subjective viewpoints of the respondents; hence, some of the themes across these categories were dependent on the various individual, psychological, emotional, and societal aspects specific to the given time window. The sudden appearance of COVID-19 and its rapid spread called for appropriately rapid responses. Considering psychological factors of egocentrism, information availability, social/group confirmation, individual motivation, and emotional affect as foundations of that rapid decision-making process, it is easily possible to misjudge and/or misperceive the key elements of the reasoning arising from the complexity of the situation [46]. However, although there were 27 individual respondents from 17 different countries, our results did not show country-specific differences. Hence, our findings can contribute to the development of strategies that will maximize the promotion of strengths and opportunities while minimizing weaknesses and threats globally.

The identified themes are consistent with the research on this topic [36,38]. The most prominent theme in the existing literature, which was also present in our study, explores the effective medical and public health system measures mapping on the key strengths identified herein. This shows how preparation and prevention strategies work, and how they can be used as a base of the powerful pushback against the spread of COVID-19. Moreover, a positive attitude toward vaccination has been defined as a strength in similar studies in India and Zimbabwe [36,38].

The application of the SWOT framework to complex societal processes can also be seen as a source of confusion. For example, a "strength" is considered as an internal aspect of the process, which can be understood to relate to the vaccination campaign itself. From this perspective, the attitude toward vaccination does not seem to be an internal component but rather an external aspect of SWOT and thus should be more appropriately classified as an opportunity rather than a strength. However, application of the SWOT framework in such complex scenarios requires consideration of the vaccination campaign as part of a sociotechnical system, thus incorporating vital elements of the social environment within the situated practice of vaccination. Combining our findings obtained from individuals from 17 different countries with previous research, it can be concluded that good organization that addresses the availability of vaccines coupled with an engaging societal discussion would represent a key strength/opportunity of the vaccination process.

A lack/shortage of vaccines combined with various logistical challenges have been reported as major issues for the success of vaccination campaigns within previous research [47,48]. The demand-supply gap combined with lack of knowledge and supporting infrastructures have been reported as particular weaknesses [36,38]. Compounding unequal vaccine distribution with unknown disease progression and an uncertain response to the vaccine seems to be the biggest barrier in the vaccine rollout [13]. Similarly, the respondents of this study recognized the practical issues of availability and fairness of distribution, and coupled these issues with the related attitudes and social division. This points to the fact that social distrust needs to be addressed within a vaccination plan as a major barrier. For both strengths and weaknesses, no clear geographical divide was present.

Increasing the public awareness about the vaccine effects through transparent communication and promotion stood out as a key opportunity-related theme. Communication reports on the widespread acceptance of COVID-19 vaccines have shown to be effective tools to further increase vaccine acceptance [49]. Moreover, in an attempt to promote vaccination, some public figures have been vaccinated on television [49]. It is interesting to see that people who used mainstream media outlets as their major source of information on health were more likely to get vaccinated [50]. Our data support the notion that transparent information-sharing about biological mechanisms, efficacy, as well as side effects of the vaccines motivates people to join vaccination programs. Previous research has identified the potentially influential role of media in increasing people's trust in vaccines when they hear politicians, celebrities, or other famous people talking positively about them [36]. Trust in vaccines, medical science, and medical professionals—together with other involved stakeholders, including government and policy makers—was highlighted in the analyzed contributions as an important factor. These findings align with previous research that found lack of communication from trusted providers and community leaders as one of the main reasons for low COVID-19 vaccination rates [44]. Communication of vaccine information and promotion of its uptake in the digital era includes the use of social networks [51]. However, the use of social networks is also associated with risks due to the wildfire-like dynamics of rumors in the digital environment and issues with unknown algorithms used by for-profit entities filtering information [52,53]. Social networks are expected to drive healthy public debates; however, they instead frequently reinforce like-minded "bubbles" and increase polarization [53,54].

Discussions on matters of autonomy (an individual's right to choose) and state power have always been at the center of public health ethical dilemmas [29]. In the specific case of COVID-19 vaccines, besides the tensions between public health and individual interest/autonomy, other ethical challenges relate to the rapid design and testing of vaccines and who gets the vaccines (first) [55]. Public health authorities need to implement efficient, flexible, responsive, and resilient strategies to successfully fight the pandemic and raise awareness of all of the dangers arising with this disease [56]. Surprisingly, in our findings, the question of one's autonomy did not crystalize as a theme. Instead, other ethical controversies and spreading of disinformation were found to be the most frequently reported themes within the threats element. In the present digital era, information accessibility is at its peak; however, it is important be aware of the source of the information given that rumors and fake news are rampant [17,57].

When discussing the threats element in the SWOT framework, it was interesting that unforeseen side effects of the vaccine have not been considered as the most prominent threat theme, whereas other research shows that the most common reason for vaccine hesitancy or refusal is due to the concerns related to the side effects/safety [50,58,59]. The emergence of new virus strains was mentioned as a threat, since they decrease the efficacy of the vaccine and hence can contribute to the further spread of COVID-19. As people were already worried about the lack of information about safety, testing, and efficacy of COVID-19 vaccines, the new variants were seen as a contributor to the negative perception of the vaccines in society [15,60]. The synthesizing category for the threat element of *Societal divide* implies that social polarizations have the potential to paralyze a society when facing a complex public health crisis. Here, it should be stated that silencing the controversies is certainly not the path to avoid such an outcome. A society where controversies are not openly discussed is not without these controversies, but rather this situation would give rise to potentially dangerous and isolation subcultures. Although *Societal divide* was recognized as a threat in our sample, there were no clear examples where this has significantly directly influenced the vaccination process. Consequently, although the awareness of controversies as a potential threat was voiced, if the social environment is developed within the context of *Transparent communication and promotion* (opportunity), the *Social divide* may never reach the level of polarization to create adverse effects on public health campaigns.

Study Limitations

The collected responses represent the subjective viewpoints of experts who volunteered to take part in the study. Therefore, extrapolation to the national level must be drawn out with caution. In addition, due to the lack of research using this same methodology and implementing it on a multinational level, there were no relevant studies to make direct comparisons with and contrast conclusions. Moreover, SWOT analysis was not performed in its original form addressing organizational dynamics. Instead, this thematic analysis of expert viewpoints was only inspired by the SWOT framework. Therefore, the results of this study should be further examined and more research is needed on this topic in general. Further studies could consider interdisciplinary and multinational frameworks to find the best practice in public health policies that could yield improved vaccination rollout results globally.

Conclusion

This study was based on a collection of short responses to a specifically designed questionnaire, written by researchers from different countries and fields of expertise, thus bringing together multidisciplinary and cross-national opinions on vaccination rollout. This represents the first analysis of the vaccination process in 17 different countries inspired by the SWOT framework. The obtained results highlight the connection between organizational aspects of the vaccination rollout and corresponding societal response, both being related to the strengths and weaknesses of the process. The opportunities and threats corresponded to external societal factors, better public communication of vaccination-related issues, ethical controversies, and the spread of disinformation. The inventory of 25 SWOT-related themes and the resulting 2×2 SWOT matrix represents an approximate best-practice viewpoint for the successful implementation of public health policies—as represented by this multidisciplinary team—in the fight against COVID-19.

Acknowledgments

This study represents an activity of the interdisciplinary network Navigating Knowledge Landscapes. The study was supported by the European Union through the European Regional Development Fund, as the Scientific Centre of Excellence for Basic, Clinical and Translational Neuroscience, under grant agreement KK.01.1.1.07.0071, project “Synergy of molecular markers and multimodal in vivo imaging during preclinical assessment of the consequences of the ischemic stroke (SineMozak)” (to SG).

Data Availability

The data sets supporting the results presented in this study can be found in the online repository [39].

Authors' Contributions

VK, KN, MV, and SG designed the study. VK, KN, and LM performed data acquisition, organization and analysis, and wrote the first version of the manuscript. VK, KN, LM, LL, ZT, MV, HM, ALS, and SG contributed to the interpretation of the data collected, framed the results, and critically revised the manuscript. All authors approved the submission to the journal.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Share of people from countries in our sample who received at least one dose of COVID-19 vaccine (taken from Our World in Data. Coronavirus (COVID-19) Vaccinations. 2021 [42]).
[PNG File , 438 KB - humanfactors_v11i1e44258_app1.png]

Multimedia Appendix 2
Average percentage of opinions covering strengths-related themes.
[PNG File , 151 KB - humanfactors_v11i1e44258_app2.png]

Multimedia Appendix 3
Average percentage of opinions covering weaknesses-related themes.
[PNG File , 126 KB - humanfactors_v11i1e44258_app3.png]

Multimedia Appendix 4
Average percentage of opinions covering opportunities-related themes.
[PNG File , 140 KB - humanfactors_v11i1e44258_app4.png]

Multimedia Appendix 5
Average percentage of opinions covering threats-related themes.
[PNG File , 134 KB - humanfactors_v11i1e44258_app5.png]

References

1. Archived: WHO Timeline - COVID-19. World Health Organization. 2020 Apr. URL: https://www.who.int/news/item/27-04-2020-who-timeline---covid-19 [accessed 2023-07-03]
2. Zhou P, Yang X, Wang X, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020 Mar;579(7798):270-273 [FREE Full text] [doi: 10.1038/s41586-020-2012-7] [Medline: 32015507]
3. Usher K, Durkin J, Bhullar N. The COVID-19 pandemic and mental health impacts. Int J Ment Health Nurs 2020 Jun;29(3):315-318 [FREE Full text] [doi: 10.1111/inm.12726] [Medline: 32277578]
4. Satici B, Saricali M, Satici SA, Griffiths MD. Intolerance of uncertainty and mental wellbeing: serial mediation by rumination and fear of COVID-19. Int J Ment Health Addict 2022;20(5):2731-2742 [FREE Full text] [doi: 10.1007/s11469-020-00305-0] [Medline: 32427165]
5. Indirili M, Di Maio V, Martinelli L. An analysis of Italian resilience during COVID-19 pandemic: first phase from January to June 2020. In: De Pascale F, editor. Geographies of the Anthropocene. Lago, Italy: Il Sileno Edizione; 2021.
6. Martinelli L, Kopilaš V, Vidmar M, Heavin C, Machado H, Todorović Z, et al. Face masks during the COVID-19 pandemic: a simple protection tool with many meanings. Front Public Health 2020 Jan 13;8:606635 [FREE Full text] [doi: 10.3389/fpubh.2020.606635] [Medline: 33520918]
7. Kopilaš V, Hasratian AM, Martinelli L, Ivkić G, Brajković L, Gajović S. Self-perceived mental health status, digital activity, and physical distancing in the context of lockdown versus not-in-lockdown measures in Italy and Croatia: cross-sectional

study in the early ascending phase of the COVID-19 pandemic in March 2020. Front Psychol 2021 Feb 4;12:621633 [FREE Full text] [doi: 10.3389/fpsyg.2021.621633] [Medline: 33613398]
8. Forster P, Forster L, Renfrew C, Forster M. Phylogenetic network analysis of SARS-CoV-2 genomes. Proc Natl Acad Sci U S A 2020 Apr 28;117(17):9241-9243 [FREE Full text] [doi: 10.1073/pnas.2004999117] [Medline: 32269081]
9. Thanh Le T, Andreadakis Z, Kumar A, Gómez Román R, Tollefsen S, Saville M, et al. The COVID-19 vaccine development landscape. Nat Rev Drug Discov 2020 May;19(5):305-306. [doi: 10.1038/d41573-020-00073-5] [Medline: 32273591]
10. Saleh A, Qamar S, Tekin A, Singh R, Kashyap R. Vaccine development throughout history. Cureus 2021 Jul;13(7):e16635 [FREE Full text] [doi: 10.7759/cureus.16635] [Medline: 34462676]
11. Randolph HE, Barreiro LB. Herd immunity: understanding COVID-19. Immunity 2020 May 19;52(5):737-741 [FREE Full text] [doi: 10.1016/j.immuni.2020.04.012] [Medline: 32433946]
12. Goel RK, Jones JR, Saunoris JW. Explaining vaccine hesitancy: a COVID-19 study of the United States. MDE Manage Decis Econ 2023;44(2):1073-1087 [FREE Full text] [doi: 10.1002/mde.3732] [Medline: 36247212]
13. Coccia M. Optimal levels of vaccination to reduce COVID-19 infected individuals and deaths: a global analysis. Environ Res 2022 Mar;204(Pt C):112314 [FREE Full text] [doi: 10.1016/j.envres.2021.112314] [Medline: 34736923]
14. Patel V, Shirbhate E, Rajak H. Coronavirus reinfections: an outlook on evidences and effects. In: Kaklauskas A, Abraham A, Okoye K, Guggari S, editors. Lessons from COVID-19: impact on healthcare systems and technology. Amsterdam: Elsevier Academic Press; 2022:19-40.
15. Coccia M. COVID-19 vaccination is not a sufficient public policy to face crisis management of next pandemic threats. Public Organiz Rev 2022 Oct 17;23(4):1353-1367. [doi: 10.1007/s11115-022-00661-6]
16. Leigh JP, Moss SJ, White TM, Picchio CA, Rabin KH, Ratzan SC, et al. Factors affecting COVID-19 vaccine hesitancy among healthcare providers in 23 countries. Vaccine 2022 Jul 29;40(31):4081-4089 [FREE Full text] [doi: 10.1016/j.vaccine.2022.04.097] [Medline: 35654620]
17. Teitler-Regev S, Hon-Snir S. COVID-19 vaccine hesitancy in Israel immediately before the vaccine operation. Yale J Biol Med 2022 Jun;95(2):199-205 [FREE Full text] [Medline: 35782475]
18. Gur-Arie R, Jamrozik E, Kingori P. No jab, no job? Ethical issues in mandatory COVID-19 vaccination of healthcare personnel. BMJ Glob Health 2021 Feb 17;6(2):e004877 [FREE Full text] [doi: 10.1136/bmjgh-2020-004877] [Medline: 33597280]
19. de Figueiredo A, Simas C, Larson HJ. COVID-19 vaccine acceptance and its socio-demographic and emotional determinants: a multi-country cross-sectional study. Vaccine 2023 Jan 09;41(2):354-364 [FREE Full text] [doi: 10.1016/j.vaccine.2022.10.051] [Medline: 36414475]
20. Schwarzinger M, Watson V, Arwidson P, Alla F, Luchini S. COVID-19 vaccine hesitancy in a representative working-age population in France: a survey experiment based on vaccine characteristics. Lancet Public Health 2021 Apr;6(4):e210-e221 [FREE Full text] [doi: 10.1016/S2468-2667(21)00012-8] [Medline: 33556325]
21. Verger P, Peretti-Watel P. Understanding the determinants of acceptance of COVID-19 vaccines: a challenge in a fast-moving situation. Lancet Public Health 2021 Apr;6(4):e195-e196 [FREE Full text] [doi: 10.1016/S2468-2667(21)00029-3] [Medline: 33556329]
22. Machingaidze S, Wiysonge CS. Understanding COVID-19 vaccine hesitancy. Nat Med 2021 Aug;27(8):1338-1339. [doi: 10.1038/s41591-021-01459-7] [Medline: 34272500]
23. Forni G, Mantovani A, COVID-19 Commission of Accademia Nazionale dei Lincei, Rome. COVID-19 vaccines: where we stand and challenges ahead. Cell Death Differ 2021 Feb;28(2):626-639 [FREE Full text] [doi: 10.1038/s41418-020-00720-9] [Medline: 33479399]
24. Rosenthal S, Cummings CL. Influence of rapid COVID-19 vaccine development on vaccine hesitancy. Vaccine 2021 Dec 20;39(52):7625-7632 [FREE Full text] [doi: 10.1016/j.vaccine.2021.11.014] [Medline: 34802786]
25. Kashte S, Gulbake A, El-Amin Iii SF, Gupta A. COVID-19 vaccines: rapid development, implications, challenges and future prospects. Hum Cell 2021 May 07;34(3):711-733 [FREE Full text] [doi: 10.1007/s13577-021-00512-4] [Medline: 33677814]
26. Coccia M. High health expenditures and low exposure of population to air pollution as critical factors that can reduce fatality rate in COVID-19 pandemic crisis: a global analysis. Environ Res 2021 Aug;199:111339 [FREE Full text] [doi: 10.1016/j.envres.2021.111339] [Medline: 34029545]
27. Mathieu E, Ritchie H, Ortiz-Ospina E, Roser M, Hasell J, Appel C, et al. A global database of COVID-19 vaccinations. Nat Hum Behav 2021 Jul;5(7):947-953. [doi: 10.1038/s41562-021-01122-8] [Medline: 33972767]
28. Coccia M. COVID-19 pandemic over 2020 (withlockdowns) and 2021 (with vaccinations): similar effects for seasonality and environmental factors. Environ Res 2022 May 15;208:112711 [FREE Full text] [doi: 10.1016/j.envres.2022.112711] [Medline: 35033552]
29. Harrison EA, Wu JW. Vaccine confidence in the time of COVID-19. Eur J Epidemiol 2020 Apr;35(4):325-330 [FREE Full text] [doi: 10.1007/s10654-020-00634-3] [Medline: 32318915]
30. Pickton DW, Wright S. What's SWOT in strategic analysis? Strat Change 1998 Mar;7(2):101-109. [doi: 10.1002/(sici)1099-1697(199803/04)7:2<101::aid-jsc332>3.0.co;2-6]

31. Wang J, Wang Z. Strengths, Weaknesses, Opportunities and Threats (SWOT) analysis of China's prevention and control strategy for the COVID-19 epidemic. Int J Environ Res Public Health 2020 Mar 26;17(7):2235 [FREE Full text] [doi: 10.3390/ijerph17072235] [Medline: 32225019]
32. Zhou Y, Bai L, Guo H, Guo S, Han X, Yue NJ, et al. SWOT analysis and preliminary study on prevention and control management of temporary integrated isolation ward during COVID-19 outbreak. Front Public Health 2021 Mar 15;9:558565 [FREE Full text] [doi: 10.3389/fpubh.2021.558565] [Medline: 33791264]
33. Miranda AE, Rosadas C, Assone T, Pereira GFM, Vallinoto ACR, Ishak R. Strengths, Weaknesses, Opportunities and Threats (SWOT) analysis of the implementation of public health policies on HTLV-1 in Brazil. Front Med 2022 Apr 7;9:859115 [FREE Full text] [doi: 10.3389/fmed.2022.859115] [Medline: 35462992]
34. Jasiulewicz-Kaczmarek M. SWOT analysis for Planned Maintenance strategy-a case study. IFAC-PapersOnLine 2016;49(12):674-679. [doi: 10.1016/j.ifacol.2016.07.788]
35. Asgary NH, McNulty RE. Contributions and challenges in the struggle to end poverty: the case of Kiva. Inf Technol Develop 2017 Apr 06;23(2):367-387. [doi: 10.1080/02681102.2016.1233856]
36. Sharma P, Pardeshi G. Rollout of COVID-19 vaccination in India: a SWOT analysis. Disaster Med Public Health Prep 2022;16(6):2310-2313 [FREE Full text] [doi: 10.1017/dmp.2021.111] [Medline: 33820581]
37. Thakur H. A strengths, weaknesses, opportunities, and threats analysis of public health in India in the context of COVID-19 pandemic. Indian J Community Med 2021;46(1):1. [doi: 10.4103/ijcm.ijcm_1053_20]
38. Murewanhema G, Musuka G, Denhere K, Chingombe I, Mapingure MP, Dzinamarira T. The landscape of COVID-19 vaccination in Zimbabwe: a narrative review and analysis of the strengths, weaknesses, opportunities and threats of the programme. Vaccines 2022 Feb 09;10(2):262 [FREE Full text] [doi: 10.3390/vaccines10020262] [Medline: 35214720]
39. Kopilaš V, Buzas N, Carmesund U, Ceausu I, Chenhall R, de FC, et al. The collection of narratives on strategies, ethics and public response of COVID-19 national vaccination process written by members of the scholarly association Navigating Knowledge Landscapes Network. Mendeley Data. 2022. URL: https://data.mendeley.com/datasets/bwdf6v5wgr/1 [accessed 2022-09-06]
40. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
41. Byrne D. A worked example of Braun and Clarke's approach to reflexive thematic analysis. Qual Quant 2021 Jun 26;56(3):1391-1412. [doi: 10.1007/s11135-021-01182-y]
42. Coronavirus (COVID-19) Vaccinations. Our World in Data. URL: https://ourworldindata.org/covid-vaccinations [accessed 2023-07-03]
43. Plotkin S. History of vaccination. Proc Natl Acad Sci U S A 2014 Aug 26;111(34):12283-12287 [FREE Full text] [doi: 10.1073/pnas.1400472111] [Medline: 25136134]
44. Razai MS, Chaudhry UAR, Doerholt K, Bauld L, Majeed A. Covid-19 vaccination hesitancy. BMJ 2021 May 20;373:n1138. [doi: 10.1136/bmj.n1138] [Medline: 34016653]
45. Moltu C, Stefansen J, Svisdahl M, Veseth M. Negotiating the coresearcher mandate - service users' experiences of doing collaborative research on mental health. Disabil Rehabil 2012;34(19):1608-1616. [doi: 10.3109/09638288.2012.656792] [Medline: 22489612]
46. Blattman C. Why we fight: the roots of war and the paths to peace. Westminister, London: Penguin Books; 2023.
47. Torjesen I. Covid-19 vaccine shortages: what is the cause and what are the implications? BMJ 2021 Mar 19;372:n781. [doi: 10.1136/bmj.n781] [Medline: 33741547]
48. Liu K, Lou Y. Optimizing COVID-19 vaccination programs during vaccine shortages. Infect Dis Model 2022 Mar;7(1):286-298 [FREE Full text] [doi: 10.1016/j.idm.2022.02.002] [Medline: 35233475]
49. Moehring A, Collis A, Garimella K, Rahimian MA, Aral S, Eckles D. Providing normative information increases intentions to accept a COVID-19 vaccine. Nat Commun 2023 Jan 09;14(1):126. [doi: 10.1038/s41467-022-35052-4] [Medline: 36624092]
50. Viswanath K, Bekalu M, Dhawan D, Pinnamaneni R, Lang J, McLoud R. Individual and social determinants of COVID-19 vaccine uptake. BMC Public Health 2021 Apr 28;21(1):818 [FREE Full text] [doi: 10.1186/s12889-021-10862-1] [Medline: 33910558]
51. Cascini F, Pantovic A, Al-Ajlouni YA, Failla G, Puleo V, Melnyk A, et al. Social media and attitudes towards a COVID-19 vaccination: a systematic review of the literature. EClinicalMedicine 2022 Jun;48:101454 [FREE Full text] [doi: 10.1016/j.eclinm.2022.101454] [Medline: 35611343]
52. Gottlieb M, Dyer S. Information and disinformation: social media in the COVID-19 crisis. Acad Emerg Med 2020 Jul;27(7):640-641 [FREE Full text] [doi: 10.1111/acem.14036] [Medline: 32474977]
53. Kopilaš V, Gajović S. Wildfire-like effect of a WhatsApp campaign to mobilize a group of predominantly health professionals with a university degree on a health issue: infodemiology study. J Med Internet Res 2020 Aug 10;22(8):e17051 [FREE Full text] [doi: 10.2196/17051] [Medline: 32442138]
54. Xie L, Wang D, Ma F. Analysis of individual characteristics influencing user polarization in COVID-19 vaccine hesitancy. Comput Human Behav 2023 Jun;143:107649 [FREE Full text] [doi: 10.1016/j.chb.2022.107649] [Medline: 36683861]

55. Raus K, Mortier E, Eeckloo K. Ethical issues in managing the COVID-19 pandemic. Bioethics 2021 Jul 05;35(6):581-588 [FREE Full text] [doi: 10.1111/bioe.12877] [Medline: 33951206]
56. Coccia M. Pandemic prevention: lessons from COVID-19. Encyclopedia 2021 May 31;1(2):433-444. [doi: 10.3390/encyclopedia1020036]
57. Roy DN, Azam MS, Biswas M, Islam E. Potential factors influencing COVID-19 vaccine acceptance and hesitancy among university students in Bangladesh: a cross-sectional comparative study. Epidemiol Infect 2022 Dec 20;151:e11 [FREE Full text] [doi: 10.1017/S0950268822001820] [Medline: 36537315]
58. Solís Arce JS, Warren SS, Meriggi NF, Scacco A, McMurry N, Voors M, et al. COVID-19 vaccine acceptance and hesitancy in low- and middle-income countries. Nat Med 2021 Aug;27(8):1385-1394 [FREE Full text] [doi: 10.1038/s41591-021-01454-y] [Medline: 34272499]
59. Hong L, Jin Z, Xu K, Shen G, Zou Y, Li R, et al. COVID-19 vaccine uptake and vaccine hesitancy in rural-to-urban migrant workers at the first round of COVID-19 vaccination in China. BMC Public Health 2023 Jan 20;23(1):139 [FREE Full text] [doi: 10.1186/s12889-023-15068-1] [Medline: 36658507]
60. Abou-Arraj NE, Maddah D, Buhamdan V, Abbas R, Jawad NK, Karaki F, et al. Perceptions of, and obstacles to, SARS-CoV-2 vaccination among adults in Lebanon: cross-sectional online survey. JMIR Form Res 2022 Dec 14;6(12):e36827 [FREE Full text] [doi: 10.2196/36827] [Medline: 36383635]

Abbreviations

NKL: Navigating Knowledge Landscape
SWOT: strengths, weaknesses, opportunities, and threats

Edited by A Kushniruk; submitted 12.11.22; peer-reviewed by M Coccia, B Hausman, Y Mejova; comments to author 01.06.23; revised version received 16.07.23; accepted 26.09.23; published 19.02.24.

Please cite as:
Kopilaš V, Nasadiuk K, Martinelli L, Lhotska L, Todorovic Z, Vidmar M, Machado H, Svalastog AL, Gajović S
Perspectives on the COVID-19 Vaccination Rollout in 17 Countries: Reflexive Thematic and Frequency Analysis Based on the Strengths, Weaknesses, Opportunities, and Threats (SWOT) Framework
JMIR Hum Factors 2024;11:e44258
URL: https://humanfactors.jmir.org/2024/1/e44258
doi:10.2196/44258
PMID:38373020

Original Paper

The Role of Social Media in the Experiences of COVID-19 Among Long-Hauler Women: Qualitative Study

Camryn Garrett[1], MPH; Atefeh Aghaei[1], MA, PhD; Abhishek Aggarwal[1], BA; Shan Qiao[1], MA, PhD

[1]Department of Health Promotion, Education, and Behavior, Arnold School of Public Health, University of South Carolina, Columbia, SC, United States

Corresponding Author:
Camryn Garrett, MPH
Department of Health Promotion, Education, and Behavior
Arnold School of Public Health
University of South Carolina
915 Greene Street
Columbia, SC, 29208
United States
Phone: 1 803 777 6844
Email: camrynmg@email.sc.edu

Abstract

Background: The extant literature suggests that women are more vulnerable to COVID-19 infection and at higher risk for developing long COVID. Due to pandemic mitigation recommendations, social media was relied upon for various aspects of daily life, likely with differences of usage between genders.

Objective: This study aimed to explore the role and functions of social media in the lives of long-hauler women.

Methods: Participants were purposively snowball-sampled from an online health promotion intervention for long-hauler women with COVID-19 from March to June 2021. During this time, one-on-one, semistructured interviews were conducted online until data saturation was agreed to have been achieved (ie, 15 interviews). Interview transcripts and field notes were analyzed using an emergent, inductive approach.

Results: In total, 15 women were enrolled. The main roles of social media included facilitating support group participation, experience sharing, interpersonal connections, and media consumption. Emergent themes demonstrated that participants rely on social media to fulfill needs of emotional support, social engagement, spirituality, health planning, information gathering, professional support, and recreationally for relaxation. As long-hauler women turn to social media to discuss symptom and health management as well as the intention to vaccinate, this study demonstrates both the associated benefits (ie, decreased isolation) and challenges (ie, misinformation, rumination, resentment, jealousy).

Conclusions: The public health implications of these findings support the development of gender-tailored health promotion interventions that leverage the benefits of social media, while mitigating the negative impacts, for women with long COVID.

(JMIR Hum Factors 2024;11:e50443) doi:10.2196/50443

KEYWORDS

COVID-19; long COVID; long-haulers; women; gender; social media; digital media; qualitative study

Introduction

The COVID-19 Context

The COVID-19 pandemic, caused by SARS-CoV-2, was declared on March 11, 2020, and was accompanied by recommendations to implement prevention measures, such as masking, vaccination, testing, social distancing, isolation, and quarantine [1-5]. In the United States, as of June 16, 2023, there had been more than 103.4 million cases of COVID-19, of which more than 6.1 million required hospitalization and more than 1.1 million resulted in death [6,7]. Of those who become infected, about 30% develop postacute sequela SARS-CoV-2 (PASC), also known as chronic COVID-19 or long COVID, characterized by symptoms of varying severity that persist for 4 weeks or more after infection (eg, chronic fatigue, pain, cognitive dysfunction, muscle deconditioning, impaired concentration, and persistent ageusia and anosmia); these patients are commonly referred to as long-haulers [8-13]. Overall, women have been found to be more likely than men to develop long COVID (ie, 9.4% vs 5.5%) [14-16]. This disproportionate trend requires further investigation into the differential experiences of long COVID among women.

Long COVID Among Women

The literature shows that there may be an association between biological sex and COVID-19 infection and recovery; however, this fails to consider the role of gender, the social environment, and gendered social norms [17]. For instance, women primarily constitute social assistance and health care workforces and face increased expectations of caregiving in the family setting, increasing the risk of COVID-19 infection [18]. As long COVID results from COVID-19 infection, differential exposure and incidence among women predisposes them to the risk for developing persisting symptoms [19]. Persistent symptoms experienced more so by long-hauler women include fatigue, difficulty breathing, muscle pain, and cognitive dysfunction, as well as the negative psychosocial outcomes of anxiety, depression, and posttraumatic stress disorder (PTSD), particularly among those who have been hospitalized [20,21].

Social Media and COVID-19

Due to stay-at-home orders and the prioritization of social distancing as the primary means to prevent the spread of COVID-19, in the online environment, social media emerged as a key tool to adjust to the new normal, necessitating research on its roles and functions. Overall, in the United States, 97% of Americans indicate owning a cell phone of any kind, 85% indicate owning a smartphone, and 85% of US households have a broadband internet connection [22,23]. Using these technologies, 85% of US adults indicate going online at least once a day and 31% indicate that they are online "almost constantly" [22]. Among mixed findings in the COVID-19 literature, women report higher usage of social media compared to men, with assumedly differential motivations for engagement and use of platform functions [24]. Due to the prevalence of individuals being online for work and personal use, it is necessary to evaluate the role of technology and social media within the context of the pandemic and, specifically, the experiences of long-hauler women [25].

Among long-haulers specifically, social media played a vital role in developing the long-hauler identity and encouraging clinical acknowledgement. The term "long COVID" originates from social media users' posts online [26]. Posts sharing long COVID experiences typically include a diagnosis or test result, the symptoms experienced, the length of time symptoms have persisted, an emotional response, and information and resources [27]. The growing conversations among long-haulers on social media shifted the experience of long COVID from anecdotal, exposing an invisible disability, to clinical [26,27]. With the creation of a shared identity, long-haulers were able to identify one another and further subdivided themselves into categories accounting for their intersecting identities (eg, long-hauler, woman, and mother).

Social media has been used to mitigate the impacts of lost social connections, social distancing, and isolation [21,22]. In the literature, social media has been demonstrated as a key tool used to maintain social connections, while adhering to social distancing recommendations, limiting feelings of isolation [28]. Additionally, as loneliness has been associated with decreased use of healthful coping behaviors, social media has been found to mediate the association during periods of isolation, such as during the COVID-19 pandemic [29]. There are linkages between daily use of social media and lowered measures of social isolation, as well as inversely with infrequent use of social media and higher measures of isolation [30]. Despite the positive outcomes associated with using social media during the pandemic, downfalls remain. For example, in a study focusing on older adults, internet use reflected coping efforts but did not necessarily enhance or sufficiently improve well-being [31]. Due to the mixed effects and roles of social media, throughout the pandemic, there is a unique opportunity for researchers to investigate the role of social media in social connections, isolation, and support, as well as in perpetuating access to information or misinformation among long-hauler women [32].

Overall, the literature suggests that social media sites impact users' ability to maintain social connections, seek social support, and access information, as well as affect isolation, social comparison, and the spread of misinformation [33,34]. According to the information systems literature, gender is associated with differential motivations to use social media sites (eg, relational uses for women vs information gathering for men) and differential perceptions of information shared online [35,36]. To the best of our knowledge, despite the gendered associations relevant to social media use in other fields, few studies have assessed the differential role of social media during the pandemic, by gender, from the public health perspective. Due to an overwhelming focus on women's experiences as essential workers and with reproductive care during the COVID-19 pandemic, there is scant literature more broadly centering on women. The experiences of women were chosen as a focus in this work due to their disproportionate burden of long COVID and their higher rates of activity and gender-specific engagement patterns on social media sites [37]. This work therefore aimed to fill a gap in the extant literature by investigating the role of social media in the experiences of long-hauler women alone.

Methods

Study Design

The data used in this study were derived from an online health promotion intervention for long-hauler women with COVID-19. Participants were recruited using snowball and purposive sampling through 2 social media sites, Facebook and Slack; the participants were recruited from 16 Facebook groups and 1 Slack group, as well as 2 websites of organizations for long-hauler women. Those eligible to participate in the study met the inclusion criteria of living in the United States, being aged 18 years or older, who spoke English, and who self-identified as long-haulers due to persistent COVID-19 symptoms for 4 weeks or more after infection.

Ethical Considerations

The University of South Carolina Institutional Review Board (Pro00109358) reviewed and approved the study protocol.

Recruitment began after group and organization administrators approved posts including the study description, a flyer, and researcher contact information. The study was then advertised in each group.

Recruitment

The recruitment period spanned 2.5 months from March to June 2021. After screening for eligibility and receiving informed consent, a total of 15 semistructured, one-on-one interviews were conducted from April to June 2021 using the online videoconferencing software Zoom [38]. Each interview lasted between 30 and 50 minutes. Participant demographics were collected through the interview process. All interviews were audio-recorded and, upon completion, field notes were written. Each participant was compensated with a US $30 e-gift card for their time and effort spent participating in the study. Data saturation was agreed to have been reached, by the 2 researchers involved in interview coding, after 15 interviews.

Data Collection

Data were collected on the participants' self-reported long-hauler status, the impact of persistent COVID-19 symptoms on their lives, coping strategies, and overall experiences. In these conversations, discussions of the roles of technology and social media arose organically following the semistructured interview guide. All interviews were recorded and transcribed using the service Otter.ai [39]. All artificial intelligence–derived transcripts were reviewed and verified by members of the research team. Interviewer field notes were used as additional data.

Data Analysis

The data were analyzed following a predominately inductive approach for the thematic analysis of the interviews, as the themes identified were derived directly from the data [40-42]. The analysis process was comprised of 6 stages beginning with data familiarization and preliminary code construction, followed by the obtaining, revising, labeling, and reporting of key themes [43]. MAXQDA software was used to analyze interview transcripts [44]. In the initial phase of the thematic analysis, 2 members of the research team independently coded the transcripts, following an open coding scheme, to identify emergent themes [45-47]. The initial development of the codebook was performed after half of the interviews (ie, 7) were coded. We discussed at length the creation of the codebook to ensure accuracy of the initial codes, themes, domains, definitions, exemplar quotes, and organization. Once the codebook was finalized, the same 2 members of the research team continued to independently code the remaining transcripts. We then engaged in a collaborative review process to confirm alignment with the final codebook and to ensure consistency in the application of codes. In comparing themes, we identified similarities, differences, and interactions between themes. We used an axial coding approach to categorize the main themes and subthemes, which then guided the selection of direct quotes to demonstrate the key findings. Peer debriefing and intercoder agreement techniques were used to ensure reliability throughout the data analysis [48,49].

Results

Participant Details

The study participants, in alignment with the inclusion criteria, all identified as women. The participants were primarily aged between 36 and 65 years (n=12, 80%), served as essential workers (n=9, 60%), and lived with others (n=13, 87%) in the eastern region of South Carolina (n=10, 67%). Table 1 lists the participant details.

Table 1. Demographic characteristics of long-hauler women (N=15).

Characteristics	Participants, n (%)[a]
Age (years)	
20-35	2 (13)
36-50	6 (40)
51-65	6 (40)
>65	1 (7)
Occupation	
Health care provider	5 (33)
Educator	4 (27)
Business owner	4 (27)
Student	1 (7)
Retiree	1 (7)
Living situation	
Living with others	13 (87)
Living alone	2 (13)
Regional location	
East	10 (67)
Central	3 (20)
West	2 (13)

[a]The percentages might add up to more than 100 because of rounding.

Benefits and Challenges

Long-hauler women indicated that their most used social media features included participating in support groups, posting, commenting, connecting with others, and consuming media. They used these features of social media sites to fulfill needs such as emotional support, social engagement, spirituality, health planning, information gathering, professional support, and recreation. The different functions of social media also resulted in a variety of benefits and challenges throughout the participants' coping with long COVID. Tables 2 and 3 present the benefits and challenges related to the themes and subthemes identified and exemplar quotes.

Table 2. Emergent themes and subthemes of the beneficial roles of social media identified by long-hauler women.

Themes and subthemes	Exemplar quotes
Social connection	
Group membership	“So I said, ‘Well, maybe if I am part of something…it is gonna be like a, like a, like a motivation for me to go through something for me. Um, because again, we are always thinking about others, you know, like, what, why am I complaining.”
Social support	“I may not exactly know what you’re going through, but I am here to help and here to listen because a lot of times you just want a hearing—somebody to hear you.”
Network building	“I also was posting and doing that, which, like, kept me motivated [to continue] sharing and connecting with other people…”
Belonging	“I probably gravitated more towards that group…and I would talk about that Facebook group a lot. Like, it felt like that was like a support group, and it felt like, you know, I am not crazy, like some other people are having it, too. And I would be active in, like, commenting on, like, you know, answering people’s questions or, like, sharing, like, a connection that I have with another person that wrote on there.”
Religiosity and spirituality	
Prayer	“I join[ed] an online group for praying.”
Fellowship	The loss of a group member highlighted a sense of duty and belongingness toward one another in the online prayer group.
Online worship	“I will go to one of my favorite pastors on YouTube and listen.”
Meditation	“I am in a meditation group that I go to online, and we do meditation together. And then, there is, like, headspace and calm, those apps. So, there is a wide variety of different things. Like chakras, and then there is, you know, just all different kinds of relaxation and tension. Like, you squeeze your arms and look at your feet.”
Information gathering	
Long hauler–shared information	“It is kind of, like, you form your own little support groups of people that had COVID. And, you know, their symptoms vary, and you are like, ‘Oh, what did you do for this?’ Or like the hair loss. That is another thing—hair loss. My hair is still not well, or whatever. And then, you know, people debating, like, ‘Are you taking the vaccine? Are you not getting the vaccine?’ So having those little groups to talk—it is good.”
Symptom management	“I downloaded an app on my phone, and I am monitoring, like, I am documenting all of my activities for the day every day so that I can document, like, different symptoms that I am having and, like, what is, like, a trigger.”
Physician-shared information	In reference to streaming YouTube videos: “…the different doctors and, like, what their findings are, what their recommendations are.”
Recreation	
Entertainment	“I will allow myself; it does not happen every day, but, like, just to play some mind games, you know, a game of solitaire or a game…on my phone just to give myself a break.”
Relaxation	“To help go to sleep at night. They try to, kind of, get me to relax, or whatever. And so, I think the biggest thing for me is disconnecting from all the things that I have going on, and I just…I struggle with that.”

Table 3. Emergent themes and subthemes of the challenging roles of social media identified by long-hauler women.

Theme and subthemes	Exemplar quotes
Social connection	
Anxiety	"They can really increase anxiety."
Resentment	"…it was hard. There would be resentment, and there is resentment now with the group, too, as terrible as it sounds, even jealousy, because I will see people that will write on Facebook, like, 'Oh, like, I had COVID in December 2020' or 'I had COVID in January 2021,' and a part of me just, like, would hate it because it is…like you knew, like people were advising you not to travel. And that would be what it was, especially around the holiday season, hearing people talk about, you know, having so-and-so over from, like, California, and then they got sick afterwards. And, like, it just makes you go crazy because there is so much more now. But I am trying not to, like, think like that because I mean, I do not know everyone's experiences, and maybe, they really did avoid it or did their best to not get it, and they got it because we are in a pandemic."
Information gathering	
Misinformation and health literacy	"Sometimes, I can understand a lot of the stuff, but there is some things that I am not as familiar with…"
Oversaturation and pandemic fatigue	"…after a while, like, even that [social media use] got to be so overwhelming because, again, like, everyone is, like, posting the same thing."

Social Connection

A majority of the women interviewed highlighted the role of social media in reducing social isolation by providing social connection. Social connections were found to be fostered through multiple functions of social media, such as personal networks, following networks, and group membership, as well as more broadly through engagement with other users, known or unknown. Long-hauler women emphasized the importance of social media as a tool to maintain connection with their social networks when unable to be present in person. Emergent subthemes related to the main theme of social connection included the benefits of belonging fostered through social support, group membership, and network building, as well as the challenges of anxiety and resentment.

Group Membership and Network Building

One participant noted the function of social media in mediating "the loss of family time and not being able to be together and doing the things we have always done as a family." Another participant described the stress associated with physical, in-person gatherings in the time of COVID-19:

> *I tried to host a barbecue out in our little place at the lake, and it caused me so much anxiety. I could not even eat my birthday barbecue, could not really interact with people.*

At a time when minimizing physical contact with others was recommended, the online environment was found to aid in maintaining social health. Networking, a distinct feature of social media platforms, connecting individuals with others they may or may not be geographically close to, emerged as instrumental to long-hauler women's social connection and, further, social support. Participants shared motivations for seeking membership and experiences as members of online support groups for COVID-19 long-haulers. One participant noted:

> *I was looking for, you know, for common ground, for folks that were experiencing some of those same things that I was, and I was also looking to support them with what I knew about my mind, body, [and] skills…*

Facebook emerged as a popular social media platform among long-hauler women due to its functionality to host support groups. Many long-hauler women reported using Facebook groups to build their social networks, while also providing social and emotional support to other long-haulers. Upon reflecting on her participation in online social support groups, a participant shared:

> *I probably gravitated more toward that group…and I would talk about that Facebook group a lot. Like, it felt like that was like a support group, and it felt like, you know, I am not crazy, like some other people are having it, too. And I would be active in, like, commenting on, like, you know, answering people's questions or, like, sharing, like, a connection that I have with another person that wrote on there.*

Long-hauler women demonstrated the role of online groups in expanding their social networks to include other long-haulers outside their direct networks. As a result of their group membership, the majority of the participants indicated providing and receiving emotional and instrumental social support through connections fostered by membership in online support groups.

Social Support

Further, participants explained the role of online groups in facilitating social support from connections because "[they] are experiencing similar things that I am experiencing, so I know that it is not just me." One participant described her role in providing emotional social support through online social connections:

> *I may not exactly know what you are going through, but I am here to help and here to listen because a lot of times you just want a hearing—somebody to hear you.*

Participants demonstrated the crucial role of validation and affirmation as emotional social support when received from other group members regarding their emotions, symptoms, and overall experiences. Participants indicated receiving validation and affirmation when posting, commenting, and being active within their support groups. One participant discussed the benefits of continued engagement in these groups:

> *I also was posting and doing that, which, like, kept me motivated [to continue] sharing and connecting with other people…*

As a result of providing and receiving social support within their online networks of long-hauler women, the majority of the participants report a positive effect on their sense of belonging.

Belonging

We found that because social media is able to connect users, participation on the platforms and in groups aids in maintaining social health through online belongingness, while also adhering to public health recommendations (eg, social distancing, isolation, quarantine). In bolstering social connections and aiding in emotional regulation, another long-hauler explained that social media motivates her to remain strong and encourages resilience. She explained:

> *So I said, well, maybe if I am part of something…it is gonna be like a, like a, like a motivation for me to go through something for me. Um, because again, we are always thinking about others, you know, like, what, why am I complaining.*

Another participant noted the benefit of belonging to a support group:

> *I joined the COVID long-haulers' Facebook group because another new thing with my shortness of breath is I noticed if I eat a lot at one time, I am way shorter of breath, and I do not know why. So, I was, like, "Oh, I am gonna see if anybody else has had these symptoms. So, I actually, like, made a post about it. And I liked that group because it makes you realize, like, you are not alone. There is all these other people that also do not have answers and also have similar symptoms as you.*

Participants described how their group membership and sense of belonging decreased their feelings of loneliness and isolation, particularly when sharing experiences and symptoms with other long-haulers.

Anxiety and Resentment

Despite the potential benefits of participating in online groups, there remain potential consequences of participation as well. Although the findings indicated social media aids in mitigating feelings of loneliness and creating a sense of belonging, they also indicated increasing anxiety and resentment among long-hauler women. One participant noted that "they can really increase anxiety." As related to seeing the posts of others within their social networks and in groups, a participant said:

> *…it was hard. There would be resentment, and there is resentment now with the group, too, as terrible as it sounds, even jealousy, because I will see people that will write on Facebook like "Oh, like, I had COVID in December 2020" or "I had COVID in January 2021," and a part of me just, like, would hate it because it is…like you knew, like people were advising you not to travel. And that would be what it was, especially around the holiday season, hearing people talk about, you know, having so-and-so over from, Like, California, and then they got sick afterwards. And, like, it just makes you go crazy because there is so much more now. But I am trying not to, like, think like that because I mean, I do not know everyone's experiences, and maybe, they really did avoid it or did their best to not get it, and they got it because we are in a pandemic. But it is stuff like that. Like, I feel like I am more, like, insecure with my experience. I get jealous of other people's experiences. There's just, like, a lot of negative-ness with it…*

In sharing this anecdote, the participant voiced her frustration toward and resentment of those who, after participating in high-risk activities, shared their COVID-19 experiences online. Engagement with such individuals and their posts then led to this participant's insecurity in their own experiences.

Religiosity and Spirituality

In addition to the impacts of social media on social health, participants highlighted its role in also maintaining their spiritual health. In addition to joining online groups topically centered around COVID-19, a participant indicated, "I join[ed] an online group for praying." She detailed the group, demonstrating its resemblance to that of other support groups, albeit not solely related to COVID-19, with the added element of religion. Overall, the participant's sentiments indicated that the group positively impacted her overall well-being. When describing the loss of a member of the group, she highlighted the role of fellowship and connection in the group as they lifted one another up in prayer and, in doing so, created belongingness, community, and strength.

Another participant demonstrated the role of social media as related to religiosity and spirituality by noting her use of video-streaming platforms to seek spiritual support. She described her engagement as, "I'll go to one of my favorite pastors on YouTube and listen." During a time when physically gathering with others, as in the case of congregating for religious observances, was considered high risk, social media provided an avenue through which long-haulers could maintain their spiritual practices. Relatedly, participants indicated using social media to engage in guided meditations. One shared her daughter's role in encouraging her participation:

> *She gave me some resources online, in an app, and then my daughter uses a different…she uses Spotify. So, she gave me that information, and so I kind of just went off of those suggestions, and now, I have my favorite guided meditations that I use on Spotify, and they are effective.*

Other participants said that they similarly engage in guided meditations but also participate in groups specific for meditation and relaxation. One participant described the meditation group and smartphone apps used:

> *I am in a meditation group that I go to online, and we do meditation together. And then, there is, like, headspace and calm—those apps. So, there is a wide variety of different things. Like chakras, and then there is, you know, just all different kinds of relaxation and tension. Like you squeeze your arms and look at your feet.*

Information Gathering

Apart from social networking, one of the most prominent functions of social media is the sharing of news and information. Within the context of the COVID-19 pandemic, social media served as a conduit for sharing COVID-19 news, government policies and announcements, updated prevention guidelines, and general information. The findings demonstrated that long-hauler women used social media to seek information related to COVID-19 vaccines, symptoms, and symptom management strategies, as well as to follow news related to emerging treatments.

Long Hauler–Shared Information

Regarding COVID-19 information, topics of interest were primarily related to symptoms and health management. Long-hauler women indicated turning to online support groups to gather information from those with similar experiences. One participant illustrated the symptom and health discussions within these groups:

> *It is kind of like you form your own little support groups of people that had COVID. And, you know, their symptoms vary, and you are like, "Oh, what did you do for this?" Or like the hair loss. That is another thing—hair loss. My hair is still not well, or whatever. And then, you know, people debating, like, "Are you taking the vaccine? Are you not getting the vaccine?" So having those little groups to talk—it is good.*

Alongside using support groups for discussion, long-haulers indicated also using smartphone apps to track symptoms and create health care plans. One long-hauler discussed her experience:

> *I downloaded an app on my phone, and I am monitoring, like, I am documenting all of my activities for the day every day so that I can document, like, different symptoms that I am having and, like, what is, like, a trigger.*

Due to the persistent nature of COVID-19 symptoms experienced by long-haulers, monitoring symptoms is in the interest of patients to aid in symptom management and for use with health care providers in creating treatment plans.

Physician-Shared Information

In addition to sharing information across networks of long-haulers on social media, participants also noted gathering information through online interactions with physicians and mental health professionals. One participant indicated obtaining pandemic-related information from physicians on YouTube as she watched "…the different doctors and, like, what their findings are, what their recommendations are." Due to the increasing burden of mental health challenges coupled with physical symptoms, as expressed by the participants, social media offers a platform for mental health resource sharing, at a time when many cannot access needed services. One participant detailed these difficulties:

> *I had been looking for, like, counseling, and a lot of the counseling in our plan has, like, basically stopped taking people. Like, I think it is, like, kind of like, overwhelmed right now, and, like, I would call, like, a whole list, and I would go through the whole list, and, like, they are not taking new patients. So, I just have to be persistent about it.*

In coping with barriers (ie, wait lists, cost) to accessing mental health services, participants indicated using social media as a tool to gain information from professionals. For instance, a participant said that she "join[ed] a group…they had a list of faculty members that were starting groups…you did not have to pay for it." This participant was able to engage in mental health services through a free and accessible online group operated by mental health professionals. This function of social media is valuable in responding to increasing mental health needs by addressing barriers to accessing professional psychological support.

Misinformation and Health Literacy

The potential consequences of users obtaining information from social media, particularly that which must be scientifically based, include a lack of or difficulty in understanding, as well as the distribution of and access to unvalidated content or misinformation. Illustrating the difficulty in understanding sought-out information, a participant shared:

> *Sometimes I can understand a lot of the stuff, but there is some things that I am not as familiar with…*

Due to the evolving nature of scientific discovery over the course of the pandemic, there were difficulties in grasping timelines and emergent findings that inhibited understanding and perpetuated misunderstanding. In the case of long-haulers, their increased need for health care exposes them to complex medical jargon that may require a higher level of health literacy to mitigate misunderstanding. Overall, due to the need for regularly updated COVID-19 information, social media functions as both a benefit and a hindrance to its dissemination. Social media provides users with increased access to information, while also providing a platform through which misinformation may be widely shared.

Oversaturation and Pandemic Fatigue

Further, despite the benefits of engaging in support groups and accessing pandemic-related information online, participants indicated differing perspectives on the amount of information shared. Referencing a long-hauler Facebook support group, a participant noted:

> *And, like, the nice thing about it is they share loads of information.*

Alternatively, another participant shared that due to the sameness and sheer volume of pandemic-related content on social media:

> *…after a while, like, even that [social media use] got to be so overwhelming because, again, like everyone is, like, posting the same thing.*

Due to oversaturation and misinformation, a participant noted that she is "disappointed with social media." This disappointment has kept the participant from participating in COVID-19 and long-hauler groups.

Recreation

In addition to the networking and information-gathering functionalities of social media, participants also indicated leveraging social media for entertainment, recreation, and relaxation. In coping with their diagnosis, symptoms, anxiety, and the state of the world, long-hauler women indicated using social media and smartphone apps to play games, watch videos, and listen to music. One participant described consuming content on social media as a method to cope with the anxiety of attending post–COVID-19 appointments. Another participant shared:

> *I will allow myself; it does not happen every day, but, like, just to play some mind games, you know, a game of solitaire or a game…on my phone just to give myself a break.*

Another participant indicated using the social media site YouTube as a way "to help go to sleep at night."

> *They try to, kind of, get me to relax or whatever. And so, I think the biggest thing for me is disconnecting from all the things that I have going on, and I just…I struggle with that.*

These findings suggest that social media is a method by which participants seek entertainment, recreate, and relax. These functions serve as social media–based coping mechanisms to alleviate mental health burdens.

Discussion

Principal Findings

Long-hauler women identify engagement in online support groups to be a primary use of social media during the COVID-19 pandemic. These groups are typically disease specific and can be described as communities where individuals can congregate and engage in broader group discussions as a form of social connection [50]. Support groups function to allow members to affirm their long-hauler identity, maintain connections, combat isolation, seek support, compare experiences, share remedies, and coruminate [51-53]. Long-hauler women seek reassurance through channels of connection with others who share their disease-specific identity and to cope with a social environment characterized by mortality, unemployment, resource loss, and psychological burdens of prevention measure adherence and disease [54].

Online support groups have been previously assessed in various disease contexts in the literature. A systematic review of the role of online support groups for patients with prostate cancer found that the groups not only aided in participant decision-making through their dissemination and exchange of information but also provided participants with social support [55]. A review of support groups for patients with breast cancer demonstrated that the benefits or consequences of participation in social support groups are inconclusive [56]. A systematic review of studies assessing the impacts of social support groups on patients with chronic conditions found that they demonstrate a wide array of support group implementation and outcome measurements that complicate their use in the context of the COVID-19 pandemic [57]. Within the context of COVID-19, online support groups act as a tool to comply with social distancing guidance, while maintaining connections and combating isolation, depression, and anxiety [51,52]. Additionally, a systematic review of COVID-19–specific social support groups demonstrated that although they are effective in addressing participants' psychological and psychosocial needs, due to their responsiveness to the emerging needs and challenges faced by participants, there remains a need for further research [58].

This study presented both benefits and challenges associated with the participation of long-hauler women in COVID-19–specific social support groups. The benefits include the validation of shared experiences, decreased isolation, and motivation to pursue symptom management and recovery. The challenges for long-hauler women's participation include experiences of increased anxiety due to rumination within the groups, resentment and jealousy due to others' posts of unsafe pandemic activities or recovery, and an insecurity of experiences as a result of comparison. Additionally, there is the complication of pandemic fatigue, as instigated by the overwhelming amount of posts within social support groups. Within the extant literature, overexposure to pandemic news may act as a disaster stressor that acts as a risk factor for negative psychosocial outcomes [59]. As demonstrated through these findings, related to social support and network building, social media presents an opportunity for individuals to receive support and engagement with others, while also facing potential, associated challenges.

Additional engagement on social media revolved around spirituality, entertainment, recreation, and relaxation. Beyond disease-specific support groups, long-hauler women reported relying on groups that specifically serve to maintain spiritual health. Digital media, more broadly, allows long-hauler women to engage in spiritual practices alone or with others, as desired. These novel functions are significant as spiritual health has been identified as a key coping mechanism to facilitate resilience [60]. Further, digital media has demonstrated its usefulness in the coping of long-hauler women, as they noted its use for entertainment, recreation, and relaxation through audio and visual content. The emerging pandemic literature has sought to assess the complex benefits and consequences of media usage, motivations, stress, and psychosocial outcomes that have been found to differ by demographics [61]. Despite the complex mechanisms of coping within the literature, long-hauler women in this study identified digital and social media used for entertainment to be a positive coping strategy.

In addition to the features of social media facilitating coping, long-hauler women also relied on networking sites to access pandemic-related information. One unique feature of social media is the unprecedented speed with which information can be shared, particularly evolving pandemic information, but it also presents the risk of misinformation and associated difficulties in mitigating its negative impacts [62,63]. A key consequence, due to the nature of social media, is the tendency for information that sparks outrage, typically containing misinformation, to move the most quickly through social channels, likely stifling needed, correct information [63]. Therefore, as outrage impacts the visibility of trending topics, depending on the content, it can alter individuals' risk perceptions [63]. This study corroborated these trends due to participants' disappointment with the distribution of information and with social media overall. Additionally, when evaluating what constitutes misinformation, it is necessary to consider nuances in the perspectives of various key players (eg, patients, providers, scientists), as well as their potential contributions to the knowledge base (eg, symptomology, diagnostic criteria).

As social media content allows misinformation to trend due to its outrage-evoking characteristics, the COVID-19 pandemic is seen as syndemic with an infodemic. Within the infodemic, long-hauler women expressed experiencing difficulty in understanding health information presented online as related to vaccines, symptom management, and news. In addition to the threats posed by misinformation to public health prevention efforts, social media presents users with a plethora of information that operates to mitigate the associated negative effects [33]. Social and cultural factors influencing the perceptions of and responses to health information and risk communication are related to personal control, uncertainty, trust in institutions, and trust in media, as well as an overall sense of immediacy [63].

Associated with COVID-19 information, long-hauler women turn to social media and online social support groups to discuss symptom and health management, as well as the intention to vaccinate. Digital media, beyond social media, has benefited women with chronic COVID-19, allowing them to document and track their symptoms. As chronic COVID-19 is characterized by persistent symptoms, symptom management support and remedy information sharing were found to be salient uses of social media among women with long COVID. This finding aligns with evidence within the extant literature where social media has been used throughout the pandemic, by broader populations, to share medication strategies, anonymously seek information, crowdsource information, and engage in advocacy [64-67]. Overall, social media is used by long-hauler women to cope, exchange social support, maintain spirituality, and seek entertainment, while also disseminating information relevant to the long-hauler experience.

Strengths and Limitations

Our findings are in alignment with "uses and gratifications theory," which posits motivations for social media use as revolving around meeting certain needs, including social connection, knowledge, and relaxation, among others [68,69]. This study contributes to the sparce, evolving literature, with findings focusing on the social media usage of long-hauler women specifically.

Our study is also subject to several limitations. First, there were constraints on the analysis due to a small sample size with limited demographic variability. Second, the structure of the questions asked restricted our ability to identify patterns of usage by platform, relying, rather, on broader trends. Additionally, as support groups were used for recruitment, the findings may not be representative of the experiences of women not engaged on social media or in online support groups.

Public Health Implications

Although this study is additive to the evolving literature, strengthening the present evidence base beyond quantitative, descriptive analyses that do not account for gendered experiences, it demonstrates a need for further research. Future deductive work should consider, concurrently, comparing the uses of social media across the spectrum of gender and age based on known differences in usage. Due to the reliance on social media platforms to gather knowledge, further work is necessitated to evaluate the content and quality of information shared within online discussions and support groups. Future research should use an intersectional framework to assess the role of social media across a variety of additional identities women hold (eg, race/ ethnicity, preexisting conditions, socioeconomic status).

Conclusion

The findings of this study support the development of gender-tailored health promotion interventions that leverage the benefits of social media, while mitigating the consequences, for women with chronic COVID. As social media serves as a pandemic mitigation tool, there is a need to better understand patterns and experiences of usage [70-72]. Informed by our findings, long-hauler women should be met where they are, through the platforms and functions that they currently use, in order for public health interventions to aid them in managing long COVID and its associated effects.

Acknowledgments

This work was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health (grant award number R01AI127203-5S1).

Conflicts of Interest

None declared.

References

1. Yakut ZAC, Sakarcan S, Sener G. What we know about COVID-19 and its treatment. J Res Pharm 2020 Sep 11;24(5):602-616. [doi: 10.35333/jrp.2020.215]
2. Baloch S, Baloch MA, Zheng T, Pei X. The coronavirus disease 2019 (COVID-19) pandemic. Tohoku J Exp Med 2020 May;250(4):271-278 [FREE Full text] [doi: 10.1620/tjem.250.271] [Medline: 32321874]
3. COVID-19 and your health. Centers for Disease Control and Prevention. URL: https://tinyurl.com/mpj37sp8 [accessed 2024-05-01]
4. Advice for the public: coronavirus disease (COVID-19). World Health Organization. URL: https://tinyurl.com/ys6r622x [accessed 2024-05-01]
5. Coronavirus disease 2019 (COVID-19). Centers for Disease Control and Prevention. URL: https://tinyurl.com/ysa769wc [accessed 2024-05-01]
6. COVID data tracker. Centers for Disease Control and Prevention. URL: https://covid.cdc.gov/covid-data-tracker [accessed 2024-05-01]
7. United States of America: WHO coronavirus disease (COVID-19) dashboard with vaccination data. World Health Organization. URL: https://covid19.who.int [accessed 2024-07-01]
8. Bull-Otterson L, Baca S, Saydah S, Boehmer TK, Adjei S, Gray S, et al. Post–COVID conditions among adult COVID-19 survivors aged 18–64 and ≥65 years — United States, March 2020–November 2021. MMWR Morb Mortal Wkly Rep 2022 May 27;71(21):713-717. [doi: 10.15585/mmwr.mm7121e1]
9. Post-COVID conditions. Centers for Disease Control and Prevention. URL: https://tinyurl.com/yc5x7e78 [accessed 2024-07-01]
10. Wong J, Kudla A, Pham T, Ezeife N, Crown D, Capraro P, et al. Lessons learned by rehabilitation counselors and physicians in services to COVID-19 long-haulers: a qualitative study. Rehabil Couns Bull 2021 Dec 14;66(1):25-35. [doi: 10.1177/00343552211060014]
11. Del Rio C, Collins LF, Malani P. Long-term health consequences of COVID-19. JAMA 2020 Dec 03;324(17):1723-1724 [FREE Full text] [doi: 10.1001/jama.2020.19719] [Medline: 33031513]
12. Pincherle A, Jöhr J, Pancini L, Leocani L, Dalla Vecchia L, Ryvlin P, et al. Intensive care admission and early neuro-rehabilitation. Lessons for COVID-19? Front Neurol 2020 Aug 25;11:880 [FREE Full text] [doi: 10.3389/fneur.2020.00880] [Medline: 32982916]
13. Tenforde MW, Kim SS, Lindsell CJ, Billig Rose E, Shapiro NI, Files DC, CDC COVID-19 Response Team, et al. Symptom duration and risk factors for delayed return to usual health among outpatients with COVID-19 in a multistate health care systems network — United States, March-June 2020. MMWR Morb Mortal Wkly Rep 2020 Jul 31;69(30):993-998 [FREE Full text] [doi: 10.15585/mmwr.mm6930e1] [Medline: 32730238]
14. Nearly One in Five American Adults Who Have Had COVID-19 Still Have "Long COVID". Centers for Disease Control and Prevention. URL: https://www.cdc.gov/nchs/pressroom/nchs_press_releases/2022/20220622.htm [accessed 2024-10-31]
15. Bai F, Tomasoni D, Falcinella C, Barbanotti D, Castoldi R, Mulè G, et al. Female gender is associated with long COVID syndrome: a prospective cohort study. Clin Microbiol Infect 2022 May;28(4):611.e9-611.e16 [FREE Full text] [doi: 10.1016/j.cmi.2021.11.002] [Medline: 34763058]
16. Huang Y, Pinto MD, Borelli JL, Asgari Mehrabadi M, Abrahim HL, Dutt N, et al. COVID symptoms, symptom clusters, and predictors for becoming a long-hauler looking for clarity in the haze of the pandemic. Clin Nurs Res 2022 Dec 24;31(8):1390-1398 [FREE Full text] [doi: 10.1177/10547738221125632] [Medline: 36154716]
17. The sex, gender and COVID-19 project. Global Health 50/50. URL: https://tinyurl.com/2h3asj2k/ [accessed 2024-05-01]
18. Carli LL. Women, gender equality and COVID-19. Gender Manag Int J 2020 Sep 30;35(7/8):647-655. [doi: 10.1108/gm-07-2020-0236]
19. Yaksi N, Teker AG, Imre A. Long hospitalized COVID-19 patients: a retrospective cohort study. Iran J Public Health 2022 Jan 10;51(1):88-95 [FREE Full text] [doi: 10.18502/ijph.v51i1.8297] [Medline: 35223629]
20. Aiyegbusi OL, Hughes SE, Turner G, Rivera SC, McMullan C, Chandan JS, TLC Study Group. Symptoms, complications and management of long COVID: a review. J R Soc Med 2021 Oct 15;114(9):428-442 [FREE Full text] [doi: 10.1177/01410768211032850] [Medline: 34265229]
21. Torjesen I. COVID-19: middle aged women face greater risk of debilitating long term symptoms. BMJ 2021 Mar 25;372:n829. [doi: 10.1136/bmj.n829] [Medline: 33766927]
22. Mobile fact sheet. Fact sheets: tech adoption trends. Pew Research Center. URL: https://tinyurl.com/ye5yzd4d [accessed 2024-05-01]
23. Computer and internet use in the United States: 2018. U.S. Census Bureau. URL: https://tinyurl.com/5746fjpd [accessed 2024-05-01]
24. Chidiac M, Ross C, Marston HR, Freeman S. Age and gender perspectives on social media and technology practices during the COVID-19 pandemic. Int J Environ Res Public Health 2022 Oct 27;19(21):13969 [FREE Full text] [doi: 10.3390/ijerph192113969] [Medline: 36360853]
25. Parker K, Horowitz JM, Minkin R. How the coronavirus outbreak has – and hasn't – changed the way Americans work. Pew Research Center. 2020 Dec 9. URL: https://tinyurl.com/2setyh4b [accessed 2024-05-01]

26. Callard F, Perego E. How and why patients made long Covid. Soc Sci Med 2021 Jan;268:113426 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113426] [Medline: 33199035]
27. Miyake E, Martin S. Long Covid: online patient narratives, public health communication and vaccine hesitancy. Digit Health 2021 Nov 29;7:20552076211059649 [FREE Full text] [doi: 10.1177/20552076211059649] [Medline: 34868622]
28. David ME, Roberts JA. Smartphone use during the COVID-19 pandemic: social versus physical distancing. Int J Environ Res Public Health 2021 Jan 25;18(3):1034 [FREE Full text] [doi: 10.3390/ijerph18031034] [Medline: 33503907]
29. Moore K, March E. Socially connected during COVID-19: online social connections mediate the relationship between loneliness and positive coping strategies. J-STAR 2022 Jan 10;1(1):1-20. [doi: 10.55319/js.v1i1.9]
30. Hajek A, König HH. The association between use of online social networks sites and perceived social isolation among individuals in the second half of life: results based on a nationally representative sample in Germany. BMC Public Health 2019 Jan 09;19(1):40 [FREE Full text] [doi: 10.1186/s12889-018-6369-6] [Medline: 30626369]
31. Changes in internet use when coping with stress: older adults during the COVID-19 pandemic. Elsevier Enhanced Reader. URL: https://pubmed.ncbi.nlm.nih.gov/32771312/ [accessed 2024-05-01]
32. Garfin DR. Technology as a coping tool during the coronavirus disease 2019 (COVID-19) pandemic: implications and recommendations. Stress Health 2020 Oct 13;36(4):555-559 [FREE Full text] [doi: 10.1002/smi.2975] [Medline: 32762116]
33. Cato S, Iida T, Ishida K, Ito A, Katsumata H, McElwain KM, et al. Social media infodemics and social distancing under the COVID-19 pandemic: public good provisions under uncertainty. Glob Health Action 2021 Jan 01;14(1):1995958 [FREE Full text] [doi: 10.1080/16549716.2021.1995958] [Medline: 34806565]
34. Clark JL, Algoe SB, Green MC. Social network sites and well-being: the role of social connection. Curr Dir Psychol Sci 2017 Dec 19;27(1):32-37. [doi: 10.1177/0963721417730833]
35. Lin X, Wang X. Examining gender differences in people’s information-sharing decisions on social networking sites. Int J Inf Manag 2020 Feb;50:45-56. [doi: 10.1016/j.ijinfomgt.2019.05.004]
36. Krasnova H, Veltri NF, Eling N, Buxmann P. Why men and women continue to use social networking sites: the role of gender differences. J Strateg Inf Syst 2017 Dec;26(4):261-284. [doi: 10.1016/j.jsis.2017.01.004]
37. SPercentage of adults in the United States who use social networks as of February 2021, by gender. Statista. 2019. URL: https://www.statista.com/statistics/471345/us-adults-who-use-social-networks-gender/ [accessed 2024-05-01]
38. Video conferencing, web conferencing, webinars, screen sharing. Zoom. URL: https://us06web.zoom.us/ [accessed 2024-07-01]
39. Otter.ai. URL: https://otter.ai/ [accessed 2024-01-31]
40. Boyatzis RE. Transforming qualitative information: thematic analysis and code development. Thousand Oaks, CA: SAGE Publications; 1998:xvi, 184.
41. Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. Int J Qual Methods 2016 Nov 29;5(1):80-92. [doi: 10.1177/160940690600500107]
42. Guest G, Bunce A, Johnson L. How many interviews are enough? Field Methods 2016 Jul 21;18(1):59-82. [doi: 10.1177/1525822X05279903]
43. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
44. All-in-one qualitative data analysis software. MAXQDA. URL: https://www.maxqda.com/qualitative-data-analysis-software [accessed 2024-07-01]
45. Burnard P. A method of analysing interview transcripts in qualitative research. Nurse Educ Today 1991 Dec;11(6):461-466. [doi: 10.1016/0260-6917(91)90009-y] [Medline: 1775125]
46. Polit DF, Beck CT. Nursing research: principles and methods. Philadelphia, PA: Lippincott Williams & Wilkins; 2004.
47. Corbin J, Strauss A. Basics of qualitative research: techniques and procedures for developing grounded theory. Fourth edition. Thousand Oaks, CA: SAGE Publications; 2014.
48. Krippendorff K. Content analysis: an introduction to its methodology. Fourth edition. Thousand Oaks, CA: SAGE Publications; 2019.
49. Padgett DK. Qualitative methods in social work research. Third edition. Thousand Oaks, CA: SAGE Publications; 2008.
50. Leist AK. Social media use of older adults: a mini-review. Gerontology 2013 Apr 16;59(4):378-384 [FREE Full text] [doi: 10.1159/000346818] [Medline: 23594915]
51. Pelletier J, Houle J, Goulet M, Juster R, Giguère CÉ, Bordet J, et al. Online and recovery-oriented support groups facilitated by peer support workers in times of COVID-19: protocol for a feasibility pre-post study. JMIR Res Protoc 2020 Dec 18;9(12):e22500 [FREE Full text] [doi: 10.2196/22500] [Medline: 33259326]
52. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Lancet 2020 Mar 14;395(10227):912-920 [FREE Full text] [doi: 10.1016/S0140-6736(20)30460-8] [Medline: 32112714]
53. Stone LB, Veksler AE. Stop talking about it already! Co-ruminating and social media focused on COVID-19 was associated with heightened state anxiety, depressive symptoms, and perceived changes in health anxiety during spring 2020. BMC Psychol 2022 Mar 07;10(1):22 [FREE Full text] [doi: 10.1186/s40359-022-00734-7] [Medline: 35130965]

54. Lee SA, Jobe MC, Mathis AA, Gibbons JA. Incremental validity of coronaphobia: coronavirus anxiety explains depression, generalized anxiety, and death anxiety. J Anxiety Disord 2020 Aug;74:102268 [FREE Full text] [doi: 10.1016/j.janxdis.2020.102268] [Medline: 32650221]
55. Ihrig A, Karschuck P, Haun MW, Thomas C, Huber J. Online peer-to-peer support for persons affected by prostate cancer: a systematic review. Patient Educ Couns 2020 Oct;103(10):2107-2115. [doi: 10.1016/j.pec.2020.05.009] [Medline: 32475711]
56. McCaughan E, Parahoo K, Hueter I, Northouse L, Bradbury I. Online support groups for women with breast cancer. Cochrane Database Syst Rev 2017 Mar 10;3(3):CD011652 [FREE Full text] [doi: 10.1002/14651858.CD011652.pub2] [Medline: 28278559]
57. Thompson DM, Booth L, Moore D, Mathers J. Peer support for people with chronic conditions: a systematic review of reviews. BMC Health Serv Res 2022 Mar 31;22(1):427 [FREE Full text] [doi: 10.1186/s12913-022-07816-7] [Medline: 35361215]
58. Margherita G, Muzii B, Caffieri A, Di Francia A, Somma B. 'Isolated together': online group treatments during the COVID-19 pandemic. A systematic review. Res Psychother 2022 Dec 04;25(3):639 [FREE Full text] [doi: 10.4081/ripppo.2022.639] [Medline: 36373389]
59. Zhao N, Zhou G. Social media use and mental health during the COVID-19 pandemic: moderator role of disaster stressor and mediator role of negative affect. Appl Psychol Health Well Being 2020 Dec 17;12(4):1019-1038 [FREE Full text] [doi: 10.1111/aphw.12226] [Medline: 32945123]
60. Roberto A, Sellon A, Cherry ST, Hunter-Jones J, Winslow H. Impact of spirituality on resilience and coping during the COVID-19 crisis: a mixed-method approach investigating the impact on women. Health Care Women Int 2020 Oct 21;41(11-12):1313-1334. [doi: 10.1080/07399332.2020.1832097] [Medline: 33085579]
61. Pahayahay A, Khalili-Mahani N. What media helps, what media hurts: a mixed methods survey study of coping with COVID-19 using the media repertoire framework and the appraisal theory of stress. J Med Internet Res 2020 Aug 06;22(8):e20186 [FREE Full text] [doi: 10.2196/20186] [Medline: 32701459]
62. Venegas-Vera AV, Colbert GB, Lerma EV. Positive and negative impact of social media in the COVID-19 era. Rev Cardiovasc Med 2020 Dec 30;21(4):561-564 [FREE Full text] [doi: 10.31083/j.rcm.2020.04.195] [Medline: 33388000]
63. Malecki K, Keating JA, Safdar N. Crisis communication and public perception of COVID-19 risk in the era of social media. Clin Infect Dis 2021 Feb 16;72(4):697-702 [FREE Full text] [doi: 10.1093/cid/ciaa758] [Medline: 32544242]
64. Thompson CM, Rhidenour KB, Blackburn KG, Barrett AK, Babu S. Using crowdsourced medicine to manage uncertainty on Reddit: the case of COVID-19 long-haulers. Patient Educ Couns 2022 Feb;105(2):322-330 [FREE Full text] [doi: 10.1016/j.pec.2021.07.011] [Medline: 34281723]
65. Koss J, Bohnet-Joschko S. Social media mining of long-COVID self-medication reported by Reddit users: feasibility study to support drug repurposing. JMIR Form Res 2022 Oct 03;6(10):e39582 [FREE Full text] [doi: 10.2196/39582] [Medline: 36007131]
66. De Choudhury M, Morris MR, White RW. Seeking and sharing health information online: comparing search engines and social media. 2014 Presented at: CHI '14: SIGCHI Conference on Human Factors in Computing Systems; April 26-May 1, 2014; Toronto, Canada p. 1365-1376. [doi: 10.1145/2556288.2557214]
67. Pavalanathan U, De Choudhury M. Identity management and mental health discourse in social media. Proc Int World Wide Web Conf 2015 May;2015(Companion):315-321 [FREE Full text] [doi: 10.1145/2740908.2743049] [Medline: 27376158]
68. Katz E, Haas H, Gurevitch M. On the use of the mass media for important things. Am Sociol Rev 1973 Apr;38(2):164. [doi: 10.2307/2094393]
69. Korhan O, Ersoy M. Usability and functionality factors of the social network site application users from the perspective of uses and gratification theory. Qual Quant 2015 Jun 14;50(4):1799-1816. [doi: 10.1007/s11135-015-0236-7]
70. Hajek A, König HH. Social isolation and loneliness of older adults in times of the COVID-19 pandemic: can use of online social media sites and video chats assist in mitigating social isolation and loneliness? Gerontology 2021;67(1):121-124 [FREE Full text] [doi: 10.1159/000512793] [Medline: 33264778]
71. YU RP, MCCAMMON RJ, ELLISON NB, LANGA KM. The relationships that matter: social network site use and social wellbeing among older adults in the United States of America. Ageing Soc 2015 Jul 02;36(9):1826-1852. [doi: 10.1017/s0144686x15000677]
72. Latikka R, Koivula A, Oksa R, Savela N, Oksanen A. Loneliness and psychological distress before and during the COVID-19 pandemic: relationships with social media identity bubbles. Soc Sci Med 2022 Jan;293:114674 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114674] [Medline: 34959045]

Edited by A Kushniruk; submitted 30.06.23; peer-reviewed by L Laestadius, J Himstedt; comments to author 20.11.23; revised version received 15.12.23; accepted 18.02.24; published 23.04.24.

Please cite as:
Garrett C, Aghaei A, Aggarwal A, Qiao S
The Role of Social Media in the Experiences of COVID-19 Among Long-Hauler Women: Qualitative Study
JMIR Hum Factors 2024;11:e50443
URL: https://humanfactors.jmir.org/2024/1/e50443
doi:10.2196/50443
PMID:38652515

Original Paper

Understanding the Use of Mobility Data in Disasters: Exploratory Qualitative Study of COVID-19 User Feedback

Jennifer Lisa Chan[1*], MD, MPH; Sarah Tsay[2*], MPH, DrPH; Sraavya Sambara[3*], AB; Sarah B Welch[4*], MPH

[1]Department of Emergency Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[2]Department of Emergency Management, University of San Diego Health, San Diego, CA, United States

[3]Harvard University, Boston, MA, United States

[4]Buehler Center for Health Policy and Economics, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[*]all authors contributed equally

Corresponding Author:
Jennifer Lisa Chan, MD, MPH
Department of Emergency Medicine
Feinberg School of Medicine
Northwestern University
211 E. Ontario Street312-694-7000
Chicago, IL, 60611
United States
Phone: 1 312 694 7000
Email: jennifer-chan@northwestern.edu

Abstract

Background: Human mobility data have been used as a potential novel data source to guide policies and response planning during the COVID-19 global pandemic. The COVID-19 Mobility Data Network (CMDN) facilitated the use of human mobility data around the world. Both researchers and policy makers assumed that mobility data would provide insights to help policy makers and response planners. However, evidence that human mobility data were operationally useful and provided added value for public health response planners remains largely unknown.

Objective: This exploratory study focuses on advancing the understanding of the use of human mobility data during the early phase of the COVID-19 pandemic. The study explored how researchers and practitioners around the world used these data in response planning and policy making, focusing on processing data and human factors enabling or hindering use of the data.

Methods: Our project was based on phenomenology and used an inductive approach to thematic analysis. Transcripts were open-coded to create the codebook that was then applied by 2 team members who blind-coded all transcripts. Consensus coding was used for coding discrepancies.

Results: Interviews were conducted with 45 individuals during the early period of the COVID-19 pandemic. Although some teams used mobility data for response planning, few were able to describe their uses in policy making, and there were no standardized ways that teams used mobility data. Mobility data played a larger role in providing situational awareness for government partners, helping to understand where people were moving in relation to the spread of COVID-19 variants and reactions to stay-at-home orders. Interviewees who felt they were more successful using mobility data often cited an individual who was able to answer general questions about mobility data; provide interactive feedback on results; and enable a 2-way communication exchange about data, meaning, value, and potential use.

Conclusions: Human mobility data were used as a novel data source in the COVID-19 pandemic by a network of academic researchers and practitioners using privacy-preserving and anonymized mobility data. This study reflects the processes in analyzing and communicating human mobility data, as well as how these data were used in response planning and how the data were intended for use in policy making. The study reveals several valuable use cases. Ultimately, the role of a data translator was crucial in understanding the complexities of this novel data source. With this role, teams were able to adapt workflows, visualizations, and reports to align with end users and decision makers while communicating this information meaningfully to address the goals of responders and policy makers.

(JMIR Hum Factors 2024;11:e52257) doi:10.2196/52257

KEYWORDS

mobility data; disasters; surveillance; COVID-19; qualitative; user feedback; policy making; emergency; pandemic; disaster response; data usage; situational awareness; data translation; big data

Introduction

Background

The COVID-19 pandemic has had a significant impact on the world, with more than 6.82 million deaths and more than 670 million cases globally as of February 2023 [1]. Public health measures, specifically physical distancing policies, were implemented throughout the world in 2020, playing a major role in the early phases of the COVID-19 global pandemic [2-4]. In general, these policies were based on the assumption that individual and population movement dynamics of those infected with COVID-19 were likely to impact the degree of spread. Studies have drawn a strong association between public health measures (eg, stay-at-home orders, physical distancing requirements) and reductions in population movement. Implementing distancing policies during the prevaccine era had the potential to mitigate the spread of disease and ultimately decrease demand on health care systems [5,6].

Nations took varying approaches to enacting COVID-19 public health measures. The People's Republic of China implemented travel restrictions and widespread population-based quarantine measures until the end of 2022. Italy, affected by COVID-19 in February 2020, instituted a nationwide quarantine until mid-2021 [7]. The United States took a patchwork approach to social distancing policies, which varied across the country by jurisdiction. Some cities had strict social distancing policies while others did not institute a stay-at-home order. During the latter part of 2020 and into 2021, the lack of a coordinated nationwide approach left cities and states to independently determine when to relax or reinstitute policies during subsequent epidemiologic waves due to new COVID-19 variants.

Using data in disasters involves a wide array of processes and skills in order to result in effective use and impact on decision-making. The process of using data includes data collection, processing, analyses, product creation (eg, visualization), and sharing [8,9]. An emerging body of evidence now recognizes the importance of collaboration, communication, and data literacy, all human factors, as important to effective use of data [10]. For government agencies around the world, understanding near real-time movement of individuals and groups was extremely difficult [11]. Key decision makers had to establish timely policies with limited, evolving information [12]. Physical distancing policies faced the same, if not more nuanced, challenges in using timely evidence, as the data and information that represented population-based movements were emerging from novel data sources not previously used for a pandemic of unprecedented proportions. Human mobility data, which are passively generated by digital devices that track information over both time and geography, provide a unique data source that reflects human movement in disaster settings [13]. In 2010, mobility data predicted population movement after the Haiti earthquake, and early studies in 2020 showed that stay-at-home orders were associated with decreased population movement [14].

In response to the opportunities that mobility data presented for decision-making in the early phase of the COVID-19 pandemic, these data were analyzed by researchers from a wide range of disciplines with the intention of understanding disease dynamics as well as the impact of physical distancing policies on human movement at the national, regional, and local levels [15]. Klein et al [16] characterized physical distancing behaviors in the United States during the autumn of 2020, highlighting urban-rural differences and concluding that physical distancing policies were associated with changes in movement patterns. Much of the current research has focused on analyzing such data and anticipating the influence on policy making, including how the applications with these quantitative data were able to predict disease spread or identify future outbreak locations [17,18]. Although movement data had been used by governments in the past, use during the COVID-19 pandemic posed new challenges [19].

A network of global researchers, response planners, and policy makers collaborated in early 2020 with the aim of translating mobility data in order to positively impact response planning and policy making. With access to anonymized and privacy-preserving mobility data from Facebook (Meta), the COVID-19 Mobility Data Network (CMDN) facilitated the responsible and meaningful use of widely available human mobility data by policy makers around the world [20,21]. Mobility data sets were shared with trusted partners (eg, researchers) to collaborate with the public sector to inform response and recovery. Facebook also created a differential privacy framework to protect the privacy of individuals in aggregated data sets, informing public sector response to the COVID-19 pandemic. Over 150 researchers from around the world joined the network, contributing their scientific and analytic skills to develop methods necessary to transform these new data sources to meet the needs of their government counterparts. Both researchers and policy makers assumed that collecting, analyzing, and sharing mobility data would provide insights to help both policy makers and response planners guide their activities to mitigate the human, economic, and social impacts of COVID-19.

Despite the ability to collect, analyze, and share mobility data results around the world, there lacked evidence for how mobility data were used in response planning and policy making. There also lacked evidence and knowledge of which processes or steps in this applied research were more or less effective. Recent evaluations of mobility studies in COVID-19 have shown that applied research, or translational research, is often neglected, limiting the ability of decision makers and the public to take action and benefit from these efforts. In the outbreak modeling community, Nixon et al [19] reviewed 136 papers but found that only 1 in 4 papers evaluated the performance of the models and few disclosed the uncertainty of the data and transparency of the methodology. In addition, more than 50% of the studies focused on predictions at the national level when policy makers and planners more frequently focus on local-level decision-making, and these approaches were least useful for decision-making [19]. Although quantitative methods can, to some degree, provide evidence on the accuracy of mobility data, complementary qualitative methods can provide evidence of their use in the social dynamic environments of disasters.

Study Objectives

This study focused on advancing the understanding of how researchers, policy makers, and response planners used human mobility data during the COVID-19 pandemic for the purpose of supporting decision-making related to stay-at-home orders and social distancing policies. Using qualitative methods to investigate the near real-time and lived experiences and perceptions of those using mobility data, the study used a constructivist point of view to describe experiences of researchers and practitioners who used these data in response planning and policy making. The study aimed to identify themes related to what would make the data more usable and valuable for public health practitioners, response planners, and researchers in disaster and emergency settings.

This study was framed within the information management and disaster response fields. The role of data and information management in disaster contexts is tightly linked to situational awareness. As described by Vieweg [22] and others [23,24], situational awareness refers to understanding the disaster environment, including hazards and emerging threats; they also described that improved situational awareness enables informed decisions related to response and recovery activities. In addition, data, whether traditional or novel, are perceived as valuable and useful when results can improve an individual's or group's situational awareness of the disaster environment. Situational awareness is assumed to be dynamic and often constantly changing, and uncertainty of the disaster environment is common.

Specific aims of the study were to explore the following:

- How groups use human mobility data and the results in response planning and policy making
- What processes and ways of working were used to analyze human mobility data and visualize or communicate the results
- Human factors related to processes described in the previous bullets, including various purposes for using mobility data, communication dynamics, and translational activities that enable the use of the data

Methods

Ethical Considerations

This project used preexisting interview information acquired for Crisis Ready's User Feedback Project and collected for the purposes of project assessment and evaluation. As such, the initial data collection was not reviewed by an institutional review board and did not collect informed consent from interview participants. This exploratory study used these data to understand how mobility data in the CMDN were being used during a large-scale global public health disaster. It was reviewed by the Northwestern University institutional review board and deemed not human research, as it was secondary data analysis with no identifiers (STU00214214).

Approach

Our study was grounded in a phenomenology approach to understand how individuals were making sense of novel mobility data and incorporating the data into preexisting workflows. The unprecedented and highly uncertain and dynamic environment experienced by the interviewees both due to the COVID-19 pandemic and novel data source was best aligned with this methodological approach. Our analytical approach used thematic analysis to identify patterns of behavior and opinion as well as understand how individuals navigated the many new inputs during this time. The analysis team consisted of 3 individuals (JLC, ST, SBW). JLC is an emergency medicine physician, humanitarian practitioner, data specialist, and researcher with qualitative experience. ST is an emergency management professional with over a decade of experience at the intersection of public health, health care, and emergency management. SBW is a public health evaluator and research project manager with experience in qualitative methods.

Data Collection

Study data included interviews completed with individuals participating in the CMDN and selected using a combination of convenience and purposive sampling to have global representation and to enroll participants from different sectors (ie, public health practitioners, response planners, or researchers) to include a variety of experiences and perspectives. Interviews were deidentified and professionally transcribed before being shared with the study team, who uploaded the data to Dedoose Version 9.0.17. Study data were defined as narrative data from interviews and demographic information.

Analysis

We used thematic analysis with an inductive approach to understand our data. ST and SBW independently open-coded 2 transcripts each, assigning codes to all concepts identified in the interview data. The entire team (JLC, ST, SBW) met to review the initial codebook and refine it by combining similar concepts from approaches derived from the literature [25-27]. ST and SBW then exchanged transcripts and blind double-coded them using the initial codebook. The entire team refined the codebook further and blind triple-coded a fifth transcript. The codebook was refined and finalized (Multimedia Appendix 1).

ST and SBW blind double-coded all interview transcripts in a systematic manner guided by coding practices outlined by MacQueen et al [26]. Coding discrepancies were documented and discussed, and the code was changed to reflect the agreed-upon code and placement. If ST and SBW could not come to an agreement, JLC was engaged in the conversation, and discussion continued until consensus between the entire coding team was reached using previously published codebook methods for team approaches to coding [28,29]. All coded excerpts were reviewed by code. Team members reviewed each quote excerpt and summarized and organized them into themes and concepts to look for common experiences, examples, and opinions.

To increase the validity, initial project findings were presented to stakeholders of participants in CMDN and peer researchers before the conclusion of analysis. This activity included peer review activities and a hybrid member-check and data party [30-32]. The analysis team conducted 2 meetings over the course of 2 days where project overview, transcript excerpts, and findings were presented to the group. Participants were asked if the information resonated with them, made sense, and reflected their understanding based on their experience; they were also asked to share their thoughts for next steps for using these data for application in future emergencies. Feedback from this meeting was used to check the themes developed in the final analysis. There were no additional themes identified based on the feedback.

Results

Participants

The anonymized and deidentified data set consisted of 33 interviews conducted with 45 researchers and practitioners involved in CMDN from June 2020 to September 2020. The majority of interviewees were from North America (28/45, 62%), with a predominance from the United States (26/45, 58%), followed by Asia (11/45, 24%; Table 1). Of the interviewees, 62% (28/45) were men, and 38% (17/45) were women. In addition, 47% (21/45) of the interviewees had practitioner roles, while 42% (19/45) had primary research roles. Specific interviewees' roles included state-level Chief Information Officer, consultant supporting a state-level response, data scientist in city government, emergency manager, and assistant professor of epidemiology. A few interviewees described their roles as hybrid, encompassing both researcher and practitioner activities.

Table 1. Participant demographics (n=45).

Characteristics	Results, n (%)
Gender	
Male	28 (62)
Female	17 (38)
Continent	
Africa	1 (2)
Asia	11 (24)
Europe	2 (4)
North America	28 (62)
South America	3 (7)
Professional field	
Practitioner	21 (47)
Researcher	19 (42)
Other	5 (11)

Principal Findings

Themes

Our principal findings are organized into the following 3 sections: (1) ways in which mobility data were used in response planning and policy making, (2) how data were processed, and (3) human factors that contributed to successful use (Table 2). The findings presented are resultant of triangulation from interviews from different groups (ie, researchers, practitioners) and reflexivity in the determination of themes.

Table 2. Taxonomy of themes by research question.

Research question	Themes
How groups use human mobility data and its results in response planning and policy making.	• There was a perception that mobility data analyses could or did improve situational awareness. • There was uncertainty on how results were ultimately used.
How mobility data were processed and ways of working	• Misaligned data purpose with priorities was common. • The team had difficulty interpreting meaning from results. • Some teams determined the results were of limited value.
Human factors enabling or hindering the use of data	• Human factors play a significant role in how mobility data were used. • Researcher-practitioner teams that were more successful had open communication styles, enabling near real-time learning. • Adapting processes, which was dependent upon communication, was often required to use the data successfully.

Use in Response Planning and Policy Making

Researchers and policy makers in the CMDN used mobility data to improve understanding of population movement; influence resource provision; and influence policies, specifically stay-at-home orders. Themes emerged among these narratives related to (1) the perception that mobility data analyses could or did improve situational awareness, (2) uncertainty on how to interpret mobility data results, and (3) the perceived lack of value despite understanding the results.

Response planning and policy making are often related but distinctly different activities. Response planning involves tactical, action-oriented activities such as daily planning and logistics, while policy making focuses on planning and articulation of actions to guide involved stakeholders who are often required to abide by or execute these policies. This qualitative study revealed that, for both response planning and policy making, there were no common uses of these data across user groups. In general, those interviewed in the study sought to better understand where people were moving as well as understand situational changes in their specific locality (eg, national, county, and city levels).

Mobility data helped many teams better understand specific group movements and helped others allocate resources such as personal protective equipment (PPE) and plan food distribution sites. One team used the data to further understand the movement of essential workers with the intention of better understanding disease spread in a large city. The mobility data helped them rank neighborhoods that had larger percent changes of movement during the citywide stay-at-home order compared with movement patterns before the pandemic (ie, baseline).

> *The mobility data has been used, and especially to identify the neighborhood that has [a] high proportion of essential service workers who had to report to work during [a] surge of COVID-19 cases. [We] calculated percent change of the commuting patterns during Week 13–the morning commute time, which is almost the end of March and all of April, the exact time the[city] experienced [a] huge surge of COVID-19 cases.* [Practitioner, United States]

Other government offices described using mobility data for PPE distribution along with other data sources to help with situational awareness and planning.

> *The [County] Office of Emergency Management is building an early warning system of areas that could become hotspots with the potential to overwhelm the medical and public health systems. Density and mobility are key indicators in this customized early warning system. We infuse the mobility data from these reports into a larger system that helps anticipate areas of potential viral spread, thus allowing for a more proactive response in regards to PPE supply distribution, and targeted community testing.* [Researcher, United States]

Interviewees mostly described how they intended to use mobility data, rather than reporting policy-related actions taken as a result of the data analyses. They also described how they used the data to plan for reopening or relaxation of existing stay-at-home orders. Many other interviewees expressed uncertainty about how the data were actually used in policy making.

> *I share[d] it with our health director every week and with the rest of our data group. And I think the thing that they're just mostly looking [at] when we move to phase one of the reopening, like how much did people start moving around more or were people still staying at home. And I think from this, we concluded that people were starting to move more.* [Practitioner, United States]

> *So, one of the things we can do with it is show that if you allow unrestricted travel during the Eid holidays, this outbreak, it's gonna spread really fast and it's gonna spread everywhere." And so, they got the CDR (call detail records) data. We did some modeling. We showed what would happen and I'm not gonna say that that's what influenced them, but they did extend the lockdown so that people couldn't travel during the Eid holidays.* [Researcher (located in the United States), India]

Improving Situational Awareness

Although the Use in Response Planning and Policy making section describes specific use cases, interviewees often described how the data were used for general situational awareness. Many believed that the data could help them understand movement of people during stay-at-home orders or during relaxation of these orders. Some groups felt the data improved their situational awareness of localized outbreaks of COVID-19 and future COVID-19 transmission monitoring.

> *Facebook has actually given us the ability to not just look at overall levels of movement, but to be able to look at where people are moving from and to over the space of time. Because we've had these fairly localized point outbreaks.* [Researcher, Australia]

> *We've been able to use that data to put together essentially risk maps saying, okay, given that we saw transmission there last week, let's go back and look up what the transmission patterns in and out of that area, wherever the proceeding couple of weeks.* [Researcher, Australia]

These are 2 examples where mobility data enabled understanding of the movement of people over space and time, as well as transform their situational awareness into risk maps, a form of visualization. Although many described the data influencing their situational awareness, there were no findings that revealed accounts of explicit changes in decision-making.

Uncertainty

This section presents the multifaceted nature of not only understanding how mobility data were used in this disaster but also the uncertainty that individuals faced that became barriers to data use. There was significant uncertainty around how to transition from the idea that mobility data could be useful to actually transforming the data to assist in response planning and policy making. For example, even though practitioners and policy makers believed that mobility data could improve situational awareness of when and where people were moving, there was uncertainty around how to formulate a specific way in which mobility results could influence response planning.

> *I mean, [public health scientist] has told us repeatedly, this is very helpful for them to know...Exactly how to then use it for their response, I think that is a much harder question.* [Researcher, India]

> *And, so, they came up with their recommendations… and unfortunately, I was not involved with the direct communication with the data requesters…so my answer to your question is, I don't know how the result of the analysis has shaped any policy decisions and any change of the course of actions.* [Practitioner, United States]

> *So, I don't know exactly how he's been using it. I know he's been monitoring it just to see whether there are any places that raise any flags.*" [Practitioner, United States]

These excerpts reflect how both researchers and practitioners felt they did not know or lacked confidence on the exact use of the data for response planning or policy making. It remains unclear whether this uncertainty is due to limited skills in problem formulation with novel data, unfamiliarity with analytical methods, or challenges with communicating results.

Processes and Ways of Working

Themes

This section describes how mobility data were processed. We present 4 ways of processing the data that respondents described: (1) identifying data purpose, (2) tools and products, (3) data preparation, and (4) results and data sharing. The themes that emerged were (1) the initial purpose for using mobility data was often misaligned with groups' priorities, (2) results sharing was often limited when the results lacked meaning, and (3) the results were deemed of little or no value.

Identifying Data Purpose

In order to use mobility data, understanding the purpose of use is one of the first crucial steps. The study revealed that each team's intention to use mobility data was linked to their perceived purpose of using the data. Although seemingly intuitive, it remains an important distinction. Many researchers and practitioners aimed to use the data to better understand where people were moving to and from and if groups were abiding by policies such as movement restrictions or stay-at-home orders. Others sought to correlate mobility data with other data sources to better understand the dynamics of COVID-19 disease transmission.

Most researchers relied upon practitioners to identify the specific purpose of using mobility data in their regions or localities and awaited direction from them. Some teams were able to identify specific questions for the data but often adapted their questions over time. In many circumstances, this was due to changing environments and, more importantly, an evolving understanding of what the mobility data could and could not answer for their specific questions.

> *Initially we were interested in...are people staying home, but then as we started to phase open, we also started tracking it even more closely to see, okay, now that we're opening, are we seeing [that] more people are moving about and things like that.* [Practitioner, United States]

> *I think the ask has changed a little bit around what we're layering the data with. The initial hope was to work with the county health department to better understand cases and different rates that we could correlate the mobility with. So, we could say, people were extremely mobile here, and we're seeing a spike in cases later on in these specific hospitals. Unfortunately, we weren't able to get that second piece of data, and with the aggregated data that we get from the county about the city as a whole we aren't seeing any correlations with that.* [Practitioner, United States]

Other teams did not ultimately share the results of the data because they were either unable to confidently interpret the results or felt that the results were not relevant to their needs and priorities.

> *We're not really sure how to move forward with the public health department I think, because we're still unclear how exactly we should use this data to prioritize any sort of EpiModel.* [Researcher, United States]

> *...But the reaction, I remember when I first saw this [visualization] was, "Cool visualization. But, what am I gonna do with this? And what is it actually going to tell me?* [Practitioner, United States]

These excerpts show the diversity of data purposes among groups. Some groups' data purpose evolved over time due to environmental changes or their ability to use the data with other sources.

Tools and Products

There was no standardized way researchers and practitioners teams used data tools, analysis software, or visualization platforms for their collaborative efforts. Ultimately, tools and products were described as vehicles with which to share information with other data teams, policy makers, and response planners. Some groups used geographic information systems (eg, ArcGIS software) and online mapping tools (eg, OpenStreet map), while others used data visualization tools such as Tableau and ShinyApp. Some products were spreadsheets (eg, csv files) that were shared by researchers to data teams within government offices. In some instances, both researcher and practitioner interviewees felt that providing analyses in these data formats helped practitioners create final reports.

> *So, I do know one outcome of that is that we requested that we get the all-city data from each timeframe and [he] was able to also upload that as CSV for us too, and that lines up with the PDF. So, at one point, we were trying to replicate what the PDFs had and couldn't, and that actually helped us lead to asking for the data.* [Practitioner, United States]

Data Preparation

The components of data preparation include data cleaning, validation, and transformation, often bringing together different data types from various sources [33]. Mobility data were a novel data type for most researchers and practitioners and required them to take additional time to familiarize themselves during the data preparation process. Some interviewees described the preparation stage of "cleaning" mobility data, such as changing formats and structures to more easily align with other data sources to make the analysis relevant for their purposes.

> *In fact, we also asked them to change a little bit in terms of the columns and the rows that we wanted them to change because they were not compatible as Interviewee 067 was mentioning. So, we have to do a lot of work by ground.* [Researcher, India]

Many groups prioritized preparing the mobility data to better meet their local environment. They often reworked the data to meet their planning geography, whether it was formal administrative units or planning geographies such as jurisdictions.

> *There's one for [city location]. There's one for [neighborhood]. All of these different areas are more canonical for people. So, the data we receive are mapped to these, and it makes a lot of sense to people used to working with this geography. So, so long as all the other sources can be mapped to the same thing, that would probably make sense to the city and decision makers.* [Volunteer, United States]

For many groups, tailoring and preparation meant just learning more about what the data represented and becoming more familiar with mobility data. The process of preparing the data was in and of itself a way to learn and gain knowledge about the data and what it could ultimately mean for policy making and response planning.

> *Yeah, so we spent about five, six days working on the data and trying to understand what it represented. So, it's probably a lot of back and forth with various people and from Facebook.* [Researcher, Thailand]

These excerpts reflect the need for teams from different contexts to further adapt the data to meet their specific needs, which took time to prepare.

Results and Data Sharing

Results were shared with national ministers, district mayors, emergency response teams, public health teams, and the European Commission, to name a few. Researchers frequently shared analytical results with practitioners in multiple formats (eg, graphs, maps) and sometimes even spreadsheets. Simplification, context alignment, and narrative text were often necessary to help readers understand what the results meant. For example, a data fellow in a city office simplified graphs to help her colleagues better interpret the results, anticipating areas of confusion and working to mitigate them. Other teams created their own charts from the data analysis to better visualize and share the results.

> *And if you can imagine a figure with a lot of noise, individuals that maybe don't have the experience that I do or people in the network have, may find that confusing. So, I did very simple things. I apply some smoothing algorithm so as to show trends rather than actual aggregated data. And that way I could mitigate the confusion regarding noise from the data...* [Researcher, United States]

> *First we look at the report that you [guys] send us. We take a look at that, understanding a little bit better. And then whatever insights that we gather from that report, I translate those into our actual report that we hand out to the city manager and use the CSV to create those charts.* [Practitioner, United States]

These excerpts reflect individuals and teams tailoring results or translating analytical findings to address "noise" or "confusion" with the aim to improve understanding and sharing with others.

Narrative statements accompanying analytical results were also common as practitioners sent reports to response planners and policy makers. Many felt that adding summaries or explanations helped transform the data results into meaningful information. Others used bullet formatting to help prioritize findings but included more lengthy results to be reviewed at the discretion of the end user.

> *And if we were sharing this report more broadly with our policy folks, that's what they're looking for, is that summary. But I think I shared our 20-some odd page weekly, right? So, we do that same thing. Here's the top five or six bullets you need to know about and validate the rest if you want.* [Practitioner, United States]

Teams needed not only to communicate and share results with one another but also the skills and time to interpret the results and understand their meaning. Some challenges laid in understanding the mobility metrics, such as how movement was measured in relationship to trips taken during a time period.

> *So, these [are] some insights...that were not provided by the original graphs. And then, another thing that I thought was very insightful is this...when you stratify by districts, you see that it's driven mainly by these three districts up here. Some districts that are behaving, if we can say that, are adhering to the social distance policy.* [Researcher, United States]

Some lacked confidence in sharing the results with response planners and policy makers. Others described uncertainty in their interpretation of results and felt that the risks of sharing this information were too great to take on. Ultimately, this prevented many interviewees from sharing results with decision makers.

> *I guess the thing about this data is that it is really difficult to know how to interpret any one city, or any one weekend change. And I would be cautious about advising changes on that until we have more data on what that change actually represents.* [Researcher, United States]

Other environmental factors such as information and data overload prevented many from not only understanding the results but also identifying their value for decision-making. For many, understanding the data and translating the findings for others required significant time, often time that many did not feel they had.

> *This is a pretty detailed report, and I think some people looking at it would be kind of intimidated by it or wouldn't have the time to really be able to go in and dig in deep, so I think we have thought a lot about how to give high-level updates that are meaningful and impactful.* [Practitioner, United States]

These excerpts highlight that additional processing was needed to distill results in order to feel comfortable sharing them. Different groups created different products to best meet the needs of their specific users and decision makers.

Human Factors

Themes

Interviewees spoke about human factors, including data translation, communication, and adaptation, that contributed to successful use of mobility data. The following 3 themes emerged: (1) human factors played a significant role in how mobility data were used, (2) researcher-practitioner teams that were more successful had open communication styles, enabling near real-time learning, and (3) adapting processes was often required to use the data successfully, which was dependent upon communication.

Data Translation

Data translation skills include behaviors such as communication skills and the ability to communicate both technical aspects of data (eg, analysis, metrics) and relevant information meaning and value of the data itself in various formats. Data translation in the context of using mobility data during the COVID-19 pandemics was tightly linked to human information processing abilities and creating products that could provide situational awareness for response planners and decision makers.

> *And so, I think for us, that was the key thing: that the [researcher] was willing to do, and work with us on, and understand that…it was a process that we all had to work through, and he, [was] just available and willing to work with us and willing to help us navigate this new type of data.* [Practitioner, United States]

Data translation helped both researchers and practitioners understand the specific purpose for which to use the data. This enabled teams to refine geographic approaches to meet city, county, and regional contexts during stay-at-home orders and the many evolving COVID-19 epidemiologic environments, including relaxation of restrictions. Although team communications were primarily remote, these social interactions were viewed as important in enabling teams to understand what data opportunities lay ahead and adapt to changing environments, evolving collective knowledge about the data's value, and optimal methods to present results to decision makers.

Communication

Teams communicated using different modalities and frequencies. Some groups had daily calls early in the collaboration, others connected on weekly Zoom meetings, and some primarily communicated via email. Most researchers did not communicate directly with response planners and policy makers but rather with data teams in government offices. The few that shared results directly did so via situation reports, updates in daily emergency operational cell meetings, or via WhatsApp and email. In South America, communications were via WhatsApp and direct voice communications from the researcher to the scientist in the country and then to national ministers.

> *I am part of [a] directory of people that are indirectly [in] discussion with the minister of science...I sent him an email directly to the minister and he ignored me. The same Professor said, well, it's easier if you [to] send me the reports and I am [in] a channel of communication with him.* [Researcher, South America]

Trusted relationships played a role in successful communications and often required some adaptations as teams determined more effective ways of sharing the results.

Although the route of communication was important, interviewees described the nature of communication including listening and back-and-forth communication as essential parts of a successful collaboration.

> *I don't know, it's so boring and simple, but I think it's just about listening, and just keep asking questions to really understand where they're at. And I think I was really helped along too by the communications director, because she was really open about asking questions when she didn't understand things.* [Researcher, United States]

Communication often led to changes in ways of working and adapting approaches, such as redefining the purpose for using mobility data, simplifying results, and helping better translate the analytical results into meaningful findings that could be shared in a more understandable way.

Adaptation

Researchers and practitioners adapted their approaches, workflows, and communications to align with the local context. Over the course of the pandemic, response activities and policies in many jurisdictions changed, requiring many groups to reassess their original purpose for using mobility data. One team adapted their initial purpose from monitoring movement to monitoring movement after various restrictions were being lifted. Some increased the frequency of reporting as one jurisdiction went into lockdown. Others adjusted their approach to analysis once they gained a deeper understanding of what the data could and could not be used for.

> *I think the teams that I've had the best experiences with in terms of the use of these data and the teams that have also gotten the most out of it have been the ones where there have been these sort of clear lines of communication, and ability to change and upgrade things as we move forward.* [Researcher, multiple locations]

> *I've been asked to provide a second report. So, I shared the presentation I gave this morning, but I also wrote a report similar to this, since they've just gone into lockdown...So, usually, they'll ask for a report on a monthly basis, – but when something happens, I try to be ready to give those as they request and have it ready to give it as quickly as I can.* [Researcher, Botswana]

These excerpts are evidence of adaptation being facilitated by strong and open communication.

Discussion

Principal Findings

This qualitative study, grounded in a phenomenology approach, analyzes the perceptions of researchers and practitioners who used mobility data during the first year of the COVID-19 pandemic. This study contributes to the emerging knowledge on how groups in disaster settings are using novel data sources for decision-making, specifically response planning and policy making. It also highlights the importance of human factors (ie, communications, data translation, and adaptation) as critical for the success of individuals and groups in defining a purpose for which to use novel data, identifying value in the data results, and ultimately using data in disaster response activities.

Although some groups used mobility data for distributing resources, most were unsure of how these data directly affected policy making. Evidence-based decision-making for policy making during COVID-19 remains an area of investigation, and published studies support our findings of limited use. A qualitative study of the perceptions of scientific experts and advisors in 11 countries described how evidence was thought to play a small role in overall policy making [34].

Humanitarian studies of data use in disaster environments have shown that, often, those who are managing, analyzing, and sharing data are often not the same individuals who are making decisions in emergency operations centers or leadership meetings [35]. The individuals we spoke to indicated this was the case in their partnerships as well. With novel data, it can take time to understand what the data represent and how the results are best communicated, requiring data translation skills or roles to help make the collaborations successful. Over time, the researchers and practitioners who were able to communicate and learn from one another were able to make the results more digestible for decision makers, likely due to improved knowledge and understanding and confidence in communicating this information.

There was a common understanding that groups were seeking new data sources to help improve situational awareness. This is in alignment with published literature on data-seeking behaviors and situational awareness. Endsley [36] described the theory of situational awareness in which individuals are described as not just recipients of data but active seekers who align this behavior with their current goals. However, with novel data, bidirectional communication between researchers and practitioners helped build collective knowledge of what the data could actually be used for in their local context. In addition, the direct collaboration between researchers and practitioners enabled teams to adapt quickly along various lines to try to meet the needs of each government group. Morss et al [37] noted that generating usable information among diverse decision makers requires working directly with these users to understand needs and context.

Human factors such as communications as well as adaptive approaches to data analyses and visual presentation were mentioned by many interviewees as steps along the way that helped them better prepare the data results for their intended users. These "soft skills" (ie, communication, interaction, creativity) described by Polese et al [38] are needed in conjunction with "hard skills" (ie, technical, analytical) to use big data. Soft skills are critical factors for the effectiveness of data interpretation and extraction of knowledge [38].

Our study identified a key role among teams—a data translator. Studies on big data note the importance of interdisciplinary and multistakeholder interaction and collaboration to enable more useful interpretation of big data [38]. Humanitarian and disaster organizations have increasingly become aware of the importance of data translators in leveraging big data opportunities [39,40].

Technology companies engaged in social good and business enterprises also recognize the role of data translators to meet their mission and goals [41-43]. Particularly when engaging with a novel type of data, this crucial role requires significant technical knowledge as well as an understanding of how the data can be best used. Often, this requires a shift in mindset for data translators to communicate simple, actionable information and communicate confidence in results. Other published studies on knowledge transfer between researchers and practitioners highlight the importance of social interactions and the social process that occurs in the environment of data, disasters, and decision-making [29,44].

Limitations

Though qualitative research using near real-time data from an unprecedented global disaster has unique value, it faces challenges and limitations in data collection, bias, and generalizability. This study was limited to the first year of the COVID-19 global pandemic and does not reflect evolutions of the policy making and response planning among those interviewed over the course of the pandemic. Although Facebook (Meta) data reflect digital movement of individuals during disasters, other mobility data were not included in this study. Facebook's history of collaborating during disasters and building the Facebook Data for Good program with anonymized and privacy-preserving data sources provided a valuable data set for our exploratory study.

Response bias may also have played a role with the interviewees, but all attempts were made to interview individuals and groups during the period of analyses to ensure information was gathered as proximate to the experience as possible. All interviews were conducted in English, which may bias the results toward the perceptions of individuals and groups who are proficient in this language. Although much research in data and disaster focuses on the impact on decision-making, this study intentionally did not explore direct decision-making. Rather, it focused on exploring and further understanding the preceding processes and human factors that could potentially facilitate or hinder the use of data. It also explored precursors of decision-making, which revealed a notable degree of uncertainty among this interview cohort on actual decision-making.

Conclusion

This phenomenological study using a constructivist perspective with data from the early COVID-19 pandemic environment contributes to the understanding of how novel data sources are used in disaster settings. This study offered a unique opportunity to investigate the application of a novel data source during an unprecedented emergency. This resulted in the challenges of working with, understanding, and communicating these data while simultaneously figuring out how to apply the data for response planning and policy making. The groups we spoke to shared a broad range of perceptions and experiences related to these data, something that is expected in an emergency or disaster setting as well as with a novel data set.

This exploratory study highlights human factors as important facilitators for understanding data purpose, eliciting meaning from the data, and using the data for response planning and policy making. Although there were a few examples of how mobility data were used among specific researcher-practitioner groups, most groups were only able to describe their journey in exploring and learning about the data and share some of the barriers to use. Ultimately, the role of a data translator was crucial for understanding the complexities of this novel data source; adapting workflows, visualizations, and reports to align with end users and decision makers; and communicating this information meaningfully to address the goals of responders and policy makers.

This study complements the larger body of literature evaluating the use of big data, including mobility data during disasters, and highlights the need for a more qualitative perspective on use and value rather than quantitative metrics. Future studies that assess the use of data during disasters should include a focus on human factors, to further understand the role of communication and learning frameworks across diverse collaborative groups, specifically to understand the working relationship between members of these teams and how they navigate roles, expertise, and power in order to achieve their goals. Finally, further qualitative studies are needed to understand more fully the transition from perceived usefulness to decision-making to action in disaster settings.

Acknowledgments

The authors would like to thank all interviewees who took time from their busy roles responding to the largest global pandemic in recent history to provide their feedback and perceptions. The authors would also like to recognize CrisisReady and specifically Caroline Buckee, PhD; Andrew Schroeder, PhD; and Satchit Balsari for their contribution to creating and coordinating the COVID-19 Mobility Data Network. This study was funded in part by the Institute for Global Health Catalyzer Grant and the Peterson Foundation grant from Northwestern University.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.
[DOCX File , 377 KB - humanfactors_v11i1e52257_app1.docx]

References

1. COVID-19 dashboard. Johns Hopkins University & Medicine Coronavirus Resource Center. URL: https://coronavirus.jhu.edu/map.html [accessed 2021-02-01]
2. Courtemanche C, Garuccio G, Le A, Pinkston J, Yelowitz A. Strong social distancing measures in the United States reduced the COVID-19 growth rate. Health Aff (Millwood) 2020 Jul;39(7):1237-1246 [FREE Full text] [doi: 10.1377/hlthaff.2020.00608] [Medline: 32407171]
3. Ferguson N, Laydon D, Nedjati-Gilani G, Natsuko I, Ainslie K, Baguelin M, et al. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College. 2020 Mar 16. URL: https://www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2020-03-16-COVID19-Report-9.pdf [accessed 2024-07-04]
4. Thunström L, Newbold SC, Finnoff D, Ashworth M, Shogren JF. The benefits and costs of using social distancing to flatten the curve for COVID-19. J. Benefit Cost Anal 2020 May 21;11(2):179-195. [doi: 10.1017/bca.2020.12]
5. Chu DK, Akl EA, Duda S, Solo K, Yaacoub S, Schünemann HJ, COVID-19 Systematic Urgent Review Group Effort (SURGE) study authors. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. Lancet 2020 Jun 27;395(10242):1973-1987 [FREE Full text] [doi: 10.1016/S0140-6736(20)31142-9] [Medline: 32497510]
6. Talic S, Shah S, Wild H, Gasevic D, Maharaj A, Ademi Z, et al. Effectiveness of public health measures in reducing the incidence of covid-19, SARS-CoV-2 transmission, and covid-19 mortality: systematic review and meta-analysis. BMJ 2021 Dec 03;375:n2997. [doi: 10.1136/bmj.n2997] [Medline: 34862160]
7. Megna R. First month of the epidemic caused by COVID-19 in Italy: current status and real-time outbreak development forecast. Glob Health Res Policy 2020 Oct 09;5(1):43 [FREE Full text] [doi: 10.1186/s41256-020-00170-3] [Medline: 33052317]
8. IM Process. Confluence. URL: https://humanitarian.atlassian.net/wiki/spaces/imtoolbox/pages/219676845/IM+Process [accessed 2024-04-23]
9. Humanitarian Coordination and Information Management Sector Update. USAID. 2015 Oct. URL: https://2017-2020.usaid.gov/sites/default/files/documents/1866/Humanitarian%20Coordination%20and%20Information%20Management%20Sector%20Update%20-%20October%202015.pdf [accessed 2024-07-04]
10. Gutiérrez M. Participation in a datafied environment: questions about data literacy. Comunicação e sociedade 2019;36:37-55 [FREE Full text]
11. Gostin LO, Wiley LF. Governmental public health powers during the COVID-19 pandemic: stay-at-home orders, business closures, and travel restrictions. JAMA 2020 Jun 02;323(21):2137-2138. [doi: 10.1001/jama.2020.5460] [Medline: 32239184]
12. Jacobsen GD, Jacobsen KH. Statewide COVID-19 stay-at-home orders and population mobility in the United States. World Med Health Policy 2020 Dec 29;12(4):347-356 [FREE Full text] [doi: 10.1002/wmh3.350] [Medline: 32837774]
13. Ren W, Tong X, Du J, Wang N, Li SC, Min G, et al. Privacy-preserving using homomorphic encryption in Mobile IoT systems. Computer Communications 2021 Jan;165:105-111. [doi: 10.1016/j.comcom.2020.10.022]
14. Lu X, Bengtsson L, Holme P. Predictability of population displacement after the 2010 Haiti earthquake. Proc Natl Acad Sci U S A 2012 Jul 17;109(29):11576-11581 [FREE Full text] [doi: 10.1073/pnas.1203882109] [Medline: 22711804]
15. Wang L, Ben X, Adiga A, Sadilek A, Tendulkar A, Venkatramanan S, et al. Using mobility data to understand and forecast COVID19 dynamics. medRxiv Preprint posted online on December 15, 2020 [FREE Full text] [doi: 10.1101/2020.12.13.20248129] [Medline: 33354685]
16. Klein B, LaRock T, McCabe S, Torres L, Friedland L, Kos M, et al. Characterizing collective physical distancing in the U.S. during the first nine months of the COVID-19 pandemic. PLOS Digit Health 2024 Feb;3(2):e0000430 [FREE Full text] [doi: 10.1371/journal.pdig.0000430] [Medline: 38319890]
17. Brown TS, Engø-Monsen K, Kiang MV, Mahmud AS, Maude RJ, Buckee CO. The impact of mobility network properties on predicted epidemic dynamics in Dhaka and Bangkok. Epidemics 2021 Jun;35:100441 [FREE Full text] [doi: 10.1016/j.epidem.2021.100441] [Medline: 33667878]
18. Chang M, Kahn R, Li Y, Lee C, Buckee CO, Chang H. Variation in human mobility and its impact on the risk of future COVID-19 outbreaks in Taiwan. BMC Public Health 2021 Jan 27;21(1):226 [FREE Full text] [doi: 10.1186/s12889-021-10260-7] [Medline: 33504339]
19. Nixon K, Jindal S, Parker F, Reich NG, Ghobadi K, Lee EC, et al. An evaluation of prospective COVID-19 modeling: from data to science translation. medRxiv Preprint posted online on April 19, 2022 [FREE Full text] [doi: 10.1101/2022.04.18.22273992]
20. COVID-19 Mobility Data Network. URL: https://www.covid19mobility.org/ [accessed 2024-07-04]
21. COVID-19 Mobility Data Network – annual report 2020. CrisisReady. 2021 Feb. URL: https://www.crisisready.io/publications/covid-19-mobility-data-network-annual-report-2020/ [accessed 2024-07-04]
22. Vieweg SE. Situational awareness in mass emergency: a behavioral and linguistic analysis of microblogged communications. University of Colorado at Boulder ProQuest Dissertations & Theses. 2012. URL: https://www.proquest.com/openview/540ee2ba902309c5ad7314438e06ea42/1 [accessed 2024-07-04]

23. Bennett KJ, Olsen JM, Harris S, Mekaru S, Livinski AA, Brownstein JS. The perfect storm of information: combining traditional and non-traditional data sources for public health situational awareness during hurricane response. PLoS Curr 2013 Dec 16;5:1 [FREE Full text] [doi: 10.1371/currents.dis.d2800aa4e536b9d6849e966e91488003] [Medline: 24459610]
24. Luokkala P, Virrantaus K. Developing information systems to support situational awareness and interaction in time-pressuring crisis situations. Safety Science 2014 Mar;63:191-203. [doi: 10.1016/j.ssci.2013.11.014]
25. DeCuir-Gunby JT, Marshall PL, McCulloch AW. Developing and using a codebook for the analysis of interview data: an example from a professional development research project. Field Methods 2010 Dec 27;23(2):136-155. [doi: 10.1177/1525822X10388468]
26. MacQueen KM, McLellan E, Kay K, Milstein B. Codebook development for team-based qualitative analysis. CAM Journal 2016 Sep 14;10(2):31-36. [doi: 10.1177/1525822X980100020301]
27. Strauss A, Corbin J. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Thousand Oaks, CA: SAGE Publications, Inc; 1990.
28. Barry CA, Britten N, Barber N, Bradley C, Stevenson F. Using reflexivity to optimize teamwork in qualitative research. Qual Health Res 1999 Jan;9(1):26-44. [doi: 10.1177/104973299129121677] [Medline: 10558357]
29. Delaney W, Ames G. Integration and exchange in multidisciplinary alcohol research. Soc Sci Med 1993 Jul;37(1):5-13. [doi: 10.1016/0277-9536(93)90311-q] [Medline: 8332924]
30. Data Party. Better Evaluation. URL: https://www.betterevaluation.org/methods-approaches/methods/data-party [accessed 2024-07-04]
31. APA Dictionary of Psychology. American Psychological Association. URL: https://dictionary.apa.org/ [accessed 2024-07-04]
32. Janesick VJ. Peer Debriefing. In: The Blackwell Encyclopedia of Sociology. Hoboken, NJ: John Wiley & Sons, Ltd; 2015.
33. Stedman C. What is data preparation? An in-depth guide. TechTarget. URL: https://www.techtarget.com/searchbusinessanalytics/definition/data-preparation [accessed 2024-07-04]
34. Vickery J, Atkinson P, Lin L, Rubin O, Upshur R, Yeoh E, et al. Challenges to evidence-informed decision-making in the context of pandemics: qualitative study of COVID-19 policy advisor perspectives. BMJ Glob Health 2022 Apr 21;7(4):e008268 [FREE Full text] [doi: 10.1136/bmjgh-2021-008268] [Medline: 35450862]
35. Paulus D, Fathi R, Fiedrich F, de Walle BV, Comes T. On the interplay of data and cognitive bias in crisis information management: an exploratory study on epidemic response. Inf Syst Front 2022 Mar 22:1-25 [FREE Full text] [doi: 10.1007/s10796-022-10241-0] [Medline: 35342331]
36. Endsley MR. Toward a theory of situation awareness in dynamic systems. Hum Factors 2016 Nov 23;37(1):32-64 [FREE Full text] [doi: 10.1518/001872095779049543]
37. Morss RE, Wilhelmi OV, Downton MW, Gruntfest E. Flood risk, uncertainty, and scientific information for decision making: lessons from an interdisciplinary project. Bull Amer Meteor Soc 2005 Nov;86(11):1593-1602 [FREE Full text] [doi: 10.1175/BAMS-86-11-1593]
38. Polese F, Troisi O, Grimaldi M, Romeo E. A big data-oriented approach to decision-making: a systematic literature review. 2019 Presented at: 22nd Excellence in Services International Conference; August 29-30, 2019; Thessaloniki, Greece URL: https://sites.les.univr.it/eisic/wp-content/uploads/2019/11/31-Polese-Troisi-Grimaldi-Romeo-1.pdf
39. Oroz ML. From big data to humanitarian-in-the-loop algorithms. UNHCR. URL: https://www.unhcr.org/innovation/big-data-humanitarian-loop-algorithms/ [accessed 2024-07-04]
40. Why data literacy is important in the aid sector. Information Management Resource Portal. 2021 Jul 16. URL: https://www.im-portal.org/blogs/why-data-literacy-is-important-in-the-aid-sector [accessed 2042-07-04]
41. Weiss H. We underestimate the power of data at our peril. This is why. World Economic Forum. 2019 Apr 02. URL: https://www.weforum.org/agenda/2019/04/we-underestimate-the-immense-power-of-data-at-our-peril-this-is-why%20/ [accessed 2024-07-04]
42. Henke N, Levine J, McInerney P. You don’t have to be a data scientist to fill this must-have analytics role. Harvard Business Review. 2018 Feb 05. URL: https://hbr.org/2018/02/you-dont-have-to-be-a-data-scientist-to-fill-this-must-have-analytics-role [accessed 2024-07-04]
43. Maynard-Atem L, Ludford B. The rise of the data translator. Impact 2020 May 03;2020(1):12-14 [FREE Full text] [doi: 10.1080/2058802X.2020.1735794]
44. Fothergill A. Knowledge transfer between researchers and practitioners. Natural Hazards Review 2000 May 01;1(2):91-98 [FREE Full text] [doi: 10.1061/(ASCE)1527-6988(2000)1:2(91)]

Abbreviations

CMDN: COVID-19 Mobility Data Network
PPE: personal protective equipment

Edited by A Kushniruk; submitted 01.09.23; peer-reviewed by M Pandin, Z Li, E Benjamin; comments to author 17.02.24; revised version received 30.04.24; accepted 22.06.24; published 01.08.24.

Please cite as:
Chan JL, Tsay S, Sambara S, Welch SB
Understanding the Use of Mobility Data in Disasters: Exploratory Qualitative Study of COVID-19 User Feedback
JMIR Hum Factors 2024;11:e52257
URL: https://humanfactors.jmir.org/2024/1/e52257
doi:10.2196/52257
PMID:39088256

Original Paper

Usability and Acceptability of a Conversational Agent Health Education App (Nthabi) for Young Women in Lesotho: Quantitative Study

Elizabeth Nkabane-Nkholongo[1*], RN, MPH; Mathildah Mpata-Mokgatle[1*], PHD; Brian W Jack[2*], MD; Clevanne Julce[3*], MPH; Timothy Bickmore[4*], PHD

[1]School of Public Health, Sefako Makgatho Health Sciences University, Pretoria, South Africa
[2]Chobanian & Avedisian School of Medicine, Boston University, Boston, MA, United States
[3]Umass Chan Medical School, University of Massachusetts, Worcester, MA, United States
[4]Khoury College of Computer Sciences, Northeastern University, Boston, MA, United States
[*]all authors contributed equally

Corresponding Author:
Elizabeth Nkabane-Nkholongo, RN, MPH
School of Public Health
Sefako Makgatho Health Sciences University
Motlalegi Street
Garankuwa
Pretoria, 0208
South Africa
Phone: 27 (012) 521 5759 ext 012
Fax: 27 (012) 560 0086
Email: elizabeth.nkholongo@gmail.com

Abstract

Background: Young women in Lesotho face myriad sexual and reproductive health problems. There is little time to provide health education to women in low-resource settings with critical shortages of human resources for health.

Objective: This study aims to determine the acceptability and usability of a conversational agent system, the Nthabi health promotion app, which was culturally adapted for use in Lesotho.

Methods: We conducted a descriptive quantitative study, using a 22-item Likert scale survey to assess the perceptions of the usability and acceptability of 172 young women aged 18-28 years in rural districts of Lesotho, who used the system on either smartphones or tablets for up to 6 weeks. Descriptive statistics were used to calculate the averages and frequencies of the variables. χ^2 tests were used to determine any associations among variables.

Results: A total of 138 participants were enrolled and completed the survey. The mean age was 22 years, most were unmarried, 56 (40.6%) participants had completed high school, 39 (28.3%) participants were unemployed, and 88 (63.8%) participants were students. Respondents believed the app was helpful, with 134 (97.1%) participants strongly agreeing or agreeing that the app was "effective in helping them make decisions" and "could quickly improve health education and counselling." In addition, 136 (98.5%) participants strongly agreed or agreed that the app was "simple to use," 130 (94.2 %) participants reported that Nthabi could "easily repeat words that were not well understood," and 128 (92.7%) participants reported that the app "could quickly load the information on the screen." Respondents were generally satisfied with the app, with 132 (95.6%) participants strongly agreeing or agreeing that the health education content delivered by the app was "well organised and delivered in a timely way," while 133 (96.4%) participants "enjoyed using the interface." They were satisfied with the cultural adaptation, with 133 (96.4%) participants strongly agreeing or agreeing that the app was "culturally appropriate and that it could be easily shared with a family or community members." They also reported that Nthabi was worthwhile, with 127 (92%) participants reporting that they strongly agreed or agreed that they were "satisfied with the application and intended to continue using it," while 135 (97.8%) participants would "encourage others to use it." Participants aged 18-24 years (vs those aged 25-28 years) agreed that the "Nthabi app was simple to use" (106/106, 100% vs 30/32, 98.8%; *P*=.01), and agreed that "the educational content was well organised and delivered in a timely way" (104/106, 98.1% vs 28/32, 87.5%; *P*=.01).

Conclusions: These results support further study of conversational agent systems as alternatives to traditional face-to-face provision of health education services in Lesotho, where there are critical shortages of human resources for health.

Trial Registration: ClinicalTrials.gov NCT04354168; https://www.clinicaltrials.gov/study/NCT04354168

(JMIR Hum Factors 2024;11:e52048) doi:10.2196/52048

KEYWORDS

preconception care; conversational agent technology; women's health education; mHealth adaptation; health information technology; health education in Africa; education; women's health; women; woman; health information; young women; survey; usability; acceptability; application; applications; app; health promotion

Introduction

Background

Digital health interventions offer considerable promise to develop new models of health care delivery and to have a large public health impact [1]. Digital channels, such as the internet, mobile phone messaging, social media, apps, voice video messaging, and telemedicine have been shown to improve the delivery of health education and care. These tools have tremendous potential to impact large-scale health promotion efforts as a cost-effective and scalable solution to address public health challenges, such as delivering sexual health education [2].

The rapid diffusion of mobile technology and advances in artificial intelligence have facilitated this trend [3]. The use of mobile devices and services has continued to increase globally, though at different rates in developed and developing countries. By the end of 2018, more than 5 billion people worldwide subscribed to mobile services, accounting for 67% of the global population, and this number is expected to exceed 70% by 2025 [4].

In Lesotho, 94% of people aged 18-29 years use smartphones, and 3G data coverage is available in almost 90% of the country [5]. This is an important group to target, as they represent the highest proportion of global consumers of mobile technology. This high penetration of mobile technologies provides an opportunity to assess the usability and acceptability of using new mobile health technologies as an alternative to the traditional face-to-face provision of health education.

Adolescents and young women continue to report low levels of sexual and reproductive health knowledge, and engage in risky sexual behaviors [6]. They also face a myriad of sexual and reproductive health problems, such as unplanned pregnancy, sexually transmitted infections, and HIV infections. Advancing sexual and reproductive health education for adolescents and young women in Africa is particularly important, as HIV accounts for 42% of new HIV infections globally [7], and 4 in 5 young people with HIV live in sub-Saharan Africa [8]. Therefore, developing new ways to provide sexual and reproductive health education in Africa is particularly important.

Lesotho is a lower middle-income country in southern Africa and has the second highest HIV prevalence in the world—at 22.7%—and one of the highest HIV incidences among adolescent girls and young women (0.33%) [9]. The maternal mortality ratio in Lesotho is the second highest in Southern African Development Community countries (544/100,000 live births) [10]. The ratio of doctors to the population is 0.9 per 10,000. For nurse-midwives, the ratio is 10.2 per 10,000, [11] which poses a challenge to the delivery of face-to-face health education.

Delivering health education via new mobile health tools has the potential to provide alternatives to traditional face-to-face provision of health education. Conversational agents are computer-based animated characters that are designed to simulate face-to-face human interactions. The human–computer interface relies only minimally on text comprehension and prioritizes conversation, thereby making it more accessible to patients with limited health literacy [12]. In health care, patient-facing conversational agents are increasingly used to deliver education, provide self-management of chronic conditions, perform routine tasks, such as appointment booking, and support health professionals' decision-making for diagnosis and triage in mental health [13,14]. These devices have the potential to automate tasks, improve access to health care services, and reduce health professionals' workload.

Prior Work

In the United States, a conversational agent named Gabby was designed to deliver preconception sexual and reproductive health information to reproductive-age African American women. Using Gabby demonstrated significant improvement in addressing reproductive health risks in randomized controlled trials [15,16].

Our research team culturally adapted Gabby to provide sexual and reproductive health education to young women in Lesotho. The newly adapted system, named the Nthabi Preconception Health Promotion App (hereafter referred to as Nthabi) is a patient-facing conversational agent that screens for sexual and reproductive health risks, and uses behavior change techniques, such as motivational interviewing and shared decision-making, to facilitate behavior change related to these risks.

The perceived appropriateness of Nthabi adaptation was studied in focus groups with young women aged 18-28 years (n=33 participants) who had used the system for 4 weeks [17]. Participants reported that adaptations were culturally appropriate, and provided relevant and culturally sensitive clinical information. They emphasized that the physical characteristics, personal and nonverbal behaviors, use of Sesotho (the local language in Lesotho) words and idioms, and clinical content were sensitively delivered and culturally appropriate. Interviews with the Ministry of Health key informants agreed that the adaptation was successful and that the system holds great potential to improve the delivery of health education content in Lesotho.

Goal of This Study

The goal of this study is to assess the perceived usability and acceptability of the Nthabi Preconception Health Promotion App among 160 young women enrolled in a clinical trial in Lesotho who had used the system for up to 6 weeks.

Methods

Study Design

In this paper, we report the results of a survey designed to assess the perceived usability and acceptability of Nthabi among the first 160 young women who used the system.

Usability is defined as the extent to which young women can use Nthabi to achieve specific goals with effectiveness, efficiency, and satisfaction [18]. Acceptability includes the satisfaction of the young women, attitudes toward using the app, and intention or willingness to continue using the app.

Study Population and Setting

The population studied was young women aged 18 to 28 years in the Leribe and Berea districts of the rural, mountainous, lower middle-income country of Lesotho in southern Africa.

Sampling

This study was conducted to assess the usability and acceptability of using Nthabi as a health education tool in Lesotho; therefore, a convenience sample of 200 young women was chosen from the population of young women in the districts of Leribe and Berea.

Recruitment

Participants were recruited in several ways. First, the research team posted messages on social media (eg, WhatsApp and Facebook) that described the study and asked potential participants to contact the research team to discuss enrolling in the study. A nongovernmental organization called Help Lesotho, which offers mentorship programs to adolescent girls and young women in the Leribe district, saw the social media posting, reached out to the research team, and offered to disseminate the recruitment announcement to their clients.

Second, the research team directly approached young women while they were waiting for consultation at the Adolescent Health Corners (clinics) and HIV and Mother and Child Health ambulatory clinical departments at the Berea and Leribe government district hospitals. Last, students were approached at the Leribe Vocational High School and the Limkokwing University of Technology to identify individuals who might be interested in participating.

Eligibility Criteria

The inclusion criteria were the following: (1) Basotho women aged 18-28 years who were from the districts of Leribe and Berea and accessed health services in these 2 districts, (2) self-reported ability to read and understand spoken English, (3) access to an Android smartphone, and (4) ability to access internet and Wi-Fi at least once at the end of the study. Those not meeting these criteria were excluded.

Enrollment

The research team assisted the participants in downloading the app on their mobile phones. Participants who were unable to download the app on their mobile phones were loaned a Lenovo Android 11 OS platform tablet to use for 6 weeks. Participants were then assisted to create a unique username and password and were shown how to log on to either their Android mobile phone or tablet and start interacting with Nthabi. Participants were encouraged to use the app at least once daily at their convenience for 6 weeks.

Baseline Data Collection

Sociodemographic information was collected (age, marital status, education level, employment status, recruitment site, and district). A total of 160 participants were enrolled. Participant contact information (phone and WhatsApp number, email address) was collected so they could be reminded to return to the recruitment site so they could access the internet when they were finished using Nthabi, to facilitate survey completion, and return the loaned tablets.

Description of the Nthabi Intervention

Nthabi was adapted in relation to physical characteristics, language, culture, and clinical content appropriate for Lesotho, as previously described (Figure 1) [17]. A description of Nthabi is found in Multimedia Appendix 1.

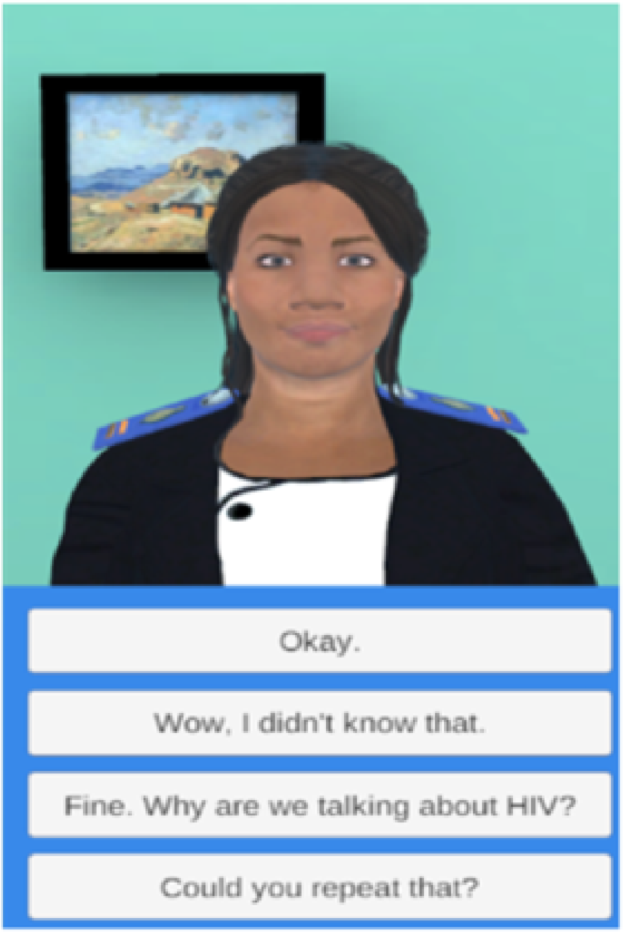
Briefly, Nthabi is an English-speaking Mosotho (person from Lesotho) nurse-midwife dressed as a professional nurse. Her hairstyle (braids), complexion (medium, similar to the local population), facial expressions (calm and gentle), and mannerisms (a humble professional with a sense of humor) were relatable to young women in Lesotho.

To establish the clinical topics to be included in the system, Ministry of Health key informants recommended 5 sexual reproductive health topics for young women (family planning, HIV, tuberculosis, healthy eating, and using folic acid). The research team then used the Lesotho National Clinical Guidelines on these topics to create evidence-based dialogue for use in Nthabi interactions.

During subsequent interactions with Nthabi, women selected the topic they wanted to discuss. Using conversational dialogue, Nthabi describes why the topic is important and offers suggestions about how to take action on it. The woman engages with the app by selecting a response from a multiple-choice menu that is updated at each turn of dialogue.

To increase the accessibility and use of the system, a decision was made that the app would be fully downloadable to the user's mobile phone, thereby enabling full content availability beyond the Wi-Fi environment. Use and information about the content discussed would be downloaded when the user was in a Wi-Fi environment. Nthabi was available from the Google Play store for downloading on mobile phones or tablets.

Figure 1. Nthabi health education interface.



Data Collection Tool

The survey instrument was based on the System Usability Scale and the Mobile App Rating Scale [19] using previous studies of Gabby adaptation [20,21] and modified for use in Lesotho. To ensure that the questions were clear and not ambiguous, the survey tool was reviewed by 12 health professionals, including nurses working in adolescent health, physicians, and district sexual reproductive health clinicians. The survey was then piloted with young women who met the eligibility criteria, to assess the respondents' understanding and interpretation of the questions. Only editorial changes, to enhance clarity, were required. The final survey contained 22 questions that elicited responses on a 4-point Likert scale (strongly agree, agree, disagree, and strongly disagree). Topics covered in the survey were usability (ease of use and reliability), satisfaction, willingness to continue, how easy it was to understand, content organization, and cultural relevance. The survey also enquired about the degree to which Nthabi helped women make health decisions and the degree to which they would encourage others to use the app.

Data Storage and Analysis

Survey data were captured on an Excel (Microsoft Corp) spreadsheet and stored on a password-protected computer. Data were analyzed using Stata software (StataCorp). Descriptive statistics were used to calculate the averages and frequencies of the variables. Inferential statistics, such as χ^2 tests, were used to determine any associations among variables. Statistical significance was set as a threshold of $P<.10$, as this was a feasibility study.

Participant Incentives

All participants received 50 Maloti (approximately US $5) to cover data costs. Participants using tablets were provided an additional 50 Maloti (approximately US $5) to cover their travel back to the recruitment sites to return the devices.

Ethical Considerations

Once the eligibility of participants was confirmed, the research team explained the purpose of the study, potential risks and benefits, compensation for travel costs, and the right to withdraw from the study at any time. After questions had been addressed, participants were asked to sign an informed consent form and were enrolled.

The study was conducted according to the Consolidated Standards of Reporting Trials (CONSORT) [22] and the adaptations for mobile health interventions [23]. Ethical clearance was obtained from the Boston University Research institutional review board (H-40268), Sefako Makgatho University of Health Sciences Ethics Review Committee (SMUREC/H/343/2021: PG), and the Lesotho Ministry of Health Research Ethics Committee (ID 145-2021). Permission was obtained from the study recruitment sites.

Results

Recruitment

The research team screened 436 young women for eligibility, as shown in the CONSORT diagram (Figure 2). Young women were recruited through social media (eg, WhatsApp and Facebook) or direct contact at Limkokwing University of Technology (n=150), Leribe Vocational School (n=88), Leribe Health Facilities (n=55), Berea Health Facilities (n=84), and Help Lesotho (n=59).

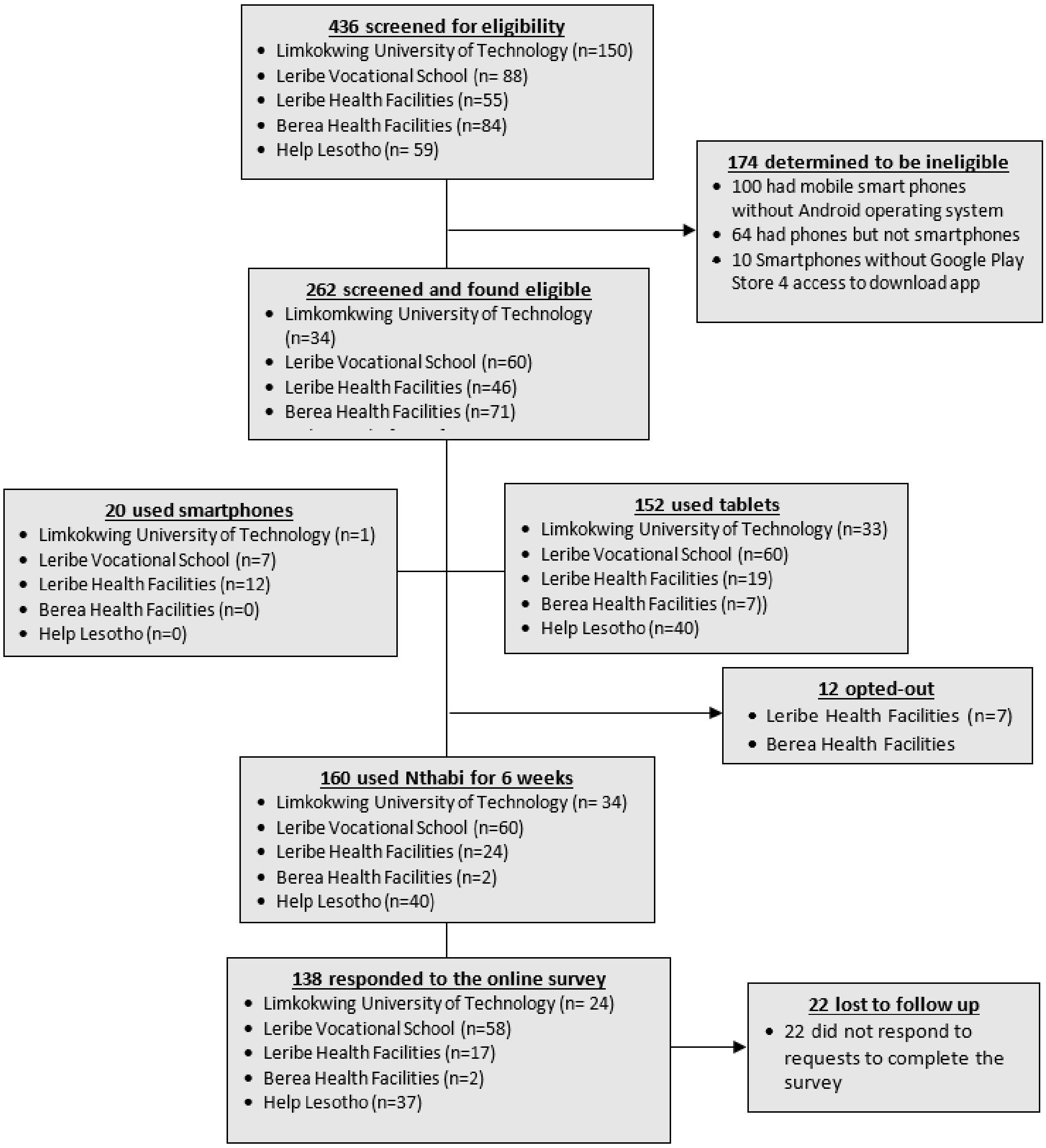
Of those screened, 174 young women were ineligible due to having smartphones without the Android operating system, while 64 young women had phones that were not smartphones, and 10 young women had Huawei Android smartphones that lacked access to the Google Play store.

Consequently, 172 participants were eligible, provided consent, and were enrolled. Those enrolled were from Limkokwing University of Technology (34 of 34 screened), Leribe Vocational School (60 of 60 participants screened), Leribe Health Facilities (31 of 46 participants screened), Berea Health Facilities (7 of 71 participants screened), and Help Lesotho (40 of 51 participants screened).

Of those enrolled, only 20 participants had sufficient memory on their phones to download the Nthabi app, and 152 participants received a tablet device to use. Of those who were able to download the Nthabi app on their mobile phones, 1 of 34 participants was from Limkokwing University of Technology, 7 of 60 participants were from Leribe Vocational School, and 12 of 31 participants were from Leribe Health Facilities.

In the weeks after enrollment, 12 participants opted out of the study because their phones froze and jammed when they tried to load the app. Therefore, 160 young women used Nthabi for up to 6 weeks, with 8 young women using phones and 152 young women using loaned tablets. At the end of 6 weeks, 138 young women responded to the survey (80 young women who had been recruited from the technology and vocational schools, 19 young women from the district health facilities, and 37 young women from the Help Lesotho program), and 22 young women did not respond to requests to complete the survey.

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



Sociodemographic Results

Table 1 shows the characteristics of the 138 participants who were enrolled and who completed the survey after 6 weeks. The mean age was 22 years (SD 2.7 years), most were unmarried, 56 (40.6%) participants had completed high school, 39 (28.3%) participants were unemployed, and 88 (63.8%) participants were students. The recruitment sites of those participants completing surveys were 24 (17.4%) participants from Limkokwing University of Technology, 58 (42%) participants from Leribe Vocational School, 17 (12.3%) participants from Leribe health facilities, 2 (1.4%) participants from Berea Health Facilities, and 37 (26.8%) participants from Help Lesotho.

Table 1. Sociodemographic characteristics of respondents (n=138).

Characteristics	Respondents, n (%)
Age (years)	
18-20	34 (24.6)
21-23	53 (38.4)
24-26	37 (26.8)
27-28	14 (10.1)
Marital status	
Married	11 (8)
Not married	127 (92)
Level of education	
Primary	2 (1.5)
High school	56 (40.6)
College	39 (28.3)
University	41 (29.7)
Employment status	
Employed	11 (8)
Unemployed	39 (28.3)
Student	88 (63.8)
Recruitment site	
Limkokwing University of Technology	24 (17.4)
Leribe Vocational School	58 (42)
Leribe Health Facilities	17 (12.3)
Berea Health Facilities	2 (1.4)
Help Lesotho	37 (26.8)

Survey Results

Table 2 shows the survey responses of the 138 young women who completed the survey. Overall, the results show that participants perceived usability and acceptability positively.

Described below are the survey responses corresponding to the components of usability (effectiveness, efficiency, and satisfaction) and acceptability (satisfaction, attitudes toward use, and intention to continue using Nthabi).

Respondents believed the app was helpful, with 134 (97.1%) participants strongly agreeing or agreeing that the app was "effective in helping them make decisions" and "could quickly improve health education and counselling."

Participants generally liked using the app, with 136 (98.6%) participants strongly agreeing or agreeing that the app was "simple to use," while 132 (95.7%) participants reported that "symbols and buttons are easy to use," 130 (94.3%) participants reported that Nthabi could "easily repeat words that were not well understood," and 128 (92.8%) participants reported that the app "could quickly load the information on the screen."

Respondents were generally satisfied with the app, with 132 (95.7%) participants strongly agreeing or agreeing that the health education content delivered by the app was "well organised and delivered in a timely way," while 133 (96.4%) participants "enjoyed using the interface."

In addition, 132 (95.7%) participants strongly agreed or agreed that they were able to complete tasks quickly using the app, while 136 (98.6%) participants reported that "I can quickly remember how to use the app after a while," and 137 (99.3%) participants reported that "it was easy to learn how to use the app."

The items rated less positively include the following: "it was easy to converse and type responses into the app" according to 95 (68.8%) participants, and "I could easily correct mistakes" according to 106 (76.8%) participants.

They also were satisfied with the cultural adaptation, with 133 (96.4%) participants strongly agreeing or agreeing that the app was "culturally appropriate and that it could be easily shared with a family or community members."

Finally, they also reported that Nthabi was worthwhile, with 127 (92%) participants reporting that they strongly agreed or agreed that they were "satisfied with the application and intended to continue using it" while 135 (97.8%) participants would "encourage others to use it."

Table 2. Survey responses on the usability and acceptability of Nthabi app (n=138).

To what extent do you agree with the following statements?	Strongly agree, n (%)	Agree, n (%)	Disagree, n (%)	Strongly disagree, n (%)
It was simple to use this app	90 (65.2)	46 (33.3)	1 (0.7)	1 (0.7)
It was easy to find the information I needed	51 (37)	68 (49.3)	13 (9.4)	6 (4.4)
It was easy to converse and type responses into this app	36 (26.1)	59 (42.8)	33 (23.9)	10 (7.3)
The information on the app screen is well-organized	59 (42.8)	76 (55.1)	1 (0.7)	2 (1.5)
It was easy to learn how to use the app	94 (68.1)	43 (31.2)	1 (0.7)	0 (0)
The symbols and buttons are easy to use	61 (44.2)	71 (51.5)	5 (3.6)	1 (0.7)
I understood how the app works the first time I used it	76 (58.1)	53 (38.4)	6 (4.4)	3 (2.2)
I can quickly remember how to use the app after a while	88 (63.8)	48 (34.8)	1 (0.7)	1 (0.7)
When I made a mistake using the app, I could easily correct the mistake	37 (26.8)	69 (50)	29 (21.1)	3 (2.2)
The app offered error messages that clearly told me how to fix the issues	22 (15.9)	57 (41.3)	52 (37.7)	7 (5.1)
The app could easily repeat words or statements that were not well understood	91 (66)	39 (28.3)	7 (5.1)	1 (0.7)
The app quickly loads the information on the screen	70 (50.7)	58 (42)	8 (5.8)	2 (1.5)
The health education content provided by the app was well-organized and delivered in a timely way	75 (54.3)	57 (41.3)	5 (3.6)	1 (0.7)
I was able to complete tasks quickly using the app	68 (49.3)	64 (46.4)	3 (2.2)	3 (2.2)
The app information was effective in helping me make decisions	74 (53.6)	60 (43.5)	1 (0.7)	3 (2.2)
The app has not stopped working or has ever closed	65 (47.1)	51 (37)	19 (13.8)	3 (2.2)
I believe the app could quickly improve health education and counseling	85 (61.59)	49 (35.51)	1 (0.72)	3 (2.1)
The app interface is nice to use	67 (48.6)	63 (45.7)	4 (2.9)	4 (2.9)
I enjoyed using the app interface	78 (56.5)	55 (39.9)	3 (2.2)	2 (1.5)
I am satisfied with the app and intend to continue using it	78 (56.5)	49 (35.5)	8 (5.8)	3 (2.2)
I want to encourage others to use the app	91 (66)	44 (31.9)	2 (1.5)	1 (0.7)
The app was culturally appropriate and I could easily share it with a family member or community member	79 (57.3)	54 (39.1)	5 (3.6)	0 (0)

Survey Responses by Age, Marital, and Education Status

Table 3 shows selected survey responses of the 138 participants who completed the survey questions by age, education, and marital status.

Participants aged 18-24 years (vs those aged 25-28 years) agreed that the "Nthabi app was simple to use" (106/106, 100% vs 30/32, 93.8%; *P*=.01), and agreed that "the educational content was well organised and delivered in a timely way" (104/106, 98.1% vs 28/32, 87.5%; *P*=.01).

Participants who were married (vs unmarried) agreed that "the educational content was well organised and delivered in a timely way" (9/11, 81.8% vs 123/127, 96.9%; *P*=.02), and agreed that "the app was nice to use" (9/11, 81.8% vs 121/127, 95.3%; *P*=.07).

Finally, young women who were in high school (vs those in tertiary education) were more likely to agree that "the app offered error messages that clearly told me how to fix the issue" (37/56, 66.1% vs 41/80, 51.3%; *P*=.02), and were "satisfied with the application and intended to continue using it" (55/56, 98.2% vs 70/80, 87.5%; *P*=.07).

Taken together, these results indicate that younger women, those in high school (and usually younger), and those unmarried (and usually younger) perceived Nthabi more positively.

Table 3. Survey responses of young women using Nthabi app by age, marital, and educational status.

	Agree, n (%)	Disagree, n (%)	*P* value
Opinions of young women and their marital status			
The health education content provided by the app was well-organized and delivered in a timely way			.02
Married (n=11	9 (81.8)	2 (18.2)	
Not married (n=127)	123 (96.9)	4 (3.2)	
The app interface is nice to use			.07
Married (n=11)	9 (81.8)	2 (18.1)	
Not married (n=127)	121 (95.3)	3 (2.4)	
Opinions of young women and their age range			
It was simple to use this app			.01
18-24 (n=106)	106 (100)	0 (0)	
25-28 (n=32)	30 (93.8)	2 (6.2)	
The health education content provided by the app was well organised and delivered in a timely way			.01
18-24 (n=106)	104 (98.1)	2 (1.9)	
25-28 (n=32)	28 (87.5)	4 (12.5)	
Opinions of young women and their educational status			
The app offered error messages that clearly told me how to fix the issues			.02
Primary school (n=2)	1 (50)	1 (50)	
High school (n=56)	37 (66.1)	19 (33.9)	
Tertiary (n=80)	41 (51.3)	39 (48.8)	
I am satisfied with the app and intend to continue using it			.07
Primary school (n=2)	2 (100)	0 (0)	
High school (n=56)	55 (98.2)	1 (1.8)	
Tertiary (n=80)	70 (87.5)	10 (12.5)	

Discussion

Principal Results

Young women in the lower middle-income country of Lesotho in southern Africa who used the newly adapted Nthabi intervention for up to 6 weeks perceived the usability and acceptability of the system very positively. Most respondents were satisfied with Nthabi and perceived it to be effective, efficient, and culturally appropriate. Participants agreed that Nthabi helped them make decisions and could improve the delivery of health education. They reported it was easy to use and well organized. Most intended to use it beyond the study period and they said they would encourage others to use it.

Improving sexual reproductive health education is a clear priority in Lesotho [9,10]. This study supports the idea that conversational agent technologies can provide sexual and reproductive health education in a rural, mountainous country like Lesotho, which has profound human resources challenges. As additional data are collected, the Ministry of Health and the health development and implementing partners should consider using Nthabi as a health promotion and education tool in Lesotho.

Comparison With Prior Work

These findings are in accordance with our previous research reporting results of focus groups of potential users who used an early version of Nthabi and key informant interviews of Ministry of Health officials. Participants reported that adaptations were culturally appropriate, and provided relevant and culturally sensitive clinical information. These qualitative data and now survey data together highlight the importance of acknowledging the local context when adapting an intervention. Nthabi was adapted to the uses, languages, interests, and realities of young people, as well as the importance of knowing what is preferred by young people as a measure of attractiveness to promote user engagement [24]. Most respondents were satisfied with the educational content and agreed that it delivered culturally appropriate and sensitive sexual and reproductive health information. Adaptations of interventions using appropriate cultural cues have a higher probability of acceptability and usability [25]. Culturally responsive interventions are effective in enhancing knowledge acquisition, attitudes, and satisfaction since they respect cultural diversity and the sociocultural factors that may affect health [26,27].

Participants agreed Nthabi could improve the delivery of health education and help them make health decisions. This finding is similar to findings from other studies conducted in lower middle-income countries, which provide evidence that a variety of mHealth apps such as voice messages and daily educational text messages can improve young people's sexual reproductive health [28] and have been shown to be feasible and acceptable for improving health education and knowledge among adolescents and young people [29]. Other studies highlighted the broad potential for digital interventions to enhance health promotion and service delivery toward better sexual health [30,31]. However, this is the first study of the acceptability and usability of potentially more engaging and effective conversational agent systems in a low- and middle-income country in southern Africa.

Younger women in this study sample appear to have more positive perceptions of Nthabi than older participants. They found the system simple to use and the content delivered in a way convenient to them. Younger women might be more familiar and comfortable with using new technologies. This is consistent with other studies of women from the global north showing their preference for digital technologies such as readily available information, and their preference for opportunities to learn more about their bodies and health status [32]. Other studies have found that younger people are not only accepting of new technologies in health care settings but are actually looking for more of these technologies to use in health settings [33,34].

Accessibility of Nthabi on Mobile Devices

In Lesotho, 94% of people aged 18-29 years use smartphones, and 3G data coverage is available in almost 90% of the country [5], yet access to public Wi-Fi and data costs remain barriers to using mobile technologies for health education. Nthabi was designed to address our concern that limited internet access would impact participants' use of Nthabi. A decision was made to download the full system to mobile devices so that participants could use the system when not in Wi-Fi environments. While this design allowed the participants to use the system at their convenience, the inclusion of all the content and most importantly, the inclusion of the system voice synthesizer, created significant difficulties for downloading and using Nthabi on most phones due to low phone memory. The finding that only 8 of 172 (4.7%) participants were able to use Nthabi on their phones demonstrates that mobile phone use is possible, though practically, only phones with sufficient available memory could be used. As it becomes increasingly possible for young women to have regular access to public Wi-Fi, it will become possible for more young women to use Nthabi in the cloud on their phones rather than downloading the full system.

Participants who were unable to download the intervention to their phones were loaned tablet devices. We purchased 20 devices (US $111 per tablet or US $14 per participant) and loaned them to participants on a rolling basis. At the end of the study, all tablets were returned. While this is a cost-effective alternative, future studies of large-scale health education programs in low-resource settings using cloud-based interventions will be possible with increased public Wi-Fi availability. We are now planning studies in which fully downloadable and Wi-Fi–enabled systems are available.

Limitations

There are several limitations to this study. First, the results are not nationally representative of women from Lesotho as participants were recruited by convenience from only 2 of the 10 districts of Lesotho. The sample included many participants recruited from the university and vocational schools, and while these participants reported residing in and receiving health services in Berea and Leribe, the results do not necessarily reflect the views of women living in rural areas. Further trials are needed to more definitively identify the perceptions of rural women. A larger study in all 10 districts of Lesotho is planned.

Second, while this study reports on perceptions of successful usability and acceptability, it does not provide evidence that the intervention improved young women's health knowledge, attitudes, and behaviors. Research to further determine the impact of knowledge of the topics discussed by Nthabi is underway.

Conclusions

Nthabi is a potentially useful intervention for providing sexual reproductive health information for young women in the rural, lower middle-income country of Lesotho with limited human resources in health. Further study of the Nthabi system is warranted to determine if the Nthabi health education content and interactive dialogue about sexual and reproductive health can improve women's knowledge, attitudes, and health behaviors.

Acknowledgments

All authors made substantial contributions to draft and critically revise the manuscript. EN-N, MM-M, and BWJ contributed to conceptualizing the study and to data acquisition, analysis, interpretation, and writing the manuscript. TB contributed to conceptualizing the study and as an expert in conversational agents. CJ contributed study design, data collection, and manuscript writing. All authors approved the final draft and agreed to be accountable for all aspects of the work. The authors are grateful to the Lesotho Ministry of Health leadership, young women who participated in this study, and research assistants Paballo Lethunya and Moleboheng Mofolo. The research reported in this publication was supported by the Fogarty International Center of the United States National Institutes of Health (award R21TW011361). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Brief Description of Nthabi Adaptation.
[DOCX File , 16 KB - humanfactors_v11i1e52048_app1.docx]

References

1. Kabir MF, Schulman D, Abdullah AS. Promoting relational agent for health behavior change in low and middle - income counntries (LMICs): issues and approaches. J Med Syst 2019;43(7):227 [FREE Full text] [doi: 10.1007/s10916-019-1360-z] [Medline: 31190131]
2. Balaji D, He L, Giani S, Bosse T, Wiers R, de Bruijn GJ. Effectiveness and acceptability of conversational agents for sexual health promotion: a systematic review and meta-analysis. Sex Health 2022;19(5):391-405 [FREE Full text] [doi: 10.1071/SH22016] [Medline: 35863761]
3. Bacchus LJ, Reiss K, Church K, Colombini M, Pearson E, Naved R, et al. Using digital technology for sexual and reproductive health: are programs adequately considering risk? Glob Health Sci Pract 2019;7(4):507-514 [FREE Full text] [doi: 10.9745/GHSP-D-19-00239] [Medline: 31874936]
4. Fernández C, Vicente MA, Carrillo I, Guilabert M, Mira JJ. Factors influencing the smartphone usage behavior of pedestrians: observational study on "Spanish smombies". J Med Internet Res 2020;22(8):e19350 [FREE Full text] [doi: 10.2196/19350] [Medline: 32667896]
5. Pule N. The state of ICT in Lesotho (2016). Commissioned by the Lesotho Communications Authority and the International Telecommunication Union. Research ICT Africa. 2017. URL: https://researchictafrica.net/wp/wp-content/uploads/2018/01/2017_The-State-of-ICT-in-Lesotho_RIA_LCA.pdf [accessed 2023-12-04]
6. Naidoo K, Adeagbo O, Pleaner M. Sexual and reproductive health needs of adolescent girls and young women in sub-Saharan Africa: research, policy, and practice. SAGE Open 2019;9(3):2158244019859951 [FREE Full text] [doi: 10.1177/2158244019859951]
7. Mwamba B, Mayers P, Shea J. Sexual and reproductive health knowledge of postgraduate students at the University of Cape Town, in South Africa. Reprod Health 2022;19(1):225 [FREE Full text] [doi: 10.1186/s12978-022-01507-6] [Medline: 36522681]
8. Rokicki S, Fink G. Assessing the reach and effectiveness of mHealth: evidence from a reproductive health program for adolescent girls in Ghana. BMC Public Health 2017;17(1):969 [FREE Full text] [doi: 10.1186/s12889-017-4939-7] [Medline: 29262823]
9. Lesotho population-based HIV impact assessment 2020. (LePHIA 2020). Lesotho Ministry of Health. 2020. URL: https://phia.icap.columbia.edu/countries/lesotho/ [accessed 2023-12-04]
10. Lesotho’s number of maternal deaths 1990–2015. World Bank. URL: https://www.ceicdata.com/en/lesotho/health-statistics/ls-number-of-maternal-death [accessed 2023-12-04]
11. World Bank. 2018. URL: https://www.worldbank.org/en/topic/health/brief/lesothos-health-sector [accessed 2023-12-04]
12. Bickmore TW, Pfeifer LM, Byron D, Forsythe S, Henault LE, Jack BW, et al. Usability of conversational agents by patients with inadequate health literacy: evidence from two clinical trials. J Health Commun 2010;15(Suppl 2):197-210 [FREE Full text] [doi: 10.1080/10810730.2010.499991] [Medline: 20845204]
13. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied conversational agents in clinical psychology: a scoping review. J Med Internet Res 2017;19(5):e151 [FREE Full text] [doi: 10.2196/jmir.6553] [Medline: 28487267]
14. Abd-Alrazaq AA, Alajlani M, Ali N, Denecke K, Bewick BM, Househ M. Perceptions and opinions of patients about mental health chatbots: scoping review. J Med Internet Res 2021;23(1):e17828 [FREE Full text] [doi: 10.2196/17828] [Medline: 33439133]
15. Jack B, Bickmore T, Hempstead M, Yinusa-Nyahkoon L, Sadikova E, Mitchell S, et al. Reducing preconception risks among African American women with conversational agent technology. J Am Board Fam Med 2015;28(4):441-451 [FREE Full text] [doi: 10.3122/jabfm.2015.04.140327] [Medline: 26152434]
16. Jack BW, Bickmore T, Yinusa-Nyahkoon L, Reichert M, Julce C, Sidduri N, et al. Improving the health of young African American women in the preconception period using health information technology: a randomised controlled trial. Lancet Digit Health 2020;2(9):e475-e485 [FREE Full text] [doi: 10.1016/S2589-7500(20)30189-8] [Medline: 33328115]
17. Nkabane-Nkholongo E, Mokgatle M, Bickmore T, Julce C, Jack BW. Adaptation of the Gabby conversational agent system to improve the sexual and reproductive health of young women in Lesotho. Front Digit Health 2023;5:1224429 [FREE Full text] [doi: 10.3389/fdgth.2023.1224429] [Medline: 37860039]
18. Manzano-Monfort G, Paluzie G, Díaz-Gegúndez M, Chabrera C. Usability of a mobile application for health professionals in home care services: a user-centered approach. Sci Rep 2023;13(1):2607 [FREE Full text] [doi: 10.1038/s41598-023-29640-7] [Medline: 36788261]
19. Hajesmaeel-Gohari S, Khordastan F, Fatehi F, Samzadeh H, Bahaadinbeigy K. The most used questionnaires for evaluating satisfaction, usability, acceptance, and quality outcomes of mobile health. BMC Med Inform Decis Mak 2022;22(1):22 [FREE Full text] [doi: 10.1186/s12911-022-01764-2] [Medline: 35081953]

20. Gardiner P, Hempstead MB, Ring L, Bickmore T, Yinusa-Nyahkoon L, Tran H, et al. Reaching women through health information technology: the Gabby preconception care system. Am J Health Promot 2013;27(3 Suppl):eS11-eS20 [FREE Full text] [doi: 10.4278/ajhp.1200113-QUAN-18] [Medline: 23286652]
21. Walker R, Drakeley S, Julce C, Sidduri N, Bickmore T, Skouteris H, et al. Assessing the potential of a virtual patient advocate to provide preconception care and health advice to women living in Australia. J Dev Orig Health Dis 2021;12(2):300-308 [FREE Full text] [doi: 10.1017/S2040174420000264] [Medline: 32308183]
22. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c332 [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]
23. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
24. Moeini B, Rezapur-Shahkolai F, Jahanfar S, Naghdi A, Karami M, Ezzati-Rastegar K. Utilizing the PEN-3 model to identify socio-cultural factors affecting intimate partner violence against pregnant women in suburban Hamadan. Health Care Women Int 2019;40(11):1212-1228 [FREE Full text] [doi: 10.1080/07399332.2019.1578777] [Medline: 31166150]
25. Arruzza E, Chau M. The effectiveness of cultural competence education in enhancing knowledge acquisition, performance, attitudes, and student satisfaction among undergraduate health science students: a scoping review. J Educ Eval Health Prof 2021;18:3 [FREE Full text] [doi: 10.3352/jeehp.2021.18.3] [Medline: 33621460]
26. Echeverri M, Chen AMH. Educational interventions for culturally competent healthcare: developing a protocol to conduct a systematic review of the rationale, content, teaching methods, and measures of effectiveness. J Best Pract Health Prof Divers 2016;9(1):1160-1177 [FREE Full text] [Medline: 34296221]
27. Martin P, Cousin L, Gottot S, Bourmaud A, de La Rochebrochard E, Alberti C. Participatory interventions for sexual health promotion for adolescents and young adults on the internet: systematic review. J Med Internet Res 2020;22(7):e15378 [FREE Full text] [doi: 10.2196/15378] [Medline: 32735217]
28. Feroz AS, Ali NA, Khoja A, Asad A, Saleem S. Using mobile phones to improve young people sexual and reproductive health in low and middle-income countries: a systematic review to identify barriers, facilitators, and range of mHealth solutions. Reprod Health 2021;18(1):9 [FREE Full text] [doi: 10.1186/s12978-020-01059-7] [Medline: 33453723]
29. Mitchell KJ, Bull S, Kiwanuka J, Ybarra ML. Cell phone usage among adolescents in Uganda: acceptability for relaying health information. Health Educ Res 2011;26(5):770-781 [FREE Full text] [doi: 10.1093/her/cyr022] [Medline: 21536715]
30. Sewak A, Yousef M, Deshpande S, Seydel T, Hashemi N. The effectiveness of digital sexual health interventions for young adults: a systematic literature review (2010-2020). Health Promot Int 2023;38(1):daac104 [FREE Full text] [doi: 10.1093/heapro/daac104] [Medline: 36757346]
31. Kouvari M, Karipidou M, Tsiampalis T, Mamalaki E, Poulimeneas D, Bathrellou E, et al. Digital health interventions for weight management in children and adolescents: systematic review and meta-analysis. J Med Internet Res 2022;24(2):e30675 [FREE Full text] [doi: 10.2196/30675] [Medline: 35156934]
32. Lupton D. Young people's use of digital health technologies in the Global North: narrative review. J Med Internet Res 2021;23(1):e18286 [FREE Full text] [doi: 10.2196/18286] [Medline: 33427684]
33. Dietrich JJ, Coetzee J, Otwombe K, Mdanda S, Nkala B, Makongoza M, et al. Adolescent-friendly technologies as potential adjuncts for health promotion. Health Education 2014;114(4):304-318 [FREE Full text] [doi: 10.1108/he-08-2013-0039]
34. Radovic A, McCarty CA, Katzman K, Richardson LP. Adolescents' perspectives on using technology for health: qualitative study. JMIR Pediatr Parent 2018;1(1):e2 [FREE Full text] [doi: 10.2196/pediatrics.8677] [Medline: 30740590]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials

Edited by A Choudhury; submitted 21.08.23; peer-reviewed by M Brown, J Xie; comments to author 07.11.23; revised version received 08.12.23; accepted 11.12.23; published 12.03.24.

Please cite as:
Nkabane-Nkholongo E, Mpata-Mokgatle M, Jack BW, Julce C, Bickmore T
Usability and Acceptability of a Conversational Agent Health Education App (Nthabi) for Young Women in Lesotho: Quantitative Study
JMIR Hum Factors 2024;11:e52048
URL: https://humanfactors.jmir.org/2024/1/e52048
doi:10.2196/52048
PMID:38470460

Original Paper

Stimulating Preconception Care Uptake by Women With a Vulnerable Health Status Through a Mobile Health App (Pregnant Faster): Pilot Feasibility Study

Sharissa M Smith[1], MD; Babette Bais[1], PhD; Hafez Ismaili M'hamdi[2], PhD; Maartje HN Schermer[2], Prof Dr; Régine PM Steegers-Theunissen[1], Prof Dr, MD

[1]Department of Obstetrics and Gynecology, Erasmus University Medical Center, Rotterdam, Netherlands

[2]Department of Medical Ethics, Philosophy and History of Medicine, Erasmus University Medical Center, Rotterdam, Netherlands

Corresponding Author:
Régine PM Steegers-Theunissen, Prof Dr, MD
Department of Obstetrics and Gynecology
Erasmus University Medical Center
Doctor Molewaterplein 40
Rotterdam, 3015 CD
Netherlands
Phone: 31 10704 ext 0704
Email: r.steegers@erasmusmc.nl

Abstract

Background: A low socioeconomic status is associated with a vulnerable health status (VHS) through the accumulation of health-related risk factors, such as poor lifestyle behaviors (eg, inadequate nutrition, chronic stress, and impaired health literacy). For pregnant women, a VHS translates into a high incidence of adverse pregnancy outcomes and therefore pregnancy-related inequity. We hypothesize that stimulating adequate pregnancy preparation, targeting lifestyle behaviors and preconception care (PCC) uptake, can reduce these inequities and improve the pregnancy outcomes of women with a VHS. A nudge is a behavioral intervention aimed at making healthy choices easier and more attractive and may therefore be a feasible way to stimulate engagement in pregnancy preparation and PCC uptake, especially in women with a VHS. To support adequate pregnancy preparation, we designed a mobile health (mHealth) app, Pregnant Faster, that fits the preferences of women with a VHS and uses nudging to encourage PCC consultation visits and engagement in education on healthy lifestyle behaviors.

Objective: This study aimed to test the feasibility of Pregnant Faster by determining usability and user satisfaction, the number of visited PCC consultations, and the course of practical study conduction.

Methods: Women aged 18-45 years, with low-to-intermediate educational attainment, who were trying to become pregnant within 12 months were included in this open cohort. Recruitment took place through social media, health care professionals, and distribution of flyers and posters from September 2021 until June 2022. Participants used Pregnant Faster daily for 4 weeks, earning coins by reading blogs on pregnancy preparation, filling out a daily questionnaire on healthy lifestyle choices, and registering for a PCC consultation with a midwife. Earned coins could be spent on rewards, such as fruit, mascara, and baby products. Evaluation took place through the mHealth App Usability Questionnaire (MAUQ), an additional interview or questionnaire, and assessment of overall study conduction.

Results: Due to limited inclusions, the inclusion criterion “living in a deprived neighborhood” was dropped. This resulted in the inclusion of 47 women, of whom 39 (83%) completed the intervention. In total, 16 (41%) of 39 participants visited a PCC consultation, with their main motivation being obtaining personalized information. The majority of participants agreed with 16 (88.9%) of 18 statements of the MAUQ, indicating high user satisfaction. The mean rating was 7.7 (SD 1.0) out of 10. Points of improvement included recruitment of the target group, simplification of the log-in system, and automation of manual tasks.

Conclusions: Nudging women through Pregnant Faster to stimulate pregnancy preparation and PCC uptake has proven feasible, but the inclusion criteria must be revised. A substantial number of PCC consultations were conducted, and this study will therefore be continued with an open cohort of 400 women, aiming to establish the (cost-)effectiveness of an updated version, named Pregnant Faster 2.

International Registered Report Identifier (IRRID): RR2-10.2196/45293

(JMIR Hum Factors 2024;11:e53614) doi:10.2196/53614

KEYWORDS

preconception care; mHealth; mobile health; pregnancy preparation; nudge; health inequality; socioeconomic status; lifestyle; women; pregnancy; pregnant women; pregnant; socioeconomic; pilot feasibility study; mHealth app; mHealth application; app; application; risk factor; nutrition; stress; chronic stress; health literacy; usability; user satisfaction; user; users

Introduction

A low socioeconomic status (SES) is associated with a vulnerable health status (VHS), which research suggests is grounded in the accumulation of risk factors, such as inadequate nutrition, smoking, and increased mental stressors [1-4]. For women, a low SES means they are more likely to have a VHS, which translates into a higher incidence of adverse pregnancy outcomes in this group [5-9]. These adverse outcomes originate at least partly in the periconception period [10], during which gametogenesis, embryonic development, and placentation take place, laying the foundation for perinatal outcomes, as well as the child's lifelong health [11,12]. For example, an accumulation of 2 or more maternal risk factors impacts embryonic growth [13], which is associated with midpregnancy fetal weight and birth weight [14]. In addition, infants born small for their gestational age are more susceptible to noncommunicable diseases, such as diabetes mellitus and cardiovascular disease [15]. The effects of adverse pregnancy outcomes therefore hit twice: once in utero and once in later life. This increases the child's chance of a VHS in adulthood, which, once again, may influence pregnancy outcomes. These transgenerational effects are further maintained by impaired health agency, which is associated with a low SES and diminishes the likelihood of seeking necessary care [16]. In accordance with these findings, research shows that women with a VHS are less likely to engage in pregnancy preparation and take up preconception care (PCC) [17].

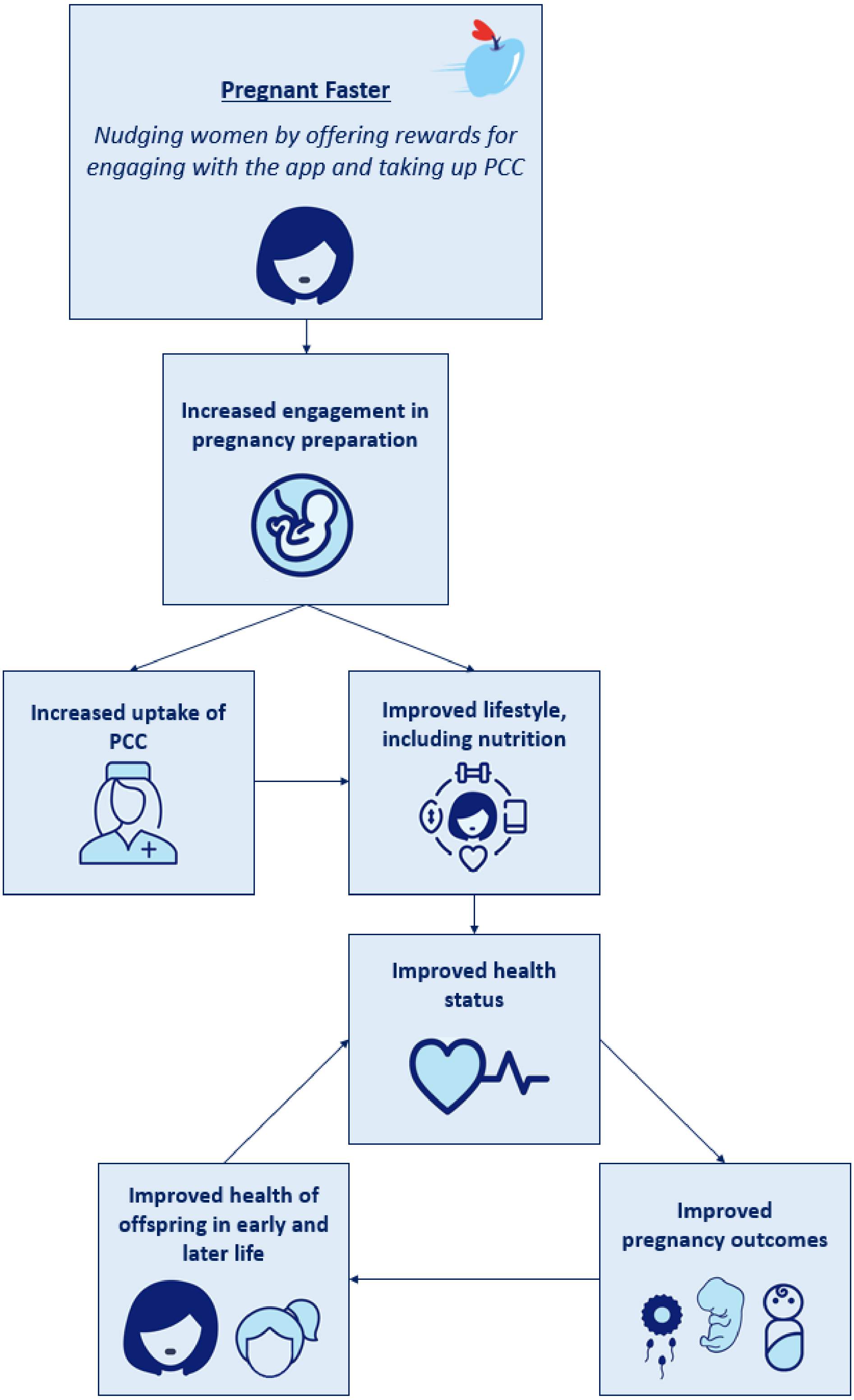
PCC is usually given by a midwife or an obstetrician and is aimed at identification of possible risk factors for adverse outcomes, ameliorating those that are modifiable prior to pregnancy [18]. This includes adopting healthy lifestyle behaviors and making beneficial choices in general that will increase the chance of having a healthy pregnancy and baby. Although ≥80% of women who wish to become pregnant have at least 1 modifiable risk factor for adverse pregnancy outcomes [19,20], the uptake of PCC remains low due to insufficient awareness of risk factors and the benefits of PCC [21]. Women with a low SES may encounter additional barriers when engaging in pregnancy preparation, as they are already burdened by the deprived circumstances in which they live. Supporting this group by making pregnancy preparation easier and attractive might be a suitable way to relieve the inequity regarding their pregnancy outcomes.

Our research group has previously developed the web-based PCC tool Smarter Pregnancy, an interactive, tailored, mobile health (mHealth) platform that offers practical coaching and customized feedback on nutrition and other lifestyle behaviors of prospective parents [22]. Smarter Pregnancy has proven to be effective in supporting healthy choices in women with a VHS, in addition to being valued highly by them [23]. To further support PCC engagement in women with a VHS, we have designed an mHealth app that especially fits their needs and preferences [24]: the app-based nudge Pregnant Faster. A nudge is an intervention that stimulates making beneficial choices by increasing the attractiveness and easiness of healthy behavior [25]. An in-depth explanation of nudge theory and its application in health policy can be found in the study by Murayama et al [26].

In the case of Pregnant Faster, participants are nudged through a loyalty program that entails collecting coins by engaging with the app and ordering rewards using those coins. The design of Pregnant Faster can be viewed as a macrolevel nudge, containing multiple microlevel nudges aimed at stimulating pregnancy preparation and encouraging the uptake of PCC. For example, the monetary value of a coin is a microlevel nudge; it varies from €0.06 to €0.26 (US $0.06-$0.26), depending on the type of reward. Healthy rewards, such as folic acid supplements, are relatively cheap, steering participants toward picking them over luxury goods, while maintaining their freedom to choose. The most important feature of the app, which also yields the highest number of coins, is the possibility to register for a PCC consultation with a nearby midwife, promoting blended care: an effective way to promote pregnancy preparation [27]. As midwives are the primary health care providers for pregnant women in the Netherlands, PCC consultations are beneficial for the bond between health care provider and client prior to and during pregnancy. The full description of Pregnant Faster's design process, detailing the imbedded nudges, has been published in *JMIR Protocols* [28].

The aim of this pilot study was to determine Pregnant Faster's feasibility pertaining to usability and user satisfaction, the number of PCC consultations booked and visited by participants, and the course of practical conduction regarding the inclusion process, reward allocation, and finalization of the study. In addition, the results of this study will be used to further develop Pregnant Faster and lay the foundation for a larger cohort study to establish its (cost-)effectiveness. Our overall ambition is that Pregnant Faster contributes to the improvement of short-term and long-term health in mothers with a VHS, their children, and future generations (Figure 1).

Figure 1. Aim of the mHealth app Pregnant Faster. mHealth: mobile health; PCC: preconception care.



Methods

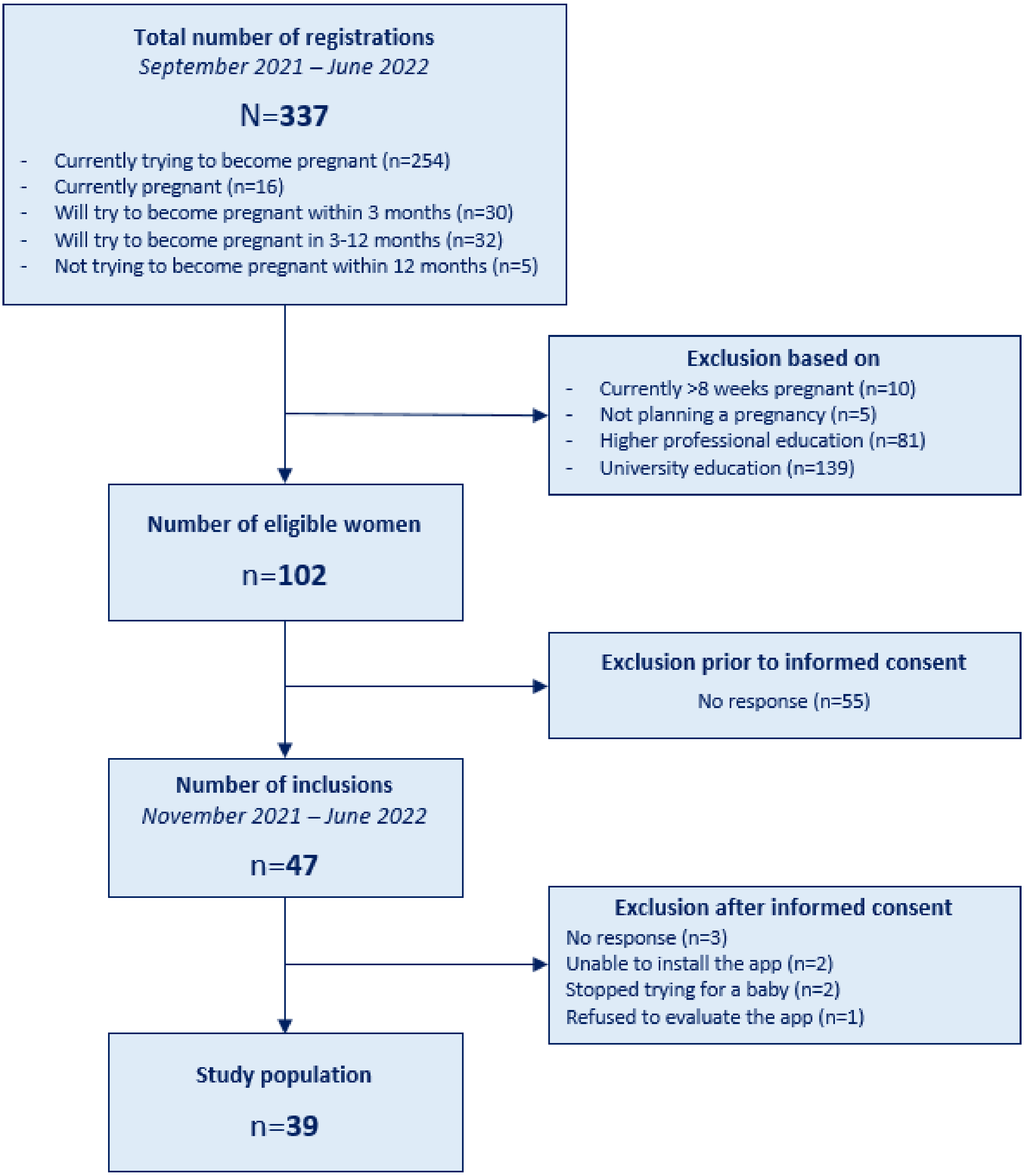
Ethical Considerations

This study was assessed and approved by the Medical Ethical Committee of the Erasmus University Medical Center, Rotterdam, The Netherlands (MEC-2020-0974). Informed consent was obtained from all participants via email. Considering the low risk of this study, composing a Data Safety Monitoring Board was deemed unnecessary.

Recruitment and Inclusion

Our aim was to include 40 participants in this study. Between September 2021 and June 2022, 337 women registered for this study, of which 102 (30.3%) were eligible for inclusion. Due to a higher-than-expected confidentiality, integrity, and availability (CIA) Triad classification (a risk score regarding user information safety [29]), additional security demands were necessary. Fulfilling these demands delayed the launch of Pregnant Faster from September 2021 to November 2021, leading to a loss of recruited eligible women. From the launch onward, 47 (46.1%) women were included, of which 39 (83%) completed the intervention (Figure 2).

Figure 2. Inclusion flowchart.



Recruitment took place through the Sneller Zwanger website [30], which was distributed through posters and flyers, advertisements on the social media platforms Facebook and Instagram, and midwifery practices that provide primary care to all pregnant women in the Netherlands. Additionally, a collaboration took place with the Dutch influencer Midwife Mother (Dutch *Verlosmoeder*) on Instagram [31-33]. Participants filled in a survey with their first name, age, telephone number, email address, zip code, educational level, and when they planned on trying to become pregnant (currently pregnant, currently trying, or trying in ≤3, >3-12, or >12 months). Pregnant women were asked to fill in their estimated or calculated due date.

Participants were selected based on the following inclusion criteria: assigned female at birth, 18-45 years old, actively trying to become pregnant within now and 12 months or pregnant with a gestational age of <8 weeks at the start of the intervention, a low-to-intermediate educational level (prevocational or vocational education), and able and willing to download and evaluate the app.

Exclusion criteria were as follows: insufficient proficiency in the Dutch language, not in possession of a smartphone or tablet suitable for the app, and refusal to download or evaluate the app. All excluded women received a free coupon for the Dutch or English version of Smarter Pregnancy [34].

Design

Pregnant Faster was developed by the Erasmus University Medical Center's research group Periconception Epidemiology at the Department of Obstetrics and Gynecology, in collaboration with TJIP The Platform Engineer and the event bureau Improve. A detailed description of the cocreation and design process of Pregnant Faster and the study protocol has been published in *JMIR Protocols* [28].

Intervention

Eligible women were sent the patient information folder, in which the intervention was explained. Inclusion was finalized after a telephone conversation in which further clarification could be provided. An email was sent with instructions on how to download and install Pregnant Faster from Apple App Store (iOS) or via a link (Android). If more than 3 days passed between inclusion and downloading, participants were approached twice by email and telephone and once by a text message to provide further support with installation.

The first log-in marked the start of the 4-week intervention. During this period, participants logged in with their email address and a password, which yielded 1 coin per day per log-in. The first log-in yielded 50 coins as a reward for installation and to immediately stimulate participants to further engage with the app. After log-in, a dashboard appeared, containing 5 buttons: (1) "Earn coins," (2) "Overview coins," (3) "See a midwife!," (4) "This study," and (5) "Rewards" (Figure 3).

Button 1, "Earn coins," led to a timeline where new blogs and tips appeared daily (Multimedia Appendix 1). Reading this information yielded 4-8 coins. In the same timeline, a daily questionnaire appeared in which participants could tick a box if they ate sufficient fruit and vegetables, exercised, and took folic acid supplements that day. Each ticked box yielded 2 coins per day. Button 2, "Overview coins," displayed when and how coins were earned and how many coins were spent on which products. Button 3, "See a midwife!," contained information regarding what PCC is and who it is for, stimulating participants to register through the app for a PCC consultation. Registering consisted of filling in their phone number, which immediately yielded 25 coins. An additional 75 coins were allocated after the visit was confirmed by the midwife. Button 4, "This study," contained information about the study itself and contact details for support. Button 5, "Rewards," contained an in-app shop where participants could order rewards, including (but not limited to) folic acid supplements, fruit, nail polish, mascara, ovulation or pregnancy tests, and newborn clothing. Rewards were sent to their home address to arrive within 5 business days. If a participant had not logged in for 7 days, they received a manually sent text message and email, encouraging them to read up on the newly offered blogs and tips, earn more coins, and order rewards.

At the end of the intervention, participants were offered a coupon for Smarter Pregnancy via the timeline, which would yield 25 coins upon use and provide them with an additional 26 weeks of coaching. Furthermore, they received an email regarding finalization of the study and available means of support. Earned coins could be spent up to 2 weeks after the intervention ended. The blogs and tips remained accessible for as long as the app remained installed. Figure 4 provides an overview of the study flow.

Figure 3. Pregnant Faster interface.

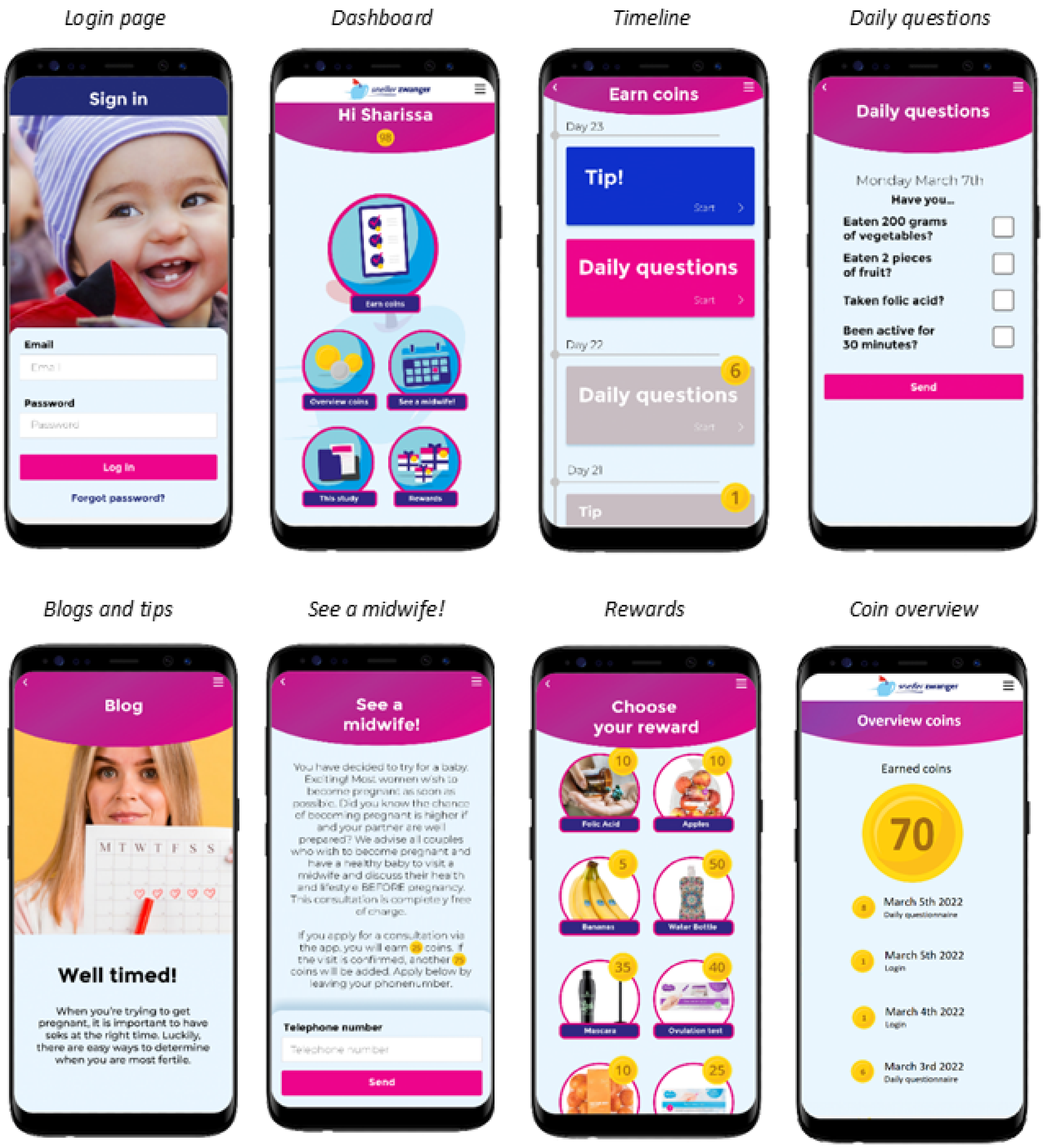
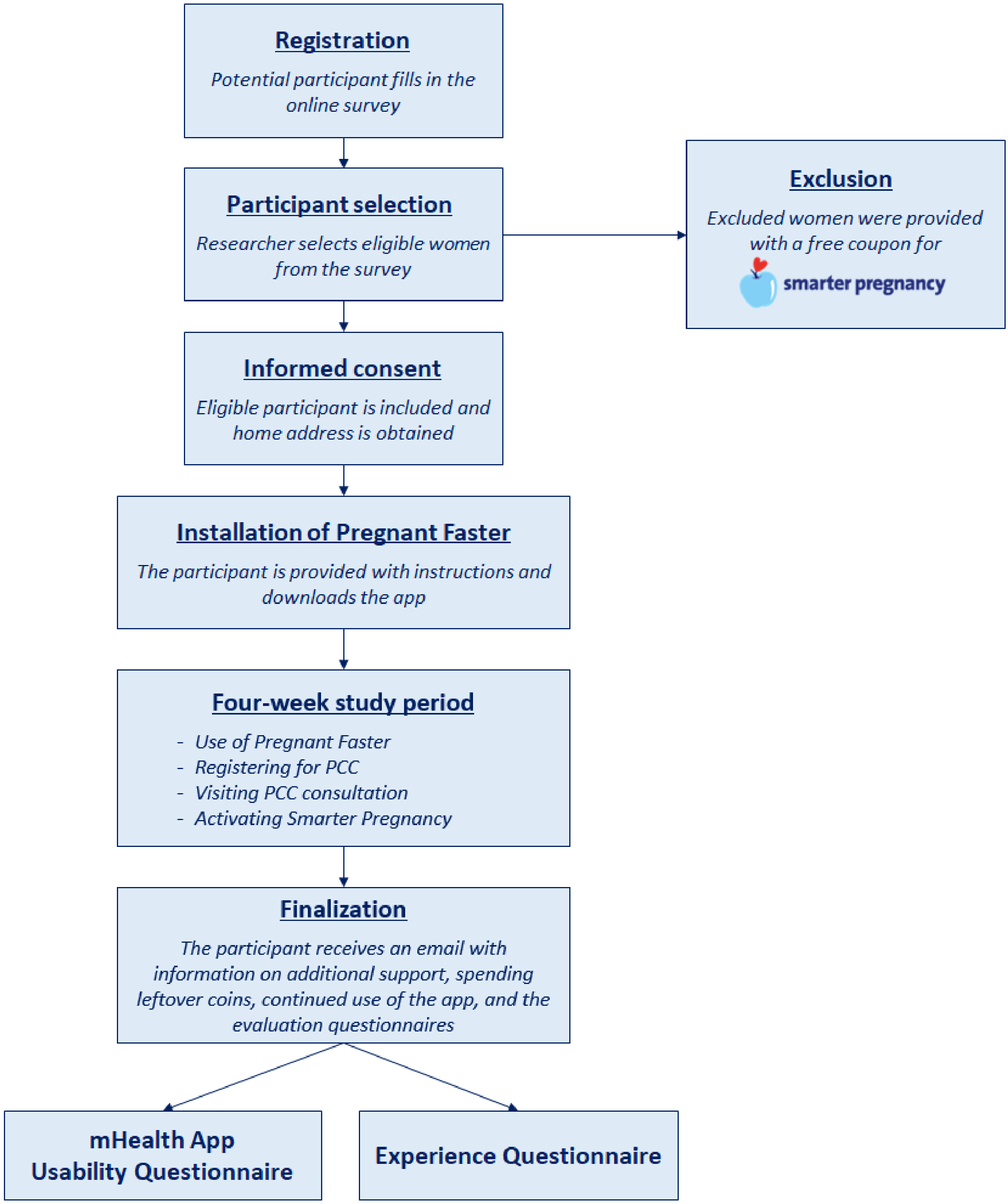


Figure 4. Flowchart of the study design. mHealth: mobile health; PCC: preconception care.



Data Collection and Outcome Measures

During the study, registration and inclusion rates were tracked and a log book was kept to note any encountered barriers and changes to the protocol.

Participants' baseline characteristics were collected prior to inclusion through the selection survey, and home addresses were collected via email after obtaining informed consent. If selected participants did not respond to attempts to include them, they were contacted twice by email, twice by telephone, and once by a text message.

After using Pregnant Faster for 4 weeks, participants filled in a modified version of the 18-item mHealth App Usability Questionnaire (MAUQ), which uses a 7-point Likert scale (Multimedia Appendix 2) [35]. In addition, the first 10 (25.6%) participants went through a semistructured interview (Multimedia Appendix 3) in which they elaborated on their experiences with the app and, if applicable, the PCC consultation. The audiotapes of these interviews were used to compose the Experience Questionnaire (ExQ; Multimedia Appendix 4) that consisted partly of questions using a 5-point Likert scale. The ExQ was offered to the remaining participants, and the first author filled in the ExQ for the first 10 (25.6%) participants, using the audiotaped data provided during the interviews. If a question in the ExQ was not clearly answered in the interview, the participant was approached by telephone to provide an answer. The answers to the open questions in the ExQ were evaluated for notable and recurring comments.

At the end of the study period, data were collected on the number of coins earned, the types of rewards that were chosen, and the number of booked and visited PCC consultations.

Data Analysis

To evaluate the inclusion strategy, the following percentages were calculated: (1) eligibility percentage, (2) inclusion percentages, and (3) intervention completion percentages. The eligibility percentage was determined by dividing the number of women eligible for inclusion by the total number of women who registered for the study. The inclusion percentages were calculated by dividing the number of included participants (who provided informed consent) by the number of total registrations and the number of eligible women who were approached for inclusion. Completion of the intervention entailed completing the evaluation of the app. The intervention completion percentages were obtained by dividing the number of participants who completed the intervention by the number of eligible women and the number of participants who provided informed consent.

The baseline of the study population is presented in tabular form using the median (IQR) for continuous data and n (%) for categorical data. Data obtained through the MAUQ and the ExQ are presented in bar charts. Notable answers to open questions are presented in narrative form. Data on feasibility from the researchers' point of view are presented as bullet points.

All calculations were carried out using SPSS Statistics 25 (IBM Corporation), charts were created using Excel 2016 (Microsoft), and figures were created in PowerPoint 2016 (Microsoft).

Results

Recruitment, Inclusion, and the Study Population

A total of 337 women registered for the intervention, of whom 102 (30.3%) were eligible for inclusion. Informed consent was signed by 47 (46.1%) women, and 39 (83%) of the 47 women participated and completed the intervention. Table 1 displays participants' baseline characteristics.

Table 1. Participants' baseline characteristics.

Characteristics	Participants (N=39)
Age (years), median (IQR)	30 (27-35)
Mean income neighborhood[a], n (%)	
Below middle	19 (48.7)
Above middle	11 (28.2)
Low to high	4 (10.3)
High	5 (12.8)
Educational level[b], n (%)	
Low	3 (7.7)
Intermediate	36 (92.3)
Trying to become pregnant, n (%)	
Currently trying	32 (82.1)
Within 3 months	5 (12.8)
Within 12 months	2 (5.1)
Mobile operating system, n (%)	
Android	17 (43.6)
iOS	22 (56.4)

[a]The median household income of a neighborhood is determined by the distribution of household income of all households in the country [28]. This table adheres to the original subdivision of the distribution of household income (year 2020): low, <€15,900 (<US $18,800); below middle, €15,900-21,000 (US $18,800-$24,800); middle, €21,000-26,800 (US $24,800-$31,700); above middle, €26,800-34,600 (US $31,700-$40,900); and high, >€34,600 (>US $40,900).

[b]Educational level [29]. Dutch educational levels are subdivided as follows: low (prevocational education, selective secondary education, or lower), intermediate (vocational education), and high (bachelor's degree, master's degree, or higher).

A 2-month gap arose between the start of recruitment and the intervention, due to the app's CIA Triad classification [29]. A low classification was expected, but the combination of 40 intended participants and their registering their first name and email address in the app warranted a slightly higher classification for confidentiality and therefore additional security demands. Despite frequent updates to keep eligible women engaged, 55 (53.9%) of 102 women did not respond when inclusion commenced. Next, we describe these events and the inclusion process in detail.

Between September 2021 and January 2022, 212 (62.9%) women from the total 337 registrations reached in June 2022 registered for the study. Of these 212 women, only 9 (4.2%) were included. To boost the registration and inclusion rates, more flyers and posters were distributed, and the choice was made to include women who were trying to become pregnant within 12 months as opposed to within 3 months, as originally intended. Furthermore, the intervention was expanded from the municipality of Rotterdam to nationwide, delivering rewards through the postal service instead of by car. Midwives throughout the Netherlands were actively approached to ask whether they were interested in participating in the study and were offered support in setting up PCC consultations in their practices. Subsequently, the collaboration with Midwife Mother was renewed, who uploaded another post and multiple stories regarding PCC and Pregnant Faster to her Instagram. All women who were previously excluded based on not living near Rotterdam were contacted and asked to participate.

These efforts showed a limited effect. By May 2022, an additional 54 (16%) registrations and a total of 16 (34%) inclusions were obtained, which led to the decision to drop the inclusion criterion of living in a deprived neighborhood, thereby lowering the chance of including women likely to have a VHS. This choice was made to allow for further development and testing of Pregnant Faster while searching for a more effective way to recruit the intended target group for the planned larger cohort. Another social media campaign was conducted, and all women who were previously excluded based on their neighborhood's median income were invited to participate.

Between May 2022 and July 2022, 71 (21%) more women registered for the study, adding up to the total of 337 registrations. Registration was closed after no new women registered for 2 weeks. From May onward, 31 participants were included, adding up to a total of 47 inclusions, of which 39 (83%) completed the intervention and 8 (17%) dropped out. Of these 8, 3 (37.5%) women who provided informed consent did not respond to instructions on how to install the app or attempts to reach them; 2 (25%) women were unable to install the app: in one case, an iPhone with Belgian settings preventing download from the Dutch Apple Store, and in the other case, the woman who had an Android device was scared by the warning prompted by download of an app outside of Google Play Store. Of the 3 (37.5%) remaining dropouts, 2 (25%) stopped trying to become pregnant and 1 (12.5%) found the questionnaires too burdensome. Table S1 in Multimedia Appendix 5 provides an overview of the eligibility, inclusion, and intervention completion percentages.

PCC Consultations

A total of 17 (43.6%) of 39 participants registered for a PCC consultation, and 16 (41%) consultations were conducted by 9 midwifery practices. One consultation was performed via telehealth by the first author because the midwife chosen by the participant had no experience in providing PCC and did not wish to implement PCC in her practice. The participant who registered for PCC but did not attend a consultation was worried about her health care insurance not covering the costs. Stating to feel overwhelmed, she declined additional support as well as a free telehealth consultation.

The most often reported reason to register for a PCC consultation was to obtain more personalized information (14/17, 82.4%), followed by being curious about what a consultation entails in practice (6/17, 35.3%). The most frequent reason not to register for a consultation was simply not being interested in doing so (6/22, 27.3%). A visual overview of participants' motivation regarding registration for PCC can be found in Figures S7 and S8 in Multimedia Appendix 6.

All participants who visited a PCC consultation agreed that registering through Pregnant Faster is easy (2/16, 12.5%, agree; 14/16, 87.5%, strongly agree) and were glad they had done so (3/16, 18.7%, agree; 13/16, 81.3%, strongly agree).

Coins and Rewards

During the study, participants could earn a maximum amount of 468 coins. Together, they earned a total of 11.791 coins (mean 284, SD 109 per participant; median 276, IQR 221-358; range 79-443). In total, 344 rewards were ordered during the study period (mean 8, SD 6 per participant; median 7, IQR 3-12; range 0-22). One participant did not wish to order rewards because she was happy with "just the app." She stated that she did not feel it was morally objectionable to be rewarded but just that she was not interested in receiving rewards.

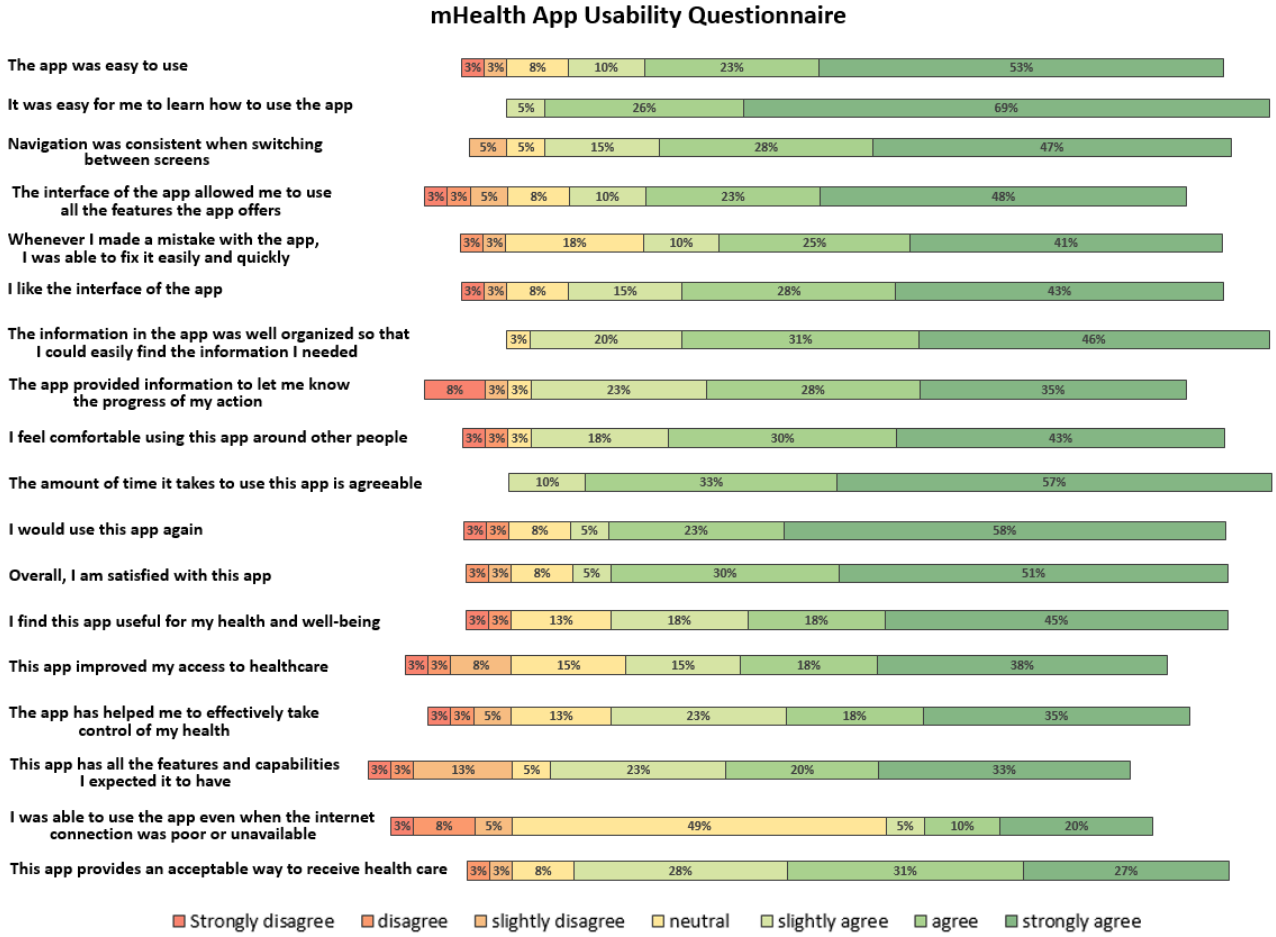
The most popular reward was a €10 (US $10) book voucher, with 19 (48.7%) of 39 participants ordering the voucher at least once. The second-most popular reward was fruit, with 17 (43.6%) participants ordering fruit at least once and a total of 87 orders (87/344, 25.3%). Bananas were the most popular fruit, amounting to 32 (36.8%) of 87 fruit orders. The third-most popular reward was a set of 2 home pregnancy tests, with 16 (41%) participants ordering this reward at least once. Most participants ordered a reward more than once, displaying their personal preferences and satisfaction regarding their previous order. Table S2 in Multimedia Appendix 5 displays all rewards and their order frequency and percentage.

Feasibility From Users' Point of View

mHealth App Usability Questionnaire

Figure 5 displays the results of the MAUQ. The participants deemed Pregnant Faster's usability satisfying, with the majority of participants agreeing with 16 (88.9%) of 18 statements. With regard to the remaining 2 statements, all participants (N=39, 100%) agreed that the amount of time the app takes is agreeable, and 19 (48.7%) were neutral about being able to use the app with a poor internet connection, indicating they may not have experienced connectivity problems.

Figure 5. Results of the MAUQ. MAUQ: mHealth App Usability Questionnaire; mHealth: mobile health.



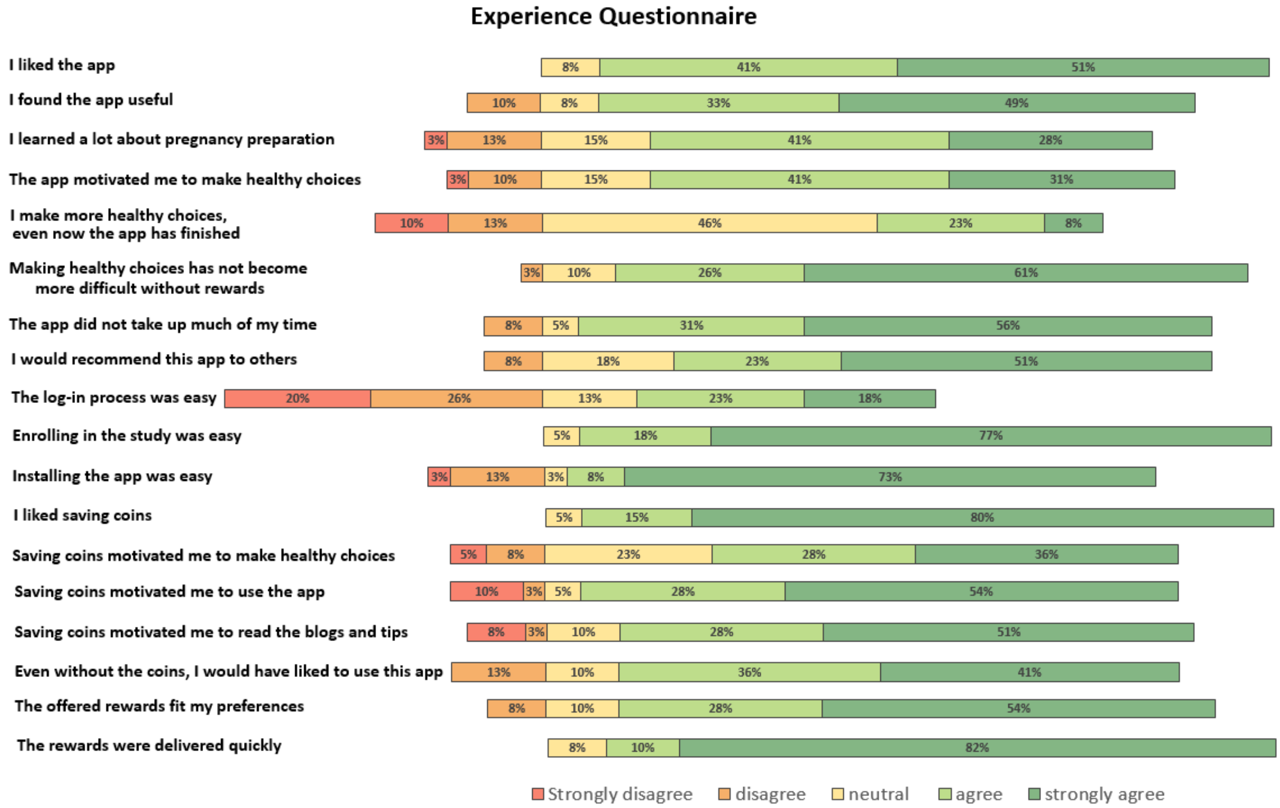
Experience Questionnaire

Figure 6 displays the results of the ExQ for which the 5-point Likert scale was used. The majority of participants agreed with 16 (88.9%) of 18 statements, conveying high user satisfaction. Regarding the log-in process, 16 (41%) of 39 participants agreed and 18 (46.2%) disagreed that it is easy, with 5 (12.8%) being neutral. In addition, 12 (30.8%) participants agreed and 9 (23.1%) disagreed with the statement regarding participants making more healthy choices after finishing the intervention. The remaining 18 (46.2%) participants were neutral.

The majority of participants stated that they used the app daily (n=18, 46.2%) or every other day (n=18, 46.2%). Most participants (n=31, 79.5%) reported to have logged in less often than they would have wanted to, the foremost reason being the requirement to log in with an email and password each time (n=14, 35.9%). Overall, participants rated Pregnant Faster 7.7 out of 10, with 10 being the best rating (mean 7.7, SD 1.0; median 8, IQR 7-8; range 5-9). Multimedia Appendix 6 contains Figures S9-S11, which provide additional results for the ExQ multiple-choice questions.

In the ExQ, participants were given the option to provide additional comments. It was notable that 8 (20.5%) participants commented that they would have liked push notifications to remind them of filling in the daily questionnaire and reading new blogs and tips. One participant recommended personalized notification settings so they would best fit her wishes regarding the subject, timing, and frequency. Furthermore, 3 (7.7%) participants commented that they would like the app to focus on their partners as well, hoping to actively involve them more in preparing for pregnancy.

Figure 6. Results of the ExQ. ExQ: Experience Questionnaire.



Feasibility From Researchers' Point of View

During the study, 11 issues were noted that should be considered before attempting to establish Pregnant Faster's (cost-)effectiveness in a larger cohort study:

- A time gap between the start of recruitment and intervention led to a significant loss of eligible participants and should be prevented.
- Limiting the study population to a local area greatly impedes inclusion rates and causes disappointment in otherwise eligible participants, which may harm the intervention's reputation.
- Using a combination of the neighborhood median income and a low-to-intermediate educational level as a proxy of low SES is not a suitable method to recruit large numbers of women with a VHS.
- Manual selection and inclusion require a significant amount of labor for which multiple researchers have to be available. The same goes for approaching individual midwifery practices for collaboration.
- Use of a classic, relatively complicated information folder and informed consent form can be overwhelming and does not lead to proper understanding of the study nor true consent.
- For participants with Android devices, installation of the app is complicated by not providing the app through the Google Play Store. Additional support is often needed.
- Fruit sent through the postal service often arrive bruised, requiring frequent checks for whether the reward is delivered in good condition, offering refunds when this is not the case. Failed deliveries result in the return of rotten fruit to the researchers. Since fruit is a popular, healthy reward, a suitable alternative should be considered, such as a voucher.
- Sending rewards daily is laborious. During this pilot, we experimented with a frequency of twice per week, clearly communicating this to participants. Afterward, no dissatisfaction regarding delivery time was noticed.
- Confirmation of PCC consultations requires the researchers to contact midwifery practices, causing a delay in coin allocation, possibly negatively impacting user satisfaction and effectiveness of the reward. Relying on participants self-reporting their visit in the app, combined with automatic coin allocation, should be considered.
- Manually keeping track of booked and confirmed consultations, in addition to manual coin allocation, requires a significant amount of time. For this reason as well, self-reporting should be considered.
- Asking participants to fill in 2 separate questionnaires causes confusion and diminishes the likeliness of completing the evaluation. It is advisable to evaluate user experiences in a succinct manner.

Discussion

Principal Findings

In this pilot study, we aimed to determine the feasibility of the app-based nudge Pregnant Faster, which is designed to fit the needs of women with a low SES and a high likelihood of having a VHS, who have a higher risk of adverse pregnancy outcomes. The aim of Pregnant Faster is to encourage these women by nudging them to adequately prepare for pregnancy through education by making healthy lifestyle choices and engaging in PCC, which will help improve their pregnancy outcomes.

Pregnant Faster has shown to be feasible from the users' point of view, showing high user satisfaction with a rating of 7.7/10 and PCC uptake by 16 (41%) of 39 participants. Notably, 27 (69.2%) participants stated to have learned a lot about pregnancy preparation and 28 (71.7%) felt motivated by the app to make healthy lifestyle choices. After the intervention ended, 12 (30.8%) participants stated that they more often make healthy choices than prior to using Pregnant Faster.

With regard to the 55 (53.9%) of 102 eligible participants who did not respond when inclusion commenced after the 2-month delay, we suspect that a loss of interest and perhaps of trust in the intervention played a role. The amount of lost eligible women suggests that time is a limiting factor, impacting women's willingness to participate in the intervention. This emphasizes the necessity of quickly responding to their willingness to participate and acceptance of offered care.

Feasibility from the researchers' point of view was satisfactory as well but only with regard to practical conduction, as adjustments to the inclusion criteria were made to up the number of inclusions. Dropping the criterion of living in a deprived neighborhood likely impacted the chance of including women who actually have a VHS. The feasibility of Pregnant Faster from the researchers' point of view can be improved by developing a new method of finding the target group, making the app available via both Apple App Store and Google Play Store, and automating (parts of) the inclusion process and coin allocation, which will limit the number of administrative tasks.

Strengths and Limitations

To the best of our knowledge, Pregnant Faster is the first mHealth intervention that aims to encourage adequate pregnancy preparation and increase the uptake of PCC, promoting blended lifestyle care, by nudging participants with a loyalty program consisting of earning and spending coins. During study conduction and after evaluation, important knowledge was gained concerning the strengths and limitations of this intervention and how best to proceed with a larger cohort study.

Despite our earlier experiences regarding the recruitment of women who likely have a VHS, we did not manage to conduct this study adhering to the original inclusion criteria [28]. It is possible, therefore, that the user feasibility would have been different had the full study population met the intended criteria.

Pregnant Faster has been designed through iterative cocreation, actively involving the target population in its development [28]. Even though adjustments were made, we consider this pilot study to be another step in the iterative cocreation process, as the results will be used for further development of the app and nearly half (19/39, 48.7%) of the study population met the criteria of living in a deprived neighborhood.

Further Development and Future Research

The insights gained through this study have prompted us to re-evaluate which characteristics to use as a proxy for a low SES and the associated VHS. To improve recruitment of the target group, we have hosted meetings with health care professionals specializing in health-related vulnerability and adverse pregnancy outcomes to gain more insight and develop new inclusion criteria for a larger cohort study. At this moment, we are researching (combinations of) different inclusion criteria based on self-reported vulnerability markers, such as high stress; financial insecurity; addiction to alcohol, drugs, and tobacco; and lack of social support, which are also associated with unfavorable health outcomes [36]. Through developing these new criteria, we aim to be more inclusive and provide support to all women with a certain degree of health-related vulnerability, instead of limiting support to those with a high likelihood of having a VHS based on the educational level and neighborhood deprivation.

We aim to continue promoting Pregnant Faster on social media platforms, such as Instagram and YouTube. These platforms have been known to use algorithms that successfully reach target audiences and prove to be effective tools with regard to providing people with support and education [37,38]. Using these platforms, therefore, will not only support recruitment and benefit the target population but also allow Pregnant Faster to contribute to pregnancy-related health in the general population.

The knowledge gained through this pilot study has inspired us to research different methods of information transfer to ensure the app fits the needs of the target group and improve Pregnant Faster's accessibility for those who experience limited literacy [39]. For the planned cohort study, for example, we have created an audio version and infographic of the patient information folder and informed consent form. Furthermore, we are currently creating additional content for the app, again focusing on multiple methods of information transfer, such as podcasts, videos, and infographics.

On a technical level, the inclusion process and content management system will be adjusted to reduce manual tasks and promote feasibility. Furthermore, we plan to change the log-in procedure to a pin code or fingerprint and enhance the app with daily notifications.

Regarding focusing more on participants' partners, we have chosen to not adhere to this suggestion at the current time, as the tips and blogs already contain information for partners and we are still in the process of establishing (cost-)effectiveness and further developing Pregnant Faster. In research concerning reproductive health, it is known that partners may sometimes take on a more passive role [40], which places the burden of preparing for pregnancy largely on the person who will carry the baby. For future development, therefore, we will consider the possibility of adding personalized settings to allow users to fill in characteristics that will adjust the app's content accordingly, such as relationship status, gender and sexual orientation, and, if applicable, gestational age and the use of donor semen.

In the future, we wish to investigate the possibility of offering Pregnant Faster to all who wish to become pregnant, possibly with rewards if cost-effectiveness is established.

Conclusion

With this pilot study, we have demonstrated that the app-based nudge Pregnant Faster provides a feasible way to stimulate the uptake of PCC and boost participants' motivation to adequately prepare for pregnancy. We will use the knowledge we have gained through this pilot study to create an updated version of the app, which will be named Pregnant Faster 2. Our next step consists of determining the (cost-)effectiveness of Pregnant Faster 2, for which we will conduct a cohort study of 400 women with a VHS based on newly devised inclusion criteria.

Acknowledgments

This research was funded by ZonMw (ZorgOnderzoek Nederland en het gebied van Medische wetenschappen, health care research in the Netherlands and the field of medical sciences; grant number 543003103). We would like to thank all women who registered for this study and especially those who participated in it. Additionally, we would like to thank all midwifery practices who performed or were willing to perform PCC consultations for this study; Anneroos van der Zande, who helped continue the inclusion process and app development during the first author's maternity leave; and Maaike Stift, owner of EAT.PEASY!, who sponsored this intervention by providing vegan meals as a reward. Lastly, we thank Djanifa da Conceicao, midwife, influencer, and owner of Verlosmoeder, who contributed greatly to our inclusion process. The table-of-content image was provided by Keira Burton via Pexels.

Data Availability

The data sets used in this study are available from the corresponding author upon reasonable request.

Authors' Contributions

RST conceived the study and gained funding. RST, SS, and BB initiated the study design. SS wrote the original draft, and BB, HIM, MS, and RST reviewed and edited the original draft. BB, RST, HIM, and MS supervised the study. All authors have contributed to refinement of the study protocol and approved the final manuscript.

Conflicts of Interest

RST is initiator and developer of the mobile health (mHealth) app Smarter Pregnancy and the ZonMw project "A Loyalty Program to Motivate Vulnerable Women to Engage in Preconception Care: From Voucher to Tablet" (reference number 543003103). Other authors declare that they have no competing interests. The funder had no role in the design of the study, the writing of the manuscript, or the decision to publish the protocol.

Multimedia Appendix 1
Example of a Pregnant Faster blog.
[DOCX File , 68 KB - humanfactors_v11i1e53614_app1.docx]

Multimedia Appendix 2
The mHealth App Usability Questionnaire, adjusted for Pregnant Faster. mHealth: mobile health.
[DOCX File , 26 KB - humanfactors_v11i1e53614_app2.docx]

Multimedia Appendix 3
The Experience interview topic list.
[DOCX File , 26 KB - humanfactors_v11i1e53614_app3.docx]

Multimedia Appendix 4
The Experience Questionnaire.
[DOCX File , 27 KB - humanfactors_v11i1e53614_app4.docx]

Multimedia Appendix 5
Additional tables.
[DOCX File , 30 KB - humanfactors_v11i1e53614_app5.docx]

Multimedia Appendix 6
Additional figures.
[DOCX File , 65 KB - humanfactors_v11i1e53614_app6.docx]

References

1. Timmermans S, Bonsel GJ, Steegers-Theunissen RPM, Mackenbach JP, Steyerberg EW, Raat H, et al. Individual accumulation of heterogeneous risks explains perinatal inequalities within deprived neighbourhoods. Eur J Epidemiol 2011 Feb;26(2):165-180 [FREE Full text] [doi: 10.1007/s10654-010-9542-5] [Medline: 21203801]
2. Hulshof KFAM, Brussaard JH, Kruizinga AG, Telman J, Löwik MRH. Socio-economic status, dietary intake and 10 y trends: the Dutch National Food Consumption Survey. Eur J Clin Nutr 2003 Jan 27;57(1):128-137. [doi: 10.1038/sj.ejcn.1601503] [Medline: 12548307]
3. Beenackers MA, Kamphuis CB, Giskes K, Brug J, Kunst AE, Burdorf A, et al. Socioeconomic inequalities in occupational, leisure-time, and transport related physical activity among European adults: a systematic review. Int J Behav Nutr Phys Act 2012 Sep 19;9(1):116 [FREE Full text] [doi: 10.1186/1479-5868-9-116] [Medline: 22992350]
4. Steptoe A, Feldman PJ. Neighborhood problems as sources of chronic stress: development of a measure of neighborhood problems, and associations with socioeconomic status and health. Ann Behav Med 2001;23(3):177-185. [doi: 10.1207/S15324796ABM2303_5] [Medline: 11495218]
5. Poeran J, Maas AFG, Birnie E, Denktas S, Steegers EAP, Bonsel GJ. Social deprivation and adverse perinatal outcomes among Western and non-Western pregnant women in a Dutch urban population. Soc Sci Med 2013 Apr;83:42-49. [doi: 10.1016/j.socscimed.2013.02.008] [Medline: 23465203]
6. Vos AA, Posthumus AG, Bonsel GJ, Steegers EAP, Denktaş S. Deprived neighborhoods and adverse perinatal outcome: a systematic review and meta-analysis. Acta Obstet Gynecol Scand 2014 Aug;93(8):727-740 [FREE Full text] [doi: 10.1111/aogs.12430] [Medline: 24834960]
7. Waelput AJM, Sijpkens MK, Lagendijk J, van Minde MRC, Raat H, Ernst-Smelt HE, et al. Geographical differences in perinatal health and child welfare in the Netherlands: rationale for the healthy pregnancy 4 all-2 program. BMC Pregnancy Childbirth 2017 Aug 01;17(1):254 [FREE Full text] [doi: 10.1186/s12884-017-1425-2] [Medline: 28764640]
8. Agyemang C, Vrijkotte TGM, Droomers M, van der Wal MF, Bonsel GJ, Stronks K. The effect of neighbourhood income and deprivation on pregnancy outcomes in Amsterdam, the Netherlands. J Epidemiol Community Health 2009 Sep;63(9):755-760. [doi: 10.1136/jech.2008.080408] [Medline: 19679715]
9. Metcalfe A, Lail P, Ghali WA, Sauve RS. The association between neighbourhoods and adverse birth outcomes: a systematic review and meta-analysis of multi-level studies. Paediatr Perinat Epidemiol 2011 May;25(3):236-245. [doi: 10.1111/j.1365-3016.2011.01192.x] [Medline: 21470263]
10. Steegers-Theunissen RPM, Twigt J, Pestinger V, Sinclair KD. The periconceptional period, reproduction and long-term health of offspring: the importance of one-carbon metabolism. Hum Reprod Update 2013;19(6):640-655 [FREE Full text] [doi: 10.1093/humupd/dmt041] [Medline: 23959022]
11. Godfrey KM, Lillycrop KA, Burdge GC, Gluckman PD, Hanson MA. Epigenetic mechanisms and the mismatch concept of the developmental origins of health and disease. Pediatr Res 2007 May;61(5 Pt 2):5R-10R. [doi: 10.1203/pdr.0b013e318045bedb] [Medline: 17413851]
12. Hanson M, Godfrey KM, Lillycrop KA, Burdge GC, Gluckman PD. Developmental plasticity and developmental origins of non-communicable disease: theoretical considerations and epigenetic mechanisms. Prog Biophys Mol Biol 2011 Jul;106(1):272-280 [FREE Full text] [doi: 10.1016/j.pbiomolbio.2010.12.008] [Medline: 21219925]
13. van Zundert SKM, van Rossem L, Willemsen SP, van der Meer L, Ernst-Smelt HE, Steegers-Theunissen RPM. Periconceptional maternal social, lifestyle and medical risk factors impair embryonic growth: the Rotterdam Periconceptional Cohort. Reprod Biomed Online 2022 Jun;44(6):1123-1133 [FREE Full text] [doi: 10.1016/j.rbmo.2022.02.011] [Medline: 35361546]
14. van Uitert EM, Exalto N, Burton GJ, Willemsen SP, Koning AHJ, Eilers PHC, et al. Human embryonic growth trajectories and associations with fetal growth and birthweight. Hum Reprod 2013 Jul;28(7):1753-1761. [doi: 10.1093/humrep/det115] [Medline: 23569080]
15. de Mendonça ELSS, de Lima Macêna M, Bueno N, de Oliveira ACM, Mello C. Premature birth, low birth weight, small for gestational age and chronic non-communicable diseases in adult life: a systematic review with meta-analysis. Early Hum Dev 2020 Oct;149:105154 [FREE Full text] [doi: 10.1016/j.earlhumdev.2020.105154] [Medline: 32799034]
16. Lastrucci V, Lorini C, Caini S, Florence Health Literacy Research Group, Bonaccorsi G. Health literacy as a mediator of the relationship between socioeconomic status and health: a cross-sectional study in a population-based sample in Florence. PLoS One 2019 Dec 23;14(12):e0227007 [FREE Full text] [doi: 10.1371/journal.pone.0227007] [Medline: 31869381]
17. M'hamdi HI, Sijpkens MK, de Beaufort I, Rosman AN, Steegers EA. Perceptions of pregnancy preparation in women with a low to intermediate educational attainment: a qualitative study. Midwifery 2018 Apr;59:62-67. [doi: 10.1016/j.midw.2018.01.004] [Medline: 29396381]
18. Wildschut HIJ, van Vliet-Lachotzki EH, Boon BM, Lie Fong S, Landkroon AP, Steegers EAP. [Preconception care: an essential part of the care for mother and child]. Ned Tijdschr Geneeskd 2006 Jun 17;150(24):1326-1330. [Medline: 16808362]
19. de Weerd S, Steegers EAP, Heinen MM, van den Eertwegh S, Vehof RMEJ, Steegers-Theunissen RPM. Preconception nutritional intake and lifestyle factors: first results of an explorative study. Eur J Obstet Gynecol Reprod Biol 2003 Dec 10;111(2):167-172. [doi: 10.1016/s0301-2115(03)00290-2] [Medline: 14597246]

20. Van Dijk MR, Huijgen NA, Willemsen SP, Laven JS, Steegers EA, Steegers-Theunissen RP. Impact of an mHealth platform for pregnancy on nutrition and lifestyle of the reproductive population: a survey. JMIR Mhealth Uhealth 2016 May 27;4(2):e53 [FREE Full text] [doi: 10.2196/mhealth.5197] [Medline: 27234926]
21. Poels M, Koster MPH, Boeije HR, Franx A, van Stel HF. Why do women not use preconception care? A systematic review on barriers and facilitators. Obstet Gynecol Surv 2016 Oct;71(10):603-612. [doi: 10.1097/OGX.0000000000000360] [Medline: 27770130]
22. van Dijk MR, Koster MPH, Oostingh EC, Willemsen SP, Steegers EAP, Steegers-Theunissen RPM. A mobile app lifestyle intervention to improve healthy nutrition in women before and during early pregnancy: single-center randomized controlled trial. J Med Internet Res 2020 May 15;22(5):e15773 [FREE Full text] [doi: 10.2196/15773] [Medline: 32412417]
23. Gootjes DV, van Dijk MR, Koster MP, Willemsen SP, Steegers EA, Steegers-Theunissen RP. Neighborhood deprivation and the effectiveness of mobile health coaching to improve periconceptional nutrition and lifestyle in women: survey in a large urban municipality in the Netherlands. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11664 [FREE Full text] [doi: 10.2196/11664] [Medline: 30973345]
24. Smith SM, van der Kleij RMJJ, Bais B, Schermer MHN, Ismaili M'hamdi H, Steegers-Theunissen RPM. Preferences of women with a vulnerable health status towards nudging for adequate pregnancy preparation as investment in health of future generations: a qualitative study. BMC Pregnancy Childbirth 2022 Jul 13;22(1):559 [FREE Full text] [doi: 10.1186/s12884-022-04887-6] [Medline: 35831786]
25. Thaler R, Sunstein C. Nudge : improving decisions about health, wealth, and happiness. New Haven, Conn. New Haven, CN: Yale University Press; 2008.
26. Murayama H, Takagi Y, Tsuda H, Kato Y. Applying nudge to public health policy: practical examples and tips for designing nudge interventions. Int J Environ Res Public Health 2023 Feb 23;20(5):3962 [FREE Full text] [doi: 10.3390/ijerph20053962] [Medline: 36900972]
27. van der Windt M, van der Kleij RM, Snoek KM, Willemsen SP, Dykgraaf RHM, Laven JSE, et al. Impact of a blended periconception lifestyle care approach on lifestyle behaviors: before-and-after study. J Med Internet Res 2020 Sep 30;22(9):e19378 [FREE Full text] [doi: 10.2196/19378] [Medline: 32996885]
28. Smith SM, Bais B, Ismaili M'hamdi H, Schermer MHN, Steegers-Theunissen RPM. Stimulating the uptake of preconception care by women with a vulnerable health status through mHealth app-based nudging (Pregnant Faster): cocreation design and protocol for a cohort study. JMIR Res Protoc 2023 Aug 09;12:e45293 [FREE Full text] [doi: 10.2196/45293] [Medline: 37556197]
29. Singh K, Rishiwal V, Kumar P. Classification of data to enhance data security in cloud computing. 2018 Presented at: IoT-SIU 2018: 3rd International Conference On Internet of Things: Smart Innovation and Usages; February 23-24, 2018; Bhimtal, India p. 23-24. [doi: 10.1109/iot-siu.2018.8519934]
30. Would you like to become pregnant? Sneller Zwanger. URL: https://www.snellerzwangerworden.nl/nl/sneller-zwanger [accessed 2024-04-11]
31. VerlosMoeder online platform. VerlosMoeder. URL: https://www.verlosmoeder.nl/ [accessed 2024-04-12]
32. VerlosMoeder. YouTube. URL: https://www.youtube.com/@VerlosMoeder [accessed 2024-04-12]
33. Djanifa da Conceicao. Instagram. URL: https://www.instagram.com/Djanifa_VerlosMoeder/ [accessed 2024-04-12]
34. Smarter Pregnancy: for couples who are or wish to get pregnant. Erasmus University Medical Center. 2019. URL: https://www.smarterpregnancy.co.uk/ [accessed 2024-04-11]
35. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500 [FREE Full text] [doi: 10.2196/11500] [Medline: 30973342]
36. Aday LA. Health status of vulnerable populations. Annu Rev Public Health 1994 May;15:487-509. [doi: 10.1146/annurev.pu.15.050194.002415] [Medline: 8054096]
37. Carpenter J, Morrison S, Craft M, Lee M. How and why are educators using Instagram? Teach Teach Educ 2020 Nov;96:103149 [FREE Full text] [doi: 10.1016/j.tate.2020.103149] [Medline: 32834464]
38. Shoufan A, Mohamed F. YouTube and education: a scoping review. IEEE Access 2022;10:125576-125599. [doi: 10.1109/access.2022.3225419]
39. Seligman HK, Wallace AS, DeWalt DA, Schillinger D, Arnold CL, Shilliday BB, et al. Facilitating behavior change with low-literacy patient education materials. Am J Health Behav 2007 Jul 01;31(1):69-78. [doi: 10.5993/ajhb.31.s1.9]
40. Pachauri S. Male involvement in reproductive health care. J Indian Med Assoc 2001 Mar;99(3):138-141. [Medline: 11478756]

Abbreviations

CIA: confidentiality, integrity, and availability
ExQ: Experience Questionnaire
MAUQ: mHealth App Usability Questionnaire
mHealth: mobile health
PCC: preconception care

SES: socioeconomic status
VHS: vulnerable health status

Edited by A Choudhury; submitted 12.10.23; peer-reviewed by S Verbiest, R Marshall; comments to author 14.02.24; revised version received 22.02.24; accepted 26.02.24; published 22.04.24.

Please cite as:
Smith SM, Bais B, Ismaili M'hamdi H, Schermer MHN, Steegers-Theunissen RPM
Stimulating Preconception Care Uptake by Women With a Vulnerable Health Status Through a Mobile Health App (Pregnant Faster): Pilot Feasibility Study
JMIR Hum Factors 2024;11:e53614
URL: https://humanfactors.jmir.org/2024/1/e53614
doi:10.2196/53614
PMID:38648092

Original Paper

Efficacy of the QuitSure App for Smoking Cessation in Adult Smokers: Cross-Sectional Web Survey

Gregory M Goldgof[1*], MD, PhD; Shweta Mishra[2], MA; Kriti Bajaj[2*], MSc

[1]Department of Laboratory Medicine, University of California, San Francisco, San Francisco, CA, United States

[2]QuitSure, Rapidkart Online Private Limited, Mumbai, India

[*]these authors contributed equally

Corresponding Author:
Kriti Bajaj, MSc
QuitSure
Rapidkart Online Private Limited
Bajaj Bhavan, 2nd Floor
226 Nariman Point
Mumbai, 400021
India
Phone: 91 2222023626
Email: kriti@quitsure.app

Abstract

Background: Cigarette smoking remains one of the leading causes of preventable death worldwide. A worldwide study by the World Health Organization concluded that more than 8 million people die every year from smoking, tobacco consumption, and secondhand smoke. The most effective tobacco cessation programs require personalized human intervention combined with costly pharmaceutical supplementation, making them unaffordable or inaccessible to most tobacco users. Thus, digital interventions offer a promising alternative to these traditional methods. However, the leading smartphone apps available in the market today have either not been studied in a clinical setting or are unable to match the smoking cessation success rates of their expensive offline counterparts. We would like to understand whether QuitSure, a novel smoking cessation app built by Rapidkart Online Private Limited, is able to bridge this efficacy gap and deliver affordable and effective smoking cessation at scale.

Objective: Our objective was to do an initial exploration into the engagement, efficacy, and safety of QuitSure based on the self-reported experiences of its users. Outcomes measured were program completion, the effect of program completion on smoking behavior, including self-reported cessation outcomes, and negative health events from using the app.

Methods: All QuitSure registered users who created their accounts on the QuitSure app between April 1, 2021, and February 28, 2022, were sent an anonymized web-based survey. The survey results were added to their engagement data on the app to evaluate the feasibility and efficacy of the app as a smoking cessation intervention. The data were analyzed using descriptive statistics (frequencies and percentages) and the χ^2 test of independence.

Results: In total, 1299 users who had completed the QuitSure program submitted the survey and satisfied the inclusion criteria of the study. Of these, 1286 participants had completed the program more than 30 days before filling out the survey, and 1040 (80.1%, 95% CI 79.1%-82.6%) of them had maintained prolonged abstinence for at least 30 days after program completion. A majority of participants (770/891, 86.4%) who were still maintaining abstinence at the time of submitting the survey did not experience any severe nicotine withdrawal symptoms, while 41.9% (373/891) experienced no mild withdrawal symptoms either. Smoking quantity prior to completing the program significantly affected quit rates (P<.001), with heavy smokers (>20 cigarettes per day) having a lower 30-day prolonged abstinence rate (relative risk=0.91; 95% CI 90.0%-96.2%) compared to lighter smokers. No additional adverse events outside of known nicotine withdrawal symptoms were reported.

Conclusions: The nature of web-based surveys and cohort selection allows for extensive unknown biases. However, the efficacy rates of survey respondents who completed the program were high and provide a case for further investigation in the form of randomized controlled trials on the QuitSure tobacco cessation program.

(JMIR Hum Factors 2024;11:e49519) doi:10.2196/49519

KEYWORDS

smoking; quit smoking; smoking cessation; smoking app; QuitSure; smoke free; quit vaping; vaping; smoker; smoke; cross-sectional study; smartphone app; tobacco consumption; tobacco; survey; nicotine; nicotine withdrawal; mobile phone

Introduction

Background

Cigarette smoking remains one of the leading causes of many premature deaths worldwide [1]. According to the World Health Organization (WHO), more than 8 million people around the world die every year, either directly or indirectly (via secondhand smoke), because of tobacco consumption. Hence, the WHO [2] has identified the tobacco epidemic as one of the world's biggest public health threats. Beyond the burden of mortality lies the burden of disease as a result of tobacco consumption. For every 1 individual who dies because of smoking, at least 30 live with a serious illness caused by smoking. Smoking causes many health issues, such as cardiovascular diseases, chronic obstructive pulmonary disease, and 12 types of cancer [3]. More than 67% of smokers face debilitating, and eventually fatal, health issues at some point in their smoking lives [4]. Health risks, as well as death risks for smokers compared to nonsmokers, have worsened, due to the deadly spread of COVID-19 across the world [5]. Meanwhile, the economic costs attributable to smoking and exposure to tobacco smoke globally have been estimated to be US $1436 per smoker, which is equivalent to around 1.8% of the world's gross domestic product [3,6].

In 2015, 68% (22.7 million) of adult smokers said that they wanted to quit smoking. In 2018, 55.1% (21.5 million) of adult smokers said that they had made a quit attempt in the past year. In 2020, 62.5% of youths (middle and high school students) who currently used tobacco products wished to quit all tobacco products, and 65.4% of youths who currently used tobacco products reported that they had stopped using all tobacco products for 1 day or longer in the past year because they were trying to quit [7]. But on the other side of the coin, a report by the National Institute of Cancer, United States [8], found that in 2020, of the 53.9% of smokers who attempted to quit smoking, only 8.5% of them were successful in doing so. In fact, research has found that it takes about 30 quitting attempts for a smoker to successfully quit [9]. The WHO, in 2022, said that without cessation support, only 4% of smokers will be able to successfully quit.

Smoking Cessation

There are several smoking cessation methods available across the world, including unassisted methods, nicotine replacement therapy (NRT), prescribed medicine (bupropion or varenicline) use, behavioral counseling, quitlines, and the use of mobile apps and websites for smoking cessation [10]. Financial incentives have gained popularity as a cessation method recently [11]. NRT, like nicotine patches, gums, and nasal sprays; medications such as bupropion and varenicline; and nonpharmacological interventions [12] are the most common smoking cessation interventions. However, NRT has shown to have success rates of only 6%-8% [13], while pharmacological interventions, despite their somewhat higher success rates of 14%-20% [14], come with the risk of side effects such as skin irritation or more serious seizures and are also very expensive [15]. Combined interventions for smokers such as behavioral interventions and long-term assistance or social support are most effective when it comes to smoking cessation [15]. However, they tend to be expensive, highly variable depending on the quality of each individual provider, accessible to only small hyperlocal communities, and cannot be scaled up to achieve population-level impact.

Smartphone Apps for Smoking Cessation

In response to the COVID-19 pandemic, the rate of smoking cessation increased from 23% to 31% [16], which creates the opportunity to encourage and support smokers to quit smoking through different smoking cessation methods. Unconventional methods such as smartphone-based apps can be more useful to increase the odds of quitting success over conventional methods because smartphone use is highly prevalent, is available 24-7, is cost-effective, requires zero-minimal human intervention, and can provide instant and constant support. Seo et al [17] found a total of 603 apps designed for smoking cessation that were available in the US, UK, Australian, and Asian markets [17]. Apps designed for smoking cessation have been downloaded 33 million times globally according to a study done by SensorTower in April 2020 (Nelson, SensorTower.com, personal communication, April 15, 2020). Users who have high engagement with smoking cessation apps have been found to be more likely to be successful in quitting [18,19].

However, literature reviews suggest that only a few apps follow the guidelines for treating tobacco dependence, and most apps use only very simple tools like calculators (41%), calendars (36%), trackers (18%), hypnosis apps (21%), and distractors (10%) [20,21]. According to the WHO, any primary care provider needs to follow the 5As (ask, advise, assess, assist, and arrange) to help a tobacco user [22]. One content analysis study suggested that 96% of the cessation apps addressed "assist" but less frequently addressed the other 4 As [21]. Another review demonstrated that only 11 (6.1%) of the 180 smoking cessation apps available in 2022 have any scientific support [23]. The review also discovered that very few apps offered evidence-based interventions such as mindfulness (18%) or cognitive behavioral therapy (CBT; 2.2%). Other reviews indicated that 88.46% of smoking cessation apps have not been updated by the developers in over a year, and 33.67% of apps have low acceptance by the market with <10,000 downloads [17,21]. Thus, the development of additional smartphone apps that have good user acceptance and are using empirically supported behavior change techniques to deliver smoking cessation interventions appears to be warranted.

The QuitSure Smoking Cessation Program

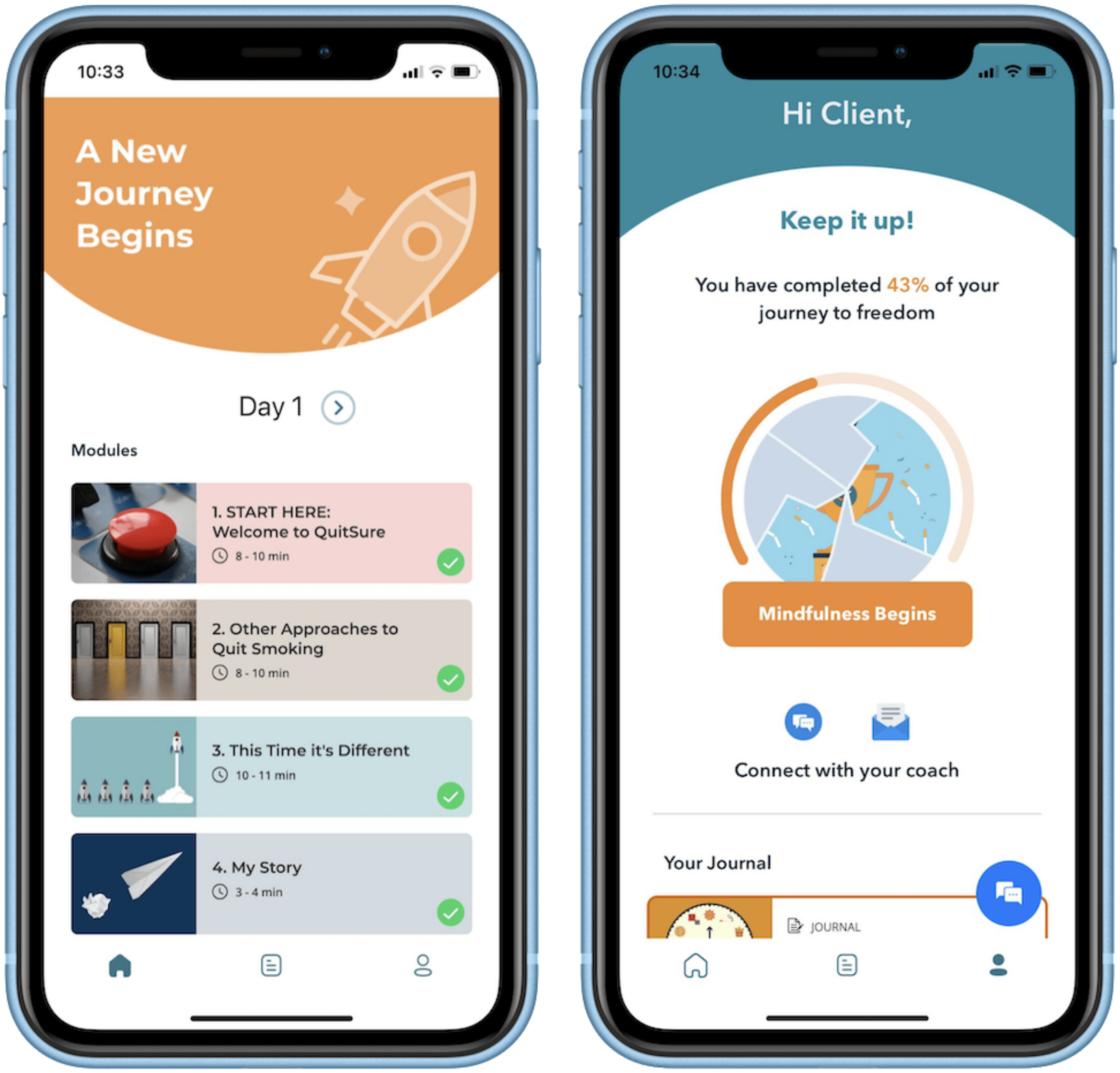
The QuitSure program (Rapidkart Online Private Limited) has been identified as one such program, which incorporates behavior change techniques like positive psychology, CBT, and mindfulness that have been shown to be effective in smoking cessation interventions [22,24], and is customized to the smoking habits and psychological needs of the user. It does not include or recommend the use of any pharmaceutical interventions, like oral supplements, medications, or NRTs.

The program follows the 5As recommended by the WHO [22]: it first asks the users about their smoking history such as quantity, patterns, past quitting attempts, and relapse reasons. This also helps the user internalize their smoking behavior. Then, it assesses if they are ready to quit and asks them about their inhibitions to quit. Unless it is a medical condition, it then advises in a clear, strong manner to quit smoking alongside a summary of the program and how it actually works. After understanding the program, willing users are then progressed to the next level, where the app provides the main content on psychology, CBT, and mindfulness [24].

CBT is incorporated by helping users question their beliefs around the positive aspects of smoking and remove them. The app does not demand any lifestyle modification. It simply helps the user accept their smoking triggers and change their response to these triggers. Under mindfulness, the app provides video exercises to teach users how to smoke mindfully by focusing on every aspect of the smoking experience. This exposes to the user the real effects of cigarettes, both while smoking and after, on their bodies and minds. All the content is delivered using empathy and without administering any guilt or blame to the user to keep the user's mind in a more calm and receptive state.

Users are required to complete the program in a very specific way with a predefined sequence of content and video exercises as shown in the screenshot of the home page in Figure 1 (left). The program requires around 6-10 hours over 6-12 days to complete. During the whole process, the app has a structured, digital journal for users to record their quitting journey and 24×7 chat-based support from trained counselors for users who have additional questions or concerns that are not addressed by the program itself as shown in Figure 1 (right). Once the user has completed the program and quit smoking, the app has postquit tools and chat support available to prevent them from relapsing. Around 12.3% (4124/33,458) of all users reach out to the counselors for support during and after the program.

Figure 1. Screenshots of the home page (left) and profile page (right) of the QuitSure app.



The program is priced at US $10 and should be affordable to most daily smokers, as the average global retail price of a pack of 20 cigarettes is approximately US $5 [25,26]. Users only have to pay on day 2 of the 6-day program. This is done, so they can understand how the program is structured and what techniques it uses before committing to it. At the end of the program, users are asked to perform a guided smoking exercise where they smoke their last cigarette and then quit cold turkey. The timestamp when they mark having performed this exercise in the app is considered as their quit date and as their program completion date. The QuitSure app has been updated an average of once every 3 weeks since its launch.

In this paper, we report the results of a retrospective cross-sectional study on users who completed the program between April 1, 2021, and February 28, 2022. Participants who started the QuitSure program but then dropped out before completing it were also surveyed to evaluate the potential areas of improvement for QuitSure. The aim of this study is to evaluate the effectiveness of the QuitSure smoking cessation program to enable smoking cessation among daily smokers. We will also examine the program's usability, feasibility, and acceptance by the market.

Methods

Design

This was a retrospective cross-sectional study to understand, through a web-based survey, smoking cessation outcomes of users who completed the QuitSure program via the QuitSure smartphone app. The survey was conducted in April 2022.

Recruitment

Users who downloaded and registered for the QuitSure quit smoking app between dates April 1, 2021, and February 28, 2022, and satisfied all the following inclusion criteria were sent the web surveys.

The study included (1) users who were daily smokers as defined by the US Centers for Disease Control and Prevention—smoked at least 1 cigarette per day before they did the program and had smoked at least 100 cigarettes before doing the program [6]; (2) adults aged 18 years and older (self-reported); (3) users who were, at minimum, proficient in the English language (since the program content is solely available in English); (4) users who had completed the entire QuitSure program; (5) and users who had a valid email address.

All registered users of the QuitSure app give consent to be contacted for the purpose of research studies at the time of registration.

Web-Based Survey

The surveys were created on the web-based Jotform tool built by Jotform Inc. The forms were set up with Jotform's Health Insurance Portability and Accountability Act (HIPAA) compliant mode, and no personally identifiable data were collected, thus protecting the confidentiality and privacy of the participants. The survey details have been reported in accordance with the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) guidelines [27].

The survey was sent via an automated email from the QuitSure server to all users who fit the inclusion criteria. The email included the relevant details for participation in the study, including the length of the survey, the aim of the study, benefits for the participant, as well as the link to the survey itself. They were encouraged to contact the investigators if they had any questions or concerns. Participation in the survey was voluntary. Participants were granted entry into a lucky draw for a US $50 Amazon voucher. This amount was chosen as being 5 times the price of the QuitSure program, and the email specified that winning the US $50 voucher did not depend on whether their experience with the QuitSure program was positive or negative.

To prevent duplication of data, each email included a hidden, unique, non–personally identifiable ID number that appeared in the survey results. For duplicate submissions, the more recent entry was kept. All questions seen by the participant, depending on the conditions, were mandatory. As a result, incomplete surveys were not recorded or used for analysis. No statistical correction methods, such as weighting of items or propensity scores, were applied to adjust for sample nonrepresentativeness.

The first part of the survey confirmed the participants' demographic data and satisfaction of the inclusion criteria. It also explicitly requested informed consent for participation in the study. Participants were shown the remaining sections in the survey only if they consented to participate in the study and met the inclusion criteria defined earlier. The second and third parts of each survey are described in the individual sections below. Prior to launching the survey, we conducted usability and technical functionality testing to ensure that participants could easily navigate and complete the electronic questionnaire. In both surveys, the questions were primarily multiple choice, with the rest requiring integer number entries. All multiple choice questions included the option for free-form "other" entries as well as the option for "none" or "choose not to share" as relevant for that question.

Survey for Program Completers

This survey, referred to as S-Completers, was sent to all users who satisfied the inclusion criteria and had also completed the QuitSure program on the QuitSure app as defined in the introduction. The second section of this survey identified the participants' smoking history, including the outcome of their completion of the QuitSure program. The third section was conditional. Each participant saw a different set of questions depending on their smoking cessation or reduction or relapse outcome after program completion. The number of questions in each section ranged from 4 to 9 depending on the responses of the participant.

Survey for Noncompleters of the Program

This survey, referred to as the S-Incompleters survey, was sent to all users who satisfied the inclusion criteria and who had started the program, but then not completed it, to get a qualitative understanding of the feasibility of the program at scale. The S-Incompleters survey also took the participants' explicit consent for participation in the study and the same demographic data and smoking history questions. The rest of the survey was to understand their reasons for leaving the program midway. The number of questions in each section ranged from 4 to 9 depending on the responses of the participant. The copy of both survey questionnaires is provided in Multimedia Appendices 1 and 2.

Study Variables

The data collected in the first part of both surveys included demographic information about the participants including gender, age, country of residence, and English proficiency. The second part of both surveys, regarding smoking history and behavior, asked what forms they consumed nicotine in before doing the program, how much and how long they had smoked, and whether they had previously tried to quit smoking using other methods. Those who had completed the QuitSure program were asked for the outcome of their most recent attempt at doing the QuitSure program and if they used any other quit smoking tools, programs, or medications or supplements during or after doing the QuitSure program.

The questions in the third part of the S-Completers survey were dependent on their outcome of doing the QuitSure program. Those who were able to quit 100% since completing the program were asked whether they experienced any mild or severe withdrawal symptoms and weight gain. They were also asked for their current level of cravings to smoke via a Likert scale ranging from 1=none to 5=unbearable. Those who quit for some time, but then relapsed, were asked how long they were able to stay quit and their reasons for relapse. Those who were only able to cut down their smoking level were asked for their new smoking rates and the reasons the program did not help them quit completely. Finally, those for whom the program had no impact, were asked why the program did not work for them. All participants who said they are still smoking were asked their current level of motivation to quit smoking on the Likert scale and whether they would use QuitSure again for their next quitting attempt.

The S-Incompleters survey asked participants why they did not complete the program and whether they were able to quit using a different method since. If they have quit since, then how they quit and what their level of cravings to smoke is. If not, then what their level of motivation to quit is and whether they will use QuitSure as their method of choice.

Ethical Considerations

This study was approved by the institutional review board of the University of California, San Francisco (IRB 21-35619, reference 331340). The email sent to the users described the study's aims and procedures as well as the security and confidentiality of their data. It also clearly stated that participation was voluntary, and they could decline to participate. The participants were given a consent form with all the details of the study including the purpose of the study. Participants younger than 18 years were not included in the study. The study observed data protection laws in effect at the time it was conducted. Participants were entered in a lucky draw to win a US $50 gift voucher for Amazon.

Statistical Analysis

Data were collected from 1299 participants from over 25 countries. The survey responses were available in Microsoft Excel (Microsoft Corp) format. The data analysis tools used were descriptive statistics (frequencies and percentages), pivot tables, as well as chi-square tests of independence.

Results

Survey for Program Completers

Of the 13,585 users who were sent the S-Completers survey email, 853 (6.3%) emails bounced. Of the 12,732 delivered emails, 5365 (42.1%) were opened. QuitSure or smoking was not mentioned in the sender or subject lines of the emails, only in the body of the email, to reduce bias based on user perception of the QuitSure program before opening the email. Therefore, we will use this number of email openers as the baseline number of users who were aware of this study. In total, 1906 (35.5%) of those who opened the email clicked on the survey link, and a final 1332 (24.8%) email openers consented to participate in the study and completed the survey. These values are all either equal to or greater than the expected opening (21.5%) and click (8%) rates based on global industry standards [28,29]. Of this set of submitters, 11 were excluded for completing the program in less than 7 days before submitting the survey. An additional 22 were excluded for submitting false data that were significantly different from actual app engagement. A final set of 1299 participants were included in the data analysis for the study. App engagement and preprogram smoking behavior data were not significantly different between those who filled out the survey versus those who did not. A flowchart representing the participant funnel for the S-Completers survey is available in Multimedia Appendix 3.

Table 1 shows the baseline characteristics of the participants, while Multimedia Appendix 4 shows their global distribution. The ratio of male to female participants was found to be 1 to 1.17. The most common age range of participants was found to be 25-34 years (n=431, 33.2%). In total, 97.2% (n=1262) of participants were cigarette smokers, while the remaining 2.6% (n=37) only consumed other forms of nicotine. Five countries, the United States, the United Kingdom, India, Canada, and Australia, had 71.8% (n=933) of the participants.

Table 1. Demographic details of participants.

	Male participants (n=587)	Female participants (n=699)	Others (n=13)	All (N=1299)
Age (years), mean (SD)[a]	35.8 (11.74)	41.83 (12.55)	27.11 (6.82)	39.5 (12.56)
Cigarettes smoked per day, mean (SD)	12.57 (8.02)	14.57 (8.56)	13.30 (5.51)	13.66 (8.83)
Median (IQR; smoking category)	10 (7-20; light)	14 (10-20; average)	13 (10-18; average)	12 (8-20; average)
Years smoked, mean (SD)	15.28 (11.40)	21.13 (12.87)	10.15 (9.69)	18.59 (12.62)
Smoking categories (cigarettes per day), n (%)				
Very light (<5)	75 (12.8)	47 (6.7)	0 (0)	122 (9.4)
Light (5-10)	119 (20.2)	109 (15.6)	3 (25)	231 (17.8)
Average (11-20)	233 (39.6)	301 (43.1)	8 (61.5)	542 (41.7)
Heavy (>20)	160 (27.2)	242 (34.6)	2 (15.4)	404 (31.1)

[a]To calculate the mean, the midpoints of each age group were considered (calculated as upper limit+lower limit/2, ie, 25+35/2 and so on).

Smoking behavior of participants prior to doing the program was grouped into 4 categories as very light (<5 cigarettes per day), light (5-10 cigarettes per day), average (11-20 cigarettes per day), and heavy (>20 cigarettes per day) [30]. The participants smoked an average of 13.66 (SD 8.83) cigarettes per day.

Effectiveness of the Smartphone App for Smoking Cessation

Participants were divided into 4 overlapping subsets based on the duration between completing the program and submitting the survey. The 4 durations chosen were 7 days, 30 days, 3 months, and 6 months, which are the commonly used smoking cessation milestones [31]. Table 2 shows the self-reported outcome of completing the program for each of these groups. Overall, 88% (1144/1299), 80.9% (1040/1286), 82.4% (991/1203), and 72.4% (725/1002) of participants had maintained prolonged abstinence for 7 days, 30 days, 3 months, and 6 months, respectively. In total, 35 of the 1203 (2.9%) participants were able to cut down on their smoking level in 30 days after the completion of the program, and 19 of the 1002 (1.9%) were able to sustain it for over 6 months after program completion.

Table 2. Prolonged abstinence after program completion.

	Participants with 7-day prolonged abstinence (1144/1299, 88%)	Participants with 30-day prolonged abstinence (1040/1286, 80.9%)	Participants with 3-month prolonged abstinence (991/1203, 82.4%)	Participants with 6-month prolonged abstinence (725/1002, 72.4%)
OR[a] (95% CI)	1.32 (86.8-89.7)	1.20 (79.4-82.3)	1.39 (80-84.6)	1.43 (69.8-75.8)
Participants by smoking behavior, n/N (%)				
Very light (<5)	107/122 (87.7)	95/121 (78.5)	84/101 (83.2)	57/87 (65.5)
Light (5-10)	209/231 (90.5)	196/228 (86)	183/207 (88.4)	132/173 (76.3)
Average (11-20)	491/542 (90.6)	446/535 (83.4)	405/466 (86.9)	319/425 (75.1)
Heavy (>20)	333/404 (82.4)	301/401 (75.1)	271/340 (79.7)	215/316 (68)
Participants by country, n/N (%)				
United States	351/393 (89.3)	324/388 (83.5)	278/357 (77.9)	242/315 (76.8)
United Kingdom	88/104 (84.6)	73/103 (70.9)	67/97 (69)	56/86 (65)
India	226/266 (84.6)	195/261 (74.7)	158/231 (68.4)	129/200 (64.5)
Canada	78/87 (90)	72/86 (83)	58/75 (77)	51/67 (76)
Australia	68/75 (91)	65/75 (87)	54/67 (81)	50/59 (85)

[a]OR: odds ratio.

The participants who were able to quit smoking as a result of the program and had maintained prolonged cessation at the time of filling out the survey (n=891) experienced varying degrees of withdrawal symptoms, as shown in Tables 3 and 4. Overall, 41.9% (n=373) experienced no mild withdrawal symptoms, and 86.4% (n=770) experienced no severe withdrawal symptoms. In total, 41.9% (n=373) experienced no weight gain after quitting with the QuitSure app; 39.6% (n=353) gained less than 5 kg, while 18.5% (n=165) gained more than 5 kg after quitting.

Table 3. Mild withdrawal symptoms reported by participants (n=891).

Symptom	Participants, n (%)
No mild withdrawal symptoms	373 (41.9)
Some mood issues	349 (39.2)
Mild sleep disturbances	196 (22)
Coughing or mild nausea	167 (18.7)
Mild digestive changes	145 (16.3)
Low energy or weakness	143 (16)
Mild headaches	131 (14.7)
Tingling of hands and feet	42 (4.7)
Others	38 (4.3)

Table 4. Severe withdrawal symptoms reported by participants (n=891).

Symptom	Participants, n (%)
No severe withdrawal symptoms	770 (86.4)
Increased depression or anxiety	90 (10.1)
Severe headaches or migraines	23 (2.6)
Severe insomnia	22 (2.5)
Severe dizziness or nausea or weakness	19 (2.1)
Strong chest pain	16 (1.8)
Others	10 (1.1)

Factors Contributing to Success

To assess whether quitting smoking via QuitSure is independent of demographic variables, 2 chi-square tests of independence were conducted as shown in Table 5. The χ^2 value for the impact of gender and age on the efficacy of the program was found to be χ^2_2=3.8 (*P*=.09) and χ^2_4=5.9 (*P*=.20), respectively. This indicates that smoking cessation via QuitSure was not dependent on the gender or age of the participants. Smoking behavior prior to starting the program, however, did affect the program's efficacy. The value was found to be χ^2_1=20.3 (*P*<.001), indicating a significant impact of smoking behavior on the quit rate. The 30-day prolonged abstinence of heavy smokers was significantly lower than that of those who smoked <20 cigarettes a day (relative risk=0.91; 95% CI 90.0%-96.2%). Country of residence also had a significant impact on program effectiveness with a value of χ^2_4=9.8 (*P*=.04) when comparing the 5 countries with the most participants. Residents of Australia had the highest 30-day prolonged abstinence rates (relative risk compared to all other participants=1.08, 95% CI 44.0%-83.0%), while residents of the United Kingdom had the lowest (relative risk for 30-day prolonged abstinence compared to all other participants=0.87, 95% CI 84.0%-99.2%).

Table 5. Factors affecting smoking cessation rates.

Factor	Chi-square (*df*)	*P* value
Gender	3.8 (2)	.09
Age groups	5.9 (4)	.20
Smoking behavior	20.3 (1)	<.001
Country of residence	9.8 (4)	.04

Factors Contributing to Failure and Relapse

Table 6 shows the reasons given for failure among participants for whom the program did not work at all (n=35), with fear of quitting (n=15, 42.9%) and lack of belief (n=11, 31.4%) being the most common reasons given. Table 7 shows the major reasons for relapse among participants who quit successfully at first but then relapsed at some point before filling out the survey (n=296). The most likely reasons for relapse were cravings for cigarettes (n=101, 34.1%) and alcohol consumption (n=91, 30.1%).

Table 6. Reasons for failure (n=112).

Reason for failure	Participants, n (%)
I was afraid of quitting	15 (42.9)
I did not believe that I could quit	11 (31.4)
I do not know	9 (25.7)
I rushed through the program and may have missed some concepts	7 (20)
I did not do all the steps of the final cigarette transformation ceremony	6 (17.1)
I smoked less than 10 cigarettes mindfully	6 (17.1)
I did not follow all the instructions	5 (14.3)
I took a break >2 days while doing the program	5 (14.3)
I did not believe in the content of the app	2 (5.1)
I did not like the content of the app	2 (5.1)

Table 7. Reasons for relapse (n=296).

Reason for relapse	Participants, n (%)
I still had bad cravings and I was unable to resist	101 (34.1)
I gave in while drinking alcohol	91 (30.1)
I faced a tragedy (eg, death of a loved one and bad breakup)	71 (24)
I became overconfident of my success	69 (23.3)
I felt self-destructive	47 (15.9)
I still believe smoking has some benefits	24 (8.1)
I did not do the program properly	18 (6.1)
My physical withdrawal symptoms were very bad	17 (5.7)
I gained a lot of weight	14 (4.7)
Other reasons (stress, peer pressure, etc)	27 (9.1)

Among those who relapsed or were unable to quit after completing the program (n=410), 80.7% (n=331) had a moderate to high motivation to quit at the time of submitting the survey. In total, 91% (n=377) said that they would consider using QuitSure for their next quitting attempt, while 46.1% (n=189) said that they will definitely use QuitSure to quit in the future.

Survey for Noncompleters of the Program

In total, 19,873 users had dropped off after starting the program and were sent the S-Incompleters survey, of which only 126 (0.6%) consented to participate in the study and submitted the survey.

Table 8 shows the reasons submitted for not completing the program (n=126). The most common reasons given for dropping off midway were a busy schedule (n=51, 40.5%), not enjoying the content of the program or having too much to read (n=25, 19.8%), and lack of belief that the program will work (n=20, 15.9%).

Table 8. Reasons for not completing the program (n=126).

Reason for not completing the program	Participants, n (%)
Busy	51 (40.5)
Did not enjoy content or too much reading	25 (19.8)
Lack of belief in the program	20 (15.9)
Quit smoking midway or cut down	16 (12.7)
Had to pay	14 (11.1)
Technical issues	11 (8.7)
Felt program was not working	9 (7.1)
Others	6 (4.7)
Was not ready	5 (4)

Discussion

Principal Findings

The purpose of the study was to understand whether the QuitSure program is an effective intervention for smoking cessation and can be implemented at scale to counteract the health and economic consequences of the tobacco epidemic. It was conducted via 2 web-based surveys. In total, 1299 participants submitted the S-Completers survey for program completers. A majority of 80.9% (1040/1286) maintained prolonged abstinence for 30 days after program completion, and 72.4% (725/1002) maintained 6-month prolonged abstinence after program completion. In total, 86.4% (770/891) of participants reported no severe withdrawal symptoms, while 41.9% (373/891) reported no withdrawal symptoms at all. Only 18.5% (165/891) experienced more than 5-kg weight gain after completing the program. Demographic variables such as gender and age did not significantly impact the program's success, but smoking quantity prior to doing the program and country of residence did have a significant impact on program efficacy. For those who relapsed, cravings and alcohol consumption were major factors, while program noncompletion was attributed to busy schedules or lack of belief in the program by the participants.

The program was able to achieve extended cessation for every category of smoker, from light to heavy with high efficacy rates, and low withdrawal symptoms after quitting. It is easy to navigate and uses simple language. The low price makes it affordable for smokers across most socioeconomic strata, and the easy-to-understand content makes it usable by anyone with a basic understanding of English. The fact that most participants who relapsed, or for whom the program did not work, continue believing in the program's potential is also a point in its favor. The difference in cessation rates in different countries indicates that the program requires some adaptations to be contextually and culturally relevant to the residents of certain countries.

When it comes to the feasibility of the program to be distributed to the population at large, the dropout rates of the program, at 59.4%, did not show improved program adherence and engagement compared to other health care apps [32]. While the very short 6-day length of the program likely increases completion rates, it requires approximately 1 hour of daily use. This high engagement requirement could be the reason why 40.5% (51/126) of participants who dropped off the program midway state being busy as the reason for noncompletion. The other major reason for dropping off the program was the length and style of the content. QuitSure could break down the program into a 30-day version with just 10-15 minutes of content per day for people who are busy or for whom the content seems too much. They could also add more graphics, videos, and design elements in the content to make it more appealing to users than plain, simple text. Lack of belief in the program's techniques was also shared as a reason for dropoff. The makers of the app can thus focus on informing the users about the scientific underpinnings of the techniques used in the program as well as include relevant references for their claims throughout the program.

The app is attempting to standardize and replicate an in-person deaddiction counseling program into a do-it-yourself app and uses many of these same psychological tools to achieve success for its users as in-person counseling [33-35]. The efficacy for those who completed the program and participated in the study seems high, indicating some degree of success. However, the program is of the do-it-yourself type and long enough that it requires high self-motivation and high intent to quit on the part of the user to complete the entire program. We do not have data on how many people dropped off even before starting because of the quantity of self-driven work required. A pre-post study analyzing dropout rates at every stage in the user journey will be required to evaluate the true feasibility of the app.

Limitations

The study had several limitations. A selection bias was created because the base sample selected was solely those who had already signed up for and completed the program, creating a closed cohort and a higher-than-normal intent to quit. Another limitation was the low response rate. Only 24.8% (1332/5365) of those who opened the email chose to submit the survey, allowing for a significant bias toward those for whom the program was successful. If we consider the program to have failed for all those who opened the email but did not submit the survey, the quit rate of QuitSure at the 30-day postprogram time point becomes just 19.4%. A recall bias may have resulted in false memories of withdrawal symptoms during the initial postprogram phase. The single measurement taken eliminates long-term cessation data of participants who only recently completed the program. Finally, the reward for filling out the survey may have motivated participants to give a false-positive response, based on an assumption that it would increase their likelihood of receiving the reward.

Ultimately, the obtained sample is not representative of the smoker population at large. To be able to understand the true feasibility and efficacy of the QuitSure program and counteract the above limitations, we would need to conduct a randomized controlled trial where the self-reported cessation of participants is confirmed via biochemical verification.

Comparison With Prior Work

Studies have shown that 46.3% of smokers who quit experience significant withdrawal symptoms ranging from anxiety, depression, irritability, and other physical symptoms [36,37]. The biggest strength of the QuitSure program is that only 13.6% (121/891) of the study participants faced severe withdrawal symptoms. Of the remaining participants, only around half faced even the milder withdrawal symptoms such as coughing and mild sleep disturbance, which are usually seen among all smokers upon quitting [38]. This could be a reflection of the program's focus on the psychological aspect of nicotine addiction via mindfulness, CBT, and reframing mental sets and beliefs, which have previously shown to reduce withdrawal symptoms after quitting [39,40]. Withdrawal symptoms are known to be a key contributor to relapse [36,37]. Therefore, it is likely that the increased effectiveness of the program and higher prolonged cessation rates are a result of these reduced withdrawal symptoms. However, QuitSure does not include any sort of NRT in its protocol. NRT is recommended by the

WHO, US Centers for Disease Control and Prevention, as well as the National Institute for Health and Care Excellence, United Kingdom [22,41,42] as an important complement to counseling and has been shown to significantly increase the success rates in psychology-based smoking cessation programs [34,43,44]. The hypothesis given is that NRT reduces withdrawal symptoms and craving levels. QuitSure could include a phased-out nicotine replacement regimen after the program to further increase its efficacy. This is especially important for heavy smokers, for whom the program was less effective.

The primary reason for relapse was due to still experiencing strong cravings for smoking. In fact, 75% (21/28) of participants who experienced greater than moderate levels of cravings after completing the program eventually relapsed. Currently, the QuitSure program does not address cravings management in any specific way after program completion, relying on the program itself to prevent the appearance of cravings at all. To address them and prevent their relapse, QuitSure can monitor craving levels after the program, with additional content and counseling for those who are struggling.

The second reason for relapse was alongside alcohol consumption. The app already recommends users not to drink any alcohol for the first week after quitting. It can extend this further and also give more guidelines on how to handle cravings when drinking alcohol.

Weight gain after quitting is another big concern among smokers, as there is evidence that nicotine reduces appetite, increases metabolism, and reduces food cravings [45,46]. Previous studies have shown that after quitting smoking, 35.4% of quitters had a weight gain over 5% of their body weight [47]. Of the study participants who were able to quit for even a brief period, 58.1% (518/891) had some weight gain. The QuitSure program should do more to specifically address the maladaptive thought patterns and beliefs connecting food, hunger, and smoking.

Previous studies have found that the higher an individual's app engagement is, the more they are likely to be able to quit smoking [19]. Thus, the QuitSure app needs to improve its engagement rates to increase program completion rates. Some tools the app developers can use to increase engagement that have previously demonstrated success are (1) gamification techniques like leaderboards, progress bars, and levels [19,48,49]; (2) small rewards to participants for every engagement milestone [48]; (3) personalizing notifications and reminders [49]; as well as (4) inclusion of a peer support group to improve program adherence and navigate postquitting withdrawal symptoms and cravings [50].

Overall, within the limitations of the study, the program shows high smoking cessation rates, low rates of withdrawal symptoms and cravings, and a generally positive experience for its users.

Conclusions

In total, 80.9% (1040/1286) of the survey respondents were able to achieve 30-day prolonged abstinence from smoking after program completion. The program also adheres to the WHO's 5As guideline for smoking cessation and includes psychological tools used in evidence-based in-person counseling protocols. However, there are many improvements in app engagement, program adherence, and postprogram support that can be made by the app developers. The high success rates, including prolonged cessation rates, among study participants are an indicator that QuitSure could be a useful tool for achieving smoking cessation at scale. Despite the severe limitations and selection biases of the study, the results make the QuitSure program a strong contender for further investigation. Health care institutions should consider and study the program's feasibility and efficacy in a more controlled setting.

Acknowledgments

The authors thank the users of the QuitSure app, whose participation and feedback played a pivotal role in this study. Their willingness to engage with the program and share their experiences provided valuable insights and enriched the findings.

Authors' Contributions

GMG and KB worked on the experimental design, data analysis, and paper writing. SM worked on data analysis and paper writing. All 3 authors reviewed the final paper.

Conflicts of Interest

GMG has no conflicts of interest. KB is a cofounder and equity holder in QuitSure. SM is an employee of QuitSure.

Multimedia Appendix 1
S-Completers survey questions.
[PDF File (Adobe PDF File), 42 KB - humanfactors_v11i1e49519_app1.pdf]

Multimedia Appendix 2
S-Incompleters survey questions.
[PDF File (Adobe PDF File), 41 KB - humanfactors_v11i1e49519_app2.pdf]

Multimedia Appendix 3

Participant funnel for the S-Completers survey.
[PDF File (Adobe PDF File), 43 KB - humanfactors_v11i1e49519_app3.pdf]

Multimedia Appendix 4
Global distribution of participants.
[PDF File (Adobe PDF File), 48 KB - humanfactors_v11i1e49519_app4.pdf]

References

1. GBD 2013 Risk Factors Collaborators, Forouzanfar MH, Alexander L, Anderson HR, Bachman VF, Biryukov S, et al. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2015;386(10010):2287-2323. [doi: 10.1016/S0140-6736(15)00128-2] [Medline: 26364544]
2. Tobacco. World Health Organization. 2021. URL: https://www.who.int/news-room/fact-sheets/detail/tobacco [accessed 2023-04-03]
3. Economic trends in tobacco. Centers for Disease Control and Prevention. 2022. URL: https://www.cdc.gov/tobacco/data_statistics/fact_sheets/economics/econ_facts/index.htm [accessed 2023-03-29]
4. More than 100 reasons to quit tobacco. Pan American Health Organization. 2022. URL: https://www.paho.org/en/more-100-reasons-quit-tobacco [accessed 2023-03-29]
5. Alqahtani JS, Oyelade T, Aldhahir AM, Alghamdi SM, Almehmadi M, Alqahtani AS, et al. Prevalence, severity and mortality associated with COPD and smoking in patients with COVID-19: a rapid systematic review and meta-analysis. PLoS One 2020;15(5):e0233147 [FREE Full text] [doi: 10.1371/journal.pone.0233147] [Medline: 32392262]
6. Goodchild M, Nargis N, Tursan d'Espaignet E. Global economic cost of smoking-attributable diseases. Tob Control 2018 Jan;27(1):58-64 [FREE Full text] [doi: 10.1136/tobaccocontrol-2016-053305] [Medline: 28138063]
7. Babb S, Malarcher A, Schauer G, Asman K, Jamal A. Quitting smoking among adults—United States, 2000-2015. MMWR Morb Mortal Wkly Rep 2017 Jan 06;65(52):1457-1464 [FREE Full text] [doi: 10.15585/mmwr.mm6552a1] [Medline: 28056007]
8. Quitting smoking. National Cancer Institute. 2022. URL: https://progressreport.cancer.gov/prevention/quitting_smoking [accessed 2023-04-04]
9. Chaiton M, Diemert L, Cohen JE, Bondy SJ, Selby P, Philipneri A, et al. Estimating the number of quit attempts it takes to quit smoking successfully in a longitudinal cohort of smokers. BMJ Open 2016 Jun 09;6(6):e011045 [FREE Full text] [doi: 10.1136/bmjopen-2016-011045] [Medline: 27288378]
10. Lund M, Kvaavik E. Methods used in smoking cessation and reduction attempts: findings from help-seeking smokers. J Smok Cessat 2021;2021:6670628 [FREE Full text] [doi: 10.1155/2021/6670628] [Medline: 34306230]
11. Notley C, Gentry S, Livingstone-Banks J, Bauld L, Perera R, Hartmann-Boyce J. Incentives for smoking cessation. Cochrane Database Syst Rev 2019 Jul 17;7(7):CD004307 [FREE Full text] [doi: 10.1002/14651858.CD004307.pub6] [Medline: 31313293]
12. Fiore MC, Jaén CR, Baker TB, Bailey WC, Benowitz NL, Curry SJ, et al. Treating tobacco use and dependence: 2008 update. US Department of Health and Human Services, Public Health Service. 2008. URL: https://www.ahrq.gov/prevention/guidelines/tobacco/index.html [accessed 2024-03-23]
13. Etter JF, Stapleton JA. Nicotine replacement therapy for long-term smoking cessation: a meta-analysis. Tob Control 2006 Aug;15(4):280-285 [FREE Full text] [doi: 10.1136/tc.2005.015487] [Medline: 16885576]
14. Cahill K, Stevens S, Perera R, Lancaster T. Pharmacological interventions for smoking cessation: an overview and network meta-analysis. Cochrane Database Syst Rev 2013 May 31;2013(5):CD009329 [FREE Full text] [doi: 10.1002/14651858.CD009329.pub2] [Medline: 23728690]
15. Stitzer ML. Combined behavioral and pharmacological treatments for smoking cessation. Nicotine Tob Res 1999;1 Suppl 2:S181-S187; discussion S207. [doi: 10.1080/14622299050012041] [Medline: 11768178]
16. Kayhan Tetik B, Gedik Tekinemre I, Taş S. The effect of the COVID-19 pandemic on smoking cessation success. J Community Health 2021 Jun;46(3):471-475 [FREE Full text] [doi: 10.1007/s10900-020-00880-2] [Medline: 32643078]
17. Seo S, Cho SI, Yoon W, Lee CM. Classification of smoking cessation apps: quality review and content analysis. JMIR Mhealth Uhealth 2022 Feb 17;10(2):e17268 [FREE Full text] [doi: 10.2196/17268] [Medline: 35175213]
18. Zeng EY, Heffner JL, Copeland WK, Mull KE, Bricker JB. Get with the program: adherence to a smartphone app for smoking cessation. Addict Behav 2016 Dec;63:120-124 [FREE Full text] [doi: 10.1016/j.addbeh.2016.07.007] [Medline: 27454354]
19. Browne J, Halverson TF, Vilardaga R. Engagement with a digital therapeutic for smoking cessation designed for persons with psychiatric illness fully mediates smoking outcomes in a pilot randomized controlled trial. Transl Behav Med 2021 Sep 15;11(9):1717-1725 [FREE Full text] [doi: 10.1093/tbm/ibab100] [Medline: 34347865]

20. Abroms LC, Lee Westmaas J, Bontemps-Jones J, Ramani R, Mellerson J. A content analysis of popular smartphone apps for smoking cessation. Am J Prev Med 2013 Dec;45(6):732-736 [FREE Full text] [doi: 10.1016/j.amepre.2013.07.008] [Medline: 24237915]
21. Hoeppner BB, Hoeppner SS, Seaboyer L, Schick MR, Wu GWY, Bergman BG, et al. How smart are smartphone apps for smoking cessation? A content analysis. Nicotine Tob Res 2016 May;18(5):1025-1031 [FREE Full text] [doi: 10.1093/ntr/ntv117] [Medline: 26045249]
22. Toolkit for delivering the 5A's and 5R's brief tobacco interventions in primary care. World Health Organization. 2014. URL: https://iris.who.int/bitstream/handle/10665/112835/9789241506953_eng.pdf [accessed 2024-03-23]
23. Bold KW, Garrison KA, DeLucia A, Horvath M, Nguyen M, Camacho E, et al. Smartphone apps for smoking cessation: systematic framework for app review and analysis. J Med Internet Res 2023 Jul 13;25:e45183 [FREE Full text] [doi: 10.2196/45183] [Medline: 37440305]
24. Hersi M, Traversy G, Thombs BD, Beck A, Skidmore B, Groulx S, et al. Effectiveness of stop smoking interventions among adults: protocol for an overview of systematic reviews and an updated systematic review. Syst Rev 2019 Jan 19;8(1):28 [FREE Full text] [doi: 10.1186/s13643-018-0928-x] [Medline: 30660199]
25. Price rankings by country of cigarettes 20 pack (Marlboro) (markets). Numbeo. 2023. URL: https://www.numbeo.com/cost-of-living/country_price_rankings?itemId=17 [accessed 2023-04-10]
26. Darden ME, Hotchkiss JL, Pitts MM. Where there's smoke…: the wage impact of smoking. Federal Reserve Bank Atlanta's Policy Hub. 2020. URL: https://www.atlantafed.org/research/publications/policy-hub/2020/10/08/11-wage-impact-of-smoking.aspx [accessed 2024-03-19]
27. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
28. What are good open rates, CTRs, and CTORs for email campaigns? Campaign Monitor. 2022. URL: https://www.campaignmonitor.com/resources/knowledge-base/what-are-good-email-metrics/#:~:text=According%20to%20our%202022%20Email,average%20between%2025%2D28%25.[accessed [accessed 2023-04-10]
29. Chung L. What is a good survey response rate for online customer surveys? Delighted. URL: https://delighted.com/blog/average-survey-response-rate [accessed 2023-04-10]
30. DeVita VT. How do I know if I'm a light, average, or heavy smoker? ShareCare. 2023. URL: https://www.sharecare.com/health/quit-smoking/how-do-i-know-if-im- [accessed 2023-04-12]
31. Japuntich SJ, Leventhal AM, Piper ME, Bolt DM, Roberts LJ, Fiore MC, et al. Smoker characteristics and smoking-cessation milestones. Am J Prev Med 2011 Mar;40(3):286-294 [FREE Full text] [doi: 10.1016/j.amepre.2010.11.016] [Medline: 21335259]
32. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. J Affect Disord 2020 Feb 15;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
33. Manjula DC, Shekar V, Reddy RCJ. Effect of behavioural counselling in tobacco cessation. J Sci Dent 2018;8(2):58-60 [FREE Full text] [doi: 10.5005/jsd-8-2-58]
34. Lindson-Hawley N, Thompson TP, Begh R. Motivational interviewing for smoking cessation. Cochrane Database Syst Rev 2015 Mar 02(3):CD006936. [doi: 10.1002/14651858.CD006936.pub3] [Medline: 25726920]
35. Blebil AQ, Sulaiman SAS, Hassali MA, Dujaili JA, Zin AM. Impact of additional counselling sessions through phone calls on smoking cessation outcomes among smokers in Penang State, Malaysia. BMC Public Health 2014 May 16;14:460 [FREE Full text] [doi: 10.1186/1471-2458-14-460] [Medline: 24886549]
36. McLaughlin I, Dani JA, De Biasi M. Nicotine withdrawal. Curr Top Behav Neurosci 2015;24:99-123 [FREE Full text] [doi: 10.1007/978-3-319-13482-6_4] [Medline: 25638335]
37. Lancaster T, Stead LF. Individual behavioural counselling for smoking cessation. Cochrane Database Syst Rev 2017 Mar 31;3(3):CD001292 [FREE Full text] [doi: 10.1002/14651858.CD001292.pub3] [Medline: 28361496]
38. Thornley S, McRobbie H, Lin RB, Bullen C, Hajek P, Laugesen M, et al. A single-blind, randomized, crossover trial of the effects of a nicotine pouch on the relief of tobacco withdrawal symptoms and user satisfaction. Nicotine Tob Res 2009 Jun;11(6):715-721. [doi: 10.1093/ntr/ntp054] [Medline: 19454549]
39. Cui ZY, Li YH, Liu Z, Li L, Nie XQ, Zhou XM, et al. The experience of tobacco withdrawal symptoms among current smokers and ex-smokers in the general population: findings from nationwide China Health Literacy Survey during 2018-19. Front Psychiatry 2022;13:1023756 [FREE Full text] [doi: 10.3389/fpsyt.2022.1023756] [Medline: 36713929]
40. Mhende J, Bell SA, Cottrell-Daniels C, Luong J, Streiff M, Dannenfelser M, et al. Mobile delivery of mindfulness-based smoking cessation treatment among low-income adults during the COVID-19 pandemic: pilot randomized controlled trial. JMIR Form Res 2021 Jul 23;5(7):e25926 [FREE Full text] [doi: 10.2196/25926] [Medline: 34033580]
41. Quit smoking. Centers for Disease Control and Prevention. 2023. URL: https://www.cdc.gov/tobacco/quit_smoking/index.htm [accessed 2023-04-13]
42. Recommendations on promoting quitting. National Institute for Health and Care Excellence. 2023. URL: https://www.nice.org.uk/guidance/ng209/chapter/Recommendations-on-promoting-quitting [accessed 2023-04-13]

43. Jódar-Sánchez F, Carrasco Hernández L, Núñez-Benjumea FJ, Mesa González MA, Moreno Conde J, Parra Calderón CL, et al. Using the Social-Local-Mobile app for smoking cessation in the SmokeFreeBrain project: protocol for a randomized controlled trial. JMIR Res Protoc 2018 Dec 06;7(12):e12464 [FREE Full text] [doi: 10.2196/12464] [Medline: 30522992]
44. Carrasco-Hernandez L, Jódar-Sánchez F, Núñez-Benjumea F, Moreno Conde J, Mesa González M, Civit-Balcells A, et al. A mobile health solution complementing psychopharmacology-supported smoking cessation: randomized controlled trial. JMIR Mhealth Uhealth 2020 Apr 27;8(4):e17530 [FREE Full text] [doi: 10.2196/17530] [Medline: 32338624]
45. Chen H, Hansen MJ, Jones JE, Vlahos R, Bozinovski S, Anderson GP, et al. Cigarette smoke exposure reprograms the hypothalamic neuropeptide Y axis to promote weight loss. Am J Respir Crit Care Med 2006 Jun 01;173(11):1248-1254. [doi: 10.1164/rccm.200506-977OC] [Medline: 16531608]
46. Bloom EL, Farris SG, DiBello AM, Abrantes AM. Smoking-related weight and appetite concerns and use of electronic cigarettes among daily cigarette smokers. Psychol Health Med 2019 Feb;24(2):221-228 [FREE Full text] [doi: 10.1080/13548506.2018.1537495] [Medline: 30346797]
47. Jeremias-Martins E, Chatkin JM. Does everyone who quit smoking gain weight? A real-world prospective cohort study. J Bras Pneumol 2019 Feb 25;45(1):e20180010 [FREE Full text] [doi: 10.1590/1806-3713/e20180010] [Medline: 30810645]
48. Garaialde D, Cox AL, Cowan BR. Designing gamified rewards to encourage repeated app selection: effect of reward placement. Int J Hum-Comput Stud 2021 Sep;153:102661 [FREE Full text] [doi: 10.1016/j.ijhcs.2021.102661]
49. Jakobsen PR, Christensen JR, Nielsen JB, Søndergaard J, Jarbøl DE, Olsen MH, et al. Identification of important factors affecting use of digital individualised coaching and treatment of type 2 diabetes in general practice: a qualitative feasibility study. Int J Environ Res Public Health 2021;18(8):3924 [FREE Full text] [doi: 10.3390/ijerph18083924] [Medline: 33917999]
50. Cheung YTD, Chan CHH, Lai CKJ, Chan WFV, Wang MP, Li HCW, et al. Using WhatsApp and Facebook online social groups for smoking relapse prevention for recent quitters: a pilot pragmatic cluster randomized controlled trial. J Med Internet Res 2015 Oct 22;17(10):e238 [FREE Full text] [doi: 10.2196/jmir.4829] [Medline: 26494159]

Abbreviations

CBT: cognitive behavioral therapy
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
HIPAA: Health Insurance Portability and Accountability Act
NRT: nicotine replacement therapy
WHO: World Health Organization

Edited by Y Senathirajah; submitted 31.05.23; peer-reviewed by D Ford, L Huang; comments to author 02.09.23; revised version received 28.11.23; accepted 28.02.24; published 06.05.24.

Please cite as:
Goldgof GM, Mishra S, Bajaj K
Efficacy of the QuitSure App for Smoking Cessation in Adult Smokers: Cross-Sectional Web Survey
JMIR Hum Factors 2024;11:e49519
URL: https://humanfactors.jmir.org/2024/1/e49519
doi:10.2196/49519
PMID:38709553

Original Paper

Digital Adherence Technologies Linked to Mobile Money Incentives for Medication Adherence Among People Living With Tuberculosis: Mixed Methods Feasibility and Acceptability Study

Angella Musiimenta[1,2], PhD; Wilson Tumuhimbise[1,2], PhD; Esther Atukunda[3], PhD; Aaron Mugaba[1,2], MBA; Sebastian Linnemayr[4], PhD; Jessica Haberer[5,6], MD

[1]Faculty of Computing and Informatics, Mbarara University of Science and Technology, Mbarara, Uganda

[2]Angels Compassion Research and Development Initiative, Mbarara, Uganda

[3]Faculty of Medicine, Mbarara University of Science and Technology, Mbarara, Uganda

[4]Rand Corporation, Santa Monica, CA, United States

[5]Harvard Medical School, Boston, MA, United States

[6]Massachusetts General Hospital Center for Global Health, Boston, MA, United States

Corresponding Author:
Angella Musiimenta, PhD
Faculty of Computing and Informatics
Mbarara University of Science and Technology
Plot 8-19 Mbarara Kabale Road
Mbarara, P.O. Box 653
Uganda
Phone: 256 776820598
Email: amusiimenta@must.ac.ug

Abstract

Background: Complementing digital adherence technologies (DATs) with mobile money incentives may improve their utility in supporting tuberculosis medication adherence, yet the feasibility and acceptability of this integrated approach remain unclear.

Objective: This study aims to describe the feasibility and acceptability of a novel DAT intervention called My Mobile Wallet composed of real-time adherence monitoring, SMS text message reminders, and mobile money incentives for tuberculosis medication adherence in a low-income setting.

Methods: We purposively recruited people living with tuberculosis from the Mbarara Regional Referral Hospital in Mbarara, Uganda, who (1) were starting tuberculosis treatment at enrollment or within the past 4 weeks, (2) owned a mobile phone, (3) were able to use SMS test messaging, (4) were aged ≥18 years, and (5) were living in Mbarara district. At study exit (month 6), we used interviews and questionnaires informed by the unified theory of acceptance and use of technology (UTAUT) to collect feasibility and acceptability data, reflecting patients' experiences of using each component of My Mobile Wallet. Feasibility also included tracking the functionality of the adherence monitor (ie, an electronic pillbox) as well as SMS text message and mobile money delivery. We used a content analytical approach to inductively analyze qualitative data and Stata (version 13; StataCorp LLC) to analyze quantitative data.

Results: All 39 participants reported that the intervention was feasible because it was easy for them to use (eg, access and read SMS text messages) and worked as expected. Almost all SMS text messages (6880/7064, 97.4%) were sent as planned. The transmission of adherence data from the monitor worked well, with 98.37% (5682/5776) of the data transmitted as planned. All participants additionally reported that the intervention was acceptable because it helped them take their tuberculosis medication as prescribed; the mobile money incentives relieved them of tuberculosis-related financial burdens; SMS text message reminders and electronic pillbox–based alarms reminded them to take their medication on time; and participants perceived real-time adherence monitoring as "being watched" while taking their medication, which encouraged them to take their medication on time to demonstrate their commitment. The intervention was perceived as a sign of care, which eventually created emotional support and a sense of connectedness to health care. Participants preferred daily SMS text message reminders (32/39, 82%) to reminders linked to missed doses (7/39, 18%), citing the fact that tuberculosis medication is taken daily.

Conclusions: The use of real-time adherence monitoring linked to SMS text message reminders and mobile money incentives for tuberculosis medication adherence was feasible and acceptable in a low-resource setting where poverty-based structural barriers heavily constrain tuberculosis treatment and care.

(JMIR Hum Factors 2024;11:e47996) doi:10.2196/47996

KEYWORDS

digital adherence technologies; real-time monitoring; SMS text message reminders; mobile money; financial incentives; tuberculosis; medication adherence; user-centered approach

Introduction

Background

Tuberculosis treatment adherence remains challenging in Uganda. Constraints to tuberculosis medication adherence include a lack of transport to the clinic to pick up the drugs and forgetfulness [1]. Digital adherence technologies (DATs) are being explored to encourage adherence to tuberculosis medication [2,3]. Recently, we showed that real-time adherence monitors linked to SMS text message reminders were potentially useful in reminding patients to take their medication and encouraging tuberculosis medication adherence in rural Uganda [4]. However, tuberculosis is well known to be a disease of poverty [5], and the lack of money may potentially limit the usefulness of DATs (eg, the inability to afford transport to pick up the medications) [6]. Although effective tuberculosis treatment has existed since the 1940s and is available for free, many people delay seeking treatment, struggle with medication adherence, or do not complete their treatment because of poverty [1]. This is because tuberculosis leads to the loss of productivity of patients and their caregivers, resulting in additional costs for patients in the form of transport to and from the clinic and may lead to loss of employment for fear of spreading the disease to other people [7]. Currently, in Uganda, 53% of patients living with tuberculosis take loans or sell property to meet the costs of their tuberculosis care [8]. Interventions are necessary to overcome the poverty-based structural barriers to tuberculosis treatment, including unconditional transport to and from the clinic. According to the End TB strategy of the World Health Organization (WHO), the use of social protection schemes (such as transport to the clinic and meals) could lower the financial burden of tuberculosis [9]. A recent systematic literature review and meta-analysis by Richterman et al [10] defines cash transfers as cash payments provided to specific beneficiaries. The review indicates that cash transfer interventions may improve treatment success among patients with pulmonary tuberculosis, although the review expresses the need for more research regarding the effectiveness of sensitive cash transfers for tuberculosis care, especially in low-income countries [10].

The use of mobile money technology (money sent, received, or saved on mobile phones) is a promising tool for delivering health-related cash transfers; for instance, mobile money enabled pregnant women to save for maternal health care in Kenya [11], while a progressive incentive scheme to reward private physicians and community health care workers enhanced identification and referral of suspected tuberculosis cases and treatment tracking in Pakistan [12]. The use of mobile money transfers to incentivize patients living with tuberculosis to take their drugs may potentially improve their adherence to medication [13]. However, the use of mobile money services in the context of health care is still in its infancy, and the limited research in this area reports mixed results [14].

My Mobile Wallet

My Mobile Wallet is a DAT intervention composed of a real-time adherence monitor, SMS text message reminders, and mobile money incentives (known as WiseCash). The financial incentives are meant to motivate participants to take their medication as well as enable them to attend their clinic appointments for pill refills. The intervention was developed through user-centered approaches [15], and we previously published formative qualitative findings indicating the anticipated benefits and challenges of using the intervention for tuberculosis medication adherence in rural Uganda [13]. In brief, participants reported that the intervention could remind them to take their medication as well as support, and motivate tuberculosis medication adherence. However, they expressed concerns about the possible unintended tuberculosis status disclosure as well as the possibility of using the money for other competing demands. This information was then used to refine and improve My Mobile Wallet.

This paper presents the feasibility and acceptability of a pilot study implementing My Mobile Wallet. Specifically, we present the practical experiences of people living with drug-sensitive tuberculosis who used the intervention during their 6-month tuberculosis treatment period.

Methods

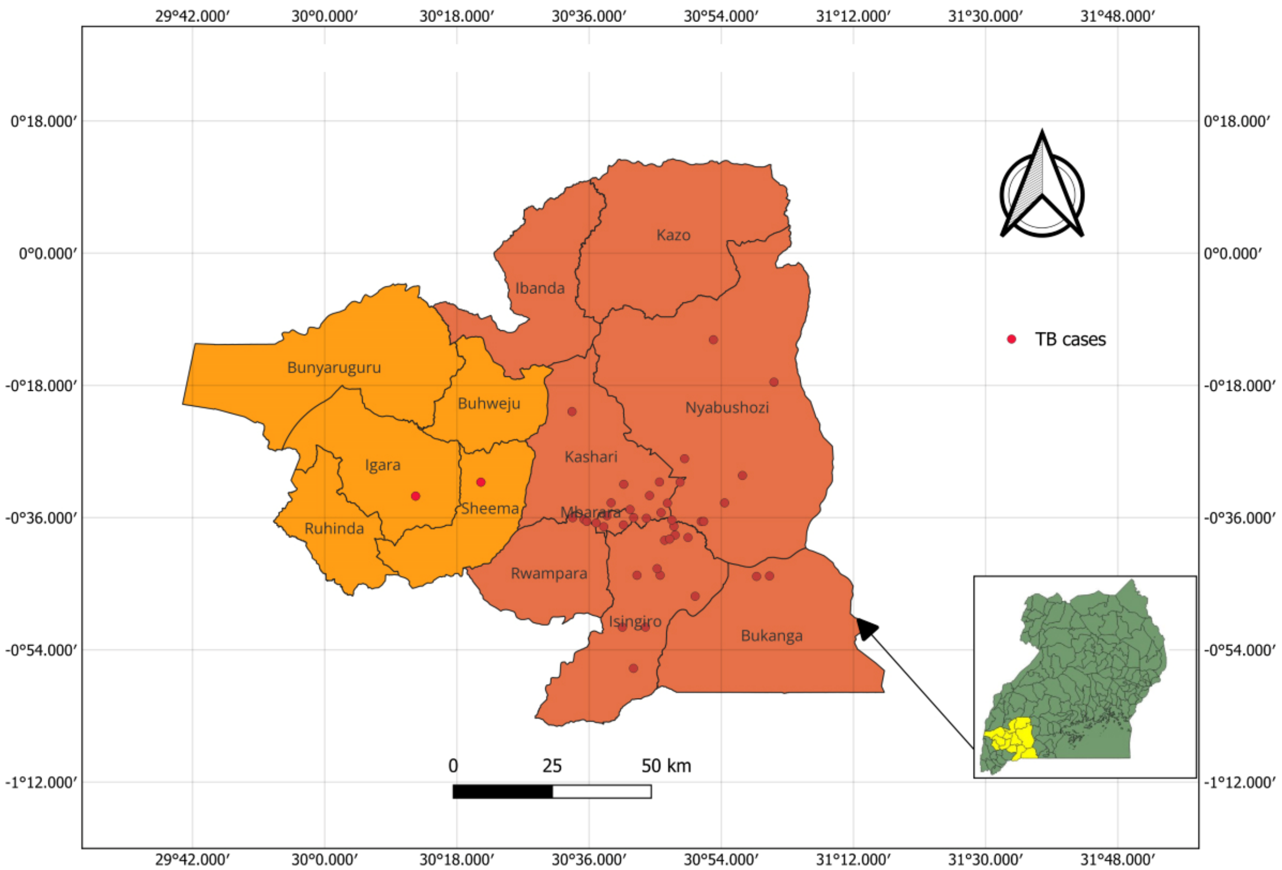
Study Design and Setting

This study used a convergent mixed methods study design. The study recruited people living with tuberculosis from the tuberculosis clinic at the Mbarara Regional Referral Hospital (MRRH) in Mbarara in southwestern Uganda. The tuberculosis clinic provides care to an estimated 600 people living with tuberculosis annually. All newly diagnosed people living with tuberculosis receive free tuberculosis medication and are counseled about the benefits of tuberculosis medication at the tuberculosis clinic. At the MRRH, the recommended directly observed therapy approach (which advises that patients should take their medication as they are watched by a health care provider or treatment supporter) is not used for monitoring medication adherence due to the costs involved for both people living with tuberculosis and the health care workers. Instead, people living with tuberculosis are treated with the 2HRZE regimen (isoniazid, rifampin, pyrazinamide, and ethambutol for 2 months) in the initiation phase, with clinic visits every 2 weeks. At the end of the 2-month period, they return to the tuberculosis clinic for a sputum conversion check. Those who become smear negative continue with the 4HR regimen (isoniazid plus rifampin for 4 months) in the continuation phase, with monthly clinic visits. Those with positive test results receive GeneXpert to exclude rifampicin resistance; subsequent treatment is then individualized. Treatment may be extended up to a full year to compensate for missed medication pick-ups or doses.

Selection of Study Participants

Between July 2022 and October 2022, we recruited participants at the MRRH according to the following inclusion criteria: (1) newly diagnosed with tuberculosis per the clinic records and starting tuberculosis treatment at enrollment or within the past 4 weeks, (2) owning a mobile phone, and (3) living in Mbarara district (Figure 1). We excluded individuals who were unwilling or unable to provide informed consent due to severe mental conditions per the clinical records and those unable to use mobile money–based SMS text messaging (we trained potential participants and tested this skill at recruitment). We purposively sampled patients to achieve relatively balanced representation by HIV status and sex to solicit diverse perspectives.

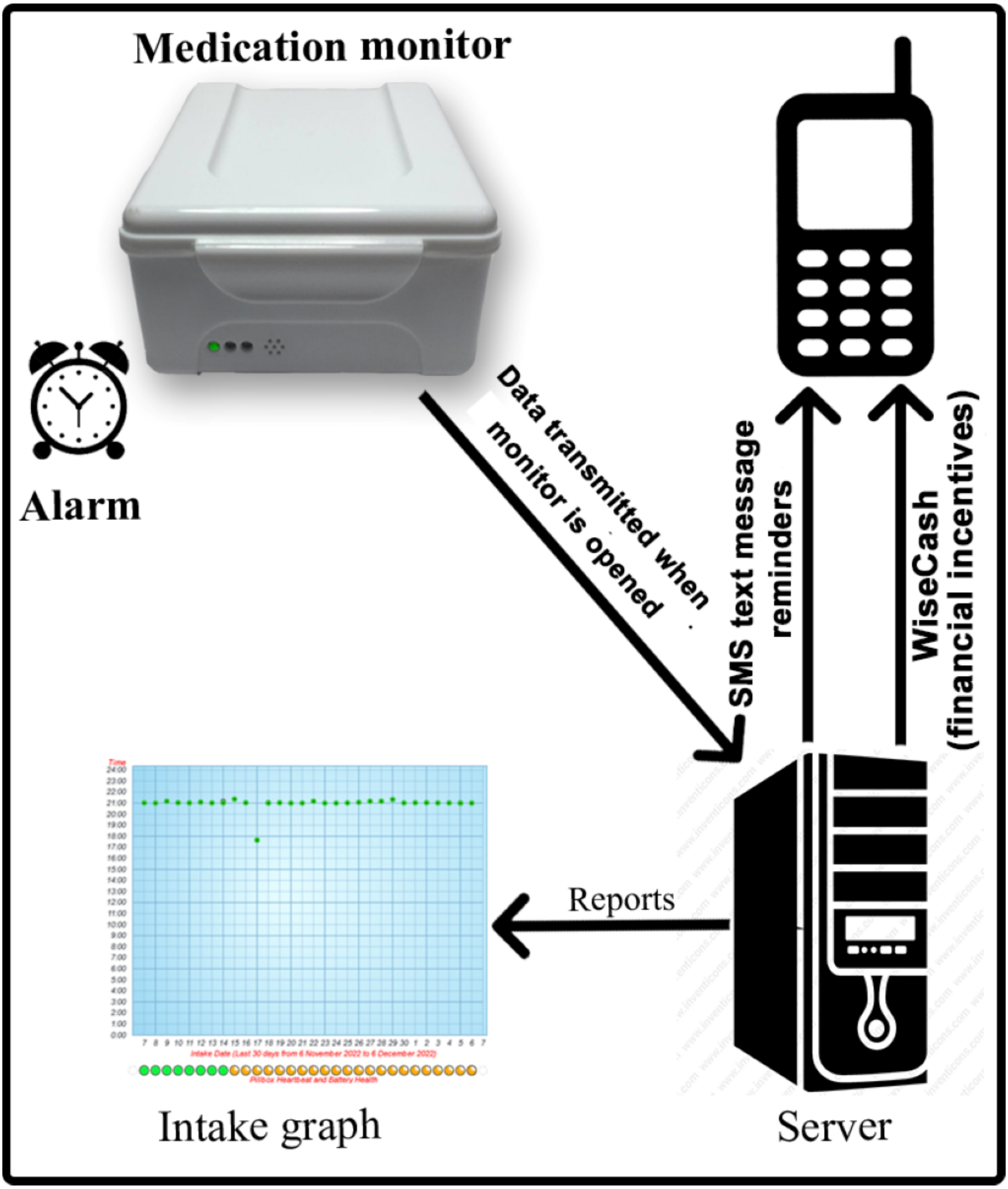
Figure 1. The study area map and geographic distribution of participants. TB: tuberculosis.



Intervention Technology

Details of the My Mobile Wallet intervention are described elsewhere [13]. Briefly, as shown in Figure 2, the intervention is composed of the following 3 components: a real-time medication monitor (Wisepill evriMED1000) to monitor medication adherence by sending signals when opened (the monitor records a date-and-time stamp as a proxy for taking medication, and it has an option to set an alarm sound to remind patients to take their medication); SMS text message reminders sent to users' mobile phones to remind them to take their medication as prescribed (reminders are sent daily for 2 months, after which they are triggered as needed by missed or delayed doses); and the WiseCash app, which uses a tailored mobile money platform for sending financial incentives for transport to the clinic and motivating medication adherence.

Figure 2. The My Mobile Wallet intervention diagram.



Study Procedures

We first oriented each participant to the My Mobile Wallet intervention components. We explained and demonstrated how the real-time adherence monitor (Wisepill) works, including how it monitors medication adherence and sends a signal to researchers every time it is opened, how the monitor makes an alarm sound to remind patients to take their medication, and how to open and close the monitor to put in or retrieve medication. Participants were then asked to demonstrate how the monitor works. Next, we explained to them how the intervention sends daily SMS text message reminders (30 minutes before medication-taking time) to remind participants to take their medication for the first 2 months. We also explained how the intervention sends SMS text message reminders for the next 4 months only if the monitor is not opened within an hour of the expected time (known as triggered SMS text message reminders). We then explained how the intervention transfers USh 28,000 (approximately equivalent to US $8; we decided upon this amount based on the transport costs, which had increased during the COVID-19 pandemic and had not reduced at the time of implementing this intervention) as an unconditional monthly mobile money incentive to mobile phones belonging to people living with tuberculosis for facilitating transport to the clinic (Figure 1 shows a visual representation of the geographic distribution of the participants) from the date of study recruitment until the end of their 6-month treatment period; furthermore, the participants were informed that USh 5250 (approximately equivalent to US $1.50) would be transferred as a monthly conditional medication adherence incentive to those with a medication adherence rate of ≥90% as ascertained from the real-time adherence monitor. The transfer of the transport incentives required patients to inform the research staff about their next date of appointment so that it could be input into the WiseCash application to allow automatic triggering of the transfer of the transport incentive a day before their next visit. We decided upon the medication adherence rate of >90% because evidence shows that adherence below this level does not yield favorable treatment outcomes [16].

Data Collection

We administered a baseline demographic and sociobehavioral questionnaire to participants at enrollment, which included age, sex, tuberculosis medication specifications (drugs and planned dosing times), and mobile phone number and use. We used the interviewer-administered approach for administering questionnaires to elicit quantitative data orally from the participants (ie, closed-ended questions read out in the participants' local language by the researcher, with participants answering the questions orally). Several validated surveys were adopted and included in this questionnaire (eg, the Duke-UNC Functional Social Support Questionnaire for measuring social support [17], the asset index scale to assess socioeconomic status [18], the depression section of the Hopkins Symptom Checklist for assessing depression [19], the Household Food Insecurity Access Scale for measuring food insecurity [20], the Alcohol Use Disorders Identification Test for assessing alcohol consumption [21], and the Internalized AIDS-Related Stigma Scale for assessing stigma [22]). Feasibility was ascertained by tracking the functionality of the monitor and SMS text message and mobile money delivery. The unified theory of acceptance and use of technology (UTAUT) model [23], given its track record of predicting a substantial portion of the acceptance of digital health interventions, provided a basis for developing surveys and interview guides for capturing participants' views on feasibility and acceptability at study exit (month 6). The UTAUT model asserts that the adoption of technology is influenced by four major constructs as perceived by an individual user: (1) performance expectancy or perceived usefulness of the intervention (in this case, My Mobile Wallet) (2) effort expectancy or perceived ease of use of the intervention, (3) social norms (how others perceive the individual's use of the intervention), and (4) facilitating conditions (the availability of technical and organizational infrastructure to support the use of the intervention). A structured exit questionnaire aimed at eliciting closed-ended information from participants regarding their experiences of using My Mobile Wallet was administered. This was a Likert scale questionnaire that sought to explore the extent to which participants liked or disliked the functionalities of the intervention. Qualitative open-ended interviews, by contrast, elicited in-depth information about participants' experiences using each component of My Mobile Wallet (the real-time monitor, SMS text message reminders and monitor alarms, and the WiseCash application), including benefits and challenges related to the technologies. Authors WT and ATM (who are trained in qualitative research and research ethics) conducted the semistructured in-depth interviews with participants in a private space at a research office near the MRRH until thematic saturation was reached at the 30th participant interview. Each interview lasted between 30 and 60 minutes and was conducted in the local language (Runyankole), digitally recorded, transcribed, and translated into English. After each interview, author AM reviewed the transcripts for quality, clarity, and detail.

Data Analysis

We followed the UTAUT model [23] to review transcripts for content related to acceptability. We then developed a coding scheme, used it to code the data, and reviewed the coded data to develop descriptive categories. We mapped the descriptive categories onto the domains of the UTAUT model's four major constructs that influence technology adoption: (1) performance expectancy or perceived usefulness, (2) effort expectancy or perceived ease of use, (3) social norms, and (4) facilitating conditions. Illustrative quotations were then selected from the coded data. After the completion of the codebook, we applied the codes using NVivo 11 (Lumivero). We followed the COREQ (Consolidated Criteria for Reporting Qualitative Research) [24] checklist in reporting qualitative results. Feasibility metrics and the quantitative assessment of acceptability were analyzed descriptively by WT and ATM using Stata 13.

Ethical Considerations

The institutional review committees of Mbarara University of Science & Technology (MUST-2021-102) and the Uganda National Council for Science and Technology (HS1688ES) approved this study. All participants provided signed informed consent before study participation.

Results

Demographic Characteristics

Of the 54 screened participants, we excluded 5 (9%) for not owning a mobile phone, 5 (9%) for not living within 60 kilometers of Mbarara district, and 4 (7%) for having mobile phone numbers that were not registered for mobile money service. Thus, 40 (74%) of the 54 screened participants were enrolled in the study and used the intervention for 6 months. Of these 40 participants, 1 (2%) was lost to follow-up (her mobile phone was unreachable). As indicated in Table 1, of the 40 participants, 24 (60%) were female, 27 (68%) had coinfection with HIV, 34 (85%) had no regular or fixed income, 18 (45%) did not study beyond primary level (typically attended by children aged 6-12 years), 40 (100%) perceived their social support to be insufficient, and 21 (53%) reported severe food insecurity. The participants' median age was 38 (IQR 28-54) years, and, before joining the study, they were on medication for a median of 4 (IQR 2.5-8) weeks.

Table 1. Baseline demographic characteristics of participants (n=40).

Characteristics	Values
Age (y), median (IQR)	38 (28-54)
Weeks on medication before joining the study, median (IQR)	4 (2.5-8)
Sex, n (%)	
Male	24 (60)
Female	16 (40)
Education, n (%)	
None	3 (8)
P1-P7[a]	18 (45)
Ordinary level	9 (22)
Advanced level	3 (8)
Tertiary level	7 (18)
Income (fixed wages or salary), n (%)	
Yes	6 (15)
No	34 (85)
Heavy alcohol consumption, n (%)	
Yes	1 (2)
No	39 (98)
Enough social support (no), n (%)	40 (100)
Food insecurity, n (%)	
Yes	21 (52)
No	19 (48)
Probable depression, n (%)	
Yes	1 (2)
No	39 (98)
Asset index scale[b], n (%)	
Lowest quartile	16 (40)
25%-100% quartiles	24 (60)
HIV status, n (%)	
Negative	13 (32)
Positive	27 (68)

[a]In the Ugandan education system, primary school (P1-P7) is often attended by children aged 6 to 12 years.

[b]Index was measured using the measure proposed by Filmer and Pritchett [18].

As indicated in Table 2, at baseline, half of the participants (20/40, 50%) did not share their mobile phones with anyone, 85% (34/40) checked their SMS text messages more frequently than *often* in a week, 75% (30/40) *often* used mobile money, 88% (35/40) preferred receiving SMS text message reminders daily because medication taking is a daily activity, 68% (27/40) preferred SMS text message reminders that are not easily related to tuberculosis (eg, “Hello today”) to avoid unwanted tuberculosis status disclosure, and 38% (15/40) preferred receiving mobile money incentives for transport to the clinic a day before the clinic visit to avoid using the money for other competing needs.

Table 2. Mobile phone use and intervention preferences at baseline (n=40).

Questions	Participants, n (%)
Who else uses your mobile phone?	
Spouse	3 (8)
Family member	16 (40)
Neighbor	1 (2)
No one else	20 (50)
Check for SMS text messages in a week	
Never	1 (2)
Less than *often*	5 (12)
More than *often*	34 (85)
Use of mobile money	
Less than *often*	10 (25)
More than *often*	30 (75)
Reasons for delay in checking for SMS text messages last week[a]	
Mobile phone not charged	20 (50)
Mobile phone was used by someone else	2 (5)
No adequate signal	9 (22)
Mobile phone not functioning	1 (2)
Used by someone else	2 (5)
Preferred frequency of receiving SMS text message reminders	
Daily	35 (88)
Weekly	5 (12)
Preferred content for SMS text message reminders	
Not easily related to TB[b] (eg “Hello today”)	27 (68)
Easily related to TB (eg, “Take your TB drugs”)	13 (32)
SMS text message language preference	
Local language	22 (55)
English	18 (45)
When to send the mobile money incentive for transport to the clinic	
1 day before the clinic visit	25 (63)
2 days before the clinic visit	10 (25)
>2 days before the clinic visit	5 (12)

[a]Reasons for delay in checking for SMS text messages last week, n=21, 52%.

[b]TB: tuberculosis.

Exit Survey Results

All participants self-reported that it was easy for them to access and read the mobile money SMS text messages as well as the medication-taking reminders (39/39, 100%) and open the Wisepill device to retrieve their medication (39/39, 100%). In addition, all participants (39/39, 100%) received the mobile money incentive for transport to the clinic as expected, received the medication adherence incentives as expected, and reported that the real-time adherence monitor worked as expected. All participants (39/39, 100%) additionally reported that the mobile money incentives, the Wisepill device, and the SMS text message reminders helped them take their tuberculosis medication on time or as prescribed.

The average adherence rate ascertained from the real-time monitors was 90.4% (SD 8.6%), and 24 (60%) of the 40 participants had an adherence rate of >90%.

As indicated in Table 3, almost all participants (38/40, 95%) opted to be reminded by both SMS text message reminders and alarms from the real-time monitor. Of the 40 participants, 2 (5%) requested study staff to switch off the alarms on their monitors at enrollment because they anticipated being inconvenienced by the sound. Participants preferred daily SMS text message reminders (32/39, 82%) to reminders linked to missed doses (7/39, 18%), citing the fact that tuberculosis medication is taken daily. All participants reported that the mobile money incentives were sent as expected. However, 4 (10%) of the 40 participants received cash (once during the study period) as refund for the transport fare instead of being sent the mobile money incentive for transport to the clinic. These participants did not inform the study staff on time about their next clinic visit date, which had to be input into the WiseCash application to allow automatic triggering of the transfer of the incentive a day before the clinic visit. Almost all the SMS text messages (6880/7064, 97.4%) were sent as planned. The transmission of adherence data from the monitor worked well, with 98.37% (5682/5776) of the data transmitted as planned. No real-time adherence monitor malfunctioned during the study period.

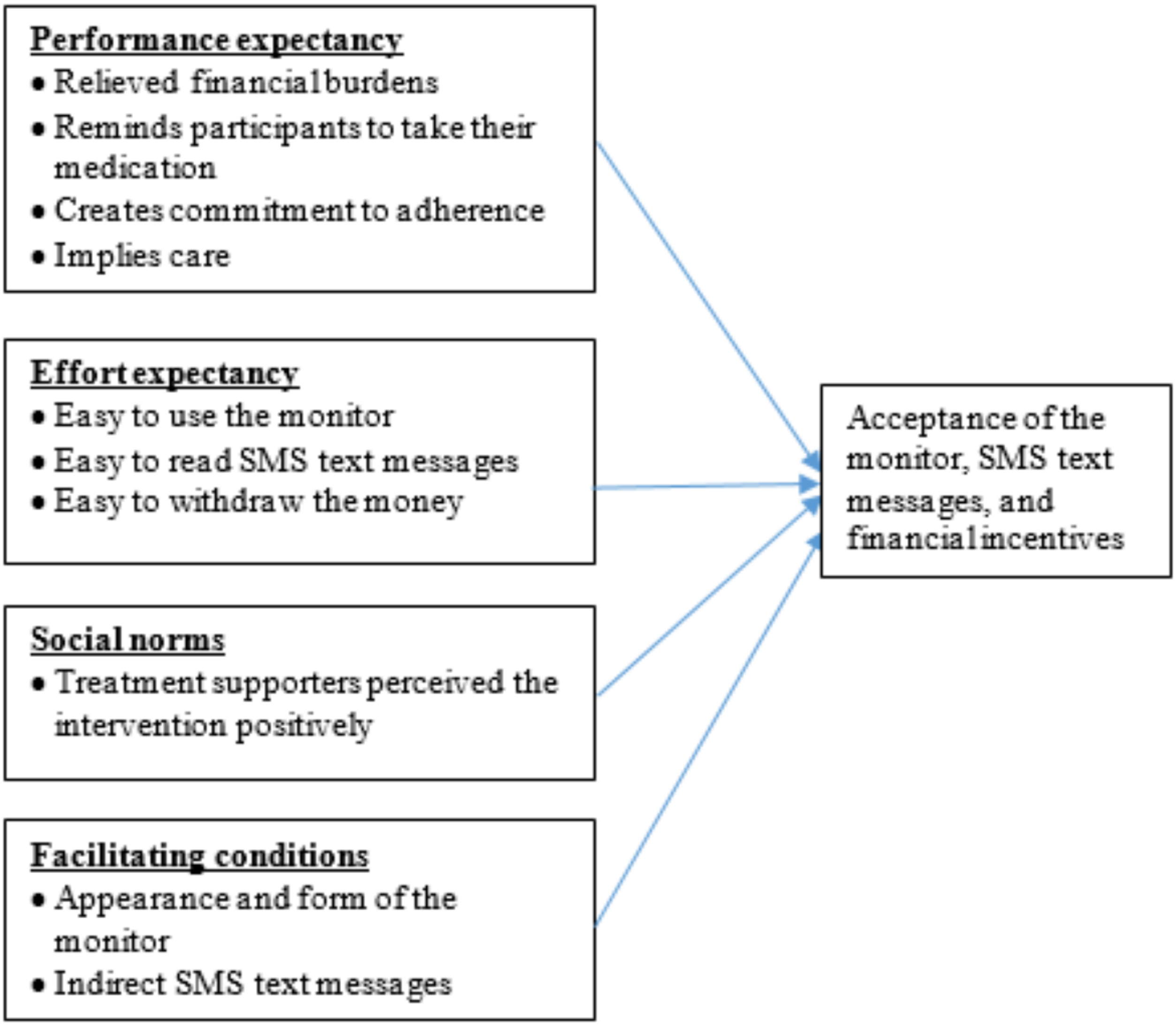
Table 3. Feasibility and acceptability of the My Mobile Wallet intervention.

Feasibility and acceptability of SMS text messages	Values
Preference of SMS text message reminders versus device-based alarms (recorded at recruitment; n=40), n (%)	
Participants who opted to be reminded by SMS text message reminders only	0 (0)
Participants who opted to be reminded by Wisepill device alarm only	2 (5)
Participants who opted for SMS text message reminders plus Wisepill device alarm	38 (95)
SMS text message reminders (automatically ascertained from the intervention; N=7064), n (%)	
Total number of SMS text message reminders sent	6880 (97.4)
SMS text message reminders not sent due to technical challenges (eg, poor network)	184 (2.6)
Mobile money incentives (automatically ascertained from the intervention; N=40), n (%)	
Transport incentives not sent	0 (0)
Transport incentives sent unnecessarily	0 (0)
Adherence incentives not sent	0 (0)
Adherence incentives sent unnecessarily	0 (0)
Feasibility and acceptability of the real-time adherence monitor, n (%)	
Data loss due to technical issues with the real-time monitors (days when the monitor was not opened; automatically ascertained from the intervention; N=5776)	94 (1.63)
Device malfunction (N=40; devices that malfunctioned and were replaced)	0 (0)
Devices successfully returned by participants (N=40; the device used by the participant who was lost to follow-up was later recovered from her treatment supporter)	40 (100)

Interview Results: Intervention Acceptability

Acceptability is presented following the constructs of the UTAUT model (Figure 3) of the performance expectancy, effort expectancy, social norms, and facilitating conditions associated with the intervention.

Figure 3. Organization of the qualitative data on acceptability following the unified theory of acceptance and use of technology (UTAUT) model.



Performance Expectancy or Perceived Usefulness

Mobile Money Incentives Relieve Participants of Tuberculosis-Related Financial Burdens

Before being enrolled in the study, some of the participants struggled to meet the basic costs of their tuberculosis care, including transport to the tuberculosis clinic and the cost of food and drinks needed to take the medication. These were mainly participants with no regular source of income, as well as those staying far from the tuberculosis clinic. They described relying on begging and borrowing money to meet their tuberculosis treatment costs. Unfortunate instances where begging and attempts to borrow money were not successful resulted in missed clinic appointments for pill refills due to lack of transport, which consequently resulted in missed medication. Others reported not taking their medication due to lack of food and drinks because they feared becoming weaker after taking their medication on an empty stomach. Participants reported that the mobile money transport incentives enabled them to meet the cost of transport to the clinic (eg, for pill refills) as well as the costs of meeting the basic tuberculosis treatment needs, such as food and drinks for taking their medication, thereby relieving them (as their treatment supporters) of the financial burdens associated with tuberculosis treatment and care. One participant stated as follows:

> *Before you started sending me mobile money, there were times when I would request people to borrow me money or help me with transport to the hospital, but sometimes they would also not be in a position to give me money, so I would miss picking my medication from the hospital...I stopped working when I got sick. I stay far from the hospital.This study helped a lot by sending me [money for] transport to the hospital so that I don't miss taking my medication on time.* [Male patient, aged 61 years]

Participants narrated how medication adherence incentives encouraged them to take their medication on time to meet the monthly target of ≥90% medication adherence, which in turn helped them meet the cost of the basic food and drinks they needed for taking medication:

> *I knew that I would be given money after getting ≥90% medication adherence, so I made sure that I was taking my medication well in order to be sent the*

> *money. This disease made me too weak to work; yet, I needed money to buy food and porridge.* [Female patient, aged 51 years]

> *Whenever I would receive a message about my adherence percentage, and it's below, it would motivate me to be serious so that next time I don't miss out again.* [Male patient, aged 39 years]

SMS Text Message Reminders and the Real-Time Adherence Monitor Enabled Participants to Take Their Medication on Time

Participants reported that the SMS text message reminders and monitor-based alarms enabled them to take their medication on time. These technologies served as medication reminders, thereby addressing forgetfulness, which was common in participants who were not yet used to taking medication regularly. They were also useful for busy participants who could easily forget their medication-taking time due to other competing demands on their time. Participants reported being able to make the necessary preparations (eg, obtaining food and drinks and going back home in case the medication had been left behind at home) after receiving the daily SMS text message reminders (which were sent 30 minutes before their medication-taking time), thereby enabling them to take their medication on time:

> *The SMS [text message] reminders were very helpful because I was still learning how to take medicine in time, so they helped me in getting used to medication taking because they were coming every day, so my body eventually got used to the time.* [Female patient, aged 35 years]

> *At times, I would get too busy at my video library and forget taking my medication, but whenever I would receive the message, I would close the business immediately, go home, eat some food, prepare a drink, and use it to take my medication on time.* [Male patient, aged 27 years]

Using a Real-Time Adherence Monitor Creates Commitment to Medication Adherence

Participants perceived real-time adherence monitoring as "being watched" while taking their medication. This perception was welcomed and encouraged them to take their medication on time to demonstrate their commitment to the health care providers who they felt were concerned about their health and would not be happy with nonadherence:

> *When I started using the device, I felt touched knowing that there are people who are concerned about my life to the extent of using the device to watch me take medication yet they are not even my relatives or friends or people I knew before. This gave me morale to swallow my medication to play my part especially because they would be seeing whether or not I am taking my medication and they would probably feel bad if I miss taking [it].* [Female patient, aged 40 years]

> *Whenever I felt like not taking the medication, I always got motivated me to take medication because I knew that you people cared for me so much to the extent that you gave me this monitor and kept texting me to remind me to take medication and even sent me money to go to the clinic. I felt encouraged because you were really interested in seeing my health condition improve.* [Female patient, aged 26 years]

Although monitoring created commitment to medication adherence, it is noteworthy that the primary motivation for taking medication on time as reported by participants was the need to recover their good health and live longer. A participant stated as follows:

> *Whenever I saw it [the device], I knew it was going to report me, so I chose to take the commitment of swallowing the medication. But, the main issue was, I really needed to recover from this disease because I loved my life and wanted to save it by getting well as soon as possible.* [Male patient, aged 57 years]

Receiving Financial Incentives and Reminders and Being Monitored Implies Care

Overview

Receiving the mobile money incentives and SMS text message or alarm reminders and being monitored via the real-time adherence monitor was perceived by participants as signs that the health care providers cared about them, which eventually created emotional support and a sense of connectedness that countered depressive feelings. A participant describes how she changed her mind about committing suicide as a result of using the technologies:

> *I was about to stop taking the medication and die because I developed self-rejection. I was in pain, and I had no one to help me, but you people encouraged me to take the medication when you put me in this study and started sending me texts, alarms, and mobile money to support me to take my medication. I dropped the idea of suicide because you people cared for me and loved me even more than I loved myself. Thank you for saving my life because I would be dead by now.* [Female patient, aged 33 years]

Effort Expectancy or Perceived Ease of Use: The Intervention Was Easy to Use

After the participants' initial orientation to using the real-time adherence monitor, they found it easy to use for taking their medication. Participants, including a few (3/40, 8%) who never went to school, reported finding it easy to read the SMS text messages sent to them. In addition, they reported that it was easy to withdraw money from mobile money agents because the agents are readily available:

> *It was very easy to use the container [the real-time adherence monitor]; you open it the same way a food box is [opened], put your medication [in], and start using it; that is all. I did not have to charge it or do any other thing with it.* [Male patient, aged 42 years]

> *Although I did not go to school, I can read messages written in my local language, so, reading the messages on [the mobile] phone was not a problem at all.* [Female patient, aged 28 years]

> *There are so many mobile money agents around. It was easy for me to withdraw my money from them.* [Male patient, aged 35 years]

Social Norms or Other People's Perceptions of the Intervention

Positive Perceptions From Treatment Supporters

Participants reported that their treatment supporters approve of them using the intervention to support their medication adherence to get well. In addition, participants stated that because of the financial incentives, their treatment supporters were relieved of the financial burden of having to take care of the financial needs of the participants:

> *When my wife saw the container and the messages, she was happy knowing that I was being supported by the hospital; she believed the support would help [me] get well soon. It was also a relief for her when you sent me money; I stopped working when I got sick, and before your assistance, I was only relying on getting money from her small shop for transport to the hospital and getting other basic needs like food.* [Male patient, aged 32 years]

Possibility of Inappropriate Use of the Financial Incentives

One participant reported how her husband initially misappropriated the financial incentives intended to pay for her transport to the clinic. Although the participant did not miss visiting the clinic, she had to keep begging her husband for money for transport to the clinic. Sometimes she would have to walk part of the distance due to insufficient transport funds provided by her husband:

> *My husband never wants me to own any money and always forces [me] to give him my money. So, whenever you would send me money on the [mobile] phone, he would force me to give him the whole of it. I would then have to go through the hassle of begging him to give me the money for my hospital visit, and sometimes the money he would give me for transport would not be enough.... But after giving you my new SIM card [details], which he did not know [about], I started receiving and using the money for transport to the hospital.* [Female patient, aged 33 years]

Facilitating Conditions: Appearance and Form of the Monitor

Participants reported that they liked the appearance and form of the real-time adherence monitor. Specifically, they liked the monitor's design, which resembled a food box, and its size, which they thought was reasonable because it accommodated all their pills; the absence of tuberculosis-related labels that could link them to the disease; and the hard outer cover that kept their medicines safe and clean, all of which motivated them to use the monitor:

> *The container looks like a food bowl, so people can easily think you have carrying some food in it; it is also big enough to carry all my medicine, and has no any TB-related word.* [Female patient, aged 35 years]

Indirect SMS Text Messages

To avoid unwanted status disclosure, participants preferred SMS text messages that could not easily link them to tuberculosis:

> *I chose the message "come and eat" because for me I knew what it reminded me to eat, but other people even if they saw it on my [mobile] phone would not know what I was going to eat.* [Male patient, aged 38 years]

Discussion

Principal Findings

Drawing on the UTAUT model, this paper describes the feasibility and acceptability of My Mobile Wallet, a DAT intervention composed of a real-time adherence monitor, SMS text message reminders, and mobile money incentives (WiseCash) for tuberculosis medication adherence in rural southwestern Uganda. Generally, we found that the intervention was technically feasible because it functioned as expected. All participants reported that it was easy for them to use the intervention; they could access and read the mobile money SMS text messages as well as the medication reminders, and they were able to open the Wisepill device to retrieve their medication. Participants reported receiving the mobile money incentives for transport to the clinic and the medication adherence incentives as expected. The SMS text messages and real-time adherence monitor also worked as expected: the SMS text messages were sent as planned, the transmission of adherence data from the monitor worked well, and no monitor malfunctioned for the entire period of the study.

Concerning acceptability, participants reported being relieved of tuberculosis-related financial burdens as a result of receiving the mobile money incentives. SMS text message reminders and real-time monitor-based alarms reminded participants to take their medication on time. Daily SMS text message reminders were preferred to reminders triggered by missed doses. Participants' preference for daily SMS text message reminders even in the treatment continuation phase (from month 4 onward) was surprising because one would assume that during this phase, they were nearly getting used to taking their medication and therefore did not require to be reminded daily. Patients' preference for daily SMS text message reminders (for taking their medications) over weekly SMS text message reminders was also reported in our previous tuberculosis study [4] and HIV study [25]. As tuberculosis medications are taken daily, daily SMS text message reminders are preferred because they are aligned with the medication-taking frequency.

Participants perceived real-time adherence monitoring as "being watched" while taking their medication, which was welcomed and encouraged them to take their medication on time to demonstrate their commitment. Receiving the mobile money incentives and SMS text message or alarm reminders and being monitored via the real-time adherence monitor was perceived by participants as signs that the health care providers cared about them. Their experiences with the intervention eventually created emotional support and a sense of connectedness that countered depressive feelings among the participants. Inappropriate use of the mobile money transport incentives was reported only rarely.

Limitations

The main limitation of this study is the possibility of social desirability bias in the data collected from interviews and surveys. The mobile money incentives in particular may have influenced participants, given the prevalence of poverty-based structural barriers to tuberculosis treatment in Uganda, a low-resource setting.

Comparison With Prior Work

We are not aware of any study that reports the impact of a DAT intervention composed of a real-time adherence monitor, SMS text message reminders, and mobile money incentives on tuberculosis medication adherence. However, it should be noted that some studies using some components of this intervention exist; for instance, a recent systematic review and meta-analysis on cash interventions to improve tuberculosis outcomes concluded that these interventions could improve tuberculosis treatment success and completion among patients in low- and middle-income countries [10]; in this review, only 1 randomized control trial in Peru [26] was identified, and the authors of the review noted that the evidence is still weak. In addition, the use of face-to-face cash transfers or transport vouchers as incentives has been reported to be acceptable in facilitating adherence to tuberculosis diagnostic evaluation in Uganda [27]. Furthermore, receiving monthly financial incentives face-to-face enabled patients living with tuberculosis in Nigeria to purchase food and get transport to the clinic [28], while, in Uganda, receiving a one-time cash transfer upon sputum submission supported tuberculosis testing completion among patients [29]. Although the receipt of unconditional cash transfers through a direct benefit scheme supported registered patients living with tuberculosis to meet their nutrition requirements in India, the scheme had no significant effect on treatment outcome [30]. In our study, individuals with no regular source of income and those living far from the clinic benefited most from the mobile money incentives; participants used the transport incentives to cover the cost of transport to the clinic, while they used the financial incentives conditional on high medication adherence to buy food and drinks required to take their medication. This approach could potentially address the financial insecurities that continue to constrain medication adherence [4]. In Uganda, 53% of the patients living with tuberculosis take loans or sell property to meet the costs of their tuberculosis care [8]. Although there is limited research in this area, our study indicates that mobile money incentives can potentially relieve the financial burden that tuberculosis places not only on patients but also on their treatment supporters. An incentive as small as US $1 can increase the tuberculosis cure rate and reduce treatment loss to follow-up in Uganda [31]. Although this study estimated an average transport cost of US $8 (based on the COVID-19–pandemic-induced transport cost increases) and provided the same amount for transport to all participants, using GPS information to estimate and provide transport costs according to each participant's distance from their home to the clinic could be a better option.

The reported practice of a husband taking the mobile money transport incentive from the wife shows the effects of poverty as well as the complexity of implementing mobile money incentives in low-resource settings and cultures where some people still believe in male-exclusive ownership of resources [32]. This scenario could result in an inappropriate use of the incentives, thus limiting the impact of the intervention. An inclusive approach that engages men in the implementation of such an intervention (eg, through awareness creation) might mitigate the risk of the incentives being misappropriated.

This is the first study to report on the feasibility of real-time monitoring linked to SMS text message reminders and mobile money incentives for tuberculosis medication adherence. In the same setting, we had previously reported that using SMS text message reminders linked to real-time monitoring is feasible and acceptable for supporting tuberculosis medication adherence [6]. This study provides insights regarding the integration of financial incentives with these technologies to support access to tuberculosis medication from the hospital and motivate medication adherence.

Participants' medication adherence ascertained from the real-time adherence monitor was quite high. The receipt of financial incentives that was conditional upon a particular adherence target (≥90%) resulted in participants taking their medication on time in order to hit the target for financial incentives. In addition, participants' awareness of the fact that their medication adherence was being watched or monitored through the real-time monitor motivated them to take their medication well in order not to disappoint those monitoring them. Notably, the reported adherence was ascertained from the monitor in the form of the monitor being opened, which was used as a proxy for medication taking. Overall, the real-time monitoring approach can potentially be more reliable than participant self-reports, which are highly subject to social desirability bias. However, although it was not reported in our findings, instances of opening the monitor without taking medication (such as accidental openings or opening the monitor to increase the chances of getting incentives) may constrain the feasibility of the intervention. In the ongoing phase of the study, we are supplementing the real-time adherence monitoring with hair analysis (assessing tuberculosis drug levels in participants' hair) to improve objectivity.

Although a few SMS text messages (184/7064, 2.6%) and some adherence data (94/5776, 1.63%) could not be sent by the SMS text messaging application and the real-time adherence monitor, respectively, mainly due to technical issues such as poor network, the intervention was otherwise feasible. The feasibility of this intervention could be attributed to the rapid evolution and adoption of mobile phone technologies in Uganda, including among populations based in rural areas and considered economically marginalized [33]. The applications for the SMS text message reminders and mobile money incentives were tailored from the existing mobile phone infrastructure, which likely facilitated use by participants who were already familiar using SMS text messaging and mobile money services in their regular routines. By leveraging the existing mobile phone infrastructure, these technologies can potentially bridge the current gaps in access to health care services between economically advantaged populations and populations considered disadvantaged, consequently contributing to equitable access to health care. In addition, the fact that all participants owned personal mobile phones, had the ability to read SMS text message reminders and mobile money SMS text messages, and had reliable mobile network (per the enrolment criteria) could have contributed to the feasibility of the intervention. Different feasibility results may be yielded if this intervention is implemented in populations with fewer resources or in privileged populations.

Concerning acceptability, participants perceived the intervention's functionalities of sending timely medication-taking reminders (through SMS text messages and monitor-based alarms), financially supporting medication adherence (through mobile money incentives), and monitoring medication taking (through the adherence monitor) as supportive and taking care of them. For the participants, this perception created a sense of connectedness with health care providers and countered depressive feelings after their tuberculosis diagnosis. It also encouraged them to adhere to taking their medication as a way of appreciating the care and proving their commitment to taking an active role in their own health with the ultimate goal of regaining their health. Such emotional support can also potentially empower patients to cope with the stigma and discrimination that are often associated with tuberculosis [34]. Importantly, there is evidence that emotional and social support can improve tuberculosis treatment success rates [35]. The use of various tuberculosis medication adherence technologies (including SMS text messages and real-time adherence monitors) was perceived by participants to reduce visits to clinics and increase access to social supporters in a variety of settings [36].

In South Africa, the use of a real-time adherence monitor (Wisepill evriMED1000) was acceptable for prompting a stepwise differentiated care approach for tuberculosis medication adherence, composed of SMS text messages, telephone calls, home visits, and motivational counseling, in response to missed doses ascertained from the monitor [37]. Other studies referencing patient experiences of using real-time adherence monitoring linked to SMS text message reminders for antiretroviral adherence support among people living with HIV in Uganda also reported perceptions of being cared for as a result of using the technologies [25]. Although there are differences between HIV and tuberculosis, they are both diseases of poverty, and result in stigma, and discrimination. The findings regarding the adherence monitor and the SMS text message aspects of the intervention were indeed similar in this and another [25] study, indicating the strong potential of the intervention in this and potentially other similar settings.

Conclusions

In sum, we found the My Mobile Wallet intervention (composed of real-time adherence monitoring linked to SMS text message reminders and mobile money incentives) for tuberculosis medication adherence to be feasible and acceptable in a low-resource setting where poverty-based structural barriers heavily constrain tuberculosis treatment and care. The intervention worked as expected, and participants found it easy to use. The intervention relieved participants of the burden of tuberculosis treatment costs, reminded them to take their medication on time, and provided emotional support that made them feel connected to care.

On the basis of the findings from this study, we are now planning a randomized controlled trial (registered on ClinicalTrials.gov; NCT05656287) for assessing the full-scale feasibility, acceptability, and impact of My Mobile Wallet.

Acknowledgments

The authors would like to acknowledge the contributions of Wisepill Technologies, Yo Uganda Limited (Yo!), and the study participants. The study was funded by a grant from the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health (R21HD107985). JH is supported through a grant from the National Institute of Mental Health (K24MH114732). EA is supported by the Fogarty International Center of the National Institutes of Health (K43TWO11004).

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. Yoeli E, Rathauser J, Rand D. The resurgence of tuberculosis is behavioral, not medical. Nudges can fix it. STAT. URL: https://www.statnews.com/2019/10/25/tuberculosis-behavior-text-based-nudges/ [accessed 2024-03-03]
2. Boutilier JJ, Yoeli E, Rathauser J, Owiti P, Subbaraman R, Jónasson JO. Can digital adherence technologies reduce inequity in tuberculosis treatment success? Evidence from a randomised controlled trial. BMJ Glob Health 2022 Dec 01;7(12):e010512 [FREE Full text] [doi: 10.1136/bmjgh-2022-010512] [Medline: 36455988]

3. Rao JS, Diwan V, Kumar AA, Varghese SS, Sharma U, Purohit M, et al. Acceptability of video observed treatment vs. directly observed treatment for tuberculosis: a comparative analysis between south and central India. Wellcome Open Res 2022 May 13;7:152. [doi: 10.12688/wellcomeopenres.17865.1]
4. Musiimenta A, Tumuhimbise W, Mugaba AT, Muzoora C, Armstrong-Hough M, Bangsberg D, et al. Digital monitoring technologies could enhance tuberculosis medication adherence in Uganda: mixed methods study. J Clin Tuberc Other Mycobact Dis 2019 Dec;17:100119 [FREE Full text] [doi: 10.1016/j.jctube.2019.100119] [Medline: 31788561]
5. Barter DM, Agboola SO, Murray MB, Bärnighausen T. Tuberculosis and poverty: the contribution of patient costs in sub-Saharan Africa--a systematic review. BMC Public Health 2012 Nov 14;12(1):980 [FREE Full text] [doi: 10.1186/1471-2458-12-980] [Medline: 23150901]
6. Musiimenta A, Tumuhimbise W, Atukunda EC, Mugaba AT, Musinguzi N, Muzoora C, et al. The feasibility, acceptability, and preliminary impact of real-time monitors and SMS on tuberculosis medication adherence in southwestern Uganda: findings from a mixed methods pilot randomized controlled trial. PLOS Glob Public Health 2023 Dec 5;3(12):e0001813 [FREE Full text] [doi: 10.1371/journal.pgph.0001813] [Medline: 38051699]
7. The Uganda tuberculosis communication strategy. Ministry of Health, Uganda. 2010. URL: http://library.health.go.ug/sites/default/files/resources/The%20Uganda%20Tuberculosis%20Communication%20Strategy.pdf [accessed 2024-04-29]
8. Direct and indirect costs due to tuberculosis and proportion of tuberculosis-affected households experiencing catastrophic costs due to TB in Uganda. Ministry of Health, Uganda. 2019. URL: https://www.health.go.ug/cause/direct-and-indirect-costs-due-to-tuberculosis-and-proportion-of-tuberculosis-affected-households-experiencing-catastrophic-costs-due-to-tb-in-uganda-febrauary-2019/ [accessed 2024-04-29]
9. Implementing the end TB strategy: the essentials (No. WHO/HTM/TB/2015.31). World Health Organization. 2015. URL: https://www.who.int/publications/i/item/9789240065093 [accessed 2024-04-29]
10. Richterman A, Steer-Massaro J, Jarolimova J, Luong Nguyen LB, Werdenberg J, Ivers LC. Cash interventions to improve clinical outcomes for pulmonary tuberculosis: systematic review and meta-analysis. Bull World Health Organ 2018 Jul 01;96(7):471-483 [FREE Full text] [doi: 10.2471/BLT.18.208959] [Medline: 29962550]
11. van Bastelaer T, Woodman B, Chatterji M, Long D. Saving for delivery through technology: an inquiry into an electronic commitment savings product for maternal healthcare in Kenya. Int J Health Plann Manage 2016 Oct 28;31(4):e254-e272. [doi: 10.1002/hpm.2292] [Medline: 25920603]
12. Mobile money for health case study compendium. United States Agency for International Development. 2015. URL: https://www.hfgproject.org/wp-content/uploads/2015/10/HFG-Mobile-Money-Compendium_October-2015.pdf [accessed 2024-04-29]
13. Musiimenta A, Tumuhimbise W, Atukunda E, Mugaba A, Linnemayr S, Haberer J. Digital adherence technologies and mobile money incentives for management of tuberculosis medication among people living with tuberculosis: mixed methods formative study. JMIR Form Res 2023 Apr 12;7:e45301 [FREE Full text] [doi: 10.2196/45301] [Medline: 37043263]
14. Lacroze E, Bärnighausen T, De Neve JW, Vollmer S, Ratsimbazafy RM, Emmrich PM, et al. The 4MOTHERS trial of the impact of a mobile money-based intervention on maternal and neonatal health outcomes in Madagascar: study protocol of a cluster-randomized hybrid effectiveness-implementation trial. Trials 2021 Oct 21;22(1):725 [FREE Full text] [doi: 10.1186/s13063-021-05694-8] [Medline: 34674741]
15. Brown T. Change by Design: How Design Thinking Transforms Organizations and Inspires Innovation. New York, NY: HarperCollins; 2019.
16. Imperial MZ, Nahid P, Phillips PP, Davies GR, Fielding K, Hanna D, et al. A patient-level pooled analysis of treatment-shortening regimens for drug-susceptible pulmonary tuberculosis. Nat Med 2018 Nov 5;24(11):1708-1715 [FREE Full text] [doi: 10.1038/s41591-018-0224-2] [Medline: 30397355]
17. Broadhead WE, Gehlbach SH, de Gruy FV, Kaplan BH. The Duke-UNC functional social support questionnaire: measurement of social support in family medicine patients. Med Care 1988 Jul;26(7):709-723. [doi: 10.1097/00005650-198807000-00006] [Medline: 3393031]
18. Filmer D, Pritchett LH. Estimating wealth effects without expenditure data-or tears: an application to educational enrollments in states of India. Demography 2001 Feb;38(1):115. [doi: 10.2307/3088292]
19. Bolton P, Wilk CM, Ndogoni L. Assessment of depression prevalence in rural Uganda using symptom and function criteria. Soc Psychiatry Psychiatr Epidemiol 2004 Jun;39(6):442-447. [doi: 10.1007/s00127-004-0763-3] [Medline: 15205728]
20. Coates J, Swindale A, Bilinsky P. Household food insecurity access scale (HFIAS) for measurement of food access: indicator guide: version 3. FANTA Project. 2007. URL: https://www.fantaproject.org/sites/default/files/resources/HFIAS_ENG_v3_Aug07.pdf [accessed 2024-04-29]
21. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory care quality improvement project (ACQUIP). Alcohol use disorders identification test. Arch Intern Med 1998 Sep 14;158(16):1789-1795. [doi: 10.1001/archinte.158.16.1789] [Medline: 9738608]
22. Kalichman SC, Simbayi LC, Cloete A, Mthembu PP, Mkhonta RN, Ginindza T. Measuring AIDS stigmas in people living with HIV/AIDS: the Internalized AIDS-related stigma scale. AIDS Care 2009 Jan 11;21(1):87-93. [doi: 10.1080/09540120802032627] [Medline: 19085224]

23. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
24. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec 16;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
25. Ware NC, Pisarski EE, Tam M, Wyatt MA, Atukunda E, Musiimenta A, et al. The Meanings in the messages: how SMS reminders and real-time adherence monitoring improve antiretroviral therapy adherence in rural Uganda. AIDS 2016 May 15;30(8):1287-1294 [FREE Full text] [doi: 10.1097/QAD.0000000000001035] [Medline: 26807967]
26. Wingfield T, Tovar MA, Huff D, Boccia D, Montoya R, Ramos E, et al. Socioeconomic support to improve initiation of tuberculosis preventive therapy and increase tuberculosis treatment success in Peru: a household-randomised, controlled evaluation. Lancet 2017 Feb;389:S16. [doi: 10.1016/s0140-6736(17)30412-9]
27. Kadota JL, Nabwire S, Nalugwa T, White JS, Cattamanchi A, Katamba A, et al. Patient perspectives and willingness to accept incentives for tuberculosis diagnostic evaluation in Uganda. Value Health Reg Issues 2021 Sep;25:48-56 [FREE Full text] [doi: 10.1016/j.vhri.2020.12.005] [Medline: 33773327]
28. Ukwaja KN, Alobu I, Mustapha G, Onazi O, Oshi DC. 'Sustaining the DOTS': stakeholders' experience of a social protection intervention for TB in Nigeria. Int Health 2017 Mar 01;9(2):112-117. [doi: 10.1093/inthealth/ihx001] [Medline: 28204499]
29. Shete PB, Kadota JL, Nanyunja G, Namale C, Nalugwa T, Oyuku D, et al. Evaluating the impact of cash transfers on tuberculosis (ExaCT TB): a stepped wedge cluster randomised controlled trial. ERJ Open Res 2023 May 01;9(3):00182-2023 [FREE Full text] [doi: 10.1183/23120541.00182-2023] [Medline: 37342088]
30. Dave JD, Rupani MP. Does direct benefit transfer improve outcomes among people with tuberculosis? - A mixed-methods study on the need for a review of the cash transfer policy in India. Int J Health Policy Manag 2022 Dec 06;11(11):2552-2562 [FREE Full text] [doi: 10.34172/ijhpm.2022.5784] [Medline: 35174676]
31. Baluku JB, Nakazibwe B, Twinomugisha B, Najjuuko R, Isabella N, Nassozi S, et al. One dollar incentive improves tuberculosis treatment outcomes in programmatic settings in rural Uganda. Sci Rep 2021 Sep 29;11(1):19346 [FREE Full text] [doi: 10.1038/s41598-021-98770-7] [Medline: 34588552]
32. Kane EW. Men's and women's beliefs about gender inequality: family ties, dependence, and agreement. Sociol Forum 1998 Dec;13(4):611-637.
33. The Collaboration on International ICT Policy for East and Southern Africa. National information technology survey report 2017/2018. National Information Technology Authority-Uganda. 2018. URL: https://s3-eu-west-1.amazonaws.com/s3.sourceafrica.net/documents/120158/National-IT-Survey-April-10th.pdf [accessed 2024-04-29]
34. Musiimenta A, Tumuhimbise W, Atukunda EC, Mugaba AT, Muzoora C, Armstrong-Hough M, et al. Mobile health technologies may be acceptable tools for providing social support to tuberculosis patients in rural Uganda: a parallel mixed-method study. Tuberc Res Treat 2020 Jan 07;2020:7401045-7401048 [FREE Full text] [doi: 10.1155/2020/7401045] [Medline: 31969998]
35. Yin J, Wang X, Zhou L, Wei X. The relationship between social support, treatment interruption and treatment outcome in patients with multidrug-resistant tuberculosis in China: a mixed-methods study. Trop Med Int Health 2018 Jun 16;23(6):668-677 [FREE Full text] [doi: 10.1111/tmi.13066] [Medline: 29691959]
36. Guzman K, Crowder R, Leddy A, Maraba N, Jennings L, Ahmed S, et al. Acceptability and feasibility of digital adherence technologies for drug-susceptible tuberculosis treatment supervision: a meta-analysis of implementation feedback. PLOS Digit Health 2023 Aug 15;2(8):e0000322 [FREE Full text] [doi: 10.1371/journal.pdig.0000322] [Medline: 37582066]
37. Mukora R, Ahumah B, Maraba N, Orrell C, Jennings L, Naidoo P, et al. Acceptability of using the medication monitor and experience of a differentiated care approach for TB treatment adherence among people living with TB in South Africa. PLOS Glob Public Health 2023 Oct 27;3(10):e0001885 [FREE Full text] [doi: 10.1371/journal.pgph.0001885] [Medline: 37889875]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
DAT: digital adherence technology
MRRH: Mbarara Regional Referral Hospital
UTAUT: unified theory of acceptance and use of technology

Edited by A Kushniruk; submitted 07.04.23; peer-reviewed by B Noort, J Wang, B Thies; comments to author 12.08.23; revised version received 30.10.23; accepted 07.04.24; published 31.05.24.

Please cite as:
Musiimenta A, Tumuhimbise W, Atukunda E, Mugaba A, Linnemayr S, Haberer J
Digital Adherence Technologies Linked to Mobile Money Incentives for Medication Adherence Among People Living With Tuberculosis: Mixed Methods Feasibility and Acceptability Study
JMIR Hum Factors 2024;11:e47996
URL: https://humanfactors.jmir.org/2024/1/e47996
doi:10.2196/47996
PMID:38819905

Original Paper

A Multicomponent Intervention (POSSIBLE) to Improve Perceived Risk for HIV Among Black Sexual Minority Men: Feasibility and Preliminary Effectiveness Pilot Study

Derek T Dangerfield II[1], PhD; Janeane N Anderson[2], MPH, PhD; Charleen Wylie[1]; Ricky Bluthenthal[3], PhD; Chris Beyrer[4], MD; Jason E Farley[5], RN, MPH, MSN, PhD

[1]Milken Institute School of Public Health, George Washington University, Washington, DC, United States
[2]University of Tennessee Health Science Center, Memphis, TN, United States
[3]Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, United States
[4]Duke Global Health Institute, Durham, NC, United States
[5]Johns Hopkins School of Nursing, Baltimore, MD, United States

Corresponding Author:
Derek T Dangerfield II, PhD
Milken Institute School of Public Health
George Washington University
950 New Hampshire Avenue NW #308
Washington, DC, 20037
United States
Phone: 1 6673555188
Email: ddanger2@gwu.edu

Abstract

Background: Increased pre-exposure prophylaxis (PrEP) use is urgently needed to substantially decrease HIV incidence among Black sexual minority men. Low perceived risk for HIV (PRH) is a key unaddressed PrEP barrier for Black sexual minority men. Peers and smartphone apps are popular intervention tools to promote community health behaviors, but few studies have used these together in a multicomponent strategy. Therefore, we designed a multicomponent intervention called POSSIBLE that used an existing smartphone app called PrEPme (Emocha Mobile Health, Inc) and a peer change agent (PCA) to increase PRH as a gateway to PrEP.

Objective: This paper aims to describe the feasibility and preliminary impact of POSSIBLE on PRH and willingness to accept a PrEP referral among Black sexual minority men.

Methods: POSSIBLE was a theoretically guided, single-group, 2-session pilot study conducted among Black sexual minority men from Baltimore, Maryland between 2019 and 2021 (N=69). POSSIBLE integrated a PCA and the PrEPme app that allows users to self-monitor sexual risk behaviors and chat with the in-app community health worker to obtain PrEP service information. PRH was assessed using the 8-item PRH scale before and after baseline and follow-up study visits. At the end of each study visit, the PCA referred interested individuals to the community health worker to learn more about PrEP service options.

Results: The average age of participants was 32.5 (SD 8.1, range 19-62) years. In total, 55 (80%) participants were retained for follow-up at month 1. After baseline sessions, 29 (42%) participants were willing to be referred to PrEP services, 20 (69%) of those confirmed scheduled appointments with PrEP care teams. There were no statistically significant differences in PRH between baseline and follow-up visits (t_{122}=–1.36; P=.17).

Conclusions: We observed no statistically significant improvement in PRH between baseline and month 1. However, given the high retention rate and acceptability, POSSIBLE may be feasible to implement. Future research should test a statistically powered peer-based approach on PrEP initiation among Black sexual minority men.

Trial Registration: ClinicalTrials.gov NCT04533386; https://clinicaltrials.gov/study/NCT04533386

(JMIR Hum Factors 2024;11:e54739) doi:10.2196/54739

KEYWORDS

pre-exposure prophylaxis; PrEP; sexual health; peers; apps; community; mobile phone; HIV; sexual minority; minority communities; minority; Black; African American; peers; patient education; self-monitoring; treatment adherence; treatment participation; community; community health; mobile health; digital health; digital technology; digital interventions; smartphones

Introduction

Increased pre-exposure prophylaxis (PrEP) use is urgently needed to substantially decrease HIV incidence among Black sexual minority men who have an estimated 50% lifetime risk of HIV acquisition [1]. Data show racial inequities in uptake and adherence among sexual minority men who meet PrEP indications [2]. Socioecological factors such as side effect concerns, stigma, low health care access, and poor clinical experiences including patient-clinician communication remain substantial PrEP barriers for this community [3-5]. Low perceived risk for HIV (PRH) is also a key unaddressed PrEP barrier [6-9]. Some Black sexual minority men have low concerns about HIV acquisition because they think their current behaviors are lower risk than their past or their peers' behaviors and they do not associate HIV status with quality of life [6,10]. Other reasons for low PRH include being in a monogamous relationship and limited sexual activity [6,9,11]. However, low PRH inadequately reflects objective risks and disease severity for Black sexual minority men, given the high-community HIV incidence, suboptimal HIV care outcomes, and negative health consequences of long-term infection [7,8]. Therefore, multicomponent interventions that address PRH and other known barriers are needed to improve PrEP uptake.

Some HIV prevention interventions leverage in-group members as peer change agents (PCAs) to disseminate health-related information within the community for behavior change [12-15]. Peers are considered a valuable resource in marginalized communities to obtain health information, discuss taboo experiences, and help group members understand why behavior change could be beneficial [6,16,17]. Peers can facilitate behavior change because they have similar experiences, can address social barriers, and can improve health literacy [12,13,16,18]. PCAs are uniquely positioned to influence health behaviors because their roles as community members, patients, and health care paraprofessionals can build trust and lead others to credible information or clarify the information [17]. PCAs have improved behavioral health [12], HIV testing [19], medication adherence, and PrEP [13-15] for HIV prevention and could be effective interventionists among Black sexual minority men.

Other interventions have used smartphone apps as electronic diaries to reduce sexual risks through self-monitoring behaviors, which facilitates reflection [20-22]. Technology-based interventions could also be effective for Black sexual minority men because many of them use apps and other mobile devices for several purposes, including partner seeking, social network development, and health information [23,24]. Using apps for interventions could help Black sexual minority men circumvent the social and structural barriers to PrEP such as perceived judgment, stigma, and discrimination from clinicians. Since peers and smartphone apps are typically used independently in interventions, they could have a stronger impact if combined into a multicomponent health communication strategy because they could hypothetically reduce socioecological barriers for Black sexual minority men simultaneously.

Motivational interviewing (MI) is a communication approach in which a professional collaborates with individuals to activate their motivations to change behavior [17,25]. Some studies use "motivational interview consistent" interventions for HIV prevention because of their cost-effectiveness, brevity, and use of client interests for behavior change [26]. However, existing MI-based interventions to improve PRH or PrEP among Black sexual minority men are limited [18]. Additionally, factors known to drive PRH or PrEP use among Black sexual minority men by PCAs may not be fully leveraged in traditional MI-based interventions [27].

We designed a multicomponent intervention called POSSIBLE that used a PCA and an existing smartphone app called PrEPme (Emocha Mobile Health, Inc) [25] to increase PRH as an important cognitive gateway to PrEP uptake among Black sexual minority men. The intervention was guided by life course theory [28,29], the health belief model [30,31], and possible selves [6,32]. Life course theory suggests that timing, major life events, and age-related vulnerabilities impact sexual health behaviors [28,33]. The health belief model posits that risk perceptions catalyze behavior change [9,30]. Possible selves suggests that ideas of what individuals could or want to become can influence behavior [6,32,34]. Taken together, we hypothesized that a PCA who represented a "future self" could influence PRH at particular points along the life course of Black sexual minority men by cueing individuals to action through a review of their sexual risks in PrEPme and successfully encourage others to use PrEP having navigated similar social challenges [6].

This paper describes the feasibility and preliminary effects of POSSIBLE on PRH and subsequent willingness to accept a PrEP referral among Black sexual minority men. More information on the feasibility and effects of a multicomponent strategy using a PCA and smartphone app is needed to advance the promise of their combined effectiveness. Given the extreme racial disparity in PrEP use among sexual minority men [35,36], strategies that can increase PrEP uptake are still needed. Findings will provide insights into the usefulness of combining 2 popular HIV prevention interventions into a multicomponent strategy and elucidate the cognitive and cultural aspects of health decision-making among this vulnerable community.

Methods

Ethical Considerations

All study procedures were approved by the Johns Hopkins School of Medicine Institutional Review Board (IRB00241244). Oral informed consent was audio-recorded and documented in study folders prior to conducting baseline study visits. Participants were given study ID numbers, and identifiable information (ie, names and data) was stored on private, password-protected servers. Participants were provided a US $50 electronic Amazon gift card for the baseline visit.

Study Design

POSSIBLE was a single-group, 2 session pilot intervention conducted between 2019 and 2021 that was refined using the ADAPT-ITT (Assessment, Decision, Administration, Production, Topical Experts, Integration, Training, and Testing) model [37]. Formative research was conducted to refine key aspects of the intervention approach such as the usefulness of the app-based diary and PCA characteristics among Black sexual minority men of different age groups [6].

Peer Change Agent

POSSIBLE incorporated a PCA who was matched with participants' key demographic and cultural characteristics as guided by theory, previous studies, and formative research identifying preferences among Black sexual minority men [6,38]. Specifically, studies suggested that the PCA should have similar experiences as Black sexual minority men (eg, navigating interactions with romantic or sexual partners, clinicians, and health insurance) and be a "future self" to whom they could aspire [6,38]. Therefore, the principal investigator (PI to DTD), used Descovy for PrEP and served the intervention as the PCA. Further details regarding the experience and dual role of the PI serving as a PrEP-using PCA have been published in an autoethnography [17].

PrEPme Smartphone App

PrEPme was designed for Maryland users to obtain statewide PrEP service information and navigation support from an app-based community health worker (CHW) [25]. PrEPme also allows users to self-monitor sexual risk behaviors, view a graph of sexual risk behaviors by week and month, and chat with the CHW in the app to obtain PrEP service information [6].

Linkage to PrEP Care

A CHW supervised by a nurse case manager within the Center for Infectious Disease and Nursing Innovation at the Johns Hopkins School of Nursing (previously known as the REACH Initiative) provided navigation services including reviewing eligibility for or access to medical insurance, identifying preferred clinic locations, and arranging appointment and scheduling activities. Occasionally, the PCA referred participants or helped schedule appointments at PrEP service organizations for individuals who wanted to avoid interfacing with additional staff associated with a medical research institution.

Study Procedures

Study Enrollment

A research assistant screened individuals who were interested in the study for eligibility via phone. Eligible individuals were emailed an informed consent form, details regarding their scheduled web-based baseline appointment, and an electronic survey assessing demographic and behavioral characteristics and PRH. Individuals were given the opportunity to ask the research team (including the PI) questions regarding the study via phone or email prior to their baseline visit.

Web-Based Baseline Study Visit

Due to COVID-19, baseline and follow-up study visits were conducted via Zoom (Zoom Video Communications). Prior to the baseline visit, the PI or PCA provided participants an additional opportunity to ask questions regarding the study and obtained oral informed consent via Zoom [39]. The script guided the PCA to obtain information regarding participants' lifestyles, personal goals and values, relative HIV risks, and PRH, then tailor health communication based on their responses to influence PRH and encourage PrEP use regardless of participants' reported behaviors [16,26,27,29]. Example questions included, "How would being diagnosed with HIV impact your goals?" and "Given that research suggests that 50% of Black sexual minority men will get HIV despite the fact that they use condoms more than other people, how likely do you think you will get it?" [6,40]. The script also provided opportunities for the PCA to address HIV or PrEP misinformation and disclose PrEP use to share experiences managing potential side effects, challenges disclosing use to romantic or sexual partners, and empathy regarding participants' stigmatizing clinical experiences [17,40].

At the end of the session, the PCA referred interested individuals to the CHW as described earlier. Individuals who declined referrals to the CHW were provided alternative service locations for PrEP care and referred upon request. Baseline visits lasted between 45 and 60 minutes (accounting for informed consent, rapport building, 15- to 20-minute conversation, and PrEP navigation for those who were interested), at the end of which participants were asked to download PrEPme to record their sexual risk behaviors in its app-based diary for 1 month. Baseline study visit procedures and effects have been published [40].

Web-Based Follow-Up Visit

In the second session, the PCA reviewed the PrEPme diary with participants and conducted another MI-consistent conversation to explore relative HIV risk behaviors, review behavioral alignment with goals and values, and reassess their PrEP interests. At the end of the session, the PCA referred interested individuals to the CHW as described earlier. Individuals who declined referrals to the CHW were provided alternative service locations for PrEP care and referred upon request. Follow-up visits lasted between 20 and 30 minutes, and participants were provided another US $50 electronic Amazon gift card for completing follow-up visits regardless of reported app use.

PrEP Referral

All participants were first referred a CHW at the Johns Hopkins School of Nursing who could help navigate them to PrEP services. Participants who were interested in case management from the CHW were linked to services of their choice. Individuals who declined referrals to the CHW were offered direct referrals by the PCA who reached out to the requested case management services to help schedule appointments.

Satisfaction Surveys

Participants completed a satisfaction survey that also assessed their PRH at the end of the baseline session prior to downloading PrEPme, then again at the end of their follow-up appointment.

Participants

Participants were recruited using a combination of active and passive strategies [39,41] in Baltimore, Maryland, and were eligible based upon the following self-reported criteria: Black or African American race, identifying as a cisgender person, being 18 years and older of age, same-sex attraction to men, HIV-negative, and having oral or anal sex with ≥1 male partner in the previous 6 months.

Measures

This concept was assessed using the 8-item PRH scale from Napper et al [42] before and after the baseline visit. Sample questions included items assessing concerns about HIV and perceived likelihoods of infection. Total possible scores ranged from 10 to 40, higher scores indicate greater PRH. The scale was found to be reliable (8 items, α=.78).

Data Analysis

Paired 1-tailed *t* tests were used to examine changes in PRH after the end of the study. Descriptive statistics were used to explore the proportion of participants who were referred to services and scheduled a PrEP appointment after baseline.

Results

A total of 291 individuals were screened for the study, 93 of whom were eligible. Among eligible individuals, 69 participated and 55 (80%) were retained for follow-up at month 1. Table 1 describes the sociodemographic characteristics and PrEP referral willingness among participants. The average age of participants was 32.5 (SD 8.1, range 19-62) years. Additionally, 52 (75%) identified as gay, 11 (16%) identified as bisexual, 51 (74%) reported being employed full-time or part-time at baseline, 58 (84%) reported having insurance coverage, 54 (78%) reported being single, and 32 (47%) reported ever having a sexually transmitted infection. After baseline sessions, 29 (42%) participants were willing to be referred to PrEP services, and 20 (69%) of them confirmed scheduled appointments with PrEP care teams.

Regarding the use of the mobile app–based diary, 17 (31%) follow-up participants reported recording an entry every week prior to their follow-up appointment, 3 (5%) reported using the app for half of the weeks, and 6 (11%) reported that they did not use the app at all. In total, 11 (20%) reported initiating PrEP prior to follow-up, and 15 (27%) of follow-up participants were willing to be referred to PrEP services. There were no statistically significant differences in mean PRH scores between baseline (21.2, SD 5.5) and follow-up (23.6, SD 5.7) visits (t_{122}=–1.36; *P*=.17).

Table 1. Sociodemographic characteristics and PrEP[a] referral willingness among Black sexual minority men in POSSIBLE 2019-2021 (N=69).

	Baseline	Month 1 follow-up
Age (years)		
Mean (SD)	32.5 (8.1)	32.7 (7.7)
Range	19-62	19-50
<35 years, n (%)	49 (71)	40 (73)
Sexual orientation, n (%)		
Homosexual, gay, same gender-loving	52 (75)	42 (76)
Bisexual	11 (16)	7 (13)
Other	6 (9)	6 (11)
Employment status, n (%)		
Full-time	44 (64)	36 (65)
Part-time	7 (10)	6 (11)
Unemployed	13 (19)	11 (20)
Other	5 (7)	2 (4)
Highest level of education, n (%)		
Grade 11 or less	5 (7)	1 (2)
Grade 12 or GED[b]	10 (14)	11 (20)
Associate degree	2 (3)	0 (0)
Some college	10 (14)	7 (13)
Bachelor degree	24 (35)	16 (29)
More than bachelor degree	18 (26)	20 (36)
Health care coverage, n (%)	58 (84)	51 (93)
Annual gross income (US $), n (%)		
Less than $20,000	15 (22)	15 (27)
Between $30,000 and $40,000	8 (11)	12 (22)
Between $40,000 and $50,000	9 (13)	7 (13)
Between $50,000 and $60,000	6 (9)	4 (7)
More than $60,000	23 (33)	17 (31)
Relationship status, n (%)		
Single or not in a relationship	54 (78)	45 (82)
In a committed relationship	9 (13)	7 (13)
Married	2 (3)	0 (0)
Other	2 (3)	3 (5)
STI[c] history past 6 months, n (%)	32 (47)	42 (76)
Drug use before sex past 6 months, n (%)	47 (68)	39 (70)
Willingness to be referred to PrEP services, n (%)	29 (42)	15 (27)
PrEP appointment scheduled	20 (69)	—[d]
Initiated PrEP prior to month 1	—	11 (20)

[a]PrEP: pre-exposure prophylaxis.
[b]GED: General Educational Diploma.
[c]STI: sexually transmitted infection.
[d]Not available.

Discussion

Principal Findings

This study explored the feasibility and preliminary effectiveness of POSSIBLE, a multicomponent intervention using a PCA and mobile app–based diary to improve PRH among Black sexual minority men. Given the high retention rate, POSSIBLE may be a feasible multicomponent strategy to implement among Black sexual minority men. We found improvements in PRH after baseline sessions [40]. However, we observed no statistically significant improvement in PRH after intervention from baseline to month 1.

We observed relatively low PRH scores at baseline and month 1 follow-up. Analyses showed that the PCA increased baseline PRH scores [40]. The effects of the intervention may have been maximized in the baseline session such that the addition of the app for reflexivity could not increase scores from baseline to month 1. Other studies have found that competing survival priorities supersede HIV-related concerns in the lives of Black sexual minority men. The shift from HIV as a "death sentence" in the early days of the epidemic to its positioning as a manageable chronic health condition could be a key reason for low PRH and for why perceived risk did not change from baseline to month 1. Black sexual minority men may also consider their current behaviors relatively safer than their past or their peers' behaviors as found in previous studies [6,10]. Some may appraise their vulnerability based upon their most recent behaviors, which may not have involved condomless sex or drug use in the month of the intervention, which was conducted partly during the height of the global COVID-19 pandemic.

We also observed relatively low use of the app-based diary in PrEPme. Mobile health interventions have been successful largely because of the convenience of the intervention within smartphone apps. However, PrEPme seemingly did not add value to the PCA intervention component. Studies suggest that aspects such as aesthetics, social networking ability, and gamification impact Black sexual minority men's use of app-based HIV prevention apps [24,43]. The self-monitoring feature of the app-based diary could be refined for gamification and cultural responsiveness. PrEPme also did not maximize the power of health communication to tailor messaging. However, we are unable to identify reasons for nonuse. Qualitative insights or participant feedback could help in identifying barriers to app use. In light of previous analyses showing baseline effects on PRH [40], adding the mobile app–based diary may not be necessary in the presence of an effective PCA.

The intervention did observe relatively high proportions of willingness to accept PrEP referral and initiation, which could be attributed to the interpersonal dynamics between participants and the PCA [17]. Studies consistently show that peers and social networks are important, trusted sources of health information and effective interventionists among marginalized communities, including Black sexual minority men [16,38,44,45]. Studies also show that Black sexual minority men are more willing to initiate PrEP if their peers are using it. The usefulness of communicating with a PCA may not have been outweighed by the convenience of technology. PRH may not necessarily be the primary motivation for PrEP referral willingness or initiation among Black sexual minority men.

Despite the feasibility of this intervention, using a PCA to catalyze PRH and PrEP initiation among Black sexual minority men is not without challenges. DTD described internal conflicts regarding honoring participants' disinterest in PrEP versus professional goals to increase uptake for HIV prevention in an autoethnography [17]. Additionally, managing discussions regarding side effects with Black sexual minority men whose health histories the PI or PCA was unfamiliar with or unqualified to discuss is important to consider in this peer-based approach. Concerns that being a PCA could overshadow professionalism as a researcher and health care professional were also salient [17]. PCAs should be trained to manage insider-outsider dynamics as an interventionist among Black sexual minority men, engage in active listening, and communicate with care [17,46]. Some qualities may not be able to be provided in training such as the shared social experiences and vulnerabilities of being Black sexual minority men.

Limitations

Study limitations include insufficient sample size to detect effect sizes. Causal inferences cannot be drawn, and effectiveness cannot be established with a pre-post single-group design. It is also possible that unstudied external factors could have produced the changes observed. All data were self-reported. A larger trial is needed to definitively establish the effects of the intervention, including biological confirmation of PrEP use beyond self-report.

Conclusions

Future research should test a statistically powered peer-based approach on PrEP initiation among Black sexual minority men. Psychometric tests should also be conducted to identify culturally relevant concepts of HIV risk and PrEP motivation for Black sexual minority men. Qualitative research should also clarify how app-based sexual risk diaries have unintended consequences of triggering self-stigma and shame versus informed decision-making [6]. Targeted studies among young Black sexual minority men younger than 35 years of age should be conducted, given their high HIV incidence and low PrEP uptake. Studies show age cohort differences regarding the needs and vulnerabilities among Black sexual minority men such that peers may be a more effective behavior change mechanism for younger men [33,47,48]. If effectively implemented, the person-centered approach of a PrEP-using PCA approach could lead to substantial community-level impact for Black sexual minority men because their needs are not the same.

Acknowledgments

This research was supported by the National Institute of Mental Health Grant (#K01MH11849; principal investigator to DTD). This research has been facilitated by the infrastructure and resources provided by the Johns Hopkins University Center for AIDS Research, a National Institutes of Health (NIH) funded program (1P30AI094189), which is supported by the following NIH cofunding and participating institutes and centers: the National Institute of Allergy and Infectious Disease, the National Cancer Institute, the National Institute of Child Health and Human Development, the National Health, Lung, and Blood Institute, the National Institute on Drug Abuse, the National Institute on Aging, the National Institute of General Medical Sciences, the National Institute of Diabetes and Digestive Kidney Diseases, and the National Institute of Minority Health and Health Disparities. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. PrEPme was developed with funding from several sources including The Johns Hopkins School of Nursing REACH Initiative, the Gilead Science Unrestricted Educational Grant, and the Baltimore City Health Department.

Conflicts of Interest

JEF holds the technology transfer license with Johns Hopkins University for PrEPme. The app was developed in collaboration with Emocha Mobile Health, Inc.

References

1. Hess KL, Crepaz N, Rose C, Purcell D, Paz-Bailey G. Trends in sexual behavior among men who have sex with men (MSM) in high-income countries, 1990-2013: a systematic review. AIDS Behav 2017;21(10):2811-2834 [FREE Full text] [doi: 10.1007/s10461-017-1799-1] [Medline: 28555317]
2. Schumacher C, Tao X, Chandran A, Fields E, Price A, Greenbaum A, IMPACT Partner Collaborative. Reaching those most at risk for HIV acquisition: evaluating racial/ethnic disparities in the preexposure prophylaxis care continuum in Baltimore City, Maryland. J Acquir Immune Defic Syndr 2021;87(5):1145-1153 [FREE Full text] [doi: 10.1097/QAI.0000000000002712] [Medline: 33883471]
3. Quinn K, Dickson-Gomez J, Zarwell M, Pearson B, Lewis M. "A Gay Man and a Doctor are Just like, a Recipe for Destruction": how racism and homonegativity in healthcare settings influence PrEP uptake among young Black MSM. AIDS Behav 2019;23(7):1951-1963 [FREE Full text] [doi: 10.1007/s10461-018-2375-z] [Medline: 30565092]
4. Cahill S, Taylor SW, Elsesser SA, Mena L, Hickson D, Mayer KH. Stigma, medical mistrust, and perceived racism may affect PrEP awareness and uptake in black compared to white gay and bisexual men in Jackson, Mississippi and Boston, Massachusetts. AIDS Care 2017;29(11):1351-1358 [FREE Full text] [doi: 10.1080/09540121.2017.1300633] [Medline: 28286983]
5. Babel RA, Wang P, Alessi EJ, Raymond HF, Wei C. Stigma, HIV risk, and access to HIV prevention and treatment services among men who have sex with men (MSM) in the United States: a scoping review. AIDS Behav 2021;25(11):3574-3604 [FREE Full text] [doi: 10.1007/s10461-021-03262-4] [Medline: 33866444]
6. Dangerfield II DT, Anderson JN, Wylie C, Arrington-Sanders R, Bluthenthal RN, Beyrer C, et al. Refining a multicomponent intervention to increase perceived HIV risk and PrEP initiation: focus group study among Black sexual minority men. JMIR Form Res 2022;6(8):e34181 [FREE Full text] [doi: 10.2196/34181] [Medline: 35947442]
7. Blumenthal J, Jain S, Mulvihill E, Sun S, Hanashiro M, Ellorin E, et al. Perceived versus calculated HIV risk: implications for pre-exposure prophylaxis uptake in a randomized trial of men who have sex with men. J Acquir Immune Defic Syndr 2019;80(2):e23-e29 [FREE Full text] [doi: 10.1097/QAI.0000000000001888] [Medline: 30422909]
8. Oostrom L, Rosentel K, Motley D, Hill BJ. Discordance in objective and self-perceived HIV risk: a potential barrier to pre-exposure prophylaxis in young gay and bisexual men. J Assoc Nurses AIDS Care 2020;31(1):103-109 [FREE Full text] [doi: 10.1097/JNC.0000000000000137] [Medline: 31613824]
9. Felsher M, Szep Z, Krakower D, Martinez-Donate A, Tran N, Roth AM. "I Don't Need PrEP Right Now": a qualitative exploration of the barriers to PrEP care engagement through the application of the health belief model. AIDS Educ Prev 2018;30(5):369-381 [FREE Full text] [doi: 10.1521/aeap.2018.30.5.369] [Medline: 30332306]
10. Gallagher T, Link L, Ramos M, Bottger E, Aberg J, Daskalakis D. Self-perception of HIV risk and candidacy for pre-exposure prophylaxis among men who have sex with men testing for HIV at commercial sex venues in New York City. LGBT Health 2014;1(3):218-224. [doi: 10.1089/lgbt.2013.0046] [Medline: 26789715]
11. Huang W, Lockard A, Kelley CF, Serota DP, Rolle CM, Sullivan PS, et al. From declining PrEP to PrEP initiation as "first nature"—what changes PrEP initiation decisions among young, Black MSM. AIDS Care 2022;34(3):284-293 [FREE Full text] [doi: 10.1080/09540121.2021.1960946] [Medline: 34369230]
12. Weeks MR, Convey M, Dickson-Gomez J, Li J, Radda K, Martinez M, et al. Changing drug users' risk environments: peer health advocates as multi-level community change agents. Am J Community Psychol 2009;43(3-4):330-344 [FREE Full text] [doi: 10.1007/s10464-009-9234-z] [Medline: 19326208]
13. Young LE, Schumm P, Alon L, Bouris A, Ferreira M, Hill B, et al. PrEP Chicago: a randomized controlled peer change agent intervention to promote the adoption of pre-exposure prophylaxis for HIV prevention among young Black men who have sex with men. Clin Trials 2018;15(1):44-52 [FREE Full text] [doi: 10.1177/1740774517730012] [Medline: 28862483]

14. Turpin RE, Hawthorne DJ, Rosario AD. Pre-exposure prophylaxis interventions among Black sexual minority men: a systematic literature review. Int J Environ Res Public Health 2022;19(4):1934 [FREE Full text] [doi: 10.3390/ijerph19041934] [Medline: 35206120]
15. Harawa NT, Schrode KM, McWells C, Weiss RE, Hilliard CL, Bluthenthal RN. Small randomized controlled trial of the new passport to wellness HIV prevention intervention for Black men who have sex With men (BMSM). AIDS Educ Prev 2020;32(4):311-324 [FREE Full text] [doi: 10.1521/aeap.2020.32.4.311] [Medline: 32897130]
16. Quinn KG, Christenson E, Spector A, Amirkhanian Y, Kelly JA. The influence of peers on PrEP perceptions and use among young Black gay, bisexual, and other men who have sex with men: a qualitative examination. Arch Sex Behav 2020;49(6):2129-2143 [FREE Full text] [doi: 10.1007/s10508-019-01593-x] [Medline: 32016815]
17. Dangerfield II DT. A principal investigator as a PrEP-using peer change agent for HIV prevention among Black gay and bisexual men: an autoethnography. Int J Environ Res Public Health 2023;20(7):5309 [FREE Full text] [doi: 10.3390/ijerph20075309] [Medline: 37047925]
18. Dennis C. Peer support within a health care context: a concept analysis. Int J Nurs Stud 2003;40(3):321-332. [doi: 10.1016/s0020-7489(02)00092-5] [Medline: 12605954]
19. Fuqua V, Chen Y, Packer T, Dowling T, Ick TO, Nguyen B, et al. Using social networks to reach Black MSM for HIV testing and linkage to care. AIDS Behav 2012;16(2):256-265. [doi: 10.1007/s10461-011-9918-x] [Medline: 21390535]
20. Hensel DJ, Fortenberry JD, Harezlak J, Craig D. The feasibility of cell phone based electronic diaries for STI/HIV research. BMC Med Res Methodol 2012;12:75 [FREE Full text] [doi: 10.1186/1471-2288-12-75] [Medline: 22691189]
21. Glick SN, Winer RL, Golden MR. Web-based sex diaries and young adult men who have sex with men: assessing feasibility, reactivity, and data agreement. Arch Sex Behav 2013;42(7):1327-1335 [FREE Full text] [doi: 10.1007/s10508-012-9984-9] [Medline: 22926939]
22. Stalgaitis C, Glick SN. The use of web-based diaries in sexual risk behaviour research: a systematic review. Sex Transm Infect 2014;90(5):374-381. [doi: 10.1136/sextrans-2013-051472] [Medline: 24723619]
23. Fields EL, Morgan A, Sanders RA, Jennings J. "I Met Him On Jack'D"—exploring HIV risk perceptions and risk mitigation strategies of young Black MSM meeting sex partners on geosocial networking apps. J Adolesc Health 2017;60(2):S8. [doi: 10.1016/j.jadohealth.2016.10.037]
24. Holloway IW, Winder TJ, Lea III CH, Tan D, Boyd D, Novak D. Technology use and preferences for mobile phone-based HIV prevention and treatment among Black young men who have sex with men: exploratory research. JMIR Mhealth Uhealth 2017;5(4):e46 [FREE Full text] [doi: 10.2196/mhealth.6436] [Medline: 28408360]
25. Farley JE, Dangerfield II DT, LaRicci J, Sacamano P, Heidari O, Lowensen K, et al. Community engagement and linkage to care efforts by peer community-health workers to increase PrEP uptake among sexual minority men. Public Health Nurs 2021;38(5):818-824 [FREE Full text] [doi: 10.1111/phn.12887] [Medline: 33749022]
26. Rollnick S, Miller WR. What is motivational interviewing? Behav Cogn Psychother 1995;23(4):325-334. [doi: 10.1017/S135246580001643X]
27. Dangerfield II DT, Davis G, Pandian V, Anderson JN. Using motivational interviewing to increase HIV PrEP initiation and adherence: a scoping review. Prev Sci 2023;24(7):1365-1375 [FREE Full text] [doi: 10.1007/s11121-023-01554-w] [Medline: 37249729]
28. Elder GH. The life course as developmental theory. Child Dev 1998;69(1):1-12. [Medline: 9499552]
29. Elder GH. Human lives in changing societies: life course and developmental insights. In: Cairns RB, Elder GH, Costello EJ, editors. Developmental Science. North Carolina: Cambridge University Press; 1996:31-62.
30. Rosenstock IM. The health belief model and preventive health behavior. Health Educ Monogr 1974;2(4):354-386. [doi: 10.1177/109019817400200405]
31. Rosenstock IM. Historical origins of the health belief model. Health Educ Monogr 1974;2(4):328-335. [doi: 10.1177/109019817400200403]
32. Markus H, Nurius P. Possible selves. Am Psychol 1986;41(9):954-969. [doi: 10.1037/0003-066X.41.9.954]
33. Dangerfield II DT, Smith LR, Anderson JN, Bruce OJ, Farley J, Bluthenthal R. Sexual positioning practices and sexual risk among Black gay and bisexual men: a life course perspective. AIDS Behav 2018;22(6):1919-1931 [FREE Full text] [doi: 10.1007/s10461-017-1948-6] [Medline: 29079948]
34. Bak W. Possible selves: implications for psychotherapy. Int J Ment Health Addict 2015;13(5):650-658 [FREE Full text] [doi: 10.1007/s11469-015-9553-2] [Medline: 26366132]
35. Chu BA, Castellanos ER, Gonzales MM, Gaither TW. Social determinants of self-reported pre-exposure prophylaxis use among a national sample of US men who have sex with men. Sex Transm Infect 2020;96(5):368-374. [doi: 10.1136/sextrans-2019-054177] [Medline: 31601642]
36. Kanny D, Jeffries WL, Chapin-Bardales J, Denning P, Cha S, Finlayson T, et al. Racial/ethnic disparities in HIV preexposure prophylaxis among men who have sex with men—23 urban areas, 2017. MMWR Morb Mortal Wkly Rep 2019;68(37):801-806 [FREE Full text] [doi: 10.15585/mmwr.mm6837a2] [Medline: 31536484]
37. Wingood GM, DiClemente RJ. The ADAPT-ITT model: a novel method of adapting evidence-based HIV interventions. J Acquir Immune Defic Syndr 2008 Mar 01;47(Suppl 1):S40-S46. [doi: 10.1097/QAI.0b013e3181605df1] [Medline: 18301133]

38. Dangerfield II DT, Harawa NT, McWells C, Hilliard C, Bluthenthal RN. Exploring the preferences of a culturally congruent, peer-based HIV prevention intervention for black men who have sex with men. Sex Health 2018;15(5):424-430 [FREE Full text] [doi: 10.1071/SH18057] [Medline: 30185352]
39. Dangerfield II DT, Wylie C, Anderson JN. Conducting virtual, synchronous focus groups among Black sexual minority men: qualitative study. JMIR Public Health Surveill 2021;7(2):e22980 [FREE Full text] [doi: 10.2196/22980] [Medline: 33427671]
40. Dangerfield II DT, Anderson JN. A scripted, PrEP-using peer change agent improves perceived risk for HIV and willingness to accept referrals quickly among Black sexual minority men: preliminary findings from POSSIBLE. AIDS Behav 2024. [doi: 10.1007/s10461-024-04313-2] [Medline: 38551719]
41. White JJ, Dangerfield II DT, Grieb SM. Methodological considerations for conducting focus groups in HIV prevention research among Black men who have sex with men. Public Health Nurs 2019;36(3):439-445 [FREE Full text] [doi: 10.1111/phn.12592] [Medline: 30761620]
42. Napper LE, Fisher DG, Reynolds GL. Development of the perceived risk of HIV scale. AIDS Behav 2012;16(4):1075-1083 [FREE Full text] [doi: 10.1007/s10461-011-0003-2] [Medline: 21785873]
43. Goldenberg T, McDougal SJ, Sullivan PS, Stekler JD, Stephenson R. Preferences for a mobile HIV prevention app for men who have sex with men. JMIR Mhealth Uhealth 2014 Oct 29;2(4):e47 [FREE Full text] [doi: 10.2196/mhealth.3745] [Medline: 25355249]
44. Kelly JA. Popular opinion leaders and HIV prevention peer education: resolving discrepant findings, and implications for the development of effective community programmes. AIDS Care 2004;16(2):139-150. [doi: 10.1080/09540120410001640986] [Medline: 14676020]
45. Kelly JA, St Lawrence JS, Diaz YE, Stevenson LY, Hauth AC, Brasfield TL, et al. HIV risk behavior reduction following intervention with key opinion leaders of population: an experimental analysis. Am J Public Health 1991;81(2):168-171. [doi: 10.2105/ajph.81.2.168] [Medline: 1990853]
46. Dangerfield II DT, Cooper J, Heidari O, Allen S, Winder TJ, Lucas GM. Nursing and health care preferences among opioid and stimulant using Black sexual minority men: an exploratory study. J Assoc Nurses AIDS Care 2021;32(5):e30-e39. [doi: 10.1097/JNC.0000000000000201] [Medline: 32809988]
47. Dangerfield II DT, Heidari O, Cooper J, Allen S, Lucas GM. Motivations for opioid and stimulant use among drug using black sexual minority men: a life course perspective. Drug Alcohol Depend 2020;215:108224 [FREE Full text] [doi: 10.1016/j.drugalcdep.2020.108224] [Medline: 32777690]
48. Dangerfield II DT, Harawa NT, Fernandez MI, Hosek S, Lauby J, Joseph H, et al. Age cohort differences in sexual behaviors among Black men who have sex with men and women. J Sex Res 2018;55(8):1012-1021 [FREE Full text] [doi: 10.1080/00224499.2017.1423016] [Medline: 29377734]

Abbreviations

ADAPT-ITT: Assessment, Decision, Administration, Production, Topical Experts, Integration, Training, and Testing
CHW: community health worker
MI: motivational interviewing
PCA: peer change agent
PI: principal investigator
PrEP: pre-exposure prophylaxis
PRH: perceived risk for HIV

Edited by C Jacob; submitted 20.11.23; peer-reviewed by S Harrison, G Sebastian, S Markham; comments to author 31.01.24; revised version received 17.02.24; accepted 22.04.24; published 11.06.24.

Please cite as:
Dangerfield II DT, Anderson JN, Wylie C, Bluthenthal R, Beyrer C, Farley JE
A Multicomponent Intervention (POSSIBLE) to Improve Perceived Risk for HIV Among Black Sexual Minority Men: Feasibility and Preliminary Effectiveness Pilot Study
JMIR Hum Factors 2024;11:e54739
URL: https://humanfactors.jmir.org/2024/1/e54739
doi:10.2196/54739
PMID:38861707

Original Paper

Mobile App–Assisted Parent Training Intervention for Behavioral Problems in Children With Autism Spectrum Disorder: Pilot Randomized Controlled Trial

JooHyun Lee[1*], BSN, MS; JaeHyun Lim[2*], MD; Soyeon Kang[3], SM; Sujin Kim[3], MD; So Yoon Jung[2], MD; Sujin Kim[2], PhD; Soon-Beom Hong[3,4,5*], MD; Yu Rang Park[1,6*], PhD

[1]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea
[2]LumanLab Inc, Seoul, Republic of Korea
[3]Division of Child and Adolescent Psychiatry, Department of Psychiatry, Seoul National University Hospital, Seoul, Republic of Korea
[4]Department of Psychiatry, Seoul National University College of Medicine, Seoul, Republic of Korea
[5]Institute of Human Behavioral Medicine, Seoul National University Medical Research Center, Seoul, Republic of Korea
[6]Department of Artificial Intelligence, Yonsei University, Seoul, Republic of Korea
[*]these authors contributed equally

Corresponding Author:
Yu Rang Park, PhD
Department of Biomedical Systems Informatics
Yonsei University College of Medicine
50-1 Yonsei-ro
Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 0222282493
Email: yurangpark@yuhs.ac

Abstract

Background: In children with autism spectrum disorder (ASD), problem behaviors play a dysfunctional role, causing as much difficulty with daily living and adjustment as the core symptoms. If such behaviors are not effectively addressed, they can result in physical, economic, and psychological issues not only for the individual but also for family members.

Objective: We aimed to develop and evaluate the feasibility of a mobile app–assisted parent training program for reducing problem behaviors in children with ASD.

Methods: This open-label, single-center, randomized controlled trial was conducted among parents of children with ASD aged 36-84 months. Participants were recruited from the Department of Psychiatry at Seoul National University Hospital. Participants were randomly assigned (1:1) by a blinded researcher. Randomization was performed using a stratified block randomization (with a block size of 4). Parents in the intervention group completed the mobile app–assisted parent training program at home over a 12-week period. They continued to receive their usual nondrug treatment in addition to the mobile app–assisted parent training program. The control group continued to receive their usual nonpharmaceutical treatment for 12 weeks without receiving the parent training program intervention. The primary outcome measure was the median change in the Korean Child Behavior Checklist (K-CBCL) scores from before to after the intervention. Lower scores on the K-CBCL indicated a decrease in overall problem behavior.

Results: Between November 9, 2022, and December 8, 2022, 64 participants were enrolled. Overall, 42 children (intervention group median age: 49, IQR 41-52.5 months; control group median age: 49, IQR 42-58 months) of the participants joined the program. The intervention group included 20 (48%) participants and the control group included 22 (52%) participants. In the intervention group, the K-CBCL total scores showed a decrease after the intervention, with a median difference of –0.5 (95% CI –4.5 to 3). Pervasive developmental disorder scores also showed a decrease, with a median difference of –2.1 (95% CI –8.5 to 2.5). However, there was no significant difference in Clinical Global Impression–Severity of Illness scores after the intervention for both the control and intervention groups. Scores on the Korean version of the Social Communication Questionnaire showed a further decrease after the intervention in the intervention group (median difference –2, 95% CI –4 to 1). Caregivers' stress evaluated using the Korean Parenting Stress Index Fourth Edition–Short Form did not show any significant differences between the control and intervention groups. There were no adverse events related to study participation.

Conclusions: The findings demonstrated the feasibility of using mobile devices for evidence-based parent training to reduce problem behaviors in children with ASD. Mobile devices' accessibility and flexibility may provide a viable alternative for offering early intervention for problem behaviors in children with ASD.

Trial Registration: CRIS KCT0007841; https://cris.nih.go.kr/cris/search/detailSearch.do?&seq=23112

(JMIR Hum Factors 2024;11:e52295) doi:10.2196/52295

KEYWORDS

autism spectrum disorder; parent training program; parent education; behavioral problems; child behavior; mobile app; feasibility; mHealth; evidence-based parent training

Introduction

Autism spectrum disorder (ASD) is a condition that affects how a person interacts with others, communicates, and behaves. It is characterized by social interaction difficulties, communication impairments, repetitive behaviors, and limited attention [1]. The prevalence and economic cost of ASD have been gradually increasing in recent years in Korea and worldwide [2,3]. Facilities for psychological evaluation and treatment, including consultation with a pediatric psychiatrist and testing for ASD, are concentrated in metropolitan areas, making them inaccessible to children in rural areas [4,5]. Additionally, even when a diagnosis is made, treatment centers for speech therapy, psychotherapy, acceptance and commitment therapy, music therapy, and occupational therapy are equally concentrated in metropolitan areas; the high cost of treatment makes it difficult for low-income children to receive appropriate treatment [4,6]. Given its enduring consequences and significant socioeconomic burden, ASD is emerging as a priority for intervention efforts.

In individuals diagnosed with ASD, problem behaviors play a dysfunctional role, causing difficulties in daily living and adjustment as much as the core symptoms [7,8]. Problem behaviors can be defined as any behavior that is aggressive toward the self or others, causes damage to the physical environment or objects, interferes with acquiring new skills, or isolates the person from society [9]. Problem behaviors that are not effectively addressed can cause physical, economic, and psychological distress to primary caregivers and other family members [10,11]. Various nonpharmacologic treatments are available to address these problems in children with ASD. Among them, nondrug treatments such as acceptance and commitment therapy and music therapy have shown promising results in addressing emotional problems and improving overall well-being, but they are limited in their ability to directly target problem behaviors [12-16]. Applied Behavior Analysis (ABA) refers to evidence-based interventions that are applied to the education of children with ASD to increase desirable behaviors and decrease or eliminate undesirable behaviors [17,18]. ABA is the science of learning and behavior; to effectively intervene and change problem behaviors, it focuses on identifying the exact cause of the behavior and implementing interventions that target that cause [19,20]. ABA-based parent training is reportedly effective in reducing problem behaviors in children developing from preschool through adolescence in general. Parent training for children with ASD is practical as a treatment model, can be used in various settings, and empowers parents to be change agents themselves [21-23].

Digital therapeutics are advanced software technologies that deliver evidence-based therapeutic interventions to patients to prevent, manage, and treat diseases or disorders [24,25]. Korea's high mobile phone use rate of 94.8% [26] highlights the potential for increasing accessibility to mobile app–assisted digital therapies. These therapies, with their scalability, accessibility, and low cost, offer the convenience of accessing treatment anytime, anywhere, making them particularly relevant for children in rural areas and low-income communities with limited access to hospitals and treatment centers [25,27].

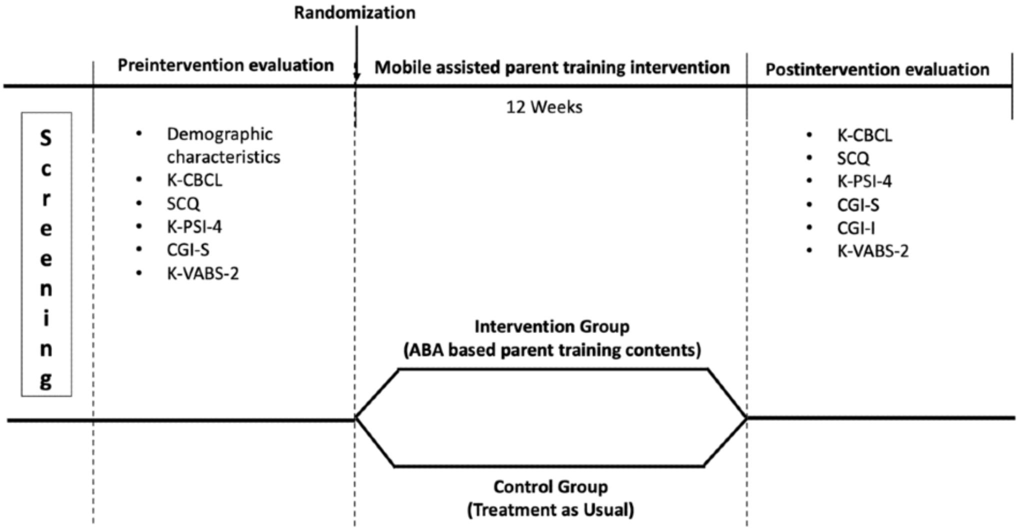
This study aimed to (1) create a mobile app–assisted parent training program using digital technology to decrease problem behaviors in children with ASD, and (2) evaluate the effectiveness of the training program in reducing problem behaviors in children with ASD by comparing pre- and postintervention outcomes.

Methods

Study Design

This study was a parallel-group, open-label, single-center, randomized controlled trial to test the effectiveness of a mobile app–assisted parent training program for reducing behavior problems in children with ASD aged 36-84 months. It followed the CONSORT (Consolidated Standards of Reporting Trials) guidelines. A parent training program based on ABA [17,18] was delivered via mobile devices. For inclusion criteria, children of participants were recruited from the Department of Psychiatry at Seoul National University Hospital and underwent the Autism Diagnostic Observation Schedule-2 (ADOS-2) and Korean Childhood Autism Rating Scale-2 (K-CARS-2), provided demographic information, and completed before the evaluation. Parenting stress levels of the parents and caregivers were also measured. The eligible participants were randomly assigned to either the intervention group or the control group at a 1:1 ratio for 12 weeks. The intervention group received a mobile app parent training program for 12 weeks, while the control group continued with their usual treatment but received no intervention. Both groups were assessed before and after the 12-week period (Figure 1).

Figure 1. Study design. ABA: Applied Behavior Analysis; CGI-S: Clinical Global Impression–Severity of Illness; K-CBCL: Korean Child Behavior Checklist; K-PSI-4: Korean version of the Parenting Stress Index 4th Edition; K-VABS-2: Korean Vineland Adaptive Behavior Scale-2; SCQ: Social Communication Questionnaire.



Participants

Participants with eligible children were recruited from November 9, 2022, to December 8, 2023, at the Department of Psychiatry at Seoul National University Hospital. The inclusion criteria for the children were as follows: (1) being aged between 36 and 84 months of age; (2) meeting the criteria for "autism" or "autism spectrum" classification according to the ADOS-2; and (3) having as a primary caregiver an adult aged at least 19 years. The following were the exclusion criteria for this study: (1) the child was currently taking psychiatric medication for behavioral regulation or similar purposes; (2) the child had a history of congenital or acquired brain damage, such as cerebral palsy; (3) the child had a history of seizure disorders or other neurological conditions; (4) the child had an uncorrected sensory impairment (eg, vision or hearing deficits); (5) the child's primary caregiver did not possess a smartphone or lacked access to smartphone apps; and (6) the primary caregiver either did not give their consent to join the study or opted to withdraw during the course of the study. All participants provided written informed consent.

Randomization and Masking

The children were divided into low- and high-severity autism groups according to ADOS-2 criteria and randomized into intervention and control groups for each severity group. Randomization was performed using a stratified block randomization method (with a block size of 4) with 1:1 assignment to either the intervention or control group. The randomization list was created using the PROC PLAN method (SAS Institute Inc) in SAS (version 9.4; SAS Institute Inc) software. The randomization process was managed and operated by an independent third party, the Medical Statistics Team of Asan Medical Center.

Sample Size

As this was an exploratory clinical trial, a formal calculation for sample size was not necessary; however, we followed the method described by Julious [28] to calculate the sample size for the pilot trial. The recommended sample size for such pilot studies is 12 persons per group. Considering the possibility of lower compliance with the use of digital therapeutics compared with taking conventional medicines, this study was conducted with a total of 60 participants, 30 in the intervention group and 30 in the control group.

Mobile App–Assisted Parent Training Program

The mobile app–assisted parent training program aims to improve the behavior of children with ASD by identifying the causes of problem behaviors through a self-report questionnaire and providing digital content about it. Previous studies have shown that mobile app or telehealth-enabled ABA interventions can be effective in managing and reducing problem behaviors and increasing parent engagement in children with autism [29-31]. The program was developed in the form of a mobile app to help parents apply ABA-based therapy to their children in their daily lives to reduce their problem behaviors [32-34]. The mobile app–based training program lasted 10-15 minutes per session and was conducted 2-5 days a week for 12 weeks. It covers topics such as an introduction to ABA, identifying the underlying reasons for problematic behavior, creating a behavioral support plan, implementing behavior intervention, practicing in different situations, individualized learning opportunities, and crisis management (Table 1). To further validate our approach and differentiate it from previous studies, we included a system in the mobile app that allows parents to record and monitor their child's behavior and progress. This real-time data collection allows for continuous monitoring and adjustment of the intervention to ensure responsiveness and effectiveness. Program content was developed with the consultation of 2 ABA experts and 1 child psychiatry specialist.

Table 1. Overview of the program curriculum of the mobile app–assisted parent training intervention for behavioral problems in children with autism spectrum disorder.

Session	Learning objectives	Description of the lesson
1	Orientation and introduction to ABA[a]	• Orientation: Introducing good time • ABA: definition, characteristics, and recent trends • Defining objective behavior and collecting data
2	Identifying the reasons behind problematic behaviors	• Functions of problem behaviors • Methodology of functional assessment: indirect functional assessment, observational functional assessment (ABC[b] functional assessment)
3	Creating a BSP[c]	• Create a BSP according to ABC functional assessment: antecedent intervention; consequences intervention; behavior intervention
4	Antecedent intervention 1: Situation (background) event intervention	• Finding and structuring environments that cause problematic behavior • Finding and structuring time environments that cause problematic behavior
5	Antecedent intervention 2	• Stimulus control • Motivating operation
6	Consequences intervention 1: Reinforcement	• Reinforcement and differential reinforcement • Behavior contract and token reinforcement
7	Consequences intervention 2: Extinction	• Extinction • Disinterestedness
8	Behavior intervention 1: Instruction following and FCT[d]	• Instruction following • FCT
9	Behavior intervention 2: Behavior shaping and chaining	• Behavior shaping • Behavior chaining
10	Situation training	• Introduce situation training • Picture-supported situation training • Word-supported situation training
11	DTT[e]	• What is DTT? • Procedure of DTT • Application of DTT
12	Crisis management: Self-harm and aggressive behavior	• Aggressive behavior • Destructive behavior, self-stimulatory behavior, and self-harm • Preparing a safety plan

[a]ABA: Applied Behavior Analysis.

[b]ABC: antecedents-behavior-consequences.

[c]BSP: behavior support plan.

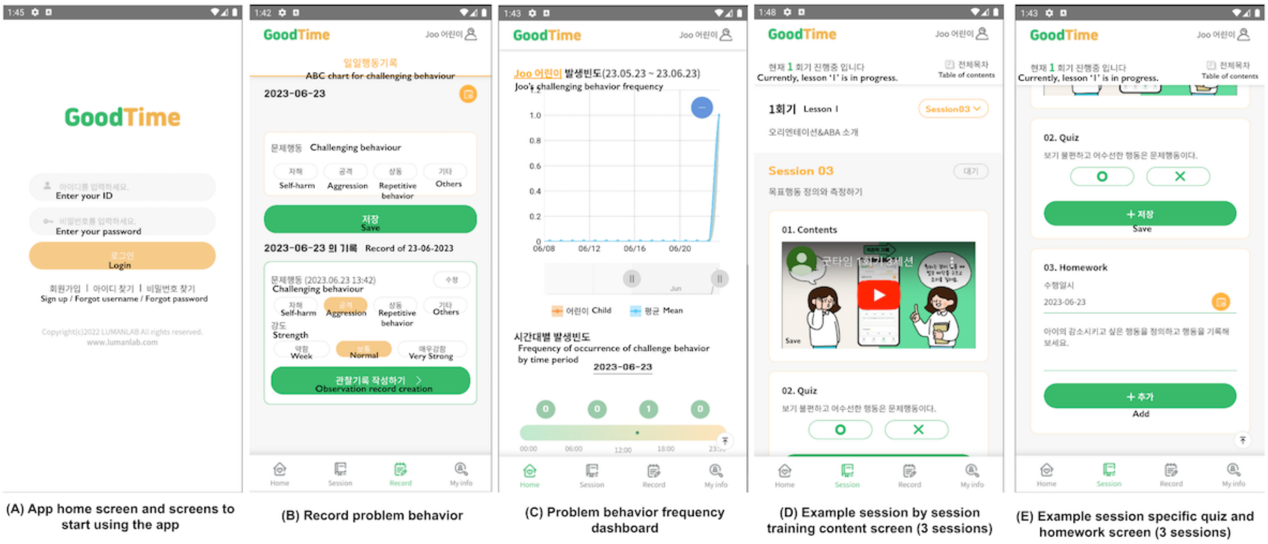
[d]FCT: functional communication training.

[e]DTT: discrete trial training.

Procedure

The mobile app–assisted parent training program was available for download from a mobile app store, and participants accessed this study's intervention using a unique, preannounced username and password. Parents in the intervention group completed the mobile app parent training program at home over a 12-week period, during which they continued to receive their usual nondrug treatment in addition to the training program. The mobile app included recording problem behaviors and viewing the frequency of child problem behaviors, education, and homework assignments to help caregivers generalize skills in a real-world setting (Figure 2). All sessions included instruction on the parent training program, video content with case examples to help parents understand the program, and quizzes and homework to check for understanding of the session. Participants could access the app at any time, and all programs delivered were relearnable.

Figure 2. Mobile app platform example. (A) App home screen and screens to start using the app; (B) record problem behavior; (C) problem behavior frequency dashboard; (D) example session by session training content screen (3 sessions); and (E) example session-specific quiz and homework screen (3 sessions).



Control Group

The control group continued to receive their usual nonpharmaceutical treatments, such as behavioral therapy, speech therapy, and play therapy, for 12 weeks, without receiving the mobile app–assisted parent training program intervention.

Measures

About ADOS-2

The ADOS-2 is a tool used to help diagnose ASD by observing children's social interactions and behaviors through play or conversation [35]. It is administered by a qualified evaluator and takes approximately 45 minutes to complete. It is used to set standards for participating children and is a semistructured tool for directly observing children [36].

About K-CARS-2

The K-CARS-2 is a tool used to differentiate between autism and other developmental disorders, and to determine the severity of autism disorders [37,38]. It can be completed in a relatively short amount of time, typically taking 20-30 minutes.

Korean Child Behavior Checklist

The Korean Child Behavior Checklist (K-CBCL) is an evaluation tool used to assess the impact of overall problem behaviors, adaptation, and social performance in children [39]. It is a standardized checklist completed by parents, who describe their child's behavioral and emotional problems. The K-CBCL comprises a social ability scale and a syndrome and total problem scale, taking 15-20 minutes to complete. It is suitable for infants and toddlers aged 18 months to 5 years and can be administered to children aged up to 6 years in kindergarten.

About K-SCQ

The Korean Version of the Social Communication Questionnaire (K-SCQ) is a useful screening tool for identifying a range of symptoms related to ASD in a short amount of time (approximately 15 min) and can be used with people of any age and language level. It comprises 2 types of tools, each containing 40 questions that ask parents or guardians about their child's symptoms related to autism (such as communication, interaction, and limited and repetitive behaviors and interests) [40]. The Social Communication Questionnaire (SCQ) current type focuses on behavior over the past 3 months, while the SCQ lifetime type focuses on behavior over a lifetime. This study used the Korean translation of the SCQ current type [41].

About Korean Vineland Adaptive Behavior Scale-2

The Korean Vineland Adaptive Behavior Scale-2 (K-VABS-2) is an evaluation tool used to assess the impact of overall problem behaviors, adaptation, and social performance in individuals of all ages. It comprises 4 areas (communication, daily life technology, socialization, and exercise functions) and 11 subareas (including language skills, coping abilities, and small muscle control). The K-VABS-2 can be administered in the form of a survey interview, taking 20-60 minutes, or a guardian rating, taking 30-60 minutes. It is suitable for individuals of all ages [42].

Clinical Global Impression–Severity of Illness

In the Clinical Global Impression–Severity of Illness (CGI-S), a physician uses a scale of 1-7 points to evaluate the severity of a disease based on the symptoms experienced by past patients diagnosed with the same disease. This scale is used to determine the degree of symptoms in currently diagnosed patients [43].

Clinical Global Impression Global Improvement

In the Clinical Global Impression Global Improvement (CGI-I), a physician uses a scale of 1-7 points to evaluate the effectiveness of therapeutic intervention in a patient with a mental disorder. The scale is used to determine whether the patient's condition has improved or worsened compared with before the intervention was initiated [43].

Korean Parenting Stress Index Fourth Edition–Short Form

The Parenting Stress Index is a self-report test used to measure the stress experienced by parents of children aged 1-12 years. It evaluates the characteristics and situational factors of children that affect parenting stress as perceived by the parents [44] and comprises 2 main areas: child and parent. The child area is further divided into 6 subscales: distraction or excessive behavior, adaptation, compensation, demand, mood, and acceptance. The parent area comprises 7 subscales: competence, isolation, attachment, health, role restriction, depression, and spouse or parenting partner relationship. The test also includes a life stress scale that measures events that can affect parenting stress. The Parenting Stress Index is available in both a general form and a short form (Korean Parenting Stress Index Fourth Edition–Short Form [K-PSI-4-SF]); the latter was used in our study [45].

Data Management

Research data were stored in a secure laboratory and personal information was stored in a separate file to ensure that personal identification is not possible through the research data. Enrollment logs were used to maintain personal identification information separately; the case report form did not include personal information except for initials and case numbers. This helped to protect participants' privacy.

Statistical Analysis

All outcomes were analyzed in participants who completed the program period and assessments at baseline and after the intervention. The analysis included participants who did not violate the protocol (per-protocol analysis set). Owing to the nonnormal distribution, nonparametric methods were used in this study. All baseline demographic variables and evaluation outcomes at baseline and after the intervention were summarized by randomized groups using the median (IQR) for continuous data or count (%) for categorical data. We estimated the median values and IQRs for each group and time point using the pre- and postintervention evaluation scores for the intervention and control groups. Before testing the effectiveness of the program, the Wilcoxon signed rank test was performed to determine whether there were differences in demographic and baseline variables between the groups. The Wilcoxon signed rank test was used to compare the effects of all outcomes. To evaluate the intervention effects between the 2 groups, the Kruskal-Wallis test was performed by judging the rejection range based on a significance level of 0.05. To assess differences in session completion rates within the intervention group, we estimated medians and IQRs and used Wilcoxon signed rank tests. To examine the correlation of completion rates with evaluation outcomes, Pearson correlation coefficient was used. Data analysis was performed using the R software (version 4.1.0; R Foundation) and Python (version 3.9.12; Python Software Foundation).

Ethical Considerations

Written informed consent was obtained from all participants. This study was approved by the Seoul National University Hospital Institutional Review Board (H-2205-158-1329).

All data were stored in a locked laboratory, and participants' personal information was stored in a separate file from the research data to prevent identification through the research data. Enrollment logs were maintained to keep personal information separate and to ensure that personal information was not exposed. Only initials and case numbers were used on case report forms to further protect participant identity.

The tests administered as part of this study were provided free of charge. Compensation was not contingent on the participant completing this study, and pretest results were provided even if the participant did not complete this study. The posttest was only administered if the participant completed this study. No other monetary compensation, including transportation, was offered.

Results

Participant Characteristics

Between November 9, 2022, and December 8, 2022, a total of 64 participants were enrolled. Further, 56 participants who met our inclusion criteria were randomly assigned into the intervention and control groups; however, after randomization, 8 participants in the intervention group and 3 in the control group declined to participate, resulting in 20 participants receiving the intervention and 25 participants in the control group. In the control group, 3 participants were excluded from the analysis; these included children who started a new treatment during the intervention period. Finally, a total of 42 participants were included in the data analysis (Figure 3).

Demographic and baseline variables were compared between the intervention and control groups before the intervention. Of the 42 participants, 20 (48%) were in the intervention group and 22 (52%) in the control group. There were no significant differences in sex, age, severity, or ethnicity between the 2 groups. Demographic variables, including the K-CARS-2 and ADOS-2, did not differ significantly between the groups. The current treatment also did not differ significantly between the intervention (n=17, 85%) and control (n=18, 81.8%) groups. Neuropsychological testing showed no significant differences between the 2 groups. The demographic and baseline variables of both groups are summarized in Table 2. Complete information is presented in Table S1 in Multimedia Appendix 1. Table S2 in Multimedia Appendix 1 provides demographic and baseline information for the full sample using an intention-to-treat (ITT) analysis approach. The results of this ITT analysis revealed a significant difference between the intervention group (14.5, 95% CI 12.5 to 16) and the control group (12, 95% CI 10 to 15) on social effect, a subscale of the ADOS, compared with the results obtained from the per-protocol analysis (*P*=.045). Except for this item, no significant differences were found for the remaining items. Multimedia Appendix 1 provides more details.

Figure 3. CONSORT flowchart of participation. CONSORT: Consolidated Standards of Reporting Trials.

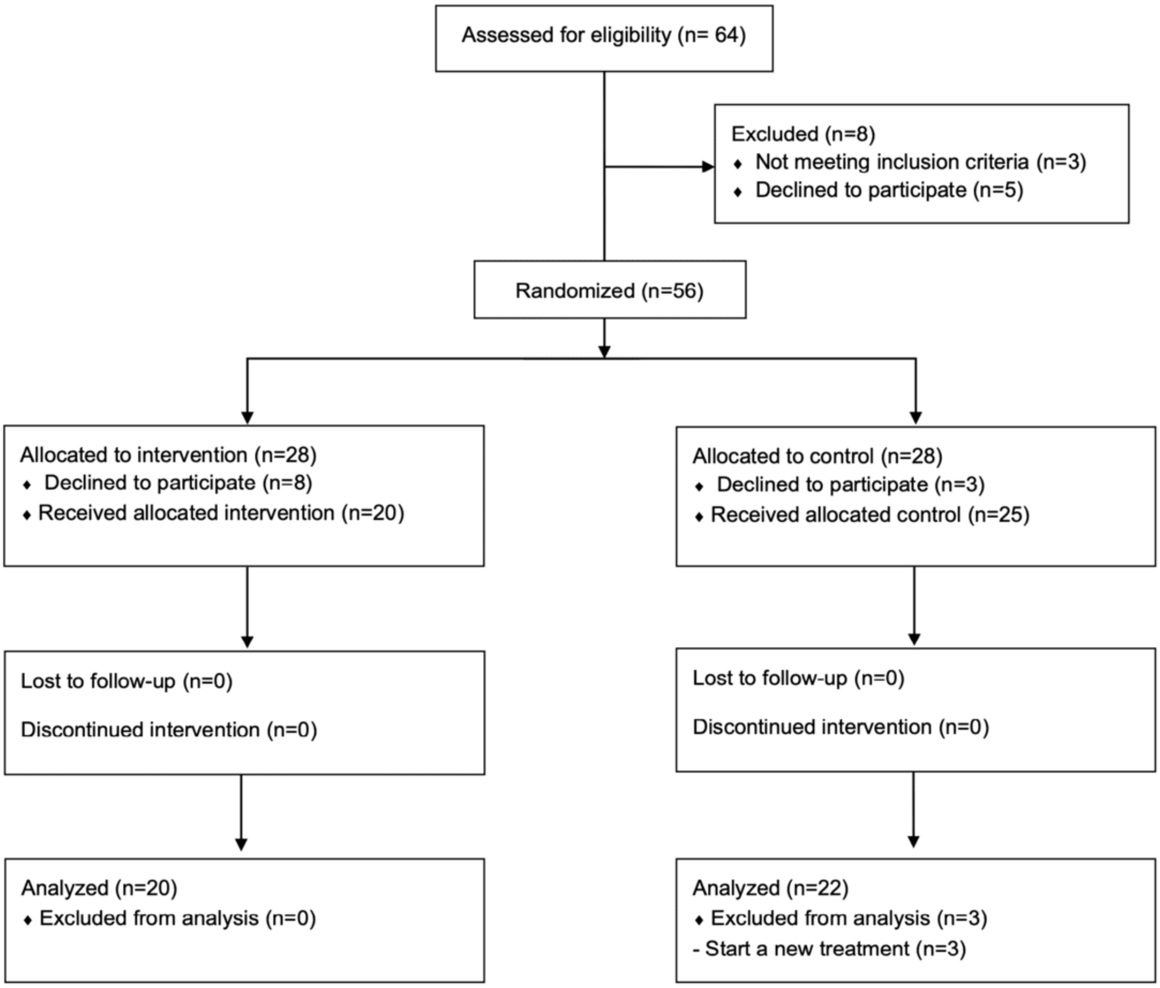


Table 2. Demographic and baseline variables for the intervention and control groups (a version of per-protocol analysis).

Characteristics	Intervention group (N=20)	Control group (N=22)
Sex , n (%)		
Male	15 (75)	18 (81.8)
Female	5 (25)	4 (18.2)
Age (months), median (IQR)	49 (41 to 52.5)	49 (42 to 58)
Severity, n (%)		
Mild	4 (20)	4 (18.2)
Severe	16 (80)	18 (81.8)
K-CARS-2[a], median (IQR)	33 (32-36.2)	35 (31-37)
ADOS[b] comparison, n (%)		
Extreme	11 (55)	11 (50)
Moderate	1 (5)	3 (13.6)
Moderate to severe	2 (10)	1 (4.5)
Severe	6 (30)	7 (31.8)
Ongoing therapy, n (%)		
No	3 (15)	4 (18.2)
Yes	17 (85)	18 (81.8)
Physician measures		
CGI-S[c], median (IQR)	3 (2 to 4)	3 (2 to 4)
Parents' measures (for children), median (IQR)		
K-CBCL[d] (total)	57.5 (50 to 63)	60.5 (55 to 64)
Emotionally reactive	50 (50 to 58)	53 (50-63)
Anxious or depressed	50 (50 to 58)	50 (50-56)
Somatic concerns	54 (50 to 59)	50 (50-54)
Withdrawn	66.5 (62 to 70)	62 (58-65)
Attention problems	61.5 (55 to 66)	59 (50-64)
Aggressive behavior	53.5 (50 to 57.5)	57 (50-64)
K-SCQ[e] (total)	18 (13 to 20)	16 (12 to 22)
K-VABS-2[f]		
Communication	64 (57 to 70)	66 (53 to 81)
Daily living skills	66 (59 to 69)	66 (59 to 80)
Socialization	51 (46.5 to 54)	50 (48 to 60)
Motor	67 (61 to 80)	68.5 (58 to 80)
Maladaptation	18 (17 to 20)	19.5 (19 to 20)
Parents' measures (for parents), median (IQR)		
K-PSI-4-SF[g]		
Total stress scale	103 (93 to 113)	95 (88 to 109)
Parental distress	34 (31.5 to 38.5)	33.5 (28 to 40)
Parent-child dysfunctional interaction	34 (30.5 to 38)	29 (26 to 38)
Difficult child	32.5 (29.5 to 38.5)	30 (26 to 37)

[a]K-CARS-2: Korean Childhood Autism Rating Scale-2.

[b]ADOS: Autism Diagnostic Observation Schedule.

[c]CGI-S: Clinical Global Impression–Severity of Illness.

[d]K-CBCL: Korean Child Behavior Checklist.

[e]K-SCQ: Korean versions of the Social Communication Questionnaire.

[f]K-VABS-2: Korean Vineland Adaptive Behavior Scale-2.

[g]K-PSI-4-SF: Korean Parenting Stress Index 4th Edition–Short Form.

Primary Outcome

The results for the K-CBCL scores are shown in Table 3. According to the K-CBCL, lower scores indicate a decrease in overall problem behaviors. The K-CBCL total score for the intervention group decreased from baseline to after the intervention (median difference –0.5, 95% CI –4.5 to 3). In comparison to the control group, the intervention group's median withdrawn score difference reduced to –2 (95% CI –8 to 4), median difference of attention decreased by –1 (95% CI –6 to 4), and median difference in pervasive developmental disorder exhibited a –1 (95% CI –5 to 2) decrease. The full table is shown in Table S3 in Multimedia Appendix 1.

Table 3. Comparison between outcome variables for the intervention and control groups (a version of per-protocol analysis).

Characteristics	Intervention				Control				Median difference, (95% CI)	*P* value
	Pre (n=20), median (IQR)	Post (n=20), median (IQR)	Median difference, (95% CI)	*P* value	Pre (n=22), median (IQR)	Post (n=22), median (IQR)	Median difference, (95% CI)	*P* value		
Primary outcomes										
K-CBCL[a] (total)[b]	57.5 (50 to 63)	58 (50 to 63)	–0.5 (–4.5 to 3)	.91	60.5 (55 to 64)	55.5 (52 to 65)	–2 (–6 to 2)	.43	2 (–4 to 6)	.50
Withdrawn	66.5 (62 to 70)	68 (60 to 71.5)	1 (–4.5 to 6.5)	.78	62 (58 to 65)	68 (62 to 70)	4 (–2.5 to 8)	.23	–2 (–8 to 4)	.50
Attention problems	61.5 (55 to 66)	61.5 (55 to 66)	0 (–4.5 to 6.5)	.95	59 (50 to 64)	61.5 (55 to 68)	1.5 (–2 to 7)	.40	–1 (–6 to 4)	.55
Pervasive developmental disorder	69 (64 to 74)	69.5 (59.5 to 72)	–2.1 (–8.5 to 2.5)	.41	69.5 (66 to 76)	69.5 (66 to 76)	1 (–4.5 to 4)	.74	–1 (–5 to 2)	.41
Physician measures										
CGI–S[c]	3 (2 to 4)	3 (2 to 3)	–0.5 (–2 to 1)	.40	3 (2 to 4)	3 (2 to 3)	–1 (–2 to 1)	.23	0 (0 to 0.5)	.70
Secondary outcomes										
K–SCQ[d] (total)[b]	18 (13 to 20)	15 (9.5 to 20)	–2 (–4 to 1)	.20	16 (12 to 22)	16.5 (9 to 20)	–1.5 (–3 to 0)	.08	0 (–3 to 3)	.90
K–VABS–2[e,f]										
Communication	64 (57 to 70)	68 (56 to 75)	4 (0 to 8)	.04	66 (53 to 81)	73 (52 to 85)	0 (–5 to 3)	.84	4 (0 to 10)	.09
Daily living skills	66 (59 to 69)	69 (63 to 74)	4.5 (–2.5 to 14.5)	.12	66 (59 to 80)	66 (49 to 76)	–5.1 (–9.5 to 0.5)	.06	6 (1 to 13)	.02
Socialization	51 (46.5 to 54)	55 (48 to 62.5)	7 (2.5 to 12)	<.01	50 (48 to 60)	51 (48 to 56)	–1.5 (–5.5 to 2)	.55	7 (2 to 13)	<.01
Parents' measures (for parents)										
K–PSI–4–SF[b,g]										
Total stress scale	103 (93 to 113)	106.5 (93 to 116)	0.5 (–4.5 to 6.5)	.88	95 (88 to 109)	93.5 (87 to 108)	–3.5 (–9 to 3)	.20	4 (–3 to 11)	.19
Parental distress	34 (31.5 to 38.5)	37 (32 to 41.5)	1.5 (–1.5 to 5)	.29	33.5 (28 to 40)	32.5 (28 to 38)	–1.5 (–3.5 to 0.5)	.15	2 (0 to 6)	.07

[a]K-CBCL: Korean Child Behavior Checklist.

[b]Lower score indicates increased ability.

[c]CGI-S: Clinical Global Impression–Severity of Illness.

[d]K-SCQ: Korean versions of the Social Communication Questionnaire.

[e]K-VABS-2: Korean Vineland Adaptive Behavior scale-2.

[f]Higher score indicates increased ability.

[g]K-PSI-4-SF: Korean Parenting Stress Index 4th Edition–Short Form.

Secondary Outcomes

Overview

A lower CGI-S score indicates less severe symptoms, a lower K-SCQ score indicates better social communication ability, a lower K-PSI-4-SF score indicates less stress, and a higher K-VABS-2 score indicates better adaptation and social performance. The results for the secondary outcomes are shown in Table 3. Table S3 and Figure S1 in Multimedia Appendix 1 show the full table and plots.

Clinical Global Impression

There was no median difference between baseline and after the intervention for the intervention and control groups on the CGI-S (0, 95% CI 0 to 0.5). On the CGI-I, the number of participants who improved after the intervention was 17 (77.3%) in the control group and 17 (85%) in the intervention group.

Social Adoption and Emotional and Behavioral Problems

K-SCQ scores in the intervention group decreased further after the intervention (median 15, IQR 9.5-20) compared with baseline (median 18, IQR 13-20). The median difference in communication on the K-VABS-2 exhibited a 4 (95% CI 0 to 8) increase after the intervention compared with the baseline, and socialization (7, 95% CI 2 to 13; *P* value <.01) increased more in the intervention group compared with the control group. Daily living skills also showed a significant difference when comparing the pre-post difference in the intervention group to the pre-post difference in the control group (6, 95% CI 1 to 13; *P* value .02).

Caregiver's Stress

Caregiver's stress was evaluated using the K-PSI-4-SF. The median difference between the intervention and control groups showed that the total stress scale (4, 95% CI –3 to 11), parental distress (2, 95% CI 0 to 6), parent-child dysfunctional interaction (1, 95% CI –2 to 4), and difficult child (0.5, 95% CI –3 to 3) were not statistically different between the intervention and control groups.

Comparison Values Based on Session Completion Rate

The mean number of sessions completed by participants in the intervention group was 13.4 of 26 (51.5%) sessions. The median pre- and post difference for K-CBCL total was –3.5 (95% CI –12 to 5) for 100% completion and –3.5 (95% CI –8.5 to 3) for 40% completion. The median difference in withdrawn was –4.5 (95% CI –14 to 5) for 100% completion and –3.1 (95% CI –11 to 6) for 40% completion, indicating a decrease in the intervention effect. For sleep problems, the median difference was –12 (95% CI not available) for 100% completion and –8.3 (95% CI –12 to –1.5) for 40% completion, indicating a decrease in the intervention effect. Communication and motor skills in the K-VABS-2 also increased from 8 (95% CI –8 to 24) and 10.8 (95% CI 2 to 22) at 100% completion to 3 (95% CI –4 to 11) and 5.5 (95% CI –0.5 to 17.5) at 40% completion, respectively. The total stress scale on the K-PSI-4-SF was 9.5 (95% CI 1.5 to 17) at 100% completion and 1.5 (95% CI –6.5 to 12) at 40% completion, indicating that higher session completion rates were associated with higher caregiver stress. All results for pre- and post median differences in evaluation scores by session completion rate are shown in Table 4 and Table S5 in Multimedia Appendix 1, and correlations for pre- and post median differences in evaluation score by session completion rate are shown in Figure S2 in Multimedia Appendix 1.

Table 4. Differences in pre- and postevaluation values by session completion rate.

Characteristics	Session completion rate				
	100%[a] (n=5), median difference, median (95% CI)	90%[b] (n=6), median difference, median (95% CI)	80%[c] (n=7), median difference, median (95% CI)	60%[d] (n=8), median difference, median (95% CI)	40%[e] (n=9), median difference, median (95% CI)
Primary outcomes					
K-CBCL (total)[f,g]	–3.5 (–12 to 5)	–3 (–8.5 to 4.5)	–3.4 (–8.5 to 4.5)	–3.5 (–8.5 to 4.5)	–3.5 (–8.5 to 3)
Anxious or depressed	5.5 (5.5 to 5.5)	6.8 (–2 to 13)	6.8 (–2 to 13)	6.8 (–2 to 13)	3.1 (–6 to 13)
Withdrawn	–4.5 (–14 to 5)	–2 (–8 to 5)	–2 (–8 to 5)	–3.5 (–10.5 to 5)	–3.1 (–11 to 6)
Sleep problems	–12 (N/A[h] to N/A)	–8.5 (–8.5 to –8.5)	–9.1 (–12 to –12)	–9.1 (–12 to –12)	–8.3 (–12 to –1.5)
Aggressive behavior	–1 (–1 to –1)	–1 (–5 to 3)	–0.4 (–5 to 3)	0.5 (–5 to 3)	–0.5 (–3 to 2.5)
Physician measures					
CGI-S[g,i]	–0.5 (–0.5 to –0.5)	–0.2 (1 to 1)	–0.4 (–0.5 to 1)	–0.4 (–0.5 to 1)	–0.4 (–0.5 to 1)
Secondary outcomes					
K-SCQ[j] (total)[g]	–5.5 (–7 to –4)	–5 (–6 to –1)	–4.5 (–6 to –0.5)	–4 (–6.5 to –0.5)	–3 (–6 to 1)
K-VABS-2[k,l]					
Communication	8 (–8 to 24)	8 (–2 to 16)	3.8 (–6 to 17)	3 (–5 to 16)	3 (–4 to 11)
Motor	10.8 (2 to 22)	10.8 (2 to 22)	8 (2 to 22)	5 (0 to 13)	5.5 (–0.5 to 17.5)
Parents' measures (for parents)					
K-PSI-4-SF[g,m]					
Total stress scale	9.5 (1.5 to 17)	6.5 (–4 to 17)	4.1 (–8.5 to 20)	2.4 (–7.5 to 15)	1.5 (–6.5 to 12)
Parental distress	5.5 (–2 to 14)	5 (–1.5 to 9.5)	1.5 (–7 to 13)	2 (–6.5 to 10)	2.5 (–3 to 9.5)
Parent-child dysfunctional interaction	1.3 (–6 to 3)	–1.5 (–6 to 2)	–1.5 (–4 to 2)	–2 (–5.5 to 2)	–1.7 (–5.5 to 2)
Difficult child	4.6 (–2 to 10)	4 (0.5 to 8)	3 (0.5 to 8)	2.9 (–5 to 9)	1.3 (–5 to 6.5)

[a]Groups with a session completion rate of 100% for the program.
[b]Groups with a session completion rate of 90% or higher for the program.
[c]Groups with a session completion rate of 80% or higher for the program.
[d]Groups with a session completion rate of 60% or higher for the program.
[e]Groups with a session completion rate of 40% or higher for the program.
[f]K-CBCL: Korean Child Behavior Checklist.
[g]Lower score indicates increased ability.
[h]N/A: not available.
[i]CGI-S: clinical global impression–severity of illness.
[j]K-SCQ: Korean versions of the Social Communication Questionnaire.
[k]K-VABS-2: Korean Vineland adaptive behavior scale-2
[l]Higher score indicates increased ability.
[m]K-PSI-4-SF: Korean Parenting Stress Index 4th Edition–Short Form.

Statement on Harm

There were no serious intervention-related adverse events that led to treatment discontinuation in either group (Table 5).

Table 5. Summary of any adverse events during the trial.

Adverse events	Intervention group (n=20), n (%)	Control group (n=22), n (%)
Yes	0 (0)	0 (0)
No	20 (100)	22 (100)

Discussion

Principal Findings

This study tested the feasibility of a mobile app–assisted parent training program by comparing pre- and postintervention outcomes over a 12-week intervention to evaluate the effectiveness of the program in reducing behavioral problems in children with autism. Specifically, this study demonstrated the potential for a mobile app–based intervention to reduce behavioral problems in children with ASD. These findings can help in the planning of early interventions and high accessibility of mobile app–assisted parent training programs for children with ASD who have behavioral issues. Problem behaviors in children with ASD cause additional challenges for parents and create uncertainty about how to manage these behaviors. Given these challenges, parents of children with ASD can handle behavioral issues through parent training [46].

Regarding this study's primary outcome, we found no clinically important differences between the intervention and control groups. A possible explanation for these results is that both groups maintained their existing treatments, such as ABA therapy, sensory integration therapy, and language therapy, which may have provided an indirect pathway for improving problem behaviors in children with ASD. However, the results can still be considered clinically significant. The pre- and posttreatment K-VABS-2 scores for the intervention group showed clinically significant effects for both communication and socialization. Daily living skills also showed clinically significant effects when comparing the intervention and control groups (Table 2). These scores represent the impact of overall problem behaviors, adaptation, and social performance in children, with similar results to previous studies [47,48]. Considering the exploratory nature of this study, we used the K-CBCL as the primary outcome measure, which provides subscale scores across various measurements. However, the actual significant effects were observed in core areas wherein individuals with ASD face difficulties, such as socialization, as was evident in the K-VABS-2, a widely used measure of adaptive behaviors in children with developmental disorders. These findings suggest the potential for conducting a confirmatory study using more focused outcome measurements.

Results on parenting stress showed that, within the intervention group, caregivers who used the training program the most reported higher levels of parenting stress (Table 3 and Figure S2 in Multimedia Appendix 1). This may imply that the more parenting stress a caregiver has, the more extensively they use the program.

Analyzing the results based on session completion rates, a more pronounced difference was observed between K-CBCL and K-VABS-2. While the scores measured by K-CBCL did not exhibit significant variations based on different completion rates, a notable difference was found in the K-VABS-2 scores. One of the key advantages of digital interventions is their relative freedom from constraints such as time, space, and cost. Therefore, a higher session completion rate would have been expected. However, considering that the session completion rate in this study was not as high as anticipated, our data suggest the need for complementary strategies to enhance user engagement. In doing so, if parental participation in this training program increases, our study's findings indicate the potential for further improvements in the K-VABS-2 scores.

Ongoing research suggests that a significant number of parents are interested in mobile app–based, evidence-based parent training programs, which do not rely on therapists because the content is standardized and can be easily accessed by parents anytime and anywhere [49]. Using these mobile app technologies can help bring key elements of evidence-based interventions to populations that may not otherwise be able to receive them.

This study had several limitations. First, it relied on parental opinions to assess children's behavior. Although the CGI-S and CGI-I were assessed by a physician, as the physician's evaluation relied on parental reports regarding the child, it was also influenced by the parents' opinions. Second, we observed a nonsignificant clinical effect of the treatment group on the primary outcome, which we attribute to indirect effects from other treatments the child was already receiving owing to this study's nature. Third, our study had low program participation rates. For studies based on mobile app interventions, there are challenges in promoting participant engagement in the program.

Conclusions

This study demonstrated the feasibility of using mobile devices for evidence-based parent training to reduce problem behaviors in children with ASD. In addition to the high prevalence of ASD and the high cost of treatment, there is a significant shortage of people to provide treatment, which is a barrier to treatment delivery. The accessibility and flexibility of mobile devices may make them a viable alternative as a parent education tool in providing early intervention for problem behaviors in children with ASD.

Acknowledgments

This work was supported by the Information & Communications Technology (ICT) research and development program of Ministry of Science and ICT/Institute for Information Communication Technology Planning and Evaluation (MSIT/IITP; 2022-0-00297, development and clinical validation of digital therapeutics for problematic behavior in children with autism).

Data Availability

The datasets produced or examined during this study are not openly accessible owing to the sensitivity of the information they contain and the potential risk of identifying individuals involved. Nonetheless, researchers who meet the necessary qualifications may be able to access these data following a structured request procedure that necessitates ethical clearance. Inquiries for data access should be directed to the corresponding author and will be evaluated individually.

Authors' Contributions

YRP and SBH were the chief investigators responsible for this study's design and execution and the decision to submit the results for publication. This study design was developed and approved by J Lee, J Lim, SBH, and YRP. J Lee, SBH, and YRP wrote the paper with approval from all authors and developed the statistical analysis plan. J Lee, SYJ, S Kim (LumanLab), J Lim, and YRP designed and wrote the delivery of the therapy. S Kang, S Kim (Seoul National University College of Medicine), and SBH ensured that the investigation was conducted per the institutional review board approval and plan. J Lee and S Kang accessed and verified the underlying data. All authors critically reviewed the paper for important intellectual content and read and approved the final paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Tables and figures for intervention and control groups and pre and post evaluation values.
[DOCX File , 219 KB - humanfactors_v11i1e52295_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 276 KB - humanfactors_v11i1e52295_app2.pdf]

References

1. Lord C, Elsabbagh M, Baird G, Veenstra-Vanderweele J. Autism spectrum disorder. Lancet 2018;392(10146):508-520 [FREE Full text] [doi: 10.1016/S0140-6736(18)31129-2] [Medline: 30078460]
2. Hong M, Lee SM, Park S, Yoon S, Kim Y, Oh I. Prevalence and economic burden of autism spectrum disorder in South Korea using national health insurance data from 2008 to 2015. J Autism Dev Disord 2020;50(1):333-339. [doi: 10.1007/s10803-019-04255-y] [Medline: 31630294]
3. Solmi M, Song M, Yon DK, Lee SW, Fombonne E, Kim MS, et al. Incidence, prevalence, and global burden of autism spectrum disorder from 1990 to 2019 across 204 countries. Mol Psychiatry 2022;27(10):4172-4180. [doi: 10.1038/s41380-022-01630-7] [Medline: 35768640]
4. Malik-Soni N, Shaker A, Luck H, Mullin AE, Wiley RE, Lewis MES, et al. Tackling healthcare access barriers for individuals with autism from diagnosis to adulthood. Pediatr Res 2022;91(5):1028-1035 [FREE Full text] [doi: 10.1038/s41390-021-01465-y] [Medline: 33767375]
5. Elder JH, Brasher S, Alexander B. Identifying the barriers to early diagnosis and treatment in underserved individuals with autism spectrum disorders (ASD) and their families: a qualitative study. Issues Ment Health Nurs 2016;37(6):412-420. [doi: 10.3109/01612840.2016.1153174] [Medline: 27070190]
6. Pickard KE, Ingersoll BR. Quality versus quantity: the role of socioeconomic status on parent-reported service knowledge, service use, unmet service needs, and barriers to service use. Autism 2016;20(1):106-115. [doi: 10.1177/1362361315569745] [Medline: 25948601]
7. Hartley SL, Sikora DM, McCoy R. Prevalence and risk factors of maladaptive behaviour in young children with autistic disorder. J Intellect Disabil Res 2008;52(10):819-829 [FREE Full text] [doi: 10.1111/j.1365-2788.2008.01065.x] [Medline: 18444989]
8. Mazurek MO, Kanne SM, Wodka EL. Physical aggression in children and adolescents with autism spectrum disorders. Res Autism Spectr Disord 2013;7(3):455-465. [doi: 10.1016/j.rasd.2012.11.004]
9. Maskey M, Warnell F, Parr JR, Le Couteur A, McConachie H. Emotional and behavioural problems in children with autism spectrum disorder. J Autism Dev Disord 2013;43(4):851-859. [doi: 10.1007/s10803-012-1622-9] [Medline: 22895777]
10. Lecavalier L, Leone S, Wiltz J. The impact of behaviour problems on caregiver stress in young people with autism spectrum disorders. J Intellect Disabil Res 2006;50(Pt 3):172-183. [doi: 10.1111/j.1365-2788.2005.00732.x] [Medline: 16430729]
11. Hayes SA, Watson SL. The impact of parenting stress: a meta-analysis of studies comparing the experience of parenting stress in parents of children with and without autism spectrum disorder. J Autism Dev Disord 2013;43(3):629-642. [doi: 10.1007/s10803-012-1604-y] [Medline: 22790429]
12. Pahnke J, Hirvikoski T, Bjureberg J, Bölte S, Jokinen J, Bohman B, et al. Acceptance and commitment therapy for autistic adults: an open pilot study in a psychiatric outpatient context. J Contextual Behav Sci 2019;13:34-41. [doi: 10.1016/j.jcbs.2019.04.002]

13. Szabo TG. Acceptance and commitment training for reducing inflexible behaviors in children with autism. J Contextual Behav Sci 2019;12:178-188. [doi: 10.1016/j.jcbs.2019.03.001]
14. Gaudiano BA, Herbert JD, Hayes SC. Is it the symptom or the relation to it? Investigating potential mediators of change in acceptance and commitment therapy for psychosis. Behav Ther 2010;41(4):543-554 [FREE Full text] [doi: 10.1016/j.beth.2010.03.001] [Medline: 21035617]
15. Geretsegger M, Elefant C, Mössler KA, Gold C. Music therapy for people with autism spectrum disorder. Cochrane Database Syst Rev 2014;2014(6):CD004381 [FREE Full text] [doi: 10.1002/14651858.CD004381.pub3] [Medline: 24936966]
16. Marquez-Garcia AV, Magnuson J, Morris J, Iarocci G, Doesburg S, Moreno S. Music therapy in autism spectrum disorder: a systematic review. Rev J Autism Dev Disord 2021;9(1):91-107. [doi: 10.1007/s40489-021-00246-x]
17. Ringdahl JE, Kopelman T, Falcomata TS. Applied behavior analysis and its application to autism and autism-related disorders. In: Handbook of Applied Behavior Analysis for Children with Autism. Autism and Child Psychopathology Series. Switzerland: Springer, Cham; 2023:37-58. [doi: https://doi.org/10.1007/978-3-031-27587-6_3]
18. Foxx RM. Applied behavior analysis treatment of autism: the state of the art. Child Adolesc Psychiatr Clin N Am 2008;17(4):821-834. [doi: 10.1016/j.chc.2008.06.007] [Medline: 18775372]
19. Cooper JO, Heron TE, Heward WL. Applied Behavior Analysis. London, UK: Pearson; 2007.
20. Alberto P, Troutman AC, Axe JB. Applied behavior analysis for teachers. London, UK: Pearson; 2006.
21. Matson ML, Mahan S, Matson JL. Parent training: a review of methods for children with autism spectrum disorders. Res Autism Spectr Disord 2009;3(4):868-875. [doi: 10.1016/j.rasd.2009.02.003]
22. Bearss K, Burrell TL, Stewart L, Scahill L. Parent training in autism spectrum disorder: what's in a name? Clin Child Fam Psychol Rev 2015;18(2):170-182 [FREE Full text] [doi: 10.1007/s10567-015-0179-5] [Medline: 25722072]
23. Sato SK. Parental stress and effective parent training for parents of individuals diagnosed with ASD. In: Matson JL, Sturmey P, editors. Handbook of Autism and Pervasive Developmental Disorder. Autism and Child Psychopathology Series. Cham, Switzerland: Springer; 2022.
24. Wang C, Lee C, Shin H. Digital therapeutics from bench to bedside. NPJ Digit Med 2023;6(1):38 [FREE Full text] [doi: 10.1038/s41746-023-00777-z] [Medline: 36899073]
25. Shah N, Velez FF, Colman S, Kauffman L, Ruetsch C, Anastassopoulos K, et al. Real-world reductions in healthcare resource utilization over 6 months in patients with substance use disorders treated with a prescription digital therapeutic. Adv Ther 2022;39(9):4146-4156 [FREE Full text] [doi: 10.1007/s12325-022-02215-0] [Medline: 35819569]
26. Jung Y, Kim Y, Oh Y. 2023 Korea Media Panel Survey. Korea Information Society Development Institute. URL: https://stat.kisdi.re.kr/kor/board/BoardList.html?board_class=BOARD15&report_div=R [accessed 2024-05-24]
27. Mandelberg J, Laugeson EA, Cunningham TD, Ellingsen R, Bates S, Frankel F. Long-term treatment outcomes for parent-assisted social skills training for adolescents with autism spectrum disorders: the UCLA PEERS program. J Ment Health Res In in Intel Dis 2013;7(1):45-73. [doi: 10.1080/19315864.2012.730600]
28. Julious SA. Sample size of 12 per group rule of thumb for a pilot study. Pharm Stat 2005;4(4):287-291. [doi: 10.1002/pst.185]
29. Allday RA, Spriggs AD, Jones ME, Noel C. Mobile devices in the context of applied behavior analysis: a multipurpose tool. In: Matson JL, editor. Handbook of Applied Behavior Analysis. Autism and Child Psychopathology Series. Cham, Switzerland: Springer; 2023:837-856. [doi: 10.1007/978-3-031-19964-6_43]
30. Gerow S, Radhakrishnan S, Davis TN, Zambrano J, Avery S, Cosottile DW, et al. Parent-implemented brief functional analysis and treatment with coaching via telehealth. J Appl Behav Anal 2021;54(1):54-69. [doi: 10.1002/jaba.801] [Medline: 33325053]
31. Zaine I, Benitez P, Rodrigues KRDH, Pimentel MDGC. Applied behavior analysis in residential settings: use of a mobile application to support parental engagement in at-home educational activities. Creat Educ 2019;10(8):1883-1903. [doi: 10.4236/ce.2019.108136]
32. SungBong L, Eunkyung K, Hyesook P, Moonbong Y, SungBong K. Applied Behavior Analysis. Seoul, Korea: Hakchisa; 2019.
33. Smith N. Parent training for disruptive behavior: the RUBI autism network, clinical manual. Karen Bearss, Cynthia R. Johnson, Benjamin L. Handen, Eric Butter, Luc Lecavalier, Tristram Smith, and Lawrence Scahill. Soc Work 2021;66(2):176-177. [doi: 10.1093/sw/swab010]
34. Behavior therapy guidelines and treatment manuals. Seoul National Hospital and Yonsei University. 2019. URL: http://www.snmh.go.kr [accessed 2024-09-25]
35. Lord C, Rutter M. Autism Diagnostic Observation Schedule, Second Edition. Torrance, CA: Western Psychological Services; 2012.
36. McCrimmon A, Rostad K. Test review: autism diagnostic observation schedule, second edition (ADOS-2) manual (Part II): toddler module. J Psychoeduc Assess 2014;32(1):88-92. [doi: 10.1177/0734282913490916]
37. Schopler E, Reichler RJ, DeVellis RF, Daly K. Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). J Autism Dev Disord 1980;10(1):91-103. [doi: 10.1007/BF02408436] [Medline: 6927682]
38. Shin MS, Kim YH. Standardization study for the Korean version of the childhood autism rating scale: reliability, validity and cut-off score. Korean J Clin Psychol 1998;17:1-15 [FREE Full text]

39. Lee H, Oh KJ, Hong KE, Ha EY. Clinical validity study of Korean CBCL through item analysis. Korean J Child Adol Psychiatr 1991;2(1):138-149 [FREE Full text]
40. Rutter M, Bailey A. Social Communication Questionnaire (SCQ). Los Angeles, CA: Western Psychological Services; 2003.
41. Kim JH, Sunwoo HJ, Park SB, Noh DH, Jung YK, Cho IH, et al. A validation study of the Korean version of social communication questionnaire. J Korean Acad Child Adolesc Psychiatry 2015;26:197-208 [FREE Full text] [doi: 10.5765/jkacap.2015.26.3.197]
42. Hwang ST, Kim J, Hong S. Standardization study of the Korean Vineland Adaptive Behavior Scales-Ⅱ (K-Vineland-II). Korean J Clin Psych 2015 Nov;34(4):851-876 [FREE Full text] [doi: 10.15842/kjcp.2015.34.4.002]
43. Busner J, Targum SD. The clinical global impressions scale: applying a research tool in clinical practice. Psychiatry (Edgmont) 2007;4(7):28-37 [FREE Full text] [Medline: 20526405]
44. Abidin RR. Parenting Stress Index - Fourth Edition (PSI-4). Lutz, FL: Psychological Assessment Resources; 2012.
45. Chung KM, Lee S, Lee C. Standardization study for the Korean version of Parenting Stress Index Fourth Edition (K-PSI-4). Korean J Gen Psychol 2019 Jun;38(2):247-273. [doi: 10.22257/kjp.2019.6.38.2.247]
46. Bearss K, Johnson C, Smith T, Lecavalier L, Swiezy N, Aman M, et al. Effect of parent training vs parent education on behavioral problems in children with autism spectrum disorder: a randomized clinical trial. JAMA 2015;313(15):1524-1533 [FREE Full text] [doi: 10.1001/jama.2015.3150] [Medline: 25898050]
47. Scahill L, Bearss K, Lecavalier L, Smith T, Swiezy N, Aman MG, et al. Effect of parent training on adaptive behavior in children with autism spectrum disorder and disruptive behavior: results of a randomized trial. J Am Acad Child Adolesc Psychiatry 2016;55(7):602-609.e3 [FREE Full text] [doi: 10.1016/j.jaac.2016.05.001] [Medline: 27343887]
48. Scahill L, McDougle CJ, Aman MG, Johnson C, Handen B, Bearss K, Research Units on Pediatric Psychopharmacology Autism Network. Effects of risperidone and parent training on adaptive functioning in children with pervasive developmental disorders and serious behavioral problems. J Am Acad Child Adolesc Psychiatry 2012;51(2):136-146 [FREE Full text] [doi: 10.1016/j.jaac.2011.11.010] [Medline: 22265360]
49. Baumel A, Pawar A, Mathur N, Kane JM, Correll CU. Technology-assisted parent training programs for children and adolescents with disruptive behaviors: a systematic review. J Clin Psychiatry 2017;78(8):e957-e969. [doi: 10.4088/JCP.16r11063] [Medline: 28493653]

Abbreviations

ABA: Applied Behavior Analysis
ADOS-2: Autism Diagnostic Observation Schedule-2
ASD: autism spectrum disorder
CGI-I: Clinical Global Impression Global Improvement
CGI-S: Clinical Global Impression–Severity of Illness
CONSORT: Consolidated Standards of Reporting Trials
ITT: intention-to-treat
K-CARS-2: Korean Childhood Autism Rating Scale-2
K-CBCL: Korean Child Behavior Checklist
K-PSI-4-SF: Korean Parenting Stress Index Fourth Edition–Short Form
K-SCQ: Korean version of the Social Communication Questionnaire
K-VABS-2: Korean Vineland Adaptive Behavior Scale-2
SCQ: Social Communication Questionnaire

Edited by A Kushniruk, E Borycki; submitted 29.08.23; peer-reviewed by C McGrath, AT Cheung; comments to author 15.05.24; revised version received 02.08.24; accepted 13.09.24; published 28.10.24.

Please cite as:
Lee J, Lim J, Kang S, Kim S, Jung SY, Kim S, Hong SB, Park YR
Mobile App–Assisted Parent Training Intervention for Behavioral Problems in Children With Autism Spectrum Disorder: Pilot Randomized Controlled Trial
JMIR Hum Factors 2024;11:e52295
URL: https://humanfactors.jmir.org/2024/1/e52295
doi:10.2196/52295
PMID:39466295

Original Paper

Characteristics, Opportunities, and Challenges of Osteopathy Based on the Perceptions of Osteopaths in Austria: Qualitative Interview Study

Jonas Manschel[1], MSc; Jan Porthun[2], MMMSc; Ulrike Winkler[3], Dr phil; Jean Marie A T Beuckels[1,4], Dr rer medic; David Martin[1,5], Dr med

[1]Institute of Integrative Medicine, Department of Medicine, Health Faculty, University Witten/Herdecke, Herdecke, Germany

[2]Norwegian University of Science and Technology, Campus Gjøvik, Gjøvik, Norway

[3]Bundeswehr University Munich, Neubiberg, Germany

[4]Department of Osteopathy, Faculty of Health and Social Sciences, Hochschule Fresenius, Munich, Germany

[5]Tübingen University Children's Hospital, Tübingen, Germany

Corresponding Author:
Jan Porthun, MMMSc
Norwegian University of Science and Technology, Campus Gjøvik
Teknologivn 22
Gjøvik, 2815
Norway
Phone: 47 95007916
Email: jan.porthun@ntnu.no

Abstract

Background: There are no uniform regulations for the osteopathic profession in Europe. It is subject to country-specific regulations defining who shall be allowed to practice osteopathy and which qualification shall be required. In recent years, legal regulations have been established in several European countries for the profession of osteopathy; however, these are also still pending for Austria. Currently, physiotherapists and physicians with osteopathic training are practicing osteopathy in Austria.

Objective: This study aims to examine the characteristics, challenges, and opportunities of osteopaths in Austria.

Methods: Guideline-based interviews with osteopaths (N=10) were conducted. The different research questions were examined using a qualitative content analysis.

Results: The study provided a differentiated insight into the professional situation of osteopaths in Austria. The most important result was that all interviewees unanimously supported a legal regulation of their profession. However, owing to their different professional self-image—on the one hand, individuals working on a structural basis, and, on the other hand, individuals working on a cranial or biodynamic basis—they were able to imagine a uniform professional regulation only to a limited extent. Additional topics for the interviewed osteopaths in Austria were the quality assurance of training and the urgent need for scientific research. Furthermore, the study also dealt with the influence of the COVID-19 pandemic on daily practice and on education and training in osteopathy.

Conclusions: This study is a pioneering study with regard to systematic basic research on osteopathy in Austria. The obtained results and the newly acquired research questions not only have the potential to serve as a basis for further studies but also provide insight into the working and professional situation of osteopaths in Austria for universities, schools, professional associations, politics, and—last but not least—all interested parties.

International Registered Report Identifier (IRRID): RR2-10.2196/15399

(JMIR Hum Factors 2024;11:e45302) doi:10.2196/45302

KEYWORDS

osteopathy; osteopath; osteopaths; osteopathic profession; health care system; Austria

Introduction

Osteopathy is a manual treatment method, the principles of which are based on its own philosophy and the consideration and treatment of special structure-function relationships in the human body [1]. Since osteopathy was established and as long as it has been applied, both its methods and the professional competence of osteopaths have been the subject of controversy among medical and therapeutic specialists. The European Committee for Standardization has defined osteopathy as a holistic, patient-centered, manual treatment method based on the interactions between the structure and function of the body and the body's self-healing ability [2].

There is no uniform European or international regulation regarding who is allowed to practice osteopathy and which qualifications are required. However, an increasing number of European countries are developing occupational laws for osteopaths. So far, 12 European countries including Cyprus, Denmark, Finland, France, Iceland, Liechtenstein, Luxembourg, Malta, Norway, Portugal, Switzerland, and the United Kingdom have adopted legal regulations regarding the practice of osteopaths [3].

In the German-speaking countries, there is no uniform picture of the profession. In contrast to Switzerland, a legal basis for the profession of osteopathy does not exist in Germany or Austria. In Austria, physicians and physiotherapists trained in osteopathy practice as osteopaths. Physicians are allowed to practice osteopathy without any restrictions, whereas physiotherapists are only allowed to practice osteopathy upon medical assignment [4]. According to the Austrian Society for Osteopathy (Österreichische Gesellschaft für Osteopathie; OEGO), approximately 2000 osteopaths practiced in Austria in 2022.

Studies about osteopathic identity are progressing internationally. However, the various legal regulations and intraprofessional conflicts make it difficult to perceive a collective identity [5]. Especially in countries where osteopathy is not regulated by law, the data about osteopathic practitioners are considered to be weak. However, quantitative studies that have surveyed the population of osteopaths with regard to work status, training, professional identity, or characteristics of clinical practice such as the typical patient profile and the use of diagnostic and treatment modalities exist already. In Austria, 2 surveys of osteopaths have been conducted in the past as part of final theses [6,7]. In 2022, the results of the Osteopathic Practitioners, Estimates, and Rates survey were also published for Austria, thus creating a solid data basis about osteopathic practitioners in Austria for the first time. The typical osteopath was defined as female, aged between 40 and 49 years, self-employed, worked before as a physiotherapist, trained in osteopathy part time, and successfully completed a master's degree [8]. However, there is a lack of studies with qualitative designs to capture and examine the work of osteopaths in German-speaking countries in more detail.

The overall aim of this study was to make a substantial contribution to the largely unexplored profession of osteopathy in the German-speaking countries. Structured, basic research was necessary to implement this project. The first steps were taken in the framework of the study, "Characteristics, Opportunities, and Challenges of Osteopathy (COCO) in the Perceptions of Osteopaths in Germany, Austria, and Switzerland: Protocol for a Comprehensive Mixed Methods Study." The study protocol was published in *JMIR Research Protocols* in 2019. The Characteristics, Opportunities, and Challenges of Osteopathy (COCO) project investigates how osteopaths in Germany, Austria, and Switzerland distinguish themselves from other medical professions and the characteristics of their work.

This study is a partial study of the COCO project, with a focus on the situation of osteopaths in Austria. Osteopaths practicing in Austria were asked about their professional profile and their professional practice. The following questions were of particular interest: (1) How do osteopaths from Austria describe osteopathy? (2) What are the challenges faced by osteopaths in Austria? and (3) What opportunities do the interviewees see for osteopathy in Austria?

Methods

Design

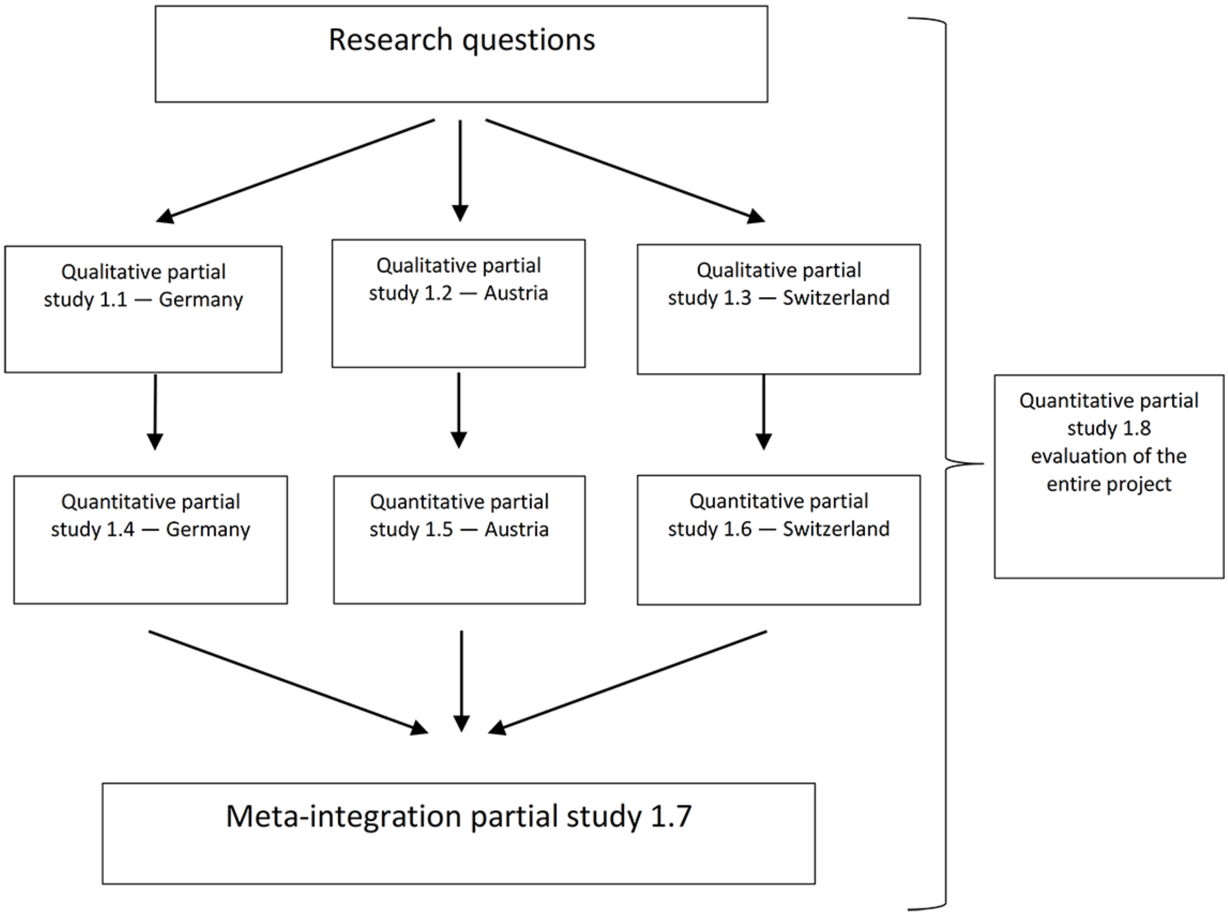
This qualitative study included the planning and implementation of guideline-based interviews with osteopaths practicing in Austria. Subsequently, a qualitative content analysis was performed according to Mayring [9]. A qualitative research design was selected to obtain questions relevant to the project that had not been considered before and views about the topic that had not yet been taken into consideration. The target of this qualitative partial study was the development of hypotheses. Accordingly, a relatively small sample of 8 to 10 participants could be used, because the results obtained shall be examined in subsequent studies with respect to their general validity using a quantitative study design [10].

To ensure the reporting quality regarding the research methodology of this qualitative study, COREQ (Consolidated Criteria for Reporting Qualitative Research) was used [11]. A checklist including the COREQ items taken into consideration has been attached to the paper (Multimedia Appendix 1). The registration identifier of the study is the International Registered Report Identifier: PRR2-10.2196/15399.

Ethical Considerations

This study (corresponding to partial study 1.2 in Figure 1), led by DM, has received ethics approval (S-287/2020) from the ethics committee of the University of Witten/Herdecke, Germany. Participants were not compensated for their participation.

Figure 1. Characteristics, Opportunities, and Challenges of Osteopathy project flowchart.



Setting and Sampling

Related to the research topic, osteopaths in Austria were questioned through guideline-based interviews. All the interviewed osteopaths (10/10, 100%) had completed at least 4 years of training as osteopaths and practiced as osteopaths in Austria. Instead of asking individual osteopaths to participate, OEGO was contacted with the study project itself, thus avoiding cold-calling. In this way, the criterion of comprehensive training in osteopathy was fulfilled, because otherwise, the participants could not be members of the professional association. This procedure ensured that the participants had provided evidence of their competence. To obtain the widest range of views, the sample was intended to show a high degree of diversity among the participants. Therefore, a further criterion for the whole group of participants was sex distribution according to the population of osteopathic practitioners in Austria. According to the OEGO's membership register, two-thirds of practicing osteopaths in Austria are female and one-third are male. Moreover, care was taken to ensure that the residences and workplaces of the participants were subject to as wide a geographical distribution as possible, so that district-specific phenomena could be excluded. As osteopathy is not an independent profession in Austria, the participants should include the different occupational groups that practice osteopathy. Inclusion and exclusion criteria were subsequently formulated (Textbox 1).

OEGO forwarded the contact details of 11 osteopaths. Appointments for an interview were made with 8 (73%) of the 11 osteopaths. No appointment could be made with 1 osteopath during the study period. From 1 other osteopath, no response to the request was received. Another 1 osteopath did not want to participate in the study; 2 new osteopaths were suggested by the osteopaths themselves. A total of 10 interviews were thus conducted.

All the participants were contacted via email and received a letter containing information about the course of the study, declaration of consent to participate in the study, and data protection declaration. The entire participation in the research project and the answering of individual questions was on a voluntary basis; nonparticipation did not lead to any disadvantages for the participants. The participants always had the option to end the survey (eg, in the case of unexpected, stressful questions). Through the format of web-based survey, any increased risk of infection for the participants owing to the COVID-19 pandemic could be excluded. Explicit cancellation criteria for the project were not set.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- The individual has completed 4 years of osteopathic training.
- The individual is currently practicing osteopathy.
- The individual has provided consent to participate in the study.

Exclusion criteria

- The individual cannot be interviewed within the examination period.
- The individual does not have the technical equipment required for participation in a web-based survey.
- The individual demands compensation.

Development of the Interview Guidelines and Data Collection

The guideline-based interview was chosen as a suitable research tool, because the aim of the data collection was to obtain concrete statements about the practice of osteopathy in Austria. In addition, the use of a guideline increased the comparability of the individual data sets [12]. Furthermore, this procedure avoided the possibility that essential aspects of the research question might be overlooked in the interviews [13]. The interview guideline was developed by JM on the basis of the research questions and 2 previous qualitative, partial studies of the COCO project [14,15]. The interview guideline was developed by JM on the basis of the research problem and 2 previous qualitative, partial studies of the COCO project [14,15].

The questions were formulated as open questions (Textbox 2) and arranged according to the groups of topics (Textbox 3). In addition, sociodemographic data about the participants were collected. Alternative questions were prepared to be able to react flexibly to the course of the interview and to respond to the potential needs of the participants. To maintain the flow of the conversation, additional questions were developed in advance. Before beginning the interviews, a test interview was conducted with a German osteopath to test the interview guideline in practice and to improve the interview technique. Finally, the questionnaire was discussed and adapted together with JP, an osteopath with experience in qualitative research.

Textbox 2. Example interview questions.

- Where did you first hear about the osteopathic profession?
- How would you define osteopathy?
- In your opinion, what differentiates osteopathy from other professions?
- What does a typical osteopathic treatment look like for you?
- Should osteopath be its own profession?
- Are there any difficulties or problems that you face in your daily work life as an osteopath?
- Has the corona pandemic changed anything in your daily practice?

Textbox 3. Contents of the interview guideline.

Training and work in osteopathy

- Educational background
- Motivation
- Training structure
- Work experience
- Acquisition structure
- Fields of activity

Characteristics of osteopathy

- Definition
- Properties
- Differentiation of professional profile
- Competences
- Features
- Limits

Challenges of osteopathy

- Health value
- Employment policy
- Obligations
- Restrictions
- Conflicts

Chances and opportunities in osteopathy

- Perspective
- Research
- Desires

Data Collection

Data were collected from November 29, 2021, to January 26, 2022. In total, 10 interviews were conducted. The interviews were conducted using the Zoom software (Zoom Video Communications; audio and video were recorded). Apart from research economy and temporal and local flexibility, interviews were primarily conducted on the web to protect the participants from infection during the COVID-19 pandemic. No other individuals were present during the interviews. No interview was repeated. The interviews were recorded as a video file for transcription. The participants agreed in writing to the archiving of the files until the end of the publication activities or up to a maximum of 5 years after data collection. The files were protected against unauthorized access and stored and evaluated on local data carriers of a password-protected computer. Only encrypted files were transferred among the study colleagues. Upon completion of the study project, the recordings of the interviews shall be deleted irrevocably. All the interviews were conducted by JM in German. The interviewer is male, holds a master of science degree, and has already published in 2019 within the COCO project. At the time of the survey, he worked independently as a physiotherapist in his own practice in Germany and was a doctoral student at the University of Witten/Herdecke. For this research project, JM was trained within a 3-part seminar at the Freie Universität Berlin regarding the collection and evaluation of qualitative data and the conduction of interviews. There was no previous personal relationship with any of the interviewees. The recordings were transcribed by JM. Transcription was performed according to pre-established rules, which were consistently observed, as there are no generally accepted transcription rules [16]. The rules were based on the transcription rules of Kuckartz and Rädiker [17] for computer-assisted evaluation. The participants were sent the transcript to gather their comments, if any, and to receive their final approval.

Data Analysis

On the basis of the results of the interview studies already conducted within the COCO project, deductive (ie, indirectly theory-driven) categories were formed first. It is indirect because the categories are descriptive and their definition is not the basis of a theory-driven description. As a first step, classical deductive codes were derived from the interview guideline (deductive category application). The transcripts were analyzed in the original language by means of content structuring. With the help of the MAXQDA 2022 software (VERBI Software) [18], quotations that were relevant for the abovementioned research questions were categorized. Owing to an extensive interest in further knowledge, the preselected segments were expanded or specified by means of inductive categorization to deduce further important aspects (deductive-inductive categorization) [17]. The quotations were summarized in categories, which were subsequently grouped into high-level categories. The main strategies of data analysis are described in Textbox 4.

Textbox 4. Data analysis strategies.

Coding

- After developing an initial category system together (first, deductive; second, inductive), 2 complete material iterations were performed by 2 separate evaluators

Discussion

- Comparison and discussion of the results, with the main focus on the integration of different perspectives and the elimination of ambiguities

Quality control

- To check the quality of the category system (intercoder reliability) with its coding rules by means of the Cohen κ coefficient

Final iteration

- After the quality control step, a final, complete material iteration was performed by an evaluator using the final category system

Shared Coding

A first interview was coded by 2 evaluators together (JM and UW), and the category system was inductively expanded. An initial category system was developed, and coding rules were determined on the basis of 2 partial studies (eg, Figure 1—partial studies 1.1 and 1.3) [14,15] and the interview questions and guidelines developed specifically for this partial study. The result was an initial, deductive category system, and the first coding rules were defined. The deductively determined categories were based on the research questions and increased based on inductive subcategories during the evaluation. Then, the first iteration was run with the entire material, followed by further inductive categorization (UW). A second evaluator (JM) ran the second iteration of the entire material based on the category system resulting from the first material iteration. The results were then compared and discussed, with the main focus being on the integration of different perspectives and the elimination of ambiguities. The code definitions and anchor examples were also revised in this step. Finally, the category system was standardized. This shared coding from the beginning of the study was intended to increase the intersubjectivity of the statement.

Quality Control

Subsequently, of the 10 interviews, 3 (30%) were selected via lot procedure as a subsample to check the quality of the category system with its coding rules by means of the Cohen κ coefficient. The determination of the Cohen κ coefficient is a method for checking the intercoder reliability [19]. UW received the 3 allotted coded interviews from JM. The segment boundaries of the encodings were retained during this iteration. In this way, the evaluator was able to recognize which text parts were encoded and arrange these according to the categories. Without random adjustment, a match of 62.1% between the 2 encodings of this subsample was reached. The randomly adjusted coefficient was 0.61. Pursuant to Altman [20], accordance is considered as “good” in the case of a κ coefficient of 61-80. After this quality control, as there was good accordance; a final complete material iteration was performed by an evaluator (JM) using the final category system.

Results

Overview

The result of this study was a system of categories in which the statements of the participants were classified and subcategorized according to the research questions regarding the groups of topics, such as characteristics, challenges, and opportunities of the osteopaths practicing in Austria. Sociodemographic data and more general statements about the practice of osteopathy were also classified in the main category, “training and work in osteopathy.” Out of current concern, the osteopaths were also asked about the influence of the COVID-19 pandemic on their work. On the basis of the interesting statements, we decided to dedicate a special category to this topic.

The final category system consists of 71 categories with a total of 783 encoded text passages. A definition was created for each individual category, and a quotation was recorded as an anchor example.

Only a part of these results could therefore be described in this paper. The printed quotations have been translated into English and serve for illustration purposes only. The question whether individual statements of the participants (O1 to O10) represent the entire population of osteopaths practicing in Austria shall be subject to further investigation.

Training and Work in Osteopathy

A total of 10 osteopaths practicing in Austria were interviewed—6 (60%) women and 4 (40%) men. Of the 10 osteopaths, 3 (30%) were physicians and 7 (70%) were physiotherapists. Of the 10 interviewees, 9 (90%) worked independently in their own practices and 1 (10%) had an employment relationship at the time of the interview. Only 10% (1/10) had other employees. The longest interview lasted 61 minutes, and the shortest interview lasted 29 minutes. The average duration of the interviews was 46 (SD 9.2) minutes. The participants' characteristics are described in Table 1.

Table 1. Participants' characteristics (N=10).

Characteristics	Participants, n (%)
Sex	
Male	4 (40)
Female	6 (60)
Profession	
Physician	3 (30)
Physiotherapist	7 (70)
Degree	
Diploma in osteopathy	3 (30)
Master of science	6 (60)
Nondegree	1 (10)
Training facility	
Vienna School of Osteopathy, Vienna, Austria	8 (80)
European College of Osteopathy, Munich, Germany	1 (10)
The International Academy of Osteopathy, Darmstadt, Germany	1 (10)
Clinical experience (y)	
1-5	2 (20)
6-15	5 (50)
>15	3 (30)

Characteristics of Osteopathy

On the basis of the descriptions provided by the participants regarding the properties of osteopathy, the following categories were developed: "definition of osteopathy," "patient profile," "anchor personalities and literature," and "limits of the treatment technique."

Definition of Osteopathy

When asked about the definition of osteopathy, several participants had difficulties in explaining the concept:

> *Yes, it's really a difficult question.* [O1; item 35]

The explanation of the term was mostly based on the manual work, origin of the word, differentiation or overlapping with other professional groups, or citation of definitions of third parties. Often, reference was made to the philosophy of osteopathy, holism of the treatment method, and activation of self-regulating forces. There was no uniform definition of osteopathy among the answers of the interviewees.

Patient Profile

Most of the interviewed osteopaths treat patients of all ages:

> *Oh, everything actually, there are patients of all ages. From...three-month-old babies to over 90-year-old men, women, so I couldn't paint a typical picture.* [O2; item 54]

Common indications for child treatments mentioned by the interviewed osteopaths are sleep disorders, torticollis, scoliosis, asthma, abdominal colic, and plagiocephaly. The treatment of adults was mainly based on the diagnosis or leading symptoms from the orthopedic area: back and neck pain and joint pain. Neurological diseases such as Parkinson disease were also mentioned. Moreover, patients were regularly treated for headache, migraine, tinnitus, chronic pain, craniomandibular dysfunction, abdominal pain, and hormonal imbalances or if wishing to become pregnant. However, internal diseases such as sinusitis, bronchitis, and cystitis were also treated by the physician O2 on the basis of osteopathic methods.

The indication of the treatment was usually given by the patients' treating physician:

> *Patients are often assigned by the doctors, meaning that the doctor writes a prescription with a recommendation to contact a certain therapist.* [O8; item 51]

A participant working as a general practitioner in addition to his osteopathic activity also acquired patients during his regular consultation as a physician:

> *And actually, many of those who go to the general practitioner's clinic in the village, they come to me, too.* [O2; item 54]

It appears that, in general, a broad medical field was covered by osteopathy. The selection of the appropriate therapist seemed to depend on personal recommendations of others, on the therapeutical possibilities in the patients' vicinity, and on the training or specializations of the osteopaths:

> *Because the patients who come to me come by word of mouth, yes.* [O5; item 47]

> *Everything else is, I think, very average, that is, all the people who come to me do so because I am in their vicinity.* [O3; item 38]

> *Patients who travel a long way mainly come because of endocrinological, metabolic problems..., gynaecological problems..., that is to say, where...the focus of my...training has mainly been during recent years.* [O3; item 38]

Anchor Personalities and Literature

It was noticeable that many participants referred to other individuals when answering questions about osteopathy in theory and practice. These "anchor personalities" seem to have a great influence on the self-image of osteopaths in Austria and have therefore been included in a separate category. Both historical personalities from the history of osteopathy and currently active osteopaths were repeatedly mentioned:

> *But there will always be people who really care about this innermost quality of osteopathy.... And just as osteopathy has developed from Still to Sutherland, Becker, Viola Frymann...and all their names..., or Mitchell and Jim Jealous now..., so it will continue to develop.* [O5; item 84]

With regard to the self-study of osteopaths, primarily only the German-language journals of osteopathy were mentioned as reading material. Of the 10 participants, only 1 (10%) indicated that they regularly read an English journal:

> *Yes, I regularly read the two journals, DO and Osteopathische Medizin.* [O5; item 88]

Limits of Osteopathy

There was agreement regarding the limits of osteopathy. The primary treatment of structural injuries or the care of patients with cancer without medical supervision were clearly mentioned by the interviewees as limits of osteopathic activity. However, patients were given osteopathic treatment nevertheless:

> *Yes, of course I see the limits in the pathologies that are there. If there is actually...an osteoarthritis that is simply there and will not vanish, osteopathy shall certainly have its limits; one can perhaps relieve the pain, but the osteoarthritis cannot be cured by osteopathy, now can it. I also see limits in some diseases.* [O7; item 56]

O9 differentiated upon request that it is not the diagnosis that is decisive for the objective of treatment when it comes to whether osteopathic treatment is indicated or contraindicated:

> *Well, that depends on the objective.... It all depends totally on the objective. If I say that I want to treat coxarthrosis curatively, I think that we shall soon reach our limits with osteopathy; if we do a control X-ray after six months, we shall see that it is still coxarthrosis.... But if I say that I want to improve the quality of life, I would treat them nevertheless. The question is what the objective is.* [O9; item 41]

O4 brought another aspect to mind:

> *We must not exceed the limits...of our own competence. That is very important. Unfortunately, many colleagues do this by suddenly giving dietary recommendations, by suddenly recommending medicines...or by talking patients out of taking medicines. Especially now when it comes to vaccination.* [O4; item 75]

Challenges of Osteopathy

The challenges mentioned for the field of osteopathy can be classified mainly into the categories, "identity problem," "disagreements within the osteopath community," "research," "training quality management," and "conflicts and difficulties."

Identity Problem

One of the great challenges of osteopathy is its unclear definition and lack of differentiation from other professional groups. As O6 clearly pointed out, osteopathy has an identity problem:

> *I would say that the identity problem is the most important issue...The definition is the most difficult question and no one can answer that. And if we can't answer that and don't deal with it, how can we argue what we are if we ourselves don't know exactly what we are.* [O6; item 49]

Disagreements Within the Osteopathic Community

The identity problem or the problem of missing a uniform professional self-image might be based on disagreements within the osteopathic community described by several participants. Overall, 2 groups can be identified among practicing osteopaths: on the one hand, the structurally working osteopaths and, on the other hand, the cranially or biodynamically working osteopaths:

> *Yes, there is really a gap between biodynamic osteopathy and structural osteopathy.* [O7; item 104]

This conflict might be decisive not only in terms of a common definition but also with respect to a possible recognition of osteopathy in terms of professional policy. O9, who was involved in professional policy, feared that these disputes might even prevent recognition:

> *The problem concerning regulation is - and that's simply the case now and that's also the elephant in the room about which no one is talking -...the problem with regulation has always been cranio.... You cannot say it openly, but it was always the problem of craniosacral therapy, no matter who I talked to.* [O9; item 93]

Those participants who worked biodynamically were more critical toward regulation:

> *If one tries now to take this out of this mental...source,... I see the risk that it is practically shifted into evidence-based, as important as that is, well, but only into evidence-based, visible and perceptible dimensions, then osteopathy shall lose its*

> *soul from my point of view.... And that's actually the greatest threat to osteopathy for me.* [O5; item 64]

Research

In this context, O1 pointed out that evidence-based research can only substantiate a certain part of osteopathy scientifically, whereas other aspects might be lost:

> *A good scientific basis in order to argue how...many benefits osteopathy has...- in the end,... academisation probably cannot be avoided and will certainly be necessary. Even if all these developments are not entirely without risks. That is, the risk of losing sight of the holistic aspects of osteopathy.* [O1; item 81]

Training Quality Management

Many osteopaths considered existing weaknesses in the training courses and their structures as a further obstacle. According to O4, a central aspect is the inadequate teaching of the skills for scientific work at osteopathy institutes and, thus, the lack of evidence-based research in osteopathy:

> *[Oh my] the training...We should learn from the beginning, not only during the last year when we have to write a Master Thesis, we should learn from the beginning what it means to work in an evidence-based way, to do research...It works in physiotherapy and is continuously getting better there, but in osteopathy...At the beginning, we never learn to deal with available studies, it is a matter of training. From the beginning, not only during the fifth year shortly before the Master Thesis, we should have the first lessons in statistics.* [O4; item 113]

Conflicts and Difficulties

The lack of clarity regarding the profession is evident in the differentiation with respect to other professional groups. With regard to the settlement for osteopathic treatment with health insurance providers, several participants also reported potential for conflicts. Osteopathic treatment is often provided on the grounds of a physiotherapeutic prescription by the physician. The reasons are the economic pressure on the practices or the social situation of the patients:

> *Many colleagues work as physiotherapists, they also charge for osteopathy as physiotherapy, and yes, they are refunded in this way. And as a result,...osteopathy is also a little...less in the focus than it should be. I've been working for 20 years now, I'm only writing osteopathic invoices.... But...of course...I understand the problem. If somebody has fewer patients...and...has to charge for...physiotherapy, I absolutely understand the situation.... But...these problems are of course...long-burning issues.* [O5; item 78]

> *No, [the bill] of course says physiotherapy and remedial massage, because otherwise the patient doesn't get his/her money from the insurance company. For the insurance company, well, this is ok or it is tolerated. I have already received the feedback from many patients that they told the company that they went to an osteopath, and the health insurance said that of course that can't be billed, [but] we shall write physiotherapy and remedial massage and then that's it.* [O8; item 55]

When asked about the challenges for osteopaths in general, the physicians working as osteopaths did not report any difficulties related to their practice. A physician and osteopath, in contrast, was aware of the potential for conflict:

> *Yes, of course I know that. I have a bonus, because I'm simply a doctor. And of course, osteopaths that aren't doctors have greater difficulties and are often rejected,...well,...because they are no medical doctors in a manner of speaking and...there are obviously difficulties.* [O1; item 101]

Another participant stated it even more clearly:

> *No, I'm a doctor, I have...no restrictions.* [O9; item 85]

Opportunities in Osteopathy

Regarding the questions about the opportunities and chances in osteopathy, most of the statements could be classified into the categories "professional profile" and "position in the health care system." A central opportunity in osteopathy is the installation of an independent profession. Almost all the osteopaths explicitly formulated the desire for their own professional profile. However, there was disagreement about the questions regarding where and how this profession should be integrated into the health care system or which competences it should include:

> *In the midst of the other health professions..., well, I don't see us as special consultants, as it is now in America, for example. But I see us as a health profession next to physiotherapists, occupational therapists...* [O4; item 81]

In this context, many possible applications were mentioned for the field of osteopathy. An osteopath saw a great opportunity in the prevention of diseases:

> *Concerning also prevention..., I believe that osteopathy has an enormous potential for people's health by simply doing something really good and also really preventing things,...follow-up problems or operations or God knows what...I see a huge opportunity there.* [O10; item 107]

O2 also attributed the potential for cost reduction to preventive osteopathy. From his point of view, examinations and medical consultations might be reduced:

> *I see a huge and very central importance of osteopathy in primary care...and I am convinced...from my daily experience that an incredible number...of diagnostic measures or specialist care...might be avoided if people were primarily also treated by osteopaths.* [O2; item 68]

Whether osteopathy actually contributes to disease prevention and can thus also lead to cost reduction or relief for the health care system is to be investigated using clinical study designs on the effectiveness of the treatment method itself. The position of osteopathy in the health care system and its differentiation from other professional groups have also not been uniformly described by practitioners in other countries.

The COVID-19 Pandemic

For current reasons, the participants were questioned about the effects of the COVID-19 pandemic, existing since March 2020, on their professional activities. Similar to a magnifying glass, crises very often reveal the weaknesses and failures of structures and concepts; however, they can also show their viability and strengths. Most respondents described the time of the COVID-19 pandemic with the lockdowns and the associated measures as a turning point in their practice. However, none of the participants described economic losses or existential fear:

> *Well, the time during Covid-19 wasn't easy at all.* [O5; item 75]

At the time of the survey, everybody had to wear a face mask, patients and osteopaths alike. The interviewees not only described the difficult communication with the patients because of the mask but also mentioned limitations during examination and treatment. Certain treatments, for example, techniques relating to the mandibular joint, could not be performed for patients wearing a face mask:

> *I believe that a lot of communication is lost through the mask, because you don't see the whole face of the patient. Of course, you have communication through the eyes, but there is still a barrier, a lot is lost...It already starts with and continues during inspection: you only see half of the face and in the case of jaw problems, I have to take down the mask first.* [O8; item 107]

Almost all the interviewees described a change in the clientele of their patients. Stress, sleep disorders, headaches, and dysfunctions of the mandibular joint were increasingly mentioned:

> *Psychosocial stress is increasing immensely...This in turn results...in sleep disorders...Mandibular joint problems due to stress, but also - and this is my own observation - because you constantly want to push around this mask if you have to wear it all day...I think that this has a huge influence...[O6; item 79]*

Teaching also seemed to be affected by the protection measures for pandemic control. An osteopath reported that teaching on inpatients at the hospital ceased. The question of whether the osteopathic treatment of inpatients in institutions was disturbed by a lack of external osteopaths remained unanswered:

> *Prior to the lockdown, our osteopathic child centre also paid visits to the neonatal ward...where we treated premature babies. Unfortunately, this is not possible at the moment.* [O1; item 47]

Owing to the cancellation of congresses and courses or their transfer to the digital world, interviewees experienced a gap in their personal training plans:

> *Of course, I...repeatedly attended courses. However, I scarcely did so in the last two years, actually...[O5; item 88]*

This can only be a small insight into the impact of the pandemic in the field of osteopathy. The effects of the COVID-19 pandemic on osteopathic care should be investigated systematically in the next few years.

Discussion

Principal Findings

This study identified numerous aspects, possibilities, and opportunities in osteopathy in Austria from the point of view of the osteopaths practicing in Austria.

In our survey, the typical osteopath presents as female and has previously worked as a physiotherapist, as previous studies have found [8]. This is consistent with other surveys from Europe regarding osteopathy. Moreover, in accordance with a study from Italy, the typical osteopath practicing in Austria works independently in their own practice and without employees [21].

The osteopaths interviewed usually found it difficult to define osteopathy. The respondents were not able to provide a uniform definition of osteopathy. Many respondents even expressed difficulties in precisely describing their profession. Nevertheless, recurring patterns can be recognized in the explanations given by the respondents.

The participants attempted to define osteopathy by drawing a distinction or differentiation from other professions and using third-party definitions. Furthermore, the philosophy of osteopathy, various osteopathic concepts or models of thought, the holistic nature of the treatment method, and activation of the patient's self-healing powers are often referred to. A possible reason for the heterogeneous attempts at explanation may lie in the difference in training and previous education. A recent study showed that only 17% of osteopaths surveyed in Austria identified themselves "exclusively" as osteopaths [8]. Therefore, there is a suspicion that, as our study also showed, the basic profession and nonregulation have a major influence on self-image. We observed a fundamental distinction between therapeutic and medical osteopaths.

From our point of view, the clear statements regarding the disagreements within the osteopath community were surprising. The conflicts do not remain in the specialist circles of osteopathy, but they even extend to the level of professional policy. The question is whether this is a country-specific observation for Austria. In their study in Australia in 2018, for example, Blaich et al [22] found disagreement about the specialization of osteopaths; however, it did not result in the splitting of osteopaths into 2 separate groups. The belief patterns and paradigms of individual treatment techniques that influence professional identity are not new in osteopathy [23]. The fact that, according to the osteopaths interviewed, these intraprofessional conflicts exist even on the political level or are the reason for nonregulation is remarkable. An increasing number of European countries regulate the professional practice of osteopathy. Therefore, it remains to be investigated whether this dispute itself has an influence on the nonregulation of osteopathy in Austria. However, conflicts and different opinions within a professional group are not inherent in osteopathy; these also exist in other medical professions such as chiropractic [24,25].

The general development of a profession is not only subject to cultural, historical, and social influences but also to the question of gender [26]. In this context, this study indicates a large influence of anchor personalities on the self-image of the interviewed osteopaths. It is remarkable that the anchor personalities mentioned are almost exclusively men. The historical context is worth noticing here, because Andrew Taylor Still, the founder of osteopathy, explicitly promoted equality between men and women already in the 19th century, in contrast to many other universities or teaching institutes during that time. He expressly included women in his courses [27]. In this context, it should be noted that in other health professions, although the practitioners are predominantly women, the leadership positions are often mainly occupied by men [28]. It is therefore not surprising that most users and practitioners of alternative medicine are women if their health needs are not being met by scientific medicine [29]. This becomes problematic when these professions are or become patriarchally dominated to match scientific standards [30].

To answer the questions about the origin of these conflicts and to deal with these in the future, we believe that a systematic and country-specific scientific analysis will be required. Conflicts in the health care system not only have the potential to weaken a profession but can also have a stimulating influence if understood as an opportunity [31].

Most of the osteopaths surveyed were in favor of a legal regulation of the profession. Under certain circumstances, osteopathically trained physiotherapists could benefit more from this, as they currently still need a physician's order to be able to practice with legal certainty.

However, nonregulation also has also some advantages—no applications for licenses, no obligation for regular further training, and unregulated pricing for treatment. With integration into the health care system, some participants fear deterioration owing to possible low or lower payment by health insurance companies.

Training quality management and studies in the subarea of osteopathy were also mentioned as challenges in this context. In Italy, Sweden, and Australia, the transfer of scientific results to the practical work of osteopaths has already been systematically investigated in a country-specific manner [32-34]. The openness to evidence-based practice (EBP) appears to exist among practicing osteopaths on a transnational basis, but the skills in dealing with the former vary from country to country. A study of EBP from Spain characterized the skills of the participants to deal with EBP as being rather low. This might be related to the lack of legal regulations and the inadequate transfer of knowledge in the training institutions [35]. The situation regarding osteopathy is similar in Austria. Additional country-specific studies are required to identify conclusions and connections.

The different situation in everyday practice owing to the COVID-19 pandemic and the respective infection protection measures also had an influence on the daily work of osteopaths. Several interviewees realized an evident change in the patients' profile. Although economic damage or fear for their professional existence were not explicitly described, most osteopaths working independently were themselves responsible for the implementation of the legal measures in their practices. The impact of the pandemic on the daily work in practice seems to have been less considerable than the impact on the field of training in osteopathy. As a large part of practical teaching occurs with patients under supervision, it is difficult to implement in a web-based format. The impact of the pandemic on clinical research at universities or universities of applied sciences remains to be examined.

In the case of further investigations in this area, we recommend a specific distinction of the participants between physicians and physiotherapists practicing as osteopaths. As there is no uniform training or legal regulation of osteopathy in Austria, only physicians and physiotherapists trained in osteopathy exclusively practice osteopathy. The results of this study suggest that there are evident differences between these 2 professional groups regarding, for example, patient acquisition, conflict management, and cooperation with other professional groups.

The extent to which the individual statements made by the interviewees represent the entirety of osteopaths practicing in Austria will be further investigated. The protocol of the COCO project describes the further procedures. The results of the qualitative partial studies (studies 1.1, 1.2, and 1.3 in Figure 1) will be combined in a following study to verify the results of the qualitative partial studies in relation to the population [36]. We will develop a standardized questionnaire as a measuring instrument.

An important feature of this study is the methodology, including 2 evaluators who completed the entire evaluation process. Through this approach, intersubjectivity increased and new, inductively formed categories were created. During this phase, many aspects of the research problems could be identified and categorized. The intercoder reliability was tested and found to be viable within this study. With another material iteration, the category system can be further refined, and the intercoder reliability can be further increased by optimized code definitions.

Limitations

First, it should be noted that the results of this study do not necessarily allow conclusions to be drawn about the entirety of osteopaths in Austria, as this is not an evaluation of representative surveys with large numbers of participants. Nevertheless, certain tendencies seem to emerge when statements by osteopaths appear to be congruent, that is, confirm each other or complement each other in a meaningful way. The sample represents the entirety of osteopaths in Austria well. Most respondents were women and physiotherapists [8]. Nevertheless, bias cannot be dismissed with such a specific sample. However, they give an idea about how osteopaths in Austria think, and the results obtained can serve as a hypothesis for large quantitative studies to test.

Conclusions

It is difficult to characterize the community of osteopaths in Austria conclusively. On the one hand, there is a great deal of agreement about the urgency regarding regulatory legislation for their profession, a necessary revision of training structures, and the specific promotion of scientific studies of osteopathy. However, when it comes to the concrete practice of osteopathy, deep trenches and even strong disputes have occurred among osteopaths.

The following question remains to be answered: what is “correct” or “true” osteopathy? If we consider that osteopathy has derived from various sources; that its founder did not give a final answer to the question about what he understood by osteopathy; and that each discipline is constantly developing, solely through the different osteopaths practicing, it appears that this question cannot be answered completely.

Apart from this issue, there is another and equally sensitive question, that is, whether and how the different parties can or even must be brought together for the regulation of their profession, which is desired by most of them. The different professional origins of osteopaths should also be considered. With regard to binding legal regulations, which would not least strengthen the professional image, mutual understanding seems to be imperative. Perhaps such an understanding might also lead to greater political weight for osteopathy, which it urgently needs, not only in terms of legal regulations but also to be able to promote important research projects.

The question arises as to whether the conflicts within osteopathy, in particular, with their possible professional-political consequences and the immense influence of the basic profession in the practice of osteopathy, are a country-specific phenomenon for Austria. However, there is a lack of studies in German-speaking countries with comparable qualitative designs to assess the work of osteopaths in more detail. We are therefore planning a meta-synthesis of qualitative studies with the aim of generating new theoretical insights from the accumulation of study results. Both the studies from the COCO project itself and other relevant literature can be used for the meta-synthesis.

To the best of our knowledge, the COCO project is the largest mixed methods study project on the osteopathic profession in German-speaking countries. The category system with its reliability check can be used as a basis for a repetition of the study. Such a research project would also be interesting if the profession was regulated formally and substantially in the near future. The results presented in this paper are not only intended to serve as a basis for further studies but also to provide universities, schools, professional associations, and politicians with an insight into the situation of osteopaths in Austria.

Conflicts of Interest

None declared.

Multimedia Appendix 1

COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.

[PDF File (Adobe PDF File), 870 KB - humanfactors_v11i1e45302_app1.pdf]

References

1. Benchmarks for training in traditional / complementary and alternative medicine: benchmarks for training in osteopathy. World Health Organization. 2010. URL: https://apps.who.int/iris/handle/10665/44356 [accessed 2023-08-19]
2. Osteopathic Education – Osteopathic International Alliance. URL: https://oialliance.org/resources/osteopathic-education/ [accessed 2023-12-24]
3. Regulation of osteopathy in Europe. European Federation & Forum for Osteopathy. URL: https://osteopathyeurope.org/regulation-2/ [accessed 2023-11-19]
4. Rechtliche situation. Österreichische Gesellschaft für Osteopathie (OEGO). URL: https://www.oego.org/home/alle-infos-uber-osteopathie/ [accessed 2023-11-19]
5. Phillips AR. Professional identity in osteopathy: a scoping review of peer-reviewed primary osteopathic research. Int J Osteopath Med 2022 Sep;45:25-37. [doi: 10.1016/j.ijosm.2022.06.005]
6. Krönke K. A questionnaire to evaluate the professional field of osteopathy in Austria pilot study 2003. Osteopathic Research Web. 2006. URL: https://www.osteopathicresearch.org/files/original/b4b7e15719626458f7ddea5b0dfea05004b992d8.pdf [accessed 2023-12-14]
7. Graf B. OsteopathInnen in Österreich Bestandsaufnahme, Vergleich mit internationalen Daten und Untersuchung der Entwicklung des Berufsbildes und -standes seit 2003 [Master-Thesis]. Krems, Vienna: Donau Universität; 2011. URL: http://tinyurl.com/26uepnmn
8. van Dun PL, Arcuri L, Verbeeck J, Esteves JE, Cerritelli F. The Austrian osteopathic practitioners estimates and RAtes (OPERA): a cross-sectional survey. PLoS One 2022 Nov 28;17(11):e0278041 [FREE Full text] [doi: 10.1371/journal.pone.0278041] [Medline: 36441692]
9. Mayring P. Qualitative Inhaltsanalyse: Grundlagen und Techniken. 12th edition. Weinheim, Germany: Beltz & Gelberg; 2015.
10. Kuckartz U. Mixed Methods: Methodologie, Forschungsdesigns und Analyseverfahren. Berlin, Germany: Springer-Verlag; 2014.

11. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
12. Deppermann A. Interviews as text vs. interviews as social interaction. Forum Qual Sozialforschung Forum Qual Soc Res 2013;14(3):2064 [FREE Full text] [doi: 10.17169/fqs-14.3.2064]
13. Flick U. Qualitative Forschung: Theorie, Methoden, Anwendung in Psychologie und Sozialwissenschaften. Berlin, Germany: Rowohlt Taschenbuch; 1999.
14. Meier MD, Porthun J. Situation, Chancen und Herausforderungen der Osteopathie in der Schweiz. DO Dtsch Z Osteopath 2020 Sep 15;18(04):26-31. [doi: 10.1055/a-1205-2394]
15. Bücker AK. Eigenschaften, Chancen und Herausforderungen der Osteopathie aus Sicht der Osteopathen in der Schweiz [Master-Thesis]. Dresden, Germany: Dresden International University; 2020.
16. Gläser J, Laudel G. Experteninterviews und qualitative Inhaltsanalyse: als Instrumente rekonstruierender Untersuchungen. 4th edition. Berlin, Germany: VS Verlag für Sozialwissenschaften; 2010.
17. Kuckartz U, Rädiker S. Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung. Weinheim, Germany: Beltz Juventa; 2022.
18. VERBI software. MAXQDA. 2022. URL: https://www.maxqda.com/about [accessed 2023-12-12]
19. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 2016 Jul 02;20(1):37-46 [FREE Full text] [doi: 10.1177/001316446002000104]
20. Altman DG. Practical Statistics for Medical Research. New York, NY: Chapman and Hall/CRC; 1991.
21. Cerritelli F, van Dun PL, Esteves JE, Consorti G, Sciomachen P, Lacorte E, OPERA-IT Group. The Italian Osteopathic Practitioners Estimates and RAtes (OPERA) study: a cross sectional survey. PLoS One 2019 Jan 25;14(1):e0211353 [FREE Full text] [doi: 10.1371/journal.pone.0211353] [Medline: 30682169]
22. Blaich R, Steel A, Clark D, Adams J. Challenges and opportunities for Australian osteopathy: a qualitative study of the perceptions of registered osteopaths. Int J Osteopath Med 2018 Dec;30:18-24. [doi: 10.1016/j.ijosm.2018.10.004]
23. Grace S, Fleischmann M, Vaughan B. "If you dont use or understand visceral osteopathy you are not a real osteopath": professional identity in Australian osteopathy through the lens of a single traditional technique. Explore (NY) 2021 Nov;17(6):535-540. [doi: 10.1016/j.explore.2020.07.001] [Medline: 32736952]
24. Pollard H. Reframing a debate in chiropractic. Chiropr Man Therap 2021 Nov 03;29(1):44 [FREE Full text] [doi: 10.1186/s12998-021-00401-5] [Medline: 34732222]
25. McGregor M, Puhl AA, Reinhart C, Injeyan HS, Soave D. Differentiating intraprofessional attitudes toward paradigms in health care delivery among chiropractic factions: results from a randomly sampled survey. BMC Complement Altern Med 2014 Feb 10;14(1):51 [FREE Full text] [doi: 10.1186/1472-6882-14-51] [Medline: 24512507]
26. Hall P. Interprofessional teamwork: professional cultures as barriers. J Interprof Care 2005 May 06;19 Suppl 1(sup1):188-196. [doi: 10.1080/13561820500081745] [Medline: 16096155]
27. Still AT, Hartmann C, MacGovern J. Das große Still-Kompendium: Autobiografie, Philosophie der Osteopathie, Philosophie und mechanische Prinzipien der Osteopathie, Forschung und Praxis. Berlin, Germany: Jolandos; 2013.
28. Feldman DE, Bellavance C, Frédérick D, Gagnon T, Lalonde C, Hudon A. Gender issues in physiotherapy in Quebec, Canada. Physiother Can 2023 May 01;75(2):169-176. [doi: 10.3138/ptc-2020-0142] [Medline: 37736386]
29. Shahvisi A. Medicine is patriarchal, but alternative medicine is not the answer. J Bioeth Inq 2019 Mar 20;16(1):99-112 [FREE Full text] [doi: 10.1007/s11673-018-9890-5] [Medline: 30570716]
30. Flesch H. Silent voices: women, complementary medicine, and the co-optation of change. Complement Ther Clin Pract 2007 Aug;13(3):166-173. [doi: 10.1016/j.ctcp.2007.03.005] [Medline: 17631259]
31. Piryani RM, Piryani S. Conflict management in healthcare. J Nepal Health Res Counc 2019 Jan 28;16(41):481-482. [doi: 10.33314/jnhrc.v16i41.1703]
32. Leach MJ, Shaw R, Austin P, Fryer G, Thomson OP, Adams J, et al. Attitudes, skills, and use of evidence-based practice: a cross-sectional survey of Swedish osteopaths. Int J Osteopath Med 2020 Dec;38:41-49 [FREE Full text] [doi: 10.1016/j.ijosm.2020.10.006]
33. Cerritelli F, Iacopini A, Galli M, Thomson OP, Sundberg T, Leach MJ, et al. Evidence-based practice among Italian osteopaths: a national cross-sectional survey. BMC Complement Med Ther 2021 Oct 07;21(1):252 [FREE Full text] [doi: 10.1186/s12906-021-03430-y] [Medline: 34620142]
34. Leach MJ, Sundberg T, Fryer G, Austin P, Thomson OP, Adams J. An investigation of Australian osteopaths attitudes, skills and utilisation of evidence-based practice: a national cross-sectional survey. BMC Health Serv Res 2019 Jul 17;19(1):498 [FREE Full text] [doi: 10.1186/s12913-019-4329-1] [Medline: 31315615]
35. Alvarez G, Justribo C, Sundberg T, Thomson OP, Leach MJ. A national cross-sectional survey of the attitudes, skills and use of evidence-based practice amongst Spanish osteopaths. BMC Health Serv Res 2021 Feb 10;21(1):130 [FREE Full text] [doi: 10.1186/s12913-021-06128-6] [Medline: 33563266]
36. Porthun J, Manschel J. Characteristics, Opportunities, and Challenges of Osteopathy (COCO) in the perceptions of osteopaths in Germany, Austria, and Switzerland: protocol for a comprehensive mixed methods study. JMIR Res Protoc 2019 Dec 18;8(12):e15399 [FREE Full text] [doi: 10.2196/15399] [Medline: 31850852]

Abbreviations

COCO: Characteristics, Opportunities, and Challenges of Osteopathy
COREQ: Consolidated Criteria for Reporting Qualitative Research
EBP: evidence-based practice
OEGO: Austrian Society for Osteopathy (Österreichische Gesellschaft für Osteopathie)

Edited by A Kushniruk; submitted 23.12.22; peer-reviewed by D Skinner, W Perveen, C Mastronardo; comments to author 20.11.23; revised version received 27.11.23; accepted 27.11.23; published 17.01.24.

Please cite as:
Manschel J, Porthun J, Winkler U, Beuckels JMAT, Martin D
Characteristics, Opportunities, and Challenges of Osteopathy Based on the Perceptions of Osteopaths in Austria: Qualitative Interview Study
JMIR Hum Factors 2024;11:e45302
URL: https://humanfactors.jmir.org/2024/1/e45302
doi:10.2196/45302
PMID:38231542

Original Paper

A Digital Behavioral Activation Intervention (JuNEX) for Pregnant Women With Subclinical Depression Symptoms: Explorative Co-Design Study

Elisa Mancinelli[1,2], BSc, MSc; Silvia Gabrielli[2], PhD; Silvia Salcuni[1], PhD

[1]Department of Developmental and Socialization Psychology, University of Padova, Padova, Italy

[2]Fondazione Bruno Kessler, Trento, Italy

Corresponding Author:
Elisa Mancinelli, BSc, MSc
Department of Developmental and Socialization Psychology
University of Padova
Via Venezia 8
Padova, 35131
Italy
Phone: 39 3342799698
Email: elisa.mancinelli@phd.unipd.it

Abstract

Background: Digital interventions are gaining increasing interest due to their structured nature, ready availability, and self-administered capabilities. Perinatal women have expressed a desire for such interventions. In this regard, behavioral activation interventions may be particularly suitable for digital administration.

Objective: This study aims to exploratorily investigate and compare the feasibility of the internet-based self-help guided versus unguided version of the Brief Behavioral Activation Treatment for Depression-Revised, an empirically supported in-person behavioral activation protocol, targeting pregnant women with subclinical depression symptoms. A user-centered design is used, whereby data are collected with the intent of evaluating how to adjust the intervention in line with pregnant women's needs. Usability and user engagement were evaluated.

Methods: A total of 11 Italian pregnant women with subclinical depressive symptoms based on the Patient Health Questionnaire-9 (scoring<15) participated in this study; of them, 6 (55%) women were randomly assigned to the guided group (age: mean 32.17, SD 4.36 years) and 5 (45%) to the unguided group (age: mean 31, SD 4.95 years). The Moodle platform was used to deliver the interventions in an e-learning format. It consisted of 6 core modules and 3 optional modules; the latter aimed at revising the content of the former. In the guided group, each woman had weekly chats with their assigned human guide to support them in the homework revisions. The intervention content included text, pictures, and videos. Semistructured interviews were conducted, and descriptive statistics were analyzed.

Results: Collectively, the data suggest that the guided intervention was better accepted than the unguided one. However, the high rates of dropout (at T6: guided group: 3/6, 50%; unguided: 4/5, 80%) suggest that a digital replica of Behavioral Activation Treatment for Depression-Revised may not be feasible in an e-learning format. The reduced usability of the platform used was reported, and homework was perceived as too time-consuming and effort-intensive. Moreover, the 6 core modules were deemed sufficient for the intervention's goals, suggesting that the 3 optional modules could be eliminated. Nevertheless, participants from both groups expressed satisfaction with the content and found it relevant to their pregnancy experiences.

Conclusions: Overall, the findings have emphasized both the intervention's merits and shortcomings. Results highlight the unsuitability of replicating an in-person protocol digitally as well as of the use of nonprofessional tools for the implementation of self-help interventions, ultimately making the intervention not feasible. Pregnant women have nonetheless expressed a desire to receive psychological support and commented on the possibilities of digital psychosocial supports, particularly those that are app-based. The information collected and the issues identified here are important to guide the development and co-design of a more refined platform for the intervention deployment and to tailor the intervention's content to pregnant women's needs.

(JMIR Hum Factors 2024;11:e50098) doi:10.2196/50098

KEYWORDS
digital intervention; behavioral activation; feasibility; pregnancy; subclinical depression symptoms

Introduction

Background

Peripartum depression refers to an episode of depression that meets the criteria for persistent or major depressive disorder, with onset occurring during the peripartum period [1]; this definition highlights the direct link between the development of the depressive condition and the bodily changes and overall characteristics inherent to the perinatal period [2]. A distinction should also be made with regard to antenatal and postnatal depression, since antenatal depression is recognized as one of the main predictors of postnatal depression, with the latter then aggravating the repercussions on the mother as well as on the child and the whole family [3]. In Europe, antenatal depression counts a mean prevalence of 17.9% [4], while postnatal depression ranges from an average of 12.91% to 16.62% [5]. In Italy, specifically, the literature highlights a prevalence ranging from 6% to 22% for antenatal depression [6-9] and from around 13% to 23% for postnatal depression [5,7,9,10]. Such percentages emphasize the necessity for early detection and the implementation of prevention programs to alleviate depression symptoms already during pregnancy. Notably, women have expressed a desire for perinatal support programs and have reported benefiting from them, both in terms of symptom reduction and increased sense of agency [11]. However, barriers to help seeking, in the context of perinatal care, have been widely recognized (*knowledge barriers*, eg, difficulty in recognizing health needs and distinguishing emotional difficulties as well as not knowing the services available; *practical barriers*, eg, time and economic constraints; and *attitudinal barriers*, eg, stigma, guilt) [11-13], contributing to the still limited availability of the services to support perinatal women's mental health [14].

Within this context, digital solutions might be particularly valuable. A recent review that specifically focused on the application of eHealth in perinatal care [15] highlighted the potential of these solutions as alternatives or supplements to standard mental health practices, both for screening and intervention. A subsequent review [16] also emphasized the beneficial role of digital solutions in addressing perinatal depression by enhancing accessibility to psychological interventions, thus promoting scalability, which could ultimately allow to work around the abovementioned barriers to help seeking. Digital interventions could thus be valuable solutions to fill in the gap between what is asked and desired by women and the logistic and economic limits on both clinical professionals and health institutions part. Notwithstanding, despite the increasing focus on peripartum depression, there is currently a scarcity of digital psychological interventions aimed at alleviating depression symptoms, particularly those grounded in empirically validated intervention protocols [16-18]. Furthermore, the prevailing focus appears to lean more toward treating rather than preventing perinatal depression. This is evident in the dearth of studies investigating digital interventions during pregnancy and exemplified by the lack of studies investigating digital interventions deployed during pregnancy and to women with subclinical depression symptoms [17,18]. In light of this, there is a need to develop theoretically grounded digital interventions tailored to pregnant women with subclinical symptoms needs and characteristics, ultimately preventing the development or worsening of clinically relevant depression symptoms during the postpartum.

Evidence-based interventions that are brief and structured such as behavioral activation (BA) interventions might be especially helpful to this end, as providing pregnant women with concrete strategies will be useful to support their adjustment. BA is an empirically supported behavioral intervention created to lessen depression symptoms [19-22]; it is based on the idea that a greater awareness of the mutual influence between behavior and emotion can ultimately encourage behavioral change by increasing participation in joyful and adaptive activities while reducing participation in maladaptive behaviors that maintain or exacerbate the depressive symptoms [19,23]. However, a recent scoping review [18] highlighted a gap in the literature: there are few digital BA interventions available during the perinatal period, and none have been specifically deployed during pregnancy. Furthermore, their usability has only been marginally evaluated.

Usability refers to the quality of the interaction occurring between the user (eg, pregnant women) and the tool used (eg, website, smartphone app, etc) [24]; a subcomponent of usability that is more specific is the *user engagement*, which includes the user's cognitive, behavioral, and affective reaction to the tool [25]. These factors are instrumental in supporting user compliance and adherence, and they should be carefully considered and addressed in the development of feasible and acceptable digital interventions [26]. The limited evaluation of these factors in the context of digital mental health solutions may be attributed to the novelty of the field, which has yet to establish a comprehensive understanding of design methods for such tools [27]. When designing these digital solutions, four components should be kept in mind: (1) the design issue and solution, (2) the context in which the design occurs, (3) the dynamics and organization of the design activity, and (4) the actors contributing to the design [28-30]. However, a recent review [27] investigating the design methods and approaches used for the design and development of digital tools for mental health stressed that *human-centered design methods* (ie, the design of digital tools not considering the engineering design and including user-centered approaches, co-design, participatory design, etc) are not yet fully integrated within the field and that reported design approaches are still mainly external, thus excluding the perspective of those for whom the tool is created.

This Study

This study aimed to investigate the feasibility of the Brief Behavioral Activation Treatment for Depression-Revised (BATD-R) [31] protocol that was structured as an internet-based self-help intervention and deployed to pregnant women with subclinical depression symptoms. Compared with other BA protocols, the BATD-R protocol specifically targets subclinical depression symptoms, is flexible (in terms of both its structuring and the population it is administered to), and can be self-administered or deployed by both specialists or nonspecialists [31].

Given that no previous study has used the BATD-R protocol for this purpose, the intervention developed and evaluated in this study serves as digital "replica" of the in-person BATD-R protocol. By adopting a *user-centered* approach, this study not only aimed to assess its initial feasibility but also sought to gather valuable feedback directly from pregnant women. This feedback will guide the adjustment of the intervention's content and thus its structure, without making assumptions beforehand about the changes required. Indeed, a user-centered design "is an approach to product development that grounds the process in information about the people who will ultimately use the product" [32]; as such, to create a well-accepted, engaging, and effective digital intervention in perinatal care, subsuming the intervention content and the mean through which it is deployed, pregnant women should be consulted in each stage of the intervention's creation and refinement, thereby ultimately allowing the co-design of the final intervention. In this regard, this study relied on the Obesity-Related Behavioral Intervention Trials model [33], which provides an iterative progressive framework guiding the development, testing, and refinement of the behavioral intervention. More specifically, it uses a user-centered design that relies on a data-driven approach to iteratively test and revise the intervention, up until it is deemed appropriate to move to further phases of development and testing, thereby going from the intervention design, its preliminary testing to investigating its efficacy and effectiveness [33].

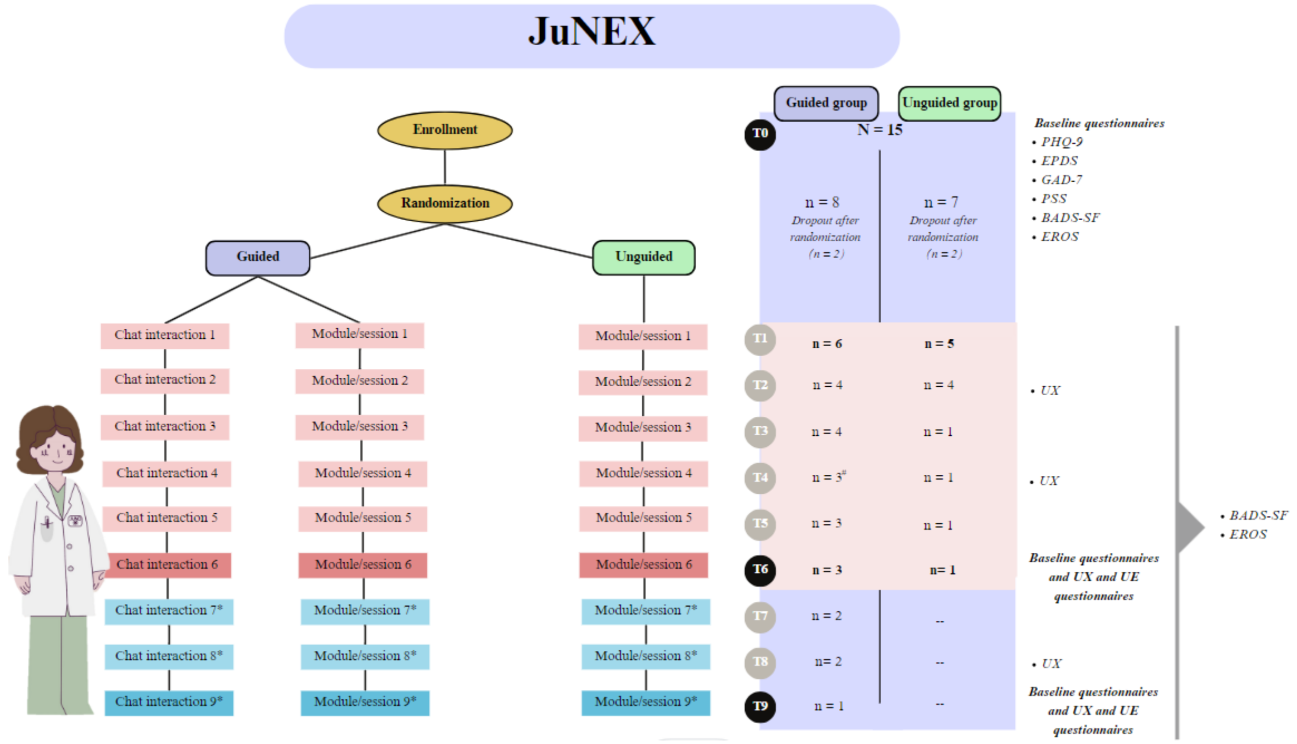
As previously reported, no digital BA intervention targeting subclinical or clinical depression symptoms among pregnant women has been developed [18]. Nonetheless, it is worth noting that among the existing digital BA interventions, many are guided interventions [18]. The guides, most often mental health specialists or trained professionals, provide additional support to women throughout the intervention, by addressing concerns or supporting them on intervention-related tasks. Mindful of this, a further aim of this study was also to explore and evaluate the role and potential benefits of including a guide as additional support in the self-help intervention. As such, the study also aimed to compare the feasibility of the guided versus unguided version of the intervention.

Methods

Recruitment

Recruitment was done through snowball sampling, using social media platforms (eg, Facebook). A Google Form survey was developed containing the informed consent and the questions and questionnaires needed to evaluate the women's eligibility for participating in the study. Specifically, eligible women complied with the following inclusion criteria: they (1) had physiological pregnancy, (2) were aged ≥18 years, (3) were between the 12th and 30th week of gestation, and (4) had subclinical depression symptoms (Patient Health Questionnaire-9 [PHQ-9] score<15) [34]. By contrast, women were excluded when (1) presenting a history of past or current mental disorders; (2) exhibiting clinically significant psychological symptoms (ie, depression symptoms: PHQ-9≥15) and suicidal ideation (PHQ-9 item 9); (3) having an obstetrically at-risk pregnancy; (4) presenting medical conditions, pregnancy-related and otherwise; (5) experiencing an artificially induced pregnancy. A total of 15 women had filled in the web-based questionnaire; all were deemed eligible, and thus none reported any of the exclusion criteria. Following randomization in either the guided or unguided group, 4 (27%) women dropped out (Figure 1) before starting the intervention. As such, the final sample is composed of 11 (73%) women.

Figure 1. Study structure and study adherence flowchart. *Indicates optional modules and # indicates that 1 participant had stopped interacting with the guide but had continued viewing the Moodle content. BADS-SF: Behavioral Activation for Depression Scale–Short Form; EPDS: Edinburgh Postnatal Depression Scale; EROS: Environmental Reward Observation Scale; GAD-7: Generalized Anxiety Disorder-7; PHQ-9: Patient Health Questionnaire-9; PSS: Perceived Stress Scale; T: time point; UE: user engagement; UX: user experience.



Procedure and Study Structure

This study is structured in 4 main phases (Figure 1). Phase 1 corresponds to time point 0 (T0), which encompasses the recruitment, enrolment (including the assessment of inclusion and exclusion criteria), and randomization processes. At this time, anamnestic information (reported in the *Recruitment* section) was collected, and standardized questionnaires were administered measuring depression and anxiety symptoms, perceived stress, current activity level, and perceived environmental reward. Phases 2 and 3 of the study correspond to T1-T6 and T7-T9, respectively. During these phases, the baseline questionnaires, along with one questionnaire assessing user engagement (UE) and another assessing user experience (UX; explained in the *Measurement Tools* section) were administered. In addition, the UX questionnaire was also administered during T2, T4, and T8. Questionnaires assessing the level of activity and related reward (explained in the *Measurement Tools* section) were administered each week, from T0 to T9. The final phase 4 involved conducting semistructured interviews that were created ad hoc, and participants who had participated up to at least T6 were interviewed. The semistructured interviews were conducted to qualitatively evaluate the women's experience with the intervention and gather further feedback necessary for refining the intervention. All questionnaires administered between T1 and T9 were created using Google Forms. Links to access these questionnaires were made available within the intervention platform and were accessible to both groups in the same manner.

The participants were informed that they could leave the study at any moment without having to provide an explanation and without incurring in any penalty. Each was assigned an alphanumeric code to ensure confidentiality.

Randomization

Alphanumeric codes were generated and allocated to each participating woman to guarantee confidentiality throughout the study. The process of creating these codes was carried out using Microsoft Excel and further randomized through a Google software [35], ensuring the unbiased assignment of participants to either the guided or unguided group before the commencement of recruitment. The results of random code assignment were kept aside, and women were only provided with their designated code once their eligibility for the study had been established. Only when eligibility was confirmed, participants were given specific information on the intervention they had to follow. The group assignment was single blinded.

The Intervention's Content

This internet-based self-help intervention originates from the BATD-R [31] protocol. Originally, this protocol consisted of 10 sessions, which included 5 main sessions and 5 additional sessions aimed at reviewing and maintaining the benefits achieved; nevertheless, it is possible to reduce the number of sessions to 5 or even expand beyond 12 sessions. However, past studies advise against exceeding 10 weeks of intervention, reporting that BATD-R interventions of up to 6 or 8 weeks allow for a significant reduction in depression symptoms [36,37]. BATD-R can be self-administered or administered by both trained and untrained staff, further emphasizing its adaptability and flexibility. The protocol begins with a psychoeducation phase focused on understanding the characteristics of depression symptoms. Furthermore, it includes homework assignments to be completed between sessions, which form the core content of the 5 main sessions. These assignments involve filling out 5 forms (refer to the original protocol by Lejuez et al [31]). The first form, called the “daily monitoring form,” should be completed throughout the intervention and requires participants to continuously evaluate their daily activities, in terms of behavioral patterns, the pleasantness of activities, and the importance of each daily action. Subsequently, the person is required to identify through a second form, called “life areas, values, and activity inventory,” what their values are within important areas of life (ie, important relationships; pleasurable activities and hobbies; work and study; mind, body, and spirituality; and daily responsibilities). This would then allow the person to identify (form 3, called “activity selection and ranking”) and plan daily activities (form 4, called “action plan”) that help them live in accordance with their values in the mentioned life areas. In this regard, it is worth noting that although the BATD-R does not include a complete functional analysis due to the brevity of the treatment approach [38], Lejuez et al [31] stressed that several components of treatment fit into a functional analytical framework. This is most noticeable when choosing activities that are closely related to values, given the dual goals of identifying the factors that maintain or reinforce depressive behavior (both positive and negative reinforcement) and the positive reinforcers that could support or strengthen healthy behavioral patterns. Along with the critical assessment of dysfunctional behavioral patterns, the BATD-R protocol includes a final form called “contracts,” which focuses on strategies to request social support. To create these “contracts,” the person is required to identify (1) an activity to perform, (2) up to 3 support persons who may be able to assist or support them, and (3) how and when each person can specifically provide said support. This activity allows the patient to identify their needs and provides a specific plan on how to seek the help they need by making concrete requests for obtaining assistance and social support to reinforce adaptive behaviors.

In this study, the intervention closely follows the structure of the original protocol but is adapted to fit a digital format. The intervention was divided into 6 weekly sessions or modules, which include the core content of the intervention. In addition, 3 optional “bust” sessions were included to reinforce and consolidate the information from previous weeks. The entire intervention spanned 9 weeks, with participants completing weekly homework assignments between sessions. While the intervention content and homework remained unchanged, it was adapted for digital delivery using text, video, and images. Information related to depression symptoms was contextualized for the pregnancy period to cater to the specific needs of the target population.

The Platforms

Overview

In this first evaluation of the intervention, the Moodle e-learning platform (Moodle 3.11; 2021) was used as the delivery method. This platform was accessible via both the web and the Moodle app on smartphones. Both the web and app versions of Moodle included a chat feature, which was exclusively used by the guided group to interact with their assigned guides. Specifically, in the guided group, guide-woman dyads were created, and they interacted once a week for the homework revision. For the unguided group, homework revision was facilitated through a written self-guide available within the Moodle platform. The forms representing the homework were structured on Google Docs, with links to access them made available within the Moodle platform so that women could directly access them both on the web and the smartphone app.

Both intervention groups were presented with the intervention as an e-learning Moodle course; the sole difference in the intervention’s content between the groups lay in the reference to the guides. The intervention was structured into modules (6 core modules and 3 optional modules), and each module could be consulted only after completing the previous one. The content of each module was delivered using illustrative videos and images, complemented by brief text information. After viewing each section within a module, participants were presented with a brief quiz comprising 3 true-or-false questions related to the content they had just reviewed. These quizzes were incorporated with the intention of fostering UE. They encouraged participants to actively engage with the material rather than to passively view it. In addition, the quizzes served as a means of assessing participants’ comprehension of the content. On the basis of the accuracy of their responses, participants received reinforcing or motivating feedback after completing each quiz.

The Guides

In the guided group, specific guide-woman dyads were randomly created. All the guides were recognized psychologists who had been trained to become psychotherapists. They underwent comprehensive training, which included the provision of detailed written information and a 2-hour in-person meeting. This training covered the intervention’s content and structure and their role as guides. The training aimed to ensure that all the guides had a consistent understanding of the intervention and its content, thus maintaining uniformity across the interactions between the guides and participants. The guides adhered to a partially defined conversational protocol, which consisted of fixed messages and information to be delivered, as well as “free” parts where they had the freedom to phrase sentences as they saw fit. This flexibility in the “free” parts allowed guides to respond adaptively to women’s answers and feedback, particularly during the homework revision part of the intervention. Furthermore, it enabled guides to support participant compliance and adherence based on their perceived motivation levels. Supervision for the guides was provided by the first author (EM) of the study, who oversaw the technical aspects of the intervention. In addition, an expert psychotherapist, the third author (SS), provided supervision for clinical matters.

To ensure a consistent participant experience, all the guides were assigned the same name, “Joy.” This uniformity aimed to minimize any potential biases that could arise from variations in the perception of the different guides. Moreover, guides were not provided with any information about the women they were assigned to, ensuring privacy and confidentiality for the participants.

Measurement Tools

PHQ-9 Tool

The PHQ-9 [34] is a unidimensional self-report tool that is widely used in the Italian context [39]. It assesses the severity of depression symptoms during the previous 2 weeks, based on the diagnostic criteria of the *Diagnostic and Statistical Manual of Mental Disorders* (*Fourth Edition*; *DSM-IV*) [40]. It consists of 9 items measured on a 4-point Likert scale (0="not at all"; 3="almost every day"). Item 9 assesses suicidal ideation. A score of ≤9 indicates mild or no symptoms of depression, and a score between 10 and 14 indicates moderate symptoms, while a score of ≥15 indicates severe symptoms of depression. The instrument shows excellent internal consistency at α=.92 [41].

Edinburgh Postnatal Depression Scale

Edinburgh Postnatal Depression Scale [42] is a unidimensional self-report tool, validated in Italy [43], which assesses the severity of depression symptoms during the previous week. Albeit developed to assess depression symptoms during the postpartum period, it is often used throughout the perinatal period. It consists of 10 items measured on a 4-point Likert scale (0="no, not at all"; 3="yes, always"). The instrument shows good internal consistency at α=.79 [43].

Generalized Anxiety Disorder-7

Generalized Anxiety Disorder-7 [44] is a unidimensional self-report tool that assesses the severity of anxiety symptoms during the previous 2 weeks. It consists of 7 items measured on a 4-point Likert scale (0="never"; 3="almost every day") and shows good psychometric indexes in the Italian context as well [41]. The instrument shows excellent internal consistency at α=.92 [41].

Perceived Stress Scale

Perceived Stress Scale [45] is a unidimensional self-report tool, validated also in Italy [46], which assesses the severity of stress symptoms in the previous month. It consists of 10 items measured on a 4-point Likert scale (0="never"; 3="quite often"). The instrument shows good internal consistency at α=.74 [46].

Behavioral Activation for Depression Scale—Short Form

Behavioral Activation for Depression Scale–Short Form [47] is a self-report tool designed to measure changes in avoidance and activation during BA interventions for depression during the previous week. It consists of 9 items measured on a 7-point Likert scale (0="not at all"; 6="completely"). The scale provides 2 scores, the first score referring to the level of BA (5 items) and the second one to the level of behavioral avoidance (5 items). Manos et al [47], the authors of the tool, advise considering the total score instead of the subscales. This questionnaire has not been translated into Italian and was therefore translated through the back translation procedure. Example items are "I am content with the amount and types of things I did" (item 2) and "Most of what I did was to escape from or avoid something unpleasant" (item 5). The instrument shows good internal consistency (total scale α=.82) [47].

Environmental Reward Observation Scale

Environmental Reward Observation Scale [48] is a unidimensional self-report tool designed to measure the level of environmental reward perceived in recent months. It consists of 10 items rated on a 4-point Likert scale (0="strongly disagree"; 4="strongly agree"). This questionnaire has not been translated into Italian and was therefore translated through the back translation procedure. Example items are "It is easy for me to find enjoyment in my life" (item 4) and "I wish that I could find more hobbies that would bring me a sense of pleasure" (item 7). The instrument shows good internal consistency (α=.87) [48].

UX Measure

Mobile Application Rating Scale (MARS) [49] is a self-report tool consisting of 23 items scored on a 5-point Likert scale (1="poor"; 5="excellent"), which assesses the quality of the app and its features (ie, the Moodle app) on 4 dimensions of objective quality: engagement (5 items), functionality (4 items), aesthetics (3 items), and information (7 items); a final scale assesses the subjective quality (4 items). The average of the scores of the 4 dimensions of objective quality provides the total scale score. The questionnaire also contains an "application-specific" section (6 items) to assess the potential impact of a particular app on domains such as users' knowledge and intentions. The total and subscale scores of the MARS have high internal consistency coefficients (α=.90 and α=.80-.89, respectively). The scale has been validated in the Italian context [49]. For this study, only the subscales related to "information," "subjective app quality," and "app-specific" sections were considered, totaling to 17 items.

Together with the MARS items, only at T6 and T9, women were also asked to prove their overall subjective opinion on the platform ("Please write below your personal opinion with respect to your experience [pros and cons] while using the platform").

UE Measure

UE Scale–Short form [25] is a short self-report tool designed to assess UE with a digital solution. It consists of 12 items based on a 5-point Likert scale (1="strongly disagree"; 5="strongly agree"). The questionnaire consists of four factors: (1) focused attention, which indicates the feeling of being immersed in the interaction; (2) perceived usability, which is the negative effect experienced due to the interaction and the effort expended; (3) aesthetic attractiveness, which represents the graphical and visual appeal concerning a digital solution; and (4) the reinforcement (reward) factor, which regards the perceived involvement and enjoyment with the digital solution. This questionnaire was not translated into Italian and was therefore translated through the back translation procedure. The 4 scales have good internal reliability, as follows: focused attention, ω=0.75; perceived usability, ω=0.70; aesthetic attractiveness, ω=0.88; and reinforcement, ω=0.79 [25].

Semistructured Interview

The semistructured interviews were conducted by the first author (EM) and featured 15 main questions developed specifically for the study, 3 (20%) of which were asked only to the guided group. This interview was conducted approximately 10 days after each woman had finished the intervention. It lasted between 15 and 20 minutes, and following the woman's consent, it was audio recorded to allow for its transcription and evaluation. For both the guided and unguided groups, the interviews investigated women's personal experience (7 questions) with the intervention and the experience (5 questions) specifically related to the use of the platform. In the guided group, women's experience with their guide and the overall chat interactions were investigated.

Data Analysis

Statistical analyses were performed using RStudio (R Foundation for Statistical Computing) [50]. Descriptive data for both categorical (n, %) and continuous (mean and SD) variables were analyzed separately for the guided and unguided groups, considering the different time points. Given the preliminary nature of the study and the small sample size, no further analyses were performed. The interviews were individually and qualitatively analyzed through thematic analysis, following the predefined semistructured interview's 3 broader themes. Thematic analysis was conducted following a modified version of the guidelines proposed by Braun and Clarke [51], which has already been used in other co-design studies [52].

Ethical Considerations

The study was conducted in compliance with the ethical guidelines of the Declaration of Helsinki [53] and the European Union law for data protection (EU General Data Protection Regulation 679/2016). The study was approved by the Ethical Committee of the Psychology Department of the University of Padova (number 4820/2022).

Results

Descriptive Information and Adherence

A total of 11 women participated in the study; 6 (55%) were part of the guided group and 5 (45%) were part of the unguided group. Descriptive information is reported in Table 1 separately for the 2 groups.

Table 1. Descriptive information (n=11).

	Guided group (n=6)	Unguided group (n=5)
Age (years), mean (SD)	32.17 (4.36)	31 (4.95)
Gestation week, mean (SD)	21.17 (5.95)	19.83 (5.75)
Living area, n (%)		
North Italy	4 (67)	4 (80)
Central Italy	2 (33)	1 (20)
South Italy	0 (0)	0 (0)
Education, n (%)		
<High-school diploma	1 (17)	0 (0)
High-school diploma	1 (17)	2 (40)
Bachelor degree	1 (17)	0 (0)
Master degree	2 (33)	2 (40)
Specialization (eg, PhD)	0 (0)	1 (20)
Marital status, n (%)		
Single	0 (0)	0 (0)
Cohabitant	3 (50)	2 (40)
Married	3 (50)	3 (60)
Past abortion, n (%)		
Yes	2 (33)	1 (20)
No	4 (67)	4 (80)
Women's occupation, n (%)		
Unemployed	1 (17)	0 (0)
Student	1 (17)	0 (0)
Student and freelance worker	1 (17)	0 (0)
Employee	3 (50)	3 (60)
Student and employee	0 (0)	1 (20)
Researcher	0 (0)	1 (20)

The descriptive statistics pertaining to psychosocial variables assessed at T0, T6, and T9 for both groups are presented in Table 2.

Table 2. Descriptive statistics at time point 0 (T0), T6, and T9 (n=11).

	Scores of the guided group (n=6)			Scores of the unguided group (n=5)	
	T0 (n=6), mean (SD)	T6 (n=3), mean (SD)	T9 (n=1)	T0 (n=5), mean (SD)	T6 (n=1)
PHQ-9[a]	6.67 (2.66)	3.33 (2.08)	3	3.4 (2.79)	1
EPDS[b]	15 (2.37)	11 (3.61)	6	10.6 (3.21)	9
GAD-7[c]	7 (1.41)	5 (2)	3	4.0 (1)	3
PSS[d]	20.50 (2.07)	18.67 (2.08)	17	18.0 (2.3)	17
BADS-SF[e]	23.17 (8.18)	16.67 (1.15)	18	19 (2.92)	27
EROS[f]	28.17 (4.17)	30.67 (2.52)	34	32 (3.39)	36

[a]PHQ-9: Patient Health Questionnaire-9.

[b]EPDS: Edinburgh Postnatal Depression Scale.

[c]GAD-7: Generalized Anxiety Disorder–7.

[d]PSS: Perceived Stress Scale.

[e]BADS-SF: Behavioral Activation for Depression Scale–Short Form.

[f]EROS: Environmental Reward Observation Scale.

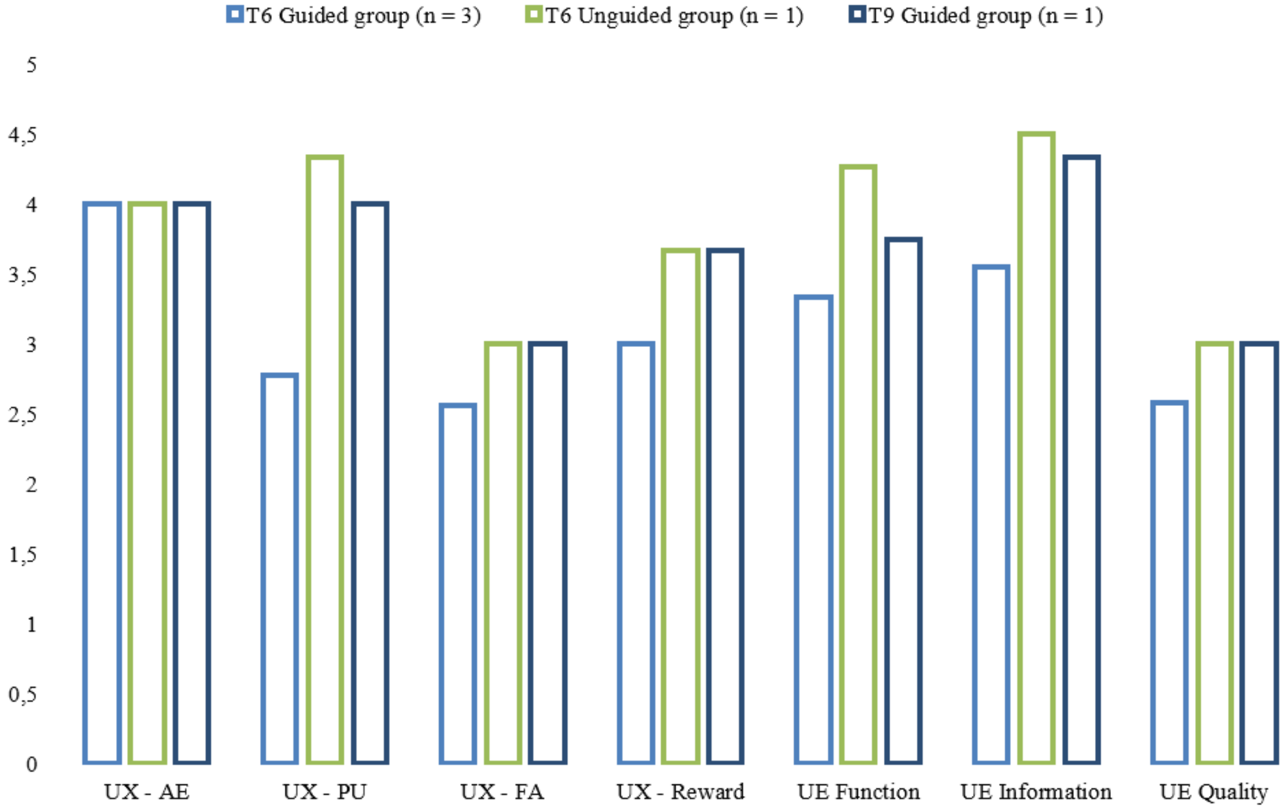
Regarding dropout rates, it was higher in the unguided group (n=5), with most participants (n=4, 80%) dropping out after completing the second module. In contrast, the dropout pattern in the guided group (n=6) was more gradual. Overall, 1 (9%) participant had dropped out because of health reasons and 3 (27%) because of the amount of time and effort (particularly related to the homework) required by the intervention, while 3 (27%) did not provide a reason for dropping out. Accordingly, 3 (50%) participants from the guided group reached T6, thereby completing the 6 core modules in Moodle, while only 1 (20%) in the unguided group reached T6. However, among the 3 participants from the guided group who reached T6, 1 (33%) ceased interactions with her guide after T3 but continued viewing the material on Moodle up to T6; the other 2 (67%) continued until T9. One of them stopped viewing the Moodle content at T8 but continued with the chat interactions with the guide.

Regarding homework completion, it is not possible to quantify adherence specifically, as, for instance, among the participants that had completed at least until T6 some (2/4, 50%) decided to handwrite the homework, instead of using Google Docs because of the low usability of the latter (ie, too many steps to go from Moodle to Google docs, which were also not well perceived and difficult to use within the smartphone). Moreover, it was considered cumbersome to write within the Google Docs.

UX and UE Measures

Descriptive statistics for UX and UE assessed at T6 and T9 are illustrated in Figure 2. At T6, when the participants were asked whether they had used primarily the app or web version of Moodle, 2 (67%) participants of the guided group and the only 1 participant of the unguided group reported using the app version, while 1 (33%) of the participants of the guided group used primarily the web version because of difficulties with using the app.

Figure 2. User experience (UX) and user engagement (UE) at time point 6 (T6) and T9. All scales' response range was from 1 to 5. AE: aesthetic appearance; FA: focused attention; PU: perceived usability; quality: subjective quality.



Participants' subjective opinion on the platform, assessed through one open question, is reported in Table 3.

Table 3. Answers to the open question "Please, write below your personal opinion with respect to your experience (pros and cons) using the platforms."

Time points and group	Participants' feedback
Time 6	
Unguided group, n=1	• "I am in my second pregnancy so I have been using the platform with a one-and-a-half-year-old [taking care of them] and I must say that being able to subdivide the time has been helpful even on an organizational level. If one fills out the daily forms from day to day it is not challenging, I must say that filling them out [the daily forms] from your cell phone though is quite inconvenient because the app [Google Docs] that opens the forms and allows you to fill them out does not always work well, so I then preferred to print them out and fill them out by hand."
Guided group, n=3	• "I think it is still very cumbersome as a platform [Moodle], not easy to use and the interaction with the Guide [was] too dry. Some things need to be revised." • "Using the platform on a practical level was quite intuitive and easy. The project itself involves a lot of effort and concentration, but it helps to feel a greater physical and psychological well-being." • "It was intuitive and fast."
Time 9	
Guided group, n=1	• "My experience in using the platform has been positive."

Semistructured Interviews Results

A total of 3 participants agreed to participate in the semistructured interview; 2 (67%) were from the guided group and had continued the intervention till T9, while 1 (33%) was from the unguided group and had participated till T6.

During the semistructured interview, all 3 (100%) participants reported that they had come to know of the study through a friend. As motivation for participating, all 3 reported "*curiosity*" as the main reason because they did not have specific expectations before starting.

Overall, participants expressed that the intervention helped them find a moment for themselves and provided them with a method or strategy to change their perspective on how they viewed and performed their daily activities. Furthermore, they noted the positive effects of engaging in more rewarding activities. All 3 (100%) participants emphasized that the intervention supported their overall well-being rather than reduced negative feelings per se. However, they also mentioned that the effort required by the intervention, particularly in terms of time and dedication, was substantial, especially regarding the homework assignments. While they appreciated the meaningful content of the homework, they found it burdensome to complete. It is important to note that part of the difficulty with homework completion was related to the reduced fluidity and usability of Google Docs on smartphones. A more thorough explanation of the findings that emerged from the semistructured interviews with verbatim examples of participants' answers has been discussed as follows, and the specific themes and subthemes that emerged are reported in Table 4.

Table 4. Semistructured interview (n=3): themes, subthemes, and examples.

Theme and subtheme	Example quotes
Personal experience with the intervention	
Self-observation and activities evaluation	• "[...]you realize things that by not doing it [the intervention] you wouldn't have realized you could have done, how you could have also handled the pregnancy period better" [Unguided group participant] • "It definitely helped me understand what are [...] let's say, [which is] the focus, where to aim to get better. I discovered some things that I had set aside[...] it helped me to not be so focused on negative things, which is kind of my problem, but more to identify something positive to do day-by-day, to be able to accomplish little goals that [might not be important for others] but for me at that moment they were important" [Guided group participant] • "[...]most definitely [I appreciated] focusing my attention on some positive activities that I had somewhat set aside and forgotten[...] when I focused on those saddest moments and I [then] realized that they were not most of my days as one thinks when one is in the sad mood. Instead, I saw that most of my days were good moments" [Guided group participant]
Effort	• "As the weeks went on, maybe even as the pregnancy progressed[...] I found it more 'burdensome.' The fact of filling out the daily forms [daily monitoring form] every day[...]. as time went on, it was challenging, in the sense that one has to stop and really take [their] time and be consistent, when as the pregnancy progresses maybe other thoughts take over and you can't quite be that consistent all the time[...]" [Guided group participant] • "[...]the part, let's say the most obnoxious, difficult, whatever we want to call it, is definitely the material to fill out during the last weeks. I have to tell you the truth, I didn't even finish them because I didn't have time[...]" [Guided group participant] • "In my experience[...]. I felt the fatigue more, maybe, here. Let's say that at the end of the six (sixth module), for myself, I felt that the intervention was -in quotes- "finished." [Guided group participant] • "[In reference to the intervention length] it probably depends on what stage of the pregnancy one is[...] I started it toward the end of my pregnancy, it would probably be better to start it before[...]"[Unguided group participant]
Learning the "method" and its application in the future	• "[...]It might be a good method at other times in life when one may experience difficulties" [Guided group participant] • "[...]mentally, I got into this mode of planning something nice to do, to be able to have that weekly commitment that I like, to ask somebody to do it with me. Maybe small things, but I'm sure it helped me for the future as well." [Guided group participant]
User experience and user engagement	
Managing issues and learnability	• "[...]when I had to write [for the homework], there was the transition from the daily monitoring [form] rather than very often I had trouble writing things down[...] I mean, I had to print them out [the forms] basically, if not I couldn't record [write] them [down]." [Unguided group participant] • "[A]t first maybe you kind of have to learn the mode of, yeah, that you get the materials out[...] It wasn't easy to tell when it saved what you had done, because it was a little dubious, sometimes[...] it was easy to use once you had gotten the hang of it, going into the week, looking at the material..." [Guided group participant]
Multimedia material	• "[...]the videos etc. were very effective in passing the message, both in terms of explanations and content, and the images, they were[...] they caught the attention[...]" [Guided group participant] • "[T]he videos[...] they are very clear, well done, they are cute" [Unguided group participant]
App interventions in the future	• "[...]with the current use of the phone and the computer in general, in my opinion an app is useful" [Unguided group participant] • "[...]the app that you have on your phone is the most useful thing. You consult it wherever you want and whenever you want" [Guided group participant]
The experience with the guide	

Theme and subtheme	Example quotes
Support	• "[...]the interaction itself was effective in the sense that it explained things to me when I had doubts, that it directed me, maybe, when I didn't quite understand the task, it directed me well[...]" • "[...]giving me suggestions, helping me even on how not to give up-because maybe there were also harder moments-thus also giving me alternatives and suggestions to see the path in a different way, make it lighter[...] I had a great time"
Personalization of the conversation	• "[...]the redundancy of the messages[...] sometimes it almost felt like a copy and paste of the messages and not an actual interaction[...] I found it a little depersonalized[...] I would have found it more enjoyable if it was personalized[...]" • "I saw that the person who was on the other side—I don't like to say 'Joy the guide!'—I still found her to be a person who each time, with respect to my mood, to how my week had gone, has put herself into my shoes, into my being, into my experience[...]"

Participant's Personal Experience With the Overall Intervention

Coherent with the *self-observation and activities evaluation* subtheme reported in Table 4, participants reported a positive personal experience as they seemed aligned with the content of the intervention, thus learning to appreciate the value of self-observation (regarding behavior and emotions) and how this can influence how they feel. For instance, the participant from the unguided group reported as follows:

> *[...]it [the intervention] was helpful, because it may not seem like it but by writing down daily what you do... you notice things that maybe normally you wouldn't notice, or [discover] free time that you can spare, which maybe you didn't even think about...*

Coherently, a participant from the guided group affirmed as follows:

> *[...]in relation to well-being, it definitely helped me understand[...] where to aim to get better... being at home during pregnancy, I rediscovered some hobbies that helped me... it helped me to not be so focused on negative things[...] to identify something positive to do day-by-day, to be able to accomplish little goals that [might not be important for others] but for me at that moment were important."*

Participants seem to have internalized the "message" that the intervention wanted to transmit. When asked whether they believed what they had done during the intervention could be useful to them in the future, participants reported already having integrated what they have learned into their daily life, for instance, by starting "[...]to keep a journal[...]" (a participant from the unguided group), even after pregnancy. Notably, a participant from the guided group reported as follows:

> *[E]ven in conditions of not pregnancy... it might be a good method at other times in life when one may experience difficulties.*

Referring to the subtheme of *learning the "method" and its application in the future*, this highlights that the intervention content was able to provide the participants with a broader method useful to support their psychological adjustment and well-being both during pregnancy and in the future:

> *[...]it's like, mentally, I got into this mode of planning something nice to do, to be able to have that weekly commitment that I like, to ask somebody to do it with me. [These] may be small things, but I'm sure it will help me in the future as well.* [Guided group participant]

In this regard, and referring to the *effort subtheme*, participants reported that 6 weeks of intervention were enough to this end, while the subsequent optional 3 weeks were perceived as a bit redundant and excessive. In addition, they stressed the effort required by the intervention overall, affirming, for instance, as follows:

> *[A]s the weeks went on[...] I found it more burdensome[...] filling out the daily form chart [daily monitoring form] every day with the activities perhaps, as time went on, was challenging, in the sense that you have to stop and really take [your] time and be consistent, when as the pregnancy progresses maybe other thoughts take over and you can't quite be that consistent all the time[...]* [Guided group participant]

On a similar note, another participant reported as follows:

> *[...]the part, let's say, the most obnoxious, difficult... is definitely the material to fill out during the last weeks. I have to tell the truth; I didn't even finish them because I didn't have time[...]* [Guided group participant]

Finally, coherent with the emerging subthemes, a further comment made by a participant from the guided group ought to be reported; she stressed that women would need to be already motivated to be able to appreciate the intervention. Indeed, she mentioned a key point, which is the need to have an adequate capacity for insight, as it might otherwise be difficult to autonomously notice the maladaptive behavioral patterns and switch to more positive ones. In this regard, the participant reported as follows:

> *[...]they [those following this intervention] must be people[...] who are capable of introspection... I imagine people who don't have so much of a way of knowing themselves[...] it's not so easy to start on such a path if you haven't done some work on yourself first[...] also, just to have self-knowledge and say*

"what are my weak points?" and say "where can I go with that?"

UX and UE Themes

Regarding UX and UE, the *managing issues and learnability* subtheme was quite prominent. All 3 (100%) women agreed that the Moodle app was not so easy to use overall, as some participants' approach has been that of learning how to work around what was not working to be able to continue the intervention. For instance, within an otherwise positive experience, a participant reported as follows:

> *...maybe the only thing that I would change, which is not really of the intervention though, [regards] more the use of the platform, is just that[...] I had trouble writing things down [for the homework...] even [the speed of the] connection when it makes you log back in to do the quiz, the page has to reload[...].* [Unguided group participant]

Similarly, a participant from the guided group reported as follows:

[Although she was a] *geeky chick* [as regards to the use of technology], *it was not easy* [to use the Moodle platform]. *It took me a while to find the material[...] I had even emailed you* [the researcher conducting the interview]*... I couldn't really understand how it worked, where they* [the materials] *were[...] I always used it from the smartphone, only a couple of times I used it from my computer[...] maybe you kind of have to learn first* [how to use the platform][...] *then I had that glitch with the quiz, and that one I found a little obnoxious, because I thought I had done the quiz[...]. it had not saved nothing. It wasn't easy to tell when it saved what you had done, because it was a little dubious, you know, sometimes[...] it was easier to use once you had gotten the hang of it[...].*

Altogether, this stresses the importance of the platform's simplicity, ease of use, and related learnability regarding UX and UE. However, it should be noted, referring to the *multimedia material* subtheme, that the aesthetic of the material present, and particularly the videos and images, were much appreciated and perceived as informative.

Furthermore, coherent with the *app interventions in the future* subtheme, all 3 (100%) participants reported that they did believe that a smartphone app, being readily available, could be a valuable tool to administer this sort of intervention saying, for instance, that "...with the current use of the phone and the computer in general...an app is useful" (the participant from the unguided group) as one can "...consult it wherever and whenever..." (a participant from the guided group). Indeed, albeit reporting difficulties with the platform, all 3 (100%) participants had already autonomously recommended the intervention to fellow pregnant women. However, the moving force, coherent with what is reported earlier, was the intervention content:

[Although deployed through a portable tool, interventions] *help in times of transition or change. This tool* [the present digital intervention] *helps because it focuses on pleasant activities, and in such a long waiting time* [the pregnancy], *with the struggles to organize exams, visits, what's going to happen tomorrow[...] it helps you a little bit to[...] focus on simpler thing, that then in itself just help you every day. So yes, I would recommend it for that* [Guided group participant]

The Experience With the Guide

Regarding the role of the guide and the guided group's overall experience with the chat interactions, the *support* subtheme has emerged, as both participants reported the positive value of having this sort of support, whether it be practical or affective. However, intersecting with this support subtheme, the *personalization of the conversation* subtheme seemed to have weighted on participant perception of the quality of the support perceived. In particular, it seems plausible to hypothesize that as participants knew that the guide was an actual person, the way of talking of the guide in the free sections of their protocol guided the participants' perception of their capacity for empathy and perception of getting in tune with them. Indeed, while 1 (33%) of the 3 participants reported the intervention to be not personalized enough, 1 (33%) reported almost the opposite. More specifically, a participant reported as follows:

> *[...]the interaction itself was effective in the sense that it explained things to me when I had doubts, that it directed me, maybe, when I didn't quite understand the task, it directed me well. Um, the part that I definitely didn't like was the redundancy of the messages because sometimes it almost felt like a copy and paste of the messages and not an actual interaction[...] I found it a little depersonalized[...] I would have found it more enjoyable if it was more personalized.*

The other, instead, reported expecting from the beginning a set of predetermined questions from the guide, particularly as some questions were indeed repeated each week; however, the participant reported about the guide as follows:

> *[S]aw that the person who was on the other side, I don't like to say "Joy the guide"- I found them to be a person who each time, with respect to my mood, to how my week had gone, has put herself into my shoes, into my being, into my experience, giving me suggestions, helping me even on how not to give up-because maybe there were also harder moments-thus also giving me alternatives and suggestions to see the path in a different way, to make it lighter. So, I had a great time!*

Such difference in the perception of the guide and of their helpfulness seems even more plausible when considering that a participant (who did not agree to the semistructured interview) had finished viewing the modules on Moodle until T6 but did not continue with interactions with the guide after T3.

Discussion

Principal Findings

This study aimed to investigate the feasibility of the BATD-R protocol [31] that was structured as an internet-based self-help intervention and deployed to pregnant women with subclinical depression symptoms, while further comparing its guided versus unguided versions. Such evaluations had the associated purpose of collecting the feedback needed to refine the intervention, thereby allowing its co-design and adaptation; moreover, it represents the first instance where the replication of the in-person BATD-R into a digital format has been empirically measured and evaluated, yielding valuable insights and shortcomings useful for future research.

Overall, results showed that this first version of the digital BA intervention, JuNEX, as a "replica" of the original BATD-R protocol [31], is not feasible to be implemented in a digital e-learning format. Specifically, data have highlighted the need to lighten the intervention and reduce the effort required for homework. Despite the perceived excessive effort, participants appreciated the content and purpose of the intervention, which allowed them to "take a moment" for themselves and understand "where to aim to get better." Indeed, albeit the intervention protocol originated from a behavioral framework, its focus on the person's everyday activities, the evaluation of one's own experiences, and how these are linked to how behaviors and emotions mutually influence each other configure in line with third-generation cognitive behavioral therapies [54]. These therapies prioritize holistic enhancement of psychological and behavioral processes related to health and well-being [55], emphasizing adaptive coping methods and increasing experiential and contextual awareness [55,56]. Such an approach is particularly relevant for nonclinical populations (which are more diverse than clinical populations), as they allow for a more transversal relevance and application of the coping methods promoted, thus making the intervention especially valuable in preventive terms. Given that this study focused on women with subclinical depression symptoms, interventions emphasizing the awareness of psychosocial functioning and the interplay between emotions and behaviors may be more beneficial than targeting specific, limited areas of functioning and distress. This broader approach could enhance women's capacity for adjustment and could be applicable to difficult situations beyond pregnancy-related challenges.

Such explanations serve to emphasize that consistent with women's feedback, the intents of the intervention per se, and thus its content as well as homework purposes, ought to be maintained as they were found pertinent to the pregnancy situation and were appreciated by women; however, data also stress the need to structure the intervention so that it can be deployed with greater ease and the need to shorten it to maximum of 6 weeks. In this regard, the data do point to satisfactory usability and UX as pivotal aspects of the intervention feasibility, which was not provided either by the Moodle platforms or the Google Docs used. These platforms were used in line with the preliminary and exploratory nature of this study, allowing the first structuring of the BATD-R protocol in a digital setting in a time- and cost-efficient manner; this has favored the co-design of the future development of the intervention by developing a more advanced and refined application after having collected some initial pregnant women's feedback.

When evaluating the usability of a digital solution, 5 main aspects are to be considered as follows [57]: (1) the simplicity of use experienced by users when learning how to use a digital tool, (2) the number of mistakes they make to do a certain action correctly, (3) the effectiveness with which users interact with a digital tool, (4) the perceived satisfaction with the UX, and (5) the memorability of how to use a digital tool after having been exposed to it. In addition to this, and particularly linked to both the effectiveness and perceived satisfaction just mentioned, is the more specific UE, thus linked to the subjective experience of the user and subsuming affective, cognitive, and behavioral components [25]. Considering the UX and UE evaluated in this study, data suggest that usability and UE were mediocre, yet not scarce. Thus, there was some appreciation for the tools used by the participants who had completed the interventions; however, the simultaneous high dropout and overall low recruitment of participants strengthen the idea that the more positive feedback given is somewhat linked to the a priori internal motivation of the participants in following the intervention. Participants themselves have indeed underlined that to follow the intervention as it is, to be able to bear the effort required by it, and to work around the limits of the platform used, women would need to be highly motivated on their own.

Coherent with the importance of the users' motivation to properly follow such interventions, data do suggest that the guided group showed greater adherence and were overall more willing to finish the intervention compared to the unguided group. A part of the guides' job was indeed to motivate women to favor compliance and adherence, and as women knew that they were interacting with a psychologist, this can be thought to have further increased their motivation. However, the practical challenges of having to set a date and time for the interactions are the limiting factors and so are the individual differences related to the guides' different writing modalities. In this regard, a step further in this direction might be the design and inclusion of a conversational agent to provide guidance and support during the digital intervention. Conversational agents can greatly favor the personalization of the user-system interactions while fostering scalability by requiring a much-reduced workforce for the intervention administration [58]. Existing literature has highlighted that conversational agents might be valuable tools to foster intervention adherence by favoring engagement and involvement, thereby supporting the overall UX [58,59] and allowing for a more immersive experience. Using a conversational agent is expected to further reduce time constraints for both patients and clinical professionals while also reducing health care costs in the long run. Within the perinatal context, future studies should thus evaluate the potentiality of including a conversational agent within digital interventions to ultimately support women's motivation and engagement as well as their compliance and adherence to the intervention, allowing them to use such digital solutions more freely while still giving the feeling of a more personalized experience.

Coherently, and in line with past evidence [60-62], women have expressed a desire for web-based solutions that support their psychological well-being. Furthermore, the data in this study seem to suggest that women might prefer app-based solutions, as almost all (3/4, 75%) participants that completed at least up to T6 used the app version instead of the web version to follow the intervention. Compared with web-based programs, interventions deployed through a smartphone app are much more readily available wherever and whenever, thus being overall easier to use and access.

Limitations

This paper reports the first-phase study's findings, highlighting valuable insights and several limitations that can guide future research and the development of digital interventions. First and foremost is the inherent challenge of having endeavored to replicate or simulate an in-person intervention digitally, a task that resulted in an ineffective outcome in our case. It is crucial to acknowledge that digital solutions offer a unique environment and set of possibilities that distinguish them from face-to-face interactions. Attempting to mirror traditional methods within this digital landscape may fall short of fully capitalizing on the advantages that digital interventions can bring. With our findings, we thus strengthened the idea that digital interventions should not be mere replicas of their in-person counterparts; rather, they should harness distinctive strengths and capabilities. This challenge underscores the need for innovation and adaptation, as well as a recognition that a direct translation of traditional methods may not always yield optimal results in the digital sphere. Future studies should first try to conduct workshops with end users in which the in-person protocol is administered to discuss feasible changes to the intervention, thus singling out the intervention's main principle and appreciated practices and then adapting them to the digital format. Furthermore, a subsequent limitation of the study overall is the limited sample size, which has prevented the possibility of computing any statistical comparisons between the 2 groups as well as generalizing findings. Nonetheless, given the preliminary and exploratory nature of this study, aimed at setting the base for the co-design of the final intervention, the data collected were still able to provide valuable insights directing the refinement and future developments of the intervention. Further limitations are attributed to the platforms used, as they are created with different purposes than what they were used for in this study. This has warranted structuring the intervention content based on the functionalities of this platform instead of creating a tool tailored to the intervention requirements; therefore, future studies are advised to avoid using such tools to administer and evaluate digital intervention, even in the first-phase studies such as this one. However, it should still be emphasized that such platforms (ie, Moodle and Google Documents) have allowed to "prototype" the intervention in a time- and cost-efficient manner while collecting the information needed to refine the intervention and to create a specific app that can meet the requirements and preferences of the users.

Conclusions

This study had the purpose of evaluating the feasibility of the BATD-R protocol [31] structured as an internet-based self-help intervention among pregnant women with subclinical depression symptoms while comparing its guided versus unguided versions. A subsequent goal was the collection of women's feedback and perceptions of the intervention content through a prototyped version based on an e-learning platform to allow the co-design of the final intervention.

Overall, the ease with which the intervention can be followed has emerged as a central component to account for in the future developments of the intervention, in terms of the intervention itself as well as the platform used to administer it. The greater effort perceived was related to the homework, whereby women did emphasize that performing them was effortful and time-consuming. Even so, comparing the 2 versions of intervention, the guided version was more well-received than the unguided version. Nonetheless, both groups expressed satisfaction with the intervention's content and felt that it was relevant to their personal experiences with pregnancy.

Acknowledgments

The authors would like to thank the University of Padova and the Scuola di Specializzazione in Psicologia del Ciclo di Vita (Specialization School of Life Cycle Psychology, Padova, Italy) for allowing the inclusion of their psychotherapist in training to work as guides in the guided intervention. Moreover, the authors would like to thank the Artigianelli Institute (Trento, Italy) for creating the videos' illustrative content.

Conflicts of Interest

None declared.

References

1. American Psychiatric Association. Diagnostic And Statistical Manual Of Mental Disorders 5Ed Dsm-5. New York, NY: American Psychological Association; 2013.
2. Batt MM, Duffy KA, Novick AM, Metcalf CA, Epperson CN. Is postpartum depression different from depression occurring outside of the perinatal period? A review of the evidence. Focus (Am Psychiatr Publ) 2020 Apr;18(2):106-119 [FREE Full text] [doi: 10.1176/appi.focus.20190045] [Medline: 33162848]
3. Howard LM, Khalifeh H. Perinatal mental health: a review of progress and challenges. World Psychiatry 2020 Oct;19(3):313-327 [FREE Full text] [doi: 10.1002/wps.20769] [Medline: 32931106]
4. Yin X, Sun N, Jiang N, Xu X, Gan Y, Zhang J, et al. Prevalence and associated factors of antenatal depression: systematic reviews and meta-analyses. Clin Psychol Rev 2021 Mar;83:101932. [doi: 10.1016/j.cpr.2020.101932] [Medline: 33176244]

5. Wang Z, Liu J, Shuai H, Cai Z, Fu X, Liu Y, et al. Mapping global prevalence of depression among postpartum women. Transl Psychiatry 2021 Oct 20;11(1):543 [FREE Full text] [doi: 10.1038/s41398-021-01663-6] [Medline: 34671011]
6. Agostini F, Neri E, Salvatori P, Dellabartola S, Bozicevic L, Monti F. Antenatal depressive symptoms associated with specific life events and sources of social support among Italian women. Matern Child Health J 2015 May 11;19(5):1131-1141. [doi: 10.1007/s10995-014-1613-x] [Medline: 25303805]
7. Cena L, Mirabella F, Palumbo G, Gigantesco A, Trainini A, Stefana A. Prevalence of maternal antenatal and postnatal depression and their association with sociodemographic and socioeconomic factors: a multicentre study in Italy. J Affect Disord 2021 Jan 15;279:217-221 [FREE Full text] [doi: 10.1016/j.jad.2020.09.136] [Medline: 33069119]
8. Corbani IE, Rucci P, Iapichino E, Quartieri Bollani M, Cauli G, Ceruti MR, et al. Comparing the prevalence and the risk profile for antenatal depressive symptoms across cultures. Int J Soc Psychiatry 2017 Nov 14;63(7):622-631. [doi: 10.1177/0020764017725543] [Medline: 28805152]
9. Giardinelli L, Innocenti A, Benni L, Stefanini MC, Lino G, Lunardi C, et al. Depression and anxiety in perinatal period: prevalence and risk factors in an Italian sample. Arch Womens Ment Health 2012 Mar 29;15(1):21-30. [doi: 10.1007/s00737-011-0249-8] [Medline: 22205237]
10. Grussu P, Quatraro RM. Prevalence and risk factors for a high level of postnatal depression symptomatology in Italian women: a sample drawn from ante-natal classes. Eur Psychiatr 2020 Apr 16;24(5):327-333. [doi: 10.1016/j.eurpsy.2009.01.006]
11. Hadfield H, Wittkowski A. Women's experiences of seeking and receiving psychological and psychosocial interventions for postpartum depression: a systematic review and thematic synthesis of the qualitative literature. J Midwifery Womens Health 2017 Nov;62(6):723-736. [doi: 10.1111/jmwh.12669] [Medline: 29210501]
12. Baldisserotto ML, Miranda Theme M, Gomez LY, Dos Reis TB. Barriers to seeking and accepting treatment for perinatal depression: a qualitative study in Rio de Janeiro, Brazil. Community Ment Health J 2020 Jan 10;56(1):99-106. [doi: 10.1007/s10597-019-00450-4] [Medline: 31512080]
13. Fonseca A, Gorayeb R, Canavarro MC. Women s help-seeking behaviours for depressive symptoms during the perinatal period: socio-demographic and clinical correlates and perceived barriers to seeking professional help. Midwifery 2015 Dec;31(12):1177-1185. [doi: 10.1016/j.midw.2015.09.002] [Medline: 26433622]
14. Fonseca A, Ganho-Ávila A, Lambregtse-van den Berg M, Lupattelli A, de la Fé Rodriguez-Muñoz M, Ferreira P, et al. Emerging issues and questions on peripartum depression prevention, diagnosis and treatment: a consensus report from the cost action riseup-PPD. J Affect Disord 2020 Sep 01;274:167-173. [doi: 10.1016/j.jad.2020.05.112] [Medline: 32469800]
15. van den Heuvel JF, Groenhof TK, Veerbeek JH, van Solinge WW, Lely AT, Franx A, et al. eHealth as the next-generation perinatal care: an overview of the literature. J Med Internet Res 2018 Jun 05;20(6):e202 [FREE Full text] [doi: 10.2196/jmir.9262] [Medline: 29871855]
16. Hussain-Shamsy N, Shah A, Vigod SN, Zaheer J, Seto E. Mobile health for perinatal depression and anxiety: scoping review. J Med Internet Res 2020 Apr 13;22(4):e17011 [FREE Full text] [doi: 10.2196/17011] [Medline: 32281939]
17. Mancinelli E, Bassi G, Gabrielli S, Salcuni S. The efficacy of digital cognitive-behavioral interventions in supporting the psychological adjustment and sleep quality of pregnant women with sub-clinical symptoms: a systematic review and meta-analysis. Int J Environ Res Public Health 2022 Aug 03;19(15):9549 [FREE Full text] [doi: 10.3390/ijerph19159549] [Medline: 35954901]
18. Mancinelli E, Dell'Arciprete G, Pattarozzi D, Gabrielli S, Salcuni S. Digital behavioral activation interventions during the perinatal period: scoping review. JMIR Pediatr Parent 2023 Mar 28;6:e40937 [FREE Full text] [doi: 10.2196/40937] [Medline: 36853756]
19. Dimidjian S, Barrera M, Martell C, Muñoz RF, Lewinsohn PM. The origins and current status of behavioral activation treatments for depression. Annu Rev Clin Psychol 2011;7:1-38. [doi: 10.1146/annurev-clinpsy-032210-104535] [Medline: 21275642]
20. Dimidjian S, Goodman SH, Sherwood NE, Simon GE, Ludman E, Gallop R, et al. A pragmatic randomized clinical trial of behavioral activation for depressed pregnant women. J Consult Clin Psychol 2017 Jan;85(1):26-36 [FREE Full text] [doi: 10.1037/ccp0000151] [Medline: 28045285]
21. Jacobson NS, Martell CR, Dimidjian S. Behavioral activation treatment for depression: returning to contextual roots. Clin Psychol Sci Pract 2001;8(3):255-270. [doi: 10.1093/clipsy.8.3.255]
22. Soucy Chartier I, Provencher MD. Behavioural activation for depression: efficacy, effectiveness and dissemination. J Affect Disord 2013 Mar 05;145(3):292-299. [doi: 10.1016/j.jad.2012.07.023] [Medline: 22884236]
23. Kanter JW, Manos RC, Bowe WM, Baruch DE, Busch AM, Rusch LC. What is behavioral activation? A review of the empirical literature. Clin Psychol Rev 2010 Aug;30(6):608-620. [doi: 10.1016/j.cpr.2010.04.001] [Medline: 20677369]
24. ISO 9241-11:2018: ergonomics of human-system interaction: part 11: usability: definitions and concepts. International Organization for Standardization. 2018. URL: https://www.iso.org/standard/63500.html [accessed 2024-04-29]
25. O'Brien HL, Cairns P, Hall MA. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. Int J Hum Comput Stud 2018 Apr;112:28-39. [doi: 10.1016/j.ijhcs.2018.01.004]

26. Davis JA, Ohan JL, Gibson LY, Prescott SL, Finlay-Jones AL. Understanding engagement in digital mental health and well-being programs for women in the perinatal period: systematic review without meta-analysis. J Med Internet Res 2022 Aug 09;24(8):e36620 [FREE Full text] [doi: 10.2196/36620] [Medline: 35943773]
27. Vial S, Boudhraâ S, Dumont M. Human-centered design approaches in digital mental health interventions: exploratory mapping review. JMIR Ment Health 2022 Jun 07;9(6):e35591 [FREE Full text] [doi: 10.2196/35591] [Medline: 35671081]
28. Dorst K, Dijkhuis J. Comparing paradigms for describing design activity. Des Stud 1995 Apr;16(2):261-274. [doi: 10.1016/0142-694x(94)00012-3]
29. Dorst K. Design beyond design. She Ji J Des Econ Innov 2019;5(2):117-127. [doi: 10.1016/j.sheji.2019.05.001]
30. Dorst K. Design research: a revolution-waiting-to-happen. Des Stud 2008 Jan;29(1):4-11. [doi: 10.1016/j.destud.2007.12.001]
31. Lejuez CW, Hopko DR, Acierno R, Daughters SB, Pagoto SL. Ten year revision of the brief behavioral activation treatment for depression: revised treatment manual. Behav Modif 2011 Mar;35(2):111-161. [doi: 10.1177/0145445510390929] [Medline: 21324944]
32. Lyon AR, Koerner K. User-centered design for psychosocial intervention development and implementation. Clin Psychol (New York) 2016 Jun 17;23(2):180-200 [FREE Full text] [doi: 10.1111/cpsp.12154] [Medline: 29456295]
33. Czajkowski SM, Powell LH, Adler N, Naar-King S, Reynolds KD, Hunter CM, et al. From ideas to efficacy: the ORBIT model for developing behavioral treatments for chronic diseases. Health Psychol 2015 Oct;34(10):971-982 [FREE Full text] [doi: 10.1037/hea0000161] [Medline: 25642841]
34. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
35. Randomizer. AnnaBet. URL: https://annabet.com/it/tools/Randomizer/ [accessed 2023-06-19]
36. Daughters SB, Braun AR, Sargeant MN, Reynolds EK, Hopko DR, Blanco C, et al. Effectiveness of a brief behavioral treatment for inner-city illicit drug users with elevated depressive symptoms: the life enhancement treatment for substance use (LETS Act!). J Clin Psychiatry 2008 Jan 15;69(1):122-129. [doi: 10.4088/jcp.v69n0116] [Medline: 18312046]
37. MacPherson L, Tull MT, Matusiewicz AK, Rodman S, Strong DR, Kahler CW, et al. Randomized controlled trial of behavioral activation smoking cessation treatment for smokers with elevated depressive symptoms. J Consult Clin Psychol 2010 Mar;78(1):55-61. [doi: 10.1037/a0017939] [Medline: 20099950]
38. Hopko DR, Lejuez CW, LePage JP, Hopko SD, McNeil DW. A brief behavioral activation treatment for depression. A randomized pilot trial within an inpatient psychiatric hospital. Behav Modif 2003 Sep 26;27(4):458-469. [doi: 10.1177/0145445503255489] [Medline: 12971122]
39. Sorveglianza dei disturbi depressivi: come lavora il Passi. Istituto Superiore di Sanità. URL: https://www.epicentro.iss.it/passi/DepressioneContesto [accessed 2023-01-29]
40. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: Dsm-IV : International Version With Icd-10 Codes. New York, NY: American Psychiatric Association; 1998.
41. Shevlin M, Butter S, McBride O, Murphy J, Gibson-Miller J, Hartman TK, et al. Measurement invariance of the Patient Health Questionnaire (PHQ-9) and Generalized Anxiety Disorder scale (GAD-7) across four European countries during the COVID-19 pandemic. BMC Psychiatry 2022 Mar 01;22(1):154 [FREE Full text] [doi: 10.1186/s12888-022-03787-5] [Medline: 35232409]
42. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh postnatal depression scale. Br J Psychiatry 1987 Jun 2;150(06):782-786. [doi: 10.1192/bjp.150.6.782] [Medline: 3651732]
43. Benvenuti P, Ferrara M, Niccolai C, Valoriani V, Cox JL. The Edinburgh Postnatal Depression Scale: validation for an Italian sample. J Affect Disord 1999 May;53(2):137-141. [doi: 10.1016/s0165-0327(98)00102-5] [Medline: 10360408]
44. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
45. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385. [doi: 10.2307/2136404]
46. Mondo M, Sechi C, Cabras C. Psychometric evaluation of three versions of the Italian Perceived Stress Scale. Curr Psychol 2019 Jan 08;40(4):1884-1892. [doi: 10.1007/s12144-019-0132-8]
47. Manos RC, Kanter JW, Luo W. The behavioral activation for depression scale-short form: development and validation. Behav Ther 2011 Dec;42(4):726-739. [doi: 10.1016/j.beth.2011.04.004] [Medline: 22036000]
48. Armento ME, Hopko DR. The Environmental Reward Observation Scale (EROS): development, validity, and reliability. Behav Ther 2007 Jun;38(2):107-119. [doi: 10.1016/j.beth.2006.05.003] [Medline: 17499078]
49. Domnich A, Arata L, Amicizia D, Signori A, Patrick B, Stoyanov S, et al. Development and validation of the Italian version of the mobile application rating scale and its generalisability to apps targeting primary prevention. BMC Med Inform Decis Mak 2016 Jul 07;16(1):83 [FREE Full text] [doi: 10.1186/s12911-016-0323-2] [Medline: 27387434]
50. RStudio Team. RStudio: integrated development for R. RStudio. URL: https://www.r-project.org/conferences/useR-2011/abstracts/180111-allairejj.pdf [accessed 2024-04-29]
51. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

52. Gabrielli S, Rizzi S, Carbone S, Donisi V. A Chatbot-based coaching intervention for adolescents to promote life skills: pilot study. JMIR Hum Factors 2020 Mar 14;7(1):e16762 [FREE Full text] [doi: 10.2196/16762] [Medline: 32130128]
53. World Medical Association. Dichiarazione di Helsinki della World Medical Association: principi etici per la ricerca biomedica che coinvolge gli esseri umani. Evidence 2013;5(10):e1000059. [doi: 10.4470/E1000059]
54. Hayes SC, Hofmann SG. "Third-wave" cognitive and behavioral therapies and the emergence of a process-based approach to intervention in psychiatry. World Psychiatry 2021 Oct 09;20(3):363-375 [FREE Full text] [doi: 10.1002/wps.20884] [Medline: 34505370]
55. Feliu-Soler A, Cebolla A, McCracken LM, D'Amico F, Knapp M, López-Montoyo A, et al. Economic impact of third-wave cognitive behavioral therapies: a systematic review and quality assessment of economic evaluations in randomized controlled trials. Behav Ther 2018 Jan;49(1):124-147 [FREE Full text] [doi: 10.1016/j.beth.2017.07.001] [Medline: 29405918]
56. Hayes SC, Hofmann SG. The third wave of cognitive behavioral therapy and the rise of process-based care. World Psychiatry 2017 Oct;16(3):245-246 [FREE Full text] [doi: 10.1002/wps.20442] [Medline: 28941087]
57. Nielsen J. Usability Engineering. Boston, MA: Academic Press; 1993.
58. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied conversational agents in clinical psychology: a scoping review. J Med Internet Res 2017 May 09;19(5):e151 [FREE Full text] [doi: 10.2196/jmir.6553] [Medline: 28487267]
59. Gaffney H, Mansell W, Tai S. Conversational agents in the treatment of mental health problems: mixed-method systematic review. JMIR Ment Health 2019 Oct 18;6(10):e14166 [FREE Full text] [doi: 10.2196/14166] [Medline: 31628789]
60. Ashford MT, Ayers S, Olander EK. Interest in web-based treatments for postpartum anxiety: an exploratory survey. J Reprod Infant Psychol 2017 Sep 02;35(4):394-409. [doi: 10.1080/02646838.2017.1320364] [Medline: 29517373]
61. Lackie ME, Parrilla JS, Lavery BM, Kennedy AL, Ryan D, Shulman B, et al. Digital health needs of women with postpartum depression: focus group study. J Med Internet Res 2021 Jan 06;23(1):e18934 [FREE Full text] [doi: 10.2196/18934] [Medline: 33404506]
62. Wenze S, Miers QA, Battle CL. Postpartum mental health care for mothers of multiples: a qualitative study of new mothers' treatment preferences. J Psychiatr Pract 2020 May;26(3):201-214. [doi: 10.1097/PRA.0000000000000469] [Medline: 32421291]

Abbreviations

BA: behavioral activation
BATD-R: Behavioral Activation Treatment for Depression-Revised
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders (Fourth Edition)
MARS: Mobile Application Rating Scale
PHQ-9: Patient Health Questionnaire-9
UE: user engagement
UX: user experience

Edited by Y Senathirajah; submitted 19.06.23; peer-reviewed by E Neter, E Toki; comments to author 02.09.23; revised version received 05.10.23; accepted 01.03.24; published 16.05.24.

Please cite as:
Mancinelli E, Gabrielli S, Salcuni S
A Digital Behavioral Activation Intervention (JuNEX) for Pregnant Women With Subclinical Depression Symptoms: Explorative Co-Design Study
JMIR Hum Factors 2024;11:e50098
URL: https://humanfactors.jmir.org/2024/1/e50098
doi:10.2196/50098
PMID:38753421

Original Paper

Trust in and Acceptance of Artificial Intelligence Applications in Medicine: Mixed Methods Study

Daria Shevtsova[1,2], MSc; Anam Ahmed[1], MSc; Iris W A Boot[1], MSc; Carmen Sanges[3], PhD; Michael Hudecek[3], PhD; John J L Jacobs[4], PhD; Simon Hort[5], MSc; Hubertus J M Vrijhoef[1], PhD

[1]Panaxea bv, Den Bosch, Netherlands

[2]Vrije Universiteit Amsterdam, Amsterdam, Netherlands

[3]Universitätsklinikum Würzburg, Würzburg, Germany

[4]Ortec bv, Zoetermeer, Netherlands

[5]Fraunhofer Institute for Production Technology, Aachen, Germany

Corresponding Author:
Anam Ahmed, MSc
Panaxea bv
Pettelaarpark 84
Den Bosch, 5216 PP
Netherlands
Phone: 31 639421854
Email: anam.ahmed@panaxea.eu

Abstract

Background: Artificial intelligence (AI)–powered technologies are being increasingly used in almost all fields, including medicine. However, to successfully implement medical AI applications, ensuring trust and acceptance toward such technologies is crucial for their successful spread and timely adoption worldwide. Although AI applications in medicine provide advantages to the current health care system, there are also various associated challenges regarding, for instance, data privacy, accountability, and equity and fairness, which could hinder medical AI application implementation.

Objective: The aim of this study was to identify factors related to trust in and acceptance of novel AI-powered medical technologies and to assess the relevance of those factors among relevant stakeholders.

Methods: This study used a mixed methods design. First, a rapid review of the existing literature was conducted, aiming to identify various factors related to trust in and acceptance of novel AI applications in medicine. Next, an electronic survey including the rapid review–derived factors was disseminated among key stakeholder groups. Participants (N=22) were asked to assess on a 5-point Likert scale (1=irrelevant to 5=relevant) to what extent they thought the various factors (N=19) were relevant to trust in and acceptance of novel AI applications in medicine.

Results: The rapid review (N=32 papers) yielded 110 factors related to trust and 77 factors related to acceptance toward AI technology in medicine. Closely related factors were assigned to 1 of the 19 overarching umbrella factors, which were further grouped into 4 categories: human-related (ie, the type of institution AI professionals originate from), technology-related (ie, the explainability and transparency of AI application processes and outcomes), ethical and legal (ie, data use transparency), and additional factors (ie, AI applications being environment friendly). The categorized 19 umbrella factors were presented as survey statements, which were evaluated by relevant stakeholders. Survey participants (N=22) represented researchers (n=18, 82%), technology providers (n=5, 23%), hospital staff (n=3, 14%), and policy makers (n=3, 14%). Of the 19 factors, 16 (84%) human-related, technology-related, ethical and legal, and additional factors were considered to be of high relevance to trust in and acceptance of novel AI applications in medicine. The patient's gender, age, and education level were found to be of low relevance (3/19, 16%).

Conclusions: The results of this study could help the implementers of medical AI applications to understand what drives trust and acceptance toward AI-powered technologies among key stakeholders in medicine. Consequently, this would allow the implementers to identify strategies that facilitate trust in and acceptance of medical AI applications among key stakeholders and potential users.

(JMIR Hum Factors 2024;11:e47031) doi:10.2196/47031

KEYWORDS

trust; acceptance; artificial intelligence; medicine; mixed methods; rapid review; survey

Introduction

Artificial Intelligence

Artificial intelligence (AI) is commonly defined as a computer system that uses statistical models, diverse algorithms, and self-modifying systems to make predictions and decisions based on its own aggregated experience. It can therefore perform tasks that usually require or even surpass the human level of intelligence [1,2]. AI has been increasingly integrated in the health care sector, where it helps with administrative workflows, diagnostic image analysis, robotic surgery, and clinical decision-making. Consequently, medical AI applications allow, amongst other things, earlier disease detection, patient-tailored treatments, and more efficient follow-ups, which should drive the health care costs down upon implementation [3]. Although medical AI applications provide various advantages to the current health care system, such as increased efficiency and improved workflows [4], there are also various challenges associated with AI implementation. For instance, a large share of potential users has concerns over privacy issues [5,6]. Equity and fairness are other important concerns, since there is a risk of perpetuating bias within data sets being adopted by AI technology [5,7,8]. Further, the implementation of AI into medical practice raises the question of accountability, since it is currently unclear whether technology developers, hospitals, or regulators should be responsible for mistakes or undesirable outcomes from the use of an AI application.

Trust and Acceptance

To ensure successful implementation of medical AI applications, it is essential to build trust in and acceptance of AI technology among its users [9,10]. In this study, a model of key drivers of trust in and acceptance of AI systems was used [11]. According to this model, trust is influenced by 4 drivers: current safeguards, job impact of AI, familiarity of AI, and AI uncertainty. Current safeguards indicate the belief that current regulations and laws are adequate for ensuring the safety of AI and the protection of people who use it. Job impact refers to the belief that there will be more jobs generated than eliminated due to AI implementation. Familiarity with AI is the level of understanding of how AI technology works and how AI applications are used. These 3 drivers have a positive influence on trust, with current safeguards being its strongest driver. The fourth driver, AI uncertainty, impacts trust in a negative way. It implies the belief that the impact of AI on society is unpredictable and the technology is still not fully explored. Overall, these drivers influence the extent to which people trust the AI system and believe it to be trustworthy. Trust, then, is a large contributor to the level of acceptance, which is the extent to which people accept or approve of AI and are willing to use it without resistance [11]. In the scientific literature, trust can be defined in different ways. In this study, we used literature-derived definitions of trust and acceptance in the context of AI implementation, namely:

- Trust is the belief of an individual that an AI application will do what it promises [12,13].
- Acceptance is the willingness of an individual to use the AI application in medicine [14].

Therefore, it can be argued that acceptance of an AI application depends on trust people have toward this technology [11,15,16]. At the same time, people can often accept their usage of technologies without necessarily trusting them [17]. Therefore, it is important to consider the 2 concepts separately as well as together.

Overall, widespread trust in and acceptance of an AI application is crucial for successful introduction and implementation of the technology. Failure to ensure trust in and acceptance of AI technology would pose the risk of “stifling innovation” and causing unnecessary “opportunity costs” [18]. The lack of trust in AI applications in medicine impedes their adoption in health care, compounded by inadequate public assurance and attention to concerns, thereby exacerbating these challenges. In addition, the anticipated benefits of AI-based innovations can coexist with significant acceptance barriers [15,18-21].

Investigating what factors contribute to trust in and acceptance of AI technology in medicine would help us understand how to make the implementation and regulatory approval of AI-powered advanced therapy manufacturing systems as efficient as possible. This can be achieved by collecting insights into stakeholders’ perspectives with regard to trust and acceptance toward medical AI applications [2,22]. Factors contributing to trust and acceptance toward medical AI applications can be attributed a different weight by various groups of stakeholders with distinct roles in AI.

Study Objectives

Since AI applications are still relatively new, users and providers are hesitant to trust and accept this new technology without restrictions. As for the future implementation of AI applications in treatment centers, it is essential that stakeholders (eg, clinicians, researchers, hospital staff) accept and trust the innovative AI-based manufacturing platform. Therefore, the aim of this study was first to identify the factors related to trust in and acceptance of AI technology in medicine and second to assess the relevance of those factors among relevant stakeholders in medicine.

Methods

Study Setting

This study is part of the European Union’s (EU) Horizon 2020 project AIDPATH (AI-driven Decentralized Production for Advanced Therapies in the Hospital; grant agreement number 101016909) [22,23]. It is an upcoming state-of-the-art AI application in hospitals, which aims to develop an AI-driven, automated chimeric antigen receptor T cell (CAR-T) manufacturing platform at the point of care as a treatment for acute leukemia and lymphoma. In CAR-T therapy, the patient’s own T cells are removed, genetically modified, and reinfused into the patient in order to find and eliminate tumour cells. Current production is characterized by laborious manual process steps, complex logistics, and a lack of process understanding. This results in long delivery times (up to 21 days) and high costs (approx €320,000, or US $347,890, per treatment) [24,25]. For this reason, AIDPATH is developing a system to fully automate the manufacturing process, from the provision of patient cells to the injection directly in the hospital. An important building block for effective and equitable manufacturing is AI. AI can provide essential process insights into the cell's characteristics and behavior. This offers a significant benefit for adaptive control of the whole process and the design of personalized process protocols. Furthermore, AI can assist cost-effective platform operation in a smart manufacturing hospital by improving manufacturing schedules and resource management [26]. In general, successful implementation of AIDPATH would serve as an example of an effective AI technology that automates the production and delivery of advanced therapy medicinal products (ATMPs). Furthermore, AI-powered technology can form the basis for a deployable platform for further pilot trials in multiple hospitals and would create a model innovation system for smart manufacturing hospitals [2,22].

In this study, to meet the study objectives, a rapid literature review was conducted, followed by a survey.

Rapid Literature Review

A rapid literature review of peer- and non-peer-reviewed publications was conducted to identify factors related to trust in and acceptance of AI applications used in medicine. As an alternative method to systematic reviews, a rapid review allows for accelerated synthesis of up-to-date evidence, while efficiently informing latest findings in recent health care research [27]. The peer- and non-peer-reviewed literature needed to be published between 2012 and 2022 in English. Data on attitudes toward AI in relation to prognosis, diagnosis, treatment, and care were included. The search was performed in PubMed/MEDLINE with the following search syntax: ((trust) OR (acceptance) OR (attitude) OR (perspective) OR (perception)) AND ((AI) OR (artificial intelligence) OR (machine learning) OR (deep learning)) AND (((prognosis) OR (diagnosis) OR (treatment) OR (care)) OR ((medic*) OR (clinic*) OR (hospital) OR (smart hospital) OR (health care)) AND ((survey) OR (questionnaire) OR (interview)). The reason for inclusion of only survey-, questionnaire-, or interview-based research in the search terms was due to their direct relevance to our research objectives.

In the non-peer-reviewed literature search, similar terms and time frame of publication were used and the first 10 pages on the Google Search engine were examined to identify other relevant papers and reports by (non)governmental and research organizations. This allowed the study findings to be applicable to a broad range of medical AI applications. Papers were screened, and data were extracted by 2 authors (DS and AA). The selected literature was analyzed to identify key trends and explanatory factors related to trust and acceptance toward medical AI applications. The factors were then grouped into 4 categories: human-related, technology-related, legal and ethical, and additional factors. These factor groups formed the basis of the survey designed to investigate factor relevance. This was performed independently by 2 authors (DS and AA).

Survey

The survey was reported in accordance with the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) guidelines [28]. The survey in English assessed the relevance of the factors related to trust in and acceptance of novel AI applications in medicine. The survey started with an introduction to AI applications in medicine and AIDPATH, followed by 7 general questions on each participant's background, including gender, age, the country they worked in, years of experience, the stakeholder group they belonged to, their familiarity with AI applications in medicine, and their general view on AI. In the last question, the following distinction was made between the answer options: "I embrace AI" meant welcoming and using AI as a constituent part of their work or life, "I approve of AI" implied that the participant agreed with the use of AI in their work or life but did not use it themselves, and "I accept AI" referred to acknowledging the use of AI in work or life but not being ready to fully approve it.

In the core section of the survey, the definitions of trust and acceptance were provided as a reference for participants. The core part also consisted of 2 identical lists of 19 factors related to trust and acceptance toward AI applications in medicine. Each factor was categorized into human-related, technology-related, legal and ethical, or additional factors. Human-related factors were linked to AI professionals assessed the relevance of the type of organization the AI professionals were affiliated to and the purpose to innovate with a specific AI application. With respect to health care professionals, the factors were related to the knowledge of AI applications and the attitude toward AI application usage in medicine. In relation to patients, the relevance of the following factors was assessed: general knowledge of AI applications in medicine, the attitude toward AI application usage in medicine, and the patient's age, gender, and level of education. Furthermore, participants were asked to evaluate the relevance of transparency between all parties involved in AI application use. Technology-related factors related to the performance of AI applications in medicine, the possibility of their integration into existing clinical workflows, a clear balance of risks and benefits of the AI applications, and the explainability and transparency of processes and outcomes. The legal and ethical factors were related to the adequacy of regulations and governance of AI applications in medicine, data use transparency, and clear accountability and responsibility for an AI application. The additional factors were concerned with the environmental sustainability of AI applications and AI's impact on job availability. For each factor, participants could indicate each factor's relevance to trust in and acceptance of AI applications from their stakeholder perspective using a Likert scale of 1-5, where 1 stood for "not relevant," 3 for "not irrelevant, nor relevant," and 5 for "relevant." Throughout the survey, "relevant" meant being highly significant for ensuring trust in or acceptance of AI applications, while "irrelevant" meant no significance. The N/A (not applicable) option was available as well for each factor. Open questions at the end of both sections allowed participants to suggest other relevant factors related to trust in or acceptance of AI applications that were not mentioned in the survey. Furthermore, the participants were invited to suggest any other factors, different from trust, deemed important for acceptance of AI applications in medicine.

Sampling

Using the convenience sampling method [29], AIDPATH Consortium members were requested to invite stakeholders in their network but outside the AIDPATH Consortium to fill in the survey on the SurveyMonkey platform. The survey was distributed by email to members of relevant stakeholder groups to capture their professional perspectives (eg, clinicians, scientists, and policy makers). Data were collected from April to May 2022 and analyzed using Microsoft Excel.

Data Collection and Analysis

After participants were asked to rate the relevance of each factor from 1 (irrelevant) to 5 (relevant), the mean score of each factor was determined by assigning each response a weight from 1 to 5. Next, means scores were calculated by finding an average of the sum of response values for each question. To visualize the survey responses and compare the mean scores for each factor included in the survey, a spider diagram was charted. This provided an overview of the factors' relevance and their relative importance in influencing both trust and acceptance toward AI applications in medicine. In addition, a scatter plot was created to obtain an overview of the interrelationship between the relevance to trust (x axis) in and acceptance (y axis) of AI applications in medicine. The plot allowed us to identify the degree of relevance of each factor in relation to both trust and acceptance. To classify the factors based on their relevance, score ranges were established. Factors with mean scores from 1 to 3 were considered to be of low relevance, while factors from 4 to 5 were deemed of high relevance. The open-question responses were considered when interpreting numerical data.

Ethical Considerations

Under Dutch law, no ethical approval was required according to Article 1b of the Dutch Medical Research in Human Subjects Act [30]. However, all participants were informed about the study objectives, their verbal consent was obtained, and all data were processed anonymously. All responses were recorded anonymously. Participants were informed of their right to withdraw from the study at any time without any consequences. They were not financially compensated.

Results

Rapid Literature Review

The literature search (Figure 1) yielded 301 hits in the PubMed database and 105 hits through gray literature search and snowballing. After screening titles and abstracts, 284 (70%) records were excluded. After full-text screening, 90 (73.8%) records were excluded primarily due to the absence of concepts of trust or acceptance and a lack of factors related to trust or acceptance in the main text or data-containing figures. As a result, 32 (26.2%) papers and reports [7,9-12,15,16,19,21,31-53] were included in the data analysis.

Overall, the rapid review identified a total of 110 factors related to trust and 77 factors related to acceptance toward medical AI technology. The full list of factors identified through the rapid review with corresponding studies can be found in Multimedia Appendix 1. Tables 1-4 show all factors from the rapid review, each with the frequency of its appearance in the literature and the corresponding overarching umbrella factors. Some factors from a single study are repeated in the same category in Tables 1 and 2 on trust and Tables 3 and 4 on acceptance or within the same category (eg, health care professionals and patients subsections of the human-related factors section). The most frequently reported human-related factors related to trust (Tables 1 and 2) in medical AI applications were knowledge and understanding of AI by health care professionals and knowledge and education of AI among patients. In terms of technology-related factors, accuracy, transparency, reliability, safety, and explainability of medical AI applications and their functioning appeared most often in the literature. Regarding legal and ethical factors, the most frequently occurring factors included fairness and equity of medical AI technology and the privacy and security of personal data handled by the AI systems. The most frequently presented human-related factors related to acceptance (Tables 3 and 4) of medical AI technology were the perceived usefulness and provision of better medical services by the AI technology. Regarding technology-related factors linked to acceptance, performance expectancy, design and output quality, and transparency were stated in the literature most often. A wide range of legal and ethical factors were mentioned in the literature, including adequate regulations of medical AI technology, protection and security of patients' data, and the allocation of accountability and responsibility for the (mal)functioning of an AI application. There were additional factors related to trust in and acceptance of medical AI technology (Tables 1-4). These included replacement of doctors by machines that lack a human touch and moral support, labor market implications, and environmental sustainability. Three studies also highlighted that acceptance of a medical AI application is directly related to trust in the AI application. Overall, there were fewer factors related to acceptance than those related to trust, whereas most of the overarching umbrella factors were fully represented in both tables. Therefore, an identical list of umbrella factors allocated within the 4 categories (human-related, technology-related, legal and ethical, and additional factors) was used in the survey for investigating the relevance of factors for both trust in and acceptance of AI applications in medicine.

Figure 1. PRISMA flow diagram of the rapid review literature screening. AI: artificial intelligence.

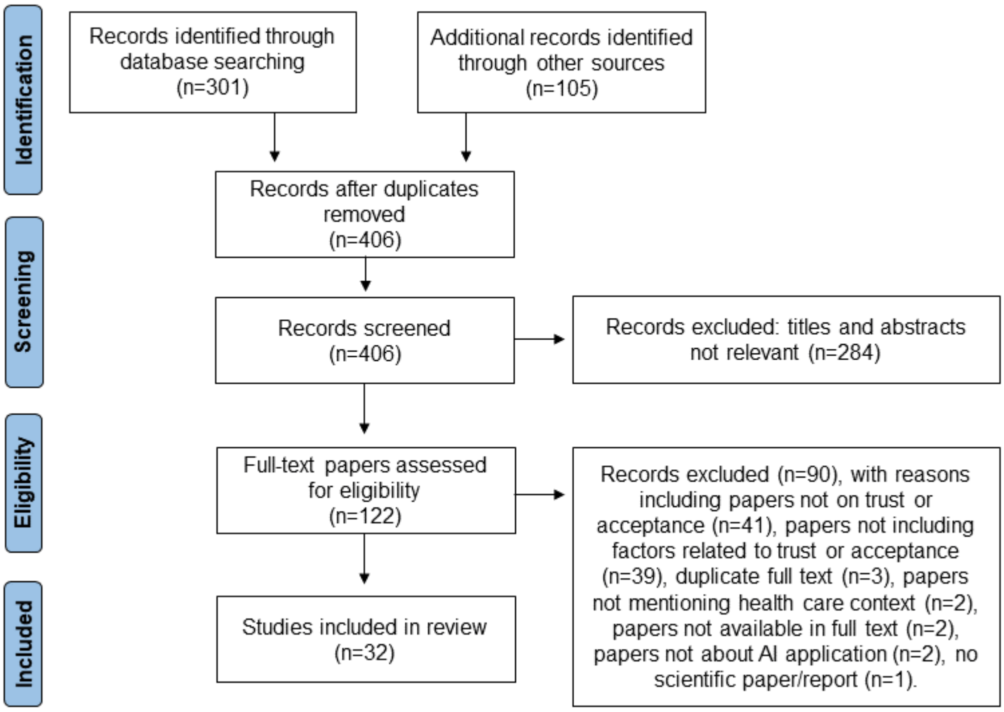


Table 1. Human-related factors related to trust (N=110) in medical AI[a] applications (22/32, 68.8%, studies).

Factor category and factors from the rapid review	Umbrella factors used in the survey
AI professionals	
AI company/provider (n=2, 9.1%)	Type of institution/organization of AI professionals (eg, university, technology company, commercial organization)
AI role (n=1, 4.5%); perceived helpfulness (n=1, 4.5%)	The purpose to innovate with a specific AI application in medicine (eg, financial vs societal)
Health care professionals	
Knowledge and understanding of AI (n=6, 27.3%); education (n=3, 13.6%)	Knowledge of AI applications in medicine (eg, by means of training and education)
Expectation of AI (n=1, 4.5%); perceived actionability (ie, clear recommendation for action; n=1, 4.5%); user's social network (n=1, 4.5%); user's media consumption (n=1, 4.5%)	Attitude toward AI application usage in medicine (eg, agreeableness, openness, conscientiousness, engagement)
Patients informed about AI application usage in the hospital	
Knowledge/education about AI (n=5, 22.7%); awareness of AI (n=2, 9.1%)	General knowledge of AI applications in medicine
Openness (to AI health care technologies and to judgments of potential benefits and harms; n=1, 4.5%); perceived benefit and lower concern (n=1, 4.5%); user's social network (n=1, 4.5%); user's media consumption (n=1, 4.5%)	Attitude toward AI application usage in medicine (eg, agreeableness, openness, conscientiousness)
Gender (n=2, 9.1%); age (n=1, 4.5%); type of user (n=1, 4.5%)	Age, gender, level of education
All parties	
Clinicians and patients interaction during AI integration (n=1, 4.5%); human agency and oversight (n=1, 4.5%)	Transparency between all involved parties (AI professionals, health care professionals, patients)

[a]AI: artificial intelligence.

Table 2. Other factors related to trust (N=110) in medical AI[a] applications (22/32, 68.8%, studies).

Factor category and factors from the rapid review	Umbrella factors used in the survey
Technology-related factors	
Accuracy (n=7, 31.8%); reliability (n=5, 22.7%); safety (n=4, 18.2%); design and output quality (n=2, 9.1%); performance expectancy (n=2, 9.1%); ability (n=1, 4.5%); perceived functionality (n=1, 4.5%); self-efficacy (n=1, 4.5%); tool itself (n=1, 4.5%)	Performance of AI applications in medicine (reproducibility of outcomes, accuracy)
Auditability (n=1, 4.5%); customizability (n=1, 4.5%); understandability (n=1, 4.5%); ease of integration into clinical workflows (n=1, 4.5%); convenience of use (n=1, 4.5%); usability (n=1, 4.5%); (over)alerting and excessive false-positive rate (n=1, 4.5%)	Possibility of integration of AI applications into existing clinical workflows
Risk and impact mitigation (n=1, 4.5%)	Clear balance of risks and benefits of the AI application
Transparency (n=6, 27.3%); explainability (n=5, 22.7%); evidence strength (n=2, 9.1%); benevolence (n=2, 9.1%); complexity (n=2, 9.1%); interpretability (n=2, 9.1%); integrity (n=1, 4.5%); predictability (n=1, 4.5%); trialability (n=1, 4.5%); trustworthiness (n=1, 4.5%)	Explainability and transparency of the processes and outcomes
Legal and ethical factors	
Fairness and equity (n=8, 36.4%); adequate regulations, legislation, and governance (n=3, 13.6%); ethical/legal implications (n=1, 4.5%)	Adequacy of the regulations and governance of AI applications in medicine
Personal data privacy and security (n=8, 36.4%); data used to train AI/cognitive bias (n=2, 9.1%); data sensitivity (n=1, 4.5%); respect and preservation of human dignity (n=1, 4.5%)	Data use transparency
Accountability (n=3, 13.6%); power-control balance (n=1, 4.5%)	Clear accountability and responsibility of the AI application (machine vs human responsibility)
Additional factors	
Environmental sustainability (n=1, 4.5%)	Environment-friendly AI application
Replacement of doctor/lack of human touch and moral support when evaluated by AI alone (n=1, 4.5%); labor market implications (n=1, 4.5%)	Impact on job availability (machines replacing humans)

[a]AI: artificial intelligence.

Table 3. Human-related factors related to acceptance (N=77) of medical AI[a] applications (14/32, 43.8%, studies).

Factor category and factors from the rapid review	Umbrella factors used in the survey
AI professionals	
AI company/provider (n=1, 7.1%); brand impact (n=1, 7.1%)	Type of institution/organization of AI professionals (eg, university, technology company, commercial organization)
Perceived usefulness (n=3, 21.4%); better medical services/ understanding of disease (n=3, 21.4%); improve the quality of people's lives (n=2, 14.3%); medical costs (n=2, 14.3%); AI role (eg, saving patients' time; n=1, 7.1%); miniaturization of hardware (n=1, 7.1%)	Purpose to innovate with a specific AI application in medicine (eg, financial vs societal)
Health care professionals	
Knowledge and understanding of AI (n=1, 7.1%)	Knowledge of AI applications in medicine (eg, by means of training and education)
Behavioral intention to use (n=2, 14.3%); effort expectancy (n=2, 14.3%); perceived ease of use (n=2, 14.3%); perceived usefulness (n=2, 14.3%); intrinsic motivation (n=1, 7.1%); interest in AI (n=1, 7.1%); professional identity (n=1, 7.1%); concerns about benefit to patient care (n=1, 7.1%); general impression of AI (n=1, 7.1%)	Attitude toward AI application usage in medicine (eg, agreeableness, openness, conscientiousness, engagement)
Patients informed about AI application usage in the hospital	
Knowledge/education about AI (n=1, 7.1%); awareness of AI (n=1, 7.1%)	General knowledge of AI applications in medicine
Behavioral intention to use (n=2, 14.3%); general impression (n=1, 7.1%); Interest in topic (n=1, 7.1%)	Attitude toward AI application usage in medicine (eg, agreeableness, openness, conscientiousness)
Age (n=1, 7.1%)	Age
All parties	
Expectations of others (n=2, 14.3%)	Transparency between all involved parties (AI professionals, healthcarehealth care professionals, patients)

[a]AI: artificial intelligence.

Table 4. Other factors related to acceptance (N=77) of medical AI[a] applications (14/32, 43.8%, studies).

Factor category and factors from the rapid review	Umbrella factors used in the survey
Technology-related factors	
Performance expectancy (n=4, 28.6%); design and output quality (n=4, 28.6%); accuracy (n=2, 14.3%); efficiency (n=1, 7.1%)	Performance of AI applications in medicine (reproducibility of outcomes, accuracy)
Perceived ease of use (n=2, 14.3%); user-friendliness (n=2, 14.3%); actual system use (n=1, 7.1%); compatibility (n=1, 7.1%); facilitating conditions (n=1, 7.1%)	Possibility of integration of AI applications into existing clinical workflows
Perceived risk (n=1, 7.1%)	Clear balance of risks and benefits of the AI application
Transparency (n=3, 21.4%); explainability (n=2, 14.3%); evidence strength (n=1, 7.1%); trustworthiness (n=1, 7.1%)	Explainability and transparency of the processes and outcomes
Legal and ethical factors	
Adequate regulations, legislation and governance (n=2, 14.3%); ethical risks (n=1, 7.1%); political support (n=1, 7.1%)	Adequacy of the regulations and governance of AI applications in medicine
Data protection/security (n=2, 14.3%); patients' consent to the continuous collection and processing of data (n=1, 7.1%)	Data use transparency
Accountability and responsibility (n=2, 14.3%); tort liability (n=1, 7.1%)	Clear accountability and responsibility of the AI application (machine vs human responsibility)
Additional factors	
Replacement of doctor/lack of human touch and moral support when evaluated by AI alone (n=1, 7.1%)	Impact on job availability (machines replacing humans)
Trust in AI applications (n=3, 21.4%)	Acceptance emerging from trust

[a]AI: artificial intelligence.

Survey

Participants

A total of 22 respondents participated in the survey, of which 18 (82%) completed the questions on trust and 15 (68%) completed the questions on acceptance. No reasons were provided for not completing the survey. Table 5 shows the characteristics of the survey participants, the majority (n=21, 95%) of whom came from European countries, were aged from 40 to 60 years, and had 0-10 or 21-30 years of professional experience.

Participants were mainly slightly (n=7, 32%) or moderately (n=8, 36%) familiar with AI-based devices used for clinical purposes (Figure 2). In thinking about AI, 9 (41%) of the participants indicated that the statement "I accept AI" best represents their view, followed by "I approve of AI" (n=6, 27%) and "I embrace AI" (n=5, 23%); see Figure 3.

Table 5. Characteristics of the participants (N=22).

Characteristic and type of participant	Participants, n (%)
Stakeholder group[a]	
Researchers	18 (82)
Technology providers	5 (23)
Hospital staff	3 (14)
Policy makers	3 (14)
Gender	
Female	8 (36)
Male	13 (59)
Prefer not to say	1 (5)
Age (years)	
≤30	2 (9)
31-39	3 (14)
40-49	6 (27)
50-59	5 (23)
≥60	6 (27)
Country of work	
Netherlands	11 (50)
Germany	3 (14)
Ireland	2 (9)
Spain	2 (9)
France	1 (5)
Hungary	1 (5)
India	1 (5)
Italy	1 (5)
Years of professional experience	
0-10	6 (27)
11-20	4 (18)
21-30	7 (32)
31-40	5 (23)

[a]Participants sometimes represented more than 1 stakeholder group.

Figure 2. Familiarity with AI-based devices used for clinical purposes (N=22). AI: artificial intelligence.

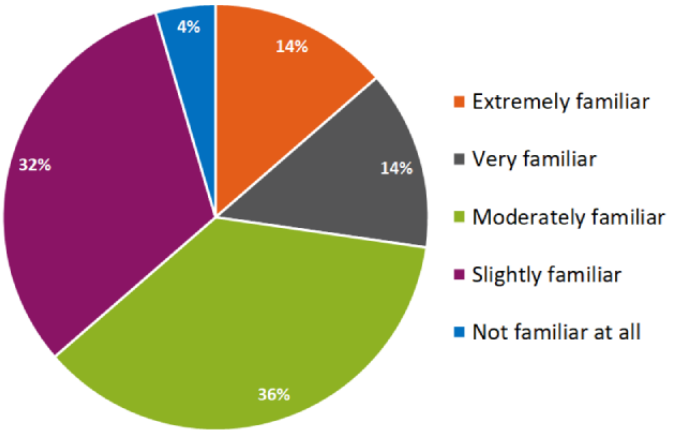
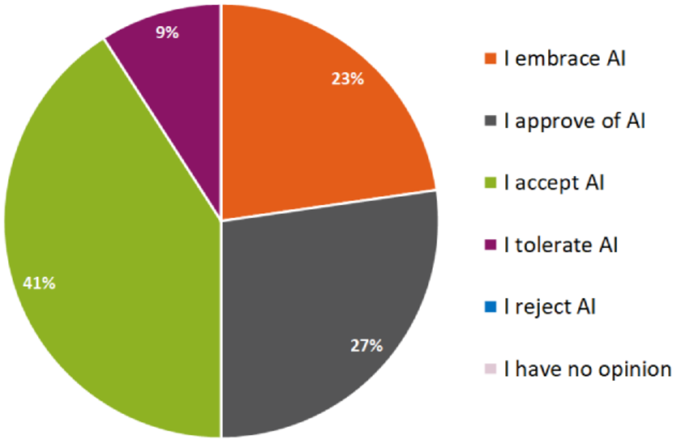


Figure 3. Statement best representing participants' view when thinking about AI (N=22). AI: artificial intelligence.



Relevance of Factors for Trust in and Acceptance of AI

In Table 6, the mean scores per factor for its relevance to trust and acceptance are shown. Figure 4 demonstrates a spider diagram with the 19 summarized statements and the corresponding mean scores of relevance to trust in and acceptance of AI applications in medicine. The degrees of relevance of the factors related to trust and to acceptance closely followed each other for all but 1 (5.3%) of the 19 factors. Only the type of AI organization was slightly more relevant to trust than to acceptance toward AI applications in medicine. In Figure 5, a scatter plot displays the combined relevance of the factors related to trust (x axis) and acceptance (y axis) toward medical AI applications. Of the 19 factors included in the survey, 3 (16%) were found to have, on average, low relevance, while the other 16 (84%) had high relevance. There were no factors relevant to acceptance and irrelevant to trust (upper-left section in the plot) and vice versa (bottom-right section in the plot).

Table 6. Mean (SD) factor relevance to trust and acceptance (N=22).

Factor	Trust	Acceptance
Type of AI[a] organization	4.72 (0.75)	4.27 (0.88)
Purpose to innovate with AI	4.33 (0.84)	4.47 (0.64)
Clinicians' knowledge about AI	4.50 (0.51)	4.73 (0.46)
Clinicians' attitude towards AI	4.50 (0.51)	4.47 (0.64)
Patients' knowledge of AI	4.17 (0.62)	4.20 (0.68)
Patients' attitude toward AI	4.28 (0.57)	4.47 (0.64)
Patients' age	3.17 (1.04)	3.47 (1.19)
Patients' gender	2.61 (1.14)	2.67 (1.05)
Patients' education level	3.50 (0.99)	3.53 (1.19)
Transparency between all parties	4.61 (0.50)	4.47 (0.64)
Performance of AI	4.83 (0.38)	4.67 (0.62)
Possibility of AI integration into existing workflows	4.56 (0.62)	4.53 (0.83)
Clear balance of AI risks and benefits	4.67 (0.49)	4.60 (0.63)
Explainability and transparency of AI processes	4.78 (0.43)	4.60 (0.63)
Adequacy of AI regulations	4.72 (0.57)	4.60 (0.83)
Data use transparency	4.61 (0.50)	4.67 (0.49)
Clear accountability and responsibility of AI	4.61 (0.61)	4.80 (0.41)
Environmental friendliness of AI	3.83 (0.79)	3.87 (0.92)
Impact on job availability	3.78 (1.11)	4.07 (0.88)

[a]AI: artificial intelligence.

Figure 4. Mean scores of factors' relevance to trust in and acceptance of AI applications in medicine (N=19). Score=1 means irrelevant; score=3 means not irrelevant, nor relevant; and score=5 means relevant. AI: artificial intelligence.

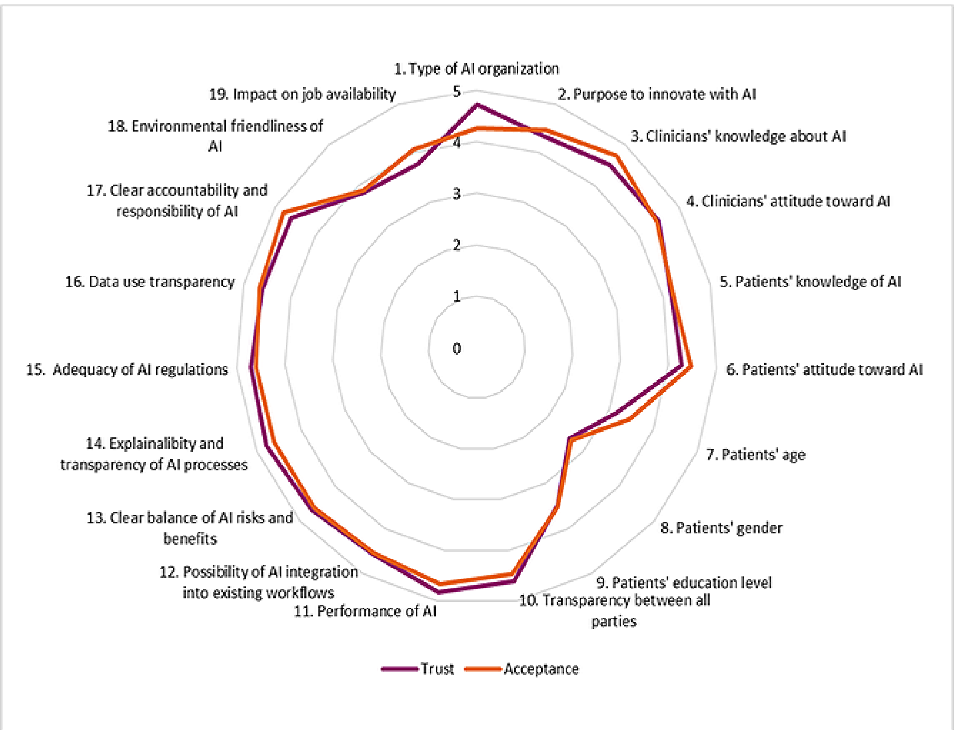


Figure 5. Overview of the relevance of factors related to trust in and acceptance of novel AI applications in medicine (1=not relevant, 5=relevant). AI: artificial intelligence.

Factors of Low Relevance

With regard to patients informed about AI application usage in the hospital, participants deemed the patient's gender, age, and educational level to be of low relevance to trust in and acceptance of novel AI applications in medicine.

Factors of High Relevance

The majority of factors were deemed highly relevant to trust in and acceptance of novel AI applications in medicine by participants. Regarding AI professionals, it was observed that the type of institution or organization where AI professionals originated from (eg, university, technology company, commercial organization) and the purpose to innovate with a specific AI application in medicine (eg, financial, societal, or clinical purpose) were considered relevant. Participants reported that the involvement of health care professionals having knowledge of the AI application (eg, by means of training and education) is highly relevant to trust and acceptance. The health care professionals' attitude toward AI application usage in medicine, comprising their agreeableness, openness, conscientiousness, and engagement, was found to be equally important. Likewise, the patients' general knowledge of and attitude toward AI application usage in medicine were found to be relevant. The transparency between all involved parties (AI professionals, health care professionals, and patients) was also deemed highly relevant. Technology-related factors were found to be highly relevant, too, in particular the performance of AI applications in medicine (eg, reproducibility and accuracy of outcomes), the possibility of integration of the AI applications into existing clinical workflows, having a clear balance of risks and benefits of the AI applications, and the explainability and transparency of the processes and outcomes. Legal and ethical factors were also considered of high relevance and concerned the adequacy of the regulations and governance of AI applications in medicine, data use transparency, and clear accountability and responsibility of the AI applications (machine vs human responsibility). Additional factors, such as AI applications being environment friendly and the impact of medical AI on job availability (eg, machines replacing human beings), were viewed as factors of high relevance.

Other Factors

Participants were able to share other factors that were not mentioned in the survey questions. Factors related to trust included solidarity and understanding the bias and interdomain knowledge of AI in software development, data science, and medicine. Other factors related to acceptance were the extent to which alternatives to AI applications are available, the length of experience, transparency about limitations, reproducibility, risks evaluation, resources, and the fear to use an AI application (ie, fear of making the wrong decision or fear of losing control).

Discussion

Principal Findings

This study aimed to identify factors related to trust and acceptance toward medical AI applications by means of a rapid review and to assess their relevance by conducting a survey. Through the rapid review, 19 key factors related to trust in and acceptance of AI-powered medical technologies were identified and subsequently grouped into 4 categories. Our survey results highlight that of all examined factors, 84% (16/19) were considered highly relevant to trust in and acceptance of novel AI applications in medicine. Only the patient's gender, age, and education level (3/19, 16%) were deemed to be of low relevance by participants.

Comparison With Prior Work

Previous studies have reported that trust in technology is mainly determined by human characteristics [54], technology-related factors [55], and environment-related factors [56], which is in line with the findings of our survey. According to Tran et al [57], who investigated patients' perceived benefits and risks of using digital and AI technology in health care, the important factors to consider are the new technologies requiring an overhaul of the current health care system as human care is being replaced by machines and health care professionals becoming sufficiently equipped with increasing knowledge of AI technology. This highlights the importance of several survey factors, including the possibility of AI integration into existing clinical workflows. Therefore, setting features such as understandability, usability, and user-friendliness (factors that frequently appeared in the rapid review) by AI professionals as key goals in the development of novel AI applications would increase the chances of successful integration of AI technology into health care systems. Tran et al [57] also highlighted the increasing importance of data use transparency toward patients and the acute need for clear accountability and responsibility (machine vs human responsibility) concerning the new technology, which also goes hand in hand with the findings from the rapid review and the survey [57]. The patient data handling must be organized in accordance with the existing data protection regulations in respective countries, with additional precautionary measures due to the sensitive nature of such medical data [57]. Shin et al [58] demonstrated that explainability of AI plays a big role in user trust and attitude toward AI. Explainability, along with transparency, was also found to be highly relevant in our study, especially in relation to the AI application processes and outcomes. In addition, Vourgidis et al [59] recommended that AI systems be regularly checked for being up to date, since today's technology is continuously evolving. This again highlights the relevance of the education of health care professionals, since they are the primary users of medical AI technology and hence need to follow the developments in the field. Yang et al [49] found that gender is not relevant to trust in AI technology in medicine. This agrees with our finding that a patient's gender has low relevance to trust in and acceptance of AI technology in medicine. Contrary to our findings, it has been reported that younger generations in general have more trust and are more likely to accept AI systems compared to older generations [11]. In our survey, the majority of participants were aged 40-60 years and above and they exhibited a solid awareness of and a positive attitude toward AI technology. Gillespie et al [11] also stated that highly educated people (university level) are more likely to trust and accept AI systems compared to those without a university degree. However, our survey showed that a patient's educational level has low relevance to trust in and acceptance of medical AI applications.

Strengths and Limitations

To the best of our knowledge, this is the first study to use a rapid review of the latest literature to identify factors related to trust in and acceptance of AI applications in medicine in order to create a survey to evaluate their relevance and the attitudes of health care stakeholders toward implementation of medical AI applications. However, the study has several limitations. Since a large number of papers and reports in the rapid review did not provide sufficient context for the factors for trust or acceptance, there could have been an increased risk of personal bias during interpretation and categorization of those factors. Furthermore, some studies did not clarify whether the reported factors were related to only trust or only acceptance, which could also lead to possible misinterpretation. To minimize the effect of such bias and misinterpretation, a third reviewer (author HJMV) was consulted in such cases. Another limitation is the relatively small number of papers included in the rapid review, given the breadth of the topic. However, this rapid review was intentionally conducted focusing on the most relevant and recent literature to provide an initial overview and highlight key themes in a time-efficient manner. We aimed to provide a starting point that formed the basis for the survey. In addition, the number of participants included in the survey can be considered relatively low, which was caused by difficulties in recruiting participants and the time-constrained nature of the study. However, sufficient diversity in participant characteristics (ie, gender, age, country of work, and years of professional experience) was achieved, which could be considered more important in terms of validity of the study findings. Even though the survey benefited from a sample with a wide diversity in participant characteristics, one of the limitations to consider is the underrepresentation of certain stakeholder groups, in particular technology providers, policy makers, and hospital staff members other than clinicians. If these groups had been included in the survey, different patterns in factor relevance might have been observed, potentially shedding light on additional concerns or challenges associated with AI applications in medicine. Moreover, when considering the relevance of factors assessed through the survey, which were predominantly highlighted by researchers, it is important to note that these factors might be readily attainable or already well established within this specific stakeholder group. As a result, these factors may not necessarily represent challenges or barriers for this particular group, as they are already well versed in the aspects related to trust and acceptance.

Recommendations for Future Research

The results of this study can be valuable for various stakeholders involved in the implementation of novel AI applications, since trust and acceptance building remains a focus point throughout the different stages, including the pilot, implementation, evaluation, and monitoring phases of the process. In the survey, participants shared other factors related to trust in and acceptance of AI applications in medicine that were not included in the survey. However, due to a lack of context, it is not entirely clear what was meant by some of these factors; since these are open to interpretation, follow-up research is required to better understand this. In addition, further research is needed to gain insight into the reasons participants considered factors to be of low or high relevance. Regarding the currently underrepresented stakeholder groups in the survey, more research is required to gain insight into the perspectives of policy regulators, technology providers, and hospital staff members. Next, once the implementation of a novel AI technology, such as the AIDPATH system, becomes clear from the trust and acceptance point of view, it would be beneficial to conduct a workshop with experts from the AI and biotechnology fields to identify technical challenges of implementation. This is crucial since, according to the survey results, the technical robustness and clarity of AI applications is a prerequisite for trust and acceptance exhibited toward this technology by stakeholders.

Recommendations for Implementation

By considering the factors that are most relevant in the AI technology adoption process, the implementers can facilitate trust in and acceptance of medical AI applications among their users and other stakeholders. Furthermore, the knowledge of the factors with high relevance to stakeholders can predict concerns the potential users might have regarding the new AI technology and act upon these concerns to implement the AI application efficiently and in a timely manner. There are several ways in how the results of the survey could be used by AI implementers, such as smart hospitals, to build trust and acceptance among various stakeholder groups. For instance, the highly relevant factor of knowledge and understanding of AI among health care professionals could be addressed by providing information about medical AI to clinicians in the form of conferences and educational workshops. These initiatives can ensure that health care professionals remain updated on significant changes in AI technology, facilitating its accurate utilization. Similarly, patients could be informed of medical AI technology through patient information initiatives in (smart) hospitals and within patient communities. The highly relevant technology-related factors could be used by technology developers and scientific researchers as guidance in the development of novel AI technology. For regulators and policy makers, it is crucial to know that users and other stakeholders consider data use transparency and fairness and equity to be of utmost importance regarding novel medical AI technology. Indeed, data privacy is a crucial and ever-so-present topic in legislation and regulations, but it needs to be constantly reviewed by policy makers due to the newness of AI in health care and the speed of its development. The legal aspects of software containing AI have been subjected to the Medical Device Regulation (MDR) [60]. For the acceptance of AI, its implementation in MDR-compliant solutions is invaluable. The tasks of policy makers could involve the risk assessment of various data breaches related to AI in medicine with continuous updating of regulations related to data security and privacy within the field of medical AI. Furthermore, both policy makers and AI professionals have to ensure the maintenance of fairness and equity of AI technology usage.

Conclusion

This study identified and assessed the relevance of factors for trust in and acceptance of AI applications in medicine. The survey demonstrated that the majority of the identified human-related, technology-related, and legal and ethical factors for trust in and acceptance of novel AI applications in medicine were considered by stakeholders to be of high relevance. Taken together, these findings and subsequent recommendations could be used by any implementers of medical AI, such as (smart) hospitals, AI technology organizations, biotechnology research institutes, and policy makers, to facilitate smooth and timely adoption of novel AI applications in medicine.

Acknowledgments

Our gratitude goes out to the AIDPATH (artificial intelligence [AI]-driven Decentralized Production for Advanced Therapies in the Hospital) Consortium partners, the Fraunhofer Institute for Cell Therapy and Immunology IZI, University College London, the Foundation for Research and Technology (FORTH)-Hellas, SZTAKI, Aglaris Cell SL, Sartorius Cell Genix GmbH, the Fundació Clínic per a la Recerca Biomèdica, IRIS Technology Solutions, Red Alert Labs, and ORTEC b.v. We would also like to thank the advisory board members of AIDPATH for providing feedback on the survey. Lastly, we thank the survey's responders.

The paper was written within the framework of the European Union's (EU) Horizon 2020 research and innovation program project AIDPATH (grant agreement number 101016909). All mentioned colleagues/companies are part of AIDPATH and therefore received funding from the EU within the scope of AIDPATH.

Data Availability

The data used and analyzed in this study are available from the corresponding author upon reasonable request.

Authors' Contributions

DS was responsible for methods design, search strategy design, data acquisition, data extraction, data analysis, interpretation of data, and design and writing of the manuscript; AA, methods design, search strategy design, data acquisition, data extraction, data analysis, interpretation of data, regularly reviewing the work, design and writing of the manuscript, and providing feedback on the manuscript; IWAB, methods design, data acquisition, and providing feedback on the manuscript; CS, methods design, regularly reviewing the work, and providing feedback on the manuscript; MH, providing feedback on the manuscript and funding acquisition; JJLJ, reviewing the work and providing feedback on the manuscript; SH, contributing to the Introduction chapter, providing feedback on the manuscript, project administration, and funding acquisition; and HJMV, concept and design of the overall study, quality assessment, interpretation of data, regularly reviewing the work, providing feedback on the manuscript, and manuscript final approval.

Conflicts of Interest

MH reports speaker honoraria from Novartis, Janssen, and Celgene/BMS and has participated in scientific advisory boards for Janssen and Celgene/BMS. MH is also listed as an inventor on patent applications and has been granted patents related to chimeric antigen receptor (CAR) technologies and CAR T cell therapy that have been filed by the Fred Hutchinson Cancer Research Center and the University of Wurzburg and that have been, in part, licensed to industry. In addition, MH is a cofounder and equity owner of T-CURX GmbH, Wurzburg, Germany. The remaining authors declare that they have no competing interests.

Multimedia Appendix 1
Identified factors and corresponding studies included in the rapid review.
[DOC File , 32 KB - humanfactors_v11i1e47031_app1.doc]

References

1. Artificial intelligence in medicine regulation. European Medicines Agency. URL: https://www.ema.europa.eu/en/news/artificial-intelligence-medicine-regulation [accessed 2022-11-25]
2. Lekadir K, Quaglio G, Tselioudis Garmendia A, Gallin C. Artificial intelligence in healthcare: applications, risks, and ethical and societal impacts. European Parliament. URL: https://www.europarl.europa.eu/stoa/en/document/EPRS_STU(2022)729512 [accessed 2023-12-04]
3. Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. Artif Intell Healthc 2020 Jun:26-60. [doi: 10.1016/b978-0-12-818438-7.00002-2]
4. Callahan G. What are the advantages of artificial intelligence? rev. URL: https://www.rev.com/blog/speech-to-text-technology/what-are-the-advantages-of-artificial-intelligence [accessed 2023-12-04]
5. Koski E, Murphy J. AI in Healthcare. Stud Health Technol Inform 2021 Dec 15:284-299. [doi: 10.3233/shti210726]
6. Murphy K, Di Ruggiero E, Upshur R, Willison DJ, Malhotra N, Cai JC, et al. Artificial intelligence for good health: a scoping review of the ethics literature. BMC Med Ethics 2021 Feb 15;22(1):14 [FREE Full text] [doi: 10.1186/s12910-021-00577-8] [Medline: 33588803]
7. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med 2022 Jan 20;28(1):31-38. [doi: 10.1038/s41591-021-01614-0] [Medline: 35058619]
8. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019 Oct 25;366(6464):447-453. [doi: 10.1126/science.aax2342] [Medline: 31649194]
9. Fujimori R, Liu K, Soeno S, Naraba H, Ogura K, Hara K, et al. Acceptance, barriers, and facilitators to implementing artificial intelligence-based decision support systems in emergency departments: quantitative and qualitative evaluation. JMIR Form Res 2022 Jun 13;6(6):e36501 [FREE Full text] [doi: 10.2196/36501] [Medline: 35699995]
10. Schwartz JM, George M, Rossetti SC, Dykes PC, Minshall SR, Lucas E, et al. Factors influencing clinician trust in predictive clinical decision support systems for in-hospital deterioration: qualitative descriptive study. JMIR Hum Factors 2022 May 12;9(2):e33960 [FREE Full text] [doi: 10.2196/33960] [Medline: 35550304]
11. Gillespie N, Lockey S, Curtis C. Trust in artificial intelligence: a five country study. KPMG. 2021. URL: https://assets.kpmg.com/content/dam/kpmg/au/pdf/2021/trust-in-ai-multiple-countries.pdf [accessed 2023-12-04]
12. Ehsan Toreini MA, Coopamootoo K, Elliott K, Gonzalez Zelaya C, van Moorsel A. The relationship between trust in AI and trustworthy machine learning technologies. 2020 Presented at: ACM FAT* 2020: 2020 Conference on Fairness, Accountability, and Transparency; January 27-30, 2020; Barcelona p. 272-283. [doi: 10.1145/3351095.3372834]
13. Ferrario A, Loi M, Viganò E. In AI we trust incrementally: a multi-layer model of trust to analyze human-artificial intelligence interactions. Philos Technol 2019 Oct 23;33(3):523-539. [doi: 10.1007/s13347-019-00378-3]
14. Gursoy D, Chi O, Lu L, Nunkoo R. Consumers acceptance of artificially intelligent (AI) device use in service delivery. Int J Inf Manag 2019 Dec;49:157-169 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2019.03.008]
15. Cornelissen L, Egher C, van Beek V, Williamson L, Hommes D. The drivers of acceptance of artificial intelligence-powered care pathways among medical professionals: web-based survey study. JMIR Form Res 2022 Jun 21;6(6):e33368 [FREE Full text] [doi: 10.2196/33368] [Medline: 35727614]
16. Joshi M, Mecklai K, Rozenblum R, Samal L. Implementation approaches and barriers for rule-based and machine learning-based sepsis risk prediction tools: a qualitative study. JAMIA open. JAMIA Open 2022 Jul;5(2):ooac022. [doi: 10.1093/jamiaopen/ooac022]
17. Rainie L, Anderson J. The fate of online trust in the next decade: theme 3: trust will not grow, but technology usage will continue to rise as a 'new normal' sets in. Pew Research Center. 2023 Jun 1. URL: https://www.pewresearch.org/internet/2017/08/10/theme-3-trust-will-not-grow-but-technology-usage-will-continue-to-rise-as-a-new-normal-sets-in/ [accessed 2023-12-04]

18. Kerasidou CX, Kerasidou A, Buscher M, Wilkinson S. Before and beyond trust: reliance in medical AI. J Med Ethics 2022 Nov 23;48(11):852-856 [FREE Full text] [doi: 10.1136/medethics-2020-107095] [Medline: 34426519]
19. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. J Med Internet Res 2020 Jun 19;22(6):e15154 [FREE Full text] [doi: 10.2196/15154] [Medline: 32558657]
20. Eltawil FA, Atalla M, Boulos E, Amirabadi A, Tyrrell PN. Analyzing barriers and enablers for the acceptance of artificial intelligence innovations into radiology practice: a scoping review. Tomography 2023 Jul 28;9(4):1443-1455 [FREE Full text] [doi: 10.3390/tomography9040115] [Medline: 37624108]
21. Matheny M, Israeli S, Ahmed M, Whicher D. Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril. Washington, DC: The National Academies Press; 2019.
22. AIDPATH. Sciencrew. URL: https://www.sciencrew.com/c/6499?title=AIDPATH [accessed 2023-12-04]
23. Artificial intelligence-driven, decentralized production for advanced therapies in the hospital. European Commission. URL: https://cordis.europa.eu/project/id/101016909 [accessed 2023-01-17]
24. Ran T, Eichmüller SB, Schmidt P, Schlander M. Cost of decentralized CAR T-cell production in an academic nonprofit setting. Int J Cancer 2020 Dec 15;147(12):3438-3445 [FREE Full text] [doi: 10.1002/ijc.33156] [Medline: 32535920]
25. Hoogendoorn K, Main H. What are ATMPs? Claryifying misconceptions. ATMP Sweden. 2020 Jun 12. URL: https://atmpsweden.se/wp-content/uploads/2020/06/What-are-ATMPs-170620.pdf [accessed 2023-12-04]
26. Hort S, Herbst L, Bäckel N, Erkens F, Niessing B, Frye M, et al. Toward rapid, widely available autologous CAR-T cell therapy – artificial intelligence and automation enabling the smart manufacturing hospital. Front Med (Lausanne) 2022 Jun 6;9:913287 [FREE Full text] [doi: 10.3389/fmed.2022.913287] [Medline: 35733863]
27. Arevalo-Rodriguez I, Tricco AC, Steingart KR, Nussbaumer-Streit B, Kaunelis D, Alonso-Coello P, et al. Challenges of rapid reviews for diagnostic test accuracy questions: a protocol for an international survey and expert consultation. Diagn Progn Res 2019 Apr 4;3(1):7 [FREE Full text] [doi: 10.1186/s41512-019-0052-y] [Medline: 31093577]
28. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
29. Polland RJ. Essentials of survey research and analysis. Psycholosphere. 1998. URL: https://www.psycholosphere.com/Essentials%20of%20Survey%20Research%20and%20Analysis%20by%20Ronald%20Polland.pdf [accessed 2023-12-04]
30. 30. Act on medical-scientific research with humans. Central Committee on Research Involving Human Subjects. URL: https://wetten.overheid.nl/BWBR0009408/2022-07-01 [accessed 2023-09-11]
31. Alaqra AS, Kane B, Fischer-Hübner S. Machine learning-based analysis of encrypted medical data in the cloud: qualitative study of expert stakeholders' perspectives. JMIR Hum Factors 2021 Sep 16;8(3):e21810 [FREE Full text] [doi: 10.2196/21810] [Medline: 34528892]
32. Allen D. Trust and artificial intelligence in healthcare. MedicalExpo e-Magazine. URL: https://emag.medicalexpo.com/trust-and-artificial-intelligence-in-healthcare/ [accessed 2023-12-04]
33. Antes AL, Burrous S, Sisk BA, Schuelke MJ, Keune JD, DuBois JM. Exploring perceptions of healthcare technologies enabled by artificial intelligence: an online, scenario-based survey. BMC Med Inform Decis Mak 2021 Jul 20;21(1):221 [FREE Full text] [doi: 10.1186/s12911-021-01586-8] [Medline: 34284756]
34. Brady D. Building public trust to successfully deploy AI in EU healthcare. European Policy Centre. 2022 Jan 26. URL: https://www.epc.eu/content/PDF/2022/CHES_PB_v2-2.pdf [accessed 2023-12-04]
35. Diprose WK, Buist N, Hua N, Thurier Q, Shand G, Robinson R. Physician understanding, explainability, and trust in a hypothetical machine learning risk calculator. J Am Med Inform Assoc 2020 Apr 01;27(4):592-600 [FREE Full text] [doi: 10.1093/jamia/ocz229] [Medline: 32106285]
36. Gille F, Jobin A, Ienca M. What we talk about when we talk about trust: theory of trust for AI in healthcare. Intell-Based Med 2020 Nov;1-2:100001-100002 [FREE Full text] [doi: 10.1016/j.ibmed.2020.100001]
37. Haggenmüller S, Krieghoff-Henning E, Jutzi T, Trapp N, Kiehl L, Utikal JS, et al. Digital natives' preferences on mobile artificial intelligence apps for skin cancer diagnostics: survey study. JMIR Mhealth Uhealth 2021 Aug 27;9(8):e22909 [FREE Full text] [doi: 10.2196/22909] [Medline: 34448722]
38. Hui CY, McKinstry B, Fulton O, Buchner M, Pinnock H. Patients' and clinicians' perceived trust in internet-of-things systems to support asthma self-management: qualitative interview study. JMIR Mhealth Uhealth 2021 Jul 16;9(7):e24127 [FREE Full text] [doi: 10.2196/24127] [Medline: 34269684]
39. Jauk S, Kramer D, Avian A, Berghold A, Leodolter W, Schulz S. Correction to: technology acceptance of a machine learning algorithm predicting delirium in a clinical setting: a mixed-methods study. J Med Syst 2021 Mar 10;45(4):52 [FREE Full text] [doi: 10.1007/s10916-021-01728-5] [Medline: 33740133]
40. Jermutus E, Kneale D, Thomas J, Michie S. Influences on user trust in healthcare artificial intelligence: a systematic review. Wellcome Open Res 2022 Feb 18;7:65. [doi: 10.12688/wellcomeopenres.17550.1]
41. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mach Intell 2019 Sep 02;1(9):389-399. [doi: 10.1038/s42256-019-0088-2]
42. Kovalchuk S, Kopanitsa G, Derevitskii I, Matveev G, Savitskaya D. Three-stage intelligent support of clinical decision making for higher trust, validity, and explainability. J Biomed Inform 2022 Mar;127:104013 [FREE Full text] [doi: 10.1016/j.jbi.2022.104013] [Medline: 35158071]

43. Shah P, Mishra D, Shanmugam M, Vighnesh M, Jayaraj H. Acceptability of artificial intelligence-based retina screening in general population. Indian J Ophthalmol 2022;70(4):1140. [doi: 10.4103/ijo.ijo_1840_21]
44. Tamori H, Yamashina H, Mukai M, Morii Y, Suzuki T, Ogasawara K. Acceptance of the use of artificial intelligence in medicine among Japan's doctors and the public: a questionnaire survey. JMIR Hum Factors 2022 Mar 16;9(1):e24680. [doi: 10.2196/24680]
45. Tucci V, Saary J, Doyle TE. Factors influencing trust in medical artificial intelligence for healthcare professionals: a narrative review. J Med Artif Intell 2022 Mar;5:4-4. [doi: 10.21037/jmai-21-25]
46. van Bussel MJP, Odekerken-Schröder GJ, Ou C, Swart RR, Jacobs MJG. Analyzing the determinants to accept a virtual assistant and use cases among cancer patients: a mixed methods study. BMC Health Serv Res 2022 Jul 09;22(1):890 [FREE Full text] [doi: 10.1186/s12913-022-08189-7] [Medline: 35804356]
47. Walter S, Gruss S, Frisch S, Liter J, Jerg-Bretzke L, Zujalovic B, et al. "What about automated pain recognition for routine clinical use?" A survey of physicians and nursing staff on expectations, requirements, and acceptance. Front Med (Lausanne) 2020 Dec 21;7:566278 [FREE Full text] [doi: 10.3389/fmed.2020.566278] [Medline: 33409286]
48. Williams J, Sim A, Lee B, Colon A, Lee R, Schick A, et al. Trustworthy AI (TAI) playbook. U.S. DEPARTMENT OF HEALTH & HUMAN SERVICES. 2021 Sep. URL: https://www.hhs.gov/sites/default/files/hhs-trustworthy-ai-playbook.pdf [accessed 2023-12-04]
49. Yang K, Zeng Z, Peng H, Jiang Y. Attitudes of Chinese cancer patients toward the clinical use of artificial intelligence. Patient Prefer Adher 2019 Nov;13:1867-1875. [doi: 10.2147/ppa.s225952]
50. Yap A, Wilkinson B, Chen E. Patients perceptions of artificial intelligence in diabetic eye screening. Asia-Pac J Ophthalmol (Philadelphia, Pa) 2022 May 1;11(3):287-293. [doi: 10.1097/apo.0000000000000525]
51. Zhang Z, Citardi D, Wang D, Genc Y, Shan J, Fan X. Patients' perceptions of using artificial intelligence (AI)-based technology to comprehend radiology imaging data. Health Inform J 2021 Apr 29;27(2):14604582211011215 [FREE Full text] [doi: 10.1177/14604582211011215] [Medline: 33913359]
52. Zhang Z, Genc Y, Wang D, Ahsen ME, Fan X. Effect of AI explanations on human perceptions of patient-facing AI-powered healthcare systems. J Med Syst 2021 May 04;45(6):64. [doi: 10.1007/s10916-021-01743-6] [Medline: 33948743]
53. Zheng B, Wu M, Zhu S, Zhou H, Hao X, Fei F, et al. Attitudes of medical workers in China toward artificial intelligence in ophthalmology: a comparative survey. BMC Health Serv Res 2021 Oct 09;21(1):1067 [FREE Full text] [doi: 10.1186/s12913-021-07044-5] [Medline: 34627239]
54. Hengstler M, Enkel E, Duelli S. Applied artificial intelligence and trust—the case of autonomous vehicles and medical assistance devices. Technol Forecast Soc Change 2016 Apr;105:105-120 [FREE Full text] [doi: 10.1016/j.techfore.2015.12.014]
55. Schaefer KE, Chen JYC, Szalma JL, Hancock PA. A meta-analysis of factors influencing the development of trust in automation: implications for understanding autonomy in future systems. Hum Factors 2016 May 22;58(3):377-400. [doi: 10.1177/0018720816634228] [Medline: 27005902]
56. Oleson K, Billings D, Kocsis V, Chen J, Hancock P. Antecedents of trust in human-robot collaborations. 2011 Presented at: 2011 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA); February 22-24, 2011; Miami Beach, FL p. 175-178. [doi: 10.1109/cogsima.2011.5753439]
57. Tran V, Riveros C, Ravaud P. Patients' views of wearable devices and AI in healthcare: findings from the ComPaRe e-cohort. NPJ Digit Med 2019 Jun 14;2(1):53 [FREE Full text] [doi: 10.1038/s41746-019-0132-y] [Medline: 31304399]
58. Shin D. The effects of explainability and causability on perception, trust, and acceptance: implications for explainable AI. Int J Hum-Comput Stud 2021 Feb;146:102551 [FREE Full text] [doi: 10.1016/j.ijhcs.2020.102551]
59. Vourgidis I, Mafuma S, Wilson P, Carter J, Cosma G. Medical Expert Systems – A Study of Trust and Acceptance by Healthcare Stakeholders. Berlin: Springer International Publishing; 2019:108-119.
60. EUR-Lex. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC. European Parliament. 2017. URL: https://eur-lex.europa.eu/eli/reg/2017/745/oj [accessed 2023-12-04]

Abbreviations

AI: artificial intelligence
AIDPATH: AI-powered Decentralized Production for Advanced Therapies in the Hospital
CAR-T: chimeric antigen receptor T cell
MDR: Medical Device Regulation

Edited by A Kushniruk, E Borycki; submitted 06.03.23; peer-reviewed by D Kohen, Y Zhuang, J Zeng; comments to author 17.06.23; revised version received 25.09.23; accepted 20.11.23; published 17.01.24.

Please cite as:
Shevtsova D, Ahmed A, Boot IWA, Sanges C, Hudecek M, Jacobs JJL, Hort S, Vrijhoef HJM
Trust in and Acceptance of Artificial Intelligence Applications in Medicine: Mixed Methods Study
JMIR Hum Factors 2024;11:e47031
URL: https://humanfactors.jmir.org/2024/1/e47031
doi:10.2196/47031
PMID:38231544

Original Paper

A Machine Learning Approach with Human-AI Collaboration for Automated Classification of Patient Safety Event Reports: Algorithm Development and Validation Study

Hongbo Chen[1], MSc; Eldan Cohen[1], PhD; Dulaney Wilson[2], PhD; Myrtede Alfred[1], PhD

[1]Department of Mechanical & Industrial Engineering, Faculty of Applied Science & Engineering, University of Toronto, Toronto, ON, Canada

[2]Department of Public Health Sciences, College of Medicine, Medical University of South Carolina, Charleston, SC, United States

Corresponding Author:
Myrtede Alfred, PhD
Department of Mechanical & Industrial Engineering
Faculty of Applied Science & Engineering
University of Toronto
27 King's College Cir
Toronto, ON, M5S 1A1
Canada
Phone: 1 4372154739
Email: myrtede.alfred@utoronto.ca

Abstract

Background: Adverse events refer to incidents with potential or actual harm to patients in hospitals. These events are typically documented through patient safety event (PSE) reports, which consist of detailed narratives providing contextual information on the occurrences. Accurate classification of PSE reports is crucial for patient safety monitoring. However, this process faces challenges due to inconsistencies in classifications and the sheer volume of reports. Recent advancements in text representation, particularly contextual text representation derived from transformer-based language models, offer a promising solution for more precise PSE report classification. Integrating the machine learning (ML) classifier necessitates a balance between human expertise and artificial intelligence (AI). Central to this integration is the concept of explainability, which is crucial for building trust and ensuring effective human-AI collaboration.

Objective: This study aims to investigate the efficacy of ML classifiers trained using contextual text representation in automatically classifying PSE reports. Furthermore, the study presents an interface that integrates the ML classifier with the explainability technique to facilitate human-AI collaboration for PSE report classification.

Methods: This study used a data set of 861 PSE reports from a large academic hospital’s maternity units in the Southeastern United States. Various ML classifiers were trained with both static and contextual text representations of PSE reports. The trained ML classifiers were evaluated with multiclass classification metrics and the confusion matrix. The local interpretable model-agnostic explanations (LIME) technique was used to provide the rationale for the ML classifier’s predictions. An interface that integrates the ML classifier with the LIME technique was designed for incident reporting systems.

Results: The top-performing classifier using contextual representation was able to obtain an accuracy of 75.4% (95/126) compared to an accuracy of 66.7% (84/126) by the top-performing classifier trained using static text representation. A PSE reporting interface has been designed to facilitate human-AI collaboration in PSE report classification. In this design, the ML classifier recommends the top 2 most probable event types, along with the explanations for the prediction, enabling PSE reporters and patient safety analysts to choose the most suitable one. The LIME technique showed that the classifier occasionally relies on arbitrary words for classification, emphasizing the necessity of human oversight.

Conclusions: This study demonstrates that training ML classifiers with contextual text representations can significantly enhance the accuracy of PSE report classification. The interface designed in this study lays the foundation for human-AI collaboration in the classification of PSE reports. The insights gained from this research enhance the decision-making process in PSE report classification, enabling hospitals to more efficiently identify potential risks and hazards and enabling patient safety analysts to take timely actions to prevent patient harm.

(JMIR Hum Factors 2024;11:e53378) doi:10.2196/53378

KEYWORDS

accident; accidents; black box; classification; classifier; collaboration; design; document; documentation; documents; explainability; explainable; human-AI collaboration; human-AI; human-computer; human-machine; incident reporting; interface design; interface; interpretable; LIME; machine learning; patient safety; predict; prediction; predictions; predictive; report; reporting; safety; text; texts; textual; artificial intelligence

Introduction

Since the publication of the seminal report on patient safety—*To Err Is Human* [1], the importance of preventing adverse events in health care has been widely recognized. Adverse events refer to unintended or unexpected incidents that occur during hospital care that cause harm to a patient [2]. Common adverse events include complications, falls, and medication errors. These events can lead to prolonged hospital stays, permanent harm to patients, life-saving interventions, or even contributing to patient deaths [2,3]. Unfortunately, adverse events remain one of the top 10 leading causes of death and disability worldwide, resulting in 251,454 deaths annually in the United States alone [4]. In Organization for Economic Cooperation and Development (OECD) countries, 15% of total hospital activity is the direct result of adverse events [5]. The global cost of adverse events has been estimated at 42 billion USD annually [6].

Patient safety event (PSE) reporting systems, also called incident reporting systems, have been widely adopted in hospitals across the world as part of their efforts to mitigate adverse events and improve patient safety [7,8]. Multiple nations, including Canada, Japan, England, and Norway, have made it mandatory for hospitals to establish and maintain a PSE reporting system, either with individual health care systems or through centralized national incident reporting platforms [9]. The primary purpose of the PSE reporting system is to provide health care organizations with a centralized system for tracking and analyzing PSEs, thereby facilitating continuous learning and maintaining a record of PSEs for risk assessment and prevention [7,10]. PSE reporting systems are tools that allow frontline health care personnel to voluntarily report adverse events, near-misses, and unsafe conditions [11]. Each PSE report includes structured data, such as event types, patient harm level, date, and location of the event, as well as unstructured data, including a free-text section that contains the factual description of the event and the patient's outcome [12]. Following submission, PSE reports are reviewed by relevant hospital staff, such as risk managers, patient safety analysts, nurse managers, physicians, and biomedical engineers, to identify areas for patient safety and quality improvement within the hospital [13].

Accurately classifying PSE reports into their appropriate event type is crucial to ensure that these reports are directed to the relevant patient safety analyst, support organizational learning, identify patterns and trends in adverse events, and ultimately prioritize measures to reduce adverse events [14,15]. An event type refers to a specific class of events that share common characteristics [16]. Examples of event types include falls, medication-related issues, and diagnosis errors [17,18]. PSE reporting systems may have upwards of 20 categories of events. The formulation of these classification taxonomies generally involves systematically grouping PSE reports based on common characteristics [19]. The descriptions of event types are not always readily accessible to PSE reporters and patient safety analysts [15]. Previous studies have found that the classification of PSE reports is inconsistent depending on the reporter's profession, interpretation of the adverse event, and understanding of the PSE classification taxonomy [15,20]. Furthermore, 25% of PSE reports are labeled with vague or nonspecific categories such as "miscellaneous" and "other" and require time-consuming retrospective analysis for reclassification [21]. These problems are further exacerbated by the growing volume of PSEs reported [18,22]. For instance, hospitals in the state of New South Wales in Australia reported close to 195,000 PSEs in 2020 [23], while there were approximately 2.3 million PSEs reported to the National Reporting and Learning System in England from April 2021 to March 2022 [24].

In light of these challenges, it is imperative to find an efficient solution to ensure the reliable classification of PSE reports. Recent studies have used static text representations and supervised machine learning (ML) techniques to automate the PSE report classification [17,25,26]. However, static text representations ignore the ordering of the words and do not account for the differences in word meaning across different contexts. These limitations may result in suboptimal classification performance. With the emergence of deep learning, contextual text representation produced from transformer-based deep learning models has achieved state-of-the-art performance on a wide range of natural language processing tasks, including text classification [27]. The contextual representation of each word is based on its surrounding context within the text, allowing for a more accurate understanding of its usage across different contexts and facilitating knowledge transfer across languages [28]. Therefore, using contextual text representation in training ML classifiers presents a promising opportunity for achieving a more precise classification of PSE reports.

The integration of ML models into PSE reporting systems has important implications for human–artificial intelligence (AI) collaboration, given the roles of the incident reporter (front end) and patient safety analyst (backend). Various approaches for using ML classifiers can be developed, including at different levels of automation; however, unifying the strengths of both human expertise and AI offers the most promising route for effective implementation [29-31]. A crucial determinant for successfully implementing the human-AI collaboration approach is decision transparency [32,33], which is often referred to as explainability. Explainability is the concept that an ML model's prediction can be explained in a way that human operators can comprehend and reconstruct the model's reasoning [33]. Incorporating explainability techniques in human-AI collaboration is paramount as it facilitates a deeper understanding of the factors influencing the predictions, thereby fostering trust and understanding between human experts and AI systems. Therefore, embedding explainability into the human-AI collaboration holds significant potential for enhancing PSE report classification.

The main aim of this study is to examine the efficacy of contextual text representation in improving the accuracy of PSE report classification. To accomplish this, we trained, evaluated, and compared various ML classifiers with both static and contextual text representations. Additionally, we developed an interface to illustrate the integration of the ML classifier in an event reporting system to support human-AI collaboration for PSE report classification. Moreover, we enhanced the explainability of the ML classifiers by using an explainable AI technique. Furthermore, we have investigated the ML classifier's performance under 2 conditions, differentiated by whether the explanation is valid for the predicted event type. Based on this analysis, we offer recommendations for optimizing human-AI collaboration in the context of PSE report classification.

Methods

Data Collection

The data set for this study was obtained from a large academic hospital located in the Southeastern United States. A total of 861 PSE reports from the labor and delivery and mother-baby units were extracted from the PSE reporting system from January 1, 2019, to December 31, 2020. Each PSE report was assigned to a single event type from a set of 25 classes, such as complication of the surgery, fall, medication-related, and supply issues. The ML classifiers were trained exclusively on PSE reports from the 7 most frequently occurring event types. This selection was intended to create a more balanced training data set to reduce sampling bias and the risk of overfitting. The selected PSE reports used for training ML classifiers constitute approximately 72.8% (627/861) of the extracted reports (Table 1).

Table 1. Prevalence of patient safety event reports by event type in this study.

Event type	Extracted reports (n=861), n (%)
Care coordination or communication	186 (21.6)
Laboratory test	122 (14.2)
Medication related	89 (10.3)
Omission or errors in assessment, diagnosis, and monitoring	67 (7.8)
Maternal	58 (6.7)
Equipment or devices	56 (6.5)
Supplies	49 (5.7)
Total	627 (72.8)

Data Preprocessing

The free-text section of PSE reports was preprocessed before feeding into ML classifiers as input features. The preprocessing procedures include text normalization, feature extraction, data splitting, and data augmentation (Multimedia Appendix 1 [28,34-39]).

Classifier Training

A range of ML classifiers, including multinomial logistic regression (MLR), support vector machine (SVM), extreme gradient boosting, light gradient boosting, random forest (RF), *k*-nearest neighbor (KNN), and multilayer perceptron, were used for the classification of PSE reports. While SVM is a binary classifier, it is also capable of performing multiclass classification using the one-versus-one strategy. This involves treating the multiclass classification problem as a series of binary classification problems, creating $n \times (n-1)/2$ binary classifiers for each pair of classes, where n represents the total number of classes, and the final classification is based on the majority vote of all binary classifiers. Extreme gradient boosting, light gradient boosting, and RF are tree-based ensemble algorithms that are commonly used in text classification tasks [17,40]. The KNN classifier predicts the class of a data point based on the majority class among its nearest neighbors in the training data set. Multilayer perceptron is a feedforward neural network consisting of multiple layers of interconnected neutrons and trained using backpropagation.

To optimize the performance of ML classifiers, we used the 5-fold cross-validation grid search technique to identify the best combination of hyperparameters. During this process, a range of values of important hyperparameters (ie, regularization strength) is assessed with 5-fold cross-validation. For each combination of hyperparameters, the training set is randomly split into 5 distinct folds, and then the ML classifier is trained and evaluated 5 times, picking a different fold for evaluation every time and training on the remaining 4 folds. The optimized combination of hyperparameters is determined based on the average performance of the classifier on the F_1-score across the 5-fold cross-validation runs.

Classifier Evaluation

We evaluated the performance of the trained classifiers on the testing set with standard classification metrics, including accuracy, precision, recall, F_1-score, and area under the receiver operating characteristic curve. We also evaluated classifiers on top-2 accuracy, which measures the proportion of predictions where the correct event type is among the top 2 highest probability event types predicted by the classifier. The definitions and mathematical formulas of the evaluation metrics are shown in Multimedia Appendix 2. Each of these metrics provides a distinct perspective on the performance of the classifier, and collectively, they offer a comprehensive understanding of how well the classifier is functioning. Since we framed PSE report classification as a multiclass text classification problem, the precision, recall, F_1-score, and area under the receiver operating characteristic curve are computed for each class and combined using a weighted average where the weights correspond to the number of data points in each class.

Development and Assessment of Explainability

As the contextual text representation is generated from transformer-based neural network, which has a black box nature, we used the local interpretable model-agnostic explanations (LIME) technique to analyze the top-performing ML classifier trained with the contextual text representation. LIME is a post hoc, local perturbation technique that provides the explanation for a single prediction. LIME generates perturbed data by randomly removing words from a text document and trains a locally explainable model with perturbed data to simulate the original classifier's prediction [41]. By measuring how the classifier's prediction changes under these perturbations, LIME reflects the contributions of each word to the prediction. The importance of each word can then be assessed for a single prediction, revealing whether the ML classifier has learned to use relevant words for classifying PSE reports. We used LIME to generate explanations for the top-performing classifier's prediction, specifically by highlighting the words that the classifier deems influential for the prediction. We presented 3 distinct cases: one where the classifier effectively leveraged relevant words for accurate prediction, another where it failed to do so, and a final case that illustrated the explanation for a misclassification. In addition, we analyzed the top 5 most prevalent words identified by LIME for each event type.

A total of 2 human factors graduate students were recruited to assess the quality of the LIME explanations. For each PSE report in the test data set, the reviewers were asked to determine independently if any of the highlighted words were relevant to the predicted event type. Based on these evaluations, the reports were then categorized into 2 distinct groups: those in which the highlighted terms were deemed relevant to the predicted event types and those where they were deemed irrelevant. Discrepancies were resolved through discussions. The interrater reliability index (Cohen κ) was calculated to quantify the level of agreement between the reviewers. The ML classifier's accuracy and F_1-score were evaluated for these 2 groups of PSE reports. A subsequent comparison will explore the influence of explanation quality on prediction reliability.

Interface Development

In the typical workflow of PSE report classification, reporters need to provide a narrative description of the event as well as key attributes such as the event type, level of harm, date, and location of the event. Subsequent to this initial classification, the patient safety analyst will review the submitted report and decide if it needs to be recategorized to better reflect the nature of the event [17,42]. To support efficient and reliable categorization, the classifier will need to provide reporters with real-time support during the reporting process. We developed a PSE reporting interface to illustrate the integration of the ML classifier and the LIME explainability technique. In the design, the ML classifier provides multiple high-probability event types along with explanations for its prediction and allows the user to select the most appropriate event type. The interface was developed in Figma [43] and designed using guidance from previous research on incident reporting systems, including question type, mandatory and optional questions, and taxonomy for event type and harm level [44,45].

Ethical Considerations

The study was approved by the Medical University of South Carolina Hospital's institutional review board (Pro00105892). Following data extraction, PSE reports were anonymized in accordance with privacy regulation guidelines.

Results

Performance Comparison

We evaluated the trained ML classifier's classification performance on both static and contextual text representations (Multimedia Appendix 3). The performance of the top-performing ML classifier trained with static and contextual text representations is shown in Table 2. Our results showed that for static text representation, the MLR classifier trained with term frequency–inverse document frequency (TF-IDF) achieved the best performance, with an F_1-score of 0.631 and an accuracy of 66.7% (84/126). On the other hand, for contextual text representation, the SVM classifier trained with RoBERTa-base outperformed others, with an F_1-score of 0.753 and an accuracy of 75.4% (95/126). The SVM classifier trained with RoBERTa-base showed a 19.3% elative improvement in F_1-score and a 13% (11/85) relative improvement in accuracy compared to the MLR classifier trained with TF-IDF for contextual text representation. In addition, we compared the accuracy (95/126, 75.4%) and top 2 accuracy (107/126, 84.9%) of the SVM classifier trained with RoBERTa-base and observed that 9.5% (12/126) of PSE reports' true event type was predicted as the second highest probability event type by the classifier, which represents 39% (12/31) of misclassified PSE reports.

Table 2. Performance of top-performing ML classifiers trained with static and contextual text representations.

Metric	Top-performing ML[a] model trained with the static text representation			Top-performing ML model trained with the contextual text representation		
	Performance	ML classifier	Text Representation	Performance	ML classifier	Text Representation
Accuracy (%)	66.67	MLR[b]	TF-IDF[c]	75.40	SVM[d]	RoBERTa-base
Top 2 accuracy (%)	85.71	MLR	TF-IDF	88.10	MLP[e]	xlm-RoBERTa-base
Precision	0.707	KNN[f]	TF-IDF	0.757	SVM	RoBERTa-base
Recall	0.667	MLR	TF-IDF	0.754	SVM	RoBERTa-base
F_1-score	0.631	MLR	TF-IDF	0.753	SVM	RoBERTa-base

[a]ML: machine learning.
[b]MLR: multinomial logistic regression.
[c]TF-IDF: term frequency–inverse document frequency.
[d]SVM: support vector machine.
[e]MLP: multilayer perceptron.
[f]KNN: *k*-nearest neighbor.

Performance on Classifying Individual Event Types

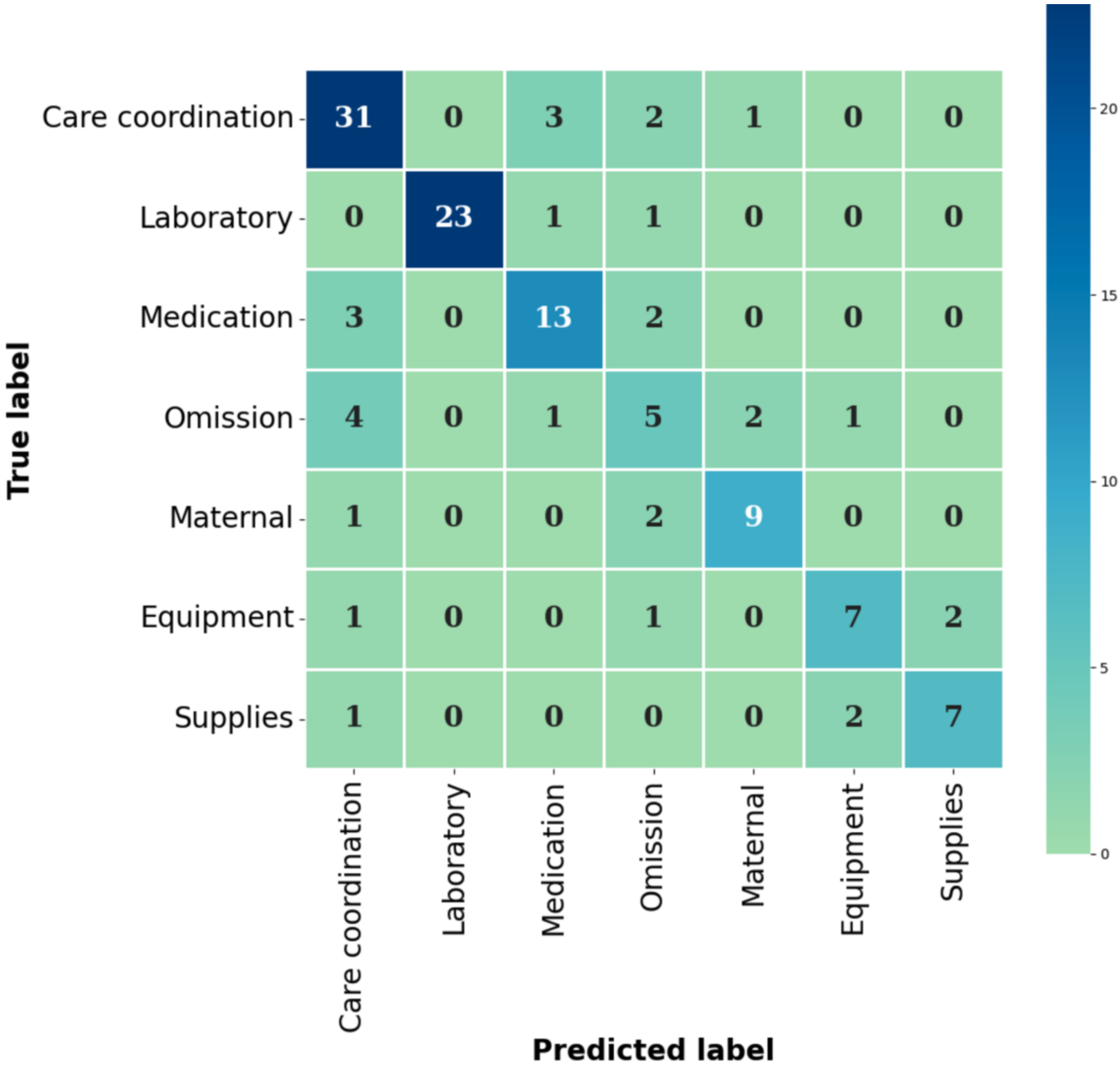
We analyzed the performance of the SVM classifier trained with RoBERTa-base on individual event types (Table 3). The F_1-score measure for different event types ranged from 0.958 (laboratory test) to 0.400 (omission or errors in assessment, diagnosis, and monitoring).

Table 3. Performance of support vector machine+RoBERTa-base on the individual event type.

Event type	Precision	Recall	F_1-score
Care coordination or communication	0.721	0.838	0.775
Laboratory test	1.000	0.920	0.958
Medication related	0.765	0.722	0.743
Omission or errors in assessment, diagnosis, and monitoring	0.417	0.385	0.400
Maternal	0.750	0.750	0.750
Equipment or devices	0.700	0.636	0.667
Supplies	0.778	0.700	0.737

Figure 1 shows the confusion matrix for the SVM classifier trained with RoBERTa-base evaluated on the test set. A confusion matrix is a table that visualizes the performance of a classifier. The main diagonal value is the number of PSE reports that have been classified as true event types, whereas off-diagonal values are the number of PSE reports that have been wrongly classified. While the classifier was able to classify the majority of event types of PSE reports correctly, there is a consistent misclassification of the omission or errors in assessment, diagnosis, or monitoring PSE report as the care coordination or communication (coordination) event type.

Figure 1. Confusion matrix for the testing set evaluation with a support vector machine classifier trained with RoBERTa-base.



LIME-Based Explainability Analysis

We used LIME to evaluate whether the SVM classifier trained with RoBERTa-base has leveraged informative words for classification. Figure 2 presents 3 examples of explanations for the classifier's predictions. At the top of Figure 2, LIME identified "ketorolac," "ibuprofen," and "doses" from the PSE report as important words for classifying the report into the medication-related event type, which is reasonable given the report's association with incorrect medication doses. Conversely, in the middle of Figure 2, LIME highlighted "our," "handle," and "or" from the text as important words for classifying the report into the equipment or device event type. Although the predicted event type was correct, the classifier relied on irrelevant words for the classification. At the bottom of Figure 2, a case of misclassification is shown. LIME highlighted "pitocin," "pump," "available," and "use" as influential words for classifying the PSE report into medication-related event type when it belongs to the equipment class. In addition, for each event type, we extract the 5 most prevalent words that were deemed important for the classifier's prediction across the whole data set (Table 4). This inclusion of stop words (ie, "was," "not," and "till") among influential terms, as shown in Table 4, demonstrated that the classifier does not always rely on relevant words for making classifications.

Figure 2. Local interpretable model-agnostic explanations of support vector machine classifiers trained with RoBERTa-base. MD: medical doctor; PSE: patient safety event; pt: patient.

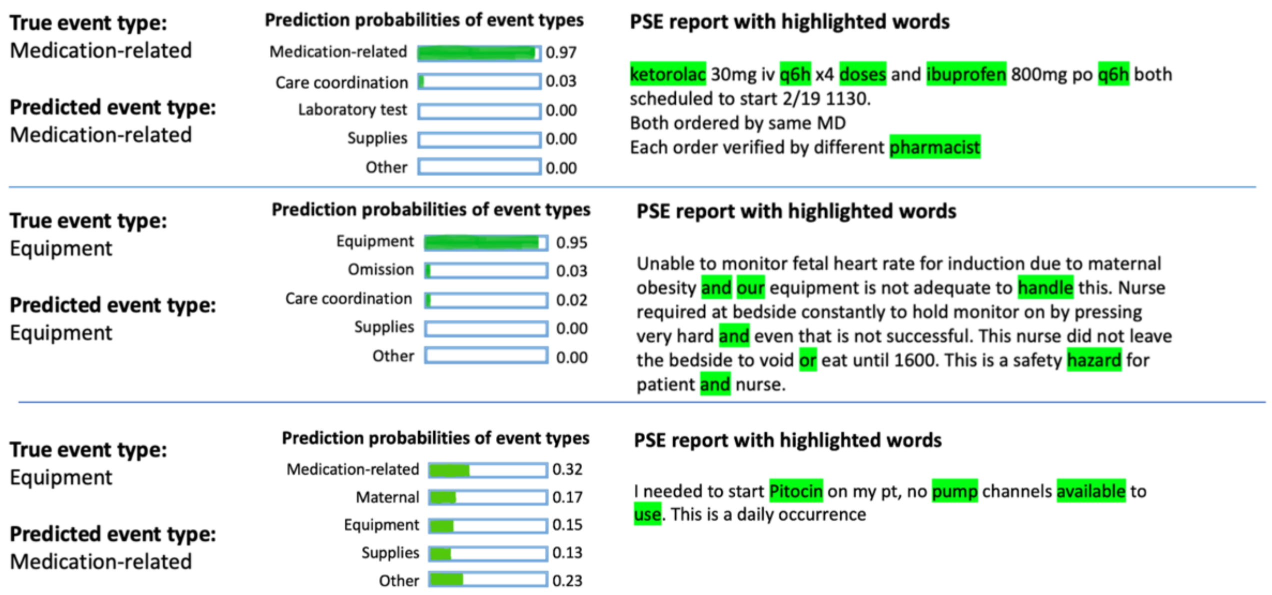


Table 4. The 5 most prevalent and important words for each event type were derived from the support vector machine classifier trained with RoBERTa-base.

Event type	Prevalent influential words highlighted by local interpretable model-agnostic explanations
Care coordination or communication	requested, delayed, patient, not, follow
Laboratory test	specimen, lab, labels, collection, results
Medication related	patches, doses, orders, medication, pitocin
Omission or errors in assessment, diagnosis, and monitoring	warning, patient, was, till, late
Maternal	baby, hysterectomy, stable, pumping, hemorrhage
Equipment or devices	instruments, trays, notified, malfunctioning, faulty
Supplies	vendor, sterile, available, needed, OR

After reviewing the LIME explanations for each PSE report in the test data set, 73.8% (93/126) of the reports were categorized into a subset where at least 1 highlighted word was deemed relevant to the predicted event type. The remaining reports comprised a second subset where no highlighted words were relevant. The interrater reliability index measured by Cohen κ between the 2 reviewers was 0.83, indicating substantial agreement. Table 5 presents the performance of the top-performing ML classifier for both subsets. For the first subset, the classifier achieved an accuracy of 84% (78/93) and an F_1-score of 0.825. In contrast, the second subset showed a classifier accuracy of 52% (17/33) and an F_1-score of 0.549.

Table 5. Performance of a top-performing machine learning classifier on reports that have relevant words highlighted and reports with irrelevant words highlighted.

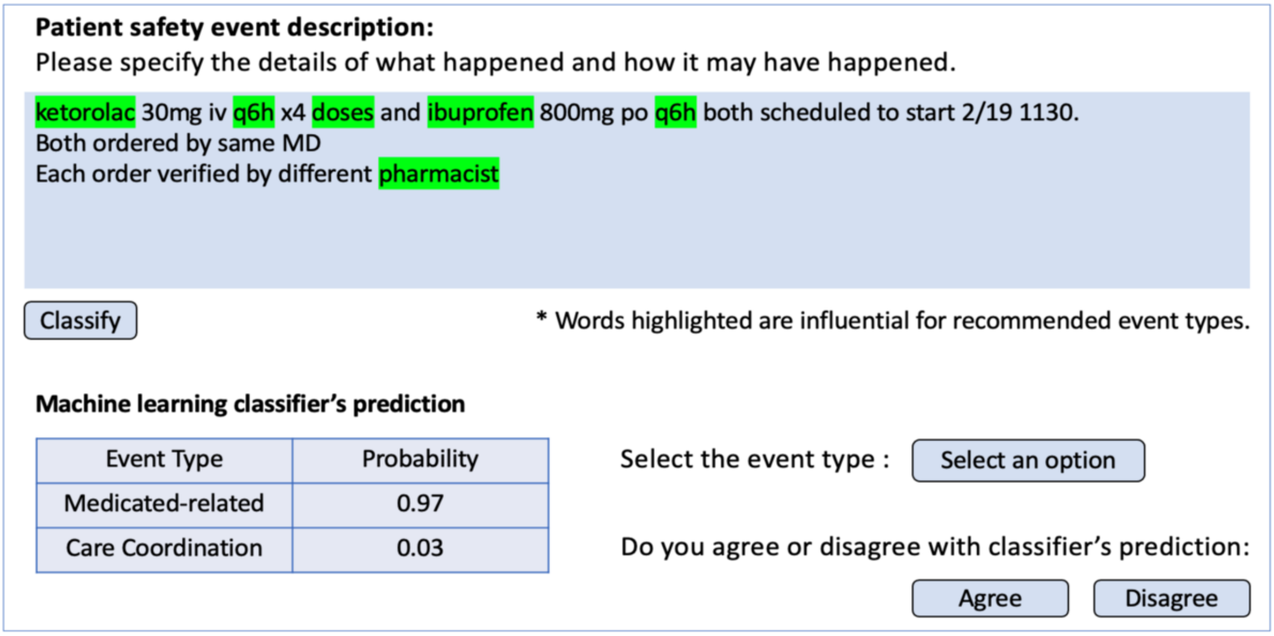
Metric	PSE[a] reports with relevant words highlighted	PSE reports with irrelevant words highlighted
Number of PSE reports, n	93	33
Percentage of test data set (%)	73.81	26.19
Accuracy (%)	83.87	51.51
F_1-score	0.825	0.549

[a]PSE: patient safety event.

PSE Reporting System Interface

We designed an event reporting interface that integrates both the ML classifier and the LIME explainability technique. Figure 3 shows the event classification screen, where reporters enter a narrative description of the event after providing the details of the event, including date, time, unit, and information about the patient and reporter. Before describing the event in narrative form, reporters also choose among factors that contributed to the incident and the level of harm experienced by the patient. Once the reporter enters their narrative and selects the “classify” button, the system activates the ML classifier. Subsequently, the interface displays the top 2 most probable event types, along with their associated probability distributions, in the lower left section. Simultaneously, the LIME technique will identify influential words that significantly contributed to the predicted event type, highlighting these words in green in the upper section of the dashboard. Based on the predicted event types and words highlighted for their influence on the prediction, the reporter may select the most suitable event type from a drop-down menu located in the lower-right section of the dashboard. Following this selection, reporters are queried on whether they agree with the classifier's prediction, and the collected data can be used to guide subsequent refinement of the ML classifier.

Figure 3. Interface visualization of a patient safety event report classifier coupled with the local interpretable model-agnostic explanations technique. MD: medical doctor.



Discussion

Overview

PSE event reporting systems are commonly used in health systems and hospitals across the world [46]. Data collected in PSE reporting systems drive quality improvement and patient safety efforts and supports regulatory reporting requirements for hospitals. The erroneous classification of PSE reports can impede the learning capabilities of the PSE reporting system, leading to suboptimal performance in detecting and preventing potential patient safety hazards [20]. It can also result in a substantial time cost for reclassifying PSE reports and compromise the integrity of a PSE database when analysts are investigating trends in events to develop effective solutions [17]. Previous studies have trained ML classifiers with static text representations for automatic PSE classification [12,17,25,26]. This study aimed to investigate whether using contextual text representations can further improve the accuracy of classifying PSE reports. We trained and evaluated a range of ML classifiers using both static and contextual text representations. To the best of our knowledge, this is the first time that contextual text representation has been used for training ML classifiers for PSE report classification. We analyzed the confusion matrix of the top-performing classifier to identify prevalent misclassified event types. Furthermore, aiming for more accurate and reliable PSE report classification, we incorporated an explainability technique to support human-AI collaboration and designed an interface to illustrate the possible integration of the ML classifier in PSE reporting systems.

Principal Findings

In this study, we extensively investigated the potential of using contextual representation for improving PSE report classification. The leading classifier trained with the static text representation (MLR trained with TF-IDF) was able to achieve an accuracy of 66.7% (84/126). This accuracy considerably exceeds the baseline accuracy of 29.4% (37/126), which involves classifying all PSE reports into the majority event type. However, using contextual text representation proved more efficacious. The SVM trained with contextual text representation (RoBERTa-base) was able to achieve an accuracy of 75.4% (95/126), reflecting a 13% (11/84) relative improvement in accuracy compared to the best-performing classifier trained with static text representation. While the achieved accuracy of 75.4% may not appear outstanding in isolation, it represents a significant advance compared with static text representations and exceeds the baseline, given the limited size of the data set. The improvement in classifier performance can be attributed to the use of contextual text representations, which can capture not only the meaning of individual words but also the complex and subtle ways in which words interact with each other in a specific context. Therefore, contextual text representation overcomes some limitations of static text representation, which relies primarily on word frequency and co-occurrence to represent text. Moreover, contextual text representation does not require explicit text normalization while also avoiding issues associated with high-dimensionality and sparsity commonly found in static text representations. Hence, when training ML classifiers for PSE reporting systems, contextual text representation should be prioritized over static text representation to ensure the highest level of accuracy in classifying PSE reports.

As part of our investigation, we evaluated the performance of the top-performing classifier trained with contextual text representation on individual event types. While the classifier demonstrated impressive performance in accurately classifying laboratory test PSE reports (F_1-score=0.958), it struggled with classifying omissions or errors in assessment, diagnosis, and monitoring PSE reports, resulting in an unsatisfactory F_1-score of 0.400. To investigate this discrepancy, we analyzed the confusion matrix for the classifier and discovered that omissions or errors in assessment, diagnosis, and monitoring PSE reports were frequently misclassified as the coordination event type. This misclassification can be attributed to the multiclass nature of PSE reports. For example, a failure to document the removal of a patient's epidural catheter (omission or errors in assessment, diagnosis, and monitoring) could lead to a medication ordered by a physician (such as Lovenox) being withheld by the pharmacy due to a complication risk (coordination). On the other hand, the laboratory test is a more distinct event type in comparison to the other event types, and the classifier was able to correctly classify the majority of these reports. The observation obtained from the confusion matrix implies that PSE reports can potentially have more than 1 event type. This finding is consistent with previous studies [25,26]. The finding also underscores the need for further refinement in the development of the PSE taxonomy to create more distinctive event types. Another potential solution for addressing the multiclass nature of PSEs is to enable multiple event-type assignments [47]. Alternatively, the ML classifier can provide several probable event types, allowing the user to select the most appropriate one. We evaluated the top 2 accuracy of the top-performing ML classifier trained with contextual text representation and observed that 39% (12/31) of misclassified PSE reports' true event type was predicted as the second-highest probability event type by the classifier. The finding suggests that there is a greater chance for the ML classifier to provide the correct event type when considering multiple options. As event reporting systems usually encompass over 20 event types, which can be difficult to memorize or access [17], narrowing down the PSE report's potential event types to a smaller range also reduces the cognitive workload for PSE reporters during the classification process [48] and enhances the efficiency of reclassifying PSE reports for patient safety analysts.

We used LIME to showcase 3 predictions' explanations and demonstrated cases where the ML classifier used informative words for classifying the PSE report and where it used irrelevant words for classification. These results highlight the importance of not solely relying on the ML classifier's prediction and underscore the need for explainability and transparency in using the ML classifier for PSE report classification. Additionally, we showed the top 5 most prevalent words the ML classifier deemed important in the PSE reports for each event type. These words are indicative of the prevalent themes and issues within specific event types. Understanding the context and relationships between these prevalent informative words and specific event types can potentially provide valuable insights into the factors contributing to different types of PSEs. Furthermore, we have evaluated the top-performing ML classifier's performance on 2 subsets of PSE reports, differentiated by whether the highlighted word by LIME is relevant to the predicted event type. Our findings reveal that the majority of PSE reports (93/126) have at least 1 relevant word highlighted, with the classifier achieving an accuracy of 84% (78/93) on these reports. Conversely, accuracy drops to 52% (17/33) when irrelevant words are highlighted. Such a disparity in performance emphasizes the necessity for additional scrutiny from reporters and patient safety analysts, particularly when dealing with PSE reports that have irrelevant words highlighted.

While previous research has focused on the development of ML classifiers, none of these previous works have investigated the potential integration of the classifier within the PSE reporting system in a manner that aligns with the workflow of the front-end reporter. We designed an interface to demonstrate the feasibility of a collaborative human-AI approach for event categorization. The interface provides the PSE reporter with multiple probable event types and associated explanations for the ML classifier's prediction. This approach aligns with the principles of level 2 automation, where ML classifiers aids human decision-making rather than fully automating it [49]. This collaboration optimally combines human expertise with ML capabilities, potentially reducing cognitive workload and memorization of the taxonomy while also reducing the risks associated with overreliance on automation. Numerous studies have shown that the human-AI collaboration approach can improve the decision-making process [50-52], indicating its potential for enhancing PSE report classification. Furthermore, the interface also integrates the LIME explainability technique, which offers real-time insights into the rationale for the probable event types. Given the role of reporters and patient safety analysts in the incident reporting process, the use of explainability techniques can also increase trust in the recommendation provided by the ML classifier as it provides transparent and interpretable reasoning for the classification decisions [50,51]. Using LIME to highlight top informative words in real time for a PSE report can assist PSE reporters by emphasizing keywords in their narratives that are linked to the proposed classification. Highlighting informative words can also facilitate patient safety analysts working at the back end by providing insights into why a specific event type was chosen for classification. Such transparency not only clarifies current recommendations but also guides analysts in identifying influential terms for future report classifications. Previous research has illustrated the value of automation transparency in supporting appropriate levels of trust in the system, including decision support systems [32]. Additionally, regularly checking the explanations of the ML classifier's prediction enables continuous monitoring of the classifier's performance, identification of issues, and refinement [52]. As we have only designed the interface, additional research is needed to test the effectiveness of this approach in PSE report classification. Assessing the interface's impact on cognitive workload and decision-making accuracy is essential for ensuring its usability and adoption in the event reporting system. We plan to undertake a usability testing study with health care professionals in a subsequent study.

Comparison With Previous Work

Research into the use of ML classifiers for the automation of PSE report classification has been relatively scarce. Wang et al [26] used logistic regression and SVM with the binary count, term frequency, and TF-IDF text representation to classify ten types of PSE reports, reaching an F_1-score as high as 0.783. However, they used a considerably larger data set (n=2860). Fong et al [17] achieved an accuracy rate of 92.0% (284/309) when they examined the usage of an ML classifier for classifying miscellaneous PSE reports using SVM, RF, and logistic regression with TF-IDF [17]. They also used a much larger data set (n=70,051). Ong et al [12] investigated the feasibility of using an ML classifier to automatically classify 2 types of PSE reports, including inadequate clinical handover and incorrect patient identification. They used Bag of Words model for text representation and trained both SVM and naive Bayes on classifying PSE reports, reaching accuracy as high as 98% (364/372). However, they framed the problem as a binary classification problem, which inherently has a higher baseline accuracy compared to our investigation. In this study, we've performed an in-depth comparative analysis with the available PSE data set and compared the established methods of classifying PSE reports and our novel method of using contextual text representations for classification. Our findings reveal that our proposed method outperforms the traditional models in terms of accuracy (ie, 84/126, 66.7% vs 95/126, 75.4%) and F_1-score (ie, 0.631 vs 0.753). This underlines the significance of our approach and its potential to advance the field of using ML classifiers for PSE report classification.

Limitations

There are several limitations to this study. First, the PSE reports used to train the ML classifiers were obtained from the maternal care units of a single hospital in the United States; therefore, the classifier might not generalize well to other settings. Second, this research's scope was constrained by the limited amount of PSE report data, and only 7 prevalent classes were incorporated for training the ML classifiers. The restricted quantity of PSE reports might also result in an underestimation of the ML classifier's actual capabilities [12]. Third, the quality of the LIME explanations was assessed by 2 graduate students; thus, further investigation is needed for a more robust validation of explanation quality. Furthermore, we have not yet empirically tested the interface for potential decision-making biases it may introduce.

Future research should investigate the performance of ML classifiers trained with contextual text representations on a larger and more diverse data set. Additionally, while we plan to refine the interface and test whether it supports event classification, future research can continue to investigate the appropriate way of incorporating the ML classifier into the reporting and reviewing workflow of PSE report classification and examine various human-AI collaboration approaches. Future studies should explore the potential biases (ie, automation bias) that the interface may introduce into the analysts' decision-making process.

Conclusions

Improving the precision of PSE report classifications is a multifaceted task, involving both the refinement of the event type taxonomy and adequate training of hospital staff on the event reporting system. Despite these challenges, ML classifiers offer substantial potential to support accurate classification throughout the reporting and reviewing process. The findings of this study contribute to the advancement of ML classifiers for PSE report classification by demonstrating the superior performance of contextual text representation over static text representations in achieving more accurate classification outcomes. The integration of explainability techniques in ML classifiers fosters trust in their usage and provides valuable insights for informed decision-making and potential adjustments to the classifier. An event reporting interface that integrates an ML classifier with collaborative decision-making capabilities offers the potential to achieve an efficient and reliable PSE report classification process. These approaches can ultimately help hospitals identify risks and hazards promptly and take timely and informed actions to mitigate adverse events and reduce patient harm.

Acknowledgments

This research was funded by the Agency for Healthcare Research and Quality (1R03HS027680).

Authors' Contributions

HC was responsible for the conceptualization, data analysis, and drafting of the manuscript. EC contributed to the conceptualization, methodology design, and review and revision of the document. DW contributed to data acquisition and funding acquisition. MA contributed to data acquisition, conceptualization, funding acquisition, and the review and revision of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data preprocessing procedures.
[DOCX File , 14 KB - humanfactors_v11i1e53378_app1.docx]

Multimedia Appendix 2
Evaluation metrics for examining patient safety event machine learning classifiers.
[DOCX File , 15 KB - humanfactors_v11i1e53378_app2.docx]

Multimedia Appendix 3
The performance of machine learning classifiers in classifying patient safety event report event type.
[DOCX File , 20 KB - humanfactors_v11i1e53378_app3.docx]

References

1. Kohn LT, Corrigan JM, Donaldson MS, Committee on Quality of Health Care in America. To Err Is Human: Building a Safer Health System. Volume 627. Washington (DC): National Academies Press (US); 2000.
2. Nebeker JR, Barach P, Samore MH. Clarifying adverse drug events: a clinician's guide to terminology, documentation, and reporting. Ann Intern Med 2004 May 18;140(10):795-801 [FREE Full text] [doi: 10.7326/0003-4819-140-10-200405180-00009] [Medline: 15148066]
3. Chopard D, Treder MS, Corcoran P, Ahmed N, Johnson C, Busse M, et al. Text mining of adverse events in clinical trials: deep learning approach. JMIR Med Inform 2021;9(12):e28632 [FREE Full text] [doi: 10.2196/28632] [Medline: 34951601]
4. Makary MA, Daniel M. Medical error-the third leading cause of death in the US. BMJ 2016;353:i2139 [FREE Full text] [doi: 10.1136/bmj.i2139] [Medline: 27143499]
5. Slawomirski L, Auraaen A, Klazinga NS. The Economics of Patient Safety: Strengthening a Value-based Approach to Reducing Patient Harm at National Level. Paris: OECD Publishing; 2017.
6. Medication without harm. World Health Organization. 2023. URL: https://www.who.int/initiatives/medication-without-harm [accessed 2023-03-08]
7. Hewitt TA, Chreim S. Fix and forget or fix and report: a qualitative study of tensions at the front line of incident reporting. BMJ Qual Saf 2015;24(5):303-310 [FREE Full text] [doi: 10.1136/bmjqs-2014-003279] [Medline: 25749025]
8. Ward JK, Armitage G. Can patients report patient safety incidents in a hospital setting? a systematic review. BMJ Qual Saf 2012;21(8):685-699 [FREE Full text] [doi: 10.1136/bmjqs-2011-000213] [Medline: 22562875]
9. Li E, Clarke J, Ashrafian H, Darzi A, Neves AL. The impact of electronic health record interoperability on safety and quality of care in high-income countries: systematic review. J Med Internet Res 2022;24(9):e38144 [FREE Full text] [doi: 10.2196/38144] [Medline: 36107486]
10. Vincent C. Incident reporting and patient safety. BMJ 2007;334(7584):51 [FREE Full text] [doi: 10.1136/bmj.39071.441609.80] [Medline: 17218667]
11. Albolino S, Tartaglia R, Bellandi T, Amicosante AMV, Bianchini E, Biggeri A. Patient safety and incident reporting: survey of Italian healthcare workers. Qual Saf Health Care 2010;19(Suppl 3):i8-i12 [FREE Full text] [doi: 10.1136/qshc.2009.036376] [Medline: 20959324]
12. Ong MS, Magrabi F, Coiera E. Automated categorisation of clinical incident reports using statistical text classification. Qual Saf Health Care 2010;19(6):e55 [FREE Full text] [doi: 10.1136/qshc.2009.036657] [Medline: 20724392]
13. Herzer KR, Mirrer M, Xie Y, Steppan J, Li M, Jung C, et al. Patient safety reporting systems: sustained quality improvement using a multidisciplinary team and "good catch" awards. Jt Comm J Qual Patient Saf 2012;38(8):339-347 [FREE Full text] [doi: 10.1016/s1553-7250(12)38044-6] [Medline: 22946251]
14. Fong A, Hettinger AZ, Ratwani RM. Exploring methods for identifying related patient safety events using structured and unstructured data. J Biomed Inform 2015;58:89-95 [FREE Full text] [doi: 10.1016/j.jbi.2015.09.011] [Medline: 26432354]
15. Brubacher JR, Hunte GS, Hamilton L, Taylor A. Barriers to and incentives for safety event reporting in emergency departments. Healthc Q 2011;14(3):57-65 [FREE Full text] [doi: 10.12927/hcq.2011.22491] [Medline: 21841378]
16. Conceptual framework for the international classification for patient safety version 1.1: final technical report January 2009. World Health Organization. 2010. URL: https://apps.who.int/iris/handle/10665/70882 [accessed 2023-03-08]
17. Fong A, Behzad S, Pruitt Z, Ratwani RM. A machine learning approach to reclassifying miscellaneous patient safety event reports. J Patient Saf 2021;17(8):e829-e833 [FREE Full text] [doi: 10.1097/PTS.0000000000000731] [Medline: 32555052]
18. Ngo J, Lau D, Ploquin J, Receveur T, Stassen K, Del Castilho C. Improving incident reporting among physicians at south health campus hospital. BMJ Open Qual 2022;11(4):e001945 [FREE Full text] [doi: 10.1136/bmjoq-2022-001945] [Medline: 36207052]
19. Schrager JD, Schuler K, Isakov AP, Wright DW, Yaffee AQ, Jacobson KL, et al. Development and usability testing of a web-based COVID-19 self-triage platform. West J Emerg Med 2020;21(5):1054-1058 [FREE Full text] [doi: 10.5811/westjem.2020.7.48217] [Medline: 32970554]
20. Gong Y, Song HY, Wu X, Hua L. Identifying barriers and benefits of patient safety event reporting toward user-centered design. Saf Health 2015;1(1):1-9 [FREE Full text] [doi: 10.1186/2056-5917-1-7]
21. Gong Y. Data consistency in a voluntary medical incident reporting system. J Med Syst 2011;35(4):609-615 [FREE Full text] [doi: 10.1007/s10916-009-9398-y] [Medline: 20703528]

22. Koike D, Ito M, Horiguchi A, Yatsuya H, Ota A. Implementation strategies for the patient safety reporting system using consolidated framework for implementation research: a retrospective mixed-method analysis. BMC Health Serv Res 2022;22(1):409 [FREE Full text] [doi: 10.1186/s12913-022-07822-9] [Medline: 35346182]
23. Biannual incident report. Clinical Excellence Commission. URL: https://www.cec.health.nsw.gov.au/Review-incidents/Biannual-Incident-Report [accessed 2023-05-16]
24. Organisation patient safety incident report up to March 2022. NHS England. 2022. URL: https://www.england.nhs.uk/publication/organisation-patient-safety-incident-report-up-to-march-2022/ [accessed 2023-03-08]
25. Evans HP, Anastasiou A, Edwards A, Hibbert P, Makeham M, Luz S, et al. Automated classification of primary care patient safety incident report content and severity using supervised Machine Learning (ML) approaches. Health Informatics J 2020;26(4):3123-3139 [FREE Full text] [doi: 10.1177/1460458219833102] [Medline: 30843455]
26. Wang Y, Coiera E, Runciman W, Magrabi F. Using multiclass classification to automate the identification of patient safety incident reports by type and severity. BMC Med Inform Decis Mak 2017;17(1):84 [FREE Full text] [doi: 10.1186/s12911-017-0483-8] [Medline: 28606174]
27. Minaee S, Kalchbrenner N, Cambria E, Nikzad N, Chenaghlu M, Gao J. Deep learning—based text classification: a comprehensive review. ACM Comput Surv 2021;54(3):1-40 [FREE Full text] [doi: 10.1145/3439726]
28. Liu Q, Kusner MJ, Blunsom P. A survey on contextual embeddings. ArXiv Preprint posted online on April 13, 2020. [doi: 10.48550/arXiv.2003.07278]
29. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. J Med Internet Res 2020;22(6):e15154 [FREE Full text] [doi: 10.2196/15154] [Medline: 32558657]
30. Sharma M, Savage C, Nair M, Larsson I, Svedberg P, Nygren JM. Artificial intelligence applications in health care practice: scoping review. J Med Internet Res 2022;24(10):e40238 [FREE Full text] [doi: 10.2196/40238] [Medline: 36197712]
31. Sezgin E. Artificial intelligence in healthcare: complementing, not replacing, doctors and healthcare providers. Digit Health 2023;9:20552076231186520 [FREE Full text] [doi: 10.1177/20552076231186520] [Medline: 37426593]
32. Hemmer P, Schemmer M, Riefle L, Rosellen N, Vössing M, Kühl N. Factors that influence the adoption of human-AI collaboration in clinical decision-making. ArXiv Preprint posted online on April 19, 2022. [doi: 10.48550/arXiv.2204.09082]
33. Amann J, Blasimme A, Vayena E, Frey D, Madai VI, Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak 2020;20(1):310 [FREE Full text] [doi: 10.1186/s12911-020-01332-6] [Medline: 33256715]
34. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. Minneapolis, Minnesota: Association for Computational Linguistics; 2019 Presented at: 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; June 2-7, 2019; Minneapolis, Minnesota p. 4171-4186.
35. Pennington J, Socher R, Manning C. GloVe: global vectors for word representation. Doha, Qatar: Association for Computational Linguistics; 2014 Presented at: 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP); October 25-29, 2014; Doha, Qatar p. 1532-1543. [doi: 10.3115/v1/d14-1162]
36. Models. Hugging Face. 2022. URL: https://huggingface.co/models [accessed 2023-03-09]
37. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: a robustly optimized BERT pretraining approach. ArXiv Preprint posted online on July 26 2019. [doi: 10.5260/chara.21.2.8]
38. Kumar P, Bhatnagar R, Gaur K, Bhatnagar A. Classification of imbalanced data:review of methods and applications. IOP Conf Ser Mater Sci Eng 2021;1099:012077 [FREE Full text] [doi: 10.1088/1757-899x/1099/1/012077]
39. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Int Res 2002;16(1):321-357 [FREE Full text] [doi: 10.1613/jair.953]
40. Wang L, Zhang Y, Chignell M, Shan B, Sheehan KA, Razak F, et al. Boosting delirium identification accuracy with sentiment-based natural language processing: mixed methods study. JMIR Med Inform 2022;10(12):e38161 [FREE Full text] [doi: 10.2196/38161] [Medline: 36538363]
41. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": explaining the predictions of any classifier. 2016 Presented at: KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17, 2016; San Francisco, California p. 1135-1144 URL: https://dl.acm.org/doi/abs/10.1145/2939672.2939778 [doi: 10.1145/2939672.2939778]
42. Palojoki S, Saranto K, Reponen E, Skants N, Vakkuri A, Vuokko R. Classification of electronic health record-related patient safety incidents: development and validation study. JMIR Med Inform 2021;9(8):e30470 [FREE Full text] [doi: 10.2196/30470] [Medline: 34245558]
43. The collaborative interface design tool. Figma. URL: https://www.figma.com/ [accessed 2023-09-24]
44. Delio J, Catalanotti JS, Marko K, Paul C, Taffel M, Ho G, et al. Integrating adverse event reporting into a free-text mobile application used in daily workflow increases adverse event reporting by physicians. Am J Med Qual 2020;35(5):374-379 [FREE Full text] [doi: 10.1177/1062860619891995] [Medline: 31795736]
45. Klemp K, Zwart D, Hansen J, Hellebek T, Luettel D, Verstappen W, et al. A safety incident reporting system for primary care. A systematic literature review and consensus procedure by the LINNEAUS collaboration on patient safety in primary care. Eur J Gen Pract 2015;21(sup1):39-44 [FREE Full text] [doi: 10.3109/13814788.2015.1043728] [Medline: 26339835]

46. Hasegawa T, Fujita S. Patient safety policies experiences, effects and priorities; lessons from OECD member states (version 2.0). Patient Safety Global Ministerial Summit 2018. 2018. URL: https://www.mhlw.go.jp/psgms2018/pdf/document/5_Document.pdf [accessed 2023-12-21]
47. Wang Y, Coiera E, Runciman W, Magrabi F. Automating the identification of patient safety incident reports using multi-label classification. Stud Health Technol Inform 2017;245:609-613. [Medline: 29295168]
48. Buettner R. Cognitive workload of humans using artificial intelligence systems: towards objective measurement applying eye-tracking technology. In: KI 2013: Advances in Artificial Intelligence Berlin. Heidelberg: Springer Berlin Heidelberg; 2013 Presented at: 36th Annual German Conference on AI, Koblenz; September 16-20, 2013; Germany p. 37-48. [doi: 10.1007/978-3-642-40942-4_4]
49. Berretta S, Tausch A, Ontrup G, Gilles B, Peifer C, Kluge A. Defining human-AI teaming the human-centered way: a scoping review and network analysis. Front Artif Intell 2023;6:1250725 [FREE Full text] [doi: 10.3389/frai.2023.1250725] [Medline: 37841234]
50. Lai Y, Kankanhalli A, Ong DC. Human-AI collaboration in healthcare: a review and research agenda. 2021 Presented at: 54th Hawaii International Conference on System Sciences (HICSS); January 5-8, 2021; Grand Wailea, Maui, Hawaii URL: https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/495a50c0-16e9-4c3e-a00b-34aeec9dff6b/content [doi: 10.24251/hicss.2021.046]
51. Reverberi C, Rigon T, Solari A, Hassan C, Cherubini P, GI Genius CADx Study Group, et al. Experimental evidence of effective human-AI collaboration in medical decision-making. Sci Rep 2022;12(1):14952 [FREE Full text] [doi: 10.1038/s41598-022-18751-2] [Medline: 36056152]
52. Lee MH, Siewiorek DP, Smailagic A, Bernardino A, Bermúdez i Badia S. A human-AI collaborative approach for clinical decision making on rehabilitation assessment. New York, NY, USA: Association for Computing Machinery; 2021 Presented at: CHI '21: CHI Conference on Human Factors in Computing Systems; May 8-13, 2021; Yokohama Japan p. 1-14 URL: https://dl.acm.org/doi/10.1145/3411764.3445472 [doi: 10.1145/3411764.3445472]

Abbreviations

AI: artificial intelligence
KNN: k-nearest neighbor
LIME: local interpretable model-agnostic explanations
ML: machine learning
MLR: multinomial logistic regression
OECD: Organization for Economic Cooperation and Development
PSE: patient safety event
RF: random forest
SVM: support vector machine
TF-IDF: term frequency–inverse document frequency

Edited by E Borycki, A Kushniruk; submitted 06.10.23; peer-reviewed by H Tibble, S Matathil; comments to author 19.11.23; revised version received 30.11.23; accepted 03.12.23; published 25.01.24.

Please cite as:
Chen H, Cohen E, Wilson D, Alfred M
A Machine Learning Approach with Human-AI Collaboration for Automated Classification of Patient Safety Event Reports: Algorithm Development and Validation Study
JMIR Hum Factors 2024;11:e53378
URL: https://humanfactors.jmir.org/2024/1/e53378
doi:10.2196/53378
PMID:38271086

Original Paper

Testing the Feasibility and Acceptability of Using an Artificial Intelligence Chatbot to Promote HIV Testing and Pre-Exposure Prophylaxis in Malaysia: Mixed Methods Study

Min Hui Cheah[1], MClinPharm; Yan Nee Gan[2], MSc; Frederick L Altice[3,4,5,6], MA, MD; Jeffrey A Wickersham[3,5,6], PhD; Roman Shrestha[5,6,7], MPH, PhD; Nur Afiqah Mohd Salleh[2], PhD; Kee Seong Ng[1], MD, PhD; Iskandar Azwa[1,6], MD; Vimala Balakrishnan[8], PhD; Adeeba Kamarulzaman[9], MBBS; Zhao Ni[5,10], PhD, BMed

[1]Department of Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia
[2]Department of Social and Preventive Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia
[3]Section of Infectious Disease, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, United States
[4]Division of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT, United States
[5]Center for Interdisciplinary Research on AIDS (CIRA), Yale University, New Haven, CT, United States
[6]Centre of Excellence for Research in AIDS (CERiA), Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia
[7]Department of Allied Health Sciences, University of Connecticut, Storrs, CT, United States
[8]Department of Information Systems, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia
[9]Monash University Malaysia, Kuala Lumpur, Malaysia
[10]School of Nursing, Yale University, Orange, CT, United States

Corresponding Author:
Zhao Ni, PhD, BMed
School of Nursing
Yale University
400 West Campus Drive
Orange, CT, 06477
United States
Phone: 1 203 737 3039
Email: zhao.ni@yale.edu

Abstract

Background: The HIV epidemic continues to grow fastest among men who have sex with men (MSM) in Malaysia in the presence of stigma and discrimination. Engaging MSM on the internet using chatbots supported through artificial intelligence (AI) can potentially help HIV prevention efforts. We previously identified the benefits, limitations, and preferred features of HIV prevention AI chatbots and developed an AI chatbot prototype that is now tested for feasibility and acceptability.

Objective: This study aims to test the feasibility and acceptability of an AI chatbot in promoting the uptake of HIV testing and pre-exposure prophylaxis (PrEP) in MSM.

Methods: We conducted beta testing with 14 MSM from February to April 2022 using Zoom (Zoom Video Communications, Inc). Beta testing involved 3 steps: a 45-minute human-chatbot interaction using the think-aloud method, a 35-minute semistructured interview, and a 10-minute web-based survey. The first 2 steps were recorded, transcribed verbatim, and analyzed using the Unified Theory of Acceptance and Use of Technology. Emerging themes from the qualitative data were mapped on the 4 domains of the Unified Theory of Acceptance and Use of Technology: performance expectancy, effort expectancy, facilitating conditions, and social influence.

Results: Most participants (13/14, 93%) perceived the chatbot to be useful because it provided comprehensive information on HIV testing and PrEP (*performance expectancy*). All participants indicated that the chatbot was easy to use because of its simple, straightforward design and quick, friendly responses (*effort expectancy*). Moreover, 93% (13/14) of the participants rated the overall chatbot quality as high, and all participants perceived the chatbot as a helpful tool and would refer it to others. Approximately 79% (11/14) of the participants agreed they would continue using the chatbot. They suggested adding a local language (ie, Bahasa Malaysia) to customize the chatbot to the Malaysian context (*facilitating condition*) and suggested that the chatbot should also incorporate more information on mental health, HIV risk assessment, and consequences of HIV. In terms of *social influence*, all participants perceived the chatbot as helpful in avoiding stigma-inducing interactions and thus could increase the frequency of HIV testing and PrEP uptake among MSM.

Conclusions: The current AI chatbot is feasible and acceptable to promote the uptake of HIV testing and PrEP. To ensure the successful implementation and dissemination of AI chatbots in Malaysia, they should be customized to communicate in Bahasa Malaysia and upgraded to provide other HIV-related information to improve usability, such as mental health support, risk assessment for sexually transmitted infections, AIDS treatment, and the consequences of contracting HIV.

(JMIR Hum Factors 2024;11:e52055) doi:10.2196/52055

KEYWORDS

artificial intelligence; acceptability; chatbot; feasibility; HIV prevention; HIV testing; men who have sex with men; MSM; mobile health; mHealth; preexposure prophylaxis; PrEP; mobile phone

Introduction

Background

HIV continues to be a global health concern causing approximately 630,000 deaths yearly worldwide [1]. In 2019, approximately 62% of the new HIV infections among adults worldwide occurred within key populations and their sexual partners [2]. Men who have sex with men (MSM) accounted for 23% of new infections of HIV, which was much higher than the percentage of new infections in other key populations, such as people who used drugs (10%), sex workers (8%), and transgender people (2%) in 2019 [3]. Malaysia is a Southeast Asian country with a population of 33.5 million, with 1 in 5 MSM living with HIV [4]. Over the past 2 decades, the mode of HIV transmission in Malaysia has shifted from needle sharing to sexual transmission, particularly among MSM [5].

HIV testing is a prerequisite to effective HIV prevention and early treatment initiation [6]; people at risk for HIV or seropositive individuals need to be tested for HIV before being linked to health care services [7]. Despite the importance of HIV testing, it is disproportionately lower among MSM in Malaysia [8]. New HIV testing guidelines recommend that MSM at high risk for HIV should be tested every 3 to 6 months, but most MSM in Malaysia do not test optimally. Studies in Malaysia have found that only 9.5% of MSM tested more than once a year. In Malaysia, engaging in same-sex sexual behavior is prohibited by both secular and Sharia laws, leading to significant levels of stigma and discrimination within society [9]. As a result, many MSM may be hesitant or unwilling to engage with health care providers and outreach workers. Therefore, designing new strategies to promote HIV testing among MSM in Malaysia is urgently needed [10].

Using portable electronic devices with software programs to deliver health care services and manage patient information is known as mobile health (mHealth) [11]. mHealth interventions could reduce barriers to HIV testing for MSM by reducing in-person contact and offering internet-based platforms for HIV testing [12,13]. Studies have demonstrated that mHealth interventions using smartphones and apps could increase the uptake of HIV testing while protecting the privacy of MSM [14-16], and MSM in Malaysia have a high acceptance of the use of mHealth for HIV testing and prevention [13,17,18]. Recent breakthroughs in artificial intelligence (AI) and machine learning can potentially automate and scale up these mHealth interventions through chatbots, a computer program that can mimic human conversation [19]. However, leveraging chatbot technology to promote HIV testing and prevention is in its infancy [15,20]. Although chatbot technology holds immense potential to prevent HIV, a lack of research in this field undermines its significance. The creation of ChatGPT has brought attention to the significance of studying chatbot technology for health care.

Our team has conducted formative research to understand HIV prevention chatbots in Malaysia and has identified the perceived benefits, limitations, and preferred features of AI chatbots for HIV testing and prevention among MSM [13]. On the basis of the study findings, we developed an HIV prevention AI chatbot prototype named Haris (a common Malaysian name) and a website called MYHIV365 (*MY* symbolizes Malaysia, *HIV* implies health care services aimed at preventing HIV, and *365* indicates the services are available every day of the year). Haris was hosted on MYHIV365 and could provide information on the 3 themes most needed by MSM: HIV testing, mental health, and pre-exposure prophylaxis (PrEP). PrEP is a highly effective HIV prevention method that involves the use of antiretroviral medication by at-risk individuals to prevent getting HIV from sex or injection drug use. Haris imitates human intelligence and can interact with users to provide support, including ordering free HIV self-testing kits, screening for depression, and recommending MSM-friendly clinics where individuals can get tested for HIV and receive PrEP.

Objectives

Despite the meticulous design and alpha testing (internal testing) of Haris among professors, experts, and community advisory board members, its feasibility and acceptability in preventing HIV among MSM is still unknown. Therefore, we conducted beta testing (testing in a real-world environment by actual users) of Haris among 14 MSM in Malaysia to address this knowledge gap. Specifically, we examined the use of the AI chatbot for delivering health information and improving linkage to HIV testing, PrEP, and care. We also investigated key strategies to refine the feasibility and acceptability of the AI chatbot in this study.

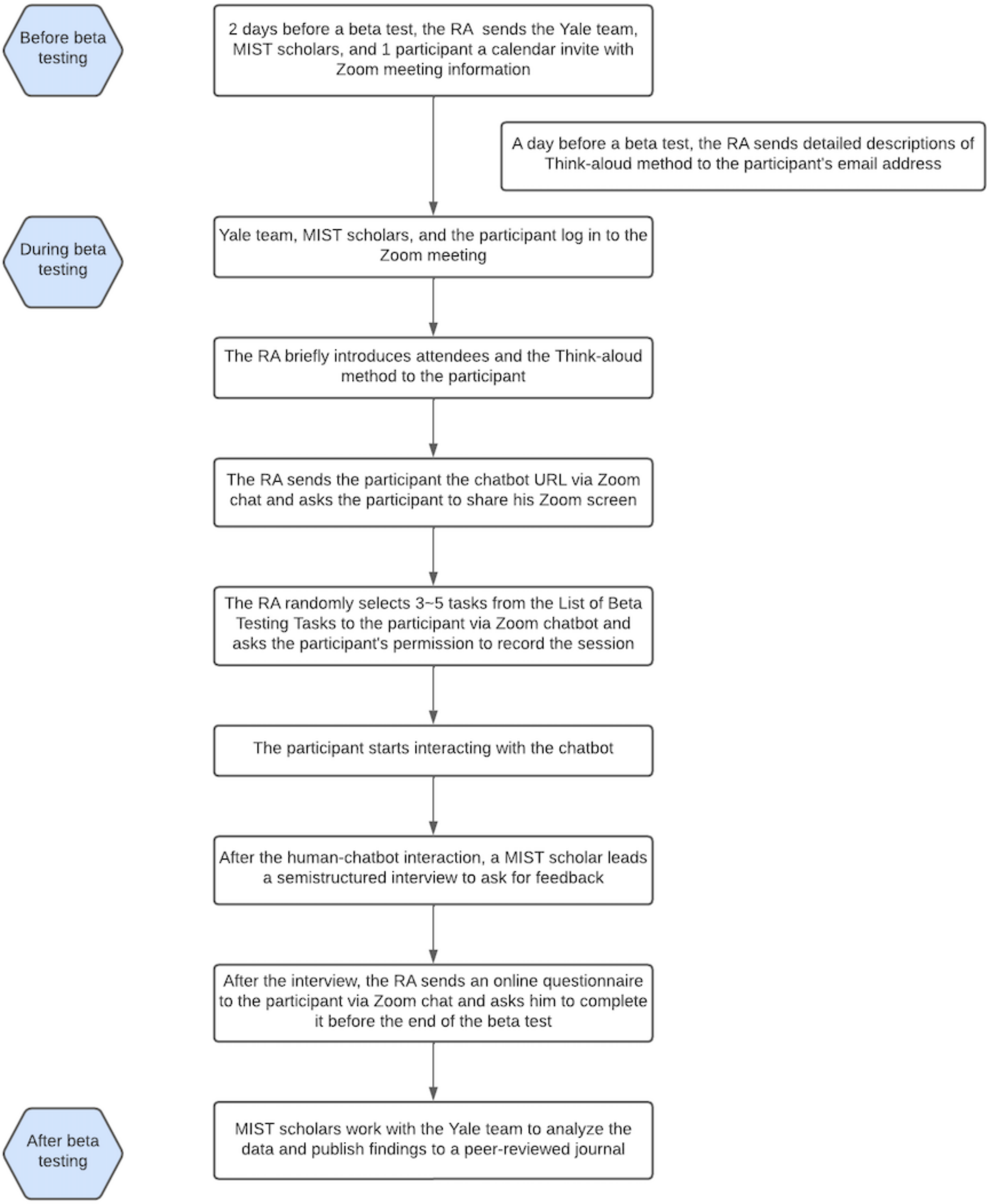
Methods

Study Design and Participants

Beta testing of the AI chatbot prototype was conducted with 14 MSM by an experienced qualitative interviewer (ZN) with expertise in chatbot development and HIV prevention in Malaysia and 4 trainees in the Malaysian Implementation Science Training program (Fogarty International Center, D43TW011324). Participants were recruited in Malaysia from February to April 2022 via social networking apps commonly used by MSM, including Grindr, Hornet, Blued, and WhatsApp. The procedures for participant recruitment have been published elsewhere [13]. A web-based screener including questions on demographic characteristics and HIV prevention practices was used. The eligibility criteria included (1) self-identification as a cisgender man, (2) age ≥18 years, (3) condomless sex with another man in the past 6 months, and (4) being HIV negative or of unknown status.

Each beta test involved the following three steps: (1) a 45-minute human-chatbot interaction using the think-aloud method [21]; (2) a 35-minute semistructured interview; and (3) a 10-minute web-based survey. The first 2 steps were conducted via Zoom (Zoom Video Communications, Inc), recorded, and transcribed verbatim. Specifically, 2 days before the test, a research assistant sent a calendar invite with Zoom meeting information to the interviewer and participant. One day before beta testing, the research assistant emailed the participant a detailed description of the human-chatbot interaction (Multimedia Appendix 1). During the human-chatbot interaction, participants were asked to share their screen via Zoom and access the chatbot through a URL sent by the research assistant. After the participants obtained access to the chatbot, the research assistant randomly selected 3 to 5 tasks from the list of beta testing tasks (Multimedia Appendix 2) and asked the participants to complete them through the chatbot. Some examples of the tasks include “find a clinic that can provide HIV testing service in Kuala Lumpur” and “find out the common symptoms of depression through the chatbot.” The study procedure is described in Figure 1.

Figure 1. Study procedure. MIST: Malaysian Implementation Science Training; RA: research assistant.



After the human-chatbot interaction, we conducted a semistructured interview (Multimedia Appendix 3 [22]) soliciting participants' feedback on two themes: (1) experience navigating the chatbot and (2) how the chatbot should be made available to a wider audience. During the interview, participants were asked several questions regarding their experience with the chatbot, such as "How was your experience with the AI chatbot?", "What feature of the AI chatbot do you like the most?", and "What information needs to be added to the AI chatbot to increase its popularity among MSM?" After the interviews concluded, the participants were provided with a survey link to assess the feasibility and acceptability of the AI chatbot. The feasibility of the chatbot was measured through 4 outcomes, including participants' ratings of the chatbot's quality, satisfaction, intention to continue using the chatbot, and willingness to refer it to others. The outcomes were measured using a 10-point rating scale, with higher scores indicating more favorable outcomes (Multimedia Appendix 4). For example, participants' satisfaction with the chatbot was measured by using the question, "How satisfied were you with the experience of interacting with the chatbot?" The score of "0" stands for not satisfactory at all and "10" stands for extremely satisfactory. The acceptability of the chatbot was measured using the standardized System Usability Scale [23] and an adjusted Chatbot Usability Scale [24]. The combination of the 2 scales provided a comprehensive evaluation of the acceptability of our chatbot.

Ethical Considerations

The participants provided electronic consent before initiating the beta testing. This study was approved by the institutional review board of Yale University (approval #2000027864) and Medical Research Ethics Committee of the University of Malaya (approval #2021112-10729). This research was conducted in accordance with the ethical standards of the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Conceptual Framework for Analysis

The Unified Theory of Acceptance and Use of Technology (UTAUT) was used as a conceptual framework to guide the analysis of the experience of MSM using the AI chatbot for HIV testing and prevention in Malaysia. UTAUT consists of four domains: (1) performance expectancy, (2) effort expectancy, (3) facilitating conditions, and (4) social influence [25]. The definitions of these 4 domains have been published elsewhere [13]. UTAUT was chosen for the following reasons. First, this AI chatbot was developed based on the findings from a formative research project that was analyzed using UTAUT [13]. Therefore, using the same theory, we can compare the results of the 2 studies on the 4 domains and are more likely to find out the feasibility and acceptability of the AI chatbot. Second, UTAUT emphasizes user-centered perspective, which allows researchers to assess the acceptance of the AI chatbot from the users' perception. Third, UTAUT has been extensively used to identify users' acceptance of technology and was reported to be effective and of high validity [26,27].

Analyses

All transcripts were cross-checked for accuracy and completeness by 7 researchers (MHC, YNG, NAMS, KSN, ZN, and 2 research assistants). Each of the 7 researchers independently coded 2 transcripts using NVivo 10 software (QSR International), compiled codes, and mapped the emerging themes from the qualitative data on the 4 domains of UTAUT, including performance expectancy, effort expectancy, facilitating conditions, and social influence. Discrepancies in codes and themes were addressed in group discussions where there was discordance in coding. We ceased the qualitative analysis when the results reached saturation, and no new themes emerged. The participants' quotes are presented throughout the results with additional quotes given in Multimedia Appendix 5. Quantitative data from the survey were analyzed using SAS (version 9.4; SAS Institute) and are presented as descriptive statistics.

Results

Participant Characteristics

The 14 participants were on average in their mid-20s (mean 25.6, SD 4.2 years), and most of them (13/14, 93%) used smartphones as the primary means to access the internet. Most participants (10/14, 71%) were Malay, followed by Chinese (3/14, 21%) and Indian (1/14, 7%). About one-third of the participants (5/14, 36%) had taken PrEP previously, and only 14% (2/14) of them were currently taking PrEP. The demographic characteristics are summarized in Table 1.

Table 1. Participant demographic details (N=14).

Characteristics	Values
Age (y), mean (SD)	25.6 (4.2)
Ethnicity, n (%)	
Malay	10 (71)
Chinese	3 (21)
Indian	1 (7)
Sexual orientation, n (%)	
Bisexual	3 (21)
Gay	11 (79)
Employment status, n (%)	
Student	6 (43)
Working full time	8 (57)
Highest level of education, n (%)	
Diploma or bachelor degree	8 (57)
Master degree or PhD	3 (21)
Secondary school	3 (21)
Average monthly income (MYR[a]; 1 MYR=US $0.21), n (%)	
<2000	6 (43)
2000-4000	5 (36)
>4000	3 (21)
Daily access to the internet, n (%)	
Yes	14 (100)
Primary device for accessing the internet, n (%)	
Smartphone	13 (93)
Laptop computer	1 (7)
Had ever taken PrEP[b], n (%)	
Yes	5 (36)
No	9 (64)
Currently taking PrEP, n (%)	
Yes	2 (14)
No	12 (86)

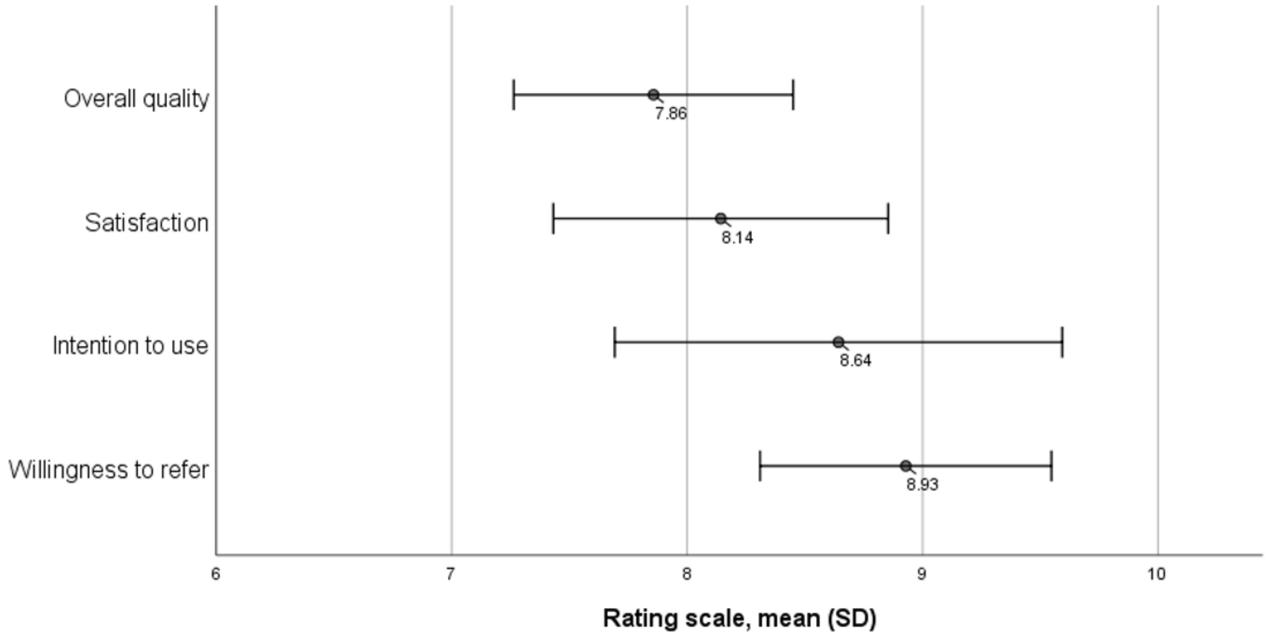
[a]MYR: Malaysian Ringgit.

[b]PrEP: pre-exposure prophylaxis.

Feasibility

The mean scores on the 4 metrics of the feasibility of the chatbot, overall quality, satisfaction, intention to continue using, and willingness to refer to others were 7.86 (SD 1.03), 8.14 (SD 1.23), 8.64 (SD 1.65), and 8.93 (SD 1.07), respectively, on a scale from 0 to 10 (Figure 2).

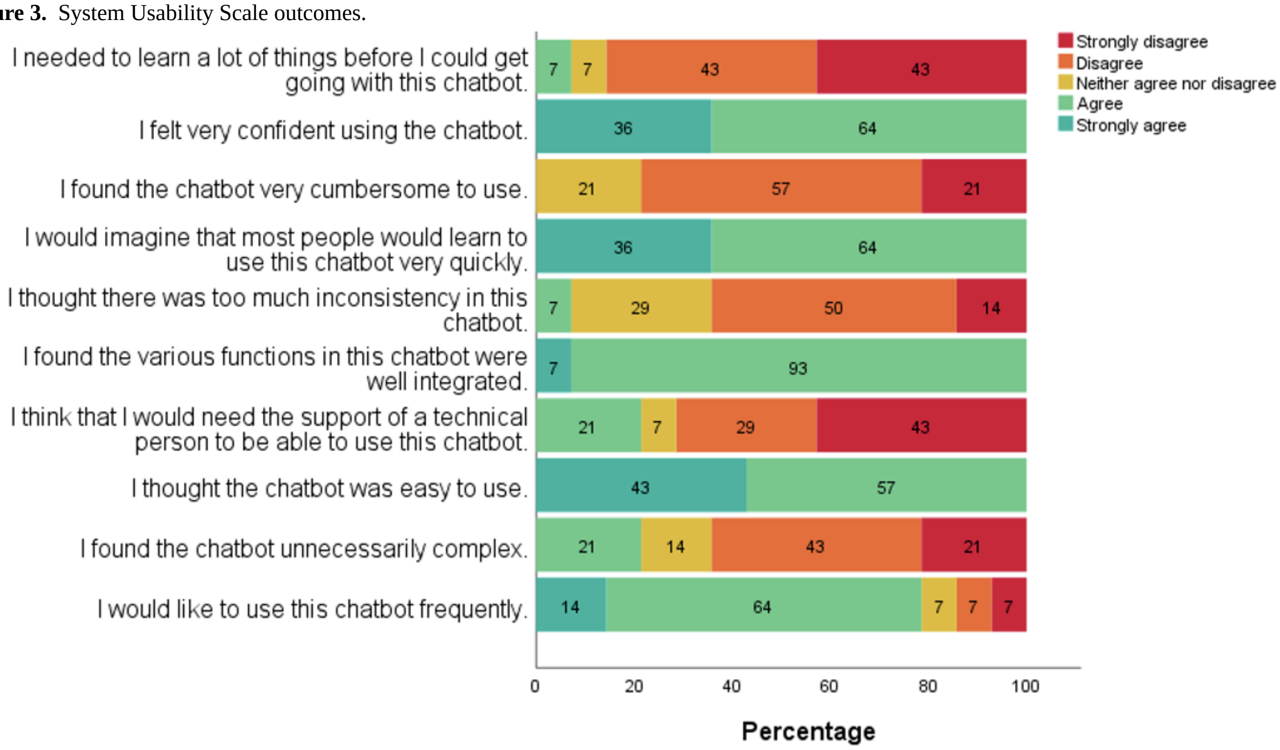
Figure 2. Feasibility ratings of the chatbot.



Acceptability

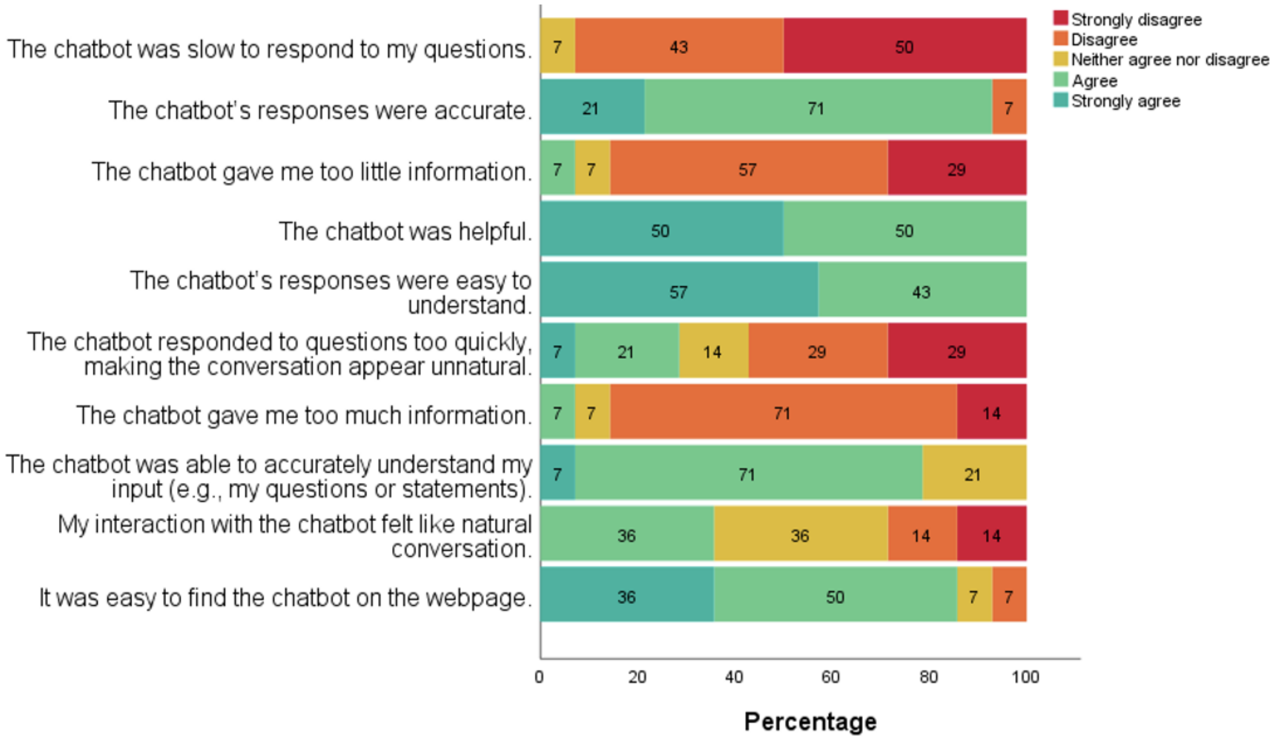
The participants found the chatbot acceptable, as it was perceived as easy to navigate and capable of providing valuable information (Multimedia Appendix 6). Specifically, all participants (14/14, 100%) expressed confidence in using the chatbot, believing that others could also quickly master its use (Figure 3). The overall mean (SD) score of the System Usability Scale was 76.07 (SD 8.19), which is greater than the recommended acceptable cutoff score of 68 [23].

Figure 3. System Usability Scale outcomes.



Moreover, 78% (11/14) of the participants agreed that the chatbot could understand their inputs accurately (Figure 4). However, only 36% (5/14) of the participants agreed that their interaction with the chatbot felt like a natural conversation.

Figure 4. Adjusted Chatbot Usability Scale outcomes.



In addition to measuring feasibility and acceptability, we summarized the study findings based on the 4 domains of UTAUT: performance expectancy, effort expectancy, facilitating conditions, and social influence.

Performance Expectancy

Overall Perception

Participants responded positively about the performance of the AI chatbot. Quantitative data analysis revealed that all participants (14/14, 100%) perceived the chatbot as a helpful tool, they would refer it to others, and 93% (13/14) of them highly rated the overall quality of the chatbot. These results were consistent with the qualitative finding that participants were satisfied with the information provided by the chatbot and deemed it a trustable source.

Contributors to Positive Performance Expectancy

The chatbot served as a reliable source of information. Participants expressed their trust in this chatbot as they believed it was developed to help people learn more about HIV. For instance, one participant stated that he trusted the chatbot more than other internet-based platforms as the following:

> *I can actually trust, trust this chatbot more than I can trust the Internet.* [Interview E]

Most participants expressed that the information provided by the chatbot was comprehensive and satisfactory. The same participant highlighted the following:

> *...everything is there, everything is informative...* [Interview E]

Participants described the chatbot's function of ordering free HIV self-test kits as one of the most useful functions. The simple and straightforward instructions from the chatbot significantly encouraged participants to perform HIV self-testing and helped prevent misuse of the self-test kit. A participant elaborated on the following:

> *...it never crossed my mind that you can do HIV self-test just using it (test stick) over the gums and without blood...* [Interview G]

In addition to the positive feedback on the chatbot's ability to order free HIV self-test kits, participants also expressed appreciation for the information provided by the chatbot on MSM-friendly local clinics where they could test for HIV and receive PrEP consultation. For example, the participant stated the following:

> *That was beyond amazing...they give [me] the addresses, contact numbers. So, I would say, if a person really needs to do the [HIV] testing, essential information like that would be very useful, so I think it's more than helpful. The options [of HIV testing clinics] are not just [limited to] one or two [clinics], you know, so that's good.* [Interview G]

The other participants stressed that the chatbot's features related to HIV self-testing and venue-based HIV testing were complementary and appreciated having access to both options through the chatbot. Although performing self-testing at home may be convenient, it may lack the human interaction that some MSM need for support during the testing process. By offering both self-testing and venue-based testing options, the chatbot gave MSM the flexibility to choose the option that best suited their individual needs and preferences. For example, a participant emphasized that the chatbot could enable his MSM friends to conduct HIV self-testing and then see a health care professional for advice on complex and sensitivity issues:

> *...the reason why [they] went to a health care facility was [that they] can have someone tell them that you know, "Being tested positive for HIV [was] not the end of the world. To reduce the [HIV] symptoms, to reduce the [HIV] effect, to reduce [their HIV] viral load... [they] can still live a normal life and so on," which might be something that the self-test kits [were] lacking.* [Interview C]

Along with the positive feedback on the chatbot's features relevant to HIV testing, the participants also expressed their favorable feelings toward the PrEP information provided by the chatbot. For example, a participant mentioned that the chatbot could send him introductory information to allow him to comfortably assess his risk level and help him decide if he needed to take PrEP. Another participant reiterated the first participant's point by emphasizing the difficulties that MSM in determining if they should take PrEP, stating the following:

> *A lot of people [MSM] are always asking themselves, "Should I get PrEP? Am I at risk? Do I need to take PrEP as a precautionary measure?"... So these are [questions] that MSM usually a bit too scared to ask the doctors.* [Interview C]

Major Concerns

Participants suggested that the chatbot would be improved if it could provide more information and resources relevant to mental health, as mental health issues were the prominent problems that MSM faced in Malaysia. Participants wanted the chatbot to provide information on strategies for managing stress, statistics about depression among MSM, peer consultation for depression, and professional health care services to prevent and treat depression. The participants also highlighted that the MYHIV365 website, where the chatbot was embedded, should provide more resources related to mental health. For example, one participant described this problem as follows:

> *The website did not have links to any information regarding mental health issues, and that is a glaring issue for it to be left out like that.* [Interview A]

The same statement was echoed by another participant, who stressed the following:

> *I just find that for the mental health, it's kind of short.* [Interview G]

Relevant Features Suggested by Participants That Are Needed to Improve the Chatbot

Although the chatbot made it easy for participants to receive an HIV self-test kit, one participant suggested that a step-by-step video demonstrating how to use the kit could be helpful for MSM who were testing for HIV. The participant stated the following:

> *If the self-test kit has got like an instructional video or something like that to kind of guide the users along the way of getting [themselves] tested, I think that'd be great because not everyone knows how to use a kit successfully.* [Interview C]

Although the AI chatbot was developed primarily for HIV prevention and to assist with HIV testing and access to PrEP, the participants pointed out that some participants may test positive for HIV and would benefit from learning more about accessing HIV care and related antiretroviral therapy services. In addition, participants suggested that the chatbot should provide more information about antiretroviral therapy so that users could better manage HIV by knowing potential drug interactions and side effects. The participants also recommended providing more information about high-risk behaviors and sexually transmitted diseases to help increase awareness about HIV and sexually transmitted diseases among MSM. One participant stated the following:

> *I think [providing more information about] HIV treatment would [be] very helpful because those who might be exposed to HIV would definitely want to know what the treatment is all about.* [Interview E]

This participant's statement was echoed by another participant, who stated the following:

> *...HIV and STDs...[are] not the same, but...I thought [they were] the same...I thought HIV and STDs were not curable...so I think it will be great if you can add STDs [to the chatbot].* [Interview E]

Effort Expectancy

Overall Perception

All surveyed participants (14/14, 100%) agreed that the chatbot was easy to use, and 86% (12/14) of the participants were satisfied with the chatbot. In the qualitative interviews, participants reported consistent feedback that the chatbot was user friendly and convenient to use, and they were satisfied with the chatbot because of its simple, straightforward design and quick, friendly responses. However, they were concerned about the technical issues, including the address input and text alignments (refer to the *Major Concerns* section). The participants also felt that tailoring the chatbot to the local context and adding a "human touch" would be helpful.

Positive Contributors to Low Effort Expectancy

Many participants expressed satisfaction with the chatbot because of its prompt response, expert information, and plain interface. Two participants commented the following:

> *...white and blue colors [are] neutral, and it [the chatbot] takes into account [of] color blindness as well, so that's great.* [Interview C]
>
> *...[I] got a quick response [from the chatbot].* [Interview I]

The individualized and user-centered features of the chatbot, which cater to users with different levels of communication skills, were highlighted among the participants. For instance, one participant stated that the chatbot offered an ideal platform for MSM who are less comfortable interacting with others. A participant stated the following:

> *As we all know, some of us didn't have the skills to communicate, so I think...[the] chatbot… will definitely help. I think it was great.* [Interview E]

Moreover, participants thought the chatbot was useful as it facilitated them to obtain culturally tailored health information. The chatbot met users' needs by providing a menu of options for users to choose from. Compared with obtaining health information in clinical settings in Malaysia, the chatbot was much simpler. A participant elaborated on the following:

> *When [the chatbot] come[s] up with three options like that, I can explore myself...I would say that [the chatbot] is more precise; it gets to the point directly.* [Interview G]

The health intervention being tailored to the local setting was highly valued by the participants. Responses from the AI chatbot that contained localized features, specifically the use of "Manglish," a less formal form of Malaysian English, were appreciated by several participants. The feature of "Manglish," which was not in the standardized form of English, has added a local flavor to the AI chatbot, which some participants found amusing. A participant stated the following:

> *The impression that this chatbot...probably comes from America. It's in English, so the moment it puts up a Malaysian style saying "Boleh"... I'm very amused with this [style].* [Interview G]

Major Concerns

Some participants spoke about the difficulty in filling in their home addresses using the current prompts on the AI chatbot when they needed to order an HIV self-test kit; the chatbot required a step-by-step input of addresses, which was counterintuitive and inefficient. Participants preferred the standard address format in Malaysia over the step-by-step input format, in which incomplete addresses would triage further prompts to ask participants to refill the HIV self-test order. For example, one participant stated the following:

> *In Malaysia, we don't use the term "line address" or "street address". We usually enter the full address with the postcode and then the city and state. The one on the chatbot seems to be how addresses are filled in the United States. That part needs to be tweaked slightly based on Malaysian cities.* [Interview C]

In addition, participants expressed that the address of the clinics provided by the chatbot needed to be tailored to Malaysian culture. For example, the district options may only be needed for certain states in Malaysia. A participant stated the following:

> *I think depends on the size of the state...we don't have to call out (provide choices for) all the districts because Perlis is already small enough, and I think...people can just go easily from one place to another in Perlis. But if...it is a big state...we need to divide it using district.* [Interview I]

Relevant Features Suggested by Participants That Are Needed to Improve the Chatbot

Although participants were satisfied with the AI chatbot, 2 participants suggested that the chatbot's interface could be improved by adding more spaces between sentences, and the alignment of sentences should be adjusted to make the chatbot look more professional. Two participants described the following:

> *...everything is tightly together with very little gap...there should be proper spacing...* [Interview L]

> *The text is not properly centered in some of the boxes, and I feel like it could [be] a better design to make it look more professional.* [Interview A]

The use of English as the only language of the AI chatbot was perceived by participants as a barrier to implementing the chatbot in Malaysia. Although all participants were proficient in English, concerns arose for the communities where English was not widely spoken. Participants suggested that the chatbot should be able to communicate in Bahasa Malaysia or Mandarin, given that the 2 languages are widely spoken in Malaysia. A participant stated the following:

> *...perhaps to have another option of language...I think that would be able to cover more people within the local population.* [Interview C]

Adding a "human touch" to the chatbot can create a more engaging and user-friendly experience for the users interacting with the chatbot. The participant described the following:

> *...ideally, we would want [the chatbot's response] to be as human as possible, and not so robotic in its responses...a nice touch to make someone feel slightly comfortable.* [Interview C]

Facilitating Conditions

Overall Perception

Participants reported 2 major facilitating conditions for the use of the AI chatbot. First, the social distancing policy adopted by the Malaysian government during the COVID-19 pandemic significantly increased the use of internet-based platforms to seek health information and consult about health issues among MSM. The participants expressed that the AI chatbot was a novel tool to promote HIV testing and prevention among MSM in Malaysia. A participant highlighted the convenience of using the chatbot as an alternative to meeting health care workers during the COVID-19 pandemic as follows:

> *...because now it's COVID, everyone is doing it in IT (information technology) format. Having an AI chatbot is definitely much more convenient than meeting people...* [Interview G]

Second, the AI chatbot's capability of referring webpages to participants where they could find mental health information, community support, and counselors was a significant facilitator for them to accept the chatbot. Many participants stated that it was much easier to obtain information through the links provided by the chatbot than searching for information via websites or mobile apps. A participant stated the following:

> *When you interact with it (the chatbot), it throws out links to you. It's easier to navigate to the particular links from there.* [Interview G]

Relevant Features Suggested by Participants That Are Needed to Improve the Chatbot

Participants suggested that the chatbot could be promoted through social media platforms, such as Facebook, Twitter, Instagram, YouTube, TikTok, and Telegram because these platforms were widely used by MSM as sources of information. Among all social networking apps, participants stated that Twitter was the best platform to advertise the AI chatbot because Twitter enabled users to post clickable links in the comments section where other users could access the chatbot. Participants further reported that Telegram was a more suitable platform for hosting the chatbot than the most popular text messaging app in Malaysia, WhatsApp. Telegram offers a more private and secure environment for MSM to ask questions or express concerns about HIV and AIDS. Participants also suggested that building a trustable relationship between the AI chatbot and the MSM community is key to implementing the AI chatbot in Malaysia. Given that there were many scams through pop-up advertisements on social media platforms, a participant described the following imagined scenario:

> *...if we play our Facebook, Twitter, Instagram, or YouTube, there are always mini advertisements, so who knows, [whether we] can add this [AI chatbot]?...I need to know about this [chatbot], and I hope this [chatbot] is not a scam.* [Interview B]

Social Influence

In terms of *social influence*, the chatbot was perceived as helpful in avoiding HIV stigma and thus could increase the HIV testing rate and PrEP uptake frequency. Quantitative data analysis found that 79% (11/14) of the participants agreed to continue using the chatbot. During the interviews, these participants reported that societal stigma and discrimination related to HIV and AIDS would make them more likely to use the chatbot. They expressed discomfort in asking people questions about HIV and AIDS as they were afraid of encountering stigma and negative attitudes from others. MSM often preferred to seek information through internet-based platforms, and the chatbot was helpful, particularly for people living in small social circles. A participant elaborated on the following:

> *...this topic [HIV] is quite sensitive to most people, it will create like a negative energy around you...personally I don't go and ask people what HIV is, I will search myself maybe on the Internet...* [Interview E]

The societal stigma and discrimination toward HIV and AIDS also facilitated participants to select HIV self-testing at home rather than testing in a clinical setting. Many participants appreciated that the chatbot offered them an opportunity to receive free HIV self-test kits while protecting their privacy. Two participants who used to be shy about discussing HIV described the following:

> *Because from the MSM community, some of us are not very comfortable of getting [HIV] test kits on site, because like...fear of the stigma, that the society will judge.* [Interview D]

> *I can directly book the test kit through the chatbot, which is very useful and informative…my identity will remain anonymous, so people don't know me.* [Interview E]

Participants deemed the AI chatbot useful and expressed their willingness to recommend the AI chatbot to others. Some participants suggested that the chatbot should be promoted among MSM who frequently use social networking apps, such as Grindr, Hornet, and Blued, to find sexual partners because those MSM were at higher risk for HIV and had greater need for HIV information. A participant stated the following:

> *I have the impression that anyone would actually need it [HIV testing]. But if we look at it from another angle, people on hookup apps like Grindr have a high tendency to hook up using those apps compared to those who don't use them...we need to introduce the chatbot to them because...they...[have] been highly exposed [to HIV].* [Interview G]

Discussion

Principal Findings

The feasibility and acceptability of leveraging AI chatbots to promote HIV testing and PrEP among MSM in Malaysia is high. Discrimination and stigma toward HIV and AIDS are major barriers for MSM to access high-quality HIV testing and prevention services in Malaysia, and they are also primary facilitators for MSM to seek health information via internet-based platforms. Our AI chatbot prototype provides a platform for MSM to order free HIV self-test kits in an MSM-friendly environment and to empower them with resources and instructions. MSM who prefer to interact with health care providers in person can also locate HIV testing clinics or PrEP clinics through the AI chatbot. MSM highlighted these functions of the AI chatbot as very useful.

Similar to other studies, AI chatbots were well received by users [28,29]. An AI chatbot could enhance engagement with the key population [30]. As contemporary social patterns increasingly involve the integration of AI into everyday routines, AI chatbots could contribute to delivering precise details regarding HIV testing to individuals actively seeking such information. A chatbot named Eli, developed by the United Nations Educational, Scientific, and Cultural Organization, received highly favorable user feedback and was widely acclaimed [29]. Eli offers a range of services, including details on HIV prevention, testing, and treatment and assistance in overcoming fears and concerns. Compared with Eli, our AI chatbot did not have information on treatment for AIDS and provided limited mental health support. Integrating these functions into our AI chatbot may support its usability. Nevertheless, our AI chatbot offers free HIV self-test kits and locates local clinics in Malaysia for HIV testing, PrEP consultation, and mental health care.

From our previous formative research, we know that factors facilitating the acceptance of an HIV prevention AI chatbot include providing useful information and having the capacity to solve problems [13]. In this study, participants reported that our AI chatbot was able to provide useful information and help solve problems. This was indicated by the results that all participants perceived this chatbot as a helpful tool, and most participants deemed the chatbot a reliable source of information with a high satisfaction score. However, one area that required significant improvement in the chatbot was its conversation flow, as only 36% (5/14) of the participants felt that their interaction with the chatbot resembled a natural conversation. This was similar to another study where the quick response of the chatbot was deemed not humanlike and perceived as a disadvantage [28]. To address this issue and advance the chatbot, improving its algorithm and continuing training it using AI and machine learning techniques based on feedback from a larger sample size is crucial. Considering that the use of AI chatbots in health care is still in its early stages, this finding holds particular significance for designing AI chatbots. To enhance usability and promote the implementation of AI chatbots in health care, the chatbots must possess the ability to initiate natural conversations with humanlike characteristics. In addition, they should be equipped to effectively address users' questions and concerns while ensuring the security and safety of users' information.

The chatbot's plain interface and simple design were popular among MSM. Digital health interventions are useful, but knowing how to navigate a digital system sometimes could be daunting for users. Through this study, we are clear that accurate and simple responses without errors and redundant information were key to the acceptability of AI chatbots among MSM. Our participants reported that the AI chatbot helped them avoid societal stigma and protected their privacy, which increased their acceptability of using the chatbot to test for HIV. This finding is consistent with our previous formative research finding that addressing sociocultural barriers can facilitate the acceptance of an AI chatbot [13]. The chatbot does not require users to provide registration information. Therefore, it can maintain participants' anonymity. However, it is still necessary for the chatbot to clarify to users that the backend researchers and engineers who have access to users' conversations and information will not expose users' information to others. This suggestion is consistent with our study findings and some previous studies showing that mHealth interventions could improve HIV testing rates if users' anonymity were guaranteed [14,17].

Some technical-related issues negatively affected the participants' experience of navigating the chatbot. The inconvenient address input process and repeated steps owing to incomplete information contributed to the inconsistency and complexity of the chatbot, prompting participants to seek technical assistance. Many of these resulted from cultural differences, as the address options were designed based on overseas settings. This signifies the importance of tailoring the chatbot to the local context to improve usability. In addition, using localized language could also enhance the participants' satisfaction with the chatbot. Despite the challenges inherent in adopting novel technology, the advantages of using chatbots to connect with high-risk populations could significantly impact the efforts to address public health emergencies.

In line with other studies conducted in Malaysia, MSM are keen to peruse the information on PrEP and mental health, particularly the information on where the PrEP and mental health clinics are located [14]. Most participants in our study felt that they would like to see more information through the chatbot introducing AIDS, its treatment, mental health issues, and sexually transmitted infections to better understand and manage AIDS, including how to prevent high-risk behaviors and where to seek timely help [12]. In Malaysia, professional and MSM-friendly care for mental health needs to be developed as most MSM reported that culturally sensitive information and resources regarding mental health issues were difficult to obtain. Interestingly, researchers have identified several obstacles to the adoption of AI chatbots for mental health care among users [31]. These include concerns related to privacy risks, restricted conversational engagement, negative user perceptions of personality traits (such as rudeness, lack of empathy, patronization, and being judgmental), and a lack of trust in the app's creators. Nevertheless, Eli chatbot overcame all these challenges by having a language that merges expertise and respect for the user, ensuring speech that is gender neutral and devoid of stigmatizing elements [29]. Our AI chatbot also had a similar language as Eli, which warrants future support on mental health issues.

To increase the use of the AI chatbot, it needs to be embedded in social media platforms that MSM frequently use. The geosocial networking apps where MSM find sexual partners, such as Grindr, Hornet, and Blued, and websites owned by nongovernmental organizations and MSM-friendly clinics are important venues to advertise the AI chatbot. MSM preferred these platforms because they are trusted and frequently used by MSM. Dissemination of the AI chatbot should be promoted among young MSM who use geosocial networking apps to find sexual partners because they are at a higher risk for HIV. Through this study, we found that to embed an AI chatbot into an internet-based platform for health promotion, researchers and engineers must consider the platform's characteristics, including its target population, level of privacy, and user-friendliness. Findings from this study will be used to improve the AI chatbot before testing on a larger scale through a national observational study in Malaysia. AI chatbots are a promising tool for promoting HIV testing and prevention. The AI chatbot must be made visible to MSM to increase its usability among MSM. Adopting the right dissemination strategies is key to increasing the visibility of AI chatbots and bringing significant impact to the MSM community. In addition, it is important for researchers to consider the sustainability of AI chatbots for MSM care in a context where sex-same sexual behaviors are criminalized. The policies and laws in Malaysia pose significant challenges on the sustainability of leveraging AI and securing funding for MSM care research. In such a political environment, it is crucial for researchers to collaborate with local nongovernmental organization and MSM-friendly clinics that operate within the existing Malaysian legal framework. Future research should focus on developing innovative and culturally tailored AI interventions to combat HIV among MSM, promote public health in Malaysia, and advocate for changes in discriminatory policies and laws to enhance the testing, implementation, and sustainability of these AI interventions.

Limitations

Testing the AI chatbot among its end users (ie, MSM) was an important step in determining its feasibility and acceptability in Malaysia and collecting feedback to improve the chatbot further. Although this study contributed important scientific knowledge, it had several limitations. One of the limitations is that we only included MSM who can read English, as the AI chatbot is currently only available in this language. Thus, the reach of the AI chatbot may be limited only to those fluent in English, which is not the case for most MSM in Malaysia. Therefore, our findings may not be generalizable to MSM who cannot read English. Considering that Malaysia is a multilingual country with Bahasa Malaysia as the official language, the chatbot must be improved to communicate in Bahasa Malaysia or Mandarin to reach a wider audience and promote greater access to HIV self-testing and PrEP. In addition, our participants were highly educated; this may lead to bias as they might possess a certain level of knowledge and health literacy, thus facilitating their interactions with the chatbot. Therefore, the findings may differ in the less educated or literate group. In addition, our study only included MSM aged ≥18 years; therefore, the study findings do not capture the perceptions of younger MSM who are typically more tech-savvy and susceptible to HIV. Although obtaining consent from younger MSM in Malaysia for HIV-related research is a significant challenge, future studies should consider conducting surveys and interviews with MSM aged <18 years who can provide insights into the experiences and needs of the younger MSM.

Conclusions

The AI chatbot was found to be feasible and acceptable among MSM, highlighting features, such as being informative, being able to respond to users' questions, and having a simple and user-friendly interface. Adapting the AI chatbot to local cultures, including support for other languages, and providing additional information such as mental health support, risk assessment for sexually transmitted infections, AIDS treatment, and the consequences of contracting HIV would contribute to the successful implementation and dissemination of the AI chatbot in Malaysia.

Acknowledgments

This work was supported by research grants from the National Institutes of Health (R21TW011663 and R33TW011663 for ZN and FLA; R21AI152927 for JAW, FLA, and ZN; and D43TW011324 for FLA) and the Fund for Lesbian and Gay Studies Research Award from Yale University for ZN.

Authors' Contributions

MHC and ZN wrote the first draft of the manuscript. YNG, NAMS, and KSN assisted in both developing the qualitative analysis and writing the qualitative results. All authors contributed to the development of the study protocol, revised the subsequent version of the manuscript, and approved the submitted version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
An email that the research assistant sent to participants introducing the human-chatbot interaction.
[PDF File (Adobe PDF File), 64 KB - humanfactors_v11i1e52055_app1.pdf]

Multimedia Appendix 2
The list of beta testing tasks.
[PDF File (Adobe PDF File), 478 KB - humanfactors_v11i1e52055_app2.pdf]

Multimedia Appendix 3
The guide on chatbot beta testing.
[PDF File (Adobe PDF File), 336 KB - humanfactors_v11i1e52055_app3.pdf]

Multimedia Appendix 4
The scales used for measuring outcome variables.
[PDF File (Adobe PDF File), 195 KB - humanfactors_v11i1e52055_app4.pdf]

Multimedia Appendix 5
Participants' insights with illustrative quotes.
[DOCX File , 32 KB - humanfactors_v11i1e52055_app5.docx]

Multimedia Appendix 6

Screenshots demonstrating the chatbot's simple conversation flow.
[PDF File (Adobe PDF File), 958 KB - humanfactors_v11i1e52055_app6.pdf]

References

1. HIV and AIDS. World Health Organization. 2022. URL: https://www.who.int/news-room/fact-sheets/detail/hiv-aids [accessed 2023-12-26]
2. New HIV infections increasingly among key populations. Joint United Nations Programme on HIV/Aids. 2020. URL: https://www.unaids.org/en/resources/presscentre/featurestories/2020/september/20200928_new-hiv-infections-increasingly-among-key-populations [accessed 2023-12-26]
3. Seizing the moment: tackling entrenched inequalities to end epidemics. Joint United Nations Programme on HIV/Aids. 2020. URL: http://tinyurl.com/4zyedprp
4. Suleiman A, Tze CP. Country progress report on HIV/AIDS 2019 Malaysia. HIV/STI/Hepatitis, C Section of Ministry of Health, Malaysia. 2019. URL: https://www.moh.gov.my/moh/resources/Penerbitan/Laporan/Umum/Report_GAM_2019_(Final).pdf [accessed 2023-12-22]
5. Global AIDS monitoring: country progress report-Malaysia. Ministry of Health, Malaysia. 2021. URL: https://www.moh.gov.my/moh/resources/Penerbitan/Laporan/Umum/MYS_country_report_2021.pdf [accessed 2023-12-26]
6. Montaner JS, Lima VD, Harrigan PR, Lourenço L, Yip B, Nosyk B, et al. Expansion of HAART coverage is associated with sustained decreases in HIV/AIDS morbidity, mortality and HIV transmission: the "HIV treatment as prevention" experience in a Canadian setting. PLoS One 2014;9(2):e87872 [FREE Full text] [doi: 10.1371/journal.pone.0087872] [Medline: 24533061]
7. Gardner EM, McLees MP, Steiner JF, Del Rio C, Burman WJ. The spectrum of engagement in HIV care and its relevance to test-and-treat strategies for prevention of HIV infection. Clin Infect Dis 2011 Mar 15;52(6):793-800 [FREE Full text] [doi: 10.1093/cid/ciq243] [Medline: 21367734]
8. Risher K, Mayer KH, Beyrer C. HIV treatment cascade in MSM, people who inject drugs, and sex workers. Curr Opin HIV AIDS 2015 Nov;10(6):420-429 [FREE Full text] [doi: 10.1097/COH.0000000000000200] [Medline: 26352393]
9. Shrestha R, Wickersham JA, Khati A, Azwa I, Ni Z, Kamarulzaman A, et al. Clinic-integrated mobile health intervention ("JomPrEP" App) to improve uptake of HIV testing and pre-exposure prophylaxis among men who have sex with men in Malaysia: protocol for an intervention development and multiphase trial. JMIR Res Protoc 2022 Dec 21;11(12):e43318 [FREE Full text] [doi: 10.2196/43318] [Medline: 36542425]
10. Beyrer C, Sullivan PS, Sanchez J, Dowdy D, Altman D, Trapence G, et al. A call to action for comprehensive HIV services for men who have sex with men. Lancet 2012 Jul 28;380(9839):424-438 [FREE Full text] [doi: 10.1016/S0140-6736(12)61022-8] [Medline: 22819663]
11. Ni Z, Wu B, Yang Q, Yan LL, Liu C, Shaw RJ. An mHealth intervention to improve medication adherence and health outcomes among patients with coronary heart disease: randomized controlled trial. J Med Internet Res 2022 Mar 09;24(3):e27202 [FREE Full text] [doi: 10.2196/27202] [Medline: 35262490]
12. Aliabadi N, Carballo-Dieguez A, Bakken S, Rojas M, Brown 3rd W, Carry M, et al. Using the information-motivation-behavioral skills model to guide the development of an HIV prevention smartphone application for high-risk MSM. AIDS Educ Prev 2015 Dec;27(6):522-537 [FREE Full text] [doi: 10.1521/aeap.2015.27.6.522] [Medline: 26595265]
13. Peng ML, Wickersham JA, Altice FL, Shrestha R, Azwa I, Zhou X, et al. Formative evaluation of the acceptance of HIV prevention artificial intelligence chatbots by men who have sex with men in Malaysia: focus group study. JMIR Form Res 2022 Oct 06;6(10):e42055 [FREE Full text] [doi: 10.2196/42055] [Medline: 36201390]
14. Shrestha R, Lim SH, Altice FL, Copenhaver M, Wickersham JA, Saifi R, et al. Use of smartphone to seek sexual health information online among Malaysian men who have sex with men (MSM): implications for mHealth intervention to increase HIV testing and reduce HIV risks. J Community Health 2020 Feb;45(1):10-19 [FREE Full text] [doi: 10.1007/s10900-019-00713-x] [Medline: 31375976]
15. Marcus JL, Sewell WC, Balzer LB, Krakower DS. Artificial intelligence and machine learning for HIV prevention: emerging approaches to ending the epidemic. Curr HIV/AIDS Rep 2020 Jun;17(3):171-179 [FREE Full text] [doi: 10.1007/s11904-020-00490-6] [Medline: 32347446]
16. Shrestha N, Kukkonen-Harjula KT, Verbeek JH, Ijaz S, Hermans V, Pedisic Z. Workplace interventions for reducing sitting at work. Cochrane Database Syst Rev 2018 Jun 20;6(6):CD010912 [FREE Full text] [doi: 10.1002/14651858.CD010912.pub4] [Medline: 29926475]
17. Shrestha R, Maviglia F, Altice FL, DiDomizio E, Khati A, Mistler C, et al. Mobile health technology use and the acceptability of an mHealth platform for HIV prevention among men who have sex with men in Malaysia: cross-sectional respondent-driven sampling survey. J Med Internet Res 2022 Jul 25;24(7):e36917 [FREE Full text] [doi: 10.2196/36917] [Medline: 35877172]
18. Rosen AO, Wickersham JA, Altice FL, Khati A, Luces JR, Ni Z, et al. Adaptation of a theory-based, clinic-affiliated smartphone app to improve HIV testing and pre-exposure prophylaxis uptake among gay, bisexual, and other men who have sex with men in Malaysia. iproc 2022 Jul 6;8(1):e39358 [FREE Full text] [doi: 10.2196/39358]

19. Caldarini G, Jaf S, McGarry K. A literature survey of recent advances in chatbots. Information 2022 Jan 15;13(1):41 [FREE Full text] [doi: 10.3390/info13010041]
20. Ahlström MG, Ronit A, Omland LH, Vedel S, Obel N. Algorithmic prediction of HIV status using nation-wide electronic registry data. EClinicalMedicine 2019 Dec;17:100203 [FREE Full text] [doi: 10.1016/j.eclinm.2019.10.016] [Medline: 31891137]
21. Eccles DW, Arsal G. The think aloud method: what is it and how do I use it? Qual Res Sport Exerc Health 2017 May 26;9(4):514-531 [FREE Full text] [doi: 10.1080/2159676x.2017.1331501]
22. Nick B. Everything you need to know about beta testing. Adobe. 2019. URL: https://xd.adobe.com/ideas/process/user-testing/everything-you-need-to-know-about-betatesting/ [accessed 2023-12-26]
23. System usability scale (SUS). U.S. General Services Administration. URL: https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html [accessed 2023-09-22]
24. Borsci S, Malizia A, Schmettow M, van der Velde F, Tariverdiyeva G, Balaji D, et al. The chatbot usability scale: the design and pilot of a usability scale for interaction with AI-based conversational agents. Pers Ubiquit Comput 2021 Jul 21;26(1):95-119 [FREE Full text] [doi: 10.1007/s00779-021-01582-9]
25. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478 [FREE Full text] [doi: 10.2307/30036540]
26. Venkatesh V, Thong JY, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157-178 [FREE Full text] [doi: 10.2307/41410412]
27. Gupta B, Dasgupta S, Gupta A. Adoption of ICT in a government organization in a developing country: an empirical study. J Strateg Inf Syst 2008 Jun;17(2):140-154 [FREE Full text] [doi: 10.1016/j.jsis.2007.12.004]
28. Ntinga X, Musiello F, Keter AK, Barnabas R, van Heerden A. The feasibility and acceptability of an mHealth conversational agent designed to support HIV self-testing in south Africa: cross-sectional study. J Med Internet Res 2022 Dec 12;24(12):e39816 [FREE Full text] [doi: 10.2196/39816] [Medline: 36508248]
29. Chatbot answers young people's questions about HIV, health and relationships. Joint United Nations Programme on HIV/AIDS. URL: https://www.unaids.org/en/resources/presscentre/featurestories/2020/october/20201015_chatbot [accessed 2023-08-15]
30. Garett R, Young SD. Potential application of conversational agents in HIV testing uptake among high-risk populations. J Public Health (Oxf) 2023 Mar 14;45(1):189-192 [FREE Full text] [doi: 10.1093/pubmed/fdac020] [Medline: 35211740]
31. Prakash AV, Das S. Intelligent conversational agents in mental healthcare services: a thematic analysis of user perceptions. Pac Asia J Assoc Inf Syst 2020 Jun 30;12:1-34 [FREE Full text] [doi: 10.17705/1thci.12201]

Abbreviations

AI: artificial intelligence
mHealth: mobile health
MSM: men who have sex with men
PrEP: pre-exposure prophylaxis
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by A Kushniruk, E Borycki; submitted 21.08.23; peer-reviewed by S Comulada, S John, R Saifi; comments to author 19.09.23; revised version received 22.09.23; accepted 12.12.23; published 26.01.24.

Please cite as:
Cheah MH, Gan YN, Altice FL, Wickersham JA, Shrestha R, Salleh NAM, Ng KS, Azwa I, Balakrishnan V, Kamarulzaman A, Ni Z
Testing the Feasibility and Acceptability of Using an Artificial Intelligence Chatbot to Promote HIV Testing and Pre-Exposure Prophylaxis in Malaysia: Mixed Methods Study
JMIR Hum Factors 2024;11:e52055
URL: https://humanfactors.jmir.org/2024/1/e52055
doi:10.2196/52055
PMID:38277206

Viewpoint

The Temperature Feature of ChatGPT: Modifying Creativity for Clinical Research

Joshua Davis[1,2*], BS; Liesbet Van Bulck[3,4*], RN, PhD; Brigitte N Durieux[1], BS; Charlotta Lindvall[1,5,6], MD, PhD

[1]Department of Psychosocial Oncology and Palliative Care, Dana-Farber Cancer Institute, Boston, MA, United States

[2]Albany Medical College, Albany, NY, United States

[3]KU Leuven Department of Public Health and Primary Care, KU Leuven-University of Leuven, Leuven, Belgium

[4]Research Foundation Flanders (FWO), Brussels, Belgium

[5]Department of Medicine, Brigham and Women's Hospital, Boston, MA, United States

[6]Harvard Medical School, Harvard University, Boston, MA, United States

[*]these authors contributed equally

Corresponding Author:
Charlotta Lindvall, MD, PhD
Department of Psychosocial Oncology and Palliative Care, Dana-Farber Cancer Institute
450 Brookline Ave
Boston, MA, 02215
United States
Phone: 1 617 632 6464
Email: charlotta_lindvall@dfci.harvard.edu

Abstract

More clinicians and researchers are exploring uses for large language model chatbots, such as ChatGPT, for research, dissemination, and educational purposes. Therefore, it becomes increasingly relevant to consider the full potential of this tool, including the special features that are currently available through the application programming interface. One of these features is a variable called temperature, which changes the degree to which randomness is involved in the model’s generated output. This is of particular interest to clinicians and researchers. By lowering this variable, one can generate more consistent outputs; by increasing it, one can receive more creative responses. For clinicians and researchers who are exploring these tools for a variety of tasks, the ability to tailor outputs to be less creative may be beneficial for work that demands consistency. Additionally, access to more creative text generation may enable scientific authors to describe their research in more general language and potentially connect with a broader public through social media. In this viewpoint, we present the temperature feature, discuss potential uses, and provide some examples.

(JMIR Hum Factors 2024;11:e53559) doi:10.2196/53559

KEYWORDS

artificial intelligence; ChatGPT; clinical communication; creative; creativity; customization; customize; customized; generation; generative; language model; language models; LLM; LLMs; natural language processing; NLP; random; randomness; tailor; tailored; temperature; text; texts; textual

Introduction

ChatGPT [1] is a large language model developed by OpenAI that currently has over 100 million users [2]. As its popularity continues to grow, clinicians and researchers are among many considering its potential applications in health care and academia. In a short time, ChatGPT has been extensively published [3], with clinical researchers exploring its potential utility for a variety of tasks, including answering patient questions [4,5], generating clinical summaries [6], and abstracting data from important documentation (eg, computed tomography reports) [7].

When using ChatGPT, one can interact through the website by providing a single prompt or engaging in a conversation. In addition to this more well-known web-based version of ChatGPT, there is also an application programming interface (API) that allows for more customization and flexibility. With the API, users can programmatically interact with ChatGPT and modify features for their specific use case. Although this approach may currently require more technical expertise for clinicians to use, its features may become available on the web interface in future iterations of the tool. Therefore, these features are important to understand and relevant to discuss in terms of their meaning for clinicians and researchers in advance of their more widespread use. Additionally, they have direct implications for introducing greater reproducibility in use cases where this matters.

The Temperature Feature of ChatGPT

ChatGPT generates text through a probabilistic language modeling approach, where it writes responses word by word, calculating the most likely next word in the sequence. A key feature that influences this behavior is called temperature [8, 9]. In this context, temperature is a value from 0 to 2 that adjusts how random each subsequent word in the chat output is. A value of 0 will give the most probable word and, thus, the least variability. As the value increases toward and beyond 1, the next word becomes less probable, leading to more randomness and "creativity" in the response. This feature can currently be adjusted in the API, where the default value is 1 [9].

The ability to adjust the "creativity" of ChatGPT output should also be of interest to clinicians and medical researchers using the tool. By accounting for temperature, large language models such as ChatGPT can be tailored for different use cases. Lowering ChatGPT's creativity level would be preferable for tasks that require more consistent outputs; for clinicians and researchers, tasks of this sort may include summarizing patient data (eg, symptoms and medications) or streamlining administrative tasks (eg, billing inquiries and patient registrations). Alternatively, increasing the creativity level may provide better outputs for less structured tasks and may specifically hold relevance for improving clinicians' and researchers' digital communication with other health care workers, patients, and a wider audience.

Currently, many clinicians may find it difficult to engage with the broader public when it comes to disseminating findings. A lack of public accessibility can limit the impact of research. Creativity proves to be a powerful ally in communication, as it simplifies complex concepts, fosters critical thinking, and enhances the accessibility of intricate information. These factors collectively contribute to increased audience engagement and more effective message transmission. This feature has value for developing patient education materials, medical courses, science communication, and public research dissemination [10]. Increasing the creativity level could be useful when trying to report research results on social media [11], at conferences, or in scientific publications. Among other things, researchers could use a more creative ChatGPT to help come up with tags that apply to their work, brainstorm analogies or metaphors for scientific concepts, and check the tone of their writing.

Illustrating This Feature: Example

As an illustration of the "creativity" feature in ChatGPT, we have provided ChatGPT with the abstract of a recently published article on the development of a keyword library for capturing symptom talk in oncology [12]. While setting the temperature to low (0.0), medium (0.5), and high (1.0), we asked ChatGPT to create a tweet for a public audience, a title for a scientific journal article, and a title for a keynote address. Example ChatGPT-generated responses can be found in Table 1.

Table 1. ChatGPT-generated responses on the prompts for the different temperatures. For each temperature level, the same prompts were run 10 times in a ChatGPT model (text-davinci-003).

Temperature	Low (0.0)	Medium (0.5)	High (1.0)
Tweet for public audience	We developed a keyword library to detect symptom talk in transcribed clinical conversations. Our findings suggest that physical symptoms are more easily detected than psychological symptoms, and that more advanced computational models are needed to detect symptom expressions. #AI #Healthcare #SymptomDetection	Have you heard about the new study that developed and evaluated a natural language processing keyword library to detect symptom talk in oncology conversations? The findings highlight the need for more advanced computational models to detect symptom expressions in transcribed clinical conversations. #Oncology #SymptomTalk #NLP #ComputationalMethods	Check out how computational methods can help us detect patient symptoms in clinical conversations! A recent study developed & evaluated a keyword library for capturing expressed symptoms, highlighting the need for more advanced models for real-time detection of symptoms at scale. #AI @healthinformatics
Title for scientific journal	Detecting Symptom Talk in Transcribed Outpatient Oncology Conversations: Developing and Evaluating a Natural Language Processing Keyword Library	Keyword-Based Detection of Symptom Talk in Oncology Conversations: Developing a Natural Language Processing Library and Exploring Opportunities for Advanced Computational Models	Unlocking the Ambiguity of Symptom Talk: Developing and Evaluating a Natural Language Processing Library for Detecting Outpatient Oncology Conversations
Title of a keynote address	Unlocking the Potential of Computational Methods for Detecting Symptoms: Achieving Better Patient Care Through Natural Language Processing	Harnessing the Power of Computational Methods to Detect Patient Suffering: A Study of Symptom Talk in Oncology Conversations	The Ambiguity of Symptom Expressions: Utilizing Computational Methods to Better Attend to Patient Suffering

The examples shown in Table 1 illustrate that when the creativity level in ChatGPT is adjusted, slightly different responses are generated; these different creativity levels may provide more useful output depending on the task at hand. For example, a tweet created with a high level of creativity includes an exclamation mark and directly addresses the audience. Therefore, it may be more engaging compared to tweets with a low or medium creativity level. The title of the article and keynote generated with high "creativity" were more surprising and potentially less useful for these tasks, although this may depend on the context, setting, and personality of the user. For these tasks, the low- and medium-creativity titles were more straightforward. Importantly, these lower values do translate to more consistent responses. We ran each of these prompts 10 times, and at a temperature level of 0, all responses were identical. Given ChatGPT's normally variable output, this feature holds exciting implications for scenarios where consistency and reproducibility are preferred.

In addition to the results reported above, we have also experimented with adjustments in temperature level using other ChatGPT models (ie, gpt-3.5-turbo-1106, gpt-3.5-turbo-instruct, and gpt-4-1106-preview). All outputs appear in Multimedia Appendix 1**.** In contrast to what we found when using the ChatGPT model "text-davinci-003," some other models showed some variability, even at a temperature level of 0. Regardless, the relative variability of outputs is still modified by temperature, with a higher temperature increasing creativity. Users should consider and test how temperature impacts outputs within the model they are using.

In the examples provided above, we have demonstrated how adjusting the level of creativity can enhance science communication, making it more engaging. However, it is crucial to also acknowledge the potential risks associated with increasing creativity, especially for clinical cases. Using ChatGPT with high creativity settings in clinical contexts, such as for summarizing patient medical data, can be problematic. Excessive creativity might lead to the embellishment or misrepresentation of crucial information, either by omitting vital details or interpreting data too liberally. Such inaccuracies could impact patient treatment and outcomes. Therefore, it is advisable to lower the creativity level of ChatGPT in clinical applications. By doing so, we ensure that the summarized information remains faithful to the original data, thereby prioritizing accuracy and reliability over creative expression.

In summary, the temperature feature of ChatGPT allows users to adjust the level of "creativity." Although no previous articles have discussed or investigated this feature for its use in clinical research, it shows promising potential for clinicians and researchers. Both high and low creativity levels could have interesting applications for health care and may broaden the ways clinicians and researchers consider using artificial intelligence (AI) tools to close gaps in areas such as digital communication. ChatGPT documentation suggests using a temperature value of 0 to 0.2 for more focused (less creative) tasks and 0.8 to 1 for more random (more creative) tasks [9]. As large language models are variable and use case dependent, we strongly suggest testing and validating the proper temperature level for your specific use case. While this feature is a powerful tool that could be useful for creating easy-to-understand summaries, captivating social media posts, or making complex information more accessible to a wider audience, the parameters need to be carefully tweaked to find a balance between coherence and creativity and to tailor to specific needs. Looking ahead, as AI continues to advance in the health care sector, the temperature feature can play a pivotal role in health care applications in generative AI, unlocking the potential for more accurate, empathetic, or creative interactions between AI and health care stakeholders.

Acknowledgments

This work was supported by the Research Foundation Flanders (FWO; grants 1159522N and K210723N to LVB) and by the National Institutes of Health, National Institute on Aging Medical Student Training in Aging Research Program (grant 5T35AG038027-13 to JD). The funding sources had no involvement in the study design, analyses, or interpretation of the data; writing of the report; or decision to submit the paper for publication. During the preparation of this work, the authors used ChatGPT in order to generate the examples reported in the paper and to assist with the writing process. The use of a large language model has greatly assisted us in rephrasing and ensuring the clarity and effectiveness of our language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Data Availability

All data generated or analyzed during this study are included in this published article and Multimedia Appendix 1.

Authors' Contributions

JD and LVB contributed equally and share first authorship. LVB, JD, and BND contributed to the conception and design of the study and drafted the paper. CL critically revised the paper. JD and LVB both accessed and verified the underlying data reported in the manuscript. All authors approved the final version of the manuscript and had full responsibility for the decision to submit for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

All data presented in this article, ChatGPT outputs for tests of 3 prompts across 3 temperature values for 3 different models (gpt-3.5-turbo-1106, gpt-3.5-turbo-instruct, and gpt-4-1106-preview; 100 runs for each test), and a summary document describing the multiple model tests.

[ZIP File (Zip Archive), 241 KB - humanfactors_v11i1e53559_app1.zip]

References

1. ChatGPT. OpenAI. URL: https://openai.com/chatgpt [accessed 2024-02-16]
2. Hu K. ChatGPT sets record for fastest-growing user base—analyst note. Reuters. 2023. URL: https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/ [accessed 2023-10-05]
3. Ward E, Gross C. Evolving methods to assess chatbot performance in health sciences research. JAMA Intern Med 2023;183(9):1030-1031. [doi: 10.1001/jamainternmed.2023.2567] [Medline: 37459112]
4. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 2023;183(6):589-596. [doi: 10.1001/jamainternmed.2023.1838] [Medline: 37115527]
5. Van Bulck L, Moons P. What if your patient switches from Dr. Google to Dr. ChatGPT? a vignette-based survey of the trustworthiness, value, and danger of ChatGPT-generated responses to health questions. Eur J Cardiovasc Nurs 2024;23(1):95-98 [FREE Full text] [doi: 10.1093/eurjcn/zvad038] [Medline: 37094282]
6. Nayak A, Alkaitis MS, Nayak K, Nikolov M, Weinfurt KP, Schulman K. Comparison of history of present illness summaries generated by a chatbot and senior internal medicine residents. JAMA Intern Med 2023;183(9):1026-1027. [doi: 10.1001/jamainternmed.2023.2561] [Medline: 37459091]
7. Fink MA, Bischoff A, Fink CA, Moll M, Kroschke J, Dulz L, et al. Potential of ChatGPT and GPT-4 for data mining of free-text CT reports on lung cancer. Radiology 2023;308(3):e231362. [doi: 10.1148/radiol.231362] [Medline: 37724963]
8. Best practices for prompt engineering with OpenAI API: how to give clear and effective instructions to GPT-3 and Codex. Shieh J. OpenAI. 2023. URL: https://help.openai.com/en/articles/6654000-best-practices-for-prompt-engineering-with-openai-api [accessed 2023-12-11]
9. API reference. OpenAI. 2023. URL: https://platform.openai.com/docs/api-reference/chat [accessed 2023-12-11]
10. Heng JJY, Teo DB, Tan LF. The impact of Chat Generative Pre-Trained Transformer (ChatGPT) on medical education. Postgrad Med J 2023;99(1176):1125-1127. [doi: 10.1093/postmj/qgad058] [Medline: 37466157]
11. Grajales FJ, Sheps S, Ho K, Novak-Lauscher H, Eysenbach G. Social media: a review and tutorial of applications in medicine and health care. J Med Internet Res 2014;16(2):e13 [FREE Full text] [doi: 10.2196/jmir.2912] [Medline: 24518354]
12. Durieux BN, Zverev SR, Tarbi EC, Kwok A, Sciacca K, Pollak KI, et al. Development of a keyword library for capturing PRO-CTCAE-focused "symptom talk" in oncology conversations. JAMIA Open 2023;6(1):ooad009 [FREE Full text] [doi: 10.1093/jamiaopen/ooad009] [Medline: 36789287]

Abbreviations

AI: artificial intelligence
API: application programming interface

Edited by E Borycki, A Kushniruk; submitted 11.10.23; peer-reviewed by L Chinthala, C Li, S Minshall; comments to author 20.11.23; revised version received 11.12.23; accepted 24.01.24; published 08.03.24.

Please cite as:
Davis J, Van Bulck L, Durieux BN, Lindvall C
The Temperature Feature of ChatGPT: Modifying Creativity for Clinical Research
JMIR Hum Factors 2024;11:e53559
URL: https://humanfactors.jmir.org/2024/1/e53559
doi:10.2196/53559
PMID:38457221

Original Paper

The Impact of Performance Expectancy, Workload, Risk, and Satisfaction on Trust in ChatGPT: Cross-Sectional Survey Analysis

Avishek Choudhury[1], PhD; Hamid Shamszare[1], MSc

Industrial and Management Systems Engineering, Benjamin M. Statler College of Engineering and Mineral Resources, West Virginia University, Morgantown, WV, United States

Corresponding Author:
Avishek Choudhury, PhD
Industrial and Management Systems Engineering
Benjamin M. Statler College of Engineering and Mineral Resources
West Virginia University
321 Engineering Sciences Building
1306 Evansdale Drive
Morgantown, WV, 26506
United States
Phone: 1 3042939431
Email: avishek.choudhury@mail.wvu.edu

Abstract

Background: ChatGPT (OpenAI) is a powerful tool for a wide range of tasks, from entertainment and creativity to health care queries. There are potential risks and benefits associated with this technology. In the discourse concerning the deployment of ChatGPT and similar large language models, it is sensible to recommend their use primarily for tasks a human user can execute accurately. As we transition into the subsequent phase of ChatGPT deployment, establishing realistic performance expectations and understanding users' perceptions of risk associated with its use are crucial in determining the successful integration of this artificial intelligence (AI) technology.

Objective: The aim of the study is to explore how perceived workload, satisfaction, performance expectancy, and risk-benefit perception influence users' trust in ChatGPT.

Methods: A semistructured, web-based survey was conducted with 607 adults in the United States who actively use ChatGPT. The survey questions were adapted from constructs used in various models and theories such as the technology acceptance model, the theory of planned behavior, the unified theory of acceptance and use of technology, and research on trust and security in digital environments. To test our hypotheses and structural model, we used the partial least squares structural equation modeling method, a widely used approach for multivariate analysis.

Results: A total of 607 people responded to our survey. A significant portion of the participants held at least a high school diploma (n=204, 33.6%), and the majority had a bachelor's degree (n=262, 43.1%). The primary motivations for participants to use ChatGPT were for acquiring information (n=219, 36.1%), amusement (n=203, 33.4%), and addressing problems (n=135, 22.2%). Some participants used it for health-related inquiries (n=44, 7.2%), while a few others (n=6, 1%) used it for miscellaneous activities such as brainstorming, grammar verification, and blog content creation. Our model explained 64.6% of the variance in trust. Our analysis indicated a significant relationship between (1) workload and satisfaction, (2) trust and satisfaction, (3) performance expectations and trust, and (4) risk-benefit perception and trust.

Conclusions: The findings underscore the importance of ensuring user-friendly design and functionality in AI-based applications to reduce workload and enhance user satisfaction, thereby increasing user trust. Future research should further explore the relationship between risk-benefit perception and trust in the context of AI chatbots.

(JMIR Hum Factors 2024;11:e55399) doi:10.2196/55399

KEYWORDS

ChatGPT; chatbots; health care; health care decision-making; health-related decision-making; health care management; decision-making; user perception; usability; usable; usableness; usefulness; artificial intelligence; algorithms; predictive models; predictive analytics; predictive system; practical models; deep learning; cross-sectional survey

Introduction

ChatGPT (OpenAI) [1] is a powerful tool for a wide range of tasks, from entertainment and creativity to health care queries [2]. However, there are potential benefits associated with this technology. For instance, it can help summarize large amounts of text data [3,4] or generate programming code [5]. There is also the notion that ChatGPT may potentially assist with health care tasks [6-9]. However, the risks associated with using ChatGPT can hinder its adoption in various high-risk domains. These risks include the potential for inaccuracies and lack of citation relevance in scientific content generated by ChatGPT [10], ethical issues (copyright, attribution, plagiarism, and authorship) [11], the risk of hallucination (inaccurate information that sounds plausible scientifically) [12], and the possibility of biased content and inaccurate information due to the quality of training data sets generated prior to the year 2021 [4].

In the discourse concerning the deployment of ChatGPT and similar artificial intelligence (AI) technologies, it is sensible to recommend their use primarily for tasks a human user can execute accurately. Few studies have advocated using the technology under human supervision [13,14]. Encouraging users to rely on such tools for tasks beyond their competence is risky, as they may need help to evaluate the AI's output effectively. The strength of ChatGPT lies in its ability to automate more straightforward, mundane tasks, freeing human users to invest their time and cognitive resources into critical tasks (not vice versa). This approach to technology use maintains a necessary balance, leveraging AI for efficiency gains while ensuring that critical decision-making remains within the purview of human expertise.

As we transition into the subsequent phase of ChatGPT deployment, establishing realistic performance expectations and understanding users' perceptions of risk associated with its use are crucial in determining the successful integration of this AI technology. Thus, understanding users' perceptions of ChatGPT becomes essential, as these perceptions significantly influence their usage decisions [2]. For example, suppose users believe that ChatGPT's capabilities surpass human knowledge. In that case, they may be tempted to use it for tasks such as self-diagnosis, which could lead to potentially harmful outcomes if the generated information is mistaken or misleading. Conversely, a realistic appraisal of the limitations and strengths of technology would encourage its use in low-risk, routine tasks and foster a safer, more effective integration into our everyday lives.

Building upon the importance of user perceptions and expectations, we must also consider that the extent to which users trust ChatGPT hinges mainly on the perception of its accuracy and reliability. As users witness the technology's ability to perform tasks effectively and generate correct, helpful information, their trust in the system grows. This, in turn, allows them to offload routine tasks to the AI and focus their energies on more complex or meaningful endeavors. Similarly, instances where the AI generates inaccurate or misleading information can quickly erode users' perception of the technology. Users may become dissatisfied and lose trust if they perceive the technology as unreliable or potentially harmful, particularly if they have previously overestimated its capabilities. This underlines the importance of setting realistic expectations and accurately understanding the strengths and limitations of ChatGPT, which can help foster a healthy level of trust and satisfaction among users. Ultimately, establishing and maintaining trust and satisfaction are not a onetime event but an ongoing process of validating the AI's outputs, understanding and acknowledging its limitations, and making the best use of its capabilities within a framework of informed expectations and continuous learning. This dynamic balance is pivotal for the effective and safe integration of AI technologies such as ChatGPT into various sectors of human activity.

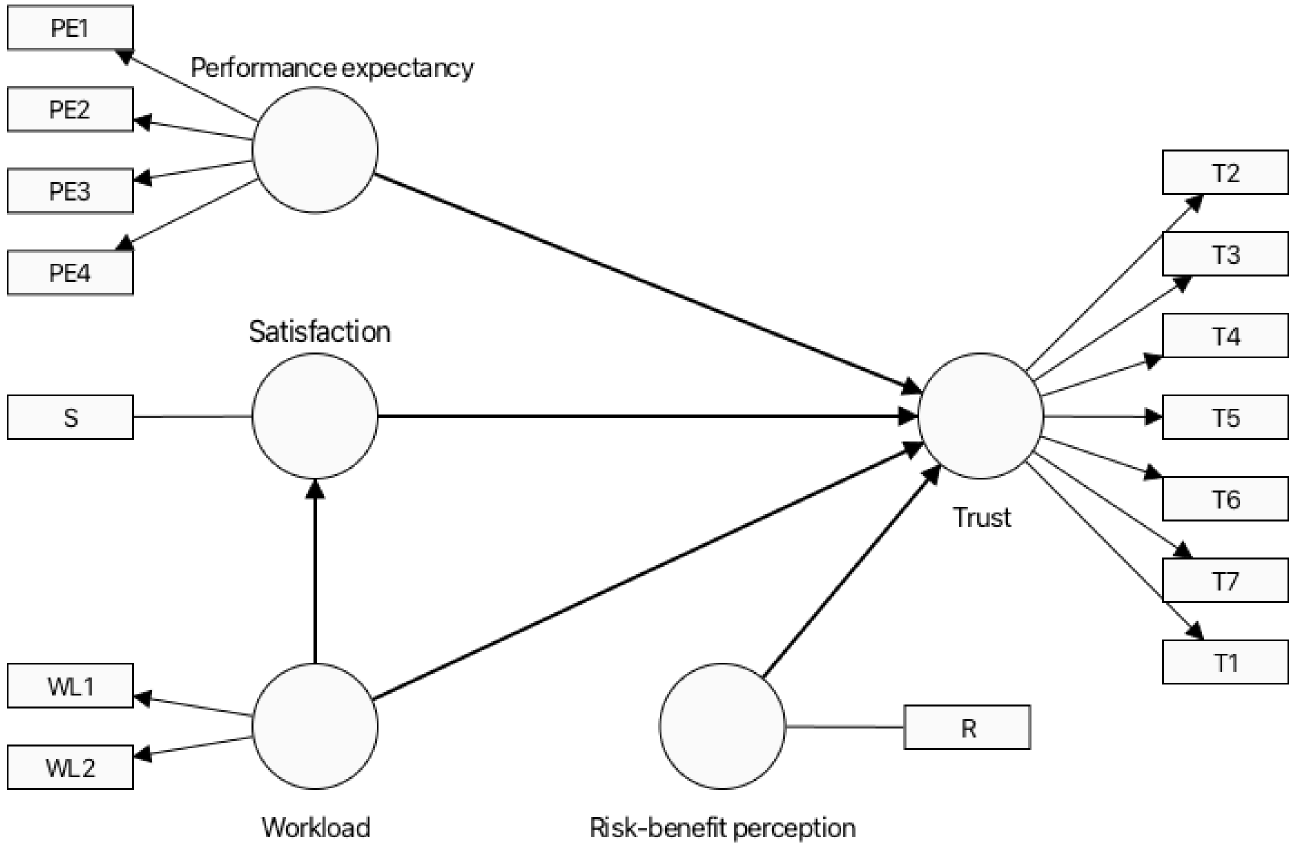
In our prior work, we explored the impact of trust in the actual use of ChatGPT [15]. This study aims to explore a conceptual framework delving deeper into the aspects influencing user trust in ChatGPT.

As shown in Figure 1, the proposed conceptual model is grounded in the well-established theories of technology acceptance and use, incorporating constructs such as performance expectancy, workload, satisfaction, risk-benefit perception, and trust to comprehensively evaluate user interaction with technology. Performance expectancy, derived from the core postulates of the technology acceptance model (TAM) [16] and the unified theory of acceptance and use of technology (UTAUT) [17], posits that the perceived use of the technology significantly predicts usage intentions. Workload, akin to effort expectancy, reflects the perceived cognitive and physical effort required to use the technology, where a higher workload may inversely affect user satisfaction—a construct that encapsulates the fulfillment of user expectations and needs through technology interaction. The risk-benefit perception embodies the user's assessment of the technology's potential advantages against its risks, intricately influencing both user satisfaction and trust. Trust, a pivotal determinant of technology acceptance [15], signifies the user's confidence in the reliability and efficacy of the technology. This theoretical framework thus serves to elucidate the multifaceted process by which users come to accept and use a technological system, highlighting the critical role of both cognitive appraisals and affective responses in shaping the technology adoption landscape.

We explore the following hypotheses:

- Hypothesis 1: Perceived workload of using ChatGPT negatively correlates with user trust in ChatGPT.
- Hypothesis 2: Perceived workload of using ChatGPT negatively correlates with user satisfaction with ChatGPT.
- Hypothesis 3: User satisfaction with ChatGPT positively correlates with trust in ChatGPT.
- Hypothesis 4: User trust in ChatGPT is positively correlated with the performance expectancy of ChatGPT.
- Hypothesis 5: The risk-benefit perception of using ChatGPT is positively correlated with user trust in ChatGPT.

Figure 1. A conceptual model of technology acceptance illustrating trust (T) as the dependent outcome variable, with performance expectancy (PE), workload (WL), and risk-benefit perception (R) as direct predictors. Satisfaction (S) is depicted as a mediating variable that moderates the impact of workload on trust.



Methods

Ethical Considerations

The study obtained ethics approval from West Virginia University, Morgantown (protocol 2302725983). The study was performed in accordance with relevant guidelines and regulations. No identifiers were collected during the study, and all users were compensated for completing the survey through an audience paneling service. In compliance with ethical research practices, informed consent was obtained from all participants before initiating the survey. Attached to the survey was a comprehensive cover letter outlining the purpose of the study, the procedure involved, the approximate time to complete the survey, and assurances of anonymity and confidentiality. It also emphasized that participation was completely voluntary, and participants could withdraw at any time without any consequences. The cover letter also included the contact information of the researchers for any questions or concerns the participants might have regarding the study. Participants were asked to read through the cover letter information carefully and were instructed to proceed with the survey only if they understood and agreed to the terms described, effectively providing their consent to participate in the study.

Study Design

A semistructured, web-based questionnaire was disseminated to adult individuals within the United States who engaged with ChatGPT (version 3.5) at least once per month. Data collection took place between February and March 2023. The questionnaire was crafted using Qualtrics (Qualtrics LLC), and its circulation was handled by Centiment (Centiment LLC), a provider of audience-paneling services. Centiment's services were used due to their extensive reach and ability to connect with a diverse and representative group via their network and social media. Their fingerprinting technology, which uses IP address, device type, screen size, and cookies, was used to guarantee the uniqueness of the survey respondents. Prior to the full-scale dissemination, a soft launch was carried out with 40 responses gathered. The purpose of a soft launch, a limited-scale trial of the survey, is to pinpoint any potential problems, such as ambiguity or confusion in questions, technical mishaps, or any other factors that might affect the quality of data obtained. The survey was made available to a larger audience following the successful soft launch.

Table 1 shows the descriptive statistics of the survey questions used in this study. We developed 3 latent constructs based on the question: trust, workload, and performance expectancy, and 2 single question variables: satisfaction and risk-benefit perception. Participant responses to all the questions were captured using a 4-point Likert scale ranging from 1=strongly disagree to 4=strongly agree. These questions were adapted from constructs used in various models and theories such as the TAM, the theory of planned behavior, UTAUT, and research on trust and security in digital environments.

- Trust: Questions T1-T7 related to trust in AI systems were adapted from the trust building model [18].

- Workload: WL1 and WL2 questions from the National Aeronautics and Space Administration Task Load Index for measuring perceived workload [19].
- Performance expectancy: PE1-PE4 are about the perceived benefits of using the system, which is a central concept in TAM and UTAUT.
- Satisfaction: The single item relates to overall user satisfaction, a common measure in information systems success models [20].
- Risk-benefit perception: Question addresses the user’s assessment of benefits relative to potential risks, an aspect often discussed in the context of technology adoption and use [21].

These references provide a starting point for understanding the theoretical underpinnings of the survey used in this study. They are adapted from foundational works in information systems, human-computer interaction, and psychology that address trust, workload, performance expectancy, satisfaction, and the evaluation of benefits versus risks in technology use.

- Trust: Questions T1-T7 related to trust in AI systems were adapted from the trust building model [18].
- Workload: WL1 and WL2 questions from the National Aeronautics and Space Administration Task Load Index for measuring perceived workload [19].
- Performance expectancy: PE1-PE4 are about the perceived benefits of using the system, which is a central concept in TAM and UTAUT.
- Satisfaction: The single item relates to overall user satisfaction, a common measure in information systems success models [20].
- Risk-benefit perception: Question addresses the user’s assessment of benefits relative to potential risks, an aspect often discussed in the context of technology adoption and use [21].

These references provide a starting point for understanding the theoretical underpinnings of the survey used in this study. They are adapted from foundational works in information systems, human-computer interaction, and psychology that address trust, workload, performance expectancy, satisfaction, and the evaluation of benefits versus risks in technology use.

Table 1. Study variables and latent construct (N=607).

Survey items	Value, mean (SD)
Trust (T)	
T1: ChatGPT is competent in providing the information and guidance I need	3.20 (0.83)
T2: ChatGPT is reliable in providing consistent and dependable information	3.16 (0.80)
T3: ChatGPT is transparent	3.12 (0.86)
T4: ChatGPT is trustworthy in the sense that it is dependable and credible	3.17 (0.84)
T5: ChatGPT will not cause harm, manipulate its responses, or create negative consequences for me	3.10 (0.88)
T6: ChatGPT will act with integrity and be honest with me	3.19 (0.82)
T7: ChatGPT is secure and protects my privacy and confidential information	3.27 (0.81)
Workload (WL)	
WL1: Using ChatGPT was mentally demanding	3.21 (0.75)
WL2: I had to work hard to use ChatGPT	2.20 (0.98)
Performance expectancy (PE)	
PE1: ChatGPT can help me achieve my goals	3.24 (0.77)
PE2: ChatGPT can reduce my workload	3.22 (0.78)
PE3: ChatGPT improves my work efficiency	3.21 (0.84)
PE4: ChatGPT helps me make informed and timely decisions	3.26 (0.79)
Satisfaction (S)	
S: I am satisfied with ChatGPT	3.24 (0.76)
Risk-benefit perception (R)	
R: The benefits of using ChatGPT outweigh any potential risks	3.20 (0.80)

Statistical Analysis and Model Validation

To test our hypotheses and structural model, we used the partial least squares structural equation modeling (PLS-SEM) method, a widely used approach for multivariate analysis. PLS-SEM enables the estimation of complex models with multiple constructs, indicator variables, and structural paths, without making assumptions about the data’s distribution [22]. This method is beneficial for studies with small sample sizes that involve many constructs and items [23]. PLS-SEM is a suitable method because of its flexibility and ability to allow for interaction between theory and data in exploratory research [24]. The analyses were performed using the *SEMinR* package in R (R Foundation for Statistical Computing) [25]. We started by loading the data set collected for this study using the *reader* package in R. We then defined the measurement model. This consisted of 5 composite constructs: trust, performance expectancy, workload, risk-benefit perception, and satisfaction. Trust was measured with 7 items (T1 through T7), performance expectancy with 4 items (PE1 through PE4), and workload with 2 items (WL1 and WL2), while risk-benefit perception and satisfaction were each measured with a single item. We also evaluated the convergent and discriminant validity of the latent constructs, which we assessed using 3 criteria: factor loadings (>0.50), composite reliability (>0.70), and average variance extracted (>0.50). We used the Heterotrait-Monotrait ratio (<0.90) to assess discriminant validity [26].

Next, we defined the structural model, which captured the hypothesized relationships between the constructs. The model included paths from risk-benefit perception, performance expectancy, workload, satisfaction to trust, and a path from workload to satisfaction. We then estimated the model's parameters using the partial least squares method. This was done with the *estimate_pls* function in the *seminar* package. The partial least squares method was preferred due to its ability to handle complex models and its robustness to violations of normality assumptions. We performed a bootstrap resampling procedure with 10,000 iterations to obtain robust parameter estimates and compute 95% CIs. The bootstrapped model was plotted to visualize the estimates and their 95% CIs.

Results

Of 607 participants who completed the survey, 29.9% (n=182) used ChatGPT at least once per month, 26.1% (n=158) used it weekly, 24.5% (n=149) accessed it more than once per week, and 19.4% (n=118) interacted with it almost daily. A substantial portion of the participants held at least a high school diploma (n=204, 33.6%), and the majority had a bachelor's degree (n=262, 43.1%). The primary motivations for participants to use ChatGPT were for acquiring information (n=219, 36%), amusement (n=203, 33.4%), and addressing problems (n=135, 22.2%). Some participants used it for health-related inquiries (n=44, 7.2%), while a few others (n=6, 1%) used it for miscellaneous activities such as brainstorming, grammar verification, and blog content creation. Table 2 shows the factor loading of the latent constructs in the model.

The model explained 2% and 64.6% of the variance in "satisfaction" and "trust," respectively. Reliability estimates, as shown in Table 3, indicated high levels of internal consistency for all 5 latent variables, with Cronbach α and ρ values exceeding the recommended threshold of 0.7. The average variance extracted for the latent variables also exceeded the recommended threshold of 0.5, indicating that these variables are well-defined and reliable. Based on the root mean square error of approximation (RMSEA) fit index, our PLS-SEM model demonstrates a good fit for the observed data. The calculated RMSEA value of 0.07 falls below the commonly accepted threshold of 0.08, indicating an acceptable fit. The RMSEA estimates the average discrepancy per degree of freedom in the model, capturing how the proposed model aligns with the population covariance matrix. With a value below the threshold, it suggests that the proposed model adequately represents the relationships among the latent variables. This finding provides confidence in the model's ability to explain the observed data and support the underlying theoretical framework.

Table 4 shows the estimated paths in our model. Hypothesis 1 postulated that as the perceived workload of using ChatGPT increases, user trust in ChatGPT decreases. Our analysis indicated a negative estimate for the path from workload to trust (–0.047). However, the *T* statistic (–1.674) is less than the critical value, and the 95% CI straddles 0 (–0.102 to –0.007), suggesting that the effect is not statistically significant. Therefore, we do not have sufficient evidence to support hypothesis 1.

Hypothesis 2 stated that perceived workload is negatively correlated with user satisfaction with ChatGPT. The results supported this hypothesis, as the path from workload to satisfaction showed a negative estimate (–0.142), a *T* statistic (–3.416) beyond the critical value, and a 95% CI (–0.223 to –0.061).

The data confirmed this relationship for hypothesis 3, which proposed a positive correlation between satisfaction with ChatGPT and trust in ChatGPT. The path from satisfaction to trust had a positive estimate (0.165), a *T* statistic (4.478) beyond the critical value, and a 95% CI (0.093-0.237).

Hypothesis 4 suggested that user performance expectations of ChatGPT increase with their trust in the technology. The analysis supported this hypothesis. The path from performance expectancy to trust displayed a positive estimate (0.598), a large *T* statistic (15.554), and a 95% CI (0.522-0.672). Finally, we examined hypothesis 5, which posited that user trust in ChatGPT increases as their risk-benefit perception of using the technology increases. The path from risk-benefit perception to trust showed a positive estimate (0.114). The *T* statistic (3.372) and the 95% CI (0.048-0.179) indicating this relationship is significant, but the positive sign suggests that as the perceived benefits outweigh the risks, the trust in ChatGPT increases. Therefore, hypothesis 5 is supported. Figure 2 illustrates the structural model with all path coefficients.

Table 2. Bootstrapped loadings: model analysis estimates the relationship between various constructs and their indicators.

Bootstrapped loadings	Loadings	*T* statistic	95% CI
Trust (T)			
T1	0.788	41.998	0.750-0.823
T2	0.753	33.795	0.706-0.794
T3	0.773	40.293	0.733-0.808
T4	0.732	28.772	0.679-0.779
T5	0.673	21.066	0.607-0.732
T6	0.799	46.065	0.763-0.831
T7	0.779	38.088	0.736-0.816
Performance expectancy (PE)			
PE1	0.809	49.231	0.775-0.839
PE2	0.733	29.360	0.681-0.779
PE3	0.802	44.968	0.766-0.835
PE4	0.777	34.198	0.729-0.818
Workload (WL)			
WL1	0.856	28.883	0.789-0.905
WL2	0.913	44.872	0.869-0.950

Table 3. Convergent reliability.

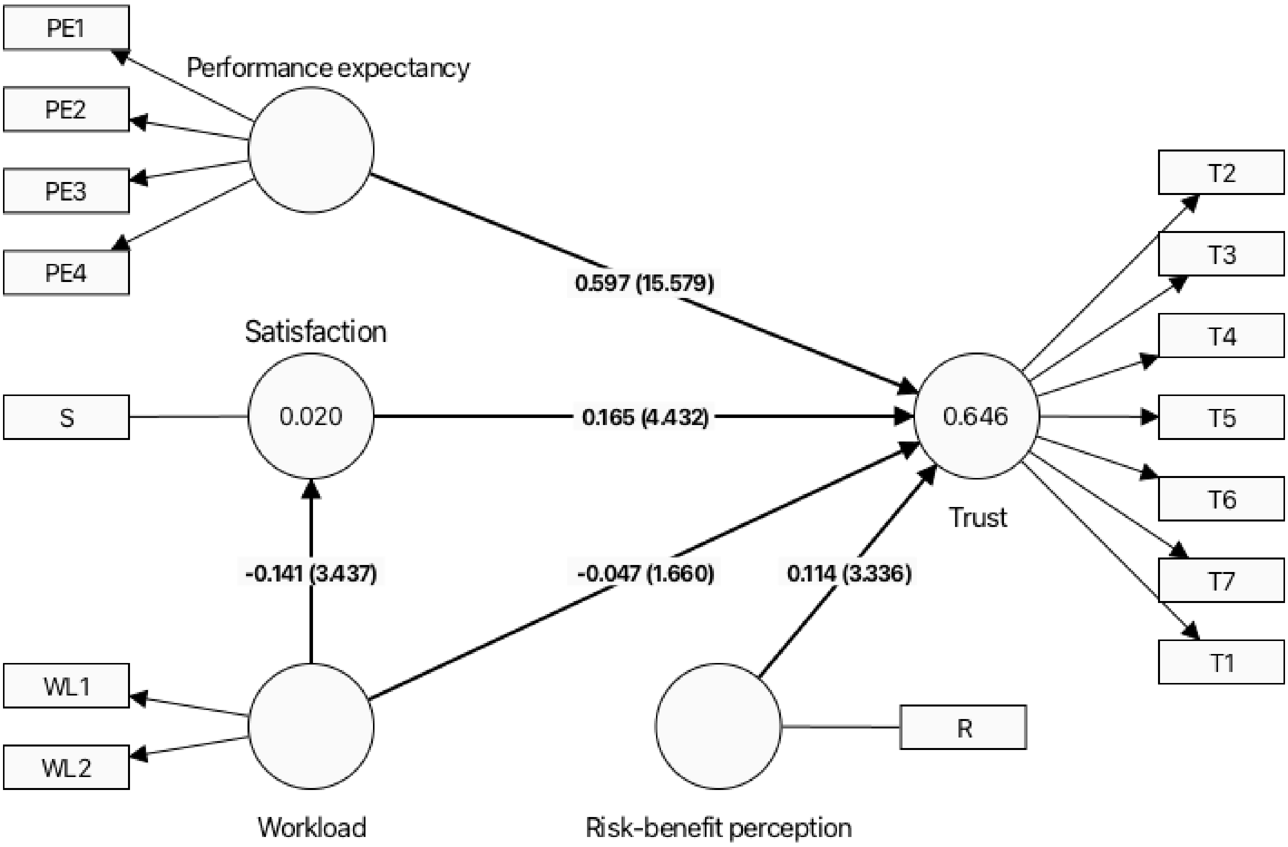
Construct	Cronbach α	ρ C	AVE[a]	ρ A
Performance expectation	0.787	0.862	0.610	0.610
Workload	0.729	0.870	0.771	0.968
Trust	0.876	0.904	0.575	0.880

[a]AVE: average variance extracted.

Table 4. Bootstrapped structural path estimates.

Direct path	Bootstrap mean standard estimate (SD)	*T* statistic	95% CI
Risk-benefit perception→trust	0.114 (0.034)	3.372	0.048 to 0.179
Performance expectancy→trust	0.598 (0.038)	15.554	0.522 to 0.672
Workload→satisfaction	–0.142 (0.041)	–3.416	–0.223 to –0.061
Workload→trust	–0.047 (0.028)	–1.674	–0.102 to 0.007
Satisfaction→trust	0.165 (0.037)	4.478	0.093 to 0.237

Figure 2. The significant paths connecting trust (T) in ChatGPT, performance expectancy (PE), satisfaction (S), workload (WL), and risk-benefit perception (R). T1 through T7: factors for trust; PE1 through PE4: factors for performance expectancy; and WL1 and WL2: factors for workload. The inner model shows the path coefficient and T statistic values.



Discussion

Main Findings

This study represents one of the initial attempts to investigate how human factors such as workload, performance expectancy, risk-benefit perception, and satisfaction influence trust in ChatGPT. Our results showed that these factors significantly influenced trust in ChatGPT, with performance expectancy exerting the strongest association, highlighting its critical role in fostering trust. Additionally, we found that satisfaction was a mediator in the relationship between workload and trust. At the same time, a positive correlation was observed between trust in ChatGPT and the risk-benefit perception. Our findings align with the May 23, 2023, efforts and initiatives of the Biden-Harris Administration to advance responsible AI research, development, and deployment [27]. The Administration recognizes that managing its risks is crucial and prioritizes protecting individuals' rights and safety. One of the critical actions taken by the administration is the development of the artificial intelligence risk management framework (AI RMF). The AI RMF builds on the importance of trustworthiness in AI systems and is a framework for strengthening AI trustworthiness and promoting the trustworthy design, development, deployment, and use of AI systems, contributing to the need for our research [28]. Our findings reveal the importance of performance expectancy, satisfaction, and risk-benefit perception in determining the user's trust in AI systems. By addressing these factors, AI systems can be designed and developed to be more user-centric, aligning with the AI RMF's emphasis on human-centricity and responsible AI.

Workload and Trust in ChatGPT

Moreover, we found that reducing user workload is vital for enhancing user satisfaction, which in turn improves trust. This finding aligns with the AI RMF's focus on creating AI systems that are equitable and accountable and that mitigate inequitable outcomes. Additionally, our research emphasizes the need for future exploration of other factors impacting user trust in AI technologies. Such endeavors align with the AI RMF's vision of managing AI risks comprehensively and holistically, considering technical and societal factors. Understanding these factors is crucial for fostering public trust and enhancing the overall trustworthiness of AI systems, as outlined in the AI RMF [28].

This study also extends and complements existing literature. Consistent with the observed patterns in studies on flight simulators, dynamic multitasking environments, and cyberattacks [29-31], we also found that higher perceived workload in using ChatGPT led to lower levels of trust in this technology. Our findings align with the existing research indicating a negative correlation between workload and user satisfaction [32]. We observed that as the perceived workload of using ChatGPT increased, user satisfaction with the technology decreased. This outcome echoes the consensus within the literature that a high workload can lead to user dissatisfaction, particularly if the technology requires too much effort or time [33]. The literature reveals that perceived workload balance significantly influences job satisfaction in work organizations [25], and similar patterns are found in the well-being studies of nurses, where perceived workload negatively impacts satisfaction with work-life balance [34]. While this study does not directly involve the workplace environment or work-life balance, the parallels between workload and satisfaction are evident. Furthermore, our research parallels the study suggesting that when providing timely service, AI applications can alleviate perceived workload and improve job satisfaction [35]. ChatGPT, as an AI-powered chatbot, could potentially contribute to workload relief when it performs effectively and efficiently, thereby boosting user satisfaction.

Satisfaction and Trust in ChatGPT

Our findings corroborate with existing literature, suggesting a strong positive correlation between user satisfaction and trust in the technology or service provider [23,24,26,36-38]. We found that the users who expressed higher satisfaction with ChatGPT were more likely to trust the system, strengthening the premise that satisfaction can predict trust in a technology or service provider. Similar to the study on digital transaction services, our research indicates that higher satisfaction levels with ChatGPT corresponded with higher trust in the AI system [37]. This suggests that when users are satisfied with the performance and results provided by ChatGPT, they tend to trust the technology more. The research on mobile transaction apps mirrors our findings, where we also discovered that satisfaction with ChatGPT use was a significant predictor of trust in the system [36]. This showcases the importance of ensuring user satisfaction in fostering trust using innovative technologies like AI chatbots. The study on satisfaction with using digital assistants, where a positive relationship between trust and satisfaction was observed [26], further aligns with our study. We also found a positive correlation between trust in ChatGPT and user satisfaction with this AI assistant.

Performance Expectancy and Trust in ChatGPT

Our findings concerning the strong positive correlation between performance expectancy and trust in ChatGPT serve as an extension to prior literature. Similar findings have been reported in previous studies on wearables and mobile banking [39,40], where performance expectancy was positively correlated with trust. However, our results diverge from the observations of a recent study that did not find a significant impact of performance expectancy on trust in chatbots [41]. Moreover, the observed mediating role of satisfaction in the relationship between workload and trust in ChatGPT is a notable contribution to the literature. While previous studies have demonstrated a positive correlation between workload reduction by chatbots and trust, as well as between trust and user satisfaction [42-44], the role of satisfaction as a mediator between workload and trust has not been explored. Finally, the positive correlation between the risk-benefit perception of using ChatGPT and trust aligns with the findings of previous studies [45-47]. Similar studies on the intention to use chatbots for digital shopping and customer service have found that trust in chatbots impacts perceived risk and is affected by the risk involved in using chatbots [46,47]. Our study adds to this body of research by confirming the same positive relationship within the context of ChatGPT.

Limitations

Despite the valuable insights provided by this study, limitations should be acknowledged. First, our research focused explicitly on ChatGPT and may not be generalizable to other AI-powered conversational agents or chatbot technologies. Different chatbot systems may have unique characteristics and user experiences that could influence the factors affecting trust. Second, this study relied on self-reported data from survey responses, which may be subject to response biases and limitations inherent to self-report measures. Participants' perceptions and interpretations of the constructs under investigation could vary, leading to potential measurement errors. Third, this study was cross-sectional, capturing data at a specific point in time. Longitudinal studies that track users' experiences and perceptions over time provide a more comprehensive understanding of the dynamics between trust and the factors investigated. Finally, the sample of participants in this study consisted of individuals who actively use ChatGPT, which may introduce a self-selection bias. The perspectives and experiences of nonusers or individuals with limited exposure to AI-powered conversational agents may differ, and their insights could provide additional valuable perspectives.

Conclusions

This study examined the factors influencing trust in ChatGPT, an AI-powered conversational agent. Our analysis found that performance expectancy, satisfaction, workload, and risk-benefit perceptions significantly influenced users' trust in ChatGPT. These findings contribute to understanding trust dynamics in the context of AI-powered conversational agents and provide insights into the factors that can enhance user trust. By addressing the factors influencing trust, we contribute to the broader goal of fostering responsible AI practices that prioritize user-centric design and protect individuals' rights and safety. Future research should consider longitudinal designs to capture the dynamics of trust over time. Additionally, incorporating perspectives from diverse user groups and examining the impact of contextual factors on trust would further enrich our understanding of trust in AI technologies.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

AC, the lead researcher, was responsible for the study's conceptualization, the survey's development, figure illustration, data collection and analysis, and manuscript writing. HS, the student author, was responsible for manuscript writing and conducting the literature review. Both authors collaborated throughout the research process and approved the final version of the manuscript for submission.

Conflicts of Interest

None declared.

References

1. OpenAI: Models GPT-3. OpenAI. URL: https://platform.openai.com/docs/models/gpt-4 [accessed 2024-05-07]
2. Shahsavar Y, Choudhury A. User intentions to use ChatGPT for self-diagnosis and health-related purposes: cross-sectional survey study. JMIR Hum Factors 2023;10:e47564 [FREE Full text] [doi: 10.2196/47564] [Medline: 37195756]
3. Editorial N. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. Nature 2023;613(7945):612 [FREE Full text] [doi: 10.1038/d41586-023-00191-1] [Medline: 36694020]
4. Moons P, Van Bulck L. ChatGPT: can artificial intelligence language models be of value for cardiovascular nurses and allied health professionals. Eur J Cardiovasc Nurs 2023;22(7):e55-e59 [FREE Full text] [doi: 10.1093/eurjcn/zvad022] [Medline: 36752788]
5. Aljanabi M, Ghazi M, Ali AH, Abed SA. ChatGpt: open possibilities. Iraqi J Sci Comput Sci Math 2023;4(1):62-64. [doi: 10.52866/20ijcsm.2023.01.01.0018]
6. D'Amico RS, White TG, Shah HA, Langer DJ. I asked a ChatGPT to write an editorial about how we can incorporate chatbots into neurosurgical research and patient care…. Neurosurgery 2023;92(4):663-664 [FREE Full text] [doi: 10.1227/neu.0000000000002414] [Medline: 36757199]
7. Holzinger A, Keiblinger K, Holub P, Zatloukal K, Müller H. AI for life: trends in artificial intelligence for biotechnology. N Biotechnol 2023;74:16-24 [FREE Full text] [doi: 10.1016/j.nbt.2023.02.001] [Medline: 36754147]
8. Sharma G, Thakur A. ChatGPT in drug discovery. ChemRxiv. 2023. URL: https://chemrxiv.org/engage/chemrxiv/article-details/63d56c13ae221ab9b240932f [accessed 2024-05-07]
9. Mann DL. Artificial intelligence discusses the role of artificial intelligence in translational medicine: a interview with ChatGPT. JACC Basic Transl Sci 2023;8(2):221-223 [FREE Full text] [doi: 10.1016/j.jacbts.2023.01.001] [Medline: 36908674]
10. Chen TJ. ChatGPT and other artificial intelligence applications speed up scientific writing. J Chin Med Assoc 2023;86(4):351-353 [FREE Full text] [doi: 10.1097/JCMA.0000000000000900] [Medline: 36791246]
11. Liebrenz M, Schleifer R, Buadze A, Bhugra D, Smith A. Generating scholarly content with ChatGPT: ethical challenges for medical publishing. Lancet Digit Health 2023;5(3):e105-e106 [FREE Full text] [doi: 10.1016/S2589-7500(23)00019-5] [Medline: 36754725]
12. Shen Y, Heacock L, Elias J, Hentel KD, Reig B, Shih G, et al. ChatGPT and other large language models are double-edged swords. Radiology 2023;307(2):e230163 [FREE Full text] [doi: 10.1148/radiol.230163] [Medline: 36700838]
13. Lubowitz JH. ChatGPT, an artificial intelligence chatbot, is impacting medical literature. Arthroscopy 2023;39(5):1121-1122 [FREE Full text] [doi: 10.1016/j.arthro.2023.01.015] [Medline: 36797148]
14. Jianning L, Amin D, Jens K, Jan E. ChatGPT in healthcare: a taxonomy and systematic review. medRxiv Preprint posted online on March 30, 2023. [doi: 10.1101/2023.03.30.23287899]
15. Choudhury A, Shamszare H. Investigating the Impact of User Trust on the Adoption and Use of ChatGPT: Survey Analysis. J Med Internet Res 2023 Jun 14;25:e47184 [FREE Full text] [doi: 10.2196/47184] [Medline: 37314848]
16. Hu PJ, Chau PYK, Sheng ORL, Tam KY. Examining the technology acceptance model using physician acceptance of telemedicine technology. J Manage Inform Syst 2015;16(2):91-112 [FREE Full text] [doi: 10.1080/07421222.1999.11518247]
17. Williams MD, Rana NP, Dwivedi YK. The unified theory of acceptance and use of technology (UTAUT): a literature review. J Enterp Inf Manag 2015;28(3):443-448 [FREE Full text] [doi: 10.1108/jeim-09-2014-0088]
18. McKnight DH, Choudhury V, Kacmar C. The impact of initial consumer trust on intentions to transact with a web site: a trust building model. J Strateg Inf Syst 2002;11(3-4):297-323 [FREE Full text] [doi: 10.1016/S0963-8687(02)00020-3]
19. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Adv Psychol 1988;52:139-183 [FREE Full text] [doi: 10.1016/S0166-4115(08)62386-9]
20. Delone WH, McLean ER. The DeLone and McLean model of information systems success: a ten-year update. J Manag Inf Syst 2014;19(4):9-30 [FREE Full text] [doi: 10.1080/07421222.2003.11045748]
21. Featherman MS, Pavlou PA. Predicting e-services adoption: a perceived risk facets perspective. Int J Hum Comput Stud 2003;59(4):451-474 [FREE Full text] [doi: 10.1016/s1071-5819(03)00111-3]
22. Hair JF, Sarstedt M, Ringle CM, Gudergan SP. Advanced Issues in Partial Least Squares Structural Equation Modeling. 2nd Edition. Thousand Oaks, CA: SAGE Publications; 2017.
23. Eriksson K, Hermansson C, Jonsson S. The performance generating limitations of the relationship-banking model in the digital era—effects of customers' trust, satisfaction, and loyalty on client-level performance. Int J Bank Mark 2020;38(4):889-916 [FREE Full text] [doi: 10.1108/ijbm-08-2019-0282]
24. Al-Ansi A, Olya HGT, Han H. Effect of general risk on trust, satisfaction, and recommendation intention for halal food. Int J Hosp Manag 2019;83:210-219 [FREE Full text] [doi: 10.1016/j.ijhm.2018.10.017]

25. Inegbedion H, Inegbedion E, Peter A, Harry L. Perception of workload balance and employee job satisfaction in work organisations. Heliyon 2020;6(1):e03160 [FREE Full text] [doi: 10.1016/j.heliyon.2020.e03160] [Medline: 31938748]
26. Marikyan D, Papagiannidis S, Rana OF, Ranjan R, Morgan G. "Alexa, let's talk about my productivity": the impact of digital assistants on work productivity. J Bus Res 2022;142:572-584 [FREE Full text] [doi: 10.1016/j.jbusres.2022.01.015]
27. FACT SHEET: Biden- Harris Administration takes new steps to advance responsible artificial intelligence research, development, and deployment. The White House. 2023. URL: https://tinyurl.com/bdfnb97b [accessed 2024-05-07]
28. Tabassi E. Artificial Intelligence Risk Management Framework (AI RMF 1.0). National Institute of Standards and Technology. 2023. URL: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf [accessed 2024-05-07]
29. Sato T, Yamani Y, Liechty M, Chancey ET. Automation trust increases under high-workload multitasking scenarios involving risk. Cogn Tech Work 2019;22(2):399-407 [FREE Full text] [doi: 10.1007/s10111-019-00580-5]
30. Karpinsky ND, Chancey ET, Palmer DB, Yamani Y. Automation trust and attention allocation in multitasking workspace. Appl Ergon 2018;70:194-201 [FREE Full text] [doi: 10.1016/j.apergo.2018.03.008] [Medline: 29866311]
31. Gontar P, Homans H, Rostalski M, Behrend J, Dehais F, Bengler K. Are pilots prepared for a cyber-attack? A human factors approach to the experimental evaluation of pilots' behavior. J Air Transp Manag 2018;69:26-37 [FREE Full text] [doi: 10.1016/j.jairtraman.2018.01.004]
32. Tentama F, Rahmawati PA, Muhopilah P. The effect and implications of work stress and workload on job satisfaction. Int J Sci Technol Res 2019;8(11):2498-2502 [FREE Full text]
33. Kim C, Mirusmonov M, Lee I. An empirical examination of factors influencing the intention to use mobile payment. Comput Human Behav 2010;26(3):310-322 [FREE Full text] [doi: 10.1016/j.chb.2009.10.013]
34. Holland P, Tham TL, Sheehan C, Cooper B. The impact of perceived workload on nurse satisfaction with work-life balance and intention to leave the occupation. Appl Nurs Res 2019;49:70-76 [FREE Full text] [doi: 10.1016/j.apnr.2019.06.001] [Medline: 31375315]
35. Nguyen TM, Malik A. A two-wave cross-lagged study on AI service quality: the moderating effects of the job level and job role. Br J Manag 2021;33(3):1221-1237 [FREE Full text] [doi: 10.1111/1467-8551.12540]
36. Kumar A, Adlakaha A, Mukherjee K. The effect of perceived security and grievance redressal on continuance intention to use M-wallets in a developing country. Int J Bank Mark 2018;36(7):1170-1189 [FREE Full text] [doi: 10.1108/ijbm-04-2017-0077]
37. Chen X, Li S. Understanding continuance intention of mobile payment services: an empirical study. J Comput Inf Syst 2016;57(4):287-298 [FREE Full text] [doi: 10.1080/08874417.2016.1180649]
38. Fang Y, Qureshi I, Sun H, McCole P, Ramsey E, Lim KH. Trust, satisfaction, and online repurchase intention. MIS Q 2014;38(2):407-428 [FREE Full text]
39. Gu Z, Wei J, Xu F. An empirical study on factors influencing consumers' initial trust in wearable commerce. J Comput Inf Syst 2015;56(1):79-85. [doi: 10.1080/08874417.2015.11645804]
40. Oliveira T, Faria M, Thomas MA, Popovič A. Extending the understanding of mobile banking adoption: when UTAUT meets TTF and ITM. Int J Inform Manage 2014;34(5):689-703 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2014.06.004]
41. Mostafa RB, Kasamani T. Antecedents and consequences of chatbot initial trust. Eur J Mark 2021;56(6):1748-1771 [FREE Full text] [doi: 10.1108/ejm-02-2020-0084]
42. Wang X, Lin X, Shao B. Artificial intelligence changes the way we work: a close look at innovating with chatbots. J Assoc Inf Sci Technol 2022;74(3):339-353 [FREE Full text] [doi: 10.1002/asi.24621]
43. Hsiao KL, Chen CC. What drives continuance intention to use a food-ordering chatbot? An examination of trust and satisfaction. Libr Hi Tech 2021;40(4):929-946 [FREE Full text] [doi: 10.1108/lht-08-2021-0274]
44. Pesonen JA. 'Are You OK?' Students' trust in a chatbot providing support opportunities. : Springer; 2021 Presented at: Learning and Collaboration Technologies: Games and Virtual Environments for Learning: 8th International Conference, LCT 2021, Held as Part of the 23rd HCI International Conference, HCII 2021; July 24-29, 2021; Virtual Event. [doi: 10.1007/978-3-030-77943-6_13]
45. Dwivedi YK, Kshetri N, Hughes L, Slade EL, Jeyaraj A, Kar AK, et al. Opinion paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. Int J Inform Manage 2023;71:102642 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2023.102642]
46. Silva SC, De Cicco R, Vlačić B, Elmashhara MG. Using chatbots in e-retailing—how to mitigate perceived risk and enhance the flow experience. Int J Retail Distrib Manag 2022;51(3):285-305 [FREE Full text] [doi: 10.1108/ijrdm-05-2022-0163]
47. Nordheim CB, Følstad A, Bjørkli CA. An initial model of trust in chatbots for customer service? Findings from a questionnaire study. Interact Comput 2019;31(3):317-335 [FREE Full text] [doi: 10.1093/iwc/iwz022]

Abbreviations

AI: artificial intelligence
AI RMF: artificial intelligence risk management framework
PLS-SEM: partial least squares structural equation modeling
RMSEA: root mean square error of approximation

TAM: technology acceptance model
UTAUT: unified theory of acceptance and use of technology

Edited by A Kushniruk, E Borycki; submitted 11.12.23; peer-reviewed by P Radanliev, G Farid; comments to author 17.01.24; revised version received 25.03.24; accepted 07.04.24; published 27.05.24.

Please cite as:
Choudhury A, Shamszare H
The Impact of Performance Expectancy, Workload, Risk, and Satisfaction on Trust in ChatGPT: Cross-Sectional Survey Analysis
JMIR Hum Factors 2024;11:e55399
URL: https://humanfactors.jmir.org/2024/1/e55399
doi:10.2196/55399
PMID:38801658

Review

Human Factors in AI-Driven Digital Solutions for Increasing Physical Activity: Scoping Review

Elia Gabarron[1,2], PhD; Dillys Larbi[2,3], MSc, MPH; Octavio Rivera-Romero[4], PhD; Kerstin Denecke[5], PhD

[1]Department of Education, ICT and Learning, Østfold University College, Halden, Norway

[2]Norwegian Centre for eHealth Research, University Hospital of North Norway, Tromsø, Norway

[3]Department of Clinical Medicine, The University of Tromsø-The Arctic University of Norway, Tromsø, Norway

[4]Department of Electronic Technology, Universidad de Sevilla, Sevilla, Spain

[5]AI for Health, Institute Patient-centered Digital Health, Department of Engineering and Computer Science, Bern University of Applied Sciences, Bern, Switzerland

Corresponding Author:
Elia Gabarron, PhD
Department of Education, ICT and Learning
Østfold University College
B R A Veien 4
Halden, 1757
Norway
Phone: 47 94863460
Email: elia.gabarron@hiof.no

Abstract

Background: Artificial intelligence (AI) has the potential to enhance physical activity (PA) interventions. However, human factors (HFs) play a pivotal role in the successful integration of AI into mobile health (mHealth) solutions for promoting PA. Understanding and optimizing the interaction between individuals and AI-driven mHealth apps is essential for achieving the desired outcomes.

Objective: This study aims to review and describe the current evidence on the HFs in AI-driven digital solutions for increasing PA.

Methods: We conducted a scoping review by searching for publications containing terms related to PA, HFs, and AI in the titles and abstracts across 3 databases—PubMed, Embase, and IEEE Xplore—and Google Scholar. Studies were included if they were primary studies describing an AI-based solution aimed at increasing PA, and results from testing the solution were reported. Studies that did not meet these criteria were excluded. Additionally, we searched the references in the included articles for relevant research. The following data were extracted from included studies and incorporated into a qualitative synthesis: bibliographic information, study characteristics, population, intervention, comparison, outcomes, and AI-related information. The certainty of the evidence in the included studies was evaluated using GRADE (Grading of Recommendations Assessment, Development, and Evaluation).

Results: A total of 15 studies published between 2015 and 2023 involving 899 participants aged approximately between 19 and 84 years, 60.7% (546/899) of whom were female participants, were included in this review. The interventions lasted between 2 and 26 weeks in the included studies. Recommender systems were the most commonly used AI technology in digital solutions for PA (10/15 studies), followed by conversational agents (4/15 studies). User acceptability and satisfaction were the HFs most frequently evaluated (5/15 studies each), followed by usability (4/15 studies). Regarding automated data collection for personalization and recommendation, most systems involved fitness trackers (5/15 studies). The certainty of the evidence analysis indicates moderate certainty of the effectiveness of AI-driven digital technologies in increasing PA (eg, number of steps, distance walked, or time spent on PA). Furthermore, AI-driven technology, particularly recommender systems, seems to positively influence changes in PA behavior, although with very low certainty evidence.

Conclusions: Current research highlights the potential of AI-driven technologies to enhance PA, though the evidence remains limited. Longer-term studies are necessary to assess the sustained impact of AI-driven technologies on behavior change and habit formation. While AI-driven digital solutions for PA hold significant promise, further exploration into optimizing AI's impact on PA and effectively integrating AI and HFs is crucial for broader benefits. Thus, the implications for innovation management involve conducting long-term studies, prioritizing diversity, ensuring research quality, focusing on user experience, and understanding the evolving role of AI in PA promotion.

(JMIR Hum Factors 2024;11:e55964) doi:10.2196/55964

KEYWORDS

machine learning; ML; artificial intelligence; AI; algorithm; algorithms; predictive model; predictive models; predictive analytics; predictive system; practical model; practical models; deep learning; human factors; physical activity; physical exercise; healthy living; active lifestyle; exercise; physically active; digital health; mHealth; mobile health; app; apps; application; applications; digital health; digital technology; digital intervention; digital interventions; smartphone; smartphones; PRISMA

Introduction

Physical activity (PA) has been recognized as a cornerstone of a healthy lifestyle since it has demonstrated numerous benefits for both physical and mental well-being [1,2]. Engaging in regular PA has been associated with preventing and managing a range of health conditions, including obesity, diabetes, cardiovascular disease, and multiple sclerosis [2,3]. However, the global population's engagement in regular PA is often low, with many individuals failing to meet the recommendations necessary for health benefits. This persistent challenge necessitates innovative approaches to motivate and facilitate increased PA participation, and mobile health (mHealth) technologies have emerged as a promising avenue for intervention [4].

The availability of mobile devices and the increasing mobile penetration provide an unprecedented opportunity to leverage mHealth solutions to promote PA [5,6]. Mobile technologies offer persuasive and ubiquitous systems. Equipped with built-in sensors that can monitor and encourage PA in real time, they can facilitate sending personalized reminders and motivational messages [7-10], which have been proven to significantly increase PA [10-12]. However, the effectiveness of mHealth interventions in promoting PA has been limited by the challenge of sustaining engagement over the medium and long term. Mönninghoff et al [13] found that mHealth "can foster small to moderate increases in PA," and the effects are even maintained long-term, but "the effect size decreases over time." This is where the integration of artificial intelligence (AI) holds immense promise. AI technology has the potential to deliver effective interventions to promote PA [11,14].

AI can enrich mHealth solutions by offering personalized, adaptive, and tailored interventions that cater to individual preferences and needs. For example, an optimal exercise plan for an individual can be suggested by AI algorithms to help maximize the long-term health utility of the user [15]. This level of customization has the potential to enhance user experience (UX), which in turn could result in increased motivation to engage in PA. Motivation is a critical factor in driving behavior change, especially when adopting and maintaining a physically active lifestyle. AI can also gamify fitness by setting challenges, goals, and rewards, motivating users to increase PA through points, competition, and achievements [16]. Moreover, research indicates that the human-likeness of conversational agents increases adherence to chatbots [17] and compliance with their recommendations [18].

In this context, human factors (HFs) play a pivotal role in the successful integration of AI into mHealth solutions aimed at promoting PA. Understanding and optimizing the interaction between individuals and AI-driven mHealth apps is essential for achieving the desired outcomes [19]. HFs, in the context of AI, involve considerations related to human cognition, behavior, and ergonomics, which are crucial for designing effective and user-friendly mHealth interventions. Bergevi et al [20] explored users' perceptions of acceptability, engagement, and usability of mHealth solutions that promote PA, healthy diets, or both. They concluded that mHealth services targeting increased PA "should be personalized, dynamic, easily manageable, and reliable." This study is distinguished from their work by focusing on AI-driven digital solutions.

This research underscores the critical role of PA in promoting overall health and well-being while highlighting the persistent challenge of low engagement in regular PA globally. It emphasizes the potential of mHealth technologies, augmented by AI, to effectively motivate and facilitate increased PA participation. By leveraging AI, mHealth solutions can offer personalized, adaptive interventions tailored to individual preferences and needs, thereby enhancing the UX and motivation. However, the successful integration of AI into mHealth solutions relies on understanding and optimizing HFs, encompassing cognition, behavior, and ergonomics, to ensure effective and user-friendly interventions. Specifically, this study aims to address the following research question, what are the key HFs influencing the effectiveness and adoption of AI-driven digital solutions aimed at promoting PA? Our objective is to review and describe the current evidence on the HFs in AI-driven digital solutions for increasing PA.

Methods

Overview

We have conducted a scoping review to capture current evidence on HFs in AI-driven digital solutions for increasing PA. A scoping review is a systematic approach used to map and synthesize existing literature on a broad topic, providing an overview of key concepts, sources, and knowledge gaps. Our review followed the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [21].

Search Strategy

We have searched for publications including keywords related to PA (ie, "physical activity;" "exercise;" "active lifestyle;" "sedentary behaviour;" "inactivity;" "resistance training;"

"exergaming;" "walking;" "swimming;" "jogging;" "climbing"), artificial intelligence (ie, "artificial intelligence;" "AI;" "machine learning;" "deep learning;" "natural language processing;" "neural networks;" "sentiment analysis"), and human factors (ie, "usability;" "task performance;" "satisfaction;" "workload;" "human errors;" "user perception;" "cognitive factors;" "mental model;" "context awareness;" "automation bias;" "teamworking;" "user experience;" "acceptance;" "acceptability;" "task analysis;" "handover;" "patient interaction;" "human factors;" "ergonomics") in their titles and abstracts. No language or year limitations were used. The full search strategy can be found in Multimedia Appendix 1.

The data search was performed on August 29, 2023. The database search was done by a single author (EG) and covered PubMed, Embase, and IEEE Xplore. Another author (DL) carried out a search on Google Scholar and selected the first 100 entries. Finally, DL used a snowballing approach to identify additional relevant studies cited in only the included publications.

Eligibility and Selection Process

Inclusion and exclusion criteria are presented in Textbox 1.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Primary studies that described an artificial intelligence (AI)–based digital solution
- AI-based digital solutions aimed at increasing physical activity
- Publications that reported results from testing the AI-based solution related to physical activity behavior

Exclusion criteria

- Publications that did not meet all 3 inclusion criteria

All references were uploaded to EndNote (version 20.6; Clarivate) [22] and Rayyan (Qatar Computing Research Institute) [23]. After duplicates were removed, 2 authors (EG and DL) independently assessed the eligibility of the remaining publications by checking their titles and abstracts. Two additional authors (KD and OR-R) checked the full text of the eligible papers after the title and abstract screening. After the full-text screening, the selected papers were included in a qualitative synthesis.

Data Items and Data Extraction

Two authors (KD and OR-R) extracted the following data: bibliographic information (publication year and country); study characteristics (study design, type of evaluation, research methods, primary and secondary measures, materials, and theoretical foundations); population (number of participants, age, and gender); intervention (intervention design, duration, and follow-ups); comparison (control group or groups and pre-post evaluation or other); outcomes (primary and secondary outcomes); and AI-related information (technology type, main purpose, platform, and HFs).

OR-R identified and assigned codes representative of the main purpose of the AI model implemented in each of the systems studied. The 3 main purposes of the AI models implemented in the studied systems were identified as personalization, communication, and human activity recognition. Personalization includes all AI models analyzed whose main purpose was to adapt the digital solution or intervention to the patient's needs, conditions, and preferences. The second group includes models that enabled a communication pathway with patients. Finally, human activity recognition includes all models that enable the detection of user behaviors, particularly PA. OR-R and KD reviewed the assigned codes and created a classification of these by consensus.

Certainty of the Evidence

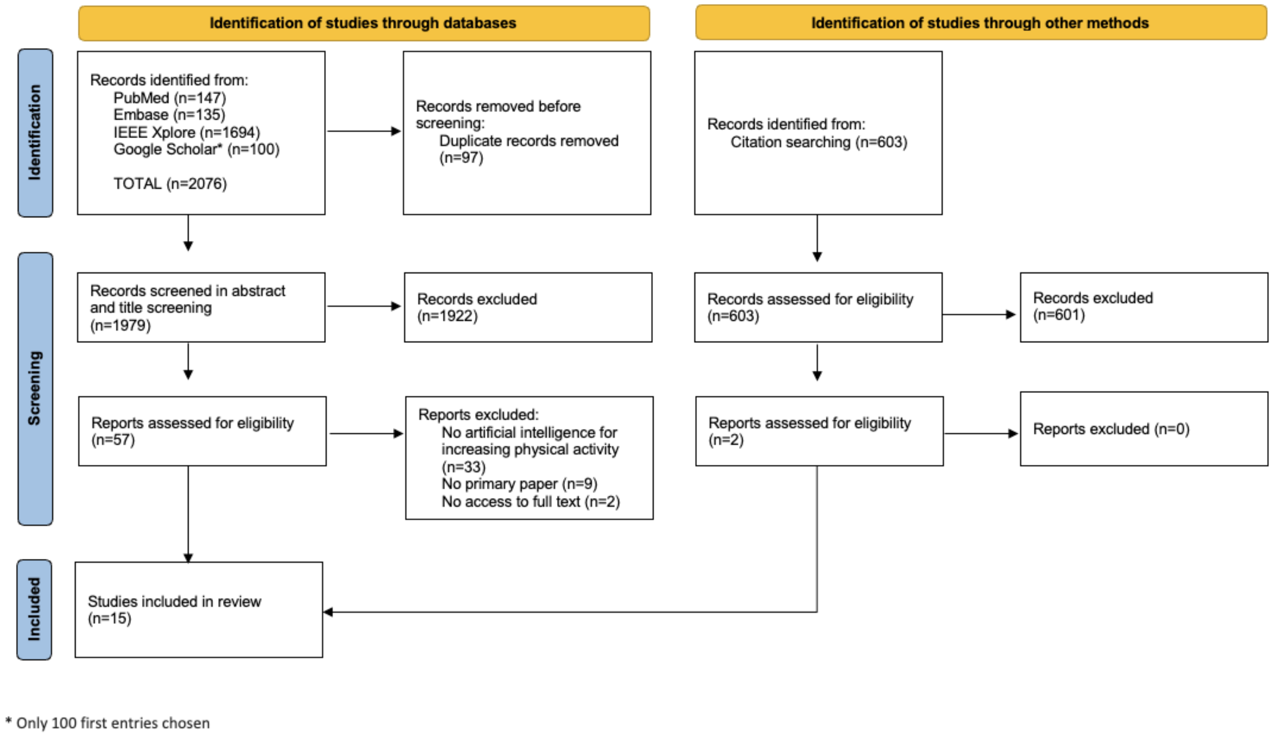
The certainty of the evidence on the outcomes was assessed by a single author (EG) by drawing on the GRADE (Grading of Recommendations Assessment, Development, and Evaluation) criteria [24] and verified by the rest of the coauthors.

Results

Study Selection

A total of 2076 articles were identified in the data search. After removing duplicates, 1979 titles and abstracts were screened for eligibility. Of those, 13 publications met the inclusion criteria [25-37]. The snowballing approach identified 2 additional publications [38,39]; therefore, the final number of publications included in this review was 15 (Figure 1 shows the PRISMA [Preferred Reporting Items for Systematic Reviews and Meta-Analyses] flowchart).

Figure 1. Flowchart diagram of the selection process.



The list of publications excluded during the full-text search and the reasons for their exclusion are reported in Multimedia Appendix 2.

Description of the Included Publications

The 15 included articles were published between 2015 and 2023. Countries of origin of these studies were: United States (n=3) [25,26,36], Australia (n=2) [32,38], South Korea (n=2) [29,35], the Netherlands (n=1) [31], Italy (n=1) [27], Belgium and Italy (n=1) [30], Thailand (n=1) [37], and Taiwan (n=1) [33]. A total of 3 publications did not specify in which country the study was performed [28,34,39].

Regarding the study design, 8 publications followed a quasi-experimental approach [26-29,32,35,38,39], 5 were randomized controlled trials [25,30,33,36,37], and 2 were exploratory studies [31,34].

Only 4 of the 15 included publications explicitly mentioned their theoretical foundations. The following theoretical approaches were cited in these 4 publications: the Fogg Model for behavior change [25,31], Capability, Opportunity, and Motivation model of Behavior [32]; learning theory [25], social cognitive theory [25], and the Transtheoretical Model [39].

The main technical characteristics of the 15 included publications are presented in Table 1.

Table 1. Main technical characteristics of the included artificial intelligence (AI)–based solutions.

Author and year	AI tech type	AI purpose	AI techniques	System platform	Human factors
Rabbi et al (2015) [25]	RS[a]	Personalization	Gaussian mixture model	MyBehavior (Mobile app)	User experience
Rabbi et al (2018) [26]	RS	Personalization and HAR[b]	Data clustering algorithm and sequential decision-making algorithm (multiarmed bandit)	MyBehaviorCBP (Mobile app)	Acceptability
Fadhil et al (2019) [27]	CA[c]	Communication	Fine state machine and multi class support vector machine	Chatbot stand-alone	Acceptability
Davis et al (2020) [28]	CA	Communication	Unknown	IBM Watson digital assistant AI software running on Slack	User experience
Luštrek et al (2021) [30]	RS	Personalization and HAR	Random forest	App with wristband	Acceptability, attitude, and performance expectancy
Joo et al (2021) [29]	RS+HAR	Personalization and HAR	Feature point extraction and part affinity fields (machine learning technology with top-down segmentation)	Weelo (web-based fitness program)	Satisfaction, usability, and usefulness
Pelle et al (2021) [31]	RS	Personalization and proposes challenging, achievable, and tailored goals	Machine learning compromising a dynamic model (contextual multiarmed bandit approach)	A stand-alone mobile health app	Usability
To et al (2021) [32]	CA	Personalization and communication	Unknown	DialogFlow (Google), Fitbit Flex, and messenger app	Usability and acceptability
Lin et al (2022) [33]	RS	Personalization and provides a personal training program	Decision tree	AIoT[d], mobile app, and web application	Usability
Park et al (2022) [34]	HAR	HAR	Convolutional neuronal networks	Mobile app	Satisfaction, acceptability, and task performance
Seok et al (2022) [35]	RS	Communication	Large-scale modular behavior networks with inferred contexts and probabilistic model and Russel’s arousal-variance model	TouchCare system: wearable watch, touchpad sensors, TouchCare app, and context-aware AI	Satisfaction
Bates et al (2023) [36]	RS	Personalization and real-time feedback on form and resistance for each task in the training program	Unknown	Tonal AI (commercially available product)	Satisfaction
Thiengwittayaporn et al (2023) [37]	RS	Personalization and patient disease stage	Decision tree classification	Mobile app	Satisfaction
Maher et al (2020) [38]	CA	Communication and personalization	Unknown	IBM Watson	Acceptability

[a]RS: recommender system.

[b]HAR: human activity recognition.

[c]CA: conversational agent.

[d]AIoT: artificial intelligence of things.

AI-Driven Technology and HFs

The most common AI technology type was recommender systems, described in 10 of the 15 included publications [25,26,29-31,33,35-37,39]. In addition to the recommender system, one of these publications also included computer vision [29]. Conversational agents were the second most used AI technology, as described in 4 publications [27,28,32,38]. One of them was integrated into a social media platform, namely Slack [28]. One study tested human activity recognition [34]. Details of the AI technology, systems, or platforms used in the included studies are summarized in Table 1.

Regarding the considered HFs, the most commonly evaluated were acceptability [26,32,34,38,39] and satisfaction [29,34-37], both reported in 5 publications. Usability was the next most considered and evaluated HF, as reported in 4 papers [29,31-33]. Usefulness was assessed in 2 publications [26,29]. Other considered and evaluated HFs were engagement [38], UX or individual perception [25], and task performance [34].

The studies that resulted in increased PA and had a moderate certainty of evidence were chatbot systems with integrated recommender systems. Although the usability of some of those systems was considered poor [32], they were perceived positively. Several papers gained interesting results regarding HFs. Given the variety of systems, a generalization for all 15 studies is difficult.

The automated collection of data needed for personalization and recommendation is an important aspect. In total, 5 systems involved fitness trackers [26,30,32,35,38] to enable automated data collection. They can be grouped into mobile-based activity tracking using movement sensors in the phone [25], dedicated fitness trackers [30,32,38], specifically an accelerometer in the wristband [30], Fitbit Flex 1 activity tracker (Fitbit LLC) [32], and Garmin Vivofit4 tracker (Garmin) [38], and smartwatches [35]. Rabbi et al [25] concluded that automated data collection would be useful. The studies involving chatbots concluded that users have high expectations regarding the chatbot's knowledge and capabilities [28]. Human likeness is reported as a success factor of such systems. Relevant aspects leading to the efficacy of the system include the human-like qualities of the chatbot and the personalization of the suggestions [28,32,39], that is, chatbots or digital assistants should have a personality, have humor, be able to act with spontaneous behavior, and in a diverse, nonrepetitive manner [28,32]. They should provide the correct answers. For successful recommendations, it is essential to learn the personal preferences of users so that suggestions can be made that fit into personal routines and lifestyles [39].

Even a combination of human agents and digital agents was reported to be better accepted than pure virtual support [27]. Beyond that, access to a system anywhere and anytime is well perceived [31]—and this is reflected by the fact that most systems included in this study are delivered as mobile apps (instead of desktop apps). Exercises and recommendations are successful in this setting when they can be easily integrated into the daily lives of the users [31].

Population, Interventions, and Comparison

A total of 899 individuals participated in the included publications. Of those, 60.7% (546) were female participants. The reported average ages of these participants ranged from 18.7 to 84.4 years. In total, 6 out of the 15 studies tested their solutions on participants with mean ages of around 50 years or older [28,31-33,37,38], while 6 studies predominantly included participants with a mean age of 40 years or younger [25,27,29,34,36,39]. Two studies did not specify the gender or age of participants [30,35].

The intervention of the included studies lasted between 2 and 26 weeks.

Prepost evaluations were carried out in 6 of the publications to evaluate the impact of the AI-driven intervention [26,29,32,35,38,39]. In 4 publications, control groups were used to assess the impact [25,30,34,36]. In 5 of the publications, the comparison methods used to assess the impact of the AI-driven intervention on increasing PA were not clearly reported [27,28,31,33,37].

Outcomes and Certainty of the Evidence

The effectiveness of AI-driven technologies for increasing PA was shown in 5 publications [28,32,36,38,39]. Three of these publications tested conversational agents [28,32,38], while the other 2 focused on recommender systems [36,39]. The analysis, based on GRADE guidelines, found moderate certainty in the evidence supporting this statement. Further details about the proven effect of these studies and the certainty of the evidence on these findings are reported in Table 2.

Table 2. Certainty of the evidence (artificial intelligence for increasing physical activity [PA]).

Outcome	Effect	Participants (studies)	Certainty of the evidence (GRADE[a])[b]	Comment
• Increased number of steps, distance walked, or time spent on PA • Follow-up: mean 9.4 weeks	• Increased walked distance [36] • exceeded step goal [28] • more steps [32] • increased walking minutes [39] • increased time spent on PA [38]	• n=260 (3 pre-post studies, 1 RCT[c], and 1 observational study)	B: moderate	We have a moderate level of confidence that the actual impact closely aligns with the estimated effect.
• Change in PA behavior and abilities to perform behavior • Follow-up: mean 13.3 weeks	• Feeling more stimulated to engage in PAs [30] • change in walking behaviors [25] • improved behaviors related to PA [35] • improved ability to do sports [37]	• n=98 reported (number not explicitly reported in 2 studies) • 3 RCTs, 1 pre-post study	D: very low	We have a very low level of confidence in the estimated effect.

[a]GRADE: Grading of Recommendations Assessment, Development, and Evaluation.

[b]Scale of 4 degrees, where A denotes the highest quality and D denotes the lowest quality.

[c]RCT: randomized controlled trial.

In total, 4 of the included articles also showed that AI technologies have an effect on changing PA behavior (ie, feeling more stimulated to engage in PAs, change in walking behavior, improved behavior related to PA, or improved ability to perform

PA) [25,30,35,37]. All these publications were recommender systems [25,30,35,37] and found a positive effect of AI-driven technology on changing PA behavior. However, the analysis, based on GRADE guidelines, found very low certainty evidence supporting this statement.

Discussion

Principal Results

In this scoping review, we aimed to identify and describe the current evidence on HFs in AI-driven digital solutions for increasing PA. The results showed that the most common AI technology used in digital solutions for PA was recommender systems, followed by conversational agents. User acceptability and satisfaction were the most commonly evaluated HFs in the included studies. Some studies also evaluated the usability of AI-driven digital solutions for PA.

We have identified studies that provide evidence that AI-driven digital technologies have the potential to increase PA (eg, number of steps, distance walked, or time spent on PA). Furthermore, AI-driven technology, particularly recommender systems and chatbots, seems to have the potential to influence changes in PA behavior. Although these studies offer valuable insights by demonstrating positive outcomes through various AI-driven technologies for enhancing PA, the evidence is still very limited. The main findings are presented in Table 3.

Table 3. Summary of main findings.

Included in review	Findings (N=15 studies; covering a total of 899 study participants)
Interventions duration	• Interventions lasted between 2 and 26 weeks
Used AI[a] technologies	• Recommender systems (described in 10/15 studies) • Conversational agents (described in 4/15 studies) • Human activity recognition (described in 1 study)
Human factors	• Acceptability (evaluated in 5/15 studies) • Satisfaction (evaluated in 5/15 studies) • Usability (evaluated in 4/15 studies) • Usefulness (evaluated in 2/15 studies) • Engagement (evaluated in 1 study) • User experience (evaluated in 1 study) • Task performance (evaluated in 1 study)
Effectiveness of AI-driven technologies for increasing PA[b]	• Moderate evidence: AI-driven digital technologies have the potential to increase PA (eg, number of steps, distance walked, or time spent on PA) • Very low evidence: Recommender systems and chatbots, seems to have the potential to influence changes in PA behavior

[a]AI: artificial intelligence.

[b]PA: physical activity.

Comparison With Previous Work

In the included studies, we recognized several benefits of AI integrated into digital solutions for increasing PA, such as the ability to adapt the solution to the patient's physical capacity, current activity, and psychological profile [8,11,30]. AI can monitor activity and inactivity and predict bodily occurrences, which is especially relevant for older people [40]. AI can also simulate the role of a personal trainer, provide guidance, form correction, and motivation [37] through voice- or text-based interactions. Users can receive real-time feedback and support during their workouts [8,10] which would be difficult to achieve with non-AI digital solutions. AI algorithms can analyze user data such as fitness levels, health conditions, and preferences and provide personalized exercise recommendations [11]. The activities or other suggestions are tailored to the specific needs and goals of the user, increasing the likelihood of adherence. Real-time feedback can be shared with the user. Previous studies found that activity tracking combined with real-time, personalized text messages can significantly increase PA and further affirm text messaging as an effective health behavior modifier [10-12]. However, in our review, researchers concluded that their solution did not achieve sufficient adherence to the exercise program [28,30]. The entire potential of personalization techniques has not yet been implemented in the solutions, as Luštrek et al [30] concluded that personalization, simplicity, ease of use, and avoiding information overload could be improved.

AI algorithms can continuously learn from user interactions and feedback to refine and improve the UX. This iterative process leads to more effective and engaging solutions over time. For example, the continuous interaction that chatbots can provide was reported to be useful in helping users increase regular PA and in helping them stay motivated to participate in PA [32]. Studies have already found that the human-likeness or anthropomorphism of a chatbot increase the likelihood that users comply with the chatbot's recommendations [18]. Roy and Naidoo [17] found that human qualities like warmth and competence are contributing to a positive UX and possibly to an increased adherence to the digital solution [17].

We only found 5 studies involving sensors to measure PA [26,30,32,35,38]. Dedicated fitness trackers seem to be more prominent to be involved in solutions increasing PA. Mobile-based activity tracking and smartwatches were only implemented in one solution. A reason might be that users prefer to use systems they already use; that is, integration with existing tools like fitness trackers is desired by users, as found by Wang et al [41]. The landscape of wearables and sensors that could be used for PA tracking is much larger than was found in our research [42]. The integration of sensors with AI could help analyze the data streams and promote an increase in PA [43]. Additionally, it could assist in monitoring PA among individuals affected by health conditions [44-46]. We hypothesize that existing research focuses on sensors that are well-known, not very intrusive, and therefore probably more accepted by users of solutions for increasing PA.

AI can gamify the fitness experience by setting challenges, goals, and rewards. Users are motivated to increase PA by earning points, competing with friends, or unlocking achievements. Xu et al [16] found in their review that gamification interventions could increase PA participation. Interestingly, none of our included studies explicitly reported about gamification elements.

Do We Have Enough Evidence on AI's Effectiveness in Increasing PA?

In total, 5 of the included studies provide moderate evidence of AI's effectiveness for increasing PA [28,32,36,38,39]. However, these studies involve short interventions lasting from 6 to 12 weeks. Hence, the significant effect might be influenced by this brief follow-up period, similar to other mHealth interventions [13]. The estimated time needed to form habits of complex behaviors such as exercise behavior is 12 weeks [47]. Thus, longer intervention studies are needed to assess the potential long-term effectiveness of AI-driven technologies for increasing PA.

Out of the 260 participants in these 5 studies [28,32,36,38,39], 72.3% (188) of them were women, and the majority were aged between 40 and 50 years. Further studies are needed to investigate the effects of these AI-driven technologies on participants with different sociodemographic characteristics, as well as those with health conditions for which exercise aids in managing the disease and preventing complications [1,2,4].

There is very limited and low-quality evidence supporting the impact of AI-driven technologies on changing PA behavior and the ability to perform such behavior [25,30,35,37]. In these cases, the durations of the interventions varied, ranging from as short as 3 to 4 weeks [25,37], to as long as 20 to 26 weeks [30,35]. Similar to previous cases, the majority of participants in these 4 studies were women, comprising 83.7% (82/98) reported participants. While research indicates that gender is one of the factors influencing the use of health-related technologies [48,49], technologies aimed at increasing PA should be tested, personalized, and accessible for all demographic groups.

What Is the Role of HFs on the AI-Based PA Solutions?

Most of the included AI-based PA systems showed positive results in terms of HFs related to their use. However, no study aimed to evaluate how the AI component could independently influence HFs such as user acceptance, perceived ease of use, or perceived usefulness. Many studies used AI techniques to personalize the PA system based on the authors' assumptions about the persuasive power of personalization that could lead to greater motivation and thus result in greater intention to use, adoption, and engagement. However, no study has tested these hypotheses. In this regard, more research is still needed to identify the role of AI components in HFs affecting PA systems. In addition, no study has focused on whether the inclusion of AI could lead to a change in the role of HFs, as has been the case with traditional technologies.

Limitations

There were some identified limitations in this scoping review. Even though we did not have a language limitation in the search strategy, all the included studies were in English. Therefore, we could have missed relevant AI-driven solutions published in other languages. The included studies were mainly from diverse high-income countries, restricting generalization to low- and middle-income countries. In addition, the studies included in the scoping review had an intervention period of a maximum of 26 weeks, showing only the short-term effect of the AI-driven solutions. All studies were included in the review, irrespective of the assessed quality of the evidence. However, the results of the included studies were reported separately according to the quality of the evidence, minimizing misinterpretation of the data.

Conclusions

This study synthesized current evidence on the effectiveness and potential of AI-driven digital solutions for increasing PA. Although the included studies offer valuable insights by demonstrating positive outcomes through various AI-driven technologies for enhancing PA, the evidence is still very limited. While some studies demonstrated moderate evidence of AI's effectiveness in increasing PA, these interventions were typically short-term. Longer-term studies are necessary to assess the sustained impact of AI-driven technologies on behavior change and habit formation. Additionally, further research is needed to investigate the effects of AI-driven interventions on diverse populations, including individuals with varying sociodemographic characteristics and conditions. Moreover, the evidence regarding the impact of AI-driven technologies on changing PA behavior remains limited and of low quality. There is a need for rigorous studies to evaluate the effectiveness of these interventions, particularly in terms of their ability to induce long-term behavior change. Furthermore, while most AI-based PA systems demonstrated positive results in terms of UX, there is a lack of research focusing on the independent influence of AI components on HFs, such as user acceptance and perceived usefulness. Additionally, more investigation is required to understand how the inclusion of AI may alter the role of HFs in PA systems compared to traditional technologies.

In conclusion, while AI-driven digital solutions hold significant promise for promoting PA and improving public health outcomes, addressing these limitations and challenges will be crucial for maximizing their effectiveness and accessibility. Continued research efforts in these areas are essential for advancing our understanding of the role of AI in PA promotion and ensuring the development of evidence-based interventions that benefit diverse populations.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full search strategy.
[DOCX File , 16 KB - humanfactors_v11i1e55964_app1.docx]

Multimedia Appendix 2
Excluded papers in full text eligibility phase.
[DOCX File , 22 KB - humanfactors_v11i1e55964_app2.docx]

Multimedia Appendix 3
PRISMA checklist.
[PDF File (Adobe PDF File), 84 KB - humanfactors_v11i1e55964_app3.pdf]

References

1. Schuch FB, Vancampfort D. Physical activity, exercise, and mental disorders: it is time to move on. Trends Psychiatry Psychother 2021;43(3):177-184. [doi: 10.47626/2237-6089-2021-0237] [Medline: 33890431]
2. Global action plan on physical activity 2018-2030: more active people for a healthier world. World Health Organization. 2018. URL: https://apps.who.int/iris/bitstream/handle/10665/272722/9789241514187-eng.pdf?sequence=1&isAllowed=y [accessed 2024-06-06]
3. Rhodes RE, Janssen I, Bredin SS, Warburton DE, Bauman A. Physical activity: health impact, prevalence, correlates and interventions. Psychol Health 2017;32(8):942-975. [doi: 10.1080/08870446.2017.1325486] [Medline: 28554222]
4. McGarrigle L, Todd C. Promotion of physical activity in older people using mHealth and eHealth technologies: rapid review of reviews. J Med Internet Res 2020;22(12):e22201 [FREE Full text] [doi: 10.2196/22201] [Medline: 33372894]
5. McIntosh J, Jay S, Hadden N, Whittaker P. Do E-health interventions improve physical activity in young people: a systematic review. Public Health 2017;148:140-148. [doi: 10.1016/j.puhe.2017.04.001] [Medline: 28482322]
6. Bickmore TW, Schulman D, Sidner C. Automated interventions for multiple health behaviors using conversational agents. Patient Educ Couns 2013;92(2):142-148 [FREE Full text] [doi: 10.1016/j.pec.2013.05.011] [Medline: 23763983]
7. Laranjo L, Ding D, Heleno B, Kocaballi B, Quiroz JC, Tong HL, et al. Do smartphone applications and activity trackers increase physical activity in adults? Systematic review, meta-analysis and metaregression. Br J Sports Med 2021;55(8):422-432. [doi: 10.1136/bjsports-2020-102892] [Medline: 33355160]
8. An R, Shen J, Wang J, Yang Y. A scoping review of methodologies for applying artificial intelligence to physical activity interventions. J Sport Health Sci 2024;13(3):428-441 [FREE Full text] [doi: 10.1016/j.jshs.2023.09.010] [Medline: 37777066]
9. Catellani P, Biella M, Carfora V, Nardone A, Brischigiaro L, Manera MR, et al. A theory-based and data-driven approach to promoting physical activity through message-based interventions. Front Psychol 2023;14:1200304 [FREE Full text] [doi: 10.3389/fpsyg.2023.1200304] [Medline: 37575427]
10. Vandelanotte C, Trost S, Hodgetts D, Imam T, Rashid M, To QG, et al. Increasing physical activity using an just-in-time adaptive digital assistant supported by machine learning: a novel approach for hyper-personalised mHealth interventions. J Biomed Inform 2023;144:104435 [FREE Full text] [doi: 10.1016/j.jbi.2023.104435] [Medline: 37394024]
11. Hassoon A, Baig Y, Naiman DQ, Celentano DD, Lansey D, Stearns V, et al. Randomized trial of two artificial intelligence coaching interventions to increase physical activity in cancer survivors. NPJ Digit Med 2021;4(1):168 [FREE Full text] [doi: 10.1038/s41746-021-00539-9] [Medline: 34887491]
12. Martin SS, Feldman DI, Blumenthal RS, Jones SR, Post WS, McKibben RA, et al. mActive: A randomized clinical trial of an automated mHealth intervention for physical activity promotion. J Am Heart Assoc 2015;4(11):e002239 [FREE Full text] [doi: 10.1161/JAHA.115.002239] [Medline: 26553211]
13. Mönninghoff A, Kramer JN, Hess AJ, Ismailova K, Teepe GW, Tudor Car L, et al. Long-term effectiveness of mHealth physical activity interventions: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2021;23(4):e26699 [FREE Full text] [doi: 10.2196/26699] [Medline: 33811021]
14. Oh YJ, Zhang J, Fang M, Fukuoka Y. A systematic review of artificial intelligence chatbots for promoting physical activity, healthy diet, and weight loss. Int J Behav Nutr Phys Act 2021;18(1):160 [FREE Full text] [doi: 10.1186/s12966-021-01224-6] [Medline: 34895247]
15. Fang J, Lee VCS, Ji H, Wang H. Enhancing digital health services: a machine learning approach to personalized exercise goal setting. ArXiv, arXiv:2204.00961 2022. [doi: 10.48550/arXiv.2204.00961]

16. Xu L, Shi H, Shen M, Ni Y, Zhang X, Pang Y, et al. The effects of mHealth-based gamification interventions on participation in physical activity: systematic review. JMIR Mhealth Uhealth 2022;10(2):e27794 [FREE Full text] [doi: 10.2196/27794] [Medline: 35113034]
17. Roy R, Naidoo V. Enhancing chatbot effectiveness: the role of anthropomorphic conversational styles and time orientation. Journal of Business Research 2021;126:23-34. [doi: 10.1016/j.jbusres.2020.12.051]
18. Adam M, Wessel M, Benlian A. AI-based chatbots in customer service and their effects on user compliance. Electron Markets 2020;31(2):427-445. [doi: 10.1007/s12525-020-00414-7]
19. Kushniruk AW, Borycki EM. Human factors in healthcare IT: management considerations and trends. Healthc Manage Forum 2023;36(2):72-78 [FREE Full text] [doi: 10.1177/08404704221139219] [Medline: 36847593]
20. Bergevi J, Andermo S, Woldamanuel Y, Johansson U, Hagströmer M, Rossen J. User perceptions of eHealth and mHealth services promoting physical activity and healthy diets: systematic review. JMIR Hum Factors 2022;9(2):e34278 [FREE Full text] [doi: 10.2196/34278] [Medline: 35763339]
21. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
22. EndNote 2024. Clarivate. URL: https://endnote.com/ [accessed 2024-06-06]
23. Rayyan 2022. Rayyan. URL: https://www.rayyan.ai/ [accessed 2024-06-06]
24. Murad MH, Mustafa RA, Schünemann HJ, Sultan S, Santesso N. Rating the certainty in evidence in the absence of a single estimate of effect. Evid Based Med 2017;22(3):85-87 [FREE Full text] [doi: 10.1136/ebmed-2017-110668] [Medline: 28320705]
25. Rabbi M, Pfammatter A, Zhang M, Spring B, Choudhury T. Automated personalized feedback for physical activity and dietary behavior change with mobile phones: a randomized controlled trial on adults. JMIR Mhealth Uhealth 2015;3(2):e42 [FREE Full text] [doi: 10.2196/mhealth.4160] [Medline: 25977197]
26. Rabbi M, Aung MS, Gay G, Reid MC, Choudhury T. Feasibility and acceptability of mobile phone-based auto-personalized physical activity recommendations for chronic pain self-management: pilot study on adults. J Med Internet Res 2018;20(10):e10147 [FREE Full text] [doi: 10.2196/10147] [Medline: 30368433]
27. Fadhil A, Wang Y, Reiterer H. Assistive conversational agent for health coaching: a validation study. Methods Inf Med 2019;58(1):9-23. [doi: 10.1055/s-0039-1688757] [Medline: 31117129]
28. Davis CR, Murphy KJ, Curtis RG, Maher CA. A process evaluation examining the performance, adherence, and acceptability of a physical activity and diet artificial intelligence virtual health assistant. Int J Environ Res Public Health 2020;17(23):9137 [FREE Full text] [doi: 10.3390/ijerph17239137] [Medline: 33297456]
29. Joo S, Lee C, Joo N, Kim C. Feasibility and effectiveness of a motion tracking-based online fitness program for office workers. Healthcare (Basel) 2021;9(5):584 [FREE Full text] [doi: 10.3390/healthcare9050584] [Medline: 34068929]
30. Luštrek M, Bohanec M, Cavero Barca C, Ciancarelli MC, Clays E, Dawodu AA, et al. A personal health system for self-management of congestive heart failure (HeartMan): development, technical evaluation, and proof-of-concept randomized controlled trial. JMIR Med Inform 2021;9(3):e24501 [FREE Full text] [doi: 10.2196/24501] [Medline: 33666562]
31. Pelle T, van der Palen J, de Graaf F, van den Hoogen FHJ, Bevers K, van den Ende CHM. Use and usability of the dr. Bart app and its relation with health care utilisation and clinical outcomes in people with knee and/or hip osteoarthritis. BMC Health Serv Res 2021;21(1):444 [FREE Full text] [doi: 10.1186/s12913-021-06440-1] [Medline: 33971861]
32. To QG, Green C, Vandelanotte C. Feasibility, usability, and effectiveness of a machine learning-based physical activity chatbot: quasi-experimental study. JMIR Mhealth Uhealth 2021;9(11):e28577 [FREE Full text] [doi: 10.2196/28577] [Medline: 34842552]
33. Lin CC, Kuo LC, Lin YS, Chang CM, Hu FW, Chen YJ, et al. AIoT-based ergometer for physical training in frail elderly with cognitive decline: a pilot randomized control trial. J. Med. Biol. Eng 2022;42(6):909-921. [doi: 10.1007/s40846-022-00759-8]
34. Park J, Chung SY, Park JH. Real-time exercise feedback through a convolutional neural network: a machine learning-based motion-detecting mobile exercise coaching application. Yonsei Med J 2022;63(Suppl):S34-S42 [FREE Full text] [doi: 10.3349/ymj.2022.63.S34] [Medline: 35040604]
35. Seok JW, Kwon Y, Lee H. Feasibility and efficacy of TouchCare system using application for older adults living alone: a pilot pre-experimental study. BMC Geriatr 2022;22(1):799 [FREE Full text] [doi: 10.1186/s12877-022-03482-w] [Medline: 36474186]
36. Bates NA, Huffman A, Goodyear E, Nagai T, Rigamonti L, Breuer L, et al. Physical clinical care and artificial-intelligence-guided core resistance training improve endurance and patient-reported outcomes in subjects with lower back pain. Clin Biomech (Bristol, Avon) 2023;103:105902. [doi: 10.1016/j.clinbiomech.2023.105902] [Medline: 36805199]
37. Thiengwittayaporn S, Wattanapreechanon P, Sakon P, Peethong A, Ratisoontorn N, Charoenphandhu N, et al. Development of a mobile application to improve exercise accuracy and quality of life in knee osteoarthritis patients: a randomized controlled trial. Arch Orthop Trauma Surg 2023;143(2):729-738 [FREE Full text] [doi: 10.1007/s00402-021-04149-8] [Medline: 34453570]

38. Maher CA, Davis CR, Curtis RG, Short CE, Murphy KJ. A physical activity and diet program delivered by artificially intelligent virtual health coach: proof-of-concept study. JMIR Mhealth Uhealth 2020;8(7):e17558 [FREE Full text] [doi: 10.2196/17558] [Medline: 32673246]
39. Rabbi M, Aung M, Zhang M, Choudhury T. MyBehavior: automatic personalized health feedback from user behaviors and preferences using smartphones. JAPAN: OSAKA; 2015:707-718.
40. Bhatia M. An AI-enabled secure framework for enhanced elder healthcare. Engineering Applications of Artificial Intelligence 2024;131:107831. [doi: 10.1016/j.engappai.2023.107831]
41. Wang B, Asan O, Zhang Y. Shaping the future of chronic disease management: insights into patient needs for AI-based homecare systems. Int J Med Inform 2024;181:105301. [doi: 10.1016/j.ijmedinf.2023.105301] [Medline: 38029700]
42. Kazanskiy NL, Khonina SN, Butt MA. A review on flexible wearables – recent developments in non-invasive continuous health monitoring. Sensors and Actuators A: Physical 2024;366:114993. [doi: 10.1016/j.sna.2023.114993]
43. Wang J, Yang Y, Liu H, Jiang L. Enhancing the college and university physical education teaching and learning experience using virtual reality and particle swarm optimization. Soft Comput 2023;28(2):1277-1294. [doi: 10.1007/s00500-023-09528-4]
44. Gupta N, Kasula V, Sanmuganathan P, Panico N, Dubin AH, Sykes DA, et al. SmartWear body sensors for neurological and neurosurgical patients: a review of current and future technologies. World Neurosurg X 2024;21:100247 [FREE Full text] [doi: 10.1016/j.wnsx.2023.100247] [Medline: 38033718]
45. Carter LG, Ford CD. Promoting physical activity in clinical practice through wearable technology. J Am Assoc Nurse Pract 2023;35(12):765-769. [doi: 10.1097/JXX.0000000000000892] [Medline: 37249382]
46. Coccia M. Deep learning technology for improving cancer care in society: new directions in cancer imaging driven by artificial intelligence. Technology in Society 2020;60:101198. [doi: 10.1016/j.techsoc.2019.101198]
47. Lally P, van Jaarsveld CHM, Potts HWW, Wardle J. How are habits formed: modelling habit formation in the real world. Euro J Social Psych 2009;40(6):998-1009. [doi: 10.1002/ejsp.674]
48. Yang Y, Koenigstorfer J. Determinants of fitness app usage and moderating impacts of education-, motivation-, and gamification-related app features on physical activity intentions: cross-sectional survey study. J Med Internet Res 2021;23(7):e26063 [FREE Full text] [doi: 10.2196/26063] [Medline: 34255656]
49. Lupton D, Maslen S. How women use digital technologies for health: qualitative interview and focus group study. J Med Internet Res 2019;21(1):e11481 [FREE Full text] [doi: 10.2196/11481] [Medline: 30681963]

Abbreviations

AI: artificial intelligence
GRADE: Grading of Recommendations Assessment, Development and Evaluation
HF: human factor
mHealth: mobile health
PA: physical activity
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
UX: user experience

Edited by A Kushniruk, E Borycki; submitted 31.12.23; peer-reviewed by G Caggianese, M Coccia, G Farid; comments to author 18.02.24; revised version received 02.04.24; accepted 05.05.24; published 03.07.24.

Please cite as:
Gabarron E, Larbi D, Rivera-Romero O, Denecke K
Human Factors in AI-Driven Digital Solutions for Increasing Physical Activity: Scoping Review
JMIR Hum Factors 2024;11:e55964
URL: https://humanfactors.jmir.org/2024/1/e55964
doi:10.2196/55964
PMID:

Original Paper

An Artificial Intelligence–Based App for Self-Management of Low Back and Neck Pain in Specialist Care: Process Evaluation From a Randomized Clinical Trial

Anna Marcuzzi[1,2], PhD; Nina Elisabeth Klevanger[1,2], PhD; Lene Aasdahl[1,3], PhD; Sigmund Gismervik[1], PhD; Kerstin Bach[4], PhD; Paul Jarle Mork[1], PhD; Anne Lovise Nordstoga[5], PhD

[1]Department of Public Health and Nursing, Norwegian University of Science and Technology, Trondheim, Norway
[2]Department of Physical Medicine and Rehabilitation, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway
[3]Unicare Helsefort Rehabilitation Center, Rissa, Norway
[4]Department of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway
[5]Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology, Trondheim, Norway

Corresponding Author:
Anna Marcuzzi, PhD
Department of Public Health and Nursing
Norwegian University of Science and Technology
Håkon Jarls gate 11
Trondheim, 7030
Norway
Phone: 47 73412378
Email: anna.marcuzzi@ntnu.no

Abstract

Background: Self-management is endorsed in clinical practice guidelines for the care of musculoskeletal pain. In a randomized clinical trial, we tested the effectiveness of an artificial intelligence–based self-management app (selfBACK) as an adjunct to usual care for patients with low back and neck pain referred to specialist care.

Objective: This study is a process evaluation aiming to explore patients' engagement and experiences with the selfBACK app and specialist health care practitioners' views on adopting digital self-management tools in their clinical practice.

Methods: App usage analytics in the first 12 weeks were used to explore patients' engagement with the SELFBACK app. Among the 99 patients allocated to the SELFBACK interventions, a purposive sample of 11 patients (aged 27-75 years, 8 female) was selected for semistructured individual interviews based on app usage. Two focus group interviews were conducted with specialist health care practitioners (n=9). Interviews were analyzed using thematic analysis.

Results: Nearly one-third of patients never accessed the app, and one-third were low users. Three themes were identified from interviews with patients and health care practitioners: (1) overall impression of the app, where patients discussed the interface and content of the app, reported on usability issues, and described their app usage; (2) perceived value of the app, where patients and health care practitioners described the primary value of the app and its potential to supplement usual care; and (3) suggestions for future use, where patients and health care practitioners addressed aspects they believed would determine acceptance.

Conclusions: Although the app's uptake was relatively low, both patients and health care practitioners had a positive opinion about adopting an app-based self-management intervention for low back and neck pain as an add-on to usual care. Both described that the app could reassure patients by providing trustworthy information, thus empowering them to take actions on their own. Factors influencing app acceptance and engagement, such as content relevance, tailoring, trust, and usability properties, were identified.

Trial Registration: ClinicalTrials.gov NCT04463043; https://clinicaltrials.gov/study/NCT04463043

(JMIR Hum Factors 2024;11:e55716) doi:10.2196/55716

KEYWORDS

low back pain; neck pain; self-management; smartphone app; process evaluation; focus group; focus groups; musculoskeletal; mHealth; mobile health; app; apps; applications; usage; interview; interviews; qualitative; engagement

Introduction

Low back pain and neck pain are the main causes of disability worldwide [1]. Up to 30% of patients with acute or recurrent disability develop persistent pain [2,3]. Patients with persistent pain are often work-disabled and might need specialist assessment, which further increases health care and societal costs [4]. Given the highly prevalent and costly nature of low back and neck pain, enabling patients to self-manage constitutes an important strategy for reducing the individual and societal burden.

Self-management is commonly defined as an individual's ability to actively monitor own health condition, adapt to physical and psychological demands, and implement lifestyle changes [5]. While self-management is endorsed in clinical practice guidelines to manage musculoskeletal pain [6], self-management support offered in clinical practice, for example, primary and specialist care, remains suboptimal [7,8]. Digital interventions such as smartphone apps can be a viable mode for delivering self-management support as an add-on to usual care due to their accessibility and possibility of making evidence-based advice easily available to patients.

We recently reported results from a randomized clinical trial (RCT) testing the effectiveness of an artificial intelligence (AI)–based self-management app (selfBACK) as an adjunct to usual care for patients with low back and neck pain in specialist care [9]. The app uses the case-based reasoning methodology, which is a branch of knowledge-driven AI [10] providing individually tailored self-management recommendations to users. Although individual tailoring is considered as an important feature for engagement in self-management interventions [11], the RCT did not show the SELFBACK app to be more effective than usual care alone or a web-based self-management intervention in improving self-reported musculoskeletal health. The aim of this study was to explore patients' engagement and experiences with the selfBACK app and specialist health care practitioners' views on adopting such digital self-management tools in their clinical practice.

Methods

Study Design and Context

This study is a process evaluation carried out in parallel with the RCT [9]. The qualitative part of the study is reported according to the consolidated criteria for reporting qualitative research (COREQ) [12].

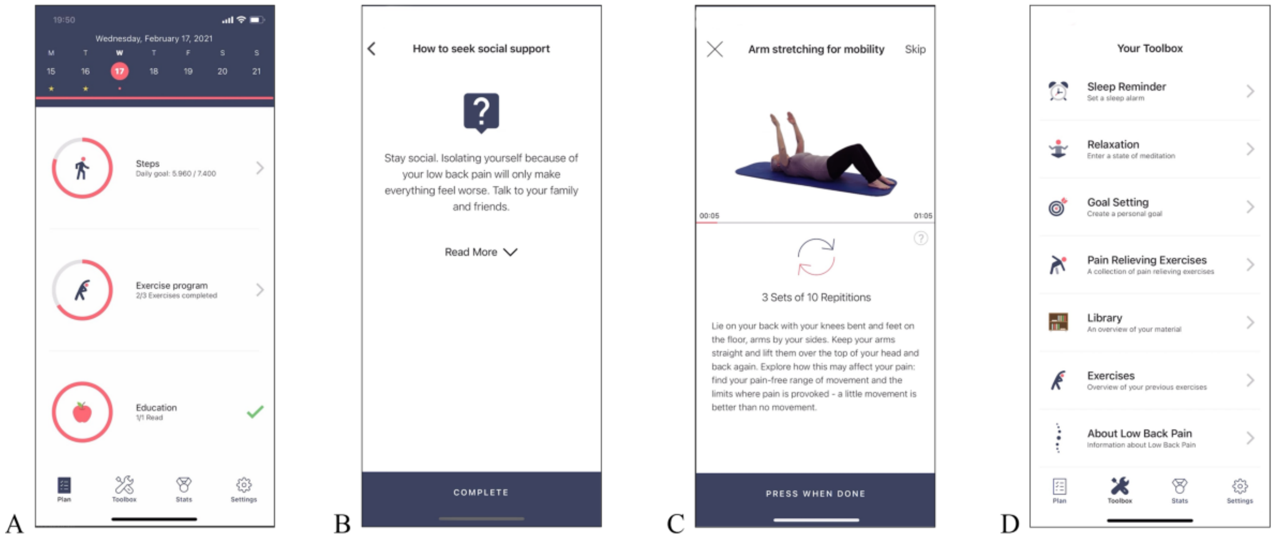
Recruitment for the RCT took place at the multidisciplinary outpatient clinic for back-, neck-, and shoulder pain at St Olavs Hospital, Trondheim, Norway. Patients with low back and neck pain who were referred and on a waiting list for a consultation at the clinic were invited to the study via SMS text message. Interested and eligible patients were subsequently randomized to (1) the selfBACK app adjunct to usual care (99/294, 33.7%); (2) the e-Help (a self-management website) adjunct to usual care (98/294, 33.3%); and (3) usual care only (97/294, 33.0%). Usual care consisted of a waiting period of approximately 6-8 weeks before a consultation, including a clinical examination, followed by recommendations for suitable treatment. The recommendations could vary from no further treatment and adjusted recommendations for primary care treatment to outpatient multimodal rehabilitation or referral for surgery.

The process evaluation consisted of descriptive data analytics on app usage and semistructured interviews involving patients allocated to the selfBACK intervention. In addition, health care practitioners at the outpatient clinic were invited to participate in focus group interviews with the purpose of exploring their views on adopting digital tools for self-management support in their clinical practice. While health care practitioners were aware of the trial, they were not provided any specific instructions in relation to the trial conduct.

SELFBACK App as an Adjunct to Usual Care

The SELFBACK intervention was developed using intervention mapping [13] and underwent iterative pilot-testing before the final version was released [14,15]. The SELFBACK is an AI-based self-management app that provides users with weekly and individually tailored plans encompassing physical activity recommendations, strength and flexibility exercises, and educational messages (updated daily). In addition, the app contains a toolbox, which is a static component of the app containing, for example, goal-setting tool, mindfulness audios, pain-relieving exercises, and sleep reminders that patients can access at their own convenience [16] (Figure 1). The tailoring of patient recommendations delivered via the app relies on the application of case-based reasoning [10], a knowledge-driven AI methodology. In this methodology, knowledge from previous similar successful patient cases is reused to offer patient-centered and tailored recommendations. Thereby, new and similar patient cases receive recommendations based on what has or has not been successful in previous patient cases [17]. The AI system uses weekly reports (eg, symptom progression) and information collected through the app (eg, exercise completion and number of steps) to personalize the self-management recommendations. Patients can collect badges and rewards within the app by adhering to weekly recommendations. Push notifications are triggered by patients' self-management behavior (eg, completion of exercises) and sent via the app to motivate and reinforce the desired self-management behavior.

Figure 1. SELFBACK app screen views. (A) Home screen containing the 3 main components of the app, that is, steps achieved, exercise completion, and educational messages read. (B) Example of an educational message. (C) Example of exercise. (D) Toolbox screen containing additional resources.



The selfBACK app was offered as an adjunct to usual care. Patients randomized to the SELFBACK group were sent an SMS with a link to download the app. They were also provided with an installation guide and contact information of the research team if they had any access issues. Instructions on how to use the app and its content was provided within the app and patients had unrestricted access throughout the 6-month study period.

App Usage Data Analytics

Data on app usage consisted of information about number of weekly plans generated, number of app access per week, and specific content visited. For weekly plans to be generated, the patient needed to access the app and complete the weekly short-tailoring questionnaire (eg, questions on pain intensity, self-efficacy level, and fear avoidance level). The number of weekly plans generated was used to dichotomize patients into moderate or high users and low users as basis for a purposive recruitment for interviews. Moderate or high use was defined as generating at least 6 plans during the first 12 weeks after first access of the app, while low use was defined as generating less than 6 plans as described previously [9]. This information together with the number of app access per week was retrieved from the back end of the AI system, which has information when users actively interact with the app (completing exercises, tailoring sessions, or similar). Information about number of days a specific content was visited per week (eg, exercises, educational component, and toolbox) was retrieved from Matomo [18], a free and open-source software that records whenever a user accesses a screen of the app.

Interviews With Patients and Health Care Practitioners

A purposive sample of 15 patients was contacted by phone and invited for interviews according to their app usage (ie, number of weekly plans generated). Of these, 3 declined participation and 1 did not answer. A total of 11 patients were interviewed (aged 27-75 years, 8 female), of whom 7 were moderate or high app users and 4 were low users.

Health care practitioners from the multidisciplinary outpatient clinic (n=11) were informed about the study during a regular staff meeting and invited to participate in focus group interviews. Overall, 9 health care practitioners expressed an interest in participating and were included. Two focus groups were formed based on the role the health care practitioners had at the clinic. One focus group (focus group 1) included 3 physiotherapists and 2 social workers (aged 32-51 years, 4 female) with 5-13 years of working experience at the outpatient clinic. The other group (focus group 2) included 4 physicians (aged 32-42 years, 3 female) with working experience at the outpatient clinic ranging from 1 week to 7 years.

The interviews with patients and health care practitioners were performed by a research assistant with prior experience of conducting qualitative interviews and no prior relationship with patients or health care practitioners. Patients were interviewed individually via Skype between January and February 2021. The interview guide was semistructured and developed using the Normalization Process Theory [19,20]. The questions included background information, the motivation to join the study, how pain was managed before the study, what facilitated or hindered the use of the app, how the app was integrated in daily life, future intentions to use the app, and general thoughts about self-management. Questions were adapted when needed and follow-up questions added where appropriate. Each interview lasted approximately 45 minutes and was recorded with the patient's permission. One interview was repeated due to failure of the recording.

The focus groups took place digitally via the Zoom (Zoom Video Communications) platform in February 2021. Prior to the focus groups, health care practitioners were provided with an overview of the selfBACK app by the research team and access to the app. They were asked about their initial impressions of the selfBACK app, their views on digital tools to support self-management, whether and how they would use them in clinical practice, what potential benefits and risks such tools entailed, whether they believed that using them could affect their professional autonomy, and whether such digital tools could be trusted. Each focus group lasted approximately 90 minutes and was facilitated by 2 research assistants (one acting as an observer). At the end of each focus group, the 2 researchers exchanged experiences on the interaction among health care practitioners and these were annotated. Both interviews and focus groups were audio recorded with the permission of all participants and transcribed verbatim. Data were de-identified during transcription and used thereafter for data analysis.

Data Analysis

Baseline characteristics of all patients were reported descriptively. Interviews and focus groups data were analyzed using thematic analysis [20]. First, ALN and AM read and coded the interview transcripts with the support from the research assistant who transcribed the interviews to ensure that coding was reflective of the material. The codes were then grouped into themes by a process of constantly deliberating their content and boundaries, resulting in 2 coding trees for the patients and health care practitioners, respectively. These were subsequently discussed with 2 researchers in the team (NK and LA) who had read all the interview transcripts. As the 2 coding trees were found to largely contain complementary themes, they were combined before writing up the results. Quotations were added to either exemplify or nuance the analytic text.

Ethical Considerations

The RCT was registered in ClinicalTrials.gov (NCT04463043), and the protocol, including the description of the process evaluation, was published [21]. Ethics approval was granted by the Regional Committee for Medical and Health Research Ethics in Central Norway (reference 64084) and the Norwegian Medicines Agency (reference 20/10329-10). All patients provided written informed consent before entering the study.

Health care practitioners were provided with oral information about the study and verbal informed consent was obtained from them before the focus groups were conducted.

Results

App Usage

The demographic and clinical characteristics of the 99 patients allocated to the selfBACK intervention stratified by app usage are shown in Table 1. Overall, patients' characteristics were similar across groups, although a greater proportion of patients who never accessed the app or who were low users reported having daily pain as well as having both neck and back pain compared with moderate or high users who reported pain less frequently and predominantly pain at the lower back (Table 1).

Table 1. Demographic and clinical characteristics of the 99 patients allocated to SELFBACK intervention, stratified by app usage.

	Never accessed app	Low usage group[a]	Moderate or high usage group[b]
Participants, n (%)	29 (29.3)	32 (32.3)	38 (38.4)
Age (years), mean (SD)	52.3 (11.5)	47.8 (14.2)	50.8 (16.4)
Women, n (%)	17 (58.6)	19 (59.4)	24 (63.2)
Education (years), n (%)			
>12	16 (26.2)	19 (31.2)	26 (42.6)
10-12	9 (32.1)	10 (35.7)	9 (32.1)
<10	4 (40.0)	3 (30.0)	3 (30.0)
Full-time or part-time employment, n (%)	19 (65.5)	25 (78.1)	26 (68.4)
Married or living with partner, n (%)	20 (69.0)	23 (71.9)	29 (76.3)
Pain localization, n (%)			
Low back pain	13 (25.0)	16 (30.8)	23 (44.2)
Neck pain	6 (24.0)	11 (44.0)	8 (32.0)
Neck and low back pain	10 (45.5)	5 (22.7)	7 (31.8)
Days with pain past year, n (%)			
≤30 days	3 (42.9)	1 (14.2)	3 (42.9)
>30 days but not every day	7 (18.4)	10 (26.3)	21 (55.3)
Every day	19 (35.2)	21 (38.9)	14 (25.9)
Use of pain medication (days per week), n (%)			
None	8 (28.6)	9 (32.1)	11 (39.3)
1-2 days	8 (40.0)	5 (25.0)	7 (35.0)
3-5 days	7 (31.8)	6 (27.3)	9 (40.9)
Daily	6 (20.7)	12 (41.4)	11 (37.9)
Musculoskeletal Health Questionnaire (score range 0-56), mean (SD)	30.9 (9.0)	28.4 (8.6)	30.8 (9.6)
Average pain intensity level past week[c] (score range 0-10), mean (SD)	5.7 (2.4)	5.5 (1.6)	4.8 (2.0)
Worst pain intensity level past week[c] (score range 0-10), mean (SD)	6.9 (2.3)	7.3 (1.6)	6.3 (2.1)
Health-related quality of life[d] (score range 0-100), mean (SD)	52.7 (20.6)	56.3 (14.1)	57.5 (19.2)
Pain Self-Efficacy Questionnaire (score range 0-60), mean (SD)	37.8 (13.7)	36.5 (14.0)	39.3 (12.1)

[a]Low usage group comprises patients who generated less than 6 of 12 weekly plans.

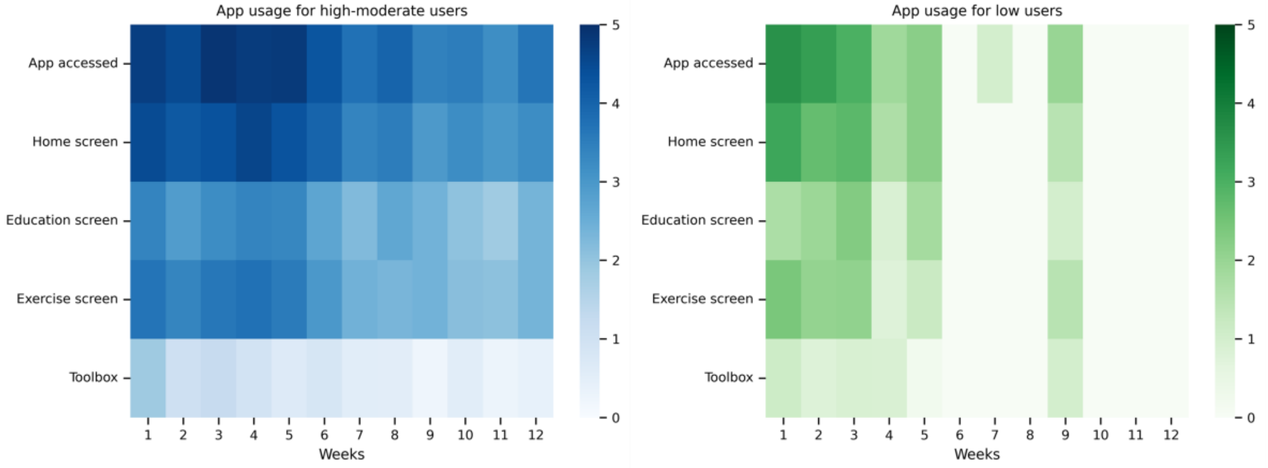
[b]Moderate or high usage group comprises patients who generated 6 of 12 weekly plans.

[c]Measured by the Numeric Rating Scale.

[d]Measured by the Visual Analogue Scale in the EQ-5D.

Figure 2 shows usage of each app components in the first 12 weeks stratified by usage group. Patients in the moderate or high usage group accessed the app, on average, most days of the week (ie, 4-5 days) for the first 5 weeks and somewhat reduced the frequency of weekly app access thereafter. The most visited content was the exercises, followed by the educational messages which were accessed, on average, more than once a week throughout the 12 weeks. The toolbox (ie, the static component of the app) was visited the least (Figure 2). Patients in the low usage group accessed the app, on average, fewer days per week from the beginning and mostly discontinued use after 5 weeks (Figure 2).

Figure 2. Average number of days each app component was visited in the first 12 weeks. Moderate or high users are shown on the left and low users are shown on the right. The bar to the right of each graph indicates the shading according to the average number of days each component was visited. App-accessed row shows any interaction with the app. Home screen, education screen, exercise screen, and toolbox rows represent the main components of the app shown in Figures 1A-D. Only those users who accessed the app at least once (n=70) are shown.



Interviews With Patients and Health Care Practitioners

We found that the interviews with patients and health care practitioners generated valuable information on their overall impression of the app, its perceived value, and suggestions for future use.

Theme 1: Overall Impression of the App

Patients discussed the interface and content of the app, reported on usability issues, and described their app usage.

App Interface and Content

Most patients described the app's user interface as simple and intuitive, with a user-friendly layout that was easy to navigate. Both patients and health care practitioners appreciated how the information was presented, for example, short and structured, and receiving a weekly plan was a convenient feature some patients valued.

> *You get a full program. You don't have to make it yourself. You are reminded and guided through the exercises so that you do not have to think and remember and count.* [Female, 50 years, low user]

When describing the content of the app, patients in both usage groups talked extensively about the exercise component. Indeed, being instructed on specific exercises was one of the main motivations for joining the study for most, and the integrated exercise videos were a feature they highly valued. Many also described that while they were often familiar with the educational content in the app, this nevertheless served as valuable reminders in their daily life. However, being familiar with the exercises and the educational content in the selfBACK app also resulted in some patients becoming unmotivated. In addition, some patients found the app's content to be tailored to their individual needs, while others believed that the content was of low relevance to them.

> *I thought that the exercises were very neutral, and I didn't get an exercise that suited my injury.* [Female, 46 years, low user]

Usability

Many patients in both user groups reported various technical problems when using the app. In the low usage group, some found the onboarding procedure burdensome and lost motivation to use the app. Several patients also had technical difficulties with step synchronization (ie, between the selfBACK app and health apps registering steps), which made them frustrated as notifications encouraging physical activity felt inappropriate when they had been active, and since they did not get validation when achieving their goals. In addition, some patients mentioned difficulties in scoring the weekly questions about their symptoms and function or not finding them relevant, although this issue did not prevent them from using the app. One patient also reported that the audio in the mindfulness section was difficult to hear due to a hearing impairment, and even though this feature felt relevant, they could not use it.

Usage Behavior

When describing how they used the app, most patients in the high usage group said that they used it regularly at the beginning and then gradually discontinued use. Some reported that, over time, they felt less need to access the app content and register their activity, mainly since their pain symptoms subsided.

> *My back improved, and I don't have the same need to use it and getting that recognition for progression and so on.* [Male, 27 years, high user]

Some patients commented that they performed the exercises in the app in combination with other exercises they already knew from before or with pain-relieving exercises in the toolbox (a static library of exercises within the app).

Theme 2: Perceived Value

Patients and health care practitioners described the primary value of the app as providing reassurance to patients by offering trustworthy information, thereby empowering them to self-manage. They also described the potential of the app to supplement usual care.

Providing Reassurance

Several patients described feeling reassured by the information in the app and experiencing exercises as manageable and not harmful.

> *These three things [components on the main page of the app] together give you some input at least during a period where you're unsure of what you can do, because it hurts really bad. And then the app comes with a little advice: okay, even if you're in pain then it doesn't get worse or, yes, it's unlikely to get much worse. And I believe that has helped.* [Female, 54 years, high user]

Most patients reflected on how recommendations encouraging activity in the selfBACK app also aligned with advice from chiropractors and physiotherapists they had consulted previously. As such, the coherence between the information in the app and health care personnel reassured them that the information and exercises could be trusted. On the contrary, 1 patient experienced that her health care practitioner did not endorse the advice provided by the app and, as a result, she discontinued using it, feeling insecure about its appropriateness for her situation.

> *I showed my therapist [chiropractor] the exercise, and he said right away: "you shouldn't do that exercise. Because then you make it worse for yourself".* [Female, 46 years, low user]

Health care practitioners underlined the need to present patients with reassuring language. The app was described as an opportunity to prevent patients from getting information from unreliable sources on the internet and provide patients with up-to-date, reliable, and consistent information that reinforced their message.

> *Knowing that the information is given via the health service can be reassuring, and they [the patients] may be more confident that it is nuanced and correct. They can go back and see "yes, that's consistent with what the doctor said".* [Female, focus group 2]

In line with this, some patients also described how their confidence in digital tools would increase if a health care practitioner or other trustworthy sources had developed or endorsed them in contrast with commercial parties.

One patient also described valuing how an app was made specifically for her health condition and that digital tools in general made her feel taken care of and included.

> *You can say that even if it's a robot, just the fact that you get a message, one feels taken care of in some way. And feeling like you're not alone in what you're struggling with, the app becomes a symbol that there are many others who are struggling with it [musculoskeletal pain].* [Female, 50 years, low user]

Empowering Patients to Self-Manage

Some patients described that using the app supported them on the road to becoming more active and that their confidence and thoughts about self-management increased while using it.

> *I'm just going to have to try on my own now. What's working and what's not working. It's not that I'm afraid something's dangerous anymore.* [Female, 44 years, high user]

Some also pointed out that although the information about self-management was perhaps well known to them, it nevertheless encouraged the thought of being able to act on one's own.

> *It has helped to think a little more positively and, yes, that you can do a lot yourself. This kind of things you know deep down, but it's about getting a little help to put your thoughts on the right track.* [Female, 54 years, high user]

Some patients described features such as reminders and activity tracking as positive influences on motivation to be active, and these also served as a reminder of how much they achieved. Health care practitioners similarly believed that digital tools such as the selfBACK app could help encourage patients' active participation in rehabilitation. The interactive features (eg, goal setting) and accessibility of the app were also mentioned as elements that can promote and reinforce self-management behaviors.

> *A nice thing is that you could make your own personal goal […], you have to take a position on some questions. For example, you are asked a lot about this goal, whether it is realistic, how long it should last. So, you have to make some active choices.* [Male, focus group 1]

One physiotherapist described this active approach as taking responsibility instead of clientification, a point also reflected in the statements of many patients.

> *There is something that we hope [to achieve] in collaboration with the patient, which is to make them accountable, so that the patient sits in the driver seat.* [Female, focus group 1]

Supplementing Usual Care

Both patients and health care practitioners believed that the selfBACK app could be a valuable supplement to usual care. When reflecting on how digital tools can help in taking responsibility for one's health, several described it as a necessity and solution to the increasing pressure on health care services.

Adopting an app such as selfBACK in clinical practice was described by both patients and health care practitioners as valuable to compensate for current organizational constraints (eg, long waiting time for the first consultation at the clinic, short consultations, limited service for people living in remote areas, and as a supplement to the physiotherapy service). One physician added how being active in advance could be helpful during consultations since the patients would then find it easier to explain their difficulties. Another physician described how the app positively impacted patients' health while awaiting health care assessment.

> *There is a long waiting time to get an appointment with us. Some patients have already gotten much better when they meet at the first appointment with*

> *me because they have started with physical activity and exercises on their own through the app […]. I think it is a great way to get them started.* [Female, focus group 2]

Others underlined how the app could also help maintain continuity for patients between consultations or cut down on the number of consultations, and some described the app as a way to support patients benefitting from less-intensive treatment. Therefore, the app was seen as a potential aid in allocating health care resources more appropriately to those needing it the most.

> *We have talked a lot about who is the right patient for us. And then we concluded that the so-called simple patients, those who can get help, for example, from an app by taking some steps in their lives that enable them to function, might not be the right patients for us. […] So, if we have a tool that we feel is good and can help some of these patients, it is fantastic. Then we handle the more complex cases, where an app is not sufficient.* [Female, focus group 1]

However, patients and health care practitioners emphasized that digital tools such as the selfBACK app should be regarded only as a supplement to usual care and not as a replacement. One patient explained how she believed that severe conditions should be ruled out first, which 1 physician also underlined when describing how many patients are not reassured by solely receiving information, for instance, when experiencing radiating pain. Another physician also described how normalizing pain in some, yet very few cases, can be inappropriate, and that the app should be combined with a health care consultation in such cases.

Several patients also underlined how their trust in and enthusiasm for technological tools did not imply that it could substitute the human contact offered by consultations with health care practitioners due to the value such interpersonal relationships represent. Similarly, the health care practitioners commented that they did not feel that the selfBACK app would interfere with their professional autonomy, seeing their role as essential.

> *Even if the app, based on how the patients respond to questionnaires, is customised and makes individual adaptations, it will never be able to do what a physician might do, see the patient in a larger perspective.* [Female, focus group 2]

Theme 3: Suggestions for Future Use

Although patients and health care practitioners felt positive about the possibility of adopting tools such as the selfBACK app and making progress on their own, they addressed several aspects they believed would determine acceptance. Many suggestions for change aligned with the difficulties described in the overall impression of the app regarding usability and content.

Some health care practitioners mentioned that the start-up process should be made more efficient if patients were to adopt the selfback app. Although patients and health care practitioners highlighted the possibility of replacing exercises as a valuable feature, some patients wished for an opportunity to provide feedback or point out the issues they experienced. Some patients also felt that they would have benefitted from more instructions within the app, for example, describing the frequency and purpose of the exercises, a point also reflected on by some physiotherapists.

Physiotherapists also said that the extensive focus on exercises was less beneficial than instructing patients to find an activity they liked to start being active. In addition, they pointed out that some exercise descriptions potentially undermined the main message of the app that activity is not harmful by communicating the opposite impression.

> *It [the app] uses words like "careful, controlled movement". Then you are communicating that you can potentially destroy something.* [Female, focus group 1]

Some physiotherapists and social workers also suggested that information on how other aspects, such as anxiety and depression, contribute to the feeling of pain should be highlighted within the app. In addition, 1 physiotherapist found the goal setting in the app so important to patients' rehabilitation processes that they suggested that it should be made mandatory to fill it in to proceed further.

Health care practitioners also reflected on aspects facilitating implementation in clinical practice. One stated how it would be beneficial to refer patients to something specific, such as the app, instead of a general call for "being active." Physiotherapists and social workers commented that having access to patients' interaction with the app would enable them to integrate it into their clinical routine. This point was also reflected by a statement from a patient.

> *It's nice if you have such an app, which you can choose to get and use yourself. Then you might get a follow-up with a doctor by phone or something like that asking: "What is the status now?". And the doctor might also be able to see the updates in the app and what you have posted. It is a tool for both the doctor and the patient if both have access to the results of such an app.* [Female, 59 years, low user]

Discussion

Principal Findings

This study explored the engagement and experiences with an app-based self-management support system (selfback) for patients with low back and neck pain referred to specialist care, and health care practitioners' views on adopting such digital tools in their clinical practice. Overall, patients' experiences and health care practitioners' perception of the app largely overlapped. Both had a positive attitude toward adopting app-based self-management support in this setting and saw a large potential in the selfback app to supplement usual care. Both described how the app can reassure patients by providing trustworthy information, thereby empowering them to take action on their own. Usability properties, content relevance, and the role of health care professionals were identified as important elements influencing acceptance and further engagement with the app.

While patients and clinicians were positive about the adoption of app-based self-management support for low back and neck pain as a supplement to usual care, the uptake of the intervention across patients enrolled in the RCT was relatively low. This somewhat differed from the SELFBACK trial in primary care, where nearly two-thirds of patients sustained use throughout 12 weeks [22,23]. Such differences might be partly explained by the onboarding procedures used, as well as by the study setting, that is, patients waiting for further consultation after referral to specialist care might be less prone or motivated to explore self-management interventions. Furthermore, low or nonusers in our study reported greater pain frequency and more widespread pain than moderate or high users. More burdensome health conditions, for example, having comorbidities or high symptom severity, have been suggested to be a barrier for engaging with self-management interventions [24,25]. Thus, understanding how the heterogeneity in clinical features might affect the uptake and engagement of self-management interventions should be explored further, particularly since greater symptom burden does not seem to modify the effect of such interventions [26-28].

Successful implementation of digital interventions relies, at least in part, on patients' acceptance of the intervention. Patients indicated that factors promoting the app's adoption included that it was easy to use, convenient, and provided structured and tailored information (eg, weekly plans). Furthermore, some patients described that knowing that the app was coming from a trustworthy source (ie, health care system or university) facilitated acceptance. On the contrary, technical difficulties, perceiving the content as irrelevant or not new, and lack of endorsement from the health care practitioner hindered some users from adopting the app. These elements are in line with existing acceptance models positing that perceived ease of use, perceived usefulness, and trusting beliefs in health care providers (or vendors) are, among other factors [29], significant predictors of behavioral intention in digital interventions [30,31].

Ensuring adequate engagement is a prerequisite for the effectiveness of app-based interventions [32]. Some patients described how notifications, activity tracking, and rewards helped them stay engaged with the app's content, emphasizing the importance of interactive, tailored support for sustaining self-management behaviors [11]. Such reminders seemed particularly relevant in the first phase of use, as some patients reported not having the need to log their activities or getting the recognition of achievements once the symptoms subsided. This use pattern has been described in previous digital interventions for low back pain [22,25]. The fact that some patients perceived the app as a supporter and tailored to one's individual needs suggests a form of therapeutic alliance with the digital interaction [33], which is an important enabler of self-management [24] and has been linked with increased engagement [34]. Conversely, perceiving the app as too general and irrelevant to one's health condition, as reported by other patients, might prevent the establishment of such a bond [33]. Technological and human-like design features, for example, AI chatbot, avatars, social forums, and peer support, can potentially foster digital therapeutic alliance further [35,36], which could be interesting to explore in future developments.

While the selfBACK app was designed to be self-explanatory, some patients indicated the need for more instructions. This was partly reflected by the fact that the exploration of the app was mostly limited to the main components of exercises and physical activity (ie, step count). Other components in the toolbox (static component within the app) containing additional self-management resources (eg, goal setting, pacing, relaxation techniques, and mindfulness) were less explored, as reflected by the usage data and the interview data. Although suggestions to access these resources were somewhat integrated into the educational content, they were not very prominent in the design of the weekly plan algorithm compared with the exercises. This might have limited the exposure and practice of relevant self-management skills linked to the promotion of self-efficacy, in turn influencing long-term behavioral change [24]. A few patients also mentioned the necessity of customizing some elements within the app (beyond changing exercises) and the ability to provide feedback. This need for greater self-tailoring of the content aligns with the concept of autonomy support, whereby taking individual preferences into account and enabling patients' perceived active control foster autonomous motivation, which is important for the maintenance of behavior change [37].

Offering a self-management app for patients on a waiting list for specialist care can be an easy and inexpensive approach to initiate cognitive and behavioral processes by providing evidence-based and tailored content. Some patients described being reassured by the educational content and exercises and developing greater awareness and confidence about the possibility of self-managing while using the app. As such, the app can increase patients' feeling of empowerment, which is important to achieve competence to manage pain and enable lifestyle changes [38]. The clinical value of embedding a self-help intervention in this phase was further highlighted by health care practitioners who stated that priming patients with such content would enhance patient-clinician communication, thus facilitating shared decision-making during the clinical encounter. However, both patients and health care practitioners often mentioned the need for clinical involvement to enable engagement with self-management advice, mostly due to diagnostic uncertainty in this patient group. Previous research has shown that health care professional support, even when remote or minimal, can increase the effectiveness of self-management interventions [39]. Thus, combining digital and human support could be a useful approach to enhance adoption of self-management, particularly in the specialist health care setting with long waiting time.

Both patients and health care practitioners widely emphasized the necessity for taking responsibility for one's own health conditions, indicating that digital interventions such as the SELFBACK app hold a large potential in mitigating current health care shortage challenges. However, while digital interventions are useful and wanted by many patients, our findings suggest that not all patients can benefit from such interventions. Since the patient group in this study is highly heterogeneous and pain management styles and preferences vary, further research should look into how to further optimize tailoring of self-management support to increase patients' feeling of relevance and usefulness over time. In addition, patients' needs, abilities, and preferences for autonomy should be considered when implementing digital interventions within the health care setting [40]. This should come with the awareness that assuming patients' responsibility for self-management could lead to stigmatization of some patients, potentially those with higher needs for health care services [41].

Strengths and Limitations

While our findings might not be generalizable to other contexts (eg, different health care systems) or patient groups with other chronic health challenges, they nonetheless provide useful insights into patients' and health care practitioners' experiences with digital self-management interventions. Since back pain complaints are one of the main causes of years lived with disability worldwide [42] and practitioners' acceptance of app-based interventions has been recognized as a global tendency [43], the need and value of digital self-management support transcend the regional setting of this study.

A strength of this study was that researchers from different backgrounds, that is, physiotherapy, medicine, anthropology, and exercise physiology read the interviews, and results were discussed thoroughly among them. Another strength was the inclusion of patients with different levels of app usage (ie, the number of plans generated), ensuring a balanced view of patients' experiences with the app, including those who might have been less satisfied with it. In addition, data on how much the users accessed different content in the app were available for all patients allocated to the SELFBACK app. Integrating the views of health care practitioners with patients' experiences allowed a better understanding of acceptability and needs from both sides, which are important for future implementation. However, some limitations need to be considered. Although health care practitioners were invited to get acquainted with the app for some weeks prior to the focus groups in addition to receiving an overview of its functionality, only a few tried the app and were familiar with the entire content. A greater firsthand experience could have increased the specificity of their views regarding adopting selfBACK in this context. Furthermore, patients were most likely interviewed when they already received first consultation and initiated treatment in specialist care (ie, 3-4 months after inclusion), and this might have affected their views on self-management and the use of digital interventions. Finally, while we interviewed patients with different app usage levels, we did not interview patients who had never accessed it. This could have provided better insight into factors related to the onboarding procedure and uptake of digital interventions in this setting.

Conclusions

Both patients and health care practitioners supported the adoption of app-based self-management support for low back and neck pain in specialist care. The selfback app was reported by some patients and health care practitioners to provide reassurance and empowering patients to take actions for their health problem on their own. Acceptance and engagement with the app-based intervention can be influenced by various factors, such as content structure and relevance, tailoring, trust, and usability properties. Digital self-help combined with human support might be necessary to enhance adoption of self-management, particularly in specialist health care settings.

Acknowledgments

This study was supported by the European Union's Horizon 2020 research and innovation program (grants 777090, Back-UP; and 689043, SELFBACK) and innovation funding allocated to St Olavs Hospital by the Central Norway Regional Health Authorities (grant 2019/34-629/2020; the e-Help project).

Conflicts of Interest

The authors declare that no conflicts of interest exist. The overall aim of this project was to test a digital decision support system and a smartphone app to support patients to self-manage their low back and neck pain. The results and experiences from this project will inform further developments of the app, which may be introduced into a commercial market.

References

1. GBD 2015 Disease and Injury Incidence and Prevalence Collaborators . Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016;388(10053):1545-1602 [FREE Full text] [doi: 10.1016/S0140-6736(16)31678-6] [Medline: 27733282]
2. Chou R, Shekelle P. Will this patient develop persistent disabling low back pain? JAMA 2010;303(13):1295-1302. [doi: 10.1001/jama.2010.344] [Medline: 20371789]
3. Carroll LJ, Hogg-Johnson S, van der Velde G, Haldeman S, Holm LW, Carragee EJ, et al. Course and prognostic factors for neck pain in the general population: results of the bone and joint decade 2000-2010 task force on neck pain and its associated disorders. J Manipulative Physiol Ther 2009;32(2 Suppl):S87-S96. [doi: 10.1016/j.jmpt.2008.11.013] [Medline: 19251079]
4. Breivik H, Eisenberg E, O'Brien T. The individual and societal burden of chronic pain in Europe: the case for strategic prioritisation and action to improve knowledge and availability of appropriate care. BMC Public Health 2013;13:1229 [FREE Full text] [doi: 10.1186/1471-2458-13-1229] [Medline: 24365383]

5. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48(2):177-187. [doi: 10.1016/s0738-3991(02)00032-0] [Medline: 12401421]
6. Lin I, Wiles L, Waller R, Goucke R, Nagree Y, Gibberd M, et al. What does best practice care for musculoskeletal pain look like? Eleven consistent recommendations from high-quality clinical practice guidelines: systematic review. Br J Sports Med 2020;54(2):79-86. [doi: 10.1136/bjsports-2018-099878] [Medline: 30826805]
7. Hutting N, Oswald W, Staal JB, Heerkens YF. Self-management support for people with non-specific low back pain: a qualitative survey among physiotherapists and exercise therapists. Musculoskelet Sci Pract 2020;50:102269 [FREE Full text] [doi: 10.1016/j.msksp.2020.102269] [Medline: 33039797]
8. Kamper SJ, Logan G, Copsey B, Thompson J, Machado GC, Abdel-Shaheed C, et al. What is usual care for low back pain? A systematic review of health care provided to patients with low back pain in family practice and emergency departments. Pain 2020;161(4):694-702. [doi: 10.1097/j.pain.0000000000001751] [Medline: 31738226]
9. Marcuzzi A, Nordstoga AL, Bach K, Aasdahl L, Nilsen TIL, Bardal EM, et al. Effect of an artificial intelligence-based self-management app on musculoskeletal health in patients with neck and/or low back pain referred to specialist care: a randomized clinical trial. JAMA Netw Open 2023;6(6):e2320400 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.20400] [Medline: 37368401]
10. Leake D. Problem solving and reasoning: case-based. In: Smelser NJ, Baltes PB, editors. International Encyclopedia of the Social & Behavioral Sciences. Oxford, United Kingdom: Pergamon; 2001:12117-12120.
11. Lustria MLA, Noar SM, Cortese J, Van Stee SK, Glueckauf RL, Lee J. A meta-analysis of web-delivered tailored health behavior change interventions. J Health Commun 2013;18(9):1039-1069. [doi: 10.1080/10810730.2013.768727] [Medline: 23750972]
12. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
13. Svendsen MJ, Sandal LF, Kjær P, Nicholl BI, Cooper K, Mair F, et al. Using intervention mapping to develop a decision support system-based smartphone app (selfBACK) to support self-management of nonspecific low back pain: development and usability study. J Med Internet Res 2022;24(1):e26555 [FREE Full text] [doi: 10.2196/26555] [Medline: 35072645]
14. Nordstoga AL, Bach K, Sani S, Wiratunga N, Mork PJ, Villumsen M, et al. Usability and acceptability of an app (SELFBACK) to support self-management of low back pain: mixed methods study. JMIR Rehabil Assist Technol 2020;7(2):e18729 [FREE Full text] [doi: 10.2196/18729] [Medline: 32902393]
15. Sandal LF, Øverås CK, Nordstoga AL, Wood K, Bach K, Hartvigsen J, et al. A digital decision support system (selfBACK) for improved self-management of low back pain: a pilot study with 6-week follow-up. Pilot Feasibility Stud 2020;6:72 [FREE Full text] [doi: 10.1186/s40814-020-00604-2] [Medline: 32489674]
16. Mork PJ, Bach K, selfBACK Consortium. A decision support system to enhance self-management of low back pain: protocol for the selfBACK project. JMIR Res Protoc 2018;7(7):e167 [FREE Full text] [doi: 10.2196/resprot.9379] [Medline: 30030208]
17. Bach K, Szczepanski T, Aamodt A, Gundersen O, Mork P. Case representation and similarity assessment in the self BACK decision support system. presented at; 2016 Presented at: International Conference on Case-Based Reasoning; 2016 Sep 19; Norway. [doi: 10.1007/978-3-319-47096-2_3]
18. Matomo. URL: https://matomo.org/ [accessed 2023-11-17]
19. May CR, Cummings A, Girling M, Bracher M, Mair FS, May CM, et al. Using normalization process theory in feasibility studies and process evaluations of complex healthcare interventions: a systematic review. Implement Sci 2018;13(1):80 [FREE Full text] [doi: 10.1186/s13012-018-0758-1] [Medline: 29879986]
20. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
21. Marcuzzi A, Bach K, Nordstoga AL, Bertheussen GF, Ashikhmin I, Boldermo NO, et al. Individually tailored self-management app-based intervention (selfBACK) versus a self-management web-based intervention (e-Help) or usual care in people with low back and neck pain referred to secondary care: protocol for a multiarm randomised clinical trial. BMJ Open 2021;11(9):e047921 [FREE Full text] [doi: 10.1136/bmjopen-2020-047921] [Medline: 34518253]
22. Svendsen MJ, Nicholl BI, Mair FS, Wood K, Rasmussen CDN, Stochkendahl MJ. One size does not fit all: participants' experiences of the selfBACK app to support self-management of low back pain-a qualitative interview study. Chiropr Man Therap 2022;30(1):41 [FREE Full text] [doi: 10.1186/s12998-022-00452-2] [Medline: 36192724]
23. Sandal LF, Bach K, Øverås CK, Svendsen MJ, Dalager T, Stejnicher Drongstrup Jensen J, et al. Effectiveness of app-delivered, tailored self-management support for adults with lower back pain-related disability: a selfBACK randomized clinical trial. JAMA Intern Med 2021;181(10):1288-1296 [FREE Full text] [doi: 10.1001/jamainternmed.2021.4097] [Medline: 34338710]
24. Devan H, Hale L, Hempel D, Saipe B, Perry MA. What works and does not work in a self-management intervention for people with chronic pain? Qualitative systematic review and meta-synthesis. Phys Ther 2018;98(5):381-397. [doi: 10.1093/ptj/pzy029] [Medline: 29669089]

25. Geraghty A, Roberts L, Stanford R, Hill JC, Yoganantham D, Little P, et al. Exploring patients' experiences of internet-based self-management support for low back pain in primary care. Pain Med 2020;21(9):1806-1817 [FREE Full text] [doi: 10.1093/pm/pnz312] [Medline: 31841156]
26. Nordstoga AL, Aasdahl L, Sandal LF, Dalager T, Kongsvold A, Mork PJ, et al. The role of pain duration and pain intensity on the effectiveness of app-delivered self-management for low back pain (selfBACK): secondary analysis of a randomized controlled trial. JMIR Mhealth Uhealth 2023;11:e40422 [FREE Full text] [doi: 10.2196/40422] [Medline: 37656023]
27. Øverås CK, Nilsen TIL, Nicholl BI, Rughani G, Wood K, Søgaard K, et al. Multimorbidity and co-occurring musculoskeletal pain do not modify the effect of the SELFBACK app on low back pain-related disability. BMC Med 2022;20(1):53 [FREE Full text] [doi: 10.1186/s12916-022-02237-z] [Medline: 35130898]
28. Rughani G, Nilsen TIL, Wood K, Mair FS, Hartvigsen J, Mork PJ, et al. The selfBACK artificial intelligence-based smartphone app can improve low back pain outcome even in patients with high levels of depression or stress. Eur J Pain 2023;27(5):568-579 [FREE Full text] [doi: 10.1002/ejp.2080] [Medline: 36680381]
29. Or CKL, Karsh BT. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009;16(4):550-560 [FREE Full text] [doi: 10.1197/jamia.M2888] [Medline: 19390112]
30. Klein R. Internet-based patient-physician electronic communication applications: patient acceptance and trust. e-Service J 2007;5(2):27-52. [doi: 10.2979/esj.2007.5.2.27]
31. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989;13(3):319-339. [doi: 10.2307/249008]
32. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
33. Berry K, Salter A, Morris R, James S, Bucci S. Assessing therapeutic alliance in the context of mHealth interventions for mental health problems: development of the Mobile Agnew Relationship Measure (mARM) questionnaire. J Med Internet Res 2018;20(4):e90 [FREE Full text] [doi: 10.2196/jmir.8252] [Medline: 29674307]
34. Clarke J, Proudfoot J, Whitton A, Birch M, Boyd M, Parker G, et al. Therapeutic alliance with a fully automated mobile phone and web-based intervention: secondary analysis of a randomized controlled trial. JMIR Ment Health 2016;3(1):e10 [FREE Full text] [doi: 10.2196/mental.4656] [Medline: 26917096]
35. Müssener U. Digital encounters: human interactions in mHealth behavior change interventions. Digit Health 2021;7:20552076211029776 [FREE Full text] [doi: 10.1177/20552076211029776] [Medline: 34262783]
36. Tremain H, McEnery C, Fletcher K, Murray G. The therapeutic alliance in digital mental health interventions for serious mental illnesses: narrative review. JMIR Ment Health 2020;7(8):e17204 [FREE Full text] [doi: 10.2196/17204] [Medline: 32763881]
37. Bol N, Høie NM, Nguyen MH, Smit ES. Customization in mobile health apps: explaining effects on physical activity intentions by the need for autonomy. Digit Health 2019;5:2055207619888074 [FREE Full text] [doi: 10.1177/2055207619888074] [Medline: 31807312]
38. Brew-Sam N, Chib A. Theoretical advances in mobile health communication research: an empowerment approach to self-management. In: Technology and Health: Promoting Attitude and Behavior Change. Cambridge, MA: Elsevier Academic Press; 2020:151-177.
39. Cuijpers P, Noma H, Karyotaki E, Cipriani A, Furukawa TA. Effectiveness and acceptability of cognitive behavior therapy delivery formats in adults with depression: a network meta-analysis. JAMA Psychiatry 2019;76(7):700-707 [FREE Full text] [doi: 10.1001/jamapsychiatry.2019.0268] [Medline: 30994877]
40. Morley J, Floridi L. The limits of empowerment: how to reframe the role of mHealth tools in the healthcare ecosystem. Sci Eng Ethics 2020;26(3):1159-1183 [FREE Full text] [doi: 10.1007/s11948-019-00115-1] [Medline: 31172424]
41. Lawn S, McMillan J, Pulvirenti M. Chronic condition self-management: expectations of responsibility. Patient Educ Couns 2011;84(2):e5-e8. [doi: 10.1016/j.pec.2010.07.008] [Medline: 20705412]
42. GBD 2021 Low Back Pain Collaborators. Global, regional, and national burden of low back pain, 1990-2020, its attributable risk factors, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021. Lancet Rheumatol 2023;5(6):e316-e329 [FREE Full text] [doi: 10.1016/S2665-9913(23)00098-X] [Medline: 37273833]
43. Newsome A, Reed R, Sansone J, Batrakoulis A, McAvoy C, W. Parrott M. 2024 ACSM worldwide fitness trends: future directions of the health and fitness industry. ACSM's Health Fitness J 2024;28(1):14-26. [doi: 10.1249/fit.0000000000000933]

Abbreviations

AI: artificial intelligence
COREQ: consolidated criteria for reporting qualitative research
RCT: randomized clinical trial

Edited by A Kushniruk, E Borycki; submitted 22.12.23; peer-reviewed by A Batrakoulis, E Ray Chaudhuri, JD Robinson; comments to author 17.02.24; revised version received 25.04.24; accepted 26.04.24; published 09.07.24.

Please cite as:
Marcuzzi A, Klevanger NE, Aasdahl L, Gismervik S, Bach K, Mork PJ, Nordstoga AL
An Artificial Intelligence–Based App for Self-Management of Low Back and Neck Pain in Specialist Care: Process Evaluation From a Randomized Clinical Trial
JMIR Hum Factors 2024;11:e55716
URL: https://humanfactors.jmir.org/2024/1/e55716
doi:10.2196/55716
PMID:38980710

Original Paper

Making Co-Design More Responsible: Case Study on the Development of an AI-Based Decision Support System in Dementia Care

Dirk R M Lukkien[1,2], MSc; Sima Ipakchian Askari[1,3], PhD; Nathalie E Stolwijk[1], MSc; Bob M Hofstede[1,3], MSc; Henk Herman Nap[1,3], PhD; Wouter P C Boon[2], Prof Dr; Alexander Peine[4], Prof Dr; Ellen H M Moors[2], Prof Dr; Mirella M N Minkman[1,5], Prof Dr

[1]Vilans Centre of Expertise of Long Term Care, Utrecht, Netherlands
[2]Copernicus Institute of Sustainable Development, Utrecht University, Utrecht, Netherlands
[3]Human Technology Interaction, Eindhoven University of Technology, Eindhoven, Netherlands
[4]Faculty of Humanities, Open University of The Netherlands, Heerlen, Netherlands
[5]Tilburg Institute for Advanced Studies School for Business and Society, Tilburg University, Tilburg, Netherlands

Corresponding Author:
Dirk R M Lukkien, MSc
Vilans Centre of Expertise of Long Term Care
Churchilllaan 11
Utrecht, 3505 RE
Netherlands
Phone: 31 612416513
Email: d.lukkien@vilans.nl

Abstract

Background: Emerging technologies such as artificial intelligence (AI) require an early-stage assessment of potential societal and ethical implications to increase their acceptability, desirability, and sustainability. This paper explores and compares 2 of these assessment approaches: the responsible innovation (RI) framework originating from technology studies and the co-design approach originating from design studies. While the RI framework has been introduced to guide early-stage technology assessment through anticipation, inclusion, reflexivity, and responsiveness, co-design is a commonly accepted approach in the development of technologies to support the care for older adults with frailty. However, there is limited understanding about how co-design contributes to the anticipation of implications.

Objective: This paper empirically explores how the co-design process of an AI-based decision support system (DSS) for dementia caregivers is complemented by explicit anticipation of implications.

Methods: This case study investigated an international collaborative project that focused on the co-design, development, testing, and commercialization of a DSS that is intended to provide actionable information to formal caregivers of people with dementia. In parallel to the co-design process, an RI exploration took place, which involved examining project members' viewpoints on both positive and negative implications of using the DSS, along with strategies to address these implications. Results from the co-design process and RI exploration were analyzed and compared. In addition, retrospective interviews were held with project members to reflect on the co-design process and RI exploration.

Results: Our results indicate that, when involved in exploring requirements for the DSS, co-design participants naturally raised various implications and conditions for responsible design and deployment: protecting privacy, preventing cognitive overload, providing transparency, empowering caregivers to be in control, safeguarding accuracy, and training users. However, when comparing the co-design results with insights from the RI exploration, we found limitations to the co-design results, for instance, regarding the specification, interrelatedness, and context dependency of implications and strategies to address implications.

Conclusions: This case study shows that a co-design process that focuses on opportunities for innovation rather than balancing attention for both positive and negative implications may result in knowledge gaps related to social and ethical implications and how they can be addressed. In the pursuit of responsible outcomes, co-design facilitators could broaden their scope and reconsider the specific implementation of the process-oriented RI principles of *anticipation* and *inclusion.*

(JMIR Hum Factors 2024;11:e55961) doi:10.2196/55961

KEYWORDS
responsible innovation; co-design; ethics; decision support systems; gerontechnology; dementia; long-term care

Introduction

Background

In the long-term care for older adults with frailty, caregivers and clients are increasingly being assisted by artificial intelligence (AI)–based technologies [1-5]. AI-based technologies can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or web-based environments, thereby using machine or human-based data and input [6]. For instance, AI is being used in decision support systems (DSSs) that acquire relevant data about care needs or processes; present the relevant data to users (eg, caregivers); and translate raw data into actionable information, such as alerts, risk assessments, or recommendations about care strategies [7-10]. Notwithstanding the opportunities and advantages, it is broadly acknowledged that the use of AI-based technologies entails societal and ethical implications. The long-term data collection in the context of monitoring older people's health and well-being and the mediating or even leading role of algorithms in interpreting these data to arrive at care-related decisions pose implications related to, among others, undermining people's privacy, autonomy, and self-determination; the discrimination and stigmatization of old age; and surveillance capitalism [1,11-15].

Due to the impact technologies such as DSSs have on people's lives and the potential resistance that might emerge during implementation, an early-stage assessment of their implications is called for. This paper explores and compares 2 of these assessment approaches: the responsible innovation (RI) framework originating from technology studies and the co-design approach originating from design studies. The term RI refers to the aim to ensure the ethical acceptability, societal desirability, and sustainability of innovation processes and outcomes [16,17]. To guide RI into practice, Owen et al [17] suggest that four process-oriented principles should guide technology research and development: (1) anticipation of the potential positive and negative implications; (2) inclusion of users and other stakeholders; (3) reflexivity of actors upon their own practices, assumptions, values, and interests; and (4) responsiveness to insights that emerge during the innovation process.

Co-design can be used as an umbrella term for approaches that actively involve users and other stakeholders of innovations in any stage of the design process to ensure that the outcomes meet their needs [18,19]. It is a commonly accepted approach in the development of technologies to support the long-term care for older adults [20-22]. On a conceptual level, co-design resonates with RI. Both approaches share a focus on developing technologies to match human needs and abilities, similar to research fields such as human factors, human-computer interaction, and cognitive engineering. In fact, co-design has increasingly received attention as a way to support RI [23]. Similar to RI, the co-design approach describes a research and development process in which innovators *inclusively deliberate* and *reflect* on the needs and values of different stakeholders and *iteratively* design and *adapt* innovations based on these insights [23]. However, in contrast to RI, co-design does not explicitly impose on innovators the need to *anticipate* potential societal and ethical implications (henceforth, abbreviated as "implications"). Co-design can yield insights into potential unintended side effects and value creation that stakeholders do not want from innovation, but this is generally not an explicit aim in co-design. Against this background, this paper empirically explores how the explicit *anticipation* of implications can complement co-design.

More specifically, this paper presents a case study on an international collaborative project that focuses on the development of a DSS to support formal caregivers involved in long-term dementia care. A co-design process involving intended users and other stakeholders (henceforth, abbreviated as "users") is central to the development of the DSS. In addition, a separate line of research of the project under investigation explicitly anticipated implications of using DSSs in dementia care, along with strategies to address these implications, thereby fostering RI in AI-assisted decision-making. This so-called RI exploration largely took place in parallel to (ie, not as part of) the co-design activities and focused on soliciting the perspectives of project members (PMs) rather than those of users. This paper describes the empirical exploration of how the co-design process of an AI-based DSS for dementia caregivers is complemented by the explicit anticipation of implications.

The Healthy Ageing Eco-System for People With Dementia Project

The case presented in this paper is the Healthy Ageing Eco-system for People With Dementia (HAAL) project, which is part of the European Active and Assisted Living (AAL) program (AAL Europe, 2021; project AAL-2020-7-229-CP). In HAAL, an international consortium comprising care organizations, research institutes, and commercial firms from the Netherlands, Italy, Taiwan, and Denmark collaborates on the co-design, development, testing, and commercialization of a DSS that is intended to provide actionable information to formal caregivers of people with dementia, with the aim of reducing their workload and increasing the quality of care [24]. The DSS developed in HAAL concerns a dashboard that integrates various types of data about the physical activity, eating and sleeping patterns, cognitive functioning, mood, social contact, and medication intake of people with dementia. These data can be collected via several digital technologies (henceforth, "HAAL technologies") throughout various stages of dementia. Besides integrating the data from HAAL technologies into 1 dashboard, possibilities to provide caregivers only the most relevant data in the form of summary overviews, alerts, predictions about emergency situations, and recommendations about care strategies were explored. To this end, both preprogrammed, rule-based algorithms and data-driven algorithms rooted in machine learning are used to process data.

With these predefined directions as a starting point, a series of iterative co-design activities involving dementia caregivers, or more correctly “proxy users” who represent these eventual users (see the study by Stewart and Hyysalo [25]), and other stakeholders were organized to feed the actual design and development of the dashboard. The co-design activities focused on exploring the relevance and possibilities of translating the data from HAAL technologies into useful information and prioritizing data that are relevant to be presented in the dashboard [24,26]. In addition, the co-design activities focused on determining functionalities of the dashboard and designing and evaluating different pages of the dashboard’s user interface.

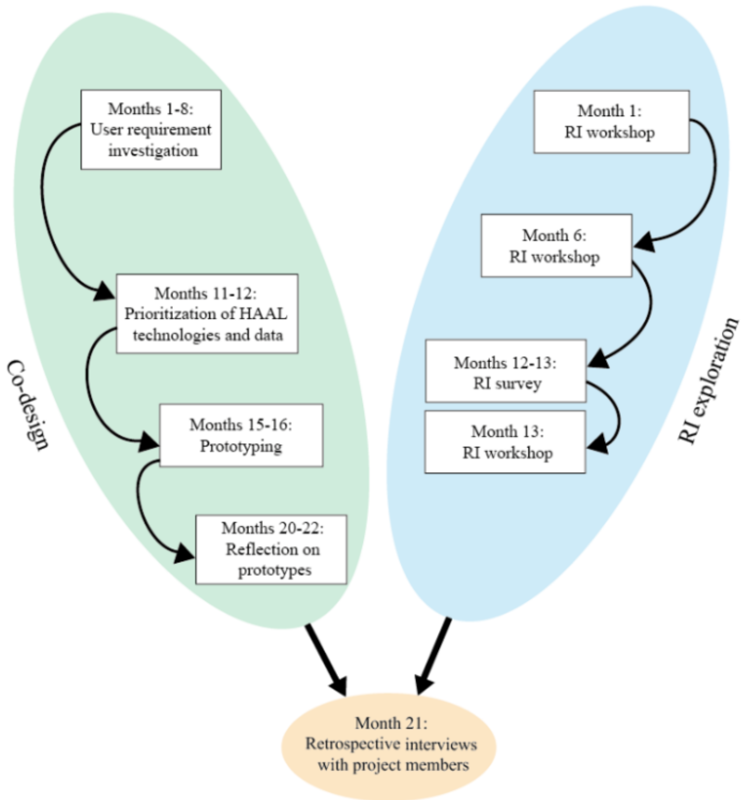
The RI exploration in HAAL, which took place largely in parallel to the co-design activities, initially focused on raising PMs’ general awareness about RI and exploring their perspectives on both positive and negative implications of using the HAAL dashboard, along with strategies to address these implications.

Methods

Overview

For this case study, results from the co-design process and RI exploration within the HAAL project were incorporated and analyzed. In addition, retrospective interviews were held with individual PMs to reflect on the co-design process and RI exploration. Because the co-design process and RI exploration were largely organized in parallel, the HAAL project provided sufficient data within a specific time and context to perform a retrospective analysis on how the explicit anticipation of implications can complement co-design. Figure 1 shows a timeline of activities.

Figure 1. Timeline of the co-design process, responsible innovation (RI) exploration, and retrospective interviews. "Month" refers to the month (count in project) in which the activity took place. HAAL: Healthy Ageing Eco-system for People With Dementia.



Co-Design Process

Table 1 describes the 4 specific steps taken in the co-design process. The co-design activities in HAAL were conducted in 4 countries: the Netherlands, Italy, Taiwan and Denmark. The organizations from Denmark are unsubsidized partners in the HAAL project and did not participate in co-design steps 3 to 4. Despite differences in dementia care systems across these countries, such as types of caregivers involved in home-based and institutionalized care settings, formal caregivers of people with dementia were perceived as the primary target group for (using) the dashboard in all countries. Hence, a variety of formal caregivers of people with dementia, such as (homecare) nurses, case managers, psychologists, psychotherapists, social workers, and specialists in the care of older adults, were involved in the co-design activities. In addition, other stakeholders, such as innovation staff, data analysts at care organizations, and people working in (care) alarm centrals, were involved in some steps of the co-design process to broadly explore requirements for the dashboard. As indicated in Table 1, two intermediate steps were taken without the direct involvement of users. Further, at the end of step 4, participants were implicitly asked about RI-related themes (autonomy and transparency). Throughout the co-design activities, data were collected in the form of notes, audio and video recordings, photos, drawings, and (web-based) canvasses and by conducting surveys.

Table 1. Steps taken in the co-design process.

Step	Methods	Research focus	Participants
1[a]	Focus group sessions (3 web based, 1 hybrid, and 16 physical)	• User requirement investigation: insights were gathered into different stakeholders' attitudes toward the HAAL[b] technologies, their ideas about the added value and functionalities that are envisioned in the integration of these technologies in 1 dashboard, and for which stakeholders such a dashboard may be most relevant.	Nurses, day-care workers, psychologists, physiotherapists, technical stakeholders, innovation managers and directors of care organizations, representatives from various municipalities, people with dementia, and informal caregivers (n=146; the Netherlands: n=18, 12.3%; Italy: n=18, 12.3%; Taiwan: n=108, 74%; Denmark: n=2, 1.4%).
2	Demonstration, try-outs, and survey (7 physical and 1 hybrid)	• Prioritization of HAAL technologies and data: based on the MoSCoW[c] technique, all HAAL technologies and corresponding data were categorized into 4 ascending categories (must have, should have, could have, and won't have this time), indicating what best fits the needs of people with dementia and their caregivers [27,28]. After a demonstration and try-outs of the HAAL technologies, participants completed a digital prioritization survey.	Nurses, day-care workers, psychologists, physiotherapists, data specialists, and innovation staff and directors from care organizations (n=48; the Netherlands: n=6, 12%; Italy: n=9, 19%; Taiwan: n=30, 62%; Denmark: n=3, 6%).
3[d]	Co-design sessions (3 physical and 2 web based)	• Prototyping: three dashboard pages were preselected to be co-designed with participants: (1) client profile with detailed information on specific clients, (2) overall list of clients, and (3) an overview of urgent situations. This resulted in insights (ie, through sketches and design by participants) into the kind of information to be displayed in the dashboard and how the information could be visualized. Finally, the results were compared with those of the preliminary mock-up.	Data specialists and innovation staff, including part-time nurses (n=21; the Netherlands: n=6, 29%; Italy: n=4, 19%; Taiwan: n=11, 52%).
4	Usability study (8 physical sessions, including survey)	• Reflection on prototypes: insights were gained into the usability and heuristics of the clickable mock-up. More specifically, after first performing 6 tasks in the mock-up, participants completed a survey. In the survey, the HUBBI[e] questionnaire [29] was used to determine usability, and heuristics were evaluated using the issue categories of Bastien and Scapin [30] and Nielsen's severity ranking [31]. After completing the survey, participants engaged in a group discussion on the overall added value and functioning of the dashboard. At the end of the discussion, participants were also asked to reflect on 2 RI[f] themes: • Autonomy: do you think the degree to which the dashboard guides your decision-making as a caregiver is adequate, too low, or too high, and why do you think so? • Transparency: what would you choose, if you had to choose between either the accuracy of the information (ie, recommendations) provided by the dashboard or the understandability of the information, and why would you choose it?	Formal caregivers, digital care ambassadors, alarm centralists, and innovation staff (n=33; the Netherlands: n=9, 27%; Italy: n=14, 42%; Taiwan: n=10, 30%).

[a]Intermediate step: after analyzing results from step 1, user personas and desired dashboard functionalities were defined and translated into a preliminary mock-up for the dashboard (iteration 1). The motivational goal model of Taveter et al [32] was used for this translation.

[b]HAAL: Healthy Ageing Eco-system for People With Dementia.

[c]MoSCoW: must have, should have, could have, and won't have this time.

[d]Intermediate step: after analyzing results from step 3, insights about user requirements were again plotted on the motivational goal model to define design requirements. These design requirements were used to translate the preliminary mock-up into a clickable mock-up (iteration 2).

[e]HUBBI: eHealth usability benchmarking instrument.

[f]RI: responsible innovation.

RI Exploration

The RI exploration was primarily based on a qualitative survey among PMs, which was preceded by 2 workshops and followed by a third workshop with PMs. The first 2 workshops with PMs were held in a hybrid setting (web based and physical) during collective consortium meetings. The goal of the first workshop was to explain the notion of RI to PMs and discuss their thoughts about the relevance of and ways to address RI in HAAL. In the second workshop, based on the guidance ethics approach of Verbeek and Tijink [33], potential positive and negative implications of using the envisioned HAAL dashboard were explored, along with ways to address these implications.

Next, a dedicated qualitative RI survey was developed and conducted among PMs (Multimedia Appendix 1). The goal of the RI survey was to reveal PMs' viewpoints on how to responsibly develop AI-based analytical functionalities and the dashboard user interface in the HAAL project. The survey first explained that AI, as in the HAAL dashboard, provides opportunities for descriptive, diagnostic, predictive, and prescriptive analyses with differing levels of complexity and automation [34,35]. Next, questions were asked in relation to 2 distinct imaginary scenarios that outline different roles for AI within the HAAL dashboard. The first scenario (A) described a descriptive and largely rule-based dashboard through which users can assess the data from HAAL technologies and how the situations of clients have changed over time. This scenario was inspired by the dashboard that was aimed to be developed in the HAAL project. The second scenario (B) took a more speculative turn and described a proactive and partially self-learning dashboard that automatically translates the data into diagnostic, predictive, and prescriptive information to prompt caregivers to take certain actions. The scenarios were used as input to inspire respondents about directions the project could take in terms of developing AI and to enable them to articulate their expectations and considerations regarding the opportunities and implications of an advanced AI-based DSS (see also the study by Noortman et al [36]). After presenting each scenario, questions were asked about the positive and negative implications of using the respective dashboard. Thereafter, respondents were asked which scenario they preferred in terms of ethical acceptability, societal desirability, and technical feasibility and why they preferred it. Next, the survey introduced six principles for responsible AI innovation, adopted from guidelines from the World Health Organization: (1) protecting human autonomy; (2) promoting human well-being and safety and the public interest; (3) ensuring transparency, explainability, and intelligibility; (4) fostering responsibility and accountability; (5) ensuring inclusiveness and equity; and (6) promoting AI that is responsive and sustainable [37]. Respondents were asked how these principles might be relevant to and could be applied in the HAAL project. The survey was completed by 12 respondents representing 7 different organizations from all 4 countries. In addition, 5 respondents partially filled in the survey anonymously.

Finally, the RI survey was followed by a third hybrid workshop in which PMs were invited to jointly discuss what they learned from answering the RI survey.

Retrospective Interviews With PMs

In addition to the co-design activities and RI exploration, semistructured interviews were held with 6 PMs: 4 co-design facilitators (n=1, 25% working in the Netherlands; n=2, 50% working in Taiwan; and n=1, 25% working in Italy) and 2 software developers (working in Italy). The goal of the interviews was to uncover possible rationales behind the co-design process, choices made throughout the co-design process, and input given by co-design participants. All interviews lasted between 30 and 40 minutes and were fully transcribed by a professional transcription service.

Analysis

The analysis of data was performed by DRML, SIA, NES, and BMH. The data collected during the co-design activities and RI exploration were first analyzed independently by these 4 researchers. While the co-design data were previously analyzed by HAAL PMs to learn about the dashboard requirements, they were analyzed again for the purposes of this paper. Taking the 6 responsible AI principles from the World Health Organization guidelines [37] as a starting point, the researchers performed an inductive thematic analysis [38] to uncover conditions for the responsible design and deployment of the HAAL dashboard, including potential negative implications and strategies to address them. In doing so, they examined how certain insights regarding these conditions emerged in the co-design activities, the RI exploration, or both. In other words, the analysis focused, first, on identifying themes common within and between the co-design and RI exploration results and, second, on examining how the results from the RI exploration complement those from the co-design activities, or vice versa, in terms of RI. Subsequently, the transcripts of the retrospective interviews were analyzed independently by DRML, SIA, and BMH to uncover new conditions for RI and explore the complementarity between the co-design process and RI exploration. An additional focus was on why certain insights about conditions for RI may have emerged less explicitly in either the co-design process or the RI exploration. While analyzing the data, the researchers applied open coding and kept track of their reflections by writing them down as memos. After the data were independently analyzed by the researchers, the findings and memos were regularly discussed and reviewed by the researchers to reconcile major discrepancies in the coding and to reach agreement on the final coding scheme. Both physical and digital meetings were held to ensure the consistency of the analysis and reach convergence.

Ethical Considerations

The authors of this study followed the guidelines in the Declaration of Helsinki and the Dutch code of conduct for scientific integrity. Ethical approval for the interviews, not subject to the medical scientific research act involving human subjects, was granted by an independent board of the lead author's department (Vilans), including a privacy officer and legal expert [39].

For each co-design step, general information about the goal and procedure was provided, and the participants were asked to read and sign an informed consent form. The original consent covers secondary analysis of the data for the purposes of this study. The data gathered through the co-design steps and RI exploration were pseudonymized before analysis. Study participants did not receive any financial compensation.

Results

Overview

Seven overarching and interlinked themes representing conditions for the responsible development and deployment of the HAAL dashboard were extracted: (1) develop a proactive dashboard, (2) prevent cognitive overload, (3) protect privacy, (4) provide transparency, (5) empower caregivers to be in control, (6) safeguard accuracy, and (7) train users. We explicate how insights related to each theme emerged in the co-design activities, the RI exploration, or both. In addition, insights from the interviews with PMs are provided. In doing so, for each theme, we discuss how the explicit anticipation of implications (ie, the RI exploration) complements the co-design process in the HAAL project. Textbox 1 excerpts the results.

Textbox 1. Analysis of complementarities between the co-design process and responsible innovation (RI) exploration per theme.

1. Develop a proactive dashboard
   - The co-design results clearly indicate a perceived need for a proactive dashboard and provide concrete arguments to this end. The RI exploration also indicated the need for a proactive dashboard, albeit with less concrete arguments. Besides, limitations were raised regarding the short-term feasibility of a proactive dashboard.
2. Prevent cognitive overload
   - The co-design process and RI exploration yielded similar insights, that is, that too much data in one place would overload caregivers' cognitive workload and that focus of the dashboard should be on providing actionable and only the most relevant information. However, this insight only emerged late in the co-design process (step 4 of 4).
3. Protect privacy
   - The need for privacy protection emerged strongly in the co-design process, and participants clearly pointed to the need for a proactive dashboard in privacy terms. The theme was discussed only briefly in the RI exploration, although some practical suggestions were provided, such as the use of encryption and passwords.
4. Provide transparency
   - While the importance of the transparency of the dashboard's information emerged in the co-design process, practical suggestions on how to provide transparency (eg, training users in correctly interpreting information and explanations) were given only in the RI exploration.
5. Empower caregivers to be in control
   - The main contribution from co-design was the proposition to gradually expand the application of artificial intelligence (AI) functions in practice so that users can get used to an increasing role of AI. In comparison, the RI exploration yielded more in-depth insights and suggestions. The RI exploration stressed that it is important for caregivers not to become too reliant on the results of AI and to have a critical mindset and keep the context in mind.
6. Safeguard accuracy
   - During co-design, the importance of accurate dashboard information was mentioned but not discussed in depth. In the RI exploration, concrete suggestions were made to ensure accuracy, such as including feedback buttons for users.
7. Train users
   - The importance of training, also in relation to other themes such as empowering caregivers to be in control and safeguarding accuracy, frequently appeared in the RI exploration but was raised by only one of the participants in the co-design process. In the RI exploration, suggestions were also provided regarding the focus of training, for instance, on creating awareness about the mediating role of AI in decision-making.

Theme 1: Develop a Proactive Dashboard

The co-design participants generally agreed that the HAAL dashboard should support decision-making proactively, by actively generating and pointing users to relevant insights, rather than passively, by merely showing data from the HAAL technologies. In contrast, the results from the RI survey showed varying viewpoints among PMs regarding the dashboard's required level of proactiveness with regard to supporting decision-making.

Co-design steps 1 and 2 showed that the data from HAAL technologies could be potentially useful for both daily caregivers and caregivers who are less frequently involved (eg, general practitioners). In these co-design steps, there was limited reflection on the possibilities of a dashboard beyond data integration. However, in co-design steps 3 and 4, most participants expressed an interest in a dashboard that also interprets data to provide new information and inspire users. That is, participants suggested that the dashboard should provide insights into or predictions about outliers from usual patterns and distinguish between urgent (eg, a fall) and nonurgent (eg, a deviation in sleeping pattern) outliers to prompt caregivers to take appropriate action. As one of the caregivers at a Taiwanese care center argued, “What I would like is an alert service, more centered on urgency than on daily, routine patient follow-up.” In addition, the dashboard was seen as a way to encourage caregivers to consider signs that might otherwise have been neglected or perceived too late. Besides, some participants of co-design steps 3 and 4 proposed that the dashboard could provide recommendations on how to prevent or address certain deviations from usual patterns.

In the RI survey, most PMs shared pros and cons related to both a descriptive dashboard (scenario A) and a proactive dashboard (scenario B). Most PMs argued that a proactive dashboard could potentially add the most value, especially in terms of enhancing prevention and reducing caregivers’ cognitive load (see also theme 2). At the same time, all PMs expressed doubts about the feasibility of developing a proactive dashboard due to the complexity and relatively limited time span of the HAAL project. Some PMs stressed that the initial acceptance and adoption of a proactive dashboard by caregivers might be low, arguing that the more proactive the dashboard is, the more it may infringe on job satisfaction. As one of the PMs explained, “Caregivers might enjoy the part in their work where they investigate the status of the client, and this is then (partially) taken over by machines.” However, although market introduction was questioned, some PMs advocated exploring possibilities for and experimenting with the more progressive concept of a proactive dashboard to iteratively learn and generate ideas and lessons for future research and development. In an interview, one of the PMs explained, “We know that we could do bigger, smarter things with AI, but you cannot start with high-level AI...But I think that these kinds of projects are useful also to build knowledge and literacy, by making people consider what technology and artificial intelligence could do.”

Theme 2: Prevent Cognitive Overload

The need to prevent cognitive overload was another recurring argument for developing a proactive dashboard in both the co-design process and RI survey. In co-design step 4, it was stressed by multiple participants that too much data or information in one place could exceed caregivers’ cognitive load and cause problems regarding the prioritization of which client, or what aspect of a client’s life, needs attention first. Similarly, in the RI survey, PMs suggested multiple times that a descriptive dashboard may require additional time for caregivers in terms of checking the data, rather than save time, and increase mental strain. As one of the PMs stated, “Adding more data in one place without elaborating on it would not really reduce the caregiver burden.”

Theme 3: Protect Privacy

While privacy was a prominent theme throughout all co-design steps, it was only briefly discussed in the RI exploration. During co-design step 3, multiple participants suggested that from a privacy perspective, a (proactive) dashboard that provides only the most relevant data patterns, notifications, and alerts may be preferred over a (descriptive) dashboard that directly discloses all data about the evolving status of clients in relation to various indicators. This link between the need for a proactive dashboard (scenario B) and privacy concerns was not discussed in the RI survey.

Further, privacy concerns raised in the co-design activities were related to the storage of large amounts of data collected about people with dementia and how these data would be handled. As one of the participants stated, “A lot of personal information is gathered, so you can get to know a lot about people.” In line with this, most participants stated that compliance with the European General Data Protection Regulation should be ensured, and some practical suggestions were made, for instance, to show the client’s home address or room number in the dashboard rather than their names in case of alarms.

The importance of privacy protection was mentioned by various PMs in the first 2 RI workshops, but in the RI survey, only 3 (18%) of the 17 PMs provided input on privacy issues. One of the PMs stated that ways must be found to balance the benefits of large-scale and long-term data collection (eg, in terms of prevention) with downsides such as a feeling of intrusion. Complementary to the co-design process, PMs also provided practical suggestions on privacy protection in the RI survey, such as using a private log-in to the dashboard for caregivers, encryption, or even facial recognition to protect data.

Another privacy concern, raised during co-design step 3, was data accessibility. Several participants reported about who should have access to the dashboard. Some participants proposed that access should be limited to specific caregivers with the specific assignment to learn from the dashboard. In contrast, others reported that all caregivers, including informal carers (eg, family), should have access to the dashboard, if desired. There was no consensus among participants about whether a distinction should be made between different users who are able to see different client data.

Theme 4: Provide Transparency

In co-design step 4, participants proposed that a condition for the use of a proactive dashboard is that users need to understand the reasons (eg, data patterns) behind information provided by the dashboard. In this respect, one of the PMs discussed in an interview that caregivers should not be overloaded with too many details about how specific dashboard information comes about (see also theme 2). In contrast, some co-design participants stressed that users should always be able to examine all data from the different HAAL technologies. Hence, this could be in conflict with the previously discussed insight from co-design that making all data available may be less preferable from a privacy perspective (see also theme 3).

The co-design participants also made various remarks regarding the context specificity of transparency needs. Multiple participants expressed that a need for transparency may not always, or for every user, mean the same. For instance, in case of alarms about certain urgent situations, it may be irrelevant or even distracting to immediately show all data that triggered the alarm. However, users may want to view all the data at a later stage to gain insights into the context and possible causes for the urgent situation, for instance, for training and prevention purposes. A similar insight was raised in the RI exploration, where it was, for example, suggested that in-depth explanations could be provided but only after users ask for it, for instance, by clicking through.

Further, during co-design step 4, it was suggested that once caregivers have built a certain level of trust in the dashboard, less detailed explanations clarifying how the dashboard reaches its conclusions might be sufficient. However, as one of the PMs added in an interview, in the long run, excessive trust might lead to caregivers making certain decisions too easily based on the dashboard's information without critical reflection: "The long-term risk is that users end up trusting the system too much" (see also theme 5).

Although co-design participants highlighted the importance of transparency in HAAL, they did not provide practical suggestions about ways to provide transparency. In the retrospective interviews, various possible explanations were given. For instance, 2 PMs argued that issues such as transparency may have been discussed with limited depth throughout co-design because they pertain more to the backend of the system (ie, algorithms and web services) than the front end (ie, interface) with which users directly interact and because participants may place a certain degree of trust in developers to deal with such issues. Besides, 2 PMs discussed that it may have been hard for co-design participants to formulate requirements regarding transparency during early phases of design because the dashboard concept was still relatively abstract. As suggested, gaining in-depth insights into issues such as these may be easier when practically demonstrating and testing the dashboard in field tests, as users can then actually experience the system and its limitations.

While practical suggestions on providing transparency in HAAL were absent in the co-design results, they were discussed in the RI survey. For instance, PMs suggested (1) showing which specific data were included by algorithms to provide certain information; (2) creating abstractions easy to understand for users to explain the logic behind data analyses, for instance, by giving explanatory examples of common use cases; and (3) training users in interpreting the information and their explanations (see also theme 7).

Theme 5: Empower Caregivers to Be in Control

It was raised in co-design step 4 that people should be in charge of decision-making, regardless of whether human decisions are in line with the dashboard's information. In the same line, multiple PMs argued in the RI survey that people (ie, caregivers) should always be making the final decisions, and they should make these decisions only after carefully valuing the dashboard's information in light of the specific context. It was also suggested during co-design that caregivers may at first instance not be ready yet to get extensive advice from a dashboard. A gradual expansion of AI-functions in real practice was suggested. For instance, in the beginning, the dashboard could provide only generic insights (eg, patterns), alarms, and predictions. In a later stage, when reliability has improved and trust in and experience with the system have been gained, recommendations or conclusions about follow-up steps could be provided. Apart from the above, the importance of people making the final decisions was not further reported by co-design participants.

In contrast, the importance of caregivers being and remaining to be in control of decision-making was more prominent in the RI exploration. In the RI survey, 3 PMs suggested that the long-term use of a proactive dashboard might slowly deprive the intuition of caregivers and maintain an automated and predefined focus whereby one might overlook the person (ie, person with dementia) behind the data. One of the PMs even stated, "There may be a tendency to rely more on AI than own observations and assessments because 'the computer is always right.'" To encourage caregivers to make autonomous decisions while using the dashboard, training was put forward as an important factor by several PMs (see also theme 7).

Theme 6: Safeguard Accuracy

The importance of accurate dashboard information was reflected to a limited extent in the co-design process. During all co-design steps, participants reported a couple of times that the accuracy of the data and data analyses should be regularly evaluated. However, in an interview, a PM suggested that co-design participants mainly shared this requirement as a general condition that must be met before the dashboard could be put into practice, rather than giving concrete ideas on how to achieve this.

The importance of accurate dashboard information and ways to achieve this were more prominently discussed in the RI survey. Multiple PMs argued that information provided by the dashboard should not lead to any faulty judgments by caregivers and that both the data and the algorithms processing data should, therefore, be accurate, without significant biases. For instance, one of the PMs stated, "The dashboard should not give unnecessary warnings to caregivers because the false warning could stimulate the caregivers to impose unnecessary boundaries to people with dementia." One of the PMs explicitly linked accuracy to being sensitive toward the diversity among clients and suggested that the dashboard be fed with data from heterogeneous clients to reduce bias. In contrast to the co-design process, PMs also provided practical suggestions about particular ways of involving users to safeguard accuracy, such as enabling users to (1) provide feedback on data or insights through a button, (2) personalize certain thresholds for alarms to the individual client, (3) keep track of their responses and follow-up actions on the dashboard's information, (4) report nonplausible suggestions and malfunctions, and (5) periodically evaluate the dashboard's functioning. Again, training was put forward as an important factor in this case for users to be able to be involved (see also theme 7).

Theme 7: Train Users

During the co-design activities, one of the participants commented that the proper use of the dashboard would require training and practical learning. In the RI survey, multiple PMs pointed out that training users is an important measure to tackle challenges related to the autonomy of users and the accuracy of the dashboard's information (see also themes 5 and 6). It was suggested that the training should focus on making the users become acquainted with the HAAL technologies; data types; and information provided by the dashboard, including underlying data analyses, and on understanding the impact that the use of the dashboard might have on decision-making. One of the PMs said, "Caregivers should be taught that they will always in some degree be influenced by the information on the dashboard, and be recommended to make their own judgements first." Another PM argued that training should prevent caregivers to become overreliant on the dashboard. In addition, training was suggested to prepare some users for active involvement in maintaining the accuracy of the dashboard information (see also theme 6).

Discussion

Principal Findings

This paper empirically explores how the co-design process of an AI-based DSS for dementia caregivers is complemented by the explicit anticipation of implications. A total of 7 overarching and interlinked themes representing conditions for the responsible development and deployment of the DSS were extracted: develop a proactive dashboard, prevent cognitive overload, protect privacy, provide transparency, empower caregivers to be in control, safeguard accuracy (eg, by reducing false positives), and train users. Because these conditions are interlinked, it is essential for various actors, including developers and users of the DSS, to work together to cohesively address them in practice. Moreover, some conditions, such as to develop a proactive dashboard and empower caregivers to be in charge or to provide transparency through detailed information and prevent cognitive overload, can be at odds with each other and need to be carefully balanced. To gain a deeper understanding about appropriate and responsible levels of proactivity by the DSS, where the contributions of AI and human input in decision-making are balanced, future studies could expand upon prior research in fields such as human factors by exploring and contextualizing notions such as automation bias [40,41] and human automation coordination [42,43] in the context of AI-assisted decision-making in long-term dementia care. Scenarios that may lead to excessive reliance on the automated execution of functions, such as AI-driven data interpretation, could be anticipated, and strategies could be devised to mitigate such scenarios [40].

As our analysis points out, the general expectation of both co-design participants and PMs was that a dashboard that proactively supports decision-making would be most valuable to dementia caregivers. To this regard, the perspectives of co-design participants were fairly aligned; there was a consensus that the dashboard should not show all available data from care technologies. Rather, it should focus on information about significant changes in the data that, for instance, indicate a deterioration of well-being. AI itself was positioned as a technical fix (see also the study by Wehrens et al [44]) to mitigate specific risks related to the remote technology-based monitoring of people with dementia, that is, the infringement of clients' privacy and cognitive overload of caregivers. This is in line with previous studies that show that too much information [45-47] and insufficient time can lead to information overload [48]. The same suggestion of using AI to actually support the responsible embedding of technology in care practice was also found in a scoping review on practical approaches to responsible AI innovation in the context of long-term care [49].

In comparison to the co-design results, the perspectives of PMs in the RI exploration were less unanimous; some PMs shared doubts about the short-term feasibility and acceptance of a proactive dashboard. This discrepancy between results may have been owing to the co-design process being focused on exploring opportunities for innovation, while the RI exploration explicitly invited PMs to reflect on opportunities as well as risks of AI-based analytical functionalities.

Throughout both the co-design process and the RI exploration, various conditions were defined for the responsible development and deployment of a proactive DSS. Similar conditions emerged in the co-design process and RI exploration. However, despite considering and addressing usability requirements, such as minimizing memory load [31,50], in the co-design process, co-design participants generally went into less detail. Compared to PMs in the RI exploration, co-design participants provided fewer practical suggestions on how to meet the RI conditions, except for conditions related to privacy protection. In addition, multiple conditions (ie, preventing cognitive overload, empowering caregivers to be in control, and safeguarding accuracy) emerged in a relatively late stage of the co-design process, once prototyping and reflection on prototypes stood central. Relevant input on implications and conditions for RI emerged more naturally in these phases of co-design, regardless of 2 RI questions related to autonomy and transparency being asked at the end of the last co-design step. Again, these differences in results could potentially be explained by the focus of co-design activities being mainly on opportunities, while the RI exploration was focused on both opportunities and risks.

Hence, the explicit anticipation of implications (ie, the RI exploration) was found to complement the insights from the co-design process in the project under investigation. At the same time, a number of deficiencies can be mentioned regarding the insights that have been gained about social and ethical implications of the DSS. For instance, potential tensions were found between conditions set by different co-design participants. More specifically, to protect privacy, some co-design participants proposed to limit access to information provided by the DSS to specific caregivers. Other participants advocated more transparency and data availability. It is premature to draw conclusions from such contrasting insights. However, it can be stated that insufficient insights were gained into people's individual views on such matters, the interrelatedness of conditions, and potential trade-offs between them. Further, it stood out that both the co-design process and RI exploration yielded limited insights into the dependency of different conditions on context (eg, time, place, and culture). Although it was indicated that trust in the dashboard and transparency needs may change over time, limited insights were gained into how conditions for RI may depend on other contextual factors, such as place and culture. Despite the co-design activities being carried out in multiple countries, no cross-country differences in conditions for the responsible design and deployment of the dashboard were found.

Practical Implications

As argued by Fischer et al [22], differences regarding who is involved in the co-design of care technologies, and how, when, and why they are involved, result in different types of outcomes. To this respect, we discuss 4 considerations that designers and co-design facilitators could take into account to increase the potential for co-design processes to contribute to ethically acceptable, societally desirable, and sustainable deployments of AI-based care technologies.

First, one could strive for balanced attention on both positive and negative implications throughout co-design processes. The co-design process in this case study was focused mostly on functional (ie, what the technology must do) and nonfunctional (eg, usability and reliability) requirements. However, rather than merely eliciting information on the needs, preferences, and requirements of users, co-design processes should go back and forth between needs and opportunities for innovation on the one hand and associated implications on the other hand. In addition, RI necessitates striking a balance in co-design practices between focusing on design aspects, such as usability and esthetics, and considering ethical and social implications. Adhering to specific design standards holds importance to meaningful field tests and the implementation of innovations in practice. However, excessive emphasis on these aspects during early phases of innovation may detract from fostering the innovation's desirability and acceptability. Although research and development projects that *integrate* anticipatory elements into co-design may yield more in-depth insights and be able to more flexibly adapt to insights than projects that anticipate implications separate from the co-design process, a few remarks can be made here. For instance, implications of innovation may need to be anticipated and addressed not only as part of co-design but also in parallel to and beyond the co-design process through methods such as impact assessments, ethical reviews, and foresight exercises. Besides, caution should be exercised to prevent co-design processes from becoming dominated by the anticipation of long-term and wider societal implications, as this may go at the expense of fast iterative design cycles exploring and addressing requirements and direct benefits for users. Further, Sumner et al [21] argued that co-design may require the commitment of a significant amount of time and resources and that some projects may have to rationalize limited resources. Naturally, the same applies to anticipating implications as part of or in parallel to co-design.

Second, one could engage with the perspectives of people who are willing and able to imagine how their interests and their role as users of technology evolve over time (ie, future users), rather than merely involve people from contemporary care practices in co-design. Innovators should not just examine the needs of current users because they may then be insufficiently able to respond to future needs [51]. For instance, in the context of the HAAL project, which was investigated in this study, this could concern the involvement of progressive and technology-savvy dementia caregivers who reflect on how the adoption of increasingly advanced DSSs and other AI technologies will change their work.

Third, one could deliberate on which stakeholders, apart from users, should actually participate in co-design and regularly evaluate how their views guide the underlying direction of innovation. Due to the focus of co-design often being on the needs, expectations, and contexts of individual users, innovators may fail to address potential negative implications, especially implications for other stakeholders or in the long run [52]. Accordingly, it might be relevant to involve certain stakeholders such as intermediary user organizations or social advocacy groups in co-design to articulate societal demands and consider societal implications from a systemic perspective [25,53,54]. For instance, in the context of the HAAL project, this could concern involving nongovernmental organizations that are committed to the privacy interests of older people.

Fourth, one could not only invite but also actively enable users to contribute to the anticipation of implications in co-design. As users are often no experts in (responsible) innovation, they may have difficulties in explicating implications and how they could be addressed, even if explicitly asked for. In this case study, it became more natural for co-design participants to come up with implications in the later phases of co-design (ie, steps 3 and 4) when the dashboard concept had become more tangible. To enable the anticipation of implications *early* in the co-design process, it may be useful to develop inspirational tools that use, for instance, examples of negative impacts of AI technologies [55], envisioning cards [56], or design fiction [36,57] to evoke consideration of the possible intended and unintended short- and long-term effects of future technologies. In addition, in the context of AI-based innovation, one could ensure through training that co-design participants have a basic understanding of what AI can do and how its behavior may be unpredictable and change over time while accumulating data [58,59].

In sum, for co-design processes to result in more RI outcomes, designers and co-design facilitators may need to broaden their scope and reconsider the specific implementation of the process-oriented RI principles of *anticipation* and *inclusion* [17,60]. Even though there are still many uncertainties about the potential uses and consequences of technology during early phases of co-design and before users can "experience" the technology in practice, the anticipation of implications with users ideally starts early, before the technology design has been locked in and change becomes difficult, time-consuming, and expensive [61]. Besides, anticipation should be a recurring element of the innovation process, as people's values and perspectives on what is *responsible* may evolve over time and under the influence of technological innovation [62].

Limitations and Suggestions for Future Research

Given that this paper studies merely a single case, our aim is not to generalize, but rather to illustrate a typical co-design process of an AI-based technology to support the care for older adults and contribute to building a nuanced view on the relation between co-design and RI [63]. Although we use a broad definition for co-design, we acknowledge that there are multiple ways, methods, and instruments to integrate users into the innovation process [21]. Therefore, our findings about the role of anticipating implications in co-design are not generally applicable to co-design. For instance, it is plausible that projects that adopt the value-sensitive design approach yield different results, as this approach aims to explicitly consider the values of users and other stakeholders and how these values are affected by the envisioned technology [64-66]. In other words, some approaches to co-design may in themselves impose on facilitators to explore the values at stake and thereby the implications of innovation. Future research could examine to what extent such approaches support RI.

Further, we recognize that there are limitations to the RI exploration that was part of our study and thus to the insights gained into conditions for the responsible development and deployment of DSSs in dementia care. Our RI exploration initially focused on the perspectives of PMs to stimulate and facilitate whole-team participation in exploring how RI could be addressed throughout the HAAL project. The underlying assumption was that RI cannot be prescribed to innovators but needs to be conceptualized and addressed "in context" by those who actually perform the research, design, development, and testing with users [67,68]. However, soliciting PMs' perspectives provided neither a complete nor necessarily an accurate picture about implications and ways they can be addressed. To this end, future studies could consider embedding trained ethicists in the research team who can provide *top-down* guidance and inspiration (eg, contextualized ethics principles) during *bottom-up* engagement with users and other stakeholders [49,69]. Besides, future research could explore the perspectives of users on RI in the context of AI-based care technologies, such as DSSs, for instance, what values come to matter most to them, what positive and negative implications they foresee, how they perceive the urgency of (other) known implications in their context, and how they look at certain strategies to address implications (eg, see the study of Lukkien et al [70]). In doing so, the perspectives of stakeholders from different care contexts (eg, care organizations or countries) can be captured with sufficient detail and be compared to learn how to account for the context specificity of values in technology design and deployment [71,72]. In addition, the perspectives of people with dementia should be clarified, even when they are only a passive user of the technology (as is often the case with DSSs), and despite these people often having difficulties in expressing their needs [73,74].

Finally, even though all co-design activities and the RI exploration had already been completed by the time the objectives for this case study were established, the RI exploration had a minor effect on the co-design process. For instance, some co-design researchers were also participants in the RI exploration, which could have affected the co-design activities. Besides, at the request of DRML (who led the RI exploration), the usability study (co-design step 4) included 2 RI-related questions. In our results, we explicated that co-design participants already discussed more implications before these 2 questions were asked. Without this minor effect, there may have been a greater knowledge gap between the results from the co-design process and RI exploration in HAAL. However, to gain more robust results into the role of the anticipation of implications in co-design, future research could study co-design processes completely separately from an exploration of associated implications.

Conclusions

In this paper, we explored how the co-design process of an AI-based DSS for dementia caregivers is complemented by the explicit anticipation of social and ethical implications. Co-design is an essential means to feed the development and deployment of AI-based care technologies with insights about needs of targeted users and collectively translate these needs into requirements for technology design. Besides, as found in this empirical study, certain implications and strategies to address these implications may be naturally anticipated in co-design, even though users may not necessarily think in terms of implications or risks, but rather in terms of conditions before the technology can be used. At the same time, this case study indicates that a co-design process that focuses on opportunities rather than balancing attention for both positive and negative implications may result in knowledge gaps related to implications and how they can be addressed. In the pursuit of responsible outcomes, co-design facilitators could consider broadening the scope of co-design processes, for instance, by moving back and forth between opportunities and associated implications of innovation, involving future users and social advocacy groups in such an inquiry, and ensuring that co-design participants are provided with inspiration and have basic knowledge and skills to contribute to anticipating implications. Explicit *anticipation* of implications in co-design and broader *inclusion* of stakeholders in doing so increase opportunities for innovators to start addressing implications of innovation before the technology design has been locked in.

Acknowledgments

The authors gratefully acknowledge support from the Active and Assisted Living (AAL) program, cofinanced by the European Commission through the Horizon 2020 Societal Challenge Health, Demographic Change, and Wellbeing. In particular, the work reported here has been supported by the AAL Healthy Ageing Eco-system for People With Dementia (HAAL) project (AAL-2020-7-229-CP). In addition, the authors thank their HAAL project partners in the Netherlands, Italy, Taiwan, and Denmark for organizing and participating in the research activities that provided the basis for this study.

Authors' Contributions

DRML contributed to conceptualization, methodology, validation, investigation, formal analysis, writing the original draft, reviewing and editing the manuscript, and funding acquisition. SIA contributed to methodology, investigation, formal analysis, and reviewing and editing the manuscript. NES contributed to methodology, investigation, and reviewing and editing the manuscript. BMH contributed to formal analysis and reviewing and editing the manuscript. HHN contributed to conceptualization, methodology, validation, reviewing and editing the manuscript, project administration, and funding acquisition. WPCB contributed to conceptualization, methodology, and reviewing and editing the manuscript. AP contributed to conceptualization, methodology, and reviewing and editing the manuscript. EHMM contributed to conceptualization, methodology, and reviewing and editing the manuscript. MMNM contributed to conceptualization and methodology. All authors contributed to writing (original draft).

Conflicts of Interest

None declared.

Multimedia Appendix 1

The responsible innovation survey.

[DOCX File , 161 KB - humanfactors_v11i1e55961_app1.docx]

References

1. Rubeis G. The disruptive power of artificial intelligence. Ethical aspects of gerontechnology in elderly care. Arch Gerontol Geriatr 2020;91:104186 [FREE Full text] [doi: 10.1016/j.archger.2020.104186] [Medline: 32688106]
2. Ienca M, Jotterand F, Elger B, Caon M, Scoccia Pappagallo A, Kressig RW, et al. Intelligent assistive technology for Alzheimer's disease and other dementias: a systematic review. J Alzheimers Dis 2017;60(1):333. [doi: 10.3233/JAD-179005] [Medline: 28869482]
3. Mukaetova-Ladinska EB, Harwoord T, Maltby J. Artificial Intelligence in the healthcare of older people. Arch Psychiatr Ment Health 2020;4:7-13 [FREE Full text] [doi: 10.29328/journal.apmh.1001011]
4. Loveys K, Prina M, Axford C, Domènec Ò, Weng W, Broadbent E, et al. Artificial intelligence for older people receiving long-term care: a systematic review of acceptability and effectiveness studies. Lancet Healthy Longev 2022 Apr;3(4):e286-e297 [FREE Full text] [doi: 10.1016/S2666-7568(22)00034-4] [Medline: 35515814]
5. Xie BO, Tao C, Li J, Hilsabeck RC, Aguirre A. Artificial intelligence for caregivers of persons with Alzheimer's disease and related dementias: systematic literature review. JMIR Med Inform 2020 Aug 20;8(8):e18189 [FREE Full text] [doi: 10.2196/18189] [Medline: 32663146]
6. Yeung K. Recommendation of the council on artificial intelligence (OECD). Int Leg Mater 2020 Mar 12;59(1):27-34 [FREE Full text] [doi: 10.1017/ilm.2020.5]
7. Akbar S, Lyell D, Magrabi F. Automation in nursing decision support systems: a systematic review of effects on decision making, care delivery, and patient outcomes. J Am Med Inform Assoc 2021 Oct 12;28(11):2502-2513 [FREE Full text] [doi: 10.1093/jamia/ocab123] [Medline: 34498063]
8. Lee S. Features of computerized clinical decision support systems supportive of nursing practice: a literature review. Comput Inform Nurs 2013 Oct;31(10):477-496. [doi: 10.1097/01.NCN.0000432127.99644.25] [Medline: 23958964]
9. Middleton B, Sittig DF, Wright A. Clinical decision support: a 25 year retrospective and a 25 year vision. Yearb Med Inform 2018 Mar 06;25(S 01):S103-S116 [FREE Full text] [doi: 10.15265/iys-2016-s034]
10. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020;3:17 [FREE Full text] [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
11. Berridge C, Grigorovich A. Algorithmic harms and digital ageism in the use of surveillance technologies in nursing homes. Front Sociol 2022;7:957246 [FREE Full text] [doi: 10.3389/fsoc.2022.957246] [Medline: 36189442]
12. Stypinska J. AI ageism: a critical roadmap for studying age discrimination and exclusion in digitalized societies. AI Soc 2023;38(2):665-677 [FREE Full text] [doi: 10.1007/s00146-022-01553-5] [Medline: 36212226]
13. Zuboff S. Big other: surveillance capitalism and the prospects of an information civilization. J Inf Technol 2015 Mar 01;30(1):75-89 [FREE Full text] [doi: 10.1057/jit.2015.5]
14. Morley J, Machado CC, Burr C, Cowls J, Joshi I, Taddeo M, et al. The ethics of AI in health care: a mapping review. Soc Sci Med 2020 Sep;260:113172 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113172] [Medline: 32702587]
15. Peine A, Neven L. From intervention to co-constitution: new directions in theorizing about aging and technology. Gerontologist 2019 Jan 09;59(1):15-21 [FREE Full text] [doi: 10.1093/geront/gny050] [Medline: 29850812]
16. von Schomberg R. A vision of responsible research and innovation. In: Owen R, Bessant J, Heintz M, editors. Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society. Chichester, UK: John Wiley & Sons; 2013:51-74.
17. Owen R, Stilgoe J, Macnaghten P, Gorman M, Fisher E, Guston D. A framework for responsible innovation. In: Owen R, Bessant J, Heintz M, editors. Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society. Chichester, UK: Wiley-Blackwell; 2013:27-50.
18. Sanders EB, Stappers PJ. Co-creation and the new landscapes of design. CoDesign 2008 Mar;4(1):5-18 [FREE Full text] [doi: 10.1080/15710880701875068]
19. Vargas C, Whelan J, Brimblecombe J, Allender S. Co-creation, co-design, co-production for public health - a perspective on definition and distinctions. Public Health Res Pract 2022 Jun 15;32(2):3222211 [FREE Full text] [doi: 10.17061/phrp3222211] [Medline: 35702744]

20. Merkel S, Kucharski A. Participatory design in gerontechnology: a systematic literature review. Gerontologist 2019 Jan 09;59(1):e16-e25 [FREE Full text] [doi: 10.1093/geront/gny034] [Medline: 29788319]
21. Sumner J, Chong LS, Bundele A, Wei Lim Y. Co-designing technology for aging in place: a systematic review. Gerontologist 2021 Sep 13;61(7):e395-e409 [FREE Full text] [doi: 10.1093/geront/gnaa064] [Medline: 32506136]
22. Fischer B, Peine A, Östlund B. The importance of user involvement: a systematic review of involving older users in technology design. Gerontologist 2020 Sep 15;60(7):e513-e523 [FREE Full text] [doi: 10.1093/geront/gnz163] [Medline: 31773145]
23. Jansma SR, Dijkstra AM, de Jong MD. Co-creation in support of responsible research and innovation: an analysis of three stakeholder workshops on nanotechnology for health. J Responsible Innov 2021 Nov 23;9(1):28-48 [FREE Full text] [doi: 10.1080/23299460.2021.1994195]
24. Nap HH, Lukkien DR, Lin CC, Lin CJ, Chieh HF, Wong YT, et al. HAAL: a healthy ageing eco-system for people with dementia. Gerontechnology 2022;21:1-5 [FREE Full text] [doi: 10.4017/gt.2022.21.s.587.sp4]
25. Stewart J, Hyysalo S. Intermediaries, users and social learning in technological innovation. Int J Innov Manag 2011 Nov 20;12(03):295-325 [FREE Full text] [doi: 10.1142/S1363919608002035]
26. Barbarossa F, Amabili G, Margaritini A, Morresi N, Casaccia S, Marconi F, et al. Design, development, and usability evaluation of a dashboard for supporting formal caregivers in managing people with dementia. In: Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments. 2023 Presented at: PETRA '23; July 5-7, 2023; Corfu, Greece p. 154-161 URL: https://doi.org/10.1145/3594806.3594820 [doi: 10.1145/3594806.3594820]
27. Kuhn J. Decrypting the MoSCoW analysis. ItSM Solut. 2009. URL: http://www.itsmsolutions.com/newsletters/DITYvol5iss44.htm [accessed 2024-04-29]
28. Bradbury K, Watts S, Arden-Close E, Yardley L, Lewith G. Developing digital interventions: a methodological guide. Evid Based Complement Alternat Med 2014;2014:561320 [FREE Full text] [doi: 10.1155/2014/561320] [Medline: 24648848]
29. Broekhuis M, van Velsen L. Improving usability benchmarking for the eHealth domain: the development of the eHealth UsaBility Benchmarking instrument (HUBBI). PLoS One 2022;17(2):e0262036 [FREE Full text] [doi: 10.1371/journal.pone.0262036] [Medline: 35176033]
30. Bastien JM, Scapin DL. A validation of ergonomic criteria for the evaluation of human ‐ computer interfaces. Int J Hum Comput Interact 1992 Apr;4(2):183-196 [FREE Full text] [doi: 10.1080/10447319209526035]
31. Nielsen J. Reliability of severity estimates for usability problems found by heuristic evaluation. In: Proceedings of the Posters and Short Talks of the 1992 SIGCHI Conference on Human Factors in Computing Systems. 1992 Presented at: CHI '92; May 3-7, 1992; Monterey, CA p. 129-130 URL: https://doi.org/10.1145/1125021.1125117 [doi: 10.1145/1125021.1125117]
32. Taveter K, Sterling L, Pedell S, Burrows R, Taveter EM. A method for eliciting and representing emotional requirements: two case studies in e-Healthcare. In: Proceedings of the 27th International Requirements Engineering Conference Workshops. 2019 Presented at: REW '19; September 23-27, 2019; Jeju, South Korea p. 100-105 URL: https://doi.org/10.1109/REW.2019.00021 [doi: 10.1109/rew.2019.00021]
33. Verbeek PP, Tijink D. Guidance ethics approach: an ethical dialogue about technology with perspective on actions. ECP: Platform voor de InformatieSamenleving. 2020. URL: https://ris.utwente.nl/ws/portalfiles/portal/247401391/060_002_Boek_Guidance_ethics_approach_Digital_EN.pdf [accessed 2024-04-29]
34. Mosavi NS, Santos MF. How prescriptive analytics influences decision making in precision medicine. Procedia Comput Sci 2020;177:528-533 [FREE Full text] [doi: 10.1016/j.procs.2020.10.073]
35. El Morr C, Ali-Hassan H. Healthcare, data analytics, and business intelligence. In: El Morr C, Ali-Hassan H, editors. Analytics in Healthcare: A Practical Introduction. Cham, Switzerland: Springer; 2019:1-13.
36. Noortman R, Schulte BF, Marshall P, Bakker S, Cox AL. HawkEye - deploying a design fiction probe. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19; May 4-9, 2019; Glasgow, UK p. 1-14 URL: https://doi.org/10.1145/3290605.3300652 [doi: 10.1145/3290605.3300652]
37. Ethics and governance of artificial intelligence for health: WHO guidance. World Health Organization. URL: https://www.who.int/publications/i/item/9789240029200 [accessed 2024-04-29]
38. Braun V, Clarke V. Thematic analysis. In: Cooper H, Camic PM, Long DL, Panter AT, Rindskopf D, Sher KJ, editors. APA Handbook of Research Methods in Psychology, Vol. 2. Research Designs: Quantitative, Qualitative, Neuropsychological, and Biologica. New York, NY: American Psychological Association; 2012:57-71.
39. Your research: Is it subject to the WMO or not? Central Committee on Research Involving Human Subjects. URL: https://english.ccmo.nl/investigators/legal-framework-for-medical-scientific-research/your-research-is-it-subject-to-the-wmo-or-not [accessed 2024-07-23]
40. Parasuraman R, Riley V. Humans and automation: use, misuse, disuse, abuse. Hum Factors 2016 Nov 23;39(2):230-253 [FREE Full text] [doi: 10.1518/001872097778543886]
41. Goddard K, Roudsari A, Wyatt JC. Automation bias: empirical results assessing influencing factors. Int J Med Inform 2014 May;83(5):368-375 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.01.001] [Medline: 24581700]
42. Dekker SW, Woods DD. MABA-MABA or abracadabra? progress on human–automation co-ordination. Cogn Technol Work 2002;4:240 [FREE Full text] [doi: 10.1007/S101110200022]

43. Dafoe A, Bachrach Y, Hadfield G, Horvitz E, Larson K, Graepel T. Cooperative AI: machines must learn to find common ground. Nature 2021 May;593(7857):33-36 [FREE Full text] [doi: 10.1038/d41586-021-01170-0] [Medline: 33947992]
44. Wehrens R, Stevens M, Kostenzer J, Weggelaar AM, de Bont A. Ethics as discursive work: the role of ethical framing in the promissory future of data-driven healthcare technologies. Sci Technol Hum Values 2021 Nov 08;48(3):606-634 [FREE Full text] [doi: 10.1177/01622439211053661]
45. Hiltz SR, Turoff M. Structuring computer-mediated communication systems to avoid information overload. Commun ACM 1985;28:680-689 [FREE Full text] [doi: 10.1145/3894.3895]
46. Hettinger LJ, Nelson WT, Haas MW. Applying virtual environment technology to the design of fighter aircraft cockpits: pilot performance and situation awareness in a simulated air combat task. Proc Hum Factors Ergon Soc Annu Meet 2016 Nov 06;38(1):115-118 [FREE Full text] [doi: 10.1177/154193129403800123]
47. Rosekind MR, Gander PH, Miller DL, Gregory KB, Smith RM, Weldon KJ, et al. Fatigue in operational settings: examples from the aviation environment. Hum Factors 1994 Jun;36(2):327-338 [FREE Full text] [doi: 10.1177/001872089403600212] [Medline: 8070796]
48. Schick AG, Gordon LA, Haka S. Information overload: a temporal approach. Account Organ Soc 1990 Jan;15(3):199-220 [FREE Full text] [doi: 10.1016/0361-3682(90)90005-f]
49. Lukkien DR, Nap HH, Buimer HP, Peine A, Boon WP, Ket JC, et al. Toward responsible artificial intelligence in long-term care: a scoping review on practical approaches. Gerontologist 2023 Jan 24;63(1):155-168 [FREE Full text] [doi: 10.1093/geront/gnab180] [Medline: 34871399]
50. Chew JC, Molich R. Heuristic evaluation of user interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1990 Presented at: CHI '90; April 1-5, 1990; Washington, DC p. 249-256 URL: https://doi.org/https://doi.org/10.1145/97243.97281 [doi: 10.1145/97243.97281]
51. Christensen CM. Innovator's Dilemma: When New Technologies Cause Great Firms to Fail. Boston, MA: Harvard Business Review Press; 2013.
52. van Velsen L, Ludden G, Grünloh C. The limitations of user-and human-centered design in an eHealth context and how to move beyond them. J Med Internet Res 2022 Oct 05;24(10):e37341 [FREE Full text] [doi: 10.2196/37341] [Medline: 36197718]
53. Boon WP, Moors EH, Kuhlmann S, Smits RE. Demand articulation in emerging technologies: intermediary user organisations as co-producers? Research Policy 2011 Mar;40(2):242-252 [FREE Full text] [doi: 10.1016/j.respol.2010.09.006]
54. Kivimaa P, Boon W, Hyysalo S, Klerkx L. Towards a typology of intermediaries in sustainability transitions: a systematic review and a research agenda. Res Policy 2019 May;48(4):1062-1075 [FREE Full text] [doi: 10.1016/j.respol.2018.10.006]
55. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019 Oct 25;366(6464):447-453 [FREE Full text] [doi: 10.1126/science.aax2342] [Medline: 31649194]
56. Friedman B, Hendry DG. The envisioning cards: a toolkit for catalyzing humanistic and technical imaginations. In: Proceedings of the 2012 SIGCHI Conference on Human Factors in Computing Systems. 2012 Presented at: CHI '12; May 5-10, 2012; Austin, TX p. 1145-1148 URL: https://doi.org/10.1145/2207676.2208562 [doi: 10.1145/2207676.2208562]
57. Dunne A, Raby F. Speculative Everything: Design, Fiction, and Social Dreaming. New York, NY: MIT press; 2013.
58. Bratteteig T, Verne G. Does AI make PD obsolete?: exploring challenges from artificial intelligence to participatory design. In: Proceedings of the 15th Participatory Design Conference: Short Papers, Situated Actions, Workshops and Tutorial - Volume 2. 2018 Presented at: PDC '18; August 20-24, 2018; Hasselt and Genk, Belgium p. 1-5 URL: https://doi.org/10.1145/3210604.3210646 [doi: 10.1145/3210604.3210646]
59. Donia J, Shaw JA. Co-design and ethical artificial intelligence for health: an agenda for critical research and practice. Big Data Soc 2021 Dec 17;8(2):205395172110652 [FREE Full text] [doi: 10.1177/20539517211065248]
60. Fraaije A, Flipse SM. Synthesizing an implementation framework for responsible research and innovation. J Responsible Innov 2019 Nov 21;7(1):113-137 [FREE Full text] [doi: 10.1080/23299460.2019.1676685]
61. Genus A, Stirling A. Collingridge and the dilemma of control: towards responsible and accountable innovation. Res Policy 2018 Feb;47(1):61-69 [FREE Full text] [doi: 10.1016/j.respol.2017.09.012]
62. Kudina O, Verbeek PP. Ethics from Within: Google Glass, the collingridge dilemma, and the mediated value of privacy. Sci Technol Human Values 2018 Aug 21;44(2):291-314 [FREE Full text] [doi: 10.1177/0162243918793711]
63. Flyvbjerg B. Five misunderstandings about case-study research. Qual Inq 2016 Jun 29;12(2):219-245 [FREE Full text] [doi: 10.1177/1077800405284363]
64. van der Hoven J, Manders-Huits N. The need for a value-sensitive design of communication infrastructures. In: Miller KW, Taddeo M, editors. The Ethics of Information Technologies. London, UK: Routledge; 2020:329-332.
65. Friedman B, Kahn Jr PH, Borning A, Huldtgren A. Value sensitive design and information systems. In: Doorn N, Schuurbiers D, van de Poel I, Gorman ME, editors. Early Engagement And New Technologies: Opening Up the Laboratory. Cham, Switzerland: Springer; 2013:55-95.
66. Smits M, Nacar M, D S Ludden G, van Goor H. Stepwise design and evaluation of a values-oriented ambient intelligence healthcare monitoring platform. Value Health 2022 Jun;25(6):914-923 [FREE Full text] [doi: 10.1016/j.jval.2021.11.1372] [Medline: 35525831]

67. Forsberg EM, Thorstensen E, Dias Casagrande F, Holthe T, Halvorsrud L, Lund A, et al. Is RRI a new R and I logic? a reflection from an integrated RRI project. J Responsib Technol 2021 May;5:100007 [FREE Full text] [doi: 10.1016/j.jrt.2020.100007]
68. Stahl BC. Responsible innovation ecosystems: ethical implications of the application of the ecosystem concept to artificial intelligence. Int J Inf Manage 2022 Feb;62:102441 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2021.102441]
69. McLennan S, Fiske A, Celi LA, Müller R, Harder J, Ritt K, et al. An embedded ethics approach for AI development. Nat Mach Intell 2020 Jul 31;2(9):488-490 [FREE Full text] [doi: 10.1038/S42256-020-0214-1]
70. Lukkien DRM, Stolwijk NE, Ipakchian Askari S, Hofstede BM, Nap HH, Boon WPC, et al. AI-Assisted Decision-Making in Long-Term Care: Qualitative Study on Prerequisites for Responsible Innovation. JMIR Nurs 2024 Jul 25;7:e55962 [FREE Full text] [doi: 10.2196/55962] [Medline: 39052315]
71. ÓhÉigeartaigh SS, Whittlestone J, Liu Y, Zeng Y, Liu Z. Overcoming barriers to cross-cultural cooperation in AI ethics and governance. Philos Technol 2020 May 15;33(4):571-593 [FREE Full text] [doi: 10.1007/s13347-020-00402-x]
72. Pollock N, Williams R, D'Adderio L. Global software and its provenance: generification work in the production of organizational software packages. Soc Stud Sci 2016 Jun 29;37(2):254-280 [FREE Full text] [doi: 10.1177/0306312706066022]
73. Papoutsi C, Wherton J, Shaw S, Morrison C, Greenhalgh T. Putting the social back into sociotechnical: case studies of co-design in digital health. J Am Med Inform Assoc 2021 Feb 15;28(2):284-293 [FREE Full text] [doi: 10.1093/jamia/ocaa197] [Medline: 33043359]
74. Suijkerbuijk S, Nap HH, Cornelisse L, IJsselsteijn WA, de Kort YA, Minkman MM. Active involvement of people with dementia: a systematic review of studies developing supportive technologies. J Alzheimers Dis 2019 Jun 18;69(4):1041-1065 [FREE Full text] [doi: 10.3233/jad-190050]

Abbreviations

AAL: Active and Assisted Living
AI: artificial intelligence
DSS: decision support system
HAAL: Healthy Ageing Eco-system for People With Dementia
PM: project member
RI: responsible innovation

Edited by A Kushniruk, E Borycki; submitted 31.12.23; peer-reviewed by M Coccia, S Mitra, A Hidki; comments to author 18.02.24; accepted 02.06.24; published 31.07.24.

Please cite as:
Lukkien DRM, Ipakchian Askari S, Stolwijk NE, Hofstede BM, Nap HH, Boon WPC, Peine A, Moors EHM, Minkman MMN
Making Co-Design More Responsible: Case Study on the Development of an AI-Based Decision Support System in Dementia Care
JMIR Hum Factors 2024;11:e55961
URL: https://humanfactors.jmir.org/2024/1/e55961
doi:10.2196/55961
PMID:

Review

Barriers to and Facilitators of Artificial Intelligence Adoption in Health Care: Scoping Review

Masooma Hassan[1], BCom, MSc; Andre Kushniruk[1], BSc, BA, MSc, PhD; Elizabeth Borycki[1], RN, HBScN, MN, PhD

Department of Health Information Science, University of Victoria, Victoria, BC, Canada

Corresponding Author:
Masooma Hassan, BCom, MSc
Department of Health Information Science
University of Victoria
HSD Building, A202
Victoria, BC, V8W 2Y2
Canada
Phone: 1 6472876274
Email: masooma.d.hassan@gmail.com

Abstract

Background: Artificial intelligence (AI) use cases in health care are on the rise, with the potential to improve operational efficiency and care outcomes. However, the translation of AI into practical, everyday use has been limited, as its effectiveness relies on successful implementation and adoption by clinicians, patients, and other health care stakeholders.

Objective: As adoption is a key factor in the successful proliferation of an innovation, this scoping review aimed at presenting an overview of the barriers to and facilitators of AI adoption in health care.

Methods: A scoping review was conducted using the guidance provided by the Joanna Briggs Institute and the framework proposed by Arksey and O'Malley. MEDLINE, IEEE Xplore, and ScienceDirect databases were searched to identify publications in English that reported on the barriers to or facilitators of AI adoption in health care. This review focused on articles published between January 2011 and December 2023. The review did not have any limitations regarding the health care setting (hospital or community) or the population (patients, clinicians, physicians, or health care administrators). A thematic analysis was conducted on the selected articles to map factors associated with the barriers to and facilitators of AI adoption in health care.

Results: A total of 2514 articles were identified in the initial search. After title and abstract reviews, 50 (1.99%) articles were included in the final analysis. These articles were reviewed for the barriers to and facilitators of AI adoption in health care. Most articles were empirical studies, literature reviews, reports, and thought articles. Approximately 18 categories of barriers and facilitators were identified. These were organized sequentially to provide considerations for AI development, implementation, and the overall structure needed to facilitate adoption.

Conclusions: The literature review revealed that trust is a significant catalyst of adoption, and it was found to be impacted by several barriers identified in this review. A governance structure can be a key facilitator, among others, in ensuring all the elements identified as barriers are addressed appropriately. The findings demonstrate that the implementation of AI in health care is still, in many ways, dependent on the establishment of regulatory and legal frameworks. Further research into a combination of governance and implementation frameworks, models, or theories to enhance trust that would specifically enable adoption is needed to provide the necessary guidance to those translating AI research into practice. Future research could also be expanded to include attempts at understanding patients' perspectives on complex, high-risk AI use cases and how the use of AI applications affects clinical practice and patient care, including sociotechnical considerations, as more algorithms are implemented in actual clinical environments.

(JMIR Hum Factors 2024;11:e48633) doi:10.2196/48633

KEYWORDS

artificial intelligence; governance; health information systems; artificial intelligence adoption; system implementation; health care organizations; health services; mobile phone

Introduction

Background

The onset of the 2020 COVID-19 pandemic has particularly triggered health care organizations across the globe to consider transforming their health delivery models. According to the 2024 Global Health Care Sector Outlook report published by Deloitte, hospitals and health care organizations are addressing challenges by turning toward novel technologies such as cloud computing, artificial intelligence (AI), 5G telecommunications, and interoperable data and analytics to enable care via digital models [1]. It was not too long ago, in 2017, when the Canadian government created the Pan-Canadian Artificial Intelligence Strategy and announced an investment plan of CAD $125 million (US $96 million) for research in AI. The Canadian Institute for Advanced Research was mandated to lead this strategy forward with a 5-year plan to enhance Canada's AI innovation profile on the international stage. Health care is one of the 4 sectors on which the Canadian Institute for Advanced Research is focusing for the advancement of AI research [2]. However, health care has seen slow success in the implementation of AI use cases.

Objective

The objective of this review was to investigate what is known from existing literature about the barriers to and facilitators of AI adoption in health care and propose recommendations on approaches that would address barriers to adoption.

We begin this paper by (1) defining AI (before providing some context for AI's use in health care), (2) describing the most prominent applications of AI in health care, (3) outlining the value that AI is expected to provide, and (4) providing a rationale for this review.

History and Definitions

AI is not necessarily a new concept; rather, the exploration of this innovation goes as far back as 10th century China, when mechanical engineer Yan Shi presented to Emperor Zhou mechanical men capable of independently moving their bodies. In the 12th century, al Jazari, who was a polymath, an inventor, and a mechanical engineer, developed humanoid robots. Furthermore, in the 15th century European Renaissance, Leonardo da Vinci similarly developed a knight robot that was able to move different parts of its body on its own. The definition of AI has changed over time, from referring to robotic machines to much more sophisticated technologies capable of mimicking human decision-making processes and behaviors. The advancement of computer systems and languages in the more recent decades has made it possible to progress toward AI systems. The definition that most fits today's application of AI and is referenced in this paper was coined by John McCarthy in 1956. McCarthy defined AI as "the science and engineering of making intelligent machines" [3]. It is unclear what definition is consistently used; however, what is clear is that today's AI encompasses various techniques aimed at mimicking humanlike intelligence and behavior to allow for the emergence of intelligent technologies capable of problem-solving and decision-making. In this way, AI should be skilled at processing large amounts of information, should arrive at a conclusion through reasoning, and have the ability to learn and solve problems on its own [4]. Various analytic techniques are used to allow for this, with the most prominent ones falling under machine learning (ML) and natural language processing (NLP) [5].

Large data sets are needed to develop effective AI algorithms and enable AI's maturity to arrive at intelligent outputs. In health care, the sources of data for NLP are primarily unstructured data, for example, free-text clinical notes from electronic medical records (EMRs). ML techniques use structured data such as diagnostic images and genomic data. ML uses two primary types of algorithms: (1) supervised and (2) unsupervised. Supervised learning provides more clinically relevant results; hence, AI applications mostly use supervised learning. There are several techniques in supervised learning, with neural networks and support vector machines being the most popular of the techniques [5]. The most modern extension of the neural network technique is called deep learning (DL). DL has been made possible due to the increasing availability of large amounts of complex data. This technique has become more popular because of the number of layers of data it can translate. NLP can be used to convert unstructured data into structured data. Therefore, both NLP and ML, along with additional data, are required to train the AI continuously. The more data that are fed into the AI, the "smarter" it becomes. In health care, data sets can be available from various sources, such as electronic health records (EHRs), laboratory tests, diagnostic imaging, electrodiagnosis, genetic diagnosis, and mass screening [5]. In 2022, the release of ChatGPT (OpenAI) brought to light the power of large language models. This type of chatbot-style generative AI is being considered to enable extracting data from EMRs and converting them into meaningful outputs that can be useful for clinicians by lowering their administrative burden [6].

Current State of AI Research and Health Care Use Cases

Research in AI has been exponentially increasing, with bibliometric reporting of published articles on the topic of health care having increased at an annual growth rate of 5.12% over the past 28 years. As of 2021, the most significant increases in bibliometric reporting took place in the 3 years before 2021 [7]. According to Tran et al [8], the disciplines with the highest number of publications at the intersection of AI and health include cancer, heart diseases and stroke, ophthalmology, Alzheimer disease, and depression. Most publications on the types of AI used reported on robotics, ML, and DL.

In health care, publications of AI applications are concentrated around operational or administrative efficiency as well as patient care improvement, including better outcomes through improved diagnosis and treatment [9]. AI enhances operational and administrative efficiency by providing administrative support to health professionals and improving performance across the organization. AI can achieve this through, for example, its ability to consolidate and provide the latest and most validated research findings that can support clinicians with up-to-date evidence-based decision-making while providing care and its ability to leverage EHR data to predict data heterogeneity between various hospitals and clinics [7]. Emergency departments are largely found to have successfully applied AI to optimize resource planning and crowd management [10,11]; for example, the Hospital for Sick Children and Humber River Hospital in Ontario, Canada, are using AI to improve emergency department operations by predicting patient surges in the emergency waiting room [12,13].

Use cases aimed at improving care outcomes include predictive analytics around disease outcome prediction or prognosis evaluation and clinical decision support systems [7]. Examples of such cases are found in cardiology and include the early detection of atrial fibrillation via a smartphone-based electrocardiogram or cardiovascular risk assessment via patient records. Other promising areas include neurology, specifically stroke prediction and diagnosis [5]. Gastroenterology AI applications have also been successfully tested, where algorithms are used to predict outcomes in cases of esophageal cancer and metastasis in colorectal cancer [14]. Image-based diagnosis is considered the most successful use of AI applications in health care, largely supporting radiology, dermatology, ophthalmology, and pathology [15]. In a review conducted on the literature on AI use in the emergency department, Kirubarajan et al [11] reported that 50% of the studies found that AI interventions were better able to diagnose various ailments, such as acute cardiac events and hyperkalemia, among other health conditions.

As mentioned earlier, AI requires large amounts of data to learn and apply sophisticated reasoning and accurate problem-solving. In addition to the race toward researching AI use cases, a surge in health care data is further setting the stage to allow for accelerated AI innovations [16]. EMR data; wearable sensor technology; and genomic, pharmaceutical, and research databases offer opportunities to apply AI to the analysis of health data. Approximately 30% of the world's data volume is generated by the health care industry. The compound annual growth rate of data for health care is expected to reach 36% by 2025 [17]. This growth of data volume in health care is faster than that in manufacturing, financial services, and media and entertainment industries [18]. This sets the stage well for developing AI technologies that can be integrated into health care practice, as algorithms now have more data to provide increasingly sophisticated outputs.

Contributions of the Research

It is clear that the changing landscape, increasing evidence on AI use cases, and increasing availability of data in health care are setting the path toward realizing real-life applications of AI. However, successful utility requires successful adoption, and a number of studies have reported on the challenges encountered with implementing AI in health care [19-21]. Health care organizations are especially complex and can be resistant to change due to various reasons associated with legacy structures, a shortage of resources, and high demand. An estimated 70% of health IT projects fail [22], and an important characteristic of successful technological implementation is tied to its adoption, which is why adoption is a key component of frameworks such as the unified theory of acceptance and use of technology and nonadoption, abandonment, scale-up, spread, and sustainability theory. These frameworks are used to evaluate and study the acceptance of technologies. For example, the unified theory of acceptance and use of technology framework, which integrates all the available theories about technology adoption, suggests several factors that help understand users' intention to adopt and use a technology. It looks at all the available theories about technology adoption to evaluate use of information systems [23]. Similarly, the nonadoption, abandonment, scale-up, spread, and sustainability framework has incorporated multiple theories to help study factors influencing "non-adoption, abandonment and challenges to scale-up, spread and sustainability of technology-supported change efforts" [24]. Both emphasize the importance of studying adoption to support the successful uptake of technologies beyond implementation. With these reasons in mind, it is important that organizations understand the barriers to and facilitators of AI adoption to ensure successful AI implementation. In reference to the widely known work of Everett Rogers, famously known as the Rogers diffusion of innovation theory, Cresswell and Sheikh [25] have defined implementation as "the consideration and the introduction of HIT applications," whereas adoption is defined as "the acceptance and incorporation of HIT applications into everyday practice."

An initial search was performed to identify whether any consolidated reviews, such as scoping reviews, were already conducted to understand the barriers to and facilitators of AI adoption in health care. During this search, it was found that a majority of the literature seemed to report on a specific area, such as radiology, in a specific setting (hospital or community), and a number of studies were reporting on implementation findings and not necessarily adoption. A few literature reviews on the determinants of and barriers to AI adoption have been conducted, such as the review by Radhakrishnan and Chattopadhyay [26]. However, these reviews span across multiple industries. For health care, 1 systematic review on the *barriers* to AI adoption in health care has been conducted by Assadullah [27]. However, there are no consolidated reviews that consider both the *barriers to and facilitators* of AI adoption in health care at large. Therefore, this review has attempted to explore the latter to provide considerations for health care organizations looking to successfully implement AI technologies via increased adoption.

Methods

Overview

This review was guided by the methodology and reporting structure outlined for scoping reviews by the Joanna Briggs Institute as well as Arksey and O'Malley [28]. The stages defined by Arksey and O'Malley [28] were followed to conduct this scoping review: (1) identifying the research question; (2) identifying relevant studies; (3) selecting studies for inclusion; (4) charting the data; and (5) collating, summarizing, and reporting the results

Stage 1: Identifying the Research Question

Because adoption is a key element of successful cost-benefit realization of technological investments, a general question was formed using the "population, concept, and context" approach [29]. The first component, "population," included users or potential users of the AI system, such as patients, providers, health care leaders, researchers, and those who were involved with implementing AI systems in various settings. The second component, "concept," consisted of barriers to and facilitators of any AI technology. The third component, "context," was centered on barriers to and facilitators of any AI technology in *any health care setting*, leaving this as broad as possible to maintain the paradigms of a true scoping review. A generic question developed was as follows: What are the barriers to and facilitators of AI adoption in health care?

Stages 2 and 3: Identifying Relevant Studies and Study Selection

In commencing the research, eligibility criteria were defined (as described in the Eligibility section). Once the eligibility criteria were defined, the search strategy was identified, and a search for articles was conducted in the selected databases.

Eligibility (Inclusion and Exclusion Criteria)

All published studies and gray literature that reported implementation findings related to adoption or reported factors impacting adoption were considered in this review. Therefore, studies with various designs, including quantitative and qualitative studies, literature reviews, thought articles, conference papers, and reports, were included in the initial search and review. "Health care organizations" were defined as organizations that are engaged in providing care to patients or involved in some aspect of providing agency to health care players. Health care players were defined as anyone involved in the process of providing or receiving care, including policy makers; administrative professionals; clinicians; physicians; and, most importantly, patients and their families. All types of AI technologies were considered in this review. Articles were not excluded based on variations in settings (hospital vs community setting) or countries where the research was conducted. Only articles in English were included. Due to the speed at which the landscape for AI is advancing, only articles that were published between January 2011 and December 2023 (when the search was conducted) were included.

Search Strategy

This review is intended to synthesize findings from publications that reported on the barriers to and facilitators of the adoption of AI implementations. A search was conducted on MEDLINE, ScienceDirect, and IEEE Xplore in December 2023. Keywords were selected in reference to the question identified to formulate the scope. Keywords included "artificial intelligence"; "healthcare" or "health care"; "hospital," "health services," or "health facilities"; "adoption"; "barriers"; "obstacles"; "challenges"; "facilitators"; and "enablers" (Textbox 1).

Textbox 1. Search query.

Query

- "artificial intelligence" AND healthcare or health care or hospital or health services or health facilities AND adoption AND barriers or obstacles or challenges or facilitators or enablers
- "artificial intelligence" AND health AND adoption AND (Barrier OR Facilitator)

Study Selection

A total of 2 reviewers independently screened the articles from the initial search by reviewing their titles and abstracts. Articles meeting the inclusion criteria were identified. Articles that did not meet the inclusion criteria were excluded. Any discrepancies were resolved through discussion.

Of the articles identified, the full text of the semifinal set of articles was reviewed to further refine selected articles. This process was iterative, and some exclusions were made during the writing phase, as the findings evolved. An Excel spreadsheet (Microsoft Corp) was used to record the articles identified. Recordings included the following details: the name of the article, authors, journal, whether the article was peer reviewed, type of paper, discipline, country, region, method, population, end users, and type of AI application (if specified). Duplicate studies were identified and removed to ensure there was no overlap.

Stages 4 and 5: Charting the Data and Collating, Summarizing, and Reporting the Results

A conventional content analysis approach was used to review the articles, chart the data, and identify themes [30]. Publications meeting the inclusion criteria were reviewed in detail, and an inductive approach was used to identify themes. First, the articles were read in full for the author to immerse into the content. This was followed by carefully reading each article and highlighting key concepts around barriers and facilitators that appeared to repeat across all the articles. These initial key concepts were recorded as themes, and this process helped identify many themes that were further categorized and grouped based on similarity. All data were charted in an Excel table to help with the analysis.

Ethical Considerations

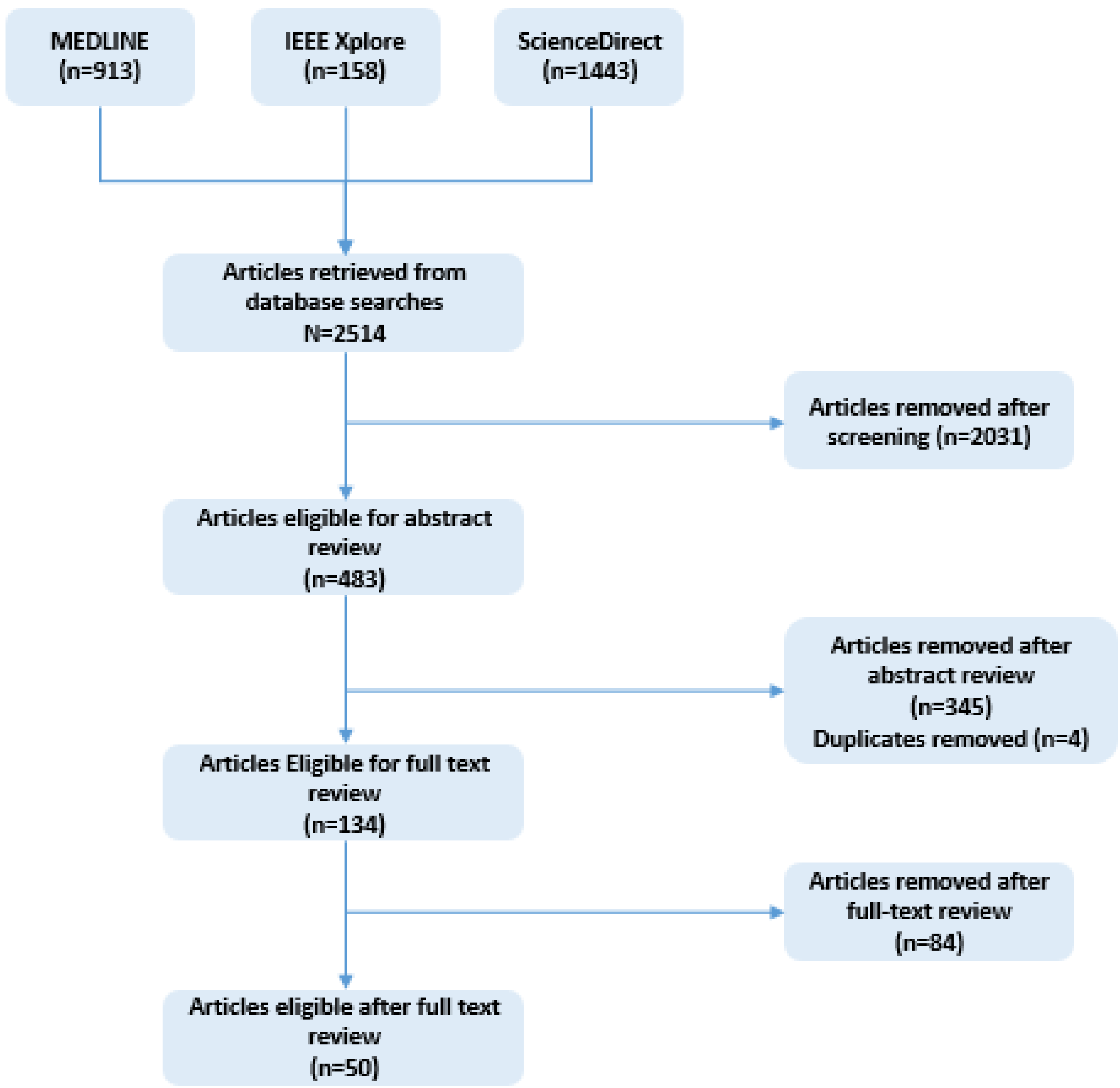
This work received an ethics exemption from the University of Victoria ethics board due to the nature of research being a literature review.

Results

Overview

The initial search from MEDLINE, IEEE Xplore, and ScienceDirect provided cumulative results of 2514 publications. After screening the results, 483 (19.21%) publications were included for abstract review based on the title of the study. After abstract review, 134 (27.7%) publications were identified for further text review, further excluding 345 (71.4%) publications, including 4 (0.8%) duplicate articles. Out of the 134 studies, 50 (37.3%) went through a thorough and more detailed review and thematic analysis. Figure 1 presents the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. The breakdown of studies is based on the country of origin, types of articles selected, and health care discipline or area covered. Overall, 11 articles were from the United States, 7 from China, 5 each from the United Kingdom and Canada, and 3 each from Germany and the Netherlands; the remainder of the articles were from Australia, France, Germany, India, Indonesia and Taiwan, Italy, New Zealand, Saudi Arabia, Singapore, Sweden, Switzerland, and other European countries. In some cases, multiple countries or regions collaborated to publish the articles together, including different European countries or the United Kingdom and United States. A total of 13% of the articles were literature reviews and 8% were mixed methods studies. The rest of the articles were cross-sectional studies, ethnographic or qualitative studies, case studies, white papers, and thought articles. In terms of setting, the majority of the articles discussed AI in health care in general with a majority of the articles reporting from the field of radiology or oncology. The setting of the remainder of the articles were academic hospitals, ophthalmology clinics, hospital, primary care, and dermatology clinics.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram showing the study selection process.



Thematic Analysis

Overview

On the basis of the conventional content analysis approach used (as described in the Methods section), a total of 18 categories of barriers and facilitators were identified (Figure 2). Interestingly, the themes were found to provide perspective on both facilitators and barriers. For example, if the theme explainability was identified as a barrier, the same theme was tabulated as a facilitator to capture what the articles recommended for overcoming challenges with explainability to increase adoption. As such, the reporting of the results for each theme provided perspective on the theme being both a barrier and facilitator, with the exception of governance, which was entirely noted as a key facilitator.

Figure 2. Themes identified. AI: artificial intelligence.

Transparency and Explainability

Explainability can be defined as the ability to deconstruct an algorithm to understand the mechanism by which it arrived at the output. Explainability has gained prominence due to the fast-paced growth of ML algorithms such as DL. These algorithms are labeled “black box” due to the difficulty in interpreting and tracing the techniques used by the AI models, thereby impacting trust and demanding the need for transparency [31,32]. According to Holzinger et al [33], “explainable AI deals with the implementation of transparency and traceability of statistical black-box machine learning methods, particularly deep learning.” While transparency has a wider definition and explainability is a component of transparency [34], most studies have noted explainability in the context of algorithmic transparency; therefore, findings from these 2 interrelated concepts have been discussed together.

Several studies noted a lack of explainability as a barrier to adoption. Baxter et al [35] reported concerns from adopters around the lack of explainability regarding the prediction of the AI algorithm embedded in the EHR to predict unplanned readmission; specifically, the lack of explainability regarding what features of the algorithm were driving the output was an impediment to trust among adopters. Other studies noted that the way the data were being used to train algorithms was not clear. The lack of traceability and logical understanding of how the algorithm arrived at a recommendation contradicted a key foundation of evidence-based medicine, which relies on high standards of explainability. Clinicians expressed the need to understand both the scientific and clinical bases of the recommendations provided by the AI to confidently validate and apply the decision [19,20,36-38]. Morrison [36] particularly quoted stakeholders seeking clarity on the extent to which there is a need to provide transparency on AI output to patients and how this is directed by legislation or data protection laws. In a study conducted by Nadarzynski et al [39], users of an AI-enabled chatbot reported hesitancy to use the technology due to a lack of transparency on how the chatbot accurately arrives at responses to health inquiries.

According to Holzinger et al [33], “explainability is an important element for consideration in order to enhance trust of medical professionals.” To facilitate adoption, improving algorithmic transparency will be a key consideration to change attitudes and build the trust of adopters [35,40]. Additional recommendations around facilitating adoption were to have processes in place to support clinicians in case of disagreements on decisions due to a lack of transparency and explainability [41]. Furthermore, revealing the process of how the algorithm was developed, who was involved in the development process, whether clinicians were consulted, and how the data were processed would enable acceptability [42,43].

In health care, causality is especially important when using automated decision systems; therefore, Holzinger et al [33] emphasized that AI systems should support the understanding and explanation of the causal models as opposed to simply solving through pattern recognition. Similarly, Gillner [44] also noted that this opacity of the output is not aligned with the “medical ethos.” Weinert et al [45] recommended that investing into explainable AI that produces a transparent and understandable AI could help address the issue of acceptability. Moorman [38] reported that successful adoption was achieved by publishing evidence on the algorithm’s underpinnings and providing clinicians with details on how data elements interacted within the algorithm to produce the predictive output.

Algorithm Bias, Equity, or Fairness

A prominent theme that came through was around the prevention of algorithm bias to ensure equity and fairness and avoid concealed discrimination. Algorithmic bias has been defined by Panch et al [46] as “the instances when the application of an algorithm compounds existing inequities in socioeconomic status, race, ethnic background, religion, gender, disability or sexual orientation to amplify them and adversely impact inequities in health systems.” Such biases have been visibly found in glomerular filtration rate and pulmonary function and have continued to persist despite efforts to address them. Inaccurate and underrepresentative training data sets for AI models can cause bias, misleading predictions, adverse events, and large-scale discrimination, causing barriers to adoption [32,41,47,48]. Baxter et al [35] and Chua et al [49] reported clinical stakeholders' concerns around the relevance of the AI model's output, especially because the algorithm did not consider social determinants of health to predict risk outcomes for readmissions. Similar concerns were raised by participants from other studies around the risk of algorithm bias as a challenge for adoption. Others expressed dissatisfaction that the AI algorithm may not be representative of the patient population among whom it is implemented or may have been trained with a biased training data set that has been retrofitted to produce certain results, therefore not providing a representative outcome of interest [36,49-51]. AI not accounting for patients' health determinants was noted as a "grand challenge" [19].

Inadequate data from representative groups, algorithms designed to represent a majority, and missing variables that impact predictions are components that contribute to bias [21]. This can be addressed by engaging clinicians in the design and development of the algorithm to ensure that appropriate measures are taken to address bias before the AI algorithm is deployed.

Functionality (Accuracy and Usefulness)

One of the major themes that came through was around the value, usefulness, and accuracy of an AI algorithm. Accuracy and quality of the AI algorithm's output were primary reasons for adoption hesitancy in the context of the functionality of the AI. In some studies, patients reported the need to assess the usefulness of the AI before using it and had concerns around the quality and accuracy of the output, thereby questioning the value of AI as a whole [39,42,52-55]. Clinicians also reported concerns around the usefulness of the output based on the lack of accuracy and inactionable output contributing to a low likelihood of use. This also included dissatisfaction if recommendations were too similar, inappropriate, or not useful [35,37,38,40,56]. Ease of use; complex interfaces; and inconsistent performance, for example, due to false positives and negatives add burden to the workflow, creating more work for clinicians, thereby impeding adoption [42,52,57-59]. Morrison [36] identified a lack of an agreed standard and benchmark for accuracy (how accurate does an AI tool need to be before it is approved for clinical practice) as an impediment to implementation and, subsequently, adoption, as, if a standard existed, it could provide rationale for the accuracy. Temsah et al [60] recommend the application of evidence-based oversight mechanisms to ensure accuracy and dependability. Finally, Choudhury [61] noted that if an algorithm is not performing up to standards or adds more work to the clinicians, this can impact adoption, as clinicians perceive it as a high risk.

Perceived benefit, perceived usefulness, usability, ease of use, usefulness, accuracy, and reliability of the output of the AI are key contributors to adoption [19,45,62-67]. In particular, usability and acceptability should be assessed with the intended user in mind [41]. Perceived benefit is especially important when it contributes to improved efficiency in clinical processes [62].

From a patient's perspective, there is value if an AI technology can be used from home for minor consultations, such as skin cancer detection using an AI-enabled app [42]. Ease of use of the technology and accessibility to information for minor health concerns [39] are especially seen as valuable, as they negate the need for a visit to the physician; however, in the event that a visit is required based on the AI's recommendation, then the integration of the technology with the health care system is considered beneficial [42]. It is essential that as the AI system matures, it is designed such that it can "adopt and challenge contradictory rules and behaviours" [43].

Risk of Harm

Patient safety concerns causing adverse effects were noted by Mlodzinski et al [48] and Vijayakumar et al [59]. The lack of accuracy of AI output was also considered to pose a potential risk of harm by both clinicians and patients, as, in some cases where the algorithm may output false negative results, it may provide an incorrect sense of reassurance and cause a delay in treatment. However, in cases where the algorithm is too sensitive, thereby providing false positive results, it may add work and costs to the treatment process [39,42,57,61,68]. The risk of harm can be lowered if AI algorithms are developed with the 5 rights (in the case of an automated decision system), similar to other clinical decision support tools [40]. In addition, Sangers et al [42] proposed that in some cases, AI applications should provide only risk indication instead of a diagnosis to reduce the risk of harm. Chen et al [68] reported that policies and mechanisms to safeguard professionals could address challenges associated with a potential risk of harm due to a lack of output accuracy.

Trust

Lack of accuracy; doubts about unsafe results; privacy breaches; patients' perceptions and acceptance of automation; and uncertainty about developers' reliability, availability, usability, and perceived usefulness were found to be obstacles to gaining trust [39,42,47,50,53,54,69,70]. Clinicians expressed fear around having to reframe their professional identity and responsibilities [57]. Fear around what AI really meant was noted as another barrier [36,42].

Facilitators of trust included the endorsement of the technology by experts as well as academically backed clinicians, including regulating bodies such as the government; evidence of output accuracy based on the evaluation of AI; and positive opinions from trusted thought leaders in the respective clinical fields [42,58,60]. Another facilitator of trust is the engagement of patients in the development of AI. This could facilitate trust in the public and address concerns around trust in data sharing [71]. Overall, trust was largely found to be associated with perceived usefulness; however, one study noted that "if peoples' confidence and beliefs are improved, they will use the product despite its usefulness" [52]. In addition, Fan et al [72] reported that initial trust is a key indicator of the intention to use the AI application and noted that if the confidence to use the AI application is high based on performance expectancy, then this will increase trust in using the AI.

Human-AI Teaming

The lack of human intervention was found to be a barrier to adoption for both clinicians and patients. From a clinician standpoint, physicians expressed that they would be less likely to use an AI, given their familiarity with the patient's condition and the value of intuition in clinical decision-making [32,35]. Mlodzinski et al [48] reported concerns about potential systemic bias present in the AI that could impact the patient-provider relationship. From a patient standpoint, the lack of human presence was seen as a limitation due to a lack of empathy and emotional connectivity with another human or simply not having the presence of a human physician to verbally communicate and discuss, such as when using an AI app or chatbot [51,64]. By contrast, the lack of human presence was, in some cases, seen as a benefit due to the anonymity in sharing intimate or uncomfortable health concerns [39,42]. Hemphill et al [73] reported increased confidence in patients when AI was combined with clinical interpretation.

Aligning Strategic Components

Several studies particularly highlighted the importance of strategic alignment with initiatives. Baxter et al [35] reported on how a lack of alignment among different organizational initiatives led to varying outcomes and disjointed communication. Strohm et al [57], on the other hand, talked about how strategic alignment could lead to better dispersing of funds across different departments.

Sun and Medaglia [50] pointed out the necessity of outlining a comprehensive "top-down strategy" that would include organizations' goals and resource distribution for AI implementation. Weinert et al [45] elaborated that to overcome the barrier to including AI initiatives in the organizational strategy, the German government introduced a new law that supported organizations with financial assistance to implement innovative digital technologies such as AI, as there was hesitancy among organizations to include expensive AI implementations as part of their strategy due to a lack of funds.

Use Case–Driven or Problem-Driven AI

Several studies noted that to start the journey of implementing AI, there is a need to identify a problem and not merely use data to come up with a solution. Therefore, use cases should be identified based on notable problems that can be addressed by AI solutions. One particular study mentioned how the lack of a use case affected the implementation of the AI model [35,36].

End-User Engagement or Co-Design

A lack of sufficient buy-in from end users and a lack of endorsement from organizational leadership emerged as barriers, including not engaging stakeholders early in the process. It is critical to incorporate clinicians and other stakeholders, such as patients, in the development life cycle, especially the testing phase with the application of a user-centered design and testing approaches. This may be time intensive but proves to be an effective approach to enable adoption [39,40]. As Ongena et al [69] pointed out in their findings, patients should be involved when developing AI systems, specifically for diagnostic, treatment planning, and prognostic purposes. Pou-Prom et al [74] reported the usefulness of engaging end users in designing, deploying, and refining the AI solution. Moorman [38] recommended maximizing buy-in and engagement at all levels of stakeholders, from leadership to users, and especially engaging a clinician leader from the onset. Finally, Goldstein et al [75] noted the inclusion of champions at the leadership and clinical levels to achieve successful implementation.

Workflow Integration

This review found that the lack of integration of the AI system into the workflow can be a barrier to successful implementation and adoption. For example, Baxter et al [35] reported impacts on success due to variations in existing workflows for risk assessments and readmission scores across different areas. Similarly, Strohm et al [57] mentioned how the lack of integration and standardization of workflows led to variations in workflows. For other types of AI solutions, such as apps, the data not being integrated into the health care system or workflow was seen as a barrier for patients to adopt the solution [42,53]. Helenason et al [58] and Schepart et al [56] both noted the importance of conformity with the workflow when integrating AI tools into the environment where they would be used.

Recommendations included the following: ensuring that AI applications easily integrate with existing IT systems, integrating data from patients' use of AI-enabled apps into the health care system and workflow, and considering the integration of the AI into the clinical workflow but maintaining autonomy for clinicians to have the final say [19,20,38,42,49,57,75]. In situations where the AI system is deployed in different areas, using a common model may improve alignment in workflows [35,36]. Chen et al [68] reported that clinicians saw AI integration into the workflow as a positive if the system was seen as potentially eliminating routine work, allowing them to focus on other tasks. Moorman [38] reported that it was helpful to assess existing unit workflows, communications, escalation, and event management processes before implementation to address challenges brought up by clinicians concerned about added work.

Awareness and Training

Training refers to educating the users on various aspects of the technology, such as the outcomes of the AI model, its benefits, and how it supports the clinical workflow, and providing new skills such as technical and data science skills to staff, especially laggards and champions, to assist with the adoption [36,41,57,62]. Kelly et al [21] particularly noted that to provide clinicians with clarity on how an algorithm could improve patient care, approaches such as using a decision curve analysis that would provide quantified benefits of using a model to inform actions that need to be taken would be helpful. Skepticism and a lack of understanding were seen as barriers to AI adoption [48,51]. Chen et al [68] found that AI adoption increased among radiologists who were more familiar with AI. Sun [76] recommended that implementation teams should consider influencing clinicians by sharing AI knowledge via more informal communications, such as social media communication or in-person communication. Moorman [38] found it helpful to develop educational material with input from clinicians to tailor it to the clinical role and hospital culture. Training and awareness should include building an understanding of the technology; providing clarity around language such as the definition of AI; education around data use in health care; and building awareness on the value of AI, including breaking down concepts that dispel fear

[36,37,39,42,50,52,71]. Clinicians' awareness and knowledge of AI before using it contributed to its successful acceptance [59,73]. In addition, users feel that the technology is "unqualified" based on their perception of the premature nature of the technology [39]. Misunderstanding of the capabilities of AI technologies in the general public and the health care sector is a challenge to adoption, with gaps in awareness around the value, advantages, and high expectations of AI [37,50]. Overall, Alsobhi et al [70] emphasized the urgency of accelerating AI adoption through the dissemination of AI training.

Resources and Infrastructure

Shortages of personnel with the required skills were reported as barriers, along with the quality of IT infrastructure available for AI implementation [32,35,36,40,45,50]. Hickman et al [71], in particular, noted the lack of technological (infrastructure) maturity to allow for the integration of AI. The presence of AI experts, perhaps a multidisciplinary team, particularly with clinical scientists, data science and subject matter experts with AI skills, an innovation manager, AI experts to provide training, and local champions within departments involved in the end-to-end process, was considered an important element for adoption [32,57,59,62,71]. Goldstein et al [75] noted that for scalability, where the AI application would be deployed in multiple sites, having designated resources from the onset of the project with clear roles and responsibilities was seen with success. Yang et al [77] recommended cultivating talent with both high-level medical and technology knowledge and understanding how the 2 domains can be used to meet patients' needs. In a different perspective on the shortage of resources, Chen et al [68] noted that radiographers and radiologists held more positive attitudes toward the adoption of AI, as it would help address workforce shortages in the radiology field in the United Kingdom.

Evaluation and Validation

The need for evaluation on multiple fronts was noted by various studies. Studies indicated that the technical evaluation of AI is necessary as a first step to validation. Technical evaluation must be followed by clinical validation (based on established methods in clinical research) and economic validation. Wolff et al [78] particularly noted the lack of clinical and economic measurements as a barrier to practical implementation. Evaluations should be tailored toward digital technologies to gather empirical evidence surrounding the value of AI's use [53,78]. However, the cost of conducting an empirical evaluation and a quantified clinical trial–type evaluation may be a deterrent to the pace of developing the technology. Therefore, this should be considered when selecting the type of evaluation to be conducted. A focus on assessing the effectiveness and accuracy was duly highlighted. Implementers should consider validating or testing the algorithm with synthetic data [40-42,57]. Establishing a standard methodology for the validation of AI algorithms and the overall evaluation of AI will be critical to gaining confidence from adopters [50]. Hickman et al [71] suggested that having structures in place for the continued monitoring of standards that impact AI (eg, regulatory standards) and ensuring that infrastructure is in place to evaluate and monitor algorithms continuously are necessary.

Data Security, Ownership, Quality, and Availability

Data quality, security, ownership, and storage were prominent themes in the reviewed studies. In terms of data quality and integrity, several issues were identified as barriers to developing good AI models that provide value to users. These were issues around variability, the nature of unstructured data, incompleteness of data, the data not representing the reality of clinical care, and the absence of data standards (specifically around how and what data are collected). Having metadata standards, terminologies, data quality metrics, and common data models were identified as facilitators in resolving some of these issues [40,43,45,50]. Fragmented access to data and limited sources, such as the availability of data only from EHRs or data silos, were also noted as barriers [78].

Data access, integrity, and provenance are key to the development of models. Institutions that were the most successful in implementing AI were thoughtful about how to guarantee data integrity [40]. Wolff et al [78] noted the challenges with data silos and fragmented access to medical data, including limitations on the availability of data only from the EHR, to enable the development of robust AI applications. In terms of data ownership, the dilemma around who owns the data, whether it would be the government, institution (eg, hospital), or patient, is a barrier to adoption, as it leaves questions around how data would be integrated or accessible for future AI advancement [37,50]. While this may not present a direct adoption challenge, it does indirectly impact the availability of data required for development and to produce meaningful outputs, which is an impediment to adoption, as identified earlier.

Data security was identified as a major contributor toward hesitancy, with concerns around cybersecurity relating to both training and testing data as well as fear of trackers and spyware obtaining unsolicited data [39,41,42,45,48,77]. Furthermore, the ability of deidentified data to be reidentified poses a major risk for the individuals and institutions providing their data. This further contributes to resistance to sharing the data that could help expand data sets for AI training. Several strategies for preventing security breaches have been proposed, and these could be helpful in securing data, especially health data that are accessible over the web [20,42].

Concerns around data processing include the lack of understanding of how data are stored, processed, and accessed; the establishment of protocols; and compliance with existing privacy policies, such as the General Data Protection Regulation and the Health Insurance Portability and Accountability Act [20,36]. A survey of patients conducted by Ongena et al [69] indicated the need for patients to be informed about how and specifically which data are processed. One study explicitly highlighted the issue of data ownership. Concerns over data ownership were particularly evident when patients linked data ownership to trust in the technology [47].

Ethics and Privacy

Concerns about privacy and ethics were focused on maintaining confidentiality, ensuring processes are in place to obtain consent, and having informed consent with clarity on how the data will be processed [32,39,47,56]. Having clear consent processes related to how data are generated through the use of AI and how these data flow as well as defining the meaning of consent and transparency on strategies to maintain privacy are seen as facilitators of adoption. This is especially applicable to "clinical and epidemiological use cases of ML in both decision support and automation categories, as data from patients or the public are essential to train algorithms in these areas" [20]. In addition, Sun and Medaglia [50] particularly pointed out the unethical use of data, such as data being used by commercial organizations. Ensuring transparency to end users, especially patients involved in the ethical and legal frameworks that guide the development of AI systems, could be helpful [69,73]. Weinert et al [45] particularly identified ethical issues, specifically as they relate to liability, as a barrier to AI adoption. Wolff et al [78] noted that integrating "privacy-by-design" technologies into AI applications that incorporate advanced data protection features could mitigate such challenges.

Governance

Governance was primarily noted as a key facilitating factor, playing the role of enabling the full cycle of AI. It is critical to have a governance structure in place to oversee the development and rollout of AI from conception to implementation, with governance tools providing guidance on various stages of the process. Governance should include diverse professionals with clear articulation of accountability, including nuances in reactions to accountability [35,40,55,58,64]. Isbanner et al [64] noted the importance of articulating accountability. This is especially important in health care because "ethical and governance challenges matter to the public." Wolff et al [78] recommended outlining specific responsibilities for different stakeholders to delineate accountability-based steps in the process, for example, identifying who would review an x-ray image analysis and identifying liability and culpability (eg, obligatory human check of a decision obtained by an AI application). According to Sunarti et al [47], the governance body should include "developers of software, government officials, health care, medical practitioners and advocacy for patients groups." The lack of accountability in the decision-making process is a challenge; therefore, framing this in the governance model could be a way to address adoption issues related to accountability [37,50,73]. Other functions of the governance body would be to ensure funding and connectivity to the wider data science community, ensure alignment with strategic initiatives in the organization, and act as a long-term centralized knowledge repository for performance oversight. A governance model should have mechanisms and systems in place to facilitate changes impacting AI technologies in development or use based on cyclical changes in the technology or changes in the external landscape, such as the ones initiated by regulatory bodies [32,41,79]. Formalized analysis of ethical considerations in the development and use of AI should be a key component of governance. Governance should also be linked to the data governance committees for various data processing, quality, and integrity oversights [43]. One particular solution proposed by Morrison [36] was to have national-level governance templates that would facilitate national data protection via the implementation of impact assessments. In contrast, governance can hinder data sharing. Therefore, governance bodies should maintain a rigorous process without becoming a constraint [36].

Regulatory and Legal Frameworks

A lack of regulation and policies from the government, including uncertainty around legal direction or law, was presented as a barrier to the application of AI technologies [37,45,57,60,77,78]. Other researchers noted that there was no clarity in the area of regulatory structures with regard to which regulatory body should be consulted for AI developments and deployments [36,53]. Therefore, it would be essential for governments to establish regulatory bodies and legal frameworks to provide guidance on various aspects of AI development and application [20,41]. In addition, ambiguity surrounding malpractice liability policy as it relates to physicians' legal responsibilities, for example, in case of diagnostic errors, remains a barrier to AI adoption [49].

Funding and Cost

The lack of and uncertainty surrounding funding are presented as barriers to implementation [32,51,56]. Researchers have suggested that there is a need to have costs identified from the start-up stage all the way to scalability. Funding can especially be a barrier if an AI technology is lacking in evidence, with little to show for the value it provides. Sun and Medaglia [50] and Xing et al [53] reported that financial barriers in the context of cost and benefits, the lack of a sustainable business model, and insufficient funding to meet public demands should also be considered. Sun and Medaglia [50] additionally noted challenges associated with the adoption of IBM Watson in China due to patients having to pay high fees for the service. Finally, funding should be cohesively considered not only for the development of the technology but also for resources required to implement the technology, such as technical subject matter experts, project managers, and champions [36,40,41,51,57]. Weinert et al [45] noted that to overcome the barrier of lack of resources and meet the financial investment demands of AI implementations, the German government introduced a new law that could help organizations bridge funding gaps; however, they could not conclude whether this would facilitate any progress, as the announcement was just made.

Discussion

Principal Findings

The principal findings of this study imply several factors impacting the adoption of AI systems, and for each barrier identified, there are corresponding facilitators. Ethics, bias, and transparency or explainability are core considerations in developing trustworthy and adoption-centric AI systems. Furthermore, the barriers identified should be holistically synergized within a governance framework, one that ideally oversees the entire end-to-end process, from ideation to the implementation and sustainability of AI systems.

Trust emerged as one of the most critical elements of AI adoption. This review revealed that trust can either be facilitated or impacted by almost all the themes identified in this scoping review. More specifically, fairness, explainability, and ethics seem to be the centerfold of barriers to AI adoption. Therefore, our discussions have focused on these 3 domains with recommendations on how organizations can address these domains to facilitate adoption.

Transparency and Explainability

Our findings revealed that explainability in the context of algorithmic transparency is a significant barrier to adoption. Various studies noted that limitations due to the opacity of an algorithm may inhibit clinicians from relying on ML outputs in clinical settings. This leads to ambiguity on whether the ML output can be trusted enough for the clinician to move forward with the clinical decision-making or should be overridden due to a lack of certitude or misalignment with traditional clinical judgment [80]. AI explainability (XAI) is an entire field dedicated to ensuring trustworthy and explainable AI. There are numerous publications from this field. For example, Markus et al [81] noted that explanations are crucial to involving a human in the process of verifying the decision of the algorithm, for example, by revealing what features were used in training the AI algorithm. Adadi and Berrada [82] have conducted a comprehensive review of existing evidence on explainability approaches and organized them from different perspectives. They specifically outline 4 guidelines for the need for explainable AI. Explain to justify: the decisions made by using an underlying model should be explained to increase their justifiability. Explain to control: explanations should enhance the transparency of a model and its functioning, allowing its debugging and the identification of potential flaws. Explain to improve: explanations should help improve the accuracy and efficiency of their models. Explain to discover: explanations should support the extraction of novel knowledge and the learning of relationships and patterns to manage social interaction and create a shared meaning of the decision-making process.

There are several techniques that organizations can adopt when aiming to achieve explainable AI. These include explainable modeling, evaluating for explainability, or following an explainability framework, as proposed by Markus et al [81]. Preece [83] and Vilone and Longo [31] have done a thorough analysis of evaluation approaches for explainable AI. The inclusion of the combination of these techniques from XAI could be useful for organizations to address adoption barriers associated with explainability. It is prudent that organizations developing and implementing AI incorporate various explainable modeling approaches, include explainability frameworks, and consider explainability evaluation in their AI life cycle. In addition, part of this process should include equipping clinicians with knowledge about what the AI takes as input, how the input is processed, and what the AI produces as output, along with the training process used. In this way, clinician engagement is essential to the process of developing and validating AI algorithms and outputs. This approach will also empower clinicians to discuss these transparencies with patients, thereby contributing toward building trust on all fronts.

Bias

In terms of equity and fairness, our findings have demonstrated that algorithm bias is a critical factor in not only gaining trust but also having meaningful outputs that can be applied to diverse patients. Specific concerns related to adoption include models being trained on data not representative of the patient population or not containing diverse data as related to social determinants of health [19,35,36]. Bias in AI systems can be introduced due to biased data, algorithms being trained on the biased data, limitations in the model itself, small training size, lack of user participation, and other unseen factors [84]. There are a number of examples of specific issues related to bias in AI systems. Buolamwini and Gebru [85] reported that an artificial vision algorithm was unable to identify dark-skinned individuals, as >80% of the individuals in a reference data set were light-skinned individuals. Another failed case is found in the field of anesthesiology, where data from 40 institutions revealed that Black patients received inferior care (with respect to postoperative nausea and vomiting prophylaxis) at nearly every single center [86]. Seyyed-Kalantari et al [87] noted that convolutional neural networks will frequently underdiagnose Hispanic patients at a disproportionate rate due to the potential lack of access to health care and insurance type. In the field of mental health, specifically concerning schizophrenia, a meta-analysis found that risk-flagging models trained on European populations have reduced performance in East Asian populations [88].

According to Panch et al [46], several challenges need to be addressed when addressing algorithmic bias. They include a lack of clear definitions and standards of “fairness,” insufficient contextual specificity, and the “black-box” nature of algorithms. These can be addressed by developing algorithms based on where they will be deployed by first establishing and identifying these contexts and ensuring processes are in place to address these challenges. There are numerous solutions to address bias that emphasize the risk of bias mitigation techniques to be applied at each stage of model development. For example, Chen et al [89] and O’Reilly-Shah [90] recommend that at the preprocessing stage, where the data may have internalized biases, techniques such as reweighing data samples of marginalized groups or resampling based on the population that algorithm output would be applied to. These techniques could help address the adoption barriers identified in this review, particularly those around underrepresentative and inaccurate training data sets. Similarly, at the postprocessing stage, similar thresholds for different representative groups could be set for the model to be monitored, adjusted, and trained. The design of the algorithm should consider equity via training data sets that are either diverse or focused to fit the localized population. Another major concern is data set drift, which means that the data set the model was trained with is different from the test data set. There are various techniques to mitigate data set shift, and these techniques should be considered in model development. Another mitigation technique, as recommended by Chen et al [89], is federated learning, where a model is trained on a global server. This technique allows for models to be trained on large data sets without sharing sensitive information. Aside from more quantitative techniques to address the risk of bias, there are assessments available that can be used as a checklist during each stage of AI development. While these tools can address statistical bias, it is much more challenging to identify social bias that can intrude into the data. Frameworks such as the one developed by Landers and Behrend [91] comprehensively outline questions that would be asked at each stage of AI development. These questions focus on information, perceptions, and other social and cultural components. Such tools, when integrated into the AI development process, would help gather evidence that could be shared with clinicians and patients on how bias mitigation has been considered in the end-to-end development process, thereby addressing adoption concerns around bias identified in this review.

Ethics

Gerke et al [92] have discussed four primary ethical challenges that need to be addressed to realize the full potential of AI in health care: (1) informed consent to use, (2) safety and transparency, (3) algorithmic fairness and biases, and (4) data privacy. These challenges resonate with our findings around barriers to adoption and, interestingly, tie in these elements of barriers to adoption under the ethics domain.

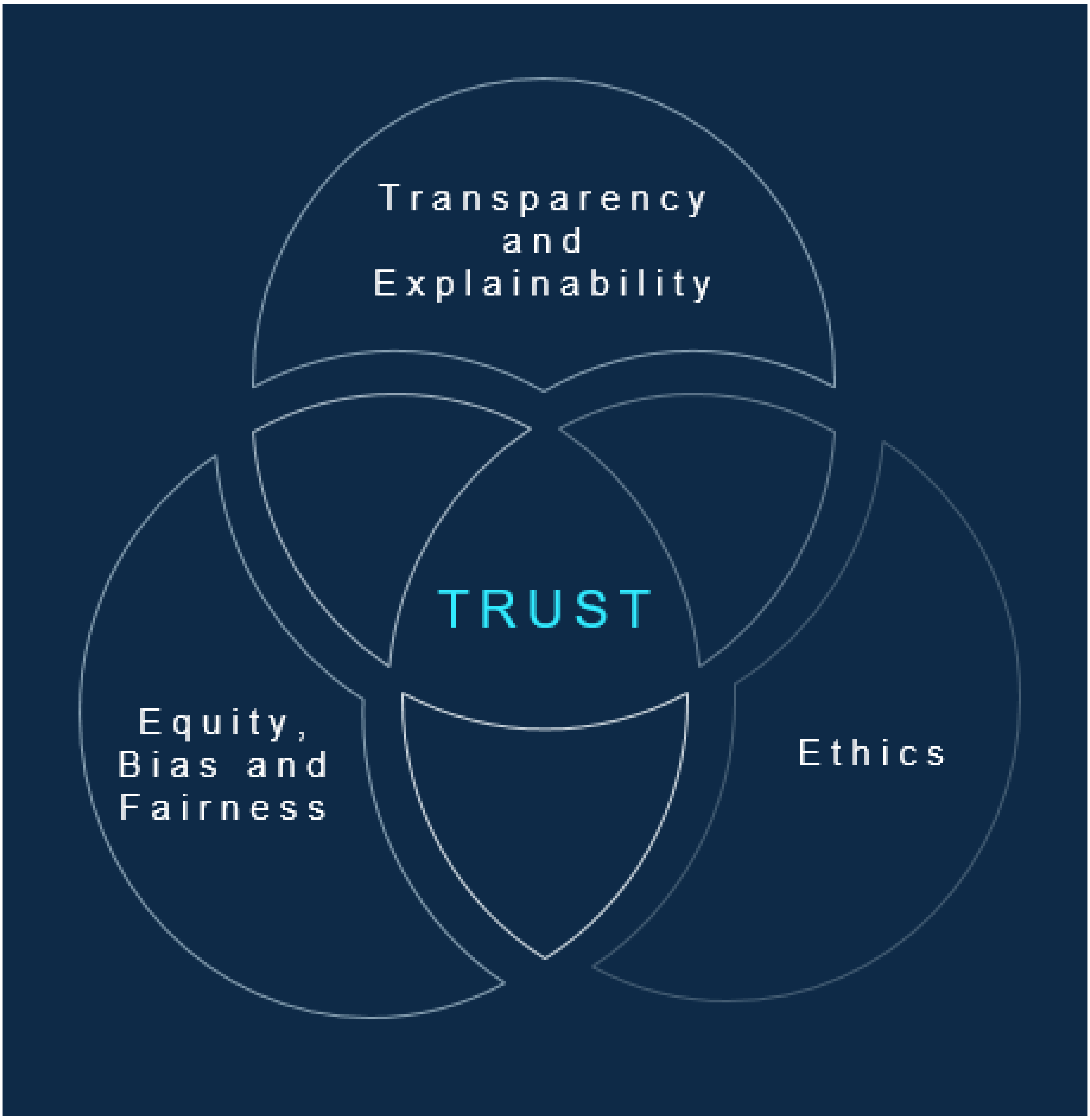
There are several cases of ethical concerns that highlight the need for ethics. In the context of ethical concerns around informed consent and data privacy, in 2017, the personal data of approximately 1.6 million patients were provided to Google DeepMind by Royal Free National Health Service Foundation Trust without the patients' consent. The data were to be used to test a new way of detecting kidney injuries [93]. From a clinician's perspective, there are concerns around what the clinician's responsibility is in informing patients about the use of AI in their care [92]. In the context of algorithmic safety and bias, Buolamwini and Gebru [85] and Liao [94] note ethical concerns around algorithms not detecting dark-skinned individuals for skin cancer detection due to the fact that the algorithm was trained on light-skinned individuals. Similarly, an algorithm that is widely used in US hospitals to identify patients who need additional care was found to use the cost expenditure by patients as a means to identify those who need extra care. This was discriminatory toward Black patients, as they generally spend less than White patients on health care, resulting in false conclusions [95].

In terms of safety and transparency, Liao [94] provides a good explanation as to why the lack of safety and transparency is an issue. They provide an example of a prediction of a 70% chance for a supposed patient's tumor to become malignant in 5 years; however, the algorithm does not necessarily provide detailed reasons as to how it arrived at the conclusion. From an ethical standpoint, this is an issue because humans need to know how a decision is reached; specifically, in health care, not being able to understand and trust a decision is problematic.

Which and, more importantly, how can organizations address ethical concerns? According to Liao [94], there are >80 ethical frameworks that have been proposed for AI. Many of these draw on the 4 principles of biomedical ethics, namely, autonomy, beneficence, nonmaleficence, and justice. Some of these frameworks provide practical checklists that organizations can use to conduct an ethics deliberation. For example, Solanki et al [96] developed a comprehensive framework for AI developers that includes ethics oversight during different phases of the AI development life cycle. Similarly, Rogers et al [97] shared a very practical approach to how they evaluated an AI model for ethics. Such tools are practical methods of assessing AI algorithms for ethical principles. Despite these practical approaches, Goirand et al [98] note that operationalizing ethical frameworks for AI is challenging, as there is a need for contextualization due to different ethical issues present in different environments. Therefore, organizations have to consider these nuances and determine an ethics approach that would work best for their organization when evaluating each AI model.

These frameworks and tools to develop trustworthy AI by addressing various barriers to adoption are also just beginning to emerge and be applied in real-life cases; however, they are a good start to the implementation journey of AI, especially those applied in clinical settings. Overall, our findings demonstrate that the adoption of an AI system has to be considered from its onset, when the system is being conceptualized, to when it is implemented and sustained. The existing technology implementation and acceptance models may not be all encompassing of adoption factors; therefore, adding additional frameworks around trust, bias, explainability, and ethics will be necessary to foresee the success of an AI innovation. A governance model may address concerns around risk, safety, and adoption barriers identified in this paper by facilitating the overall development process of AI and ensuring various checks and balances are in place. Figure 3 is a visual depiction of the core elements that were found to impact trust, as discussed in this section.

Figure 3. Adoption barriers, as related to trust.



Limitations

Given the limited application of AI in health care at the time of this research, only a few number of papers that reported on implementation barriers and facilitators were reviewed to identify AI adoption barriers and facilitators. As the application of AI and types of AI systems in health care grows, a follow-up on adoption barriers and facilitators to assess for additional barriers and facilitators suitable to future environments may be necessary.

Comparison With Prior Work

At the time this search was conducted, a few literature reviews on the determinants of and barriers to AI adoption were conducted, such as the review by Radhakrishnan and Chattopadhyay [26]. However, these studies span across multiple industries. For health care, one systematic review on the *barriers* to AI adoption in health care was conducted by Assadullah [27]. However, there is less work that considers both the *barriers to and facilitators* of AI adoption in health care at large. Therefore, this review has attempted to explore the latter to provide considerations for health care organizations looking to successfully implement AI technologies via increased adoption. Our findings are validated due to the replication of several themes identified in similar, previous research by Assadullah [27]. Common themes identified around barriers to adoption included explainability, trust issues centered on privacy, challenges around data ownership, lack of regulatory standards, issue of bias, and lack of accountability. Overall, the issue of trust was found to be centered on bias, ethics, and explainability, which led to a lack of accountability and an inability to evaluate. Other issues impeding trust included impacts on model performance leading to inaccurate results. These findings around trust resonate with results from this research, reinforcing the barriers to adoption identified in both studies.

Conclusions

This literature review revealed that trust is impacted by a number of elements identified as barriers and that trust is a significant catalyst of adoption. A governance structure can be a key facilitator in ensuring all the elements identified as barriers are addressed appropriately. The findings demonstrate that the implementation of AI in health care is still in many ways dependent on the establishment of regulatory and legal frameworks. Further research around the combination of governance and implementation frameworks, models, or theories to enhance trust that would specifically enable adoption is needed to provide the necessary guidance to those translating AI research into practice. Future research could also be expanded to include attempts at understanding patients' perspectives on complex, high-risk AI use cases and how the use of AI applications affect clinical practice and patient care, including sociotechnical considerations, as more algorithms are implemented in actual clinical environments.

Acknowledgments

EB received a Michael Smith Health Research BC Health Professional Investigator Award (grant HPI-2018-2045). This project received no specific grant from any funding agency in the public, commercial, or not-for-profit sector.

Conflicts of Interest

AK is editor in chief of *JMIR Human Factors*, but was not involved in the processing of this article. EB is the editor in chief of *JMIR Nursing*, but was not involved in the processing of this article.

Multimedia Appendix 1
PRISMA checklist.
[PDF File (Adobe PDF File), 166 KB - humanfactors_v11i1e48633_app1.pdf]

References

1. 2022 global health care outlook. Deloitte. 2024 May 27. URL: https://www.deloitte.com/global/en/Industries/life-sciences-health-care/perspectives/global-health-care-sector-outlook.html [accessed 2024-05-28]
2. Key strategic priorities. CIFAR. URL: https://cifar.ca/ai/key-strategic-priorities/#topskipToContent [accessed 2022-02-15]
3. Hamet P, Tremblay J. Artificial intelligence in medicine. Metabolism 2017 Apr;69S:S36-S40. [doi: 10.1016/j.metabol.2017.01.011] [Medline: 28126242]
4. The age of AI: what exactly is AI? Deloitte Malta. URL: https://www2.deloitte.com/mt/en/pages/rpa-and-ai/articles/mt-age-of-ai-2-what-is-it.html [accessed 2024-04-29]
5. Tekkeşin A. Artificial intelligence in healthcare: past, present and future. Anatol J Cardiol 2019 Oct;22(Suppl 2):8-9 [FREE Full text] [doi: 10.14744/AnatolJCardiol.2019.28661] [Medline: 31670713]
6. Yang X, Chen A, PourNejatian N, Shin HC, Smith KE, Parisien C, et al. A large language model for electronic health records. NPJ Digit Med 2022 Dec 26;5(1):194 [FREE Full text] [doi: 10.1038/s41746-022-00742-2] [Medline: 36572766]
7. Secinaro S, Calandra D, Secinaro A, Muthurangu V, Biancone P. The role of artificial intelligence in healthcare: a structured literature review. BMC Med Inform Decis Mak 2021 Apr 10;21(1):125 [FREE Full text] [doi: 10.1186/s12911-021-01488-9] [Medline: 33836752]
8. Tran BX, Vu GT, Ha GH, Vuong Q, Ho M, Vuong T, et al. Global evolution of research in artificial intelligence in health and medicine: a bibliometric study. J Clin Med 2019 Mar 14;8(3):360 [FREE Full text] [doi: 10.3390/jcm8030360] [Medline: 30875745]
9. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019 Jun 13;6(2):94-98 [FREE Full text] [doi: 10.7861/futurehosp.6-2-94] [Medline: 31363513]
10. Liu N, Zhang Z, Wah Ho AF, Ong ME. Artificial intelligence in emergency medicine. J Emerg Crit Care Med 2018 Oct;2:82. [doi: 10.21037/jeccm.2018.10.08]
11. Kirubarajan A, Taher A, Khan S, Masood S. Artificial intelligence in emergency medicine: a scoping review. J Am Coll Emerg Physicians Open 2020 Dec 07;1(6):1691-1702 [FREE Full text] [doi: 10.1002/emp2.12277] [Medline: 33392578]
12. Varner C. Can artificial intelligence predict emergency department crowding? Healthy Debate. URL: https://healthydebate.ca/2020/03/topic/predicting-emergency-department-crowding-artificial-intelligence-mar2020/ [accessed 2024-04-29]
13. Cuttler M. Transforming health care: how artificial intelligence is reshaping the medical landscape. CBC News. 2019. URL: https://www.cbc.ca/news/health/artificial-intelligence-health-care-1.5110892 [accessed 2024-04-29]
14. Briganti G, Le Moine O. Artificial intelligence in medicine: today and tomorrow. Front Med (Lausanne) 2020 Feb 5;7:27 [FREE Full text] [doi: 10.3389/fmed.2020.00027] [Medline: 32118012]
15. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng 2018 Oct 10;2(10):719-731. [doi: 10.1038/s41551-018-0305-z] [Medline: 31015651]

16. Chang AC. Big data in medicine: the upcoming artificial intelligence. Prog Pediatr Cardiol 2016 Dec;43:91-94. [doi: 10.1016/j.ppedcard.2016.08.021]
17. The healthcare data explosion. RBC Capital Markets. URL: https://www.rbccm.com/en/gib/healthcare/episode/the_healthcare_data_explosion [accessed 2024-04-29]
18. Reinsel D, Gantz J, Rydning J. The digitization of the world: from edge to core. Seagate. 2018. URL: https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf [accessed 2024-04-29]
19. Romero-Brufau S, Wyatt KD, Boyum P, Mickelson M, Moore M, Cognetta-Rieke C. A lesson in implementation: a pre-post study of providers' experience with artificial intelligence-based clinical decision support. Int J Med Inform 2020 May;137:104072. [doi: 10.1016/j.ijmedinf.2019.104072] [Medline: 32200295]
20. Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial intelligence and the implementation challenge. J Med Internet Res 2019 Jul 10;21(7):e13659 [FREE Full text] [doi: 10.2196/13659] [Medline: 31293245]
21. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med 2019 Oct 29;17(1):195 [FREE Full text] [doi: 10.1186/s12916-019-1426-2] [Medline: 31665002]
22. Leviss J. HIT or Miss – studying failures to enable success. Appl Clin Inform 2017 Dec 16;02(03):345-349. [doi: 10.4338/aci-2011-03-ie-0020]
23. Venkatesh N, Morris N, Davis N, Davis N. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]
24. Greenhalgh T, Abimbola S. The NASSS framework - a synthesis of multiple theories of technology implementation. Stud Health Technol Inform 2019 Jul 30;263:193-204. [doi: 10.3233/SHTI190123] [Medline: 31411163]
25. Cresswell K, Sheikh A. Organizational issues in the implementation and adoption of health information technology innovations: an interpretative review. Int J Med Inform 2013 May;82(5):e73-e86. [doi: 10.1016/j.ijmedinf.2012.10.007] [Medline: 23146626]
26. Radhakrishnan J, Chattopadhyay M. Determinants and barriers of artificial intelligence adoption – a literature review. In: Proceedings of the 2020 International Conference on Transfer and Diffusion of IT, TDIT. 2020 Presented at: TDIT '20; December 18-19, 2020; Tiruchirappalli, India p. 89-99 URL: https://link.springer.com/chapter/10.1007/978-3-030-64849-7_9 [doi: 10.1007/978-3-030-64849-7_9]
27. Assadullah MM. Barriers to artificial intelligence adoption in healthcare management: a systematic review. SSRN Journal. Preprint posted online March 4, 2020 [FREE Full text] [doi: 10.2139/ssrn.3530598]
28. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
29. Hosseini M, Jahanshahlou F, Akbarzadeh MA, Zarei M, Vaez-Gharamaleki Y. Formulating research questions for evidence-based studies. J Med Surg Public Health 2024 Apr;2:100046. [doi: 10.1016/j.glmedi.2023.100046]
30. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
31. Vilone G, Longo L. Notions of explainability and evaluation approaches for explainable artificial intelligence. Inf Fusion 2021 Dec;76:89-106. [doi: 10.1016/j.inffus.2021.05.009]
32. Brennan HL, Kirby SD. Barriers of artificial intelligence implementation in the diagnosis of obstructive sleep apnea. J Otolaryngol Head Neck Surg 2022 Apr 25;51(1):16 [FREE Full text] [doi: 10.1186/s40463-022-00566-w] [Medline: 35468865]
33. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip Rev Data Min Knowl Discov 2019 Apr 02;9(4):e1312 [FREE Full text] [doi: 10.1002/widm.1312] [Medline: 32089788]
34. Larsson S, Heintz F. Transparency in artificial intelligence. Internet Policy Rev 2020;9(2). [doi: 10.14763/2020.2.1469]
35. Baxter SL, Bass JS, Sitapati AM. Barriers to implementing an artificial intelligence model for unplanned readmissions. ACI open 2020 Jul 19;4(2):e108-e113 [FREE Full text] [doi: 10.1055/s-0040-1716748] [Medline: 33274314]
36. Morrison K. Artificial intelligence and the NHS: a qualitative exploration of the factors influencing adoption. Future Healthc J 2021 Nov 02;8(3):e648-e654 [FREE Full text] [doi: 10.7861/fhj.2020-0258] [Medline: 34888459]
37. Laï MC, Brian M, Mamzer M. Perceptions of artificial intelligence in healthcare: findings from a qualitative survey study among actors in France. J Transl Med 2020 Jan 09;18(1):14 [FREE Full text] [doi: 10.1186/s12967-019-02204-y] [Medline: 31918710]
38. Moorman LP. Principles for real-world implementation of bedside predictive analytics monitoring. Appl Clin Inform 2021 Aug 22;12(4):888-896 [FREE Full text] [doi: 10.1055/s-0041-1735183] [Medline: 34553360]
39. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: a mixed-methods study. Digit Health 2019;5:2055207619871808 [FREE Full text] [doi: 10.1177/2055207619871808] [Medline: 31467682]
40. Watson J, Hutyra CA, Clancy SM, Chandiramani A, Bedoya A, Ilangovan K, et al. Overcoming barriers to the adoption and implementation of predictive modeling and machine learning in clinical care: what can we learn from US academic medical centers? JAMIA Open 2020 Jul;3(2):167-172 [FREE Full text] [doi: 10.1093/jamiaopen/ooz046] [Medline: 32734155]

41. Li JP, Liu H, Ting DS, Jeon S, Chan RP, Kim JE, et al. Digital technology, tele-medicine and artificial intelligence in ophthalmology: a global perspective. Prog Retin Eye Res 2021 May;82:100900 [FREE Full text] [doi: 10.1016/j.preteyeres.2020.100900] [Medline: 32898686]
42. Sangers TE, Wakkee M, Kramer-Noels EC, Nijsten T, Lugtenberg M. Views on mobile health apps for skin cancer screening in the general population: an in-depth qualitative exploration of perceived barriers and facilitators. Br J Dermatol 2021 Nov 05;185(5):961-969 [FREE Full text] [doi: 10.1111/bjd.20441] [Medline: 33959945]
43. Liyanage H, Liaw ST, Jonnagaddala J, Schreiber R, Kuziemsky C, Terry AL, et al. Artificial intelligence in primary health care: perceptions, issues, and challenges. Yearb Med Inform 2019 Aug 25;28(1):41-46 [FREE Full text] [doi: 10.1055/s-0039-1677901] [Medline: 31022751]
44. Gillner S. We're implementing AI now, so why not ask us what to do? - how AI providers perceive and navigate the spread of diagnostic AI in complex healthcare systems. Soc Sci Med 2024 Jan;340:116442 [FREE Full text] [doi: 10.1016/j.socscimed.2023.116442] [Medline: 38029666]
45. Weinert L, Müller J, Svensson L, Heinze O. Perspective of information technology decision makers on factors influencing adoption and implementation of artificial intelligence technologies in 40 German hospitals: descriptive analysis. JMIR Med Inform 2022 Jun 15;10(6):e34678 [FREE Full text] [doi: 10.2196/34678] [Medline: 35704378]
46. Panch T, Mattie H, Atun R. Artificial intelligence and algorithmic bias: implications for health systems. J Glob Health 2019 Dec;9(2):010318 [FREE Full text] [doi: 10.7189/jogh.09.020318] [Medline: 31788229]
47. Sunarti S, Fadzlul Rahman F, Naufal M, Risky M, Febriyanto K, Masnina R. Artificial intelligence in healthcare: opportunities and risk for future. Gac Sanit 2021;35 Suppl 1:S67-S70 [FREE Full text] [doi: 10.1016/j.gaceta.2020.12.019] [Medline: 33832631]
48. Mlodzinski E, Wardi G, Viglione C, Nemati S, Crotty Alexander L, Malhotra A. Assessing barriers to implementation of machine learning and artificial intelligence-based tools in critical care: web-based survey study. JMIR Perioper Med 2023 Jan 27;6:e41056 [FREE Full text] [doi: 10.2196/41056] [Medline: 36705960]
49. Chua IS, Gaziel-Yablowitz M, Korach ZT, Kehl KL, Levitan NA, Arriaga YE, et al. Artificial intelligence in oncology: path to implementation. Cancer Med 2021 Jun 07;10(12):4138-4149 [FREE Full text] [doi: 10.1002/cam4.3935] [Medline: 33960708]
50. Sun TQ, Medaglia R. Mapping the challenges of artificial intelligence in the public sector: evidence from public healthcare. Gov Inf Q 2019 Apr;36(2):368-383. [doi: 10.1016/j.giq.2018.09.008]
51. Zhang M, Scandiffio J, Younus S, Jeyakumar T, Karsan I, Charow R, et al. The adoption of AI in mental health care-perspectives from mental health professionals: qualitative descriptive study. JMIR Form Res 2023 Dec 07;7:e47847 [FREE Full text] [doi: 10.2196/47847] [Medline: 38060307]
52. Ye T, Xue J, He M, Gu J, Lin H, Xu B, et al. Psychosocial factors affecting artificial intelligence adoption in health care in China: cross-sectional study. J Med Internet Res 2019 Oct 17;21(10):e14316 [FREE Full text] [doi: 10.2196/14316] [Medline: 31625950]
53. Xing F, Peng G, Zhang B, Li S, Liang X. Socio-technical barriers affecting large-scale deployment of AI-enabled wearable medical devices among the ageing population in China. Technol Forecast Soc Change 2021 May;166:120609. [doi: 10.1016/j.techfore.2021.120609]
54. Petitgand C, Motulsky A, Denis JL, Régis C. Investigating the barriers to physician adoption of an artificial intelligence-based decision support system in emergency care: an interpretative qualitative study. Stud Health Technol Inform 2020 Jun 16;270:1001-1005. [doi: 10.3233/SHTI200312] [Medline: 32570532]
55. Wang B, Asan O, Mansouri M. Perspectives of patients with chronic diseases on future acceptance of AI-based home care systems: cross-sectional web-based survey study. JMIR Hum Factors 2023 Nov 06;10:e49788 [FREE Full text] [doi: 10.2196/49788] [Medline: 37930780]
56. Schepart A, Burton A, Durkin L, Fuller A, Charap E, Bhambri R, et al. Artificial intelligence-enabled tools in cardiovascular medicine: a survey of current use, perceptions, and challenges. Cardiovasc Digit Health J 2023 Jun;4(3):101-110 [FREE Full text] [doi: 10.1016/j.cvdhj.2023.04.003] [Medline: 37351333]
57. Strohm L, Hehakaya C, Ranschaert ER, Boon WP, Moors EH. Implementation of artificial intelligence (AI) applications in radiology: hindering and facilitating factors. Eur Radiol 2020 Oct 26;30(10):5525-5532 [FREE Full text] [doi: 10.1007/s00330-020-06946-y] [Medline: 32458173]
58. Helenason J, Ekström C, Falk M, Papachristou P. Exploring the feasibility of an artificial intelligence based clinical decision support system for cutaneous melanoma detection in primary care - a mixed method study. Scand J Prim Health Care 2024 Mar 20;42(1):51-60 [FREE Full text] [doi: 10.1080/02813432.2023.2283190] [Medline: 37982736]
59. Vijayakumar S, Lee VV, Leong QY, Hong SJ, Blasiak A, Ho D. Physicians' perspectives on ai in clinical decision support systems: interview study of the CURATE. AI personalized dose optimization platform. JMIR Hum Factors 2023 Oct 30;10:e48476 [FREE Full text] [doi: 10.2196/48476] [Medline: 37902825]
60. Temsah M, Aljamaan F, Malki KH, Alhasan K, Altamimi I, Aljarbou R, et al. ChatGPT and the future of digital health: a study on healthcare workers' perceptions and expectations. Healthcare (Basel) 2023 Jun 21;11(13):1812 [FREE Full text] [doi: 10.3390/healthcare11131812] [Medline: 37444647]

61. Choudhury A. Factors influencing clinicians' willingness to use an AI-based clinical decision support system. Front Digit Health 2022 Aug 16;4:920662 [FREE Full text] [doi: 10.3389/fdgth.2022.920662] [Medline: 36339516]
62. Victor Mugabe K. Barriers and facilitators to the adoption of artificial intelligence in radiation oncology: a New Zealand study. Tech Innov Patient Support Radiat Oncol 2021 Jun;18:16-21 [FREE Full text] [doi: 10.1016/j.tipsro.2021.03.004] [Medline: 33981867]
63. Goswami M, Jain S, Alam T, Deifalla AF, Ragab AE, Khargotra R. Exploring the antecedents of AI adoption for effective HRM practices in the Indian pharmaceutical sector. Front Pharmacol 2023 Nov 14;14:1215706 [FREE Full text] [doi: 10.3389/fphar.2023.1215706] [Medline: 38034991]
64. Isbanner S, O'Shaughnessy P, Steel D, Wilcock S, Carter S. The adoption of artificial intelligence in health care and social services in Australia: findings from a methodologically innovative national survey of values and attitudes (the AVA-AI study). J Med Internet Res 2022 Aug 22;24(8):e37611 [FREE Full text] [doi: 10.2196/37611] [Medline: 35994331]
65. Mahlknecht A, Engl A, Piccoliori G, Wiedermann CJ. Supporting primary care through symptom checking artificial intelligence: a study of patient and physician attitudes in Italian general practice. BMC Prim Care 2023 Sep 04;24(1):174 [FREE Full text] [doi: 10.1186/s12875-023-02143-0] [Medline: 37661285]
66. Kleine AK, Kokje E, Lermer E, Gaube S. Attitudes toward the adoption of 2 artificial intelligence-enabled mental health tools among prospective psychotherapists: cross-sectional study. JMIR Hum Factors 2023 Jul 12;10:e46859 [FREE Full text] [doi: 10.2196/46859] [Medline: 37436801]
67. Choudhury A, Shamszare H. Investigating the impact of user trust on the adoption and use of ChatGPT: survey analysis. J Med Internet Res 2023 Jun 14;25:e47184 [FREE Full text] [doi: 10.2196/47184] [Medline: 37314848]
68. Chen Y, Stavropoulou C, Narasinkan R, Baker A, Scarbrough H. Professionals' responses to the introduction of AI innovations in radiology and their implications for future adoption: a qualitative study. BMC Health Serv Res 2021 Aug 14;21(1):813 [FREE Full text] [doi: 10.1186/s12913-021-06861-y] [Medline: 34389014]
69. Ongena YP, Haan M, Yakar D, Kwee TC. Patients' views on the implementation of artificial intelligence in radiology: development and validation of a standardized questionnaire. Eur Radiol 2020 Feb 08;30(2):1033-1040 [FREE Full text] [doi: 10.1007/s00330-019-06486-0] [Medline: 31705254]
70. Alsobhi M, Sachdev HS, Chevidikunnan MF, Basuodan R, Dhanesh Kumar KU, Khan F. Facilitators and barriers of artificial intelligence applications in rehabilitation: a mixed-method approach. Int J Environ Res Public Health 2022 Nov 29;19(23):15919 [FREE Full text] [doi: 10.3390/ijerph192315919] [Medline: 36497993]
71. Hickman SE, Baxter GC, Gilbert FJ. Adoption of artificial intelligence in breast imaging: evaluation, ethical constraints and limitations. Br J Cancer 2021 Jul 26;125(1):15-22 [FREE Full text] [doi: 10.1038/s41416-021-01333-w] [Medline: 33772149]
72. Fan W, Liu J, Zhu S, Pardalos PM. Investigating the impacting factors for the healthcare professionals to adopt artificial intelligence-based medical diagnosis support system (AIMDSS). Ann Oper Res 2018 Mar 19;294(1-2):567-592. [doi: 10.1007/s10479-018-2818-y]
73. Hemphill S, Jackson K, Bradley S, Bhartia B. The implementation of artificial intelligence in radiology: a narrative review of patient perspectives. Future Healthc J 2023 Mar;10(1):63-68 [FREE Full text] [doi: 10.7861/fhj.2022-0097] [Medline: 37786489]
74. Pou-Prom C, Murray J, Kuzulugil S, Mamdani M, Verma AA. From compute to care: lessons learned from deploying an early warning system into clinical practice. Front Digit Health 2022 Sep 5;4:932123 [FREE Full text] [doi: 10.3389/fdgth.2022.932123] [Medline: 36133802]
75. Goldstein J, Weitzman D, Lemerond M, Jones A. Determinants for scalable adoption of autonomous AI in the detection of diabetic eye disease in diverse practice types: key best practices learned through collection of real-world data. Front Digit Health 2023 May 18;5:1004130 [FREE Full text] [doi: 10.3389/fdgth.2023.1004130] [Medline: 37274764]
76. Sun TQ. Adopting artificial intelligence in public healthcare: the effect of social power and learning algorithms. Int J Environ Res Public Health 2021 Dec 01;18(23):12682 [FREE Full text] [doi: 10.3390/ijerph182312682] [Medline: 34886404]
77. Yang J, Luo B, Zhao C, Zhang H. Artificial intelligence healthcare service resources adoption by medical institutions based on TOE framework. Digit Health 2022 Oct 05;8:20552076221126034 [FREE Full text] [doi: 10.1177/20552076221126034] [Medline: 36211801]
78. Wolff J, Pauling J, Keck A, Baumbach J. Success factors of artificial intelligence implementation in healthcare. Front Digit Health 2021 Jun 16;3:594971 [FREE Full text] [doi: 10.3389/fdgth.2021.594971] [Medline: 34713083]
79. Farič N, Hinder S, Williams R, Ramaesh R, Bernabeu MO, van Beek E, et al. Early experiences of integrating an artificial intelligence-based diagnostic decision support system into radiology settings: a qualitative study. J Am Med Inform Assoc 2023 Dec 22;31(1):24-34 [FREE Full text] [doi: 10.1093/jamia/ocad191] [Medline: 37748456]
80. Lang M, Bernier A, Knoppers BM. Artificial intelligence in cardiovascular imaging: "unexplainable" legal and ethical challenges? Can J Cardiol 2022 Feb;38(2):225-233. [doi: 10.1016/j.cjca.2021.10.009] [Medline: 34737036]
81. Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies. J Biomed Inform 2021 Jan;113:103655 [FREE Full text] [doi: 10.1016/j.jbi.2020.103655] [Medline: 33309898]

82. Adadi A, Berrada M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access 2018;6:52138-52160. [doi: 10.1109/access.2018.2870052]
83. Preece A. Asking 'why' in AI: explainability of intelligent systems – perspectives and challenges. Intell Sys Acc Fin Mgmt 2018 Apr 19;25(2):63-72. [doi: 10.1002/isaf.1422]
84. Kumar A, Aelgani V, Vohra R, Gupta SK, Bhagawati M, Paul S, et al. Artificial intelligence bias in medical system designs: a systematic review. Multimed Tools Appl 2023 Jul 22;83(6):18005-18057. [doi: 10.1007/s11042-023-16029-x]
85. Buolamwini J, Gebru T. Gender shades: intersectional accuracy disparities in commercial gender classification. In: Proceedings of the 2018 International Conference on Machine Learning and Machine Intelligence. 2018 Presented at: MLMI '18; September 28-30, 2018; Ha Noi, VietNam p. 1-15 URL: https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf
86. White RS, Andreae MH, Lui B, Ma X, Tangel VE, Turnbull ZA, Multicenter Perioperative Outcomes Group Collaborators. Antiemetic administration and its association with race: a retrospective cohort study. Anesthesiology 2023 Jun 01;138(6):587-601 [FREE Full text] [doi: 10.1097/ALN.0000000000004549] [Medline: 37158649]
87. Seyyed-Kalantari L, Liu G, McDermott M, Chen IY, Ghassemi M. CheXclusion: fairness gaps in deep chest X-ray classifiers. In: Proceedings of the 2021 Pacific Symposium. 2021 Presented at: Biocomputing '21; January 3-7, 2021; Kohala, HI p. 232-243 URL: https://www.worldscientific.com/doi/abs/10.1142/9789811232701_0022 [doi: 10.1142/9789811232701_0022]
88. Lam M, Chen CY, Li Z, Martin AR, Bryois J, Ma X, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Indonesia Schizophrenia Consortium, Genetic REsearch on schizophreniA neTwork-Chinathe Netherlands (GREAT-CN), et al. Comparative genetic architectures of schizophrenia in East Asian and European populations. Nat Genet 2019 Dec 18;51(12):1670-1678 [FREE Full text] [doi: 10.1038/s41588-019-0512-x] [Medline: 31740837]
89. Chen RJ, Wang JJ, Williamson DF, Chen TY, Lipkova J, Lu MY, et al. Algorithmic fairness in artificial intelligence for medicine and healthcare. Nat Biomed Eng 2023 Jun 28;7(6):719-742 [FREE Full text] [doi: 10.1038/s41551-023-01056-8] [Medline: 37380750]
90. O'Reilly-Shah VN, Gentry KR, Walters AM, Zivot J, Anderson CT, Tighe PJ. Bias and ethical considerations in machine learning and the automation of perioperative risk assessment. Br J Anaesth 2020 Dec;125(6):843-846 [FREE Full text] [doi: 10.1016/j.bja.2020.07.040] [Medline: 32838979]
91. Landers RN, Behrend TS. Auditing the AI auditors: a framework for evaluating fairness and bias in high stakes AI predictive models. Am Psychol 2023 Jan;78(1):36-49. [doi: 10.1037/amp0000972] [Medline: 35157476]
92. Gerke S, Minssen T, Cohen IG. Ethical and legal challenges of artificial intelligence-driven health care. SSRN Journal. Preprint posted online May 26, 2020 [FREE Full text] [doi: 10.2139/ssrn.3570129]
93. Google DeepMind NHS app test broke UK privacy law. BBC News. 2017. URL: https://www.bbc.com/news/technology-40483202 [accessed 2024-04-29]
94. Liao SM. Ethics of AI and health care: towards a substantive human rights framework. Topoi 2023 Apr 12;42(3):857-866. [doi: 10.1007/s11245-023-09911-8]
95. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019 Oct 25;366(6464):447-453. [doi: 10.1126/science.aax2342] [Medline: 31649194]
96. Solanki P, Grundy J, Hussain W. Operationalising ethics in artificial intelligence for healthcare: a framework for AI developers. AI Ethics 2022 Jul 19;3(1):223-240. [doi: 10.1007/s43681-022-00195-z]
97. Rogers WA, Draper H, Carter SM. Evaluation of artificial intelligence clinical applications: detailed case analyses show value of healthcare ethics approach in identifying patient care issues. Bioethics 2021 Sep 28;35(7):623-633. [doi: 10.1111/bioe.12885] [Medline: 34046918]
98. Goirand M, Austin E, Clay-Williams R. Implementing ethics in healthcare AI-based applications: a scoping review. Sci Eng Ethics 2021 Sep 03;27(5):61. [doi: 10.1007/s11948-021-00336-3] [Medline: 34480239]

Abbreviations

AI: artificial intelligence
DL: deep learning
EHR: electronic health record
EMR: electronic medical record
ML: machine learning
NLP: natural language processing
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by C Jacob; submitted 01.05.23; peer-reviewed by J McMurray, I Soares-Pinto, Y Zhuang, M Wright, L Weinert; comments to author 03.06.23; revised version received 28.02.24; accepted 12.06.24; published 29.08.24.

Please cite as:
Hassan M, Kushniruk A, Borycki E
Barriers to and Facilitators of Artificial Intelligence Adoption in Health Care: Scoping Review
JMIR Hum Factors 2024;11:e48633
URL: https://humanfactors.jmir.org/2024/1/e48633
doi:10.2196/48633
PMID:

Patient Preferences for Direct-to-Consumer Telemedicine Services: Replication and Extension of a Nationwide Survey

Julia Ivanova[1], PhD; Hattie Wilczewski[1], BSc; Farina Klocksieben[2], MPH; Mollie Cummins[1,3], PhD; Hiral Soni[1], PhD; Triton Ong[1], PhD; Janelle Barrera[1,4], MPH, CHES, CPH; Jillian Harvey[5], PhD; Nathaniel O'Connell[6], PhD; James McElligott[7], MD; Brandon Welch[1,8], PhD; Brian Bunnell[1,4], PhD

[1]Doxy.me Research, Doxy.me, Inc., 3445 Winton Pl, #114, Rochester, NY, United States

[2]Research Methodology and Biostatistics Core, Morsani College of Medicine, University of South Florida, Tampa, FL, United States

[3]Department of Biomedical Informatics, College of Nursing and Spencer Fox Eccles School of Medicine, University of Utah, Salt Lake City, UT, United States

[4]Department of Psychiatry and Behavioral Neurosciences, Morsani College of Medicine, University of South Florida, Tampa, FL, United States

[5]Healthcare Leadership and Management, College of Health Professions, Medical University of South Carolina, Charleston, SC, United States

[6]Department of Biostatistics and Data Science, Wake Forest School of Medicine, Winston-Salem, NC, United States

[7]Department of Pediatrics, Medical University of South Carolina, Charleston, SC, United States

[8]Biomedical Informatics Center, Medical University of South Carolina, Charleston, SC, United States

Corresponding Author:
Julia Ivanova, PhD
Doxy.me Research, Doxy.me, Inc., , 3445 Winton Pl, #114, Rochester, NY, , United States

Abstract

Background: A 2017 survey of patient perspectives showed overall willingness and comfort to use telemedicine, but low actual use. Given recent growth and widespread exposure of patients to telemedicine, patient preferences are likely to have changed.

Objective: This study aimed to (1) identify demographic trends in patient preferences and experiences; (2) measure ease of use and satisfaction of telemedicine; and (3) measure changes in telemedicine use, willingness, and comfort since 2017.

Methods: We replicated a 2017 study with a nationwide survey of US adults. The survey, an extended version of the previous study, measured patient health care access as well as knowledge, experiences, and preferences regarding telemedicine encounters. We recruited participants using SurveyMonkey Audience in July 2022. We used descriptive statistics and generalized estimating equations to measure change and identify trends.

Results: We accrued 4577 complete responses. Patient experience with telemedicine was substantially higher in 2022 than in 2017, with 61.1% (vs 5.3%) of participants aware that their primary care provider offered telemedicine and 34.5% (vs 3.5%) reporting use of telemedicine with their primary care provider. This study also reported ease of use and satisfaction rates to be similar to in-person visits, while overall willingness and comfort in using telemedicine increased from 2017. Individuals at the poverty line were significantly less likely to report satisfaction with telemedicine visits. We found increased interpersonal distance in a patient and health care professional relationship significantly reduced patient ease of use, willingness, and comfort in using telemedicine.

Conclusions: This study identified an association between income and patient satisfaction, conveying the importance of understanding telemedicine in relation to health care access and equity. Telemedicine ease of use and satisfaction were comparable to in-person visits. Individuals reported greater use and higher positive perceptions of telemedicine willingness and comfort since 2017.

(JMIR Hum Factors 2024;11:e51056) doi:10.2196/51056

KEYWORDS

telemedicine; survey; patient preferences; direct-to-consumer telemedicine; patient-provider relationship; inequity; consumer; patient experience; willingness; income; association; satisfaction; mobile phone

Introduction

Telemedicine has allowed for the growth of new technologies and platforms for asynchronous and synchronous health care. Most patients, though, imagine direct-to-consumer, synchronous visits when considering “telemedicine” [1]. While lower costs and higher ease of access had long been recognized in the biomedical literature [2,3], telemedicine did not come into widespread use until the 2020 pandemic caused by

SARS-CoV-2. Public health recommendations to use telemedicine greatly expanded its use and created opportunities for expanded research measuring stakeholder satisfaction, as well as barriers and facilitators [4,5]. Given the profound increase in use, it is necessary to understand whether patient preferences, experiences, and needs related to telemedicine have changed. A current understanding of the patient experience can help us to effectively develop and implement telemedicine programs within health care, especially with the push toward hybrid care [6-8].

According to the 2021 National Center for Health Statistics survey, more than one-third (37%) of US adults had used telemedicine [9]. The same study showed demographic trends that may point to differences in preferences or even unequal access to telemedicine based on income, gender, and race [9]. Though studies generally show high satisfaction of telemedicine use among patients [10-12], specific measurement of patient comfort in using telemedicine and willingness to use telemedicine is difficult to find in literature, and most studies were conducted before widespread adoption of telemedicine in the year 2020 [13-15]. Our team, Welch et al [14], conducted a 2017 study measuring aspects of willingness and comfort in direct-to-consumer telemedicine. In a nationwide representative sample of over 4300 US adults, we found that only 5.3% of patients knew their primary care health care professional (PCP) offered telemedicine and 3.5% had ever used telemedicine for their PCP visits [14]. Overall, over 50% of patients were willing and comfortable to use telemedicine with their own PCP [14]. Health care professional scenario types appeared to affect patient willingness and comfort, though: for example, patients were mostly willing (51.9%) to see their own PCP using telemedicine but less so for a different PCP within the health organization (34.9%) or a different PCP from a different organization (18.6%). Similar results were seen regarding comfort with 53.7% of patients being comfortable in using telemedicine with their own PCP but only 18.6% with a different PCP from a different organization [14]. Additionally, over 56% of patients reported that having an established relationship with a health care professional before having a telemedicine visit is important [14]. While this survey study did not include direct measures regarding satisfaction and ease of use, it showed that patients in the United States had a positive disposition toward telemedicine, even with little direct telemedicine exposure. Inclusion of additional assessments within this survey, including measures of satisfaction, would have helped to determine a more holistic understanding of patient preferences and experiences in telemedicine.

Telemedicine satisfaction has been assessed in myriad ways, from brief questions to full-length, validated questionnaires [10,16-19]. Generally, literature shows telemedicine satisfaction can be reliably measured through survey research, comparing satisfaction of telemedicine and in-person visits [17,19]. This approach to understanding telemedicine preferences can reasonably be extended to determine satisfaction with telemedicine in a hybrid care environment. For example, while Welch et al [14] may have measured willingness and comfort in using telemedicine, the addition of specific patient satisfaction questions for both telemedicine and in-person visits would have provided an additional dimension in assessing patient preferences [17]. While previous studies have shown little difference in satisfaction between in-person and telemedicine visits for specific types of visits or specialties [12,20,21], consumer experience with telemedicine has grown enormously in recent years, and their perspectives may have changed [22,23].

One of the largest changes to telemedicine occurred within the realm of compliance and regulation in 2020. With the public health emergency providing the flexibility necessary to implement telemedicine in a widespread fashion [24], telemedicine was provided to the US population in ways previously impossible. Patients could visit with health care professionals across state lines, use modes of telemedicine that were not necessarily Health Insurance Portability and Accountability Act (HIPAA)–compliant, and have the cost of telemedicine visits covered by insurance or other payors [24]. Indeed, with the new flexibilities and regulatory changes, many long-standing barriers to telemedicine usage were placed in the spotlight for public discussion. For example, 2015 guidelines from the American College of Physicians promoted that a patient first meets their health care professional in person prior to any telemedicine visits [25]. Loosened regulations in 2020 placed these expectations under debate [26]. With new flexibilities due to the public health emergency, patients could still access telemedicine if they were unable or unwilling to first meet a health care professional in person [26]. Compliance and regulatory shifts have resulted in critical changes to pathways of telemedicine implementation. As a result, revisiting patient experiences, preferences, and satisfaction with telemedicine is critical in determining ideal implementation strategies of telemedicine. Additionally, understanding patient perceptions of comfort and satisfaction regarding telemedicine may help identify needed changes to national and state policies relating to successful telemedicine practice within the United States.

In prior work, the 2017 Welch et al [14] study, we assessed patients' willingness and comfort to use telemedicine. The results may be used as a baseline to determine how patient willingness and comfort to use direct-to-consumer telemedicine have changed since 2017, and after large-scale population exposure to telemedicine. Here, we conducted a survey of the US population, replicating the items used in the prior study, to understand how comfort and willingness to use telemedicine has changed. This study aimed to (1) identify potential demographic trends in US patient preferences and experiences in telemedicine visits; (2) measure the current state of patient ease of use and satisfaction with using telemedicine; and (3) measure change regarding telemedicine use, comfort, and willingness among patients since 2017. By assessing the current patient experience in finer resolution, we can better understand patient needs, preferences, and potential barriers to successful use of telemedicine.

Methods

Sample and Procedures

We conducted a cross-sectional survey of a national sample of adults from July 1 to July 2, 2022, using SurveyMonkey

Audience, an online market research platform [27]. To emulate the Welch et al [14] study, we chose to recruit a similar sample size: we recruited 4500 adults (aged 18 years or older) representative of the US adult population based on age, gender, income, and regionality. Based on sample size calculations, we would be able to report results with a 99.99% confidence level and 2.75% margin of error. The anonymous survey was estimated to take 10 minutes. After providing informed consent, participants answered questions about their PCP, previous use of telemedicine, willingness to use telemedicine, and comfort using telemedicine. Participants received a US $0.50 donation to a nonprofit of their choice in return for their participation. SurveyMonkey Audience uses fraud and bot detection to ensure the integrity of their survey results [27].

Ethical Considerations

Study procedures were approved as exempt by the Institutional Review Board at the University of South Florida (IRB#4255). Participants were provided with informed consent and were able to opt out of the survey if they chose.

Survey and Methods

See Multimedia Appendix 1 for a complete list of survey questions and response options. We replicated and extended a previous 2017 survey, developed and reviewed by health care professionals and researchers, to directly compare results [14]. The adapted 10-minute survey comprised 13 multiple-choice questions, which included 7 Likert-scale matrix questions. The outcome measures included participant perceptions of willingness to use telemedicine, comfort in using telemedicine, ease of meeting with health care professionals, and satisfaction in meeting with providers. SurveyMonkey Audience provided 5 multiple-choice screening questions to ensure a representative US sample by gender, age, income, device type, and region.

We retained multiple-choice items from the 2017 survey related to duration (Q1) and frequency (Q2) of visits to a PCP, and added items asking participants of their level of use of telemedicine (Q3) [14]. Next, we asked participants, who responded to having used telemedicine with their PCP, about the satisfaction (Q4), ease of meeting with their health care professionals (Q5), and whether they would be disappointed if telemedicine were no longer offered (Q6). As an extension of the original study, the questions about satisfaction and ease of meeting with providers in person and via telemedicine were anchored on a 5-point Likert scale from 1=very unsatisfied or very difficult to 5=very satisfied or very easy. Additionally, the disappointment question was anchored on a 4-point scale from 1=very disappointed to 5=not disappointed. We then asked participants whether they have used telemedicine with any other health care professionals (Q7) following with the same set of questions regarding satisfaction (Q8), ease (Q9), and disappointment (Q10). All participants were asked about their willingness (Q11) and comfort (Q12) of using telemedicine in different scenario types related to their level of relationship with a PCP (own PCP, different PCP from the same health care organization, and different PCP from different health care organization). Willingness and comfort questions were anchored on 5-point Likert scales similar to the 2017 study from 1=very unwilling or very uncomfortable to 5=very willing or very comfortable. Lastly, we asked participants of their level of agreement (Q13) regarding statements about the importance of (1) having telemedicine as an option, (2) switching to new health care professionals offering telemedicine, (3) having an established relationship with a telemedicine health care professional, and (4) one's health care professional having access to health records. These importance questions were anchored on a 5-point Likert scale from 1=strongly disagree to 5=strongly agree similar to the 2017 study [14].

Data Analysis

In order to directly compare the current survey results with the 2017 results, we replicated the analysis detailed in Welch et al [14], using SPSS (version 29; IBM Corp).

We measured each of the four outcomes by the self-reported, 5-point Likert scale; however, for the purposes of analysis, these variables were dichotomized in a similar fashion as seen in Welch et al [14] (eg, very willing, willing, and neutral grouped as willing; and unwilling and very unwilling grouped as unwilling). We reported all four outcome measures in relation to health professional scenario types, income, gender, and age. We determined a reference for each demographic or scenario type for the purposes of analysis (see *Results* section for references).

We computed descriptive statistics for demographic variables and generalized estimating equation (GEE) models for multivariate data modeling of parameter estimates [14,28]. A GEE shows how the average of a response variable of an individual changes with covariates. Additionally, a GEE allows us to view this correlation between repeated measurements on the same individual [29,30].

We used GEE models with the logit function and an exchangeable correlation matrix. For each GEE model, the four demographic predictors were used as independent variables. As the outcome measures were dichotomized, we used binomial logistic models. Independent variables included scenario type (ie, own PCP, different PCP from the same organization, and different PCP from a different organization), age, income, and gender. Each set of independent variables included a reference that yielded adjusted odds ratios (ORs), determining the odds of individuals responding other than the reference category, with the dependent variable in mind (see *Results* section). References were chosen based on the limiting impact of relationship for scenario types and impact of socioeconomic accessibility issues (eg, higher income and age).

Results

Overview

We conducted the survey from July 1 to July 2, 2022, and accrued a sample of 4639 participants, resulting in 4577 completed surveys (98.66% completion rate).

Patient Demographics and Preferences

We present complete descriptive statistics for survey items in Table 1. Patient demographics were representative of the national population by gender and age. Household income displayed a top-heavy trend with 31.07% (1422/4577) reporting they make more than US $75,000.

Table . Sample characteristics (N=4577).

Demographics		Participants, n (%)
Age (years)		
	18 - 29	1109 (24.2)
	30 - 44	1076 (23.5)
	45 - 60	1268 (27.7)
	>60	1124 (24.6)
Sex		
	Female	2374 (51.9)
	Male	2203 (48.1)
Household income (US $)		
	$0-$9.999	356 (7.8)
	$10,000-$24,999	569 (12.4)
	$25,000-$49,999	1020 (22.3)
	$50,000-$74,999	731 (16)
	$75,000-$99,999	505 (11)
	$100,000-$124,999	346 (7.6)
	$125,000-$149,999	203 (4.4)
	$150,000-$174,999	106 (2.3)
	$175,000-$199,999	80 (1.7)
	$200,000+	182 (4)
	Prefer not to answer	479 (10.5)
Region		
	East North Central	555 (12.1)
	East South Central	256 (5.6)
	Middle Atlantic	660 (14.4)
	Mountain	333 (7.3)
	New England	224 (4.9)
	Pacific	861 (18.8)
	South Atlantic	844 (18.4)
	West North Central	288 (6.3)
	West South Central	464 (10.1)
Device type		
	iOS phone or tablet	2501 (54.6)
	Android phone or tablet	1076 (23.5)
	Windows desktop or laptop	314 (6.9)
	MacOS desktop or laptop	110 (2.4)
	Other	29 (0.6)

Patients reported a length of relationship with their current PCP of less than 6 months (424/4577, 9.26%), 6 months to a year (527/4577, 11.51%), 1 to 3 years (996/4577, 21.76%), 3 to 5 years (574/4577, 12.54%), or 5 years or more (1618/4577, 35.35%). Some patients (438/4577, 9.57%) reported not having a PCP. Patients noted the frequency of visits to a PCP in the last 12 months: 1 time (1033/4577, 22.57%), 2 times (1197/4577, 26.15%), 3 times (752/4577, 16.43%), 4 times (422/4577, 9.22%), 5 or more times (459/4577, 10.03%), or none (714/4577, 15.6%).

Fewer than half of patients (1909/4577, 41.71%) agreed (1288/4577, 28.14%) or strongly agreed (621/4577, 13.57%) that it is important that their current health care professional offers telemedicine visits. Some (1286/4577, 28.10%) agreed (936/4577, 20.45%) or strongly agreed (350/4577, 7.65%) that they would consider switching to a new health care professional, who offers telemedicine. Most (2864/4577, 62.60%) agreed (1843/4577, 40.27%) or strongly agreed (1021/4577, 22.31%) that it is important to have an established relationship with the health care professional they are having a telemedicine visit with. Most patients (3235/4577, 70.68%) agreed (1706/4577, 37.27%) or strongly agreed (1529/4577, 33.41%) that it is important that their health care professional has access to their health records.

Measures of Satisfaction and Ease of Using Telemedicine

About 35% (1588/4577) of patients reported having had a telemedicine visit with their PCP and 26.59% (1231/4577) reported that their PCP offers telemedicine, but they have not had a telemedicine visit. Of these, 70.34% (1117/1588) reported being satisfied (651/1588, 40.99%) or very satisfied (466/1588, 29.35%) with telemedicine for online video visits, and 77.83% (1236/1588) reported feeling satisfied (578/1588, 36.40%) or very satisfied (658/1588, 41.44%) with in-person visits at clinic (Figure 1). Regarding ease and difficulty of visits, 71.28% (1132/1588) reported telemedicine visits as being easy (612/1588, 38.54%) or very easy (520/1588, 32.75%), and 62.9% (999/1588) reported seeing their health care professional in person at a clinic as easy (581/1588, 36.59%) or very easy (418/1588, 26.32%; Figure 2). If they no longer had the option to meet with their PCP using telemedicine, 69.9% (1110/1588) reported that they would be somewhat to very disappointed (Figure 3). Some patients reported that their PCP did not offer telemedicine as an option (551/4577, 12.04%), or that they were not sure whether their PCP offered telemedicine (974/4577, 21.28%).

Figure 1. Satisfaction meeting with your PCP and other health care professionals. PCP: primary care health care professional.

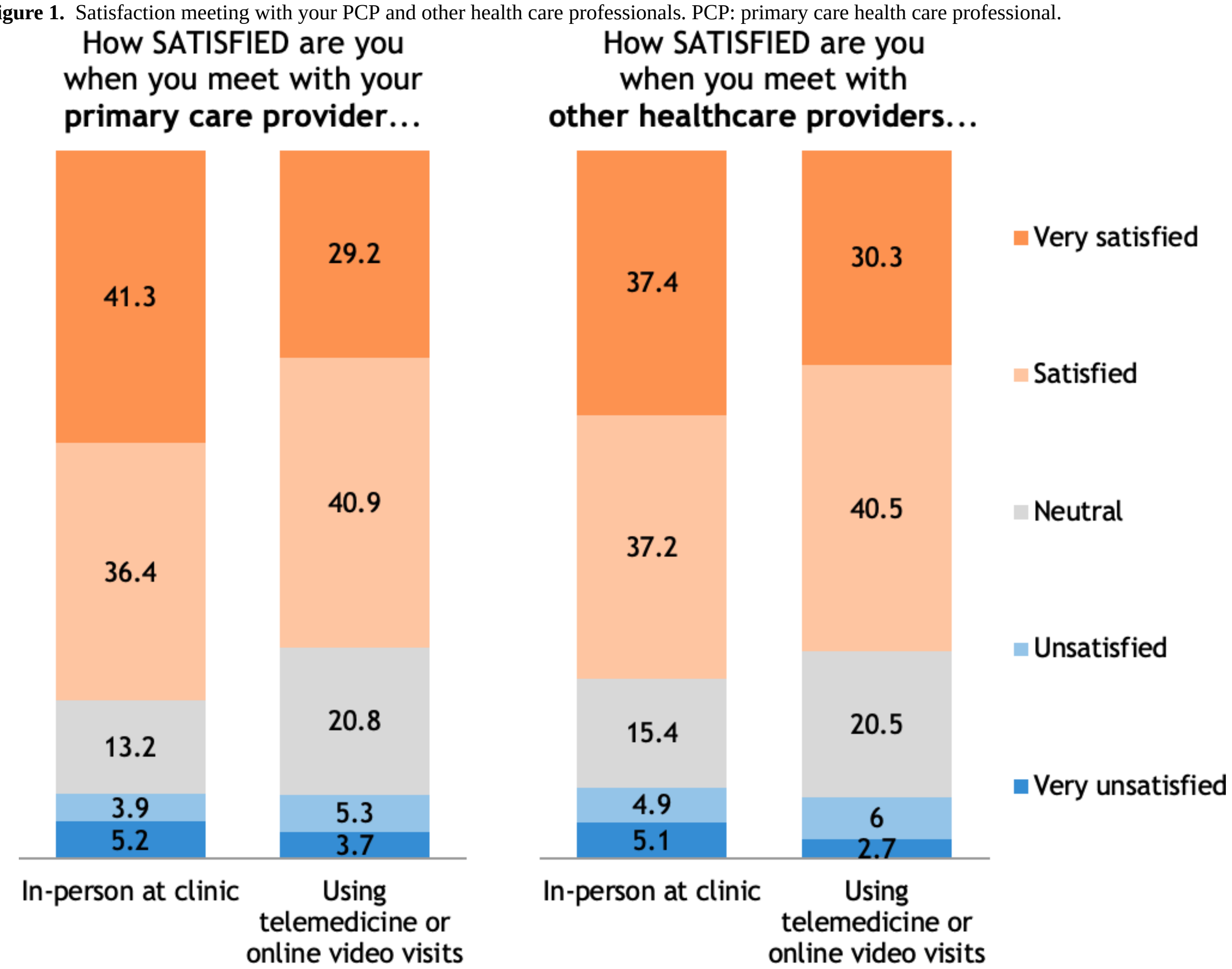


Figure 2. Ease of meeting with your PCP and other health care professionals. PCP: primary care health care professional.

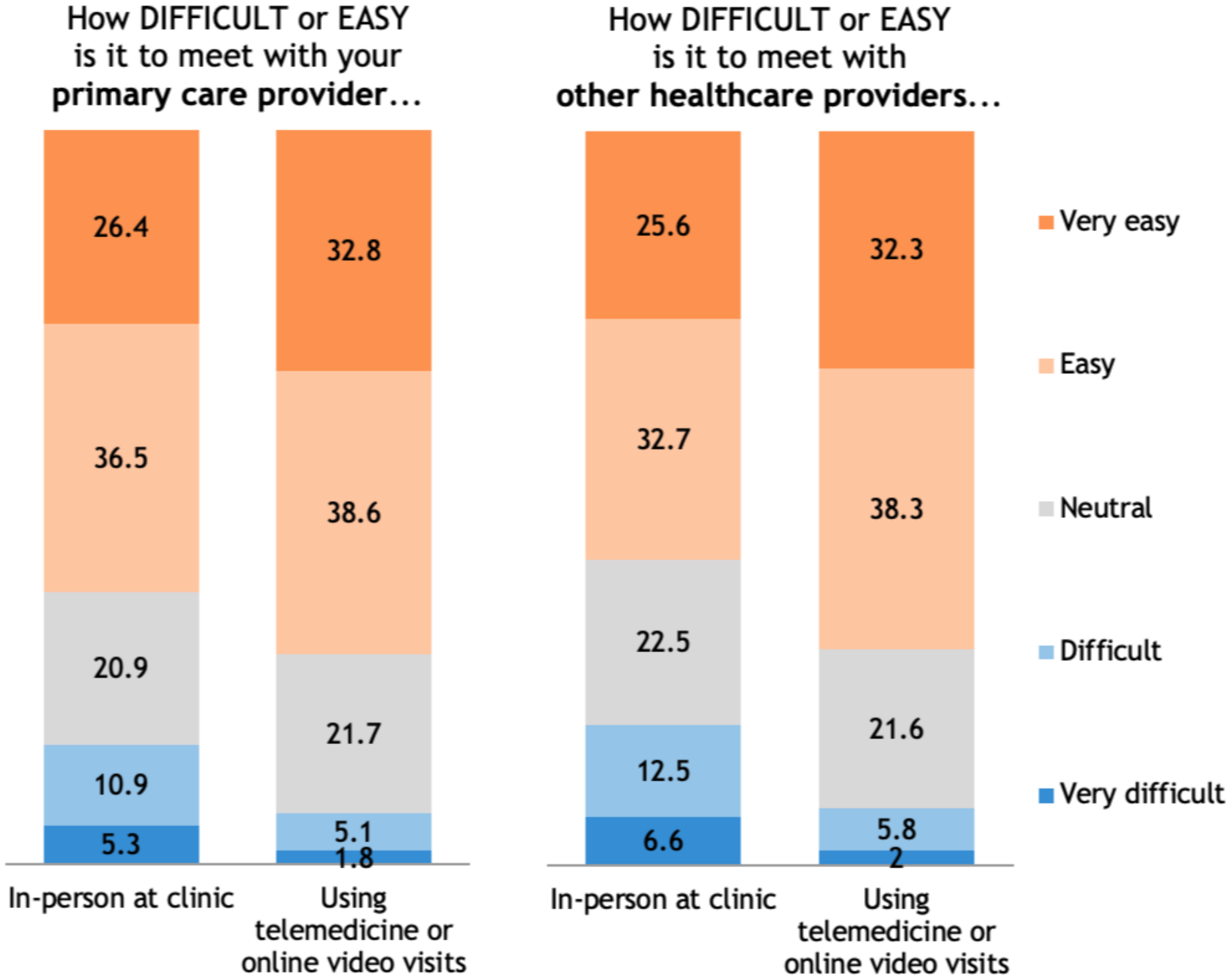
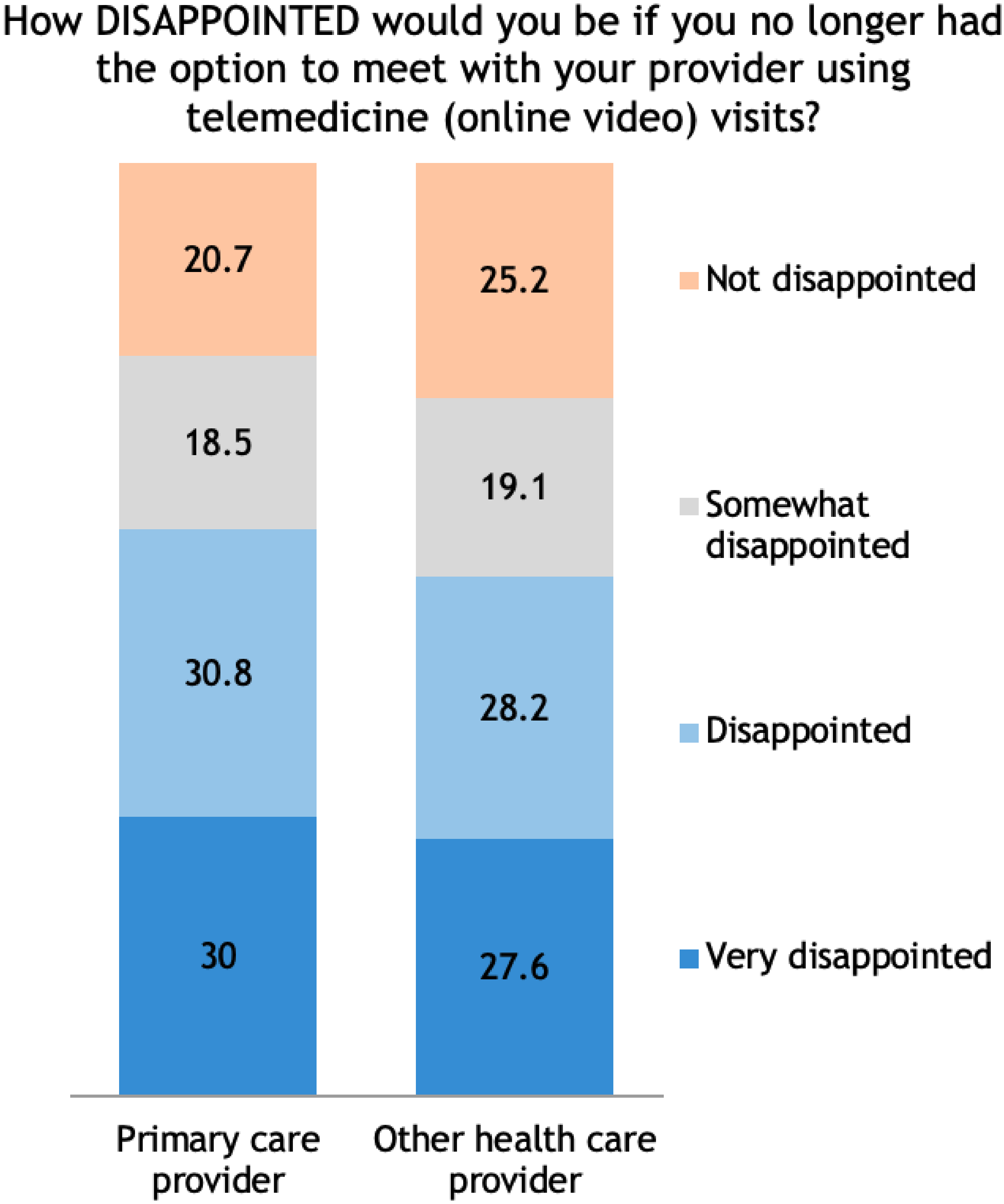


Figure 3. Disappointment if telemedicine was no longer an option to meet with your health care professional.



About one-third (1513/4577, 33.06%) of patients reported having had a telemedicine visit with any health care professional (eg, mental health professional, cardiologist, or dermatologist), while 25.54% (1169/4577) reported that their other health care professional offered telemedicine services, but that they had not had a telemedicine visit. Of the 33.06% (1513/4577) that have had a telemedicine visit, 70.79% (1071/1513) reported being satisfied (612/1513, 40.45%) or very satisfied (459/1513, 30.34%) with telemedicine visits, and 74.55% (1128/1513) reported being satisfied (562/1513, 37.14%) or very satisfied (566/1513, 37.41%) with in-person visits at the clinic (Figure 1). Regarding ease and difficulty of visits, 70.65% (1069/1513) reported telemedicine visits with other health care professionals as being easy (581/1513, 38.40%) or very easy (488/1513, 32.25%), and 58.36% (883/1513) reported seeing their other health care professionals in person at a clinic as easy (497/1513, 32.84%) or very easy (386/1513, 25.51%; Figure 2). If they no longer had the option to meet with their other health care professionals using telemedicine, 72.44% (1096/1513) reported that they would be somewhat to very disappointed (Figure 3). Some patients reported that other health care professionals did not offer telemedicine as an option (683/4577, 14.92%) or that they were not sure if their other health care professionals offered telemedicine (1255/4577, 27.42%).

We fitted GEE models for patients who had experienced telemedicine visits and considered scenario type, age, gender, and income. These models suggested patients had lower odds of reporting ease when meeting in person with an established health care professional or new professional from the same organization compared to a telehealth visit with a new provider from a different organization (Table 2). Patients aged between 18 - 29 years (OR 0.51, 95% CI 0.37 to 0.71, *P*<.001) and 30 - 44 years (OR 0.61, 95% CI 0.44 to 0.84, *P*=.003) had lower odds of reporting ease when meeting in person with their PCP, when compared to patients aged 60 years and older. Individuals who reported an income of US $0 - $24,999 had lower odds of reporting ease when meeting in person with their PCP (OR 0.69, 95% CI 0.50 to 0.95, *P*=.02) as compared to those with an income of US $100,000 or over. Gender did not appear to impact ease of meeting with a health care professional.

Patients reporting poverty-level income (US $0 - $24,999) had lower odds of reporting satisfaction with in-person health care professional meetings when compared to those with an income of US $100,000 or more (OR 0.53, 95% CI 0.37 to 0.75, *P*<.001). Scenario type (eg, own PCP), gender, and age had no statistically significant impact on reported satisfaction in meeting with health care professionals.

Table . Odds ratios from generalized estimating equation models predicting ease to meet with health care professional and satisfaction of using telemedicine.

		Ease to meet with health care professional		Satisfaction meeting with health care professional	
		Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value
Scenario					
	PCP[a]—in person	0.404 (0.32 to 0.51)	<.001	0.96 (0.76 to 1.23)	.75
	PCP—telehealth	1.097 (0.85 to 1.41)	.47	0.97 (0.78 to 1.21)	.79
	Other health care professional—in person	0.344 (0.28 to 0.43)	<.001	0.86 (0.69 to 1.07)	.17
	Other health care professional—telehealth	Reference	—[b]	Reference	—
Income (US $)					
	$0 - $24,999	0.69 (0.50 to 0.95)	.02	0.53 (0.37 to 0.75)	<.001
	$25,000 - $49,999	0.82 (0.60 to 1.12)	.21	0.77 (0.53 to 1.11)	.16
	$50,000 - $74,999	0.94 (0.66 to 1.34)	.74	0.84 (0.58 to 1.22)	.37
	$75,000 - $99,999	0.78 (0.52 to 1.15)	.21	0.74 (0.48 to 1.16)	.19
	$100,000+	Reference	—	Reference	—
Gender					
	Male	1.07 (0.86 to 1.33)	.56	0.87 (0.68 to 1.11)	.26
	Female	Reference	—	Reference	—
Age (years)					
	18 - 29	0.51 (0.37 to 0.71)	<.001	0.83 (0.58 to 1.19)	.31
	30 - 44	0.61 (0.44 to 0.84)	.003	0.98 (0.69 to 1.38)	.91
	45 - 60	0.81 (0.58 to 1.12)	.20	1.22 (0.87 to 1.71)	.24
	60+	Reference	—	Reference	—

[a]PCP: primary care health care professional.

[b]Not applicable.

Measures of Telemedicine Use, Willingness, and Comfort

Over half (2836/4577, 61.96%) of patients would be willing (1573/4577, 34.37%) or very willing (1263/4577, 27.59%) to have a telemedicine visit with their health care professional. About half (2278/4577, 49.77%) would be willing (1645/4577, 35.94%) or very willing (633/4577, 13.83%) to have a telemedicine visit with a different health care professional from the same health care organization. About one-third (1550/4577, 33.86%) would be willing (1037/4577, 22.66%) or very willing (513/4577, 11.21%) to have a telemedicine visit with a different health care professional from a different health care organization (Figure 4).

The odds of patients being willing to use telemedicine with their own PCP (OR 2.49, 95% CI 2.27 to 2.73, *P*<.001) or a different PCP from the same organization (OR 2.21, 95% CI 2.05 to 2.38, *P*<.001) were statistically significantly higher compared to a different PCP from a different organization (Table 3). Age showed an impact on willingness to use telemedicine, with the age group 30 - 44 years having the highest odds of willingness compared to those aged 60 and older (OR 2.33, 95% CI 1.96 to 2.76, *P*<.001). Income and gender did not appear to impact willingness to use telemedicine.

Most patients (2819/4577, 61.59%) reported being comfortable (1672/4577, 36.53%) or very comfortable (1147/4577, 25.06%) having a telemedicine visit with their health care professional. Most (2333/4577, 50.97%) would be comfortable (1729/4577, 37.78%) or very comfortable (604/4577, 13.20%) having a telemedicine visit with a different PCP in the same health care organization. Approximately one-third of patients (1533/4577, 33.49%) reported being comfortable (1050/4577, 22.94%) or very comfortable (483/4577, 10.55%) having a telemedicine visit with a different health care professional from a different health care organization (Figure 5).

The odds of patients feeling more comfortable in using telemedicine with their own PCP (OR 2.72, 95% CI 2.47 to 2.99, *P*<.001) or a different PCP from the same organization (OR 2.56, 95% CI 2.37 to 2.77, *P*<.001) were significantly higher than those with a different PCP from a different organization (Table 3). Age had a significant effect on comfort in using telemedicine, with the age group of 30 - 44 years reporting the highest odds of comfort in using telemedicine compared to those aged 60 years and older (OR 2.35, 95% CI 1.98 to 2.80, *P*<.001). Individuals reporting an income of US $0 - $24,999 had lower odds (OR 0.81, 95% CI 0.67 to 0.97, *P*=.02) of being comfortable using telemedicine in reference to those who made US $100,000 or more. Gender did not appear to affect comfortability in using telemedicine.

Figure 4. Willingness to have a telemedicine visit.

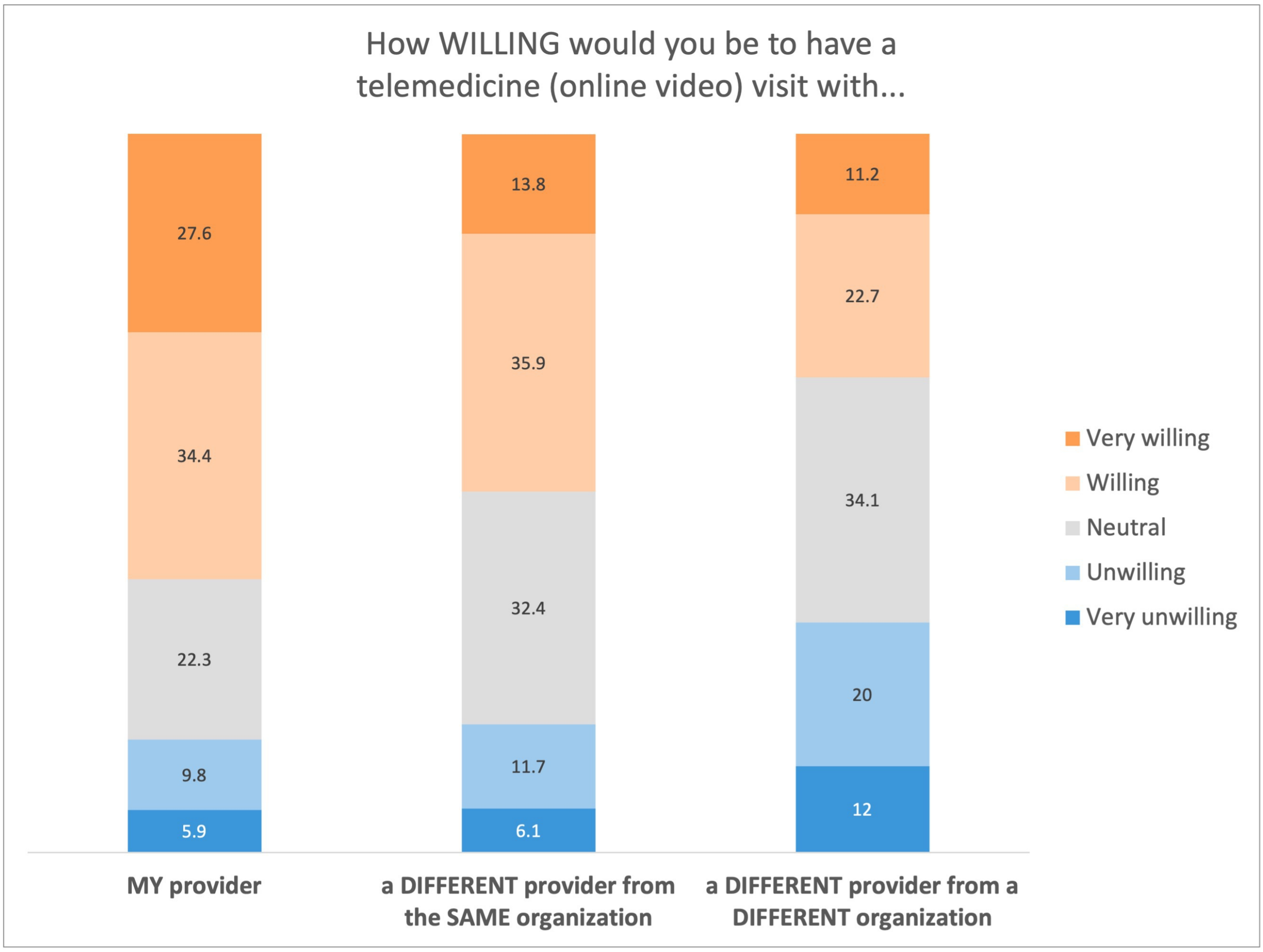


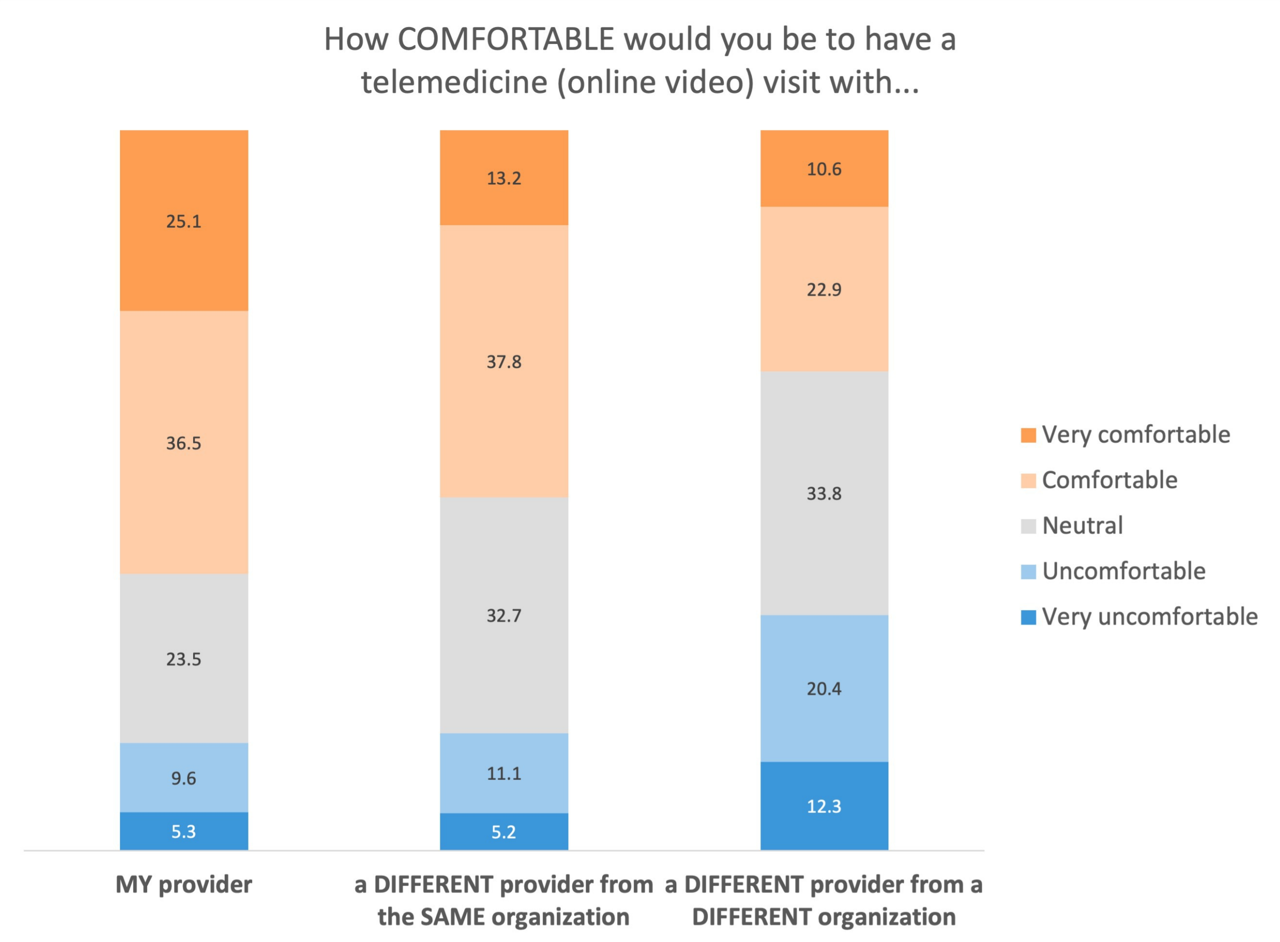
Table . Odds ratios from generalized estimating equation models predicting willingness and comfortability of using telemedicine.

		Willingness to use telemedicine		Comfort using telemedicine	
		Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value
Scenario					
	Own PCP[a]	2.49 (2.27 to 2.73)	<.001	2.72 (2.47 to 2.99)	<.001
	Different PCP or same organization	2.21 (2.05 to 2.38)	<.001	2.56 (2.37 to 2.77)	<.001
	Different PCP or different organization	Reference	—[b]	Reference	—
Income (US $)					
	$0 - $24,999	0.97 (0.81 to 1.17)	.76	0.81 (0.67 to 0.97)	.02
	$25,000 - $49,999	0.95 (0.79 to 1.13)	.53	0.93 (0.77 to 1.11)	.40
	$50,000 - $74,999	1.02 (0.84 to 1.23)	.84	0.96 (0.79 to 1.17)	.70
	$75,000 - $99,999	1.01 (0.82 to 1.25)	.90	0.92 (0.74 to 1.13)	.42
	$100,000+	Reference	—	Reference	—
Gender					
	Male	0.99 (0.88 to 1.12)	.91	1.07 (0.94 to 1.21)	.30
	Female	Reference	—	Reference	—
Age (years)					
	18 - 29	1.61 (1.37 to 1.90)	<.001	1.81 (1.53 to 2.13)	<.001
	30 - 44	2.33 (1.96 to 2.76)	<.001	2.35 (1.98 to 2.80)	<.001
	45 - 60	1.94 (1.65 to 2.29)	<.001	1.97 (1.67 to 2.32)	<.001
	60+	Reference	—	Reference	—

[a]PCP: primary care health care professional.

[b]Not applicable.

Figure 5. Comfortability having a telemedicine visit.



Discussion

Principal Results

This study aimed to assess demographic trends in patient preferences in telemedicine visits; measure patient ease of use and satisfaction in using telemedicine; and measure changes in patient use, willingness, and comfort with telemedicine since 2017. The results showed that telemedicine has become an accepted part of many people's health care. We found individuals reporting incomes at or below poverty levels tended to find telemedicine more difficult to use, were less comfortable using it, and were ultimately less satisfied with a telemedicine visit. The level of patient–health care professional relationship (scenario types) significantly affected willingness and comfort in using telemedicine, with closer relationships showing higher levels of comfort and willingness to use telemedicine. This study leveraged the baseline data from a study published in 2017 to understand current patient preferences and the potential impact of increased exposure to telemedicine [14].

Comparison With Prior Work

We compared public use, knowledge, and perceptions of telemedicine from 2017 and 2022. Importantly, patient telemedicine preferences and knowledge had changed dramatically with 61.1% of patients knowing their PCP offered telemedicine compared to 5.3% in 2017 and 34.69% had used telemedicine with their PCP as compared to only 3.5% in 2017 [14]. Interestingly, nearly a quarter (21.28%) and over a quarter (27.42%) of patients did not know whether their PCP or other health care professionals, respectively, offered telemedicine. While only 12.04% and 14.92% of patients had PCPs or other health care professionals not offering them telemedicine, these numbers contrast greatly when noting that 67.7% of patients in 2017 reported that their PCPs did not offer telemedicine at all. Though an increase was expected, these numbers highlight the drastic increase telemedicine use saw in a short period of time. To put these numbers in perspective regarding general health care visits, most (69.65%) individuals reported being with their PCP for at least one year, and nearly half (48.72%) went to their PCP annually or biannually. While little change was seen in general health care access from 2017, major changes were indeed visible in telemedicine knowledge and use [14].

Overall, satisfaction in the use of telemedicine was found to be comparable (70.34%) to visits in person (77.83%) with a PCP. While in-person visits trended to be higher in satisfaction, telemedicine visits trended to be higher (71.28%) in ease for visits with a PCP than in-person visits (62.90%). About 70% of patients reported they would feel some level of disappointment should their PCP no longer offer telemedicine options. These results conveyed that while people view telemedicine visits similarly to in-person visits regarding satisfaction and ease, having the option for telemedicine access is now expected. Patients and health care professionals now use telemedicine as an additional feature, intervention mode, or aspect of treatment [31].

While the health care professional scenario type impacted the odds of individuals reporting higher ease of using telemedicine for visits, this study indicated patients also generally find telemedicine visits to be easier than in-person visits. Interestingly, a similar result is not seen for the satisfaction measure. Indeed, while age and income had a significant impact on whether patients found telemedicine easier or harder, GEE models showed that patients with an income of US $0 - $24,999 were the only group significantly more likely to report lower satisfaction with health care professional visits when compared to the reference group. Such a result is not unexpected, as those at or below the poverty line would have the most difficulty in accessing telemedicine for health care visits due to the costs associated with having the right hardware (computer, tablet, or smartphone), access to reliable internet, and availability of privacy for a visit [32,33].

Literature supports this study's results that patients at poverty level incomes are less comfortable using telemedicine, find telemedicine more difficult to use, and are less satisfied with telemedicine than those in higher income brackets. Though telemedicine increases access to health care for many [32], Curtis et al [33] show that digital access disparities can exacerbate current health care inequalities, leaving those most vulnerable even further behind. While telemedicine can be provided from the comfort and convenience of one's home [34], such a scenario does not explicitly mean a patient is comfortable to use telemedicine [35,36]. For example, a patient who may need to borrow a smartphone for a telemedicine visit may feel uncomfortable asking for the device. While some studies have shown that both health care professionals and patients feel telemedicine may improve access to care due to convenience, there is a need for better assessment of actual patient satisfaction, comfort, and willingness in using telemedicine [14,37]. Considering that the US $0 - $24,999 income group was significantly more likely to report difficulty in meeting with a PCP over telemedicine, this study clearly shows the limitations individuals face due to economic status. Understanding such patient preferences informs approaches to increase telemedicine accessibility [35].

While the development of the patient–health care professional relationship is a critical part of medicine, it is even more scrutinized in telemedicine [38-40]. Though federal laws allow health care professionals to establish a relationship with a new patient via telemedicine, some states pose specific stipulations that severely restrict this process [41]. The three scenarios used in the patient survey provide insight into key aspects of the patient–health care professional relationship—necessary insight for context of how telemedicine is provided in terms of visit environment and communication. The importance of the patient-provider relationship scenarios became evident when looking at reported willingness and comfort in using telemedicine by patients. The results highlighted that people are significantly more willing and comfortable in using telemedicine with their own PCP or at least a PCP from their health organization than an unknown PCP from a different organization altogether. Welch et al [14] found that patients became less willing to use telemedicine as they became further detached from their own health care professional in 2017, our study showed a similar trend for both willingness and comfort. While trust in health care professionals may be measured in many ways, the Trust in Physician scale has shown that patient trust increased with length of relationship with a provider and that patients who were able to choose their providers exhibited higher trust scores [42]. One study by Orrange et al [43] surveyed 368 patients during March-April 2020 regarding satisfaction, trust, and concerns related to telemedicine visits. Patients reported being very satisfied (47.4%) or satisfied (35.3%) with their telemedicine visits, significant correlation with Trust in Physician scores and technical issues, concerns over privacy and cost, satisfaction with convenience, and the amount of time spent in a visit [43]. As the interpersonal patient–health care professional relationship becomes further scrutinized with the new era of telemedicine, knowledge of patient perceptions regarding satisfaction and trust are important for successful implementation of telemedicine and telemedicine policy. The future of health care seems to include a combination of telemedicine and in-person care access, which means improving telemedicine with such factors in mind are key for success.

Though general trends show that willingness and comfort in using telemedicine have increased since 2017, the level of change may provide insight toward actual telemedicine access. Individuals were more likely to report greater comfort and willingness to use telemedicine depending on their age, income, and scenario type. While income did not factor into willingness to use telemedicine, individuals reporting an income at or below poverty levels were significantly less likely to be comfortable using telemedicine. Some literature shows that income may not relate to certain aspects of comfort such as technological literacy; yet, there is a definite need in understanding how income inequalities may affect feelings of comfort [44]. Though all patients may be equally willing to try telemedicine based on income, differences in comfort highlighted current limitations for telemedicine visits in the United States. For example, while patients may want to use telemedicine, they may find themselves uncomfortable in using the technology or asking others for aid [45]. Such differences between willingness and comfort levels have become easier to see: in the 2017 study, willingness and comfort mirrored one another, while this study saw a distinct separation of the two when looking at the comfort of using telemedicine with a wholly new health care professional—18.6% for both comfort and willingness in 2017, and 33.49% comfort and 33.86% willingness now [14]. To implement telemedicine successfully, stakeholder experience and needs must be considered (eg, why is patient comfort lagging behind willingness to use telemedicine?) to make changes in policy and organizational processes.

Prior studies have shown the importance of the patient–health care professional relationship in patient satisfaction and trust [43]. This study showcased that satisfaction and willingness are affected by this relationship and bolstered prior findings. This knowledge is critical in developing successful telemedicine guidelines for health care professionals and patients. As changes to policy and regulation occur for telemedicine [46], a careful consideration of patient preferences and needs would help develop more effective, timely policies. With the knowledge that health care is moving toward a hybrid care model, successful implementation of telemedicine requires that both patient and health care professional preferences are considered. The public health emergency gave telemedicine more regulatory flexibility for both stakeholders, thus helping to lead to the sustained use of telemedicine [47]. Now, policy makers are under pressure to ensure patient and health care professional needs are being met regarding telemedicine and ensuring patient safety and health information security. Understanding that more of the population feel telemedicine is important (41.71% from 19.8% in 2017) showed that telemedicine is now an expected and significant part of health care.

Limitations

We obtained a national sample using SurveyMonkey Audience, in part to emulate sampling from the 2017 study. However, the platform no longer includes race or education as part of reported demographics or as sampling criteria. Without data describing race and education, it is difficult to compare or fully understand socioeconomic differences between the original 2017 survey and this study. However, we remained able to examine age, gender, and income in relation to willingness, comfort, satisfaction, and ease of use of telemedicine within the United States. It is important to note that though sampling was carried out using the same methods for the purpose of consistency, bias is possible in online survey recruitment platforms. For example, those surveyed already have access to technology in the forms of the devices used and internet. This study leverages the 2017 results as baselines where possible and relevant. As a result, not all methods and analyses reflect the prior study as descriptive and frequency statistics were determined to provide a satisfactory mode for comparison. Future work should include further measures of socioeconomic status, such as race, ethnicity, and education, to further understand how telemedicine is currently used and what can be carried out to improve its accessibility and levels of success.

Conclusion

With the increased public exposure to telemedicine, there has been a visible difference in telemedicine use and perceptions among the US population. Willingness and comfort to use telemedicine have increased since 2017. Further, the patient–health care professional relationship appears to influence willingness, comfort, and ease of using telemedicine. Additionally, this study highlighted the significant negative effect income has for individuals regarding comfort, satisfaction, and ease of use of telemedicine; this result is especially important when considering that telemedicine is touted as a breakthrough for health care access. Overall, there has been a large increase in telemedicine usage since 2017, with more finding telemedicine an important part of their health care.

Conflicts of Interest

FK, JH, NO, and JM have no conflicts to disclose. BW is a shareholder, and all other authors not listed earlier are employees of Doxy.me Inc, a commercial telemedicine company. The authors declare no other conflicts of interest.

Multimedia Appendix 1
Survey questions.
[DOCX File, 20 KB - humanfactors_v11i1e51056_app1.docx]

References

1. Allely EB. Synchronous and asynchronous telemedicine. J Med Syst 1995 Jun;19(3):207-212. [doi: 10.1007/BF02257174] [Medline: 7643019]
2. Brotman JJ, Kotloff RM. Providing outpatient telehealth services in the United States: before and during coronavirus disease 2019. Chest 2021 Apr;159(4):1548-1558. [doi: 10.1016/j.chest.2020.11.020] [Medline: 33245875]
3. Uscher-Pines L, Mulcahy A, Cowling D, Hunter G, Burns R, Mehrotra A. Access and quality of care in direct-to-consumer telemedicine. Telemed J E Health 2016 Apr;22(4):282-287. [doi: 10.1089/tmj.2015.0079] [Medline: 26488151]
4. Cieślik B, Kuligowski T, Cacciante L, Kiper P. The impact of personality traits on patient satisfaction after telerehabilitation: a comparative study of remote and face-to-face musculoskeletal rehabilitation during COVID-19 lockdown. Int J Environ Res Public Health 2023 Mar 12;20(6):5019. [doi: 10.3390/ijerph20065019] [Medline: 36981927]
5. Wu TC, Ho CTB. Barriers to telemedicine adoption during the COVID-19 pandemic in Taiwan: comparison of perceived risks by socioeconomic status correlates. Int J Environ Res Public Health 2023 Feb 16;20(4):3504. [doi: 10.3390/ijerph20043504] [Medline: 36834205]
6. Qian AS, Schiaffino MK, Nalawade V, et al. Disparities in telemedicine during COVID-19. Cancer Med 2022 Feb;11(4):1192-1201. [doi: 10.1002/cam4.4518] [Medline: 34989148]
7. Sanayei N, Albrecht MM, Martin DC, et al. Outcomes of a hybrid ophthalmology telemedicine model for outpatient eye care during COVID-19. JAMA Netw Open 2022 Aug 1;5(8):e2226292. [doi: 10.1001/jamanetworkopen.2022.26292] [Medline: 36006645]
8. Adepoju OE, Dang P, Valdez MR. Comparing in-person only, telemedicine only, and hybrid health care visits among older adults in safety-net clinics. Telemed Rep 2023;4(1):93-99. [doi: 10.1089/tmr.2023.0003] [Medline: 37283857]

9. Lucas JW, Villarroel MA. Telemedicine Use Among Adults: United States 2021. : Centers for Disease Control and Prevention; 2022. [doi: 10.15620/cdc:121435]
10. Gudipudi R, Jungbauer WN Jr, Brennan E, Pecha PP. Parent and provider satisfaction of telehealth in pediatric surgical subspecialty care: a systematic review. J Telemed Telecare 2024 Jul;30(6):905-917. [doi: 10.1177/1357633X221110368] [Medline: 35850559]
11. Eldaly AS, Maniaci MJ, Paulson MR, et al. Patient satisfaction with telemedicine in acute care setting: a systematic review. J Clin Transl Res 2022 Dec 29;8(6):540-556. [Medline: 36518201]
12. Alhajri N, Simsekler MCE, Alfalasi B, et al. Exploring quality differences in telemedicine between hospital outpatient departments and community clinics: cross-sectional study. JMIR Med Inform 2022 Feb 15;10(2):e32373. [doi: 10.2196/32373] [Medline: 34978281]
13. Duan-Porter W, Van Houtven CH, Mahanna EP, et al. Internet use and technology-related attitudes of veterans and informal caregivers of veterans. Telemed J E Health 2018 Jul;24(7):471-480. [doi: 10.1089/tmj.2017.0015] [Medline: 29252110]
14. Welch BM, Harvey J, O'Connell NS, McElligott JT. Patient preferences for direct-to-consumer telemedicine services: a nationwide survey. BMC Health Serv Res 2017 Nov 28;17(1):784. [doi: 10.1186/s12913-017-2744-8] [Medline: 29183372]
15. Davies C, Vas P, Oyibo SO. Telephone follow-up for the management of thyrotoxicosis: a patient satisfaction survey. J Telemed Telecare 2013 Jan;19(1):29-32. [doi: 10.1177/1357633X12474737] [Medline: 23390211]
16. Mekhjian H, Turner JW, Gailiun M, McCain TA. Patient satisfaction with telemedicine in a prison environment. J Telemed Telecare 1999;5(1):55-61. [doi: 10.1258/1357633991932397] [Medline: 10505370]
17. Nguyen M, Waller M, Pandya A, Portnoy J. A review of patient and provider satisfaction with telemedicine. Curr Allergy Asthma Rep 2020 Sep 22;20(11):72. [doi: 10.1007/s11882-020-00969-7] [Medline: 32959158]
18. Pogorzelska K, Chlabicz S. Patient satisfaction with telemedicine during the COVID-19 pandemic-a systematic review. Int J Environ Res Public Health 2022 May 17;19(10):6113. [doi: 10.3390/ijerph19106113] [Medline: 35627650]
19. Rose S, Hurwitz HM, Mercer MB, et al. Patient experience in virtual visits hinges on technology and the patient-clinician relationship: a large survey study with open-ended questions. J Med Internet Res 2021 Jun 21;23(6):e18488. [doi: 10.2196/18488] [Medline: 34152276]
20. Champion A, Congiusta A, Yagnik A. Comparison of patient satisfaction measures between in-person and telemedicine postoperative appointments following third molar surgery. Int J Oral Maxillofac Surg 2021 Jun;50(6):830-834. [doi: 10.1016/j.ijom.2020.12.012] [Medline: 33414036]
21. Aashima, Nanda M, Sharma R. A review of patient satisfaction and experience with telemedicine: a virtual solution during and beyond COVID-19 pandemic. Telemed e-Health 2021 Dec 1;27(12):1325-1331. [doi: 10.1089/tmj.2020.0570]
22. Cascella M, Coluccia S, Grizzuti M, et al. Satisfaction with telemedicine for cancer pain management: a model of care and cross-sectional patient satisfaction study. Curr Oncol 2022 Aug 4;29(8):5566-5578. [doi: 10.3390/curroncol29080439] [Medline: 36005177]
23. Reider-Demer M, Jalilian L, Roy S, et al. Implementation and evaluation of a neurology telemedicine initiative at a major academic medical center. Telemed J E Health 2022 Feb;28(2):158-166. [doi: 10.1089/tmj.2020.0502] [Medline: 33913758]
24. Assistant Secretary for Public Affairs (ASPA). HHS fact sheet: telehealth flexibilities and resources and the COVID-19 public health emergency. US Department of Health & Human Services. 2023 May 10. URL: https://www.hhs.gov/about/news/2023/05/10/hhs-fact-sheet-telehealth-flexibilities-resources-covid-19-public-health-emergency.html [accessed 2024-10-25]
25. Daniel H, Sulmasy LS, Health and Public Policy Committee of the American College of Physicians. Policy recommendations to guide the use of telemedicine in primary care settings: an American College of Physicians position paper. Ann Intern Med 2015 Nov 17;163(10):787-789. [doi: 10.7326/M15-0498] [Medline: 26344925]
26. ATA overview on in-person requirements. American Telemedicine Association. 2021. URL: https://www.americantelemed.org/policies/ata-overview-on-in-person-requirements-for-telehealth-mental-health-services/ [accessed 2024-10-25]
27. Reach your target audience with. SurveyMonkey. URL: https://www.surveymonkey.com/market-research/audience-targeting/ [accessed 2024-10-25]
28. Bakhtiyar K, Beiranvand R, Ardalan A, et al. An investigation of the effects of infertility on women's quality of life: a case-control study. BMC Womens Health 2019 Sep 4;19(1):114. [doi: 10.1186/s12905-019-0805-3] [Medline: 31484531]
29. Conroy A. The influence of relationship power dynamics on HIV testing in rural Malawi. J Sex Res 2014;52(3):347-359. [doi: 10.1080/00224499.2014.883590] [Medline: 24670263]
30. Huang FL. Analyzing cross-sectionally clustered data using generalized estimating equations. J Educ Behav Stat 2022 Feb;47(1):101-125. [doi: 10.3102/10769986211017480]
31. Gass F, Halle M, Mueller S. Telemedicine acceptance and efficacy in the context of preventive cardiology interventions: a systematic review. Digit Health 2022;8:20552076221114186. [doi: 10.1177/20552076221114186] [Medline: 35874861]
32. Akintunde TY, Akintunde OD, Musa TH, et al. Expanding telemedicine to reduce the burden on the healthcare systems and poverty in Africa for a post-coronavirus disease 2019 (COVID-19) pandemic reformation. Glob Health J 2021 Sep;5(3):128-134. [doi: 10.1016/j.glohj.2021.07.006] [Medline: 36338822]

33. Curtis ME, Clingan SE, Guo H, Zhu Y, Mooney LJ, Hser YI. Disparities in digital access among American rural and urban households and implications for telemedicine-based services. J Rural Health 2022 Jun;38(3):512-518. [doi: 10.1111/jrh.12614] [Medline: 34355427]
34. Beck CA, Beran DB, Biglan KM, et al. National randomized controlled trial of virtual house calls for Parkinson disease. Neurology (ECronicon) 2017 Sep 12;89(11):1152-1161. [doi: 10.1212/WNL.0000000000004357] [Medline: 28814455]
35. Cottrell MA, Hill AJ, O'Leary SP, Raymer ME, Russell TG. Patients are willing to use telehealth for the multidisciplinary management of chronic musculoskeletal conditions: a cross-sectional survey. J Telemed Telecare 2018 Aug;24(7):445-452. [doi: 10.1177/1357633X17706605] [Medline: 28449620]
36. Russell TG, Gillespie N, Hartley N, Theodoros D, Hill A, Gray L. Exploring the predictors of home telehealth uptake by elderly Australian healthcare consumers. J Telemed Telecare 2015 Dec;21(8):485-489. [doi: 10.1177/1357633X15606264] [Medline: 26391512]
37. Hall TL, Connelly L, Staton EW, et al. Points of concordance, points of discordance: a qualitative examination of telemedicine implementation. J Am Board Fam Med 2022;35(3):517-526. [doi: 10.3122/jabfm.2022.03.210325] [Medline: 35641043]
38. Bulik RJ. Perspectives on the patient-provider relationship in primary-care telemedicine. Telemed J E Health 2004;10(4):466-468. [doi: 10.1089/tmj.2004.10.466] [Medline: 15689652]
39. Andreadis K, Muellers K, Ancker JS, Horowitz C, Kaushal R, Lin JJ. Telemedicine impact on the patient-provider relationship in primary care during the COVID-19 pandemic. Med Care 2023 Apr 1;61(Suppl 1):S83-S88. [doi: 10.1097/MLR.0000000000001808] [Medline: 36893423]
40. Reed ME, Huang J, Parikh R, et al. Patient-provider video telemedicine integrated with clinical care: patient experiences. Ann Intern Med 2019 Aug 6;171(3):222-224. [doi: 10.7326/M18-3081] [Medline: 31035294]
41. American Medical Association. 50-state survey: establishment of a patient-physician relationship via telemedicine. American Medical Association. 2018. URL: https://www.ama-assn.org/system/files/2018-10/ama-chart-telemedicine-patient-physician-relationship.pdf [accessed 2024-10-26]
42. Thom DH, Ribisl KM, Stewart AL, Luke DA, The Stanford Trust Study Physicians. Further validation and reliability testing of the Trust in Physician Scale. Med Care 1999 May;37(5):510-517. [doi: 10.1097/00005650-199905000-00010] [Medline: 10335753]
43. Orrange S, Patel A, Mack WJ, Cassetta J. Patient satisfaction and trust in telemedicine during the COVID-19 pandemic: retrospective observational study. JMIR Hum Factors 2021 Apr 22;8(2):e28589. [doi: 10.2196/28589] [Medline: 33822736]
44. Irfan A, Lever JM, Fouad MN, et al. Does health literacy impact technological comfort in cancer patients? Am J Surg 2022 Apr;223(4):722-728. [doi: 10.1016/j.amjsurg.2021.08.006] [Medline: 34384588]
45. Galvin E, Desselle S, Gavin B, et al. Patient and provider perspectives of the implementation of remote consultations for community-dwelling people with mental health conditions: a systematic mixed studies review. J Psychiatr Res 2022 Dec;156:668-678. [doi: 10.1016/j.jpsychires.2022.10.051] [Medline: 36399859]
46. American Medical Association. CARES act: AMA COVID-19 pandemic telehealth fact sheet. American Medical Association. 2020. URL: https://www.ama-assn.org/delivering-care/public-health/cares-act-ama-covid-19-pandemic-telehealth-fact-sheet [accessed 2024-10-25]
47. Rossano A, Crijns T, Ring D, Reichenberg J. Clinician preferences for current and planned future use of telemedicine. Telemed J E Health 2022 Sep;28(9):1293-1299. [doi: 10.1089/tmj.2021.0486] [Medline: 35007442]

Abbreviations

GEE: generalized estimating equation
HIPAA: Health Insurance Portability and Accountability Act
OR: odds ratio
PCP: primary care health care professional

Edited by E Borycki; submitted 19.07.23; peer-reviewed by E Vashishtha, M Kluge; revised version received 17.01.24; accepted 23.09.24; published 27.11.24.

Please cite as:
Ivanova J, Wilczewski H, Klocksieben F, Cummins M, Soni H, Ong T, Barrera J, Harvey J, O'Connell N, McElligott J, Welch B, Bunnell B
Patient Preferences for Direct-to-Consumer Telemedicine Services: Replication and Extension of a Nationwide Survey
JMIR Hum Factors 2024;11:e51056
URL: https://humanfactors.jmir.org/2024/1/e51056
doi:10.2196/51056

Original Paper

Accelerometer-Based Physical Activity and Health-Related Quality of Life in Korean Adults: Observational Study Using the Korea National Health and Nutrition Examination Survey

Sujeong Han[1*], MD; Bumjo Oh[1,2*], MD, MPH; Ho Jun Kim[3], MD; Seo Eun Hwang[4], MD, PhD; Jong Seung Kim[1,2], MD, PhD

[1]Department of Family Medicine, SMG-SNU Boramae Medical Center, Seoul, Republic of Korea

[2]College of Medicine, Seoul National University, Seoul, Republic of Korea

[3]Department of Family Medicine, Hongseoung Medical Center, Chungcheongnam-do, Republic of Korea

[4]Department of Family Medicine, Seoul National University Hospital, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Jong Seung Kim, MD, PhD
Department of Family Medicine
SMG-SNU Boramae Medical Center
20, Boramae-ro 5-gil, Dongjak-gu
Seoul, 07061
Republic of Korea
Phone: 82 2 870 2681
Email: atenae68@nate.com

Abstract

Background: Health-related quality of life (HRQoL) reflects an individual's perception of their physical and mental health over time. Despite numerous studies linking physical activity to improved HRQoL, most rely on self-reported data, limiting the accuracy and generalizability of findings. This study leverages objective accelerometer data to explore the association between physical activity and HRQoL in Korean adults.

Objective: The objective of this study is to analyze the relationship between objectively measured physical activity using accelerometers and HRQoL among Korean adults, aiming to inform targeted interventions for enhancing HRQoL through physical activity.

Methods: This observational study included 1298 participants aged 19-64 years from the Korea National Health and Nutrition Examination Survey (KNHANES) VI, who wore an accelerometer for 7 consecutive days. HRQoL was assessed using the EQ-5D questionnaire, and physical activity was quantified as moderate-to-vigorous physical activity accelerometer-total (MVPA-AT) and accelerometer-bout (MVPA-AB). Data were analyzed using logistic regression to determine the odds ratio (ORs) for low HRQoL, adjusting for socioeconomic variables and mental health factors.

Results: Participants with higher HRQoL were younger, more likely to be male, single, highly educated, employed in white-collar jobs, and had higher household incomes. They also reported less stress and better subjective health status. The high HRQoL group had significantly more participants meeting MVPA-AB ≥600 metabolic equivalents (*P*<.01). Logistic regression showed that participants meeting MVPA-AB ≥600 metabolic equivalents had higher odds of high HRQoL (OR 1.55, 95% CI 1.11-2.17). Adjusted models showed consistent results, although the association weakened when adjusting for mental health factors (OR 1.45, 95% CI 1.01-2.09).

Conclusions: The study demonstrates a significant association between HRQoL and moderate to vigorous physical activity sustained for at least 10 minutes, as measured by accelerometer. These findings support promoting physical activity, particularly sustained moderate to vigorous activity, to enhance HRQoL. Further interventional studies focusing on specific physical activity domains such as occupational, leisure-time, and commuting activities are warranted.

(JMIR Hum Factors 2024;11:e59659) doi:10.2196/59659

KEYWORDS

Health-Related Quality of Life (HRQoL); physical activity; Accelerometer; Korea National Health and Nutrition Examination Survey (KNHANES); mobile phone

Introduction

Physical activity (PA) is defined as any bodily movement produced by skeletal muscles that results in energy expenditure [1]. It includes all activities such as work, housework, commuting, and leisure. PA is widely recognized as a critical component of healthy lifestyle, contributing significantly to the prevention and management of various chronic diseases such as cardiovascular disease, diabetes, and obesity [2-4]. Despite various guidelines for PA based on many studies, the rate of aerobic PA in Korea has been continuously decreasing, with the rate among women being less than half (44%) [5].

Quality of life (QoL) is defined as an individual's perception of their position in life in the context of the culture and value systems where they live and in relation to their goals, expectations, standards, and concerns [6]. It is divided into health-related quality of life (HRQoL) and non-HRQoL [7]. As medical technology and accessibility advances, life expectancy increases, and major health problems become chronic diseases; not only the interest in treatment but also the management and prevention of diseases is increasing. HRQoL is a concept used to help determine an individual's physical and mental health to help prevent disease and make treatment decisions [8]. It is important because it can be applied to actual clinical treatment through research.

Previous studies have shown a positive association between PA and HRQoL, emphasizing the importance of maintaining active lifestyle for overall well-being. For instance, a study by Scarabottolo et al [9] found that different domains of PA (occupational PA and leisure-time sports practice) were significantly associated with improved HRQoL. Similarly, Puciato et al [10] reported that PA positively influenced the QoL in working-age people in Poland. However, most of the studies have predominantly focused on elderly populations or patients with specific diseases [11-15]. Furthermore, the majority of these studies have relied on self-reported questionnaires to measure PA [9,10,13,15-17].

In recent years, the use of accelerometers to objectively measure PA levels has gained popularity, offering more accurate and reliable data compared to self-reported measures [18-21]. This advancement provides deeper insights into the relationship between PA and various health outcomes. However, studies investigating the association between objectively measured PA using devices such as accelerometers and HRQoL are relatively scarce, especially in Korea.

Therefore, the aim of this study was to analyze the association between objective PA, as measured by accelerometers, and HRQoL in Korean adults using data from the Korea National Health and Nutrition Examination Survey (KNHANES). By analyzing this relationship, we anticipate to contribute to the development of evidence-based strategies to promote PA and enhance HRQoL.

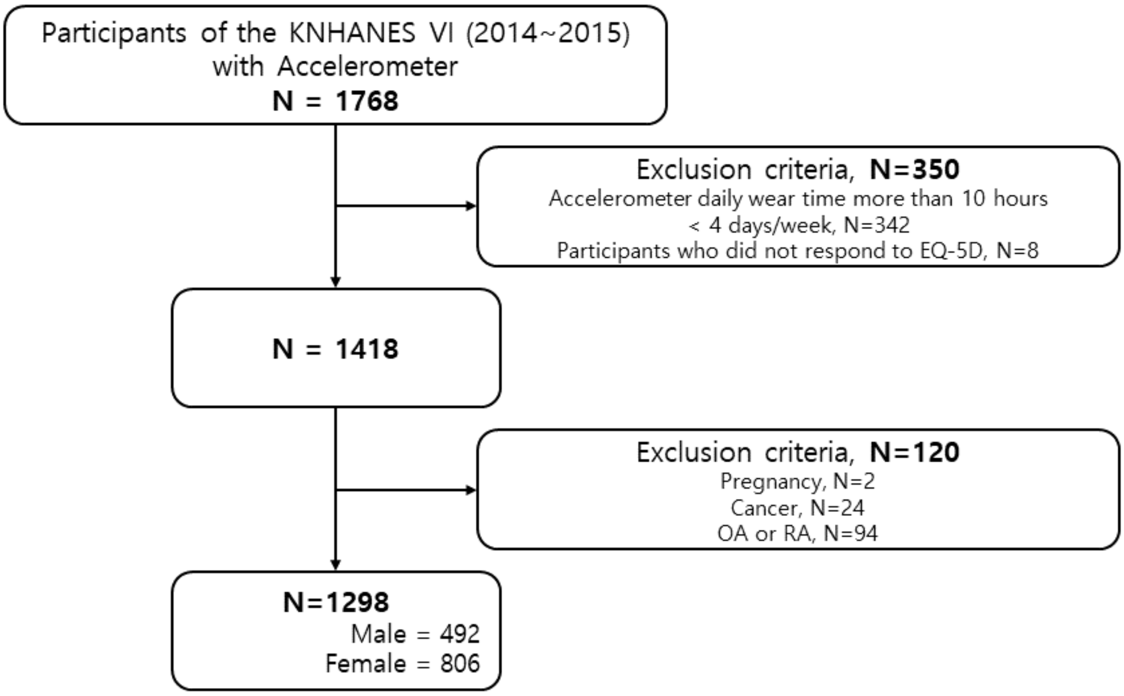
Methods

Data Source and Study Population

This study was based on data from the sixth KNHANES (KNHANES VI) conducted from 2014 to 2015 by the Korea Centers for Disease Control and Prevention (KCDC). KNHANES is a legal survey conducted annually on the level of health of the public, health-related awareness and behaviors, chronic diseases, and food and nutrition intake.

Among the participants of the KNHANES VI (2014-2015), 1827 people participated in the accelerometer survey, but 59 people were excluded due to loss of the accelerometer (9 people), nonwearers (47 people), and mechanical errors (3 people). A total of 1418 people were selected first, excluding 342 people who did not meet the minimum wearing time and number of days of the accelerometer used in the previous study for more than 10 hours a day, 4 days a week, and 8 people who did not respond to the health-related quality of life survey (EQ-5D). Among them, 1298 people were selected as participants for final analysis, excluding 2 pregnant women, 24 people receiving cancer treatment, and 94 people with osteoarthritis or rheumatoid arthritis that could affect their PA (Figure 1).

Figure 1. Flowchart of the study populations. KNHANES: Korea National Health and Nutrition Examination Survey; OA: Osteoarthritis; RA: Rheumatoid arthritis.



Ethical Considerations

It was approved by the Research Ethics Review Committee of the KCDC (Approval No. 2013-12 EXP-03-5C, 2015-01-02-6C).

Variables and Measurements: Health-Related Quality of Life

Tools to evaluate HRQoL include World Health Organization Quality of Life Brief Version (WHOQOL-BREF) [6], Quality of Well-Being Scale [22], 36-Item Short Form [23], and EQ-5D [24]. Among them, the reliability and validity of the Korean version of EQ-5D reliability and validity have been verified in several studies [25,26].

EQ-5D, a self-report questionnaire, was used for measurement of HRQoL. The survey questions were answered on a three-step scale (no problem, moderate problem, severe problem) with 5 questions: mobility, self-care, usual activity, pain/disability, and anxiety/depression. The KCDC calculated the EQ-5D index using the following formula to express it as a single index.

$$\text{EQ-5D index} = 1-(0.050+0.096\times M2+0.418\times M3+0.046\times SC2+0.136\times SC3+0.051\times UA2+0.208\times UA3+0.037\times PD2+0.151\times PD3+0.043\times AD2+0.158\times AD3+0.050\times N3)$$

Based on previous research methods [8], the EQ-5D index was divided into 2 groups, below average (low quality of life, low QoL) and above average (high quality of life, high QoL).

PA

PA Measured by Accelerometer

The accelerometer used in KNHANES is the wGT3X+ (ActiGraph LLC), and it is a device that converts the acceleration of movement displayed by PA into an electrical signal. The accelerometer and a written consent form were provided to adults 19-64 years old who consented to wear the accelerometer. Measurements were programmed to record from midnight (midnight) on the next day after delivery, and they were instructed to wear the accelerometer on the waist for 7 consecutive days except for swimming, showering, and sleeping.

The following criteria were applied with reference to previous studies [18,27]: (1) data summary cycle (60 seconds); (2) intensity of PA (count per minute [CPM], sedentary behavior <100; 2000 ≤moderate PA ≤5998; vigorous PA ≥5999); (3) determination algorithm of accelerometer wearing or nonwearing time (If CPM is 0 and lasts longer than 60 minutes, it is considered nonwearing time. However, if CPM is less than 100 and lasts less than 2 minutes, it is acceptable); (4) accelerometer minimum wearing time and days (10 hours per day, 4 days per week); and (5) criteria for meeting PA guidelines (600 metabolic equivalent [MET]-minutes/week: [(moderate PA × 4 METs) + (vigorous PA × 8 METs)] ≥ 600 METs)

PA measured with an accelerometer was quantified in two ways: (1) Moderate to vigorous PA accelerometer-total (MVPA-AT, for at least 1 minute), and (2) Moderate to vigorous PA accelerometer-bout (MVPA-AB, for at least 10 minutes; if the time when the corresponding strength number of cutting points has not been reached is less than 2 minutes, it is acceptable). One minute of vigorous PA was counted as 2 minutes of moderate PA.

Self-Reported PA

Self-reported PA was collected by using the Global Physical Activity Questionnaire (GPAQ), divided into 3 categories: leisure, occupation, and commuting PA. Since the PA measured with the accelerometer was collected for most of the time while wearing the accelerometer, all 3 physical activities (leisure, occupation, comminuting) were summed up. After that, the same criteria for the PA measured by accelerometer were applied.

Covariates

Socioeconomic factors such as age, sex, marital status, education level, employment, household income, and residence were investigated. Employment was classified according to occupational reclassification and unemployment and economic inactivity status codes. After that, it was divided into 3 groups: white collar (managers, experts, related workers, and office workers), blue collar (service/sales workers, skilled workers in agriculture, forestry and fisheries, craftsmen, workers in machine operation/assembly, and simple labor workers), and unemployed (housewives, students, etc). Household income level was classified into 2 groups: the bottom 50% (Low) and the top 50% (High) of household income. Residence was classified into *dong* (urban) and *eup*/*myeon* (rural).

Health-related lifestyle factors including smoking, drinking, and average daily sleep time were investigated. Smoking was classified into 3 groups: nonsmoker (person who never smoked or smoked less than 5 packs or 100 cigarettes in their lifetime), past smoker (person who smoked in the past but not now), and smoker (person who currently smokes). Alcohol use was classified into 3 groups according to the WHO high-risk drinking standards: heavy drinker (14 or more drinks per week for men/10 or more drinks per week for women), adequate drinker (annual drinker, not heavy drinker), and abstainer (those who have not drunk alcohol in their lifetime).

Mental health factors including stress perception rate and subjective health status were investigated. Stress was classified into 2 categories with answers to the question “How much stress do you usually feel in your daily life?”: Stressful (I feel a lot, I feel a little), little stress (I hardly feel it). Subjective health status was classified into 3 categories with answers to the question “How do you usually feel about your health?”: good (very good, good), normal (average), and poor (bad, very bad).

Variables related to chronic disease, the prevalence of cardiovascular disease, diabetes, and depression was investigated. Cardiovascular disease includes stroke, angina, and myocardial infarction. The prevalence of each disease was classified into those diagnosed by a doctor.

Statistical Analysis

Data are presented as the mean (SD) for continuous variables, and presented as number and percent for categorical variables. We analyzed the study participants’ characteristics according to the EQ-5D index, using *t* test to compare continuous variables, chi-square test for categorical variables.

Additionally, adjusted odds ratio (OR) and 95% CIs for the risk of low HRQoL according to PA measured by the accelerometer and GPAQ were calculated using logistic regression after adjusting the socioeconomic variables (age, sex, marital status, education, employment, and income) for model 1, covariates in model 1 plus mental health-related variables (stress, subjective health status, and depression) for model 2.

All statistical analyses were performed using STATA (version 18.0; StataCorp) and *P* values of <.05 were considered to indicate statistical significance.

Results

General Characteristics

Table 1 shows the baseline characteristics of the below-average (low QoL) and above-average (high QoL) groups of the EQ-5D index. Age was significantly lower in the high QoL group. In addition, males, singles, highly educated people, office workers, and high-income earners were more common in the high QoL group. Those who usually feel less stress and those whose subjective health was good were more common in the high QoL group. There was no significant difference between the 2 groups in some health-related lifestyle variables such as smoking, drinking, and average sleep time. Among chronic disease-related variables, cardiovascular disease and diabetes were not significantly different between the 2 groups.

Table 1. Baseline characteristics of the study population by quality of life.

Variable	Low QoL[a] (n=293)	High QoL (n=1005)	*P* value[b]
Age in years, mean (SD)	45.43 (11.72)	43.18 (12.34)	.006
Sex, n (%)			<.001
Male	84 (28.67)	408 (40.60)	
Female	209 (71.33)	597 (59.40)	
Marital status, n (%)			<.001
Single	46 (15.70)	216 (21.51)	
Married	216 (73.72)	740 (73.71)	
Separated, divorced, or widowed	31 (10.58)	48 (4.78)	
Education, n (%)			<.001
Middle school or less	67 (22.08)	139 (13.83)	
High school	119 (40.61)	410 (40.80)	
College or more	107 (36.52)	456 (45.37)	
Employment, n (%)			.003
White collar	68 (23.29)	339 (33.80)	
Blue collar	118 (40.41)	361 (35.99)	
Unemployed	106 (36.30)	303 (30.21)	
Household income, n (%)			.009
Low	110 (37.54)	296 (29.51)	
High	183 (62.46)	707 (70.49)	
Residence, n (%)			.45
Urban	237 (80.89)	832 (82.79)	
Rural	56 (19.11)	173 (17.21)	
Smoking, n (%)			.27
Nonsmoker	218 (74.66)	702 (69.85)	
Past smoker	38 (13.01)	151 (15.02)	
Smoker	36 (12.33)	152 (15.12)	
Alcohol use, n (%)			.55
Abstainer	60 (20.55)	234 (23.28)	
Adequate drinker	197 (67.47)	664 (66.07)	
Heavy drinker	35 (11.99)	107 (10.65)	
Sleep duration (h), mean (SD)	6.82 (1.33)	6.86 (1.18)	.60
Stress, n (%)			<.001
Little stressful	176 (60.27)	788 (78.41)	
Stressful	116 (39.73)	217 (21.59)	
Subjective health status, n (%)			<.001
Poor	89 (30.38)	79 (7.86)	
Normal	157 (53.58)	545 (54.23)	
Good	47 (16.04)	381 (37.91)	
Cardiovascular disease, n (%)			.91
No	290 (98.98)	994 (98.91)	
Yes	3 (1.02)	11 (1.09)	
Diabetes, n (%)			.74

Variable	Low QoL[a] (n=293)	High QoL (n=1005)	P value[b]
No	281 (95.90)	968 (96.32)	
Yes	12 (4.10)	37 (3.68)	
Depression, n (%)			<.001
No	267 (91.13)	979 (97.41)	
Yes	26 (8.87)	26 (2.59)	
Sedentary time (min), mean (SD)	3194.75 (766.72)	3258.27 (764.79)	.21

[a]QoL: quality of life.

[b]*P* value is from *t* test for continuous variables and Chi-square test for categorical variables.

PA

Table 2 presents PA measured by the accelerometer and GPAQ in the 2 groups divided by the average of the EQ-5D index. There were significantly more persons who met MVPA-AB ≥600 METs in the high QoL group (*P*<.10). There was no significant difference between high and low QoL groups for MVPA-AT and self-reported PA.

Table 2. Comparison of physical activity by health-related quality of life.

Variable	Low QoL[a] (n=293), n (%)	High QoL (n=1005), n (%)	P value[b]
MVPA[c]-AB[d] (10 min bouts)			.01
<600 METs	243 (82.9)	762 (75.82)	
≥600 METs	50 (17.1)	243 (24.18)	
MVPA-AT[e] (Total bouts)			.27
<600 METs[f]	127 (43.3)	400 (39.80)	
≥600 METs	166 (56.7)	605 (60.20)	
MVPA-S (Self-reported)			.73
<600 METs	127 (43.3)	447 (44.48)	
≥600 METs	166 (56.7)	558 (55.52)	

[a]QoL: Quality of Life.

[b]*P* values obtained using chi-square tests.

[c]MVPA: moderate to vigorous physical activity.

[d]AB: accelerometer-bout.

[e]AT: accelerometer-total.

[f]METs: metabolic equivalents.

Association Between the Domains of PA and HRQoL

In Table 3, logistic regression analysis was performed to determine the correlation between HRQoL and PA measured by the accelerometer and GPAQ, and marked with an OR and 95% CI. Compared with the low QoL group, the crude OR of the HRQoL was 1.55 (95% CI 1.11-2.17) in MVPA-AB ≥600 METs. After adjusting for the socioeconomic variables (model 1), the OR of the HRQoL was 1.60 (95% CI 1.13-2.27). After adjusting model 1 for the mental health–related variables (model 2), the OR was 1.45 (95% CI 1.01-2.09), which weakened the statistical significance compared with model 1. There was no statistical significance between the 2 QoL groups for MVPA-AT and self-reported PA.

Table 3. Association between domains of physical activity and health-related quality of life.

Domains of physical activity	Number	Crude OR[a] (95% CI)	Model 1[b] (95% CI)	Model 2[c] (95% CI)
MVPA[d]-AB[e] (10-min bouts)				
<600 METs[f]	1005	1	1	1
≥600 METs	239	1.55 (1.11-2.17)	1.60 (1.13-2.27)	1.45 (1.01-2.09)
MVPA-AT[g] (Total bouts)				
<600 METs	527	1	1	1
≥600 METs	771	1.15 (0.89-1.51)	1.07 (0.82-1.41)	0.91 (0.68-1.22)
MVPA-S (Self-reported)				
<600 METs	574	1	1	1
≥600 METs	724	0.96 (0.73-1.24)	0.86 (0.66-1.13)	0.79 (0.59-1.06)

[a]OR: odds ratio.

[b]Adjusted for age (continuous), sex, marital status, education, employment, income.

[c]Further adjusted for stress, subjective health status, depression.

[d]MVPA: moderate to vigorous physical activity.

[e]AB: accelerometer-bout.

[f]MET: metabolic equivalent.

[g]AT: accelerometer-total.

Discussion

Principle Findings

This study was designed to confirm the relationship between HRQoL and PA measured by an accelerometer using national survey data. As a result of the analysis, HRQoL was significantly associated with the group that engaged in more MVPA for at least 10 minutes, as measured by the accelerometer. However, there was no statistical significance with HRQoL for MVPA measured for at least 1 minute by the accelerometer. In the previous KNHANES, only self-report questionnaires were used to measure PA [8]. However, in the sixth KNHANES (2014-2015), PA measured with an accelerometer was provided, making it possible to conduct studies [28-30] like ours.

Our findings align with previous research demonstrating the beneficial effects of PA on HRQoL across various populations [31,32]. For example, multiple studies have reported that increased PA is associated with improved HRQoL. A study by Brown et al [16] found that adults who engaged in regular MVPA had significantly higher HRQoL scores compared to those who were inactive. Similarly, a review by Marquez et al [33] synthesized evidence from numerous studies and concluded that PA interventions can effectively enhance HRQoL across various populations. PA has been shown to enhance self-efficacy, physical self-esteem, and positive affect [32,34], which collectively contribute to improvements in HRQoL. Overall, our results are consistent with the majority of previous literature on the positive relationship between PA and HRQoL. This study further supports these findings by using objective measures of PA, thus providing a more accurate assessment of its impact on HRQoL.

This study can serve as a basis for recommending at least 10 minutes of MVPA to improve HRQoL. However, the phrase “at least 10 minutes” was deleted from the PA guidelines published by the US Department of Health and Human Services in 2018, as MVPA for less than 10 minutes can benefit health. Therefore, to compare this with other health-related variables not considered in this study, it is necessary to measure MVPA for at least 1 minute. Nevertheless, since the PA questionnaire used in the KNHANES specifies “usually continued PA for at least 10 minutes during a week,” it is more appropriate to compare it with PA for more than 10 minutes, as measured with an accelerometer in this study [35].

Earlier, it was mentioned that the GPAQ calculates PA that has been continued for at least 10 minutes, so it is appropriate to compare it with PA measured with an accelerometer for at least 10 minutes. However, previous studies [28,36-38] found no correlation between accelerometer data and the questionnaire's satisfaction with the PA guidelines, and the results were not consistent in this study as well. This inconsistency may arise because the questionnaire’s criteria for “PA continued for at least 10 minutes” are subjective or may include PA lasting less than 10 minutes. Additionally, there is a possibility of bias in the self-report questionnaire, as people engaging in MVPA may over-report their activity levels [37-39]. Furthermore, since the accelerometer was provided after filling out the questionnaire, the data measurement periods differ, and the GPAQ may misclassify location movement PA as MVPA without considering movement speed. Therefore, to measure PA more accurately, a specific method that can complement both approaches is necessary.

In previous studies [40-42], stress and subjective health status perception had a negative correlation with HRQoL, and in 1 study [43]; it was reported that stress had the greatest influence on the deterioration of women's QoL. In this study, when socioeconomic variables were corrected for the group with a lot of PA measured with an accelerometer, the statistical significance was stronger than before the correction. But when stress and subjective health status perception were additionally corrected, the degree of statistical significance decreased. Similar to previous studies, this can be interpreted as having a greater effect on stress and subjective health status perception on HRQoL.

Limitations

This study has the following limitations. First, it is difficult to generalize the study results since it is a nonprobability sample composed of study subjects selected through convenience sampling. Second, variables including EQ-5D are investigated through a self-report questionnaire, so there is a possibility of social desirability bias, misclassification bias, and recall bias. Third, since it was conducted as an observational cross-sectional study, it is not possible to know the causal relationship between HRQoL and variables. Nevertheless, it is significant that the study was conducted using national health and nutrition survey data of about 1300 adults aged 19-64. Also, since it is a study using data obtained using an accelerometer, it has an advantage in being able to compare it with other domestic and foreign studies conducted in a similar way. Research using accelerometers began to increase in the 2000s, and since the beginning of 2010, nearly 100 related studies have been published annually [44,45]. In addition, the era has come when it is possible to measure PA using wearable devices such as smartphones and smart watches. It is thought that various follow-up studies using tools that can quantify PA are possible in the future. Based on this, we expect the growth of fields related to PA aimed at promoting health.

Conclusions

Our study shows the association between HRQoL and MVPA, which is significantly higher in the group with more PA for at least 10 minutes, using the KNHANES. The study emphasizes the need for promoting PA and interventions focusing specifically on continuing at least 10 minutes. We expect further interventional studies to focus on the specific PA time period such as occupational PA, leisure-time PA, and commuting PA.

Conflicts of Interest

None declared.

References

1. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Rep 1985;100(2):126-131. [Medline: 3920711]
2. Warburton DER, Nicol CW, Bredin SSD. Health benefits of physical activity: the evidence. CMAJ 2006;174(6):801-809 [FREE Full text] [doi: 10.1503/cmaj.051351] [Medline: 16534088]
3. Lee IM, Paffenbarger RS. Associations of light, moderate, and vigorous intensity physical activity with longevity. The harvard alumni health study. Am J Epidemiol 2000;151(3):293-299. [doi: 10.1093/oxfordjournals.aje.a010205] [Medline: 10670554]
4. Kohl HW, Craig CL, Lambert EV, Inoue S, Alkandari JR, Leetongin G, Lancet Physical Activity Series Working Group. The pandemic of physical inactivity: global action for public health. Lancet 2012;380(9838):294-305 [FREE Full text] [doi: 10.1016/S0140-6736(12)60898-8] [Medline: 22818941]
5. KNHANES. Korea National Health and Nutrition Examination Survey FACT SHEET. 2020. URL: https://knhanes.kdca.go.kr/knhanes/sub04/sub04_04_05.do [accessed 2024-06-26]
6. NA. What quality of life? The WHOQOL group. World health organization quality of life assessment. World Health Forum 1996;17(4):354-356. [Medline: 9060228]
7. Farquhar M. Definitions of quality of life: a taxonomy. J Adv Nurs 1995;22(3):502-508. [doi: 10.1046/j.1365-2648.1995.22030502.x] [Medline: 7499618]
8. Kim H, Kwon H, Jang M. The association between health-related quality of life and leisure-time, occupational and commuting physical activity in Korean workers: the sixth Korea national health and nutrition examination survey, 2014 and 2015. Korean J Fam Pract 2018;8(4):616-622.
9. Scarabottolo CC, Tebar WR, Gobbo LA, Ohara D, Ferreira AD, da Silva Canhin D, et al. Analysis of different domains of physical activity with health-related quality of life in adults: 2-year cohort. Health Qual Life Outcomes 2022;20(1):71 [FREE Full text] [doi: 10.1186/s12955-022-01981-3] [Medline: 35488344]
10. Puciato D, Rozpara M, Borysiuk Z. Physical activity as a determinant of quality of life in working-age people in Wroclaw, Poland. Int J Environ Res Public Health 2018;15(4):623 [FREE Full text] [doi: 10.3390/ijerph15040623] [Medline: 29596312]
11. Groessl EJ, Kaplan RM, Rejeski WJ, Katula JA, Glynn NW, King AC, et al. Physical activity and performance impact long-term quality of life in older adults at risk for major mobility disability. Am J Prev Med 2019;56(1):141-146 [FREE Full text] [doi: 10.1016/j.amepre.2018.09.006] [Medline: 30573142]
12. Perez-Cruzado D, Cuesta-Vargas AI, Vera-Garcia E, Mayoral-Cleries F. The relationship between quality of life and physical fitness in people with severe mental illness. Health Qual Life Outcomes 2018;16(1):82 [FREE Full text] [doi: 10.1186/s12955-018-0909-8] [Medline: 29720196]

13. Xu H, Tang L, Hu Z, Gao F, Yang Y, Qin L, et al. Association between physical activity and health-related quality of life in elderly individuals with pre-diabetes in rural Hunan Province, China: a cross-sectional study. BMJ Open 2018;8(4):e019836 [FREE Full text] [doi: 10.1136/bmjopen-2017-019836] [Medline: 29654016]
14. Gavilán-Carrera B, Segura-Jiménez V, Acosta-Manzano P, Borges-Cosic M, Álvarez-Gallardo IC, Delgado-Fernández M. Patterns of sedentary time and quality of life in women with fibromyalgia: cross-sectional study from the al-Ándalus project. JMIR Mhealth Uhealth 2020;8(3):e14538 [FREE Full text] [doi: 10.2196/14538] [Medline: 32191211]
15. Aune D, Markozannes G, Abar L, Balducci K, Cariolou M, Nanu N, et al. Physical activity and health-related quality of life in women with breast cancer: a meta-analysis. JNCI Cancer Spectr 2022;6(6):pkac072 [FREE Full text] [doi: 10.1093/jncics/pkac072] [Medline: 36474321]
16. Brown DR, Carroll DD, Workman LM, Carlson SA, Brown DW. Physical activity and health-related quality of life: US adults with and without limitations. Qual Life Res 2014;23(10):2673-2680 [FREE Full text] [doi: 10.1007/s11136-014-0739-z] [Medline: 24952110]
17. Bădicu G. Physical activity and health-related quality of life in adults from Brasov, Romania. Education Sciences 2018;8(2):52. [doi: 10.3390/educsci8020052]
18. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3] [Medline: 18091006]
19. Esliger DW, Tremblay MS. Physical activity and inactivity profiling: the next generation. Can J Public Health 2007;98 Suppl 2:S195-S207. [Medline: 18213949]
20. Lee I, Shiroma EJ. Using accelerometers to measure physical activity in large-scale epidemiological studies: issues and challenges. Br J Sports Med 2014;48(3):197-201 [FREE Full text] [doi: 10.1136/bjsports-2013-093154] [Medline: 24297837]
21. Doherty A, Jackson D, Hammerla N, Plötz T, Olivier P, Granat MH, et al. Large scale population assessment of physical activity using wrist worn accelerometers: the UK Biobank study. PLoS One 2017;12(2):e0169649 [FREE Full text] [doi: 10.1371/journal.pone.0169649] [Medline: 28146576]
22. Andresen EM, Rothenberg BM, Kaplan RM. Performance of a self-administered mailed version of the quality of well-being (QWB-SA) questionnaire among older adults. Med Care 1998;36(9):1349-1360. [doi: 10.1097/00005650-199809000-00007] [Medline: 9749658]
23. Ware JE. SF-36 health survey update. Spine (Phila Pa 1976) 2000;25(24):3130-3139. [doi: 10.1097/00007632-200012150-00008] [Medline: 11124729]
24. Kind P, Brooks R, Rabin R. EQ-5D concepts and methods : a developmental history. Netherlands: Springer Dordrecht; 2005:1-240.
25. Kim M, Cho Y, Uhm W, Kim S, Bae S. Cross-cultural adaptation and validation of the Korean version of the EQ-5D in patients with rheumatic diseases. Qual Life Res 2005;14(5):1401-1406. [doi: 10.1007/s11136-004-5681-z] [Medline: 16047514]
26. Lim N, Lee I, Lee E, Lee K, Cho K, Rhee S, et al. A validation study of EQ-5D in the patients with osteoarthritis. Journal of muscle and joint health 2010;17(2):203-211. [doi: 10.5953/jmjh.2010.17.2.203]
27. Lee M, Lee H, Choi J. Error rates of prediction equations and cut-points of actigraph GT3X+. Korean Soc Meas Eval Phys Edu Sports Sci 2016;18(1):17-29. [doi: 10.21797/ksme.2016.18.1.002]
28. Lim J, Sung H, Lee O, Kim Y. Physical activity in South Korea measured by accelerometer: The Korea national health and nutrition examination survey VI 2014-2015. Korean J Sport Sci 2020;31(2):169-179. [doi: 10.24985/kjss.2020.31.2.169]
29. Shin SW, Kim H, Kang YH, Kim J. Accelerometer-derived physical activity and sedentary behavior patterns among Korean adults. Phys Act Nutr 2023;27(2):25-33 [FREE Full text] [doi: 10.20463/pan.2023.0015] [Medline: 37583069]
30. Park J, Nam H, Cho S. Association between accelerometer-derived physical activity and depression: a cross-sectional study using isotemporal substitution analysis. BMJ Open 2024;14(4):e078199 [FREE Full text] [doi: 10.1136/bmjopen-2023-078199] [Medline: 38684272]
31. Rejeski WJ, Mihalko SL. Physical activity and quality of life in older adults. J Gerontol A Biol Sci Med Sci 2001;56 Spec No 2:23-35. [doi: 10.1093/gerona/56.suppl_2.23] [Medline: 11730235]
32. Joseph RP, Royse KE, Benitez TJ, Pekmezi DW. Physical activity and quality of life among university students: exploring self-efficacy, self-esteem, and affect as potential mediators. Qual Life Res 2014;23(2):659-667 [FREE Full text] [doi: 10.1007/s11136-013-0492-8] [Medline: 23928820]
33. Marquez DX, Aguiñaga S, Vásquez PM, Conroy DE, Erickson KI, Hillman C, et al. A systematic review of physical activity and quality of life and well-being. Transl Behav Med 2020;10(5):1098-1109 [FREE Full text] [doi: 10.1093/tbm/ibz198] [Medline: 33044541]
34. Elavsky S, McAuley E, Motl RW, Konopack JF, Marquez DX, Hu L, et al. Physical activity enhances long-term quality of life in older adults: efficacy, esteem, and affective influences. Ann Behav Med 2005;30(2):138-145. [doi: 10.1207/s15324796abm3002_6] [Medline: 16173910]
35. Keating XD, Zhou K, Liu X, Hodges M, Liu J, Guan J, et al. Reliability and concurrent validity of global physical activity questionnaire (GPAQ): a systematic review. Int J Environ Res Public Health 2019;16(21):4128 [FREE Full text] [doi: 10.3390/ijerph16214128] [Medline: 31717742]

36. Skender S, Ose J, Chang-Claude J, Paskow M, Brühmann B, Siegel EM, et al. Accelerometry and physical activity questionnaires - a systematic review. BMC Public Health 2016;16:515 [FREE Full text] [doi: 10.1186/s12889-016-3172-0] [Medline: 27306667]
37. Dyrstad SM, Hansen BH, Holme IM, Anderssen SA. Comparison of self-reported versus accelerometer-measured physical activity. Med Sci Sports Exerc 2014;46(1):99-106. [doi: 10.1249/MSS.0b013e3182a0595f] [Medline: 23793232]
38. Cerin E, Cain KL, Oyeyemi AL, Owen N, Conway TL, Cochrane T, et al. Correlates of agreement between accelerometry and self-reported physical activity. Med Sci Sports Exerc 2016;48(6):1075-1084 [FREE Full text] [doi: 10.1249/MSS.0000000000000870] [Medline: 26784274]
39. Lee J, Lee C, Min J, Kang D, Kim J, Yang HI, et al. Development of the Korean global physical activity questionnaire: reliability and validity study. Glob Health Promot 2020;27(3):44-55. [doi: 10.1177/1757975919854301] [Medline: 31375056]
40. Kim E, Oh K, Ha E. The vulnerability of married women on depression: focused on life stress and coping processes. Korean J Woman Psychol 1999;4(1):1-14.
41. Grasaas E, Skarstein S, Mikkelsen HT, Småstuen MC, Rohde G, Helseth S, et al. The relationship between stress and health-related quality of life and the mediating role of self-efficacy in Norwegian adolescents: a cross-sectional study. Health Qual Life Outcomes 2022;20(1):162 [FREE Full text] [doi: 10.1186/s12955-022-02075-w] [Medline: 36482450]
42. Okwuosa LN, Onu DU, Onyedibe MC. Perceived stress and health-related quality of life in cancer patients: the mediating role of religious coping. Curr Psychol 2023;43(4):3166-3174. [doi: 10.1007/s12144-023-04510-7]
43. Jun EM, Cho DS. The relationship of symptoms of stress, cancer prevention behavior and the quality of life in women. Korean J Women Health Nurs 2005;11(2):156-162. [doi: 10.4069/kjwhn.2005.11.2.156]
44. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. Br J Sports Med 2014;48(13):1019-1023 [FREE Full text] [doi: 10.1136/bjsports-2014-093546] [Medline: 24782483]
45. Zeng Y, Lin X, Zhang Z, Chen Y, Fan G, Guo H. Visual atlas analysis of acceleration sensor research literature based on citespace bibliometrics. SET 2019;6(1):70. [doi: 10.11114/set.v6i1.4209]

Abbreviations

CPM: count per minute
GPAQ: Global Physical Activity Questionnaire
HRQoL: health-related quality of life
KCDC: Korea Centers for Disease Control and Prevention
KNHANES: Korea National Health and Nutrition Examination Survey
MET: metabolic equivalent
MVPA-AB: moderate to vigorous physical activity with an accelerometer for at least 10 minutes
MVPA-AT: moderate to vigorous physical activity with an accelerometer for at least 1 minute
MVPA-S: moderate to vigorous physical activity with self-report questionnairePA: physical activity
OR: odds ratio
QoL: quality of life
WHOQOL-BREF: World Health Organization Quality of Life Brief Version

Edited by A Kushniruk; submitted 22.04.24; peer-reviewed by A AL-Asadi; comments to author 21.05.24; revised version received 04.06.24; accepted 04.08.24; published 03.09.24.

Please cite as:
Han S, Oh B, Kim HJ, Hwang SE, Kim JS
Accelerometer-Based Physical Activity and Health-Related Quality of Life in Korean Adults: Observational Study Using the Korea National Health and Nutrition Examination Survey
JMIR Hum Factors 2024;11:e59659
URL: https://humanfactors.jmir.org/2024/1/e59659
doi:10.2196/59659
PMID:39226099

Original Paper

Understanding Morning Emotions by Analyzing Daily Wake-Up Alarm Usage: Longitudinal Observational Study

Kyue Taek Oh[1*], PhD; Jisu Ko[2*], MS; Nayoung Jin[2], BS; Sangbin Han[3], MD; Chan Yul Yoon[4], BBA; Jaemyung Shin[4], MS; Minsam Ko[1], PhD

[1]Department of Human-Computer Interaction, University of Hanyang, Ansan, Gyeonggi-do, Republic of Korea

[2]Department of Applied Artificial Intelligence, University of Hanyang, Ansan, Republic of Korea

[3]Seoul Asan Medical Center, Seoul, Republic of Korea

[4]Delightroom, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Minsam Ko, PhD
Department of Human-Computer Interaction
University of Hanyang
ERICA Campus, 55 Hanyangdeahak-ro, Sangnok-gu
Ansan, Gyeonggi-do, 15588
Republic of Korea
Phone: 82 31 400 1071
Email: minsam@hanyang.ac.kr

Abstract

Background: Morning emotions can significantly affect daily wellness. While many studies have analyzed daily survey responses to identify factors influencing morning emotions, these methods require additional time and effort from individuals for emotional monitoring.

Objective: This study aims to identify daily alarm usage patterns related to morning emotions.

Methods: We recruited 373 users of the Alarmy app (DelightRoom) in the United States and South Korea and surveyed their demographics and usual behaviors related to morning emotions. Participants described their morning emotions over a 2-week period, during which we collected daily alarm app logs. We used a generalized estimating equation (GEE) method to identify factors affecting morning emotions.

Results: The findings indicate that varied alarm usage is related to morning emotions. Alarm set time was positively associated with feelings of peacefulness and refreshment in the morning, while task-based alarms were related to nervousness. The time taken to deactivate the alarm after it rang was negatively correlated with happiness. In addition, usual behaviors and demographic factors were found to be related to morning emotions, consistent with previous studies.

Conclusions: The study reveals that daily alarm usage is related to morning emotions, suggesting that daily alarm logs can supplement survey methods to facilitate daily emotion monitoring.

(JMIR Hum Factors 2024;11:e50835) doi:10.2196/50835

KEYWORDS

morning emotion; wake-up alarm usage; morning context; emotion monitoring; longitudinal observational study

Introduction

Background

The mental or emotional state of an individual in the morning significantly impacts daily wellness and is associated with 6 dimensions of holistic health: physical, mental, spiritual, emotional, social, and environmental health [1]. Rothbard and Wilk [2] identified a correlation between employees' morning moods and their emotional states throughout the day. Employees who began their day with positive emotions tended to maintain those emotions, positively influencing their interactions with customers. Conversely, morning fatigue adversely affects daily life by reducing the quantity and intensity of physical activities [3]. Studies suggest that daily performance can be enhanced by activities that positively influence one's morning emotional state, such as listening to music [4] or taking a shower [5].

Numerous studies have attempted to understand and modify morning emotions. For instance, Sonnentag et al [6] reported that evening relaxation could promote positive emotional states the following morning. Nandy et al [7] found that regular exercise can enhance positive morning emotions. Among the various factors influencing morning mood, sleep quality is paramount. High-quality sleep is significantly associated with the suppression of negative emotions and the maintenance of positive emotions into the next day [8,9]. Early bedtimes have been shown to improve social interactions and emotional experiences [10]. Furthermore, the mood experienced in dreams can affect the emotional state the next morning [11].

However, to our knowledge, previous studies have not sufficiently addressed the impact of morning context—specific circumstances and environments experienced during the morning hours—on morning emotions. The morning period possesses unique characteristics, such as sleep inertia and preparation for daily routines, distinguishing it from other times of the day. Therefore, understanding awakening contexts could provide deeper insights into morning emotions. Furthermore, contextual factors are generally recognized as important in understanding specific behaviors [12-14]. Similarly, Oh et al [15] highlighted the necessity of considering morning contexts to comprehend changes in morning behavior.

Objectives

This study aimed to explore the contextual factors that affect morning emotions. To achieve this, we analyzed alarm usage logs to evaluate the contextual information surrounding the moment of waking up. An alarm usage log reflects various aspects of waking behavior, such as the time and regularity of waking. In addition, monitoring and analyzing alarm usage logs is relatively easy because no additional user action is required. Previous studies to understand morning emotions have mainly relied on survey data that requires responses from the user. This study investigates the feasibility of a new method for monitoring morning emotions based on alarm logs by analyzing the relationship between alarm usage log data and morning emotions. If a significant relationship is found, this method could be expanded to monitor morning emotions using alarm usage log data. This approach could complement traditional methods that rely heavily on surveys, potentially providing a more efficient and unobtrusive way to understand and track morning emotions.

Methods

Participants

We recruited participants for this study from users of the Alarmy app (DelightRoom), focusing on the United States and South Korea, where the app is highly popular. Recruitment posts detailing the experiment's background, objectives, and procedures, along with a web URL for the application, were disseminated through the app's notice board and push notifications. In total, 373 subjects were recruited: 232 from the United States and 141 from South Korea.

The decision to use Alarmy as the focal application for this study was based on several compelling reasons. First, Alarmy is one of the most downloaded alarm apps globally, making it a widely used tool across diverse demographics and cultural contexts. Second, its unique features, such as mission-based alarms and loud alarm options, allow for a comprehensive analysis of wake-up behaviors and morning routines, providing insights that extend beyond basic alarm functions.

A key aspect of this study was to ensure that the findings were representative of the broader population of alarm users. To address potential concerns about generalizability, we compared the demographic characteristics of our sample to the larger Alarmy user population. The comparison revealed no significant differences in major usage statistics between our dataset and those used in previous studies analyzing large-scale Alarmy usage logs [16,17]. Furthermore, the age and gender distribution of our sample closely mirrored that of the overall Alarmy user base, suggesting that our findings are robust and applicable to a wide range of users.

Furthermore, a collaborative research agreement with Alarmy facilitated access to detailed log data, enhancing the robustness of the study. This agreement allowed us to leverage a large-scale dataset, ensuring that our sample was representative of the entire user base. Consequently, while the study exclusively recruited users of the Alarmy app, the demographic similarities between our sample and the broader population of Alarmy users support the generalizability of our conclusions.

All subjects participated voluntarily and received compensation of US $43 for their commitment. The compensation amount was carefully chosen to reflect the effort needed for daily participation over 2 weeks and the specific requirement to respond during morning hours. This ensured adequate participant motivation while minimizing any potential bias due to compensation. The study's compensation strategy was designed to provide sufficient incentive to maintain participant engagement without compromising the integrity of the results. Evidence suggests that appropriately calibrated compensation is unlikely to bias experimental outcomes. According to a review by Singer and Couper [18], empirical studies typically find little evidence of payment influencing participants' behavior in a way that significantly affects the results, provided that the payment is not coercively large. Our compensation of US $43 was benchmarked against typical study payments for similar durations and task demands, ensuring it was reasonable yet not overly influential.

Experimental Procedure

We conducted a longitudinal observational study to explore the relationship between daily waking behaviors and morning emotions. Given that an individual's demographics and characteristics could influence morning emotions, we categorized factors into three groups: (1) demographic, (2) usual, and (3) daily. Demographic factors included age and gender, while usual factors encompassed self-assessments of one's typical morning state. Daily factors were derived from analyzing alarm app usage. As dependent variables, we specified 9 emotions commonly experienced in the morning—hopeful, happy, peaceful, refreshed, annoyed, tired, depressed, nervous, and neutral—along with a morning wellness score. These variables were measured through a presurvey, daily surveys, and alarm app usage logs.

The main experiment proceeded with participants first completing an 18-question presurvey covering demographics and their usual morning state. Over the course of 2 weeks, participants responded to daily surveys about their morning emotions and the time they went to sleep the previous night. These surveys were implemented and conducted using TypeForm, a web-based platform. To ensure accurate data collection on morning emotions, the initial survey link was sent 30 minutes after the participant's alarm was set. If there was no response, a follow-up link was sent at 10 AM. This method aimed to capture participants' emotions soon after waking while providing a second opportunity for those who missed the initial prompt. Alarm app usage data during the study period was shared for analysis.

After 2 weeks, the experimental procedure concluded, and participants were compensated for their involvement. This comprehensive approach, combining presurvey data, daily surveys, and usage logs, allowed for an in-depth analysis of the relationship between waking behaviors and morning emotions.

Independent Variables

The independent variables in this study were classified into three categories: (1) demographic, (2) usual, and (3) daily. Table 1 provides detailed descriptions of each independent variable within these categories.

Table 1. Descriptions of independent variables.

Factors	Variable	Description
Demographics		
Sex	demo_gender	Sex of the participant (male=0, female=1)
Age	demo_age	Age of the participant
Country	demo_country	Country of residence (United States=0, South Korea=1)
Usual		
Morning state	usual_state_alert	Degree of alertness after waking up (Not at all alerts=1, Slightly alert=2, Fairly alert=3, Very alert=4)
Morning state	usual_state_tired	Degree of tiredness after waking up (Very refreshed=1, Fairly refreshed=2, Slightly tired=3, Very tired=4)
Morning state	usual_state_score	Usual morning condition, rated on a 1-7 scale, assessed at the start of the experiment.
Habit	usual_habits_morningActivity	Habit of engaging in positive morning activities, such as exercise or meditation (No=0, Yes=1)
Theory of Planned Behavior	usual_tpb_attitude	Attitude toward morning waking (Strongly disagree=1, Strongly agree=7)
Theory of Planned Behavior	usual_tpb_ subjectiveNorm	Subjective norm toward morning waking (Strongly disagree=1, Strongly agree=7)
Theory of Planned Behavior	usual_tpb_intention	Intention toward morning waking (Strongly disagree=1, Strongly agree=7)
Theory of Planned Behavior	usual_tpb_control	Control toward morning waking (Strongly disagree=1, Strongly agree=7)
Daily		
Alarm usage	daily_usage_dayOfWeek	Alarm usage on weekdays vs. weekends (weekday=0, weekends=1)
Alarm usage	daily_usage_ringCount	Number of alarms used to wake up in the morning
Alarm usage	daily_usage_ring2dismiss	Time elapsed from the alarm ringing to deactivation
Alarm usage	daily_usage_continuity	Number of days the alarm was used in the past week
Alarm usage	daily_first_ringTime	Time of the first alarm in the morning
Alarm usage	daily_last_typeInBed	Percentage of wake-up tasks performed in bed after the last alarm
Alarm usage	daily_last_typeOutOfBed	Percentage of wake-up tasks performed out of bed after the last alarm
Alarm usage	daily_last_isLoud	Use of a loud sound for the last alarm
Alarm usage	daily_last_isLabel	Use of a label for the last alarm
Time to sleep	daily_sleep_time	Bedtime the previous night

Demographic Factors

A total of 3 demographic variables were used: age, gender, and country, denoted as "demo_age," "demo_gender," and "demo_country," respectively. Gender, being a categorical variable, was represented using a single dummy variable (male=0, female=1). Similarly, the participant's country was coded as a binary variable (United States=0, South Korea=1). All demographic variables were measured in a preliminary survey.

Usual Factors

We examined variables related to the usual morning state through a presurvey, assessing three types of individual characteristics: (1) usual morning state, (2) habits, and (3) planned behavior.

Participants' self-assessments of their usual morning state included aspects such as alertness and tiredness. These were measured using 4-point Likert scale questions adapted from the morning-evening wellness questionnaire [19]. Specifically, participants were asked, "How alert do you feel during the first half-hour after you wake up in the morning?" and "During the first half-hour after you wake up in the morning, how tired do you feel?" These responses were designated as "usual_state_alert" and "usual_state_tired," respectively. In addition, participants rated their overall sense of well-being in the morning on a 7-point scale ("Rate your usual morning on a 7-point scale."), which we denoted as "usual_state_score."

We also considered the impact of participants' morning habits on their emotions, as indicated by previous studies [7]. Participants reported their engagement in frequent dynamic activities (eg, exercise) or static activities (eg, meditation) in the morning. These activities were recorded as "usual_habits_morningActivity."

Finally, we explored behavioral factors based on the theory of planned behavior (TPB) [20]. TPB variables are widely used to describe purposive human behaviors [21-23]. We developed a questionnaire with 4 questions on a 7-point Likert scale to evaluate participants' usual target waking hours and morning behavioral patterns. The questions included: "Waking up on time in the morning is important to me," "I intend to wake up on time in the morning," "Other people think that I should wake up earlier in the morning," and "I am confident that I am capable of waking up on time in the morning." The responses to these questions were denoted as "usual_tpb_attitude," "usual_tpb_intention," "usual_tpb_subjectiveNorm," and "usual_tpb_control," respectively.

Daily Factors

In this study, we analyzed the Alarmy app usage logs of participants during the experiment period to track their daily morning behavior. Given the reported differences in sleep patterns between weekdays and weekends [24], we denoted the days of the week with the variable "daily_usage_dayOfWeek," coding weekdays as 0 and weekends as 1. Holidays were excluded from this classification. To examine the continuity of alarm use, the number of days the alarm was used in the previous week was recorded as "daily_usage_continuity." In addition, we gathered information on how quickly users woke up in the morning. The number of alarms required to wake them up was denoted as "daily_usage_ringCount," and the time from the ringing of the first alarm to its deactivation was denoted as "daily_usage_ring2dismiss."

We further focused on the first and last alarms used in the morning. The time of the first alarm, indicating the targeted waking time, was marked as "daily_first_ringTime." We also recorded the type of tasks required to dismiss the alarm. Alarms that could be deactivated without moving out of bed were indicated as "daily_last_typeInBed," while those requiring movement out of bed were indicated as "daily_last_typeOutOfBed." The use of a loud alarm sound for the last alarm was denoted as "daily_last_isLoud," and the labeling of the last alarm was indicated by "daily_last_isLabel," as the alarm label is the first text the user notices upon waking.

Finally, we included the time at which participants went to bed the night before. The ordinal variable "daily_sleep_time" was divided into 30-minute increments from 9 PM to 3 AM, with a total of 14 values numbered from 0 to 13.

Dependent Variables: Morning Emotions

We collected overall condition scores on a 7-point scale each morning using the question, "Rate your morning on a 7-point scale," and used these responses as the variable "daily_overall_state." In addition, we considered 9 specific categories for morning emotions. Previous studies have defined emotions in various ways. For instance, Ekman [25] identified 6 basic emotions, while Plutchik [26] presented a detailed diagram of 8 core emotions. Russell [27] developed a circular emotion model using arousal and valence as the two axes. To specify the emotions people feel upon waking, we conducted a pilot survey with 29 participants before the main experiments. Respondents were asked to describe their feelings in the morning hours, and their responses were classified into 9 representative emotions: 4 positive emotions (hopeful, happy, peaceful, and refreshed) and 4 negative emotions (annoyed, tired, depressed, and nervous). The ninth category was "no emotion" (none).

Analysis Methods

This study used generalized estimating equations (GEE) analysis [28] to effectively account for temporal changes in morning emotions. The GEE method is a widely used statistical technique for analyzing variables that are repeatedly measured over time [29,30]. While repeated-measures ANOVA is a common approach for analyzing repeated measurement data, the GEE method was deemed more appropriate for this study due to the presence of more than 2 independent variables.

Ethical Considerations

To ensure participant confidentiality, all study data were deidentified. In addition, the experiment received an exemption for informed consent from the institutional review board of Hanyang University (HYUIRB-202205-011).

Results

User Statistics

Table 2 presents the descriptive statistics of the independent variables. Initially, we collected demographic information from participants based on 3 factors. The gender ratio was 49.6% (185/373 male participants). The average age of the participants was 23.94 (SD 9.65) years. Of the participants, 62.2% (232/373) resided in the United States, while the remainder lived in South Korea.

Table 2. Statistics of independent variables.

Factor	Variable	Mean (base)	SD
Demographic			
Sex	demo_gender	0.49 (male)	0.50
Age	demo_age	23.94	9.65
Country	demo_country	0.62 (US)	0.48
Usual			
Morning state	usual_state_alert	2.27	0.86
Morning state	usual_state_tired	3.30	0.68
Morning state	usual_state_score	4.29	1.39
Habit	usual_habit_morningActivity	0.60	0.48
Theory of planned behavior	usual_tpb_attitude	6.36	1.07
Theory of planned behavior	usual_tpb_subjectiveNorm	5.72	1.62
Theory of planned behavior	usual_tpb_intention	6.45	1.03
Theory of planned behavior	usual_tpb_control	4.43	1.86
Daily			
Alarm usage	daily_usage_dayOfWeek	0.79 (Weekday)	0.24
Alarm usage	daily_usgae_ringCount	1.72	1.46
Alarm usage	daily_usage_ring2dismiss	721.10	1496.27
Alarm usage	daily_usage_continuity	3.78	1.67
Alarm usage	daily_first_ringTime	25900.05	4570.43
Alarm usage	daily_last_typeInBed	0.51	0.46
Alarm usage	daily_last_typeOutOfBed	0.16	0.35
Alarm usage	daily_last_isLoud	0.05 (True)	0.18
Alarm usage	daily_last_isLabel	0.29 (True)	0.43
Time to sleep	daily_sleep_time	7.55	3.06

We then examined usual lifestyle factors related to the morning state using a 7-point Likert scale. In the self-assessment of their usual morning condition within 30 minutes of waking, the overall health condition averaged 4.29 out of 7. The scores for alertness and tiredness were 2.27 and 3.30 out of 4, respectively. In addition, 60.3% (225/373) of participants reported that they typically engage in active exercises in the morning. The average scores for the four variables in the TPB category were as follows: attitude, 6.36; subjective norm, 5.72; intention, 6.45; and control, 4.43.

We also collected data on daily user behavior based on alarm use. Overall, 79% of the total usage occurred on weekdays, with participants using alarms an average of 3.787 days per week. On days when alarms were used, an average of 1.7 alarms were set due to participants either snoozing or setting multiple alarms. Approximately 12 minutes passed before deactivating the alarm. The average ring time of the first alarm of the day was 7:11:40 AM (25,900 seconds). The most commonly used alarm mission was the touch-a-button mission, with participants using this feature an average of 32.4% of the time. On average, 51.2% of wake-up tasks were performed in bed (eg, typing, math, memory, and shaking), while 16.3% of tasks required getting out of bed (eg, walking, squats, photo, and barcode). In addition, on average, participants selected loud alarm sounds 5.8% of the time and used alarm labels 29.9% of the time. Regarding usual bedtime, most respondents reported going to bed between 12 AM and 12:30 AM.

GEE Analysis Results

Our GEE analysis identified unique wake-up context factors for each emotion. The comprehensive results of the GEE analysis are presented in Multimedia Appendix 1. In this section, we describe the variables that showed statistically significant results, including the odds ratio (OR), 95% CI of the original GEE coefficient values, and *P* values.

Daily Overall State

Our analysis identified several variables significantly associated with the overall morning condition (daily_overall_state). Fewer alarms used per day (daily_usage_ringCount: OR 0.901, 95% CI –0.170 to –0.039; *P*=.002) and a greater number of days with alarm usage (daily_usage_continuity: OR 1.063, 95% CI 0.005-0.119; *P*=.03) were linked to higher overall scores. This suggests that consistency in alarm usage and reduced dependence on multiple alarms can enhance morning well-being.

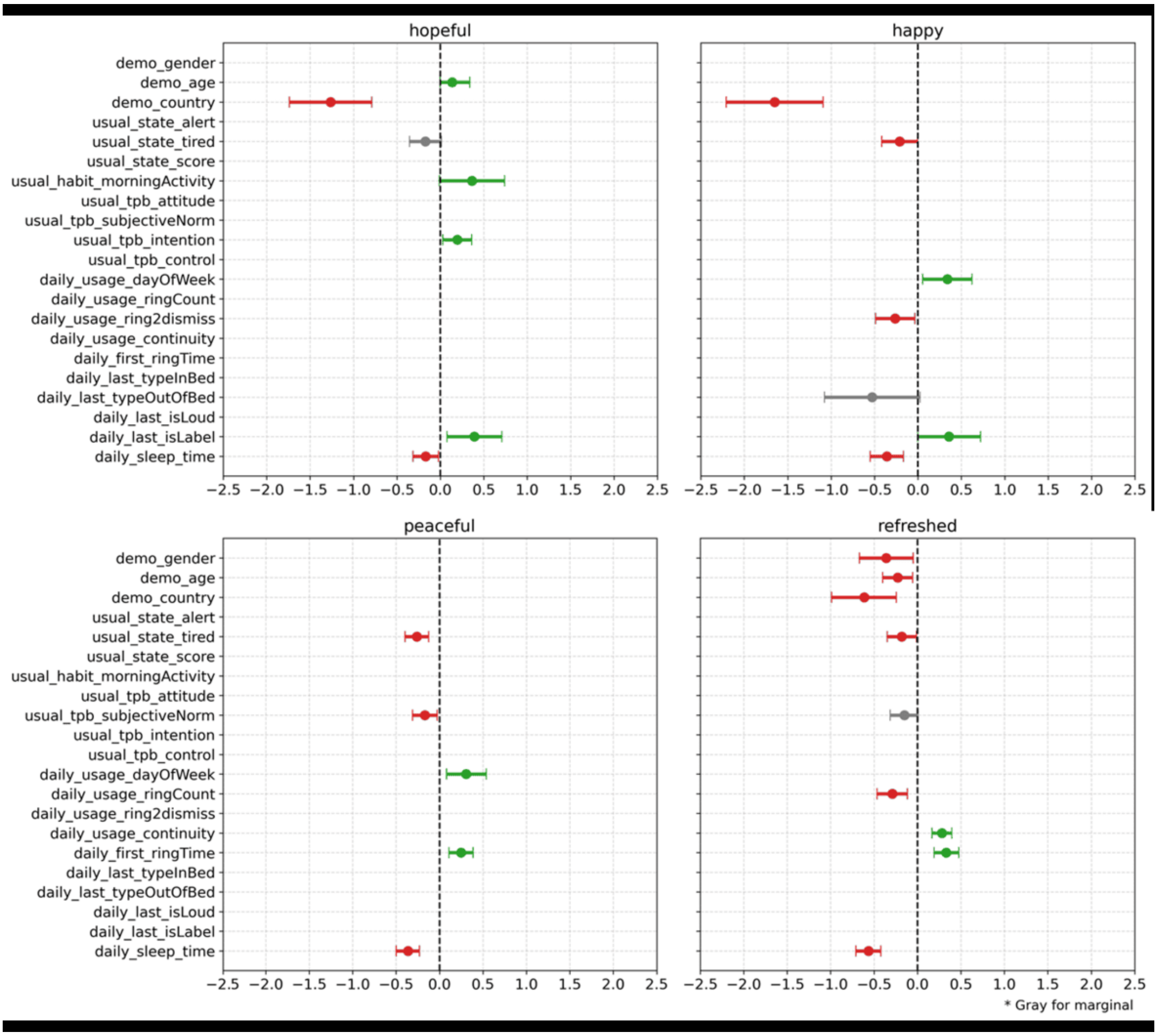
The time of the first alarm (daily_first_ringTime) showed a positive correlation with the overall state (OR 1.133, 95% CI 0.052-0.199; *P*=.001), indicating that waking up later in the morning can contribute to a better overall morning state. Conversely, bedtime (daily_sleep_time) exhibited a negative relationship with the overall score (OR 0.743, 95% CI –0.370 to –0.224; *P*<.001), suggesting that going to bed earlier significantly improves morning well-being.

Furthermore, individuals who perceived their usual morning state positively (usual_state_score) tended to have better daily morning states (daily_overall_state) (OR 1.437, 95% CI 0.208-0.519; *P*<.001), highlighting the importance of a consistently good morning routine. Among the TPB variables, only attitude (usual_tpb_attitude) showed a positive correlation with the overall morning state of the day (OR 1.202, 95% CI 0.050-0.319, *P*=.007), implying that a positive outlook toward waking up is beneficial for morning well-being.

Positive Emotions

Figure 1 displays the GEE analysis results for positive emotions, including the β coefficient and CI for each significant independent variable.

Figure 1. Results of generalized estimating equations analysis on positive emotions.



Hopefulness

We identified several significant factors influencing the emotion of hopefulness in the morning. Using a label on the last alarm (daily_last_isLabel) increased the odds of feeling hopeful (OR 1.479, 95% CI 0.076-0.709; *P*=.01), suggesting that personalized alarm labels can positively impact morning emotions.

Conversely, a late bedtime (daily_sleep_time) the previous day decreased the odds of feeling hopeful (OR 0.845, 95% CI –0.313 to –0.023; *P*=.02), indicating the importance of an early bedtime for a positive morning start. Age was positively correlated with hopefulness (demo_age: OR 1.181, 95% CI –0.004 to 0.339], *P*=.05), and Korean users had significantly lower odds of feeling hopeful compared with American users (demo_country: OR 0.282, 95% CI –1.738 to –0.789; *P*<.001). Feeling tired after waking (usual_state_tired) negatively influenced hopefulness (OR 0.842, 95% CI –0.353 to 0.011; *P*=.06), whereas engaging in physical activities in the morning (usual_habits_morningActivity) tended to increase hopefulness (OR 1.437, 95% CI –0.013-0.740; *P*=.05). Among the TPB factors, only intention (usual_tpb_intention) significantly increased the odds of feeling hopeful (OR 1.215, 95% CI 0.028-0.362; *P*=.02).

Happiness

The use of alarms significantly affected morning happiness. The elapsed time from the first alarm ring to its deactivation (daily_usage_ring2dismiss) was negatively associated with happiness (OR 0.768, 95% CI –0.489 to –0.036; *P*=.02), suggesting that quicker responses to alarms may improve morning mood. Alarm use on weekends (daily_usage_dayOfWeek) was 1.4 times more likely to increase happiness compared to weekdays (OR 1.402, 95% CI 0.054-0.623; *P*=.02). This may be because individuals who maintain a regular rhythm without distinguishing between weekdays and weekends tend to feel happier. Performing wake-up tasks out of bed after the last alarm (daily_last_typeOutOfBed) marginally decreased happiness (OR 0.590, 95% CI –1.077 to 0.023; *P*=.06), implying that remaining in bed for initial tasks might contribute to better mood. Using a label on the last alarm (daily_last_isLabel) was positively correlated with happiness (OR 1.430, 95% CI –0.002 to 0.719; *P*=.05), while a late bedtime (daily_sleep_time) significantly decreased happiness (OR 0.699, 95% CI –0.548 to –0.166; *P*<.001), highlighting the importance of an early bedtime for positive morning emotions. Morning tiredness (usual_state_tired) significantly reduced happiness (OR 0.809, 95% CI –0.418 to –0.005; *P*=.04), indicating the negative impact of fatigue on morning mood.

Peacefulness

Using alarms on weekends (daily_usage_dayOfWeek) was more likely to result in peaceful emotions in the morning compared to weekdays (OR 1.360, 95% CI 0.080-0.538; *P*=.008). Setting the alarm for a later time (daily_first_ringTime) was associated with increased feelings of peace (OR 1.281, 95% CI 0.109-0.388; *P*<.001). Like other positive morning emotions, a late bedtime (daily_sleep_time) was positively and significantly correlated with peacefulness (OR 0.696, 95% CI –0.497 to –0.229; *P*<.001). Those who usually felt tired upon waking (usual_state_tired) experienced less peace (OR 0.770, 95% CI –0.396 to –0.126; *P*<.001). In addition, participants who were more conscious of the people around them (usual_tpb_subjectiveNorm) tended not to feel peaceful (OR 0.845, 95% CI –0.310 to –0.026; *P*=.02), suggesting that social pressures may detract from morning peace.

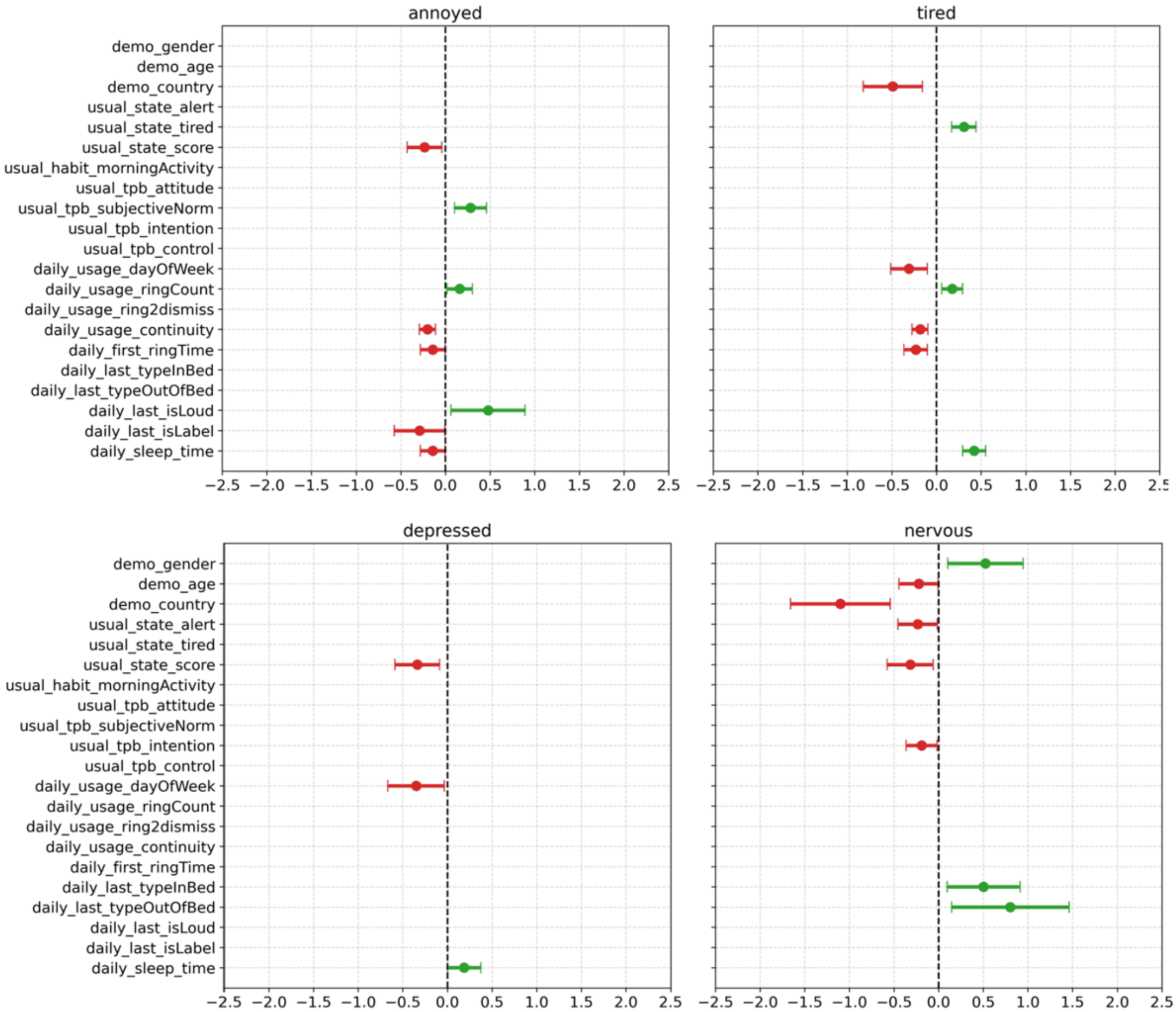
Refreshed

The number of alarms used in a day (daily_usage_ringCount) negatively affected the likelihood of feeling refreshed (OR 0.748, 95% CI –0.464 to –0.117; *P*=.001). Regular alarm use (daily_usage_continuity) positively influenced feelings of refreshment (OR 1.324, 95% CI 0.167 to 0.395; *P*<.001). Early bedtimes (daily_sleep_time: OR 0.568, 95% CI –0.708 to –0.421; *P*<.001) and later wake-up times (daily_first_ringTime: OR 1.392, 95% CI 0.190-0.474; *P*<.001) were significantly associated with feeling refreshed. Younger users (demo_age) felt more refreshed (OR 0.796, 95% CI –0.399 to –0.055; *P*=.01), as did males (demo_gender: OR 0.698, 95% CI –0.669 to –0.050; *P*=.02) and American users (demo_country: OR 0.541, 95% CI –0.988 to –0.242; *P*=.001). Usual tiredness after waking (usual_state_tired) was negatively correlated with feeling refreshed (OR 0.835, 95% CI –0.349 to –0.011; *P*=.03). The subjective norm in TPB (usual_tpb_subjectiveNorm) marginally decreased the feeling of refreshment (OR 0.858, 95% CI –0.313 to 0.008; *P*=.06), indicating that awareness of others can negatively affect the feeling of refreshment.

Negative Emotions

Our analysis identified several significant variables affecting negative morning emotions, as shown in Figure 2.

Figure 2. Results of generalized estimating equations analysis on negative emotions.



Annoyance

Participants who used alarms more frequently in a day (daily_usage_ringCount) were more likely to feel annoyed (OR 1.173, 95% CI 0.018-0.303; *P*=.02). Conversely, a higher number of consistent alarm use days (daily_usage_continuity) reduced the tendency for annoyance (OR 0.817, 95% CI –0.293 to –0.11; *P*<.001). Setting an alarm for a later time (daily_first_ringTime) was associated with decreased annoyance (OR 0.868, 95% CI –0.279 to –0.004; *P*=.04), suggesting that waking up later can improve mood. Loud alarm sounds (daily_last_isLoud) increased annoyance (OR 1.611, 95% CI 0.062-0.892; *P*=.02), whereas using alarms with label texts (daily_last_isLabel) was negatively related to annoyance (OR 0.748, 95% CI –0.573 to –0.007; *P*=.04). The need to wake up earlier (daily_first_ringTime) also increased annoyance (OR 1.221, 95% CI –0.279 to –0.004; *P*=.008). Participants' usual morning state (usual_state_score) showed a negative correlation with morning annoyance (OR 0.791, 95% CI –0.427 to –0.041; *P*=.01). In addition, the subjective norm variable in TPB (usual_tpb_subjectiveNorm) was significantly and positively related to annoyance (OR 1.323, 95% CI 0.102-0.459; *P*=.002), suggesting that social pressures contribute to feelings of irritation.

Tiredness

Alarms used on weekends (daily_usage_dayOfWeek) resulted in significantly lower tiredness than alarms used on weekdays (OR 0.734, 95% CI –0.513 to –0.103; *P*=.003). More frequent alarm use in a day (daily_usage_ringCount) increased tiredness (OR 1.190, 95% CI 0.058-0.291; *P*=.003), whereas regular alarm use throughout the week (daily_usage_continuity) was associated with reduced tiredness (OR 0.831, 95% CI –0.274 to –0.096; *P*<.001). Later bedtimes (daily_sleep_time) increased the likelihood of feeling tired (OR 1.523, 95% CI 0.295-0.550; *P*<.001), as did earlier wake-up times (daily_first_ringTime: OR 0.792, 95% CI –0.365 to –0.101; *P*=.001). Participants in the United States felt significantly less tired in the morning compared with those in South Korea (demo_country: OR 0.612, 95% CI –0.822 to –0.158; *P*=.004). Those who reported usually feeling tired in the presurvey (usual_state_tired) were more likely to feel tired during the experiment (OR 1.357, 95% CI 0.168-0.444; *P*<.001), indicating a consistent pattern of fatigue.

Depression

Participants with lower scores on the usual morning state in the presurvey (usual_state_score) were more likely to feel depressed during the experiment (OR 0.713, 95% CI –0.588 to –0.088; *P*=.008). Using alarms on weekends (daily_usage_dayOfWeek) was negatively related to morning depression (OR 0.703, 95% CI –0.669 to –0.035; *P*=.02). Later bedtimes (daily_sleep_time) increased the likelihood of feeling depressed the following day (OR 1.205, 95% CI 0.002-0.373; *P*=.04).

Nervousness

Morning nervousness varied depending on the type of wake-up tasks performed. In-bed wake-up tasks (daily_last_typeInBed: OR 1.655, 95% CI 0.096-0.912; *P*=.01) and out-of-bed tasks (daily_last_typeOutOfBed: OR 2.229, 95% CI 0.146-1.460; *P*=.01) were both associated with increased nervousness, suggesting that the nature of the wake-up activity impacts stress levels. Usual morning state (usual_state_score: OR 0.727, 95% CI –0.575 to –0.061; *P*=.01) and usual waking level (usual_state_alert: OR 0.776, 95% CI –0.456 to –0.015; *P*=.03) were negatively related to nervousness. Participants in the United States were significantly more likely to feel nervous compared with those in South Korea (demo_country: OR 0.333, 95% CI –1.657 to –0.542; *P*<.001). Men were 1.68 times less likely to feel nervous than women (demo_gender: OR 1.687, 95% CI 0.101 to 0.946; *P*=.01), and younger users tended to feel more nervous (demo_age: OR 0.800, 95% CI –0.442 to –0.005; *P*=.04). Among the TPB factors, only intention (usual_tpb_intention) significantly decreased nervousness in the morning (OR 0.826, 95% CI –0.364 to –0.019; *P*=.02).

No Emotions

Our analysis identified a single variable significantly related to the absence of emotions in the morning. South Korean participants were significantly more likely to report no specific emotions upon waking compared to participants in the United States (demo_country: OR 4.083, 95% CI 0.859-1.956; *P*<.001). This finding may be influenced by cultural differences, as previous research has indicated that Western individuals tend to express their emotions more openly than their Eastern counterparts [31].

Discussion

Daily Alarm Usage Related to Morning Emotions

The findings of this study provide insightful correlations between mobile alarm usage and morning emotions. Several key factors influencing morning emotions were identified, and these results can be interpreted and applied to improve daily wake-up experiences.

First, the significant influence of alarm set time on morning emotions can be explained by the alignment with natural circadian rhythms. Waking up later allows individuals to complete their sleep cycles, leading to feelings of peacefulness and refreshment. Conversely, a late bedtime disrupts these cycles, suppressing positive emotions and augmenting negative ones. This underscores the importance of maintaining consistent sleep and wake times for emotional well-being. Similar results have been demonstrated in previous studies dealing with sleep duration and its effect on emotions. For example, Demasi et al [32] reported that sleep duration has a positive relationship with mental health. Insufficient sleep duration deteriorates emotional health conditions [33,34], such as depression [35]. Wrzus et al [36] identified the influence of the appropriate duration of sleep on mental health the following day by age. Also, a high quality of sleep fostered positive emotions on the morning of the next day [37].

Second, the method of alarm deactivation impacts morning emotions, likely due to the physical and cognitive demands placed on individuals immediately upon waking. Active wake-up tasks, such as taking pictures or performing squats, can induce stress and discomfort, resulting in nervousness and unhappiness. Previous study results show that those who do not wake up on time usually prefer hard alarms based on wake-up tasks [16]. These tasks are often preferred by individuals who struggle to wake up on time, further exacerbating their negative morning emotions.

Third, the detailed alarm settings, such as loudness and sound type, affect post-wake emotions due to their direct impact on sensory processing upon waking. Loud and jarring alarms can cause abrupt awakenings, leading to annoyance and a negative start to the day. Han et al [38] analyzed the effects of the types of alarm sounds on postwake emotions and noted that loud sounds could cause bad morning emotions. In contrast, using alarms with pleasant sounds and moderate volume can create a smoother transition from sleep to wakefulness. In addition, an alarm used on weekends had a significant effect on positive emotions, such as happiness and peace in the morning, while negative emotions, such as tiredness and depression, decreased.

Finally, the relationship between the wake-up process and morning emotions highlights the role of sleep inertia—a state characterized by reduced cognitive and sensory-motor performance after waking. A prolonged period between the alarm ringing and deactivation can prolong this inertia, decreasing morning happiness. Similarly, multiple alarms can increase annoyance and tiredness due to repeated disruptions in the waking process. This indicates that morning emotion can be related to sleep inertia [39], which refers to the transitional state experienced after waking and lowers cognitive and sensory-motor performance [40-42]. The findings of this study are consistent with previous study results that sleep inertia impairs positive emotions and encourages negative emotions [43,44]. In addition, using labeled alarms (eg, self-motivating to wake up on time or specifying the purpose) tended to induce feelings of hope and happiness and reduce annoyance in the morning.

Based on these interpretations, practical suggestions for optimizing alarm usage to improve morning emotions are as follows:

- Optimal alarm timing: Set alarms for later times in the morning, if possible, to align with natural circadian rhythms. Aim for consistent wake times that allow for the completion of sleep cycles, promoting peacefulness and refreshment.
- Consistent sleep schedule: Maintain a regular bedtime and wake time, even on weekends, to ensure sufficient sleep

duration and prevent the suppression of positive emotions. A consistent sleep schedule supports overall emotional health and well-being.

- Gentle wake-up methods: Choose alarm deactivation methods that require minimal physical activity and are less intrusive. Avoid active tasks like taking pictures or performing exercises immediately upon waking, as these can increase stress and negative emotions.
- Moderate alarm volume and pleasant sounds: Use alarms with moderate volume and pleasant sounds to avoid abrupt awakenings and reduce annoyance. Select sounds that are calming and conducive to a gentle wake-up experience.
- Labeling alarms for motivation: Use labeled alarms with motivational messages or specified purposes to foster feelings of hope and happiness. This can make the wake-up process more intentional and positive.
- Minimize multiple alarms: Limit the number of alarms used in the morning to reduce repeated disruptions and minimize sleep inertia. Aim to wake up with a single alarm to enhance morning happiness and reduce tiredness.

In conclusion, while this study provides valuable insights into the relationship between alarm usage and morning emotions, future research with extended durations could offer a more comprehensive understanding. By adopting these practical suggestions, individuals can improve their morning emotional states and overall well-being.

Nonalarm Factors Influencing Morning Emotions

Our study identified various nonalarm factors that influence morning emotions. First, demographic variables demonstrated significant differences in experiencing morning emotions. For example, male participants tended to be less nervous upon waking, consistent with previous studies indicating that females report more negative emotions than males [45,46]. In addition, significant correlations were found between age and feelings of nervousness and hope in the morning, aligning with previous findings [47]. Older participants reported lower levels of feeling refreshed upon waking, which may be related to decreased sleep quality with age [48]. Participants in the United States experienced a wider range of morning emotions compared to Korean participants, highlighting the need to consider cultural factors in studies of human emotion [47].

The usual morning states reported by participants also explained their morning emotions. For instance, alertness and tiredness within the first 30 minutes of waking were significantly related to negative emotions. Every negative emotion, except tiredness, tended to decrease when the usual morning score was higher. These findings align with previous studies [49,50], which show that individuals with low self-assessment scores tend to have lower quality of life in terms of physical and mental health, experiencing higher levels of pain, depression symptoms, and lower daily activity scores.

Previous research has emphasized the importance of behavioral patterns and lifestyles in understanding human emotions [51-53]. For example, Peluso and Guerra de Andrade [54] examined the relationship between emotion and physical activity in improving mental health, while Penedo and Dahn [55] explored the benefits of physical activity on emotional well-being. Studies have also used apps or web-based services to investigate the effects of physical activity on emotions [56,57]. Similarly, our results indicated that participants who engaged in physical activities after waking experienced more positive emotions, such as happiness.

Finally, we found that TPB variables related to waking on time significantly affect morning emotions, corroborating previous studies on morning behavior change [15]. For example, a positive intention to wake up on time was positively related to hopeful emotions and negatively related to nervousness. This suggests that a strong willingness to wake up on time fosters feelings of hope and reduces nervousness in the morning. The subjective norm variable was significantly positively correlated with annoyance and negatively correlated with feelings of peace and refreshment. This indicates that relying on external cues to wake up can hinder positive morning emotions.

Unexplored Influences and Future Research

In our study, we ventured beyond traditional research paradigms by exploring how waking context, particularly alarm usage behaviors, influences morning emotions. This approach is significant as it sheds light on specific daily behaviors that can substantially impact emotional states at the start of the day. Notably, our research highlights the role of the Alarmy app's various functions, such as mission-based alarms, in shaping individuals' initial emotional responses. These findings underscore the importance of considering practical daily interactions with technology when studying emotional dynamics.

However, our investigation did not encompass all possible variables that could influence morning emotions. Factors such as family relationships, work-life balance, and health conditions can significantly influence individuals' morning emotions [7,58]. Research indicates that behaviors like physical activity, social interactions, diet, and medication compliance play a crucial role in shaping morning emotional states, particularly in populations with mental health issues and histories of chronic homelessness [59]. In addition, loneliness has been associated with lower positive affect and higher distressed affect throughout the day, emphasizing the impact of social relationships on emotional well-being [60]. Furthermore, work-related stress and exhaustion can lead to physical and mental health problems, affecting overall living conditions and emotional states [61].

Cultural factors also play a significant role in influencing the positivity or negativity of morning emotions. For example, individuals from different countries exhibit distinct morningness patterns, highlighting strong national differences in morningness scores and factorial structures [47]. Cultural factors between the US and South Korea can significantly influence emotions, as evidenced in various studies. The emotional tendencies of Korean women in the US and Korea were influenced more by personal characteristics than cultural differences [62]. In addition, Korean American adolescents often face challenges due to the clash between traditional Korean culture and Americanization demands, impacting their emotional well-being and behavior in school settings [63]. Differences in aesthetic perceptions between US and South Korean participants highlight how cultural contexts shape emotional responses to website aesthetics [64]. Furthermore, US-aided construction projects in

South Korea post-Korean War influenced the sociopolitical context and emotional landscape of the region by promoting democratic citizenship, private enterprise, and the American way of life [65].

The cultural dimensions proposed by Hofstede [66] further illuminate the distinctions between collectivistic and individualistic societies, which can deeply influence emotional responses. In collectivistic societies, individuals prioritize social harmony and often avoid behaviors that might discomfort others. This cultural trait could influence morning emotions, as people may wake feeling more harmonious or stressed based on their perception of social obligations or conflicts. Conversely, in individualistic cultures, where personal autonomy and self-expression are valued, transgressions of social norms can induce feelings of guilt, impacting morning emotional states.

For future research, it would be pertinent to incorporate these variables into the study design. A broader examination that includes diverse family dynamics, parenting styles, and a more extensive cultural representation would likely enhance the understanding of morning emotional states. In addition, longitudinal studies could provide deeper insights into how consistent interactions with alarm features, like those offered by Alarmy, influence long-term emotional and psychological well-being.

Such an integrated approach would not only validate the findings from this initial study but also expand our understanding of the intricate web of factors that influence morning emotions. Future research could also explore the impact of technology-based interventions tailored to different cultural contexts to improve morning emotional states. This could lead to more targeted interventions aimed at enhancing emotional well-being right from the start of the day, considering both cultural and familial contexts.

Limitations

This study presents several limitations that should be considered when interpreting the overall findings within the specific context in which the research was conducted.

First, this research examined factors related to morning emotions by analyzing the usage logs of the Alarmy app from 373 users across two countries, emphasizing the significance of the wake-up context. Previous studies have seldom addressed app usage logs or waking conditions in relation to morning emotions, making the use of actual log data notably uncommon. However, other known factors, such as family relationships, work-life balance, and health conditions, which also influence morning emotions, were not included in this study. The practical constraints of conducting research in real-world settings, as opposed to a laboratory environment, limited the consideration and control of all variables. Future studies could benefit from incorporating these additional factors to offer a more comprehensive understanding of morning emotions.

Second, this study used actual alarm usage logs and daily survey data collected over a 2-week period. Consequently, it did not capture long-term usage patterns. Further research is necessary to observe long-term effects or pattern changes. Despite the short duration, the study gathered daily data from 373 real users, yielding a substantial number of sample instances that reflect real-life scenarios. The findings from this study will be valuable for designing future long-term studies.

Third, while this study analyzed the relationships between various variables and morning emotions, it primarily focused on correlations rather than causations. This means that although significant associations between variables and morning emotions were identified, direct causal relationships cannot be conclusively determined. However, understanding these correlations can still be valuable. Depending on the nature of the variables, these correlations can provide useful insights for predicting morning emotions and potentially guiding interventions.

Finally, this study concentrated on the Alarmy app among mobile alarm applications. Alarmy is widely used globally and provides a variety of wake-up alarm functions, and its log data was accessible for analysis. Therefore, the findings of this study should be interpreted within the context of the Alarmy app. Future research comparing different types of wake-up alarm apps or methods could offer insights into varying user experiences and usage patterns.

Conclusions

This study aimed to comprehensively understand morning emotions by examining the context of morning waking through alarm usage. We categorized morning emotions into 9 distinct categories and identified factors affecting each. Specifically, we analyzed daily alarm usage logs, including ring times and deactivation types. The GEE analysis results revealed several factors related to each emotion. In addition to traditional factors such as demographics and usual state, daily alarm usage played a crucial role in understanding morning emotions.

This study demonstrates the feasibility of using waking alarms to monitor daily emotions. Existing methods for monitoring emotions primarily depend on explicit user data, such as survey responses [67,68]. While some studies have proposed monitoring methods based on implicit data from mobile sensors [69,70] or social activities from social media (eg, Twitter) [71-73], their practical application may be limited due to sensitive data collection and complex analyses. We believe that the findings of this study contribute to the improvement of future emotion monitoring by illustrating the relationship between waking alarm usage and specific emotions.

Acknowledgments

This work was supported by Culture, Sports and Tourism R&D Program through the Korea Creative Content Agency grant funded by the Ministry of Culture, Sports and Tourism in 2023 (Project Name: Cultivation of Next-generation Convergence Culture Technology [CT] Experts through Education, Developments, and Demonstration Projects for the Real-time Adaptive Performance and Exhibition Technologies to the User Reactions in the Metaverse, Project Number RS-2023-00224524) and by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2020S1A5B8104091).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Generalized estimating equations analysis results.
[DOCX File , 71 KB - humanfactors_v11i1e50835_app1.docx]

References

1. Dunn HL. High-level wellness for man and society. Am J Public Health Nations Health 1959;49(6):786-792. [doi: 10.2105/ajph.49.6.786] [Medline: 13661471]
2. Rothbard NP, Wilk SL. Waking up on the right or wrong side of the bed: start-of-workday mood, work events, employee affect, and performance. AMJ 2011;54(5):959-980 [FREE Full text] [doi: 10.5465/amj.2007.0056]
3. Vetrovsky T, Omcirk D, Malecek J, Stastny P, Steffl M, Tufano JJ. Morning fatigue and structured exercise interact to affect non-exercise physical activity of fit and healthy older adults. BMC Geriatr 2021;21(1):179 [FREE Full text] [doi: 10.1186/s12877-021-02131-y] [Medline: 33711945]
4. Hayashi M, Uchida C, Shoji T, Hori T. The effects of the preference for music on sleep inertia after a short daytime nap. Sleep and Biological Rhythms 2004;2(3):184-191. [doi: 10.1111/j.1479-8425.2004.00142.x]
5. Kräuchi K, Cajochen C, Wirz-Justice A. Waking up properly: is there a role of thermoregulation in sleep inertia? J Sleep Res 2004;13(2):121-127. [doi: 10.1111/j.1365-2869.2004.00398.x] [Medline: 15175091]
6. Sonnentag S, Binnewies C, Mojza EJ. "Did you have a nice evening?" A day-level study on recovery experiences, sleep, and affect. J Appl Psychol 2008;93(3):674-684. [doi: 10.1037/0021-9010.93.3.674] [Medline: 18457495]
7. Nandy RR, Nandy K, Hébert ET, Businelle MS, Walters ST. Identifying behaviors predicting early morning emotions by observing permanent supportive housing residents: an ecological momentary assessment. JMIR Ment Health 2019;6(2):e10186 [FREE Full text] [doi: 10.2196/10186] [Medline: 30730296]
8. Sonnentag S, Binnewies C. Daily affect spillover from work to home: detachment from work and sleep as moderators. Journal of Vocational Behavior 2013;83(2):198-208. [doi: 10.1016/j.jvb.2013.03.008]
9. Fritz C, Caughlin DE, Auten D. When a good day makes for a good night: antecedents of sleep and morning states. J Bus Psychol 2022;37(6):1235-1253. [doi: 10.1007/s10869-022-09793-4]
10. Totterdell P, Reynolds S, Parkinson B, Briner RB. Associations of sleep with everyday mood, minor symptoms and social interaction experience. Sleep 1994;17(5):466-475. [doi: 10.1093/sleep/17.5.466] [Medline: 7991960]
11. Mallett R, Picard-Deland C, Pigeon W, Wary M, Grewal A, Blagrove M, et al. The relationship between dreams and subsequent morning mood using self-reports and text analysis. Affect Sci 2022;3(2):400-405 [FREE Full text] [doi: 10.1007/s42761-021-00080-8] [Medline: 36046002]
12. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011;6(1):42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
13. Hankonen N, Hardeman W. Developing behavior change interventions. In: The Handbook of Behavior Change. Cambridge, England: Cambridge University Press; 2020:300-317.
14. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. CAIS 2009;24(1):28. [doi: 10.17705/1cais.02428]
15. Oh KT, Ko J, Shin J, Ko M. Using wake-up tasks for morning behavior change: development and usability study. JMIR Form Res 2022;6(9):e39497 [FREE Full text] [doi: 10.2196/39497] [Medline: 36129742]
16. Oh KT, Shin J, Kim J, Ko M. Analysis of a Wake-Up Task-Based Mobile Alarm App. Applied Sciences 2020;10(11):3993. [doi: 10.3390/app10113993]
17. Oh KT, Shin J, Kim J, Song IS, Lee U, Park E, et al. Wake-Up task: understanding users in task-based mobile alarm app. : Association for Computing Machinery; 2019 Presented at: Proceedings of the Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems; May 8, 2020; Glasgow, Scotland UK. [doi: 10.1145/3290607.3312765]
18. Singer E, Couper MP. Do incentives exert undue influence on survey participation? Experimental evidence. J Empir Res Hum Res Ethics 2008;3(3):49-56 [FREE Full text] [doi: 10.1525/jer.2008.3.3.49] [Medline: 19385770]
19. Terman M, Janet W, TM W. Morningness-eveningness questionnaire: self-assessment version (meq-sa).: Center for Environmental Therapeutics Internet; 2008. URL: https://cet.org/wp-content/uploads/2019/11/MEQ-SA-EN.pdf [accessed 2024-10-30]
20. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
21. Lee MC. Factors influencing the adoption of internet banking: an integration of TAM and TPB with perceived risk and perceived benefit. Electronic Commerce Research and Applications 2009;8(3):130-141. [doi: 10.1016/j.elerap.2008.11.006]

22. Shah AS, Sayuti N. Applying the theory of planned behavior (TPB) in halal food purchasing. International journal of Commerce and Management 2011;21(1):8-20. [doi: 10.1108/10569211111111676]
23. Liao C, Chen JL, Yen DC. Theory of planning behavior (TPB) and customer satisfaction in the continued use of e-service: an integrated model. Computers in Human Behavior 2007;23(6):2804-2822. [doi: 10.1016/j.chb.2006.05.006]
24. Monk TH, Buysse DJ, Rose LR, Hall JA, Kupfer DJ. The sleep of healthy people--a diary study. Chronobiol Int 2000;17(1):49-60. [doi: 10.1081/cbi-100101031] [Medline: 10672433]
25. Ekman P. An argument for basic emotions. Cognition and Emotion 2008;6(3-4):169-200. [doi: 10.1080/02699939208411068]
26. Plutchik R. A general psychoevolutionary theory of emotion. Theories of emotion 1980:3-33. [doi: 10.1016/b978-0-12-558701-3.50007-7]
27. Russell JA. A circumplex model of affect. Journal of Personality and Social Psychology 1980;39(6):1161-1178. [doi: 10.1037/h0077714]
28. Diggle P, Heagerty P, Liang KY, Zeger S. Analysis of Longitudinal Data. United Kingdom: Oxford University Press; 2002.
29. Amini P, Najafi-Vosough R, Ghaleiha A, Mahjub H. Evaluation of bipolar disorder in several relapses over time using generalized estimating equations. Iran J Psychiatry 2017;12(3):182-187 [FREE Full text] [Medline: 29062369]
30. Nam JH, Lim MS, Choi HK, Kim JY, Kim SK, Oh SS, et al. Factors increasing the risk for psychosocial stress among Korean adults living in rural areas: using generalized estimating equations and mixed models. Ann Occup Environ Med 2017;29(1):53 [FREE Full text] [doi: 10.1186/s40557-017-0209-5] [Medline: 29118991]
31. Matsumoto D. Cultural similarities and differences in display rules. Motiv Emot 1990;14(3):195-214. [doi: 10.1007/bf00995569]
32. DeMasi O, Feygin S, Dembo A, Aguilera A, Recht B. Well-Being tracking via smartphone-measured activity and sleep: cohort study. JMIR Mhealth Uhealth 2017;5(10):e137 [FREE Full text] [doi: 10.2196/mhealth.7820] [Medline: 28982643]
33. Kortesoja L, Vainikainen MP, Hotulainen R, Rimpelä A, Dobewall H, Lindfors P, et al. Bidirectional relationship of sleep with emotional and behavioral difficulties: a five-year follow-up of finnish adolescents. J Youth Adolesc 2020;49(6):1277-1291 [FREE Full text] [doi: 10.1007/s10964-020-01203-3] [Medline: 32086723]
34. Dinges DF, Pack F, Williams K, Gillen KA, Powell JW, Ott GE, et al. Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4–5 hours per night. Sleep 1997;20(4):267-277. [doi: 10.1093/sleep/20.4.267]
35. Berger AT, Wahlstrom KL, Widome R. Relationships between sleep duration and adolescent depression: a conceptual replication. Sleep Health 2019;5(2):175-179 [FREE Full text] [doi: 10.1016/j.sleh.2018.12.003] [Medline: 30928118]
36. Wrzus C, Wagner GG, Riediger M. Feeling good when sleeping in? Day-to-day associations between sleep duration and affective well-being differ from youth to old age. Emotion 2014;14(3):624-628. [doi: 10.1037/a0035349] [Medline: 24512244]
37. Triantafillou S, Saeb S, Lattie EG, Mohr DC, Kording KP. Relationship between sleep quality and mood: ecological momentary assessment study. JMIR Ment Health 2019;6(3):e12613 [FREE Full text] [doi: 10.2196/12613] [Medline: 30916663]
38. Han W, Wang J, Hu X, Cai H, Cheng J, Ning Z. The impact of digital alarm sound to human emotions: a case study. : IEEE; 2018 Presented at: Proceedings of the 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC); January 17, 2019; Miyazaki, Japan p. 1903-1908. [doi: 10.1109/smc.2018.00329]
39. Tassi P, Muzet A. Sleep inertia. Sleep Med Rev 2000;4(4):341-353. [doi: 10.1053/smrv.2000.0098] [Medline: 12531174]
40. Jeanneret PR, Webb WB. Strength of grip on arousal from full night's sleep. Percept Mot Skills 1963;17(3):759-761. [doi: 10.2466/pms.1963.17.3.759]
41. Jewett ME, Wyatt JK, Ritz-De Cecco A, Khalsa SB, Dijk DJ, Czeisler CA. Time course of sleep inertia dissipation in human performance and alertness. J Sleep Res 1999;8(1):1-8. [doi: 10.1111/j.1365-2869.1999.00128.x] [Medline: 10188130]
42. Wilkinson RT, Stretton M. Performance after awakening at different times of night. Psychon Sci 2013;23(4):283-285. [doi: 10.3758/bf03336113]
43. Kubo T, Takahashi M, Takeyama H, Matsumoto S, Ebara T, Murata K, et al. How do the timing and length of a night-shift nap affect sleep inertia? Chronobiol Int 2010;27(5):1031-1044. [doi: 10.3109/07420528.2010.489502] [Medline: 20636214]
44. Santhi N, Groeger JA, Archer SN, Gimenez M, Schlangen LJM, Dijk DJ. Morning sleep inertia in alertness and performance: effect of cognitive domain and white light conditions. PLoS One 2013;8(11):e79688 [FREE Full text] [doi: 10.1371/journal.pone.0079688] [Medline: 24260280]
45. Fujita F, Diener E, Sandvik E. Gender differences in negative affect and well-being: the case for emotional intensity. J Pers Soc Psychol 1991;61(3):427-434. [doi: 10.1037//0022-3514.61.3.427] [Medline: 1941513]
46. Costa PT, Zonderman AB, McCrae RR, Cornoni-Huntley J, Locke BZ, Barbano HE. Longitudinal analyses of psychological well-being in a national sample: stability of mean levels. J Gerontol 1987;42(1):50-55. [doi: 10.1093/geronj/42.1.50] [Medline: 3794196]
47. English T, Carstensen LL. Emotional experience in the mornings and the evenings: consideration of age differences in specific emotions by time of day. Front Psychol 2014;5:185 [FREE Full text] [doi: 10.3389/fpsyg.2014.00185] [Medline: 24639663]

48. Johnson JE. Progressive relaxation and the sleep of older men and women. J Community Health Nurs 1993;10(1):31-38. [doi: 10.1207/s15327655jchn1001_4] [Medline: 8496723]
49. Pandey R, Choubey A. Emotion and health: an overview. SIS Journal of Projective Psychology & Mental Health 2010;17(2):135-152 [FREE Full text]
50. Hu J, Gruber KJ. Positive and negative affect and health functioning indicators among older adults with chronic illnesses. Issues Ment Health Nurs 2008;29(8):895-911. [doi: 10.1080/01612840802182938] [Medline: 18649214]
51. Quiroz JC, Geangu E, Yong MH. Emotion recognition using smart watch sensor data: mixed-design study. JMIR Ment Health 2018;5(3):e10153 [FREE Full text] [doi: 10.2196/10153] [Medline: 30089610]
52. Hung GCL, Yang PC, Chang CC, Chiang JH, Chen YY. Predicting negative emotions based on mobile phone usage patterns: an exploratory study. JMIR Res Protoc 2016;5(3):e160 [FREE Full text] [doi: 10.2196/resprot.5551] [Medline: 27511748]
53. Sultana M, Al-Jefri M, Lee J. Using machine learning and smartphone and smartwatch data to detect emotional states and transitions: exploratory study. JMIR Mhealth Uhealth 2020;8(9):e17818 [FREE Full text] [doi: 10.2196/17818] [Medline: 32990638]
54. Peluso MAM, Guerra de Andrade LHS. Physical activity and mental health: the association between exercise and mood. Clinics (Sao Paulo) 2005;60(1):61-70 [FREE Full text] [doi: 10.1590/s1807-59322005000100012] [Medline: 15838583]
55. Penedo FJ, Dahn JR. Exercise and well-being: a review of mental and physical health benefits associated with physical activity. Curr Opin Psychiatry 2005;18(2):189-193. [doi: 10.1097/00001504-200503000-00013] [Medline: 16639173]
56. Green J, Huberty J, Puzia M, Stecher C. The effect of meditation and physical activity on the mental health impact of COVID-19-related stress and attention to news among mobile app users in the United States: cross-sectional survey. JMIR Ment Health 2021;8(4):e28479 [FREE Full text] [doi: 10.2196/28479] [Medline: 33788698]
57. Lambert JD, Greaves CJ, Farrand P, Price L, Haase AM, Taylor AH. Web-Based intervention using behavioral activation and physical activity for adults with depression (The eMotion Study): pilot randomized controlled trial. J Med Internet Res 2018;20(7):e10112 [FREE Full text] [doi: 10.2196/10112] [Medline: 30012547]
58. Newman DB, Epel ES, Coccia M, Puterman E, Prather AA. Asymmetrical effects of sleep and emotions in daily life. Affect Sci 2022;3(2):307-317 [FREE Full text] [doi: 10.1007/s42761-022-00112-x] [Medline: 36043202]
59. Steptoe A, Leigh ES, Kumari M. Positive affect and distressed affect over the day in older people. Psychol Aging 2011;26(4):956-965. [doi: 10.1037/a0023303] [Medline: 21517182]
60. Elias MA, Navarro VL. [The relation between work, health and living conditions: negativity and positivity in nursing work at a teaching hospital]. Rev Lat Am Enfermagem 2006;14(4):517-525 [FREE Full text] [doi: 10.1590/s0104-11692006000400008] [Medline: 16967157]
61. Carciofo R. Morning affect or sleep inertia? Comparing the constructs and their measurement. Chronobiol Int 2023;40(4):458-472. [doi: 10.1080/07420528.2023.2187211] [Medline: 36912023]
62. Kim S, Dorner LM. “I won’t talk about this here in America:” sociocultural context of Korean english language learners’ emotion speech in english. L2 2013;5(2):43-67. [doi: 10.5070/l25218124]
63. Kim JA. A study on the climacteric symptoms and emotion of korean women living in the United States and Korea. J Korean Acad Fundam Nurs 1999;6(1):35-50 [FREE Full text]
64. Faiola A, Ho CC, Tarrant MD, MacDorman KF. The aesthetic dimensions of U.S. and south korean responses to web home pages: a cross-cultural comparison. International Journal of Human-Computer Interaction 2011;27(2):131-150. [doi: 10.1080/10447318.2011.537173]
65. Jung AW, Stang K, Ferko D, Han SIL. A commentary on cultural influences impacting the education of Korean American adolescents. American Secondary Education 2011;40(1):31-38 [FREE Full text]
66. Hofstede G. Dimensionalizing cultures: the Hofstede model in context. Online Readings in Psychology and Culture 2011;2(1):8. [doi: 10.9707/2307-0919.1014]
67. Rickard N, Arjmand HA, Bakker D, Seabrook E. Development of a mobile phone app to support self-monitoring of emotional well-being: a mental health digital innovation. JMIR Ment Health 2016;3(4):e49 [FREE Full text] [doi: 10.2196/mental.6202] [Medline: 27881358]
68. Helweg-Jørgensen SH, Beck Lichtenstein M, Fruzzetti AE, Møller Dahl C, Pedersen SS. Daily self-monitoring of symptoms and skills learning in patients with borderline personality disorder through a mobile phone app: protocol for a pragmatic randomized controlled trial. JMIR Res Protoc 2020;9(5):e17737 [FREE Full text] [doi: 10.2196/17737] [Medline: 32449690]
69. Hart A, Reis D, Prestele E, Jacobson NC. Using smartphone sensor paradata and personalized machine learning models to infer participants' well-being: ecological momentary assessment. J Med Internet Res 2022;24(4):e34015 [FREE Full text] [doi: 10.2196/34015] [Medline: 35482397]
70. Piskioulis O, Tzafilkou K, Economides AA. Emotion detection through smartphone's accelerometer and gyroscope sensors. 2021 Presented at: Proceedings of the UMAP '21: 29th ACM Conference on User Modeling, Adaptation and Personalization; June 21, 2021; Netherlands p. 130-137. [doi: 10.1145/3450613.3456822]
71. Liu S, Zhu M, Yu DJ, Rasin A, Young SD. Using real-time social media technologies to monitor levels of perceived stress and emotional state in college students: a web-based questionnaire study. JMIR Ment Health 2017;4(1):e2 [FREE Full text] [doi: 10.2196/mental.5626] [Medline: 28073737]

72. Ofoghi B, Mann M, Verspoor KM. Towards early discovery of salient health threats: a social media emotion classification technique. 2016 Presented at: Proceedings of the Pacific Symposium on Biocomputing; January 04, 2016; Hawaii, US p. 504-515. [doi: 10.1142/9789814749411_0046]
73. Hasan M, Rundensteiner E, Agu E. Automatic emotion detection in text streams by analyzing Twitter data. Int J Data Sci Anal 2018;7(1):35-51. [doi: 10.1007/s41060-018-0096-z]

Abbreviations

GEE: generalized estimating equations
OR: odds ratio
TPB: theory of planned behavior

Edited by E Borycki; submitted 28.08.23; peer-reviewed by H Veldandi, S Pandey, H Sun; comments to author 11.01.24; revised version received 23.05.24; accepted 01.10.24; published 29.11.24.

Please cite as:
Oh KT, Ko J, Jin N, Han S, Yoon CY, Shin J, Ko M
Understanding Morning Emotions by Analyzing Daily Wake-Up Alarm Usage: Longitudinal Observational Study
JMIR Hum Factors 2024;11:e50835
URL: https://humanfactors.jmir.org/2024/1/e50835
doi:10.2196/50835
PMID:

The Added Value of Using Video in Out-of-Hours Primary Care Telephone Triage Among General Practitioners: Cross-Sectional Survey Study

Mette Amalie Nebsbjerg[1], MD; Katrine Bjørnshave Bomholt[1], MD; Claus Høstrup Vestergaard[1], MSc; Morten Bondo Christensen[1,2], MD, PhD; Linda Huibers[1,2], MD, PhD

[1]Research Unit for General Practice, Bartholins Allé 2, Aarhus, Denmark

[2]Department of Public Health, Aarhus University, Aarhus, Denmark

Corresponding Author:
Mette Amalie Nebsbjerg, MD
Research Unit for General Practice, , Bartholins Allé 2, Aarhus, , Denmark

Abstract

Background: Many countries have introduced video consultations in primary care both inside and outside of office hours. Despite some relational and technical limitations, general practitioners (GPs) have reported the benefits of video use in the daytime as it provides faster and more flexible access to health care. Studies have indicated that video may be specifically valuable in out-of-hours primary care (OOH-PC), but additional information on the added value of video use is needed.

Objective: This study aimed to investigate triage GPs' perspectives on video use in GP-led telephone triage in OOH-PC by exploring their reasons for choosing video use and its effect on triage outcome, the decision-making process, communication, and invested time.

Methods: We conducted a cross-sectional questionnaire study among GPs performing telephone triage in the OOH-PC service in the Central Denmark Region from September 5, 2022, until December 21, 2022. The questionnaire was integrated into the electronic patient registration system as a pop-up window appearing after every third video contact. This setup automatically linked background data on the contact, patient, and GP to the questionnaire data. We used descriptive analyses to describe reasons for and effects of video use and GP evaluation, stratified by patient age.

Results: A total of 2456 questionnaires were completed. The most frequent reasons for video use were to assess the severity (n=1951, 79.4%), to increase the probability of self-care (n=1279, 52.1%), and to achieve greater certainty in decision-making (n=810, 33%) (multiple answers were possible for reasons of video use). In 61.9% (n=1516) of contacts, the triage GPs anticipated that the contact would have resulted in a different triage outcome if video had not been used. Use of video resulted in a downgrading of severity level in 88.3% (n=1338) of cases. Triage GPs evaluated the use of video as positive in terms of their decision-making process (n=2358, 96%), communication (n=2214, 90.1%), and invested time (n=2391, 97.3%).

Conclusions: Triage GPs assessed that the use of video in telephone triage did affect their triage outcome, mostly by downgrading the level of care needed. The participating triage GPs found video in OOH-PC to be of added value, particularly in communication and the decision-making process.

(JMIR Hum Factors 2024;11:e52301) doi:10.2196/52301

KEYWORDS

primary health care; after-hours care; referral and consultation; general practitioners; triage; remote consultation; telemedicine

Introduction

Health care systems are undergoing a digital transformation worldwide [1-9]. Studies on contact patterns in primary health care during the COVID-19 pandemic found an overall increase in digital consultations (e-mail, video, or telephone) [5,10,11]. A central part of this development is driven by the use of video consultations, which many countries have introduced in primary care both inside and outside office hours [2,5,6,10,12,13]. Video use in daytime general practice has been reported to range from 1% to 6.4% [14-18], but studies on video use in out-of-hours primary care (OOH-PC) are lacking.

Video use has been shown to be context dependent, such as daytime general practice versus OOH-PC setting [15,19-21]. Several studies have investigated the perspectives of general practitioners (GPs) on video use in daytime general practice [3,12,15,19-26]. The benefits of video use include faster access to health care compared to face-to-face consultations and flexibility in care delivery [3,21,22,25,26]. As drawbacks of video use, GPs have reported technical [12,20,22-25] and relational limitations, such as lack of connection with the patient and of the feeling of presence and intimacy [12,24,26], and uncertainty about which patients are eligible candidates for video use [3,15,25,26]. Therefore, many GPs prefer face-to-face consultations over video consultations [12,19,21].

Limited knowledge exists on the user experiences of video in OOH-PC. In total, 2 qualitative studies found that the use of video adds value specifically in the context of OOH-PC services [19,27]. A focus group study among GPs in British OOH-PC conducted by Payne et al [27], showed that GPs used video to support their clinical decision-making. Furthermore, a questionnaire study conducted by Gren et al [28] among Danish triage nurses found that the use of video improved patient assessment and reassurance in calls related to young children. No quantitative studies on the GP perspective have been conducted, highlighting the need for additional insights into the added value of video use in OOH-PC, such as the effect on patient flows. This knowledge is essential to ensure optimal utilization and further implementation of video technology.

This study aimed to investigate triage GPs' perspectives on video use in GP-led telephone triage in OOH-PC by exploring their reasons for choosing video use and its effect on triage outcome, the decision-making process, communication, and invested time. Based on literature, we expected high satisfaction levels among GPs. In addition, we hypothesized a moderate effect of video use on triage outcomes, as video theoretically provides the triage GP with more information, enabling them to make a better-informed decision.

Methods

Design

We conducted a cross-sectional questionnaire study among GPs conducting telephone triage in the OOH-PC service in the Central Denmark Region from September 5, 2022, until December 21, 2022.

Setting

In Denmark, health care is publicly funded and free of charge for all residents, including all care delivered by GPs. The health care system is centrally regulated, but most services, including OOH-PC, are provided by the local governments of the 5 Danish regions. OOH-PC services are open on weekdays from 4 PM to 8 AM and for 24 hours during weekends and holidays. GPs and GP trainees (hereafter referred to collectively as triage GPs) are obligated to cover shifts in their regional OOH-PC service.

The study was conducted in the Central Denmark Region. This region has 2 call centers handling all telephone triage contacts, along with 13 locations for clinic consultations. The number of triage GPs present depends on weekday and time of day, ranging from 2 to 15 triage GPs. Patients are obligated to call the OOH-PC, being triaged by telephone. They cannot show up physically at the OOH-PC service without a prior telephone triage call. Triage GPs perform telephone triage with the use of video (video contact) or without the use of video (telephone contact), but without a clinical decision support tool. When answering a telephone triage contact, the triage GP gathers information about the reason for the encounter. Based on this information and their clinical experience, the triage GP decides if the telephone triage contact is suitable for video use. If so, a video link is sent to the caller in an SMS text message, and, when activated by the caller, a one-way video connection is established. In Denmark, video contacts were rapidly implemented at the start of the COVID-19 pandemic, without clear guidelines for their use. Triage GPs are paid a fee-for-service through standardized remuneration codes for telephone contacts (€10, US $10.89) or video contacts (€31, US $33.76). Triage GPs also decide upon the triage outcome: self-care, referral to face-to-face consultation (clinic consultation or home visit), or referral to hospital. Out of 5 Danish regions, 3 are organized with GP-led triage (including the Central Denmark Region), whereas 2 regions have nurse-led or mixed triage.

Questionnaire Development

The development of the questionnaire followed a thorough process. The content was based on results from studies investigating video use in a general practice setting [1,3,4,7,8,12,15,16,22,25,29-35] and on experiences with video use shared by triage GPs and local stakeholders. The research group conducted several internal feedback rounds, which were followed by external feedback from 8 GPs with expertise in both research and triage in OOH-PC. Both internal and external feedback concerned the wording, construction, and clinical relevance of included items. In addition, short interviews were conducted with several triage GPs from the regional OOH-PC service to compile a comprehensive list of relevant response categories to the defined questions. Furthermore, cognitive interviews were conducted with 2 GPs from the external feedback panel to assure understanding of the meaning and relevance of the questions and answering categories. Finally, we performed a 1-day pilot test among triage GPs in the regional OOH-PC service, receiving responses from 19 different triage GPs, which resulted in 40 completed questionnaires. The test provided us with further information to consider before initiation of data collection (eg, relevance of response categories, response rate, and ceiling effect).

Final Questionnaire

The final questionnaire consisted of 7 questions. One question covered reasons for video use and offered predefined response categories with multiple responses possible as well as a free-text response option. One question covered the actual triage outcome (self-care, referral to face-to-face consultation, or referral to hospital), and one question covered the anticipated triage outcome without video use to assess if the use of video may influence the triage outcome. In total, 3 questions covered the added effect of video on the decision-making process (5-point Likert scale), communication (5-point Likert scale), and invested time (predefined response categories). One final question covered the reason for encounter (data not presented in this article). An English version of the questionnaire is available in Multimedia Appendix 1.

Data Collection

The questionnaire was integrated into the electronic patient registration system of the regional OOH-PC service and was set to pop-up after every third video contact. This setup automatically linked background data on the contact, patient, and GP to the questionnaire data (date and time of contact, and sex and age of both patient and triage GP). As the questionnaire was integrated into the electronic patient registration system, triage GPs had to answer the questionnaire between patient contacts. To affect the workflow and telephone waiting time as little as possible, answering the questionnaire should take no more than 1 minute, which was confirmed in the pilot test.

To investigate the representativeness of patients and triage GPs, we included data on all video contacts in the regional OOH-PC service in the study period. Triage GPs were invited to participate in the study when logging into the electronic patient registration system at the beginning of a telephone triage shift. Participation was voluntary, and the triage GPs were remunerated for participation (€4 [US $4.36] per completed questionnaire). To avoid the sample being dominated by a few very active triage GPs, we allowed each individual triage GP to participate in a maximum of 5 telephone triage shifts.

Data Handling and Statistical Analyses

Contacts were categorized according to whether the triage GP participated in the study. Triage GPs who had already participated in 5 triage shifts were categorized as participating, since it is reasonable to assume that these triage GPs would also have accepted participation if they received an invitation for subsequent triage shifts.

On September 6, 2022, a server crash prevented questionnaire data generation. Therefore, all video contacts on this date were excluded from the analysis (n=173; 0.8%). Furthermore, 59 blank questionnaires were excluded from the analysis. In total, 5 questionnaires had missing answers to the question on estimated triage outcome without video use; these questionnaires were not excluded since the remaining parts had been completed (Multimedia Appendix 2).

Patient age and sex were calculated through the Danish personal identification number. A total of 28 (0.3%) video contacts had missing personal identification number and were excluded from the analysis. For the purposes of stratification, patient age was categorized in 3 groups: 0 - 4 years, 5 - 18 years, and >18 years. This stratification was chosen because of very few older adult patients receiving a video contact (Multimedia Appendix 3).

Descriptive analyses were used to describe the study population and the questionnaire data. We performed stratified analyses according to patient age and used the Pearson chi-square test to compare groups. To investigate possible selection bias, we compared contacts handled by participating triage GPs to contacts handled by nonparticipating triage GPs. The reasons for video use were presented with predefined response categories, and answers were stratified by patient age. To investigate the possible effect of video use on the actual triage outcome, we tabulated answers to the question regarding the anticipated triage outcome against answers to the question on the actual triage outcome. Finally, to present the triage GPs' evaluation of video use, we tabulated the answers to the questions regarding the decision-making process, communication, and time invested.

Ethical Considerations

The Committee on Health Research Ethics in the Central Denmark Region approved the data collection from the electronic patient records in the OOH-PC registration system (1-45-70-15-22). All GPs taking shifts in the regional OOH-PC service were informed about the study prior to the beginning of the data collection. During the study period, all triage GPs were informed about the project and the ability to opt out in accordance with the General Data Protection Regulation of the European Union. Answering questionnaires was voluntary, and triage GPs gave their consent for participation at the beginning of every triage shift. Triage GPs were paid €4 (US $ 4.24) per questionnaire. The study was listed in the record of processing activities at the Research Unit for General Practice in Aarhus, and data were anonymized in accordance with the provisions of the General Data Protection Regulation. Finally, the study was endorsed by the Danish Organization of General Practitioners in the Central Denmark Region.

Results

Participation and Questionnaire Completion Rates

In the study period, a total of 189,229 telephone triage contacts were conducted in the regional OOH-PC service. Of these, 22,093 (11.7%) were video contacts, of which 2456 resulted in a completed pop-up questionnaire (Multimedia Appendix 2).

In total, 649 different triage GPs were invited 2597 times when logging into the electronic patient registration system at the beginning of a telephone triage shift. Of these, 1330 invitations were accepted (participation rate: 51.2%), of which 430 invitations (32.3%) did not produce completed pop-up questionnaires as the triage GP used video less than 3 times during the triage shift. The remaining 900 accepted invitations resulted in 2456 completed questionnaires. The average number of questionnaires completed by a triage GP was 0.36 (IQR 0.14 - 0.50) questionnaires per hour, corresponding to 2.8 (IQR 1.12 - 4) questionnaires on an 8-hour triage shift.

Patient characteristics for all video contacts during the study period were stratified according to whether (or not) the contact resulted in a completed pop-up questionnaire. Patient characteristics were comparable for both groups (Multimedia Appendix 3).

Characteristics of Triage GPs

Table 1 shows GP characteristics for all video contacts during the study period. Video contacts were stratified according to whether the contact was handled by a participating or a nonparticipating triage GP. Compared to nonparticipating triage GPs, participating GPs were significantly more often male (54.6% vs 65.1%) and under the age of 50 years (46.4% vs 63.5%) (Table 1).

Table . Distribution of triage GP[a] characteristics in all video contacts, stratified for whether (or not) the triage GP participated in the study.

Triage GP characteristics	Video contacts (n=22,093), n (%)	
	Nonparticipating triage GPs (n=5802)	Participating triage GPs (n=16,291)
Sex [b]		
Female	2634 (45.4)	5691 (34.9)
Male	3168 (54.6)	10,600 (65.1)
Age (years) [b]		
31 - 40	968 (16.7)	4104 (25.2)
41 - 50	1722 (29.7)	6248 (38.4)
51 - 60	2000 (34.5)	4013 (24.6)
61 - 70	1083 (18.6)	1830 (11.2)
>70	29 (0.5)	96 (0.6)

[a]GP: general practitioner.
[b]Significant difference between groups (P<.001) when using the Pearson χ^2 test.

Reasons for Video Use

The most frequent reasons for video use were to assess the severity (79.4%), to increase the probability of self-care (52.1%), and to achieve greater certainty in decision-making (33%) (Table 2). When stratifying for patient age, significant differences between age groups were found; to assess the severity and to achieve greater certainty in decision-making were more often used for contacts concerning children, whereas to better understand what the encounter was about was more often used for contacts concerning adults.

Table . Triage GPs'[a] reasons for using video according to predefined response categories, stratified for patient age groups.

Reasons for using video[b]	Age groups, n (%)			Total (n=2456)
	0 - 4 years (n=974[c])	5 - 18 years (n=461[c])	>18 years (n=1021[c])	
To better assess the severity of the condition and the described symptoms[d]	821 (84.3)	365 (79.2)	765 (74.9)	1951 (79.4)
To increase the probability of being able to finish the patient by phone	490 (50.3)	251 (54.5)	538 (52.7)	1279 (52.1)
To achieve greater certainty in decision-making about the triage outcome[d]	364 (37.4)	144 (31.3)	302 (29.6)	810 (33.0)
To better understand what the encounter was about[d]	79 (8.1)	47 (10.2)	120 (11.8)	246 (10.0)
To meet the patient's needs (eg, long distance to consultation room, requested by the patient)	41 (4.2)	22 (4.8)	52 (5.1)	115 (4.7)
To ensure that the triage of the patient could be completed in shorter time	20 (2.1)	3 (0.7)	12 (1.2)	35 (1.4)

[a]GP: general practitioner.
[b]Multiple answers allowed (total number of answers: 4436).
[c]n: number of completed questionnaires within the age group.
[d]Significant difference between age groups (P<.001) when using the Pearson χ^2 test.

Effect on Triage Outcomes

Of the 2456 included video contacts, triage GPs anticipated that the contact would have resulted in a different triage outcome if video had not been used in 1516 contacts (61.9%) (Table 3). With the use of video, 1338 of these contacts (88.3%) were downgraded to a lesser severity level. Of the 650 contacts that triage GPs anticipated to have ended with self-care prior to video use, 120 contacts were upgraded to either a face-to-face consultation (16.9%) or a hospital referral (1.5%). Additionally, video use changed the outcome of some of the 1719 contacts anticipated to end with a face-to-face consultation prior to video use; video downgraded 76.4% of these contacts to self-care, whereas 2.3% were upgraded to a hospital referral. Prior to video use, triage GPs would have ended 82 contacts with a hospital referral, but video use downgraded 23.2% of these contacts to face-to-face consultation and 30.5% to self-care. When we stratified for patient age, we found that young children aged 0 - 4 years (974 contacts) were more often upgraded from self-care to a hospital referral (2.6%) compared to older children (461 contacts) (0%) and adults (1016 contacts) (1.2%) (data not shown in the table).

Table . Anticipated triage outcome (vertical) versus actual triage outcome (horizontal).

		Actual triage outcome (%)			Total[a], n
		Self-care	Face-to-face consultation	Hospital referral	
Anticipated triage outcome (%)					
	Self-care	81.5[b]	16.9[c]	1.5[c]	650
	Face-to-face consultation	76.4[d]	21.4[b]	2.3[c]	1719
	Hospital referral	30.5[d]	23.2[d]	46.3[b]	82

[a]Total number of completed questionnaires (n=2451), as 5 questionnaires had missing data regarding estimated triage outcome without video use.

[b]Unchanged triage outcome.

[c]Upgraded triage outcome.

[d]Downgraded triage outcome.

Effect on the Decision-Making Process, Communication, and Invested Time

According to triage GPs, video use contributed to a better decision-making process and better communication with the patient (Table 4). Only 5 questionnaires (0.2%) had a negative assessment (3 regarding the decision-making process and 2 regarding communication). Triage GPs generally assessed that video use was worth the time spent, but disagreement was reported in 49 questionnaires (2%). Triage GPs assessed that their decision-making process and the time spent on video use primarily benefitted the youngest patients (0 - 4 years) (data not shown in the table).

Table . Triage GPs'[a] assessment of video use regarding the decision-making process, communication, and invested time.

Question and response categories		Response, n (%)[b]
How did the use of video influence your decision-making process?		
	It got much better	1349 (54.9)
	It got better	1009 (41.1)
	There was no difference	95 (3.9)
	It got worse	2 (0.1)
	It got much worse	1 (0.0)
How did the use of video influence your communication with the patient?		
	It got much better	1143 (46.5)
	It got better	1071 (43.6)
	There was no difference	240 (9.8)
	It got worse	2 (0.1)
	It got much worse	0 (0)
Was the use of video worth the time spent?		
	Yes, to a large extent	1887 (76.8)
	Yes, to a lesser extent	504 (20.5)
	No	49 (2.0)
	Don't know	16 (0.7)

[a]GP: general practitioner.

[b]Total number of completed questionnaires for each question (n=2456).

Discussion

Principal Results

The most frequent reasons for video use were to assess the severity, to increase the probability of self-care, and to achieve greater certainty in decision-making. In 61.9% of contacts, the triage GPs anticipated that the contact would have resulted in a different triage outcome if video had not been used. Use of video resulted in a downgrading of severity level in 88.3% of cases. Triage GPs found the use of video to have a positive influence on the decision-making process, communication, and invested time.

Limitations

To our knowledge, our study is the first to quantitatively investigate GPs' perspectives on video use in OOH-PC. The study is based on a large dataset that combines data from questionnaires and register data. The risk of recall bias was reduced as triage GPs answered the questionnaire immediately after a video contact. Furthermore, the questionnaire was answered by many different triage GPs, which increased the representativeness of the results.

Our study also had some limitations. Selection bias may be present because triage GPs who declined participation may differ from participating triage GPs in their way of using and assessing video. Participating GPs were younger and more often male compared to those who declined. In addition, triage GPs conducting many video contacts were overrepresented in our data set, as they completed more questionnaires. These triage GPs are likely to differ from triage GPs having only a few video contacts in terms of their attitude toward video use. Furthermore, we found a ceiling effect of answers regarding communication, the decision-making process, and time invested (with mainly very positive assessments). This ceiling effect points toward indication bias, as the triage GPs first decided if the use of video was suitable and then assessed the quality of this decision. Furthermore, the answering categories in the question regarding time invested do not have the same degree of differentiation as the other questions, which may have skewed the results in a positive direction. Telephone triage by GPs is uncommon in most countries with comparable OOH-PC services [36], and GPs may differ in their use and assessment of video. Finally, we investigated 1-way video, whereas 2-way video may have produced different findings.

Comparison With Prior Work

We found that triage GPs most often used video in telephone triage contacts to assess the severity, to increase the probability of self-care, and to achieve greater certainty in decision-making. These findings are in line with results from 2 qualitative studies investigating video use in OOH-PC [27,28]. Gren et al [28] interviewed triage nurses using video in contacts concerning young children and found that nurses identified several children who appeared more ill on video than presented by telephone. Payne et al [27] conducted a focus group study among GPs working in OOH-PC. They found that GPs used video to support decision-making about which patients could be managed safely with self-care advice and which patients were seriously ill and needed further care. Many contacts to OOH-PC concern reassurance [28]. Seeing the patient, thereby not only relying on verbal information from the patient, is likely to increase reassurance and the likelihood of ending with self-care advice. In addition, a scoping review found that GPs in daytime general practice considered video to be effective in achieving successful decision-making [21].

In our study, the use of video appeared to improve the patient flow due to more self-care advice and direct referrals to the hospital. In line with our findings, the 2 abovementioned qualitative studies found that video facilitated more effective handling of calls [27,28]. Improving the call handling might offer substantial advantages to the health care system and clinicians, primarily by reducing the use of health care resources and optimizing resource allocation. More efficient call handling may also enhance patient safety through the facilitation of direct hospital referrals when needed [28].

Danish triage GPs found video to be valuable in communicating with the patient at OOH-PC. Similarly, Payne et al [27] highlighted that GPs experienced a closer personal connection with patients when using video compared to telephone contacts. In line with this, Gren et al [28] found that triage nurses assessed their interactions to become more humane when using video. However, findings from studies conducted in daytime general practice present a more negative picture. Some studies have indicated that GPs perceive relational limitations due to inaccurate reading of nonverbal signals and challenges with expressing emotions when not being physically present [12,24,26]. Yet, video use may also help build relationships with patients who might otherwise be reluctant to seek help [26]. The difference found between daytime general practice and OOH-PC could be due to many factors, including contextual factors and variations in patient population and reasons for encounter [37-40]. Furthermore, in daytime general practice, video consultations are often compared with in-clinic consultations, whereas contacts in OOH-PC video are compared with telephone contacts [25,27]. This discrepancy could contribute to the difference in evaluation of video use between the 2 settings. At OOH-PC, video is used as an additional triage tool to assess the level of care needed.

Implications for Practice and Future Research

Our study suggests that great added value of video use in OOH-PC telephone triage. Using video may optimize the utilization of limited medical resources during periods of large demands by reducing the number of face-to-face consultations and increasing the number of self-care advices. However, the real potential of video use should be investigated, addressing the scope of use and the number of suitable telephone triage contacts. Further implementation of video in OOH-PC could be recommended, although future studies are needed to support our results.

Future studies should examine whether the outcomes of this study remain consistent across varying OOH-PC models and varying triage professionals. Moreover, the added value of 2-way video in OOH-PC should be investigated, specifically the quality of communication. Research on patient safety associated with the use of video in OOH-PC is also highly relevant, as we found a considerable level of downgrading of severity level.

Conclusions

Triage GPs assessed that the use of video in telephone triage did affect their triage outcome, mostly by downgrading the level of care needed. The participating triage GPs found video in OOH-PC of added value, in particular in the communication and in the decision-making process. However, future research is needed to assess the full potential of video use and to define the best scope of use.

Acknowledgments

We gratefully acknowledge the financial support provided by the Danish health insurance foundation Sygeforsikringen Danmark, the General Practice Research Foundation (Praksisforskningsfonden) of the Central Denmark Region, the Department of Public Health at Aarhus University, the Multipractice Study Committee (Multipraksisudvalget) of the Danish College of General Practitioners, and the Committee for Quality Improvement and Continuing Medical Education (Kvalitets- og Efteruddannelsesudvalget) of the Central Denmark Region. The funding bodies had no role in the study design, data collection, data analysis, data interpretation, writing of the manuscript, or submission of the final article.

Authors' Contributions

All authors contributed to the study design, interpretation of results, and drafting and revising of the manuscript. The first author (MAN) was responsible for data management and the statistical analysis. All authors agreed to the final submitted version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire on video use in out-of-hours primary care (final version).
[DOCX File, 23 KB - humanfactors_v11i1e52301_app1.docx]

Multimedia Appendix 2
Flowchart presenting data collection.
[PPTX File, 36 KB - humanfactors_v11i1e52301_app2.pptx]

Multimedia Appendix 3
Distribution of patient characteristics.
[DOCX File, 14 KB - humanfactors_v11i1e52301_app3.docx]

References

1. Drerup B, Espenschied J, Wiedemer J, Hamilton L. Reduced no-show rates and sustained patient satisfaction of telehealth during the COVID-19 pandemic. Telemed J E Health 2021 Dec;27(12):1409-1415. [doi: 10.1089/tmj.2021.0002] [Medline: 33661708]
2. Green MA, McKee M, Katikireddi SV. Remote general practitioner consultations during COVID-19. Lancet Digit Health 2022 Jan;4(1):e7. [doi: 10.1016/S2589-7500(21)00279-X] [Medline: 34952678]
3. Johnsen TM, Norberg BL, Kristiansen E, et al. Suitability of video consultations during the COVID-19 pandemic lockdown: cross-sectional survey among Norwegian general practitioners. J Med Internet Res 2021 Feb 8;23(2):e26433. [doi: 10.2196/26433] [Medline: 33465037]
4. Saint-Lary O, Gautier S, Le Breton J, et al. How GPs adapted their practices and organisations at the beginning of COVID-19 outbreak: a French national observational survey. BMJ Open 2020 Dec 2;10(12):e042119. [doi: 10.1136/bmjopen-2020-042119] [Medline: 33268433]
5. Sigurdsson EL, Blondal AB, Jonsson JS, et al. How primary healthcare in Iceland swiftly changed its strategy in response to the COVID-19 pandemic. BMJ Open 2020 Dec 7;10(12):e043151. [doi: 10.1136/bmjopen-2020-043151] [Medline: 33293329]
6. Wherton J, Greenhalgh T, Shaw SE. Expanding video consultation services at pace and scale in Scotland during the COVID-19 pandemic: national mixed methods case study. J Med Internet Res 2021 Oct 7;23(10):e31374. [doi: 10.2196/31374] [Medline: 34516389]
7. Bertelsen P, Stub Petersen L. Danish citizens and general practitioners' use of ICT for their mutual communication. Stud Health Technol Inform 2015;216:376-379. [Medline: 26262075]
8. Kampik T, Larsen F, Bellika JG. Internet-based remote consultations - general practitioner experience and attitudes in Norway and Germany. Stud Health Technol Inform 2015;210:452-454. [Medline: 25991185]
9. Bradford NK, Caffery LJ, Smith AC. Telehealth services in rural and remote Australia: a systematic review of models of care and factors influencing success and sustainability. Rural Remote Health 2016;16(4):27817199. [Medline: 27744708]
10. Ramerman L, Rijpkema C, Bos N, Flinterman LE, Verheij RA. The use of out-of-hours primary care during the first year of the COVID-19 pandemic. BMC Health Serv Res 2022 May 21;22(1):679. [doi: 10.1186/s12913-022-08096-x] [Medline: 35597939]
11. Morreel S, Philips H, Verhoeven V. Organisation and characteristics of out-of-hours primary care during A COVID-19 outbreak: a real-time observational study. PLoS One 2020;15(8):e0237629. [doi: 10.1371/journal.pone.0237629] [Medline: 32790804]
12. Due TD, Thorsen T, Andersen JH. Use of alternative consultation forms in Danish general practice in the initial phase of the COVID-19 pandemic - a qualitative study. BMC Fam Pract 2021 Jun 2;22(1):108. [doi: 10.1186/s12875-021-01468-y] [Medline: 34078281]
13. Joy M, McGagh D, Jones N, et al. Reorganisation of primary care for older adults during COVID-19: a cross-sectional database study in the UK. Br J Gen Pract 2020 Aug;70(697):e540-e547. [doi: 10.3399/bjgp20X710933] [Medline: 32661009]
14. Assing Hvidt E, Christensen NP, Grønning A, Jepsen C, Lüchau EC. What are patients' first-time experiences with video consulting? A qualitative interview study in Danish general practice in times of COVID-19. BMJ Open 2022 Apr 15;12(4):e054415. [doi: 10.1136/bmjopen-2021-054415] [Medline: 35428624]
15. Scott A, Bai T, Zhang Y. Association between telehealth use and general practitioner characteristics during COVID-19: findings from a nationally representative survey of Australian doctors. BMJ Open 2021 Mar 24;11(3):e046857. [doi: 10.1136/bmjopen-2020-046857] [Medline: 33762248]
16. Murphy M, Scott LJ, Salisbury C, et al. Implementation of remote consulting in UK primary care following the COVID-19 pandemic: a mixed-methods longitudinal study. Br J Gen Pract 2021;71(704):e166-e177. [doi: 10.3399/BJGP.2020.0948] [Medline: 33558332]
17. Chang JE, Lindenfeld Z, Albert SL, et al. Telephone vs. video visits during COVID-19: safety-net provider perspectives. J Am Board Fam Med 2021;34(6):1103-1114. [doi: 10.3122/jabfm.2021.06.210186] [Medline: 34772766]
18. Dai Z, Sezgin G, Hardie RA, et al. Sociodemographic determinants of telehealth utilisation in general practice during the COVID-19 pandemic in Australia. Intern Med J 2023 Mar;53(3):422-425. [doi: 10.1111/imj.16006] [Medline: 36624629]
19. Greenhalgh T, Ladds E, Hughes G, et al. Why do GPs rarely do video consultations? Qualitative study in UK general practice. Br J Gen Pract 2022 May;72(718):e351-e360. [doi: 10.3399/BJGP.2021.0658] [Medline: 35256385]

20. Meurs M, Keuper J, Sankatsing V, Batenburg R, van Tuyl L. “Get used to the fact that some of the care is really going to take place in a different way”: general practitioners’ experiences with e-health during the COVID-19 pandemic. Int J Environ Res Public Health 2022 Apr 22;19(9):5120. [doi: 10.3390/ijerph19095120] [Medline: 35564519]
21. Thiyagarajan A, Grant C, Griffiths F, Atherton H. Exploring patients’ and clinicians’ experiences of video consultations in primary care: a systematic scoping review. BJGP Open 2020;4(1):32184212. [doi: 10.3399/bjgpopen20X101020] [Medline: 32184212]
22. Koch S, Guhres M. Physicians’ experiences of patient-initiated online consultations in primary care using direct-to-consumer technology. Stud Health Technol Inform 2020 Jun 16;270:643-647. [doi: 10.3233/SHTI200239] [Medline: 32570462]
23. Mold F, Hendy J, Lai YL, de Lusignan S. Electronic consultation in primary care between providers and patients: systematic review. JMIR Med Inform 2019 Dec 3;7(4):e13042. [doi: 10.2196/13042] [Medline: 31793888]
24. Nordtug M, Assing Hvidt E, Lüchau EC, Grønning A. General practitioners’ experiences of professional uncertainties emerging from the introduction of video consultations in general practice: qualitative study. JMIR Form Res 2022 Jun 14;6(6):e36289. [doi: 10.2196/36289] [Medline: 35653607]
25. Randhawa RS, Chandan JS, Thomas T, Singh S. An exploration of the attitudes and views of general practitioners on the use of video consultations in a primary healthcare setting: a qualitative pilot study. Prim Health Care Res Dev 2019 Jan;20:e5. [doi: 10.1017/S1463423618000361] [Medline: 29909798]
26. Norberg BL, Getz LO, Johnsen TM, Austad B, Zanaboni P. General practitioners’ experiences with potentials and pitfalls of video consultations in Norway during the COVID-19 lockdown: qualitative analysis of free-text survey answers. J Med Internet Res 2023 Mar 20;25:e45812. [doi: 10.2196/45812] [Medline: 36939814]
27. Payne RE, Clarke A. How and why are video consultations used in urgent primary care settings in the UK? A focus group study. BJGP Open 2023 Sep;7(3):37068795. [doi: 10.3399/BJGPO.2023.0025] [Medline: 37068795]
28. Gren C, Egerod I, Linderoth G, et al. “We can’t do without it”: parent and call-handler experiences of video triage of children at a medical helpline. PLoS One 2022;17(4):e0266007. [doi: 10.1371/journal.pone.0266007] [Medline: 35421109]
29. Alsnes IV, Munkvik M, Flanders WD, Øyane N. How well did Norwegian general practice prepare to address the COVID-19 pandemic? Fam Med Com Health 2020 Dec;8(4):e000512. [doi: 10.1136/fmch-2020-000512]
30. Hanna L, Fairhurst K. Using information and communication technologies to consult with patients in Victorian primary care: the views of general practitioners. Aust J Prim Health 2013;19(2):166-170. [doi: 10.1071/PY11153] [Medline: 22950961]
31. Jiwa M, Meng X. Video consultation use by Australian general practitioners: video vignette study. J Med Internet Res 2013 Jun 19;15(6):e117. [doi: 10.2196/jmir.2638] [Medline: 23782753]
32. Kludacz-Alessandri M, Hawrysz L, Korneta P, Gierszewska G, Pomaranik W, Walczak R. The impact of medical teleconsultations on general practitioner-patient communication during COVID- 19: a case study from Poland. PLoS One 2021;16(7):e0254960. [doi: 10.1371/journal.pone.0254960] [Medline: 34270587]
33. Nguyen AD, Frensham LJ, Baysari MT, Carland JE, Day RO. Patients’ use of mobile health applications: what general practitioners think. Fam Pract 2019 Mar 20;36(2):214-218. [doi: 10.1093/fampra/cmy052] [Medline: 29873708]
34. Sankaran J, Menachery SM, Bradshaw RD. Patient interest in video integration for after-hours telemedicine. J Am Board Fam Med 2020;33(5):765-773. [doi: 10.3122/jabfm.2020.05.190362] [Medline: 32989071]
35. Torrent-Sellens J, Díaz-Chao Á, Soler-Ramos I, Saigí-Rubió F. Modeling and predicting outcomes of eHealth usage by European physicians: multidimensional approach from a survey of 9196 general practitioners. J Med Internet Res 2018 Oct 22;20(10):e279. [doi: 10.2196/jmir.9253] [Medline: 30348628]
36. Steeman L, Uijen M, Plat E, Huibers L, Smits M, Giesen P. Out-of-hours primary care in 26 European countries: an overview of organizational models. Fam Pract 2020 Nov 28;37(6):744-750. [doi: 10.1093/fampra/cmaa064] [Medline: 32597962]
37. Moth G, Huibers L, Christensen MB, Vedsted P. Out-of-hours primary care: a population-based study of the diagnostic scope of telephone contacts. Fam Pract 2016 Oct;33(5):504-509. [doi: 10.1093/fampra/cmw048] [Medline: 27328678]
38. Moth G, Olesen F, Vedsted P. Reasons for encounter and disease patterns in Danish primary care: changes over 16 years. Scand J Prim Health Care 2012 Jun;30(2):70-75. [doi: 10.3109/02813432.2012.679230] [Medline: 22643150]
39. Huibers L, Moth G, Bondevik GT, et al. Diagnostic scope in out-of-hours primary care services in eight European countries: an observational study. BMC Fam Pract 2011 May 13;12:30. [doi: 10.1186/1471-2296-12-30] [Medline: 21569483]
40. Khan R, Khan A, Mohammed I, et al. Investigating the leading reasons for primary health care encounters and its implications for health care in Trinidad and Tobago. A systematic review. J Family Med Prim Care 2022 Sep;11(9):4949-4980. [doi: 10.4103/jfmpc.jfmpc_1803_21] [Medline: 36505547]

Abbreviations

GP: general practitioner
OOH-PC: out-of-hours primary care

Edited by A Kushniruk, E Borycki; submitted 30.08.23; peer-reviewed by J Keuper, L Getz; revised version received 13.08.24; accepted 13.09.24; published 15.11.24.

Please cite as:
Nebsbjerg MA, Bomholt KB, Vestergaard CH, Christensen MB, Huibers L
The Added Value of Using Video in Out-of-Hours Primary Care Telephone Triage Among General Practitioners: Cross-Sectional Survey Study
JMIR Hum Factors 2024;11:e52301
URL: https://humanfactors.jmir.org/2024/1/e52301
doi:10.2196/52301

Original Paper

The Role of Coherent Robot Behavior and Embodiment in Emotion Perception and Recognition During Human-Robot Interaction: Experimental Study

Laura Fiorini[1,2], PhD; Grazia D'Onofrio[3], PhD; Alessandra Sorrentino[1], PhD; Federica Gabriella Cornacchia Loizzo[2], MSc; Sergio Russo[4], MSc; Filomena Ciccone[3], PsyD; Francesco Giuliani[4], MSc; Daniele Sancarlo[5], MD; Filippo Cavallo[1,2], PhD

[1]Department of Industrial Engineering, University of Florence, Firenze, Italy

[2]The BioRobotics Institute, Scuola Superiore Sant'Anna, Pontedera (Pisa), Italy

[3]Clinical Psychology Service, Health Department, Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo (Foggia), Italy

[4]Innovation & Research Unit, Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo (Foggia), Italy

[5]Complex Unit of Geriatrics, Department of Medical Sciences, Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo (Foggia), Italy

Corresponding Author:
Laura Fiorini, PhD
Department of Industrial Engineering
University of Florence
Via Santa Marta 3
Firenze, 50139
Italy
Phone: 39 0552758663
Email: laura.fiorini@unifi.it

Abstract

Background: Social robots are becoming increasingly important as companions in our daily lives. Consequently, humans expect to interact with them using the same mental models applied to human-human interactions, including the use of cospeech gestures. Research efforts have been devoted to understanding users' needs and developing robot's behavioral models that can perceive the user state and properly plan a reaction. Despite the efforts made, some challenges regarding the effect of robot embodiment and behavior in the perception of emotions remain open.

Objective: The aim of this study is dual. First, it aims to assess the role of the robot's cospeech gestures and embodiment in the user's perceived emotions in terms of valence (stimulus pleasantness), arousal (intensity of evoked emotion), and dominance (degree of control exerted by the stimulus). Second, it aims to evaluate the robot's accuracy in identifying positive, negative, and neutral emotions displayed by interacting humans using 3 supervised machine learning algorithms: support vector machine, random forest, and K-nearest neighbor.

Methods: Pepper robot was used to elicit the 3 emotions in humans using a set of 60 images retrieved from a standardized database. In particular, 2 experimental conditions for emotion elicitation were performed with Pepper robot: with a static behavior or with a robot that expresses coherent (COH) cospeech behavior. Furthermore, to evaluate the role of the robot embodiment, the third elicitation was performed by asking the participant to interact with a PC, where a graphical interface showed the same images. Each participant was requested to undergo only 1 of the 3 experimental conditions.

Results: A total of 60 participants were recruited for this study, 20 for each experimental condition for a total of 3600 interactions. The results showed significant differences ($P<.05$) in valence, arousal, and dominance when stimulated with the Pepper robot behaving COH with respect to the PC condition, thus underlying the importance of the robot's nonverbal communication and embodiment. A higher valence score was obtained for the elicitation of the robot (COH and robot with static behavior) with respect to the PC. For emotion recognition, the K-nearest neighbor classifiers achieved the best accuracy results. In particular, the COH modality achieved the highest level of accuracy (0.97) when compared with the static behavior and PC elicitations (0.88 and 0.94, respectively).

Conclusions: The results suggest that the use of multimodal communication channels, such as cospeech and visual channels, as in the COH modality, may improve the recognition accuracy of the user's emotional state and can reinforce the perceived emotion. Future studies should investigate the effect of age, culture, and cognitive profile on the emotion perception and recognition going beyond the limitation of this work.

(JMIR Hum Factors 2024;11:e45494) doi:10.2196/45494

KEYWORDS

social robot; emotion recognition; human emotion perception; human-robot interaction; robot cospeech gestures evaluation

Introduction

Background

During the last decade, there has been increasing interest in research on socially assistive robotics aimed at realizing intelligent robotic solutions for health care and social assistance. We experience an evolution of social robot applications; indeed, they moved from the role of *concierge* and *helper* [1] toward the role of *companion* and *therapist* [2,3]. Social robots have the potential to contribute to the greater good of society; indeed, it has been demonstrated that they can support everyday life as companions and the health care system from *logistics* to *assistance and rehabilitation*. Thinking to include social robots in the care chain, they can be used to reduce stress, anxiety, and pain in children [4]; they can be integrated into conventional behavioral and cognitive therapies for both children and adults who struggle with social anxiety [5]; or they can be used to promote mental health [6]. A review by Hung et al [7] showed evidence that Paro robots can reduce negative emotions in patients, promoting a positive mood and improving social engagement. Rossi et al [8] demonstrated that social robots are effective in decreasing stress in children accessing the emergency room. As the complexity of the robot task increases, social robots are required to perform more complex perceptual, cognitive, and interactive functionalities. This is the case in long-term interactions in which robots and users should establish meaningful communication, emotional awareness, and reliable engagement.

In this context, the human-robot interaction (HRI) field has become crucial, and it is now compelling to better understand how humans perceive, interact with, or accept these machines in social and real contexts. Researchers are also debating on defining the factors that can influence the perceived social capabilities and intelligence of a robot [9,10]. De Graaf et al [11] highlighted the significance of the robot's social capability, emphasizing the importance of 2-way interaction where a robot is expected to respond to humans in a social manner. In addition, De Graaf et al [11] underlined that a social robot should *also display thoughts and feelings* and should be *socially aware of the environment*, among other issues. When a robot failed to perform this 2-way interaction, people were disappointed and experienced a sense of dissonance. In other words, when interacting with a social robot, especially a humanoid robot, we expect to use the same mental structure and social rules that guide us in human-human communication, expecting empathetic interaction because they are perceived as social actors [12].

From a roboticist or engineering point of view, these concepts are translated into the design and development of behavioral models that can guarantee an efficient and reliable 2-way interaction [13,14]; they should *perceive* and *show* emotions (and social norms) and thus be understood by humans with whom they are interacting. The key challenge in this field is to provide robots with cognitive and affective capabilities, developing architectures that allow them to establish empathetic relationships with users, which can foster long-term and meaningful interactions. From an implementation perspective, the design and the deployment of a socially capable social robot comprises 2 essential parts. The first is devoted to designing and implementing a consistent and congruent emotional behavioral architecture that makes the robot react or act to the environment (ie, *display thoughts and feelings*). The capabilities of a user to understand the emotions displayed by a robot have been explored in different settings [15,16]. Examples of actions can include the expression of congruent cues such as facial expressions [17], changes in the color of the eyes, movement of the upper limbs [16,18], or smart navigation strategies [19]. In contrast, the other part is more focused on the robot's perception of the user's emotional response to these behaviors [20], with special attention to contextualizing its action and reaction according to the living contexts and habits or preferences of the person with whom it is interacting (ie, *being socially aware of the environment*) [21].

Related Work on Emotion and Social Robots

The ability of a robot to perceive the nonverbal cues of the user, which convey user emotion and intent, plays a key role in the development of social robots capable of performing meaningful interactions [22,23]. In this sense, humans' gaze, body posture, cospeech gestures, and facial expressions play a leading role in defining the context of the interaction, helping the robot to correctly classify the experience, and associating it with informative content [21]. The development of such abilities, for a researcher in the field of robotics, translates into the use of multimodal sensor modality and the implementation of several complex algorithms to endow robots with different cognitive and social capabilities. The visual modality is the most commonly used [24] because it can detect nonverbal behaviors that are representative of the emotional state of users without requiring them to wear any external sensor. Alternatively, wearable sensors [25] can be used, also using a multimodal approach, to overcome the problems related to occlusion and low light. Other algorithms or modules were implemented to perform multiperson tracking [26], speech recognition [27,28], and automatic engagement detection [29]. A recent review paper [24] provides a deep insight into the most used methods and approaches.

For the *showemotion* part, robots must exploit several channels (ie, auditory, visual, cospeech, and gestures) and mechanisms (eg, body posture, facial expressions, vocal prosody, touch, and gaze) to communicate their "internal emotional status" and intentions authentically and clearly [30]. Thus, the capabilities of a user to understand the emotions displayed by a robot have been explored in several settings [31]. Over the last few years, several attempts have been made using both video-simulated robots and real robots. Guo et al [20] showed participants 5 different emotions using the humanoid robot called Alpha2, and they were asked to rate the perceived emotion using the Self-Assessment Manikin questionnaire (SAM; only valence and arousal dimensions) [32]. In contrast, Barchard et al [33] conducted a web-based study to evaluate the perception of a robot's social intelligence by showing videos of robot interactions. However, the embodiment and the appearance of social robots play important roles in the perception of the robot; therefore, video-based elicitation could introduce some bias in the analysis of perceived emotion. This is why other research has relied on investigating the emotion perceived during a real HRI. This is the case of Bagheri et al [34], who asked participants to watch 6 performances of America's Got Talent Show on Pepper's tablet that are expected to evoke the 6 basic emotions. Rossi et al [35] and Staffa et al [36] relied on movie trailers to evoke emotions. However, they used nonstandard videos, making it challenging to identify the target emotion in a recognized and standardized manner, as the elicited emotion through the video clips is not known a priori, and consequently, it is difficult to define the role of the robot (and its embodiment) in the elicitation process.

Research groups have recently begun to study the effects of multimodal channels on communication. Studies conducted with embodied conversational agents showed that incongruent emotional stimuli (eg, auditory and visual stimuli) can result in adverse consequences on user rating; conversely, congruent stimuli can facilitate the recognition of emotions [37]. Other researchers have also studied the role of nonverbal behavioral cues while interacting with robots. Movie clips showing coherent and incoherent robot behaviors are often used to elicit emotional responses from users with respect to those induced by movie clips [15,16,18,35]. For instance, Rossi et al [16] investigated how an incoherent nonverbal robot's behavior with respect to the presented emotion can produce a type of humorous effect. Tsiourti et al [18] investigated how contextual incongruence (ie, a robot's reaction conflicts with the socioemotional context) can confuse the observers, decreasing the accuracy of the perceived emotion. Nevertheless, such a cospeech robot's behavior was used in addition to a nonstandard method of emotion elicitation, as previously remarked; thus, it is not easy to understand the role of the robot's behavior with respect to the emotional context. Therefore, it is important to understand how the robot's nonverbal behavior might shape the human perception of the showed emotion elicited through standard emotionally labeled visual data sets and, at the same time, observe the robot's emotion recognition accuracy rate. Although previous studies have shown a correlation between the robot's nonverbal action and perceived emotion, there is a lack of use of standard elicitation modalities.

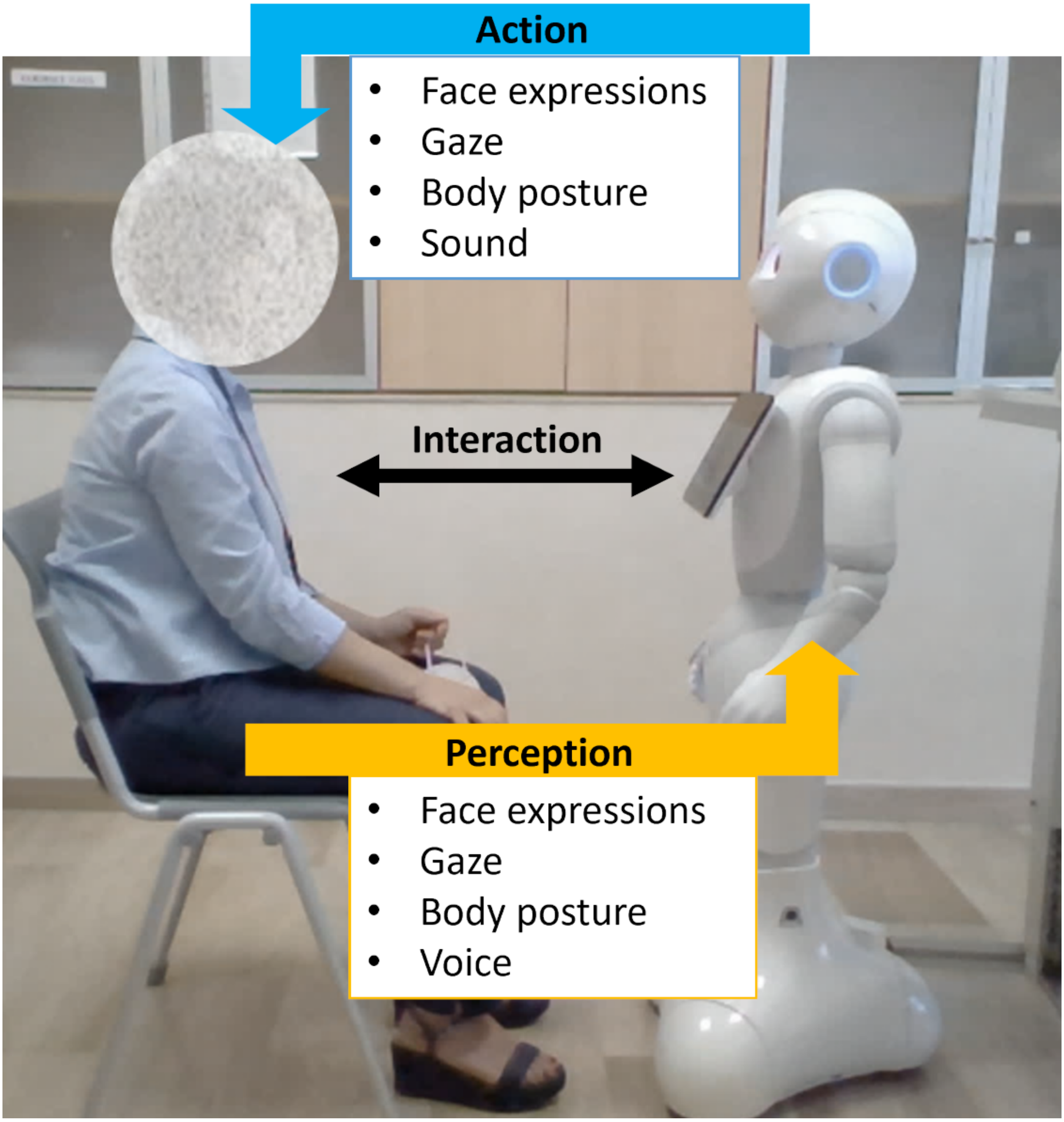
Therefore, in this work, we present the results of 3 experimental sessions to observe the performance of the robot in recognizing users' emotions as well as to investigate the difference (if any) in eliciting emotions in humans when using a social robot (with or without coherent behavior) rather than a PC. We plan to use a standard data set of pictures, namely, the International Affective Picture System (IAPS) [38], to elicit emotions in users. Particularly, the robot will use a multimodal behavior (ie, head movements, vocal reinforcement, and body gestures) to interact with the participants while showing the graphical emotions by establishing social binding, whereas the PC will provide emotion elicitation only through a graphical interface. The 2 graphical interfaces have been designed to provide the same information to the user but using different communication channels. In this context, the aim of this work is dual. First, it aims to investigate the increase in the user's emotional perception during the interaction with a robot with respect to a PC (Figure 1, blue arrow). In particular, this work investigates the role of the robot's coherent nonverbal behavior in emotion perception by consequently assessing the impact of robot embodiment and, eventually, its coherent behavior. Robot nonverbal cues are manipulated with respect to a mapping between the main associated emotion and cospeech gestures that can be generated on the robot. At the end of each interaction, the participants were asked to self-assess their perceived emotions. In this study, we used the emotion classification proposed by Russel et al [39], which relies on 3 variables, namely, valence, arousal, and dominance. Valence describes the degree to which a stimulus causes a positive or negative emotion, arousal refers to the intensity or level of energy invested in the emotion, and dominance reflects the extent of perceived control over the emotional response when facing the stimulus. The collected answers were analyzed to answer the following research questions (RQs):

1. RQ1: Emotion elicited through a humanoid robot interacting with coherent emotional behavior is rated higher than emotions elicited by a web application in terms of emotional valence, arousal, and dominance.
2. RQ2: There are significant differences in terms of emotional valence, arousal, and dominance between a robot showing coherent behavior rather than a robot that it is not moving at all (static condition).
3. RQ3: The embodiment of the humanoid robot will not affect the emotion perception compared with the web application.

Second, this study aims to assess the accuracy of the robot in recognizing the elicited emotion in the participants (Figure 1, yellow arrow). The ability to infer and interpret emotions plays a key role in establishing intuitive and engaging HRIs. On the one hand, a robot endowed with emotion recognition skills can adapt its behavior based on the detected user emotion [22]. On the other hand, a robot expressing recognizable emotions positively influences the evaluation of its capabilities [40]. In particular, features related to facial expressions were extracted, preprocessed, and analyzed with 3 supervised machine learning techniques to verify the following RQ:

1. RQ4—There is no difference in the robot emotion recognition accuracy despite the elicitation modalities (robot or web application).

Figure 1. Two-way interaction proposed in this study. To improve the human-robot interaction, the robot should perceive the user's behavior (yellow arrow) and plan appropriate action (blue arrow).



In our previous studies [41,42], we evaluated the perceived acceptance and the recognition rate of having a robot that acts coherently and incoherently despite the standard emotion showed with respect to the standard elicitation modality. In contrast, in this study, we focus only on coherent behavior by comparing it with a standard web application that runs on a PC. In addition, instead of focusing on evaluating how the robot's acceptance is modulated according to the elicitation modality, we focused on the perceived emotion evoked.

Methods

Instrumentation

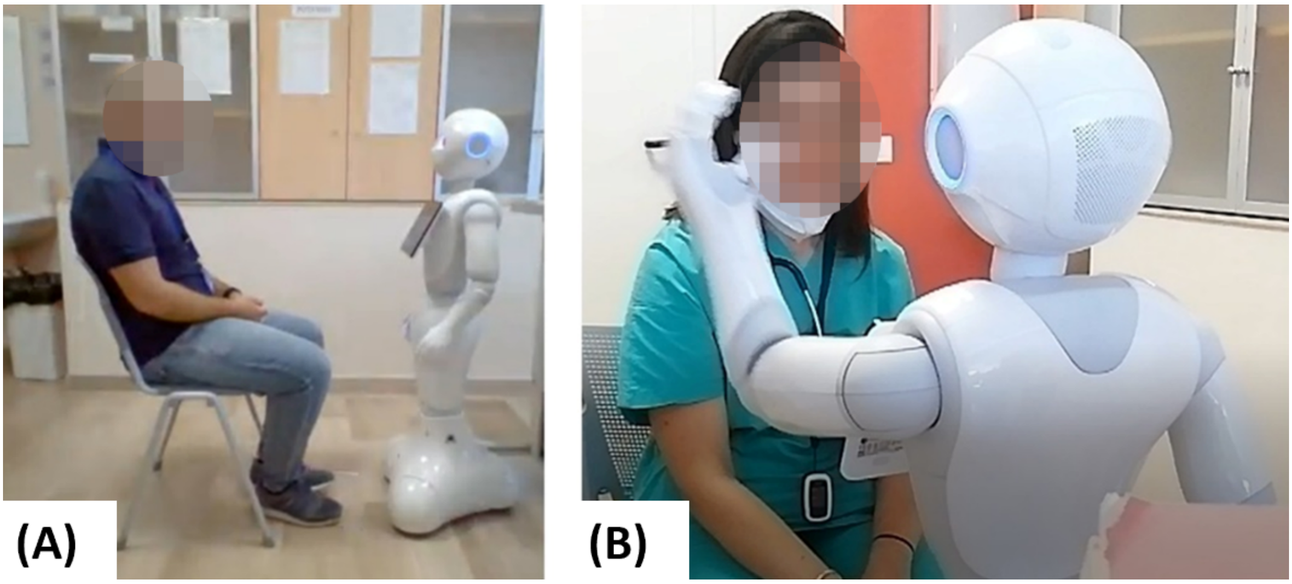
The instrumentation is composed of the following elements: (1) a Pepper robot (Aldebaran, United Robotics Group) or a PC, (2) the RoboMate (Behaviour Labs) interface for cospeech gestures, (3) a custom interface that contains pictures from the IAPS for eliciting emotion, and (4) an external camera placed on Pepper to record the participants' emotions during the interaction. Pepper is a humanoid robot that is widely used for experimentation in socially assistive robotics. It is 120 cm tall, weighs 28 kg, and has 20 df, including 1 head, 2 arms, and 1 wheeled base. In addition, it has a tablet on the front. Robot coherent behavior was managed through the RoboMate interface [43] to animate Pepper, when necessary, selecting among the behaviors classified as "positive social stimulus" or "negative social stimulus." The selected stimulus was modeled by a psychologist using 3 modalities: body gestures (upper limb and head), gaze, and sound. IAPS is a database of images devoted to eliciting standardized emotions [44]. It was developed by the Center for Emotion and Attention at the University of Florida. This database is commonly used in psychological studies on emotions and attention. Each image in the data set is labeled with the corresponding emotion, thus enabling researchers to properly select the stimulus. In this study, 60 images were selected from the team of psychologists of the hospital "Casa Sollievo della Sofferenza." According to the IAPS valence dimension, 21 of the selected images were rated as positive, 19 as negative, and 20 as neutral. A customized web-based interface was developed to standardize the emotional stimulation when using 2 different communication channels (a robot and a PC).

Experimental Setup

A psychologist welcomed the participant, briefly explaining the experimental setup, including how to use the evaluation tool. It is important to emphasize that the participant was not aware of the real objective of the experimentation, thus avoiding interference with the experience. To properly investigate the RQs, each participant underwent 1 of the following elicitation modalities.

1. Static (STA) behavior: Pepper robot has its arms along the body in a neutral position (Figure 2A). Pepper's face was looking at the participant but without any animacy. Pepper displayed IAPS images on its tablet through the customized web application.
2. Coherent (COH) behavior: Similar to the STA condition, the IAPS images were shown on Pepper's tablet. Using the RoboMate application, the psychologist assigned a coherent behavior to Pepper with the shown images. In particular, the psychologist can choose and combine 3 modalities for elicited emotions: body gesture (upper limb and head), gaze, and sound, which are available on the RoboMate application (Figure 2B). For example, in the case of positive emotion, Pepper's gestures were chosen to look friendly; it should look to the user direction, and the voice gave positive reinforcements.
3. PC: For this experimental condition, we used a PC instead of the Pepper robot. Participants were asked to evaluate the images shown on a PC through the customized web application.

Figure 2. Experimental setup. The participants were interacting with Pepper robot during the experimentation. (A) Participants were asked to sit in front of the robot and watch the images on its tablet. (B) If the participant belonged to the coherent elicitation modality group, the Pepper robot would move its arms, eyes, and head.



The participant was asked to sit in front of the technology (ie, Pepper robot or PC). If the user interacts with Pepper, Pepper is placed 0.5 to 0.6 m far from the user (ie, personal distances [45]); in the case of interaction with the PC, the user is requested to sit and interact with the computer as he or she will commonly do.

Each stimulus was shown for 7 seconds, and at the end, the participant was asked to fill out the SAM [32], as adapted in the study by Gatti et al [46] directly on the robot or on the computer after each picture. SAM is an emotion assessment tool that uses graphic scales, depicting cartoon characters expressing 3 emotional elements (valence, arousal, and dominance). Each participant was asked to rate the domains by selecting an image that corresponded to a score between 1 and 9. A picture of the interface is presented in Multimedia Appendix 1.

At the end of the experimental session, each participant completed 60 SAM questionnaires. The psychologist was present during the test, and she or he was ready to intervene in case of necessity. All the tests were performed at the "Casa Sollievo della Sofferenza" research hospital.

Ethical Considerations

The approval of the study for experiments using human participants was obtained from the local Ethics Committee on Human Experimentation (register code 3038/01DG). All participants signed an informed consent form before participating in this study, and pictured participants provided written informed consent to allow their image to be published. The data were pseudoanonymized and stored on a GDPR-compliant server.

Participants

Participants were recruited from July 2020 to February 2021 from employees and staff of the "Casa Sollievo della Sofferenza" research hospital located in Apulia (San Giovanni Rotondo, Foggia) using convenience sampling. Participants were excluded if they had a hearing or visual impairment. Recruited participants were then randomly assigned to undergo 1 of the 3 experimental conditions (ie, STA, COH, and PC). Sociodemographic information (age, education, and sex) was collected to verify the similarities between the groups.

Data Analysis

Overview

Owing to the sample size of each cohort, the nonparametric statistic was used, particularly the Kruskal-Wallis test and chi-square test, to investigate significant differences between participants' groups in terms of age, sex, and educational level. The significance level was set at $P=.05$. The following paragraphs describe the analysis performed on the SAM questionnaires and the data collected from camera sensors.

Emotion Perception Analysis

A total of 60 SAM questionnaires were collected for each participant. The average values of the valence, arousal, and dominance domains were computed for each selected image of each group of elicitation modality (ie, STA, COH, and PC). Differences were analyzed with the Kruskal-Wallis test ($P<.05$) and post hoc evaluated with the Mann-Whitney *U* test (with Bonferroni correction) used to identify between which pair of elicitation modes the difference has occurred.

Emotion Recognition Analysis

Data from the camera were processed and examined offline. The recordings were initially analyzed [47] to ensure that only the frames featuring the face of the person performing the test were included in the study. Then the recordings were segmented, providing short videos that corresponded to the user's reaction to each image proposed, totaling 60 videos per user. The OpenFace toolkit [48] was used to extract 150 features related to gaze and facial expression from each video as well as the quality (ie, confidence) of the extracted features. The data were filtered according to the confidence score (frames with a confidence score <0.90 were discarded). The data were then labeled based on the IAPS-defined emotions (ie, positive, negative, and neutral). Data were normalized and selected. Only features with a correlation coefficient of <0.85 were picked from the initial data set, avoiding those with a high correlation coefficient (which may represent redundant information). The data of the merged data set were then separated into sub–data sets (one for each participant), and emotion classification was performed using the selected features. In this study, we rely on state-of-the-art methods used for emotion recognition [24] to facilitate a comparison with other works. The 3 supervised classifiers used are support vector machine (SVM), random forest (RF), and K-nearest neighbor (KNN). To classify the data by participant, a 10-fold cross-validation procedure was applied, and the outputs were organized in a confusion matrix. The classification performance was assessed in terms of accuracy, precision, recall, and F-measure [49]. The calculations were computed using MATLAB 2020a. More details on emotion recognition analysis are available in Multimedia Appendix 2 [24,47-49].

Results

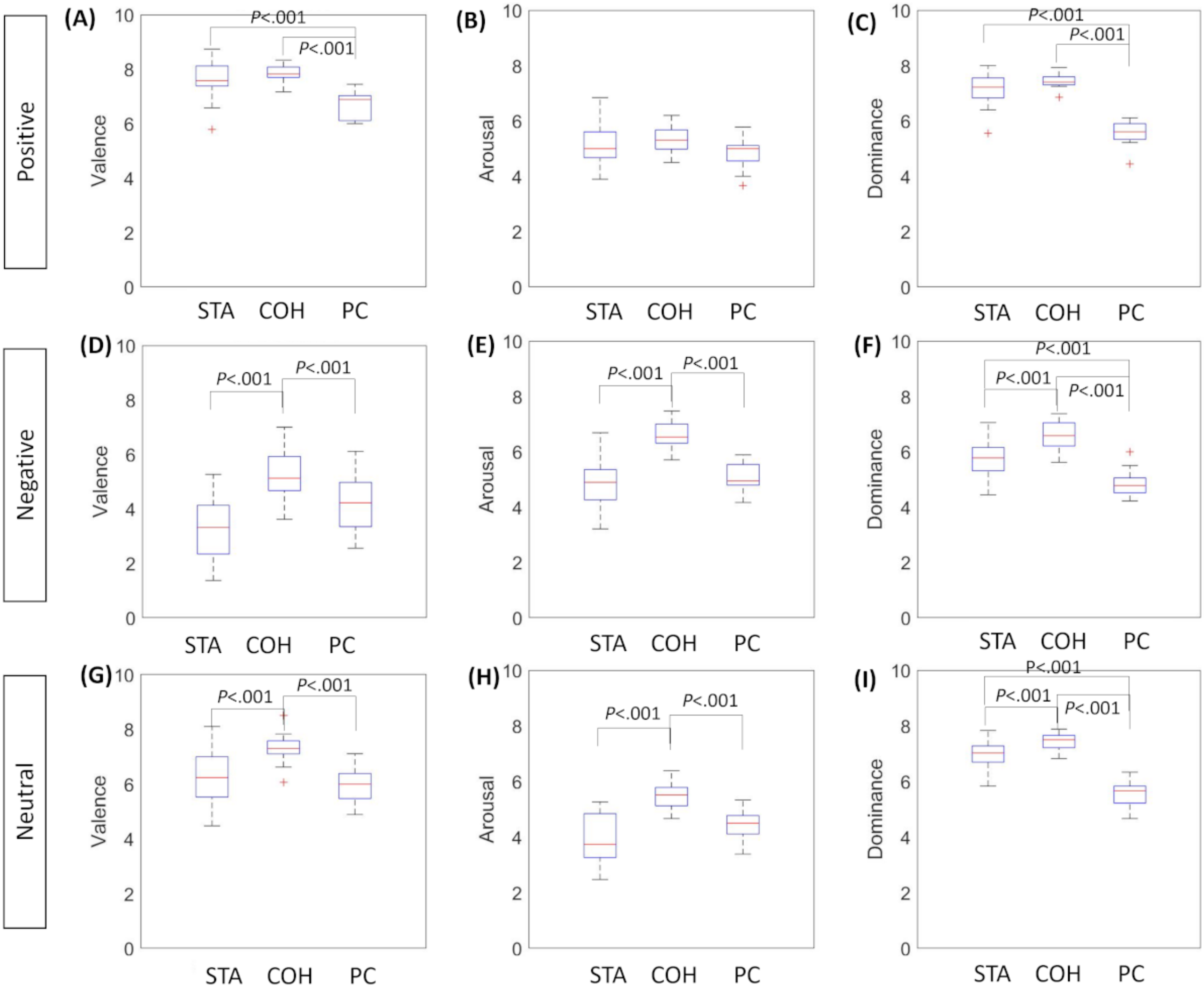
Description of the Participant Cohort

A total of 60 participants were involved in this study, 20 for each modality, resulting in 3600 interactions with technologies. In total, 3 participants were excluded from the analysis of perceived emotion because not all SAM evaluations were correctly saved after each elicitation. In case of missing SAM values, these ratings were removed from the analysis of average values. Finally, 57 participants were included in these subgroups of analyses linked to RQ1, RQ2, and RQ3. Regarding the recognition of emotion using machine learning techniques (linked to RQ4), a total of 53 participants were included in the analysis. A total of 7 participants were excluded because of technical problems related to the quality of the recorded images. The statistical tests did not indicate any difference between the 3 participant cohorts regarding age, sex, and educational level. The participant demographics and educational analyses are reported in Multimedia Appendix 3.

Participants' Perceived Emotion Results

The results underline significant differences ($P<.001$) in the perceived emotions according to the different elicitation modalities, except for the arousal elicited with the positive images (Figure 3). The median and IQR values are fully reported in Multimedia Appendix 4. As for valence, the robot with coherent behavior elicited significant differences ($P<.001$) and higher values in terms of valence, arousal, and dominance domains compared with the other 2 modalities for negative and neutral emotions. In terms of negative valence, the participants perceived fewer negative emotions with the coherent robot than with the other 2 modalities. For positive valence, elicitation with the web application is significantly different from that with the robot ($P<.001$).

Figure 3. Self-Assessment Manikin Questionnaire results for the 3 elicitation modalities. Boxplot matrix (A), (B), and (C) denote valence, arousal, and dominance for the positive elicitation, respectively; (D), (E), and (F) denote valence, arousal, and dominance for the negative elicitation, respectively; (G), (H), and (I) denote valence, arousal, and dominance for the neutral elicitation, respectively; asterisks on boxplot remark the significant differences evaluated with the Mann-Whitney post hoc test corrected with Bonferroni. COH: coherent; STA: static.



Regarding arousal, the coherent robot was rated higher than the other 2 modalities, but there were significant differences (*P*<.001) only for negative and neutral emotions, whereas for positive arousal, the results, depicted in Figure 3, highlight only a trend. All the *P* values are reported in Multimedia Appendix 4.

The participants stimulated using the robot rated significantly higher dominance across all 3 emotions rather than the cohort that used the PC in the test. As for positive elicitation, we found significant differences (*P*<.001) between the cohort stimulated with the PC and those stimulated with the robot (ie, static behavior and coherent behavior). Indeed, the participants rated the emotions (in terms of valence and arousal) elicited by the robot more than the ones elicited using the PC. All *P* values are reported in Multimedia Appendix 4.

Robot's Emotion Recognition Results

Because of technical issues 1848 frames pertaining to the PC modality were removed from the analysis during the preprocessing. At the end, the total number of samples included in this study was 296,677 for the STA modality, 228,170 for the COH modality, and 103,758 for the PC modality. The number of columns in each data set corresponded to the number of features selected using the correlation analysis method. The following features were selected (Figure 4):

1. The x-, y-, and z-coordinates of the eye gaze direction vector for eye 0 (3 features).
2. The z-coordinate of the eye gaze direction vector for eye 1 (1 feature).
3. The x- and y-coordinates of the location of the landmark 8 (the leftmost in the image) of the eye 0 (2 features).

The 53 data sets were fed into 3 classifiers (SVM, RF, and KNN) [24]. The data sets were uniformly distributed across the 3 groups, as presented in Table 1.

Figure 4. Selected features. (A) Face and (B) eye landmarks extracted with OpenFace software. The landmark 8 in panel B was chosen after the feature selection.

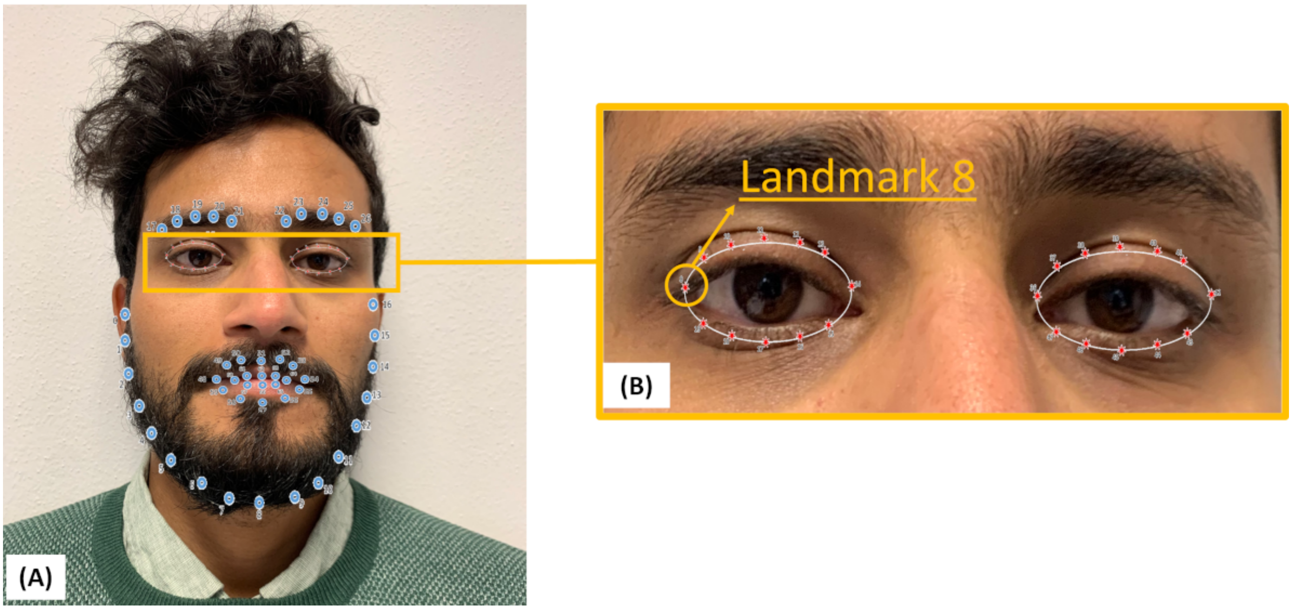


Table 1. Distribution of data set instances.

Group	Positive, n (%)	Negative, n (%)	Neutral, n (%)
Static (n=296,677)	103,992 (35.03)	94,257 (31.77)	98,499 (33.2)
Coherent (n=228,170)	70,710 (30.99)	74,383 (32.6)	83,077 (36.41)
PC (n=103,758)	35,195 (33.92)	32,072 (30.91)	36,492 (35.17)

Accuracy, precision, F-measure, and recall were calculated as the mean values from the participants in the same experimental cohort. According to the findings, the KNN classifier offers the best classification results, with an accuracy of up to 0.88 for STA behavior, 0.97 for COH, and 0.94 for PC. The SVM classifiers, in contrast, had the lowest results (accuracy of up to 0.57, 0.67, and 0.68 for STA, COH, and PC, respectively); hence, they were excluded from further research. Compared with the RF classifier, the KNN classifier has the best F-measure (>0.88).

Table 2 presents the complete results for the KNN and RF classifiers, including the accuracy, F-measure, precision, and recall for each group. According to the overall trend, the COH modality achieves a high level of accuracy when compared with the STA and PC elicitations. In terms of the other indicators, the COH was better with the KNN classifier and slightly worse with the RF classifier when it came to elicitation with the PC.

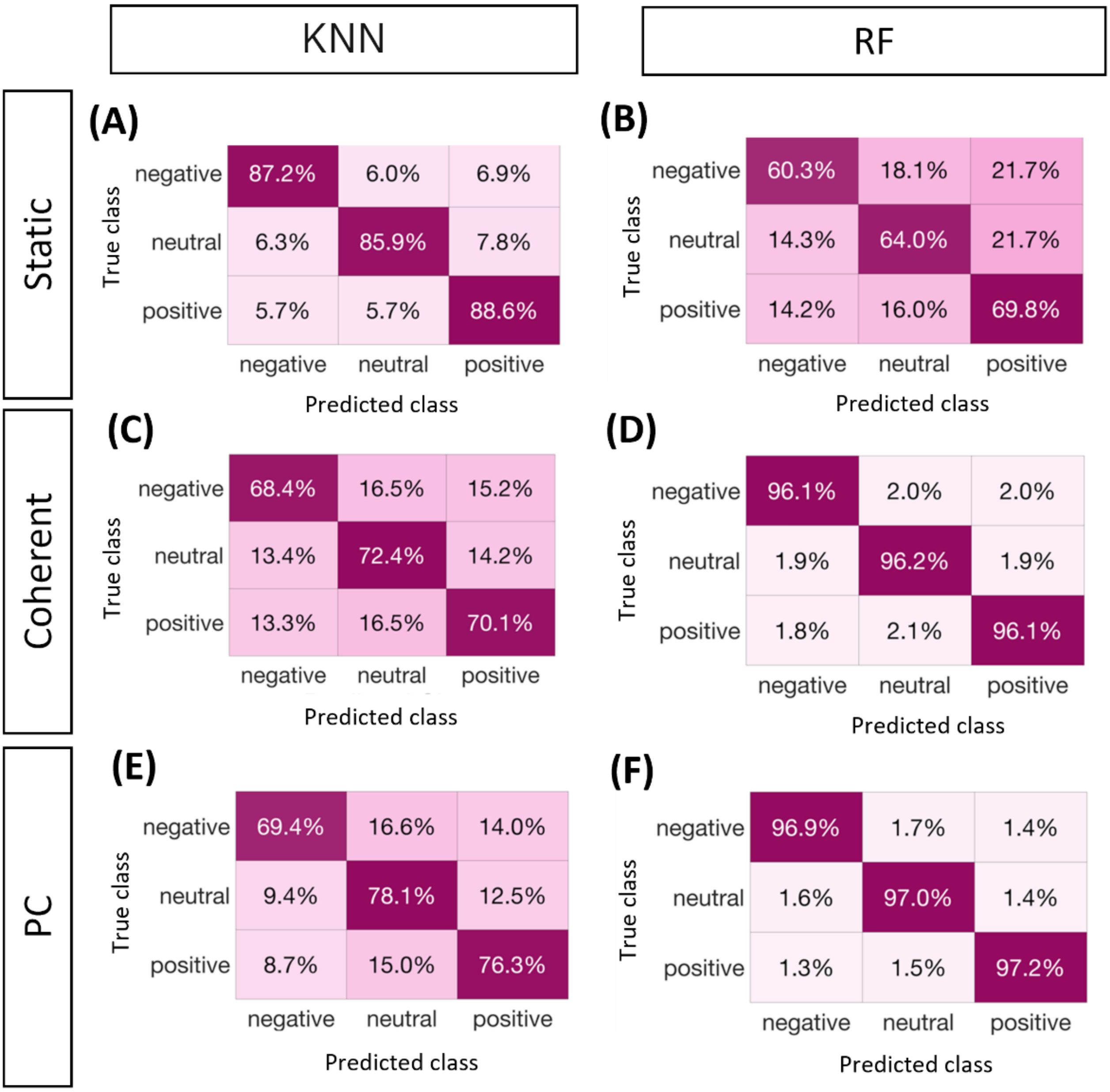
Table 2. Performance of K-nearest neighbor (KNN) and random forest (RF) classifiers[a].

Group	Accuracy		Precision		F-measure		Recall	
	KNN	RF	KNN	RF	KNN	RF	KNN	RF
Static	0.88	0.65	0.88	0.65	0.88	0.65	0.88	0.65
Coherent	0.97	0.73	0.96	0.72	0.96	0.72	0.96	0.72
PC	0.94	0.74	0.94	0.74	0.94	0.74	0.94	0.74

[a]Mean values are used to calculate the results.

Confusion matrices (Figure 5) for the 3 elicitation modalities were generated to investigate the performance of the classifiers in recognizing the 3 selected emotions. The positive emotion was often better identified, whereas the negative emotion was the least recognized. When the user is stimulated with the robot with coherent modality and the PC, the RF classifier performs better than the KNN classifiers in distinguishing emotions. The KNN classifier appeared to perform better in the static modality than in the other 2.

Figure 5. Confusion matrices for K-nearest neighbor (KNN) and random forest (RF) classifiers. The confusion matrices obtained for the 3 elicitation modalities (ie, static, coherent, PC) are reported considering only the KNN and the RF classifiers.



Discussion

Principal Findings

The results confirm RQ1 (“A humanoid robot interacting with coherent emotional behavior is rated higher in terms of emotional valence, arousal, and dominance compared to the web application”) because the COH robot is rated significantly higher for all SAM dimensions (except positive arousal) with respect to the PC condition (Figure 3). However, it is worth noting that when speaking of negative elicitation, receiving a higher rating of valence means that the stimulus with the COH condition was perceived less negatively than the ones elicited with the others. RQ2 (“There are significant differences in terms of emotional valence, arousal, and dominance between the static robot compared to the robot that shows movement”) is confirmed for the 3 dimensions for negative and neutral emotions (Figure 3). It is worth noting that these results confirm that the robot’s movements cause the negative emotion to be perceived as less negative (STA valence median value=3.32; COH valence median value=5.13). As for the positive emotion, there were no significant differences, which could suggest that the robot’s behavior per se did not affect the perception of the positive emotion.

The presented results did not confirm the RQ3 (“The embodiment of humanoid robot will not affect the emotion perception compared to the web application”) for all elicited emotion and SAM constructs. Indeed, there were no significant differences between the STA and the PC elicitation for valence and arousal measured during negative and neutral elicitation (Figure 3). Conversely, COH and STA differed significantly from PC in terms of positive elicitation. These results suggest that robot embodiment per se has a role in the perception of dominance associated with negative and neutral emotions with respect to a standard web interface. On the contrary, as for the positive emotion, embodiment seems to play a key role because both COH and STA elicitations differ from the web application in terms of valence and dominance.

The ability to recognize user emotions is a fundamental step in the development of socially aware robots (RQ4). The emotions were recognized with an average accuracy >0.88 over the 3 elicitation conditions. In addition, the amount of gaze also depends on the interpersonal dynamics between the partner and their personalities and on the intent of using gaze to communicate their internal state. Therefore, it is important to measure it during interactions. As shown in Table 2, the accuracy of COH stimulation was higher than that of the other 2 methods. In addition, the results in the confusion matrices were aligned with the perceived emotion (Figure 5). According to the SAM results, the valence ratings for positive elicitation elicited with PC were significantly different from the other 2 with lower median values. This trend is reflected in the confusion matrices obtained using RF classifiers.

Comparison With Prior Work

Previous qualitative studies have pointed out how incoherent behavior can generate hilarious reactions in humans [16]. The presented results suggest that we can observe something similar, even if the stimulus is coherent. It appears that the robot's behavior somewhat distracts from perceiving negative emotions, even if the behavior is aligned with the shown emotion. In addition, as confirmation, positive emotion was perceived significantly more positively than the PC modality, suggesting that robot movements make the robot more positive. Consequently, these results suggest that it is important to tailor the reaction of the robot appropriately to elicit a specific emotion. Indeed, if we need to stimulate—for a certain reason—negatively the users, we need to reduce the robot's body expression because they can decrease the perception of negative emotions. Alternatively, if we need to provide positive feedback to users, the combined actions of both verbal and nonverbal communication can be used.

A previous study [36] compared robots and web applications that focused on investigating preferences and acceptance, and they did not find any significant deviation in the quantitative results. In contrast, in this study, we focus on human emotion perception, and this perception seems to be influenced or biased by the emotion itself and the robot's movement. This finding highlights the significance of not just robot embodiment but also its cospeech gestures in designing social agents, particularly when evaluating all dimensions of emotions. Methodologically, the presented findings carry significant implications for the design of experimental protocols. Evaluating HRI cannot rely solely on videos, as they overlook the importance of physical interaction. In the literature, some papers [33] provide a user impression without direct interaction with a robot; the collected results can be biased because the participant missed the contribution in the perception related to embodiment. Take, for instance, the scenario where you are testing a new game application or software on a tablet meant for eventual integration into a robot. Particularly when assessing emotions, it's crucial to approach the generalization of results with caution. In this sense, the result could be altered because the emotions elicited could not be directly applicable when interacting with an embodied agent.

The results obtained for the STA robots with the KNN and RF classifiers were slightly improved with respect to the results obtained in our previous work [42] (average accuracy was equal to 0.85 with KNN and 0.98 with RF), where we used them in combination with encoders. It is also worth noting that after the feature selection process, only the features related to gaze were retained in the analysis. Gaze is extremely important in managing interpersonal interaction and also during human-robot conversation; indeed, it can be correlated with user engagement during conversation or mutual tasks [50,51].

Limitations of the Study

The limitations of this study were mainly related to the cohort of recruited participants. First, both cognitive and cultural backgrounds are factors that can influence the perception of emotions [52]. Some neurological pathologies (eg, Parkinson disease) can affect facial expressions, whereas others can affect body gestures and language (eg, autism spectrum disorders and apathy); consequently, emotion recognition accuracy in such cases can change. The RQs do not focus on investigating their role in emotion perception; consequently, we recruited cohorts of people comparable for cultural background and cognitive status to limit the impact of these factors. The second limitation of this study refers to how the emotion is evaluated; in this study, we evaluated each SAM dimension separately. The third limitation of this study relies on the supervised machine learning techniques used. In this study, we rely on standard supervised methods because our main RQs are not focused on learning methods; therefore, we apply the most used techniques.

Future Directions

In this context, by applying the findings and implications of this paper in the health care context, we can conclude that it is important to tailor the reaction of the robot properly; indeed, if we need to stimulate—for a certain clinical reason—the users negatively, we need to reduce the robot's body expression because they can decrease the perception of negative emotions. Alternatively, if we need to give positive feedback to the users, for instance, during an exercise, we can use the combined action of both verbal and nonverbal communication. To overcome the limitations of this study, future research can be planned to extend the study to include a different group of participants with some cognitive and physical disorders and different cultural backgrounds to evaluate the effect of these factors on emotion perceptions. Future studies should also investigate whether there are differences in combining valence-arousal domains, as proposed in other studies [16,53]. Finally, the data could be analyzed using also deep learning and reinforcement learning techniques.

Conclusions

This study aimed to investigate the role of robot embodiment and its behavior in emotion perception and recognition using a standard elicitation model. In total, 4 RQs were investigated to understand how the robot's nonverbal behavior might shape the human perception of the showed emotion elicited through a standard data set and, at the same time, to observe the robot's emotion recognition accuracy rate. This study presents an experimental setup in which 60 participants were asked to interact with 2 embodied agents (ie, a robot or tablet) that acted as emotion facilitators by showing them 60 standard pictures. The results underline the good recognition accuracy of the perception modules of the robot. Indeed, we can correctly classify the valence of the emotion (ie, positive, neutral, and negative) with an accuracy of up to 0.97 in the best case. According to the results, robot embodiment affects the perception of dominance significantly compared with web applications, which means that participants' emotions were less controlled when they were interacting with an embodied agent.

Acknowledgments

The authors would like to thank all the people involved in the study. This study was funded by “An adapted behavioral robot model with advanced cognitive/physical interaction capabilities for assessment and rehabilitation of neurodegenerative diseases (DESTINI)” founded by Unione Europea—NextGenerationEU (CUP: B55F21007810001).

Data Availability

The data sets generated and analyzed during this study are not publicly available because of the scope of the consent signed by the patient participating in the study but are in part (no video recordings) available from the corresponding author on reasonable request.

Authors' Contributions

The conceptualization was done by LF, GDO, and F Cavallo. Data curation was conducted by GDO, F Ciccone and AS. LF, FGCL, and AS were responsible for the data analysis. LF acquired the funding. The methodology was developed by GDO, LF, F Cavallo, and FG. The investigation was carried out by GDO and F Ciccone. AS and SR handled the software. DS, FG, and F Cavallo provided supervision. LF was responsible for the original draft of writing, while all authors contributed to the writing, review, and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Self-Assessment Manikin questionnaire.
[PDF File (Adobe PDF File), 134 KB - humanfactors_v11i1e45494_app1.pdf]

Multimedia Appendix 2
Emotion recognition analysis.
[PDF File (Adobe PDF File), 194 KB - humanfactors_v11i1e45494_app2.pdf]

Multimedia Appendix 3
Participants' description.
[PDF File (Adobe PDF File), 120 KB - humanfactors_v11i1e45494_app3.pdf]

Multimedia Appendix 4
Median and IQR values computed for each elicited emotion.
[PDF File (Adobe PDF File), 143 KB - humanfactors_v11i1e45494_app4.pdf]

References

1. Cavallo F, Esposito R, Limosani R, Manzi A, Bevilacqua R, Felici E, et al. Robotic services acceptance in smart environments with older adults: user satisfaction and acceptability study. J Med Internet Res 2018 Oct 21;20(9):e264 [FREE Full text] [doi: 10.2196/jmir.9460] [Medline: 30249588]
2. Mannion A, Summerville S, Barrett E, Burke M, Santorelli A, Kruschke C, et al. Introducing the social robot MARIO to people living with dementia in long term residential care: reflections. Int J Soc Robot 2019 Jun 14;12(2):535-547. [doi: 10.1007/s12369-019-00568-1]
3. Pfadenhauer M, Dukat C. Robot caregiver or robot-supported caregiving? Int J Soc Robot 2015 Jan 30;7(3):393-406. [doi: 10.1007/s12369-015-0284-0]
4. Dawe J, Sutherland C, Barco A, Broadbent E. Can social robots help children in healthcare contexts? A scoping review. BMJ Paediatr Open 2019 Jan 31;3(1):e000371 [FREE Full text] [doi: 10.1136/bmjpo-2018-000371] [Medline: 30815587]

5. Rasouli S, Gupta G, Nilsen E, Dautenhahn K. Potential applications of social robots in robot-assisted interventions for social anxiety. Int J Soc Robot 2022 Jan 25;14(5):1-32 [FREE Full text] [doi: 10.1007/s12369-021-00851-0] [Medline: 35096198]
6. Scoglio AA, Reilly ED, Gorman JA, Drebing CE. Use of social robots in mental health and well-being research: systematic review. J Med Internet Res 2019 Jul 24;21(7):e13322 [FREE Full text] [doi: 10.2196/13322] [Medline: 31342908]
7. Hung L, Liu C, Woldum E, Au-Yeung A, Berndt A, Wallsworth C, et al. The benefits of and barriers to using a social robot PARO in care settings: a scoping review. BMC Geriatr 2019 Aug 23;19(1):232 [FREE Full text] [doi: 10.1186/s12877-019-1244-6] [Medline: 31443636]
8. Rossi S, Santini SJ, Di Genova D, Maggi G, Verrotti A, Farello G, et al. Using the social robot NAO for emotional support to children at a pediatric emergency department: randomized clinical trial. J Med Internet Res 2022 Jan 13;24(1):e29656 [FREE Full text] [doi: 10.2196/29656] [Medline: 34854814]
9. Henschel A, Laban G, Cross ES. What makes a robot social? A review of social robots from science fiction to a home or hospital near you. Curr Robot Rep 2021 Feb 04;2(1):9-19 [FREE Full text] [doi: 10.1007/s43154-020-00035-0] [Medline: 34977592]
10. Onyeulo EB, Gandhi V. What makes a social robot good at interacting with humans? Information 2020 Jan 13;11(1):43. [doi: 10.3390/info11010043]
11. de Graaf MM, Ben Allouch S, van Dijk JA. What makes robots social?: a user's perspective on characteristics for social human-robot interaction. In: Proceedings of the Social Robotics. 2015 Presented at: ICSR 2015; October 26-30, 2015; Paris, France. [doi: 10.1007/978-3-319-25554-5_19]
12. Horstmann AC, Krämer NC. Great expectations? Relation of previous experiences with social robots in real life or in the media and expectancies based on qualitative and quantitative assessment. Front Psychol 2019 Apr 30;10:939 [FREE Full text] [doi: 10.3389/fpsyg.2019.00939] [Medline: 31114525]
13. Rosenthal-von der Pütten AM, Krämer NC, Herrmann J. The effects of humanlike and robot-specific affective nonverbal behavior on perception, emotion, and behavior. Int J Soc Robot 2018 Jan 23;10(5):569-582. [doi: 10.1007/s12369-018-0466-7]
14. Mavridis N. A review of verbal and non-verbal human–robot interactive communication. Robot Auton Syst 2015 Jan;63:22-35. [doi: 10.1016/j.robot.2014.09.031]
15. Rossi S, Ruocco M. Better alone than in bad company. Interact Stud 2019 Nov 18;20(3):487-508. [doi: 10.1075/is.18066.ros]
16. Rossi S, Cimmino T, Matarese M, Raiano M. Coherent and incoherent robot emotional behavior for humorous and engaging recommendations. In: Proceedings of the 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). 2019 Presented at: 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN); October 14-18, 2019; New Delhi, India. [doi: 10.1109/ro-man46459.2019.8956260]
17. Canamero L, Fredslund J. I show you how I like you - can you read it in my face? [robotics]. IEEE Trans Syst Man Cybern Part A Syst Humans 2001 Sep;31(5):454-459. [doi: 10.1109/3468.952719]
18. Tsiourti C, Weiss A, Wac K, Vincze M. Multimodal integration of emotional signals from voice, body, and context: effects of (in)congruence on emotion recognition and attitudes towards robots. Int J Soc Robot 2019 Feb 4;11(4):555-573. [doi: 10.1007/s12369-019-00524-z]
19. Sorrentino A, Khalid O, Coviello L, Cavallo F, Fiorini L. Modeling human-like robot personalities as a key to foster socially aware navigation. In: Proceedings of the 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN). 2021 Presented at: 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN); August 08-12, 2021; Vancouver, BC. [doi: 10.1109/ro-man50785.2021.9515556]
20. Guo F, Li M, Qu Q, Duffy VG. The effect of a humanoid robot's emotional behaviors on users' emotional responses: evidence from pupillometry and electroencephalography measures. Int J Hum Comput Interact 2019 Mar 29;35(20):1947-1959. [doi: 10.1080/10447318.2019.1587938]
21. Umbrico A, Cesta A, Cortellessa G, Orlandini A. A holistic approach to behavior adaptation for socially assistive robots. Int J Soc Robot 2020 Jan 28;12(3):617-637. [doi: 10.1007/s12369-019-00617-9]
22. Spezialetti M, Placidi G, Rossi S. Emotion recognition for human-robot interaction: recent advances and future perspectives. Front Robot AI 2020 Dec 21;7:532279 [FREE Full text] [doi: 10.3389/frobt.2020.532279] [Medline: 33501307]
23. Yan F, Iliyasu AM, Hirota K. Emotion space modelling for social robots. Eng Appl Artif Intell 2021 Apr;100:104178. [doi: 10.1016/j.engappai.2021.104178]
24. Cavallo F, Semeraro F, Fiorini L, Magyar G, Sinčák P, Dario P. Emotion modelling for social robotics applications: a review. J Bionic Eng 2018 Mar 23;15(2):185-203. [doi: 10.1007/s42235-018-0015-y]
25. Fiorini L, Loizzo FG, Sorrentino A, Kim J, Rovini E, Di Nuovo A, et al. Daily gesture recognition during human-robot interaction combining vision and wearable systems. IEEE Sensors J 2021 Oct 15;21(20):23568-23577. [doi: 10.1109/jsen.2021.3108011]
26. Manzi A, Fiorini L, Limosani R, Dario P, Cavallo F. Two‐person activity recognition using skeleton data. IET Comput Vis 2017 Oct 20;12(1):27-35. [doi: 10.1049/iet-cvi.2017.0118]
27. Albanie S, Nagrani A, Vedaldi A, Zisserman A. Emotion recognition in speech using cross-modal transfer in the wild. In: Proceedings of the 26th ACM international conference on Multimedia. 2018 Presented at: MM '18; October 22-26, 2018; Seoul, Republic of Korea. [doi: 10.1145/3240508.3240578]

28. Juszkiewicz L. Improving speech emotion recognition system for a social robot with speaker recognition. In: Proceedings of the 19th International Conference on Methods and Models in Automation and Robotics (MMAR). 2014 Presented at: 19th International Conference on Methods and Models in Automation and Robotics (MMAR); September 02-05, 2014; Miedzyzdroje, Poland. [doi: 10.1109/mmar.2014.6957480]
29. Anzalone SM, Boucenna S, Ivaldi S, Chetouani M. Evaluating the engagement with social robots. Int J Soc Robot 2015 Apr 17;7(4):465-478. [doi: 10.1007/s12369-015-0298-7]
30. Nocentini O, Fiorini L, Acerbi G, Sorrentino A, Mancioppi G, Cavallo F. A survey of behavioral models for social robots. Robotics 2019 Jul 09;8(3):54. [doi: 10.3390/robotics8030054]
31. Rossi S, Ferland F, Tapus A. User profiling and behavioral adaptation for HRI: a survey. Pattern Recognit Lett 2017 Nov;99:3-12. [doi: 10.1016/j.patrec.2017.06.002]
32. Bradley MM, Lang PJ. Measuring emotion: the self-assessment manikin and the semantic differential. J Behav Ther Exp Psychiatry 1994 Mar;25(1):49-59. [doi: 10.1016/0005-7916(94)90063-9] [Medline: 7962581]
33. Barchard KA, Lapping-Carr L, Westfall RS, Fink-Armold A, Banisetty SB, Feil-Seifer D. Measuring the perceived social intelligence of robots. ACM Trans Hum Robot Interact 2020 Sep 11;9(4):1-29. [doi: 10.1145/3415139]
34. Bagheri E, Esteban PG, Cao HL, De Beir A, Lefeber D, Vanderborght B. An autonomous cognitive empathy model responsive to users' facial emotion expressions. ACM Trans Interact Intell Syst 2020 Nov 09;10(3):1-23. [doi: 10.1145/3341198]
35. Rossi S, Staffa M, Tamburro A. Socially assistive robot for providing recommendations: comparing a humanoid robot with a mobile application. Int J Soc Robot 2018 Jan 31;10(2):265-278. [doi: 10.1007/s12369-018-0469-4]
36. Staffa M, Rossi S. Recommender interfaces: the more human-like, the more humans like. In: Proceedings of the Social Robotics. 2016 Presented at: ICSR 2016; November 1-3, 2016; Kansas City, MO. [doi: 10.1007/978-3-319-47437-3_20]
37. Becker C, Prendinger H, Ishizuka M, Wachsmuth I. Evaluating affective feedback of the 3D agent max in a competitive cards game. In: Proceedings of the Affective Computing and Intelligent Interaction. 2005 Presented at: ACII 2005; October 22-24, 2005; Beijing, China URL: https://link.springer.com/chapter/10.1007/11573548_60 [doi: 10.1007/11573548_60]
38. Lang PJ, Bradley MM, Cuthbert BN. International affective picture system (IAPS): affective ratings of pictures and instruction manual. Technical report A-8. University of Florida. 2008. URL: https://gitlab.pavlovia.org/rsaitov/experimental-psycholoy-ltu-final/raw/563e2b213be7eb54829116f05b83f0124362d71b/IAPS.TechManual.1-20.2008(1).pdf [accessed 2020-02-01]
39. Russell JA, Mehrabian A. Evidence for a three-factor theory of emotions. J Res Pers 1977 Sep;11(3):273-279 [FREE Full text] [doi: 10.1016/0092-6566(77)90037-X]
40. Salem M, Rohlfing K, Kopp S, Joublin F. A friendly gesture: investigating the effect of multimodal robot behavior in human-robot interaction. In: Proceedings of the 2011 RO-MAN. 2011 Presented at: 2011 RO-MAN; July 31-August 03, 2011; Atlanta, GA. [doi: 10.1109/roman.2011.6005285]
41. Fiorini LF, Loizzo FG, D'Onofrio G, Sorrentino A, Ciccone F, Russo S, et al. Can I feel you? Recognizing human's emotions during human-robot interaction. In: Proceedings of the Social Robotics. 2022 Presented at: ICSR 2022; December 13-16, 2022; Florence, Italy. [doi: 10.1007/978-3-031-24667-8_45]
42. D'Onofrio G, Fiorini L, Sorrentino A, Russo S, Ciccone F, Giuliani F, et al. Emotion recognizing by a robotic solution initiative (EMOTIVE project). Sensors (Basel) 2022 May 08;22(8):2861 [FREE Full text] [doi: 10.3390/s22082861] [Medline: 35458845]
43. Social robot. Behaviour Labs. URL: https://www.blabs.eu/robomate [accessed 2023-03-30]
44. Lang PJ, Bradley MM, Cuthbert BN. International affective picture system (IAPS): technical manual and affective ratings. Federal University of São Paulo. 1997. URL: https://www2.unifesp.br/dpsicobio/adap/instructions.pdf [accessed 2023-12-31]
45. Hall ET. The Hidden Dimension. New York, NY: Knopf Doubleday Publishing Group; 1966.
46. Gatti E, Calzolari E, Maggioni E, Obrist M. Emotional ratings and skin conductance response to visual, auditory and haptic stimuli. Sci Data 2018 Jul 26;5(1):180120 [FREE Full text] [doi: 10.1038/sdata.2018.120] [Medline: 29944144]
47. Avidemux. Uptodown. URL: https://avidemux.it.uptodown.com/windows [accessed 2023-03-30]
48. Baltrusaitis T, Zadeh A, Lim YC, Morency LP. OpenFace 2.0: facial behavior analysis toolkit. In: Proceedings of the 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). 2018 Presented at: 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018); May 15-19, 2018; Xi'an, China. [doi: 10.1109/fg.2018.00019]
49. Lara OD, Labrador MA. A survey on human activity recognition using wearable sensors. IEEE Commun Surv Tutor 2012 Nov 29;15(3):1192-1209. [doi: 10.1109/SURV.2012.110112.00192]
50. Admoni H, Scassellati B. Social eye gaze in human-robot interaction: a review. J Hum Robot Interact 2017 May;6(1):25-63. [doi: 10.5898/JHRI.6.1.Admoni]
51. Kompatsiari K, Bossi F, Wykowska A. Eye contact during joint attention with a humanoid robot modulates oscillatory brain activity. Soc Cognit Affect Neurosci 2021;16(4):383-392. [doi: 10.31234/osf.io/eam3d]
52. Fiorini L, Coviello L, Sorrentino A, Sancarlo D, Ciccone F, D'Onofrio G, et al. User profiling to enhance clinical assessment and human-robot interaction: a feasibility study. Int J Soc Robot 2023;15(3):501-516 [FREE Full text] [doi: 10.1007/s12369-022-00901-1] [Medline: 35846164]

53. Fiorini L, Mancioppi G, Semeraro F, Fujita H, Cavallo F. Unsupervised emotional state classification through physiological parameters for social robotics applications. Knowl Based Syst 2020 Feb;190:105217. [doi: 10.1016/j.knosys.2019.105217]

Abbreviations

COH: coherent
HRI: human-robot interaction
IAPS: International Affective Picture System
KNN: K-nearest neighbor
RF: random forest
RQ: research question
SAM: Self-Assessment Manikin questionnaire
STA: static
SVM: support vector machine

Edited by A Choudhury; submitted 04.01.23; peer-reviewed by O Bendel, M Amith; comments to author 21.03.23; revised version received 24.04.23; accepted 29.11.23; published 26.01.24.

Please cite as:
Fiorini L, D'Onofrio G, Sorrentino A, Cornacchia Loizzo FG, Russo S, Ciccone F, Giuliani F, Sancarlo D, Cavallo F
The Role of Coherent Robot Behavior and Embodiment in Emotion Perception and Recognition During Human-Robot Interaction: Experimental Study
JMIR Hum Factors 2024;11:e45494
URL: https://humanfactors.jmir.org/2024/1/e45494
doi:10.2196/45494
PMID:38277201

Original Paper

Reducing the Number of Intrusive Memories of Work-Related Traumatic Events in Frontline Health Care Staff During the COVID-19 Pandemic: Case Series

Veronika Kubickova[1], DClinPsy; Craig Steel[1,2], PhD; Michelle L Moulds[3], PhD; Marie Kanstrup[4,5], PhD; Sally Beer[6], PG Cert; Melanie Darwent[6], BM BCh; Liza Keating[7], MB ChB, MSc; Emily A Holmes[8], PhD; Lalitha Iyadurai[9], PhD

[1]Oxford Institute of Clinical Psychology Training and Research, Oxford Health NHS Foundation Trust, Oxford, United Kingdom
[2]Department of Experimental Psychology, University of Oxford, Oxford, United Kingdom
[3]School of Psychology, UNSW Sydney, Sydney, Australia
[4]Department of Psychology, Uppsala University, Uppsala, Sweden
[5]Behavioral Medicine, Theme Women's Health and Allied Health Professionals, Karolinska University Hospital, Stockholm, Sweden
[6]Emergency Medicine Research Oxford, Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom
[7]University Department of Emergency Medicine, Royal Berkshire NHS Foundation Trust, Reading, United Kingdom
[8]Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden
[9]Department of Psychiatry, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Craig Steel, PhD
Oxford Institute of Clinical Psychology Training and Research
Oxford Health NHS Foundation Trust
Isis Education Centre
Warneford Hospital
Oxford, OX3 7JX
United Kingdom
Phone: 44 1865226366
Email: craig.steel@hmc.ox.ac.uk

Abstract

Background: Frontline health care staff are frequently exposed to traumatic events as part of their work. Although this study commenced before the emergence of COVID-19, levels of exposure were heightened by the pandemic. Many health care staff members report intrusive memories of such events, which can elicit distress, affect functioning, and be associated with posttraumatic stress disorder symptoms in the long term. We need evidence-based interventions that are brief, preventative, nonstigmatizing, suitable for the working lives of frontline health care staff, and effective for repeated trauma exposure. A brief, guided imagery-competing task intervention involving a trauma reminder cue and Tetris gameplay may hold promise in this regard, given evidence that it can prevent and reduce the number of intrusive memories following trauma across various settings.

Objective: This case series aims to investigate the impact of a brief imagery-competing task intervention on the number of intrusive memories, general functioning, and symptoms of posttraumatic stress, anxiety, and depression, and examine the feasibility and acceptability of the intervention for UK National Health Service frontline health care staff. The intervention was delivered with guidance from a clinical psychologist.

Methods: We recruited 12 clinical staff from the UK National Health Service, specifically from emergency departments, the intensive care unit, and the ambulance service. We evaluated the intervention using an AB single-case experimental design, where the baseline (A) was the monitoring-only phase and the postintervention (B) period was the time after the intervention was first administered. Methods were adapted once the COVID-19 pandemic began.

Results: There was a decrease (59%) in the mean number of intrusive memories per day from baseline (mean 1.29, SD 0.94) to postintervention (mean 0.54, SD 0.51). There was a statistically significant reduction in the number of intrusive memories from baseline to postintervention, as shown by an aggregated omnibus analysis with a small effect size (τ-U=–0.38; P<.001). Depression, anxiety, and posttraumatic stress symptoms all significantly reduced from preintervention to postintervention.

Participants also reported improvements in functioning based on both quantitative and qualitative measures. The intervention was feasible to deliver and rated as acceptable by participants.

Conclusions: These preliminary findings suggest that this brief therapist-guided imagery-competing task intervention offers a potential approach to mitigating the impact of work-related traumatic events in frontline health care staff, both during a pandemic and beyond. Randomized controlled trials will be an important next step.

(JMIR Hum Factors 2024;11:e55562) doi:10.2196/55562

KEYWORDS

intrusive memories; digital intervention; psychological trauma; remote delivery; health care staff; COVID-19; case series

Introduction

Background

Frontline health care staff, including emergency department (ED) staff, intensive care unit (ICU) staff, and paramedics, frequently encounter highly stressful and traumatic clinical events, such as patient death, resuscitations, and treating patients with severe injuries [1]. These experiences often lead to intrusive memories, which are involuntary sensory recollections, primarily visual but sometimes involving other senses, such as smell and sound [2-4]. These memories can evoke strong negative emotions [5], disrupt concentration [6], and impair daily functioning [7], making them a critical treatment target.

Intrusive memories are a core symptom of posttraumatic stress disorder (PTSD) and acute stress disorder [8], and they may drive posttraumatic stress symptomatology [9]. Reducing intrusive memories could lead to a broader reduction in PTSD symptoms [10], suggesting a potential "therapeutic cascade" [11].

PTSD rates were already high among National Health Service (NHS) health care staff before the COVID-19 pandemic and increased substantially during it, with prevalence ranging from 21% among health care workers across 21 countries to 40% among ICU staff in the United Kingdom [12-17]. Unrelenting workloads, burnout, and stress are major drivers of nurses leaving the profession [18]. Furthermore, investigating new initiatives to improve ED staff engagement, resilience, and retention has been highlighted as a research priority for emergency medicine in the United Kingdom [19]. Therefore, developing effective treatments to protect the mental well-being of frontline health care staff is essential [20].

Frontline health care staff may experience intrusive memories without meeting the full diagnostic criteria for PTSD. Many may prefer shorter, more targeted interventions due to the stigma surrounding mental health issues among health care professionals and the substantial workload imposed by the pandemic [21,22]. It is crucial to develop novel interventions that are effective, brief, accessible, scalable, and nonstigmatizing, and that can be used repeatedly to manage ongoing trauma exposure.

A novel, brief, low-intensity, imagery-competing task intervention has been developed to target intrusive memories of trauma [23-25]. The intervention, which lasts approximately 25 minutes, can be administered in a single researcher-assisted session and self-administered thereafter. It aims to reduce the number of intrusive memories by using principles from cognitive science, specifically dual tasking, the properties of mental imagery [26], and the theories of memory consolidation and reconsolidation [27].

The intervention's mechanism involves performing a visuospatial task while the trauma memory is still labile. This is theorized to interfere with the memory consolidation process and prevent the emergence of intrusive memories [28]. This approach for recent trauma has been supported by 3 randomized controlled trials involving women who experienced traumatic childbirth [29], motor vehicle accident survivors in the ED [30], and a mixed trauma sample of individuals presenting to the ED [25]. In these trials, participants who received the intervention reported significantly fewer intrusive memories in the first week after exposure to trauma compared with control groups, with effect sizes ranging from small to medium (Cohen d=0.43 to Cohen d=0.67) [25,29,30]. In addition, the intervention significantly reduced intrusion-related distress and vividness [25].

Evidence also supports that the intervention can reduce older, established trauma memories, with the effect informed by theories of memory reconsolidation interference. Promising results, including a decreased frequency of intrusive memories and improvements in functioning, have been shown in case series involving refugees [31], inpatients with complex PTSD [32], and women with childhood trauma [33,34].

A digital version of the intervention, delivered remotely with researcher support, was tested among health care staff members during the COVID-19 pandemic in a pilot study and was found to be feasible and acceptable [35]. This brief remotely delivered digital intervention was tested in a randomized controlled trial with the UK NHS ICU staff during the COVID-19 pandemic [36,37]. Participants who received immediate access to the intervention (with an initial guided session) reported significantly fewer intrusive memories at week 4 compared with those with delayed access. Intrusive memories also significantly decreased at week 8 (postintervention) compared with week 4 (preintervention) in the delayed access arm [37]. Furthermore, the immediate (vs delayed) intervention group reported reduced PTSD symptoms, insomnia, anxiety, posttrauma distress, and burnout, and improved work functioning and general well-being. Participants found the intervention to be acceptable and safe. The study provided strong evidence of the intervention's efficacy for reducing intrusive memories among this group of health care staff [36]. While the intervention has been tested among hospital staff, its effectiveness among prehospital staff, such as paramedics, who often face unpredictable, high-intensity trauma in chaotic and risk-prone settings, remains unknown [38-40].

Objectives

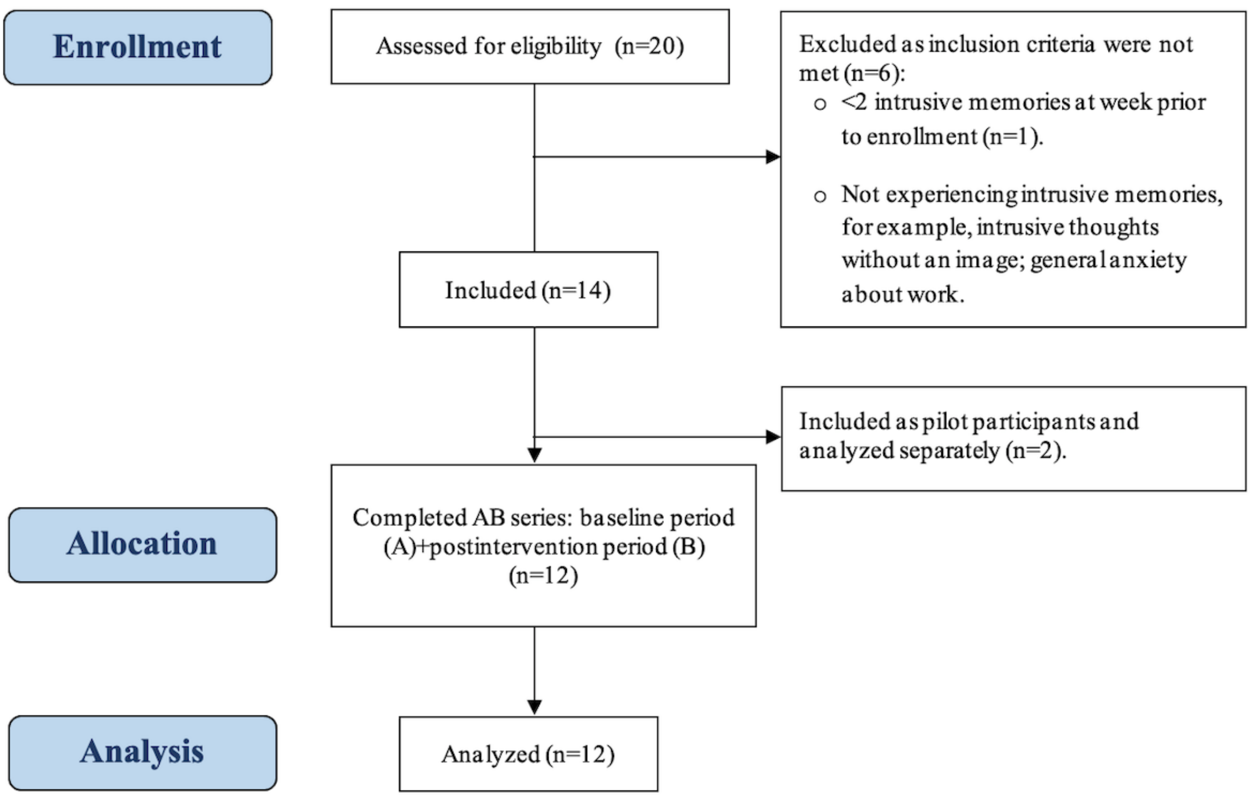
Given these promising findings and the need for novel approaches to support health care personnel beyond ICU staff, our study aimed to test the efficacy, feasibility, and acceptability of the brief guided imagery-competing task intervention in a sample of frontline prehospital and hospital NHS health care staff exposed to traumatic events in the course of their work. The primary aim was to determine whether the intervention reduced the number of intrusive memories of work-related traumatic events. We predicted a significant decrease in the number of intrusive memories from baseline to postintervention. Secondary aims included assessing changes in intrusive memory characteristics (eg, vividness and distress), their impact on daily functioning (eg, sleep and concentration), and other mental health symptoms (eg, posttraumatic stress, depression, and anxiety). In addition, we evaluated the intervention's feasibility and acceptability among NHS health care staff.

Methods

Participants

In total, 12 frontline NHS health care staff members, including doctors, nurses, and paramedics, working in hospital (eg, ED and ICU) and prehospital departments (eg, ambulance services team) at Oxford University Hospitals, Royal Berkshire, and South Central Ambulance Service NHS Foundation Trusts participated in the study (recruited from February 2020 to September 2020). Inclusion criteria were aged ≥18 years; ability to read, write, and speak in English; being able and willing to provide informed consent and complete study procedures; employed as NHS hospital or prehospital clinical staff; having experienced at least 2 intrusive memories (self-rated as problematic) of a work-related traumatic incident in the previous 7 days; being able and willing to talk about the intrusive memories; ability to complete a web-based daily intrusive memory diary over a 2-to-3-week period; being capable of playing Tetris on a handheld device; and not currently undergoing treatment for PTSD. Participants were excluded if they had <2 intrusive memories per week during the baseline period or had started undergoing treatment for PTSD. The study was piloted on 2 participants (Multimedia Appendix 1). Figure 1 depicts an adapted CONSORT (Consolidated Standards of Reporting Trials) flow diagram for the study.

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



Design

This case series used an AB single-case experimental design [41]. The primary outcome was the change in the number of intrusive memories of work-related traumatic events from baseline (A) to postintervention (B). The baseline was the period before the first administration of the intervention, that is, monitoring-only, and the postintervention period was anytime thereafter. The study consisted of a 1-week baseline period, a 2-week postintervention period following a single researcher-assisted intervention session, and a follow-up 4 weeks after the intervention session.

Following the single researcher-assisted intervention session, participants received instructions for using optional intervention boosters as needed in their daily lives. The intervention boosters could be researcher-assisted (ie, by scheduling a guided intervention session with the researcher) or self-administered (ie, playing Tetris for 25 minutes *immediately* after experiencing an intrusive memory, excluding a brief trauma reminder cue procedure required in the researcher-assisted intervention sessions). Participants were invited to use the intervention boosters for any new intrusive memories of work-related traumatic events they experienced after enrolling in the study; however, these were not included in the analysis. Participants could use the intervention boosters as many times as needed until week 4 (postintervention).

Due to COVID-19-related national restrictions, all participant interactions and intervention sessions were conducted remotely (ie, via videocall).

Procedure

Training to Deliver the Intervention

To ensure proper intervention delivery and protocol adherence, the primary researcher, VK (a trainee clinical psychologist in the final year of clinical training), received comprehensive training, feedback, and monitoring from clinical psychologists experienced in delivering the intervention. This training included a 1-day web-based workshop covering the theoretical foundations, key components, and wider protocol aspects of the intervention, including identifying and recording intrusive memories, the trauma reminder cue procedure, delivering “mental rotation” (that is, visualizing in the mind’s eye how to rotate and move the Tetris blocks to make horizontal lines) instructions for Tetris gameplay, and collecting primary outcomes. VK participated in role-playing sessions during the workshop and subsequently with LI and MK, with filmed role-plays evaluated for protocol fidelity by an independent rater.

A written standard protocol was used by the researcher in all researcher-assisted intervention sessions to ensure intervention fidelity. Criteria for intervention fidelity were completion of an arousal level manipulation check before and after the trauma reminder cue and after Tetris gameplay; administration of the trauma reminder cue and checks to ensure the visual or perceptual details were in the participant’s mind; delivery of Tetris gameplay instructions with an emphasis on mental rotation; ensuring participants played Tetris for at least 25 minutes with at least 1 continuous period of 10 minutes. For the pilot and the first study participants, LI provided supervision and protocol fidelity checks via phone or video call after each session.

Regular supervision meetings addressed protocol fidelity and adaptations. During recruitment and data collection, VK participated in fortnightly video calls with other researchers to share experiences, best practices, and receive feedback.

Eligibility and Baseline Assessment Meeting: Week 1 (Preintervention)

Potential participants met with VK to assess eligibility based on the inclusion and exclusion criteria. During this meeting, the study’s purpose, procedures, risks, and benefits were explained, and verbal informed consent was obtained due to COVID-19 social distancing measures. All participants were provided with information about local occupational and mental health support services.

After giving their informed consent, participants completed the baseline assessment, including self-report measures administered via Qualtrics [42], a web-based survey platform. The “hotspots” form was then used to gather information about the number and timing of traumatic events, as well as a brief description of participants’ intrusive memories. Participants were verbally given the following definition of intrusive memories: “vivid, emotional memories of the incident that ‘pop’ into mind without warning, often taking the form of visual pictures in the mind’s eye, for example, a snapshot image or a film clip.” Participants were also informed that intrusive memories can include other senses (such as sounds and smells), that they may or may not be triggered by something the person is aware of (eg, telling someone about the incident or being back at the scene), and that intrusive memories can be “very short, fleeting, and broken up.” Participants were told that deliberately thinking about an incident, mulling it over, or having general thoughts about it without an image did not constitute an intrusive memory. Participants identified the most frequent and bothersome intrusive memory to be targeted during the intervention. Participants were instructed on how to complete a web-based daily intrusive memory diary using Qualtrics and were given clear guidelines on how to document each occurrence of an intrusive memory.

Intrusive Memory Monitoring and Weekly Questionnaires

Participants were asked to complete the intrusive memory diary at least once daily for 3 weeks. They received automated reminders via email and text message to ensure compliance with daily diary entries. In addition to the daily diary, participants completed weekly secondary outcome measures from week 1 (preintervention) through to week 4 (postintervention). These measures assessed various aspects of participants’ mental health and functioning (Measures and Materials section).

Intervention Session: Week 2 (Preintervention)

The intervention, with guided delivery by VK, targeted the most frequent and bothersome intrusive memory identified during the baseline assessment. The brief single-session imagery-competing task intervention consisted of 3 components: (1) a trauma reminder cue to bring the specific trauma memory to mind; (2) engaging in a visuospatial interference task, that is, playing the computer game Tetris for 25 minutes; and (3) using specific mental rotation instructions to play the game to optimize visuospatial demand along set timing parameters.

The trauma reminder cue consisted of participants briefly bringing the targeted intrusive memory to mind and writing down its contents (in the first person) on a blank piece of paper (similar to the method of Kessler et al [32], although with a briefer description). Participants were asked to write only as much detail as was necessary to briefly recall the memory to avoid it becoming too emotionally overwhelming. Of note, this brief trauma reminder cue is focused on specific sensory-perceptual aspects of the trauma memory, and contrary to reliving procedures included in exposure-based therapies for PTSD (eg, Shearing et al [43]), it does not include a focus on the details of the event or emotional and cognitive aspects. Furthermore, also in contrast to reliving procedures, it is brief, only just enough to bring the memory image to mind before engaging in the gameplay part of the intervention. To prevent unintended reminders, this written description was discarded by the participant immediately after the intervention session; its contents were not read by or discussed with VK, as the aim of the procedure was solely to bring the intrusive memory into working memory (ie, to activate it).

Participants then received instructions on how to play the computer game Tetris using mental rotation. To increase the visuospatial demands of the game, participants were asked to plan and work out where to place the next blocks coming up, as well as the block that was currently presented [7,24]. After receiving instructions and a demonstration from VK, participants were given the opportunity to practice the game by completing 1 or 2 lines of Tetris blocks using mental rotation before beginning the timed portion of the intervention. The intervention was delivered via the official Tetris website [44], accessed on participants' mobile phone, or tablet computer, and the game was set to "marathon mode" and "ghost piece" off.

Participants played Tetris for at least one uninterrupted period of 10 minutes and for approximately 25 minutes in total [25]. Participants were asked to restart the game if "game over" was reached [32]. VK was present for the duration of the 25-minute gameplay to encourage participants to maintain engagement with the game and occasionally remind them about the mental rotation instructions.

Follow-Up: Week 4 (Postintervention)

At week 4 (postintervention), participants completed the final set of secondary outcome measures. They also completed a feedback questionnaire assessing the acceptability and feasibility of the intervention and study procedures. Participants were provided with information about the study aims and reimbursed £50 (US $65.3) for their time and effort in participating in the study.

Measures and Materials

Baseline Information Regarding Traumatic Events and Intrusive Memories

Participants' descriptions of their intrusive memories were recorded in a web-based "hotspots" form (Multimedia Appendix 2) in Qualtrics [42] using the "share screen" option.

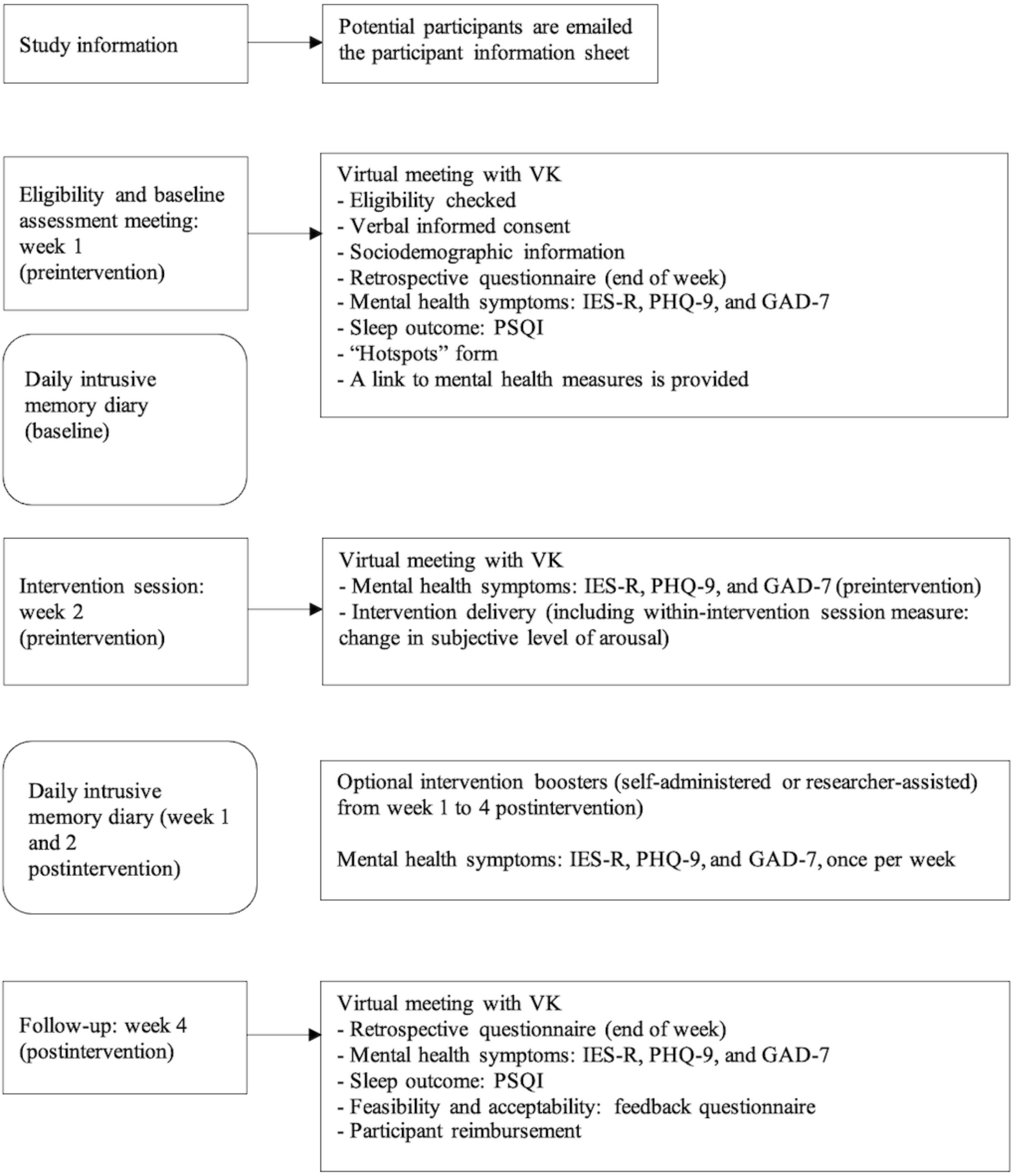
Participants indicated how many work-related traumatic events they were currently experiencing intrusive memories of and how long ago each event took place (ie, years and months). Participants were given the following verbal instruction: "Please briefly describe the worst moments of the traumatic event(s), e.g., an image or a sound. It is fine to summarize them in just a few words." These descriptions were then used to identify established intrusive memories, rather than only "hotspots" of trauma as in Kanstrup et al [25] and helped ensure that what participants noted were indeed intrusive memories (rather than ruminative thoughts, etc). We note that this process did not involve a detailed discussion of the traumatic event, or a focus on emotional aspects, but rather a brief description primarily focused on visuo-sensory content (ie, just a few words, such as "seeing the pattern on the cardiac monitoring machine," as per Hoppe et al [45]). Participants were asked to label each intrusive memory by selecting a few keywords (eg, "cardiac monitoring machine") to describe its content. These keywords were used to aid the intrusive memory diary completion as well as to identify which intrusive memory to target as part of the intervention procedure.

Primary Outcome Measure: Number of Intrusive Memories

Participants recorded the primary outcome, the number of intrusive memories experienced each day, in a diary, which was adapted from paper diaries used in previous studies [24,25,30,32] and here delivered using a web-based format (via Qualtrics [42]; Multimedia Appendix 3).

Participants received daily reminders through secure email links and automated text messages to ensure diary completion. In the diary, participants were instructed to select the intrusive memories they had experienced since their last entry from a drop-down menu populated with the intrusive memories they had previously identified on the "hotspots" form. They were also asked to indicate how many times they had experienced that intrusive memory by typing in the relevant number. If they experienced a new intrusive memory (ie, not one that was identified on the "hotspots" form), they had the option to select "other." Participants were asked to complete the diary as soon as possible after they experienced an intrusive memory, but at least once per day. If no intrusive memories were experienced that day, they were asked to indicate this by recording 0. Figure 2 depicts an overview of study meetings and when each measure was administered.

Figure 2. Overview of study meetings: the eligibility and baseline assessment occurred at week 1 (preintervention), followed by the intervention session at week 2 (preintervention). The final study meeting took place 4 weeks postintervention. Participants completed a daily diary for 3 weeks (1 week baseline and 2 weeks postintervention). GAD-7: Generalized Anxiety Disorder Scale; IES-R: Impact of Events Scale-Revised; PHQ-9: Patient Health Questionnaire Depression Scale; PSQI: Pittsburgh Sleep Quality Index.



Secondary Outcome Measures

Intrusive Memory Characteristics and Impact on Functioning: Daily Intrusive Memory Diary

In the intrusive memory diary, participants rated how vivid and distressing each intrusive memory was (0=not at all to 10=extremely), how much the intrusive memory disrupted their concentration and disrupted the task at hand (0=not at all to 10=a great deal), the length of time that the intrusive memory bothered them (<1 min, 1-5 min, 6-10 min, 11-30 min, 31-60 min, and >60 min), and sleep quality over the past 24 hours (very good, fairly good, fairly bad, and very bad). Participants indicated whether they were at work when they experienced the intrusive memory, the approximate time at which this intrusive memory occurred, and how many times they used the intervention since the last diary entry (ie, to record the number of intervention booster sessions).

Mental Health Symptoms: Impact of Events Scale-Revised

PTSD symptoms were assessed with the Impact of Event Scale-Revised (IES-R) [46], a 22-item measure containing 3 subscales: intrusion, avoidance, and hyperarousal. Each item ranges from 0=not at all to 4=extremely. The IES-R has high internal consistency (Cronbach α=0.96) [47] and good test-retest reliability (ranging from 0.89 to 0.94) [48] and is sensitive to a general construct of traumatic stress in populations with lower symptom levels [48].

Mental Health Symptoms: Patient Health Questionnaire Depression Scale

Depression symptoms were assessed with the Patient Health Questionnaire-9 (PHQ-9) [49], a 9-item self-report measure of depression symptoms severity. Each item ranges from 0=not at all to 3=nearly every day. A score of ≥10 has sensitivity of 88% and specificity of 88% for major depression. The PHQ-9 has excellent internal reliability (Cronbach α=0.89) and test-retest reliability (r=0.84) [49], and its validity has been demonstrated in a nonclinical population [50].

Mental Health Symptoms: Generalized Anxiety Disorder Scale

Anxiety symptoms were assessed with the Generalized Anxiety Disorder-7 (GAD-7) [51], a brief self-report measure of symptoms of general anxiety disorder and their severity. Each item ranges from 0=not at all to 3=nearly every day. The GAD-7 has excellent internal consistency (Cronbach α=0.92) and good test-retest reliability (r=0.83) [51].

Mental Health Symptoms: Pittsburgh Sleep Quality Index

Participants rated their sleep quality with the Pittsburgh Sleep Quality Index (PSQI) [52], a self-report measure of sleep quality over a 1-month time interval. The PSQI has good internal consistency (Cronbach α=0.83) and test-retest reliability (r=0.85) [53].

Intrusive Memory Characteristics and Impact on Functioning: Retrospective Questionnaire (End of Week)

Participants rated the characteristics of their intrusive memories and their impact on functioning, retrospectively (ie, "over the past week;" Multimedia Appendix 4), using a 10-item rating scale. Participants provided ratings of the number of intrusive memories (none (0), some (1-4), quite a few (5-10), lots (11-20), very many (21-30), a large amount (31-50), and more than 50), as well as their vividness and associated distress (0=not at all to 10=extremely), the extent to which they disrupted concentration, interfered with the task at hand, affected night's sleep, and impacted their ability to function in daily life (0=not at all to 10=a great deal). In addition, participants reported the duration of time the intrusive memories were bothersome (<1 min, 1-5 min, 6-10 min, 11-30 min, 31-60 min, and >60 min) and described how their ability to function in daily life was affected by intrusive memories.

Change in Subjective Level of Arousal: Within-Intervention Session Measure

Participants rated their subjective level of arousal before and after the trauma reminder cue and after playing Tetris during all researcher-assisted intervention sessions on an 11-point scale (0=calm to 10=maximum arousal). This manipulation check was done to assess changes in arousal in response to the trauma reminder cue and to check for potential immediate effects of the intervention.

Acceptability and Feasibility: Feedback Questionnaire

Participants completed an 11-item feedback questionnaire (adapted from Iyadurai et al [30]) to assess the acceptability of the intervention and feasibility of the study procedures (Multimedia Appendix 5). The questionnaire included items assessing participants' experience of the intervention (ie, how easy, helpful, and burdensome they found it), their willingness to use the intervention if it was offered to them in the future, and their confidence in recommending the intervention to a colleague who was experiencing intrusive memories (0=not at all to 10=extremely). Participants also described their experience of the intervention and taking part in the study (eg, suggestions for improvements; other comments about the intervention or the study).

Intrusive Memory Diary Adherence

Adherence to completing the daily intrusive memory diary was assessed with the rating, "How accurately do you think you completed the diary? (0=not at all to 10=extremely)."

Treatment Adherence

Adherence to treatment was assessed by recording whether participants completed the trauma reminder cue procedure, received mental rotation instructions, and the duration of Tetris gameplay during the researcher-assisted intervention sessions. Adherence to mental rotation and 25-minute gameplay in self-administered sessions was not assessed.

Data Analysis

Participant Characteristics

Sociodemographic data and baseline information regarding traumatic events and intrusive memories were summarized using descriptive statistics.

Primary Outcome Analyses

We calculated the mean number of intrusive memories per day for each participant to assess the change in the mean number of intrusive memories from baseline to postintervention. The per-day unit was chosen over per-week for greater measurement accuracy. Intrusive memories of new work-related traumatic events that occurred after participants enrolled in the study (marked as "other" in the daily intrusive memory diary) were excluded from the analyses and reported separately using descriptive statistics (Table S1 in Multimedia Appendix 1).

Baseline and postintervention daily intrusive memory means were calculated considering the exact timing of the intervention delivery within a 24-hour period. Baseline time was determined as the number of complete baseline days plus the number of hours before intervention delivery divided by 24. The baseline mean was the number of baseline intrusive memories divided by baseline time. The postintervention mean was the number of postintervention intrusive memories divided by postintervention time. The percentage change in intrusive memory frequency from baseline to postintervention was calculated as (1−[mean number per day postintervention/mean number per day at baseline]×100) [32]. A “global” percentage change in intrusive memories was calculated across all participants.

Visual inspection of individual time-series graphs is fundamental to case series methodology [54]. The time-series graphs were created using a website for single-case data analysis [55]. Visual inspection was conducted to identify patterns of change in intrusive memory frequency from baseline to postintervention [54,56].

In addition, the τ-*U* statistic was used to analyze the intervention’s impact on between-phase differences (baseline vs postintervention) [57]. Multimedia Appendix 1 gives further details regarding τ-*U* analyses.

Secondary Outcome Analyses

Changes in intrusive memory characteristics (ie, vividness and distress) and their impact on functioning (ie, concentration and task disruption) from baseline to postintervention were analyzed using means, tests of difference, and effect sizes. Categorical data (ie, sleep quality and length of time intrusive memories were bothersome) were presented using descriptive statistics.

We calculated means, tests of difference, and effect sizes from week 1 (preintervention) to week 4 (postintervention) to assess changes in depression, anxiety, and posttraumatic stress symptoms, global sleep, and overall sleep quality.

For retrospective ratings of intrusive memory characteristics and impact on functioning, we calculated means, SDs, and effect sizes for week 1 (preintervention) and week 4 (postintervention) ratings. Categorical data (number of intrusive memories in the past week and length of time intrusive memories were bothersome) were summarized using descriptive statistics. Qualitative data on the impact of intrusive memories on everyday functioning were presented as anonymized quotes.

Changes in subjective arousal levels (pre- to posttrauma reminder cue and pre- to post-Tetris gameplay) were reported as means and SDs, with tests of difference calculated. These findings are detailed in Multimedia Appendix 1.

Feedback questionnaire ratings were summarized using descriptive statistics to assess the intervention’s acceptability. Open-ended responses regarding the acceptability and feasibility of the intervention and study procedures were analyzed for themes, with anonymized quotes presented as examples.

Treatment adherence, including the duration of Tetris gameplay, subjective accuracy ratings for intrusive memory diary completion, and outcome measure completion rates, was summarized using descriptive statistics (Multimedia Appendix 1 provides further details).

Missing Data

Participants with missing data were excluded from the analysis for the specific outcome measure in which data were missing.

Data Accuracy Checks

Outcome data accuracy was verified by an independent rater.

Ethical Considerations

The study required ethics approval due to the involvement of NHS staff members as participants. Ethics approval was granted by the Health Research Authority and the University of Oxford Medical Sciences Inter-Divisional Research Ethics Committee (approval number R64738/RE001). An amendment to obtain verbal consent due to COVID-19 social distancing measures was approved on April 23, 2020 (approval number R64738/RE004). Trust management approval was provided by the Oxford University Hospitals NHS Foundation Trust, the Royal Berkshire NHS Foundation Trust, and the South Central Ambulance Service NHS Foundation Trust. For public record, the study was retrospectively registered on ClinicalTrials.gov (NCT04769999) following data collection but preanalysis (Multimedia Appendix 1 provides further details).

Results

Intrusive Memory Diary Adherence

The mean subjective accuracy rating for daily intrusive memory diary completion across all entries and participants was 8.50 (SD 0.90; range 0-10).

Treatment Adherence

Treatment adherence was 100% for all researcher-assisted sessions: every participant completed all 3 components of the intervention protocol (trauma reminder cue, receiving mental rotation instructions, and playing Tetris for at least 25 minutes). In these sessions, all participants played Tetris for a minimum of 25 minutes (mean 25.17, SD 0.49 min; range 25-27 min).

Rates of Outcome Measure Completion

The completion rate for the primary outcome measure was 100%. For secondary outcome measures, the IES-R, PHQ-9, and GAD-7 completion rate was 99% (87/88; week 5 measures were missing for 1 participant). The PSQI completion rate was 96% (23/24; missing data for 1 participant). All participants completed the study feedback questionnaire.

Attrition

All participants remained in the study for its full duration.

Sociodemographic Information

Sociodemographic information is presented in Table 1.

Table 1. Participant sociodemographic characteristics (N=12).

Characteristic	Value
Sex, n (%)	
Female	8 (67)
Male	4 (33)
Intersex	0 (0)
Prefer not to say	0 (0)
Age (y), mean (SD; range)	32.92 (7.39; 22-49)
Education (years from first grade), mean (SD; range)	18.04 (2.75; 14-22)
Ethnicity, n (%)	
White British	8 (67)
Any other White background	3 (25)
Pakistani	1 (8)
Relationship status, n (%)	
Single	7 (58)
Married or cohabitating	4 (33)
Divorced or separated	1 (8)
Job role, n (%)	
Nursing	4 (33)
Medical doctor	1 (8)
Student (medical and nursing)	2 (17)
Ambulance service team (eg, paramedic and emergency care assistant)	5 (42)
Department of employment, n (%)	
Emergency department	3 (25)
Adult intensive care unit	1 (8)
Neonatal intensive care unit	2 (17)
Other inpatient hospital ward	1 (8)
Ambulance service	5 (42)
Length of time working in current job role (mo), mean (SD; range)	60.67 (71.39; 3-240)
NHS[a] banding level, n (%)	
Band 3	2 (17)
Band 5	4 (33)
Band 6	4 (33)
Band 7	1 (8)
Not applicable (medical student)	1 (8)

[a]NHS: National Health Service.

Baseline Information Regarding Traumatic Events and Intrusive Memories

The mean number of work-related traumatic events reported at the week 1 preintervention meeting was 2.58. Of all reported work-related traumatic events, 39% (12/31) occurred between January and September 2020, during the COVID-19 pandemic. On average, participants reported 5.33 work-related traumatic event “hotspots.” Participants reported a range of intrusive memory content; examples include patients with fatal injuries, patients dying, extremely distressed family members, and medical procedures (Table S2 in Multimedia Appendix 1 gives further details regarding baseline information).

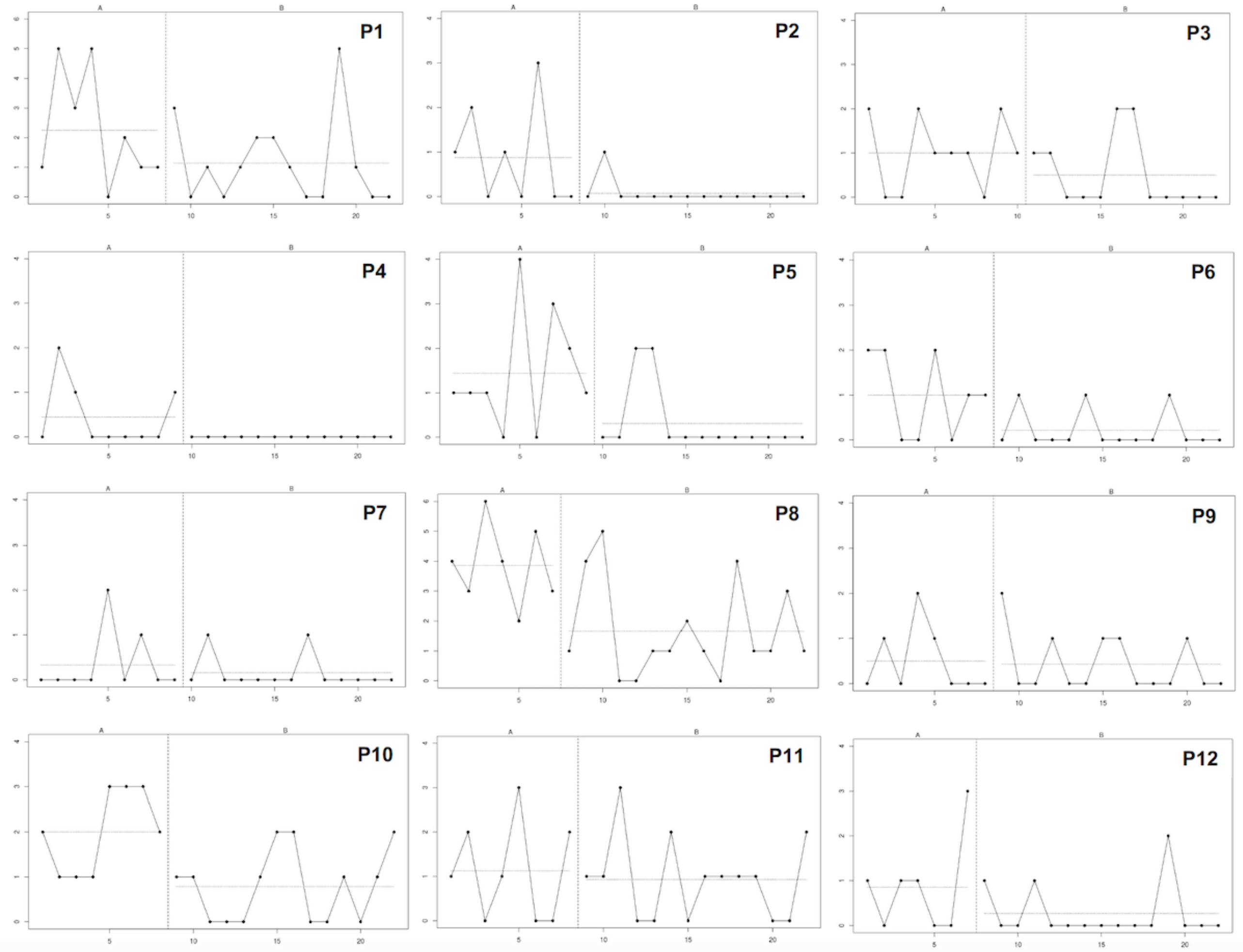
Following the baseline period, 85 intervention sessions were delivered across all participants. Of these, 86% (73/85) were intervention boosters, with 14% (10/73) being researcher-assisted and 86% (63/73) self-administered. Table S3 in Multimedia Appendix 1 gives details of the researcher-assisted and self-administered booster sessions (ie, following the first researcher-assisted session).

Primary Outcome Measure: Number of Intrusive Memories

Across all participants, the mean number of intrusive memories per day decreased by 59% from baseline (mean 1.29, SD 0.94) to postintervention (mean 0.54, SD 0.51). Individual-level data showed that 83% (10/12) of participants experienced reductions between 51% and 100% (Table S4 in Multimedia Appendix 1 gives individual percentage changes).

Visual inspection of individual time-series graphs [54,56] indicated a reduction in the number of intrusive memories following the intervention for all participants, as evidenced by overall lower measures of central tendency from baseline to postintervention (Figure 3). Generally, the number of intrusive memories decreased either immediately or soon after the initial intervention session.

Figure 3. Graphs showing the primary outcome data (number of intrusive memories) for all participants (N=12). The y-axis represents the number of intrusive memories per day, and the x-axis represents each day of the study period. The horizontal dashed lines represent a measure of central tendency for the baseline (A) and postintervention (B) periods.



Participants P2 to P5 maintained 0 intrusive memories for the final 5 consecutive days of the study. However, participants P1, P9, and P11 showed a reemergence of intrusive memories postintervention, similar to their baseline levels. Participant P1 reported several intrusive memories during a particularly stressful period at work, coinciding with the postintervention period. This period included a taxing day during the COVID-19 pandemic, including redeployment to a COVID-19 ward, associated with a sharp increase in intrusive memories on day 19. Participants P1 and P9 reported functional improvements at work and in their social lives, despite modest reductions in the number of intrusive memories.

τ-*U* analysis yielded significant (*P*<.05) medium effect sizes for 3 participants (P5, P8, and P10). The aggregated omnibus analysis showed a significant (*P*<.001) small effect size (τ-*U*=–0.38; Table S4 in Multimedia Appendix 1 gives individual participant-level data).

Secondary Outcome Measures

Intrusive Memory Characteristics and Impact on Functioning: Daily Intrusive Memory Diary

A total of 216 intrusive memories were recorded by participants in the daily intrusive memory diary during the study period, with 31.9% (69/216) occurring while they were at work.

There were no statistically significant reductions in any ratings of intrusive memory characteristics (ie, vividness and distress), functioning (ie, concentration and task disruption), length of time intrusive memories were bothersome, or sleep quality, as recorded in the daily intrusive memory diary (Table S5 in Multimedia Appendix 1).

Mental Health Symptoms: IES-R, PHQ-9, GAD-7, and PSQI

Mental health outcomes, including depression, anxiety, and posttraumatic stress symptoms, showed significant reductions from week 1 (preintervention) to week 4 (postintervention). However, global sleep and overall sleep quality scores did not change significantly over this period (Table 2).

Table 2. Mental health symptoms and continuous outcomes of the retrospective ratings of intrusive memory characteristics (end of week) at week 1 (preintervention) and week 4 (postintervention), with analyses and effect sizes (N=12).

	Week 1 (preintervention), mean (SD)	Week 4 (postintervention), mean (SD)	*z* score	*t* test (*df*)	*P* value	Effect size, Cohen *d* (95% CI)
Mental health outcome measures						
IES-R[a]	41.08 (15.30)	12.58 (12.72)	–3.06	—[b]	.002	—
PHQ-9[c]	8.83 (7.54)	3.83 (4.11)	—	3.46 (11)	.005	1.00 (0.28-1.68)
GAD-7[d]	7.67 (6.11)	3.58 (4.50)	–2.76	—	.006	—
PSQI global[e,f]	8.27 (5.10)	7.36 (5.63)	—	1.01 (10)	.34	0.31 (–0.31-0.90)
PSQI overall sleep quality[f,g]	1.64 (0.92)	1.18 (0.98)	—	2.19 (10)	.05	0.66 (–0.01-1.30)
Continuous outcomes[h]						
Vividness[i]	7.31 (1.75)	4.50 (1.51)	—	3.72 (7)	.007	1.32 (0.33-2.26)
Distress[j]	5 (1.69)	2.75 (1.58)	—	3 (7)	.02	1.06 (0.16-1.92)
Concentration disruption[k]	5.75 (2.38)	2.38 (1.77)	—	4.34 (7)	.003	1.53 (0.46-2.56)
Task disruption[l]	4 (2.39)	1.63 (1.60)	—	3.49 (7)	.01	1.24 (0.27-2.15)
Night's sleep interference[m]	3.86 (3.48)	2 (3.83)	–1.84		.07	—
Daily functioning interference[n]	4 (2.65)	1 (1.53)	–2.23		.03	—

[a]IES-R: Impact of Events Scale-Revised (scores ranging from 0-88).

[b]Not available.

[c]PHQ-9: Patient Health Questionnaire (scores ranging from 0-27).

[d]GAD-7: Generalized Anxiety Disorder Scale (scores ranging from 0-21).

[e]PSQI global: Pittsburgh Sleep Quality Index (global PSQI scores ranging from 0-21).

[f]n=11 (excluding P3 who had missing PSQI data for week 4 postintervention).

[g]During the past month, how would you rate your sleep quality? very good=0, fairly good=1, fairly bad=2, and very bad=3.

[h]For continuous outcomes, N=8. Data at week 1 (preintervention) have been excluded for those participants who reported no intrusive memories at week 4 (postintervention). Night's sleep interference and daily functioning interference are missing for P1, hence N=7 for these items.

[i]How vivid were your intrusive memories? 0=not at all and 10=extremely.

[j]How distressing were your intrusive memories? 0=not at all and 10=extremely.

[k]How much did they disrupt your concentration? 0=not at all and 10=a great deal.

[l]How much did they disrupt the tasks you were doing? 0=not at all and 10=a great deal.

[m]How much did your intrusive memories interfere with your night's sleep? 0=not at all and 10=a great deal.

[n]How much have your intrusive memories affected your ability to function in your daily life? 0=not at all and 10=a great deal.

Intrusive Memory Characteristics and Impact on Functioning: Retrospective Questionnaire (End of Week)

All retrospective ratings of intrusive memory characteristics showed significant reductions from week 1 (preintervention) to week 4 (postintervention), except for ratings of how much intrusive memories interfered with sleep. Participants reported fewer intrusive memories and a shorter duration of being bothered by them from week 1 (preintervention) to week 4 (postintervention; Tables 2 and 3).

Table 3. Discrete outcomes of the retrospective ratings of intrusive memory characteristics (end of week) at week 1 (preintervention) and week 4 (postintervention; N=12).

	Week 1 (preintervention), n (%)	Week 4 (postintervention), n (%)
Discrete outcome[a]		
None	0 (0)	4 (33)
Some (1-4)	1 (8)	7 (58)
Quite a few (5-9)	4 (33)	1 (8)
Lots (10-20)	3 (25)	0 (0)
Very many (21-30)	2 (17)	0 (0)
A large amount (31-50)	2 (17)	0 (0)
More (>50)	0 (0)	0 (0)
Length of time that intrusive memories were bothersome[b] (min)		
<1	3 (25)	4 (50)
1-5	5 (42)	3 (38)
6-10	2 (17)	1 (13)
11-30	1 (8)	0 (0)
31-60	0 (0)	0 (0)
>60	1 (8)	0 (0)

[a]How many intrusive memories did you have? None (0), some (1-4), quite a few (5-10), lots (10-20), very many (21-30), a large amount (31-50), more (>50).

[b]Approximately how long did your intrusive memories bother you for? <1 minutes, 1-5 minutes, 6-10 minutes, 11-30 minutes, 31-60 minutes, and >60 minutes.

Responses to the open-ended question, "How have intrusive memories affected your ability to function in your daily life in the past week?" revealed that intrusive memories impacted various functional domains before the intervention, including occupational, cognitive, social and home life, and emotional functioning. For example, one participant stated:

> *I was previously reluctant to go to the next medical rotation when I have had memories about the event connected to the rotation, but now it's felt very much in my control. It couldn't have been a better timing to do the study.*

Another participant reported, "I can focus on the tasks at hand whilst at work." Another participant felt that they are "no longer struggling to get to sleep or waking up startled." One participant explained, "I'm more conversive again with people; I'm more chatty with my colleagues." Another reported:

> *The frustration within the family about my mentioning intrusive memories that I've been having...it's not happening anymore—I've not been needing to mention anything to my wife at all.*

One participant noted, "[My] manager noticed a difference—I'm more myself now than at the beginning of the study." One participant observed global changes at work, "At work I'm less stressed, less moody, and happier."

About 83% (10/12) of participants reported improved functioning at postintervention. For example, some indicated that when intrusive memories occurred, they no longer interfered with functioning. One participant noted, "The intrusive memories aren't as vivid and they don't last as long, so I can focus on the tasks at hand whilst at work." Another said, "I'm still getting some of the intrusive memories, but they're a lot less bothersome, and I feel able to manage them better." One participant noted global changes in the impact of intrusive memories, "I don't feel like [intrusive memories] are impacting me in any way now."

Acceptability and Feasibility: Feedback Questionnaire

Participants generally found playing Tetris at work helpful and not very burdensome. There was considerable variability in participants' ratings of how easy they found playing Tetris at work. Overall, participants found taking part in the study easy and not burdensome and reported feeling confident in suggesting playing Tetris to another staff member (Table S6 in Multimedia Appendix 1).

Participants generally found taking part in the study helpful, describing the intervention as accessible. A common theme was that playing Tetris enabled them to focus on something other than work: “It takes your concentration somewhere else.” Some found it a helpful “distraction” after an intrusive memory, “I play it from time to time following intrusive memories of other events in my life. It’s a massive distraction.”

Several participants noted a reduction in the emotional impact of their intrusive memories:

> *When I had the intrusive memory this week, it didn’t have such a strong emotional potency and didn’t have such a hold of me. You can deal with the memory better when it doesn’t have such a strong emotion attached to it. I can really see the benefits.*

Regarding the feasibility of playing Tetris while at work, participants mentioned challenges, such as lack of time or opportunity, as for some, breaktimes were the only times they were able to access the intervention. Even then, it was sometimes perceived as antisocial:

> *During breaktimes is the only time that it’s possible [to play Tetris], however, even then, the staffroom is very busy and it’s hard to concentrate. You might have others asking you to join in a conversation, and you don’t want to be seen as antisocial.*

Participants also highlighted potential challenges to the intervention’s acceptability in their work environments, such as: “I didn’t feel comfortable telling my nurse in charge that I need to go play Tetris,” and:

> *When colleagues describe intrusive memories, the expectation is that we sit there and listen. Culturally, we are not there yet to do something that might distract away from the memory. It might come across insensitive or that I’m not interested.*

Finally, participants recommended improvements, such as having Tetris on an app, a version without advertisements, shorter play sessions, and regular reminders to play Tetris.

Discussion

Principal Findings

This single-case series examined the efficacy, feasibility, and acceptability of a brief imagery-competing task intervention to reduce intrusive memories in frontline NHS health care staff, including both prehospital and hospital staff, exposed to workplace trauma. Overall, participants reported a substantial reduction (59%) in the number of intrusive memories per day from baseline to postintervention. All participants experienced a decrease in daily intrusive memories from baseline to postintervention, with reductions ranging from 51% to 100% in 10 of 12 (83%) participants. In addition, there was a notable improvement in depression, anxiety, and posttraumatic stress symptoms over the intervention period, although no change in sleep patterns was observed. Qualitative feedback indicated enhanced cognitive, behavioral, and emotional functioning across social, home life, and work-related domains following the intervention, emphasizing its feasibility and acceptability. Adherence to treatment protocols was excellent.

This study supports previous findings that indicate the effect of the intervention in reducing intrusive memories among trauma-exposed populations (eg, inpatients with complex PTSD [32], refugees [31], and ICU staff [36,37]), extending to various frontline health care staff employed in prehospital contexts, such as ambulance services, as well as hospital staff (eg, in the ED and ICU). Alongside reduced intrusive memory frequency, improvements in psychopathology symptoms (eg, posttraumatic stress symptoms) and overall functioning were observed, in accordance with evidence that intrusive symptoms are centrally linked to other PTSD symptoms [10].

Participants also reported significant reductions in intrusive memory–related distress and vividness, along with positive changes in their appraisals of intrusive memories postintervention, as assessed by the retrospective questionnaire (end of week). For example, following the intervention, participants rated their intrusive memories as less “bothersome,” and in qualitative responses, some reported feeling better able to manage intrusive memories when they did come to mind. However, significant changes were not observed when the characteristics of intrusive memories were measured in the daily intrusive memory diary. The discrepancies between the daily and retrospective measures of intrusive memory characteristics may be due to differences in data collection methods. Daily diary entries capture immediate, specific experiences, while retrospective ratings reflect a broader, overall assessment of intrusive memories over the past week, potentially encompassing global changes in participants’ perceptions and impact on functioning. Thus, although the primary objective of the intervention is to reduce the number of intrusive memories, our findings suggest possible broader clinical impact. Future evaluations should consider measuring outcomes beyond the frequency of intrusive memories, such as their qualities, associated appraisals, symptoms of psychopathology, and functional impacts, to capture the broader potential effects of the intervention.

The COVID-19 pandemic had an unavoidable impact on this research, necessitating several modifications to the study design and procedures after the study had commenced. First, the study duration was reduced, and it was not possible to proceed with the originally planned randomization to a 3- or 5-week baseline, which we acknowledge as a limitation. Second, due to the national lockdown, study procedures were adapted for remote rather than in-person delivery. Nevertheless, delivering the intervention according to the protocol despite these modifications underscores its flexibility and potential for remote delivery—2 key features of a scalable intervention.

Study Limitations

We acknowledge some important limitations of the study that should be considered when interpreting the findings. First, we used an AB single-case experimental design. Further research has since included randomized procedures to test the effectiveness and feasibility of the intervention for NHS staff, as demonstrated by Ramineni et al [36] and Iyadurai et al [37] for ICU staff. In addition, future studies could benefit from incorporating additional time-related controls, such as the stability of daily intrusive memories over time, the typical frequency of intrusive memories in the week before the study, and the long-term effects of the intervention. Finally, this study design does not allow us to distinguish the effects of specific intervention components, such as discarding the written descriptions of the intrusive memories. Further research with varied designs is needed to isolate and understand the impact of each individual component of the procedure.

Conclusions

In conclusion, this case series supports the efficacy of an imagery-competing task intervention in reducing the number of intrusive memories and improving mental health and functioning among frontline NHS prehospital and hospital staff, aligning with findings in other populations. Important next steps have included further optimizing the digital delivery of the intervention across various NHS health care staff groups and settings, using rigorous randomized controlled designs (eg, with an active control), and investigating longer-term effects of the intervention [36,37]. These and similar future trials will further enhance the development of a remotely delivered, scalable intervention aimed at mitigating the impact of work-related traumatic events on health care staff globally.

Acknowledgments

The authors would like to thank the participants who gave their time to this study, as well as staff from the Oxford University Hospitals National Health Service (NHS) Foundation Trust, the Royal Berkshire NHS Foundation Trust, and the South Central Ambulance Service NHS Foundation Trust, who provided input in adapting study materials for frontline health care staff and assisted with the recruitment process. This project was mainly supported by a grant to LI from the University of Oxford John Fell Fund (grant 0006754). VK was supported by funding from the Oxford Institute of Clinical Psychology Training and Research, EAH receives grants from the Swedish Research Council (grant 2020-00873), the Oak Foundation (grant OKAY-18-442), AFA Insurance (grant 200342), and LI was supported by National Institute of Health Research Capability Funding. EAH's salary was part funded by Wellcome Trust (grant 223016/Z/21/Z) via consultancy to P1vital Products Ltd. Sponsorship was provided by the University of Oxford Clinical Trials Research Governance.

Authors' Contributions

LI conceived the study. VK and LI designed the study with substantial contributions from CS, EAH, SB, MD, and LK. SB, MD, and LK additionally facilitated recruitment. EAH, MLM, and MK contributed to interpretation of the data and supported intervention training and delivery. EAH developed the imagery-competing task intervention for intrusive memories and training in using it (ANEMONE). All the authors were involved in drafting or revising the work and have approved the final version.

Conflicts of Interest

EAH is on the board of trustees of the MQ Foundation. EAH receives book royalties from Guilford Press and the Oxford University Press and receives occasional honoraria for conference keynotes and clinical workshops.

Multimedia Appendix 1

Data pertaining to pilot participants 1 and 2, further information for primary and secondary outcome analyses, further information regarding ethics approval, results from the within-intervention session measure, and supplementary tables.

[PDF File (Adobe PDF File), 449 KB - humanfactors_v11i1e55562_app1.pdf]

Multimedia Appendix 2

"Hotspots" form.

[PDF File (Adobe PDF File), 23 KB - humanfactors_v11i1e55562_app2.pdf]

Multimedia Appendix 3

Intrusive memory diary.

[PDF File (Adobe PDF File), 567 KB - humanfactors_v11i1e55562_app3.pdf]

Multimedia Appendix 4

Retrospective questionnaire (end of week).

[PDF File (Adobe PDF File), 78 KB - humanfactors_v11i1e55562_app4.pdf]

Multimedia Appendix 5

Feedback questionnaire.

[PDF File (Adobe PDF File), 53 KB - humanfactors_v11i1e55562_app5.pdf]

References

1. Mealer ML, Shelton A, Berg B, Rothbaum B, Moss M. Increased prevalence of post-traumatic stress disorder symptoms in critical care nurses. Am J Respir Crit Care Med 2007 Apr 01;175(7):693-697. [doi: 10.1164/rccm.200606-735OC] [Medline: 17185650]
2. Kleim B, Bingisser MB, Westphal M, Bingisser R. Frozen moments: flashback memories of critical incidents in emergency personnel. Brain Behav 2015 Jul;5(7):e00325 [FREE Full text] [doi: 10.1002/brb3.325] [Medline: 26221567]
3. Grey N, Holmes EA. "Hotspots" in trauma memories in the treatment of post-traumatic stress disorder: a replication. Memory 2008 Oct;16(7):788-796. [doi: 10.1080/09658210802266446] [Medline: 18720224]
4. Ehlers A, Clark DM, Hackmann A, McManus F, Fennell M. Cognitive therapy for post-traumatic stress disorder: development and evaluation. Behav Res Ther 2005 Apr;43(4):413-431. [doi: 10.1016/j.brat.2004.03.006] [Medline: 15701354]
5. Ehlers A, Steil R. Maintenance of intrusive memories in posttraumatic stress disorder: a cognitive approach. Behav Cogn Psychother 1995 Jul;23(3):217-249 [FREE Full text] [doi: 10.1017/S135246580001585X] [Medline: 21241539]
6. Holmes EA, Ghaderi A, Eriksson E, Lauri KO, Kukacka OM, Mamish M, et al. 'I can't concentrate': a feasibility study with young refugees in Sweden on developing science-driven interventions for intrusive memories related to trauma. Behav Cogn Psychother 2017 Mar;45(2):97-109 [FREE Full text] [doi: 10.1017/S135246581600062X] [Medline: 28229806]
7. Iyadurai L, Visser RM, Lau-Zhu A, Porcheret K, Horsch A, Holmes EA, et al. Intrusive memories of trauma: a target for research bridging cognitive science and its clinical application. Clin Psychol Rev 2019 Apr;69:67-82 [FREE Full text] [doi: 10.1016/j.cpr.2018.08.005] [Medline: 30293686]
8. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5). 5th edition. New York, NY: American Psychiatric Association; 2013.
9. Solberg Ø, Birkeland MS, Blix I, Hansen MB, Heir T. Towards an exposure-dependent model of post-traumatic stress: longitudinal course of post-traumatic stress symptomatology and functional impairment after the 2011 Oslo bombing. Psychol Med 2016 Nov;46(15):3241-3254. [doi: 10.1017/S0033291716001860] [Medline: 27609412]
10. Bryant RA, Creamer M, O'Donnell M, Forbes D, McFarlane AC, Silove D, et al. Acute and chronic posttraumatic stress symptoms in the emergence of posttraumatic stress disorder: a network analysis. JAMA Psychiatry 2017 Feb 01;74(2):135-142. [doi: 10.1001/jamapsychiatry.2016.3470] [Medline: 28002832]
11. McNally RJ. The ontology of posttraumatic stress disorder: natural kind, social construction, or causal system? Clin Psychol Sci Pract 2012 Sep;19(3):220-228 [FREE Full text] [doi: 10.1111/cpsp.12001]
12. Clohessy S, Ehlers A. PTSD symptoms, response to intrusive memories and coping in ambulance service workers. Br J Clin Psychol 1999 Sep;38(3):251-265. [doi: 10.1348/014466599162836] [Medline: 10532147]
13. Czaja AS, Moss M, Mealer M. Symptoms of posttraumatic stress disorder among pediatric acute care nurses. J Pediatr Nurs 2012 Aug;27(4):357-365 [FREE Full text] [doi: 10.1016/j.pedn.2011.04.024] [Medline: 22703683]
14. Laposa JM, Alden LE. Posttraumatic stress disorder in the emergency room: exploration of a cognitive model. Behav Res Ther 2003 Jan;41(1):49-65. [doi: 10.1016/s0005-7967(01)00123-1] [Medline: 12488119]
15. Greene T, Harju-Seppänen J, Adeniji M, Steel C, Grey N, Brewin CR, et al. Predictors and rates of PTSD, depression and anxiety in UK frontline health and social care workers during COVID-19. Eur J Psychotraumatol 2021 Mar 10;12(1):1882781 [FREE Full text] [doi: 10.1080/20008198.2021.1882781] [Medline: 33968317]
16. Li Y, Scherer N, Felix L, Kuper H. Prevalence of depression, anxiety and post-traumatic stress disorder in health care workers during the COVID-19 pandemic: a systematic review and meta-analysis. PLoS One 2021;16(3):e0246454 [FREE Full text] [doi: 10.1371/journal.pone.0246454] [Medline: 33690641]
17. Greenberg N, Weston D, Hall C, Caulfield T, Williamson V, Fong K. Mental health of staff working in intensive care during COVID-19. Occup Med (Lond) 2021 Apr 09;71(2):62-67 [FREE Full text] [doi: 10.1093/occmed/kqaa220] [Medline: 33434920]
18. COVID-19 pandemic one year on: ICN warns of exodus of experienced nurses compounding current shortages. International Council of Nurses. 2021 Mar 11. URL: https://www.icn.ch/news/covid-19-pandemic-one-year-icn-warns-exodus-experienced-nurses-compounding-current-shortages [accessed 2021-05-10]
19. Cottey L, Shanahan TA, Gronlund T, Whiting C, Sokunbi M, Carley SD, James Lind Alliance (JLA) Emergency Medicine (EM) Priority Setting Partnership (PSP) Refresh Steering Group. Refreshing the emergency medicine research priorities. Emerg Med J 2023 Sep;40(9):666-670 [FREE Full text] [doi: 10.1136/emermed-2022-213019] [Medline: 37491155]
20. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. Lancet Psychiatry 2020 Jun;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]
21. Post-traumatic stress disorder NICE guideline [NG116]. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/guidance/ng116/chapter/Recommendations#care-for-people-with-ptsd-and-complex-needs [accessed 2021-05-10]
22. Knaak S, Mantler E, Szeto A. Mental illness-related stigma in healthcare: barriers to access and care and evidence-based solutions. Healthc Manage Forum 2017 Mar;30(2):111-116 [FREE Full text] [doi: 10.1177/0840470416679413] [Medline: 28929889]

23. Holmes EA, James EL, Coode-Bate T, Deeprose C. Can playing the computer game "Tetris" reduce the build-up of flashbacks for trauma? A proposal from cognitive science. PLoS One 2009;4(1):e4153 [FREE Full text] [doi: 10.1371/journal.pone.0004153] [Medline: 19127289]
24. James EL, Bonsall MB, Hoppitt L, Tunbridge EM, Geddes JR, Milton AL, et al. Computer game play reduces intrusive memories of experimental trauma via reconsolidation-update mechanisms. Psychol Sci 2015 Aug;26(8):1201-1215 [FREE Full text] [doi: 10.1177/0956797615583071] [Medline: 26133572]
25. Kanstrup M, Singh L, Göransson KE, Widoff J, Taylor RS, Gamble B, et al. Reducing intrusive memories after trauma via a brief cognitive task intervention in the hospital emergency department: an exploratory pilot randomised controlled trial. Transl Psychiatry 2021 Jan 11;11(1):30 [FREE Full text] [doi: 10.1038/s41398-020-01124-6] [Medline: 33431807]
26. Baddeley AD, Andrade J. Working memory and the vividness of imagery. J Exp Psychol Gen 2000 Mar;129(1):126-145. [doi: 10.1037//0096-3445.129.1.126] [Medline: 10756490]
27. Monfils MH, Holmes EA. Memory boundaries: opening a window inspired by reconsolidation to treat anxiety, trauma-related, and addiction disorders. Lancet Psychiatry 2018 Dec;5(12):1032-1042. [doi: 10.1016/S2215-0366(18)30270-0] [Medline: 30385214]
28. Singh L, Espinosa L, Ji JL, Moulds ML, Holmes EA. Developing thinking around mental health science: the example of intrusive, emotional mental imagery after psychological trauma. Cogn Neuropsychiatry 2020 Sep;25(5):348-363. [doi: 10.1080/13546805.2020.1804845] [Medline: 32847486]
29. Horsch A, Vial Y, Favrod C, Harari MM, Blackwell SE, Watson P, et al. Reducing intrusive traumatic memories after emergency caesarean section: a proof-of-principle randomized controlled study. Behav Res Ther 2017 Jul;94:36-47 [FREE Full text] [doi: 10.1016/j.brat.2017.03.018] [Medline: 28453969]
30. Iyadurai L, Blackwell SE, Meiser-Stedman R, Watson PC, Bonsall MB, Geddes JR, et al. Preventing intrusive memories after trauma via a brief intervention involving Tetris computer game play in the emergency department: a proof-of-concept randomized controlled trial. Mol Psychiatry 2018 Mar;23(3):674-682 [FREE Full text] [doi: 10.1038/mp.2017.23] [Medline: 28348380]
31. Kanstrup M, Kontio E, Geranmayeh A, Olofsdotter Lauri K, Moulds ML, Holmes EA. A single case series using visuospatial task interference to reduce the number of visual intrusive memories of trauma with refugees. Clin Psychol Psychother 2021 Jan;28(1):109-123. [doi: 10.1002/cpp.2489] [Medline: 32525244]
32. Kessler H, Holmes EA, Blackwell SE, Schmidt A, Schweer JM, Bücker A, et al. Reducing intrusive memories of trauma using a visuospatial interference intervention with inpatients with posttraumatic stress disorder (PTSD). J Consult Clin Psychol 2018 Dec;86(12):1076-1090. [doi: 10.1037/ccp0000340] [Medline: 30507232]
33. Thorarinsdottir K, Holmes EA, Hardarson J, Stephenssen ES, Jonasdottir MH, Kanstrup M, et al. Using a brief mental imagery competing task to reduce the number of intrusive memories: exploratory case series with trauma-exposed women. JMIR Form Res 2022 Jul 20;6(7):e37382 [FREE Full text] [doi: 10.2196/37382] [Medline: 35857368]
34. Thorarinsdottir K, Holmes EA, Hardarson J, Hedinsdottir U, Kanstrup M, Singh L, et al. Reducing intrusive memories of childhood trauma using a visuospatial intervention: case study in Iceland. JMIR Form Res 2021 Nov 04;5(11):e29873 [FREE Full text] [doi: 10.2196/29873] [Medline: 34734830]
35. Singh L, Kanstrup M, Depa K, Falk A, Lindström V, Dahl O, et al. Digitalizing a brief intervention to reduce intrusive memories of psychological trauma for health care staff working during COVID-19: exploratory pilot study with nurses. JMIR Form Res 2021 May 26;5(5):e27473 [FREE Full text] [doi: 10.2196/27473] [Medline: 33886490]
36. Ramineni V, Millroth P, Iyadurai L, Jaki T, Kingslake J, Highfield J, et al. Treating intrusive memories after trauma in healthcare workers: a Bayesian adaptive randomised trial developing an imagery-competing task intervention. Mol Psychiatry 2023 Jul;28(7):2985-2994 [FREE Full text] [doi: 10.1038/s41380-023-02062-7] [Medline: 37100869]
37. Iyadurai L, Highfield J, Kanstrup M, Markham A, Ramineni V, Guo B, et al. Reducing intrusive memories after trauma via an imagery-competing task intervention in COVID-19 intensive care staff: a randomised controlled trial. Transl Psychiatry 2023 Sep 01;13(1):290 [FREE Full text] [doi: 10.1038/s41398-023-02578-0] [Medline: 37658043]
38. Lawn S, Roberts L, Willis E, Couzner L, Mohammadi L, Goble E. The effects of emergency medical service work on the psychological, physical, and social well-being of ambulance personnel: a systematic review of qualitative research. BMC Psychiatry 2020 Jul 03;20(1):348 [FREE Full text] [doi: 10.1186/s12888-020-02752-4] [Medline: 32620092]
39. Shabanikiya H, Kokabisaghi F, Mojtabaeian M, Sahebi T, Varmaghani M. Global prevalence of workplace violence against paramedics: a systematic review and meta-analysis. Health Emerg Disasters Q 2021 Jul 01;6(4):205-216 [FREE Full text] [doi: 10.32598/hdq.6.4.259.3]
40. Hogan KA, Fothergill-Bourbonnais F, Brajtman S, Phillips S, Wilson KG. When someone dies in the emergency department: perspectives of emergency nurses. J Emerg Nurs 2016 May;42(3):207-212. [doi: 10.1016/j.jen.2015.09.003] [Medline: 26435352]
41. Barlow DH, Nock M, Hersen M. Single Case Experimental Designs: Strategies for Studying Behavior Change. 3rd edition. Washington, DC: Pearson Prentice Hall; 2008.
42. Home page. Qualtrics XM. URL: https://www.qualtrics.com/uk/ [accessed 2021-05-10]

43. Shearing V, Lee D, Clohessy S. How do clients experience reliving as part of trauma-focused cognitive behavioural therapy for posttraumatic stress disorder? Psychol Psychother 2011 Dec;84(4):458-475. [doi: 10.1111/j.2044-8341.2010.02012.x] [Medline: 22903886]
44. The addictive puzzle game that started it all!. Tetris. URL: https://tetris.com/ [accessed 2024-07-02]
45. Hoppe JM, Walldén YS, Kanstrup M, Singh L, Agren T, Holmes EA, et al. Hotspots in the immediate aftermath of trauma - mental imagery of worst moments highlighting time, space and motion. Conscious Cogn 2022 Mar;99:103286 [FREE Full text] [doi: 10.1016/j.concog.2022.103286] [Medline: 35220032]
46. Weiss DS, Marmar CR. The impact of event scale: revised. In: Wilson JP, Tang CS, editors. Cross-Cultural Assessment of Psychological Trauma and PTSD. New York, NY: The Guilford Press; 1997:219-238.
47. Rash CJ, Coffey SF, Baschnagel JS, Drobes DJ, Saladin ME. Psychometric properties of the IES-R in traumatized substance dependent individuals with and without PTSD. Addict Behav 2008 Aug;33(8):1039-1047 [FREE Full text] [doi: 10.1016/j.addbeh.2008.04.006] [Medline: 18501524]
48. Creamer M, Bell R, Failla S. Psychometric properties of the impact of event scale - revised. Behav Res Ther 2003 Dec;41(12):1489-1496. [doi: 10.1016/j.brat.2003.07.010] [Medline: 14705607]
49. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
50. Martin A, Rief W, Klaiberg A, Braehler E. Validity of the brief Patient Health Questionnaire mood scale (PHQ-9) in the general population. Gen Hosp Psychiatry 2006;28(1):71-77. [doi: 10.1016/j.genhosppsych.2005.07.003] [Medline: 16377369]
51. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
52. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
53. Germain A, Buysse DJ, Shear MK, Fayyad R, Austin C. Clinical correlates of poor sleep quality in posttraumatic stress disorder. J Trauma Stress 2004 Dec;17(6):477-484. [doi: 10.1007/s10960-004-5796-6] [Medline: 15730066]
54. Morley S. Single Case Methods in Clinical Psychology: A Practical Guide. Hoboken, NJ: Routledge; 2017.
55. De TK, Michiels B, Vlaeyen JW, Onghena P. Shiny SCDA. KU Leuven. URL: https://ppw.kuleuven.be/mesrg/software-and-apps/shiny-scda [accessed 2024-06-24]
56. Kazdin AE. Single-case experimental designs. Evaluating interventions in research and clinical practice. Behav Res Ther 2019 Jun;117:3-17. [doi: 10.1016/j.brat.2018.11.015] [Medline: 30527785]
57. Parker RI, Vannest KJ, Davis JL, Sauber SB. Combining nonoverlap and trend for single-case research: Tau-U. Behav Ther 2011 Jun;42(2):284-299. [doi: 10.1016/j.beth.2010.08.006] [Medline: 21496513]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
ED: emergency department
GAD-7: Generalized Anxiety Disorder-7
ICU: intensive care unit
IES-R: Impact of Event Scale-Revised
NHS: National Health Service
PHQ-9: Patient Health Questionnaire-9
PSQI: Pittsburgh Sleep Quality Index
PTSD: posttraumatic stress disorder

Edited by K Adapa; submitted 01.01.24; peer-reviewed by A Horsch, G Colville; comments to author 12.03.24; revised version received 02.07.24; accepted 13.09.24; published 18.11.24.

Please cite as:
Kubickova V, Steel C, Moulds ML, Kanstrup M, Beer S, Darwent M, Keating L, Holmes EA, Iyadurai L
Reducing the Number of Intrusive Memories of Work-Related Traumatic Events in Frontline Health Care Staff During the COVID-19 Pandemic: Case Series
JMIR Hum Factors 2024;11:e55562
URL: https://humanfactors.jmir.org/2024/1/e55562
doi:10.2196/55562
PMID:

Original Paper

The Effect of a Video-Assisted Health Education Program Followed by Peer Education on the Health Literacy of COVID-19 and Other Infectious Diseases Among School Children: Quasi-Randomized Controlled Trial

Xiaojuan Zhang[1,2], BMed; Yingkun Justin Wen[1,3], BMed; Ning Han[4*], BMed; Yawen Jiang[1*], PhD

[1]School of Public Health (Shenzhen), Sun Yat-sen University, Shenzhen, Guangdong, China

[2]School of Public Health, Sun Yat-sen University, Guangzhou, China

[3]Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden

[4]Institute of Public Health Supervision of Longgang District, Shenzhen, China

[*]these authors contributed equally

Corresponding Author:
Yawen Jiang, PhD
School of Public Health (Shenzhen)
Sun Yat-sen University
Room 533, West Wing of Medical Complex #1, Sun Yat-sen University
66 Gongchang Road, Guangming District Shenzhen
Shenzhen, Guangdong
China
Phone: 86 13632974660
Email: jiangyw26@mail.sysu.edu.cn

Abstract

Background: To improve the engagement and effectiveness of traditional health programs, it is necessary to explore alternative models of health education including video-assisted lectures and peer education.

Objective: This study aimed to evaluate the effects of a combination of video-assisted lectures and peer education on health literacy related to infectious diseases among students.

Methods: Third-grade classes from 11 pilot schools in Longgang District of Shenzhen, China, were randomized to the intervention and control groups. In the intervention group, a video-assisted interactive health education program was conducted twice over a time span of 5 months. Each of the 2 sessions included a 40-minute lecture on COVID-19 and other common infectious diseases in schools and a 5-minute science video. In addition, 5 "little health supervisors" at the end of the first session were elected in each class, who were responsible for helping class members to learn health knowledge and develop good hygiene habits. Students answered the same quiz before the first and after the second session. Models based on item response theory (IRT) were constructed to score the students' knowledge of infectious diseases based on the quiz.

Results: In total, 52 classes and 2526 students (intervention group: n=1311; control group: n=1215) were enrolled. Responses of the baseline survey were available for 2177 (86.2%; intervention group: n=1306; control group: n=871) students and those of the postintervention survey were available for 1862 (73.7%; intervention group: n=1187; control group: n=675). There were significant cross-group differences in the rates of correctly answering questions about influenza symptoms, transmission, and preventive measures; chicken pox symptoms; norovirus diarrhea symptoms; mumps symptoms; and COVID-19 symptoms. Average IRT scores of questions related to infectious diseases in the intervention and control groups were, respectively, –0.0375 (SD 0.7784) and 0.0477 (SD 0.7481) before the intervention (P=.01), suggesting better baseline knowledge in the control group. After the intervention, the average scores of the intervention and control groups were 0.0543 (SD 0.7569) and –0.1115 (SD 0.7307), respectively (P<.001), suggesting not only significantly better scores but also greater improvement in the intervention group.

Conclusions: After the health education project, the correct answer rate of infectious disease questions in the intervention group was higher than that of the control group, which indicates significant effects of the combination of video-assisted lectures and peer education for the promotion of health literacy. In addition, the intervention effect of the first session persisted for at least 4 months up to the second session. As such, the proposed program was effective in improving the health literacy of school children in relation to infectious diseases and should be considered for massive health promotion campaigns during pandemics.

Trial Registration: ISRCTN ISRCTN49297995; https://www.isrctn.com/ISRCTN49297995

(JMIR Hum Factors 2024;11:e43943) doi:10.2196/43943

KEYWORDS

infectious diseases; primary school students; quasi-randomized controlled trial; video-assisted health education; peer education; item response theory; IRT

Introduction

Primary school students are vulnerable to emerging and common infectious diseases such as COVID-19, influenza, mumps, and intestinal infectious diseases [1]. A survey on the reasons for sick leaves in primary and secondary schools in Shenzhen showed that the top 5 causes were common cold, gastrointestinal diseases, unexplained or other illness, influenza, and chicken pox [2]. In addition, the importance of acquiring essential knowledge regarding the prevention and control of COVID-19 cannot be overstated during the pandemic. Accordingly, it is critical to embed health promotion into the school education of primary school students. To that end, the outline of "Healthy China 2030" emphasizes the importance of fortifying health education among school children. In particular, primary schools were integral to the life cycle of the health education curriculum to the extent that early-life exposure to information on diseases and health behaviors is associated with improved future health outcomes [3].

Despite its importance, health education was highly restricted in its delivery forms. Conventionally, the most prevalent approach of health education of infectious diseases for school students was, arguably, classroom lectures aided with paper-based materials, in which the teaching contents are usually compiled by school teachers and researchers [4]. Traditional health education is also reported to have a limited duration of effects. Hampered by the collective challenges faced in traditional health education, most schools lack systematic health education programs [5]. To increase students' interest in healthy behaviors and to extend the duration of education effects, researchers have been exploring alternative media for health education. Among the various new models, two of the prevailing approaches are video-assisted health education and interactional peer education [5,6].

In professional medical education, video-assisted lectures are useful tools for students to acquire basic clinical skills. When delivered in bundle with in-person lectures, video-based materials are often preferred by students [7]. In addition, video-assisted health education has been shown to be more effective than oral education in facilitating postoperative recovery of patients [8].

The effects of health education are not necessarily limited to the immediate recipients of the program themselves. Students may also help to shape the opinions and behaviors of their classmates by becoming peer educators of health and hygiene. Peer education is defined as "sharing experiences and learning among people with something in common," such as a similar age, living environment, and culture [9]. There is substantial evidence that peer education is highly effective in specific areas of medical and health education, including professional medical training, chronic disease prevention, and sexual health behaviors [6,10,11]. Incorporating peer effects into the design of health education programs could, therefore, strengthen the programs' impacts on behavioral change.

However, evidence on the effects of video-assisted lectures and peer education on health literacy among school children is still lacking. Given its substantial potential for public health practice, we designed a health education package that combined video-assisted classroom teaching and peer education and tested the effectiveness of this program. This program, which we anecdotally refer to as the "Little Health Supervisors" project, was anticipated to improve the health literacy of students over an array of infectious diseases.

Methods

Trial Design

The "Little Health Supervisors" project is jointly enacted by the Longgang District Bureau of Health and the Longgang District Bureau of Education as an administrative task. Third-grade classes from 11 pilot schools in Longgang district of Shenzhen, China, were randomized to the intervention and control groups. Our aim was to allocate equal numbers of third-grade classrooms to the intervention and control groups within each school. However, schools with an odd total number of classes inevitably resulted in uneven groups; hence, one group might outnumber another eventually. This project enclosed 2 health education sessions 4 months apart in Dec 2021 and Apr 2022 in Longgang District, Shenzhen City in the Guangdong Province of China, which is a district with approximately 4 million residents and 0.4 million school students.

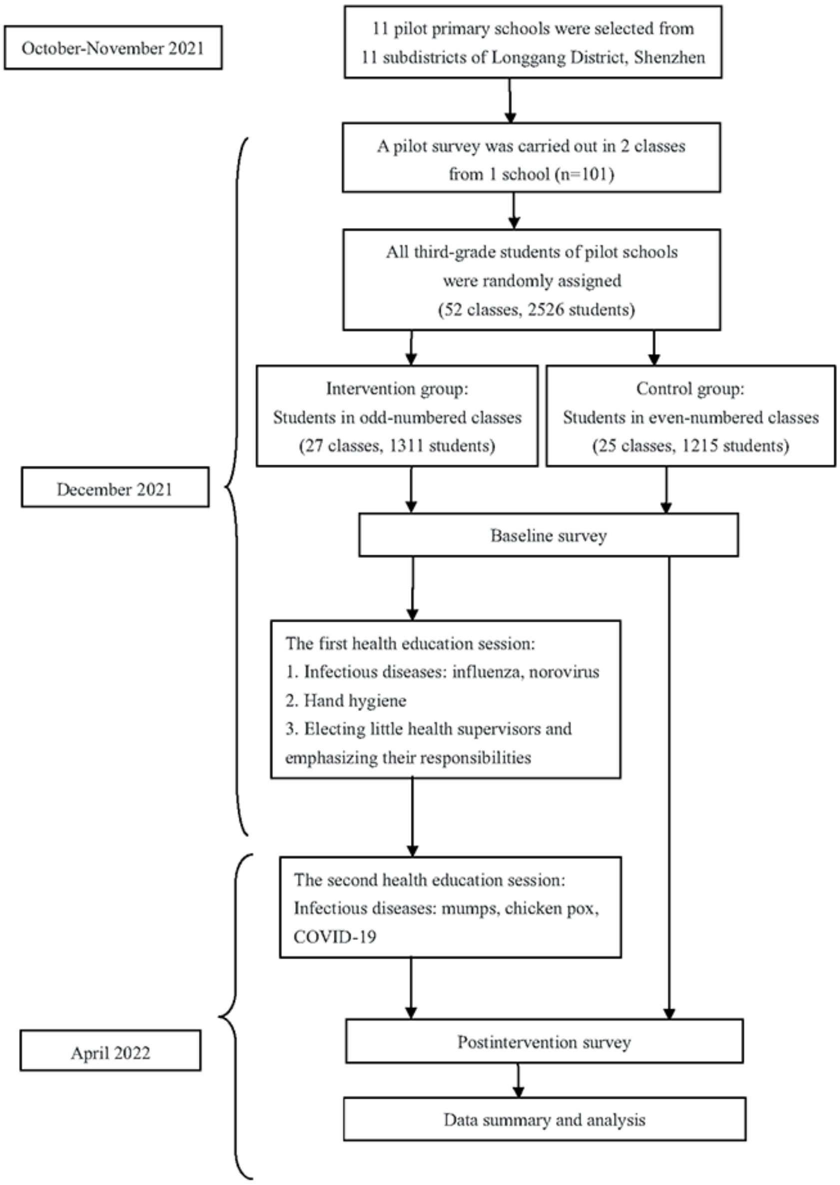
Ethical Considerations

The "Little Health Supervisors" project was launched by the district government as a public service project. The study protocol was approved by the Biomedical Research Ethics Review Committee, School of Public Health (Shenzhen), Sun Yat-sen University [2021(056)] and was registered with Longgang District Bureau of Health (Figure S1 in Multimedia Appendix 1). Informed consent was obtained from all students and their parents who met the inclusion criteria and were willing to participate. Confidentiality of information was maintained.

Recruitment

In the first step of sample enrollment, considering the feasibility of the project's implementation, the Longgang District Bureau of Health and the Longgang District Bureau of Education recommended 1 primary school based on the willingness to participate for each of the 11 subdistricts of this district. Second, all third-grade students in the 11 schools were eligible for participation if they met the following requirements: (1) they were not taking a leave of absence from school at the time of enrollment; (2) they agreed (or their guardians agreed) to spend time on attending lectures; (3) they had access to a computer, tablet, or smartphone with an internet connection; (4) they had sufficient knowledge to use mobile devices or computers (assistance allowed); and (5) they were able to read and interpret Chinese characters. Next, as decided by the researchers, the eligible students were assigned to the intervention and control groups using the class number as the randomizer. Specifically, odd-numbered classes were assigned to the intervention group; even-numbered classes, the control group. The flowchart of participant enrollment is illustrated in Figure 1.

Figure 1. Flowchart of the "Little Health Supervisors" project (a cluster randomized controlled trial) from December 2021 to April 2022.



Data Collection

To standardize students' knowledge of infectious diseases both before and after the education program, a questionnaire containing a quiz on COVID-19 and selected infectious diseases with relatively high local incidences was curated, which included influenza, chicken pox, norovirus diarrhea, and mumps. The questionnaire also collected demographic characteristics (school, class, student number, sex, and date of birth) and COVID-19 vaccination status. In addition, we delivered a separate questionnaire to a parent of each student who collected parental assent to vaccinate their children against COVID-19. Moreover, family socioeconomic information was also collected in the parent questionnaire, which included monthly household income and the parents' education level [12].

Questionnaires were distributed via a web-based survey platform (Wenjuanxing, Changsha Ranxing Information Technology Co, Ltd). In the baseline survey, students completed the questionnaires in a computer laboratory with the instructions of either the computer teachers or the class advisors. To collect parents' responses, the teachers arranged a meeting with each family using previously connected social media to select a representative for questionnaire responses. Due to COVID-19 outbreaks during the planned time period of the second session, the postintervention survey was distributed on the web. Simultaneously, the researchers also collected the questionnaire from the control group.

Interventions

The intervention was developed by both researchers and the local health department. Details of the development process and the content of the intervention are provided in Table S1 and Figures S2-S3 in Multimedia Appendix 1 [13,14]. Students randomized to the intervention group had access to 2 free sessions of health education during the study. Each session included a 40-minute lecture on the transmission and prevention of different infectious diseases, followed by a 5-minute science video. To incentivize learning, students were informed that there would be interactional question-and-answer sections during the lecture, for which the participating students were eligible for prizes.

In December 2021, the baseline survey and the first health education session were conducted, with the former preceding the latter. The in-person lecture and the videos of the first session pertained to influenza, norovirus diarrhea, and hand hygiene. At the end of this session, 5 little health supervisors were elected by the teachers from each class. They were naturally assumed as opinion leaders, showcasing their ability to effectively convey knowledge and could supervise the learning of health knowledge and the development of good hygiene habits of their classmates. The teachers also handed out brochures, armbands, and stickers to the 5 little health supervisors. In addition, the teachers encouraged all students to take health knowledge home and improve the family's health literacy by way of "small hands holding big hands," which aimed to exploit the power of two-step flow theory of communication for information transmission. Originating from political science, the two-step flow theory asserts that information can be conveyed through the chain of media-opinion leaders-audience. Students may also help to shape the opinions and behaviors of their family members by becoming an opinion leader of health and hygiene [15,16].

In April 2022, the second health education session and the postintervention survey were carried out. However, the order of education and survey was reversed in relation to the first session. The lecture and the videos of the second session pertained to chicken pox, mumps, and COVID-19 symptoms. Affected by a local COVID-19 outbreak, students had to take the web-based classes at home, so the health education sessions had to be conducted in the form of recorded course videos. In the intervention group, students were required to watch the video, and the teachers also encouraged all students to distribute health knowledge to the people around them.

As for the control group, the students only received routine health education at school, which included health tips on influenza from school doctors and 1 or 2 public welfare courses conducted by the local health department or hospitals every semester. These routine health education sessions were balanced between the 2 groups.

Outcomes

The primary outcomes of this trial were the score in the original scale (hereafter referred to as "crude score") and item response theory (IRT) score of questions related to infectious diseases, the correct answer rates of questions related to infectious diseases, and the pre-post changes in the correct answer rates after the intervention. The secondary outcomes were the COVID-19 vaccination rates. For those who did not receive COVID-19 vaccines at baseline or at the end of the program, we also exploratively asked about their willingness to get vaccinated and the reasons for not being vaccinated.

Statistical Analysis

To gain an overview of students' characteristics, their families' demographic data were collected. Monthly household income (in ¥) was categorized into 4 levels (<¥5000 [US $702.97], ¥5000 [US $702.97]~¥10,000 [US $1405.94], ¥10,000 [US $1405.94]~¥20,000 [US $2811.88], and ≥¥20,000 [US $2811.88]). Parent's education was grouped into 3 levels (junior high or below, secondary school [including technical secondary school], and college and above]. For the questions related to infectious diseases, multiple answers were regarded as correct only if all the correct answers were selected. Correctly answered questions contributed 1 point, and incorrectly answered questions contributed 0 points. The crude score of questions related to infectious diseases ranged from 0 to 7, with a higher score indicating higher knowledge of infectious diseases. For the item of willingness to be vaccinated against COVID-19, we assigned 1, 2, 3, 4, and 5 points respectively to the 5 options of very reluctant, reluctant, neutral, willing, and very willing. To comprehensively evaluate the students' knowledge of infectious diseases, IRT was used to fit the model of 7 items of the questionnaire. Frequently used in studies on education examinations, IRT is a set of psychometric models used to measure unobservable characteristics of the respondents and the development of scoring scales [17-19]. IRT can be used to explain the relationship between a latent trait (eg, the health literacy of school children related to infectious diseases) and observable characteristics and items (eg, questionnaire answers). IRT has at least 3 model specifications. The one parameter logistic model takes item difficulty into account when evaluating individual ability, whereas the two parameter logistic model additionally considers differential discrimination of items [19,20]. In addition to these 2 models, the three-parameter model (TPM) allows the possibility of guessing [19,20]. In this study, a TPM was selected to calculate the IRT score (Table S2 in Multimedia Appendix 1). To score the students' latent health literacy, we fitted TPM using the R package "ltm: Birnbaum's three parameter model" to the 7 questions related to the knowledge of infectious diseases [20]. A higher score meant higher health literacy. We plotted the estimated IRT score of questions related to infectious diseases to visualize the students' performance (Figures S4-S7 in Multimedia Appendix 1). Although not directly related to our main analyses, we also plotted the item characteristic curves, item information curves, and the test information curve to provide some information regarding the difficulty of the test (Figures S8-S10 in Multimedia Appendix 1).

Finally, to summarize categorial sociodemographic characteristics, the correct answer rates of answering the questions, the pre-post changes in the correct answer rates after the intervention, the COVID-19 vaccination rate, the reasons for nonvaccination, and the percentages of the corresponding variables were calculated. We used mean and SD to describe the crude score, the IRT score of questions related to infectious diseases, and the willingness to vaccinate against COVID-19. We used *t* tests to compare the crude score, the IRT score, and the willingness to be vaccinated against COVID-19 across groups. Regarding the willingness to be vaccinated between 2 groups, we also conducted a stratified analysis based on the parents' sex. Chi-square tests were carried out on the basis of the correct answer rate, the COVID-19 vaccination rate, and the reason for nonvaccination to investigate differences between the 2 groups. The pre-post changes in the correct answer rates after the intervention were compared between study groups, using the *z* test. Furthermore, since we used class as our intervention unit, we also conducted an additional analysis using class as the primary unit of analysis. This was undertaken to ensure that our class-based examination would yield coherent findings as well (Tables S3-S5 in Multimedia Appendix 1). A *P* value less than .05 was considered significant. All data were analyzed using SPSS (version 26; IBM Corp) and R (version 4.2.0; The R Foundation).

Power

We calculated the power of this study on the basis of the sample size of the intervention and on the primary outcome. To calculate power, we used the sample size of 1862 (intervention group: n=1187; control group: n=675), an acceptable probability for type I error of .05, a pooled SD of 0.767, and a minimal difference in the infectious disease knowledge scores between the 2 groups of 0.166 (ie, μ_1–μ_2). The power of this study was 99.43%.

Data Exclusion

First, when an intervention group student decided to quit or was lost to follow-up, the student was excluded from the primary analysis. Second, the researchers checked information such as IP address, birth date, sex, and school and class codes to identify duplicates.

Results

Study Population

In the baseline survey, 2177 (intervention group: n=1306; control group: n=871) student questionnaires and 2496 (intervention group: n=1430; control group: n=1066) parent questionnaires were collected, amounting to response rates of 86.2% and 98.8%, respectively. In the postintervention survey, 1862 (intervention group: n=1187; control group: n=675) student questionnaires and 1799 (intervention group: n=1076; control group: n=723) parent questionnaires were retrieved, yielding response rates of 73.7% and 71.2%, respectively (Tables S6-S9 in Multimedia Appendix 1). In the intervention group, 2493 (intervention group: n=1306; control group: n=1187) student questionnaires were collected, with a response rate of 95.1%. In the control group, 1546 (baseline survey: n=871; postintervention survey: n=675) student questionnaires were collected, with a response rate of 63.6%.

There were no significant differences in baseline characteristics between the intervention and control groups (Table 1). In the intervention group, there were 691 male and 615 female students; the corresponding numbers in the control group were 459 and 412, respectively. The proportion of households earning less than ¥5000 (US $702.97) was relatively small in both groups (9.8% and 8.9%). Finally, the proportions of students whose parents had college education and above was 72.7% in both groups.

Table 1. Sociodemographic characteristics of third-grade students from 11 pilot schools in Longgang District of Shenzhen, China.

Characteristics	Intervention group, n/n (%)	Control group, n/n (%)
Sex		
Male	691/1306 (52.9)	459/871 (52.7)
Female	615/1306 (47.1)	412/871 (47.3)
Monthly household income (¥[a])		
<5000	140/1430 (9.8)	95/1066 (8.9)
5000~10,000	359/1430 (25.1)	267/1066 (25.0)
10,000~20,000	403/1430 (28.2)	311/1066 (29.2)
≥20,000	528/1430 (36.9)	393/1066 (36.9)
Parent's educational level		
Junior high or below	116/1430 (8.1)	82/1066 (7.7)
High school or technical secondary school	275/1430 (19.2)	209/1066 (19.6)
College and above	1039/1430 (72.7)	775/1066 (72.7)

[a]¥1=US $0.1445.

Correct Answer Rates of Questions Related to Infectious Diseases

At baseline, the correct answer rates for questions related to influenza symptoms, influenza preventive measures, and norovirus diarrhea symptoms were different between the intervention and control groups. Specifically, the correct answer rate was higher in the control group (Table 2). In terms of the correct answer rates for questions regarding influenza transmission, chicken pox symptoms, mumps transmission, and COVID-19 symptoms, there were no significant differences between the 2 groups (Table 2). After the intervention, the differences between the 2 groups in the correct answer rates for questions regarding influenza symptoms, influenza preventive measures, and norovirus diarrhea symptoms were no longer observed (Table 2). By contrast, the differences in the correct answer rates for questions regarding chicken pox symptoms, mumps transmission, and COVID-19 symptoms between the 2 groups at the end point were significant, such that intervention group outperformed the control group (Table 2).

Table 2. The correct answer rates for questions related to infectious diseases in the intervention and control groups.

Questions	Total, %	Intervention group, %	Control group, %	*P* value
Baseline				
Influenza symptoms	67.57	65.39	70.84	.008
Influenza transmission	81.86	81.47	82.43	.57
Influenza preventive measures	84.66	83.08	87.03	.01
Norovirus diarrhea symptoms	56.41	54.21	59.70	.01
Chicken pox symptoms	28.34	28.79	27.67	.57
Mumps transmission	6.89	7.27	6.31	.39
COVID-19 symptoms	29.54	28.33	31.34	.13
End point				
Influenza symptoms	86.09	86.77	84.89	.26
Influenza transmission	78.30	78.69	77.63	.60
Influenza preventive measures	92.91	93.09	92.59	.69
Norovirus diarrhea symptoms	72.93	73.80	71.41	.26
Chicken pox symptoms	43.18	47.01	36.44	<.001
Mumps transmission	10.15	13.23	4.74	<.001
COVID-19 symptoms	49.14	52.40	43.41	<.001

Regarding the pre-post changes in the correct answer rates after the intervention, the differences between the 2 groups were significant for all items (Table 3). Specifically, the correct answer rates for questions regarding influenza symptoms, influenza preventive measures, norovirus diarrhea symptoms, chicken pox symptoms, and COVID-19 symptoms increased in both groups (for all, P<.001). However, the correct answer rates of the intervention group increased more than those of the control group. In the intervention group, the correct answer rate for questions regarding mumps transmission increased in the intervention group but decreased slightly in the control group. Compared with that before the intervention, the correct answer rate for questions regarding influenza transmission decreased slightly after the intervention (Table 3).

Table 3. Pre-post changes in the correct answer rates after the intervention in the intervention and control groups.

Questions	Total, %	Intervention group, %	Control group, %	*P* value
Influenza symptoms	18.52	21.38	14.05	<.001
Influenza transmission	–3.56	–2.78	–4.80	.02
Influenza preventive measures	8.25	10.01	5.56	<.001
Norovirus diarrhea symptoms	16.52	19.59	11.71	<.001
Chicken pox symptoms	14.84	18.22	8.77	<.001
Mumps transmission	3.26	5.96	–1.57	<.001
COVID-19 symptoms	19.60	24.07	12.07	<.001

Crude and IRT Scores for Questions Related to Infectious Diseases

Before the intervention, there was a significant difference in the mean scores for questions regarding infectious disease knowledge between the 2 groups. The mean IRT score of the intervention group (–0.0375, SD 0.7784) was significantly lower (*P*=.01) than that of the control group (0.0477, SD 0.7481). After the intervention, the mean IRT score of the intervention group (0.0543, SD 0.7569) surpassed that of the control group (–0.1115, SD 0.7307). Notably, the postintervention mean score of the intervention group increased from that at baseline, whereas the control group displayed an opposite trend (Table 4). The situation is similar for the crude score (Table 4).

Table 4. The crude and item response theory (IRT) score of questions related to infectious diseases in the intervention and control groups.

	Crude score, mean (SD)			IRT-based score, mean (SD)		
	Intervention group	Control group	*P* value	Intervention group	Control group	*P* value
Baseline	3.49 (1.628)	3.65 (1.552)	.02	–0.0375 (0.7784)	0.0477 (0.7481)	.01
End point	4.45 (1.469)	4.11 (1.420)	<.001	0.0543 (0.7569)	–0.1115 (0.7307)	<.001

COVID-19 Vaccination Rates

The COVID-19 vaccination rates of the intervention and the control groups at baseline were 94.8% and 93.2%, respectively; by the end of the program, they increased slightly to 97.6% and 96.6%, respectively. The differences, however, were not significant (Table 5).

Table 5. The COVID-19 vaccination rates of third-grade students before and after the intervention.

	Intervention group, n/n (%)	Control group, n/n (%)	*P* value
Baseline	1238/1306 (94.8)	812/871 (93.2)	.13
End point	1158/1187 (97.6)	652/675 (96.6)	.23

Willingness to Get Vaccinated and the Reasons for Not Being Vaccinated

Among the study participants who have not been vaccinated against COVID-19, the differences between students' and parents' willingness to receive the vaccine in the 2 groups were not significant (Table S10 in Multimedia Appendix 1). After stratifying by parents' sex, the differences between the 2 groups were still not significant (Table S11 in Multimedia Appendix 1). For students who had not been vaccinated against COVID-19 after the intervention, the students and their parents were worried about side effects among many other reasons (Table S12 in Multimedia Appendix 1).

Discussion

Principal Results

Using a quasi-randomized controlled design, this study assessed the effectiveness of a video-assisted health education program sequenced by peer education on infectious disease health literacy among school students. The results suggest that the proposed multicomponent model of health education improved the knowledge of infectious diseases among students, and are consistent with those of previous randomized controlled trials in health education among primary school students [12,21,22]. Moreover, this study not only showcases an innovative approach to raise awareness of disease prevention by incorporating technology and behavioral elements, but also represents a preliminary effort to test the effectiveness of an infectious disease health education program using IRT-based scores.

Our results encapsulate important implications for the practice of health education and healthy behavior promotion. First, the inexpensive and convenient innovative health education approach proposed in this study represents a viable approach to improve student health literacy during pandemics and should be considered in future programs of healthy behavior promotion among school students. The fact that the program was effective among third-grade students does not restrict the potential of this approach since senior students are likely to capture the contents of the program better than third-grade students. Second, the results from the second session of this study partially indicate that web-based teaching may also be an effective tool to promote student engagement in health education, which has been highlighted in previous studies but not confirmed [7].

The possible long-term effects of the first session from our findings should not be ignored. The postintervention survey was carried out immediately after the second education session (including chicken pox, mumps, and COVID-19) and 4 months after the first education session (including influenza and norovirus diarrhea). Despite the time elapsed, the correct answer rates of questions related to infectious diseases that were of focus in the first session were still higher in the intervention group than in the control group. Therefore, third-grade primary school students may endure the impact of health education for at least 4 months. Given the low likelihood of frequently setting up health education sessions in schools, the slow waning of the program's effects is a desirable feature. However, the cross-over effect from the second session could not be ruled out. For example, the learning of COVID-19 may strengthen the students' previous understanding of influenza and increase the effect of intervention in influenza. In addition, the second session may sensitize the students in the intervention group. They may review the knowledge of the first session to prepare for the postintervention quiz, which may also enhance the effect of the first session.

It is noteworthy that there was some difference in response rates between the interventional and control groups. The difference in response rates might be attributable to an absence of treatment blinding. In fact, the intervention in this study could not be blinded due to its physical nature, in which case, the intervention group students might be motivated by the education sessions to meet the expectation of the educators to respond to the surveys.

In addition, there was no significant difference in the correct answer rate for questions related to flu transmission routes before or after the intervention, but the pre-post changes in the correct answer rates was different between the 2 groups, and the intervention group performed better than the control group. Owing to countrywide vaccination campaigns, the COVID-19 vaccination rates between the intervention and control groups were not significantly different. The results of the 2 questionnaire surveys showed that the vaccination rates of the 2 groups increased, which was related to the local epidemic and the country's policy encouragement for vaccination.

Limitations

Several limitations of the study should be noted when interpreting the results. First, we did not collect data on the incidence of related infectious diseases before and after the intervention. A previous study reported that in areas with a high incidence of infectious disease, the health education package had no overall effect in preventing infections. However, the intervention was effective in preventing infections in areas where the baseline prevalence was relatively low [21]. Further studies are needed to explore the impact of our composite intervention on preventing infections. Second, this study was limited in its ability to evaluate component-specific versus composite effects of the educational video, the didactic lessons, the cooperative learning exercises, and peer engagement. The 2-arm trial design could not parse out the influence of each element. Future work should incorporate multiple comparison arms to better isolate the impacts of intervention components. Third, we regret that we did not measure changes in attitudes and behaviors after the intervention, as the health education package is hypothesized to influence these aspects. This is a gap that exists in our study, which future research could explore. Fourth, we used a self-rating questionnaire to collect data. Although self-reporting is a common and accepted method, we could not completely rule out the possibility of measurement error. However, the reliability and validity of self-reporting among children aged >8 years have been shown to be good in health-related questionnaires [23,24]. Fifth, the contamination in this study may underestimate the effect of our intervention. We adopted a clustered quasi-randomized controlled trial design to mitigate within-class person-to-person contamination, although interclass contamination caused by students and teachers could not be eliminated. However, the contamination, if any, happened more likely to the first session rather than the second session since students were physically isolated during the latter. Sixth, as we did not receive the questionnaire from the students lost to follow-up, the primary analysis was not intent-to-treat. Seventh, the second session of health education originally scheduled to enter the campus was changed to web-based classes owing to the serious local epidemic. Therefore, the students were required to fill in the web-based questionnaire at home, which affected the independence of the participants in answering questions; hence, the correct answer rates of the 2 groups were generally higher than those at baseline. Besides, the recovery of the questionnaire was decreased probably due to the lack of the teachers' supervision outside the schools. However, the missing rates were balanced between the 2 groups, thereby reducing the chances of influencing our conclusions. Moreover, the effect of the health education provided herein may be underestimated because this missing group of students and parents might have lower health literacy, in which case, the intervention would have incremental value.

Comparison With Prior Work

Despite these limitations, the primary strengths of our study are that it is the first quasi-randomized controlled trial to evaluate the effect of a video-assisted health education program sequenced by peer education on the health literacy of COVID-19 and other infectious diseases among school children, and it is also the first to report IRT scores for questions related to the infectious diseases. Additionally, while our study is quasi-randomized, the allocation process likely achieved reasonable randomization, effectively balancing confounding factors across study arms as evidenced by the systematic allocation of students to intervention or control groups based on their odd or even class numbers, as outlined in Table 1. Importantly, the allocation of students to odd or even classes was not based on systematically different characteristics, as the Ministry of Education of the People's Republic of China does not permit students to be segregated into different classes based on specific attributes. Therefore, the grouping of students based on class number parity can be considered to approximate the effects of randomization. Moreover, the sample size in this study allowed minimal chances of underpowered analyses. Previous studies might have engaged nonrandomized designs such that mixed results were reported [12,21,22,25-30]. Although most studies demonstrated that the health intervention is effective in improving health knowledge and health literacy, a quasi-randomized controlled trial in China found that the intervention's effect was not significant among primary school students [25]. Moreover, a number of studies adopted self-control, or observational designs, based on which solid conclusions are difficult to derive [3-5,26-30].

Conclusions

Our study confirmed that the combination of video-assisted and peer education in a health education program had significant effects on school children. In addition, the effect of the first health education session may endure after 4 months. As such, the proposed program was effective in improving health literacy related to infectious diseases among school children and should be considered for en masse health promotion campaigns during pandemics.

Acknowledgments

This study was carried out by the Longgang District Bureau of Health as an administrative task. This sponsor was involved in developing the intervention and study design, and was kept informed during data collection and data analysis. The corresponding authors had full access to all of the data in the study and had the final responsibility for the decision to submit the paper for publication. We also appreciate Jing Chen and Wanjing Zhang from the Longgang District Health Inspection Institute for their contribution and support to the early development of the intervention package.

Data Availability

The data presented in this study are not publicly available because of ethical requirements but are accessible upon reasonable request to the corresponding author.

Authors' Contributions

All authors conceptualized and designed the study and provided administrative, technical, and material support; they also supervised the study. XZ and YJW acquired the data. XZ analyzed and interpreted the data and drafted the manuscript. YJ critically revised the manuscript for important intellectual content. XZ carried out the statistical analyses. NH is a co-correspondent (email: 207baby@163.com).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary figures and tables.
[DOCX File , 922 KB - humanfactors_v11i1e43943_app1.docx]

Multimedia Appendix 2
CONSORT eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1240 KB - humanfactors_v11i1e43943_app2.pdf]

References

1. Dong Y, Wang L, Burgner DP, Miller JE, Song Y, Ren X, et al. Infectious diseases in children and adolescents in China: analysis of national surveillance data from 2008 to 2017. BMJ 2020 Apr 02;369:m1043 [FREE Full text] [doi: 10.1136/bmj.m1043] [Medline: 32241761]
2. Wu Y, Liao Y, Zhou L, Luo Q, Wang Y, Li B. Monitoring of student absence due to illness in the first semester of 2018 in Shenzhen. Chin J Sch Health 2019;40(10):1584-1586. [doi: 10.16835/j.cnki.1000-9817.2019.10.044]
3. Qiu H, Yu Y, Liu B, Shen Y, Shao H. Study on effect of health education of hepatitis B among the primary, middle and high school students in Zhangjiagang of Jiangsu Province. J Med Pest Control 2019;35(08):751-747. [doi: 10.7629/yxdwfz201908008]
4. Liu J, Liu W. Evaluation of the effect of school-based sexuality education in reducing AIDS discrimination in primary school. Chin J Sch Health 2019;35(12):1121-1110. [doi: 10.16168/j.cnki.issn.1002-9982.2019.12.012]
5. Daoyong C, Shuo W, Tong L, Fei W, Yue L, Jixing Y. Effect evaluation of the video education on infectious disease knowledge among primary school students in Hongkou District, Shanghai. Health Promot Educ 2019;14(06):529-531. [doi: 10.16117/j.cnki.31-1974/r.201906018]
6. Layzer C, Rosapep L, Barr S. Student voices: perspectives on peer-to-peer sexual health education. J Sch Health 2017 Jul;87(7):513-523. [doi: 10.1111/josh.12519] [Medline: 28580671]
7. Coyne E, Rands H, Frommolt V, Kain V, Plugge M, Mitchell M. Investigation of blended learning video resources to teach health students clinical skills: an integrative review. Nurse Educ Today 2018 Apr;63:101-107 [FREE Full text] [doi: 10.1016/j.nedt.2018.01.021] [Medline: 29425738]
8. Li P, Li X, Meng H, Huang L, Zhang L, Wang S, et al. Video-assisted health education promotes rehabilitation training of total knee arthroplasty patients and reduces stress and burnout in nurses compared to oral education. Biomed Res Int 2021;2021:5058899 [FREE Full text] [doi: 10.1155/2021/5058899] [Medline: 34722765]
9. Seymour JE, Almack K, Kennedy S, Froggatt K. Peer education for advance care planning: volunteers' perspectives on training and community engagement activities. Health Expect 2013 Mar;16(1):43-55 [FREE Full text] [doi: 10.1111/j.1369-7625.2011.00688.x] [Medline: 21615641]
10. Benè KL, Bergus G. When learners become teachers: a review of peer teaching in medical student education. Fam Med 2014;46(10):783-787 [FREE Full text] [Medline: 25646829]

11. Mbachu C, Dim C, Ezeoke U. Effects of peer health education on perception and practice of screening for cervical cancer among urban residential women in south-east Nigeria: a before and after study. BMC Womens Health 2017 Jun 09;17(1):41 [FREE Full text] [doi: 10.1186/s12905-017-0399-6] [Medline: 28599635]
12. Tang Y, Diao H, Jin F, Pu Y, Wang H. The effect of peer education based on adolescent health education on the resilience of children and adolescents: A cluster randomized controlled trial. PLoS One 2022;17(2):e0263012 [FREE Full text] [doi: 10.1371/journal.pone.0263012] [Medline: 35108312]
13. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
14. Frank L, Basch E, Selby JV, Patient-Centered Outcomes Research Institute. The PCORI perspective on patient-centered outcomes research. JAMA 2014 Oct 15;312(15):1513-1514. [doi: 10.1001/jama.2014.11100] [Medline: 25167382]
15. Yao Q, Li Y, Song L. Construction safety knowledge sharing on YouTube from 2007 to 2021: Two-step flow theory and semantic analysis. Safety Science 2022 Sep;153:105796. [doi: 10.1016/j.ssci.2022.105796]
16. Wright J. International Encyclopedia of the Social & Behavioral Sciences (second edition). Amsterdam: Elsevier; 2015.
17. Liu SH, Juster R, Dams-O'Connor K, Spicer J. Allostatic load scoring using item response theory. Compr Psychoneuroendocrinol 2021 Feb;5:100025 [FREE Full text] [doi: 10.1016/j.cpnec.2020.100025] [Medline: 35754455]
18. Fries JF, Bruce B, Cella D. The promise of PROMIS: using item response theory to improve assessment of patient-reported outcomes. Clin Exp Rheumatol 2005;23(5 Suppl 39):S53-S57. [Medline: 16273785]
19. Nguyen TH, Han H, Kim MT, Chan KS. An introduction to item response theory for patient-reported outcome measurement. Patient 2014 Jan 9;7(1):23-35 [FREE Full text] [doi: 10.1007/s40271-013-0041-0] [Medline: 24403095]
20. Rizopoulos D. ltm: an R Package for latent variable modeling and item response analysis. J Stat Softw 2007;17 [FREE Full text] [doi: 10.18637/jss.v017.i05]
21. Mationg MLS, Williams GM, Tallo VL, Olveda RM, Aung E, Alday P, et al. "The Magic Glasses Philippines": a cluster randomised controlled trial of a health education package for the prevention of intestinal worm infections in schoolchildren. Lancet Reg Health West Pac 2022 Jan;18:100312 [FREE Full text] [doi: 10.1016/j.lanwpc.2021.100312] [Medline: 35024651]
22. Hu P, Han L, Sharma M, Zeng H, Zhang Y, Li H, et al. Evaluation of cognitive and behavioral effects of peer education model-based intervention to sun safe in children. Iran J Public Health 2014 Mar;43(3):300-309 [FREE Full text] [Medline: 25988089]
23. Riley AW. Evidence that school-age children can self-report on their health. Ambul Pediatr 2004;4(4 Suppl):371-376. [doi: 10.1367/A03-178R.1] [Medline: 15264962]
24. Rebok G, Riley A, Forrest C, Starfield B, Green B, Robertson J, et al. Elementary school-aged children's reports of their health: a cognitive interviewing study. Qual Life Res 2001;10(1):59-70. [doi: 10.1023/a:1016693417166] [Medline: 11508476]
25. Xu L, Zhou C, Zhou J. Evaluation on effect of health promotion of school construction on health literacy of students in Changsha. Bull Dis Control Prev 2021;36(05):70-73. [doi: 10.13215/j.cnki.jbyfkztb.2107033]
26. Chen M, Wu C, Tsai C, Tsai M. Enterovirus board game for elementary school children: a pilot study. Public Health Nurs 2022 Mar;39(2):500-505. [doi: 10.1111/phn.12976] [Medline: 34537984]
27. Odone A, Bricchi L, Signorelli C. COVID-19 control school-based interventions: characteristics and impact of a national-level educational programme in Italy. Acta Biomed 2022 Jan 17;92(S6):e2021495 [FREE Full text] [doi: 10.23750/abm.v92iS6.12327] [Medline: 35037911]
28. Supriatun E, Nursasi AY, Fitriyani P. Enhancement of pulmonary tuberculosis prevention behavior with role play among elementary school students. Compr Child Adolesc Nurs 2017;40(sup1):78-87. [doi: 10.1080/24694193.2017.1386974] [Medline: 29166188]
29. Wang M, Han X, Fang H, Xu C, Lin X, Xia S, et al. Impact of health education on knowledge and behaviors toward infectious diseases among students in Gansu Province, China. Biomed Res Int 2018;2018:6397340 [FREE Full text] [doi: 10.1155/2018/6397340] [Medline: 29707573]
30. Taware S, Gawai P, Chatterjee A, Thakur H. Outcome of school-based intervention program in promoting personal hygiene in primary school children of Mumbai, India. Int Q Community Health Educ 2018 Oct;39(1):31-38. [doi: 10.1177/0272684X18809487] [Medline: 30376407]

Abbreviations

IRT: item response theory
TPM: three-parameter model

Edited by P Santana-Mancilla; submitted 31.10.22; peer-reviewed by W Ni, D Laborde, M Nasiri, Y Elsanousi; comments to author 28.07.23; revised version received 10.09.23; accepted 29.12.23; published 29.01.24.

Please cite as:
Zhang X, Wen YJ, Han N, Jiang Y
The Effect of a Video-Assisted Health Education Program Followed by Peer Education on the Health Literacy of COVID-19 and Other Infectious Diseases Among School Children: Quasi-Randomized Controlled Trial
JMIR Hum Factors 2024;11:e43943
URL: https://humanfactors.jmir.org/2024/1/e43943
doi:[10.2196/43943](https://doi.org/10.2196/43943)
PMID:[38285496](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=38285496&dopt=Abstract)

Original Paper

Barriers to and Facilitators of Key Stakeholders Influencing Successful Digital Implementation of Remote Monitoring Solutions: Mixed Methods Analysis

Fahad Mujtaba Iqbal[1], BSc, MBChB, PhD; Ravi Aggarwal[1], PhD; Meera Joshi[1], PhD; Dominic King[1], MBChB; Guy Martin[1], PhD; Sadia Khan[2], MBChB; Mike Wright[3], PhD; Hutan Ashrafian[1], PhD; Ara Darzi[1], FMedSci

[1]Division of Surgery, Imperial College London, London, United Kingdom

[2]West Middlesex University Hospital, London, United Kingdom

[3]Innovation Business Partner, Chelsea and Westminster NHS Trust, London, United Kingdom

Corresponding Author:
Fahad Mujtaba Iqbal, BSc, MBChB, PhD
Division of Surgery
Imperial College London
10th Floor QEQM Building
Praed Street
London, W2 1NY
United Kingdom
Phone: 44 020 3312 6666
Email: fahad.iqbal@doctors.org.uk

Abstract

Background: Implementation of remote monitoring solutions and digital alerting tools in health care has historically been challenging, despite the impetus provided by the COVID-19 pandemic. To date, a health systems–based approach to systematically describe barriers and facilitators across multiple domains has not been undertaken.

Objective: We aimed to undertake a comprehensive mixed methods analysis of barriers and facilitators for successful implementation of remote monitoring and digital alerting tools in complex health organizations.

Methods: A mixed methods approach using a modified Technology Acceptance Model questionnaire and semistructured interviews mapped to the validated fit among humans, organizations, and technology (HOT-fit) framework was undertaken. Likert frequency responses and deductive thematic analyses were performed.

Results: A total of 11 participants responded to the questionnaire and 18 participants to the interviews. Key barriers and facilitators could be mapped onto 6 dimensions, which incorporated aspects of digitization: system use (human), user satisfaction (human), environment (organization), structure (organization), information and service quality (technology), and system quality (technology).

Conclusions: The recommendations proposed can enhance the potential for future remote sensing solutions to be more successfully integrated in health care practice, resulting in more successful use of “virtual wards.”

Trial Registration: ClinicalTrials.gov NCT05321004; https://www.clinicaltrials.gov/study/NCT05321004

(JMIR Hum Factors 2024;11:e49769) doi:10.2196/49769

KEYWORDS

implementation science; health plan implementation; mobile health; health care industry; stakeholder; COVID-19; remote monitoring; digital tools; digital health; pandemic; virtual wards; virtual ward; health care delivery; telemedicine; telehealth; wearables; wearable; technology; United Kingdom; UK; digital services

Introduction

Since the onset of the COVID-19 pandemic, adoption and implementation of novel health care pathways have accelerated globally. A key change has been transitioning beyond the traditional face-to-face model of health care delivery with the incorporation of novel remote monitoring solutions [1,2]. They offer a significant advantage in moderating viral exposure risk to health care staff, reducing community spread, and delivering quality health care remotely for exposed or infected individuals [3,4].

The integration of telemedicine and remote monitoring into medical practice is expected to expand by appropriately permitting selected individuals to continue living at home rather than admitting them into secondary care; this very premise is the foundation of "virtual wards" [5]. With the recent improvements made to wearable technology, they can support health provider assessment and clinical decision-making through collected biometric data both in secondary care and in the community [6-10].

However, successful implementation of digital technologies across complex hospital systems is seldom a smooth process [11-13]. The absence of standardized procedures for implementation and evaluation alongside the deficiency of published implementation strategies adds to these difficulties. One study in the National Health Service (NHS) that implemented wearable sensors and alerting systems in secondary care reported no improvements in clinical outcomes among patients [14,15]. The aim was to use wearable sensors to provide continuous remote monitoring to patients admitted to acute (nonintensive) wards and alert health care staff upon recognition of deterioration. Interestingly, although the digital solution was able to pick up clinical deterioration in vital signs and alert health care staff, responding to the alert was met with significant delay. This was in spite of health care staff in the NHS reporting favorable perceptions of digital solutions with potential improvements to patient safety and reduced staff burden [16]. Therefore, there is a need to further explore implementation issues.

Patients have reported high levels of acceptance, comfort, and safety and deemed such digital tools favorable [17-19]. The main concerns, from a patient perspective, surround potential overreliance on numbers with diminishing contact from clinical staff [17,20,21]. Health care staff perceptions, however, have been more mixed, with concerns regarding changing and increasing workloads, uncertainty surrounding the clinical meaningfulness of captured data, and alert fatigue [19-21]. Although mixed methods exploration of these 2 key stakeholder groups has been well documented, understanding how to integrate remote monitoring digital tools in the NHS requires further examination of cultural and management issues in the health care organization, an area where evidence is missing.

In the United Kingdom, large health informatics programs and widespread digital transformations are delivered by NHS Digital, a nondepartmental public body [22,23]. To support digitization, NHS England has formed a framework consisting of 3 ambitions: digital readiness, maturity, and data-enabled services [24]. In line with this, NHS England has supported the development and use of virtual wards, further indicating the "digital push" [5]. For policy makers, understanding the barriers and facilitators as perceived by key organizational members is crucial for the effective provision and smooth deployment of digitally enabled care. A proposed framework evaluates these aspects, incorporating the concept of fit among humans, organizations, and technology (HOT-fit) [25]. This framework offers a structured basis to examine factors that focus on alignment and compatibility across these 3 domains, thereby enhancing the effectiveness of digital health care initiatives.

Therefore, the aim of this study was to evaluate key stakeholder perspectives on an organizational level of implementing remote monitoring solutions in the NHS, identifying factors that could affect successful execution and adoption using the HOT-fit framework. In doing so, we propose a road map for implementing wearable solutions in secondary care.

Methods

Study Design

A mixed methods approach was implemented that consisted of semistructured interviews and questionnaires [26]. This was developed in accordance with recommendations from the Standards for Reporting Qualitative Research (SRQR) guidelines where appropriate [27]. The semistructured interviews were conducted with high-level stakeholders from industry and academia, as well as with health care providers who played an instrumental role in and had prior experience of implementing digital solutions. Additionally, a validated questionnaire was used to ascertain the perceived technological acceptance of new remote monitoring systems.

To ensure appropriate recruitment among all key stakeholder groups, a key informant strategy was followed for purposive recruitment [28,29]. Individuals were identified through their notable work with implementation of remote monitoring solutions in health care, including authors of impactful research in the literature, major digital technology companies, technicians involved with digital tool infrastructure development, and experts recommended by peers. This represented a variety of groups, including academics, clinicians, allied health care professionals, and employees of Google Health, who had experience with implementing digital solutions with the NHS.

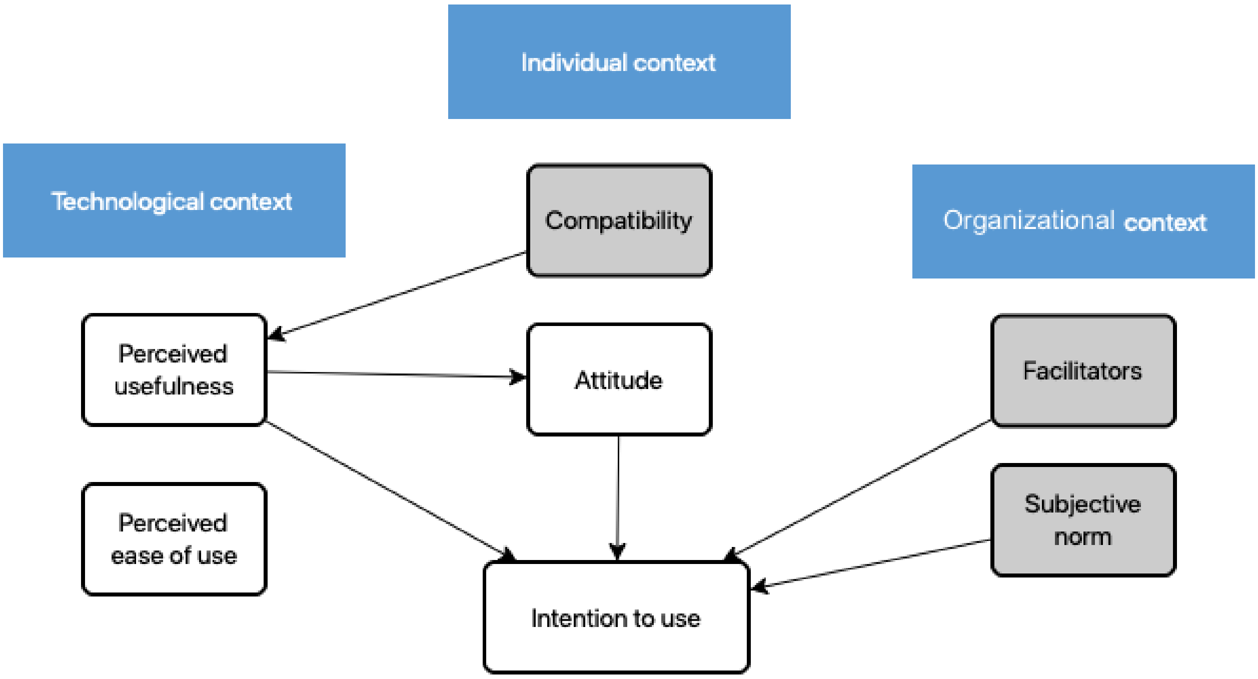
Ethical Considerations

All recruited participants provided written informed consent. Ethical approval for this study was obtained by Imperial College London's Science Engineering Technology Research Ethics Committee (20IC6331), and it was conducted in accordance with the Good Clinical Practice guidelines and the Declaration of Helsinki. Storage and handling of personal data complied with the General Data Protection Regulation. Interviews were recorded, anonymized, and transcribed.

Questionnaires

An adapted version of the Technology Acceptance Model (TAM) questionnaire was used; this validated questionnaire has shown acceptably high Cronbach α values [30]. This ensures the reliability of our findings, contributing to the robustness of the study's methodology and its implications in understanding technology acceptance dynamics. The proposed theoretical framework (information technology acceptance) is shown in Figure 1. It has been adapted from Chau and Hu [31], comprising individual context, technological context, and organizational context. Further adaptations from Gagnon et al [30], with the inclusion of theories of interpersonal behavior and reasoned action building on the TAM, proposed by Davis [32], have been included [30-34]. As such, individual context consists of compatibility (factors that affect acceptance of a new technology) and attitude (perception of the individual to adopting a technology); technological context consists of perceived usefulness and perceived ease of use of technologies. Lastly, organizational context consists of facilitators and subjective norms; the latter can be described as social (an individual's perception of a behavior) or descriptive (behavior of others).

Figure 1. Theoretical framework for the modified Technology Acceptance Model questionnaire [30].



Semistructured Interviews

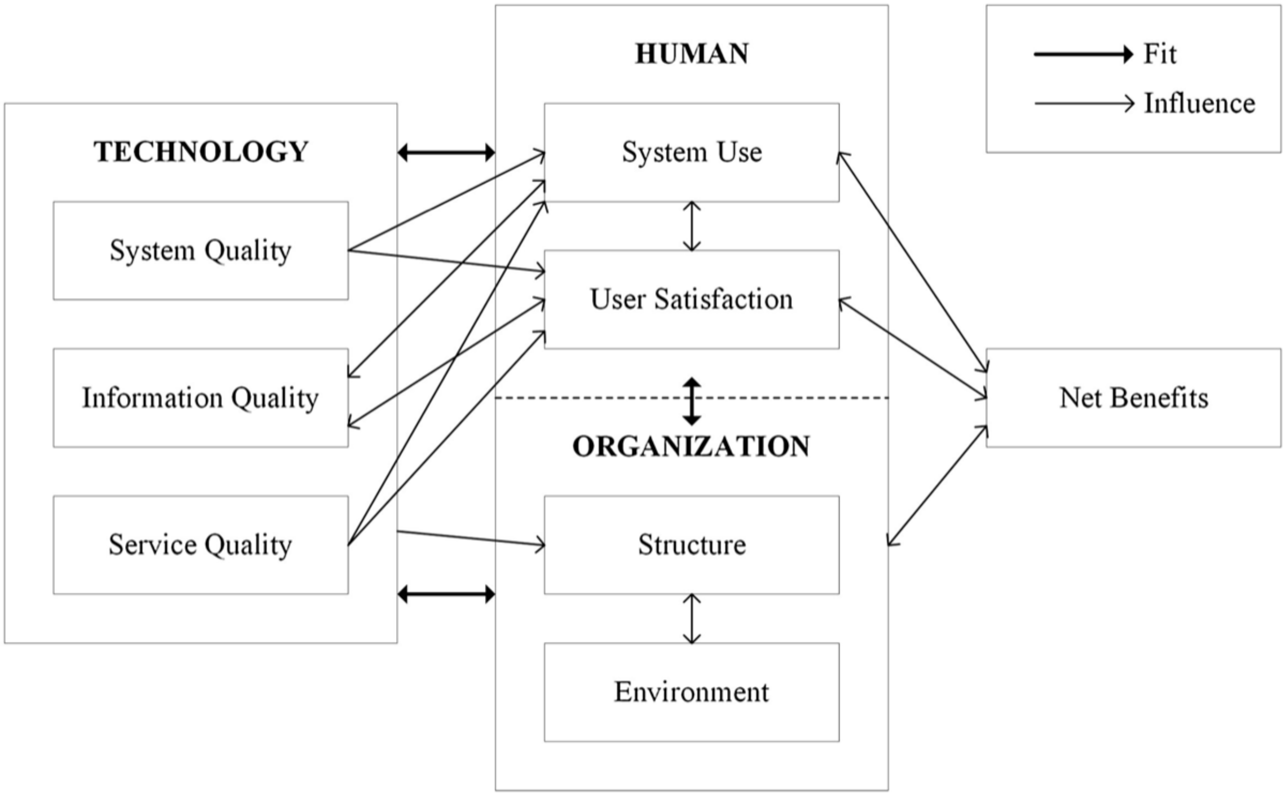
All participants were invited to take part in semistructured interviews conducted by the lead researchers. A structured topic guide was created following a literature review that drew heavily from a model proposed by Simblett et al [35] and by the HOT-fit framework [25].

Data collection was an iterative process; emerging recurring concepts were incorporated into the interview guide for further exploration with remaining participants. Interviews were recorded, anonymized, and transcribed verbatim before being entered into NVivo (version 12; QSR International) for analysis.

HOT-Fit Framework

This validated framework identifies dimensions that can be mapped onto and used as reference models for evaluating the performance, effectiveness, and impact of health systems [25,36]. A fit between human, organizational, and technological factors is required to ensure successful implementation and has been highlighted in Figure 2.

Figure 2. The fit among humans, organizations, and technology (HOT-fit) framework, adapted from Yusof et al [25].



Data Analysis

Frequency distributions were generated for the 7-point Likert scale responses to the modified TAM questionnaire using R studio (R Foundation for Statistical Computing) with the *Likert* package (Bryer and Speerschneider).

Transcribed interviews were analyzed using a broadly deductive approach [37], with the topic guide adapted as previously described [35]. This formed the basis for the initial predefined coding framework and was undertaken by 2 independent researchers to determine barriers and facilitators [37]. An iterative process of coding and data indexing occurred, ensuring key aspects were not missed from the predefined coding framework. Subsequent emerging themes were summarized and mapped to the evaluation measures corresponding to each dimension of the HOT-fit framework [25]. The results were discussed until consensus was reached.

Results

Overviews of the included participants and the reported evaluation measures are shown in Tables 1 and 2, respectively.

Table 1. Demographics of included participants.

Group	Role 1	Role 2	Role 3	Role 4	Role 5	Role 6	Role 7
Health care trusts	Director of strategy, research and innovation	Chief clinical information officer and Caldicott Guardian	Digital quality improvement lead	Project manager	Chief information officer	Systems, integration interoperability architect	Lead nurse for remote monitoring
Academics	Clinical lecturer	Clinical lecturer	Chief scientific advisor	—[a]	—	—	—
Google Health	Clinical lead	Clinical specialist	Product manager	Implementation specialist	Implementation manager	Program manager	—
Other	Programme director: innovation of health	Managing director: digital health	—	—	—	—	—

[a]—Not applicable.

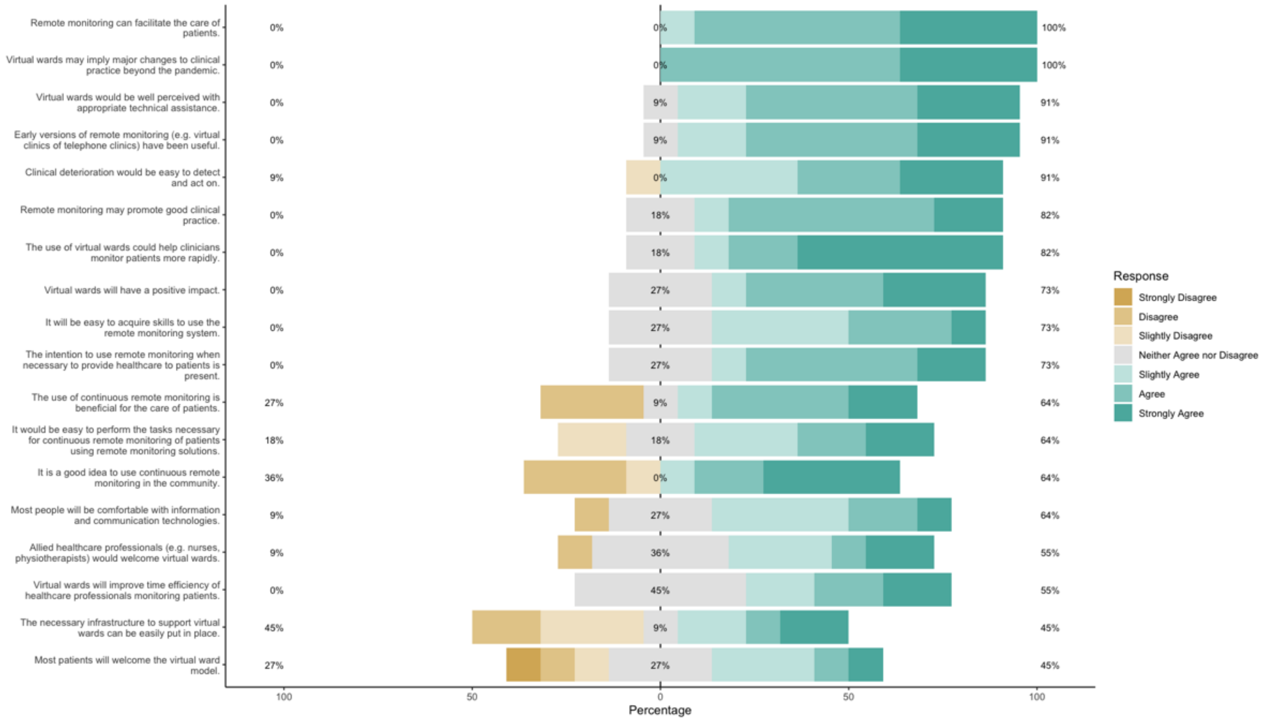
Table 2. Overview of reported evaluation measures.

Dimension and evaluation measures	Factors
System use	
Expectation and beliefs	• Improved efficiency (facilitator) • Appropriate selection of end users suitable for digital tool (facilitator)
Training, knowledge, and expertise	• Lack of troubleshooting support (barrier) • Engagement with new starters (facilitator)
Motivation	• Large data burden (barrier) • Post–COVID-19 fatigue of staff (barrier) • Finding local champions (facilitator)
User satisfaction (no evaluation measures)	• Developing relationships for feedback (facilitator) • Previous negative experiences with no feedback on benefit (barrier)
Environment (no evaluation measures)	• Overburdened National Health Service system (barrier)
Structure (clinical process)	• Clear strategic framework and partnership (facilitator)
Information and service quality (no evaluation measures)	• Poor interoperability (barrier) • Poor user interface and user engagement (barrier)
System quality	• Failure to provide added value (barrier)

TAM Questionnaire

A total of 11 participants (response rate 11/22, 50%) responded to the questionnaire; the responses are represented as a Likert plot (Figure 3). Overall, the technology surrounding remote monitoring and virtual wards was perceived well by the questioned stakeholders, who considered that it facilitated the care of patients and that these pathways, initially introduced during the pandemic, were likely to change long-term provision of health care. However, some concerns were noted regarding whether the existing infrastructure could support the technology's use and whether it would improve efficiency. Of note, there was uncertainty regarding whether most patients would welcome virtual wards or remote monitoring.

Figure 3. Likert plot displaying responses to the modified Technology Acceptance Model questionnaire. The percentages on the left and right sides of the plot represent the totals for negative and positive responses, while the percentages in the center represent neutral responses.



Semistructured Interviews

A total of 22 participants were approached, of whom 18 (response rate: 82%) participated in the semistructured interviews (Table 1). An overview of the factors, by dimension, that respondents felt were responsible for contributing to implementation is summarized in Table 2.

System Use

Expectations and Beliefs

The prospect of introducing novel remote monitoring technologies was felt to facilitate implementation through improved efficiency, particularly since the implementation of electronic health records has improved data availability and clarity:

> *...with the implementation [and] introduction of electronic health records where the data that's available is so granular. And in addition to new technologies that are coming. I think that you can do a lot more, remotely or virtually, and it does make things a lot more efficient...* [Participant 15]

Moreover, respondents also commented that for successful implementation, a selective process should be in place for patients who would benefit the most from novel technologies, rather than using the technology in cases that would not be meaningful:

> *From a patient perspective, we don't want to one size fits all approach. We need to be clear about how we personalize this and how it's relevant and meaningful.* [Participant 17]

Training, Knowledge, and Expertise

Problems with troubleshooting and available training were reported to reduce successful implementation due to a lack of support:

> *We've had problems when trying to use the remote monitoring, it came up with an error and then I have to try and sort that out, you know? It's just things like that that make extra work.* [Participant 16]

> *I know that the nurses have struggled a huge amount with remote monitoring, and I expected that...because there's a lot of upskilling.* [Participant 18]

However, engaging early with health care workers and obtaining their involvement was shown to improve implementation of remote monitoring solutions:

> *[We received] better engagement by tying the implementation with the new starters in the role and the changeover of junior doctors, because it was a new product to offer to new junior doctors.* [Participant 3]

Motivation

It was felt that motivation to engage with technologies would be impacted through the excessive availability of data acting as a deterrent:

> *We need to be mindful about the data burdens, not just for patients but for staff because this kind of remote technology follows you around. You basically could work 24/7 365 of the year.* [Participant 17]

In addition, following the pandemic, many health care workers were fatigued and unmotivated to engage in change, acting as a barrier to successful remote technology implementation:

> *Post-COVID the workforce has been decimated, been exhausted and is fatigued. It's not the only problem though, because you know as well as I do that the NHS has run this model of where it's good will. We've never had infrastructure that we needed to do stuff and we still get a huge amount done. So it's not the only driver at the moment. It's more noticeable because of where people's heads are at and obviously where their physical levels and mental levels of exhaustion are...* [Participant 17]

However, respondents also noted that finding a few motivated individuals to champion change at a local level can help implementation:

> *I asked them to self-nominate three of them who were interested in helping [implement]. So they led and supported the [technology]...* [Participant 18]

User Satisfaction

Respondents reported that previous experience with digital tools tied into user satisfaction. Feedback to end users demonstrating meaningful impact was deemed important for engagement and successful implementation:

> *Where staff or patients, for example, have been involved in projects before that they haven't had any feedback from, haven't seen any meaningful outcome from...they're like, well, why would I want to get engaged with this? That's a lot of energy and effort from me and I won't see any benefit.* [Participant 17]

> *...develop relationships, so between, if you like, supplier and developer and clinical staff so you've got these rapid cycles of feedback and learning.* [Participant 1]

Environment

Respondents reported that previous hindrance of effective implementation was because of an overburdened system unable to give the appropriate attention to integrating a digital solution in the NHS:

> *NHS is overburdened and so that level of diligence...wasn't there until it had to be, until things became mission critical...that comes down to a bandwidth problem...* [Participant 10]

Similarly, underresourcing was noted to be a barrier, particularly during the early stages, where issues would arise:

> *More resource[s] to get [things] kick started [are usually needed]...because we had to go through all the teething problems ourselves which created extra work for us.* [Participant 16]

> *We've got very limited resources, that they're very thinly spread across all of the IT projects that require integration and interoperability...just the sheer volume*

> *of work that the Trust has heaped on us over the last three or four years is the bigger constraining factor.* [Participant 4]

Lastly, organizational culture supporting digitization was a commonly reported theme, with some institutions more readily accepting of innovation than others:

> *Organisational culture can be both the barrier and facilitator. We know that there are some organisations that are much more ready and able to adopt innovation. I think from an organisational perspective, competing priorities are a huge issue...If your IT is majorly engaged in doing something else, for example an EHR implementation, its ability to support remote monitoring and other technologies is really poor.* [Participant 17]

Structure

Respondents also commented on the need for a clear process and said that developing a strategic partnership and framework would facilitate implementation and should be planned before rollout:

> *Strategic framework is crucial on things.... What does a strategic partnership look like? What is the direction that we want to jointly head in? What do we want to achieve together, and what are the different components to get there...* [Participant 1]

> *Making [the product vision and roadmap] clear as early on and getting that input right at the beginning of any kind of feature development. So that there is expectation alignment on what is being developed whether the minimal viable product meets the use cases that it needs to, and that there's a partnership in prioritizing these features and when they're delivered. As opposed to just showing a feature set a few weeks before it gets deployed.... I think that initial understanding of the vision...and getting that clinical engagement as early on helps to set the path going forward.* [Participant 12]

Information and Service Quality

Respondents noted the need for digital tools to be interoperable and usable, as poorly designed digital tools would be a barrier, hindering an overly strained NHS system:

> *The challenges are IT and interoperability…you don't want 20 bits of data…from 20 apps that don't work, so that's the usability and the accessibility and the staffing of these models because traditionally they basically get added onto someone's day job. But that day person's already overwhelmed.* [Participant 17]

System Quality

Respondents highlighted that for a digital tool to be successfully implemented, it needed to provide added value, with perceived usefulness and ease of use being crucial.

> *[What] was the added value in [this digital app]? All it did was render some of the information that we already had in a limited manner, back in the mobile device.* [Participant 8]

> *Usability, the accessibility, and the staffing of these models [are really poor] because traditionally they basically get added onto someone's day job.... The data element [is also] really poor, so you get a lot of enthusiasts doing a lot of projects. But if you then say where's your evidence that makes any difference to anything meaningful that matters to patients and staff, they can't produce that. I think the digital health tech industry has been really slow at that.* [Participant 17]

Furthermore, it was believed that the best way to implement a digital tool (eg, remote monitoring solutions) was through rapid quality improvement cycles following the plan-do-study-act (PDSA) technique, focusing on targeting user experience issues:

> *...believe the technology suffered from very poor clinical and user engagement. So I know [technological companies] will tell us they've had loads of user engagement, but actually most patients wouldn't say that, they'd say well, why is it like this? No, why is nobody been engaged in the design for this?* [Participant 17]

> *...trying to give clinical input into feasibility, usability, implementation in terms of the design of how we were going to implement stuff, so...[a] genuine PDSA type approach to implementation, and I was quite involved in some of the thinking about spread and how do you get this utilized across different parts of the Trust...* [Participant 1]

Discussion

Principal Findings

This study explored barriers and facilitators for implementing digital tools, in particular remote monitoring solutions, in the NHS, alongside the acceptance of such technology using the modified TAM questionnaire. Using the HOT-fit framework, human, organization, and technological factors were categorized, allowing for a multiple-angled approach to a multifaceted problem. Therefore, key barriers and facilitators could be mapped onto 6 dimensions, which incorporated aspects of digitization: system use (human), user satisfaction (human), environment (organization), structure (organization), information and service quality (technology), and system quality (technology).

With regards to system use, the importance of improving workflow efficiency, having appropriate troubleshooting support available for staff, finding local champions to help integration within the clinical workforce, and positively engaging with health care staff were highlighted as facilitators. To support this, young staff have been deemed the most likely to engage with and benefit from a new workflow [36,38-40]. This, in part, may be explained by more adept digital literacy skills and technical proficiencies associated with junior members [41]. In the literature, concise and tailored education surrounding implementation has been promoted as an important facilitator [42].

Key barriers relating to system use and environment included poor training and the burden of data, particularly with continuous remote monitoring of vital signs. These data may not always be clinically meaningful or because of poor resourcing may not be acknowledged appropriately, generating additional work for existing staff, who are already overburdened [14,43]. Previously, this unincentivized workflow change led to poor response times to alerts generated through alerting systems in an acute surgical ward [14]. In this study, 36% (4/11) of respondents to the modified TAM questionnaire were unsure whether allied health care professionals would welcome virtual wards (Figure 3). One study highlighted that these workers, in particular nurses and clinicians, were the most important gatekeepers for remote monitoring solutions [44]. Therefore, engaging these groups, fostering positive relationships, and delivering regular feedback would enhance user satisfaction, allow user interface and engagement issues to be proactively tackled, and subsequently enable successful implementation.

Concerning system quality, perceived usefulness and ease of use were deemed as important facilitators for successful implementation. In the literature, intuitive and user-friendly systems have been confirmed to have easier acceptance [36,39]. The modified TAM questionnaire similarly confirmed this in our cohort, particularly through questions concerning acquiring new skills and impact, emphasizing that remote monitoring technology could be readily accepted.

Limitations

This study included key stakeholders belonging to a broad selection of groups (academics, industry, and health care) in order to create a broad understanding of factors that influence implementation of remote monitoring solutions in the NHS. Given that previous studies have focused on end user testing, this study sought to provide a top-down view to give a better understanding of considerations that could influence widespread implementation [16,18]. However, in doing so, our interpretations have some limitations. First, the broad, heterogeneous sample of key stakeholders included may identify issues that are generalizable, but the nonprobabilistic sampling may have resulted in a selection bias. Moreover, the included sample size was limited. Despite this, the use of semistructured interviews yielded pertinent considerations for pragmatic implementation in hospital settings. In addition, differences between various hospitals and departments, which may have different attitudes toward digital technologies, were not explored in this study. A final limitation relates to the HOT-fit framework; although it is considered useful, the mapping of factors is a subjective undertaking and mapping to one specific measure was, at times, difficult.

Further Research and Recommendations

Although our cohort showed that there was overall acceptance of remote monitoring technology (Figure 3), there remains a deficiency with respect to successful implementation. This was noted most recently in one study where the median time to acknowledge an alert from health care staff was 111 (range 1-2146) minutes, despite early recognition of deterioration from remote sensing [14]. Therefore, further research should incorporate human factors and behavior evaluation when implementing remote monitoring solutions with the NHS; moreover, implementation frameworks such as HOT-fit should be used to ensure multiple angles have been carefully considered.

To facilitate the effective integration of remote monitoring solutions in clinical workflows, a comprehensive strategic framework is paramount. This framework should prioritize the early involvement of end users, fostering relationships that enable rapid feedback on implementation strategies, user interfaces, and user experience issues. Such engagement allows for iterative enhancements through PDSA cycles, promoting continuous improvement [45].

Industries aiming to develop remote monitoring technologies must collaborate closely with key stakeholders, ensuring the creation of products that provide significant value and feature user-friendly interfaces. This approach emphasizes the importance of a bottom-up strategy in technology implementation, valuing the autonomy and insights of end users, who play a crucial role in the successful adoption of these solutions. Crucial to this process is the establishment of robust infrastructural support prior to the deployment of remote monitoring systems. Adequate resourcing and the involvement of technical support staff are essential to facilitate seamless integration with existing information technology frameworks, thereby enhancing the prioritization and effectiveness of digital health initiatives. By adhering to these guidelines, health care organizations can enhance the integration of remote monitoring into clinical practice, leading to improved operational efficiency, patient care, and overall health care service delivery.

Conclusion

Implementation of remote monitoring solutions in the NHS remains a complex challenge. The results of this study have highlighted key stakeholder perceptions that could influence successful integration. Through the proposed recommendations, there is potential for future remote sensing solutions to be more successfully integrated into our health care practices, resulting in novel pathways expanding beyond virtual wards.

Acknowledgments

Infrastructure support for this research was provided by the National Institute for Health and Care Research (NIHR) Imperial Biomedical Research Centre and the NIHR Imperial Patient Safety Translational Research Centre.

Data Availability

The data sets used and/or analyzed during the study are available from the corresponding author on reasonable request.

Authors' Contributions

FMI drafted the manuscript. Significant amendments were made by RA, MJ, MW, SK, HA, and AD. All authors approved the final manuscript.

Conflicts of Interest

AD is chair of the Health Security initiative and HA is chief scientific officer at Flagship Pioneering UK Ltd. Flagship Pioneering had no role in the development, conduct, or analysis of the study.

References

1. Watson AR, Wah R, Thamman R. The value of remote monitoring for the COVID-19 pandemic. Telemed J E Health 2020 Sep;26(9):1110-1112. [doi: 10.1089/tmj.2020.0134] [Medline: 32384251]
2. The impact of the COVID-19 pandemic on outpatient visits: changing patterns of care in the newest COVID-19 hot spots. The Commonwealth Fund. URL: https://www.commonwealthfund.org/publications/2020/aug/impact-covid-19-pandemic-outpatient-visits-changing-patterns-care-newest [accessed 2020-09-19]
3. Xu S, Li Y. Beware of the second wave of COVID-19. Lancet 2020 Apr 25;395(10233):1321-1322 [FREE Full text] [doi: 10.1016/S0140-6736(20)30845-X] [Medline: 32277876]
4. WHO coronavirus disease (COVID-19) dashboard. World Health Organization. URL: https://covid19.who.int/ [accessed 2020-06-25]
5. Supporting information: virtual ward including hospital at home. NHS England. URL: https://www.england.nhs.uk/publication/supporting-information-virtual-ward-including-hospital-at-home/ [accessed 2022-06-07]
6. Martirosyan M, Caliskan K, Theuns D, Szili-Torok T. Remote monitoring of heart failure: benefits for therapeutic decision making. Expert Rev Cardiovasc Ther 2017 Jul;15(7):503-515. [doi: 10.1080/14779072.2017.1348229] [Medline: 28656875]
7. Annis T, Pleasants S, Hultman G, Lindemann E, Thompson JA, Billecke S, et al. Rapid implementation of a COVID-19 remote patient monitoring program. J Am Med Inform Assoc 2020 Aug 01;27(8):1326-1330 [FREE Full text] [doi: 10.1093/jamia/ocaa097] [Medline: 32392280]
8. Ford D, Harvey JB, McElligott J, King K, Simpson KN, Valenta S, et al. Leveraging health system telehealth and informatics infrastructure to create a continuum of services for COVID-19 screening, testing, and treatment. J Am Med Inform Assoc 2020 Dec 09;27(12):1871-1877 [FREE Full text] [doi: 10.1093/jamia/ocaa157] [Medline: 32602884]
9. O'Carroll O, MacCann R, O'Reilly A, Dunican EM, Feeney ER, Ryan S, et al. Remote monitoring of oxygen saturation in individuals with COVID-19 pneumonia. Eur Respir J 2020 Aug;56(2):2001492 [FREE Full text] [doi: 10.1183/13993003.01492-2020] [Medline: 32616588]
10. Medina M, Babiuch C, Card M, Gavrilescu R, Zafirau W, Boose E, et al. Home monitoring for COVID-19. Cleve Clin J Med 2020 Jun 11:1-4 [FREE Full text] [doi: 10.3949/ccjm.87a.ccc028] [Medline: 32409432]
11. Wachter R. Making IT work: harnessing the power of health information technology to improve care in England. Gov.UK. URL: https://assets.publishing.service.gov.uk/media/5a8091afe5274a2e87dba8f2/Wachter_Review_Accessible.pdf [accessed 2023-12-02]
12. Greenhalgh T, Swinglehurst D, Stones R. Rethinking resistance to 'big IT': a sociological study of why and when healthcare staff do not use nationally mandated information and communication technologies. Health Soc Care Deliv Res 2014 Nov;2(39):1-112. [doi: 10.3310/hsdr02390] [Medline: 27466649]
13. Kruse C, Betancourt J, Ortiz S, Valdes Luna SM, Bamrah IK, Segovia N. Barriers to the use of mobile health in improving health outcomes in developing countries: systematic review. J Med Internet Res 2019 Oct 09;21(10):e13263 [FREE Full text] [doi: 10.2196/13263] [Medline: 31593543]
14. Iqbal F, Joshi M, Fox R, Koutsoukou T, Sharma A, Wright M, et al. Outcomes of vital sign monitoring of an acute surgical cohort with wearable sensors and digital alerting systems: a pragmatically designed cohort study and propensity-matched analysis. Front Bioeng Biotechnol 2022;10:895973 [FREE Full text] [doi: 10.3389/fbioe.2022.895973] [Medline: 35832414]
15. Iqbal FM, Joshi M, Khan S, Ashrafian H, Darzi A. Implementation of wearable sensors and digital alerting systems in secondary care: protocol for a real-world prospective study evaluating clinical outcomes. JMIR Res Protoc 2021 May 04;10(5):e26240 [FREE Full text] [doi: 10.2196/26240] [Medline: 33944790]
16. Joshi M, Archer S, Morbi A, Ashrafian H, Arora S, Khan S, et al. Perceptions on the use of wearable sensors and continuous monitoring in surgical patients: interview study among surgical staff. JMIR Form Res 2022 Feb 11;6(2):e27866 [FREE Full text] [doi: 10.2196/27866] [Medline: 35147503]
17. Downey C, Brown J, Jayne D, Randell R. Patient attitudes towards remote continuous vital signs monitoring on general surgery wards: an interview study. Int J Med Inform 2018 Jun;114:52-56. [doi: 10.1016/j.ijmedinf.2018.03.014] [Medline: 29673603]
18. Joshi M, Archer S, Morbi A, Arora S, Kwasnicki R, Ashrafian H, et al. Short-term wearable sensors for in-hospital medical and surgical patients: mixed methods analysis of patient perspectives. JMIR Perioper Med 2021 Apr 22;4(1):e18836 [FREE Full text] [doi: 10.2196/18836] [Medline: 33885367]

19. Leenen JPL, Dijkman EM, van Dijk JD, van Westreenen HL, Kalkman C, Schoonhoven L, et al. Feasibility of continuous monitoring of vital signs in surgical patients on a general ward: an observational cohort study. BMJ Open 2021 Feb 17;11(2):e042735 [FREE Full text] [doi: 10.1136/bmjopen-2020-042735] [Medline: 33597138]
20. Weenk M, van Goor H, Frietman B, Engelen LJ, van Laarhoven CJ, Smit J, et al. Continuous monitoring of vital signs using wearable devices on the general ward: pilot study. JMIR Mhealth Uhealth 2017 Jul 05;5(7):e91 [FREE Full text] [doi: 10.2196/mhealth.7208] [Medline: 28679490]
21. Areia C, King E, Ede J, Young L, Tarassenko L, Watkinson P, et al. Experiences of current vital signs monitoring practices and views of wearable monitoring: A qualitative study in patients and nurses. J Adv Nurs 2022 Mar;78(3):810-822 [FREE Full text] [doi: 10.1111/jan.15055] [Medline: 34655093]
22. About NHS Digital. National Health Service. URL: https://digital.nhs.uk/about-nhs-digital [accessed 2020-09-19]
23. NHS Digital annual report and accounts 2019 to 2020. National Health Service. URL: https://tinyurl.com/2pcn7y6b [accessed 2020-09-19]
24. Digital framework for allied health professionals. NHS England. 2019. URL: https://www.england.nhs.uk/wp-content/uploads/2019/04/a-digital-framework-for-allied-health-professionals.pdf [accessed 2023-12-02]
25. Yusof M, Kuljis J, Papazafeiropoulou A, Stergioulas L. An evaluation framework for health information systems: human, organization and technology-fit factors (HOT-fit). Int J Med Inform 2008 Jun;77(6):386-398. [doi: 10.1016/j.ijmedinf.2007.08.011] [Medline: 17964851]
26. Iqbal FM, Joshi M, Khan S, Wright M, Ashrafian H, Darzi A. Key stakeholder barriers and facilitators to implementing remote monitoring technologies: protocol for a mixed methods analysis. JMIR Res Protoc 2022 Jul 21;11(7):e38437 [FREE Full text] [doi: 10.2196/38437] [Medline: 35862185]
27. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
28. Marshall MN. The key informant technique. Fam Pract 1996 Feb;13(1):92-97. [doi: 10.1093/fampra/13.1.92] [Medline: 8671109]
29. Campbell S, Greenwood M, Prior S, Shearer T, Walkem K, Young S, et al. Purposive sampling: complex or simple? Research case examples. J Res Nurs 2020 Dec;25(8):652-661 [FREE Full text] [doi: 10.1177/1744987120927206] [Medline: 34394687]
30. Gagnon MP, Orruño E, Asua J, Abdeljelil AB, Emparanza J. Using a modified technology acceptance model to evaluate healthcare professionals' adoption of a new telemonitoring system. Telemed J E Health 2012;18(1):54-59 [FREE Full text] [doi: 10.1089/tmj.2011.0066] [Medline: 22082108]
31. Chau P, Hu P. Examining a model of information technology acceptance by individual professionals: an exploratory study. J Manag Inf Syst 2014 Dec 23;18(4):191-229. [doi: 10.1080/07421222.2002.11045699]
32. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
33. Hill RJ, Fishbein M, Ajzen I. Belief, attitude, intention and behavior: an introduction to theory and research. Contemp Sociol 1977 Mar;6(2):244. [doi: 10.2307/2065853]
34. Triandis HC. Nebraska Symposium on Motivation, 1979: Beliefs, Attitudes and Values. Lincoln, NB: University of Nebraska Press; 1980:195-259.
35. Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrão J, et al. Barriers to and facilitators of engagement with remote measurement technology for managing health: systematic review and content analysis of findings. J Med Internet Res 2018 Jul 12;20(7):e10480 [FREE Full text] [doi: 10.2196/10480] [Medline: 30001997]
36. Kilsdonk E, Peute LW, Jaspers MWM. Factors influencing implementation success of guideline-based clinical decision support systems: A systematic review and gaps analysis. Int J Med Inform 2017 Feb;98:56-64. [doi: 10.1016/j.ijmedinf.2016.12.001] [Medline: 28034413]
37. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
38. Lugtenberg M, Pasveer D, van der Weijden T, Westert GP, Kool RB. Exposure to and experiences with a computerized decision support intervention in primary care: results from a process evaluation. BMC Fam Pract 2015 Oct 16;16(1):141 [FREE Full text] [doi: 10.1186/s12875-015-0364-0] [Medline: 26474603]
39. Varonen H, Kortteisto T, Kaila M, EBMeDS Study Group. What may help or hinder the implementation of computerized decision support systems (CDSSs): a focus group study with physicians. Fam Pract 2008 Jun;25(3):162-167. [doi: 10.1093/fampra/cmn020] [Medline: 18504253]
40. Zheng K, Padman R, Johnson MP, Diamond HS. Understanding technology adoption in clinical care: clinician adoption behavior of a point-of-care reminder system. Int J Med Inform 2005 Aug;74(7-8):535-543. [doi: 10.1016/j.ijmedinf.2005.03.007] [Medline: 16043083]
41. Mobasheri MH, King D, Johnston M, Gautama S, Purkayastha S, Darzi A. The ownership and clinical use of smartphones by doctors and nurses in the UK: a multicentre survey study. BMJ Innov 2015 Oct 07;1(4):174-181. [doi: 10.1136/BMJINNOV-2015-000062]

42. Trivedi MH, Daly EJ, Kern JK, Grannemann BD, Sunderajan P, Claassen CA. Barriers to implementation of a computerized decision support system for depression: an observational report on lessons learned in "real world" clinical settings. BMC Med Inform Decis Mak 2009 Jan 21;9:6 [FREE Full text] [doi: 10.1186/1472-6947-9-6] [Medline: 19159458]
43. Downey C, Randell R, Brown J, Jayne D. Continuous versus intermittent vital signs monitoring using a wearable, wireless patch in patients admitted to surgical wards: pilot cluster randomized controlled trial. J Med Internet Res 2018 Dec 11;20(12):e10802 [FREE Full text] [doi: 10.2196/10802] [Medline: 30538086]
44. Whitten PS, Mackert MS. Addressing telehealth's foremost barrier: provider as initial gatekeeper. Int J Technol Assess Health Care 2005;21(4):517-521. [doi: 10.1017/S0266462305050725] [Medline: 16262977]
45. Leis JA, Shojania KG. A primer on PDSA: executing plan-do-study-act cycles in practice, not just in name. BMJ Qual Saf 2017 Jul;26(7):572-577. [doi: 10.1136/bmjqs-2016-006245] [Medline: 27986900]

Abbreviations

HOT-fit: fit among humans, organizations, and technology
NHS: National Health Service
PDSA: plan-do-study-act
SRQR: Standards for Reporting Qualitative Research
TAM: Technology Acceptance Model

Edited by A Kushniruk; submitted 08.06.23; peer-reviewed by A Acharya, C Sakal; comments to author 19.01.24; revised version received 26.01.24; accepted 07.04.24; published 06.05.24.

Please cite as:
Iqbal FM, Aggarwal R, Joshi M, King D, Martin G, Khan S, Wright M, Ashrafian H, Darzi A
Barriers to and Facilitators of Key Stakeholders Influencing Successful Digital Implementation of Remote Monitoring Solutions: Mixed Methods Analysis
JMIR Hum Factors 2024;11:e49769
URL: https://humanfactors.jmir.org/2024/1/e49769
doi:10.2196/49769
PMID:37338929

Original Paper

Influence of TikTok on Body Satisfaction Among Generation Z in Indonesia: Mixed Methods Approach

Hanifa Ariana[1*], SKom; Ikmal Almuhtadi[1*], SKom; Nikita Jacey Natania[1*], SKom; Putu Wuri Handayani[1*], Prof Dr; Stéphane Bressan[2*], PhD; Pramitha Dwi Larasati[1*], MKom

[1]Faculty of Computer Science Universitas Indonesia, Depok, Indonesia

[2]School of Computing National University of Singapore, Singapore, Singapore

[*]all authors contributed equally

Corresponding Author:
Putu Wuri Handayani, Prof Dr
Faculty of Computer Science Universitas Indonesia
Kampus UI Street
Depok, 16424
Indonesia
Phone: 62 217863410
Email: Putu.wuri@cs.ui.ac.id

Abstract

Background: As social media platforms gain popularity, their usage is increasingly associated with cyberbullying and body shaming, causing devastating effects.

Objective: This study aims to investigate the impact of social media on Generation Z users' body image satisfaction. More specifically, it examines the impact of TikTok on body image satisfaction among TikTok users aged between 17 years and 26 years in Indonesia.

Methods: The methodology used mixed-method approaches. Quantitative data were obtained from 507 responses to a questionnaire and analyzed using covariance-based structural equation modeling. Qualitative data were obtained from the interviews of 32 respondents and analyzed through content analysis.

Results: This study reveals that upward appearance comparison is influenced by video-based activity and appearance motivation. Conversely, thin-ideal internalization is influenced by appearance motivation and social media literacy. Upward appearance comparisons and thin-ideal internalization comparisons detrimentally impact users' body image satisfaction.

Conclusions: The results of this study are expected to provide valuable insights for social media providers, regulators, and educators in their endeavors to establish a positive and healthy social media environment for users.

(JMIR Hum Factors 2024;11:e58371) doi:10.2196/58371

KEYWORDS

body satisfaction; social media; TikTok; Indonesia; cyber-bullying; cyberbullying; cyberbully; cyber-harassment; bullying; harassment; body shaming; objectify; objectifying; social media; social media use; social media usage; socials; social network; social networks; social networking; Tik Tok; GenZ; Gen-Z; youth; adolescent; adolescents; teen; teens; teenager; teenagers; young-adult; young-adults

Introduction

Background

In 2022, the estimated number of social media users was 4.80 billion, with an expected increase to 7 billion by the decade's end [1]. Concomitantly, the use of social media appears to be increasingly linked to cyberbullying, especially in the context of body shaming—the act of criticizing and stigmatizing someone's physical appearance [2]. Noteworthy cases involving public figures such as Kylie Jenner and Adele [3,4] underscore the fact that nobody is exempt from such incidents.

Individual perceptions of one's body can range from positive to negative [5], with positive perceptions indicating body satisfaction (BS) and negative perceptions indicating body dissatisfaction. Body image, including BS or dissatisfaction [6], plays a pivotal role in various aspects of psychological development, interpersonal relationships, and overall quality of life [6,7]. A survey of 5623 adolescents and adults in England by the UK Mental Health Foundation [8] shows that 1 in 5 individuals felt ashamed of their body image, resulting in feelings of anxiety, depression, and even suicidal thoughts. Disturbances in body image perception, such as body dissatisfaction, are pertinent factors in clinical issues, including obesity and eating disorders [6]. Generally, body image perception is an indicator of quality of life [7]. Consequently, understanding the relationship between social media use and body image is essential for addressing these significant issues.

While previous studies have explored the influence of appearance, family, peer, and media pressures on body dissatisfaction among teenage girls (eg, work by Roberts et al [9]), there has been a notable absence of specific attention to the impact of social media activities or aspects on body dissatisfaction. Conversely, Rodgers et al [10] found a connection between social media use, upward appearance comparison (UAC), and body dissatisfaction, with the latter being mediated by the internalization of an idealized appearance. Upward social comparison typically involves evaluating oneself with the aim of self-improvement by assessing the perceived advantages of the objects being compared and learning ways to enhance one's own attributes. Meanwhile, downward comparison is the process of comparing oneself with someone judged to be worse than oneself. Upward social comparison is often deemed unfavorable since the assessed object holds a higher "value" than oneself, while downward comparison is considered favorable as the object holds a lower "value" than oneself [11]. Therefore, this research will primarily focus on the dynamics of upward comparisons.

Furthermore, prior research has predominantly focused on platforms such as Snapchat, Facebook, and Instagram [9,12,13]. Studies related to TikTok delve into areas such as body dissatisfaction [14,15], BS in dancer challenges [16], body neutrality [17], and systematic literature reviews on body image [18]. TikTok has a substantial user base primarily consisting of Generation Z (Gen Z) individuals, aged between 11 years and 26 years, with approximately 37 million users worldwide. Indonesia stands as the second largest TikTok user in the world to date [19,20]. Finally, TikTok is one of the social media that focuses on user-generated videos that display the appearance of the human body, such as content in the category's selfie videos, self-portrait videos, dance, fashion, beauty and skin care, fitness or sports, and entertainment [21]. TikTok also strives to create an environment to support user's body positivity [21].

Conceptual Model

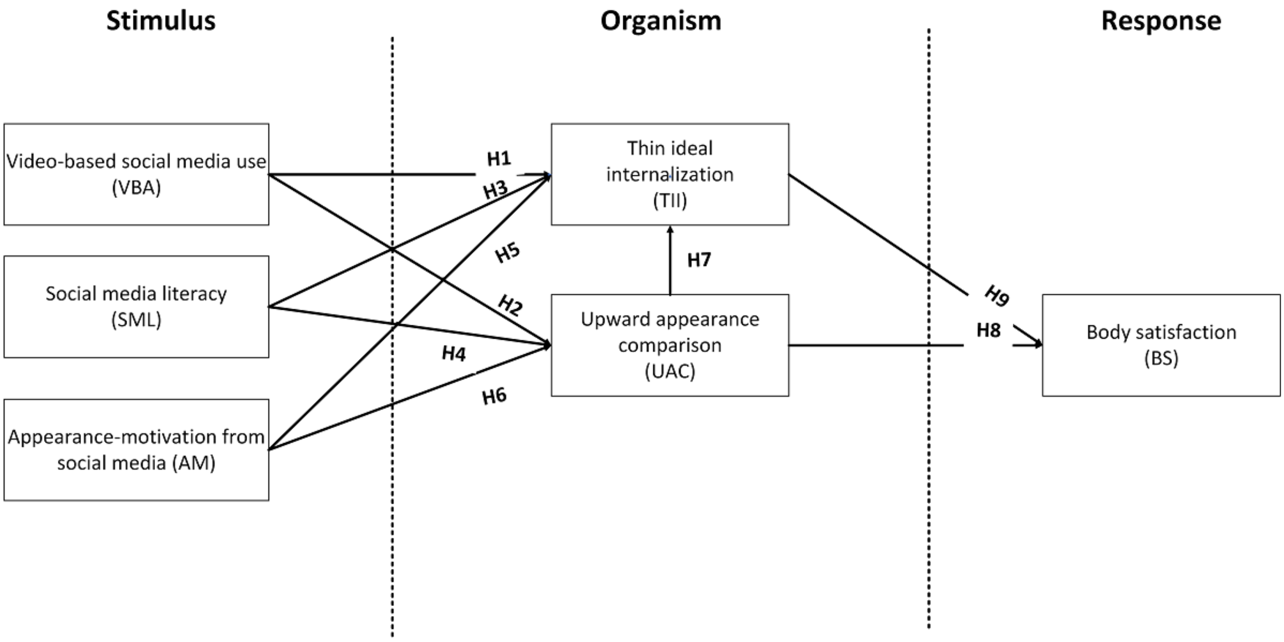
Jarman et al [12] applied the tripartite influence model (TIM) to investigate the relationship between social media intensity, frequency of use, BS, and well-being mediated by internalization and comparison in Australia. However, their study did not explicitly consider content related to appearance. This study uses the TIM, a sociocultural theory that explores the influence of family, friends, and media on body image dissatisfaction and eating disorders. To narrow the focus to social media use, we chose to emphasize 1 sociocultural aspect—social media. TIM suggests that the stimulus aspect has a notable impact on physical appearance [14]; yet, it does not thoroughly address how social media specifically affects body image. To address this gap, our focus lies on social media use involving videos prominently featuring physical appearance. Moreover, we use stimulus-organism-response (SOR) theory to categorize factors into stimulus, organism, and response. The SOR framework states that environmental conditions can present signs that can trigger (stimulus) the inner state of an individual (organism) so that the individual will produce a response (response) in either positive or negative form [22]. The development of SOR theory can help in understanding the reasons behind a person's behavior with the assumption that most human behavior reflects the stimuli we feel, so this theory is very relevant when it is related to problems related to human behavior under certain conditions. The combination of SOR and TIM offers a more comprehensive context for understanding the influence of social media use on human BS, using the explanatory power of SOR theory to delve into the reasons behind human behavior (response) in specific situations.

By focusing specifically on TikTok, we aim to identify the factors within this social media platform that influence BS among Gen Z individuals. Through this investigation, our goal is to shed light on the factors that contribute to body shaming incidents, increase user awareness of the effects of social media on body image perception, and empower individuals to take control of their usage. These insights can also serve as valuable guidance for social media developers, aiding them in creating platforms that foster a positive digital environment. Moreover, stakeholders can use these findings to inform and shape regulatory efforts, contributing to the establishment of a safer digital space.

According to TIM, pressure related to appearance is a consequence of sociocultural factors [9]. Therefore, the use of social media, especially on platforms such as TikTok, offers a means to assess these sociocultural influences. TikTok, being a platform centered around appearance, typically entails activities oriented toward videos (referred to as video-based social media use [VBA]), such as viewing, uploading, and interacting with images of oneself and others [23]. Examining specific activities to measure social media use can facilitate a deeper understanding of how social media affects body image [12].

Mink and Szymanski [14] suggested that variables of social media literacy (SML) could moderate the impact of TikTok usage on body dissatisfaction. Similarly, Tort-Nasarre et al [24] highlighted the importance of this variable in research concerning the impact of social media use on BS, noting that a high level of SML can assist individuals in developing positive body perceptions in conjunction with social media use. Furthermore, Roberts et al [9] indicated that internalization encompasses various aspects, such as thin-ideal internalization (TII), muscular internalization, and social media internalization. For our purposes, we narrow our focus to TII and social media internalization, as these variables are more inclusive of both genders than the more gender-specific muscular internalization. Thus, we propose Figure 1 as our conceptual model.

Figure 1. Proposed conceptual model. AM: appearance-motivation from social media; BS: body satisfaction; H: hypothesis; SML: social media literacy; TII: thin-ideal internalization; UAC: upward appearance comparison; VBA: video-based social media use.



Research Hypothesis

The activities examined in this study pertain to video-related actions because TikTok is a social networking site centered around video-based appearances [14]. According to Mink and Szymanski [14], these activities allow users to engage with the physical representation of other people's lives, including their appearance, through features such as share and comments. TII is an individual's cognitive level to believe and inspire that the ideal social definition of attractiveness, namely a thin appearance, is an actual attractiveness [25]. Jarman et al [12] stated that the use of social media, including the activities carried out on it, is one of the reasons for the internalization of an ideal thin appearance. Jarman et al [12] found that the more frequent use of social media led to a higher likelihood of internalizing an ideal thin appearance. In addition, Duan et al [26] found that people who are involved in activities on social media are more likely to internalize their ideal appearance; in this case, they internalize their appearance to make themselves look thin. Such internalization is made possible because videos posted on social media, such as selfies, can be seamlessly modified and edited using filters, artificial intelligence, and image processing tools and uploaded selectively by people to meet social body standards [27,28]. Considering various activities on social media, such as viewing posts or comments portraying ideal physical appearances on platforms such as TikTok, users are likely to internalize the notion of an ideal thin appearance. Therefore, we assume the following hypothesis:

H1: VBA influences TII.

Users engaging in VBA activities have several goals, such as admiring individuals whose physical appearance aligns with societal standards, maintaining relationships, and receiving visual feedback [29]. However, many studies have found that user participation in activities on social media, such as scrolling through posts adhering to ideal standards [30], uploading content with positive comments [31], and content receiving numerous likes [32], can negatively impact individual perceptions of body image. As outlined by Mink and Szymanski [14], activities within VBA platforms allow users to be involved with the physical representation of others' lives, including their appearance, through features such as likes, shares, and comments. UAC is an individual behavior where one assesses their appearance against individuals deemed to possess a superior appearance, such as celebrities or social media influencers [33]. According to Wang et al [34], social media users who engage in appearance comparisons often do so with individuals considered superior, especially those portraying an idealized life on social media. This comparison prompts individuals to seek improvement, often leading to the editing of selfies to compensate for perceived weaknesses [27]. Women, in particular, may seek quantitative proof of popularity through metrics such as likes, comments, and followers [28]. After reviewing several previous studies, we posit that UACs, including those with celebrities or influencers, are also influenced by VBA usage activities. When TikTok users encounter videos showcasing an ideal physical appearance uploaded by influencers or celebrities, accompanied by positive reactions from other users, they are likely to initiate comparisons of their own physical appearance with these influencers or celebrities. Therefore, we propose the following hypothesis:

H2: VBA influences UAC.

SML is the application of critical analysis regarding the motivations behind postings on social media and the nature of images that are constructed, generally unrealistically, when viewing images that focus on appearance on social media [35]. According to Tamplin et al [36], a high level of media literacy is characterized by the ability to think critically about media. Tamplin et al [36] explained that the ability to critique idealized performance images, considering their realism and commercial intent, helps offset the impact of viewing these images on individuals. The level of media literacy is also marked by the ability to judge whether or not an image in the media is intended to reduce the credibility and persuasive influence of the media [37]. Individuals with high levels of media literacy tend to avoid comparing themselves with ideal thin appearances in media that are considered unrealistic [37]. McLean et al [37] also found that media literacy produces a protective effect, in which research respondents who have a low level of critical thinking are easily negatively affected when viewing images that display someone who has an ideally thin appearance, and research respondents who have high critical thinking are not easily negatively affected. Therefore, we propose the following hypothesis:

H3: SML influences TII.

McLean et al [37] explained that individuals with high media literacy tend to prefer to compare themselves with ideal thin appearances in the media, which are considered unrealistic. A high level of media literacy, such as thinking that postings of someone whose physical appearance looks ideal in the media are unrealistic or manipulated, can protect users from upward comparisons, such as those with celebrities or influencers [37]. McLean et al [37] also stated that high critical thinking is a protective factor from a high upward comparison. Mink and Szymanski [14] also found that SML exacerbated the direct relationship between TikTok use and UACs. When users have low SML skills, they may not be able to determine the authenticity of the TikTok videos they see. They may need to realize that the videos have been edited or changed [14]. Therefore, we assume the following hypothesis:

H4: SML influences UAC.

Appearance motivation (AM) from social media is a feeling that arises when individuals want to look attractive and try to improve themselves because they are motivated by messages conveyed by the media [10]. The media plays an important role in spreading the message that an ideal appearance is a good thing, such as popularity, and success can be achieved if someone has an attractive appearance [38]. Attractive appearance in the media has standards, such as that women must be young, thin, and sexy [38]. When the message reaches the audience, they will become motivated and try to change their appearance to make it look more ideal [38]. In addition, Kvardova et al [39] explained that the effects of media use tend to affect the internalization of an ideal thin appearance. In line with Wilson et al [40], internalizing an ideal thin appearance can be interpreted as an indicator of a motivational approach to thin bodies. According to Graff [41], social media displays content that is somewhat unrealistic and causes users to compare their appearance, thus motivating them to engage in diet and exercise. Therefore, we propose the following hypothesis:

H5: AM from social media influences TII.

Motivation for appearance from social media is a feeling that arises when individuals want to look attractive and try to improve themselves because they are motivated by messages conveyed by the media [10]. The motivation to improve by making comparisons based on the desire to improve aspects of oneself has been associated with increased exercise activity and BS [15]. When individuals are motivated to improve their physical appearance, they tend to make upward social comparisons [42], such as following attractiveness standards by changing their physical appearance to become ideally thin [43]. In addition, when individuals are urged to improve themselves, they will compare themselves with others who are better off, such as idealized models in the media [10]. Then, we suggest the following hypothesis:

H6: AM from social media influences UAC.

UAC indicates that individuals evaluate their appearance against other individuals who are considered to have a superior appearance than themselves, such as celebrities or social media influencers [44,45]. TII or internalization of an ideal thin appearance is an individual's cognitive level to believe and inspire that the ideal social definition of attractiveness, namely a thin appearance, is an actual attractiveness [25]. Rodgers et al [10] found that social media use may be associated with body dissatisfaction through comparisons of upward appearance and internalization of thin appearance. According to Seekis et al [45], a comparison of upward appearances with celebrities or health influencers leads to the drive to be thin and dissatisfied, respectively. In addition, according to Hsu et al [46], internalization measures of ideal thin appearance include statements that assess the degree of comparison with an ideal thin body, such as "I wish I looked like a swimsuit model" and "I often read magazines such as Cosmopolitan, Vogue, and Glamor and comparing my looks to models." Therefore, we assume the following hypothesis:

H7: UAC influences TII.

Social media impacts the development of body dissatisfaction through 2 channels: the comparison of appearances on social media and the internalization of ideal appearance [12]. According to Wang et al [34], social media users who compare themselves may engage in upward comparisons, considering the appearance and ideal life posted on social media, which is likely to have a detrimental effect on BS and well-being. According to the TIM, exposure to the media can increase the likelihood that individuals will adopt an ideal appearance as their standard and contribute to a greater tendency to compare their physical appearance with someone's unrealistic uploads in the media [12]. Individuals who compare themselves with social media's ideal physical appearance standards will likely feel that they do not conform to them, reducing their BS [12]. Mink and Szymanski [14] showed that the more often individuals use TikTok, the more they engage in appearance comparisons and the more body dissatisfaction they experience. Today, individuals have access to people worldwide, meaning that comparisons can occur at an unprecedented level and scale [47]. Thus, we assume the following hypothesis:

H8: UAC influences BS.

TII or internalization of an ideal thin appearance is an individual's cognitive level to believe and inspire that the ideal social definition of attractiveness, namely a thin appearance, is an actual attractiveness [36]. Individuals often post carefully edited and curated content about their physical appearance, which presents unrealistic content to make them appear attractive [25]. Rodgers et al [10] stated that individuals involved in internalizing an ideal appearance have a lower level of performance satisfaction. Individuals use media content as a source of information about how to improve their physical appearance and compare their physical appearance with media models to set standards that must be met [10]. Jarman et al [12] also found that internalizing an ideal thin appearance is inversely related to BS. Exposure to images on social media internalizing an ideal thin appearance causes dissatisfaction with the body and face of the individual [32]. Thus, we propose the following hypothesis:

H9: TII influences BS.

Methods

Study Design

This study used a mixed methods approach using questionnaires (quantitative approach) and interviews (qualitative approach). A readability test was carried out on the questionnaire to verify the appropriateness of the questions within the research context. This assessment encompassed an examination of the diction and grammar used in the questionnaire to ensure clarity. The primary aim was to ascertain that all questions and instruments in the questionnaire could be easily and clearly understood by the respondents. The readability test was carried out for 6 days with a total of 13 respondents. This evaluation took place through both digital and offline channels. We conducted a pilot study of 33 respondents where the Cronbach α (CA) value for each construct was more than 0.7 [48].

The questionnaire link was distributed through various social media, such as Instagram, Twitter, Line, and WhatsApp, from February 28 to March 29, 2023, with a total duration of 30 days. Respondents involved in this study are TikTok users who are Gen Z (aged between 17 years and 26 years). Most of the respondents are in Greater Jakarta and are bachelor students. Most users' frequent activities on TikTok are liking videos and they often view content related to physical appearance as well as seeing body shaming comments on TikTok. This study obtained 507 respondents, and Table 1 summarizes the demographics of the respondents. After obtaining quantitative data, we collected qualitative data by conducting interviews with 32 TikTok users in the age range of 17-26 years.

Table 1. Demographics of respondents who filled out the questionnaire.

Demographics	Respondents, n (%)
Gender	
Men	134 (26)
Women	373 (74)
Age (years)	
17-19	76 (14.99)
20-22	339 (66.86)
22-26	92 (18.15)
Domicile	
Greater Jakarta	312 (61.54)
Java Island outside Greater Jakarta	125 (24.65)
Outside Java Island	70 (13.81)
Education	
High school students	93 (18.34)
Diploma students	45 (8.88)
Bachelor students	355 (70.02)
Master students	14 (2.76)
Reasons to use TikTok	
Entertainment	421 (47.09)
Increase creativity	129 (14.43)
Know the latest information	301 (33.67)
Add friendship	29 (3.24)
Others	14 (1.57)
Most frequent activities on TikTok	
Upload a video	128 (18.47)
Comment on a video	132 (19.05)
Like a video	433 (62.48)
Frequency of viewing content related to physical appearance on TikTok	
Never	15 (2.96)
Rarely, occasionally when using TikTok	244 (48.13)
Often, every time TikTok is used	248 (48.92)
Frequency of seeing body shaming comments on TikTok	
Never	51 (10.06)
Rarely, occasionally when using TikTok	220 (43.39)
Often, every time TikTok is used	236 (46.55)

Quantitative data for this study were processed using the covariance-based structural equation method (CB-SEM) using the AMOS 24 (IBM) app, SPSS Statistics (version 27; IBM Corp), Google Sheets, and Excel (Microsoft Corp). CB-SEM aimed to confirm the relationship between variables defined theoretically in a model, attempting to minimize the differences between the covariance matrices implied in the model and the covariance matrices obtained from research samples [48]. The CB-SEM procedure included 3 stages: model specification, measurement model evaluation, and structural model evaluation [48]. In addition, content analysis was used to process qualitative data.

To analyze in more detail the results of hypothesis testing, interviews were conducted with respondents aged between 17 years and 26 years who had at least actively used the TikTok application. A total of 32 respondents were interviewed from April 19, 2023, to May 13, 2023, digital and offline. All interview results obtained were then processed and analyzed using one of the qualitative data processing methods, namely content analysis. This stage produces an interpretation of all patterns resulting from the interview process so that it can support the explanation of the hypothesis results in this research.

Research Instruments

The questionnaire was composed of 4 parts. The first part contained instructions for filling out the questionnaire, the respondent's consent, and validation that the respondent is a TikTok application user and is aged 17-26 years. The second part contained questions on the demographics of the respondents. The third section contained general information about the TikTok application, physical appearance content, and body shaming. The fourth section contained 32 measurement items. Multimedia Appendix 1 describes the questionnaire used in this study, and Multimedia Appendix 2 explains the interview questions for each hypothesis.

Ethical Considerations

This study has obtained approval from the research unit of the Faculty of Computer Science, University of Indonesia (letter S-1/UN2.F11.D1.5/PPM.00.00/2024). In accordance with the Decree of the Board of Trustees of the University of Indonesia, based on the research ethics policies applicable to members of the faculty of the University of Indonesia, each research unit or department implements research supervision guidelines, and oversees and adheres to the responsibilities prescribed therein. In line with university policy, the Research and Community Service Department, Faculty of Computer Science, University of Indonesia, adhere to the guidelines and procedures established by the faculty and provided ethics approval for this study. All respondent data were anonymized and exclusively used for the purposes outlined in this research. All questionnaire respondents provided written informed consent, and all the interview respondents provided verbal informed consent to participate in this study.

Results

All variables in Table 2 surpass the 0.50 threshold of average variance extracted values, indicating the successful representation of the variables by their respective indicators. Additionally, the reliability of the measurements was assessed using CA and construct reliability. All CA and construct reliability values met the recommended minimum threshold of 0.7, signifying strong reliability [48]. Goodness-of-fit index values are outlined in Table 3, while Table 4 provides an overview of the R^2 values.

Table 2. CR[a], CA[b], and AVE[c] values.

Variable	CR	CA	AVE
Video-based social media use	0.858	0.896	0.559
Social media literacy	0.788	0.826	0.712
Appearance motivation	0.917	0.888	0.948
Thin-ideal internalization	0.910	0.936	0.857
Upward appearance comparison	0.881	0.913	0.750
Body satisfaction	0.922	0.933	0.834

[a]CR: construct reliability.

[b]CA: Cronbach α.

[c]AVE: average variance extracted.

Table 3. Goodness-of-fit index values.

GoF[a] index	Cutoff value	Value	Description
CMIN/df[b]	<2	1.213	Good fit
RMSEA[c]	0.08	0.021	Good fit
NFI[d]	0.9	0.967	Good fit
CFI[e]	0.9	0.994	Good fit
GFI[f]	0.9	0.95	Good fit
TLI[g]	0.9	0.992	Good fit
RMR[h]	0.05	0.05	Good fit

[a]GoF: goodness-of-fit.
[b]CMIN/df: minimum discrepancy function by degrees of freedom divided.
[c]RMSEA: root-mean-square error of approximation.
[d]NFI: normed fit index.
[e]CFI: comparative fit index.
[f]GFI: goodness-of-fit index.
[g]TLI: Tucker-Lewis index.
[h]RMR: root-mean-square residual.

Table 4. R^2 values.

Dependent variable	R^2
Upward appearance comparison	0.451
Thin-ideal internalization	0.452
Body satisfaction	0.099

Next, the hypothesis testing carried out is a 2-tailed test. Hypothesis testing can be done by looking at the significance value or *P* value. Since the test is carried out in 2 directions, this test uses a significance of 5%, which means that a hypothesis with a *P* value of >.005 is rejected and vice versa is accepted [48]. Based on Table 5, 2 of the 9 hypotheses were rejected because they had a *P* value of >.005.

Table 5. Hypothesis testing results.

Hypothesis	Estimate	*P* value	Description
H1: VBA[a] influences TII[b]	–0.082	.19	Rejected
H2: VBA influences UAC[c]	0.469	.002	Accepted
H3: SML[d] influences TII	–0.14	.001	Accepted
H4: SML influences UAC	–0.083	.06	Rejected
H5: AM[e] influences TII	0.261	.002	Accepted
H6: AM influences UAC	0.28	.001	Accepted
H7: UAC influences TII	0.534	.003	Accepted
H8: UAC influences BS[f]	–0.204	.003	Accepted
H9: TII influences BS	–0.168	.006	Accepted

[a]VBA: video-based social media use.
[b]TII: thin-ideal internalization.
[c]UAC: upward appearance comparison.
[d]SML: social media literacy.
[e]AM: appearance motivation.
[f]BS: body satisfaction.

This study concludes that there is no significant relationship between the activity of using VBA and the internalization of an ideal thin appearance. TikTok shares this observation by expressing its commitment to supporting existing content to be more inclusive and a “body-positive environment” [21]. Some of the steps taken by TikTok are increasing restrictions on ads that promote negative body image, reducing advertisements for diet products that exaggerate their promotion, working with creators to promote body positivity, and working with various organizations to help people who experience body image problems [21]. Respondents also felt that an ideal body does not mean thin or slender physical appearance: “In my opinion, a body that is still classified as ideal is if the body is slightly heavier than BMI. If a thin body is not ideal in my opinion” (interviewee 27) and “Body ideal is a body that is not thin and not fat” (interviewee 2). Then, respondents who feel their physical appearance is thin enough or feel their physical appearance is good enough have no desire to be thinner anymore: “Not really wanting to change their appearance to be thinner, but more motivated to change their lifestyle to be healthier” (interviewee 4). This is also encouraged by the growth of TikTok content, which promotes self-love or body positivity toward physical appearance and can improve the mental health and well-being of its users [49]. This movement changes the perspective of its users in viewing their physical appearance positively [49]. They realize that they are not alone when they feel insecure and know that many other users on TikTok are spreading this positive movement [49].

In addition, the relationship between the activity of using VBA influences the UACs. This is in line with Mink and Szymanski [14], who stated that the longer an individual uses TikTok, the more involved they are in comparing appearance and the more they will watch their body. In addition, Jarman et al [12] stated that when individuals are exposed to content on social media, the ideal body for them is the body they see. The impact is that they will compare their appearance with this ideal body perception and tend to produce negative self-evaluations [12]. Furthermore, the results of this hypothesis were strengthened by the results of qualitative interviews, which explained that respondents felt that the longer they opened TikTok, the more they would be increasingly exposed to physical appearance content that displays ideal body standards according to users. At first, they just admired, but over time, the respondents wanted to look as good as the people in the content they saw on TikTok: “If you look at a tall physical appearance, especially men, it will affect self-assessment and comparison with that person” (interviewee 20). In addition, this is exacerbated by the TikTok algorithm, which recommends videos according to the number of viewing likes: “Because I often see physical appearance content, especially for women who like to upload dance content on TikTok, I wonder why people that person is prettier than me. That person has clean skin, and I am also influenced to use skin care so that I look as good as the person I see on TikTok” (interviewee 32). This is also directly supported by TikTok, which explains that their recommendation system works after looking at user preferences through various interactions, such as comments or following certain accounts [49].

This study found that SML influences the internalization of an ideal thin appearance, which aligns with McLean et al [37]. In addition, Tamplin et al [36] explained that the ability to critique idealized performance images, considering their realism and commercial intent, helps offset the impact of viewing these images on individuals. Furthermore, the results of this hypothesis are strengthened by the results of qualitative interviews, which explain that respondents tend to be more aware of content that promotes a thin body as the ideal body, especially in less realistic content, for example, the TikTok trend to show a thin body: “If you have to the point that it is unrealistic, it is very toxic and destructive. If it is more realistic, in my opinion it is quite good, such as the promotion of a healthy lifestyle” (interviewee 19) and “This trend is not good because it sets new standards that lead to body shaming if you don't meet these standards” (interviewee 27). In addition, respondents believe more in content that promotes a slim body is an ideal body if the content creator has credibility, for example, a nutritionist and bodybuilder: “If the content is made by a health professional and can prove it, I agree with the content” (interviewee 26). This is in line with Mink and Szymanski [14], who explained that it is easy for individuals to determine the authenticity of the content they see when they have high SML.

Moreover, this study observed that there is no significant impact of SML on UACs. This is in line with Mink and Szymanski [14], where individuals who have strong SML skills can inadvertently make appearance expectations and appearance evaluations more prominent in women’s minds. Hence, women must change their physical appearance by using filters or editing to look as good as people they find attractive. The results of this hypothesis are strengthened by the results of qualitative interviews, which explain that respondents still compare their physical appearance with people they consider better looking in TikTok even though their level of SML is already high, for example, by knowing the use of filters on the videos they watch: “Ever compared because they saw other people use filters too and are curious about how they look on their own faces” (interviewee 7) and “I’ve compared if it matches my preferences. There was also a thought that I wanted to be like him because it’s more in line with preferences and just looks better” (interviewee 29).

Furthermore, this study revealed that AM from social media influences the internalization of an ideal thin appearance, and this result is in line with Trekels and Eggermont [38], where attractive appearance in the media has standards, such as that women must be young, thin, and sexy. When the message reaches individuals, they will become motivated and change their appearance to make it look better [38]. Furthermore, the results of this hypothesis are strengthened by the results of qualitative interviews, which explain that respondents are motivated after watching content that promotes better physical appearance, such as exercise and a healthy lifestyle: ”Feel motivated to look more attractive when viewing makeup, skin care, or sports content” (interviewee 5) and “Ever motivated after seeing only gym content but wasn't interested in looking thin like a model” (interviewee 1). After being motivated, the interviewees felt moved to change their physical appearance by doing the same thing to have an equally good physical appearance. In line with Trekels and Eggermont [38], the media is important in spreading the message that an ideal appearance is a good thing. For example, popularity and success can be achieved if someone has an attractive appearance.

This study affirmed that social media AM influences the UAC, which is in line with Rodgers et al [10,42], where when individuals have the drive to improve themselves, they will compare themselves with other people who are better looking for them, such as idealized models in the media. In addition, when individuals are motivated to improve their physical appearance, they tend to make upward social comparisons [42,43]. Furthermore, the results of these hypotheses were strengthened by the results of qualitative interviews, which explained that 24 out of 32 (75%) respondents felt that if they watched a video that presented an ideal physical appearance in their opinion, the respondents wanted to know whether their physical appearance was as good as the ideal standard derived from the video they watched: “Once motivated to eat healthy food made by people on TikTok, where they have more fit bodies, so they are influenced to make that food” (interviewee 5). In line with Grabe et al [43], when individuals feel motivated to improve their physical appearance, they follow attractiveness standards by changing their physical appearance.

In addition, respondents also felt that content displaying an ideal physical appearance on TikTok received validation, such as praise comments. Hence, respondents felt motivated to change their physical appearance to be more ideal: “Very motivated to look cool and handsome because people also commented that interesting in the content” (interviewee 20) and “Motivated to look more ideal because if you upload a video later the audience will definitely like it and there will be lots of laudatory comments” (interviewee 3). This aligns with Trekels and Eggermont [38], in which media messages that focus on appearance can be strengthened through validation from others, such as comments. Rodgers et al [10] also explained that comments from users on social media on content that focus on physical appearance explicitly describe appearance comparisons and can encourage other users to engage in physical appearance comparisons.

The comparison of upward appearance is found to influence the internalization of an ideal thin appearance. This is in line with the research of Seekis et al [45], where when individuals have the desire to look as good as celebrities and influencers, there is an urge to change their physical appearance to become thinner. Furthermore, the results of the hypothesis are strengthened by the results of qualitative interviews, which explain that respondents try to look as good as people they consider ideal on the TikTok app: “Ever compared with thinner people. Feeling a thin body is interesting and has quite an impact on my thoughts towards an ideal body because it indicates the person has a healthy lifestyle” (interviewee 28). They usually assess to measure how far their score is from the ideal value so that when they feel they have not met these ideal standards, they tend to internalize new ideal standards: “Ever compared and tried to measure whether my body already looks the same as his, because almost every I watch TikTok every day and see interesting people, so it has quite an influence on the perception of an ideal body and beauty standard” (interviewee 3) and “Because I compare things, I can know what the ideal point is like. If I’m not ideal, I will do certain activities so that I can be as ideal as that” (interviewee 26). This is in line with Hockey et al [50] and Shen et al [51], who stated that the effect of using social media is associated with a high level of comparison and ultimately leads to internalization of the ideal thin appearance of its users.

Next, individuals who compare themselves with ideal physical appearance standards on social media will most likely feel that they do not conform to these standards, thereby lowering their level of BS. This is in line with Weinstock [47], where individuals not only compare themselves with others, but they also compare themselves with other people’s optimized versions thanks to video editing apps, which can worsen mental health and lower the individual’s level of satisfaction. Conversely, the respondents have no desire to have the same physical appearance as the videos they watch on the TikTok app because their BS tends to be high: ”Does not affect self-satisfaction because they are already satisfied with their current physical appearance“ (interviewee 25) and “Comparing physical appearance to the one in the video doesn’t really affect self-satisfaction because you feel enough with what you are now” (interviewee 18). They already feel sufficient and confident with their current physical appearance: “I don’t compare with other people; I don’t even care about the scales either. Now, I just look in the mirror and observe whether I have approached the ideal body or not. I feel quite satisfied when, for example, I wear clothes and look good when worn” (interviewee 5). This is also in line with Rodgers et al [10], who stated that high self-confidence or self-esteem tends to result in lower levels of dissatisfaction, especially among women.

This study found that internalizing an ideal thin appearance is inversely related to BS. When individuals are not affected by internalizing the thin appearance of the content they see on social media, they tend to be satisfied with their current physical appearance [12]. Furthermore, the results of this hypothesis are strengthened by the results of qualitative interviews, which explain that 17 out of 32 (53%) respondents internalize the ideal thin body standard with what they see on the TikTok app, so most of their BS levels are also affected because they feel they have not met the ideal body standard. This is in line with Tiggemann et al [32] about exposure to images on social media related to the internalization of an ideal thin appearance, causing dissatisfaction with the body and face of the individual. We found a pattern suggesting that the influence depends on the level of self-acceptance of the respondents. With this internalization, if the current body condition is not per the ideal standard and respondents have low self-acceptance, then they will tend to be affected because they have not accepted the current body condition. Conversely, if respondents already have self-acceptance, they tend to be more accepting of all their body conditions and produce a stable level of BS. This is also in line with Rodgers et al [10], who stated that high self-confidence or self-esteem in individuals tends to result in lower levels of dissatisfaction, especially among women.

Discussion

Principal Findings

The desire to resemble celebrities seen on TikTok (UAC) is shaped by various activities on the platform, including watching videos, perusing comments and likes, and uploading content (video-based activity). TikTok exposes users to diverse physical appearances that are often perceived as more appealing, encompassing facial features, skin type, color, and body shape. Exposure to such diverse physical attributes prompts unconscious self-evaluation, fostering a longing to emulate those deemed more attractive. This desire is further influenced by the heightened motivation to present oneself attractively on social media (AM). Greater motivation heightens the aspiration to mirror the perceived attractiveness of these individuals [38].

Moreover, this study found the development of TII is influenced by AM, SML, and UAC. The drive to present attractively on TikTok (AM) shapes the perception of an ideal body. Regular exposure to videos emphasizing a "thin body as the ideal" (TII) fosters the internalization of this concept. Conversely, users with a higher level of SML can discern realistic content, thus avoiding the internalization of potentially harmful ideals. Although the TikTok algorithm aids in content curation, personal discernment becomes crucial; failure to discern can lead to increased exposure to harmful content. The aspiration to resemble individuals observed on TikTok (UAC) prompts a comparative self-assessment, often yielding unfavorable outcomes, as users perceive those they watch as possessing more attractive physical attributes. This perpetuates an ideal that equates thinness with attractiveness, further reinforcing TII.

Consequently, the intense desire to emulate others on TikTok (UAC) contributes to a diminished self-assessment, fostering a heightened awareness of one's perceived inadequacies and resulting in body dissatisfaction. Similarly, strong adherence to thin body standards establishes unrealistic benchmarks, potentially leading to dissatisfaction when one's appearance fails to meet these standards.

Notably, the shift from the original appearance pressure variable to the AM variable revealed its influence on TII and UAC. Findings from the qualitative analysis indicate that users are motivated by positive comments on content, driving a desire to emulate the praised appearance of content creators, aligning with Trekels and Eggermont [38], who highlighted the reinforcement of appearance-focused media messages through social validation. Furthermore, this study indicates that users' SML levels impact TII but do not significantly affect UAC. Despite users' adeptness in discerning social media techniques, they still compare themselves with physically attractive individuals, as seen in their social media feeds. Notably, video-based activity, an original stimulus variable in the TIM, does not significantly impact TII. This contrasts with previous findings [14,26,28], supported by statements from other sources, including research results [50] and statements from social media platforms, such as TikTok [21] and Instagram, signaling a changing trend in the perception of the ideal body, emphasizing self-love and body positivity.

This study establishes the significant impact of organismal variables, namely UAC and TII, on the response variable, BS. Prior research has highlighted the negative impact of comparing oneself with others on social media, leading to decreased BS [9,12,14]. Conversely, individuals exhibiting self-acceptance tend to maintain stable levels of BS [10]. Similarly, previous studies emphasize that internalizing the ideal thin appearance from TikTok content leads to dissatisfaction with one's physical appearance, while self-acceptance fosters stable BS [10,32].

Practically, these findings offer crucial insights for TikTok, a platform that frequently features content related to physical appearance, aiming to mitigate its potential negative impact on users' BS. Ong and Sündermann [52] found that a self-guided mHealth app could improve body image concerns and self-compassion in young adult university students. Moreover, a social media–based, fictional 6-episode video series with self-guided web-based activities for improving body image could increase trait BS and mood and decrease internalization of appearance ideals [53]. Therefore, this research encourages the development of a healthier social media environment, fostering users' mental well-being and comfort.

Limitations

This study has two limitations. First, respondents mostly live in the greater Jakarta area which is the biggest city in Indonesia. Second, the respondents for this study are predominantly undergraduate students who are either currently pursuing or had completed their undergraduate education. Therefore, the background of the area and current activities of these respondents may impact the results of this study which are only based on the perspectives of students and users in the big cities in Indonesia. Future studies could be carried out to involve respondents from urban areas and other users in smaller cities. However, it is important to acknowledge that these limitations align with a survey conducted by the IDN Research Institute [22]. The survey indicated that over 70% of Gen Z individuals in Indonesia are still in junior high school and senior high school, while the percentage of Gen Z individuals currently pursuing education beyond high school is only 10.36% [22]. Future studies are expected to provide comparisons between the results of this study and studies involving other generations and other social media.

Acknowledgments

We want to convey our gratitude to the University of Indonesia for the Grant Hibah PUTI Hibah Publikasi Terindeks Internasional/PUTI Q2 Number NKB-570/UN2.RST/HKP.05.00/2023.

Data Availability

The data sets generated or analyzed during this study are not publicly available due to the lack of authority to share data.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire.
[DOCX File , 25 KB - humanfactors_v11i1e58371_app1.docx]

Multimedia Appendix 2
Interview questions.
[DOCX File , 16 KB - humanfactors_v11i1e58371_app2.docx]

References

1. Chaffey D. Global social media statistics research summary 2023. URL: https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/ [accessed 2023-06-23]
2. Fadli R. Body shaming - Pengertian, Jenis dan Cara Mengatasinya, 2023. URL: https://www.halodoc.com/kesehatan/body-shaming [accessed 2023-06-23]
3. Andrews F. Adele opens up about body shaming after weight loss: that hurt my feelings. 2021. URL: https://www.thenationalnews.com/lifestyle/wellbeing/2021/10/09/adele-opens-up-about-body-shaming-after-weight-loss-that-hurt-my-feelings/ [accessed 2023-06-23]
4. Bazaar U. Kylie Jenner angkat bicara mengenai komentar body shaming. 2020. URL: https://harpersbazaar.co.id/articles/read/4/2020/10727/kylie-jenner-angkatbicara-mengenai-komentar-body-shaming [accessed 2023-06-23]
5. National Eating Disorder Collaboration. Body image. 2022. URL: https://nedc.com.au/eating-disorders/eating-disorders-explained/body-image/ [accessed 2023-06-23]
6. Markey CN. Invited commentary: why body image is important to adolescent development. J Youth Adolesc 2010;39(12):1387-1391. [doi: 10.1007/s10964-010-9510-0] [Medline: 20339908]
7. Nayir T, Uskun E, Yürekli MV, Devran H, Çelik A, Okyay RA. Does body image affect quality of life?: A population based study. PLoS One 2016;11(9):e0163290 [FREE Full text] [doi: 10.1371/journal.pone.0163290] [Medline: 27649389]
8. UK Mental Health Foundation. Body image report—executive summary. UK Mental Health Foundation. 2019. URL: https://www.mentalhealth.org.uk/explore-mental-health/articles/body-image-report-executive-summary [accessed 2023-06-23]
9. Roberts SR, Maheux AJ, Hunt RA, Ladd BA, Choukas-Bradley S. Incorporating social media and muscular ideal internalization into the tripartite influence model of body image: towards a modern understanding of adolescent girls' body dissatisfaction. Body Image 2022;41:239-247 [FREE Full text] [doi: 10.1016/j.bodyim.2022.03.002] [Medline: 35306356]
10. Rodgers RF, Slater A, Gordon CS, McLean SA, Jarman HK, Paxton SJ. A biopsychosocial model of social media use and body image concerns, disordered eating, and muscle-building behaviors among adolescent girls and boys. J Youth Adolesc 2020;49(2):399-409. [doi: 10.1007/s10964-019-01190-0] [Medline: 31907699]
11. Grogan S. Body image: understanding body dissatisfaction in men, women, and children. Choice Reviews Online 1999;37(01):370628. [doi: 10.5860/choice.37-0628]
12. Jarman HK, Marques MD, McLean SA, Slater A, Paxton SJ. Social media, body satisfaction and well-being among adolescents: a mediation model of appearance-ideal internalization and comparison. Body Image 2021;36:139-148. [doi: 10.1016/j.bodyim.2020.11.005] [Medline: 33285385]
13. Rousseau A, Rodgers RF, Eggermont S. A biopsychosocial model for understanding media internalization and appearance dissatisfaction among preadolescent boys and girls. Commun Res 2020;47(3):346-372. [doi: 10.1177/0093650217739996]
14. Mink DB, Szymanski DM. TikTok use and body dissatisfaction: examining direct, indirect, and moderated relations. Body Image 2022;43:205-216. [doi: 10.1016/j.bodyim.2022.09.006] [Medline: 36191378]
15. Pryde S, Prichard I. TikTok on the clock but the #fitspo don't stop: the impact of TikTok fitspiration videos on women's body image concerns. Body Image 2022;43:244-252. [doi: 10.1016/j.bodyim.2022.09.004] [Medline: 36194987]
16. Joiner R, Mizen E, Pinnell B, Siddique L, Bradley A, Trevalyen S. The effect of different types of TikTok dance challenge videos on young women's body satisfaction. Comput Human Behav 2023;147:107856. [doi: 10.1016/j.chb.2023.107856]
17. Seekis V, Lawrence RK. How exposure to body neutrality content on TikTok affects young women's body image and mood. Body Image 2023;47:101629 [FREE Full text] [doi: 10.1016/j.bodyim.2023.101629] [Medline: 37742535]
18. Harriger JA, Wick MR, Sherline CM, Kunz AL. The body positivity movement is not all that positive on TikTok: a content analysis of body positive TikTok videos. Body Image 2023;46:256-264. [doi: 10.1016/j.bodyim.2023.06.003] [Medline: 37379612]

19. Santika EF. Kelompok anak muda jadi pengguna terbesar TikTok, usia berapa mereka?. 2023. URL: https://databoks.katadata.co.id/datapublish/2023/09/27/kelompok-anak-muda-jadi-pengguna-terbesar-tiktok-usia-berapa-mereka [accessed 2023-10-01]
20. Dixon SJ. Gen Z: Selected social media platforms in the U.S. 2020-2025. 2023. URL: https://www.statista.com/statistics/1276021/instagram-snapchat-tiktok-gen-z-users/ [accessed 2023-10-01]
21. TikTok. Coming together to support body positivity on TikTok | TikTok Newsroom. 2020. URL: https://newsroom.tiktok.com/en-us/coming-together-to-support-body-positivity-on-tiktok [accessed 2023-06-23]
22. IDN Research Institute. Indonesia Gen Z Report. IDN Times. 2022. URL: https://cdn.idntimes.com/content-documents/indonesia-gen-z-report-2022.pdf [accessed 2024-06-11]
23. Anderson M, Jiang J. Teens, social media and technology. 2018. URL: https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/ [accessed 2023-06-23]
24. Tort-Nasarre G, Pollina Pocallet M, Artigues-Barberà E. The meaning and factors that influence the concept of body image: systematic review and meta-ethnography from the perspectives of adolescents. Int J Environ Res Public Health 2021;18(3):1140 [FREE Full text] [doi: 10.3390/ijerph18031140] [Medline: 33561092]
25. Takamura A, Yamazaki Y, Omori M. Developmental changes in fat talk to avoid peer rejection in Japanese girls and young women. Health Psychol Open 2019;6(1):2055102919854170 [FREE Full text] [doi: 10.1177/2055102919854170] [Medline: 31210957]
26. Duan C, Lian S, Yu L, Niu G, Sun X. Photo activity on social networking sites and body dissatisfaction: the roles of thin-ideal internalization and body appreciation. Behav Sci (Basel) 2022;12(8):280 [FREE Full text] [doi: 10.3390/bs12080280] [Medline: 36004851]
27. Chae J. Virtual makeover: Selfie-taking and social media use increase selfie-editing frequency through social comparison. Comput Human Behav 2017;66:370-376. [doi: 10.1016/j.chb.2016.10.007]
28. Chua THH, Chang L. Follow me and like my beautiful selfies: Singapore teenage girls' engagement in self-presentation and peer comparison on social media. Comput Human Behav 2016;55:190-197. [doi: 10.1016/j.chb.2015.09.011]
29. Bell BT. "You take fifty photos, delete forty nine and use one": a qualitative study of adolescent image-sharing practices on social media. Int J Child Comput Interact 2019;20:64-71. [doi: 10.1016/j.ijcci.2019.03.002]
30. Fardouly J, Holland E. Social media is not real life: the effect of attaching disclaimer-type labels to idealized social media images on women's body image and mood. New Media Soc 2018;20(11):4311-4328. [doi: 10.1177/1461444818771083]
31. Tiggemann M, Barbato I. "You look great!": the effect of viewing appearance-related Instagram comments on women's body image. Body Image 2018;27:61-66. [doi: 10.1016/j.bodyim.2018.08.009] [Medline: 30138768]
32. Tiggemann M, Hayden S, Brown Z, Veldhuis J. The effect of instagram "likes" on women's social comparison and body dissatisfaction. Body Image 2018;26:90-97. [doi: 10.1016/j.bodyim.2018.07.002] [Medline: 30036748]
33. O'Brien KS, Caputi P, Minto R, Peoples G, Hooper C, Kell S, et al. Upward and downward physical appearance comparisons: development of scales and examination of predictive qualities. Body Image 2009;6(3):201-206. [doi: 10.1016/j.bodyim.2009.03.003] [Medline: 19447692]
34. Wang JL, Wang HZ, Gaskin J, Hawk S. The mediating roles of upward social comparison and self-esteem and the moderating role of social comparison orientation in the association between social networking site usage and subjective well-being. Front Psychol 2017;8:771 [FREE Full text] [doi: 10.3389/fpsyg.2017.00771] [Medline: 28553256]
35. Paxton SJ, McLean SA, Rodgers RF. "My critical filter buffers your app filter": social media literacy as a protective factor for body image. Body Image 2022;40:158-164. [doi: 10.1016/j.bodyim.2021.12.009] [Medline: 34968853]
36. Tamplin NC, McLean SA, Paxton SJ. Social media literacy protects against the negative impact of exposure to appearance ideal social media images in young adult women but not men. Body Image 2018;26:29-37. [doi: 10.1016/j.bodyim.2018.05.003] [Medline: 29807299]
37. McLean SA, Paxton SJ, Wertheim EH. Does media literacy mitigate risk for reduced body satisfaction following exposure to thin-ideal media? J Youth Adolesc 2016;45(8):1678-1695. [doi: 10.1007/s10964-016-0440-3] [Medline: 26880285]
38. Trekels J, Eggermont S. Me, myself, and my favorite media figure: an objectification perspective on the role of media and peers in early adolescents' self-sexualization. Mass Commun Soc 2020;24(1):31-56. [doi: 10.1080/15205436.2020.1817943]
39. Kvardova N, Machackova H, Smahel D. The direct and indirect effects of online social support, neuroticism, and web content internalization on the drive for Thinness among women visiting health-Oriented websites. Int J Environ Res Public Health 2020;17(7):2416 [FREE Full text] [doi: 10.3390/ijerph17072416] [Medline: 32252325]
40. Wilson S, Benning SD, Racine SE. Examining relationships among thin-ideal internalization, eating pathology, and motivational reactions to high- and low-calorie food. Appetite 2022;178:106258. [doi: 10.1016/j.appet.2022.106258] [Medline: 35921868]
41. Graff M. Does social media influence the motivation to exercise?. 2022. URL: https://www.psychologytoday.com/us/blog/love-digitally/202201/does-social-media-influence-the-motivation-exercise [accessed 2023-06-23]
42. Meier A, Schäfer S. Positive side of social comparison on social network sites: how envy can drive inspiration on instagram. Cyberpsychol Behav Soc Netw 2018;21(7):411-417. [doi: 10.1089/cyber.2017.0708] [Medline: 29995526]
43. Grabe S, Ward LM, Hyde JS. The role of the media in body image concerns among women: a meta-analysis of experimental and correlational studies. Psychol Bull 2008;134(3):460-476. [doi: 10.1037/0033-2909.134.3.460] [Medline: 18444705]

44. Barnhart WR, Sun H, Lin Z, Lu C, Han X, He J. Integrating the tripartite influence, minority stress, and social comparison theories to explain body image and disordered eating in Chinese sexual minority men and women. Body Image 2022;43:95-106. [doi: 10.1016/j.bodyim.2022.08.012] [Medline: 36095853]
45. Seekis V, Bradley GL, Duffy AL. Appearance-related social networking sites and body image in young women: testing an objectification-social comparison model. Psychol Women Q 2020;44(3):377-392. [doi: 10.1177/0361684320920826]
46. Hsu JL, Hung RTL, Antoine M. Investigating the linkages between BMI, body image, and SATAQ among young asian females. Int J Environ Res Public Health 2021;18(14):7460 [FREE Full text] [doi: 10.3390/ijerph18147460] [Medline: 34299911]
47. Weinstock T. How do I stop comparing myself to others on Instagram? | Vogue India. 2020. URL: https://www.vogue.in/wellness/content/why-cant-i-stop-comparing-myself-to-other-people-on-instagram [accessed 2023-06-23]
48. Hair JF, Ringle CM, Sarstedt M. Partial least squares: the better approach to structural equation modeling? Long Range Plann 2012;45(5-6):312-319. [doi: 10.1016/j.lrp.2012.09.011]
49. Olivia S. How TikTok's body positivity movement overshadows body shaming. 2021. URL: https://www.torchonline.com/opinion/2021/12/14/how-tiktoks-body-positivity-movement-overshadows-body-shaming/ [accessed 2023-06-23]
50. Hockey A, Milojev P, Sibley CG, Donovan CL, Barlow FK. Body image across the adult lifespan: a longitudinal investigation of developmental and cohort effects. Body Image 2021;39:114-124. [doi: 10.1016/j.bodyim.2021.06.007] [Medline: 34271529]
51. Shen J, Chen J, Tang X, Bao S. The effects of media and peers on negative body image among Chinese college students: a chained indirect influence model of appearance comparison and internalization of the thin ideal. J Eat Disord 2022;10(1):49 [FREE Full text] [doi: 10.1186/s40337-022-00575-0] [Medline: 35413877]
52. Ong WY, Sündermann O. Efficacy of the mental health app "intellect" to improve body image and self-compassion in young adults: a randomized controlled trial with a 4-week follow-up. JMIR mHealth uHealth 2022;10(11):e41800 [FREE Full text] [doi: 10.2196/41800] [Medline: 36409542]
53. Garbett KM, Haywood S, Craddock N, Gentili C, Nasution K, Saraswati LA, et al. Evaluating the efficacy of a social media-based intervention (Warna-Warni Waktu) to improve body image among young indonesian women: parallel randomized controlled trial. J Med Internet Res 2023;25:e42499 [FREE Full text] [doi: 10.2196/42499] [Medline: 37010911]

Abbreviations

AM: appearance motivation
BS: body satisfaction
CA: Cronbach α
CB-SEM: covariance-based structural equation method
Gen Z: Generation Z
SML: social media literacy
SOR: stimulus-organism-response
TII: thin-ideal internalization
TIM: tripartite influence model
UAC: upward appearance comparison
VBA: video-based social media use

Edited by K Adapa; submitted 14.03.24; peer-reviewed by M Mardiana, L Reid-Chassiakos; comments to author 29.04.24; revised version received 02.05.24; accepted 23.05.24; published 06.09.24.

Please cite as:
Ariana H, Almuhtadi I, Natania NJ, Handayani PW, Bressan S, Larasati PD
Influence of TikTok on Body Satisfaction Among Generation Z in Indonesia: Mixed Methods Approach
JMIR Hum Factors 2024;11:e58371
URL: https://humanfactors.jmir.org/2024/1/e58371
doi:10.2196/58371
PMID:

Uptake of Digital Health Interventions for Cardiometabolic Disease in British South Asian Individuals: Think Aloud Study

Mel Ramasawmy[1,2], PhD; Dan Roland Persson[3], PhD; David Sunkersing[1], PhD; Paramjit Gill[4], DM; Kamlesh Khunti[5]; Lydia Poole[6], PhD; Wasim Hanif[7], MD; Ann Blandford[8], PhD; Madiha Sajid[9], MA; Fiona Stevenson[10], PhD; Nushrat Khan[11], PhD; Amitava Banerjee[1,12], DPhil

[1]Institute of Health Informatics, University College London, 222 Euston Road, London, United Kingdom
[10]Department of Primary Care and Population Health, University College London, London, United Kingdom
[11]Department of Primary Care and Public Health, Imperial College London, London, United Kingdom
[12]Department of Cardiology, Barts Health National Health Service Trust, London, United Kingdom
[2]Department of Applied Mathematics and Computer Science, Wolfson Institute of Population Health, Queen Mary University of London, London, United Kingdom
[3]Technical University of Denmark, Kongens Lyngby, Denmark
[4]Warwick Medical School, University of Warwick, Coventry, United Kingdom
[5]Diabetes Research Centre, Leicester General Hospital, University of Leicester, Leicester, United Kingdom
[6]Department of Psychological Interventions, School of Psychology, University of Surrey, Guildford, United Kingdom
[7]Department of Diabetes, University Hospital Birmingham, Birmingham, United Kingdom
[8]University College London Interaction Centre, University College London, London, United Kingdom
[9]Patient and Public Involvement Representative, Digital Interventions for South Asians with Cardiometabolic Disease Study, London, United Kingdom

Corresponding Author:
Mel Ramasawmy, PhD
Institute of Health Informatics, University College London, , 222 Euston Road, London, , United Kingdom

Abstract

Background: Digital health interventions (DHIs) could support prevention and management of cardiometabolic disease. However, those who may benefit most often experience barriers to awareness and adoption of these interventions.

Objective: Among South Asian individuals, we evaluated user experience of DHIs for prevention and management of cardiometabolic disease, aiming to understand barriers and facilitators to initial and ongoing use.

Methods: Among South Asian individuals recruited via primary care, community organizations, and snowball methods (n=18), we conducted "think-aloud" interviews using a reflective and reactive approach. Participants included nonusers, as well as those that used a range of DHIs as part of monitoring and improving their health. Participants were asked to think aloud while completing a task they routinely do in a familiar DHI, as well as while setting up and completing a search task in a novel DHI; they were encouraged to behave as if unobserved.

Results: Lack of cultural specificity was highlighted as reducing relevance and usability, particularly relating to dietary change. Preferred features reflected individual health beliefs and behaviors, digital skills, and trust in DHIs. For example, tracking blood glucose was considered by some to be positive, while for others it caused distress and anxiety. Similarly, some users found the novel DHI to be extremely simple to set up and use, and others grew frustrated navigating through initial interfaces. Many participants raised concerns about data privacy and needing to agree to terms and conditions that they did not understand. Participants expressed that with information and support from trusted sources, they would be interested in using DHIs as part of self-management.

Conclusions: DHIs may support South Asians to prevent and manage cardiometabolic disease, but it is important to consider the needs of specific user groups in DHI development, design, and implementation. Despite motivation to make health changes, digital barriers are common. Cultural appropriateness and trusted sources (such as health care providers and community organizations) have roles in increasing awareness and enabling individuals to access and use DHIs.

(JMIR Hum Factors 2024;11:e57338) doi:10.2196/57338

KEYWORDS

digital health; cardiometabolic disease; cardiology; cardiovascular risk; health inequality; health disparity; usability; user experience; think aloud; cultural barriers; digital divide; digital literacy

Introduction

Use of digital health interventions (DHIs) for prevention and management of cardiometabolic diseases (CMD), such as diabetes, heart disease, and hypertension, has increased rapidly in the National Health Service (NHS) [1-3]. Digital health approaches, such as support with diet, activity, self-management, and remote monitoring, have shown some patient benefits, including reduction in cardiometabolic risk factors and outcomes [4,5], glycemic control [6], and reduced hospitalizations [7]. However, increasing use of DHIs may exclude some populations, particularly ethnic minorities at greater risk of health inequalities [8,9].

Although there is evidence related to improving accessibility of DHIs in African-American populations [10,11], there is limited evidence related to DHIs in South Asian populations in the United Kingdom [12,13], who face increased risk of CMD [14] and may be more likely to experience digital inequalities [15,16]. Previous research in South Asian populations in the United Kingdom considered acceptability and design issues related to SMS text messaging services [17]. Among South Asians, lack of awareness and the process of downloading and setting up DHIs have been identified as being specific barriers to ongoing engagement with digital health [18].

Approximately 25% of all apps are estimated to be uninstalled after only one use [19]. A meta-analysis of app use for chronic disease found that in real-world studies lasting between 2 weeks and 1 year, average attrition rate (negligible or ceased use) was 49% [20]. Understanding how participants react and interact with new DHIs and what features encourage them to explore further or stop use could inform recommendations for improvement. As more complex digital approaches (such as remote monitoring and virtual wards) are widely adopted, it is important to understand the needs of different population subgroups, such as South Asians. We aimed to understand the experience of DHI use by South Asian individuals with CMD, as well as how design and implementation can be improved to support uptake and use.

Methods

Ethical Considerations

Ethical approval was by the NHS London - Brent Research Ethics Committee (IRAS 261047). Informed consent was obtained from participants through a written consent form that was explained by the researcher. Participation was voluntary and participants could withdraw at any point of the study. There were no direct risks involved for participants. Data collected were deidentified and stored securely in accordance with the institution’s data management policies. Participants were reimbursed £50 (US $65.62) for their time.

Think Aloud Interviews

A mixed (reactive and reflective) “think aloud” interview approach was used [21-23]. Interviews were conducted by MR and NK, who are both experienced qualitative researchers. Recruitment was via primary care, community organizations, and snowball methods, ensuring representation across South Asian ethnicity, age, geography, generation, and gender. Participants were English-speaking, had or were at risk of diabetes and/or heart disease (ie, prediabetes was included), and had access to a suitable device (smartphone, tablet, or laptop). Participants were offered WiFi access to facilitate downloading the app, if required.

After written informed consent, interviews lasting 45 - 60 minutes were conducted by MR or NK in person, at a location convenient to participants. Participants were advised that the interview would be audiotaped, would be used for research purposes only, and would not be accessible to anyone outside the research team. It was stressed that the opinions of the participants were important, that there were no right or wrong answers, and that they could withdraw at any time. The interview guide is provided in Multimedia Appendix 1.

All participants were asked to use their device to navigate to the NHS website while explaining what they were doing as a warm-up think aloud activity. Participants who currently used a DHI were asked to navigate through it as usual, narrating what they were doing and why, for example updating and checking their step count.

All participants (including those who did not currently use a DHI) were asked to participate in the next part of the think aloud interview, involving an unfamiliar smartphone app chosen from a list provided by the research team. The 3 options were chosen from commonly used apps in a previous survey, which were in line with interview participant preferences for DHIs that were free, provided by a trusted organization (NHS), and available in iOS and Android online stores [18]. Participants were asked to choose 1 of these 3 apps ahead of the interview and to download it to their device. Support and WiFi were provided to those who needed to download the app during the interview.

Participants were asked to think aloud about the process of setting up these apps and, if they were able and willing to continue, to undertake one action on the app (eg, navigating to a specific page of the selected app). In the second half of the interview, participants were asked to describe and explain their thoughts and preferences on the use of DHIs to the interviewer, with the option of using their devices to demonstrate their views. Discussions focused on exploring use of technology as part of health, as well as barriers and facilitators to use. Participants were provided with a £50 (US $65.62) retail voucher as compensation for their time.

Recordings were professionally transcribed, and a reflexive thematic approach was taken to analysis [24]. Familiarization and initial coding was completed by MR and DRP, and merged for comparison. Through discussion between MR, DRP, NK, and DS, code definitions were revised, and transcripts further iteratively coded. These discussions drew on the team’s experiences in human-computer interactions, think aloud and qualitative methods, health inequalities, and long-term health conditions. Codes were initially grouped by MR, and themes were developed collaboratively through reflective notes and discussion between MR, DRP, DS, and NK.

Results

Participant Characteristics and Current App Use

A total of 18 participants with CMD were recruited, representing a range of gender, age, and ethnicity (Table 1). Most participants (n=11, 61%) currently used at least one DHI on their mobile phone, and a further 2 participants (11%) used their desktop, for example to access patient portals. Participants who did not use DHIs beyond receiving SMS text messages as part of appointment or other reminders were also included (n=5, 28%).

Table . Participant demographics.

	Participants, n (%)
Ethnicity	
Bangladeshi	5 (28)
Indian	7 (39)
Pakistani	6 (33)
Gender	
Female	8 (44)
Male	10 (56)
Age range (years)	
18 - 34	1 (6)
35 - 44	3 (17)
45 - 54	5 (28)
55 - 64	5 (28)
65 - 74	3 (17)
≥75	1 (6)
Education	
Secondary	4 (22)
Tertiary	14 (78)
Languages spoken	
English	18 (100)
Bengali	4 (22)
Gujarati	4 (22)
Hindi	6 (33)
Punjabi	7 (39)
Urdu	6 (33)
Other/not provided	2 (11)
Religion	
No religious beliefs/none provided	2 (11)
Hinduism	4 (22)
Islam	10 (56)
Sikhism	2 (11)
Location	
Greater London	6 (33)
Midlands East	7 (39)
West Midlands	1 (6)
Yorkshire and Humber	4 (22)
Health conditions	
Prediabetes, type 1 diabetes, or type 2 diabetes	14 (78)
Coronary heart disease, hypertension	13 (72)

Overall, participants described their DHI use as relating to information seeking; supporting exercise and well-being; accessing resources recommended by health care professionals; booking appointments or checking test results; home monitoring (eg, for atrial fibrillation); and managing diabetes (continuous glucose monitoring [CGM]). Familiar DHIs demonstrated in the first think aloud task included the following: the NHS or NHS Covid app (n=3); wearable step counter (n=3, including one without an associated app); in-built mobile app used for step tracking (n=3); gym and fitness app (n=1); CGM app (n=1); heart rate monitoring app for atrial fibrillation (n=1); and no DHIs (n=6). Apps chosen in the second think aloud task (novel DHI) included Active 10 [25] (n=11); Weight loss [26] (n=4); and Couch to 5K [27] (n=2). One participant was not able to proceed with app download but participated in the interview.

The key themes identified included facilitators and barriers to use of digital tools for health; terms and conditions, permissions, and data privacy; the changing role of DHIs in addressing health needs over time; and personalization and adaptation to meet needs. Finally, participants offered recommendations for improvement of design and implementation.

Facilitators and Barriers to Making Good Use Of Digital Tools for Health Management

All participants included in the study had a smartphone but had varying levels of engagement with digital health: "*…it's not that we are not using [smartphones]. We are still using it but maybe not as freely as others would. If we could motivate ourselves a little bit more, maybe we could all make good use of it.*" [int 14, female, 53 years old]. Several common barriers to digital uptake were described by participants who had limited or no knowledge of digital health, including lack of digital skills, fear, and previous negative experiences such as scams, viruses, or errors leading to loss of money. Participants also spoke about barriers faced by others, including language, literacy, and digital access. Together these barriers had an impact on the ability of participants to interact with health services:

> *If they're not answering the phone...they say "...make an appointment online." But what am I supposed to say? I don't know how to, you know, I haven't got the app...So everything is made difficult for us.* [int 03, female, 49 years old, DHI nonuser]

Although some described themselves as not confident digitally, they were able to use an app if they had support in setting it up, such as in the place they receive care, or from family members or friends, they would be able to continue with app use independently. However, a few highlighted that while it was easier for family members to do it, then teach them, they were afraid about what would happen when their children left home, particularly as more services became digitized:

> *There's something that as you're getting older you think, this is digital…I'm like, God what can I do? Especially when the kids leave and everything, me and my husband we're just going to be stuck…You know nobody's reached out to us. So we're just stuck where we are.* [int 03, female, 49 years old, limited use]

In the think aloud section related to setting up a novel (to participants) app, participants who considered themselves to be digitally confident found the process to be relatively easy and completed it swiftly. However, 5 of the 18 participants did not finish: 1 was unable to begin the task ("*I don't know how to download other apps*." [int 10, male, 71 years old, nonuser]) and 2 others did not have the digital skill to proceed. One participant told us: "*[I don't know] the simplest sign like swipe to the left or right or do this*" [int 03, female, 49 years old], echoing other participants' reflections on the lack of clarity or instructions about what to do when setting up or using an app for the first time.

Features that promoted engagement with the app included a simple interface, with clear presentation of available functions, and clear signposting and navigation. For example, one participant, who was using an affordable, wearable exercise tracker that did not have an app, praised the simplicity of the design and how it met her specific needs as she could "*keep pressing [a button to] look at the time and my steps and that's it*" [int 03, female, 49 years old]. The inclusion of pictures and descriptions was praised as something that made it easy for anyone to operate, including those with language barriers. However, gaps were identified, including voice recognition not being good at picking up accented English, a lack of culturally specific information, and apps not always returning relevant results. For example, in relation to an app that was able to calculate calories based on a photo, one participant described its limitations:

> *You know if it's just standard [Asian sweets] like gulab jamun…if you take a picture of that obviously you've got hundreds…of pictures of it on the internet but if it's some other Asian dessert that's not as common as that, the app might get a bit confused to what you're eating.* [int 05, male, 40 years old, DHI nonuser]

Participants who were already using DHIs reported starting using them for a number of reasons, including the following: it was recommended by friends and family, some of whom also installed the app; a need to engage with new digital modes of contact with health care; a recommendation from their health care team as part of routine management; and increased confidence from app use in other areas (such as banking). Some participants described being willing to use DHIs, but they were not aware that they existed, needed more information about their function, or wanted recommendations from trusted sources such as the NHS.

For some, while they might use digital tools in other aspects of their life, limiting their digital use for health self-management was a choice. They expressed a preference to speak to another person directly (on the phone or face-to-face); found no benefit to app use over existing actions to manage their health condition; or considered current service provision to be satisfactory and did not need to use digital to engage with health care providers: "*…I can order [my prescriptions] as well through my app but I have not done it so far…usually they already given me [paper copy]…*" [int 18, male, 77 years old, DHI user]

Design and implementation features that participants highlighted as specific barriers to initial and continued use of digital for health purposes included the affordability of DHIs, complexity of sign-ins to maintain security, and suitability to manage their health needs. Affordability was raised in particular regarding CGM, which is available through the NHS for all adults with type 1 diabetes and for some adults with type 2 diabetes (eg, if insulin-treated or other clinical need is identified) [28]. The need to set up health apps in a secure fashion, such as remembering passwords, entering a lot of information, or setting up other security features, caused some people to pause the setup process to seek help or terminate the activity:

> *But it was quite difficult actually to go through the process [of setting up the NHS app to use the COVID passport]...they ask for the information but then they weren't recognising like the face recognition...* [int 01, male, 52 years old, DHI user]

Linked to the reasons for stopping app setup, participants identified reasons that they might choose not to use the app. These include malfunctions (such as the app freezing or not syncing), needing to repeat log-in and administrative tasks, or a lack of integration with other apps—for example, participants might have to use multiple portals for appointment booking, reminders, and viewing test results. In addition, 2 participants described apps as having poor usability for individuals with complex needs (eg, cannot scroll and select all medications for reordering, or the buttons being too small on a smartphone, making entering data "tricky"), although in both these cases they opted to complete these tasks on their desktop computers rather than avoiding digital altogether.

Terms and Conditions, Permissions, and Data Privacy: Participant Concerns About Agreeing to the Unknown

In the think aloud portion of the interview, many participants agreed to the terms and conditions without reading them or after taking only a cursory look, with one participant summarizing attitudes observed across the sample: "*Yes, I'm not going to view terms and conditions, but I will agree to them because nobody reads terms and conditions*" [int 08, female, 49 years old, DHI user]. A similar approach was taken in regard to permissions associated with the app (location and motion-tracking), with some participants choosing to accept to continue rapidly onto the app.

All participants were encouraged to stop installation at any point they wished, and of the 5 participants who did not complete installation of the app in the think aloud portion, 3 who did so had concerns about permissions and privacy. Difficulty with access to the terms and conditions through an external link; the length of terms and conditions and the technical language used; and not being sure of what they should be looking for were all highlighted as specific worries:

> *But I'm just wondering what- when it's saying there's a link, why can't I find it, you know. If I put continue, it says I have to tick these two boxes. Because then you think "I might as well tick them", don't you?* [int 13, female, 64 years old, limited DHI user, did not continue with installation after this comment]

Participants did not necessarily understand the purpose and requirement for permissions. Some felt that they needed to agree to enable apps to work, for example in relation to tracking activity. Although generally attitudes toward the NHS and NHS-related DHIs were positive, 1 participant explained that for them, media coverage of issues around "track and trace" (the NHS COVID-19 contact tracing program) had undermined their trust in NHS data handling. Although the participant did not give further details, multiple instances of poor practices in relation to data handling were reported during the pandemic [29,30]. Unknowns about data sharing and privacy were raised in a number of different ways, around the purpose of sharing and the risk of potential misuse:

> *I mean my GP knows my medical history and my husband, but I don't know if I want everybody to know it, you know. And how would they actually use it, you know. If I put [health details on the app]…how would they- what would they do with it? Is it necessary to put on there? I don't know.* [int 13, female, 64 years old, limited DHI user]

A lack of clear and specific information about permissions led to concerns such as that agreeing to one set of permissions might give access to their data for another purpose: "*you don't know what type of apps come to you with the main things but actually, tracking you for other things*" [int 06, male, 59 years old, DHI on PC only]. Another participant was worried about whether data from their devices could be linked externally:

> *[…] say I have [a smartwatch] I'm sending emails and like, [from its ID] you know that it's me, right? So, then I'm sending my medical data and then it'll be known that that's my medical data. So, for somebody down the road, you know, i.e. in the database, it'll be isolated as my data. That is concerning, yes. So, the anonymity isn't there if I'm using that watch because linked to that watch is my, you know, account effectively.* [int 09, male, 55 years old, DHI user]

To mitigate these concerns, participants described being selective in the types of DHIs used (such as only for low-risk activities such as step tracking) or, in the case of 2 male participants, used multiple phones. However, they highlighted that there was a balance between concerns about privacy risks and perceived benefits, as 1 participant stated: "*But does [data sharing] really matter if you're going to get a health benefit out of it?*" [int 09, male, 55 years old, DHI user].

DHIs Play a Changing Role in Addressing Health Needs Over Time

Four main types of DHI benefits were described: accessing and keeping contact with health care providers (eg, appointment booking and reminders via SMS text messages, general practitioner platforms, or the NHS app); benefits related to COVID-19 during the pandemic (eg, information and vaccination passports); helping manage their overall health and well-being (eg, diet, exercise, and other lifestyle changes); and supporting the management of specific health conditions (eg, glucose, heart rate, or blood pressure monitoring). The discussions focused on DHIs for behavior change and specific support with health conditions, as well as the changing relevance of DHIs over time.

In relation to behavior change, participants identified that DHIs had supported their needs through features such as availability of relevant information, setting of manageable goals, tracking progress, motivation through rewards and competition with others, and prompting action through reminders and notifications.

> *I didn't know this information, it's quite helpful for me. I can walk the 10 minutes brisk walk…It's good. And it's the matter the information that people don't know about these things and that is the information gap. If the people know these little information, these very tiny things, then many people can improve their health.* [int 12, male, 55 years old, DHI user, reflecting on the NHS app Active 10]

> *Yes, it encourages me because I do like to compare with my wife…who has done more. So it's like a bit of a competition as well. So it's- I find that really good, you know. At the end of the day, evening time I'll see what I've done today. And then the next day if I've done less I'll try and do more.* [int 01, male, 52 years old, reflecting on their step tracker]

General app features, such as heart rate monitors, were also considered meaningful to those with a family history of heart disease. Specific app features that were described as helpful included a personalized home screen with key information, clear visuals around progress and rewards ("your own personal win"), and reasonable prompts. For example, while demonstrating the in-built health app on their phone, 1 participant said:

> *…what I do here I go to the option "steps" to see what I have done. So I can see today I have only done 1,127 steps…but I can go for the weekly option there…and it also actually gives me the option to see my monthly data…I do [it] regularly, normally I actually see my weekly data and see what I have done.* [int 12, male, 55 years old]

However, most participants did not know where and how to find and select the apps relevant to their needs.

Although many participants described their experiences of digital health in relation to general engagement with the health system or behavior change for overall risk reduction, a few described them as part of a program of management of CMD. There were 2 participants who spoke about the use of CGM in diabetes, 1 for themselves, and another as part of managing their relative's health. Specific benefits highlighted included real-time monitoring, ability to monitor family members remotely as part of a team with their caregivers, alerting to the need to rapidly respond, and improved understanding of how their glucose responded to their diet, which contributed to better long-term management of their diabetes.

> *I was actually able to keep my blood sugars at a constant because it become almost like, it's like you gamify it you know…Very soon you learn what makes your sugar goes up and what makes it go down. And, it's very sad to say this but the knowledge that I might go blind doesn't make as much as a difference as the fact that I've got that little graph that's telling me you're going to hit nine.* [int 08, female, 49 years old, reflecting on experience of self-funded CGM]

> *It's confidence, it's less cost…You can imagine, you know, the ailments that you get from high sugar and low sugar. It was constant chaos and that's kind of completely- almost completely taken that out. So, you're in control, and you're comfortable and relaxed about it really…* [int 09, male, 55 years old, supporting a family member with diabetes]

Additionally, changes to health over time could impact interactions with specific digital interventions. One participant described how they avoided looking at historical data as it reflected their decreased mobility: "*…it's quite upsetting knowing that a couple of years ago I did nine, ten thousand steps and I can't do it now*" [int 13, female, 64 years old, limited user]. Another highlighted how they preferred to make practical decisions based on their experience, rather than following technology that gave generic advice or was unable to adapt:

> *I think some people [become anxious and check] blood pressure like every day, two times a day…as long as I know I'm feeling alright a particular day, I don't think I need to know, you know. My legs swell up a particular day…I just take it easy that day and elevate it as much as I can, you know support it. And I think I like to do it…in a more practical way. I don't always want to be led by technology.* [int 13, female, 64 years old, limited user]

Perception of the accuracy of apps played a significant role in whether they considered them beneficial to their needs. In relation to activity trackers, a lack of explanation about how they work and how to set achievable goals meant that one participant discounted the readings as it would count when they moved their arm:

> *I don't think they're very accurate. So it's like even if I'm sat there just moving my arm around and then I'll try to get something, it just increases my tally of my steps I've done. So I know it's not true.* [int 05, male, 40 years old, limited user]

Although some participants recognized the potential benefit of DHIs for others, they felt that they had limited benefit for themselves, for example, if they were already active (and did not need a reminder or tracker) or managed their health in other ways. There were 2 participants who expressed a preference for not taking a phone to track exercise due to concerns about theft or loss, although one used a simple wearable device instead.

Anxiety around health caused by searching for health information ("*it was like my addiction*"), constant monitoring, competing with others, not meeting goals, or potential inaccuracy of readings led to a few participants choosing to use their existing DHIs less or discontinue them:

> *…One reason I'm checking is because I want to know how I can make my sugar level go down. Other reason I don't want to check is because when the sugar high, it makes me more anxious…We're relying on all these numbers and digital and all the technology, and sometimes, of course it's all the time is good, but sometimes we think, can we not just lead simple life? And sometimes I'm thinking that if machine is not working properly, then that's giving us wrong reading and making you feel even more nervous.* [int 14, female, 53 years old, DHI nonuser]

Finally, the time-consuming nature of entering exercise or diet data, or reacting to notifications, was another reason that people did not feel DHIs could meet their current needs. A busy working mother, undergoing menopause, described an overload of notifications to her phone, which acted as an interface for her duties to other people: "*too many notifications to do- too many emails, too many phone calls to call back*." However, while taking care of her health was an additional burden, she expressed an interest in trying: "*But I think maybe this…could be helpful for me at least to remind me to move. Maybe I- I will defer other notification…I need to make my health the priority, not always the other people*."

User Personalization and Adaptation of DHIs to Meet Their Needs

Participants described several ways in which they tailored their experience of DHIs to meet their needs. The first was in selection, trialing several apps before identifying the right one; for example, 1 participant with diabetes described the process of finding one for dietary management:

> *I think what has happened is over the past four, five years, I've tried to do different kinds of diet...it helps for an app to give you the information you require. So I tried [calorie counting app]...but I eventually ended up using [another] so yes, it's been a trial and error thing. So yes, one after the other.* [int 08, female, 49 years old, DHI on PC only]

Participants also limited ways in which they engaged with apps to those that directly met their needs or were easy to use. As one participant put it: "*I want it in the most simplest form and whatever I think I need that's what I've gone for*." Participants who used in-built health apps for step tracking were aware of other features but did not consider them relevant at this time.

> *...I never actually use these things...There are many things to explore but to be honest this is it...I kind of focus on how many steps or how many miles I am doing.* [int 12, male, 55 years old]

This also varied temporally, such as choosing not to wear heart rate or activity monitors overnight due to discomfort, or using DHIs only when they felt they were needed. This need included changes in symptoms, or when they felt they needed to reestablish behavior change:

> *If my symptoms start telling me something I'll check my blood sugar immediately and also if I'm feeling more thirsty, I'm going frequent to the toilet, I start noticing all these, check it on there whether I need to look and do something else.* [int 15, male, 69 years old, DHI user]

Although apps were considered useful for providing information, the actual practice of maintaining behavior change was identified as being challenging, with participants describing returning to previous behaviors after achieving change. Participants demonstrated a variety of "bargaining" strategies in how DHIs were used to manage health behaviors, such as not collecting data when the answer will be "bad" or only checking when they know it will be "good," or using more than one app to compare data and predictions and using the preferred answer.

> *If you feel like you're going off track and you're suddenly going through sort of like a binge eating period, you can kind of see that there's like a certain month period where I've not - I've just decided not to document what I'm eating because you don't want to see that I'm having... [a] whole tub of [ice-cream].*
>
> *[int 02, female, 42 years old]*

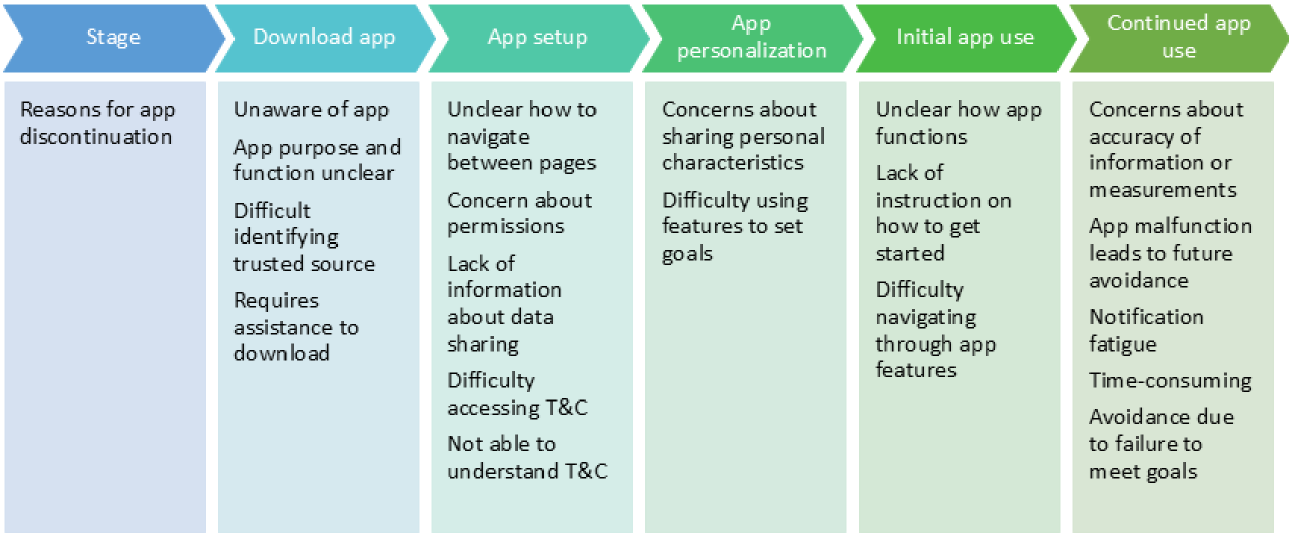
For participants who had sufficiently embedded the required behavior change to achieve stable management of their cardiometabolic condition, DHIs were identified as having a natural endpoint or a reduced role:

> *There's a certain amount of time, two or three months, that you kind of really use the [CGM] app and then after that it peters out because you're comfortable...it's almost the usefulness of the app is for that three months, right, and beyond that it's almost like- almost like an insurance policy…* [int 09, male, 55 years old]

Participant Recommendations for Improving Design and Implementation to Support DHI Uptake and Use

Participants made several recommendations aimed at improving their own experience or that of those in their wider community that they thought could benefit from DHIs with some adaptations. Figure 1 summarizes the process of setting up and using a DHI described by participants, reasons for discontinuation, and opportunities to intervene to promote uptake and use of DHIs, using apps as an example.

Figure 1. User pathways to app use for cardiometabolic disease and reasons for app discontinuation. T&C: terms and conditions.



Participants thought that the DHIs discussed had potential benefits for a wide range of people who were interested in making changes to their health. Suggestions for increasing accessibility in app features included: information provided in a range of languages, and with audio or video options; options for font size; visual aids on the page of where to click; and in-app support (such as live-chat options). They mentioned that relevance to the South Asian population could be improved through including a variety of cultural foods in diet-tracking apps, as well as including relevant celebrities, voices, or avatars in apps that included these. When introducing or implementing digital interventions, suggestions included the use of wearables or simple devices to promote exercise; installation and start-up support from health care providers; and providing safe spaces for digital upskilling and practice. Younger participants suggested utilizing social media such as WhatsApp and Facebook to engage people in issues around health as these were popular among the older generation who did have access to phones; however, there is mixed evidence supporting the use of social media for peer-to-peer information sharing [31] or as a channel for delivering diabetes education [32].

Participants also suggested engaging with the current "middle-aged" generation who could be supported to improve their health as they move into their fifties and sixties. In both app setup and ongoing use, participants wanted clear communication about the reasons for inputting personal details (ie, whether this information will be used to personalize the service and how this will be handled by third parties). Although goal setting was seen as beneficial, individuals suggested that they would benefit from instructions and guidance, such as in the form of a "how to" section or videos, as well as explaining exactly how the app works—for example, whether an activity-tracking DHI requires activation or collects information passively. The importance of keeping a range of options available was emphasized, including support for desktop versions of interfaces, and respecting the need to keep nondigital channels open for those with new concerns or for people who might choose to never use digital, including some older people.

In relation to their own use, participants wished for increased personalization and accountability, such as through digital programs that included coaching. Improved integration, particularly in accessing their health data, test results, appointments, and other NHS functions, was also seen as being beneficial. A participant who had not been offered CGM also identified that they would want something to "*do with my diabetes and I could see what my levels were. You know sometimes, because I have to do pricking every day*," suggesting that there is an appetite for CGM among populations who are not aware of it or able to ask for it. Ideal DHIs included joined-up dietary support for those with diabetes, which would suggest and track food intake, and allow for inputting exercise details. Additionally, participants suggested the idea of a "lab on a watch," which could act as a one-stop shop to monitor health.

The final areas of recommendations related to terms and conditions and data collection. Including a link to terms and conditions that navigated to a web page was confusing for some participants, and it would be helpful if it was integrated within the app. However, for a more meaningful way for a user to engage with the agreement, participants suggested that a lay summary of key messages should be included ahead of the checkboxes. Similarly, an explanation of the purpose of each permission, and how that data would be shared, would reassure people that the app was safe to use.

Discussion

Principal Findings

We present experiences of utilizing DHIs for prevention and management of CMD across a diverse group of UK individuals from a South Asian background, building on previously identified barriers and facilitators of digital acceptability, uptake, and use [18]. We highlighted willingness to use technology as part of prevention and management of CMD, but one single approach did not suit all, including the choice not to use DHIs among individuals who might otherwise use internet-based communication and entertainment, as described in other populations [33,34]. A review of nonuse of telemedicine also highlighted "other preferences" as a key aspect of attitudes toward telemedicine technologies, describing preferences for conventional solutions or other technical solutions [35].

Across all levels of digital skill, a key concern in downloading and setting up apps focused on data privacy and needing to agree to difficult-to-understand terms and conditions to use DHIs, including those for essential services. Recent studies in other contexts have also put forward ways to improve user engagement with terms and conditions, including not using hyperlinks and ensuring that they are transparently displayed early in user interaction [36]. Others have included terms of use as a factor to consider in a digital health evaluation tool for patients and clinicians [37]. However, many participants would not have been able to utilize this tool as it assumes a level of digital knowledge. There is a pressing need for improved communication and education for meaningful consent. Moreover, some concern about data safety within DHIs appears justified in light of evidence around vulnerabilities in medical devices and commercially available wearables [38-40], highlighting the need to provide assurance to the public.

The types of DHIs used by participants ranged from step counters to those with more complex features such as CGM. Features that were considered by one user to be positive (such as monitoring and tracking measurements) were reported by others to cause distress and anxiety. Regular DHI users found the NHS apps used in the "think aloud" portion of the interview to be extremely simple to set up and use, while participants who described themselves as less confident with technology grew frustrated navigating through the initial information screens. International recommendations on diabetes highlight the importance of designing apps with the level of technology proficiency of different patient populations in mind and increasing accessibility by using languages other than English, as well as increasing accessibility for people with visual impairment [41]. However, even when individuals did not face these barriers, a lack of cultural specificity in provided information reduced relevance and usability, particularly in relation to CMD, where dietary change may form a significant part of patient self-management [18,42].

Participants described dynamic changes in their interaction with and expectations of DHIs, reflecting variation in their health needs and exploration of approaches, in the short- and longer-term, both in relation to lifestyle change and management of specific CMDs. This supports previous findings in other populations such as those with diabetes, where after addressing initial acute needs after diagnosis, individuals may decrease use until a new event prompts interaction [43,44]; a similar pattern is demonstrated in people who may have a long-standing diagnosis but are newly introduced to DHIs. Digital approaches may also be seen as unsupportive or not relevant for newly diagnosed cardiometabolic conditions [18,45], which should be considered in South Asian populations as minority ethnic groups have been associated with an increased prevalence of diabetes distress [46].

Comparison With Prior Work

Our findings in this study reflect in a health setting the framework for temporality of use of a new digital tool by Karapanos and colleagues: anticipation (participant expectations prior to use); orientation (excitement and frustration in learning and exploration); incorporation (meaningful use of product in daily lives); and identification (form a relationship with a product as it becomes part of their routine and social interactions) [47].

Positive impacts of DHIs for diabetes on individuals have been suggested to include helping them understand and feel in control of their condition; construct positive identities through being experts in their disease; increase their sense of power in clinician interactions; and demonstrate their goodness [44,48]. Although our findings reflect common positive outcomes in some users, we found that without adequate consideration of individual need, DHIs may instead increase anxiety or disempower patients. This may result in users feeling that they are being "led by the technology" rather than the technology supporting their needs, and this raises questions about the suitability and acceptability of automation in health interventions [49].

Studies in older populations in the United States have highlighted that people who are actively engaged in health management are more likely to use wearables and are willing to share health data with providers [50]. DHIs can be beneficial to older individuals when adapted for use and supported by communication with medical professionals [51,52]. Participants suggested a range of ways in which the design and implementation of digital technologies in health could support a diverse population, including promoting localization to different languages and cultures, including audio options; clear design and other accessibility measures; and provision of information and instruction within the app and through health care teams. Similar recommendations have been made for DHIs relating to CMD [53], in South Asian communities in the United Kingdom and India in relation to lifestyle change [42,54], for other health conditions [55-57], and for patient populations with different needs [58]. Embedding these approaches can have benefit beyond any single patient population [59]. However, for participants (of any age) who are digitally confident, more complex and integrated interventions may be suitable. This highlights the importance of having a range of interventions that are targeted to the needs and expectations of populations who may benefit from them (eg, other ethnic or minority groups) [60].

Limitations

There is significant diversity within the South Asian population in the United Kingdom, by migration generation, country of origin, ethnicity, religion, education, occupation, and income, as well as by age and gender; with a relatively small sample size (n=18), this study cannot represent all experiences of digital access. In addition, due to the think aloud method, only participants with mobile phones or other devices able to support apps, and who were able to speak English, were recruited. In 2011 census data, most people of South Asian ethnicity in the United Kingdom were able to speak English, with nonspeakers tending to be older, female, and from Bangladeshi and Pakistani backgrounds [61]. This emphasizes the need for appropriate support to ensure these groups are able to access and benefit from health services for CMD. However, there is a significant difference between age groups; although Bangladeshi women over the age of 65 years reported the highest rate of not speaking English (44.9%), this was only 2.8% in the 25 - 44 year age group [61]. This suggests that there is a large demographic of English-speaking South Asians in the United Kingdom that could potentially benefit from digital health approaches, given the right support.

There is currently limited evidence on the experiences of digital health among South Asian populations in the United Kingdom, particularly in relation to CMD [12,13]. This study makes an important contribution toward understanding opportunities for culturally relevant DHI design and implementation. In addition, broader findings and recommendations have the potential to benefit other populations that may currently be digitally excluded. Future research should explore implementation approaches to support those new to digital to install and use apps, as well as explore how to improve communication around permissions, privacy, and terms and conditions to ensure patients can meaningfully consent to the use of DHIs.

Conclusion

In our study, we demonstrated that individuals from a South Asian background in the United Kingdom are interested in DHIs as part of prevention and management of CMD, and in addressing short- and long-term needs, including engaging with family members and carers for the benefit of those who may be otherwise digitally excluded. Initial access to technology (by which we mean awareness, downloading, and setting up a device or app) is a significant barrier. This emphasizes the importance of support from appropriate trusted sources, such as health care teams, in initiating DHI use. Suggestions to improve relevance and utility of DHIs for the South Asian population in the United Kingdom focused on inclusion of cultural foods in advice around dietary management of CMD. Participants made a number of recommendations around DHI design and implementation that would improve accessibility across user groups. This highlights that design approaches for DHIs for prevention or management of cardiometabolic disease should take into account the diverse needs of many populations for universal benefit.

Acknowledgments

The authors were funded by the National Institute of Health Research (NIHR; NIHR200937). The funding source played no role in the design of the study; collection, analysis, and interpretation of data; writing of the report; and decision to submit the paper for publication. PG was supported by the NIHR Applied Research Collaborations West Midland and is an NIHR Senior Investigator. The views expressed in this publication are those of the authors and not necessarily those of the NIHR or the UK Department of Health and Social Care. KK is supported by the NIHR Applied Research Collaboration East Midlands and NIHR Leicester Biomedical Research Centre. LP is supported by the NIHR (NIHR155654). This review was conducted on behalf of the Digital Interventions for South Asians with Cardiometabolic Disease Study consortium [62].

Authors' Contributions

The study concept was designed by A Banerjee, MR, and LP. Interviews were carried out by MR and NK. Analysis was conducted by MR and DRP with support from NK and DS. MR wrote the original draft with support from DRP, and review and edits by A Banerjee, NK, and DS. The figure was designed by MR. Additional review was carried out by PG, KK, LP, WH, A Blandford, MS, and FS for the Digital Interventions for South Asians with Cardiometabolic Disease Study consortium.

Conflicts of Interest

KK is the director of the University of Leicester Centre for Ethnic Health Research. AB, WH, and KK are trustees of the South Asian Health Foundation. WH is also a trustee of Diabetes UK.

Multimedia Appendix 1
Interview guide.
[DOCX File, 19 KB - humanfactors_v11i1e57338_app1.docx]

References

1. Ross JAD, Barron E, McGough B, et al. Uptake and impact of the English National Health Service digital diabetes prevention programme: observational study. BMJ Open Diabetes Res Care 2022 May;10(3):e002736. [doi: 10.1136/bmjdrc-2021-002736] [Medline: 35504697]
2. New digital health check to tackle deadly cardiovascular disease. GOVUK. URL: https://www.gov.uk/government/news/new-digital-health-check-to-tackle-deadly-cardiovascular-disease [accessed 2023-10-29]
3. NHS england » NHS virtual wards to treat thousands of patients with heart failure at home. NHS England. URL: https://www.england.nhs.uk/2023/10/nhs-virtual-wards-to-treat-thousands-of-patients-with-heart-failure-at-home/ [accessed 2023-10-29]
4. Widmer RJ, Collins NM, Collins CS, West CP, Lerman LO, Lerman A. Digital health interventions for the prevention of cardiovascular disease: a systematic review and meta-analysis. Mayo Clin Proc 2015 Apr;90(4):469-480. [doi: 10.1016/j.mayocp.2014.12.026] [Medline: 25841251]

5. Azizi Z, Adedinsewo D, Rodriguez F, Lewey J, Merchant RM, Brewer LC. Leveraging digital health to improve the cardiovascular health of women. Curr Cardiovasc Risk Rep 2023;17(11):205-214. [doi: 10.1007/s12170-023-00728-z] [Medline: 37868625]
6. Zimmermann G, Venkatesan A, Rawlings K, Scahill MD. Improved glycemic control with a digital health intervention in adults with type 2 diabetes: retrospective study. JMIR Diabetes 2021 Jun 2;6(2):e28033. [doi: 10.2196/28033] [Medline: 34075880]
7. Widmer RJ, Allison TG, Lerman LO, et al. Digital health intervention as an adjunct to cardiac rehabilitation reduces cardiovascular risk factors and rehospitalizations. J Cardiovasc Transl Res 2015;8:283-292. [doi: 10.1007/S12265-015-9629-1/FIGURES/5]
8. Richardson S, Lawrence K, Schoenthaler AM, Mann D. A framework for digital health equity. NPJ Digit Med 2022 Aug 18;5(1):119. [doi: 10.1038/s41746-022-00663-0] [Medline: 35982146]
9. GOV.UK. NHS COVID-19 app: early adopter evaluation report NHS test and trace programme. : Department of Health and Social Care; 2021 URL: https://www.gov.uk/government/publications/nhs-covid-19-app-early-adopter-phase-reports [accessed 2024-10-08]
10. Barber-Gumbs T, Trolle Lagerros Y, Sena LM, et al. Perspectives from underserved African Americans and their health care providers on the development of a diabetes self-management smartphone app: qualitative exploratory study. JMIR Form Res 2021 Feb 26;5(2):e18224. [doi: 10.2196/18224] [Medline: 33635279]
11. Schoenthaler A, Leon M, Butler M, Steinhaeuser K, Wardzinski W. Development and evaluation of a tailored mobile health intervention to improve medication adherence in Black patients with uncontrolled hypertension and type 2 diabetes: pilot randomized feasibility trial. JMIR mHealth uHealth 2020 Sep 23;8(9):e17135. [doi: 10.2196/17135] [Medline: 32965230]
12. Goswami A, Poole L, Thorlu-Bangura Z, et al. The use of digital health interventions for cardiometabolic diseases among South Asian and Black minority ethnic groups: realist review. J Med Internet Res 2023 Jan 6;25:e40630. [doi: 10.2196/40630] [Medline: 36607732]
13. Aldosari N, Ahmed S, McDermott J, Stanmore E. The use of digital health by South Asian communities: scoping review. J Med Internet Res 2023 Jun 12;25:e40425. [doi: 10.2196/40425] [Medline: 37307045]
14. Coles B, Zaccardi F, Ling S, Davies MJ, Samani NJ, Khunti K. Cardiovascular events and mortality in people with and without type 2 diabetes: an observational study in a contemporary multi-ethnic population. J Diabetes Investig 2021 Jul;12(7):1175-1182. [doi: 10.1111/jdi.13464] [Medline: 33206469]
15. Exploring the UK's digital divide - Office for National Statistics. ONS. 2019. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/articles/exploringtheuksdigitaldivide/2019-03-04 [accessed 2021-05-25]
16. Banerjee A. Digital health interventions and inequalities: the case for a new paradigm. BMJ EBM 2021 Jun;26(3):77-78. [doi: 10.1136/bmjebm-2019-111282]
17. Prinjha S, Ricci-Cabello I, Newhouse N, Farmer A. British South Asian patients' perspectives on the relevance and acceptability of mobile health text messaging to support medication adherence for type 2 diabetes: qualitative study. JMIR mHealth uHealth 2020 Apr 20;8(4):e15789. [doi: 10.2196/15789] [Medline: 32310150]
18. Ramasawmy M, Sunkersing D, Persson DR, et al. “If it all goes digital, we'll have to learn”: facilitators and barriers to uptake of digital health in British South Asians with cardiometabolic disease. BMC Digit Health 2024;2(1):32. [doi: 10.1186/s44247-024-00083-y]
19. Nearly 1 in 4 people abandon mobile apps after only one use. TechCrunch. URL: https://techcrunch.com/2016/05/31/nearly-1-in-4-people-abandon-mobile-apps-after-only-one-use/ [accessed 2024-10-03]
20. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020 Sep 29;22(9):e20283. [doi: 10.2196/20283] [Medline: 32990635]
21. Fischer T, Reinhardt G, Stumpf P, Schwarz PEH, Harst L. Think-aloud usability testing of an app to increase physical activity. Res Sq 2019. [doi: 10.21203/rs.2.17034/v2]
22. Alkhaldi G, Modrow K, Hamilton F, Pal K, Ross J, Murray E. Promoting engagement with a digital health intervention (HeLP-Diabetes) using email and text message prompts: mixed-methods study. Interact J Med Res 2017 Aug 22;6(2):e14. [doi: 10.2196/ijmr.6952] [Medline: 28829328]
23. Perski O, Blandford A, Ubhi HK, et al. Smokers' and drinkers' choice of smartphone applications and expectations of engagement: a think aloud and interview study. BMC Med Inform Decis Mak 2017;17:25. [doi: 10.1186/S12911-017-0422-8/TABLES/3]
24. Braun V, Clarke V. One size fits all? What counts as quality practice in (reflexive) thematic analysis? Qual Res Psychol 2021 Jul 3;18(3):328-352. [doi: 10.1080/14780887.2020.1769238]
25. Active 10 Walking Tracker. NHS. URL: https://www.nhs.uk/better-health/get-active/ [accessed 2024-10-03]
26. Weight Loss Plan. NHS. URL: https://www.nhs.uk/better-health/lose-weight/ [accessed 2024-10-03]
27. Couch to 5K. NHS. URL: https://www.nhs.uk/better-health/get-active/ [accessed 2024-10-03]
28. Glucose monitoring for patients living with diabetes. NHS England. URL: https://www.england.nhs.uk/diabetes/digital-innovations-to-support-diabetes-outcomes/flash-glucose-monitoring/ [accessed 2023-12-21]

29. Cellan-Jones R. Coronavirus: England's test and trace programme 'breaks GDPR data law'. BBC. URL: https://www.bbc.co.uk/news/technology-53466471 [accessed 2024-10-03]
30. Manthorpe R. COVID-19: Test and Trace barely used check-in data from pubs and restaurants - with thousands not warned of infection risk. Sky News. URL: https://news.sky.com/story/test-and-trace-barely-used-check-in-data-from-pubs-and-restaurants-with-thousands-not-warned-of-infection-risk-12235392 [accessed 2023-12-19]
31. Elnaggar A, Ta Park V, Lee SJ, Bender M, Siegmund LA, Park LG. Patients' use of social media for diabetes self-care: systematic review. J Med Internet Res 2020 Apr 24;22(4):e14209. [doi: 10.2196/14209] [Medline: 32329745]
32. Gabarron E, Årsand E, Wynn R. Social media use in interventions for diabetes: rapid evidence-based review. J Med Internet Res 2018 Aug 10;20(8):e10303. [doi: 10.2196/10303] [Medline: 30097421]
33. Safarov N. Personal experiences of digital public services access and use: older migrants' digital choices. Technol Soc 2021 Aug;66:101627. [doi: 10.1016/j.techsoc.2021.101627]
34. Gordon NP, Crouch E. Digital information technology use and patient preferences for internet-based health education modalities: cross-sectional survey study of middle-aged and older adults with chronic health conditions. JMIR Aging 2019 Apr 4;2(1):e12243. [doi: 10.2196/12243] [Medline: 31518291]
35. Reinhardt G, Schwarz PE, Harst L. Non-use of telemedicine: a scoping review. Health Informatics J 2021;27(4):14604582211043147. [doi: 10.1177/14604582211043147] [Medline: 34696613]
36. Kitkowska A, Högberg J, Wästlund E. Online terms and conditions: improving user engagement, awareness, and satisfaction through UI design. 2022 Apr 29 Presented at: 2022 CHI Conference on Human Factors in Computing Systems; Apr 29 to May 5, 2022; New Orleans, LA, USA p. 1-22 URL: https://dl.acm.org/doi/proceedings/10.1145/3491102 [doi: 10.1145/3491102.3517720]
37. Henson P, David G, Albright K, Torous J. Deriving a practical framework for the evaluation of health apps. Lancet Dig Health 2019 Jun;1(2):e52-e54. [doi: 10.1016/S2589-7500(19)30013-5]
38. Silva-Trujillo AG, González González MJ, Rocha Pérez LP, García Villalba LJ. Cybersecurity analysis of wearable devices: smartwatches passive attack. Sensors (Basel) 2023 Jun 8;23(12):5438. [doi: 10.3390/s23125438] [Medline: 37420605]
39. Ioannidou I, Sklavos N. On general data protection regulation vulnerabilities and privacy issues, for wearable devices and fitness tracking applications. Cryptogr 2021;5(4):29. [doi: 10.3390/cryptography5040029]
40. Bracciale L, Loreti P, Bianchi G. Cybersecurity vulnerability analysis of medical devices purchased by national health services. Sci Rep 2023 Nov 9;13(1):19509. [doi: 10.1038/s41598-023-45927-1] [Medline: 37945583]
41. Fleming GA, Petrie JR, Bergenstal RM, Holl RW, Peters AL, Heinemann L. Diabetes digital app technology: benefits, challenges, and recommendations. A consensus report by the European Association for the Study of Diabetes (EASD) and the American Diabetes Association (ADA) Diabetes Technology Working Group. Diabetologia 2020 Feb;63(2):229-241. [doi: 10.1007/s00125-019-05034-1] [Medline: 31802144]
42. Ranjani H, Nitika S, Hariharan R, et al. Systematic review and scientific rating of commercial apps available in India for diabetes prevention. J Diabetol 2021;12(3):285-292. [doi: 10.4103/JOD.JOD_68_21]
43. Klasnja P, Kendall L, Pratt W, Blondon K. Long-term engagement with health-management technology: a dynamic process in diabetes. AMIA Annu Symp Proc 2015;2015:756-765. [Medline: 26958211]
44. Persson DR, Zhukouskaya K, Wegener AMK, Jørgensen LK, Bardram JE, Bækgaard P. Exploring patient needs and designing concepts for digitally supported health solutions in managing type 2 diabetes: cocreation study. JMIR Form Res 2023 Aug 25;7:e49738. [doi: 10.2196/49738] [Medline: 37624633]
45. Poduval S, Marston L, Hamilton F, Stevenson F, Murray E. Feasibility, acceptability, and impact of a web-based structured education program for type 2 diabetes: real-world study. JMIR Diabetes 2020 Jan 6;5(1):e15744. [doi: 10.2196/15744] [Medline: 31904580]
46. Skinner TC, Joensen L, Parkin T. Twenty-five years of diabetes distress research. Diabet Med 2020 Mar;37(3):393-400. [doi: 10.1111/dme.14157] [Medline: 31638279]
47. Karapanos E, Zimmerman J, Forlizzi J, et al. User experience over time: an initial framework. Presented at: CHI '09: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; Apr 4-9, 2009; Boston, MA, USA p. 729-738. [doi: 10.1145/1518701.1518814]
48. Turnbull S, Lucas PJ, Hay AD, Cabral C. Digital health interventions for people with type 2 diabetes to develop self-care expertise, adapt to identity changes, and influence other's perception: qualitative study. J Med Internet Res 2020 Dec 21;22(12):e21328. [doi: 10.2196/21328] [Medline: 33346733]
49. Parasuraman R, Sheridan TB, Wickens CD. A model for types and levels of human interaction with automation. IEEE Trans Syst Man Cybern A Syst Hum 2000 May;30(3):286-297. [doi: 10.1109/3468.844354] [Medline: 11760769]
50. Chandrasekaran R, Katthula V, Moustakas E. Too old for technology? Use of wearable healthcare devices by older adults and their willingness to share health data with providers. Health Informatics J 2021;27(4):14604582211058073. [doi: 10.1177/14604582211058073] [Medline: 34802315]
51. Hägglund E, Strömberg A, Hagerman I, Lyngå P. Theory testing of patient perspectives using a mobile health technology system in heart failure self-care. J Cardiovasc Nurs 2019;34(6):448-453. [doi: 10.1097/JCN.0000000000000595] [Medline: 31365443]

52. Sun C, Sun L, Xi S, et al. Mobile phone-based telemedicine practice in older Chinese patients with type 2 diabetes mellitus: randomized controlled trial. JMIR mHealth uHealth 2019 Jan 4;7(1):e10664. [doi: 10.2196/10664] [Medline: 30609983]
53. Paul T, Mehawej J, Philis-Tsimikas A. Digital health tools to promote diabetes education and management of cardiovascular risk factors among under-resourced populations. Cardiovasc Digit Health J 2021 Dec;2(6):298-300. [doi: 10.1016/j.cvdhj.2021.09.005] [Medline: 35265925]
54. Taak K, Brown J, Perski O. Exploring views on alcohol consumption and digital support for alcohol reduction in UK-based Punjabi-Sikh men: a think aloud and interview study. Drug Alcohol Rev 2020 Sep 16;40(2):231-238. [doi: 10.1111/dar.13172] [Medline: 32964571]
55. Hughson JAP, Daly JO, Woodward-Kron R, Hajek J, Story D. The rise of pregnancy apps and the implications for culturally and linguistically diverse women: narrative review. JMIR mHealth uHealth 2018 Nov 16;6(11):e189. [doi: 10.2196/mhealth.9119] [Medline: 30446483]
56. Cheng VWS, Piper SE, Ottavio A, Davenport TA, Hickie IB. Recommendations for designing health information technologies for mental health drawn from self-determination theory and co-design with culturally diverse populations: template analysis. J Med Internet Res 2021 Feb 10;23(2):e23502. [doi: 10.2196/23502] [Medline: 33565985]
57. Graham-Brown MPM, Smith AC, Greenwood SA. Digital health interventions in chronic kidney disease: levelling the playing field? Clin Kidney J 2023 May;16(5):763-767. [doi: 10.1093/ckj/sfac259] [Medline: 37151422]
58. Radcliffe E, Lippincott B, Anderson R, Jones M. A pilot evaluation of mHealth app accessibility for three top-rated weight management apps by people with disabilities. Int J Environ Res Public Health 2021 Apr 1;18(7):3669. [doi: 10.3390/ijerph18073669] [Medline: 33915917]
59. Harrington CN, Ruzic L, Sanford JA. Universally accessible mHealth apps for older adults: towards increasing adoption and sustained engagement. Presented at: HCI International 2017; Jul 9-14, 2017; Vancouver, Canada p. 3-12. [doi: 10.1007/978-3-319-58700-4_1]
60. Boels AM, Vos RC, Dijkhorst-Oei LT, Rutten G. Effectiveness of diabetes self-management education and support via a smartphone application in insulin-treated patients with type 2 diabetes: results of a randomized controlled trial (TRIGGER study). BMJ Open Diab Res Care 2019 Dec;7(1):e000981. [doi: 10.1136/bmjdrc-2019-000981]
61. English language skills - GOV.UK Ethnicity facts and figures. GOV.UK. 2020. URL: https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/demographics/english-language-skills/latest/#full-page-history [accessed 2023-12-21]
62. The DISC study. Doctor as Data Scientist. URL: https://www.doctorasdatascientist.org/the-disc-study [accessed 2024-10-03]

Abbreviations

CGM: continuous glucose monitoring
CMD: cardiometabolic disease
DHI: digital health intervention
NHS: National Health Service

Edited by E Borycki, Y Senathirajah; submitted 13.02.24; peer-reviewed by E Fioratou, S Madonsela; revised version received 13.08.24; accepted 13.09.24; published 24.10.24.

Please cite as:
Ramasawmy M, Roland Persson D, Sunkersing D, Gill P, Khunti K, Poole L, Hanif W, Blandford A, Sajid M, Stevenson F, Khan N, Banerjee A
Uptake of Digital Health Interventions for Cardiometabolic Disease in British South Asian Individuals: Think Aloud Study
JMIR Hum Factors 2024;11:e57338
URL: https://humanfactors.jmir.org/2024/1/e57338
doi:10.2196/57338

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org